VOL. 4 NO. 1 DECEMBER 2018

JOURNAL OF NETWORK OPERATIONS

SCTE • ISBE STANDARDS

NETWORK OPERATIONS

scte.org • isbe.org

NETWORK OPERATIONS

SCTE · ISBE

Society of Cable Telecommunications Engineers
International Society of Broadband Experts

JOURNAL OF NETWORK OPERATIONS

VOLUME 4, NUMBER 1
December 2018

Society of Cable Telecommunications Engineers, Inc.
International Society of Broadband Experts™
140 Philips Road, Exton, PA 19341-1318

© 2018 Society of Cable Telecommunications Engineers, Inc. All rights reserved.

SCTE · ISBE
STANDARDS

Table of Contents

Editorial Correspondence: If there are errors or omissions to the information provided in this journal, corrections may be sent to our editorial department. Address to: SCTE Journals, SCTE•ISBE, 140 Philips Road, Exton, PA 19341-1318 or email journals@scte.org.

Submissions: If you have ideas or topics for future journal articles, please let us know. Topics must be relevant to our membership and fall under the technologies covered by each respective journal. All submissions will be peer reviewed and published at the discretion of SCTE•ISBE. Electronic submissions are preferred, and should be submitted to SCTE Journals, SCTE•ISBE, 140 Philips Road, Exton, PA 19341-1318 or email journals@scte.org.

Subscriptions: Access to technical journals is a benefit of SCTE•ISBE Membership. Nonmembers can join at www.scte.org/join.

From the Editors

Welcome to Volume 4 Issue 1 of the *Journal of Network Operations*, a publication of collected papers by the Society of Cable Telecommunications Engineers (SCTE) and its global arm, the International Society of Broadband Experts (ISBE). A variety of plant-related topics are included in this month's issue.

Cable operators have been deploying DOCSIS 3.1 service over the last year or so (and are looking at full duplex DOCSIS for the future). A key challenge is finding available spectrum in which to transmit a wide bandwidth orthogonal frequency division multiplexing (OFDM) signal. While D3.1 technology supports downstream OFDM channel bandwidths ranging from 24 MHz to 192 MHz, that much open bandwidth is not always available in many networks. What to do? John Ulm provides some useful guidelines in his paper, "Finding Room for D3.1 & FDX." And to get the plant itself ready for FDX, you may need to upgrade your taps and other plant passives to get the full capacity promised by the FDX specification. Mark Knowles provides you with both the challenges and new micro-coupler-based technology to address the potential interference issues that may arise when FDX is deployed at scale.

One plant impairment that has caused confusion in recent years is common path distortion, or CPD. In the days of mostly- or all-analog TV channel networks, upstream CPD was easy to identify on a spectrum analyzer display given its characteristic beat clusters spaced every 6 MHz (8 MHz in European networks). Once the transition was made to mostly- or all-digital operation, some have mistakenly assumed that CPD went away. It didn't – it just took on a different appearance. Tom Williams discussed this and more in his paper, "Common Path Distortion in Digital Networks." Another paper by Tom takes a look at three novel methods – two in-service and one out-of-service – to perform time domain reflectometer (TDR)-like characterization of coaxial cable reflections.

Energy consumption and heat dissipation in the outside plant are also key issues as newer, more power hungry actives are deployed with distributed access architectures and the new generic access platform (GAP) under development in the SCTE standards program. Heat dissipation and passive cooling in plant actives is addressed with a simple mathematical model by Bruce Jackson in this issue to guide the process of making early thermal design tradeoffs. And while initial adaptive power schemes for plant actives consider temporary elimination of higher RF carriers to save energy when traffic permits, Lamar West has considered the impact of nonlinear distortion products and offers some reasons to consider nulling lower frequencies instead of higher frequencies to save energy while maintaining plant performance overall. For the latter contribution, please see the next issue of the SCTE·ISBE *Journal of Energy Management*.

As always, when new technologies are developed and deployed, and new goals such as energy efficiency arise, we continue to revise our standard practices for plant maintenance and benchmarking. Alexander Adams takes us through an updated process for optimization of practices for HFC plant operation and maintenance, and Ronald Hasenberger shows us how we can better prepare for future network evolution via operations benchmarking.

We are grateful for the individuals who contributed to this issue of the *Journal of Network Operations*, including the authors, reviewers, and the SCTE·ISBE publications and marketing staff. We hope you enjoy this issue of the *Journal*, and that the selected papers provide inspiration for new ideas and innovations in cable network operation. If you have feedback on this issue, have a new idea, or would like to share a success story please let us know at journals@scte.org.

SCTE·ISBE *Journal of Network Operations* Senior Editors,

Ron Hranac
Technical Marketing Engineer – Cable Access Business Unit, Cisco Systems, Inc.
SCTE Fellow

Daniel Howard
Principal, Enunciant LLC.
SCTE Senior Member

Making Room for D3.1 & FDX

Leveraging Something Old that is New Again!

A Technical Paper prepared for SCTE•ISBE by

John Ulm, SCTE Member
Engineering Fellow, Broadband Systems
CTO-Network Solutions team
ARRIS

Table of Contents

List of Figures

List of Tables

1. Introduction

Broadband data usage has climbed for decades. After a couple years of steady growth, there were significant increases last year. We review our cable industry data findings and highlight how upstream growth is now on par with downstream. The rapid growth of gigabit services over D3.1 is quickly increasing the need for additional OFDM spectrum. Massive upstream growth and symmetric services will require migration to full duplex (FDX) DOCSIS sooner than later. Our traffic engineering results highlight the latest data utilization and when operators will need the spectrum.

Adding multiple 192 MHz D3.1/FDX channels puts tremendous strain on the spectrum budget. Splitting nodes hits a diminishing point of return and doesn't solve the issue. Upgrading your HFC to fiber deep 1.2 GHz plant is costly AND time-consuming. It turns out a decade old technology, switched digital video (SDV), is coming back to rescue the day!

Original SDV deployments a decade ago were limited by existing edge-QAM (EQAM) modulator costs and densities. This meant SDV service groups tended to be very large (e.g., thousands of tuners) with a handful of EQAM modulator-sourced channels. This was good enough then to free up ~96 MHz of precious spectrum for HD video or D3.0 channels. Today, the integrated-CCAP has fundamentally changed SDV dynamics. It allows SDV SG size to align with much smaller DOCSIS SG sizes and for significantly more EQAM modulator-based channels. Both factors combine to amplify the SDV spectrum savings.

Our modeling results show how operators can use SDV to free 384+ MHz of spectrum for OFDM channels. And this is done on an I-CCAP box with no outside plant changes and no re-wiring in the headend!

So, while SDV technology has been relatively dormant over the last five to ten years, the I-CCAP deployments allow operators to leverage re-born SDV to address the rapid growth of D3.1/FDX in addition to MPEG-2 to MPEG-4 transition and IP video migration strategies.

Broadband Bandwidth Trends

The Internet has been growing at a breakneck speed since its inception. And with it, we have seen a corresponding growth in dedicated network capacity. [ULM_2017] and other earlier papers have provided an overview of these trends which are highlighted and updated below.

1. Nielsen's Law and Cloonan's Curves

While Moore's Law is infamous in silicon realms, Nielsen's Law of Internet Bandwidth has become renowned in the networking world. It basically states that network connection speeds for high-end home users would increase 50% per year. This law has driven much of the traffic engineering and network capacity planning in the service provider world. It has also led to much research on those topics.

In [CLOONAN_2014, EMM_2014], this research was expanded to also include traffic utilization in addition to the network connection speed. Nielsen's Law is shown in Figure 1. Since the Y-axis is a log scale, the 50% compounded annual growth rate (CAGR) appears as a straight line. An interesting fact is that the graph starts in 1982 with a 300-baud phone modem. The industry is now in the fourth decade of closely following this trend.

While Nielsen's Law focuses primarily on downstream speeds, upstream speeds have generally followed the same growth rate, but at about one-tenth the speed. However, with more fiber to the premise (FTTP) deployments and the upcoming introduction of FDX DOCSIS, it is expected that the highest offered upstream speeds will take a step up as symmetric services become available.

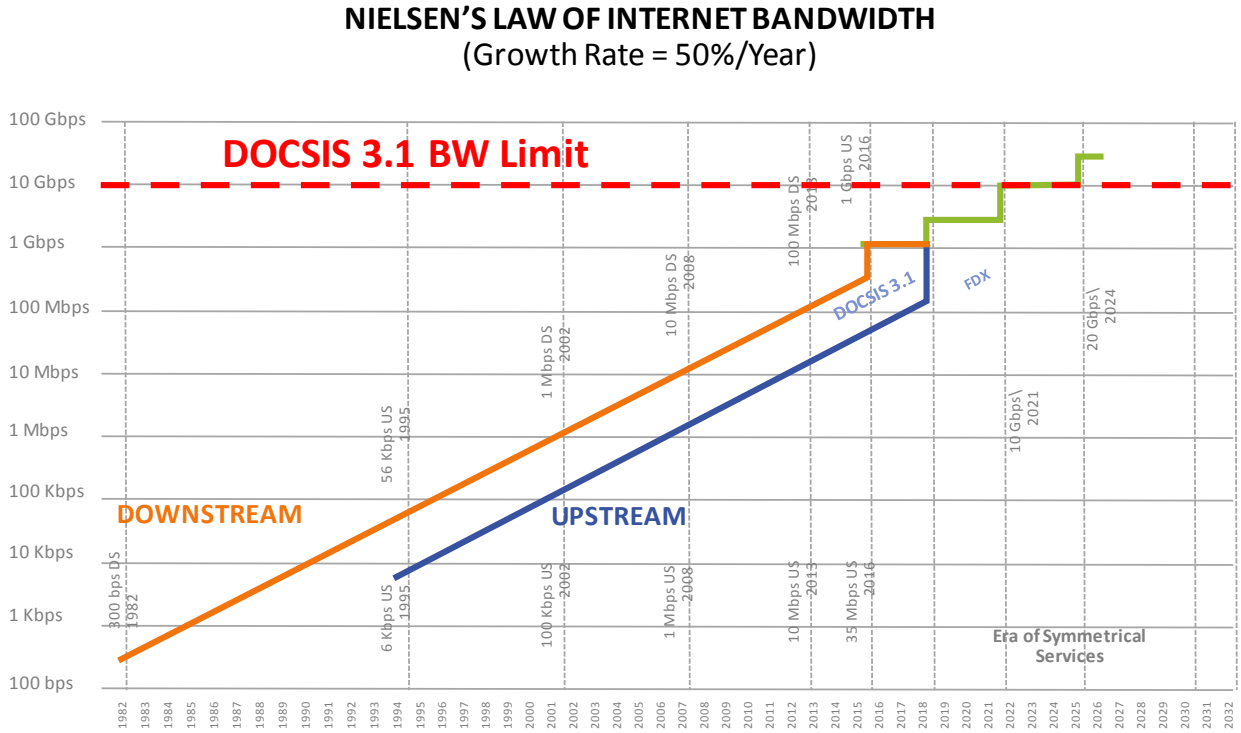

Figure 1 – Nielsen's Law – 50% CAGR

NIELSEN'S "SLOWED" LAW OF INTERNET BANDWIDTH
(Growth Rate = 30-40%/Year)

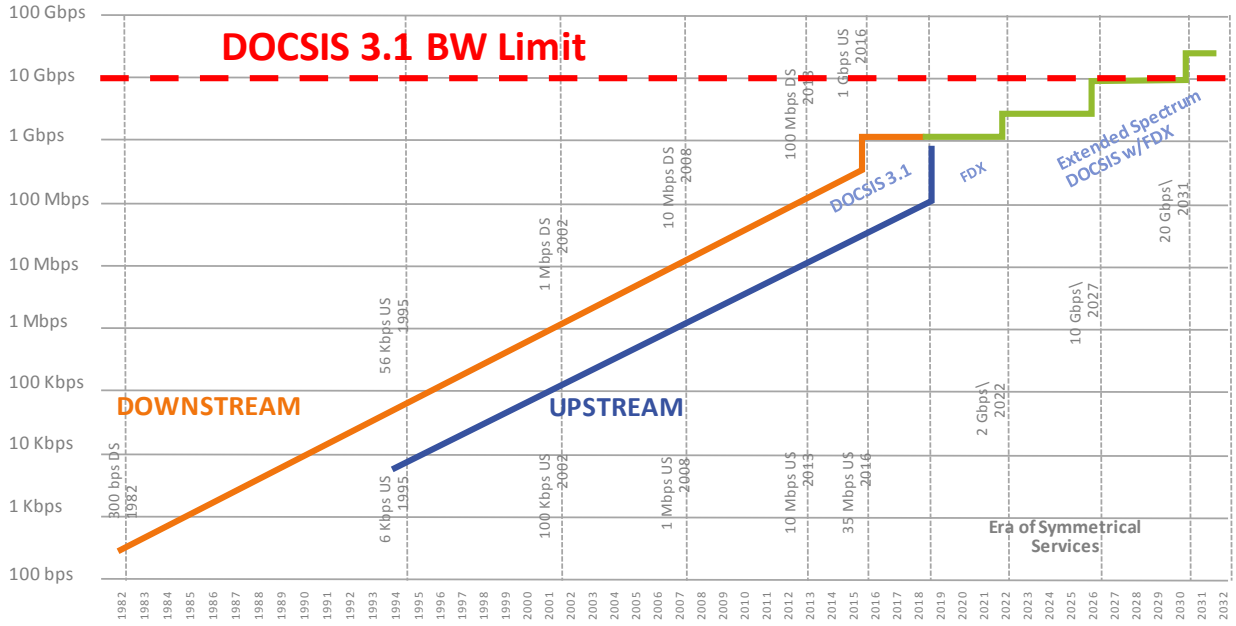

Figure 2 – Modified Nielsen's Law – 30%-40% CAGR after 2018

Will Nielsen's Law continue its 50% growth unabated for the next couple decades? In recent years, there has been some suggestions whether Moore's Law may be slowing down. Will this have a corresponding impact on Nielsen's Law? One could argue that Moore's Law is the fuel behind ever advancing customer premises equipment (CPE) that drive the need for Internet bandwidth. Figure 2 looks at a modified Nielsen's Law where the CAGR is reduced to 30%-40% going forward. This stretches the time for 10X growth from the original 5½ years up to 8 years and is highlighted in Table 1.

Table 1 - Potential Future Service Tier Growth

High SLA BW (Gbps)	Year (50% CAGR)	Year (30-40% CAGR)
1	2016	2016
2	2019	2019
10	2021	2022
20	2024	2025
100	2026	**2029**
200	2030	2032
1000	2032	**2036**

One can see from a network capacity planning perspective; the overall impacts are similar. The changes over the next decade are minimal. Longer term, the time it will take to reach the 1 Tbps milestone gets pushed out about four years, from 2032 to 2036. So, even with a slowing in Nielsen's Law, there will be minimal impact in operators' long-term network capacity planning, just a couple of additional years to execute their plan.

1.1. Selective Subscriber Tier Migration Strategy

At first glance, Nielsen's Law is a scary proposition such that HFC networks might be obsolete in five to seven years while it may take decades to build out an FTTP infrastructure. However, this is not the full story. As was shown in [ULM_2016, ULM_2014], Nielsen's Law applies to the top speed tiers which is only a very small percentage of the entire subscriber base, perhaps less than 1%. So, the key question then becomes, "What happens to the vast majority of subscribers on HFC who are not in the top speed tiers (a.k.a. billboard tiers) and when?"

The [ULM_2014] case study looked at service tier evolution at a few cable operators. Perhaps the key finding from this study is that the different service tiers are growing at different rates. While the top billboard tier continues to follow Nielsen's Law 50%, each subsequent lower speed tier is growing at a slower rate. Hence, the lower the service tier rate, the lower its CAGR.

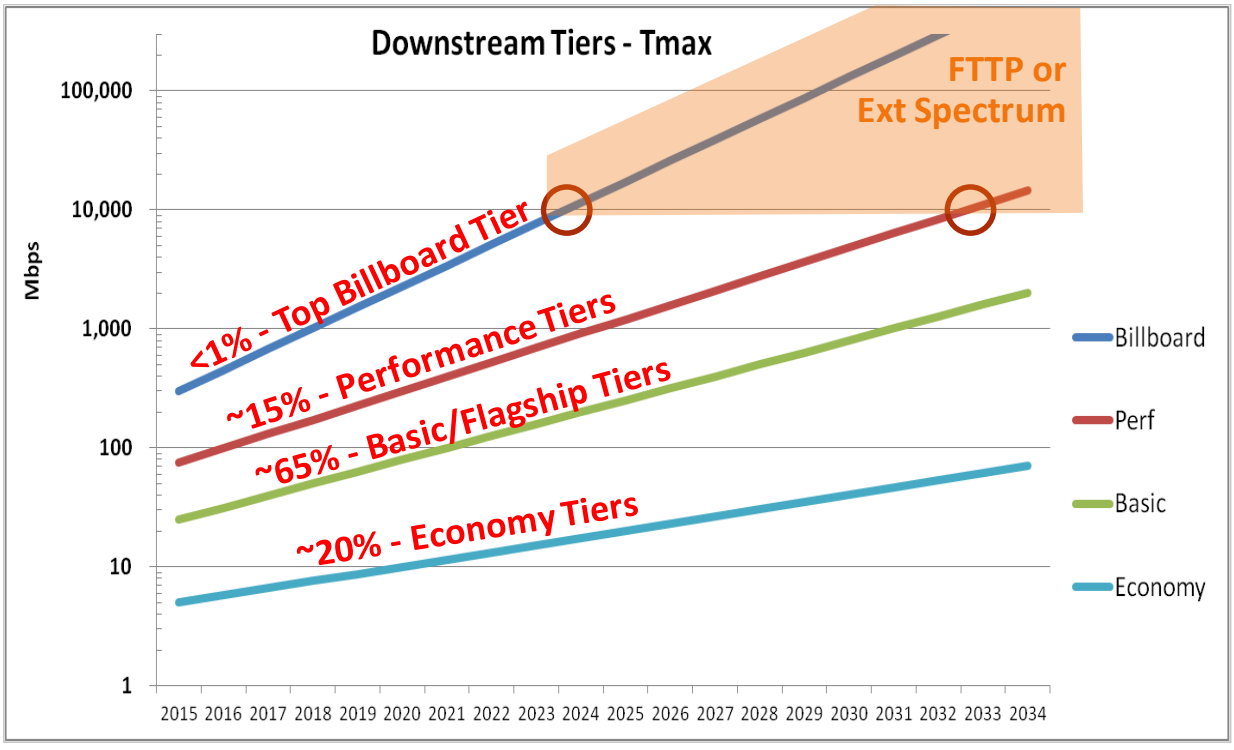

Figure 3 – Downstream Growth with Multiple Service Tiers

Figure 3 maps out a typical scenario from the case study of the various service tier growth over the next two decades. While less than 1% of subs in the top billboard tier hit 10 Gbps in ~2024, the 15% of subs in

the performance tier doesn't hit that mark until ~2032. Notice that 85% of subscribers in the flagship basic tier and economy tier stay below this mark for several decades.

It is important to note that 99% of the subscribers can still comfortably use today's DOCSIS technology on HFC a decade from now. With a selective subscriber migration strategy, it is very important from a traffic engineering perspective to understand the behavior of the individual service tiers. But with this understanding in hand, selective subscriber migration can be used to extend the life of HFC for decades to come. When the time comes to migrate the premium tiers from traditional HFC to new technologies, it might be done with either FTTP technology or maybe extended spectrum HFC if it is viable by that time.

2. Broadband Subscriber Traffic Consumption

Earlier work by Cloonan noted that the primetime average subscriber consumption (a.k.a. Tavg) has also been following this same basic trend as shown in the Figure 1. For service providers, an important metric is the traffic utilization in a service group (SG). The SG traffic utilization is a function of the number of subscribers (Nsub) times the average bandwidth per sub (Tavg). In [CLOONAN_2014, EMM_2014], this research expanded to also include traffic utilization in addition to the network connection speed. This became known as Cloonan's Curve, where SG consumption is shown in addition to Nielsen's Law.

In the early DOCSIS days, many nodes were combined together, and SG traffic was an order of magnitude higher than the maximum network connection speed (a.k.a. Tmax after the DOCSIS parameter that dictates max network rates). Over time, the SG size has been shrinking and, with it, the ratio between Nsub*Tavg to Tmax. The SG traffic will eventually approach that of Tmax. As SG sizes dip below 100 subs, then Tmax starts to dominate the traffic engineering.

ARRIS has been monitoring subscriber usage for many years now. Figure 4 shows Tavg, the average subscriber downstream consumption during peak busy hours, for a number of operators over a nine-year period. At the start of 2017, Tavg finally broke the 1 Mbps barrier.

Over the last four or five years, this group of operators has an average downstream traffic growth that has been just under 40%. This equates to roughly doubling every other year. On a yearly basis, traffic growth can be very sporadic. In 2018, there was some conflicting data for the downstream as a couple operators saw ~70% growth while the others saw ~30% growth. It is not clear yet whether this is a change in the long-term trend or just a single year aberration.

The upstream traffic had been growing at a significantly slower rate of ~20% CAGR for the previous eight-year period. However, the 2018 data showed a significant uptick in upstream usage of ~45% across the board. The upstream growth rate now appears to start rivaling the downstream growth rate.

Figure 6 shows the DS:US BW ratio of downstream to upstream traffic. For the last decade, traffic had been getting more asymmetric, but now we see that three of the four operators have leveled off around a 14:1 ratio. This may imply that upstream growth may match downstream growth going forward.

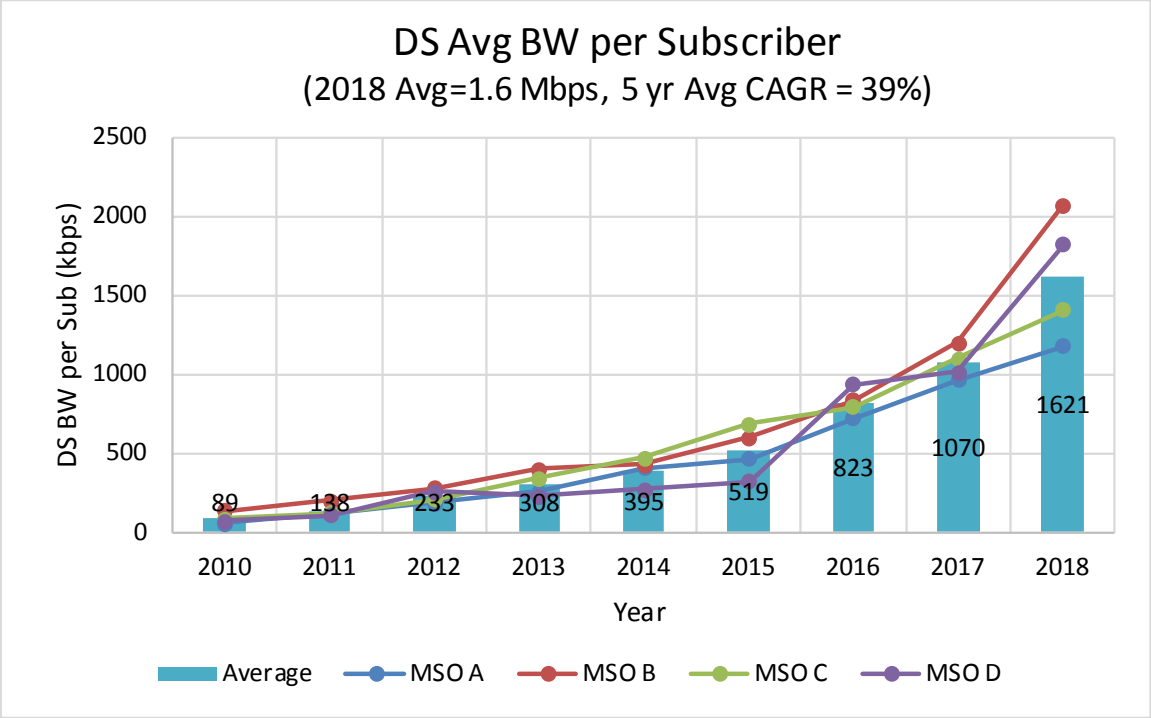

Figure 4 – Tavg, Average Subscriber Downstream Consumption

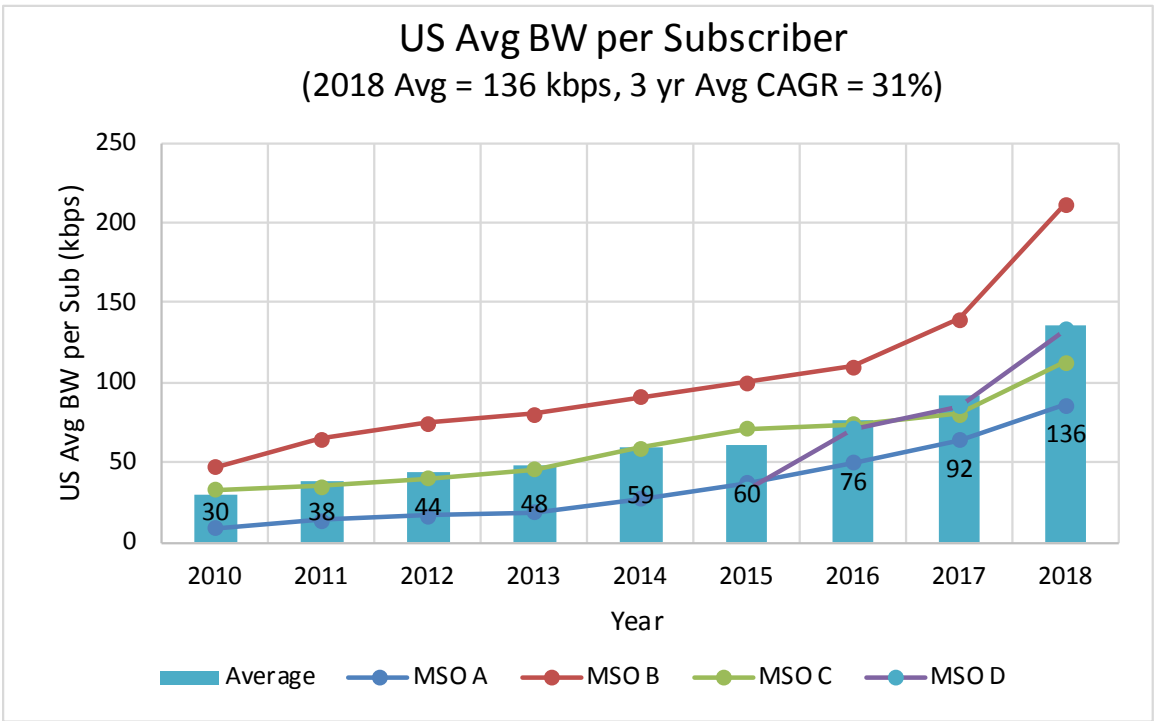

Figure 5 – Tavg, Average Subscriber Uptream Consumption

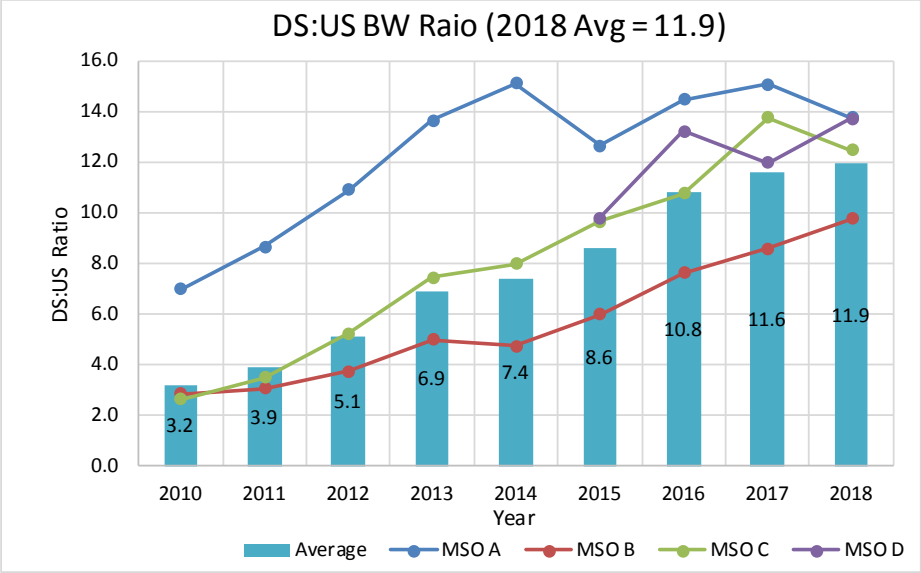

Figure 6 – Downstream:Upstream Bandwidth Ratio

Review of Broadband Traffic Engineering

Previously, [CLOONAN_2014] provided an introduction on traffic engineering and quality of experience (QoE) for broadband networks. From there, the paper went on to develop a relatively simple traffic engineering formula for cable service groups.

1. The "Simple" Traffic Engineering Formula

The "simple" formula shown below is a simple two-term equation. Its simplicity is part of its beauty. The first term (Nsub*Tavg) allocates bandwidth capacity to ensure that the aggregate average bandwidth generated by the Nsub subscribers can be adequately carried by the service group's bandwidth capacity. The first term is viewed as the "DC component" of traffic that tends to exist as a continuous flow of traffic during the peak busy period (e.g., 8:00 p.m. to 11:00 p.m.).

THE "2014" TRAFFIC ENGINEERING FORMULA (BASED ON Tmax_max):

$$C >= (Nsub*Tavg) + (K*Tmax\_max), \quad (1)$$

where:

C is the required bandwidth capacity for the service group
Nsub is the total number of subscribers within the service group
Tavg is the average bandwidth consumed by a subscriber during the busy-hour
K is the QoE constant (larger values of K yield higher QoE levels)…
where 0 <= K <= infinity, but typically 1.0 <= K <=1.2
Tmax_max is the highest Tmax (i.e., service tier) offered by the operator

There are obviously fluctuations that will occur (i.e., the "AC component" of traffic) which can force the instantaneous traffic levels to both fall below and rise above the DC traffic level. The second term (K*Tmax_max) is added in an attempt to increase the probability that all subscribers, including those with the highest Tmax values, will experience good QoE levels for most of the fluctuations that go above the DC traffic level.

The second term in the formula (K*Tmax_max) has an adjustable parameter defined by the K value. This parameter allows the operator to increase the K value and add bandwidth capacity headroom that helps provide better QoE to their subscribers within a service group. [CLOONAN_2014] goes on to provide simulation results that showed a value between K=1.0 and 1.2 would provide good QoE results for a service group of 250 subscribers. Larger SGs would need larger values of K while very small SGs might use a K value less than 1.0.

Traffic engineering research continues to advance as discussed in [ULM_2017]. However, for the remainder of this paper, the "simple" formula is used because of its clarity in showing basic concepts.

2. Examples using the "Simple" Traffic Engineering Formula

Using the simple traffic engineering formula (1), it becomes possible to develop sophisticated network capacity models. Some downstream examples are shown in Figure 7:

- 500 Mbps DS service tier being supported by 32-bonded DOCSIS 3.0
- 1 Gbps DS service tier supported by a bonded mix of DOCSIS 3.0 SC-QAM and 3.1 OFDM
- 3 Gbps DS service tier using 2x192 MHz D3.1 OFDM bonded with 32 SC-QAM 3.0 channels

While the 500 Mbps DS tier needs 192 MHz of spectrum, note that the 1 Gbps DS tier requires 224 MHz to 256 MHz of spectrum depending on the mix of SC-QAM and OFDM channels. Finally, the 3 Gbps DS tier requires a big jump to 576 MHz of spectrum. This gives the operator a sense of the amount of additional downstream spectrum that will be needed going forward with gigabit services.

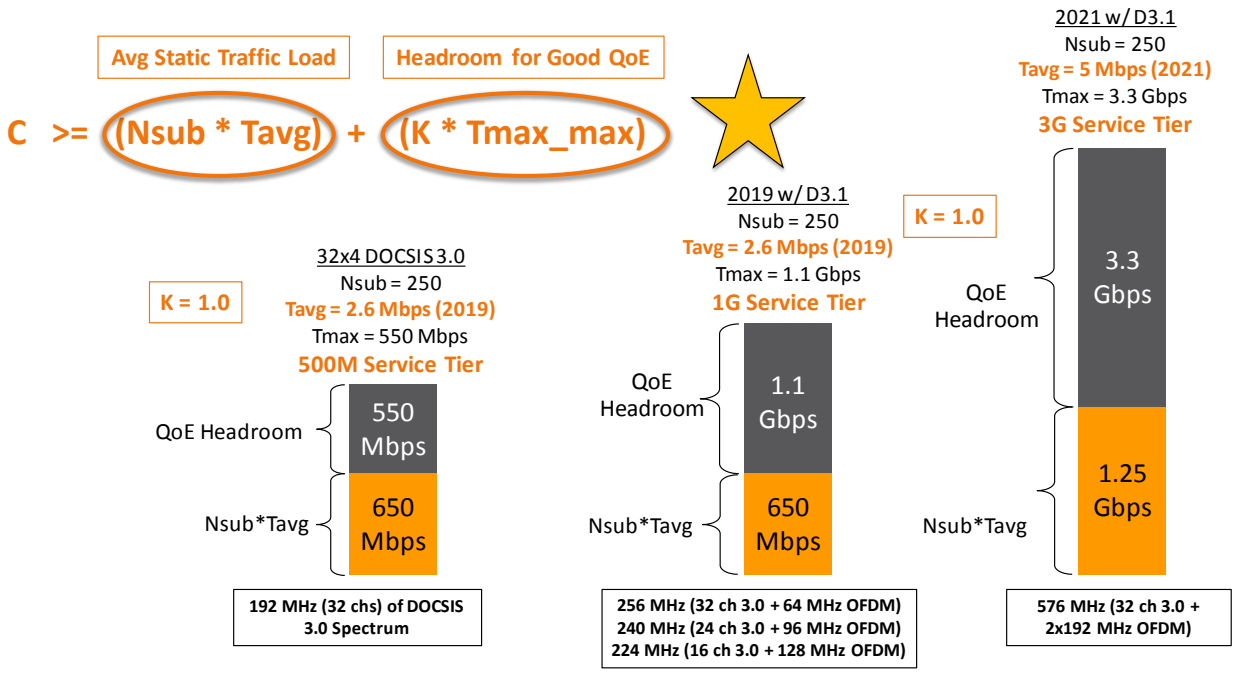

Figure 7 – Traffic Engineering Examples

One interesting observation derived from the "simple" traffic engineering formula is the impact that SG splits have on the available bandwidth capacity. For the scenario in Figure 8, it assumes a starting point in 2016 of Tmax_max = 300 Mbps, Tavg = 1 Mbps, and Nsub = 300. These requirements could be served with 16 to 18 SC-QAM D3.0 channels. This is shown by the leftmost candlesticks in Figure 8.

With a service tier CAGR = 50% and Tavg CAGR = 40%, the required capacity grows to 16 Gbps by 2025. This is represented by the second candlestick drawing. If the operator splits the SG twice, then Nsub drops to 75 subs per SG and is shown by the third candlestick drawing. It is very important to note that splitting SGs only impacts part of the traffic engineering formula (i.e., Nsub * Tavg). This means that splitting SGs reaches a point of diminishing returns. It is very important for the operator to also increase available spectrum within the SG to keep up with growing Tmax!

The fourth candlestick diagram on the right shows the impact of a selective subscriber migration strategy where the 10G tier was moved to another technology (e.g., FTTP or extended spectrum DOCSIS) and the remaining tiers (i.e., 3G and below) can comfortably live on existing HFC with DOCSIS 3.1 technology.

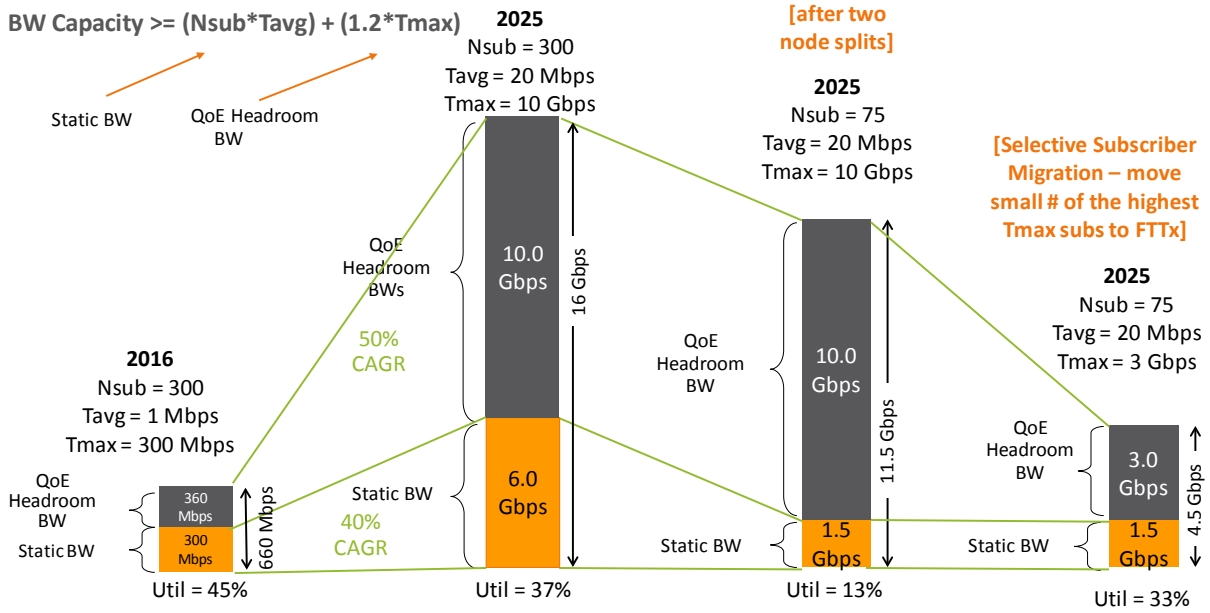

Figure 8 – Applying ARRIS QoE Traffic Engineering

3. Preparing for Symmetric Upstream Service Tiers and Growth

As was shown in the BW Trends section at the start of this paper, operators are starting to consider offering more symmetric upstream service tiers while consumers' upstream consumption rates are now rivaling or even exceeding downstream growth rates. This will put a tremendous strain on the existing 42 MHz and 65 MHz HFC return systems. Even operators that have upgraded to an 85 MHz upstream will not have enough capacity to offer a 1G upstream service tier.

There are a couple potential paths to get to gigabit upstream services. One approach is using DOCSIS 3.1 to support a 204 MHz upstream split. This provides enough spectrum to support 2x96 OFDMA channels capable of 1.5 to 1.8 Gbps. As was seen in our earlier DS traffic engineering example, this can comfortably handle the 1G US service tier. But this comes at a steep price in the downstream which now starts at 258 MHz. That means the operator must find 150 MHz of spectrum for displaced DS services.

Other operators are looking at supporting FDX DOCSIS for gigabit upstream services. While FDX maintains the 85 MHz split in a N+0 system, the FDX spectrum is limited to D3.1 OFDM channels. The FDX spectrum will go to 300 MHz initially, but then expand to 684 MHz in the future. This puts a serious crunch on existing legacy video and DOCSIS 3.0 channels. Many legacy STBs can only operate to 870 MHz while DOCSIS 3.0 hits a ceiling at 1002 MHz. As hundreds of megahertz of lower spectrum are lost to the upstream, the service provider must find a way to shrink the spectrum requirements for legacy video and DOCSIS 3.0.

SDV 2.0 – Re-inventing SDV for CCAP Era

After some initial success a decade ago, SDV fell out of favor with many operators. However, most if not all of the original barriers to SDV have been removed and the industry is seeing a resurgence of interest in using SDV to enable their gigabit service evolution.

1. Technology Options to enhance Bandwidth Capacity

SDV is one of many bandwidth tools available to cable operators. It is good to understand how SDV fits relative to the other options. A summary of possible options is shown in Table 2.

Table 2 – Technology Options to enhance Bandwidth Capacity

Typical Options for last few years:	Ease & Cost	Tmax	Nsub*Tavg	Notes
Analog Reclamation	☺	☺☺☺ (DS only)	☺	Need DTAs
Node Segmentation, splits	☺ to ☹		☺☺☺	HFC plant upgrades
DOCSIS 3.1	☺☺	☺☺		Need D3.1 CPE
More Options for coming years:				
Fiber deep – N+0, N+1; 1.2 GHz DS, 85-204 MHz US	☹	☺☺☺☺☺	☺☺☺☺☺	Long term strategic direction, often combined with DAA
MPEG-4 transition	☺☺	☺☺ (DS only)		If mostly MPEG-4 capable STB
IPTV transition	☺	☺ -> ☺☺☺ (DS only)	☺ (DS only)	Need IP capable CPE; gains depend on IPTV strategies
SDV 2.0	☺☺☺	☺☺☺ (DS only)	☺ (DS only)	Leverage existing legacy STB
Future Options:				
DOCSIS FDX	☹☹	☺☺☺ (US only)		Assumes DAA & N+0, too
FTTH, Ext Spectrum, FTTTap	☹☹☹	☺☺☺☺☺	☺☺☺☺☺	Next steps after fiber deep

Over the past decade or so, most operators have reclaimed analog spectrum and continued to do node segmentation and node splits as business as usual. It turns out that each addresses a different component in our traffic engineering formula. The analog reclamation frees up spectrum which is critical to offering higher service tier SLA (i.e., Tmax) while node segmentation and splits reduce SG size which directly addresses the Nsub*Tavg component in our traffic engineering formula.

More recently, operators have started using DOCSIS 3.1 as they migrate to 1G downstream service tiers. D3.1 operates in today's existing HFC plants without any changes. It improves spectral efficiency (i.e., bps/Hz) enabling more capacity from a given amount of spectrum. D3.1 is also robust enough to operate

in the roll-off region to additional bonus capacity. It is straightforward to implement by deploying D3.1 cable modems and some SW/HW upgrades to the existing CMTS/CCAP platforms.

Over time, the D3.1 benefits will continue to grow as the plant is improved. The spectral efficiency improves significantly with fiber deep networks and it enables operation over wider frequencies (e.g., 1218 MHz downstream, 85/204 MHz upstream).

From a strategic perspective, all operators plan to push fiber deeper until it eventually becomes fiber to the home (FTTH). However, this is a multi-decade process. Some operators are considering a fiber deep HFC (with N+0, N+1 cascades) as the next major step along the way. But this can be a monumental task with pulling fiber deeper in the plant and increasing nodes counts by a factor of 10X to 20X. So, while this option gets top marks for increasing spectrum AND significantly reducing SG size for both upstream and downstream, its major costs and complexities will force this option to be done slowly over time.

If an operator is focused on more easily increasing downstream bandwidth capacity, there are several other options available in the near term including:

- Migrating broadcast video to MPEG-4/H.264 on existing legacy STBs
- Migrating broadcast video to IP video delivery over IP STBs
- Migrating broadcast video to SDV over existing legacy STBs

The MPEG-4/H.264 option makes sense if all or the clear majority of the legacy STBs support MPEG-4 decoding. Overall, it provides a 2:1 gain in broadcast spectrum converted. Problems arise if there is still a substantial MPEG-2 only STB user base. That may limit the broadcast programs that can be converted (e.g., limited program tiers) or require those older STBs to be replaced.

The IP video migration includes a wide array of potential solutions. Some operators may only move video on demand (VOD) and select programming content to IP video while other operators may move aggressively to IP video everywhere. The potential bandwidth capacity gains from reducing legacy broadcast video can vary dramatically as well. In addition to an all new video infrastructure, the operator also needs to replace its CPE with IP video capable boxes. During the transition window, the operator must continue to support the legacy STBs, creating a bubble in their bandwidth needs.

The third option for reducing legacy broadcast spectrum is SDV. In the early days with large SGs and limited number of SDV EQAM modulators, the spectrum gains were somewhat limited (e.g., 84 MHz). However, with today's current technologies, SDV can be used more aggressively and achieve gains in excess of 400 MHz. A SDV case study is described in a following section. SDV also has the advantage over the other approaches in that it works on existing legacy STBs (MPEG-2 and/or MPEG-4) and requires minimal video infrastructure.

The final options for increased bandwidth capacity are looking much farther into the future. Operators will be considering options such as FDX DOCSIS to enhance the available upstream spectrum and network architectures such as FTTH, fiber to the tap (FTTT) and DOCSIS extended spectrum (e.g., 1.8+ GHz).

2. Leveraging SDV to make room for D3.1 &/or IP Video Migrations

2.1. SDV Early Days – Past Barriers to Deployments

SDV burst onto the scene over a decade ago. At that point in time, the cost of an EQAM channel was relatively expensive and the EQAM modulator hardware only supported a couple channels per port. The SDV systems deployed tended to have very large SGs with thousands of tuners. This really limited the spectrum gains from SDV to really long tail content that was almost never watched.

Installing the SDV narrowcast EQAM modulators was also an onerous task that required re-wiring the RF combining in the headend. So, trying to change the SDV SG size was very inflexible. Even as technology changed, many existing SDV systems continued to use these conservative parameters.

And at that time, this was new technology with a large R&D investment, so the SW costs tended to be very high. This tended to limit the solution to some of the large operators. It was hard for smaller operators and smaller sites to justify the investments for a limited bandwidth gain.

As time moved on, SDV got a stigma of being an "old" technology and potentially putting money into stranded capital. Operators would rather consider a migration using MPEG-4 or IP video as moving in the "right" direction.

2.2. SDV 2.0 – Leveraging Today's Technology to Remove Barriers

Switched digital video is an extremely powerful technology. It is effectively delivering switched IP video to our legacy STBs using cloud-based resources. This is being done without replacing any CPE; without touching the outside plant; and with very minimal changes to the video infrastructure. It has tremendous potential for saving additional spectrum and creating bandwidth capacity.

Cable technologies have changed significantly over the last decade to now allow us to take full advantage of SDV capabilities. The biggest change is brought about by the CCAP box. Not only was this the next step in the evolution of EQAM modulator density increase and cost reduction, but it integrated the EQAM modulator inside with the CMTS channels onto a single integrated RF port. This means that SDV narrowcast EQAM modulators could be turned on via SW control with absolutely no changes to the headend's RF combining network.

Besides enabling much higher counts of SDV EQAM modulator-based channels, the CCAP box also enabled the SDV SG to become aligned with the much smaller DOCSIS SG. As the following case study shows, this enables massive potential spectrum savings of hundreds of megahertz.

With SDV becoming a mature technology, the business model has also evolved. Now more cost-effective solutions are available based on a price model where the operator pays per tuner and per QAM channel. This greatly lowers the bar to entry enabling much smaller systems to now consider SDV. And since it is an investment in SW running on COTS servers, there is no stranded HW capital investment. Once SDV is no longer needed, the COTS servers can be easily repurposed for other applications.

While SDV can now be considered a viable BW option in its own right, it also compliments both the MPEG-4 and IP video migration strategies as well. As mentioned previously, the MPEG-4 strategy works great if all the legacy STBs are MPEG-4 capable but can be problematic if MPEG-2 only STBs are still in the field. SDV can be used to address the MPEG-2 only STB, moving most or all of the MPEG-2 content

onto the switched tier. As the MPEG-2 only STB penetration declines, the number of SDV narrowcast EQAM modulators also decreases and the spectrum gains increase. Once the SDV system is in place, then MPEG-4 STB can also take advantage and the MPEG-4 long tail content can be moved onto the switched tier as well.

As mentioned previously, the IP video migration will require a relatively lengthy period to changeover all the legacy STBs to IP video-capable CPE. During this period, video services must be provided to both which creates a bandwidth bubble that compounds the situation. SDV should be considered as part of the IP video migration strategy. It is effectively delivering switched IP video (to the EQAM modulator) for delivery to legacy STBs. As IP video CPE penetration increases and legacy STB penetration decreases, the gains seen from SDV continue to increase. SDV can not only eliminate the BW bubble but provide a net positive gain in spectrum saved.

The bottom line is that SDV costs should be considered as part of the solution that enables spectrum savings for gigabit services, so it is part of the overall BW creation solution. While CCAP provided the initial impetus, future distributed access architectures like remote PHY and remote MAC/PHY will continue support for the integrated EQAM modulator.

3. SDV Case Study – Freeing up Spectrum for Gigabit Services

3.1. SDV Case Study Overview

To illustrate the capabilities of SDV, a recent use case is analyzed where a service provider decided to use a combination of SDV and MPEG-4/H.264 migration to free enough capabilities to quickly offer gigabit services across their entire footprint.

This service provider is in an extremely competitive market and needed to offer gigabit services quickly across its entire subscriber base. This immediately precluded any plant upgrades due to cost and time. Initial traffic engineering estimates indicated that they needed to free a minimum additional 72 MHz downstream spectrum for a DOCSIS 3.1 OFDM channel for a 1G DS service tier. 96 MHz of spectrum would be preferred to provide some additional cushion. Eventually, they want to provide 2x192 MHz D3.1 OFDM channels so they can consider 2G and 3G DS service tiers in their competitive market.

Approximately one-sixth of the installed STB base was still MPEG-2. This precluded using only an MPEG-4 transition plan. The service provider was already using SDV but was still locked into using the older inefficient SDV parameters. This included SDV SG sizes up to 3000 STBs with only six SDV narrowcast EQAM modulators. The initial broadcast/switch program mix included 400 total SD programs of which 250 were broadcast and 150 switched; and 178 total HD programs of which 160 were broadcast and 18 were switched. This resulted in a net savings of 84 MHz (or 14 CTA SC-QAM channels).

The first step was to move the six external SDV EQAM modulators inside the integrated CCAP box. This now allows the SDV SG to be aligned with the DOCSIS SG and drop to about 600 STBs. From there, increasing the number of SDV EQAM modulators is possible through SW licensing to provide additional spectrum savings.

3.2. ARRIS Network Capacity Model predicts SDV Spectrum Gains

The ARRIS Network Capacity Modeling Tool was leveraged to estimate potential SDV gains over time. The key parameters to maximize SDV spectrum saving include:

- SDV SG size
- SDV EQAM modulator channels/SG
- Broadcast/switch program mix

Some results from the modeling are shown in the following two figures. Figure 9 shows the number of narrowcast SDV EQAM modulators that are required for various broadcast/switched configurations. Note that the X-axis is reversed and goes from a very large 3000 STB SG down to zero. Figure 10 shows the total number of CTA SC-QAM channels required for each of the various broadcast/switched configurations. This total includes both the broadcast QAMs + SDV EQAM modulators needed.

The yellow lines in both figures represents the SDV system as it is configured today. Note the yellow circles that indicate the current operating point with 3000 STBs per SDV SG. Figure 9 shows a total of six SDV EQAM modulators required while Figure 10 indicates a total of 81 QAM channels needed for video. This is 14 channels (or 84 MHz) less than the 95 QAM channels needed for 100% broadcast video.

If the SDV SG size is left alone and only the SDV EQAM modulator is increased, a point of diminishing returns is quickly reached. For example, quadrupling the SDV EQAM modulator to 24 channels allows more content to be put on the switched tier (i.e., pink line with 108 HD B-cast, 70 HD SDV). But at 3000 STBs per SDV SG, this only reduces total QAM channels to 68. A heavy price to pay for additional 13 channels of savings.

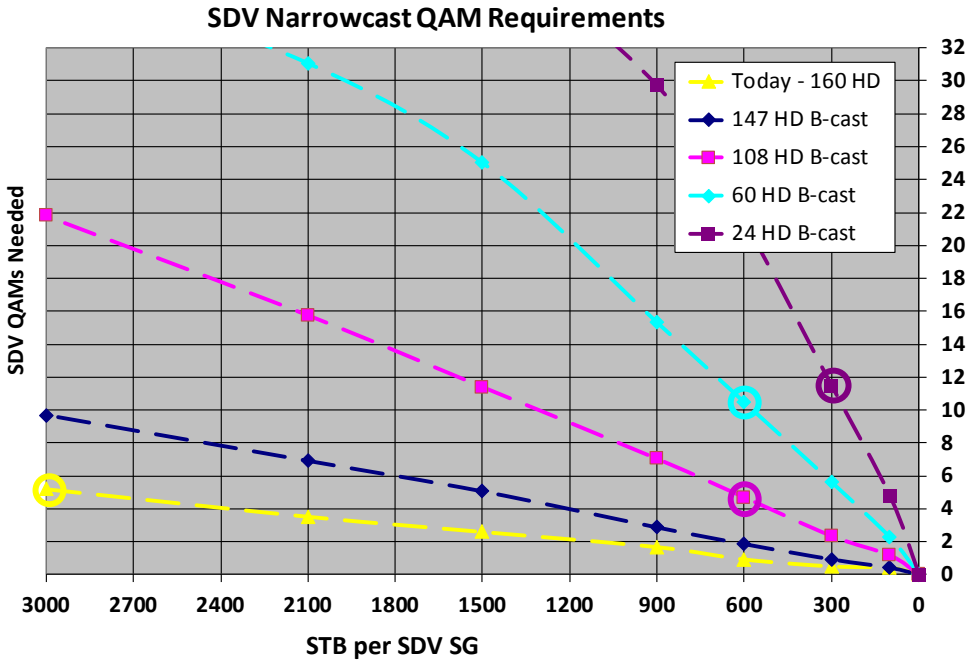

Figure 9 – SDV Narrowcast QAM Channel Requirements

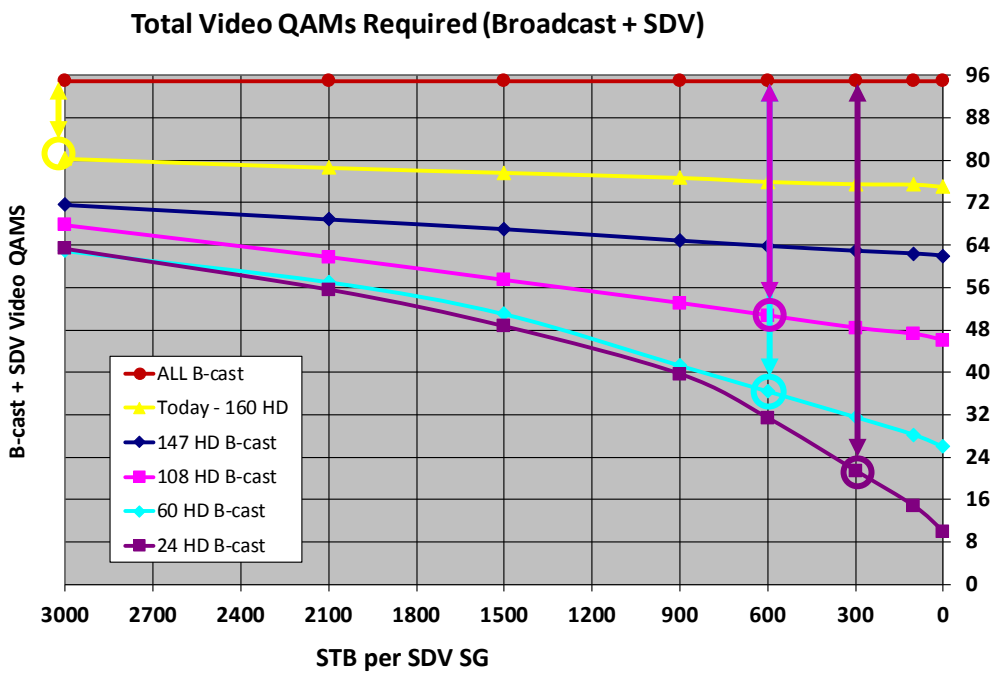

Figure 10 – Total Video QAM Channels Required (B-cast + SDV)

By aligning the SDV SG with the DOCSIS SG inside the integrated CCAP, the SDV operating point can be moved down to 600 STB per SG while keeping six SDV EQAM modulators per SG. This point is shown with the pink circle in Figure 9 and Figure 10. With much smaller SG sizes, SDV becomes much more effective. That's because the smaller viewing population is watching far fewer unique programs during peak busy hour.

Notice from Figure 10 that the total video QAM channels required is now down to 55, which is a 40 channel or 240 MHz savings over the original broadcast-only lineup and a 26 channel or 156 MHz savings over the existing SDV operating point. This more than meets the service provider's initial goals for offering 1G DS services to their customers.

Over time, the service provider can become more aggressive with SDV to get additional spectrum savings. By increasing the SDV EQAM modulator to 12 channels @ 600 STBs/SG, the service provider can now move to a new operating point shown by the light blue circle. Even more content is put onto the switched tier with only 60 HD broadcast programs and 112 HD switched programs. This provides a total savings of 57 of 95 broadcast channels for 342 MHz.

As above, there is a point of diminishing returns, even at 600 STBs/SG. If the SDV EQAM modulator count is doubled again to 24 channels, only an additional six channels are saved beyond the 57 channels of savings.

Eventually, broadband growth will lead to the SG being split, reducing the SDV SG size to 300 STBs/SG. With 12 SDV EQAM channels, the service provider can now move to a new operating point shown by the maroon circles. At this point, only 24 HD programs are being broadcast while the remaining 154 HD programs are switched. The service provider has now recovered 73 of the original 95 broadcast channels

for a savings of 438 MHz. This is more than enough to add a pair of D3.1 192 MHz OFDM channels and/or change the upstream split to 204 MHz for symmetric gigabit services.

3.3. Teamwork – SDV and MPEG-4 migration working together

The previous section shows the potential of an SDV system. But in the real world, the service provider needs to analyze actual user viewership patterns and determine which programs can be moved to the switched tier and which programs remain broadcast. On day one, the SDV system can be put into place and monitor viewership patterns while all programs are still part of the broadcast tier. Once the viewership data is available, then a plan can be put into place for migrating programs to the switched tier.

For this use case, the service provider also wanted to start migrating its legacy video infrastructure from MPEG-2 delivery to MPEG-4/H.264 delivery which doubles video capacity in the same spectrum. The problem is that one-sixth of the outstanding STBs are older MPEG-2 only boxes. For a typical SG with 600 STBs, that means ~100 STBs can only receive MPEG-2 video. The remaining ~500 STBs can receive either MPEG-2 or MPEG-4 video. Both types of STBs can convert HD content to SD outputs if needed.

With that restriction, the service provider decided to use SDV for the MPEG-2 only STBs rather than replacing all these STBs in the field. Working with the ARRIS service team, they came up with the following progression of moving content to the switched tier:

1. Reduce SDV SG tier to align with I-CCAP SG using existing SDV parameters
2. Monitor viewership patterns across all STBs
3. Move all remaining SD content to switch
4. Move most MPEG-2 only HD programs to switch
5. Move MPEG-4 HD long tail programs to switch
6. Increase SDV EQAM count from 6 to 12 channels and move more MPEG-4 programs to switch
7. Split SDV SG size and move even more MPEG-4 programs to switch

Steps 1 and 2 above are putting the necessary components into place for future expansion. Step 3 is possible since popular programs in HD broadcast can be converted to SD at the STB and less popular SD only programs make a nice fit on the switch tier.

The goal for step 4 was to move all the MPEG-2 only HD programs to the switch tier. As was seen by the network capacity analysis in the previous section, a point of diminishing returns is often reached. With ~100 MPEG-2 only STB, there may be one or two dozen MPEG-2 programs that should remain in the broadcast lineup with the majority going to the switched tier.

As time progresses, the service provider will continue to gather more viewership data. With this knowledge, more of the remaining MPEG-4 HD content can be moved to the switched tier. As the SDV EQAM channels increase and the SDV SG continues to shrink, additional spectrum gains can be achieved over time.

4. Conclusion – SDV as an Important Bandwidth Capacity Tool

Switched digital video is an extremely powerful technology. It is effectively delivering switched IP video to legacy STBs using cloud-based resources. This is being done without replacing any CPE; without

touching the outside plant; and with very minimal changes to the video infrastructure. It has tremendous potential for saving additional spectrum and creating bandwidth capacity.

Adding multiple 192 MHz D3.1/FDX channels puts tremendous strain on the spectrum budget. Splitting nodes hits a diminishing point of return and doesn't solve the issue. Upgrading your HFC to fiber deep 1.2 GHz plant is costly AND time-consuming. It turns out a decade old technology, SDV, is coming back to rescue the day!

Original SDV deployments a decade ago were limited by existing EQAM modulator costs and densities. This meant SDV service groups tended to be very large (e.g., thousands of tuners) with a handful of EQAM-based channels. This was good enough then to free up ~96 MHz of precious spectrum for HD video or D3.0 channels. Today, the integrated-CCAP has fundamentally changed SDV dynamics. It allows SDV SG size to align with much smaller DOCSIS SG sizes and for significantly more EQAM-based channels. Both factors combine to amplify the SDV spectrum savings to hundreds of megahertz.

The ARRIS Network Capacity Model shows that more than 400 MHz of spectrum can be reclaimed and repurposed for DOCSIS gigabit services, including symmetric upstream services too. And this is done on an I-CCAP box with no outside plant changes and no re-wiring in the headend!

The highlighted use case shows how MPEG-4 migration with SDV results in much more bandwidth savings than either could achieve on its own. The speed of deployment was a critical concern for this service provider. Plant upgrades occur over months or years. The SDV system was deployed over their entire footprint in a matter of weeks without ever touching the outside plant.

This identical strategy can be used with an IP video migration as well where an increasing number of subscribers have IP STBs while a decreasing number of legacy STBs use SDV to minimize network capacity devoted to legacy video.

So, while SDV technology has been relatively dormant over the last five to 10 years, the I-CCAP deployments allow operators to leverage re-born SDV to address the rapid growth of D3.1/FDX in addition to MPEG-2 to MPEG-4 transition and IP video migration strategies.

The bottom line is that SDV costs should be considered as part of the solution that enables spectrum savings for gigabit services, so it is part of the overall BW creation solution. While CCAP provided the initial impetus, future distributed access architectures like remote PHY and remote MAC/PHY will continue support for the integrated EQAM modulator.

Bibliography & References

[CLOONAN_2013] Tom Cloonan, Jim Allen, Tony Cotter, and Ben Widrevitz, "Advanced Quality of Experience Monitoring Techniques for a New Generation of Traffic Types Carried by DOCSIS," Proceedings, The NCTA Cable Show Spring Technical Forum (June 2013).

[CLOONAN_2014NCTA] Tom Cloonan, Mike Emmendorfer, John Ulm, Ayham Al-Banna, and Santhana Chari, "Predictions on the Evolution of Access Networks to the Year 2030 and Beyond," Proceedings, The NCTA Cable Show Spring Technical Forum (April 2014).

[CLOONAN_2014EXPO] "Predictions on the Evolution of Access Networks to the Year 2030 & Beyond"; T. Cloonan, M. Emmendorfer, J. Ulm, A. Al-Banna, S. Chari, The Cable Show NCTA/SCTE Technical Sessions Spring 2014

[CLOONAN_2015] "Lessons from Telco and Wireless Providers: Extending the Life of the HFC Plant with New Technologies," Tom Cloonan et. al., The NCTA Cable Show Spring Technical Forum, May, INTX 2015

[CLOONAN_2016] "Using DOCSIS to Meet the Larger Bandwidth Demand of the 2020 decade and Beyond" Tom Cloonan, Ayham Al-Banna, Frank O'Keeffe; The NCTA Cable Show Spring Technical Forum, INTX 2016

[EMM_2014] "Nielson's Law vs. Nielson TV Viewership for Network Capacity Planning," Mike Emmendorfer, Tom Cloonan; The NCTA Cable Show Spring Technical Forum, April 2014

[ULM_2014] "Is Nielsen Ready to Retire? Latest Developments in Bandwidth Capacity Planning", John Ulm, T. Cloonan, M. Emmendorfer, J. Finkelstein, JP Fioroni; 2014 SCTE Cable-Tec Expo

[ULM2_2014] "Scaling Traditional CCAP To Meet the Capacity Needs of the Next Decade" John Ulm, Tom Cloonan; 2014 SCTE Cable-Tec Expo

[ULM_2016] "Giving HFC a Green Thumb: A Case Study on Access Network and Headend Energy & Space Considerations for Today & Future Architectures" John Ulm, Zoran Maricevic; 2016 SCTE Cable-Tec Expo

[ULM2_2016] "Adding the Right Amount of Fiber to Your HFC Diet: A Case Study on HFC to FTTx Migration Strategies", John Ulm, Zoran Maricevic; 2016 SCTE Cable-Tec Expo

[ULM2_2017] "Traffic Engineering in a Fiber Deep World", John Ulm, Tom Cloonan; 2017 SCTE Cable-Tec Expo

Abbreviations

BAU	business as usual
Bcast	broadcast
BPS	bits per second
BW	bandwidth
CAA	centralized access architecture
CAGR	compounded annual growth rate
CAPEX	capital expense
CCAP	converged cable access platform
CM	cable modem
CMTS	cable modem termination system
COTS	commercial off-the-shelf
CPE	customer premises equipment
D3.1	Data-Over-Cable Service Interface Specifications version 3.1
DAA	distributed access architecture
DCA	distributed CCAP architecture
DEPI	downstream external PHY interface
DOCSIS	Data-Over-Cable Service Interface Specifications
DS	downstream
DWDM	dense wavelength division multiplexing
E2E	end to end
EPON	Ethernet passive optical network
EQAM	edge quadrature amplitude modulation [modulator]
FD	fiber deep
FDX	full duplex
FTTH	fiber to the home
FTTLA	fiber to the last active
FTTP	fiber to the premises
FTTT	fiber to the tap
FTTx	fiber to the ‘x’ where ‘x’ can be any of the above
Gbps	gigabits per second
GHz	gigahertz
HFC	hybrid fiber/coax
HP	homes passed
HSD	high speed data
I-CCAP	integrated converged cable access platform
IEEE	Institute of Electrical and Electronics Engineers
IEQ	integrated edge QAM [modulator]
IP	Internet protocol
LDPC	low density parity check
MAC	media access control
MAC/PHY	media access control/physical layer
Mbps	megabits per second
MDU	multiple dwelling unit
MHz	megahertz
MSO	multiple system operator
N+0	node plus zero
Ncast	narrowcast
NFV	network functions virtualization

NSI	network side interface
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OLT	optical line termination
ONU	optical network unit
OOB	out of band
OPEX	operating expense
OTT	over the top
PHY	physical layer
PNM	proactive network maintenance
PON	passive optical network
QAM	quadrature amplitude modulation
QoE	quality of experience
QoS	quality of service
RF	radio frequency
R-OLT	remote OLT
RPD	remote PHY device
R-MAC/PHY	remote MAC/PHY
R-PHY	remote PHY
RX	receive
SDN	software defined network
SDV	switched digital video
SG	service group
SCTE	Society of Cable Telecommunications Engineers
SNR	signal-to-noise ratio
STB	set-top box
TaFDM	time and frequency division multiplexing
Tavg	average bandwidth per subscriber
Tmax	maximum sustained traffic rate
TX	transmit
US	upstream
VOD	video on demand
WDM	wavelength division multiplexing

Common Path Distortion (CPD) in Digital Cable Networks

A Technical Paper prepared for SCTE•ISBE by

Tom Williams, Distinguished Technologist CableLabs, SCTE•ISBE Member
858 Coal Creek Circle
Louisville, CO 80027
t.williams@cablelabs.com
303-661-3486

Jason Rupe, Principal Architect, CableLabs, SCTE•ISBE Member
858 Coal Creek Circle
Louisville, CO 80027
j.rupe@cablelabs.com
303-661-3332

Table of Contents

List of Figures

1. Introduction

1.1. Background

Upstream interference caused by ingress is the most common source of data errors. Less common, but still significant, is common path distortion (CPD). CPD is a non-linear distortion impairment that is typically created by downstream signals mixing together in corrosion diodes. These corrosion diodes can be created by contact between dissimilar metals. In wireless networks this same phenomenon is known as passive intermodulation (PIM). In cable networks mixing products landing in the 5 MHz to 42 MHz band, propagate upstream toward the headend or hub, and can cause interference with upstream cable signals. Mixing products also land in other bands causing reduced performance there.

In past days of mostly- or all-analog television channel loading, the downstream video carriers were separated by 6 MHz (in North America) and the upstream distortion products appeared as sets of three beat clusters every 6 MHz. An example is illustrated in the spectral plot of Figure 1. In Figure 2 the beats are identified as second or third order distortion products. Now that the downstream analog TV carriers have been replaced by digital signals, the distinct CPD beats every 6 MHz have been replaced by an elevated "noise" floor, which are nonlinear distortion products. CPD has not gone away, it has just changed appearance, and is not as obvious to field techs.

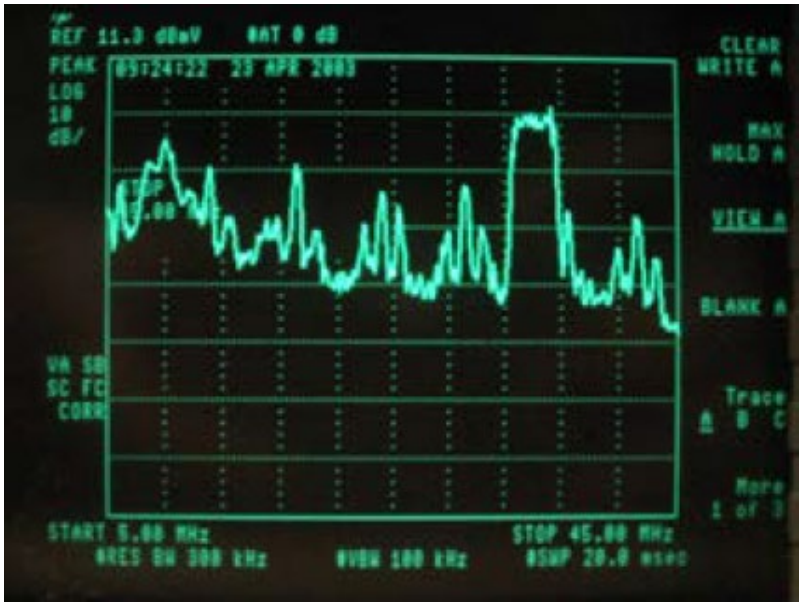

Figure 1 - Common path distortion beats every 6 MHz in the upstream (return) band, from analog downstream video carriers

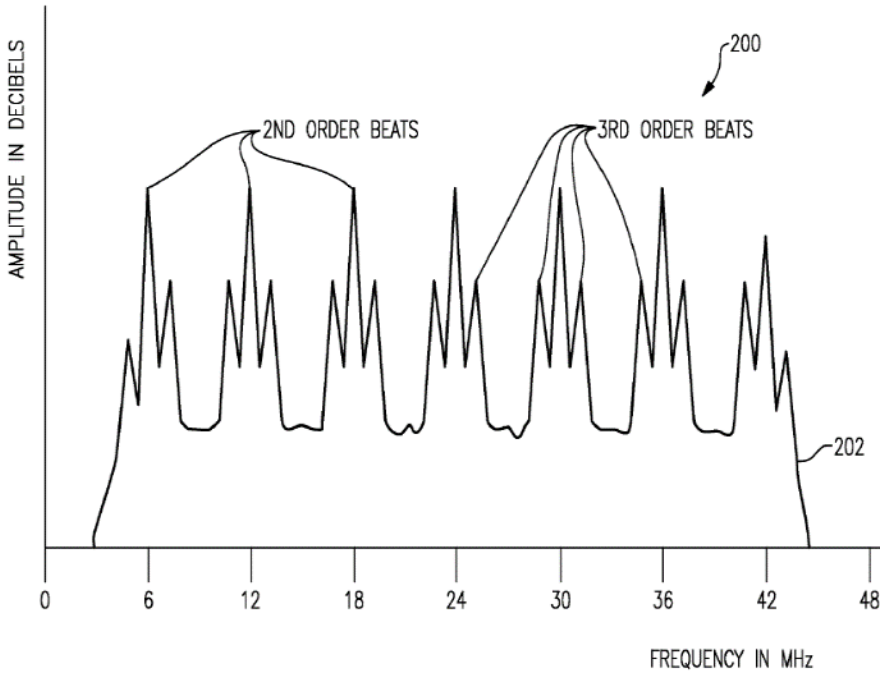

Figure 2 - Explanation of CTB distortion beats

ANSI/SCTE 109 2016 (https://www.scte.org/SCTEDocs/Standards/ANSI_SCTE%20109%202016.pdf) describes a method for testing a network or device for CPD in a lab. It consists of applying two elevated continuous wave (CW) signals to the downstream band and examining the upstream band for a CW signal at the difference frequency. For example, if two CW carriers at 800 MHz and 840.5 MHz were injected into a network and a CW difference frequency of 40.5 MHz were observed, that difference signal would be second order CPD. This procedure was tested in the field on live plant and shown to work outdoors also. Figure 3 is a spectral plot of a 40.5 MHz CW difference product observed in the headend of an operating cable network. This test method was done on about 67 active nodes, and a histogram of 40.5 MHz beat products' signal levels was made. The histogram is illustrated in Figure 4. Elevated CPD was observed on about six of the nodes. But a different phenomenon was observed on another four nodes: the CPD was modulated at a rate of 120 Hz, twice the AC line frequency. Figure 5 through Figure 8 are amplitude vs. time plots of dynamic beat products (i.e., the 120 Hz modulated CPD). The CPD products, both dynamic and static, were also observed to vary with time. That is because the corrosion diode junction is thought to be only a few molecules thick, so a slight change in temperature or mechanical stress such as vibration can cause it to disappear or reappear.

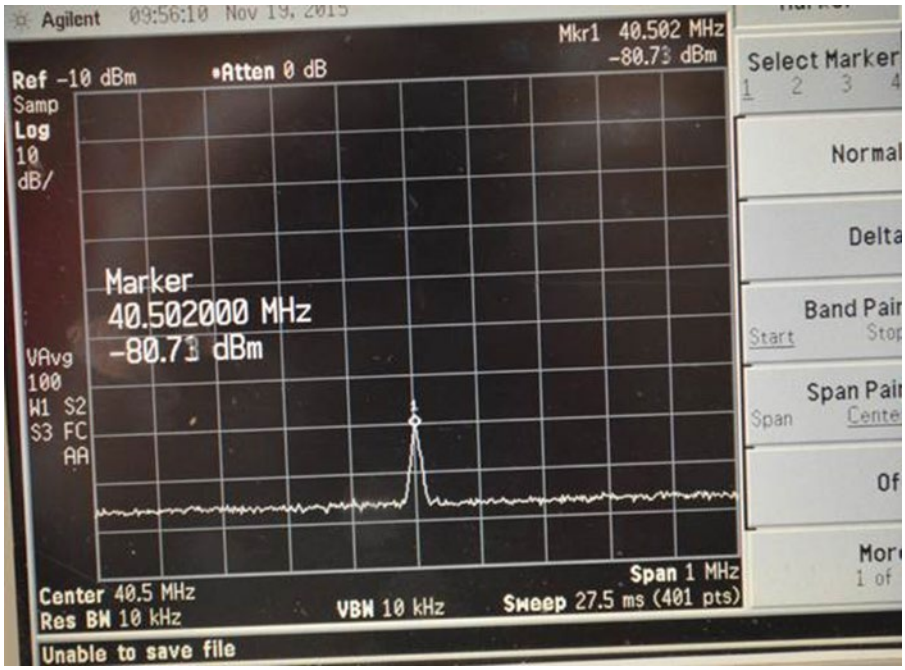

Figure 3 - A steady CW distortion beat product observed in a cable system's headend upstream test point. It was created as a difference product between two downstream test CW carriers at 840.5 MHz and 800 MHz

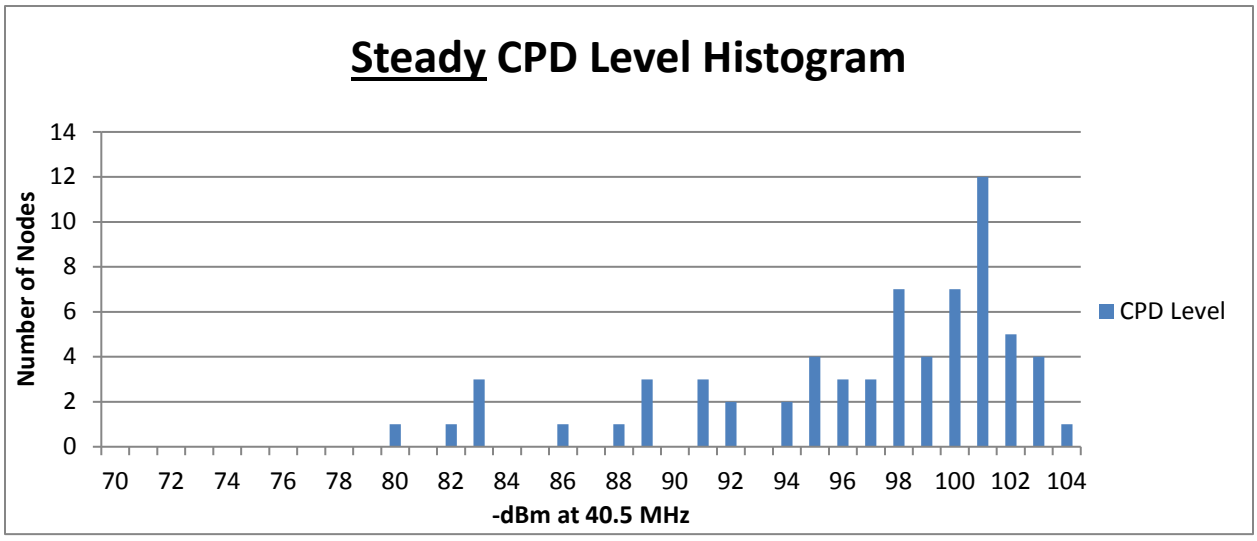

Figure 4 - A histogram of 67 beat measurements on 67 nodes. Only the seven strongest are considered to have notable CPD.

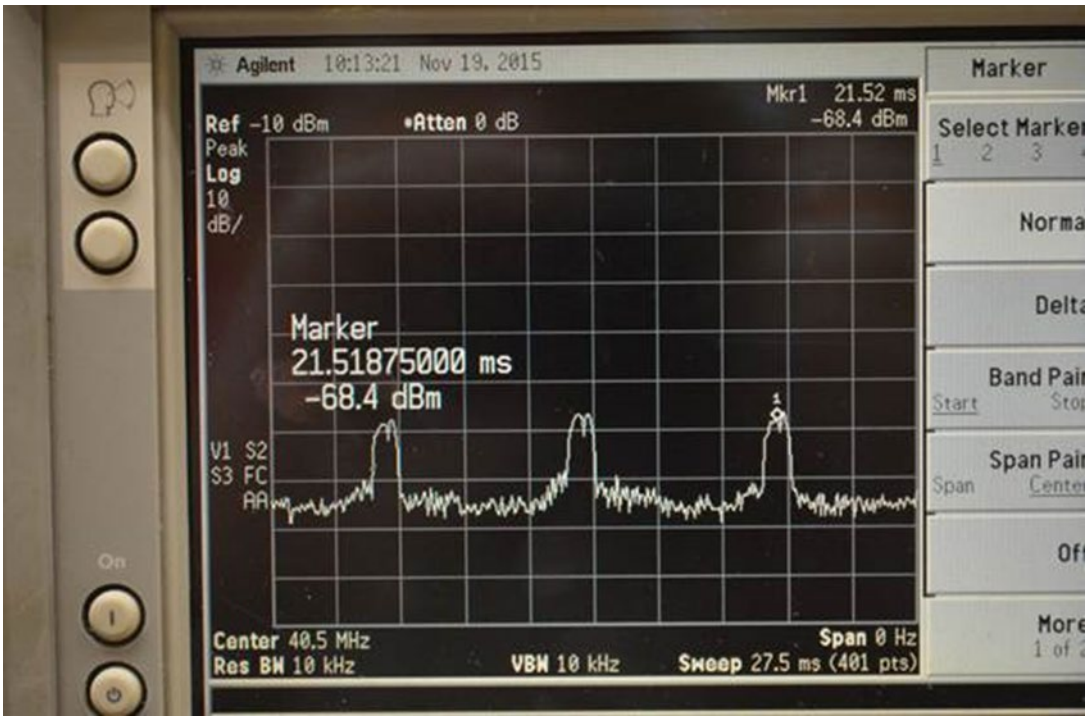

Figure 5 - First example of dynamic CPD at 120 Hz

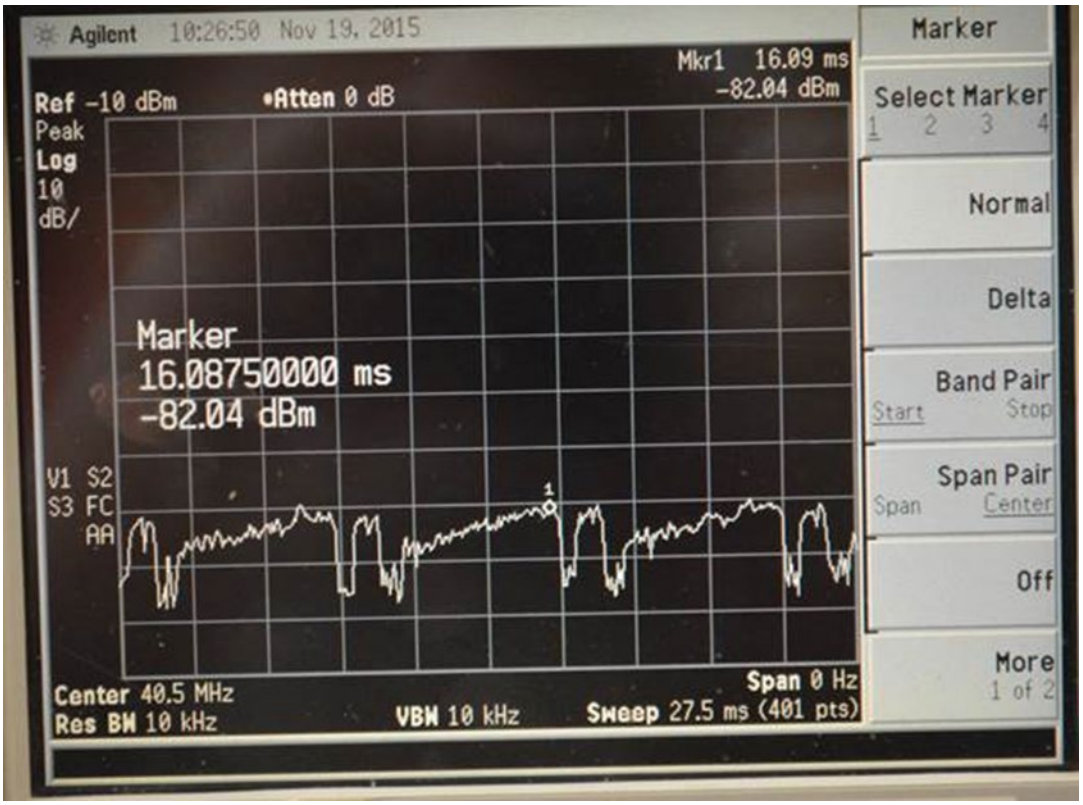

Figure 6 - Second example of dynamic CPD at 120 Hz

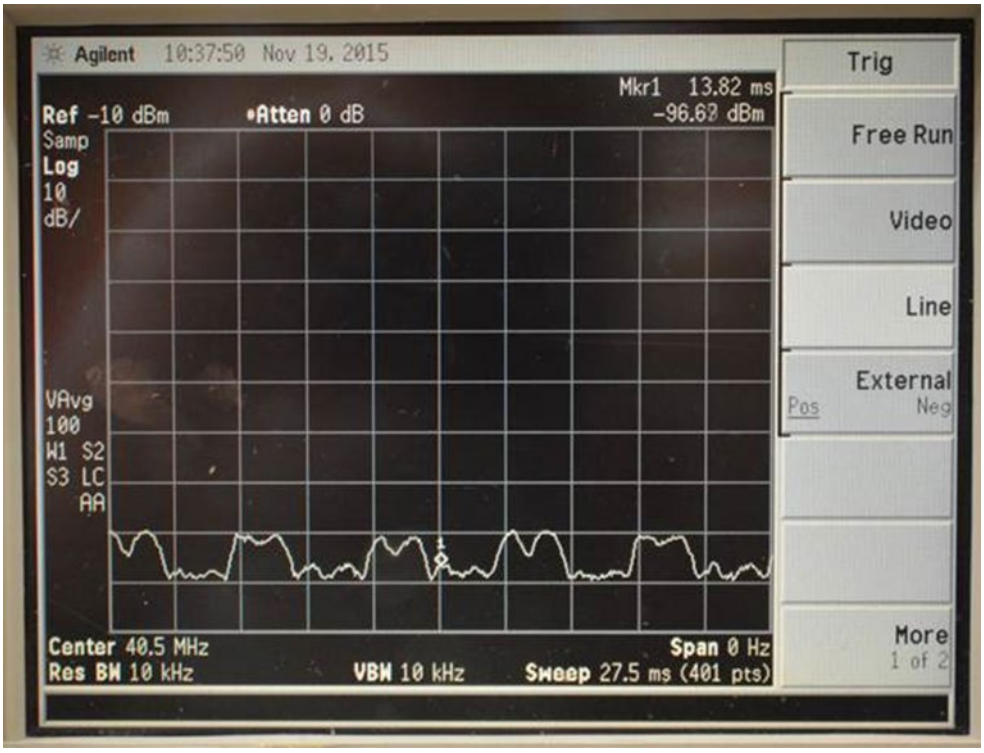

Figure 7 - Third example of dynamic CPD at 120 Hz

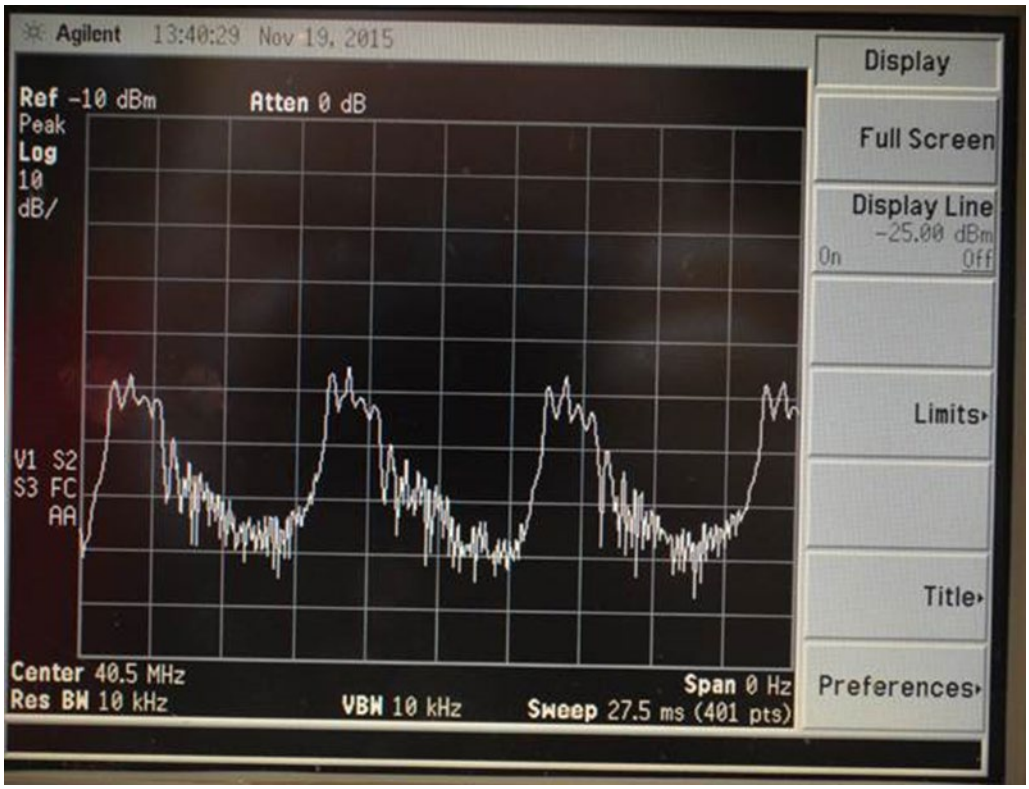

Figure 8 - Fourth example of dynamic CPD at 120 Hz

A downstream digital cable signal was captured on a 12-bit digital LeCroy HDO6104-MS oscilloscope, sampling at 2.5 gigasamples/second and then mathematically distorted to create second and third order distortions. This captured signal is illustrated in Figure 9 in the frequency domain, left pane. The vertical scales are linear voltage, and the horizontal frequency scale is 0 MHz to 1250 MHz. The frequency domain plots were created by taking a fast Fourier transform (FFT) of the time series. The middle pane is the created second order distortion, and the right pane is the created third order distortion. Observe that distortion energy is spread all over the band, not just in the observable 5 MHz to 42 MHz band. Figure 10 is a repeated test with an up-tilt applied to the captured signal before creating the nonlinear distortion. Another observation from the modeling is that the noise in the upstream band is relatively flat over frequency. The nonlinear distortion energy was created by simply squaring or cubing the sample in the time series. An equivalent method would have been to perform a double convolution on the frequency domain signal for second order distortion, and a triple convolution on the frequency domain signal for third order distortion.

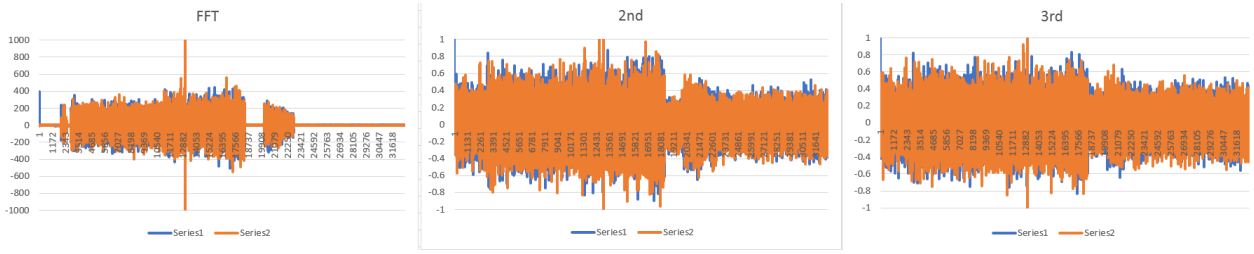

Figure 9 - Left plot is an all-digital downstream signal, linear voltage vs. frequency; middle plot is second order distortion created from the signal illustrated in the left plot; right plot Is third order distortion created from the signal illustrated in the left plot.

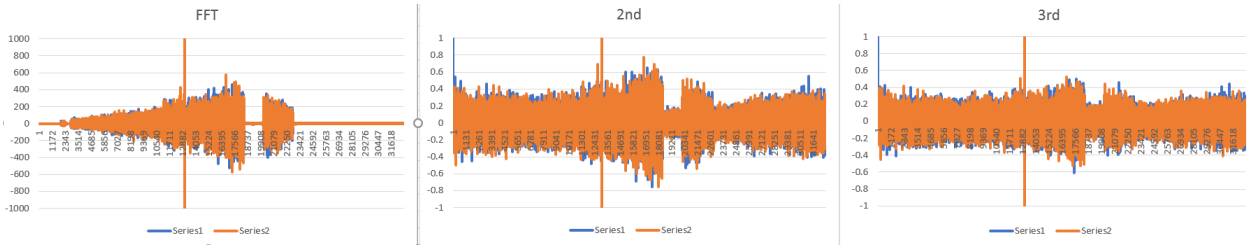

Figure 10 - Left plot is an all-digital downstream signal, linear voltage vs. frequency, but up-tilted; middle plot is second order distortion created from the signal illustrated in the left plot; right plot is third order distortion created from the signal illustrated in the left plot.

2. Noise Sources with 120 Hz Modulation

2.1. Causes for 120 Hz Modulated CPD

Figure 11 is a diagram of a seizure mechanism showing one possible way that 120 Hz modulated CPD can be created. At point “A” a resistive connection is made from a loose seizure screw, and several amperes of AC current are flowing to power the network’s amplifiers. This causes an AC voltage difference between the seizure screw and coax center conductor that is proportional to AC current. Nearby, at points “B” corrosion diodes are present. The corrosion diodes can only actively mix downstream signals while the AC voltage difference is small. Note that this model needs to be vetted in the field with additional testing and troubleshooting. Other hypotheses for the 120 Hz phenomena have been proposed and should also be investigated. The good news is that a distance (range) to a source of CPD can be determined using commercially-available test equipment designed exactly for this purpose, such as the Quiver® from Arcom Digital, so troubleshooting is readily facilitated.

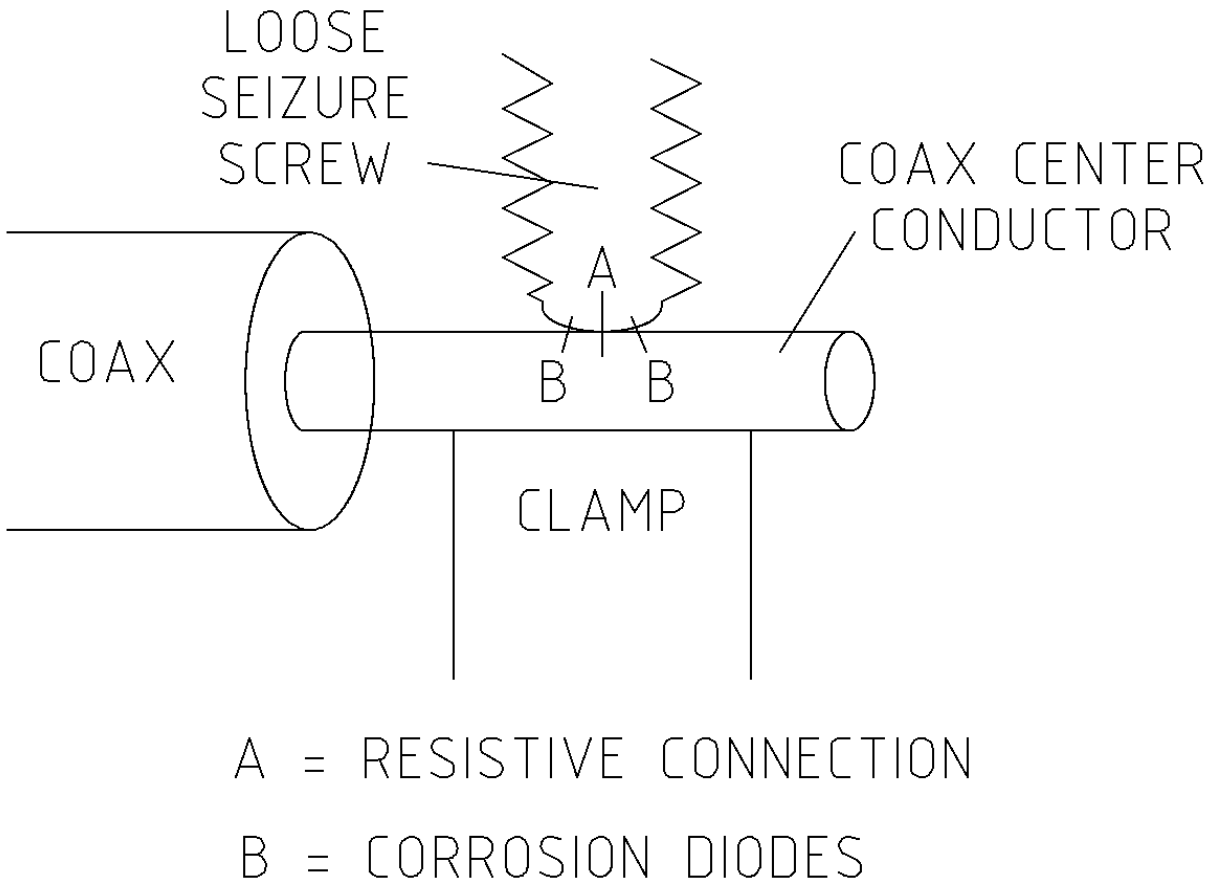

Figure 11 - One possible mechanism for 120 Hz modulated CPD. A loose seizure screw produces a resistive contact at point "A." At point "B" corrosion diodes exist.

While AC voltage is low, diodes are active. While AC voltage is high or low, there is a potential difference between the coax center conductor and seizure screw, and diodes are either forward or back-biased.

2.2 Another Source of 120 Hz Modulated Noise – Switching Regulator Noise

There is a second and different source of upstream noise that is also modulated at 120 Hz, and that is switching regulator noise. Figure 12 is a simplified block diagram of a generic switching AC power supply. A full wave rectifier bridge converts the 120 VAC into 170 VDC, and a switching regulator circuit operates from the 170 VDC to make lower DC output voltages, such as 5 volts and 12 volts. Typical switching frequencies range from 50 kHz to 250 kHz. While the diodes in the full wave bridge are conducting, they have a low resistance, and the current pulses ("I") pass easily to the AC line. But when the diodes are back-biased, the current pulses are blocked because the diodes have high resistance. Not illustrated is an AC line filter, which is generally only partially effective, and is more effective at some frequencies than others. The current pulses occur periodically in the time domain, and so appear as a comb in the frequency domain on a spectrum analyzer. Figure 13 is an oscilloscope trace captured in the field showing voltage vs. time at 10 ms per division. The upstream burst transmissions overload the oscilloscope, but a background noise can be observed to vary at 120 Hz (8.33 ms).

The commercially-available CPD ranging gear does not produce a correlation peak on switching regulator noise. The CPD test equipment is not fooled.

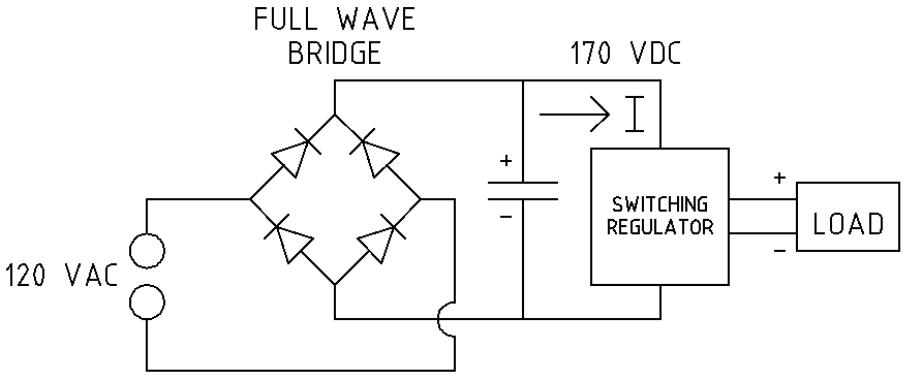

Figure 12 - Current pulses "I" from switching regulator circuit easily pass to AC line when diodes in full wave bridge are conducting, and are blocked when the diodes are back-biased.

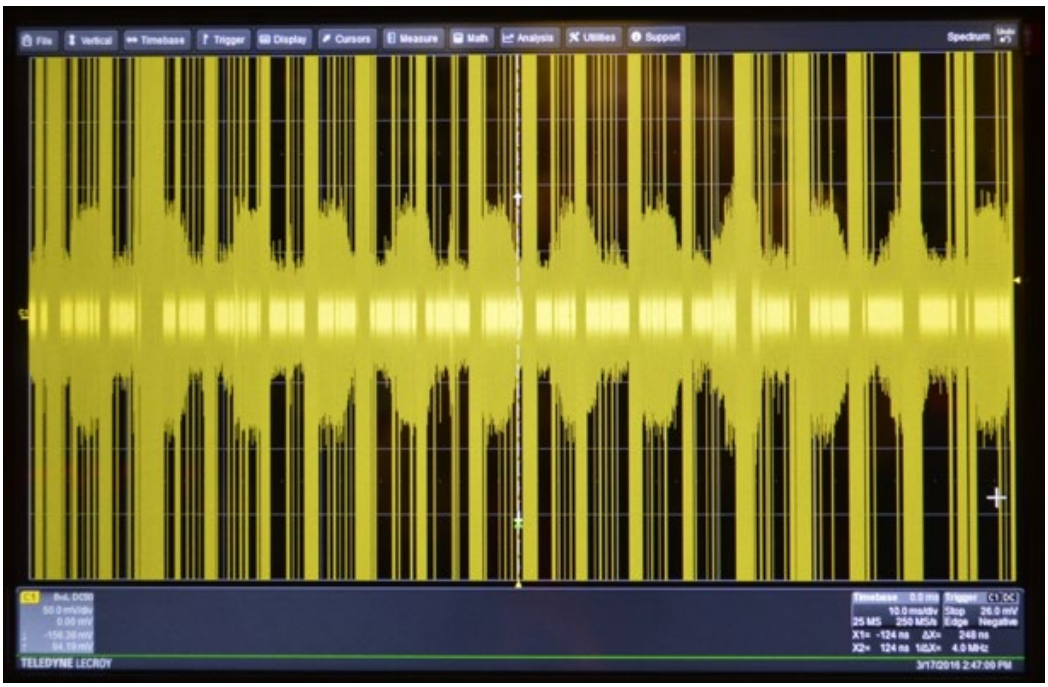

Figure 13 - A digital oscilloscope trace of an upstream DOCSIS™ signal showing background 120 Hz time-varying noise. Upstream burst transmission overloads the oscilloscope.

3. Modifying SCTE-109 Test Procedure to Locate CPD

Instead of using two fixed CW carriers to implement the SCTE-109 test procedure, if one (or both) of the two carriers is made tunable, the frequency of a returning CPD CW beat will also vary. The two downstream CW carriers are mixed at the headend (or node insertion point) to make a headend difference signal. The phase angle of the return CW beat from the cable plant is compared with the phase angle of the headend CW difference signal, and the phase angle can be observed to shift as the CW is tuned. The round-trip time delay is calculated as $\tau = d\varphi/d\omega$, so the round-trip distance to the CPD diode can be calculated from this time delay using the cable's velocity of propagation. Figure 14 is a block diagram of equipment in the headend to make this delay measurement. It can be viewed as a type of homodyne detector. The advantage of this method is that very little downstream carrier frequency change is needed to range a CPD source. In the block diagram of Figure 14, the DC voltmeter makes a sine wave every time the number of round-trip wavelengths decreases (or increases) by one.

For example, if an 800 MHz test signal is used, its wavelength in free space is:

Wavelength of test signal = c/freq. = $3x10^8$ meters per sec /$8x10^8$ Hz = 0.375 m

Assume it took 150 kHz for the DC voltmeter to swing a full cycle, then:

Test freq./Delta freq. = 800 MHz/0.15 MHz = 5333.33 wavelengths round trip

Number of cycles round-trip = 5333.33/0.375 = 2000 meters before correction for velocity of propagation of the cable.

If velocity factor (velocity of propagation expressed in decimal format) of the cable is 0.82, the round-trip distance is:

#wavelengths/vel. factor = 2000/0.82 = 2439 meters.

In other words, the test frequency is 800 MHz, the wavelength is 0.375 meters, which is 5333.33 cycles. To produce a full 360-degree cycle on the DC voltmeter, the wavelength would need to be changed to 5332.33 cycles, which is 0.375070 meters. That is a frequency change to 799.850 MHz, which is a shift of only 150 kHz.

If the CPD diode is very close to the test signal injection point, which ideally is at a fiber node, a piece of coax can be used to limit the delta frequency. If the DC voltmeter makes a non-sinusoidal signal, there may be more than one corrosion diode.

Another test method is to replace the DC voltmeter with an analog-to-digital (A-D) converter and capture voltage samples as the delta frequency is changed. The values can be inserted into an inverse fast Fourier transform (IFFT) to produce a time plot. Note that phase values can be assumed to be zero.

Another implementation of the circuit of Figure 14 is to use a complex demodulator in place of MX 2, which is fed with a sine and a cosine local oscillator to produce in-phase and quadrature voltages. As the delta frequency is swept, the I and Q values are captured and inserted into an IFFT to produce a time plot showing multiple CPD sources. Windowing the data samples improved the time samples by reducing leakage.

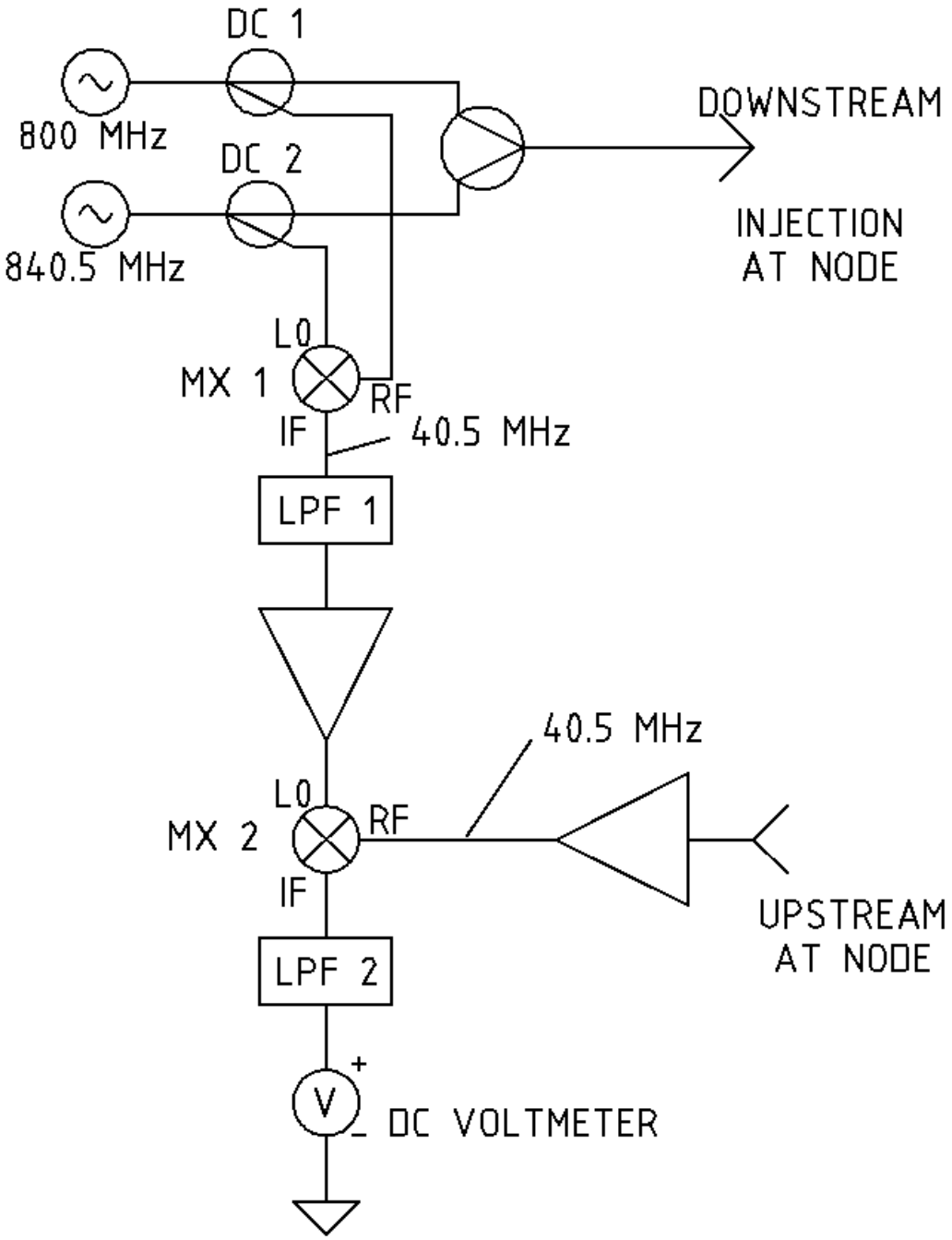

Figure 14 - A block diagram of an improved SCTE-109 test. The distance to a CPD source can be located by phase comparing a field difference (CPD) signal with a headend reference difference signal, using the relationship τ = dφ/dω (time delay is equal to the delta phase divided by delta frequency).

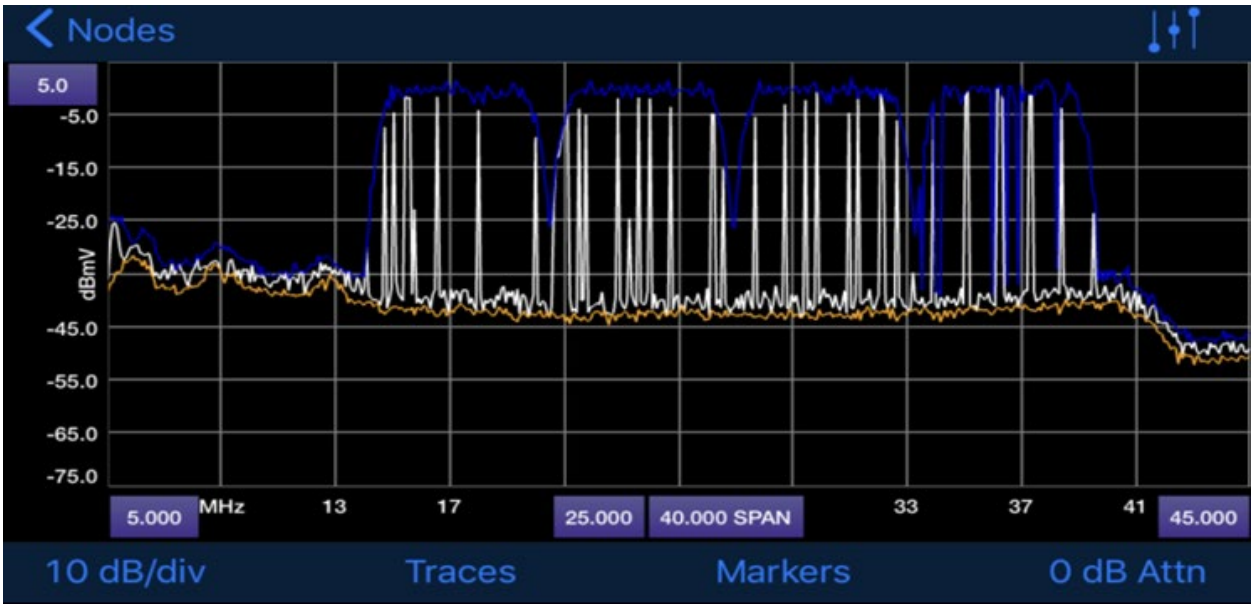

Figure 15 - A spectral plot of CPD affecting a node. Orange line is noise-like CPD

3.1.1. Field Troubleshooting Results

Figure 15 is an upstream spectral plot showing max hold (blue), min hold (orange) and current spectrum (white). This node has a high node score (bad) and was providing poor service to subscribers due to upstream noise. Note particularly the 10 dB rise in the orange trace between the diplexer roll-off (45 MHz) and the active upstream band due to CPD.

The Quiver® ranging tool was used to identify the amplifier location that was the source of the CPD. Figure 16 is a photo showing the test equipment, the amplifier, and the right-angle connector (arrow) with a seizure screw that was found to be ½ turn loose.

Figure 17 is a spectral plot of the same node taken three days after repair. The node's performance changed to a node score of 0.0 (excellent). Note that the "rise" in the noise floor (orange trace) between the diplexer roll-off (45 MHz) and the active band is greatly reduced.

Figure 16 - Location of loose seizure screw at input of amplifier-screw in right angle connector was one-half turn loose

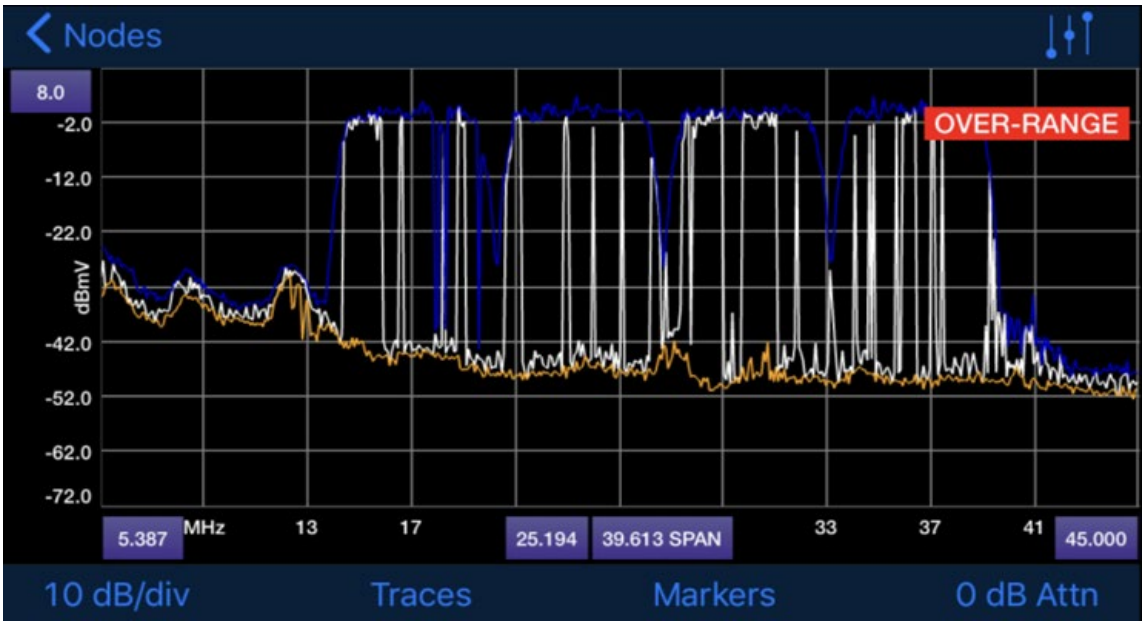

Figure 17 - A spectral plot of the node after CPD, caused by a loose seizure screw, was fixed

3.1.2. Proactive Network Maintenance Implications

One of the indicators that CPD is present is an elevated return noise floor. While the upstream noise floor is expected to vary gradually with temperature, CPD, because of its intermittent nature, does not normally change slowly over time. Figure 18 is a 3-D waterfall plot (frequency is horizontal, magnitude is vertical, and time is orthogonal to the upper right) of CPD made in the lab, as a corrosion (Schottky or hot carrier) diode is attached and removed. This plot can be created on an active node by measuring the background noise when no transmissions are present. Another PNM indicator of CPD is comparable errors on all upstream channels. Typically ingress from burst noise causes more errors on the lower-frequency carriers.

Figure 19 is a block diagram of an easy method to make CPD for lab testing, using a diplexer, a downstream signal source, and a 1N5711 Schottky diode.

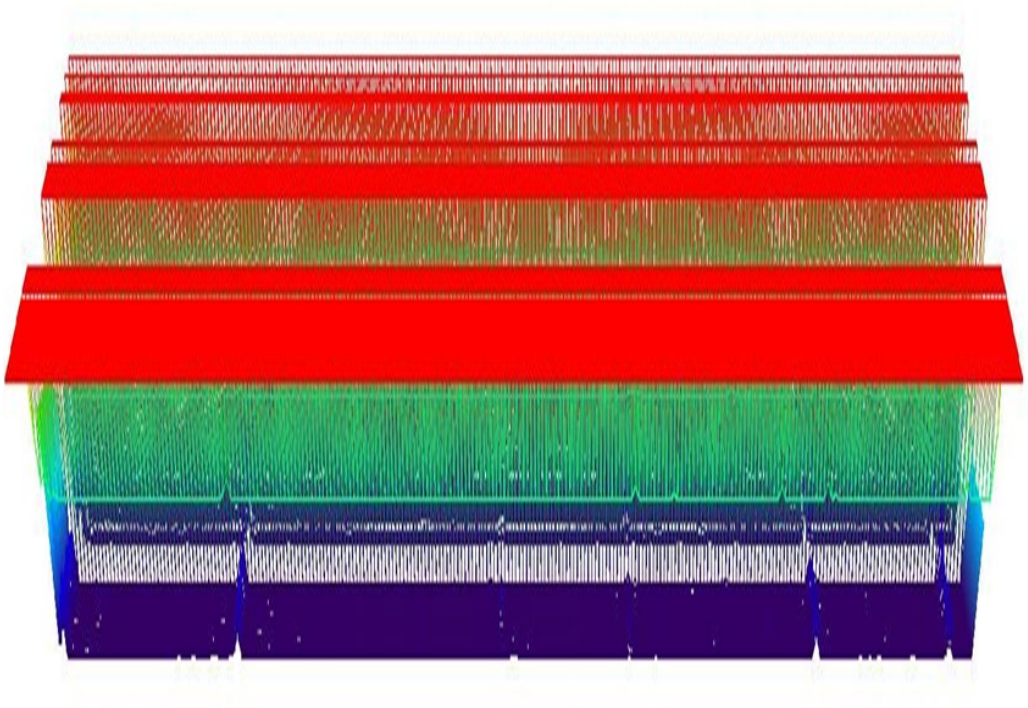

Figure 18 - A waterfall plot showing lab-created spectrally-flat CPD varying with time

3.1.3. Making CPD in a Lab

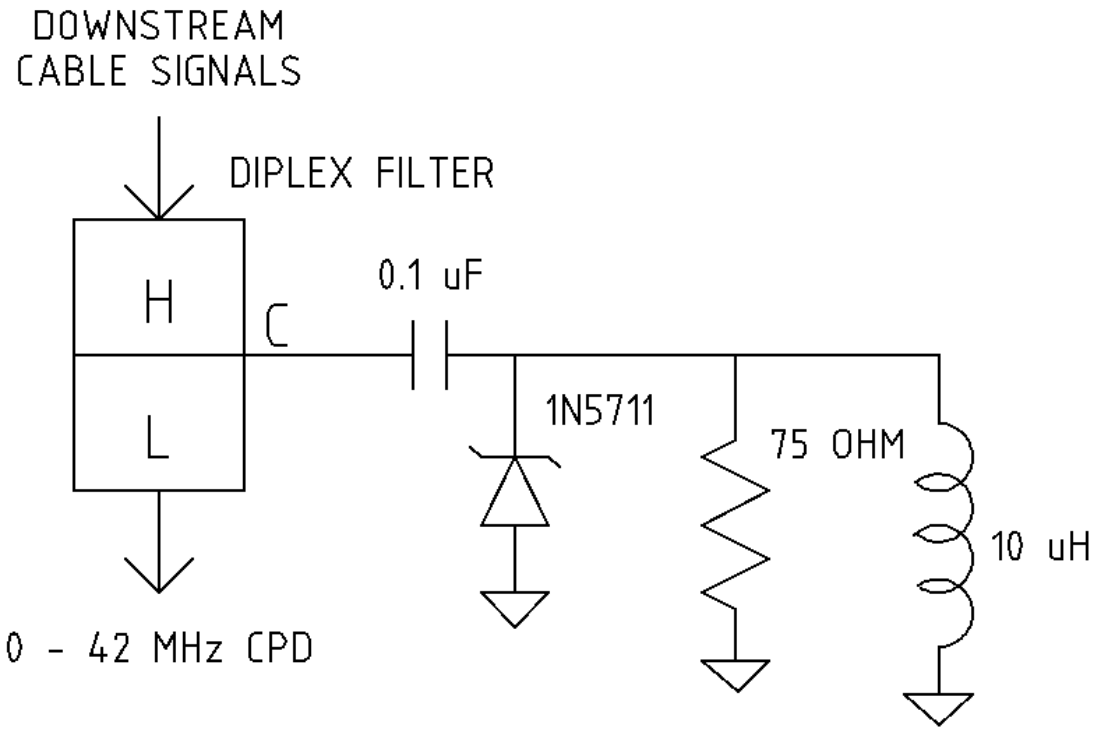

Figure 19 - A block diagram illustrating how to make CPD in the lab for testing

4. Conclusions

Upstream noise and interference are operational issues for cable systems, and while most packet errors are caused by ingress, some significant percentage is caused by CPD. With analog TV signals, CPD used to appear as three beat clusters every 6 MHz. But now, with mostly- or all-digital downstream carriers, CPD appears as an elevated upstream noise floor. It has not gone away. Some percentage of the CPD sources are steady, but a significant percentage are 120 Hz modulated at twice the frequency of the AC power line. Switching regulators also produce a different impulsive energy that is 120 Hz modulated. CPD-producing corrosion diodes are extremely delicate, and the CPD is not steady because of mechanical stress and temperature variation. CPD can readily be located with commercially-available test gear, or by the modified SCTE-109 method described in this paper. CPD reduction is an area for operational improvement and significant cost savings.

CableLabs Time Domain Reflectometer

A Technical Paper prepared for SCTE•ISBE by

Tom Williams, Distinguished Technologist /CableLabs, SCTE•ISBE Member
858 Coal Creek Circle
Louisville, CO 80027
t.williams@cablelabs.com
303-661-3486

Jason Rupe, Principle Architect/CableLabs, SCTE•ISBE Member
858 Coal Creek Circle
Louisville, CO 80027
j.rupe@cablelabs.com
303-661-3332

Table of Contents

List of Figures

1. Introduction

If a spectral plot of a wide block of digital carriers is captured, it can be analyzed for the presence of standing waves, revealing echoes caused by cable or other equipment damage in the signal path. Because of the exceedingly wide bandwidth of digital signals commonly available on downstream cable plant, the time delay accuracy (used for distance) from this measurement method is very high.

Damage to coaxial lines or defective components, such as surge-damaged or water-filled taps, affect network performance. One common characteristic of damaged lines is the creation of reflections, or echoes. Another characteristic is increased attenuation over the original design. Generally, adaptive equalizers can compensate for the echoes if they are static or slowly-changing. However, adaptive equalizers cannot compensate for excessive attenuation causing weak, noisy signals. Random noise added to a signal is unrepairable. Components, such as taps, passives, splitters, and power inserters, generally have a return loss specification that should be higher than about 16 dB at any specified frequency. Amplifiers' return loss should be higher than about 18 dB at any frequency in the operating passband. Coaxial cable attenuation is published in manufacturer data sheets and increases with both frequency and line length.

The procedure described in this paper covers two methods to determine if there are reflections in metallic cable lines for plant that is in-service and one method for plant that is out-of-service. These are done by capturing standing waves on digital carriers, which may be DOCSIS® 3.0 SC-QAM or DOCSIS 3.1 OFDM downstream carriers, or digital video carriers. In the first method, a highly-averaged wide-bandwidth spectral plot with ripples on digital carriers is processed with an inverse fast Fourier transform (IFFT) to create an impulse response. In the second method, an additional step is taken to capture an unimpaired calibration signal, resulting in a higher quality impulse response.

The procedure also covers a third out-of-service method to determine if a section of plant has too much attenuation by intentionally open-circuiting (or short-circuiting) the cable and measuring the magnitude of an expected standing wave. For example, this test could be performed on a buried subscriber drop to find water damaged cable or a pirate tap (theft-of-service). If the standing wave's peak-to-valley response is too small, the drop cable has excessive attenuation. If a reflection is discovered closer than the induced open (or short) circuit, the plant may have a pirate tap, such as a buried splitter, or other damage.

2. Standing Wave Analysis

Figure 1 is a block diagram of a standing wave test that can be performed in the field by a line technician. A high impedance probe is connected at a test point and used to measure signals traveling downstream combined with reflections of those downstream signals heading upstream. The test is not service interrupting. The bidirectional probe test point could, for instance, be a seizure screw on a tap. The probe could be a Viavi/Trilithic I-Stop® device which provides high impedance signal probing and AC power isolation. This type of standing wave from a single echo can only be observed while testing in the field because the reflected echo is absorbed by its launch amplifier and does not show up either in the home (for downstream) or the headend (for upstream). Furthermore, the standing wave cannot normally be observed on a tap port because tap ports are fed from directional couplers. However, the standing wave can be observed on the tap's seizure screws under the 5/8-24 port caps. The advantage of this type of test, besides not being service interrupting, is that the location of the impedance mismatch may be determined by echo delay time relative to the location where the probe is inserted. One use case for this test method is finding damage to hardline coaxial cable without disconnecting service. Another use case is determining

if a drop cable, particularly a buried drop cable, has been tapped into by a pirate tap, such as a splitter or directional coupler.

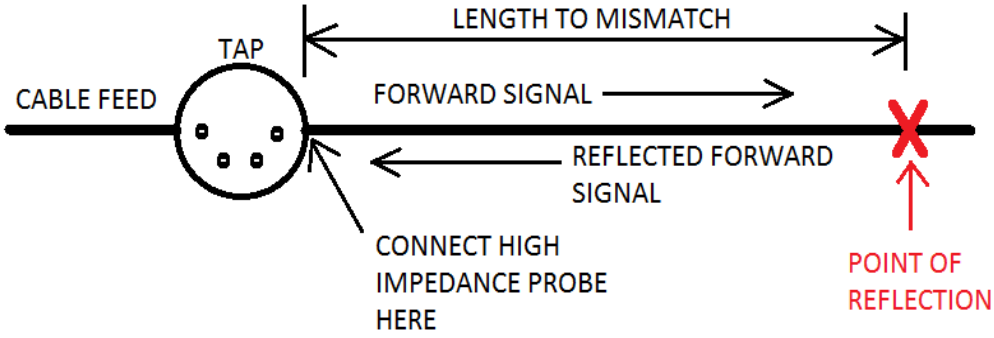

Figure 1 - In this block diagram, a high impedance probe is connected to bi-directional test point, and a broad spectrum of digital signals is captured with heavy averaging to produce a smooth spectral trace.

Figure 2 is a block diagram of an echo tunnel created by a pair of impedance mismatches (reflection points). The echoes bounce back and forth between two impedance mismatches, creating standing waves. This echo tunnel can be observed at an endpoint, either in a home or the headend. However, while the length of the echo tunnel can be determined accurately, the actual start and stop locations of the tunnel are unknown.

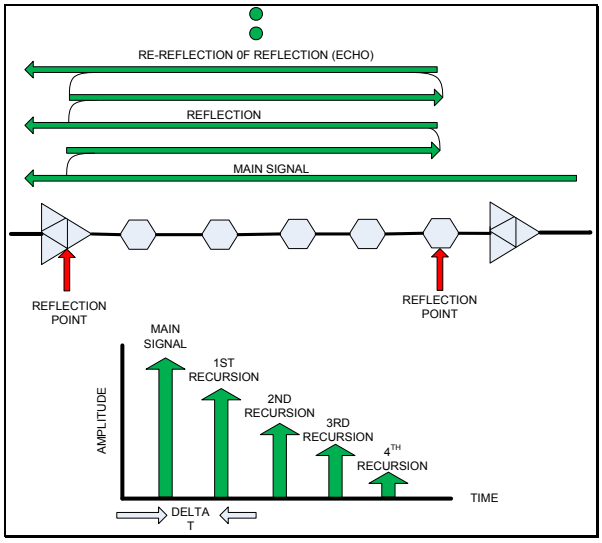

Figure 2 - An upstream signal impairment created by an echo tunnel caused by two impedance mismatches. This upstream signal with ripples may be captured in the headend; an impaired downstream signal can be captured in the home.

In both cases, the wider the test bandwidth, the greater the accuracy in measuring the echo's delay.

Figure 3 is a spectral plot of many digital carriers, showing a standing wave produced by a reflection. It was captured in the lab using the wiring diagram of Figure 1. The random data samples are heavily averaged to produce a smooth magnitude plot.

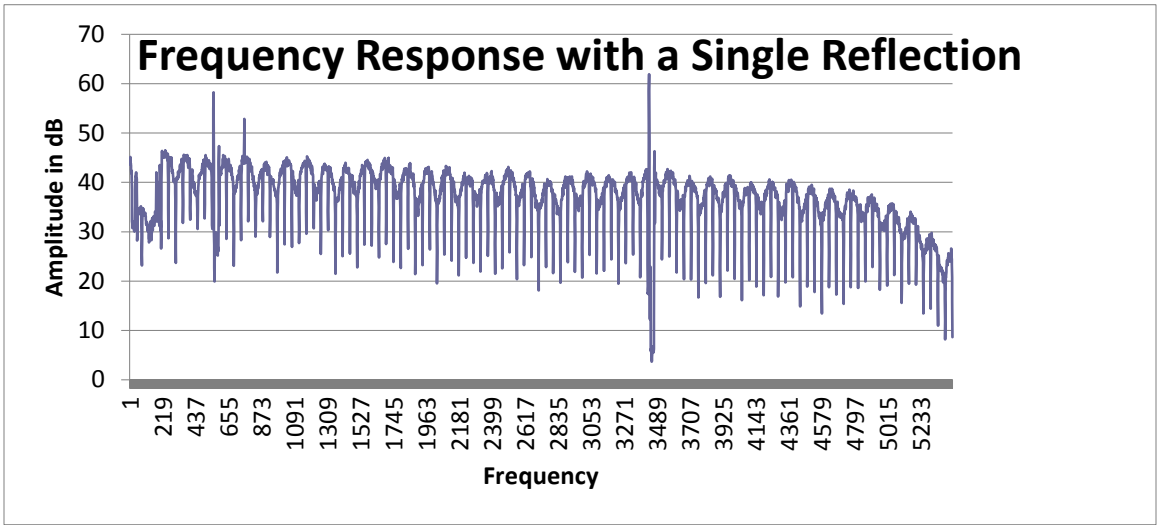

Figure 3 - A 108 MHz to 750 MHz spectral plot showing a standing wave riding on 6 MHz digital carriers.

Figure 4 is a time plot of the inverse Fourier transform of the captured data in Figure 3. The transform was produced by converting the dB frequency samples to linear voltage and using these samples for real frequency values. Zeroes were used (zero-stuffed) for all imaginary frequency values. Half of the resulting inverse transformed time domain samples are mirror images and are discarded. The standing wave is highlighted with a red arrow, and the harmonics (H1, H2, H3, etc.) were created by the dropout on the 6 MHz channel edges, so they occur every 167 ns. [$1/(6x10^6)$]. Harmonics should be ignored.

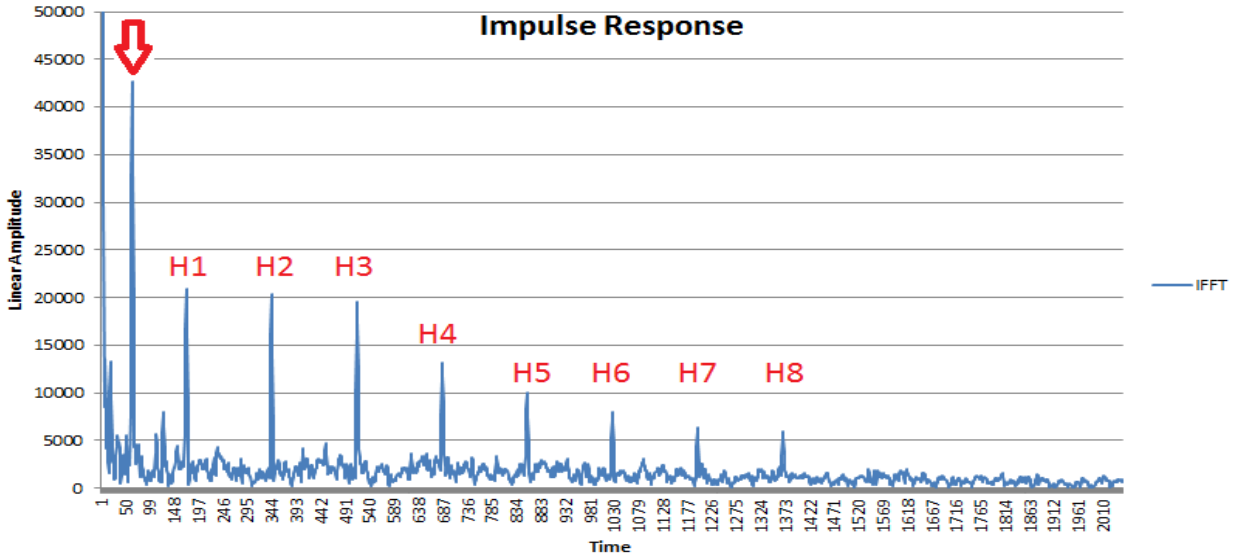

Figure 4 - A temporal plot showing the standing wave (red arrow) and the harmonics created by the 6 MHz channel plan.

2.1. Procedure

Processing Steps Method 1 (no reference plot).

1. Capture a spectral plot of a wide block of digital carriers using a high impedance probe inserted into the network at a point with standing waves. Use large averaging (e.g., >100 frequency domain samples at each frequency) to produce a smooth magnitude plot. There should be more than two frequency samples per frequency domain ripple to meet Nyquist's criteria.
2. Examine the spectrum for frequencies where the spectrum has a large voltage standing wave ratio (VSWR).
3. Using the widest bandwidth with a standing wave, insert the frequency domain samples into an IFFT as real values, and with zeroes for imaginary values. 2 raised to an integer exponent is a convenient length for fast Fourier transforms.
4. Perform the transform and discard the mirror image in the time domain.
5. Ignore the harmonics created by the haystack shape with 6 MHz dropouts (H1, H2, H3, etc.).
6. Measure the time domain impulse(s) created by reflection(s).
7. Estimate the round-trip distance from the time coefficients, factoring in the cable's velocity of propagation.
8. The magnitude of the reflection can be determined as the ratio of the forward traveling wave to the reflected wave. The accuracy of the reflection's magnitude can be improved by factoring in the round-trip transmission line loss.

Processing Steps Method 2 (using a reference plot).

1. Capture a reference spectral plot of digital carriers at a node launch point (or other unimpaired location) and store it. Use large averaging (e.g., >100 frequency domain samples) at each frequency. The signal at the launch point should contain a forward-only, reflection-free signal, so

an amplifier test point or a fiber node launch point are good locations. There should be more than two frequency samples per frequency domain ripple to meet Nyquist's criteria.

2. Capture a matching test plot at a point in the network with standing waves with the same number of frequency samples as the reference plot. Use a high impedance probe inserted upstream from the suspected cable damage.
3. Reduce the spectrum for both plots to frequencies where the spectrum has a large VSWR. Using a wider frequency band for analysis results in better time accuracy, as time accuracy is proportional to the reciprocal of the analysis bandwidth.
4. Subtract the dB values at each sample frequency of the reference plot from the test plot.
5. Convert the difference values from log to linear.
6. Insert the frequency domain samples into an IFFT as real values and use zeroes for imaginary values.
7. Perform the IFFT and discard the mirror image in the time domain.
8. Harmonics created by the haystack shape with 6 MHz dropouts are automatically eliminated by taking a reference plot. Likewise, noise is reduced.
9. Measure the time domain impulse(s) created by reflection(s).
10. Estimate the round-trip distance from the time coefficients, factoring in the cable's velocity of propagation.

Processing Steps Method 3 (testing an out-of-service span of drop or hardline cable using standing waves.)

1. Use either Method 1 or Method 2, with the end of the cable open (or short circuited) to purposefully create a standing wave. This method is particularly useful for cable which has more attenuation than expected, possibly due to water absorption or dog chews.
2. Capture a spectral plot. See Figure 5, which is linear plot of a standing wave. Note that the standing wave envelope decreases with frequency due to round-trip cable loss increasing with frequency. The larger the ratio of the peak voltage minus the minimum voltage to the average voltage, the lower the cable loss. If the cable attenuation vs. frequency measurement is nominal, the trace will reach the computed standing wave limits (red and green traces in Figure 5). The standing wave limits are computed using the type of cable (e.g., Series 6) and an estimated distance to the open circuit. If the standing wave is too small, the cable has too much loss, which may indicate water or other damage. A plot of cable loss vs. frequency can be made along with the calculated expected cable loss (not illustrated).
3. Insert averaged voltage samples into an IFFT and perform a transform.
4. The time domain impulse result after an IFFT (not illustrated) should show the distance to the open connection. If another earlier impulse is detected, there may be cable damage or a pirate tap. After getting a more accurate distance measurement to the open circuit, the measurement may be used to modify the red and green traces illustrated in Figure 5.

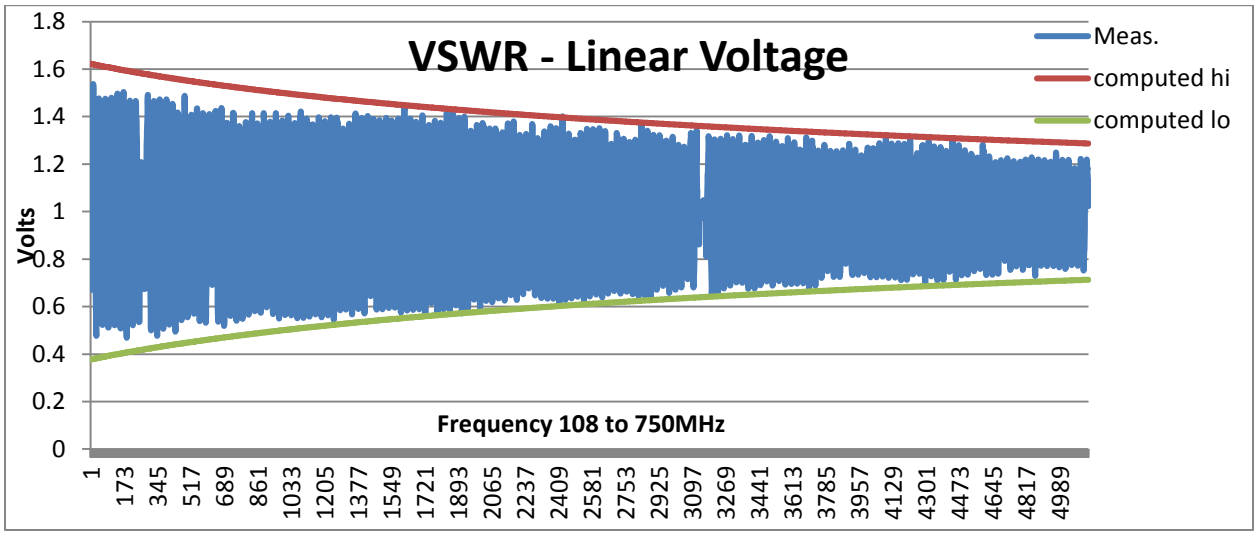

Figure 5 - A spectral plot of a standing wave from an intentionally unterminated cable.

Suggestions, alternative approaches, and enhancements.

1. If no digital carriers are available, a spectral plot can be created by other means, such as a stepped continuous wave carrier, a broadband noise source, a chirp signal or even a comb signal.
2. Spectral plots can be captured from full band capture (FBC) capable cable modem chips. Spectral plots can also be captured from software defined radios (SDRs) such as the Ettus B200, and even from analog swept scalar spectrum analyzers.
3. If the spectral plots are captured in a home, the computed distances will be the length of echo tunnels, whose locations are unknown. Nevertheless, having an accurate distance measurement and a strand map with distances can assist with troubleshooting.
4. Using interpolation between digital carriers, the harmonics (H1, H2, H3, etc.) can be reduced or eliminated. Another better approach is to capture a reference trace with no reflection and use that in the processing, as described in Method 2. The reference trace can be captured, for instance, at a directional coupler test point at an amplifier's output or a test point from a fiber node. In processing, the linear magnitude of the test signal is divided by the linear magnitude of the reference signal at each frequency point. Alternately, the dB values can be simply subtracted prior to the IFFT to achieve a same effect as division.
5. Parabolic interpolation can be used to increase time (distance) accuracy, for example, when energy is dispersed across multiple time domain samples. A better signal processing method is to zero-stuff the spectral data with zeroes; and when the data is inverse Fourier transformed, it will be smoothed, allowing time peaks to be discerned.
6. If the standing wave can be extended down to DC, using mathematical extrapolation, the standing wave may be determined to have been created by an open or a short circuit. That is, if the ripples of Figure 3 were created by a short, replacing the short with an open will invert the wave.
7. Empirically it has been found that improved results occur when a mathematical tilt is applied to a tilted signal to render it flat before the inverse Fourier transform. This tilt removes the difference "step" between the highest and lowest frequencies.
8. If an echo is extremely long, the averaged frequency domain "ripples" may appear like random noise. However, a plot of the inverse Fourier transform coefficients will clearly reveal the reflection in the time domain.

9. Testing in the upstream 5 MHz to 42 MHz band requires a signal to be supplied, as there are no continuous digital carriers in a DOCSIS upstream system. A swept signal can be inserted into a tap port and a standing wave may be observed on the seizure screw of the tap. Note that the test signal does not need to be limited to just 42 MHz (assuming no return equalizer in the tap) and return band testing is desirable for locating very long echoes because the cable loss is lower.
10. Using a frequency domain sample size of 2N samples facilitates the use of an IFFT. Likewise, zero padding or windowing may be used.

2.2. Theory (why it works and limitations)

A natural question that arises is how do you take an inverse Fourier transform when no phase information is available? That is, a Fourier transform needs complex values in and produces complex values out. The procedures in this paper have only magnitude data available.

In control theory and signal processing, a linear, time-invariant system is said to be minimum-phase if the system and its inverse are causal and stable. If certain conditions are met, most cable networks can be assumed to be minimum phase networks. In a minimum phase network, phase values can be computed from magnitude values.

Figure 6 is a phasor diagram (I-Q plot) showing a static outgoing signal with a magnitude of 1.0 and a delayed returning rotating phasor with a magnitude of "a" which is much less than 1.0. This "a" phasor adds in and out of phase as the frequency changes. You need to determine a relative magnitude of "a" which reveals echo strength, and its rate of rotation with frequency, which reveals echo delay time.

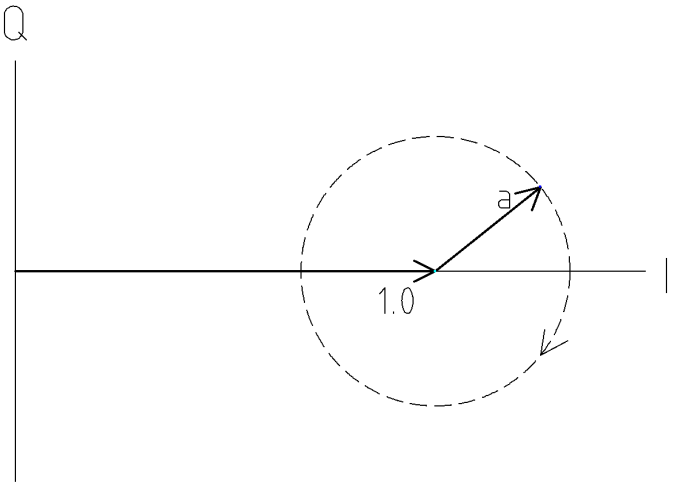

Figure 6 - A phasor diagram showing spectral components adding in and out of phase. The magnitude of the echo relative to the main signal is given by "a". If the delta phase rotation of phasor "a" is divided by a delta frequency shift "ω", the round-trip delay is computed as: τ= dφ/dω.

The information can be computed from a smoothed magnitude plot, as is contained in Figure 7B, which is an observed standing wave that varies with frequency. If you assume that this sine wave signal on a pedestal is a vector sum that is real-only (Q is always zero), the I-Q phasor plot of Figure 7A is the result. It comprises one static phasor (pedestal) with a value of 1.0, one phasor with a value of a/2 going clockwise, and another phasor with a value of a/2 going counterclockwise. Note that the Q components

cancel each other and the I components add. With digital signal processing one of the sideband phasors can be eliminated. Note that for reporting the magnitude of "a" the length of the rotating phasor a/2 must be doubled. This requirement is because half of the transformed time domain samples were discarded.

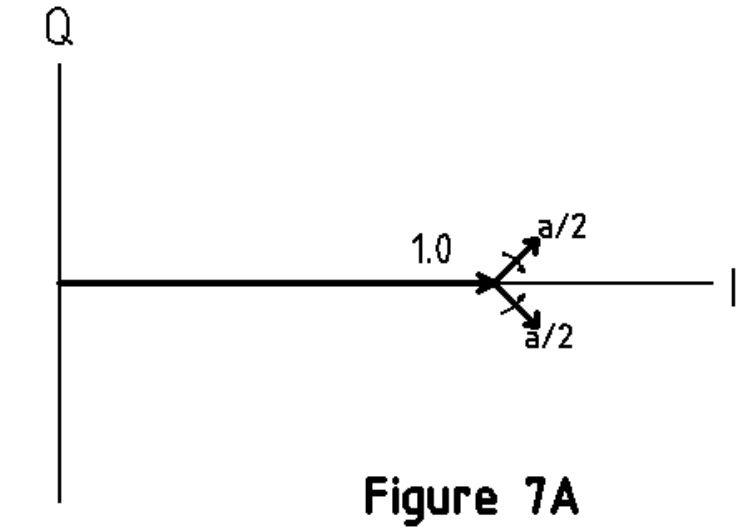

Figure 7A

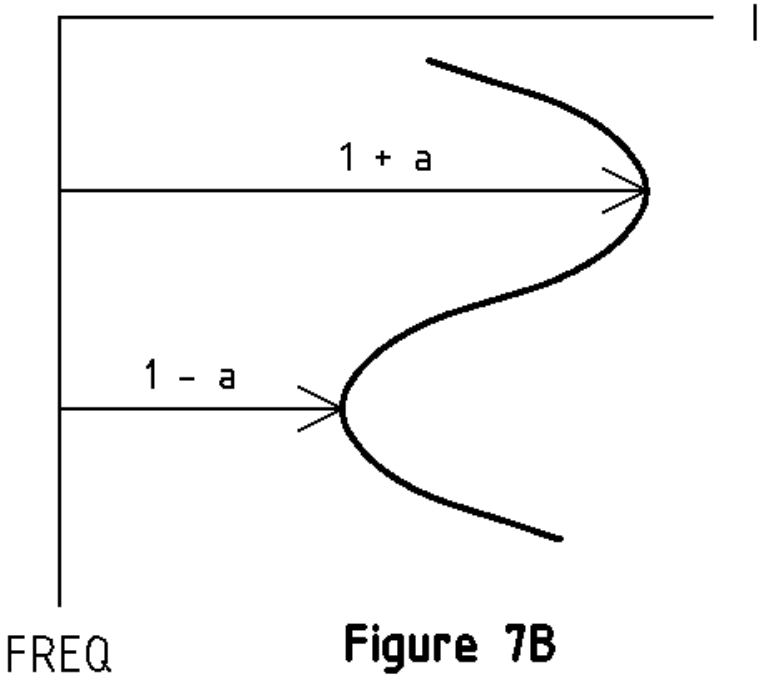

Figure 7B

Figure 7 - Figure 7A is an I-Q phasor diagram that would create a real-only (in-phase) signal that varies with frequency, which is illustrated in Figure 7B.

A vector sum in Figure 7A comprises three component phasors, one static with a value of 1.0, one with a value of a/2 rotating clockwise, and one with a value of a/2 rotating counterclockwise. The two rotating

phasors cancel imaginary values and add real values. Upper and lower sideband phasors counter-rotate. Figure 7B is a plot of in-phase voltage vs. frequency. The magnitude plot will be approximately equal to the real value, provided "a" << 1.0.

2.3. Numerical Example

Figure 8 is plotted from a data file with dB values from a FBC. In this file there were 8704 samples with a frequency separation of 118 kHz between samples. Figure 9 is a subset of 4096 selected frequency samples, chosen for a maximum ripple. This span is 483.328 MHz (4096 * 118 kHz). 1/483.328 MHz yields a time delay of 2.06898 ns for each symbol. This can be verified by noting that there are 80 time domain samples between subcarriers in Figure 9, which occur at 167 ns, the reciprocal of 6 MHz. If zero padding were desired, you could add 2048 (or more) frequency domain zeroes to the left and 2048 frequency domain zeroes to the right of Figure 8, and then perform an 8192 point inverse Fourier transform. Figure 10 is a time plot showing the duration (length) of an echo tunnel.

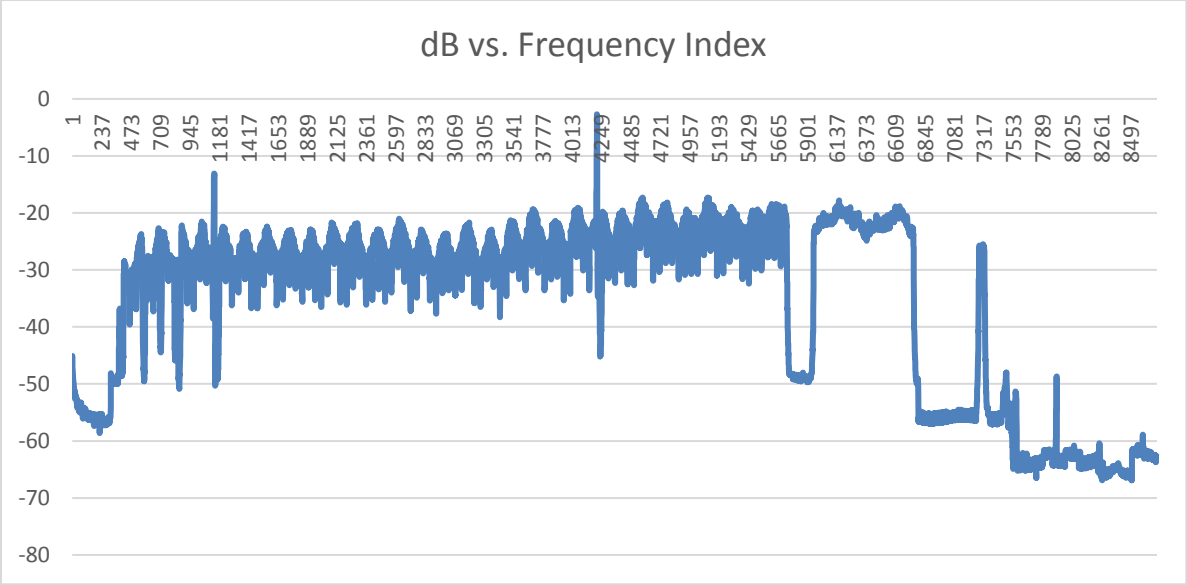

Figure 8 - Data captured by full band capture with a standing wave. 8704 points are available. Frequency samples are separated by 118 kHz.

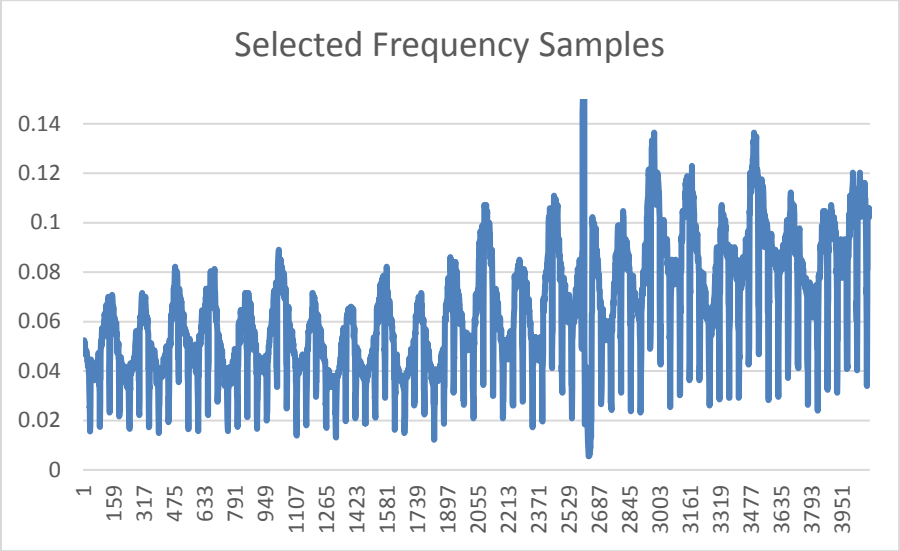

Figure 9 - 4096 frequency points selected for maximum ripple. Span is 483.328 MHz.

As a check, if the frequency sample separation is 118 kHz, the longest echo that could be detected, using a variant of the Nyquist sampling theorem (that requires every periodic sine wave component to be sampled >2X) is 1/236 kHz, or 4.23 µs. A longer echo than 4.23 µs would be aliased.

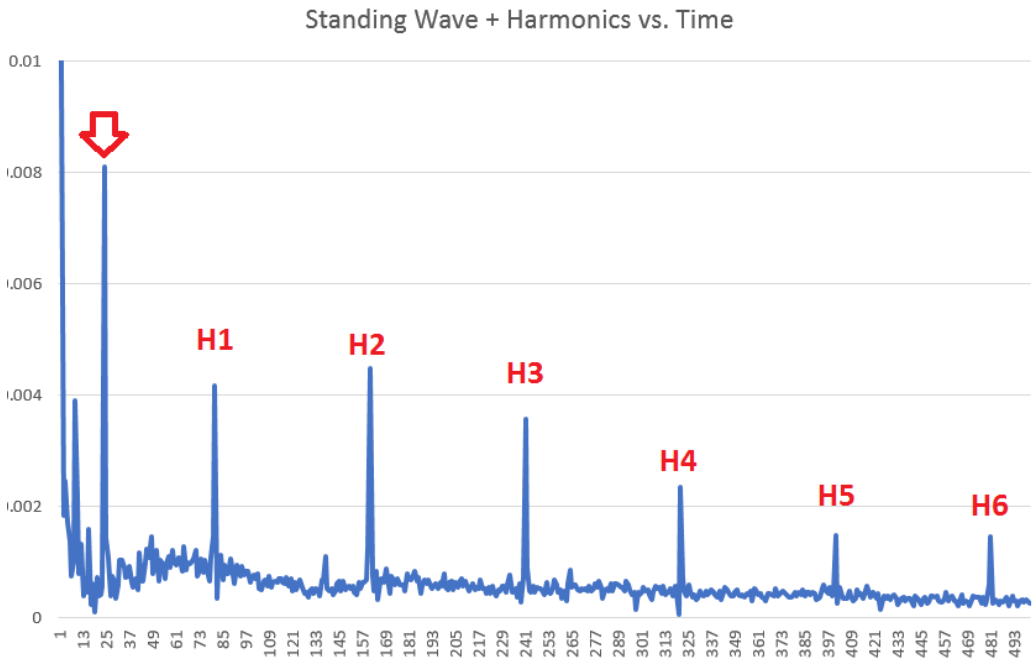

Figure 10 - Time domain plot showing standing wave plus harmonics.

3. Conclusions

Using a high impedance probe on a cable line, a highly-averaged spectral block of digital carriers may be captured and analyzed using an IFFT of the standing waves. The resulting impulse response reveals the reflection's delay and strength. This method may be improved using an unimpaired reference plot for processing. If an intentional short or open circuit is applied to a cable under test, the loss of the cable may be determined from a ripple magnitude created by a resulting standing wave.

Appendix A. Code Used for Digital Signal Processing

The intended audience for this code is software developers.

```
//Copyright (c) 2018, Cable Television Laboratories Inc.
//All rights reserved.
//
//Redistribution and use in source and binary forms, with or without
//modification, are permitted provided that the following conditions are met:
//
// 1. Redistributions of source code must retain the above copyright notice, this
//  list of conditions and the following disclaimer.
// 2. Redistributions in binary form must reproduce the above copyright notice,
//  this list of conditions and the following disclaimer in the documentation
//  and/or other materials provided with the distribution.
//
// THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND
// ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED
// WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
// DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR
// ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES
// (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES;
// LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND
// ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT
// (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS
// SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
//
// The views and conclusions contained in the software and documentation are those
// of the authors and should not be interpreted as representing official policies,
// either expressed or implied, of the FreeBSD Project.

//TDR1.cpp : main project file.
//#include "stdafx.h" //uncomment this line with Microsoft Visual Studio
//some C libraries to use
#include <math.h>
#include <stdio.h>
#include <stdlib.h>
#include <malloc.h>
#include <stdlib.h>
#include <string.h>

#define nrpts 2048 //IFFT size

#define hnrpts 1024 //half number of points

typedef struct { float real, imag; } COMPLEX;//COMPLEX holds FFT/IFFT data
extern void ifourier(COMPLEX *, int);//inverse fast Foruier transform
extern void black(COMPLEX *, int);//Blackman window

int main()
```



```
{
	int i, cnt;
	float t1[nrpts], v1[nrpts], real[nrpts], imag[nrpts], mag;
	FILE *inputdata, *outputdata;
	COMPLEX *x;
	int m = 11; //2^11 = size of FFT
	x = (COMPLEX*)calloc(nrpts, sizeof(COMPLEX));
	if (!x) { printf("\n Unable to allocate input memory.\n"); printf("\x7"); exit(1);
};

	if ((inputdata = fopen("lab1.txt", "r")) == NULL) {//get a file with magnitude
data
		printf("could not open file.\n");
		system("pause");
		exit(0);
	}

	if ((outputdata = fopen("output.txt", "w")) == NULL) {//to make plots
		printf("could not open file.\n");
		system("pause");
		exit(0);
	}

	for (cnt = 0; cnt<nrpts; cnt++) {
		fscanf(inputdata, "%f ", &t1[cnt]); //scan in the coefficients
	}

	for (cnt = 0; cnt<nrpts; cnt++) {//optional: convert dB values to linear if
necessary
		v1[cnt] = pow(10, t1[cnt] / 20);

		printf("%d\t %f\t %f\n", cnt, t1[cnt], v1[cnt]);
	}

	for (i = 0; i<hnrpts; i++) {//load upper sideband
		x[i].real = v1[i + hnrpts];
		x[i].imag = 0;//this is the key step, as you don't have complex data, only
magniude frequency data
		real[i] = x[i].real;//save freq values
		imag[i] = x[i].imag;
	}

	for (i = hnrpts; i<nrpts; i++) {//load lower sideband
		x[i].real = v1[i - hnrpts];
		x[i].imag = 0;
		real[i] = x[i].real;//save values
		imag[i] = x[i].imag;
	}
	black(x, m);//optional step to reduce leakage in IFFT
	//in this implemantation of Fourier transform, the first frequency sample is
center of channel.
	ifourier(x, m);//convert into time domain
	//because 0's were used for imaginary values, half of time samples are mirror
images
	for (i = 0; i<nrpts; i++) {
```



```
            mag = sqrt(x[i].real*x[i].real + x[i].imag*x[i].imag);
            fprintf(outputdata, "%d \t%f \t%f \t%f \t%f \t%f\n", i, real[i], imag[i],
x[i].real, x[i].imag, mag);
      }
      printf("Finished");
      fclose(inputdata);
      fclose(outputdata);

      return 0;
}// END OF MAIN

void ifourier(COMPLEX *x, int m)//takes an inverse fourier transform of size 2 to the m
power
{
      static COMPLEX *w;
      static int mstore = 0;
      static int n = 1;
      COMPLEX u, temp, tm;
      COMPLEX *xi, *xip, *xj, *wptr;
      int i, j, k, l, le, windex;
      double arg, w_real, w_imag, wrecur_real, wrecur_imag, wtemp_real;
      float scale;
      if (m != mstore) {
            if (mstore != 0) free(w);
            mstore = m;
            if (m == 0) exit(1);
            n = 1 << m;
            le = n / 2;
            w = (COMPLEX *)calloc(le - 1, sizeof(COMPLEX));
            if (!w) {
                  printf("\nUnable to allocate array\n");
                  exit(1);
            }
            arg = 4.0*atan(1.0) / le;
            wrecur_real = w_real = cos(arg);
            wrecur_imag = w_imag = sin(arg);
            xj = w;
            for (j = 1; j < le; j++) {
                  xj->real = (float)wrecur_real;
                  xj->imag = (float)wrecur_imag;
                  xj++;
                  wtemp_real = wrecur_real*w_real - wrecur_imag*w_imag;
                  wrecur_imag = wrecur_real*w_imag + wrecur_imag*w_real;
                  wrecur_real = wtemp_real;
            }
      }
      le = n;
      windex = 1;
      for (l = 0; l < m; l++) {
            le = le / 2;
            for (i = 0; i < n; i = i + 2 * le) {
                  xi = x + i;
                  xip = xi + le;
                  temp.real = xi->real + xip->real;
                  temp.imag = xi->imag + xip->imag;
```



```
                        xip->real = xi->real - xip->real;
                        xip->imag = xi->imag - xip->imag;
                        *xi = temp;
                }
                wptr = w + windex - 1;
                for (j = 1; j < le; j++) {
                        u = *wptr;
                        for (i = j; i < n; i = i + 2 * le) {
                                xi = x + i;
                                xip = xi + le;
                                temp.real = xi->real + xip->real;
                                temp.imag = xi->imag + xip->imag;
                                tm.real = xi->real - xip->real;
                                tm.imag = xi->imag - xip->imag;
                                xip->real = tm.real*u.real - tm.imag*u.imag;
                                xip->imag = tm.real*u.imag + tm.imag*u.real;
                                *xi = temp;
                        }
                        wptr = wptr + windex;
                }
                windex = 2 * windex;
        }
        j = 0;
        for (i = 1; i < (n - 1); i++) {
                k = n / 2;
                while (k <= j) {
                        j = j - k;
                        k = k / 2;
                }
                j = j + k;
                if (i < j) {
                        xi = x + i;
                        xj = x + j;
                        temp = *xj;
                        *xj = *xi;
                        *xi = temp;
                }
        }
        scale = (float)(1.0 / n);
        for (i = 0; i < n; i++) {
                x->real = scale*x->real;
                x->imag = scale*x->imag;
                x++;
        }
}//end of ifourier


void black(COMPLEX *x, int n)

{
        int  i;
        double black, factor;

        factor = 8.0*atan(1.0) / (n - 1);
        for (i = 0; i<n; ++i) {
```



```
            black = 0.42 - 0.5*cos(factor*i) + 0.08*cos(2 * factor*i);
            x->real *= black;
            x->imag *= black;
            x++;
        }
}//end of black
```


A Mathematical Model For Performing Early Thermal Design Tradeoffs With Passively-Cooled Outdoor Electronics Assemblies Having Metal Enclosures

Find Major Design Variables Quickly Without Up Front Time-Consuming and Expensive Computer Aided Engineering Models

A Technical Paper prepared for SCTE•ISBE by

H. Bruce Jackson, Principal Mechanical Engineer/Consultant,
Lancaster, NY, 14086
Jackson_Consults@outlook.com
(301) 514-3180

SCTE • ISBE STANDARDS

Table of Contents

List of Figures

List of Tables

1. Introduction

As often seen outside on the ground in fenced compounds, pole-mounted, or aerial-mounted high above the ground on roadside utility lines, utility electronic assemblies are unattractive yet important components of cable television, power, and telephone networks. Most are housed in relatively thick (≥ 0.06 inch) die-cast aluminum or sheet metal weather-tight hinged cabinet or "suitcase" enclosures with external metal fin arrays to enhance heat removal. These assemblies are typically heated only by internal power dissipation, sunlight exposure, and atmospheric radiation. Many, like cable TV broadband amplifiers are "strand-mounted" to high-tension wire ropes. The internal power dissipations are usually low enough and the exposed surface areas large enough that these enclosures are designed to cool passively by natural convection, conduction, and radiation. For these assemblies, the expense, power consumption, low reliability, and loss of seal integrity introduced by fans and vents are wisely avoided. Early in a typically chaotic design effort, dimensions and internal heat dissipation are changing so often that a sufficiently-detailed finite volume method (FVM)-based or finite element analysis (FEA)-based computational fluid dynamic (CFD) analysis is futile. This begs the question: For early design tradeoffs, what variable can be quantified to quickly compare the effects of changes in internal electronic power dissipation, ambient temperature, and solar load, along with enclosure dimensions, orientations, and finishes?

In this paper, a simple mathematical model for early thermal design tradeoffs with passively-cooled outdoor electronics assemblies having metal enclosures is presented. The model is simple in that it lacks details required for predicting electrical component junction temperatures and associated temperature distributions within the enclosure. It does, however, enable the designer to very quickly compare performance differences between a high number of designs and at least determine initial enclosure dimensions, orientation, and surface finish while considering many variables. Furthermore, it offers a reasonable initial constant-wall-temperature boundary condition for the first CFD analysis to study the electronics and air temperatures inside the enclosure. It does so by calculating the average surface temperatures of each configuration. Any rise or fall in average surface temperature will indicate a loss or gain of thermal performance, respectively. Once a configuration is quickly identified, *then* it is recommended that a CFD FVM program be used to finalize the design with external fin arrays providing sufficiently low internal component junction temperatures. Given that the power consumption and heat dissipation requirements of outside plant devices are increasing with distributed access architectures, the need for such modeling to ensure device reliability is clear.

2. Model Development

To keep pace with rapid changes in internal electronic power dissipation and enclosure: dimensions, surface orientations, solar exposure factors and finish properties, it is useful to calculate an average steady-state enclosure surface temperature considering the following:

- the enclosure dimensions,
- the total internal component power dissipation,
- a diurnal solar cycle with its solar intensity (W/m^2) curve,
- a diurnal ambient temperature cycle,
- the highest expected relative humidity,
- sunlight exposure factors for each enclosure surface,
- the hemispherical infrared (IR) emittances of the enclosure's surface finishes, and
- the total hemispherical solar reflectances of the enclosure's surface finishes.

Here we make the conservative assumption of pure passive cooling (no forced convection, a.k.a. wind) while adopting the harsh diurnal temperature cycle Curve A1 and harsh diurnal solar intensity cycle Curve W of The US Department of Defense's MIL-STD-810G Environmental Engineering Considerations and Laboratory Tests [1]. This environment is used going forward.

If ground-based equipment in full sunlight is installed with the bottom surface close to the ground, then this model may be easily simplified to suit by removing the energy terms coming and going from the bottom surface. Once a preliminary design is defined, internal temperature distributions and component junction temperatures are typically estimated using thermal resistance network models [2] [3] or FVM CFD programs tailored for electronics applications such as MENTOR GRAPHICS' FloTHERM®, MENTOR GRAPHICS' FloEFD® or ANSYS's Icepak®. CFD tools are usually required anyway to optimize fin arrays since these are usually needed to keep enclosures small and lightweight. If rib-fins are chosen, a spreadsheet model can do a fair job of factoring in their contribution but it will not prevent the user from placing fins too close together and/or too deep to cause air flow "choking"[1] and such a simple solution is possible only with lengthwise-vertical rib-fins on the enclosure sides and ends. One way to overcome the temperature sensitivity of rib-fins to orientation is to adopt cylindrical pin-fins [4]. It has also been shown that an optimized pin-fin array is more effective in all orientations for passive cooling and requires less enclosure material than an optimized rib-fin array occupying the same volume in space, and, pound-for-pound, provides more heat dissipation [4]. If pin-fins are adopted or rib-fins are non-vertical or on the bottom or top surfaces, then a CFD modeling program is unavoidable since a simple mathematical model cannot be used to solve such problems.

The following assumptions apply to the general problem. In no order of importance:

a) The enclosure is encompassed by an infinite volume containing still air at constant ambient temperature T_a. This is a good assumption with utility assemblies high above the ground.
b) The assembly is installed far enough away from other heat-dissipating assemblies that there are no significant radiative heat exchanges between them.
c) A control volume is assumed conforming to the enclosure surfaces,
d) Airflow over the assembly surfaces is laminar natural convection.
e) The enclosure is metal so is more thermally-conductive than typically more economical plastics.

[1] This can also be called "incomplete convection boundary layer development."

f) Radiative surface properties are constant from surface-to-surface.
g) The enclosure surfaces are opaque in the visible and IR spectra.
h) The air surrounding the assembly is non-participatory with regard to radiation. This means the air is assumed to be fully transparent having no airborne particles, smoke, water vapor, or chemical pollutants.
i) Surface finish total solar absorptance, total solar reflectance, IR absorptance, and IR emittance will change with weathering.
j) There is no heat conducted through the assembly mounts. If there are any suspension wire rope "strands" and/or mounting brackets, these provide no heat conduction benefit. For pole-mounted assemblies, the designer should consider reducing the heat rejection from, and the sunlight entering, the pole-facing surface.
k) No heat is conducted through cables entering and leaving the assembly.
l) Surfaces are clean.

We go on to solve the general problem of finding the average surface temperature of a finless enclosure oriented as shown in Figure 1 with free convection cooling terms $\dot{Q}_{Ci}$ leaving the i'th surface, radiative cooling terms $\dot{Q}_{Ri}$ leaving the i'th surface, solar heating terms entering the i'th surface $\dot{Q}_{sun,i}$, environmental radiation heating terms entering the side and bottom surfaces $\dot{Q}_{a,i}$ ($1 \leq i \leq 5$), and a sky radiation heating term entering the sky-facing top surface $\dot{Q}_{sky,6}$ (i=6). These terms are all fully defined going forward.

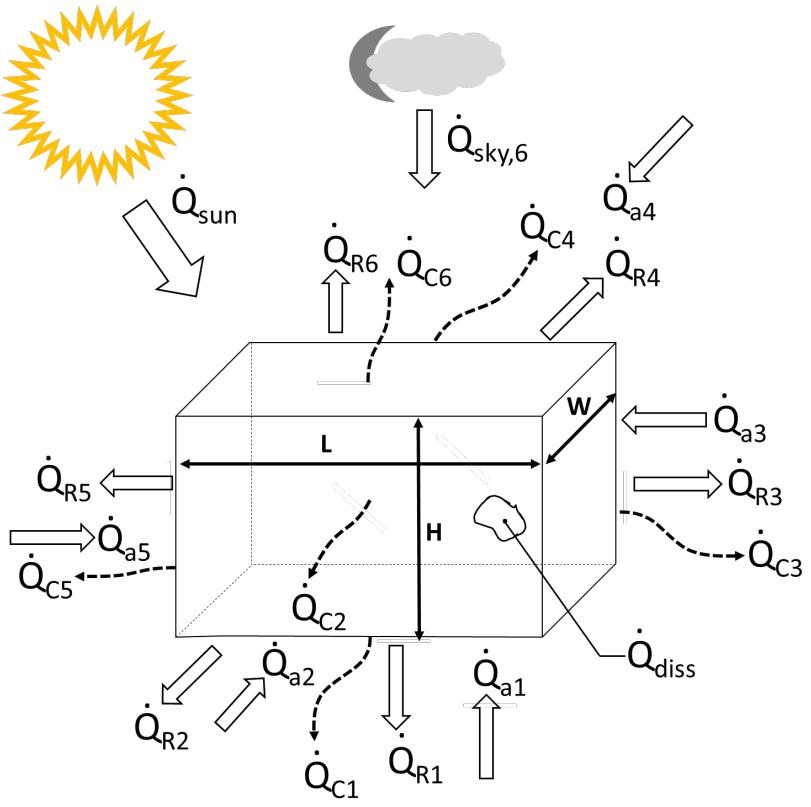

Figure 1 - Passive Cooling Terms for Finless Housing

A heat balance gives:

$$\dot{Q}_{out} - \dot{Q}_{in} = 0 \quad \text{(Equation 1)}$$

Considering first the energy leaving the assembly, $\dot{Q}_{out}$:

$$\dot{Q}_{out} = \dot{Q}_{C1} + \dot{Q}_{R1} + \dot{Q}_{C2} + \dot{Q}_{R2} + \dot{Q}_{C3} + \dot{Q}_{R3} + \dot{Q}_{C4} + \dot{Q}_{R4} + \dot{Q}_{C5} + \dot{Q}_{R5} + \dot{Q}_{C6} + \dot{Q}_{R6}$$
$$= \sum_{i=1}^{6} \dot{Q}_{Ri} + \sum_{i=1}^{6} \dot{Q}_{Ci} = \dot{Q}_R + \dot{Q}_C \quad \text{(Equation 2)}$$

It's necessary to consider each side of the enclosure one-by-one since natural convection and radiation terms depend on surface dimensions and orientations. Note there are no outgoing solar reflection terms. Reflected solar energy is not an outgoing term since it does not pass through the control volume boundary on the enclosure surfaces. Now it is assumed that the top (i = 6) is normal to the sun and sky-facing. $\dot{Q}_{out}$ consists of six convective terms and six radiative terms. Temperatures with the hat ($\widehat{\ }$) notation are in kelvin (K) and those without are in degree Celsius (°C):

$$\sum_{i=1}^{6} \dot{Q}_{Ci} = X\left(T_{surf} - T_a\right) \sum_{i=1}^{6} h_i \, A_i \quad \text{(Equation 3)}$$

$$\sum_{i=1}^{6} \dot{Q}_{Ri} = \sum_{i=1}^{6} f_{i\to\infty} A_i \, \sigma \epsilon_{h,IR} \hat{T}_{surf}^4 \quad \text{(Equation 4)}$$

leaving

$$\dot{Q}_{out} = \sigma \epsilon_{h,IR} \hat{T}_{surf}^4 \sum_{i=1}^{6} f_{i\to\infty} A_i + X\left(T_{surf} - T_a\right) \sum_{i=1}^{6} h_i \, A_i \quad \text{(Equation 5)}$$

$$X = \sqrt{\frac{\rho_{alt}}{\rho_{sl}}} = \sqrt{\frac{p_{alt}}{p_{sl}}} \quad \text{(Equation 6)}$$

$$\frac{\rho_{alt}}{\rho_{sl}} = \left(\frac{288 - 0.00198z}{288}\right)^{4.256} \quad \text{(Equation 7)}$$

$$h_i = 0.52 C_i \left(\frac{T_{surf} - T_a}{D_i}\right)^{0.25} \quad \text{(Equation 8)}$$

$$h_i = 0.80 C_i \left(\frac{T_{surf} - T_a}{D_i}\right)^{0.25} \quad \text{(Equation 9)}$$

Now on substituting the dimension-, orientation-, and temperature-dependent convection coefficients into Equation 5, noting that $\hat{T}_{surf} - \hat{T}_a = T_{surf} - T_a$, $f_{i\to\infty} = 1$, simplifying, and noting that Equation 9 commits use of mks SI metric units going forward,

$$\dot{Q}_{out} = \sigma \epsilon_{h,IR} \hat{T}_{surf}^4 \sum_{i=1}^{6} A_i + 0.80X\left(\hat{T}_{surf} - \hat{T}_a\right)^{1.25} \sum_{i=1}^{6} \frac{C_i A_i}{D_i^{0.25}} \quad \text{(Equation 10)}$$

The definitions, assumptions, units, and notes for the terms and subscripts in Equations 3-10 are included in Table 1.

Table 1 - Definitions, Assumptions, Units, and Notes for the Terms and Subscripts of Equations 3-10

Term/Subscript	Value/Units	Definition	Assumptions	Notes
σ = the Stefan-Boltzmann constant	5.6704 X 10^{-8} W/($m^2 \cdot K^4$)	See any heat transfer text like [5] or [6].	-	-
$\epsilon_{h,IR} \leq 1$	no units	The hemispherical IR emittance[2] of a clean assembly's surface.	Emittance is the same on all surfaces. Measure per [7]. For comparative purposes only, measure per [8].	Higher values of IR emittance apply to nonconductors (e.g., paints) and lower values apply to polished reflectors (e.g., polished aluminum) and may not exceed unity. Since published values of emittance rarely state what spectra they apply over or whether the emittance is *normal* or *hemispherical*, it's best to measure $\epsilon_{h,IR}$ for each finish. Many people assume commercial finishes are diffuse where $\epsilon_{h,IR} = \epsilon_{n,IR}$ but even more make the reckless assumption they are also gray where $\epsilon_h = \epsilon_n$.

[2] The USA NIST suggests that radiation properties ending in "ivity" (e.g. emissivity) should be reserved only for "optically pure" surfaces and properties for all other surfaces should end with "tance" (e.g., emittance). But NIST's own literature uses the terms emissivity and emittance interchangeably for all kinds of surfaces.

T_{surf}	°C	The steady-state average surface temperature in °C.	The same from point-to-point and from surface-to-surface regardless of sun exposure. The assembly is light enough to give it a time constant short enough to approximate a steady-state about the peak of the diurnal solar cycle.	The assembly's time constant needs to be estimated analytically or by test to ensure it is short enough to justify a steady-state assumption around the solar cycle peak. If not, then a transient solution is required.
$\hat{T}_{surf}$	K	The steady-state average surface temperature in K.	-	-
$1 \leq i \leq 6$	no units	A subscript representing one of the six surfaces of a the "finless box" enclosure of Figure 1.	-	-
1	no units	Subscript pertaining to the assembly's bottom surface.	Bottom receives scattered sunlight from below. Bottom receives environmental IR radiation from below.	-
6	no units	Subscript pertaining to the assembly's top surface.	Top sees the sun's full intensity from directly overhead.	-
2-5	no units	Subscripts pertaining to the assembly's side and end surfaces.	Sides and ends see only part of the sun's full intensity. Sides and ends receive lateral environmental IR radiation.	-

$f_{i\to\infty}$	= 1 no units	Radiation view factor from enclosure surface *i* to the surrounding infinite control volume.	Assembly is high enough off the ground and far enough from other objects at temperatures warmer or cooler than ambient temperature. This justifies the "infinite air volume" assumption: $f_{i\to\infty} = 1$.	-
A_i	m^2	The area of enclosure surface i.	Surfaces are all flat, rectangular and either horizontal, or vertical.	-
Y_i	$0 \le Y_i \le 1$ no units	Sunlight exposure factor is the fraction of full sunlight seen by surface i.	Per [9] except for the bottom surface (i=1).	It's understood that the top surface (W x L) may not be the largest since (H x L) may very well be3.
Y_1	0.25 no units	The sunlight exposure factor for the bottom is 25%.	For scattered sunlight (UV, visible and IR) from below.	-
Y_i $2 \le i \le 5$	0.50 no units	Sunlight exposure factor for the sides and ends (50%).	[9]	-
Y_6	1.0 no units	Sunlight exposure for the top (100%).	[9]	-
$0 < X \le 1$	no units	Altitude derating factor for convection cooling.	-	Accounts for thinner air and reduced convection at design altitude.

[3] Regarding solar flux (W/m^2) at times other than peak, one might ask if the sun normal to the enclosure's maximum projected area (all sides combined) at that time of day may be the worst case. Certainly, the illuminated area is greatest but not the intensity. This complicates the problem and suggests a complete transient CFD FVM solution.

$\frac{\rho_{alt}}{\rho_{sl}}$	no units	The ratio of air density ρ_{alt} at the design altitude at constant temperature to air density ρ_{sl} at sea level at the same temperature.	Used in Equation 7.	[10]
$\frac{p_{alt}}{p_{sl}}$	no units	The ratio of air pressure at the design altitude p_{alt} to the air pressure at sea level p_{sl}.	Used to compensate for design altitude in natural convection equation.	From the Ideal Gas Law, for an isothermal expansion from sea level at constant temperature to the design altitude at constant temperature: $\frac{p_{alt}}{p_{sl}} = \frac{\rho_{alt}}{\rho_{sl}}$
$0 \leq z \leq 37{,}000$	feet	The design altitude above sea level.	-	Applies to Equation 7.
C_i	no units	A coefficient depending on the orientation of surface i.	Laminar airflow.	[11]
$C_1 = 0.26$	no units	A coefficient for horizontal surfaces facing down (bottom only)	Laminar airflow.	[11]
$C_i = 0.56$	for i=2-5 no units	A coefficient for vertical surfaces (two sides and two ends)	Laminar airflow.	[11]
$C_6 = 0.52$	no units	A coefficient for horizontal surfaces facing up (top only)	Laminar airflow.	[11]

T_a	°C	The ambient air temperature in °C from [1] for this example.	Constant within the infinite volume surrounding the assembly. The diurnal ambient temperature curve lags the solar flux diurnal cycle and may be obtained by time-averaging. The ambient temperature curve peaks at the same time as the solar intensity curve.	-
$\hat{T}_a$	K	The ambient air temperature in K.	-	-
h_i	$\frac{\text{Btu}}{\text{hr} \cdot \text{ft}^2 \cdot \ ^\circ\text{F}}$	The laminar natural convection coefficient for surface i.	Airflow is laminar over the assembly surfaces.	Equation 8 per [11]
h_i	W/(m2·°C)	The laminar natural convection coefficient for surface i in mks SI metric units.	-	Equation 9. Author converted Equation 9 from Equation 8.
D_i	m	The characteristic length of the i'th convecting surface.	[11]	-

$D_1 = D_6$	$= 2\frac{LW}{L+W}$	The characteristic length of the horizontal rectangular surfaces [11] (top and bottom).	$L \geq W$	See Figure 1 for dimension L and W conventions. Note the formula for this purpose in [12] (Ellison, 1989) is in error and provides ¼ of these values.
D_i for i=2-5	$= H$ m	The characteristic length of the side and end surfaces [11].	-	See Figure 1 for dimension *H* convention.

Now for the total energy entering the enclosure Q_{in}:

$$\dot{Q}_{in} = \dot{Q}_{diss} + \dot{Q}_{sun} + \dot{Q}_a + \dot{Q}_{sky,6} \quad \text{(Equation 11)}$$

For the sunlight:

$$\dot{Q}_{sun} = \sum_{i=1}^{6} \dot{Q}_{sun,i} = \sum_{i=1}^{6} \overline{S}' \alpha_{h,sun} Y_i A_i = \overline{S}' \alpha_{h,sun} \sum_{i=1}^{6} Y_i A_i \quad \text{(Equation 12)}$$

and since

$$\alpha_{h,sun} + \rho_{h,sun} + \tau_{h,sun} = 1 \quad \text{(Equation 13)}$$

and $\tau_{h,sun}$= 0 for the opaque surfaces,

$$\alpha_{h,sun} = 1 - \rho_{h,sun} \quad \text{(Equation 14)}$$

$$\dot{Q}_{sun} = \overline{S}' \alpha_{h,sun} \sum_{i=1}^{6} Y_i A_i \quad \text{(Equation 15)}$$

Now we need the IR radiation absorbed from the surroundings $\dot{Q}_a$ (W) into the sides, bottom, and ends. By again setting $f_{i\to\infty} = 1$,

$$\dot{Q}_a = \sum_{i=1}^{5} \dot{Q}_{a,i} = \sigma \epsilon_{h,IR} \hat{T}_a^4 \sum_{i=1}^{5} A_i \quad \text{(Equation 16)}$$

The longwave radiation absorbed by the top, $\dot{Q}_{6,sky}$, is different from the sides, ends, and bottom since the top (i = 6) faces the sky at temperature $\hat{T}_{sky}$ rather than the surroundings at temperature $\hat{T}_a$. The solar intensity reaching the Earth is obviously greatest on a cloudless day. The monthly average clear sky emissivity in the USA is given by [13]:

$$\epsilon_{sky} = 0.711 + 0.56\left({}^{T_d}/_{100}\right) + 0.73\left({}^{T_d}/_{100}\right)^2 \quad \text{(Equation 17)}$$

and an adjustment for the time of day as suggested by [13]:

$$\Delta\epsilon_{sky} = 0.013\cos\left(2\pi\, {}^{t}/_{24}\right) \quad \text{(Equation 18)}$$

where t is the hour of the day ($0 \leq t \leq 24$).

The clear sky temperature is then given by [13]:

$$\hat{T}_{sky} = \hat{T}_a \sqrt[4]{\epsilon_{sky} + \Delta\epsilon_{sky}} \quad \text{(Equation 19)}$$

and finally,

$$\dot{Q}_{sky,6} = \sigma\epsilon_{h,IR} A_6 \hat{T}_{sky}^4 \quad \text{(Equation 20)}$$

To find the dew point temperature requires a relative humidity. A representative value in the daytime can come from the Mohave Desert where the daytime relative humidity varies from 10% to 30% [14]. Since high dew point temperatures come with high relative humidity, and a high sky emissivity means a higher sky temperature and more energy absorbed by the assembly, we conservatively assume a relative humidity of 30%. With a dry bulb temperature of 47 °C (117 °F) from [1], the dew point temperature is 25 °C (77 °F) [15].

Then ϵ_{sky} = 0.89, and since the peak solar intensity occurs at hour t = 12.5 per [1], $\Delta\epsilon_{sky} = -0.013$, and $\hat{T}_{sky} = (320)\sqrt[4]{0.88} = 310$ K.

The long-wave radiation absorbed into the top surface is now available using Equation 20. Finally, for the total energy entering the assembly $\dot{Q}_{in}$:

$$\dot{Q}_{in} = \dot{Q}_{diss} + \dot{Q}_{sun} + \dot{Q}_a + \dot{Q}_{sky,6} \quad \text{(Equation 21)}$$

$$\dot{Q}_{in} = \dot{Q}_{diss} + \overline{S}'\alpha_{h,sun}\sum_{i=1}^{6} Y_i A_i + \sigma\epsilon_{h,IR}\hat{T}_a^4\sum_{i=1}^{5} A_i + \sigma\epsilon_{h,IR}A_6\hat{T}_{sky}^4 \quad \text{(Equation 22)}$$

The definitions, assumptions, and notes for all new terms in Equations 11-22 are included in Table 2.

Table 2 - Definitions, Assumptions, and Notes for New Terms in Equations 11-22

Term	Value/Units	Definition	Assumptions	Notes
$\dot{Q}_{diss}$	W	The total of the assembly's internal electronic heat dissipation	-	-
$\dot{Q}_{sun}$	W	The total sunlight absorbed by the assembly surfaces.	-	-
$\dot{Q}_{a,i}$	W	The environmental IR heat absorbed by the i'th surface for i=1-5 from the surroundings.	-	-
$\dot{Q}_{a}$	W	The total environmental IR heat absorbed by the end, side, and bottom surfaces for i=1-5.	-	-
$\dot{Q}_{sky,6}$	W	The longwave IR sky radiation absorbed by the sky-facing top surface.	-	-
$\hat{T}_{sky}$	K	The clear sky temperature.	-	Equation 19
T_d	°C	The dew point temperature.	[15]	-
$\overline{S'} = 1120$	W/m^2	The time-averaged solar radiation about the peak of the solar cycle appropriate to the geography, altitude, and time-of-year for the design environment. From [1] for this example.	The assembly's thermal time constant is short enough to justify a steady-state assumption about the solar cycle's peak.	-
$\epsilon_{sky} \leq 1$	no units	The monthly average clear sky emissivity in the USA.	-	Equation 17
$\Delta\epsilon_h \leq 1$	no units	A time-of-day adjustment to the clear sky emissivity.	-	Equation 18

Term	Value/Units	Definition	Assumptions	Notes
$\rho_{h,sun} \leq 1$	no units	The total hemispherical solar reflectance[4] is the same on all surfaces.	Measure per [16]. For comparative purposes only, measure per [17].	It's unclear if *hemispherical* solar reflectance and *normal* solar reflectance are substantially different for commercial finishes (mill finishes included), so if the designer finds published *normal* values or finds values that says neither, then he/she is advised to have candidate surface finishes tested.
$\alpha_{h,sun} \leq 1$	no units	The total hemispherical solar absorptance.	-	Calculated from $\rho_{h,sun}$ using Equation 14.
$\tau_{h,sun} \leq 1$	no units	The total hemispherical solar transmittance.	$\tau_{h,sun} = 0$ for the opaque surfaces.	-

Finally, for the energy balance, recall Equation 1:

$$\dot{Q}_{out} - \dot{Q}_{in} = 0 \quad \text{(Equation 1)}$$

Substituting:

[4] It's not enough that the incident solar radiation has the right intensity G (W/m^2). The radiation must also have the right spectral distribution. For example, some people use IR light sources with an intensity G to replicate a solar spectrum of the same intensity G. The result will be highly inaccurate because commercial finishes are selective to spectral bands. For example, most paints have solar reflectances that are more selective to an IR light source than a "spectrally-correct" solar light source having the highest intensity component in the visible band. So, most paints under IR light sources will perform worse (have a lower solar reflectance) than they will in real or simulated sunlight.

$$\left[\sigma\epsilon_{h,IR}\hat{T}_{surf}^{4}\sum_{i=1}^{6}A_i + 0.80X\left(T_{surf}-T_a\right)\sum_{i=1}^{6}C_i\frac{\left(\hat{T}_{surf}-\hat{T}_a\right)^{0.25}}{D_i^{0.25}}A_i\right] - \left[\dot{Q}_{diss}+\overline{S}'\alpha_{h,sun}\sum_{i=1}^{6}Y_i\,A_i+\sigma\epsilon_{h,IR}\hat{T}_a^{4}\sum_{i=1}^{5}A_i+\sigma\epsilon_{h,IR}A_6\hat{T}_{sky}^{4}\right]=0$$

and simplifying,

$$\left[\sigma\epsilon_{h,IR}\hat{T}_{surf}^{4}\sum_{i=1}^{6}A_i + 0.80X\left(\hat{T}_{surf}-\hat{T}_a\right)^{1.25}\sum_{i=1}^{6}\frac{C_iA_i}{D_i^{0.25}}\right] - \left[\dot{Q}_{diss}+\overline{S}'\alpha_{h,sun}\sum_{i=1}^{6}Y_i\,A_i+\sigma\epsilon_{h,IR}\left(\sum_{i=1}^{5}A_i\hat{T}_a^{4}+A_6\,\hat{T}_{sky}^{4}\right)\right] = 0 \quad \text{(Equation 23)}$$

This is a single equation in one unknown $\hat{T}_{surf}$. Without knowing what finish, if any, will be used to coat the enclosure we don't know for certain if a weathered surface helps or hurts cooling. So, there are two extremes to consider: one with a new surface and one with a weathered "old" surface. For this reason, both problems must be considered. An industry-standard accelerated weathering can be achieved using [18][5] for finished and unfinished metal surfaces[6] . But it's critical to realize that the IR emittance and solar absorptance of organically-finished and unfinished metals are lowest with new, un-weathered surfaces and increase with old, weathered surfaces. So, the big question is: does a candidate surface get better or worse with age for rejecting heat? Sufficient testing and this model will answer that question. The measured solar radiation surface property is the reflectance so as reflectance reduces, absorptance increases per Equation 14. This reduction in reflectance is obvious from looking at everyday outdoor surfaces because these surfaces become duller (less reflective) with weathering.

The nonlinear Equation 23 can be solved easily by iterating $\hat{T}_{surf}$ using any math or spreadsheet program.

3. Determining the Steady-State Solar Radiation[7]

It's generally too harsh to assume a constant peak value over any complete heat-up/cool-down cycle so a reasonable time interval is needed for determining the time-averaged solar radiation.

[5] In a lab certified to ISO/TEC 17025.
[6] Per [20], [18] is used for comparing emittances and solar absorptances of unfinished metal surfaces with finished metal surfaces (although the title says "Non-Metallic Surfaces").
[7] Some industries may be obligated to follow environmental standards specifically for that industry. For example, the designers of outdoor cable TV equipment may use [19] where there is no specified solar load. That standard, rather than stating transient solar radiation and temperature profiles, imposes an artificially high constant ambient temperature of 60 °C however it says nothing about airspeed in the test chamber which dramatically influences forced convection cooling. And it says nothing about the radical

The time constant used later in the example was the time necessary for an assembly to heat from 25 °C to 41 °C when heated from a steady-state in a well-mixed chamber at 25 °C to a steady-state in a well-mixed chamber at 50 °C. Although the ambient temperature here is 47 °C, the time constant previously measured will be used for the example. The 41 °C was calculated as follows:

For a lumped mass system,

$$T = T_1 + (T_2 - T_1)\left(1 - e^{-t/\tau}\right) \quad \text{(Equation 24)}$$

and when the time constant is reached, t=τ so

$$T = T_1 + (T_2 - T_1)(1 - e^{-1}) \quad \text{(Equation 25)}$$

$$= 25 + (50 - 25)(1 - e^{-1}) = 41\ ℃$$

Our assembly was found to take 1.5 hr to reach 41 °C in the shade in a temperature chamber with well-mixed air[8] [4]. Hence the time constant for the assembly was 1.5 hr in that environment.

Since we know the assembly will reach a chamber temperature of 47 °C sooner than 1.5 hr, we know the time constant for a lower chamber temperature will be less than 1.5 hr. With sunlight exposure and still air at a constant 47 °C, the time to warm to steady-state will certainly be less than 1 hr. Assume a time constant of 1 hr and a 2 hr interval centered about the solar cycle peak.

The solar radiation curve W is symmetrical about its peak of 1120 W/m^2 occurring over time interval 1200-1300. Over this interval, the radiation spectrum goes first through the ultraviolet (UV) A band 0.28-0.32 µm containing 5 W/m^2, the UVB band 0.32-0.40 µm containing 63 W/m^2, the visible band 0.40-0.78 µm containing 560 W/m^2, and the IR band 0.78-3.0 µm containing 492 W/m^2 [1]. If the designer chooses to do in-situ testing and doesn't care about actinic (UV) effects, then if a test chamber lacks the UV spectral bands then these can be ignored because they contribute so little to heating.

Since the cycle is symmetrical about the peak at 7.5 hr, we only need to integrate over a one-hour interval leading up to the peak. With t =0 at sunrise (0500), a 6th order interpolating polynomial (from a spreadsheet scatter plot) over $0 \le t \le 7.5$ hr is:

$$S'(t) = 0.0082t^6 - 0.3728t^5 + 5.935t^4 - 47.05t^3 + 181.5t^2 - 83.41t$$

$$\bar{S}' = \text{the time averaged solar radiation} = \frac{\int_{t_1}^{t_2} S'(t)dt}{t_2 - t_1} \quad \text{(Equation 28)}$$

Integration yields $\bar{S}' = 1117\ \frac{\text{W}}{\text{m}^2}$.

differences in temperature between a shaded object and one in the sun. Lastly, since all thermal chambers are not created equal, how can results be reasonably compared from one chamber to the next without chamber standardization?

[8] A running average was collected from thermocouples placed at four representative places inside the assembly.

So, the time constant is short enough to assume a constant solar flux of $\bar{S}' = 1120\ {}^{W}/_{m^2}$ over the 2 hr interval. But for a larger, heavier enclosure (or cabinet) with a higher time constant (greater "time to reach steady-state") the time-averaged radiation will be less, and time-averaging will be more important to prevent overstating the solar load.

4. An Example Steady-State Solution

Consider an assembly and environment with the following conditions and properties:

A. Fluid: initially still air with only laminar natural convection at all times.
B. Relative humidity: 30%.
C. Length, L: 0.91 m (36in).
D. Width, W: 0.30 m (12in).
E. Height, H: 0.46 m (18 in).
F. Internal power dissipation, $\dot{Q}_{diss}$: 53 W.
G. Elevation, z: 1800 m (5906 ft).
H. Time-averaged solar flux: 1120 W/m^2.
I. Ambient (dry bulb) temperature: 47 °C (117 °F).
J. Surfaces: three types per Table 3.
K. Top, side, and end surface exposures (%) per [9].
L. Bottom surface exposure 25%.
M. Diurnal cycle per temperature curve A1 and solar curve W of [1] Figure 505.6-1. Procedure 1, Cycling Test.
N. Total hemispherical solar reflectances were measured by R&D SERVICES (Lenoir City, TN) using a DEVICES AND SERVICES COMPANY (Dallas, TX) Model SSR-ER Version 5.0 portable solar reflectometer per [17]. The test method was inexpensive and fast. Values were used for comparative purposes only and the method is not intended to supplant [16].
O. Hemispherical IR emittances were measured by R&D SERVICES (Lenoir City, TN) using a DEVICES AND SERVICES COMPANY (Dallas, TX) Model AE portable emissometer per [8]. The test method was inexpensive and fast. Values were used for comparative purposes only and the method is not intended to supplant [7].

Table 3 - A Comparison of New Enclosure Surfaces[9]

Surface Description	$\frac{\epsilon_{h,IR}}{\alpha_{h,sun}}$	$\epsilon_{h,IR}$	$\rho_{h,sun}$	$\alpha_{h,sun}$	T_{surf} °C/°F	Surface temp rise above ambient °C/°F
Clean new A360 die-cast aluminum from foundry with light oxidation	0.40	0.23	0.42	0.58	140/284	93/167

[9] Due to the conservative model assumptions, average surface temperature *differences* are more credible than average surface *temperatures*.

Clean new A360 die-cast aluminum with clear conversion coating per MIL-C-5541, Class 1A, Clear (Irridite)	0.33	0.20	0.40	0.60	149/300	102/184
Clean new A360 die cast aluminum with white TGIC polyester powder coat paint, SHERWIN WILLIAMS #PWS-8-3007	3.4	0.91	0.73	0.27	69/156	23/41

Observations from this example:

a) Simply by painting a shiny new die-cast aluminum enclosure with a white powder coat paint, the model shows the average surface temperature is lowered by 140-69=71 °C (128 °F). While this figure is based on conservative assumptions, it's still clear that an enclosure painted white is much favored over an unfinished one.
b) Some surface treatments or finishes may harm thermal performance. This example shows that a clear Irridite ® surface treatment will increase the average surface temperature by 149-140 = 9 °C (16 °F).
c) Clearly the first two surface finishes in Table 3 are problematic. If the designer is required to use either surface finish, then his/her choices would be to increase dimensions (in this example, dimension H would be the first considered), add fins, and/or incorporate sun shields.

The average surface temperature is sensitive to the ratio $\epsilon_{h,IR}/\alpha_{h,sun}$. The higher this ratio the better. With total hemispherical solar absorptance and hemispherical IR emittance varying from 0 to 1 in steps of 0.05, a 3D plot of average surface temperature vs. emittance and absorptance is provided in Figure 2 for this example[10]. As for adding rib-fins, they do have obvious advantages due to their toughness and simplicity, but they must be installed oriented lengthwise-vertical and there is no way to guarantee installers will do that. They become less effective oriented otherwise and installing them lengthwise-horizontal will do more harm than good [4].

One way to overcome fin orientation sensitivity is to switch to pin-fins. In the author's opinion, the drawbacks to pin-fins are that they are not as durable as rib-fins, so the enclosures require more careful handling and may require at least three big stacking pins for handling and storage. The tooling is costlier, the die ejection scrap rate may be higher, and pin-fins won't help your enclosures win any beauty contests! But the lower per enclosure foundry metal volume (and lower weight), lower orientation sensitivity, and higher heat rejection of cylindrical pin-fins easily outweigh their drawbacks.

[10] This required the solution of the nonlinear equation $21^2 = 441$ times.

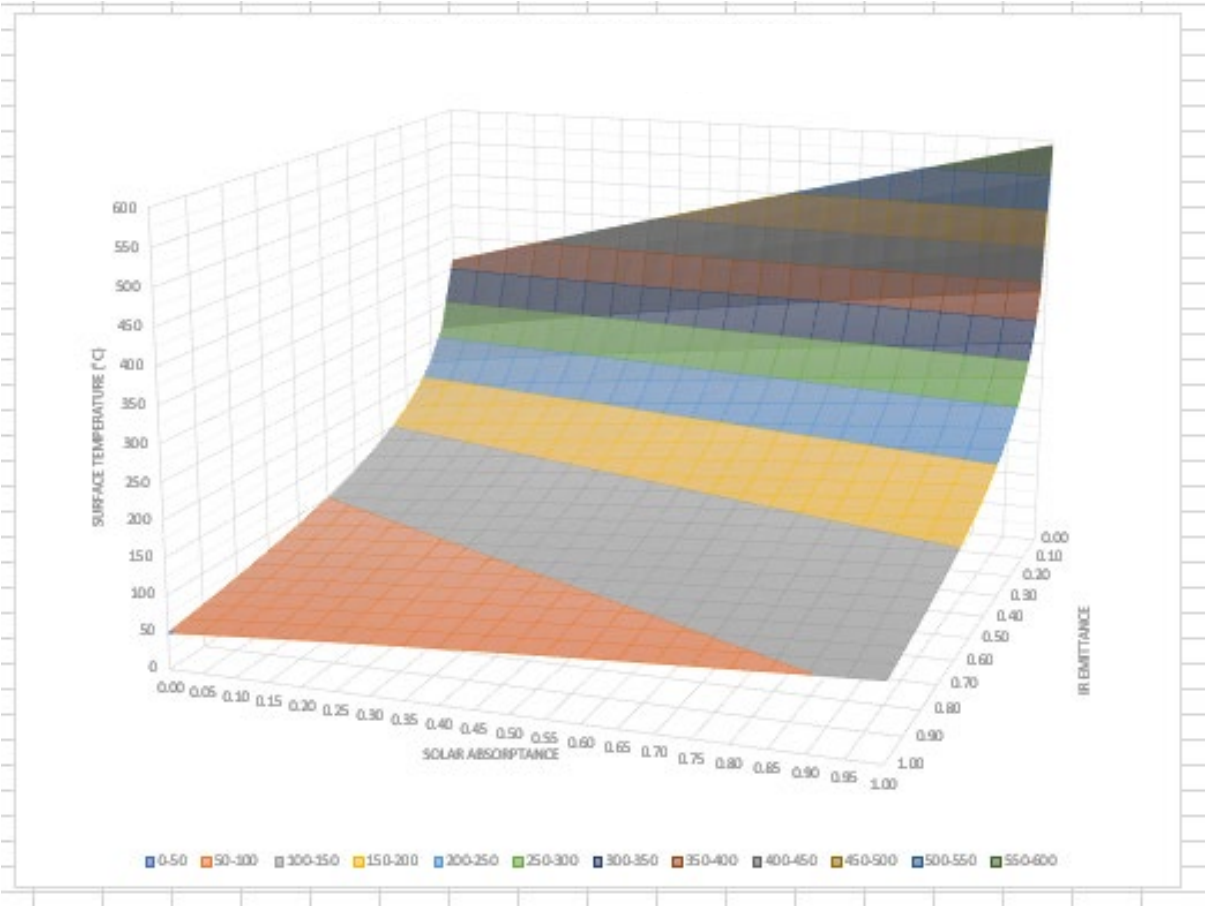

Figure 2 – Surface Temperature Sensitivity to IR Emittance and Solar Absorptance

5. Conclusions and Recommendations

In closing, while the model described by Equation 23 has limited value for predicting electrical component junction temperatures because it says little about temperature distributions inside the enclosure and ignores any benefit from external fins, it does offer the designer a way to make rapid comparisons of thermal design parameters without getting caught up in complex CAE modeling software that is unforgiving given numerous major design changes early in the design process; it empowers the designer to at least determine workable enclosure dimensions, surface orientations and surface finishes while managing many variables. But for the final design, the designer will likely need a more extensive analysis using a full-blown CFD FVM program where internal detail and external fins can be reckoned with simultaneously. Lastly, it provides a reasonable initial boundary condition (a constant-wall-temperature) for a CFD FVM analysis of the electronics and air temperatures inside the enclosure.

If the designer wishes to compare performance differences between many designs by comparing average surface temperatures, then Equation 23 can be looped where each iteration solves for one configuration. For steady state or transient analysis of internal component junction temperatures, then a CFD FVM program is favored over a thermal FEA program or thermal network resistance program [2] [3] because no guesswork is required in determining convection coefficients (which are not needed). In any event, a

CFD FVM program is a must for optimizing external fin arrays and fine-tuning the overall enclosure geometry.

Enclosure temperatures in harsh, full-sunlight desert environments are highly sensitive to enclosure surface finish radiation properties. The designer can explore various surface finishes, but testing is needed to find their radiation properties before and after weathering. The best finish is one that maintains the highest value of $\epsilon_{h,IR}/\alpha_{h,sun}$ over time. For example, due to the rougher texture it seems that with most surfaces the surface hemispherical IR emittance and total hemispherical solar absorptance increase with weathering but by how much? What happens to the ratio $\epsilon_{h,IR}/\alpha_{h,sun}$? Service providers install utility assemblies practically anywhere (not just where they were designed to go!) with little attention to thermal performance. For example, although the designers never intended assemblies to be installed this way, service providers install assemblies with rib-fins lengthwise-horizontal. This makes rib-fins do more harm than good. Many assemblies are also installed in ventilated ground pedestals or in green plastic "toaster" ground enclosures (e.g., cable TV amplifiers) where the cables are buried. These installations are worthy of analysis, but it's most likely that enclosures in direct sunlight are worst-case with respect to thermal performance if one considers only common-sense installations.

6. Abbreviations (including those used in select subscripts)

a	ambient
°C	degree Celsius
CAE	computer aided engineering
CFD	computational fluid dynamics
d	dew point
diss	dissipation
°F	degree Fahrenheit
FEA	finite element analysis
FVM	finite volume method
h	hemispherical
IR	infrared
ISBE	International Society of Broadband Experts
K	kelvin
mks	meter-kilogram-second
n	normal
NIST	National Institute of Standards and Technology
SCTE	Society of Cable Telecommunication Engineers
SI	International System of Units (*Le Système International d' Unités*)
UV	ultraviolet (with two sub-bands referenced: UVA and UVB)

7. Bibliography and References

[1] USA Department of Defense MIL-STD-810G. DEPARTMENT OF DEFENSE TEST STANDARD: ENVIRONMENTAL ENGINEERING CONSIDERATIONS AND LABORATORY TESTS. s.l. : USA Department of Defense, 2014.

[2] K & K Associates. network modeling guide from NASA Contract #NAS 9-10435. Version 97.003. Westminster : K & K Associates, 2000. p. 30. No title is provided. Title in this citation is only a description.

[3] Naughton, Mike. Thermal Software. Sauna. [Online] Thermal Solutions. [Cited: April 23, 2018.] Designed just for electronics applications. www.thermalsoftware.com.

[4] Jackson, Bruce. Project Report, Thermal R&D for CATV Applications. State College: C-COR ELECTRONICS, INC., 1997. RT0001.

[5] Incropera, Frank P. and DeWitt, David P. Fundamentals of Heat and Mass Transfer. s.l. : John Wiley & Sons, Inc., 1990.

[6] Thomas, L.C. Heat Transfer Professional Version. s.l. : Capstone Publishing Corp., 1999.

[7] ASTM International C835-06(2013)e1. Standard Test Method for Total Hemispherical Emittance of Surfaces up to 1400°C. West Conshohocken, PA, USA, PA, USA : s.n., 2013.

[8] ASTM International C1371. Standard Test Method for Determination of Emittance of Materials Near Room Temperature Using Portable Emissometers. West Conshohocken, PA, USA, PA, USA : s.n., 1997 (pre-release).

[9] Telecordia Technologies GR-487-CORE. Generic Requirements for Electronic Equipment Cabinets. s.l. : Telecordia Technologies, March 2000. 2. GR-487-CORE.

[10] Madison, Richard. Fan Engineering. Buffalo, NY : Buffalo Forge Company, 1949.

[11] Steinberg, Dave S. Cooling Techniques for Electronic Equipment. 2. s.l. : John Wiley & Sons, Inc., 1991.

[12] Ellison, Gordon N. Thermal Computations for Electronic Equipment. s.l. : Robert E. Krieger Publishing, 1989.

[13] Characteristics of Infrared Sky Radiation in the United States. Martin, M. and Berdahl, P. 3-4, s.l. : Solar Energy, 1984, Vol. 33.

[14] Liba, Angela. What Is the Humidity of the Mojave Desert? Sciencing.com. [Online] April 25, 2017. [Cited: January 5, 2018.] https://sciencing.com/humidity-mojave-desert-19526.html.

[15] Image Permanence Institute, Rochester Institute of Technology. Dew Point Calculator. [Online] [Cited: January 5, 2018.] http://www.dpcalc.org/.

[16] ASTM International E903-12. Standard Test Method for Solar Absorptance, Reflectance, and Transmittance of Materials Using Integrating Spheres. West Conshohocken, PA, USA, PA, USA : s.n., 2012.

[17] ASTM International C1549. Standard Test Method for Determination of Solar Reflectance Near Ambient Temperature Using a Portable Solar Reflectometer. West Conshohocken, PA, USA : s.n.

[18] ASTM International G155-13. Standard Practice for Operating Xenon Arc Light Apparatus for Exposure of Non-Metallic Materials. West Conshohocken, PA, USA, PA, USA : s.n., 2013.

[19] Society of Cable Telecommunications Engineers. Recommended Environmental Condition Ranges for Broadband Communications Equipment. Exton, PA, USA: Society of Cable Telecommunications Engineers, 2016. ANSI/SCTE 158 2016.

[20] Pattison, Lynn. Technical Contact for ASTM G155-13. [interv.] H. Bruce Jackson. February 12, 2018.

Optimization of Practices in Operation and Maintenance of HFC-Networks

Improved Scheduling of Field Service Deployment Using Mathematical Optimization Algorithms

A Technical Paper prepared for SCTE•ISBE by

Dr. Alexander C. Adams, Managing Director & CSO, Adams Group, SCTE•ISBE Member
Hessenring 3
Büttelborn, Germany, 64572
alexander.adams@adamsgroup.de
+49 176 1988 2023

Dr. Antonio Morsi, VP Analytics & Optimization, Adams Group
Hessenring 3
Büttelborn, Germany, 64572
antonio.morsi@adamsgroup.de
+49 170 1644 880

Dr. Björn Geißler, VP Analytics & Optimization, Adams Group
Hessenring 3
Büttelborn, Germany, 64572
bjoern.geissler@adamsgroup.de
+49 172 631 4470

Table of Contents

List of Figures

1. Introduction

In recent years, data analytics and especially machine learning (ML) have gained increased interest in the cable industry. Applications of ML range from proactive network maintenance (PNM) upstream pre-equalization and downstream spectral analysis to predictive video channel lineups, digital video recording (DVR) and customer churn prediction. For an overview on recent applications of ML in cable access networks we refer to [1]. However, besides ML, which is essentially solving classification and regression problems with advanced methods of mathematical statistics, there is another related field of applied mathematics that has evolved from mathematical research to an industry-ready technology in recent years: mathematical optimization (MO), cf., Figure 1. In a nutshell, while ML is about learning a model to predict future events, MO is about prescribing an action that leads to the best possible outcome on the basis of a model. Although less noted than ML, MO applies to the field of cable access networks as well, ranging from network design and user service profile optimization [2] to operational tasks such as scheduling a team of field service technicians. The latter is the subject of this paper.

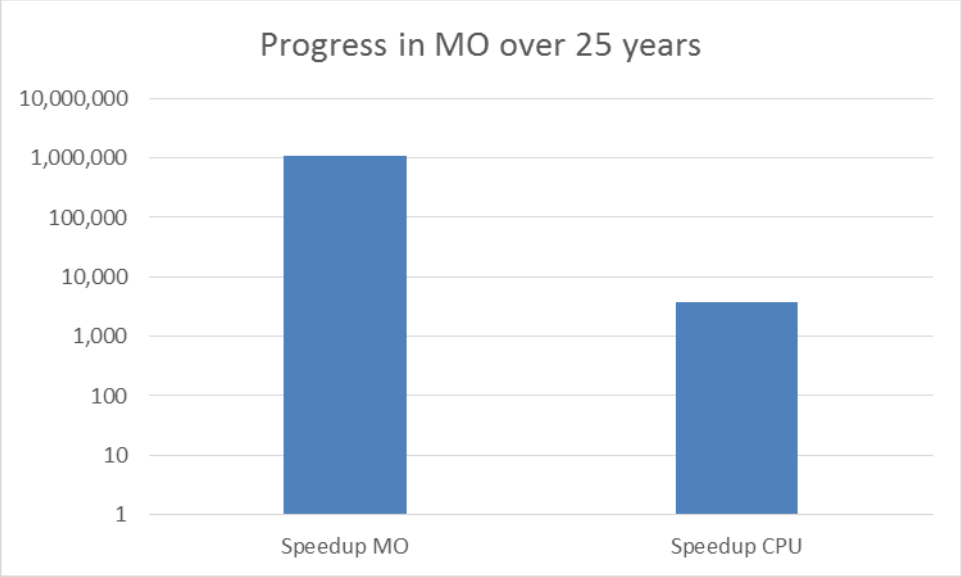

Figure 1 – Machine independent speedup of MO algorithms

The speed-up of mathematical optimization algorithms in Figure 1 relates to a set of benchmark optimization problems which were measured over the years on then up-to-date hardware utilizing the best available codes at the time. This speed-up factor is about 1.3 million in 2017. It is 300 times larger than the speed-up of computer processors (Moore's "Law") during the same time [3].

2. Scheduling Field Service Technicians

Currently the scheduling of field service technicians is often done manually or with the aid of freely available route planning tools. Computing the fastest routes between service locations helps dispatchers to make reasonable assignments of technicians to service tasks and to set proper appointments with the

customers, one by one. However, this approach lacks a global perspective on the problem to be solved and there is headroom for improvement incorporating MO, as shown in this paper.

On the other hand, there is commercial software available for solving vehicle routing problems (VRP) accounting for this global perspective by assigning a team of technicians to a large number of service tasks at once, considering service level agreements, predefined time-windows, driving distances and even technicians' qualifications. However, these tools mostly rely on so-called metaheuristic or evolutionary algorithms that cannot provide any guarantee on the quality of the computed schedule or are computationally expensive. Under these circumstances some functions, such as dynamic rescheduling, become impractical.

In contrast, the algorithm for scheduling field service technicians described in this paper utilizes the most recent mathematical optimization techniques and is thus able to (a) simultaneously take all data and constraints into account, (b) offers dynamic rescheduling capabilities, and (c) always provides a guarantee on the quality of the computed schedules.

2.1. The Application

The herein described routing and scheduling application for field service technicians is based on a geo-referenced geographical map representation for easy visualization purposes. To add a service task some basic information is needed. At least the customer's geographic location needs to be known in the form of an address or geo-location (GIS-coordinates). An expected service time, a time-window to start the service (from earliest to latest arrival time) and required technician skills to perform the task can be additionally specified. For each available technician the application requires start and end locations (typically it is the technician's home or company address) as well as the earliest departure and latest arrival time (e.g., beginning and end time of the work-shift). Optionally, maximum capacities or skills can be further added. All data can be imported from Excel/CSV, manually entered or programmatically provided by a REST API. The application data interface is depicted in Figure 2.

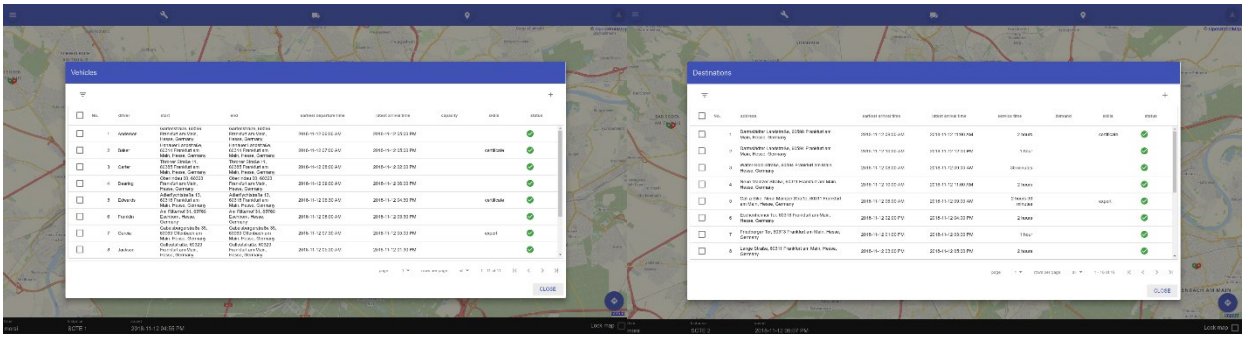

Figure 2 - Application Data Input

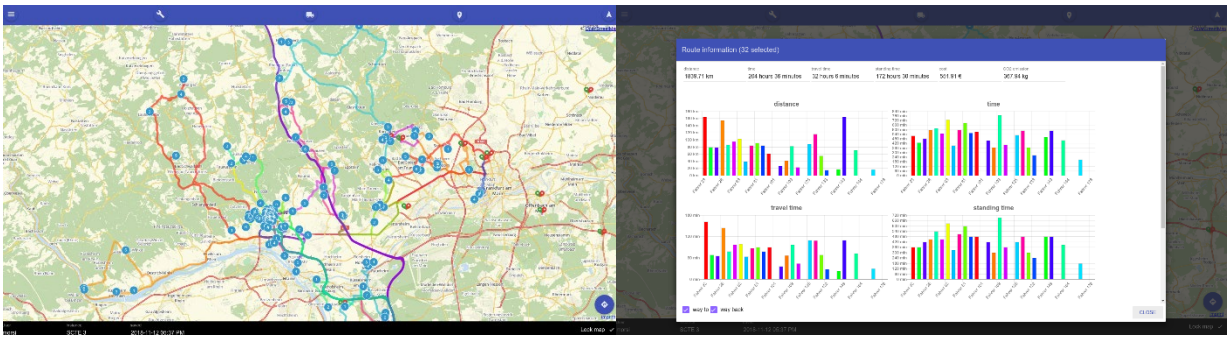

Figure 3 - Application Data Output

Travel distances/times matrices can be computed using OpenStreetMap or Google Maps. Final routes and schedules of technicians can be visualized on the map and evaluated by the user according to different key metrics such as travel time/distance/costs of service/waiting time (cf. Figure 3). Computed routes can be sent to the technicians' smartphones for route navigation, e.g., via Google Maps. Hence, direct incorporation of calculated routes into vehicle navigation systems is an option. If tracking of technicians is technically and legally possible one can think of a scenario in which current locations are sent back to the application and incorporated into the underlying geographic information system (GIS). Optimization is constantly triggered to reoptimize the schedules and routes.

2.2. The Mathematical Model

The first step towards applying MO to a specific application is to formulate a mathematical problem description. In case of vehicle routing problems [4], one of the best MO formulations is based on introducing the set of customers C to be served and also the set of all possible routes R to serve customers or subsets of customers. In this model the costs of a route $r \in R$ are denoted by c_r, where costs might be parameters such as e.g., travel time, distance or real costs. The number of visits to customer $i \in C$ on route $r \in R$ is denoted by a_{ir} (a_{ir} either zero or one). The binary decision variables are given by $x_r, r \in R$, where $x_r = 1$ if route r is selected and $x_r = 0$ otherwise.

$$\min \sum_{r \in R} c_r x_r$$

$$s.t. \sum_{r \in R} a_{ir} x_r = 1 \qquad \forall i \in C$$

$$x_r \in \{0,1\} \quad \forall r \in R.$$

This model is known as a set partitioning formulation. The first line defines the objective function, which is to minimize the total costs of all selected routes. The second row defines the first type of constraints, namely to ensure that every customer is visited exactly once. Finally, the last type of constraint forces the variables to be binary, i.e., to either fully select or deselect a route. To reflect more realistic scenarios, as in the case of scheduling field service technicians, additional constraints are necessary, such as a limited number of staff or to reflect inhomogeneous tasks and technician skills.

2.3. The Solution Approach

Both commercial and noncommercial standard software tools are available to tackle particular classes of mathematical optimization problems, e.g., when the objective function is linear and only linear constraints

and integer restrictions are present. The success of applying such standard software heavily depends on the mathematical formulation itself. This is due to more or less obvious differences, such as the number of variables/constraints and numerically inequivalent representations as well as to a hidden quality feature known as the "strength of formulation."

The VRP set partitioning formulation is very strong and hence a good formulation, but it has a very large number of variables due to the generally exponential size of the number of possible routes |R|. With no further route restrictions, such as route length or time-windows, there are $\sum_{k=0}^{n} \binom{n}{k} k!$ variables for n customers, which yields 9864101 variables for as little as n=10. Nonetheless, it is misleading to think that a model with such a large number of variables has no chance of being solved. It even might be one of the best performing MO-models to solve the VRP exactly. To overcome this drawback, an MO-technique known as column generation (CG) is being applied, meaning we start with a few variables and dynamically generate more if they have the potential to improve the objective function.

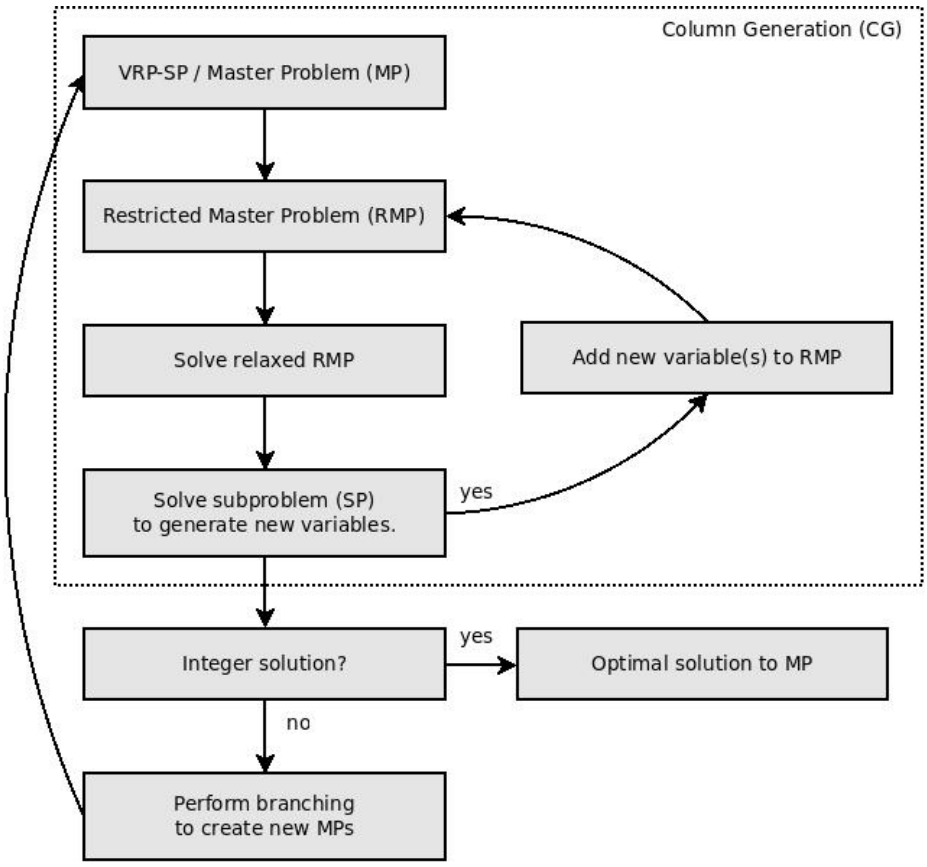

Figure 4 - Branch & Price Solution Approach to VRP

To solve the set-partitioning formulation, the branch-and-price (BP) methodology depicted in Figure 4 is applied, which is a combination of a branch-and-bound approach and column generation [5]. In the context of column generation (CG) the initial problem is typically known as the master problem (MP). Since the MP has a huge number of variables, CG is applied to solve the problem. We choose an initial

set of variables, e.g., derived from a heuristic, to set up the restricted master problem (RMP); omitted variables are implicitly handled as set to zero. In a second step, the restriction that variables must take binary assignments is relaxed to accept fractional values within the interval $[0; 1]$. This way the complexity arising from integer constraints is being deferred and we initially obtain a purely linear problem with a moderate number of variables that can be solved with a standard approach for linear problems, such as the simplex method. After the RMP has been solved, a so-called pricing subproblem (SP) is started to compute one or more new variables to augment the RMP. If a new variable having the potential to improve the objective function value is found, this variable is added to the RMP and a new iteration is initialized. Otherwise, if no augmenting variable is found, an optimal solution to MP has been computed. If the optimal solution to MP satisfies integrality ($x_r = 1$ or $x_r = 0$ for all $r \in R$) then the optimal solution to the VRP has been determined. Otherwise, the optimal solution to the MP gives only a lower bound on the optimal solution to the VRP. In that case branching is performed, i.e., a fractional variable $x_{\tilde{r}} \notin \{0,1\}$ is chosen and two new problems are created and recursively solved—one with the additional restriction $x_{\tilde{r}} = 0$ and one with $x_{\tilde{r}} = 1$.

2.4. Results

In the following, the proposed algorithm is being compared to both the schedules computed by experienced dispatchers and those computed by jsprit [6], an open source metaheuristic solver that constitutes the backend of several widely used commercial VRP software packages. For this case study we used real-world data from a network service company providing service technicians to network operators. Altogether the schedules of 23 different workdays have been optimized and compared. On each of the days between 21 and 37 technicians had to complete up to 162 service tasks within predefined time-windows ranging from 30 minutes to eight hours. All computations have been performed single-threaded on a Notebook with four Intel® Core™ i7-4702HQ CPU processors running at 2.20 GHz and 16 GB of main memory. As key performance metrics, the total travel time of all service technicians, the number of technicians needed to perform a given set of service tasks and the time needed to compute a schedule are being compared.

Compared to the manually determined schedules an average reduction of 27% in total travel time is being achieved and we were able to perform 100% of the service tasks on time with only 89% of the involved technicians. In other words, about 11% more service tasks could have been performed with the same personnel. If we additionally drop the time-window restrictions and let the algorithm compute optimal time-windows, e.g., as an initial suggestion computed some days ahead, the total travel time drops by roughly 41% and additional capacities of about 25% are being gained. Of course, not all of these optimized time-windows would have been accepted by the customers and the schedule would have been re-optimized several times with various changes in time-windows. After all, the truth lies somewhere in between 27% and 41% average reduction in total travel time and in between 11% and 25% of additional capacities, cf., Figure 5 and Figure 6.

Next, with an average computing time of approximately three seconds, the proposed algorithm is about five times faster than jsprit and the computed schedules result in consistently lower total technician travel times. On average these are four hours less.

If we allow for 30 minutes computing time in both algorithms to further improve their initially computed schedules, it turns out that the algorithm computes a schedule at least as good as the best schedule found by jsprit about 114 times faster. Vice versa, our initial schedule is better than the best solution found by

jsprit in 26% of the cases. In the remaining cases, jsprit is about 135 times slower in computing a schedule that is as least as good as our initial solution, cf., Figure 7.

As mentioned earlier, the proposed algorithm and solution always maintains a valid lower bound on the total travel time. This finally allows to conclude that for our algorithm on average 60 seconds of computing time are sufficient to obtain schedules that are optimal (meaning no better schedule exists) up to one hour of total travel time.

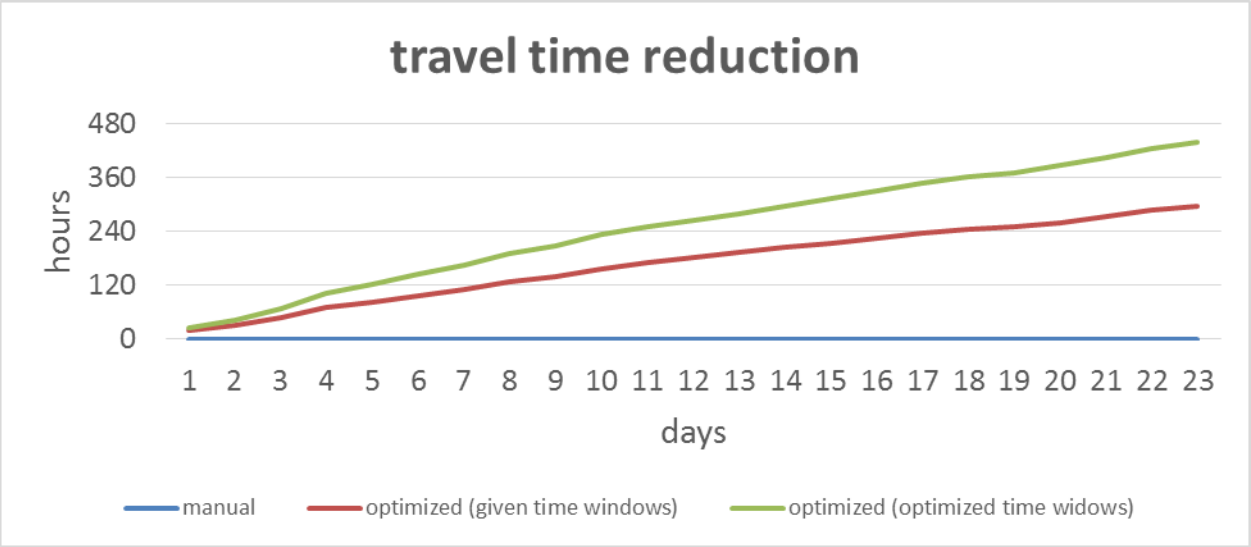

Figure 5 - Reduction of Travel Time due to Optimization

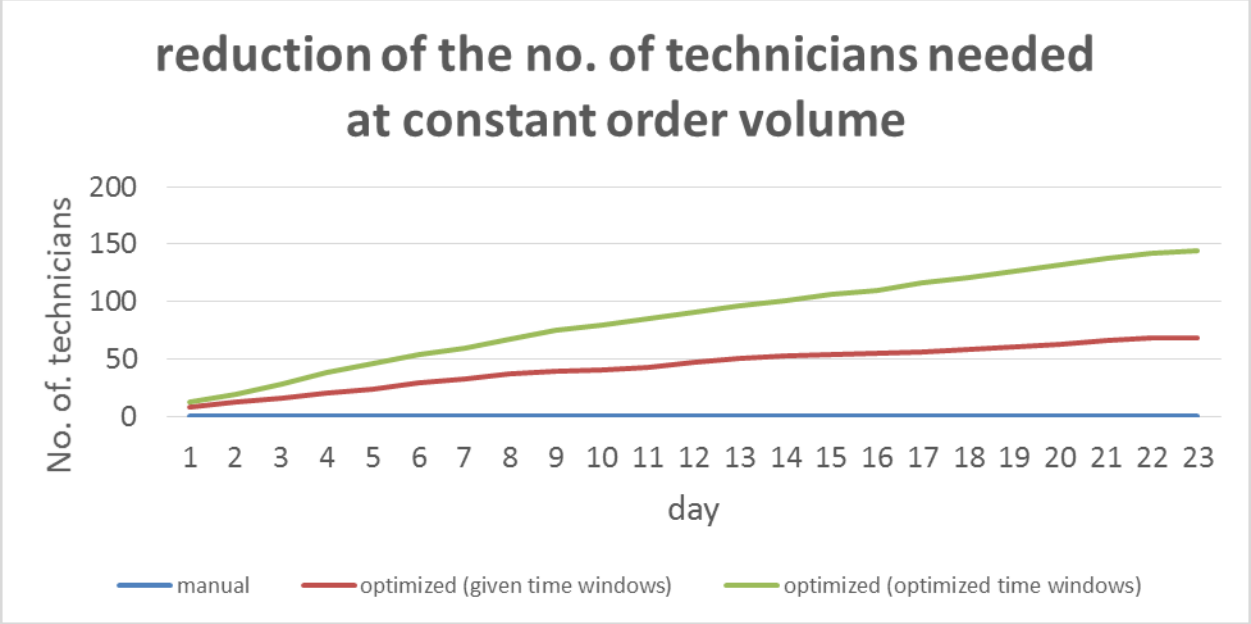

Figure 6 - Increased Capacities due to Optimization

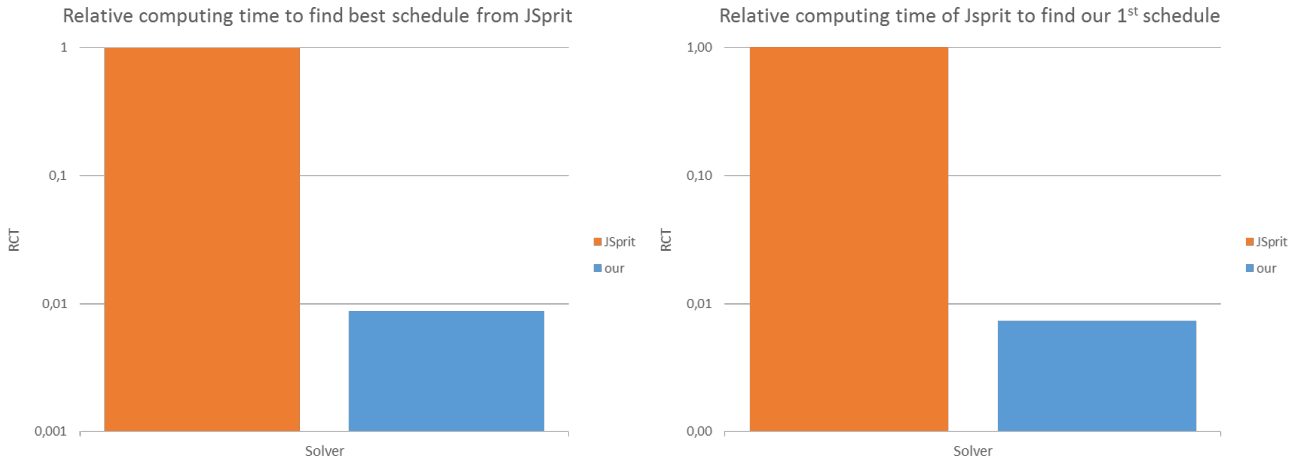

Figure 7 - Comparison of Relative Computing Times

3. Conclusions

Next to data analytics methods such as machine learning, prescriptive analytical tools—namely mathematical optimization—have numerous applications in the cable industry. In this contribution, scheduling of field service technicians in cable access networks was pointed out as an application and it was demonstrated how mathematical optimization can be applied to gain significant improvement as well as cost reduction while computing optimized schedules. In combination with geoinformation and geographical information systems, mathematical optimization can be applied to help streamline field service operations and add to the reduction of cost in network service and operation.

4. Abbreviations and Definitions

4.1. Abbreviations

BP	branch-and-price
CG	column generation
GIS	geographical information system
ML	machine learning
MO	mathematical optimization
MP	master problem
PNM	proactive network maintenance
REST-API	representational state transfer application programming interface
RMP	restricted master problem
SP	subproblem
VRP	vehicle routing problem

4.2. Definitions

branch-and-bound algorithm	An exact mathematical solution approach for a variety of mathematical optimization problems. It delivers an optimal solution and guarantees quality. The solution space is employed for logical branching and relaxation (bounding) to eliminate suboptimal solutions early in the process.
heuristic	Mathematical approach to solve an optimization problem, not including a guarantee for quality. A solution can show to be non-optimal or non-existent, albeit a solution may exist.
representational state transfer API	Interface that allows programming access to a service (here: REST-API defines how data is transferred to the web-service, how calculations are initiated and how the calculated routes are being returned to the user.)
simplex method	An exact mathematical approach to solving linear optimization problems.

5. Bibliography and References

[1] K. Sundaresan, N. Metts, G. White and A. Cabellos-Aparicio, "Applications of Machine Learning in Cable Access Networks," NCTA, 2016.

[2] G. White and K. Sundaresan, "DOCSIS 3.1 Profile Management Application and Algorithms," NCTA, 2016.

[3] Bixby, Robert E.; Gurobi Optimization, "Computational Progress in Linear and Mixed Integer Programming," 03-06 May 2017. [Online]. Available: https://ecco2017.euro-online.org/en/resources/files/invited-speakers/robertbixby.pdf. [Accessed 16 November 2018].

[4] M. Desrochers, J. Desrosiers and M. Solomon, "A New Optimization Algorithm for the Vehicle Routing Problem with Time Windows," *Operations Research, Vol. 40, No. 2,* pp. 342-354, 1992.

[5] G. Desaulniers, J. Desrosiers und M. Solomon, Column Generation, Springer US, 2005.

[6] "jsprit," [Online]. Available: https://github.com/graphhopper/jsprit. [Accessed 16 November 2018].

Making Cable Networks Ready for the Future Through Operations Benchmarking

A Technical Paper prepared for SCTE•ISBE by

Ronald Hasenberger, Senior Consultant, Bell Labs Consulting, SCTE•ISBE Member
Leonard-Bernstein-Straße 10
A-1220 Vienna
ronald.hasenberger@bell-labs-consulting.com

Michal Roszkiewicz, Cloud Automation Lead, Nokia, SCTE•ISBE Member
Rodziny Hiszpańskich 8
PL-02-685 Warszawa
michal.roszkiewicz@nokia.com

Martin Glapa, Partner, Bell Labs Fellow, Bell Labs Consulting, SCTE•ISBE Member
martin.glapa@bell-labs-consulting.com

Table of Contents

List of Figures

1. Introduction

The current telecommunications business is characterized by flat revenues and rising network demands stressing costs that are driving fierce competition between multiple system operators (MSOs) and communication service providers (CSPs) (both called "CSP" in the rest of this paper). A high percentage of cost is in network operational expenses (OpEx), which establishes a strong driver for operations optimization initiatives. Benchmarking is the key optimization enabler, helping to identify best practices and improvement areas with variations up to and larger than 50% observed between CSPs in selective network domains[1]. These quantitative representations of business performance are highly valued by CSP decision makers to help prioritize optimization initiatives.

The major challenge in quantitative benchmarking is obtaining comparable data sets. The comparisons are not straightforward because of differences in business environments. Such differentiation factors like geography, urbanization, infrastructure maturity level, local economic factors and others impact the business performance of CSP operations, making direct comparisons difficult. Therefore, normalization of measurements and performance indicators[2] (PIs) is necessary.

This paper provides a survey of measurement areas applicable to CSPs. A classification of measurements is proposed, along with mapping of differentiation factors into normalization formulas. Drill-down dimensions of the normalization outcome and synergies between cost modeling and efficiency benchmarking are discussed.

2. About Benchmarking

Essentially, benchmarking involves comparing the benchmarked entities against a reference value. As long as the benchmarked entities are simple, i.e., defined by a number of comparable, simple attributes, this operation may be done in a straightforward manner.

In the context of this paper, the benchmarked entities are either the operations organization of a CSP or managed services projects (both are referred to as Projects in this paper). Projects are very complex entities, because they are described by such attributes as scope (representation of functional content of a Project) and service level agreements (SLAs)[3]. Direct comparisons applied for Projects will unavoidably lead to errors, hence the need of normalization arises.

---

[1] Based on Bell Labs Consulting analysis.

[2] Performance Indicators can also be called "metrics." Both are in most cases the result of the combination of multiple measurements, which is a first step towards comparability by taking care of basic driver relations. As an example: the number of incidents in a project can be expected to scale linearly with the number of active network elements; number of incidents, number of network elements being considered measurements; number of incidents per network element becoming a meaningful performance indicator.

[3] SLAs represent quality levels with regards to network performance, timeliness, etc. demanded either by the CSP's customers or by the CSP from the managed services provider, whereby the latter will typically be in response to the former.
Requirements put up by the customers of a CSP, especially in the retail market, are not always explicitly agreed but can be the result of "typical expectations" in a certain market or demands of authorities.

There are two main factors that prohibit direct comparisons of Projects:

1. **Project internal design**
   Projects are very complex collections of resources and processes. Since resources and processes vary, their measurements and data are not identical. Therefore, the data reported from Projects cannot be compared directly and needs alignment.
2. **Project environment**
   Each Project is surrounded by a certain environment, characterized by regulatory environment, CSP strategy, customer contract, geography, industrialization, economy, etc. Two identically designed Projects would perform differently if embedded in two different environments. Normalization is necessary for a fair comparison.

In this paper, a holistic benchmarking system proposal is discussed. The goal is to enable comparability of Projects across the globe by identifying best practices amongst Projects (to learn from them), as well as Projects with room for improvement.

The proposed benchmarking system is based on three foundational concepts:

1. **Measurement definition framework**
   This is a structured approach to measurement definition, relying on the measurement of objects, their attributes and relationships. The main advantage of this concept is that it allows for much more formal and unambiguous definition of basic measurements, and hence helps to limit misalignments between Projects.
2. **Alignment**
   This concept is introduced to eliminate differences that stem from internal designs of various Projects. The intention is to eliminate misalignments due to differences in tools, data collection methods, reporting schemes, processes, organization structures, etc., by means of processing the measurement data.
3. **Normalization**
   This concept deals with differences coming from various environments surrounding specific Projects. Executed on aligned measurements, normalization allows mitigating the environmental impact to the Projects, hence improving Projects mutual comparability.

These three foundational concepts are discussed in more detail in the following sections.

The paper concludes with a description of the measurement and benchmarking blueprint reference system.

3. Measurement Definition Framework

The main goal of this framework is to facilitate definition of measurements. Traditionally, measurements are defined as textual descriptions and formulas. These descriptions are then translated into data collection instructions (in case of manual collection) or database queries (in case of tool-based collection).

The proposed measurement definition framework allows expressing measurement definitions as objects and their relationships, thus creating a formal relationship between measurement definition and measurement collection.

Three measurement object classes are defined:

1. Managed resources
2. Managed process artifacts
3. Managed activities

3.1. Managed Resources

The concept of managed resources was derived from the resource breakdown structure described in [1].

The real-world prototypes of these objects are tangible and quasi-static. The main attribute to the objects of this class is inventory (number of elements) calculated over an observation period. The main purpose of this class of measured objects is the volumetric reference for PI calculation. They may also be used as a basis for cost calculations.

An example of a managed resource is a "network element" (NE), e.g. a router, a switch—but could also be a wireless eNodeB[4]. The number of NEs in a given observation period (e.g., month or quarter) is a reference for such a benchmark like "number of alarms per NE," which is an indicator of network stability. Other managed resource examples are employee, vehicle, test instrument and site.

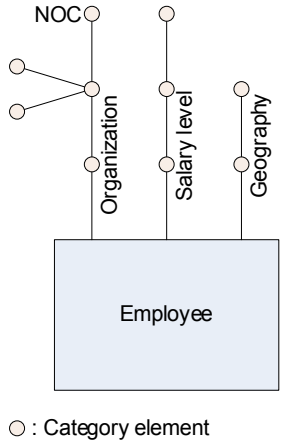

Figure 1 - Managed Resource Example

A key concept related to the managed resources is categorization. Each managed resource may be categorized along many dimensions. Categorization allows modeling the real-life properties of an object, e.g., employees may be categorized as per skill level, salary level, competences, organization, etc. Categorization complexity may vary from a very simple, one dimensional list of category elements, through tree-like structures, up to complex multi-dimensional structures.

[4] eNodeB (enhanced NodeB) is the name of the wireless basestation for LTE.

Considering the example from Figure 1 some of the above-mentioned categorization concepts are represented:

- Lists of categories
    - Salary level (low, medium, high)
    - Geography (typically on the level of regions, like Europe, Middle East, Africa - EMEA, Asia Pacific – APAC, …)
- Tree like structure in organization
    - Network operations center (NOC)
    - Technical support
    - Field services
    - Customer contact center
    - Engineering
        - Radio access network (RAN)
        - Core network
        - Billing
        - ….

Managed resources are represented as inventory records in tools.

3.2. Managed Process Artifacts

This concept was derived from artifact-centric process modeling theory, originally proposed in [2]. The most important advantage of using managed process artifacts is the ability to express processes as a set of business rules. A framework for such representation is discussed in [3].

The real-world prototypes of these objects are volatile and dynamic. They are also stateful, i.e., each of these objects is described by a specific state graph. These objects are generated (created) at discrete points in time, pass through their state graph, and finally disappear. The measured attributes for managed process artifacts are:

1. State counters: number of artifact instances that reached a given state (e.g., number of trouble tickets (TTs) that entered the “restored” state)
2. Time spent at a state: a representation of idle time when no change occurs to the artifact
3. State arrival time: a time stamp measured to allow for cycle time measurement as described in the statistical process control (SPC) approach [4]

An example of a managed process artifact is a TT. The counter of new tickets shows the workload offered to the resolution team. The discrete time distribution for time spent at various states shows where potential improvement opportunities are located and the “work in progress” queue.

Other examples are field work order, alarm, sub-ticket and a call.

Managed process artifacts are represented as data objects in tools.

3.3. Managed Activities

These abstract objects are process reference points. There is no real-world prototype to these objects. Managed activities may be considered as a representation of effort (which may be human effort, processing effort, or usage of other means, like car mileage) that is interesting to characterize the process artifact from a measurement and benchmarking point of view.

Managed activities are closely linked with process artifacts. In principle, activity completion results in a state graph transition of the related managed process artifact. As an example, completion of the activity called “restoration” results in transition of TT from “analyzed” to “restored.”

Other examples are dispatch, travel, assignment and site visit.

Managed activities are represented as attributes / flags / data records in tools[5].

3.4. Measurements and Performance Indicators

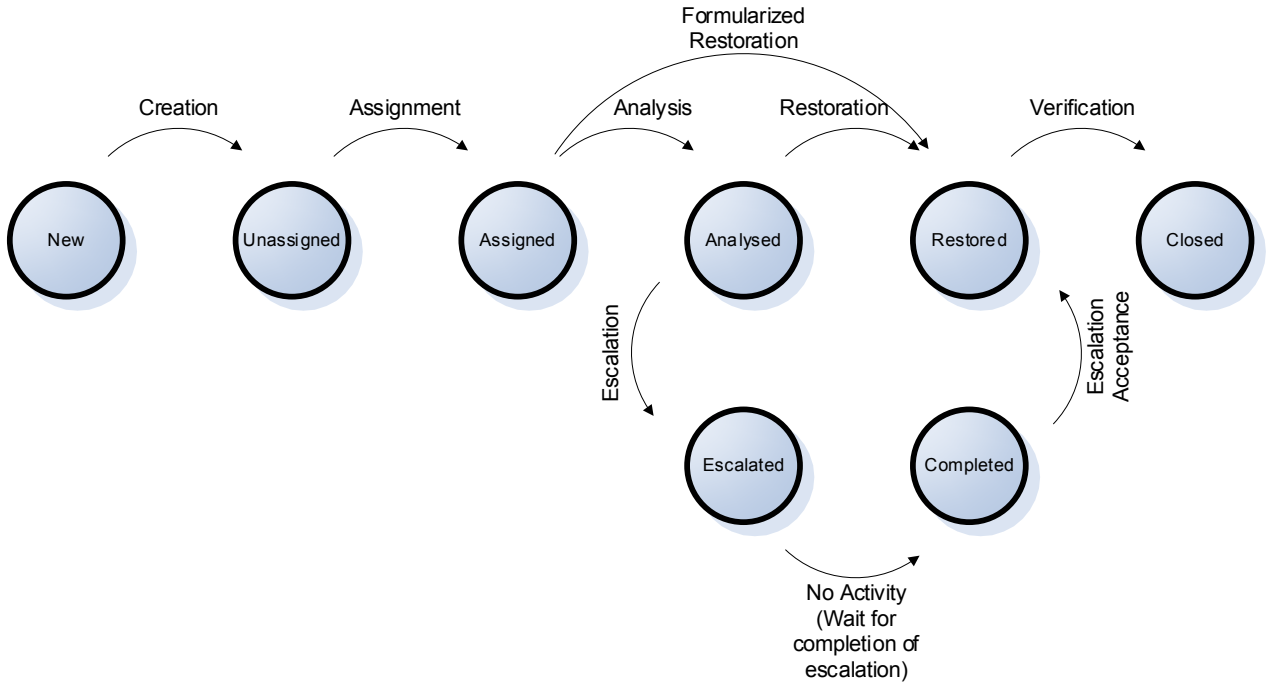

Figure 2 - Simplified Trouble Ticket State Graph

Simple measurements may be defined in a straightforward manner such as read-out the value of a required managed object attribute. As an example, to know the number of TTs restored within an observation period in a Project, the number of TTs that entered the restored state may be counted. This is depicted in Figure 2.

However, more complex measurements are often needed to build meaningful PIs. Examples may include the number of TTs resolved for a given technology, number of TTs resolved by a given team, number of alarms generated by a specific technology, etc. Such measurements span more than a single managed object attribute; hence relationships are required to define them.

[5] Examples for tools where these managed activities may be represented are a workflow management or a ticketing tool.

Most of the measurements are defined on a single instance of a managed process artifact. To determine the measurements described here, they need to be aggregated. The aggregation is usually performed by means of simple statistical aggregation, e.g., taking the mean or the median of the values. If no additional pre-processing / data cleaning steps are being performed then it is generally better to use the median, as this is the aggregation method that is more resilient against outliers (which may have been caused by measurement errors).

3.5. Relationships

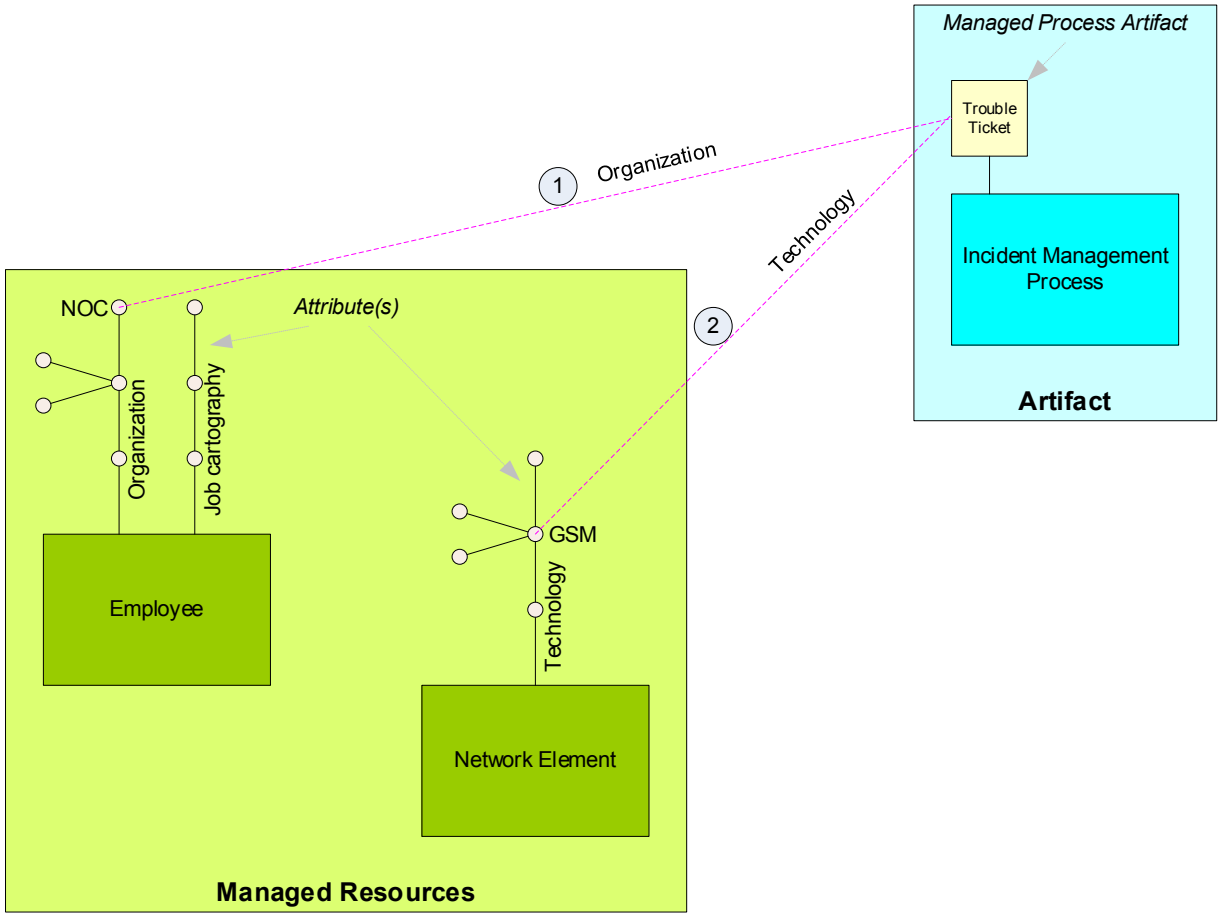

Figure 3 - Measurement Structure / Relationships

The relationship concept is demonstrated in Figure 3. In this example, managed process artifact ⇔ managed resource relationship is presented. Each managed process artifact may be assigned to one or more managed resources using resource categorization as a relationship anchor point.

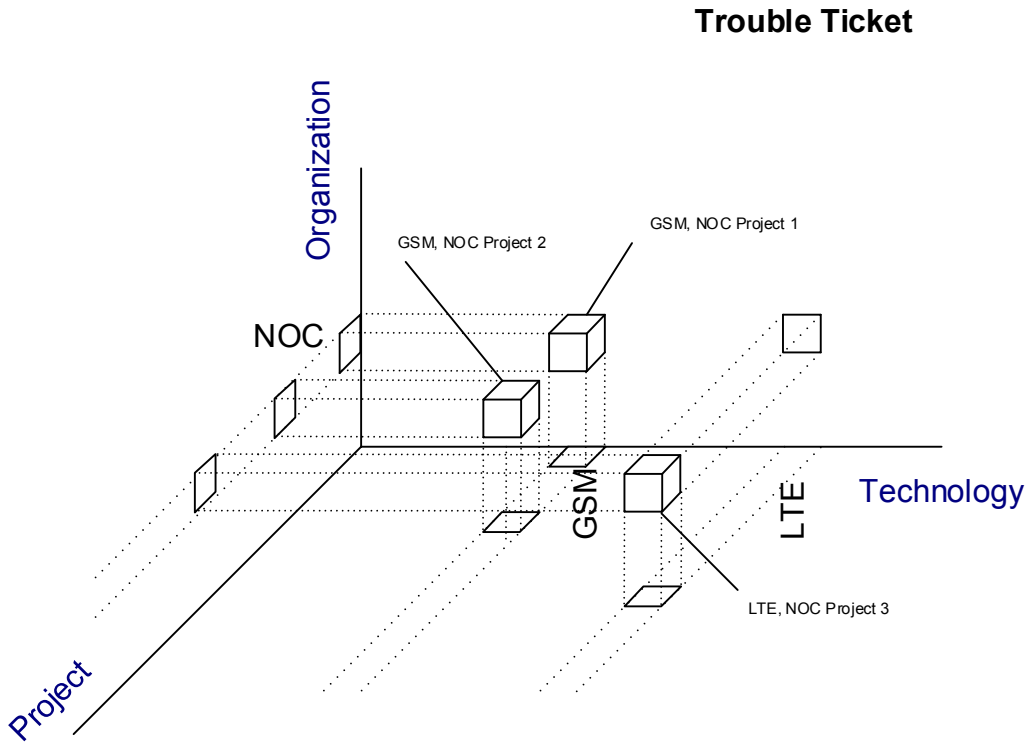

Figure 4 - Measurement Categories

A complex measurement that aims to count TTs related to GSM technology restored by an organization called a "NOC" is graphically demonstrated in Figure 4 for Projects 1 and 2, while Project 3 is having a similar origin but related to LTE as technology, thus putting this project into a different spot on the organization / technology plane which has an impact on comparability.

The relationships may be used to span an n-dimensional matrix (in the example from Figure 4, n=2 to ease graphical representation in the further steps of this paper) where each point is exactly defining a measurement category which may be interpreted as an object that describes the specifics of this measurement. When introducing the project as another dimension it becomes possible to assign project specifics to each of those measurement objects (see Figure 4). Examples for further dimensions could be:

1. Equipment vendor
2. Region
3. Network size

This structure creates a powerful tool that may be used to clearly identify different measurement categories and structure them in a way that allows easy assignment to alignment and normalization requirements.

The previous example is still a relatively simple case of a measurement with two relationships. It was selected for graphical representation of the concept. In practice, production measurements are based on a higher number of relationships. For example, an interesting measurement could be the joint effort of a group of job profiles working together to maintain a set of NE types (forming a technology) in the NOC (i.e., with TTs) and in the field (i.e., with field work orders). A traditional definition based on textual description and enumerative lists would be very complex. Thanks to the proposed Project decomposition into resources, artifacts and activities, it is possible to construct an object-oriented representation of the measured Project and to express such measurements in a pseudo-language understood either by business analysts or by administrators of a business intelligence (BI) system.

4. Alignment

The measurement definition framework described in the previous section allows for an unambiguous definition. However, real implementation of measurements always depends on a Project's internal design. Measurement frameworks like TM-Forum's Information Framework (SID) [6] are helping to avoid some ambiguity, but usually some aspects stay specific to Projects.

There are three reasons that cause measurement misalignment, all of them stemming from CSP definitions:

1. **Resources organization**
   As per the measurement definition framework, all resources are categorized. These categorizations may differ to a certain extent between Projects. For example, technologies may be split according to CSP-specific technology directory, job catalog classifications may not be aligned, etc.
2. **Processes definition**
   The state graph representation of the measured process allows for decoupling of process and measurements definition. This means that process redefinitions do not impact the measurements definition as long as the states and transitions are not changed. However, in some Projects the process deviates from the standard definition so much that a global state/transition model is no longer valid. For example, two states defined in the global model may be represented by a single state as per a Project specific process definition, making direct reporting on each of the original two state counters impossible.
3. **Tools capabilities**
   The setup of the measurements also depends on the capabilities of the tools that are used. This is a cause of major misalignments, e.g., lack of required relationships between measured objects or data sourcing capabilities different than envisaged in the global model.

Alignment operations may be categorized as:

1. **Interpolation**
   Creation of new measurement data points within the range of discrete set of known measurement data points.
2. **Extrapolation**
   Creation of new measurement data points outside of the range of discrete set of known measurement data points.
3. **Estimation**
   Creation of new measurement data points based on another set of measurement data points.

A simple example for extrapolation alignment is the case of a monthly aggregate to be calculated with only weekly measurement aggregates being available as raw data. In this case the raw data needs to be extrapolated to cover the whole month (based on the weeks best covering the time range).

Let us consider a PI that needs the number of work orders calculated for April 2018, whereas only weekly data samples are available.

April 2018 spans from Sunday (week 13) to Monday (week 18), thus embracing weeks from #14 to #17.

Therefore, the number of work orders can be linearly extrapolated according to the following formula:

$$\#WorkOrder_{Apr2018} = \left(1 + {}^{2}/_{28}\right) \cdot \sum_{n=14}^{17} \#WorkOrder_{Week_n,Apr2018}$$

In this formula, ${}^{2}/_{28}$ is representing the fact that two days out of the 30 are not covered in the 28 days (four weeks) that are included in the weekly data samples summation. Obviously, the alignment assumes flat distribution of the work orders over all days of the week including Saturday and Sunday. If such an alignment is to be done for consecutive months additional care needs to be taken to avoid modification of the overall number of work orders for longer lasting aggregates (quarterly, yearly) by using this process. This leads to an exclusion rule to be applied together with this alignment formula.

Estimation alignment is used if certain measurements cannot be gathered from the Project. An example could be issues resolved immediately by the agents of the NOC without tracking them in the ticketing system as formal TTs. It is possible to recover the missing numbers by means of estimation from experience gathered with other Projects.

In case of estimation, special care is taken with regards to the usage of the estimated data. If a PI is built using only estimated data, it will bring the same value that the Projects used as a reference to create the estimate. A PI based only on estimated measurements is most probably worthless, since it is very similar to directly estimating the PI. If the estimated measurement constitutes only a part of what data is used for PI calculation, comparability may be significantly increased. To stick to the example discussed previously, the PI showing the percentage of TTs solved by the NOC agent without assignment to another organization based on an estimated number of tickets closed by the NOC agent is most probably worthless because it has no foundation in the properties of the observed Project, because it is calculated purely on the estimation applied. If, under the same circumstances, the target is to calculate the percentage of TTs going to the field it makes sense to include an estimate of the number of TTs closed by the NOC agent without assignment to another organization in the denominator.

5. Normalization

The main reasons to do benchmarking for Projects are the identification of best practice and improvement possibilities. Put in other words, the target is to compare the Projects according to their “core performance” excluding effects that cannot be influenced by the Projects.

That leads to the requirement of normalization, where all these non-influenceable effects are normalized. For example, all projects are virtually brought to the same reference environment and compared. It is therewith important to define the environment that all the projects are being brought into.

Examples for areas where normalization is required are:

1. **Strategy**
   Different strategies lead to focusing on different aspects, thus potentially leading to pushing one PI at the cost of another one.
2. **Scope**
   Differences in scope obviously lead to different efforts in running a network.
3. **Site / Task density**
   Different site / task densities have a significant effect on field services efforts because it affects travel time.
4. **SLAs**
   Different SLAs may not necessarily change the overall amount of work required to run the network, but they will influence the required flexibility in reacting, which may cause additional efforts.
   Especially for field service, tighter SLAs may impede the possibility for optimal routing of the field engineer therewith influencing the drive time.
5. **Legal regulation**
   Legal regulations may have an influence on the benchmarking by e.g., influencing the number of people required to complete a certain job (e.g., due to safety requirements).
6. **National infrastructure capabilities**
   The national infrastructure capabilities in the form of the transportation infrastructure (streets) have an obvious impact on the field service drive times. On a similar note the availability of the electricity supply has an obvious influence on the reliability of any kind of telecommunication network, even in the case of backup power installed on site.
7. **Network condition / Incident probability**
   The condition of the network leads to different incident probabilities as well as to a different classification of incidents (in a highly redundant network the failure of a single device may be a rather low-level incident; in a non-redundant network the same incident may cause cessation of the service in large parts of the network which causes higher severity / urgency of the [technically] identical incident).
8. **General environmental conditions**
   Some environmental conditions may influence incident frequency (e.g., humidity, storms, …) as well as cause additional effort for incident solving. To some extent unfavorable environmental conditions (e.g., heat, humidity) are requiring the usage of additional equipment (e.g., air conditioning units) which by themselves are additional units that can and will break, therewith increasing the amount of activities required to keep the network fully operational.
9. **Purchasing power parity**
   As soon as any monetary measures are included in the measurements the money needs to be translated as well. The obvious approach here is the calculation according to exchange rates, but this is only half of the truth. The other half is represented by the purchasing power parity (PPP, see [8]).

Effects of Project inherent choices by the CSP (like the toolset being employed to support the project) are deliberately left out of the normalizing factors because part of the expected results out of benchmarking is to identify the optimum setup in this regard.

The following nomenclature is used in this section:

- N_x : normalization factor for area x; the normalization factor is defined as the factor to multiply the effort of the benchmarked project (BP) to get the effort, normalized for the effects of the area x.
  The effort, normalized for a certain area x, may be calculated as $E_{norm,x} = E_{BP} \cdot N_x$, the fully normalized effort as $E_{norm} = E_{BP} \cdot \prod_{x:allAreas} N_x$
- C_x : count of events
- $EE(\mathbf{S}_{\mathbf{union}}, NW, Env)$: estimated effort; estimation e.g., based on modeling; controlled by
  - Scope
  - Network properties (especially network size)
  - Environment
- $\mathbf{S}$: the vector of all scope elements present in the Project
- D : spatial density of sites and tasks (unit: sites / km^2 or tasks per day / km^2)
- p_{travel} : portion of the overall task related time that is devoted to traveling

Normalization is represented as a vector of normalization factors, each of which covers a specific relevant area ($\mathbf{N} = (N_{Scope}, N_{Environment}, N_{Legal}, ..., N_{Density})$). The whole vector is as well assigned to each of the elements in the (n+1) dimensional matrix introduced just above.

Based on this structure it is easy to implement the overall normalization structure according to Figure 5.

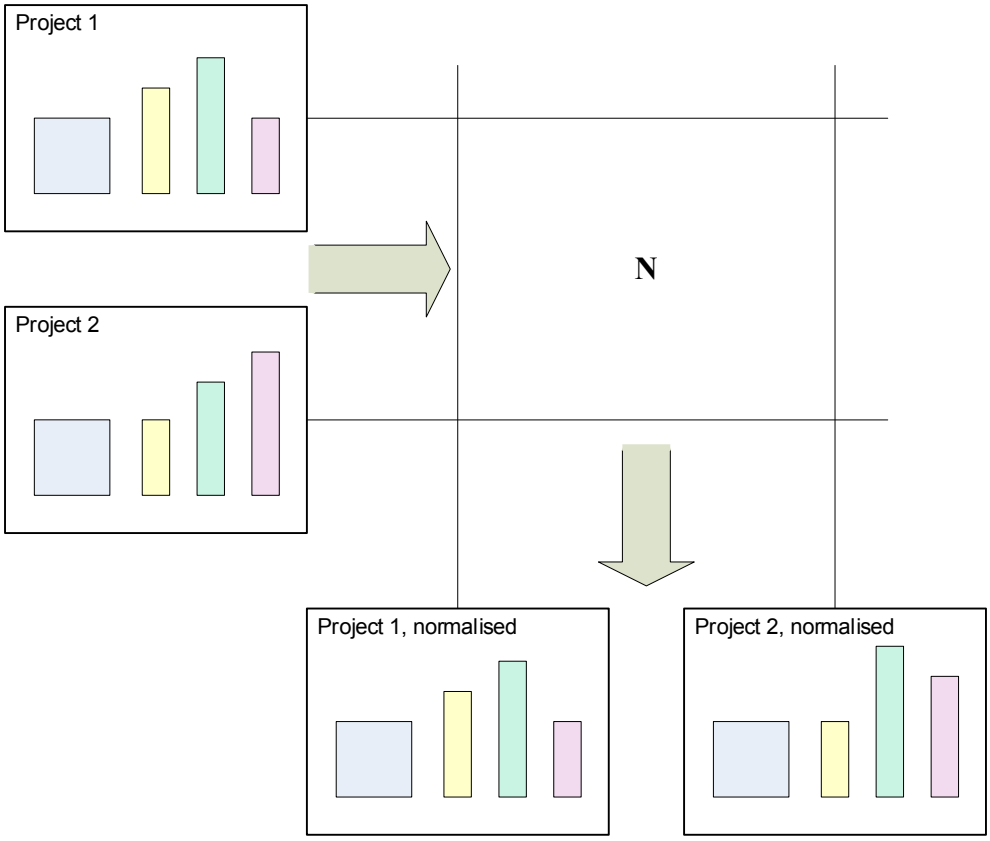

Figure 5 - Normalization

The normalization areas are discussed in detail in the following subsections.

5.1. Strategy

The strategy under which a Project operates may have a significant impact on some PIs. In most cases there is a tradeoff between different aspects, e.g., cost vs. quality. Dependent on a strategy decision, more or less focus may be put on either of these, which will become visible in the performance achieved for different PIs.

As a concrete example, take two projects, where one project follows a top-quality strategy while the other project follows a least cost strategy.

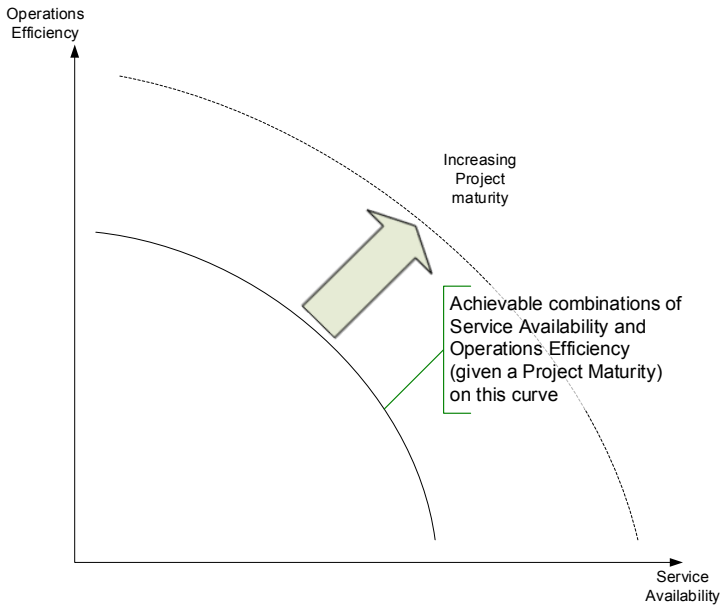

Figure 6 - Strategy Impact

Figure 6 provides a sample for the relation between cost and quality related metrics with different operating points being achieved dependent on the strategy applied in the Project, while keeping the Project's maturity constant. With increasing maturity of the Project, the whole curve is expected to move towards the upper right.

Project maturity can be impacted along dimensions such as

1. People (definition of roles, RACI, alignment of skills to roles, organization structure)
2. Processes (definition and adherence to processes)
3. Platforms (performance and alignment of platforms [tools] used to support operations)
4. PIs (availability and usage of relevant PIs)
5. Policy / data (data management and usage)

This is an effect that cannot be normalized. In some cases, it is possible to avoid the impact of strategy by moving to a higher level PI[6].

5.2. Scope

The normalization in the scope area occurs by an extrapolation of the scope of the BPs towards reference scope (which may be the superset / union of the scopes of all Projects to be benchmarked or a predefined "full Project" scope). This normalization occurs potentially separately for all other differentiated aspects (e.g., organization etc.).

Thus, the normalization may be expressed as follows:

[6] Assuming a hierarchy of PIs where high level PIs are broken down into more detailed ones. Some of the focusing aspects coming with the selection of a strategy may become invisible by moving towards higher level PIs.

$$N_{Scope} = \frac{EE_{Ref}}{EE_{BP}},$$

with EE_{Ref} being the estimated effort for the reference scope used (which may be the union of the scopes of all participated projects or a defined "full project" scope) and EE_{BP} being the estimated effort for the scope for the BP. Both efforts being estimated under the same border conditions, e.g., by using a standardized cost model.

$$EE_{Ref} = EE(\mathbf{S_{Ref}}, NW, Env)$$

$$EE_{BP} = EE(\mathbf{S_{BP}}, NW, Env)$$

It is necessary to estimate the different efforts using the same network and environmental variables for both scope variants since the scaling of the effort must be considered dependent on these aspects.

5.3. Site / Task Density

The site / task density is assumed to mainly influence field service drive time, leading to normalization being only foreseen for this area.

The average distance between different tasks scales approximately with the inverse of the square root of the site / task density in most circumstances (exceptions are coverage of geographical islands where each geographical island carries only a small number of sites and, "linear" coverage where the covered area stretches linearly, e.g., along a river or a coastline but lacks the perpendicular dimension).

Thus, the normalization for site / task density can be represented as:

$$N_{Density} = 1 + p_{travel} \cdot \left(\sqrt{\frac{D_{Ref}}{D_{BP}}} - 1 \right)$$

5.4. SLAs

For different SLAs no generalized method for normalization is currently available.

Different SLAs are expected to lead to only minor differences in effort required for solving the problem (regardless of problem severity and required timing for solving a problem it must be solved and this effort can be assumed to be the same). There will be differences with regards to

1. Number of people required to satisfy 24 / 7 availability
2. Loss of synergies
3. Additional drive time to fix a problem because an incident cannot be included in a route but must be handled separately to keep the SLA
4. Additional cost
    - Overtime payments
    - Payment for "on duty" services

5.5. Legal Regulation

For different legal regulations no method for normalization is currently available.

Legal regulation will cause effects predominantly in the following areas:

1. Number of people required to satisfy 24 / 7 availability
   In order to staff a function 24x7 a certain number of people need to be available to enable a legally admissible shift plan
2. Different safety regulations (includes differently executed safety regulations) cause different demand due to safety reasons by
   - Requirement for field engineers to work only in pairs in certain situation (e.g., height limits)
   - Limitations to work in certain circumstances, which may require duplicate travels to the problem site

5.6. National Infrastructure Capabilities

National infrastructure capabilities are relevant in three areas:

1. Telecommunication network
2. Energy supply
3. Travel conditions (street conditions / traffic)

5.6.1. Telecommunication / Transmission Network

Here the reliability of the transmission network is relevant for normalization, to the extent the transmission network is not under the Project's control. If not having the transmission network under control is a deliberate decision of the project, the resulting aspects should not be normalized.

To the extent the transmission network is not under the Project's control, the effort for rectifying transmission incidents is normalized according to the relation of the number of non-influenceable transmission incidents for this network ($C_{TxI,BP}$) in relation to the same category of incidents for the reference network ($C_{TxI,Ref}$):

$$N_{Tx} = \frac{C_{TxI,Ref}}{C_{TxI,BP}}$$

Inherent to this definition is the assumption that the average effort per incident will be the same for all different kinds of environments.

5.6.2. Energy Supply / Grid

The reliability of the energy supply is relevant for normalization, to the extent it is not under the Project's control (which is usually the case).

To the extent the energy supply is not under the Project's control, the effort for rectifying power incidents is normalized according to the relation of number of non-influenceable power incidents for this network ($C_{PI,BP}$) in relation to the same category of incidents for the reference network ($C_{PI,Ref}$):

$$N_P = \frac{C_{PI,Ref}}{C_{PI,BP}}$$

5.6.3. Travel Conditions (Street / Traffic)

Travel conditions have a similar influence on the overall effort as the site / task density. The common main factor employed is the achievable travel speed under the given circumstances. The resulting normalization is similar to the one for the site/task density:

$$N_{Travel} = 1 + p_{travel} \cdot \left(\frac{v_{Ref}}{v_{BP}} - 1\right)$$

This normalization does not account for potentially differing possibilities to optimize route planning, that is mainly influenced by the possibility of the field engineer to have a reasonable stock of spare parts, or to have the required spare parts delivered to the task location by an external company.

5.7. Network Condition / Incident Probability

The condition of the network (especially the degree of redundancy) has a significant impact on the severity of incidents.

Another aspect is the probability of incidents happening at all. Differences in this area may be caused by

1. Type of equipment
2. Age of the equipment
3. Maintenance of the equipment
4. Environmental conditions (especially temperature and humidity; handled in Section 5.9)

To the extent the aspects are not under the Project's control, the efforts are normalized according to the relation of number of non-influenceable incidents for this network ($C_{I,BP}$) in relation to the same category of incidents for the reference network ($C_{I,Ref}$):

$$N_{NC} = \frac{C_{I,Ref}}{C_{I,BP}}$$

5.8. Stability / Maturity of the Network

No method for normalization is currently known.

The amount of changes to the network (which is the intended interpretation of stability / maturity) has an obvious impact on the effort going into fulfillment activities.

Additionally, it usually appears that networks run more reliable if they run without "external disturbances" (changes of any kind).

5.9. General Environmental Conditions

There are a couple of "other" environmental conditions that may influence the failure rate of network equipment and the effort for correcting failures. Those aspects are difficult to handle globally and need to be considered on a per Project basis.

1. Environmental conditions (especially temperature and humidity)
    - Temperature has an obvious influence on the failure probability (Arrhenius' equation[7])
2. Different safety requirements cause different additional demand due to safety reasons by
    - Requirement for bodyguards to protect field engineers

6. Benchmarking

Global benchmarking means comparing many different projects embedded in various environments. On one hand, this poses the problem of comparability, discussed in depth in the previous sections. On the other hand, it is an opportunity to obtain certain reference values out of the entire population of Projects, and then to reuse these values for benchmarking specific Projects.

A useful methodology is to benchmark against the top quartile to identify candidates for best practice, and the bottom quartile to identify candidates for improvement. The potentially heavy number crunching to align and normalize PIs also creates a significant risk for introducing artifacts that need to be checked for. These checks are dependent on the "implementation maturity" of the whole process described in the paper, and are considered useful especially for the projects at the top and the bottom of the scale.

Assuming that there is one best practice[8] the next step would then be a check of what all the best performers are doing different than the remainder of the population, which would constitute reasonable candidates for best practice waiting to be implemented in the other projects. Identifying if there are consistent differences of best practice between structurally different environments (e.g., highly developed vs. emerging market) may be derived from a correlation between applied processes (process flavors) and environmental variables (at least those environmental variables going into the normalization because those are made explicit as part of the benchmarking process).

PIs in the lowest quartile of the benchmarked sample are candidates to address with improvement initiatives as the effort for improvement by a certain degree can be expected to be lower than the same improvement on a PI that is already top of the range ("law of diminishing returns").

Assuming that the improvement Projects lead to the desired effect raises the bar for the Projects in the next benchmarking round.

7. Cost (effort) Modeling

Benchmarking has the target to compare an existing operation with a reference therewith creating a statement about this Project. Cost modeling is the activity of estimating the effort and, mainly derived from the effort, the cost required to run a project, which may be done in a managed services context in

[7] Rule of thumb based on the Arrhenius' equation leads to the estimate of halving lifetime by increasing the temperature by 10 K (see also [9]).

[8] It cannot be taken as a given that there is a single best practice globally as there may be different best practices dependent on circumstances – e.g., what is best practice in a highly developed country doesn't need to be best practice in an emerging market country.

preparation for doing a bid for taking over work or when doing a resizing of an operator in preparation of launch or a major transformation. During cost modeling, only limited insight into the details of the project may be available.

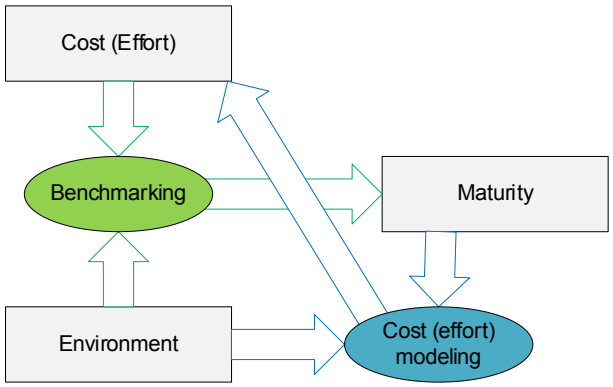

Figure 7 - Benchmarking vs. Cost (effort) modeling

Whereas benchmarking is an analytical exercise that uses observable properties to normalize, cost modeling is more a synthesizing exercise where properties that are not known in advance (Project internal design) need to be taken into consideration. The relationship between these activities is shown in Figure 7, reflecting the three main components that are present in both cases:

1. Cost (effort)
2. Environment
3. Maturity

While the environment is an input in the both activities, benchmarking is using cost (effort) to quantify maturity, cost (effort) modeling is using the expected Project maturity to determine an estimation of the cost (effort) required to run the Project.

Observing the similarities between the two domains of benchmarking and cost modeling the same concepts and similar structures can and should be utilized in both areas, which creates synergies in model creation and makes exchange/reuse between the domains easier.

The selection of a certain operating model (with corresponding investment in supporting platforms, process development and implementation, etc.) defines the maturity of the future operation which is an input into the cost modeling (see Figure 7).

The results of benchmarking may be used for cost modeling in the following areas:

1. Tuning of the expected efforts per event and occurrence parameters. This activity ensures that the cost modeling activity achieves high realism and provides efforts / costs that can be achieved in delivery.
2. Different families of environments / delivery scenarios may be identified out of analyzing similarities within the normalization requirements; i.e., if there is one cluster of Projects with similar / same normalization requirements and another cluster that is significantly different from the first one but is similar, it may be expected that these reflect specific environments and / or delivery scenarios. Within the cost modeling environment, it appears to be useful to provide “template parameterizations” for each of these scenarios to jump start the cost modeling for a specific Project.

In the other direction, properly defined cost models may be used for:

1. Estimation of unmeasured aspects of a currently analyzed Project. This may be necessary for alignment purposes, e.g., to estimate the field drive time[9] when the drive time is not measured, or to provide an objective drive time estimate, not dependent on potentially suboptimal dispatching behaviors within the Project.
2. Providing standardized basics for normalization purposes, e.g., creating a standardized measure for the drive time portion for field service activities to normalize for drive time effects.

Overall this leads to the following benchmarking-cost modeling synergy cases:

1. PIs utilized for benchmarking provide real life representation of project performance within a standardized environment. These PIs are used to derive the parameterization of the cost model
2. Clustering of normalization vectors $\mathbf{N}$ is used to identify families of environments (e.g., developed market / emerging market) / delivery scenarios. These families may be represented by parameterization templates for the families.
3. The cost model is used to estimate unmeasured aspects of running Projects.
4. The cost model is used to support normalization efforts in benchmarking.

8. Measurements and Benchmarking System Blueprint

The foundation concepts discussed in this paper may be applied in practice to construct a measurement and benchmarking system. The measurement and benchmarking system blueprint is proposed in Figure 8.

[9] Potentially different drive time estimates apply for long term plannable activities like preventive maintenance or short-term plannable activities that typically occur as part of assurance or fulfillment.

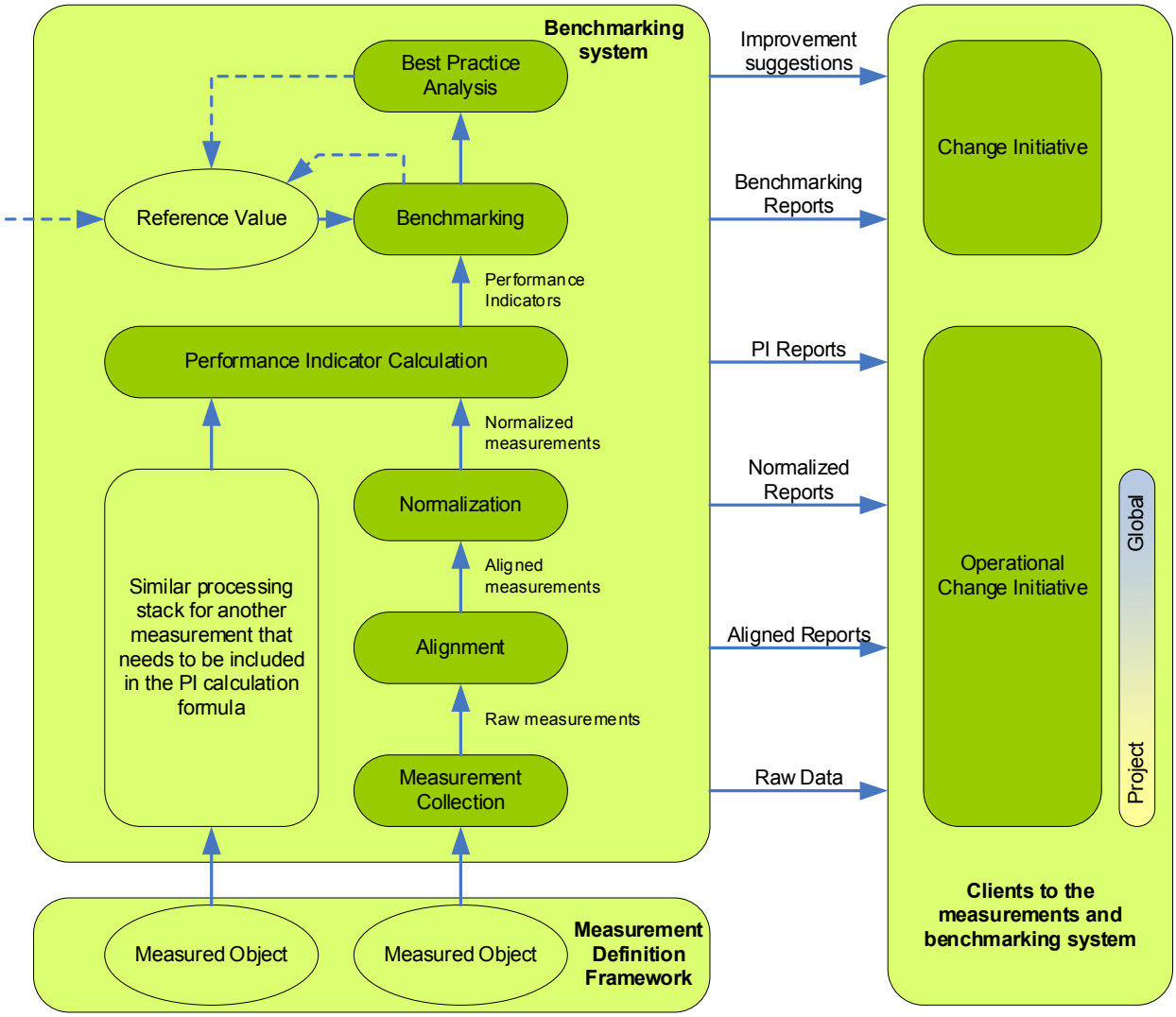

Figure 8 - Measurements and Benchmarking System Blueprint

The foundation to this system is the measurement definition framework, that allows decomposing the Project into categorized measured objects, process artifacts, and activities. All these entities may be linked with relationships, thus becoming an implementable relationship data model.

The data flow, as depicted in Figure 8, consists of several processing steps:

1. Measurement collection is a technical step that allows gathering measurement data from the project. It is based on sourcing and reporting rules expressed by use of objects defined in the measurement definition framework.
2. Raw measurements gathered in the previous step are subjected to alignment, to eliminate project-specific design issues as described in-depth in the “alignment” section. In the blueprint definition this step is expressed as a set of formula-based alignment rules and metadata to specify applicability conditions of these rules. This allows for automation of alignment operation.

3. Aligned data are then normalized, to reduce the effects of Project environment. This step is also modeled as a set of formula-based rules. Contrary to the recommended implementation of alignment, application of normalization rules is controlled manually by the business analyst.
4. Normalized data may be used to calculate PIs. Organization of the PI space is a topic widely discussed in the literature (e.g., see [5], [6]) and is not the subject of this paper. However, it is important to note that the approach presented in this paper allows supporting any PI space through appropriate modeling of measured objects, artifacts and activities.
5. PIs are then compared to reference values, thus resulting in the benchmark value.
6. These results then undergo additional analysis, e.g., trend analysis, searching for relevant correlation to derive useful information (best practice for project delivery, aspects to avoid in contracts etc.).

Thanks to the underlying object-oriented model and rule-definition of the processing steps such measurement and benchmarking, a blueprint can be implemented in the BI platforms available on the market.

The data that is created at different steps of the processing may be used for different purposes as is shown in the "clients" box of Figure 8. Data that is provided in the earlier processing steps ("raw data") may be useful in some instances on the individual Project level; potentially within the Project. Further up the processing chain, data is becoming useful as well for comparing different Projects concluding in the identification of best practices that can and should lead to change initiatives for the lagging Projects.

9. Conclusion

Globalization of the service business brings on the problem of performance data comparability. In this paper this issue is addressed stepwise at the level of measurements, alignment, and normalization. The proposed approach allows expression of measurements, alignment, and normalization by artifacts typical for BI-class tools, like object-oriented data models and data transformation rules. Therefore, implementation of this concept is straightforward. As components of the of this concept are implemented, comparability will be improved stepwise, therewith increasing the amount of data that can be used for benchmarking a given project without manual normalization. Future research work will be focused on development of a normalization error assessment methodology as well as applicability heuristics for normalization rules. In parallel, concept implementation works will comprise measurement object library definition, data model construction and alignment rules library creation.

10. Abbreviations and Definitions

10.1. Abbreviations

BI	business intelligence
BP	benchmarked project
CSP	communications service provider
DSLAM	digital subscriber line access multiplexer
eNodeB	evolved NodeB
GSM	Global System for Mobile Communications
LTE	long-term evolution
MSO	multiple service operator
NE	network element

NOC	network operations center
OpEx	operational expenses
PI	performance indicator
PPP	purchasing power parity
RACI	responsible, accountable, consulted, informed
SLA	service level agreement
SPC	statistical process control
TT	trouble ticket

11. Bibliography and References

1. Project Management Institute, Project Management Body of Knowledge (PMBOK Guide) – Fourth Edition, 2008.
2. Nigam, A. and N.S. Caswell, "Business artifacts: An approach to operational specification", IBM Syst. J., 2003, 42(3): 428-445.
3. Yongchareon, S. and Liu, Chengfei, "A Process View Framework for Artifact-Centric Business Processes," R. Meersman (ed.): OTM 2010, Part I, LNCS 6426, 26–43.
4. Oakland, J, Statistical Process Control, Butterworth-Heinemann, 2003.
5. Tsay, D. and Pariser, E., "Metrics for Comparative Analysis of Operations Competency," Bell Labs Tech. J., 10:1 (2005), 175–179.
6. TeleManagement Forum, GB922 Information Framework (SID), <https://www.tmforum.org>.
7. Jerry Z. Muller, "The Tyranny of Metrics", Princeton University Press, 2018
8. OECD, "Purchasing Power Parities", <https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm>
9. JEDEC, "Arrhenius equation (for reliability)", < https://www.jedec.org/standards-documents/dictionary/terms/arrhenius-equation-reliability>

Improving the Capacity of Full Duplex DOCSIS with Full Duplex Passives

Letter to the Editor prepared for SCTE•ISBE by

Mark Knowles, Senior Product Manager, Technetix
Innovation House, Technetix Business Park
Albourne, West Sussex, BN6 9EB
mark.knowles@technetix.com

Jan Ariesen, Chief Technology Officer, Technetix
Innovation House, Technetix Business Park
Albourne, West Sussex, BN6 9EB
Jan.Ariesen@technetix.com

1. Introduction

As consumers, we have been getting increasingly data hungry over the last few years with an insatiable appetite for faster download speeds to view that video quicker, stream our audio playlists and improve our online experiences. As content providers evolve their offerings, higher video and audio resolutions are now the norm and over-the-top services such as Netflix, YouTube and others now provide stable ultra high definition (UHD) 4K resolutions for continuous streaming. Our industry has responded well to this demand well so far with many network upgrades to 1.0 GHz and 1.2 GHz architectures already well underway.

Now we face a new challenge - the upstream. Today, we are using existing devices differently and also acquiring new technology that is increasing the use of upstream bandwidth. For the first time, the upstream traffic volume is growing faster than the downstream. We now FaceTime our friends and Skype our business colleges as daily activities and these applications are pushing the upstream demands to new levels. In addition, our lives are becoming 'smarter' with all kinds on in-home devices from thermostats to security cameras, all connected to our networks and uploading data to our phones for viewing or cloud storage services for storing. Cable operators are also growing their base of business customers who typically have higher upstream rates than residential customers. Finally, and perhaps most importantly, there are now competing services that offer symmetrical gigabit rates over fiber to the home. Cable operators must therefore have a similar offering to remain competitive.

2. Passives Solution

To meet the aforementioned market demands, CableLabs and industry partners have been developing the specification for full duplex DOCSIS (FDD, also called FDX DOCSIS) to provide a symmetrical high speed service over existing cable networks for some time. FDD is a cost-effective way to create a 10 Gbps symmetrical communication path to the home. Technologies including the scheduler, echo cancellation and smart algorithms make it feasible to drive full duplex communication over existing coax installations.

Typically, FDD is expected to be designed in conjunction with a remote physical layer (remote PHY) or a node+0 environment as shown in Figure 1.

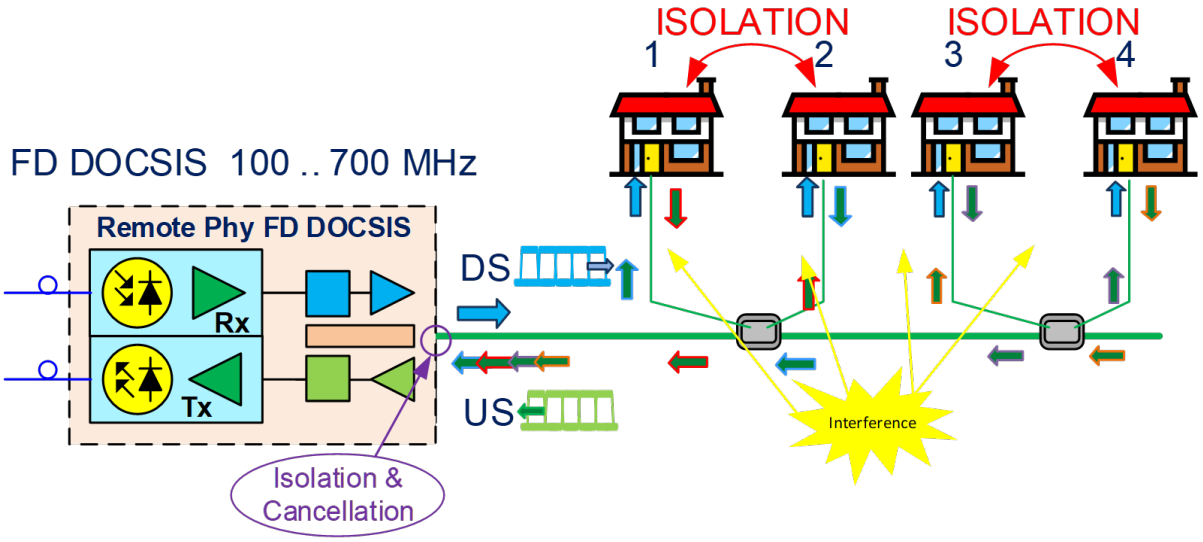

Figure 1 – Typical FDD/remote PHY architecture.

However, with traditional taps built using conventional ferrite structures, a port-to-port isolation figure of around 20 dB to 25 dB is usual. Unfortunately, this simply isn't enough to prevent neighboring homes from creating interference to other FDD services. Without a better approach, this interference would affect the FDD network design and limit capacities and capabilities of the entire network.

Some of the interference phenomena which can impact FDD operation are listed below:

- adjacent level interference (ALI)
- adjacent device interference (ADI)
- co-channel interference (CCI)
- adjacent carrier interference (ACI)
- passive intermodulation (PIM)

The solution currently offered by the industry is to create groups of homes (interference groups) and only use half duplex communication within a group. Dividing a tap cascade into three isolated groups will help to reduce interferences like ACI, CCI, ADI and ALI, but the unfortunate result is a network data capacity loss of 40%.

There are three types of outdoor taps:

- Normal taps:
  - These are the taps used at this moment with and isolation of 20-25 dB between the ports
- High isolation taps:
  - These are taps that have additional isolation between the outdoor taps, also in low values so every tap stands on its own but the isolation between the tap ports is still 20-25 dB
- FDD taps:
  - These are taps that have additional isolation between the taps and the tap ports.

Here's how the grouping works: with standard taps you need to group the low value taps, as the isolation between the taps is too low for FDD. Figure 2 shows an example of seven taps that can be combined into five groups.

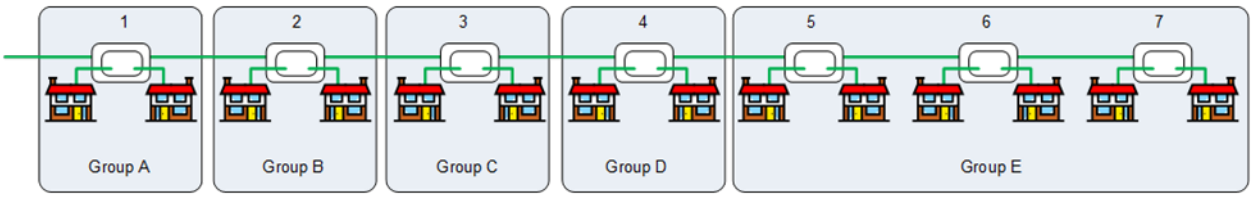

Figure 2 - Example of 7 taps in cascade with an average number of 6 ports per tap

So we have:

- 42 households passed (HP)
- Seven taps
- Five groups

FDD means that the remote PHY device (RPD) is using full duplex with echo cancelation and the modem is half duplex, meaning they do not transmit on the same frequency as they receive.

If we split the upstream of 10 Gb/s over 42 HP that means 238 Mb/s per home:

Full duplex taps: Every home has an upstream of 238 Mb/s and a downstream of (10 Gb/s – 238 Mb/s) = 9762 Mb/s

High isolation taps: Every home has an upstream of 238 Mb/s and a downstream of (10 Gb/s - (6 x 238 Mb/s)) = 8571 Mb/s

Normal taps: Every home in group A to D, has an upstream of 238 Mb/s and a downstream of (10 Gb/s - (6 x 238 Mb/s)) = 8571 Mb/s

Every home in group E has an upstream of 238 Mb/s and a downstream of (10 Gb/s - (18 x 238 Mb/s)) = 5714 Mb/s

So, the efficiency of FDD with the different taps are as follows:

Full Duplex taps: Upstream 238 Mb/s + Downstream 9,762 Mb/s = 10,000 Mb/s = 100%

High isolation taps: Upstream 238 Mb/s + Downstream 8,571 Mb/s = 8,810 Mb/s = 88%

Normal taps: Upstream 238 Mb/s + Downstream 5,714 Mb/s = 5,952 Mb/s = 60%

This capacity loss is created due to the fact we can only communicate half-duplex per group. If a frequency in that group is used, there is no option to use this frequency again in the same group. Smart algorithms must be created to avoid this capacity loss, and they are still limited.

Next to the capacity loss due to the interference there is also the effect of passive intermodulation (PIM). PIM is created by magnetism in ferrite beats. These ferrites are used in normal splitters or taps. If a ferrite gets magnetized, the center point of the ferrite transfer characteristic is moved and therefore becomes

non-linear. This non-linear behavior will cause intermodulation and because the upstream signals are much higher in level than the downstream signals, even small intermodulation products will cause a degradation of the downstream quality.

A new passive technology has been developed that will reduce the interference levels of ACI, ADI, CCI and ALI without the need to create groups. With this technology the capacity is not reduced and the maximum full duplex DOCSIS performance of 10 Gb/s remains. The problem of PIM is also solved by this technology and therefore the operator can make the maximum use of their investment to create a full duplex communication path.

To achieve the above, an alternative to ferrite couplers was developed by utilizing microstrip coupler (micro-coupler) technology shown in Figure 3. This micro-coupler technology was already used in a previous series of tilted cabinet taps. Micro-coupler technology usually requires a larger physical design in order to accommodate the tracks within the printed circuit board (PCB) design. Fitting this methodology successfully into a small Regal style horseshoe tap has been extremely challenging but has been achieved. This permits the new device to maintain low through-losses, return loss figures of >16 dB up to the full 1.2 GHz design bandwidth and most importantly a significantly higher port-to-port isolation figure. Across the FDD band up to 700 MHz, it is now possible to keep this figure over 40 dB, as depicted in Figure 4.

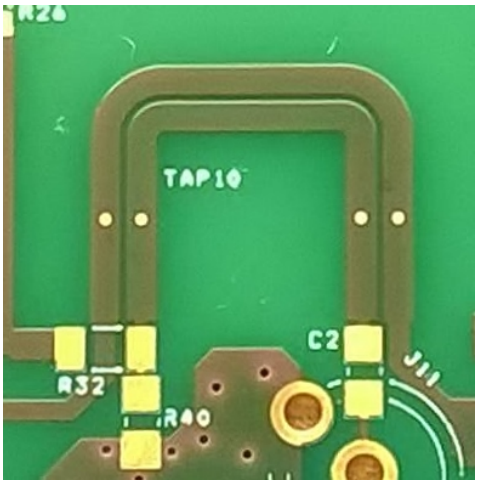

Figure 3 – Micro-coupler technology alternative to ferrite couplers.

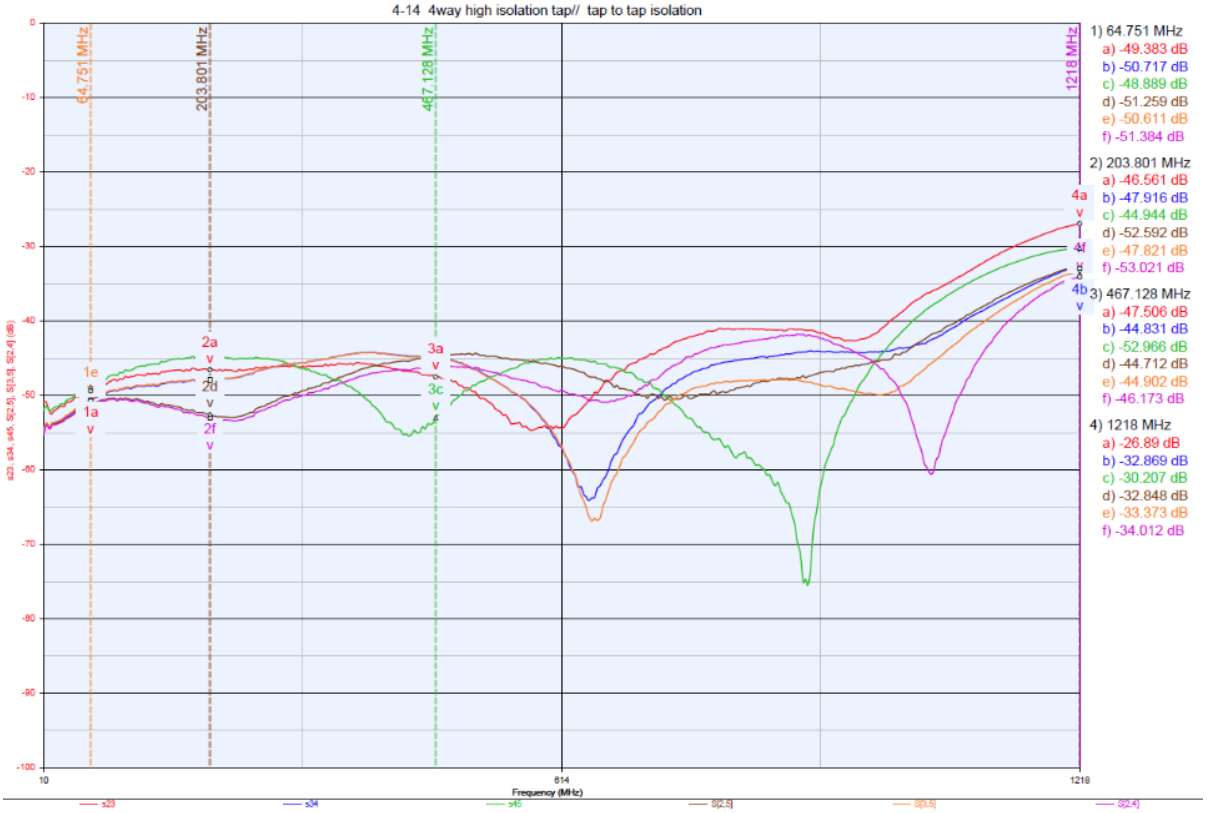

Figure 4 – Isolation measurements on taps with micro-coupler technology.

These prototype designs have been tested by the laboratory at CableLabs over a seven tap cascaded design model, which is shown in Figure 5.

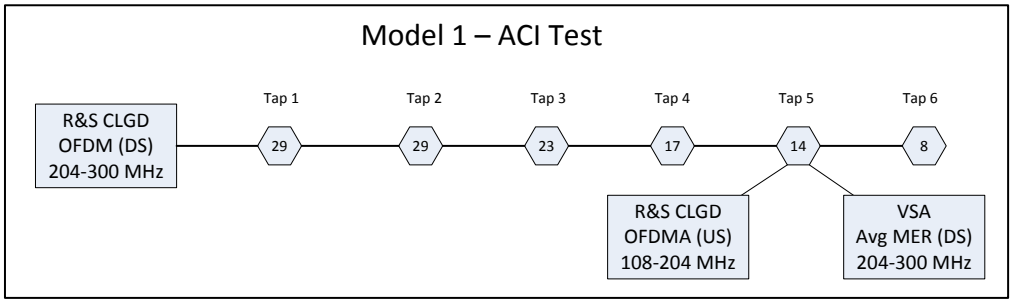

Figure 5 – Cablelabs 7 tap cascaded design model.

The new technology has proven to be even more effective than anticipated. At +62 dBmV adjacent channel levels, the micro-coupler-based FDD cascade downstream modulation error ratio (MER) was measured to be over 10 dB better than the legacy taps. A measurement of the isolation of the tap under test is shown in Figure 6.

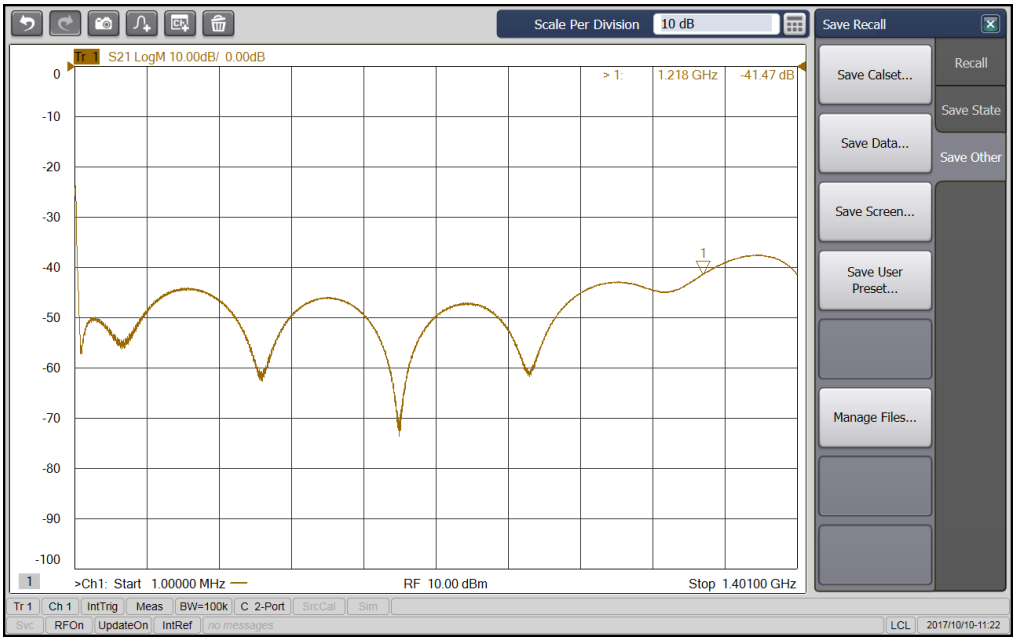

Figure 6 - Isolation of the tap under test

With the legacy taps were used in the same cascade, the test equipment could no longer decode the signals due to the strong ACI generated.

Another positive effect of using the micro-coupler technology is that since the new technology does not have a traditional ferrite structure that can get magnetized, there is no measurable PIM and, as a result no intermodulation from the upstream signals that can destroy the downstream signal quality.

3. Conclusions

Full duplex DOCSIS with a capacity of 10 Gbps in the upstream and downstream is now even more feasible with new active and passive devices based on micro-coupler technology. The new micro-coupler technology significantly reduces interference issues like ACI, CCI, ADI and ALI which slow down the data transport, and also effectively eliminates PIM. New passives with high isolation can now reduce the interference issues from the outdoor taps and therefore improve overall network capacity.

FDD as a network architecture is still evolving but one thing is certain: the ongoing quest for the fastest possible symmetrical data speeds doesn't end here. Whichever way the industry moves to implement the final solution, we now know that a bidirectional 10 Gb/s service is achievable using existing coaxial network structures. The passive part of the network has been seen as both a challenging and important 'final piece' of the FDD solution, but with new devices based on micro-coupler technology, the previous hurdles to full capacity FDD networks can be truly 'jumped.'

4. Abbreviations

ACI	adjacent channel interference
ADI	adjacent device interference
ALI	adjacent level interference

CCI	co-channel interference
dB	decibel
FDD	full duplex DOCSIS
GHz	gigahertz
ISBE	International Society of Broadband Experts
MER	modulation error ratio
MHz	megahertz
PIM	passive intermodulation
SCTE	Society of Cable Telecommunications Engineers
UHD	ultra-high definition

SCTE • ISBE
STANDARDS

NETWORK
OPERATIONS