Keysight Fundamentals of Arbitrary Waveform Generation

A high performance AWG primer

Reference Guide

KEYSIGHT TECHNOLOGIES

Notices

Copyright Notice

Manual Part Number

M8190-91050

Edition

Edition 4.0, June 2015

Published by:

Keysight Technologies
Deutschland GmbH,
Herrenberger Str. 130,
71034 Böblingen, Germany

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

ESD Sensitive Device

All front-panel connectors of the M8190A are sensitive to Electrostatic discharge (ESD). We recommend to operate the instrument in an electrostatic safe environment. There is a risk of instrument malfunction when touching a connector.

Please follow this instruction:

Before touching the front-panel connectors, discharge yourself by touching the properly grounded mainframe.

U.S. Government Rights

The Software is "commercial computer software," as defined by Federal Acquisi- tion Regulation ("FAR") 2.101. Pursuant to FAR 12.212 and 27.405-3 and Depart- ment of Defense FAR Supplement ("DFARS") 227.7202, the U.S. govern-ment acquires commercial computer software under the same terms by which the software is customarily provided to the public. Accordingly, Keysight pro- vides the Software to U.S. government customers under its standard commercial license, which is embodied in its End User License Agreement (EULA), a copy of which can be found at http://www.keysight.com/find/sweula. The li- cense set forth in the EULA represents the exclusive authority by which the U.S. government may use, modify, distribute, or disclose the Software. The EULA and the license set forth therein, does not require or permit, among other things, that Keysight: (1) Furnish technical information related to commercial computer software or commercial computer soft- ware documentation that is not custom- arily provided to the public; or (2) Relin- quish to, or otherwise provide, the gov- ernment rights in excess of these rights customarily provided to the public to use, modify, reproduce, release, perform, dis- play, or disclose commercial computer software or commercial computer soft- ware documentation. No additional government requirements beyond those set forth in the EULA shall apply, except to the extent that those terms, rights, or licenses are explicitly required from all providers of commercial computer soft- ware pursuant to the FAR and the DFARS and are set forth specifically in writing elsewhere in the EULA. Keysight shall be under no obligation to update, revise or otherwise modify the Software. With respect to any technical data as defined by FAR 2.101, pursuant to FAR 12.211 and 27.404.2 and DFARS 227.7102, the U.S. government acquires no greater than Limited Rights as defined in FAR 27.401 or DFAR 227.7103-5 (c), as appli- cable in any technical data.

Warranty

THE MATERIAL CONTAINED IN THIS DOCUMENT IS PROVIDED "AS IS," AND IS SUBJECT TO BEING CHANGED, WITHOUT NOTICE, IN FUTURE EDITIONS. FURTHER, TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, KEYSIGHT DISCLAIMS ALL WARRANTIES, EI- THER EXPRESS OR IMPLIED, WITH REGARD TO THIS MANUAL AND ANY INFORMATION CONTAINED HEREIN, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FIT- NESS FOR A PARTICULAR PUR- POSE. KEYSIGHT SHALL NOT BE LIABLE FOR ERRORS OR FOR INCIDENTAL OR CONSEQUENTIAL DAM- AGES IN CONNECTION WITH THE FURNISHING, USE, OR PERFOR- MANCE OF THIS DOCUMENT OR OF ANY INFORMATION CONTAINED HEREIN. SHOULD KEYSIGHT AND THE USER HAVE A SEPARATE WRITTEN AGREEMENT WITH WARRANTY TERMS COVERING THE MATERIAL IN THIS DOCUMENT THAT CONFLICT WITH THESE TERMS, THE WARRANTY TERMS IN THE SEPARATE AGREEMENT SHALL CONTROL.

Safety Information

CAUTION

A CAUTION notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a CAUTION notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

Safety Summary

General Safety Precautions

The following general safety precautions must be observed during all phases of operation of this instrument. Failure to comply with these precautions or with specific warnings elsewhere in this manual violates safety standards of design, manufacture, and intended use of the instrument. For safe operation the general safety precautions for the M9502A and M9505A AXIe chassis, must be followed. See: http://www.keysight.com/find/M9505A Keysight Technologies Inc. assumes no liability for the customer's failure to comply with these requirements. Before operation, review the instrument and manual for safety markings and instructions. You must follow these to ensure safe operation and to maintain the instrument in safe condition.

Initial Inspection

Inspect the shipping container for damage. If there is damage to the container or cushioning, keep them until you have checked the contents of the shipment for completeness and verified the instrument both mechanically and electrically. The Performance Tests give procedures for checking the operation of the instrument. If the contents are incomplete, mechanical damage or defect is apparent, or if an instrument does not pass the operator's checks, notify the nearest Keysight Technologies Sales/Service Office.

WARNING To avoid hazardous electrical shock, do not perform electrical tests when there are signs of shipping damage to any portion of the outer enclosure (covers, panels, etc.).

General

This product is a Safety Class 3 instrument. The protective features of this product may be impaired if it is used in a manner not specified in the operation instructions.

Environment Conditions

This instrument is intended for indoor use in an installation category II, pollution degree 2 environment. It is designed to operate within a temperature range of 0 °C – 40 °C (32 °F – 105 °F) at a maximum relative humidity of 80% and at altitudes of up to 2000 meters.

This module can be stored or shipped at temperatures between -40 °C and +70 °C. Protect the module from temperature extremes that may cause condensation within it.

Before Applying Power

Verify that all safety precautions are taken including those defined for the mainframe.

Line Power Requirements

The Keysight M8190A operates when installed in an Keysight AXIe mainframe.

Do Not Operate in an Explosive Atmosphere

Do not operate the instrument in the presence of flammable gases or fumes.

Do Not Remove the Instrument Cover

Operating personnel must not remove instrument covers. Component replacement and internal adjustments must be made only by qualified personnel. Instruments that appear damaged or defective should be made inoperative and secured against unintended operation until they can be repaired by qualified service personnel.

Safety Symbols

Table 1 Safety Symbol

Symbol	Description
	Indicates warning or caution. If you see this symbol on a product, you must refer to the manuals for specific Warning or Caution information to avoid personal injury or damage to the product.
	C-Tick Conformity Mark of the Australian ACA for EMC compliance.
	CE Marking to state compliance within the European Community: This product is in conformity with the relevant European Directives.
	General Recycling Mark

Table 2 Compliance and Environmental Information

Symbol	Description
	This product complies with the WEEE Directive (2002/96/EC) marketing requirements. The affixed label indicates that you must not discard this electrical/electronic product in domestic household waste. Product category: With reference to the equipment types in the WEEE Directive Annexure I, this product is classed as a “Monitoring and Control instrumentation” product. Do not dispose in domestic household waste. To return unwanted products, contact your local Keysight office, or see http://about.keysight.com/en/companyinfo/environment/takeback.shtml for more information.

Contents

1 Introduction

1.1 The Need for Stimuli Instrumentation

Testing an electrical/electronic device or system implies applying the right stimuli to the DUT/SUT (Device Under Test/System Under Test) and analysing the resulting behaviour (see Figure 1). In some cases, stimuli come from the real world but in most situations a set of signals must be supplied by appropriate instrumentation. These sources must be capable of generating a variety of stimuli and it must be possible to set their characteristics in a broad enough range so that the DUT's operating range may be established and validated. The traditional approach has been to use a specialized device for each stimulus category. This has resulted in a variety of instrument types and subtypes offered by T&M manufacturers:

- AC/DC Power Supplies
- Electronic Loads
- Function Generators
- RF Generators
- Pulse and Pattern Generators
- Arbitrary Waveform Generators (AWG)

Traditionally, function generators have been the most popular stimuli instrument category. They can output a limited set of waveforms (sine, square, triangle, etc.) and users can adjust many of their characteristics such as frequency, amplitude, DC offset, duty cycle, and symmetry. Many function generators implement basic internal or external modulation such as AM (Amplitude Modulation), PM (Pulse Modulation), and FM (Frequency Modulation), and some of them can even sweep the output frequency in a range of interest. Although many traditional function generators incorporate digital control of their functions, waveforms are generated using analog circuits. Available waveform shapes, number of outputs, and frequency range are some of the most important limiting factors of function generators.

Pulse generators can provide a train of pulses with controllable parameters such as PRF (Pulse Repetition Frequency), pulse duration, and “high” and “low” voltages. Edge location and rise/fall times can be controlled, sometimes independently in some instruments, so that jitter or DCD (Duty Cycle Distortion) can be emulated. Modern pulse generators are not limited to a repetitive train of pulses, but they are capable of generating a pre-defined sequence of “1” and “0” therefore becoming true serial and parallel data generators. Traditionally, some timing parameters of the pulses such as rise/fall times or jitter were controlled using analog circuitry.

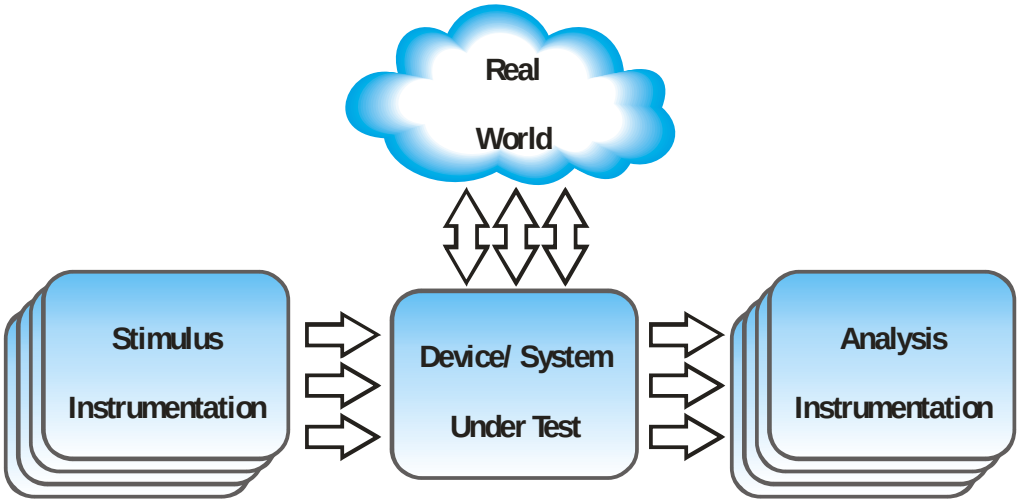


Figure 1: The stimulus-analysis test model

RF generators include many subcategories. Basic RF generators usually supply an unmodulated CW (Continuous Waveform) carrier where spectral purity is paramount. Although many generators include the capability to sweep the output frequency and to provide basic analog modulation, they cannot be used for functional testing of modern radio communication systems based in complex digital modulation schemes. A more advanced instrument class, the vector signal generator or VSG, permits bipolar modulation of two quadrature carriers (90° phase difference) so any amplitude/phase modulation state can be achieved through two baseband signals, known as the I (or In-Phase) and Q (or Quadrature) components. While some VSGs may need external I and Q sources, others can generate these signals internally by digital means. Carrier frequency range, modulation bandwidth (twice the maximum bandwidth of the acceptable I and Q signals), and modulation schemes supported are some of the limiting factors of VSGs.

1.2 The Arbitrary Waveform Generator (AWG)

All of the above generator classes are typically designed to test a specific type of system and they provide a quite limited set of stimuli. Frequently, they are designed to source clean or "perfect" signals, so unknown, uncontrolled signal distortions do not influence test results. Just as frequently, they are used to provide distortions that occur in real-life signals. Applying real-life stimuli simulating actual or extreme working conditions usually implies putting the DUT/SUT to work in a real environment where these conditions are difficult to control and many times even to characterize. Reaching extreme conditions may be difficult and may take some time, as they can be statistically infrequent. This approach also implies that testing the device in a realistic environment is only possible at the end of the design/integration process when all sub-systems are available. At this stage of any project, to solve any design flaw results in additional costs and, what is even worse, in incremental delays. Isolating the source of any problem is also challenging as unexpected interaction between modules and subsystems may arise.

Ideally, it would be better to test each module early in the development process in as realistic as possible conditions. In some cases this can be accomplished using "golden" devices that provide the right signals that may be well characterized, but often they are difficult to control. A universal signal source where any signal could be generated would be the ultimate solution. Specifically, this is the goal of arbitrary waveform generators (AWGs, also known as "arbs"). The basic architecture of these instruments is quite straight forward, and it looks very close to that of a DSO (Digital Storage Oscilloscope) where the signal flow has been inverted: a waveform memory with a numerical representation of the desired signal supplies samples sequentially to a DAC (Digital-to-Analog Converter) at a constant sampling speed (see Figure 2).

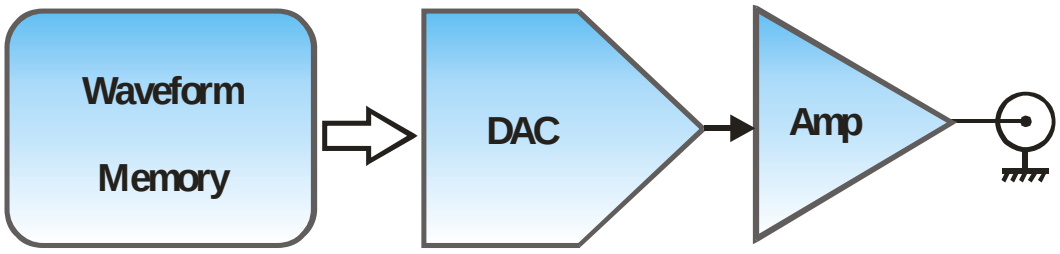


Figure 2: Basic block diagram for an Arbitrary Waveform Generator (AWG)

AWGs first appeared in the market during the eighties when available data conversion, memory, and instrument communication technologies made them possible. Most of the features and characteristics of modern high-performance AWGs were first introduced into the marketplace by Hewlett-Packard (Keysight's parent Company) with the 8770A Arbitrary Waveform Synthesizer, released in 1988. This two-channel, 12 bit DAC @ 125 MSa/s (Mega samples/second) and 50MHz bandwidth AWG featured some characteristics such as memory segmentation and sequencing, and real-world interaction through input/output digital signals common in today's instruments. Main application areas targeted were IQ modulation, radar, and magnetic storage media and device test, which are still important markets for high-performance AWGs today.

During the last 30 years, AWGs have become a major instrumentation category, and they have diversified in many markets and performance segments. Form factors available range from USB dongles up to modules for most computer and instrumentation buses such as VXI, PXI or AXIe. AWGs can be found alone in bench top instruments or integrated within other test gear such as digital oscilloscopes or vector signal generators. Nowadays low-cost instruments such as the Keysight 33600 series are capable of generating signals at hundreds of MSa/s while state-of-the art devices such as the Keysight M8195A can reach 65GSa/s simultaneously in each of its four channels.

Due to their flexibility and the availability of software packages tailored to specific applications, AWGs are replacing or absorbing many signal source categories as most of today's function and pulse generators are in fact implemented using an AWG architecture. In the RF domain, AWGs have been extensively used to generate baseband signals. Currently available sampling rates, analog bandwidths, and output signal quality, both in terms of spurious-free dynamic range (SFDR) and phase noise, allow AWGs to successfully replace RF generators in some application areas.

Waveform calculation is an important associated task to any AWG, and its complexity depends mainly on the application area. Definition of waveforms and the tools to obtain them is one of the issues to consider when evaluating and implementing signal generation solutions based on AWGs. Fortunately, a wealth of waveform creation tools including specialized software packages or general-purpose mathematical packages and simulation tools are available today to AWG users.

1.3 AWG Basic Characteristics

As in any instrument category, there are a large variety of instruments available in the market, and there are many characteristics that differentiate them. Some are related to their architecture, while others are related to their raw performance. The basic characteristics for any AWG are the following:

Sampling Rate (or Frequency): This is the maximum speed of conversion of the DAC. For most AWGs, this parameter is equal to the speed at which samples are read from the waveform memory. It can be set by the user in a certain range, so a minimum sampling rate may be also specified. Typically, a specified sampling rate is maintained during a test, and changing it requires the user to stop the output. However, some instruments include a sample clock input so sampling frequency can be changed continuously. The Nyquist sampling theorem relates this specification with the maximum meaningful frequency component that can be synthesized by the generator, one half of the sampling rate (called Nyquist Frequency).

Memory Size (or Record Length): This is the maximum number of samples that can be stored in the waveform memory. Most AWGs allow the user to use just a fraction of the available waveform memory. However, record length selection may be limited, due to memory access issues, in two different ways: minimum length and length granularity. The length granularity limits the choice as the selected record length must be a multiple of a given integer. Some AWGs may limit the record length depending on the sampling rate. Users can play with record length and sampling rate settings to reach the desired bandwidth and the necessary time window for a given application. These are constrained by the following formula:

$$\textit{Time Window} = \textit{Record Length} / \textit{Sampling Rate}$$

Vertical Resolution: This is the number of bits of the DACs in the AWG. For a given vertical resolution, N, the DAC will be capable of generating 2^N different levels. The higher the vertical resolution, the more detailed the waveform will be. Good vertical resolution and high sampling rates are often difficult to obtain in the same device.

Analog Bandwidth: This is the 3dB BW for the signals coming out of the AWG. This spec is directly related to the rise time in the step response of the generator. Analog bandwidth may be higher, equal, or lower than the Nyquist frequency.

Output Characteristics: Amplitude and DC offset ranges are important to some applications. The more extended these ranges are the more likely the signal will be usable without further processing. It is difficult to reach high amplitudes and high bandwidths simultaneously. Other important characteristics are output impedance (typically 50 Ω), the availability of differential outputs or the choice of output filters.

Number of Channels: Many AWGs incorporate more than one channel. Two or four channel devices are common in the market. While some generators share the same sampling clock and some waveform parameters (i.e. record length), others allow users to setup each channel independently so they effectively work as independent AWGs. Additionally, some AWGs include a standard or optional parallel digital output to implement the functionality of a digital pattern generator. High-performance AWGs usually implement several marker and trigger outputs associated to the waveform memory so synchronization with other test equipment or with the DUT is easier.

The influence of these parameters in the quality of the generated signal, the technology involved, and the way to extract the most of the available performance will be extensively covered in the next chapters.

2 AWG Fundamentals

2.1 Sampling Theory for AWGs

Any digital electronic system designed to process signals does not treat these signals continuously as they are in reality. Instead, it handles them in the form of samples. Samples can be defined by the instantaneous value of a signal in a given instant in time. When a signal evolves in time, as all of the interesting ones do, a series of samples must be used as a summarized image of the real-world signal. Although it is not a compulsory requirement, most Digital Signal Processing (DSP) systems use a series of equally spaced samples to define a continuous signal. Sampling pace can be expressed in terms of time, sampling period (Ts), frequency, or sampling rate (Fs). Systems such as a CD player, a digital oscilloscope, or an AWG are some good examples of this. Any sampled version of a signal will somehow lose information, as its evolution between samples is not explicitly contained in the information carried by the samples.

Sampling theory, a discipline started by Harry Nyquist and Claude Shannon in the 20th Century, establishes the conditions that a signal must meet so that no information is lost after a properly performed sampling process thus enabling the complete original signal to be recovered from the samples. Probably, the most important result of the theory is the Sampling Theorem (also known as the Nyquist Sampling Theorem). The theorem can be stated as follows:

"Given a bandwidth-limited signal with bandwidth B, it can be recovered from a sequence of equally spaced samples sampled at a speed of at least 2B (Figure 3)."

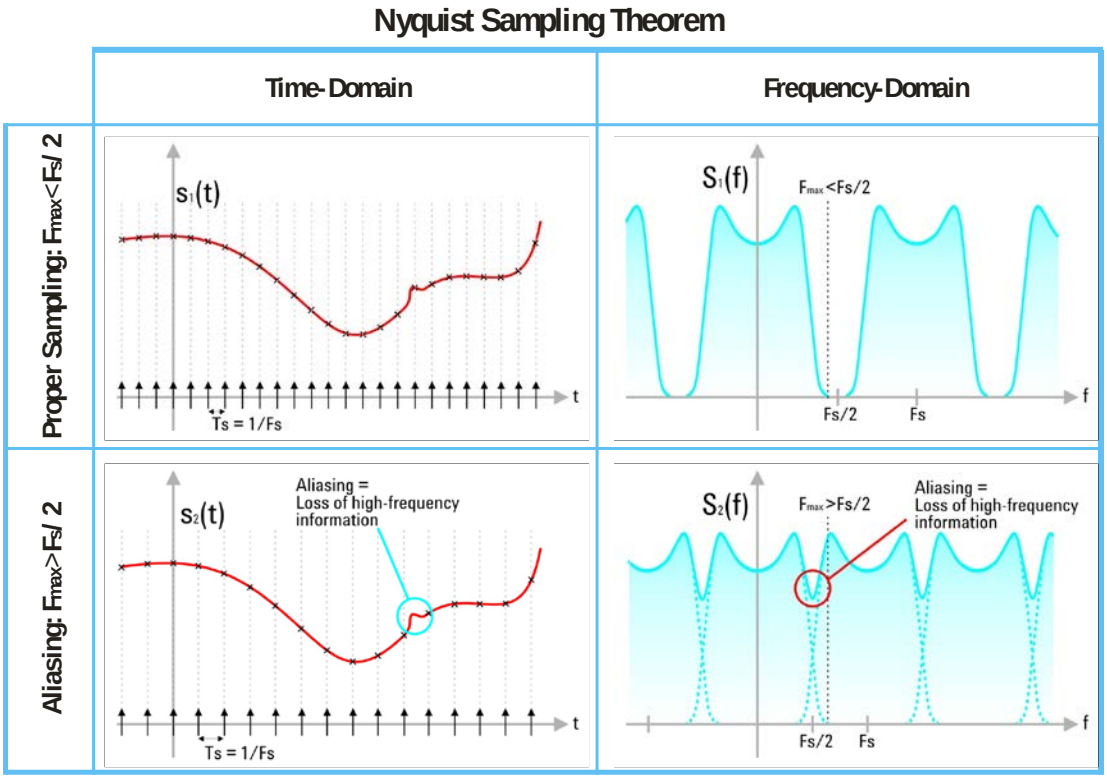


Figure 3: Nyquist sampling theorem in the time and frequency domains. Sampling rate must be at least twice the signal's maximum frequency. Otherwise the high frequency information will be lost.

While in a DSO interpolation is performed by numerical means, AWGs implement this function through a low-pass analog filter (also known as "reconstruction" or "interpolation" filter) or a combination of a digital and an analog filter.

Not meeting the sampling condition specified by the Nyquist theorem will result in an unrecoverable loss of information or even worse, the addition of false information, or aliasing, in the recovered signal. It is important to understand that just to sample a signal properly is not sufficient to recover the original signal. In order to do so, any amplitude level corresponding to any instant between samples must be reconstructed through an interpolation process (Figure 4). The way interpolation is performed will greatly influence the overall performance of the system.

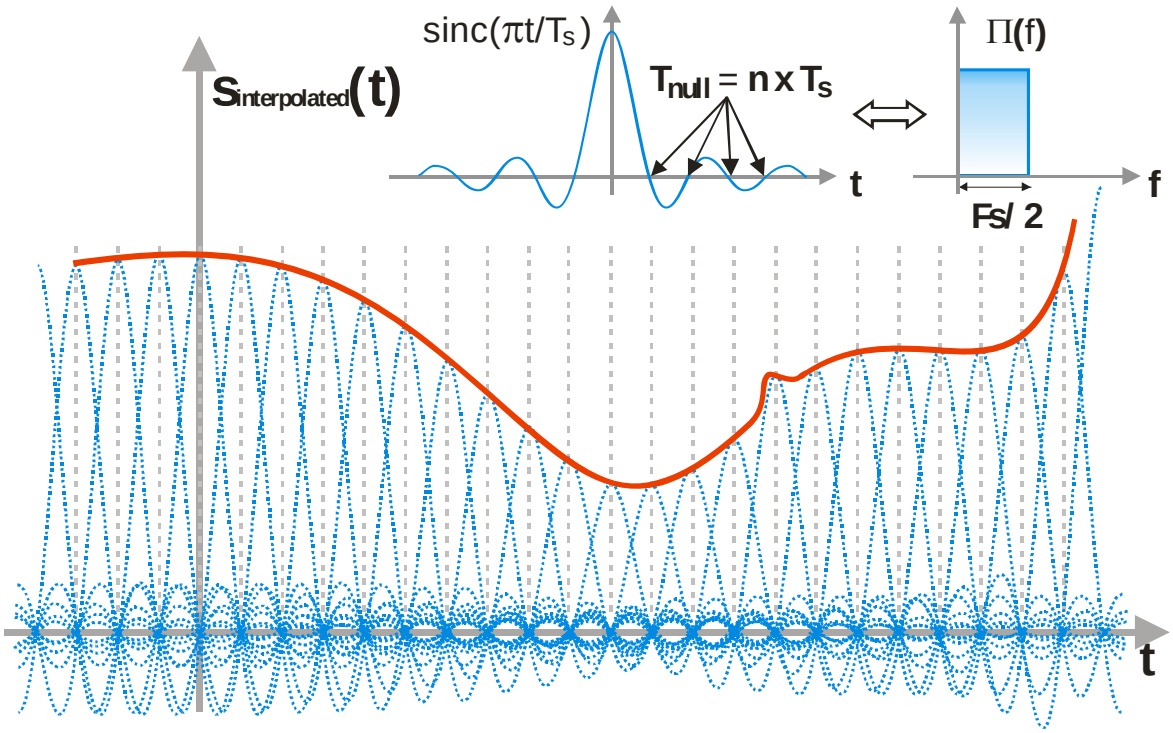


Figure 4: A bandwidth-limited signal can be fully recovered if the sampling theorem requirements are met and the right interpolation process is applied to the sample. Here, the ideal interpolator sinc(πt/Ts) is applied.

AWGs convert numerical samples into a continuous analog signal by feeding them into a DAC . Sampled signal analysis is performed by handling each sample as a Dirac delta function δ of the same amplitude and located at the sampling instant. Figure 5c shows the result of sampling a bandwidth limited signal, a train of Dirac's δ with the same amplitude as the original samples. The spectrum of that signal (Figure 5d) is the superposition of infinite copies of the original spectrum located at multiples of the sampling frequency. It is easy to see that if the Nyquist sampling theorem is met then the original signal may be recovered by simply filtering the sampled signal with a brick-wall filter with an Fs/2 cut-off frequency. The original spectrum is recovered so the original time-domain signal will be recovered as well. Filtering in the frequency domain is equivalent to convolution in the time domain. Convolution of a train of Dirac's δ with the filter impulse response is as simple as superimposing multiple copies of it centred in the location of each δ (Figure 4). The filter's impulse response is, in fact, an interpolation function also called interpolation filter. For the ideal interpolator filter, a brick-wall filter of Fs/2 bandwidth, the corresponding interpolation function is:

$$\text{sinc}(\pi t/Ts) = \sin(\pi t/Ts) / (\pi t/Ts)$$

The value of this function is 1 for t=0 while it is 0 for n x Ts (n≠0).

While a train of Dirac's δ is the best way to mathematically describe a sampled signal, this is not the kind of signal an actual DAC can generate. DACs typically update the output voltage level at the sampling instants and preserve that level in the meantime. It is quite simple to obtain the signal coming out of the DAC from the corresponding train of δ, as it is the result of convolving it with the Π(t/Ts) function, also known as a zero order hold function (Figure 5e). As convolving in the time domain means multiplying in the frequency domain, the resulting output signal's spectrum (Figure 5f) will be the result of multiplying the sampled signal spectrum with the Fourier transform of the Π function which is the sinc (f/Fs). This response can be understood as a not very selective low-pass filter with an attenuation of 3.92 dB at the Nyquist frequency, Fs/2.

Recovering the original function is still possible by applying an analog interpolation filter (Figure 5h) with a response of:

Π(2f/Fs) / sinc(f/Fs)	f<Fs/2
0	f>=Fs/2

Such a response cannot be obtained in the real world, since it is not possible to implement a real filter with such an abrupt roll-off. A real filter will require a reasonable roll-off band (Figure 5h) and, as a consequence, it will not be strictly possible to obtain aliasing-free signals with bandwidths up to Fs/2. Maximum practical bandwidth may be limited to 10 to 25% below the Nyquist frequency. Obtaining accurate interpolation filters with the required accuracy may be quite difficult and expensive and, theoretically, a different filter will be required for each intended sampling frequency. Additionally, the required amplitude response will only be possible at the expense of a relatively poor group delay response resulting in further signal distortions.

Fortunately, what matters is the overall system response, and even if the analog interpolation filter is far from perfect, the signal contained in the waveform memory may be lineally pre-distorted in magnitude and phase in order to compensate for the filter (and other) imperfections. This approach allows AWG designers to implement a limited number (sometimes just one) of simpler and cheaper interpolation filters while the overall response will be corrected by the right digital filter applied to the original samples. Obtaining the right overall response requires good characterization of the analog response of the AWG output stage in order to calculate the digital filter coefficients. Some generators (such as the Keysight M8195A 65GSa/s AWG) even incorporate the frequency response for each channel stored in non-volatile memory after careful characterization during the manufacturing process. Some AWGs can even apply digital filters to the samples stored in the waveform memory in real-time during signal generation, but for those without real-time DSP capability, this is not feasible. In this case, original samples must be pre-filtered offline once and then transferred to the target AWG waveform memory for real-time generation.

Mixed analog-digital interpolation filtering has an additional advantage as multiple interpolation functions may be implemented with the same hardware. A flat response may be desirable for most situations but, sometimes, the resulting ringing visible in fast transitions may limit the usability of the signal. An overall Gaussian response will be better to generate a fast pulse without overshoots, but it will not be as flat in the frequency domain. Digital filtering can be also used to improve the overall generation bandwidth by implementing signal pre-emphasis and can be also used to emulate (embed) or correct (de-embed) the response of external devices such as cabling, additional instrumentation, or circuit elements.

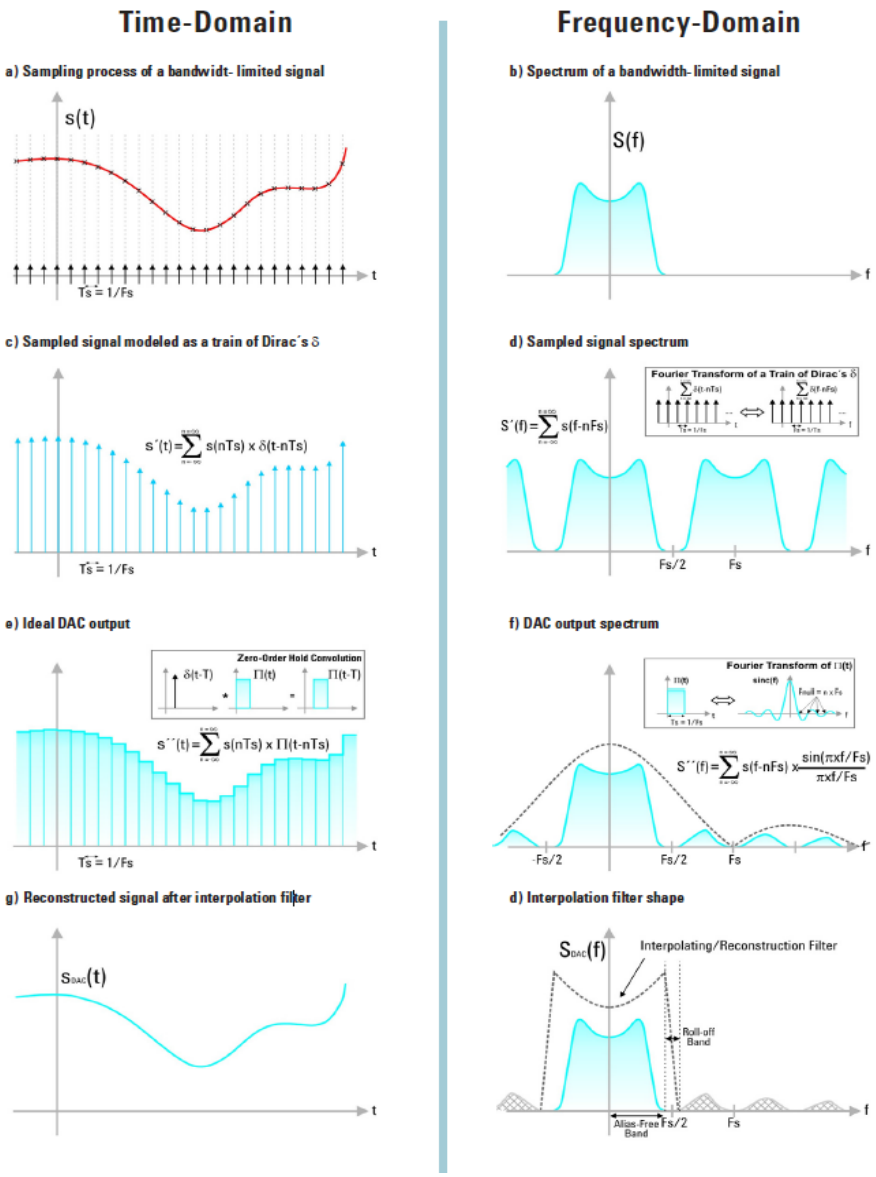


Figure 5: Signal processing flow in an ideal Arbitrary Waveform Generator in both the time and frequency domains for a properly sampled, bandwidth limited signal. Interpolation filter may be implemented with just analog circuits or using a combination of a low-pass filter and digital pre-correction of waveform data.

2.2 AWG Architectures

Although sampling rate and vertical resolution are still the key differentiators for AWGs, there are some differentiated classes based on their basic architecture rather than performance:

”True arb”: This category includes all of the generators that read samples one by one from the waveform memory and convert them to analog levels at a sampling rate set by the user (see Figure 6). The memory access speed is established by the sampling rate and access to the waveform memory is sequential.

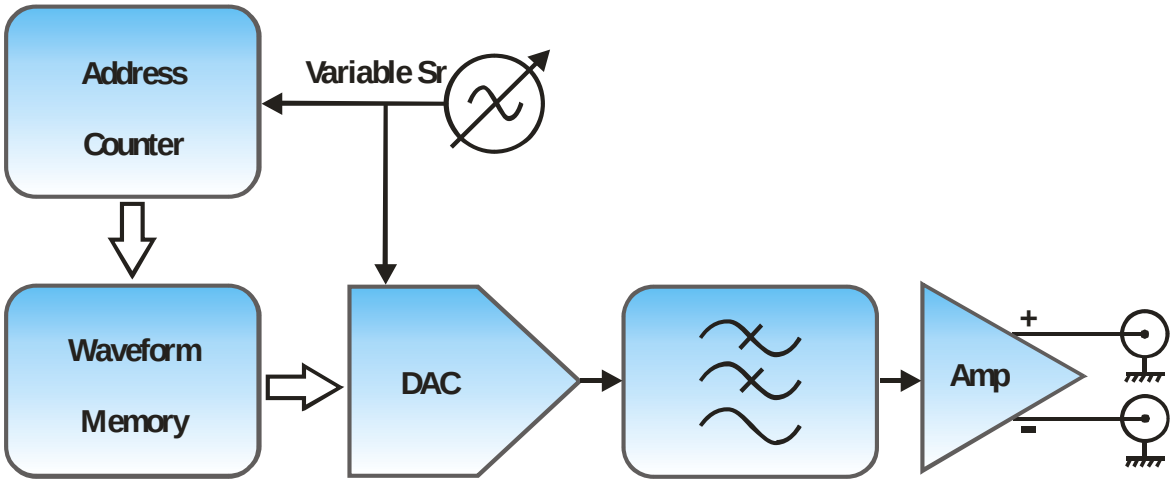


Figure 6: Block diagram for a “true-arb” architecture AWG.

DDS (Direct Digital Synthesis): Here the DAC works at a fixed sample rate while the user controls the repetition rate of the waveform stored in memory. This is accomplished by changing the value to be added to a “phase accumulator” for every DAC clock cycle (see Figure 7). The phase accumulator value is translated into a waveform memory address so samples do not have to be read sequentially. This architecture permits seamless changes to the repetition frequency of the signal stored in the waveform memory and allows for direct frequency sweep generation and PM/FM modulation. For this reason and for its simplicity, this architecture is very popular in low-cost arbitrary function generators, basically digital implementations of function generators.

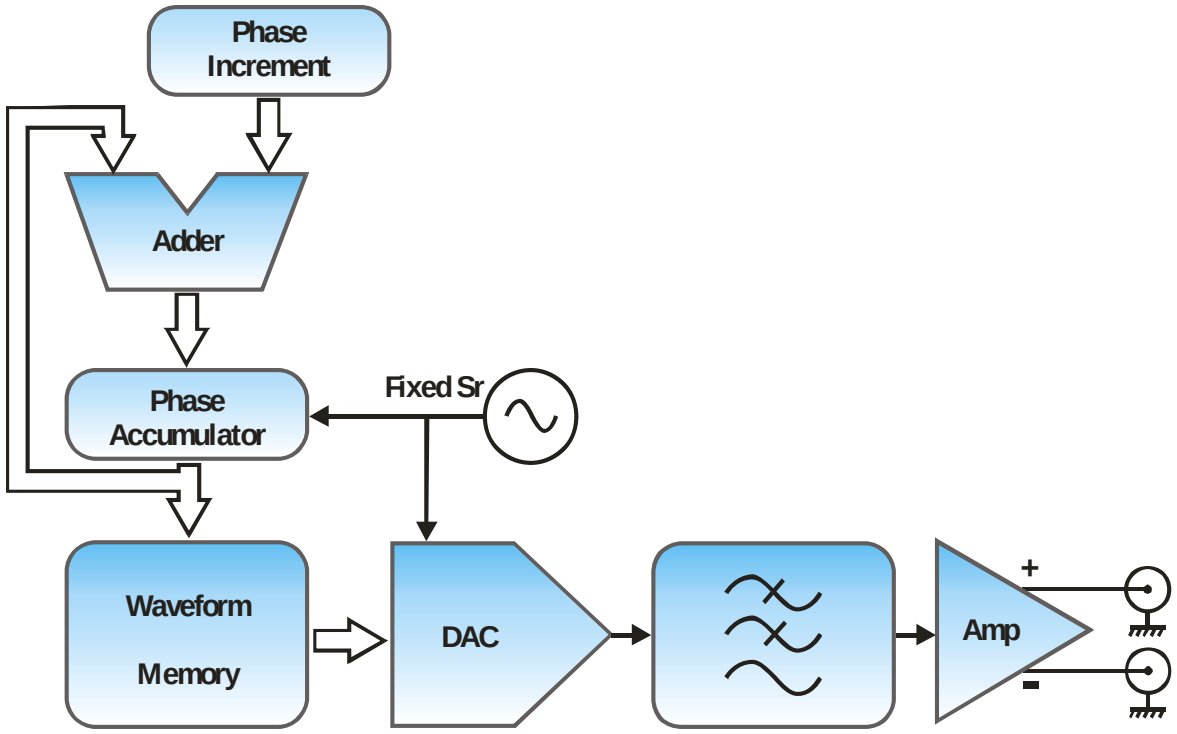


Figure 7: Block diagram for a DDS architecture AWG.

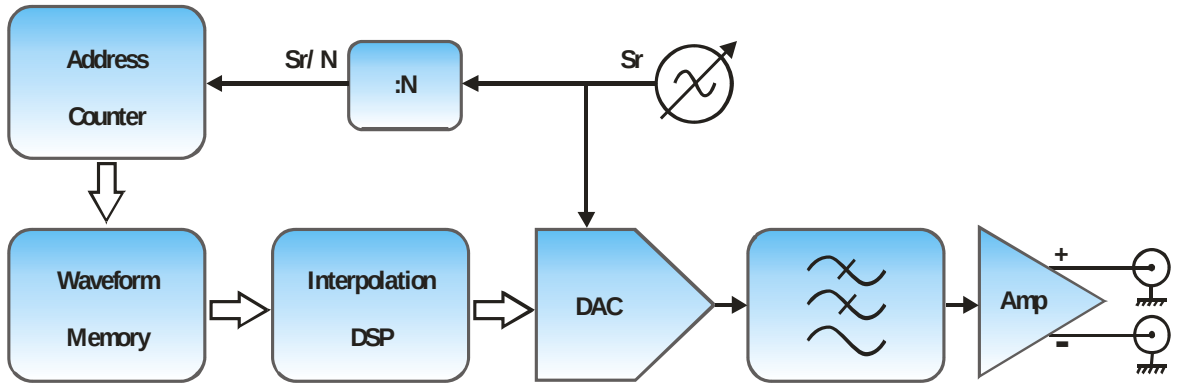


Figure 8: Block diagram for an “Interpolating DAC” architecture AWG.

Interpolating DAC: These AWGs generate signals at a higher sampling rate than the sample access rate. The interpolated samples required to "fill in the blanks" are calculated in real-time by a processing block between the waveform memory and the DAC (see Figure 8). Interpolating functions may be as simple as a straight line (triangular interpolation) or as complex as multiple cascaded FIR (finite-impulse response) low-pass filters to implement near-ideal interpolation. This architecture has some advantages in terms of cost (waveform memory does not need to be very fast) and signal quality, but the maximum frequency component of the signal is still limited by the memory access speed (the actual waveform sampling rate) rather than the DAC sampling rate.

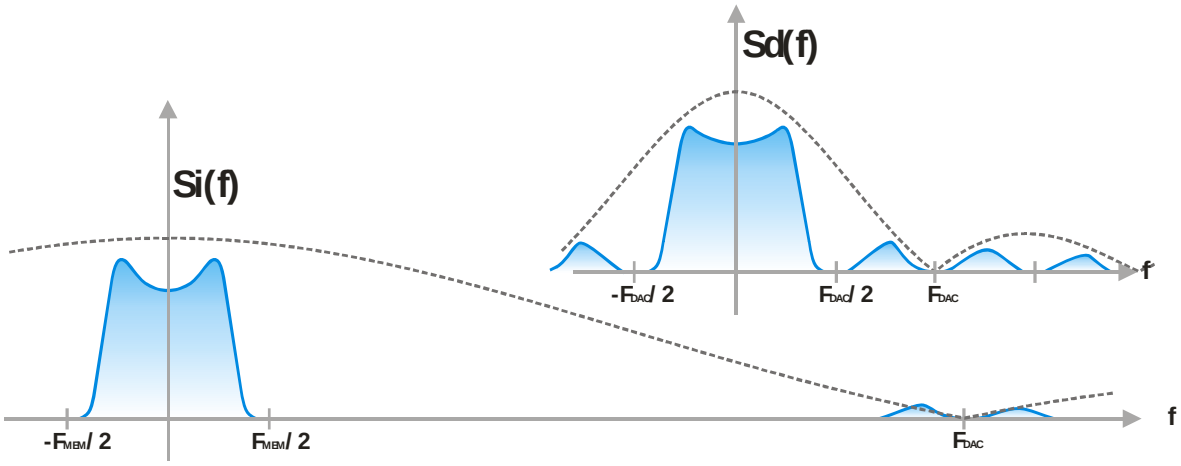


Figure 9: In an interpolating DAC architecture AWG, images are located farther away from the useful signal improving signal fidelity and signal-to-noise ratio and easing the implementation of the analog reconstruction filter.

Interpolating DAC AWGs have some performance advantages over "true-arbs" working at the same waveform memory access speed. As it can be seen in Figure 9, the interpolation/up-sampling process shifts the position of images in the spectrum to multiples of the DAC sampling rate (F_{DAC}) while the maximum achievable alias-free signal frequency component is located at half of the memory access rate ($F_{MEM}/2$). As a result, signal quality compared to a "true arb" with $F_{DAC}=F_{MEM}$ will improve for a series of reasons:

- As FDAC is higher than FMEM, the distance between the main signal and the images will be much higher than that of a corresponding \ “true arb”. This allows for a simpler and more effective implementation of the analog reconstruction filter as there is plenty of room for the filter roll-off.
- The new zero-order hold envelope sinc has its first null at FDAC. As the signal images are relatively closer to the nulls, their amplitude will be lower than in a “true arb”. Additionally, the sinc response is much flatter in the region of interest so signal linear distortion will be lower. Lower amplitude and more distant images result in smoother signals.
- DAC noise will be spread over a large frequency range although its total power depends basically on the DAC resolution. This will result in a lower noise power density and as a consequence, a better signal-to-noise ratio (SNR). This is equivalent to use a higher resolution, lower speed DAC in a “true arb” generator.

Real-Time DSP AWGs: The interpolation process implemented by the interpolating-DAC architecture can be seen as a particular case of real-time digital signal processing of the waveform stored in the AWG’s memory. The availability of faster processors makes feasible the addition of new, interesting functions to even the fastest AWGs available in the marketplace. These include the following:

- Memory size and bandwidth reduction: Memory access speed may be a limiting factor in the performance of AWGs. In some applications, such as the generation of baseband or RF/IF signals for complex modulation found in wireless and optical communication systems, the signal consists of a series of relatively low-speed symbols (compared to the intended sample rate for the final signal) that go through some sort of pulse shaping/bandwidth limiting filter. In a real-time DSP based AWG, the waveform memory could be filled with just the low-speed symbols in the sequence while the pulse-shaping filtering and even the complex modulation of a carrier can be left to the DSP block. The size of the waveform memory (or the time window that can be implemented with it), and the calculation time and complexity required to create the waveform are also impacted by this solution.

- Time and frequency domain corrections: The frequency response is not the same for each channel in an AWG (nor for each AWG unit). This is especially true for very high-speed, very high bandwidth instruments. The best way to obtain high quality results is by carefully calibrating the frequency response for each output and applying some kind of correction. A convenient way to do this in an AWG is to apply a digital filter to compensate for the original distortions. The traditional way to implement these corrections is to apply the filter offline in a computer before downloading the waveform to the AWG's memory. This means that the signal must be recalculated every time it is applied to a different channel or conditions have changed significantly. In a real-time DSP-based AWG, the same signal can be downloaded to any channel and corrections applied in real-time so there is no need to recalculate the waveform. Additionally, this approach can be used to consistently change the overall response of the instrument depending on the targeted application as some of them may require a flat frequency response (complex digital modulation or multi-carrier generation) or low ringing (i.e. pulse or serial-data generation).
- Fast and even seamless changes of signal parameters: Interactive control of the DSP can be used to change signal parameters without having to recalculate the signal. As an example, the carrier frequency or the pulse shaping filter of a complex-modulated wireless signal can be changed without having to recalculate the complete signal. If those changes can be applied without interrupting signal generation and under accurate timing control, then extremely fast (i.e. addition of phase noise to a carrier) or extremely slow (i.e. Doppler shift of the same carrier) variations can be implemented in real-time, without having to recalculate the base waveforms. Traditional, "true-arb" architecture AWGs can implement seamless, long-term variations of any parameter in a waveform by using extremely large size waveform memory sometimes combined with memory segmentation and sequencing.
- Implementation of application-specific signal generation: In many application-specific situations, specialized signal generators are preferred to general-purpose AWGs even when these are capable of generating waveforms with sufficient quality for that purpose. The DSP section can be programmed to perform application-specific tasks in addition to more general-purpose ones, such as resampling or filtering.

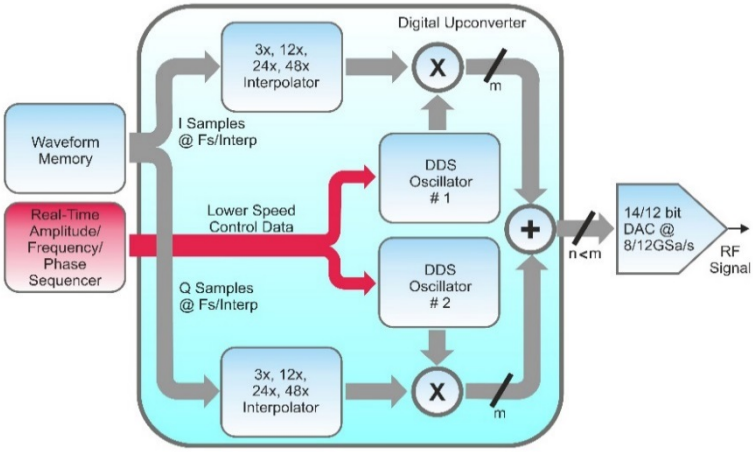


Figure 10: M8190A architecture

Given the typical sample rates found in high-speed AWGs, most DSP blocks are built around a FPGA or proprietary ASIC capable of handling and producing waveform data at the maximum conversion rate even on multiple channels simultaneously. Both the Keysight M8190A and M8195A AWGs incorporate powerful DSP blocks in their architectures.

The 12 GSa/s, two-channel M8190A AWG incorporates a powerful DSP block implemented in an FPGA (see Figure 10). The functionality of this block has been designed to fulfil the requirements of the wireless test market. Today's AWG can provide good RF performance and much more flexibility than traditional vector signal generators (VSGs) (see chapter 6). While AWGs have been very popular to generate baseband signals for quadrature modulation of carriers, most often modulation was produced by an analog circuit with all the potential quadrature imbalances and linearity issues associated with them. The M8190A with its 12 GSa/s can generate multiple modulated carriers from DC up to 6GHz, so potentially it can be used as a self-contained, fully functional RF signal generator. This capability does not come for free in traditional "true-arb" architecture devices.

The sampling rate must be at least twice the maximum carrier frequency, regardless of the symbol rate of the modulating baseband signal. This greatly influences the waveform memory requirements, the time-window that can be generated, and the calculation time for the waveform. Additionally, even if the baseband signal is the same, changing the carrier frequency, a very common need, requires stopping generation, recalculating the waveform, downloading it to the waveform memory and finally re-starting generation.

In the M8190A, the complex baseband signal may be defined as a fraction of the final sampling rate, saving waveform memory and calculation and download time. Memory savings may be up to 12 times the amount required by traditional arbs. The complex samples, read as IQ pairs from the waveform memory, are then up-sampled through an interpolator/antialiasing filter up to the final sampling rate. Finally, a digital quadrature modulator attached to two phase-controlled numerical oscillators (with a DDS architecture) supplies a complex modulated waveform at the final carrier frequency that feeds the DAC. In this scheme, changes in the carrier frequency can be applied "on the fly" without having to recalculate the waveform and without interrupting signal generation at all.

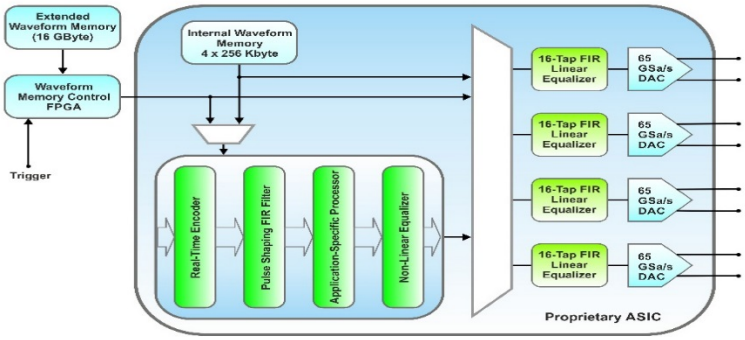


Figure 11: M8195A Architecture

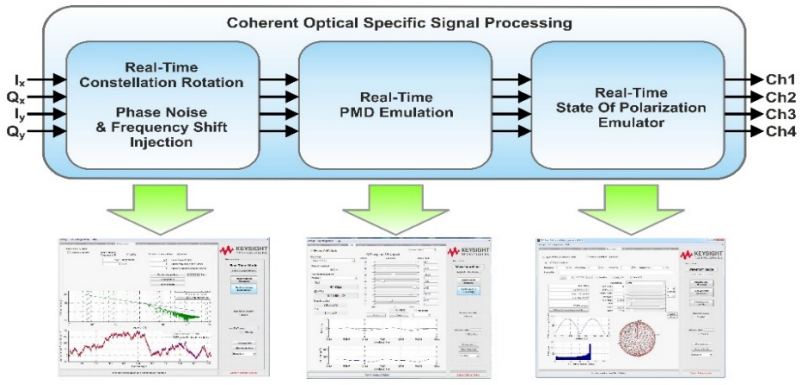
The real-time DSP architecture is even more complex in the M8195A 65 GSa/s, 4-channel AWG. The data throughput required to drive four 65 GSa/s DAC converters requires the implementation of most of the critical functions in a single chip (see Figure 11). The especially designed ASIC developed by Keysight for the M8195A series incorporates the high-speed waveform memory, the DSP block and the DAC converter for all the channels in a single device. High channel density is one of the most important features required by most applications targeted by AWGs in this speed and bandwidth category, such as high-speed serial links and coherent optical communication systems.

As an example, a coherent, dual-polarization multiplexed optical link at a single wavelength requires the generation of four simultaneous multi-level signals at signalling speeds up to 32 GBaud with extremely tight requirements for timing accuracy. An AWG's signal integrity itself is an important issue when signal bandwidths go beyond 20 GHz as is the case in the M8195A. The internal architecture of the core ASIC has been designed with all of the previous requirements in mind. The implementation of the four DACs and all of the clock distribution circuitry in a single die ensures the best possible time alignment between all of the channels while minimizing jitter. The incorporation of 1MSa of ultra-fast waveform memory in the ASIC is key to reach the goal of full-speed signal generation of four independent waveforms in a single device (the M8195A also supports an extremely large amount of external waveform memory, see chapter 3). The same processing speed requirements led to the implementation of the DSP block right in the very same ASIC.

The M8195A is a mixture of general-purpose and application-specific signal processing functions. One of the objectives of the general-purpose section is to minimize memory needs and reduce the speed of transfers between the waveform memory and the DAC. Most high-speed serial and RF/Optical baseband signals can be modelled as a series of multi-level symbols going through some pulse-shaping filter. In "true-arb" architecture generators, the final filtered signal is calculated and then transferred to the waveform memory. As a result, effective generation of the signal requires the transfer of samples to the DAC at the final conversion rate. In the M8195A, it is possible to store just the symbols in the waveform memory, or even just the data bits defining the symbols, so further memory usage reduction is obtained. In the latest case, the DSP block can map the data to symbols in a predefined or user-defined constellation. The symbols read from the waveform memory or created by the real-time encoder can then be oversampled and filtered through a real-time pulse-shaping filter. The taps for these FIR filters can be selected from a pre-defined set, or they can even be defined externally and downloaded. Another general-purpose processing block takes care of applying a non-linear equalization to the signal to feed any non-linear external device (i.e. a Mach-Zehnder optical modulator). Again, the non-linear characteristics of this block can be user-defined and even adjusted interactively for better results. The last step in the general-purpose processing chain is another FIR digital (16 tap). The main purpose for this filter is the linear equalization of each DAC output so a good level of flatness, group delay and channel-to-channel skew may be obtained.

General-purpose DSP may be useful for virtually any potential AWG application. However, it may not be enough to address some application-specific requirements. As an example, one of the target applications of the M8195A is the generation of complex modulated optical carriers found in modern coherent optical communication systems (see Figure 12). The optional coherent-optical application specific DSP block can help in the interactive control (without the need to recalculate the waveform information) and in the addition of some signal characteristics such as PMD (Polarization-Mode Dispersion), SOP (State of Polarization) and even fine-tuning of the optical carrier (through constellation rotation) or its line width (through real-time addition of phase noise in the baseband signals). The continuous, real-time control of these parameters allow the creation of scenarios that last seconds, minutes and even hours, well beyond the reach of any possible waveform memory or sequencing scheme.

Physical implementation of DSP-based architectures is very important given the amount of data to be processed and transferred. The M8195A solution is based in a single chip ASIC supporting the four channels and the associated DSP blocks. This strategy also minimizes jitter and channel-to-channel skew, while adding some extra flexibility. For instance, complexity of a DSP can be increased (i.e. by concatenating two FIR filters) when a reduced number of channels is required.



Pseudo-interleaving DAC:

Figure 12: Coherent-optical application specific DSP block

This class of AWGs combines the output of two channels to double the equivalent sampling rate. This architecture offers a higher equivalent sampling rate through simple summation (see Figure 13). Time skew between the two channels must be exactly half of the sampling period to obtain the desired boost. Odd and even samples are independently classified and written to the waveform memory associated to each channel. This technique effectively extends the usable frequency range, but signal quality is extremely sensitive to timing accuracy and channel-to-channel frequency response mismatch.

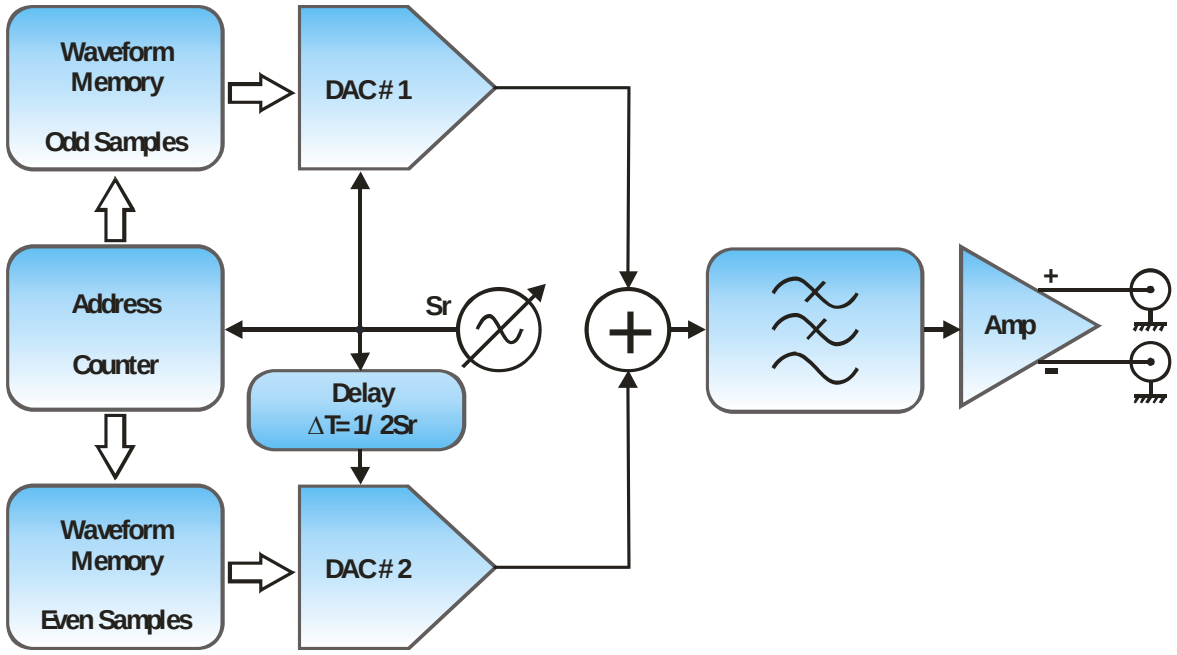


Figure 13: Block diagram for a “pseudo-interleaving DAC” architecture AWG

In this AWG category, each DAC will produce two versions of the same waveform made of interleaved samples. The spectrum of the signal coming out of each DAC will be the same for both channels except for the images around the odd multiples of F_{DAC}, which will invert their relative phase. Combination of both signals will cancel these images. As seen in Figure 14, the resulting spectrum will be similar to that of one single DAC running at twice the speed so images will be located only around the even multiples of F_{DAC}. An interesting side effect, as seen in the figure, is that components of the original signal that would be affected by aliasing in a normal DAC can be recovered through the interleaving DAC summation as the folded-down components around F_{DAC} from each DAC cancel each other. This effect allows extending the usable frequency range (the effective Nyquist frequency) by a factor of two. Addition is a linear operator and the original zero order hold frequency response for each DAC is visible in the combined signal resulting in a zero located at the new Nyquist frequency. This means that, although this architecture effectively extends the frequency coverage, in practice it cannot be extended to its theoretical maximum. Another advantage of this architecture is an improvement of the signal to noise ratio as noise from each DAC will be uncorrelated.

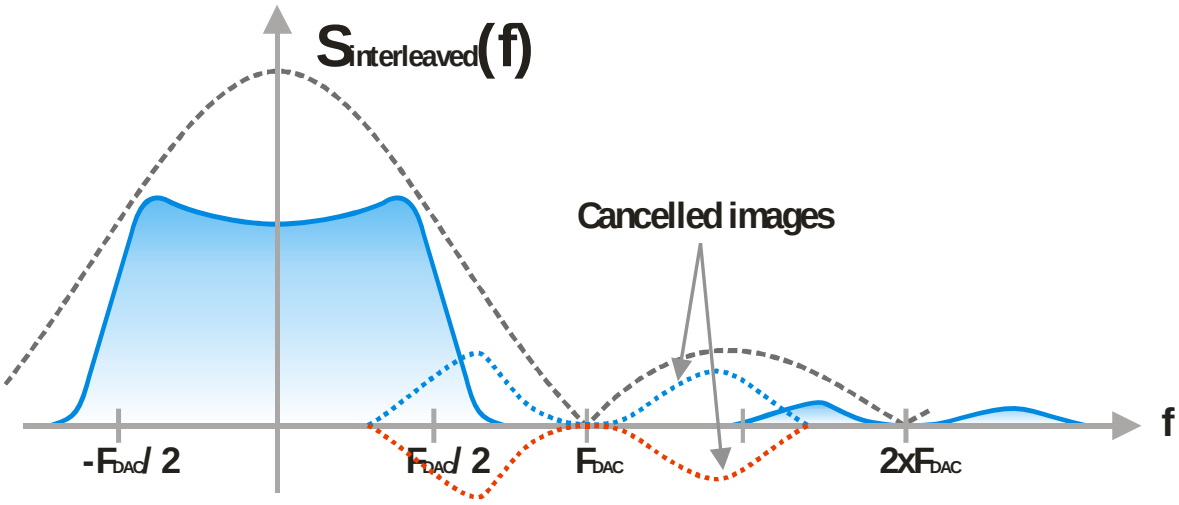


Figure 14: In a "pseudo-interleaving DAC" architecture AWG, signal images from each DAC located around odd multiples of the sampling frequency cancel each other as they have 180° relative phase. This allows the generation of signals with frequency components beyond the Nyquist region for one isolated DAC

The pseudo-interleaving DAC architecture is very sensitive to channel-to-channel mismatches. Differences in the frequency response between the two channels or an inaccurate delay will cause a poor image cancellation. These residual images will interfere with the direct signal when its frequency content goes beyond the $F_{DAC}/2$ limit. Careful alignment between channels may be required to obtain optimal results. Accurate alignment become more critical and difficult to obtain (and maintain) as the sample rate increases.

Although a monolithic AWG instrument can internally implement this architecture, it is also possible to use two independent channels (or even two separate instruments) through the usage of an external combiner if both channels are properly amplitude and time aligned. Timing alignment of two or more channels is a straightforward operation as fine delay adjustment controls are typically available in today's high-performance AWGs.

2.3 DDS versus 'True Arb'

DDS-based and "true-arb" architectures are by far the most popular among AWG users. It is important then to understand the advantages and disadvantages of each class. The DDS architecture allows for a simple and low-cost implementation of an AWG as it requires just a fixed sampling clock and, as a consequence, a single interpolation low-pass filter. Changes in the phase increment value result in an instantaneous, seamless change in the pace the waveform memory is scanned so the output waveform frequency changes accordingly.

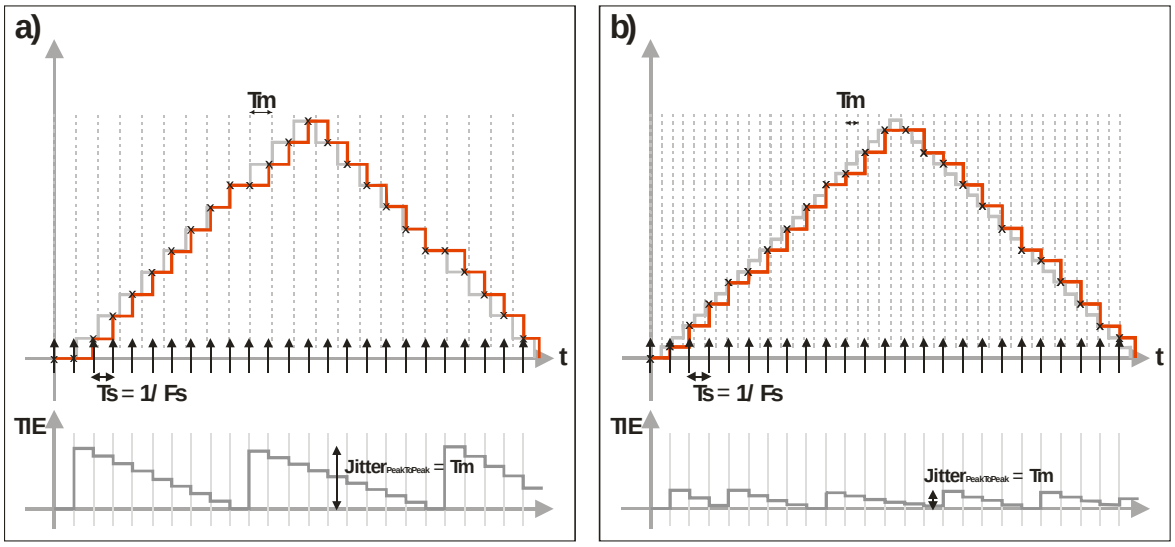


Figure 15: Generation of a triangular wave in a DDS architecture AWG with two different waveform memory settings. Time resolution in a) is lower than in b). TIE (Time Interval Error) graphs for each case shows the lower distortion of the highest resolution waveform although the output sampling rate remains the same.

Depending on the phase increment value, a given sample may be read more than once, or one or more contiguous samples may be skipped. Generally speaking, the phase accumulator is a pointer to the waveform memory that contains fractional values so the actual pointer (to the sample fed to the DAC) should be the result of rounding its current contents. While this behaviour may be acceptable for relatively low frequency, smooth waveforms, it can be unacceptable when spectral purity and timing accuracy (both are interlocked) are important given that the DDS architecture inherently produces jitter in the output waveform.

In a DDS-based AWG, a sufficiently detailed mathematical representation of the signal is written into the waveform memory. Waveform memory can be seen as a look-up table where the generator finds the corresponding amplitude value for a given instant in time. Instead of adjusting the sampling rate for the output, as done in "true-arbs", DDS-based AWGs use a fixed value and the user has to set the desired duration of one occurrence of the signal, T_{signal} (or its inverse, the repetition frequency). The instrument will then compute the corresponding phase increment, $\Delta\phi$, required to obtain that repetition rate:

$$\Delta\phi = 360^{o} \times Ts / T_{signal}$$

The phase increment can be translated into samples:

$$\Delta n = Ts \times RL / T_{signal}\ ,\ RL = \text{Waveform memory length}$$

Δn will not, in general, be an integer number. Waveform memory pointer for DAC sample #i and continuous playback can be expressed as:

$$\text{Pointer} = \text{round}((\Delta n \times i) \bmod RL)$$

The modulus operator results in a continuous playback without phase discontinuities. Generally speaking, each iteration of the signal will not be played back in exactly the same way as the samples read from the waveform memory and the timing distance relative to the DAC sampling instant will change.

A good way to understand the DDS behaviour is by considering the generation of a triangular wave, as shown in Figure 15. For simplicity, timing for the target waveform will be chosen so the waveform memory will contain exactly one cycle. In this example, repetition rate will be set first so the phase increment Δn (expressed in waveform memory samples) will be slightly smaller than one (Figure 15a), so the DAC sampling period, Ts, is shorter than the time resolution, Tm, of the waveform stored in the memory (Ts<Tm). When this signal is applied to a DDS architecture AWG, the first sample will be output as the value of sample #0 while the next sample will be defined by rounding phase increment to the nearest sample in the waveform memory and so on. As it can be seen, each sample coming out of the DAC will consist of the nearest one in the waveform memory. Although in this example all of the samples stored in the waveform memory will be used at least once in a single pass, the triangular wave looks distorted by the jitter introduced by the sample mapping process. A simple analysis of the potential relative location of the output samples versus the ideal time location of the samples in the waveform memory will give a peak-to-peak jitter of Tm with a uniform statistical distribution. Increasing the time resolution for the waveform stored in the AWG memory will then reduce the jitter and, as a consequence, the signal distortion. The signal resolution has been doubled in Figure 15b and, as expected, the signal distortion has been reduced as well as the peak-to-peak jitter. This has been accomplished at the expense of doubling the size of the waveform memory. Given the ratio between Ts and Tm, not all of the stored samples will be used in a single pass now, but, potentially all of them may be used in some iteration during continuous generation.

The previous discussion leads to the conclusion that the smaller the timing resolution of the signal stored in memory, the better results the results that will be obtained. Ideally, Tm should be much shorter than Ts for all the intended signal repetition rates. This may not be a viable solution in actual instruments as it can result in memory sizes much higher than available or time windows much shorter than expected. As a rule of thumb, the timing resolution for the samples in the waveform memory should be much better than the AWG's fixed sampling rate applied to the DAC for any intended repetition rate.

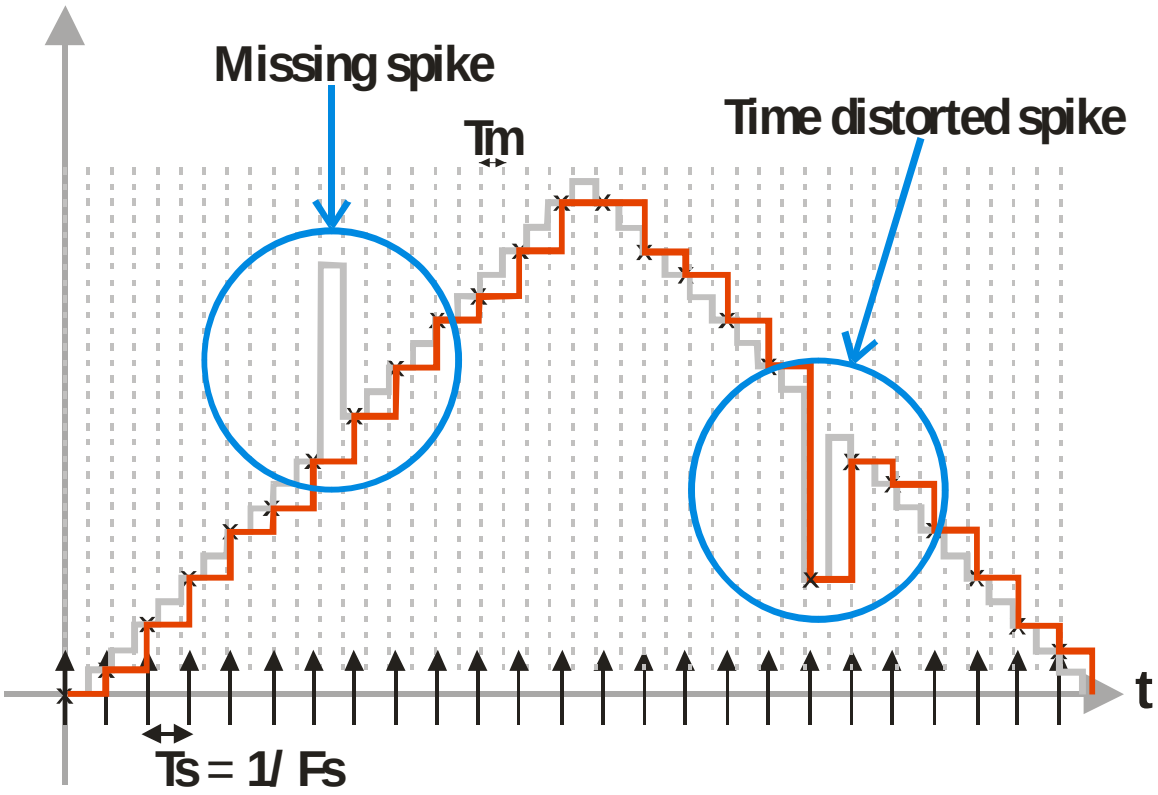


Figure 16: Fast signal features may disappear or be distorted randomly due to an insufficient output sampling rate

Another undesired effect appears when Ts is larger than Tm as not all of the samples in the waveform memory will be used in a single occurrence. This effect can be easily understood by analyzing the signal in Figure 16 where two, one-sample spikes, have been added in each slope of a triangular wave. The first spike will not be visible at all in the output as the sample representing this value will not be read in this particular iteration, while the second spike will show-up although its width will not be accurate. In different occurrences of the waveform, these two spikes may appear or disappear in a random way depending on the relative location of the DAC sampling instants.

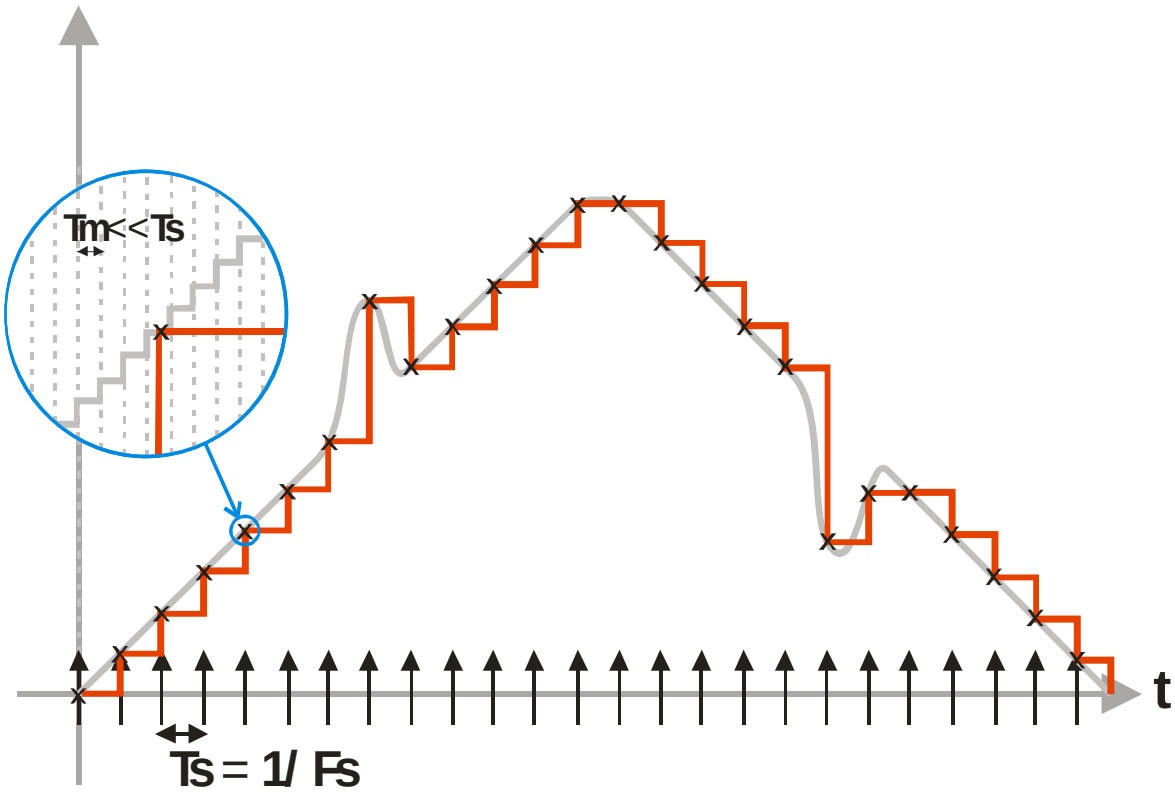


Figure 17: Increasing the timing resolution of the waveform memory and adapting the signal BW to the output's sampling rate will reduce jitter and improve the waveform playback consistency.

The basic problem is that the DAC's sampling rate is not fast enough to faithfully reconstruct the signal as it is stored in memory, a potential issue for any AWG regardless of its architecture. However, the apparent randomness of the spikes at the output is a unique unwanted side effect of the DDS architecture.

This problem may be minimized if the waveform stored in memory is smooth enough (low-pass filtered) to make sure all of the signal details can be reconstructed. Figure 17 shows the same signal after applying a low-pass filter while it has been rendered in much greater detail into the waveform memory. Most of the jitter is gone as there is a wealth of samples in the memory close enough for any potential location of the DAC samples. Additionally, any detail of the signal will be reproduced properly as the low-pass filtering smoothes the signal so the spikes now will be wider than Ts. The high-frequency sections of the triangular wave itself, such as the top corner, will be smoothed as well.

The above problems do not show up in “true-arb” AWGs as every sample will be used only once in each pass. Although changing the timing of the signal will require either changing the sampling clock, loading a faster or slower version of the waveform into the generation memory, or both. In every situation the generator will stop the output for a certain period of time as these changes cannot be made “on the fly”. One trick to obtain repeatable, jitterless signals with a DDS architecture AWG is taking special care when defining the waveform samples and the instrument settings so Δn is an integer. In this way a DDS AWG behaves exactly as a “true-arb” architecture generator, but most of the DDS AWG’s signal generation flexibility and features are lost in this way.

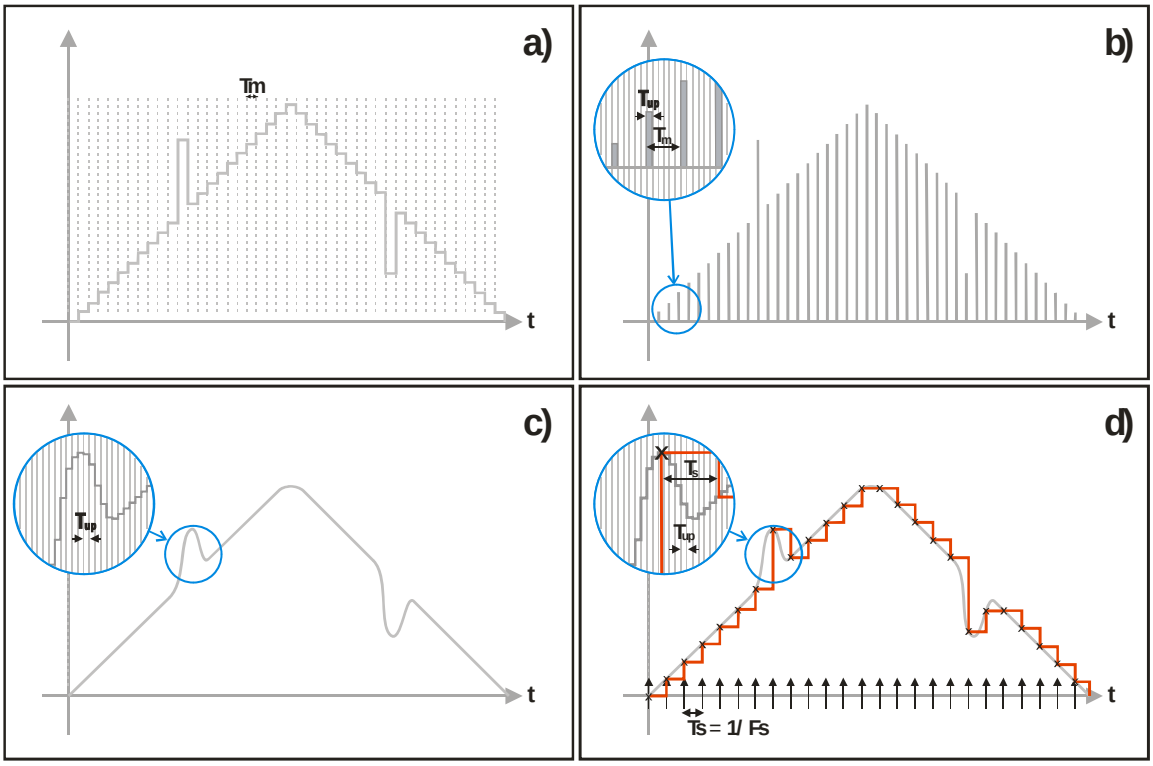


Figure 18: Waveform processing in a True*form* AWG. The original signal contained in the waveform memory (a) is up-sampled through “zero-padding” (b), then a digital low-pass filter interpolates the intermediate samples (c) so they can be decimated and applied to the fixed rate DAC used by True*form* AWG generators (d).

2.4 The True*form* Architecture

Keysight, with the 33500 and 33600 series of function/arbitrary waveform generators, has introduced a new AWG architecture known as True*form*, with all of the flexibility and advantages of DDS-based AWGs, but without any of the drawbacks (Figure 19). In this architecture, the Ts>>Tm condition is met by interpolating the samples stored in the waveform memory (Figure 18) in real-time through a DSP block including a FIR low-pass filter. In this way, the jitter associated to time mapping of the samples stored in the waveform memory can be virtually eliminated without having to improve their original resolution, so valuable waveform memory is saved and potential time windows extended. The filter cut-off frequency and shape are adjusted so the frequency contents of the resulting signal can be reproduced faithfully by the instrument's DAC. Finally, the filtered, up-sampled signal will be decimated to match the DAC fixed sampling rate. This arrangement can be seen as a real-time re-sampling system, and it will always use all of the available samples in the process, as all of them will be fed to the processing block. Fast features in the signal that would be skipped randomly in traditional DDS-based generators will be now consistently shown with very small jitter. This architecture offers an equivalent much longer record length thanks to the real-time interpolation process.

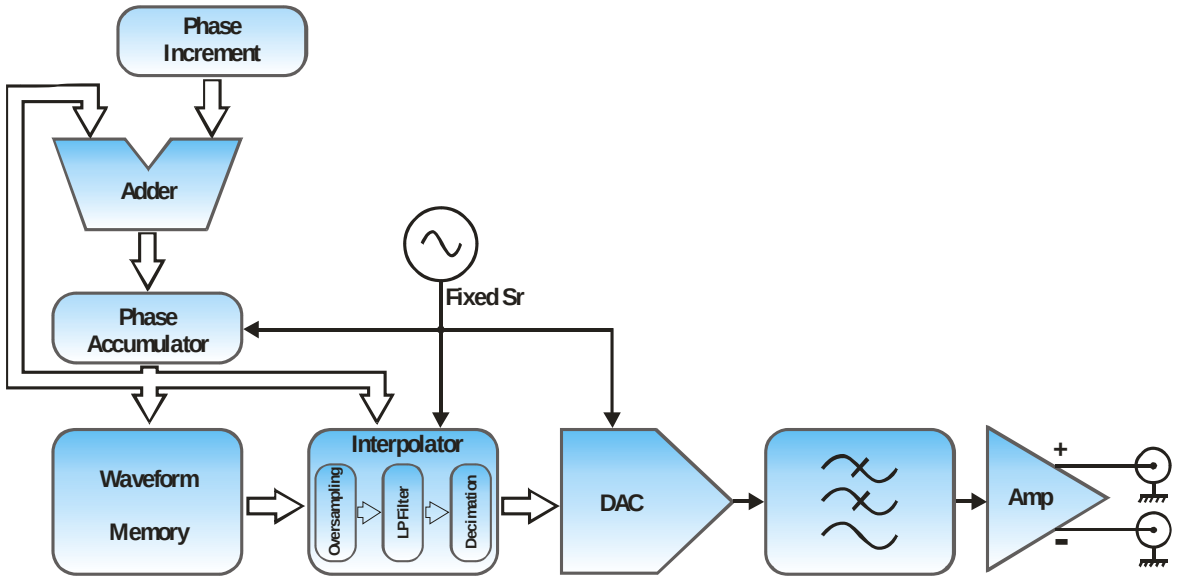


Figure 19: Block diagram for a (True*form*) architecture AWG

The low-pass digital real-time filter can also be used to improve the time or the frequency response of the output. As an example, in the Keysight 33500 and 33600 series, the combination of a high sampling rate related to the instrument bandwidth (oversampling) and a good analog interpolation filter at the output, results in very clean, image-free signal. Its digital filter is designed to compensate the DAC's frequency response and two filtering modes can be chosen: "brick-wall" for flat frequency response (i.e. for IQ, multi-tone, or IF signal generation) or "gaussian" to obtain a step response with fast rise-times but without any ringing (i.e. for pulse or pattern generation).

2.5 Digital to Analog Convertor (DAC) Quantization Noise

Probably, the most critical component in any AWG is the Digital-to-Analog Converter or DAC. It is in the DAC where a mathematical description of a signal becomes a real-world analog entity. It is also in the DAC where the performance of the instrument and the quality of the generated signals are influenced the most. The goal of any DAC is to convert a numerical code into an analog electrical magnitude, typically a voltage, and to do it with the highest accuracy. There are many possible circuit designs capable of implementing this basic function. However, this document will focus on those with the capability of producing analog samples at speeds of tens, hundreds, or even thousands of Mega samples/s (MSa/s), which are the typical conversion rates found in today's AWGs.

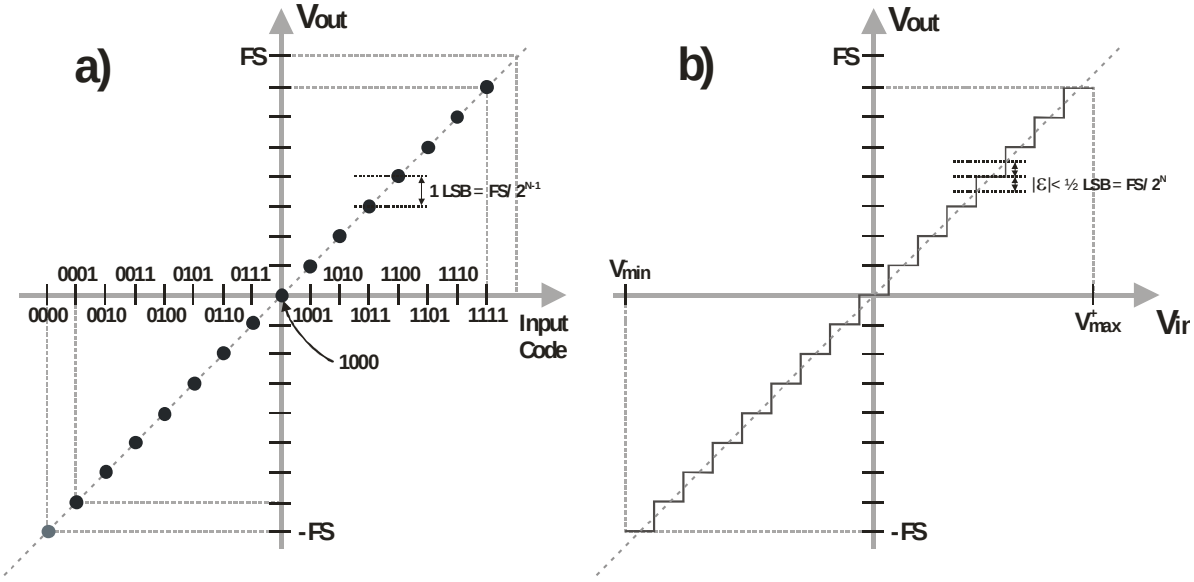


Figure 20: Transfer functions for a 4-bit bipolar DAC as a function of the input code (a) and the target voltage (b). For any target voltage there will be a quantization error ε limited to ±1/2 LSB.

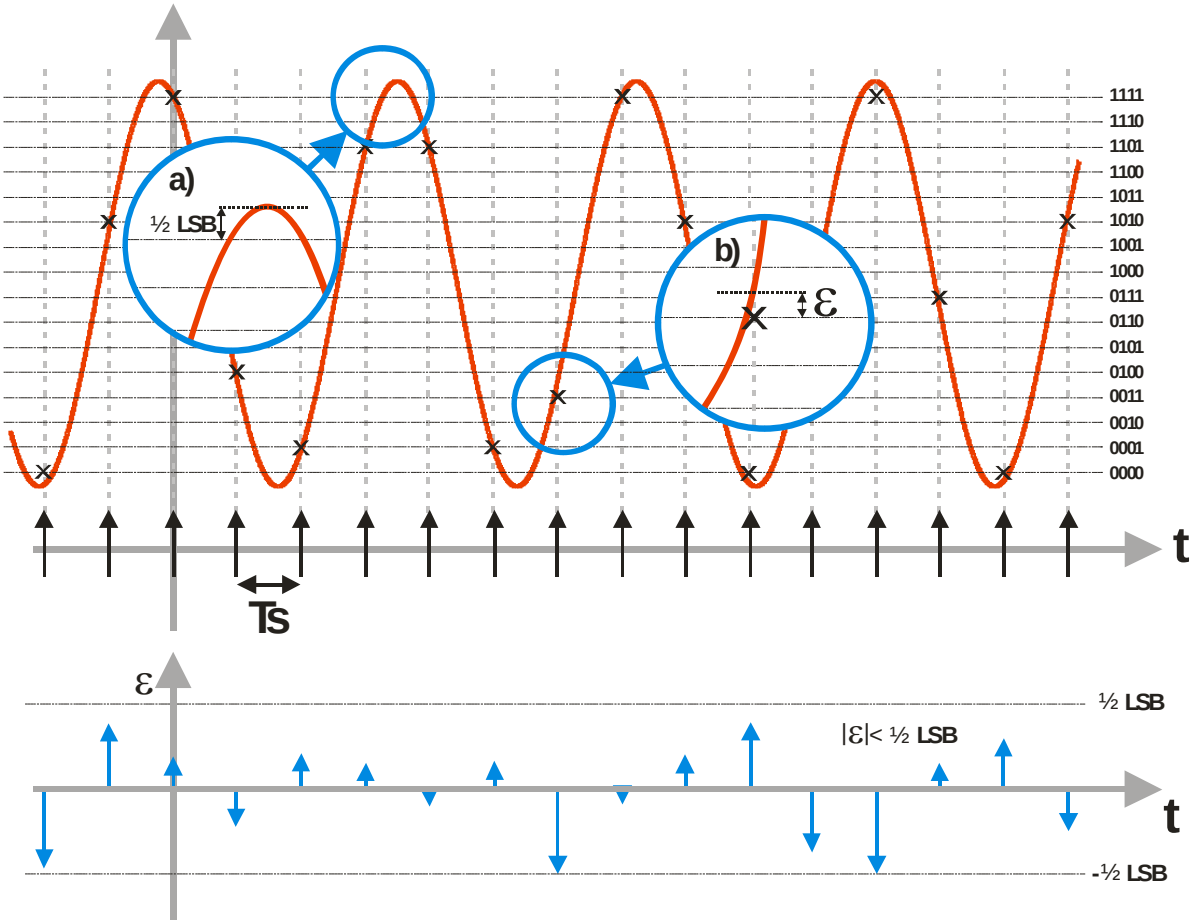


Figure 21: Quantization error can be analyzed as a sampled waveform itself. The resulting quantization noise waveform ε(N Ts) is bounded to ±1/2 LSB and the amplitude distribution uniformly over the whole range. Amplitude of the input waveform can go beyond by 1/2LSB (a) and still keep the bounded error criteria (b).

Before starting to discuss circuit architectures, compare them, and establish whether or not they are suited to the needs of real instruments, it may be interesting to first discuss the desired characteristics of an ideal DAC, and then the kind of impairments that can be found and characterized in actual devices. Typically, the desired response of a DAC should be linear. Linear means that the DAC output voltage is somehow proportional to the input code. As seen previously, the sampling process implies a quantization in the time domain.

DACs add another quantization layer, now in the amplitude domain. Most DACs use binary integer words of N bits as an input. Using all of the possible input combinations gives 2^N possible amplitude values. The N parameter is also known as the DAC resolution and it is measured in bits. It is a logarithmic parameter so increasing it one bit means doubling the number of levels the DAC can generate. Quantization level, the size of the quantum or minimum step size, for a bipolar DAC with a given voltage range or full scale (FS) will be:

$$Q_{level} = FS / 2^{N-1}$$

Quantization level is also known as 1 LSB as it is also the output change produced by the least significant bit. As seen in Figure 20a, the ideal DAC transfer function will look like a set of points over a straight line. Minimum and maximum voltages may be arbitrarily set although most converters are either unipolar (range goes from 0 to FS) or bipolar (range goes from –FS to FS) as shown in the figure. For bipolar DACs, there will be some asymmetry if the middle code is assigned to the zero voltage, as there is one more voltage level below zero than there is above zero. It would be possible to obtain a totally symmetrical linear transfer function, but then there would be no code resulting in exactly zero volts at the output (Figure 20b). Although asymmetry can be noticed easily in low resolution DACs, most AWGs have 10 or more bits of resolution so the lack of symmetry is negligible and full scale for both negative and positive maximum output level can be approximated by +FS and –FS respectively. It is also useful to consider the transfer function between the desired voltage (V_{in}) and the output voltage (V_{out}) as seen in Figure 20b. Waveforms are typically calculated by software processing numbers with a much higher resolution (i.e. floating point), than the target AWG. In some way, the original mathematical representation of the waveform may be seen as a very accurate version of the continuous, analog signal of interest. The V_{in}/V_{out} DAC transfer function gives a clear view of the amplitude quantization process and the errors involved, even in a perfect DAC.

Therefore, even ideal DACs produce errors in the signal. A simple analysis (Figure 20b) provides a limit for the error ε in a bipolar DAC:

$$|\varepsilon| < 1/2\ LSB = 1/2\ FS/2^{N-1} = FS/2^{N}$$

The output signal can be seen as the combination of the input signal and an error signal, also known as quantization noise (Figure 21). Unlike AWGN (Additive White Gaussian Noise), which is unbounded, quantization noise distribution is uniform over a limited range. A simple statistical analysis provides an rms amplitude value for it:

$$\varepsilon_{RMS} = Q_{level} / 12^{1/2}$$

For a truly random quantization noise, its spectrum is flat like white Gaussian noise, and noise power is evenly distributed between DC and FS/2. The resulting SQNR (Signal-to-Quantization Noise Ratio) will depend on the signal being generated. A useful example may be a sinusoidal signal s(t) with a FS amplitude:

$$s(t) = FS \times \sin(2\pi ft) = ((Q_{level} \times 2^N)/2) \times \sin(2\pi ft)$$

$$S_{RMS} = FS/2^{1/2} = (Q_{level} \times 2^N)/(8^{1/2})$$

Signal to noise ratio can be obtained by combining the root mean square (RMS)expressions for the sinusoidal and quantization noise signal amplitudes:

$$SNR(dB) = 20 \log_{10}(S_{RMS} / \varepsilon_{RMS}) = 20 \log_{10}(2^N) + 20\log_{10}((3/2)^{1/2}))$$

$$SNR(dB) = 6.02N + 1.76dB, \quad DC < f < Fs/2$$

The previous result implies that adding one bit of resolution will increase SQNR by about 6dB. For a bandwidth limited signal with bandwidth B, out-of-band noise could be eliminated through filtering without affecting the signal. In this case the following expression may be more meaningful:

$$SNR(dB) = 6.02N + 1.76dB + 10\log_{10}(Fs/2B), \text{ where B=bandwidth}$$

A different way to read the above equation can be used with interpolating DAC AWGs. These devices achieve interpolation by applying an oversampling/filtering process to the waveform samples in real-time. As signal bandwidth is limited by the sampling rate of the waveform in the generation memory rather than the effective DAC conversion frequency, oversampling in the DAC will result in an improvement of the SQNR (Figure 22). This is equivalent to use a higher resolution DAC. Theoretical resolution improvement (in bits) can be obtained using the following expression:

$$\Delta bits = 10\log_{10}(\text{Oversampling Factor})/6.02$$

This means that increasing the sample rate by a factor of four (4x) is equivalent to improving the DAC resolution by one bit in a non-interpolating AWG.

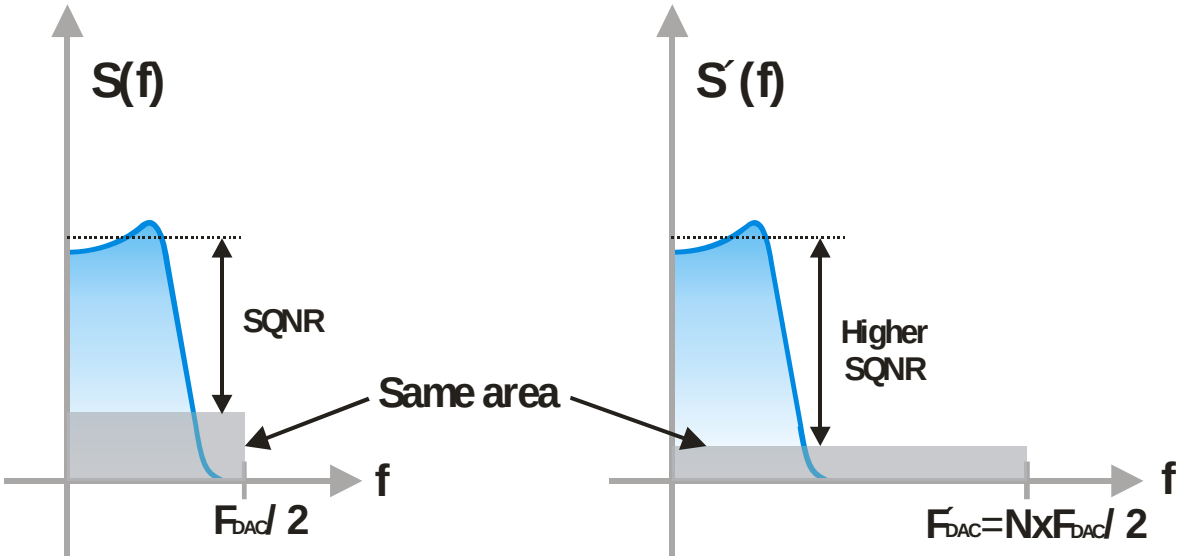


Figure 22: Quantization noise is spread uniformly over the complete Nyquist band. For a bandwidth limited signal, noise power density can be reduced by increasing the sampling rate (oversampling) as the same overall power is spread over a larger band. This effect is exploited by interpolating DAC AWGs.

DAC resolution is a very important figure of merit for any AWG and, as a general consideration, the higher the better. However, there is a trade-off between sampling rate (and signal bandwidth) and resolution, and restricting the resolution to a reasonable limit will provide good-enough performance at an acceptable cost. Quantization noise is just one of the components of the overall noise present at the generator's output. Even ideal AWGs have another unavoidable source of noise: thermal noise. Thermal noise power is proportional to bandwidth, and it is caused by the random motion of charges within any electronic device or system. Its distribution over frequency is close to uniform, and the statistical distribution of its amplitude is Gaussian, so it can be handled as AWGN (Additive White Gaussian Noise). Its RMS amplitude is given by the following expression:

$V_{thermal}=(4kTBR)^{1/2}$, k=Boltzmann's constant, B in Hz

It can be also expressed in terms of spectral density:

$V'_{thermal}=(4kTR)^{1/2}$ Volts/$Hz^{1/2}$

For a typical lab temperature (300K) and a 50Ω source impedance the above expression can be simplified to:

$V'_{thermal}=0.91nV/Hz^{1/2}$

It is clear that incorporating a DAC with a quantization noise lower than thermal noise does not make any sense, as the DAC resolution will not improve the quality of the output. A good exercise is to compare ideal quantization noise with thermal noise in terms of power density. Quantization noise power density depends on three variables: full scale amplitude, DAC resolution, and sampling rate. Figure 23 shows quantization noise power densities for a 12 and 14 bit DACs as a function of sampling rate and for a 0.7Vpp full-scale voltage (these numbers correspond to the two M8190A AWG operating modes). As the graph shows, quantization noise intercepts the thermal noise level at specific sampling rates (within the M8190A AWG sampling rate ranges for each supported DAC working modes).

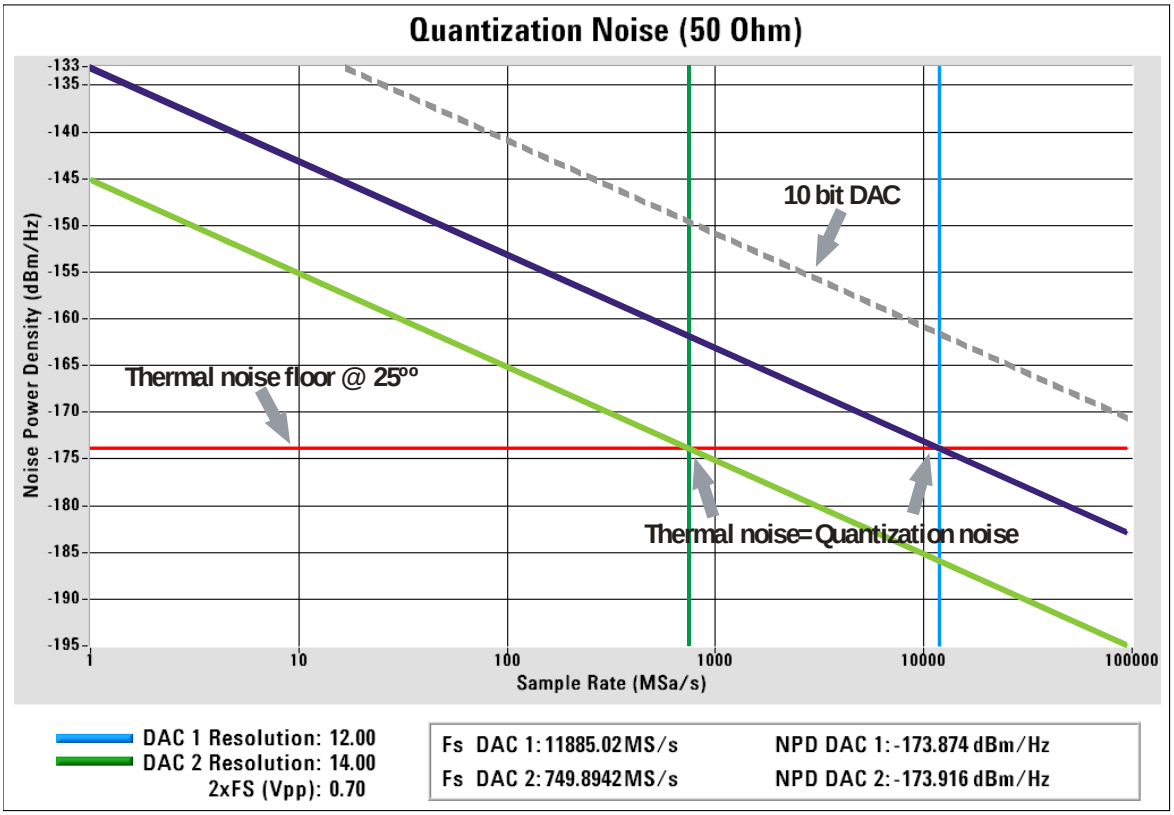


Figure 23: this plot shows quantization noise power density vs. sample rate for ideal 12 (blue trace) and 14 bit (green trace) resolution DACs and 0.7Vpp full scale output range. The horizontal red line shows the thermal noise power density (-174dBm/Hz), which is independent of sampling rate. Quantization noise and thermal noise are equal at 750MSa/s for a 14 bit DAC and at 12 GSa/s for a 12 bit DAC. The grey dashed trace corresponds to a 10 bit DAC which intercepts the thermal noise level at sample rates well beyond 100 GSa/s. In the Keysight M8190A AWG, with 14 bits resolution at 8GSa/s and 12 bits at 12 GSa/s, quantization noise is negligible in front of other noise sources, however, it may not stand true for a 10 bit instrument running at the same speed.

It is important to note that thermal noise is just the minimum unavoidable noise that cannot be reduced unless the operating temperature is lowered. Other sources of noise include clock and data feed-through, switching glitches, clock jitter, or crosstalk from other channels or the clock signal. All of these sources of noise and distortion are also part of the DAC output so resolution requirements for the high-speed DACs in use in high-performance AWGs are even less demanding. As an example, 12 bits of resolution for a 12 GSa/s AWG may be more than enough to make quantization noise insignificant in the overall SNR equation. For a 10 bit resolution ideal DAC with 0.7Vpp full-scale voltage, the intercept point is located well beyond 100 GSa/s, so quantization noise will be more than 12 dB over thermal noise @ 12 GSa/s.

2.6 AWG Non-Linearities

Real world DACs are not perfect and transfer function deviates from the ideal response. Static non-linear behaviour is quite easy to characterize as modern DMMs (Digital Multi-Meters) can measure each quantization level with resolution and accuracies much better than 1 LSB for any practical DAC resolution. Figure 24 shows a hypothetical result of such characterization for a bipolar DAC. The actual input code to output voltage (black line) does not follow precisely the ideal response (dashed grey line). Distortion may be split between linear and non-linear components. The main practical difference is that the linear component will not result in harmonic or inter-modulation distortion in the output waveform as opposed to the non-linear components.

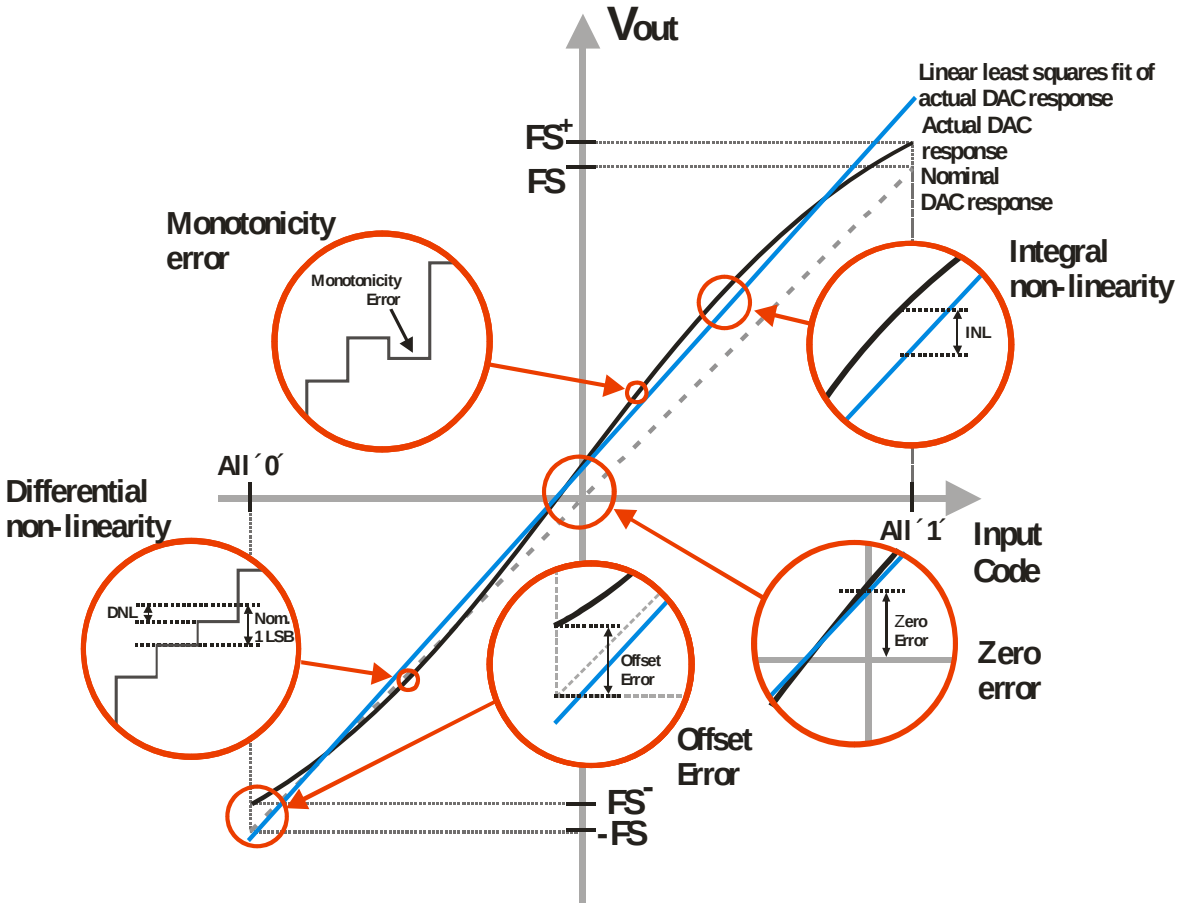


Figure 24: DAC (and AWG) non-linear response may be characterized using different concepts, depicted here. Although any deviation from the ideal DAC response is part of it, offset and gain errors must be subtracted in order to properly estimate non-linearity. This may be accomplished by obtaining a linear least squares fitting (linear regression) of the actual DAC response to be used as a new reference.

Linear distortion is composed of the following components:

- Gain Error: Difference between the slopes of the actual and ideal transfer function.
- Offset Error: Difference between the actual and ideal “all 0s” voltage levels.
- Zero Error: Actual output voltage for the theoretical “zero-voltage” code, typically the midrange code. This parameter only makes sense for bipolar converters.

The effects of linear distortion can be easily compensated for through simple changes in the AWG settings or the samples in the waveform memory so it is quite useful to extract them from the analysis of the non-linear components. To do so, the ideal response used as a reference should be replaced with a new linear model incorporating the linear impairments. As the actual response will not in general be a straight line, some method must be used to obtain a good enough linear approximation. A simple way would be to use the minimum (all zeros) and maximum (all ones) levels as the extreme points. The problem with this method is that deviations from the linear approximation depend strongly on the shape of the actual response. A better approximation may be obtained by using a least-squares linear fit (linear regression) process. A first order polynomial will be then obtained:

$$V'_{out}=A \times C_{in} + B$$

Actual gain is defined by the A coefficient so gain error can be easily established. A new normalized LSB value can be also defined from the above polynomial:

$$LSB'= (V'_{1s}-V'_{0s})/(2^N-1)$$

V'_{1s} = Interpolated “all ones” output voltage

V'_{0s} = Interpolated “all zeros” output voltage

This value can be used as a better reference to compare the actual size of each quantization level in the DAC response such that the effect of linear distortion, although typically small, is cancelled. With these new reference entities, it is possible to define the basic non-linear impairments:

- Differential non-linearity (DNL): It is the difference between the size of each step in the response and the normalized ‘LSB’ value in LSB units. For a given DAC, the absolute value of the worst case is usually specified. An interesting situation occurs when DNL<-1LSB for some input codes as this means that for those codes the DAC loses the monotonic behaviour. This condition is known as monotonicity error.
- Integral non-linearity (INL): This is the maximum difference between the actual response and the linear fit approximation. It can be specified in LSBs or as a percentage of the DAC’s full range.

The previous concepts may be applied to DACs as a component or to a complete system where DACs are incorporated such as AWGs. Within an AWG, converters are typically connected to an output stage including amplifiers, switches, and filters.

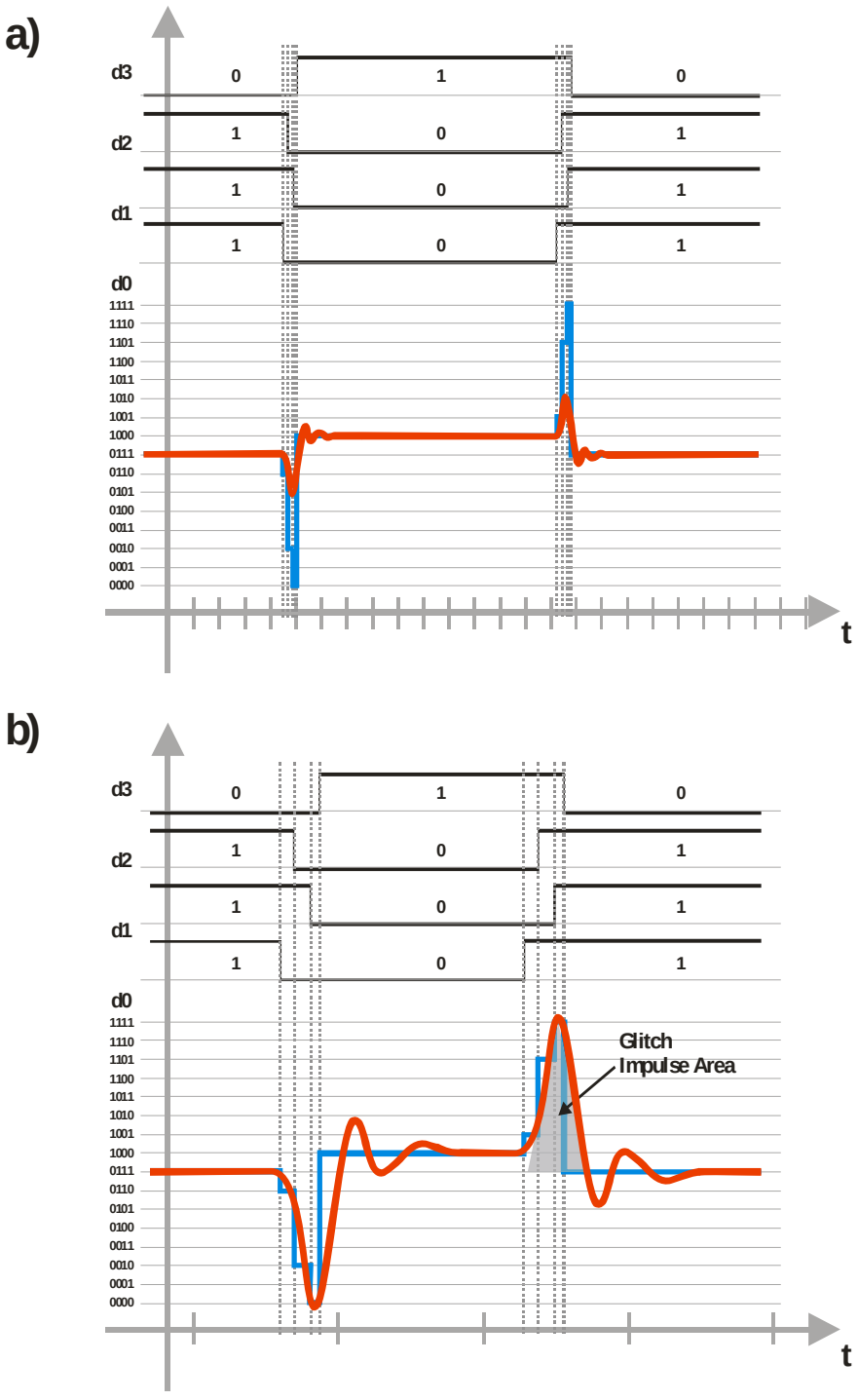


Figure 25: Switching glitches are caused by timing differences between control bits in the DAC. Here, a major-carry transition is shown for a 4-bit DAC running at low (a) and high (b) sampling rates, showing the growing importance of this distortion as sampling frequency and analog bandwidth increase.

Non-linear distortions, noise and other spurious signals (spurs) added by those additional blocks will influence the overall performance. Some high-performance AWGs incorporate a “Direct DAC” output mode to minimize non-linear distortion and other errors, while maximizing bandwidth at the potential expense of limiting control on the signal’s amplitude and bandwidth.

Static non-linearity does not provide a complete picture of the DAC’s (and AWG’s) non-linear behaviour. Dynamic non-linear behaviour is an important contributor to the AWG’s impairments, especially around transitions in the output. Glitch impulse (or switching glitch) is one of the most important sources of dynamic distortion. It is mainly associated to timing differences in the switching of the different elements in the DAC. These timing differences may be caused by skews in the waveform data parallel lines feeding the DAC, internal decoding circuitry, and/or differences in the response time of the analog sections within the DAC. Glitch impulse is very sensitive to the internal DAC architecture, and the sequence of codes that are applied to it. The amplitude of the glitch tends to be higher when more MSB bits are switched, so a worst case scenario appears in the midrange when input codes transition from 011...11 to 100...00 or vice versa. This scenario is referred to as a “major-carry transition”. It is interesting to note that when a major-carry transition occurs, the output changes by just one LSB, so a tiny amplitude change in the DACs output can potentially be associated with a worst-case glitch. Figure 25 shows such an event for a hypothetical 4-bit DAC running at two different speeds. One important conclusion is that the higher the sampling rate, the more the effect will be visible on the output signal. There are two basic reasons for this: skews will take a higher percentage of the sampling period, and the wider bandwidth reconstruction low-pass filter will extract less energy from the glitch, as most frequency components of it will be located within the Nyquist band. Glitch impulse is typically characterized by measuring the total area (in units of volts-per-second) covered by it (positive and negative portions partially compensate each other) or just for the highest impulse.

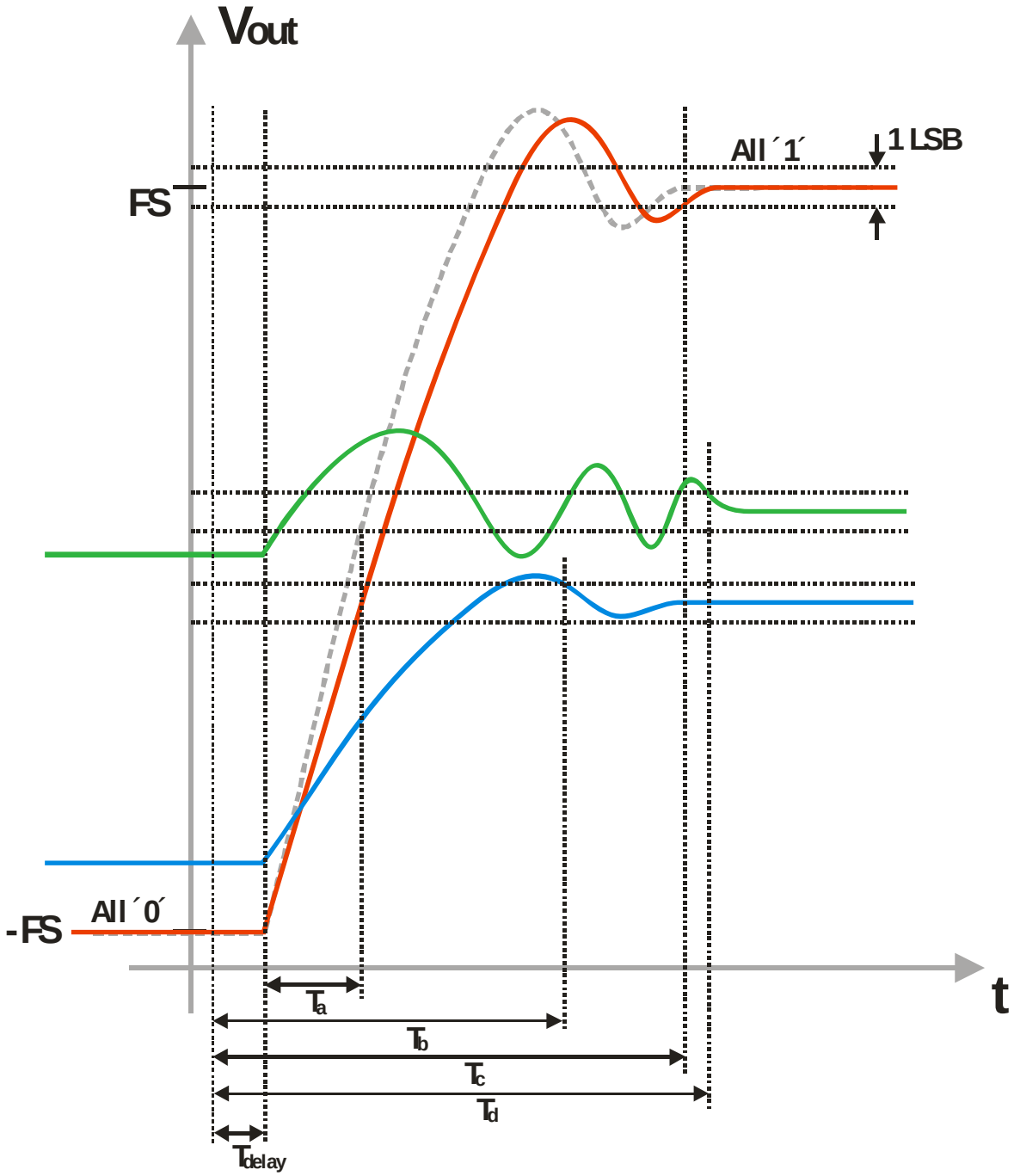


Figure 26: Settling time for a variety of transient situations. The grey dashed trace shows the linear step response for a full-scale step while the red line shows the actual response affected by slew-rate limitation (T_a). Lower amplitude swings may be not affected by slew-rate issues (blue trace) but even small amplitude changes (green trace) may be heavily distorted by switching glitches. Settling time specifications may include the delay time (T_{delay}), but typically it should not be included in the spec for high-speed AWGs as it does not affect the overall performance.

Output slew-rate limitation is another dynamic non-linear behaviour. This effect limits the maximum transition speed to a given V/s value. The observed effect is that the transition time depends on the amplitude difference between the initial and final levels. Linear transients always show the same rise or fall times regardless of the levels involved. Generating signals with fast edges and high amplitudes through very high sampling rate AWGs may be a challenge as the DAC signal must be properly amplified, and amplifiers are typically the limiting factor when it comes to slew-rate performance. Even an amplifier with an excellent 10,000V/µs slew rate specification will have problems to handle a 1 Volt transition in a 10GSa/s AWG, since the time required for the transition (100ps) is equal, in this particular case, to the sampling period.

Settling time can be defined as the time required by a DAC or an AWG to reach a new output level within a given error margin, typically +-1/2 LSB. Unlike rise (or fall) time in a linear system, settling time depends on the initial and final levels (see Figure 26). Linear distortions, such as limited bandwidth and ringing, will increase settling time as the difference between the initial and final states grows. The previously described non-linear distortions, slew-rate limitation and switching glitches, especially during major carry transitions, will also influence settling time. Settling time for a full-scale transition will be a good indicator of the worst-case effects of limited bandwidth and slew-rate limitation. Many high-speed AWGs specify rise/fall times close to or larger than the sampling period so that the final target level cannot be reached during one sample time. Relatively slow rise/fall times may be caused by limited bandwidth, which is a linear distortion. However, the effects of switching glitches and slew rate limitation can be masked by the linear step response and the resulting lack of bandwidth of the device. Under these circumstances, it may be very difficult to observe and characterize the effects of non-linearity in the time domain (i.e. through a digital storage oscilloscope, DSO).

2.7 Non-Linearities in the Frequency Domain

Many of the AWG applications areas, such as wireless signal generation, require good signal quality in the frequency domain. Non-linear behaviour is sometimes easier to characterize and categorize in the frequency domain due to the fact that it results in unwanted frequency components that can be easily identified through frequency selective instrumentation such as spectrum analyzers or FFT-capable DSOs. Figure 27 shows the hypothetical spectrum of an AWG output signal while generating a sinewave, characterized by its frequency and amplitude. The spectrum of an ideal DAC will show just one tone within the Nyquist band and the corresponding images around multiples of the sampling frequency. Several components in the spectrum can be identified:

- The fundamental sinewave: Its peak-to-peak amplitude must be equal to or lower than the full-scale range of the DAC. Although typically maximizing amplitude improves the signal-to-noise ratio, sometimes it may be better to set a lower amplitude as some non-linear effects grow with amplitude faster than the amplitude itself.
- Its harmonics: Spurious signals located at multiples of the sinewave frequency are typically caused by harmonic distortion. Odd harmonics are caused by symmetrical distortions, while even harmonics come from asymmetrical distortions. However, harmonics located beyond the Nyquist frequency will fold down to the first Nyquist band, as distortion will be sampled as well.

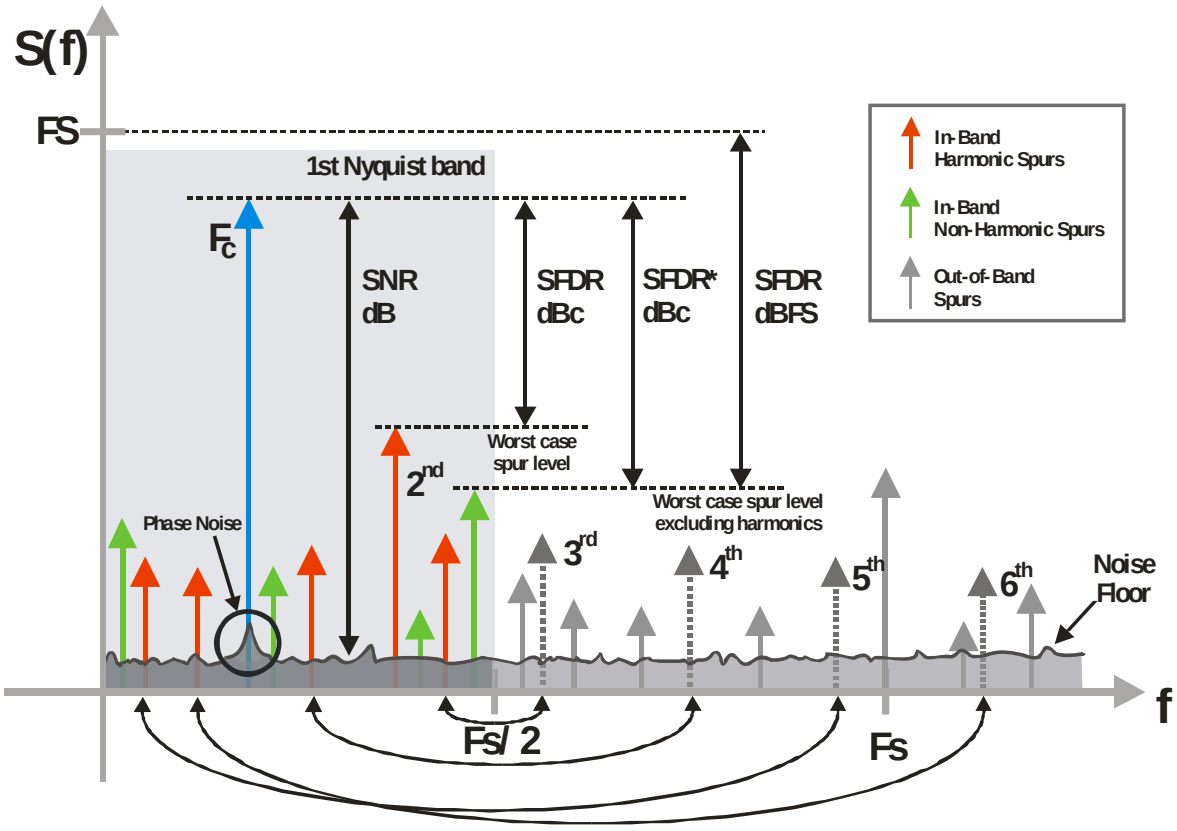


Figure 27: Non-linear behavior can be easily observed in the frequency domain by generating a pure sinewave. Nonlinearities result in unwanted components such as harmonics. Harmonics beyond the Nyquist frequency are folded down to the first Nyquist band. Other non-harmonic spurs may eventually show up in the spectrum. SFDR (Spurious-Free Dynamic Range) can be established through a spectrum analyser as well as noise floor and SNR (Signal-to-Noise Ratio).

Potential locations (Fd) for those harmonics (Fh) can be easily calculated by applying the following algorithm:

Fd=Fh;

While Fd>Fs/2

Fd=|Fd-Fs|;

- Non-harmonic spurs: These signals may come from a variety of sources such as sampling and other digital clocks (and its harmonics and sub-harmonics), switching power supplies, etc. A very important potential component may come from repetitive quantization noise. If an integer number of signal cycles include an integer number of samples, then the quantization noise will be repetitive and its power will concentrate in multiples of the repetition frequency. In particular, if one cycle of the signal is made of exactly an integer number of samples, quantization noise will show up at the harmonic frequencies so it will easily be confused with harmonic distortion. Continuous playback in AWGs requires recycling the same waveform memory section through the DAC, therefore quantization noise will be always repetitive, although its repetition rate will depend basically on the record length and sampling rate, provided that the signal does not repeat exactly in the same way within the waveform memory.
- Wideband noise: Noise floor is composed of several wideband sources such as quantization noise, noise induced by digital circuitry, thermal noise, switching noise, etc. Although quantization noise may be periodic in an AWG, it can be seen as a wideband signal (instead of a series of narrowband tones) depending on the repetition rate and the resolution bandwidth used to analyze it.
- Phase Noise: Jitter present in the sampling clock will show up as phase noise in the fundamental carrier. There are other potential sources for it such as sampling uncertainty, jitter associated to the data signals, skew between the different current sources in the DAC, etc.

The spectrum of a simple sinewave may be quite complex given the presence of the previously discussed unwanted components. The SFDR parameter is a good way to assess the performance of an AWG with one single figure of merit that summarizes many factors affecting signal quality. This parameter typically compares the amplitude of the sinewave with the strongest component of all, or a subset, of the unwanted spurious signals. SFDR specifications may or may not include harmonics. If not, an independent total harmonic distortion (THD) specification should be provided. SFDR is typically specified in dBc (dB respect to the fundamental carrier level), and its value may be quite sensitive to some signal parameters such as absolute amplitude, percentage of the DAC range being used by the signal, carrier frequency, sampling rate, and even sampling clock source.

Often several SFDR values will be provided for different carrier frequencies and, sometimes, several sampling rates. SFDR can also be specified relative to the full-scale (FS) level. In this case, the unit for the spec will be dBFS for obvious reasons. It is important to keep in mind that specifying SFDR in this way does not imply that its value has been established using a sinewave covering the full range of the DAC, so this may be a method to list a better specification in the instrument data sheet. Another important consideration for the SFDR parameter is the band covered by the specification. For AWGs, the band is usually the first Nyquist band (DC-Fs/2) or the analog bandwidth, whichever is lower. Depending on the application, this may not be the band of interest either. If oversampling is being used, any spur beyond the base signal bandwidth will not be relevant as it can be eliminated through filtering. For narrowband signals, such as digitally modulated IF/RF signals, spurs close to the signal (i.e. affecting adjacent channels) may be more important.

Harmonic distortion is the most visible consequence of non-linearity. This is why a specific analysis is often necessary. As previously stated, THD is the most popular specification for harmonic distortion. THD as a percentage can be calculated using the following formula:

$$THD_{\%}=(\Sigma H_n^2)^{1/2}/S \times 100\%,\ n=2...N,\ H_n \text{ and } S \text{ are rms values}$$

THD can be also expressed in dBc units

$$THD_{dBc}=20\log_{10}(THD_{\%}/100)$$

The spurs generated by harmonics, including those folded-down, in the band of interest must be identified. In case of doubt, changing slightly the fundamental frequency will shift the harmonics up and down while the non-harmonic spurs will stay unchanged. THD specs must include information about the test conditions and the number of harmonics (the N in the above formula) included in the analysis.

Signal-to-noise ratio (SNR) specifies the relative power in dB between the signal and the in-band wideband noise or noise floor. Noise power should be properly calculated by integrating its power in the band of interest (typically up to FS/2) while excluding narrowband spurs, both harmonic and non-harmonic. This may be a difficult measurement, as it requires accurate and sensitive instrumentation and a careful evaluation of noise.

SFDR, THD, and SNR are very useful to establish the suitability of an AWG to generate a signal with a given quality. These parameters can also be used to compare different generators. However, different instruments may be better than others in different specifications, so an instrument with a better SFDR specification may have a worse THD, etc. An additional spec combining appropriately all of the previous factors can be useful as a direct way to define the overall signal quality of a generator. SINAD (Signal-to-Noise-and-Distortion) combines all of the unwanted spectral components and compares them with the sinewave power. As the different components are uncorrelated, SINAD can be computed adding the power of all the in-band spurs and integrating the noise floor power over the same band. If the harmonic spurs dominate over the non-harmonic ones, what it is often true in actual equipment, then SINAD can be obtained using the following formula:

$$SINAD_{dBc} = 10\log_{10}(10^{-SNR/10}+10^{-THD/10}),\ THD=100xTHD_{\%}$$

Even ideal DACs will add quantization noise to any sinewave. As previously stated, SNR for a full-scale sinewave in an ideal DAC is:

$$SNR_{dBc} = 6.02N+1.76dB, \qquad DC<f<FS/2$$

The effective number of bits (ENOB) parameter can be obtained by substituting SNR with the SINAD parameter. In this way, the above formula can be used to compute the hypothetical number of bits of an ideal DAC (or AWG) resulting in a quantization noise with the same power as that of the combined unwanted components of the actual device:

$$ENOB = (SINAD-1.76)/6.02$$

Generally speaking, ENOB results in non-integer values. Again, ENOB will be a function of frequency and sampling rate among other factors. The difference between the ENOB parameter and the AWG vertical resolution is a good indicator of how performance is degraded by noise and distortion.

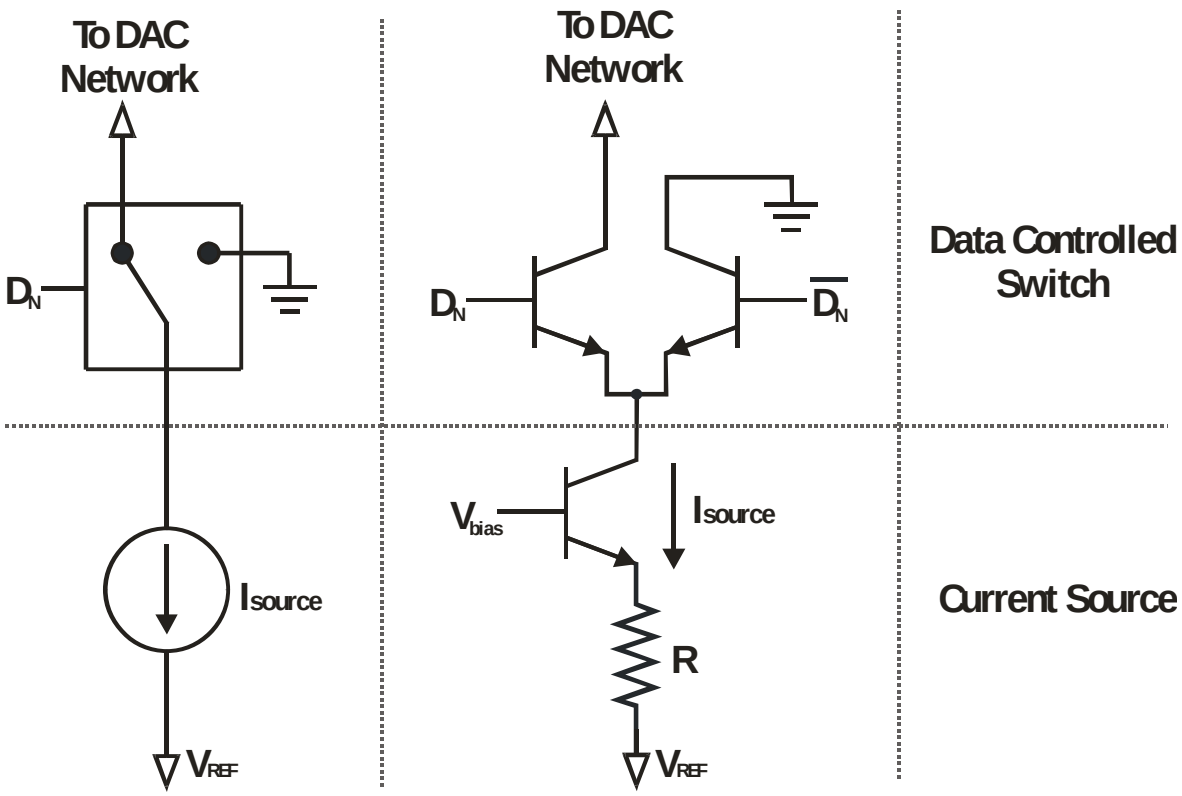
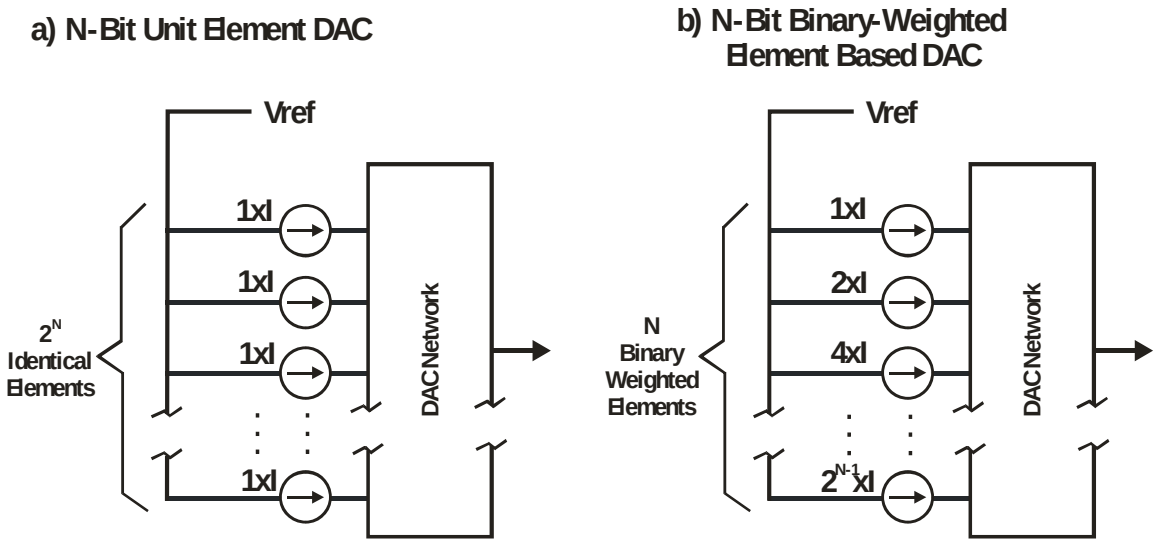


Figure 28: Most high-speed DACs use constant current sources as the basic building block. Current flow to the DAC network is regulated by a switch controlled directly or indirectly by bits in the sample word. The functional block diagram on the left shows such an arrangement. Using a SPDT (Single-Pole, Double-Throw) switch makes sure that the current flow through the source is never interrupted greatly improving dynamic behavior. The simplified schematic in the left shows a solid-state implementation.

2.8 High-Speed DAC Architecture

There are many possible implementations for a digital-to-analog converter. The chosen architectures strongly depend on factors such as the target vertical resolution and sampling rate. Both requirements are typically contradictory so obtaining a high-sampling rate and high vertical resolution in the same device is a design challenge by itself. Dynamic performance and vertical resolution are especially important for high-speed DACs used in AWGs. Good signal fidelity requires both bandwidth and accuracy, including low noise and distortion. DAC complexity and quality requirements heavily depend on the target vertical resolution therefore some architectures that provide excellent performance result in an unacceptable complexity. This section will cover some of the basic DAC architectures and those most appropriate to be applied to high-speed AWGs.

Most high-speed DACs use multiple current sources switched under the control of the digital value of the sample to be converted. In some architectures, the individual bits of the sample word can be used to control specific switches, while others require previous decoding. The accuracy of the current sources and the associated switches greatly influence the performance of the DAC itself. The simplified schematic in Figure 28 shows a possible high-speed current switching device. The switch itself is a SPDT (single pole, double throw) device so it does not interrupt current flow through the source as current is supplied from ground (real or virtual) when it has been interrupted from the DAC network. Switching the current source itself (i.e. digitally controlling the V_{bias}) would result in slower settling times and, consequently, a lower sampling rate. Additionally, a large number of current sources and/or current values will make reaching a given accuracy more difficult and most times reaching it requires laser trimming of the current source resistances.



	Unit Element DAC (N=12)	Binary Weighted Element DAC (N=12)
# of elements	2^N (4096)	N (12)
DNL (σ_ε)	$\sigma_{DNL}=\sigma_\varepsilon$ (σ_ε)	$\sigma_{DNL}=2^{N/2}\sigma_\varepsilon$ (64 σ_ε)
INL (σ_ε)	$\sigma_{INL}=2^{N/2-1}\sigma_\varepsilon$ (32 σ_ε)	$\sigma_{INL}=2^{N/2-1}\sigma_\varepsilon$ (32 σ_ε)
Switching Glitches	Moderate	High: Very non-linear vs. output level
I_N / I_1 Ratio	1 (1)	2^{N-1} (2048)

Figure 29: DAC architectures can be classified into the unit element and the binary-weighted basic structures. Each structure has some advantages and disadvantages (in red in the table). Unit element based solutions are far superior when it comes to linearity, including guaranteed monotonicity, but at the expense of a much higher complexity. The table shows linearity estimates as a function of the element's dispersion (σ_ε) and example values for a 12 bit DAC.

DAC architectures can be classified in two large groups: unit element and binary-weighted element based. In unit element based architectures, every minimum increment in the output level is accomplished by adding an additional identical voltage or current source, so an N bit vertical resolution DAC requires about 2^N identical elements and sample words must be decoded prior to conversion.

As a consequence, this kind of DAC is monotonic by nature regardless of the accuracy of each individual current element: every additional element will always increment the output. Binary-weighted element based DACs use a series of sources with 2^n, n=0...N-1, relative values. The number of elements required for an N bit DAC is just N. Bits in the sample word can control the switches directly so the complexity is even lower. However, monotonicity is not guaranteed since it depends on the accuracy of the sources. Additionally, the large difference in the source's magnitude levels makes accuracy more difficult to be achieved.

Another important differentiator of each architecture is the sensitivity to switching glitches. Unit element based DACs involve only a limited number of switches (those between the initial and the final states) for every transition, and transitions are always monotonic so glitch amplitude is really limited, and it basically consists in a slower transition. On the other hand, binary-weighted DACs may involve a switch during any transition (i.e. major-carry transitions) regardless of the difference between the initial and final levels so the glitch amplitude is not directly related to them and it is, statistically speaking, mainly a function of the DAC's full range.

Beyond monotonicity, it is possible to statistically analyze the influence on each basic architecture of the accuracy of individual elements in the overall DAC accuracy, in terms of dynamic and integral non-linearity (DNL and INL respectively). Unit element DAC's DNL does not depend on the vertical resolution since every quantization level depends only on the absolute accuracy of the element associated to that level. Statistically speaking, the largest INL will appear in the DAC midrange, and it depends both on the absolute accuracy of each element and the DAC resolution, as every extra bit of resolution will require doubling the number of elements. The same analysis for binary-weighted element DACs results in a similar value for INL as a function of the element accuracy, but DNL will be much worse since from a statistical standpoint, dynamic non-linearity is twice the INL, and it grows with the vertical resolution of the DAC. Worst-case INL can be located anywhere in the DACs range in this case. For both types, the higher the vertical resolution, the higher the accuracy requirements for of each element. However, for a given element, a much better measure of accuracy in terms of DNL will be obtained with unit element based DACs at the expense of a higher complexity. In both cases, implementing high-resolution DACs is a challenge since unit element DACs may require an unacceptable number of elements while binary-weighted DACs may require element accuracy difficult to reach.

What follows is a general description of DAC architectures and their suitability for high-speed, highly accurate AWGs:

String DAC (Kelvin Divider):

The basic architecture of a string DAC can be seen in Figure 30a. Its origin can be traced back to Lord Kelvin in the 19th century. It is a unit element based DAC, and it consists basically of a voltage divider made by 2^N identical resistors connected to the same number of switches. Only one switch is on for a given sample word so it must be properly decoded. Although it is quite simple conceptually, it is not very popular in high-speed DACs. Among other issues, output impedance changes with output voltage, and the switches must handle very different voltage levels since switch linearity influences the overall accuracy. Switching glitches will show up during transients because either none, or two switches, may be closed for a short instant. INL may be improved by applying element-matching techniques where intermediate levels in the divider are forced by applying accurate voltages.

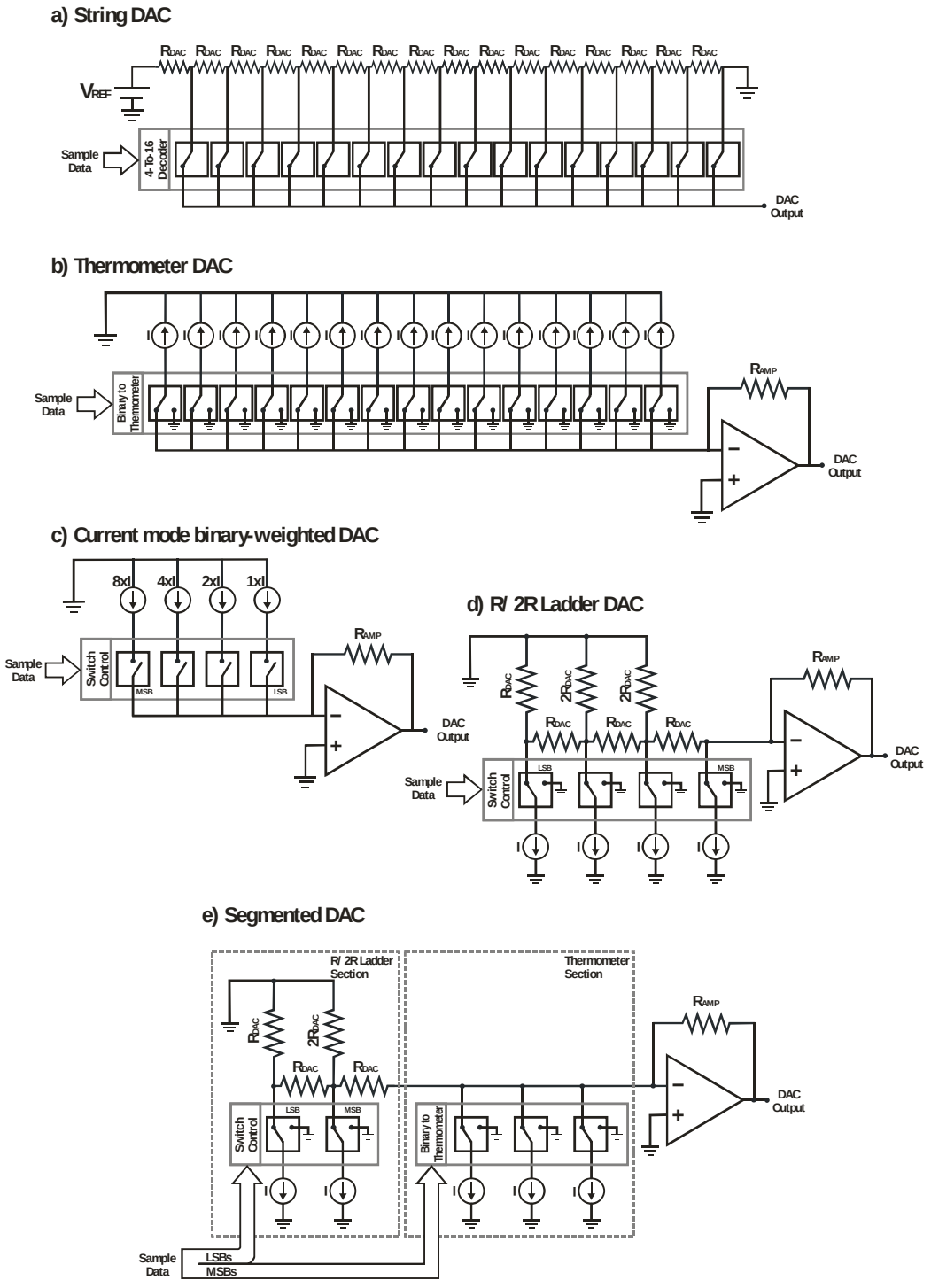


Figure 30: Different 4 bit DACs using the most popular DAC architectures. Unit-element based solutions are far more complex than binary-weighted ones, even for a limited vertical resolution.
A trade-off between complexity, accuracy, and linearity leads to segmented DACs where some MSB are associated to a thermometer DAC section while the remaining LSBs are assigned to a
binary-weighted DAC section.

Thermometer DAC: Also known as a fully decoded DAC. The block diagram in Figure 30b shows a possible implementation using identical current sources. It is also a unit element based DAC, but in this case, it only requires 2^N-1 elements and switches to output 2^N states. The N bits in the sample word must be fully decoded to 2^N-1 bits, each one attached to one of the elements. For a given input word, the number of 1s at the decoder's output will be equal to the numerical value of the word so that the number of elements switched will be proportional to the input word. The decoder output looks like a linear scale in a traditional thermometer.

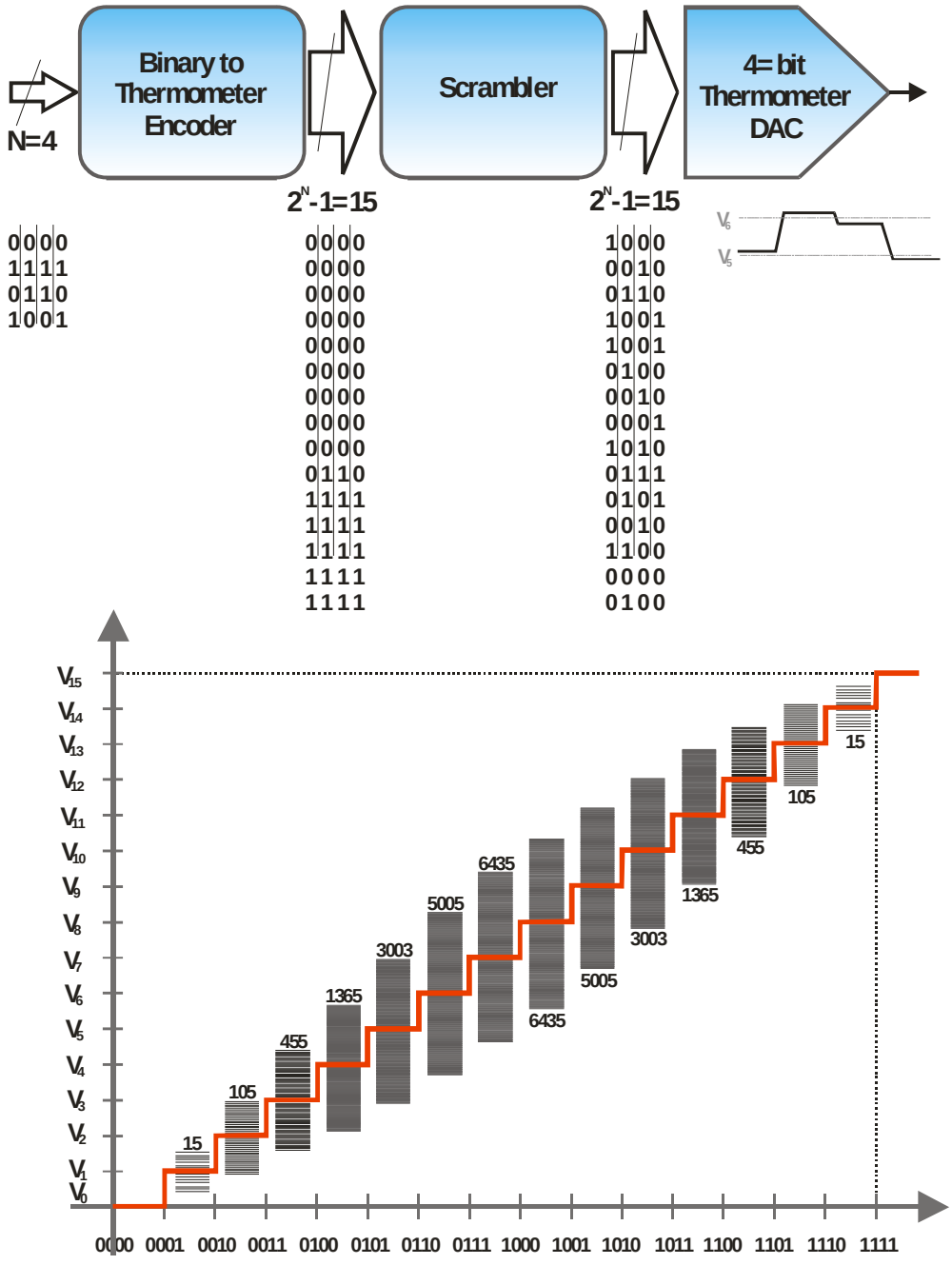


Figure 31: Scrambling the output of the binary-to-thermometer decoder so the current sources involved for each sample change, converts linearity errors into random noise (top). For each nominal output level there may be a huge number of possible outcomes, but the average of all of them over time will be exactly adjusted at a perfectly linear transfer function (bottom).

Linearity (DNL and INL) in thermometer DACs may be greatly improved through a technique known as dynamic element matching. Dynamic element matching consists of scrambling the output of the binary-to-thermometer decoding so for each occurrence of a given sample word the elements activated change, although the overall number of 1s (and 0s) remain unchanged (Figure 31). Depending on the combination of elements and their accuracy, the output level will change statistically between two extreme levels (the m highest current elements and the m lowest current elements). A simple combinational analysis shows that depending on the specific output level there will be a different number of element combinations resulting in the required number of 1s. The average level for each sample word will be that of an ideal DAC, so extremely good DNL and INL performance is obtained regardless of the basic element accuracy (Figure 31, bottom). The price paid for this improvement in linearity is degradation in SNR performance since this technique effectively transforms linearity errors into random noise. Additionally, switching glitches will develop depending on the order elements are activated or deactivated during each sample-to-sample transition, even when the output value remains unchanged.

Binary Weighted DAC:

This is the simplest implementation for a DAC since it requires the minimum number of elements and switches. Figure 30c shows a simplified block diagram of a 4-bit binary weighted DAC. Current for each source is different since it must implement the corresponding power of two weights. There is no need to decode the sample word, since there is a one-to-one relationship between the input bits and the switches. As in any binary weighted architecture, element accuracy and switch linearity must be very high to obtain good results.

R/2R Ladder DAC:

This is a particular implementation for a binary weighted DAC where the power of two relationships between the current sources associated to each bit is accomplished through a combination of an iterative resistor network (a ladder) composed by only two resistance values, R and 2R (Figure 30d). Current mode R/2R converters, as shown in the figure, use equal current sources in all of the nodes, improving the simplicity and accuracy of the implementation. It is easier to reach a good accuracy sinceit is easier to implement and adjust resistors with just two values. It may even be simpler since a 2R network can be implemented using two R resistors in series (or R resistors can be implemented with two 2R resistors in parallel) so there is only one single resistor value in the device.

Segmented (or Hybrid) DAC:

Unit element DACs and binary-weighted DACs have their own strengths and weaknesses. Hybrid (or segmented) architectures mix both solutions in a single device to try to leverage the strengths of both architectures while limiting the effects of weaknesses. Typically, a unit element section is associated to the M MSBs while N-M LSBs are attached to a binary-weighted section. The unit element section associated to the MSBs makes sure that the overall linearity is good, especially if dynamic element matching is used in the design, while switching noise is limited given the limited range of the binary weighted section, which keeps complexity under control as it limits the overall number of elements. The selection of M will depend on a variety of factors such as the acceptable number of elements, the accuracy of resistors and current sources, and the cost and complexity of laser trimming them during production. Figure 30e shows a 4-bit segmented DAC where the two MSBs are associated to the thermometer section, and the two LSBs are connected to a R/2R ladder DAC in current mode. This segmented DAC requires just one more element than a 4-bit binary weighted DAC, but only one third of those required for a thermometer DAC with the same resolution. The combination of a thermometer and an R/2R section is especially interesting since it is possible to use the same current for all of the elements. Segmented DACs offer the best combination of resolution, accuracy, and speed, and as a consequence, they are extensively used in high-speed AWGs.

As an example, the Keysight M8190A, a 14 bit, 8 GSa/s (or 12 bit, 12 GSa/s) AWG, uses a segmented DAC with a 4-bit thermometer section and a 10-bit R/2R ladder section. The Keysight N8241A, a 15 bit, 1.25 GSa/s AWG, uses a segmented DAC with a 6-bit thermometer section and a 9-bit R/2R ladder section. Both of them integrate Keysight proprietary DACs implemented in SiGe BiCMOS processes.

The 65 GS/s, 8 bit DAC in the M8195A AWG is also implemented in a segmented-DAC architecture with a 3 bit (7 elements) thermometer section and a 5 bit binary-weighted (R/2R) section. However, this DAC is implemented in a CMOS technology given the density and power requirements resulting from the need to integrate four channels and powerful DSP blocks associated to each DAC in the same chip.

2.9 Complementary Output DACs

Most high-speed signals are transmitted through differential signaling over balanced symmetrical lines given the multiple advantages of this strategy, especially regarding noise and interference resilience (including crosstalk), and its suitability to support low voltage signals. As a consequence, AWGs are often used to generate differential signals such as LVDS (Low-Voltage Differential Signalling). There are many possible ways to meet this requirement, but integrating DACs with differential outputs should be a good starting point. Fortunately, some DAC architectures allow for an easy implementation of differential signals by adding some extra circuitry to simultaneously output an additional complementary output:

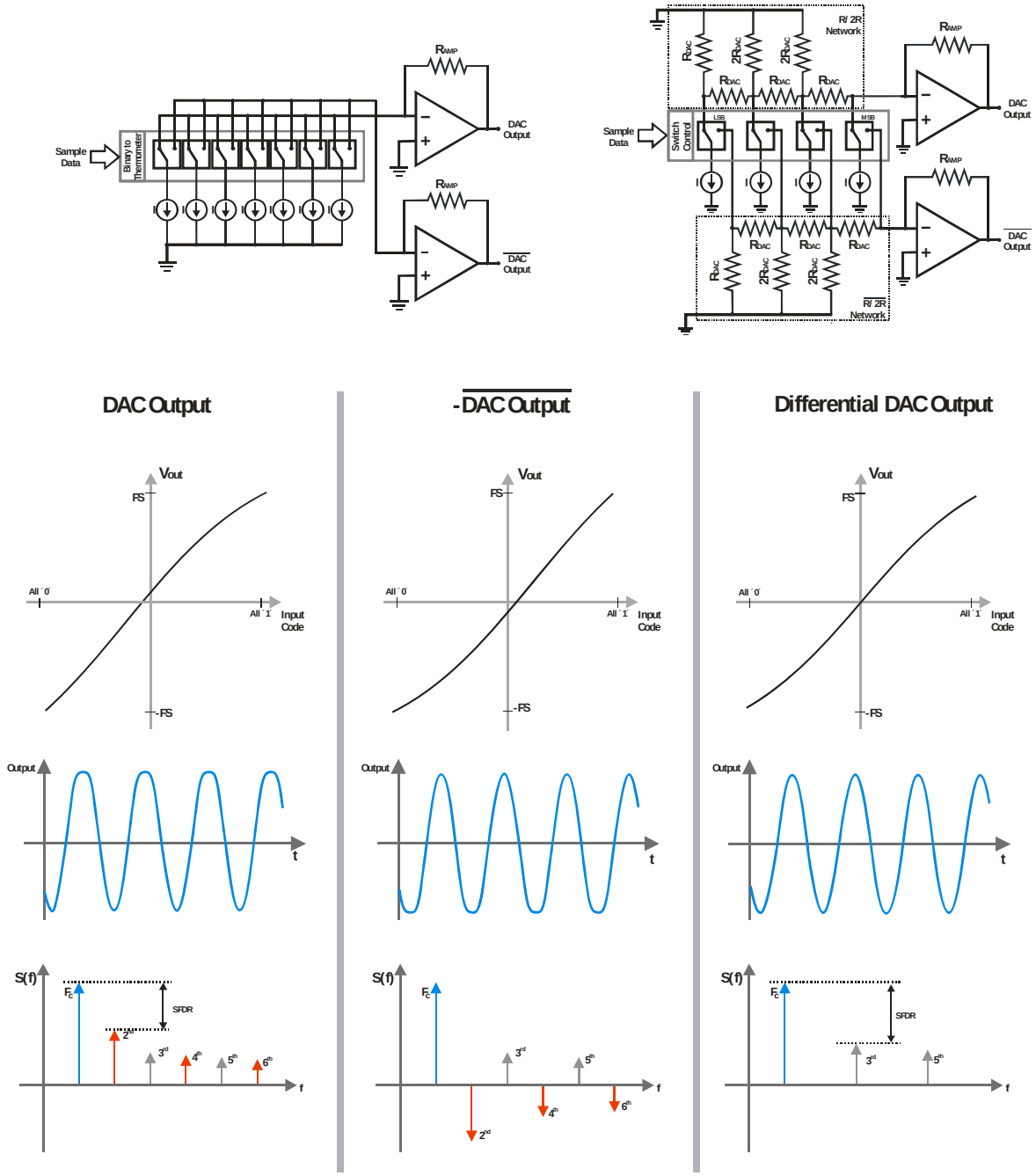


Figure 32: Complementary output DACs implement a differential output using a single set of current elements (current source + switch). Beyond the convenience of differential outputs, this strategy improves linearity as it cancels the asymmetrical component of the transfer function. Above, complementary output arrangements for a thermometer DAC (left) and an R/2R ladder DAC (right).

Direct Signal: $DAC^{+}(n) = s(nTs)$

Complementary Signal: $DAC^{-}(n) = FS-s(nTs)$

Figure 32 shows the schematics for a thermometer DAC and an R/2R ladder DAC with complementary outputs. Both solutions share the same current source elements among the direct and complementary outputs through SPDT switches, so every current source may alternatively be connected to either the direct or the complementary network. This is functionally equivalent to an additional DAC where the sample words are inverted. The beauty of this solution is that, beyond the circuit economy, sharing so many circuit blocks results in better differential signal quality since there will not be any timing skew due to sampling clock distribution issues or DAC dispersion. Additionally, INL will be greatly improved for the differential signal. For thermometer DACs, the transfer function of the complementary output will be exactly the reverse of the direct output. Therefore, when they are combined as a differential output any asymmetricity of the transfer curve will be fully removed. A hypothetical differential sinewave generated with such a device will show only the odd harmonics as even harmonics are created by the asymmetrical component of the transfer curve (Figure 32). For R/2R ladder DACs, where two different resistor networks are required, non-linearities caused by the current source will also be removed. However, some remaining effects of the inaccuracy of the resistors will still show-up. In segmented DACs where the MSBs are associated to the thermometer section, the differential arrangement results in a much better THD (and potentially SFDR) performance. If necessary, the differential signal can be converted into a single-ended AC signal through an appropriate balun so that dynamic range performance gains can be maintained.

2.10 Deglitching DACs and Distributed Resampling

Switching glitches are a very important source of noise and impairments in AWGs, especially for those with very high sampling rates and output bandwidth. In order to achieve an acceptable performance, it is necessary to eliminate glitches (deglitch) from the output of the DACs. Figure 33 shows a potential deglitching architecture.

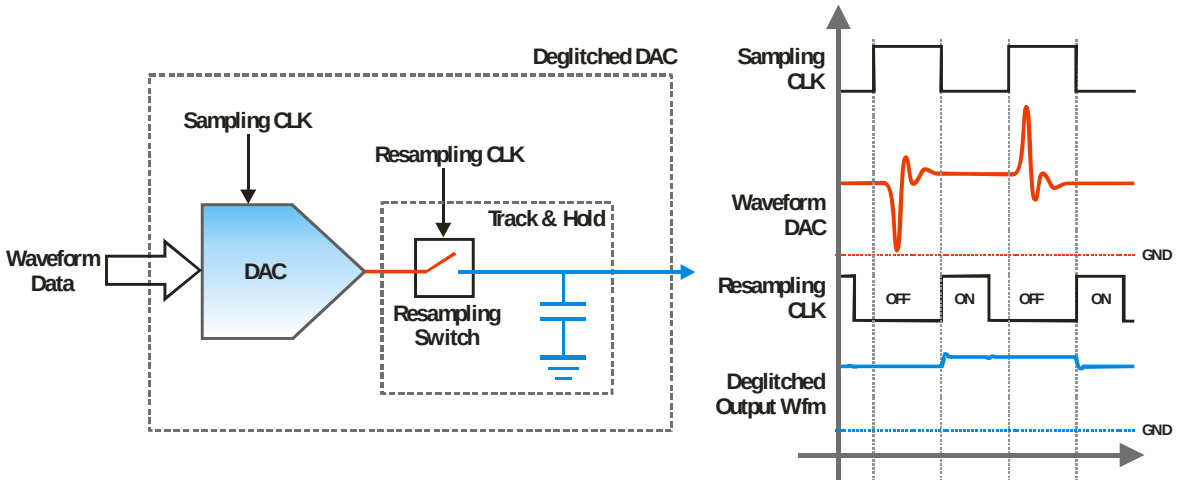


Figure 33: Resampling is the preferred strategy to deglitch DACs by isolating the output from the DAC while noisy transitions are taking place. Linearity of the resampling switch is critical since it directly influences the quality of the output. Using a track-and-hold stage maintains the output voltage while an isolated switch would result in a Return-to-Zero signal.

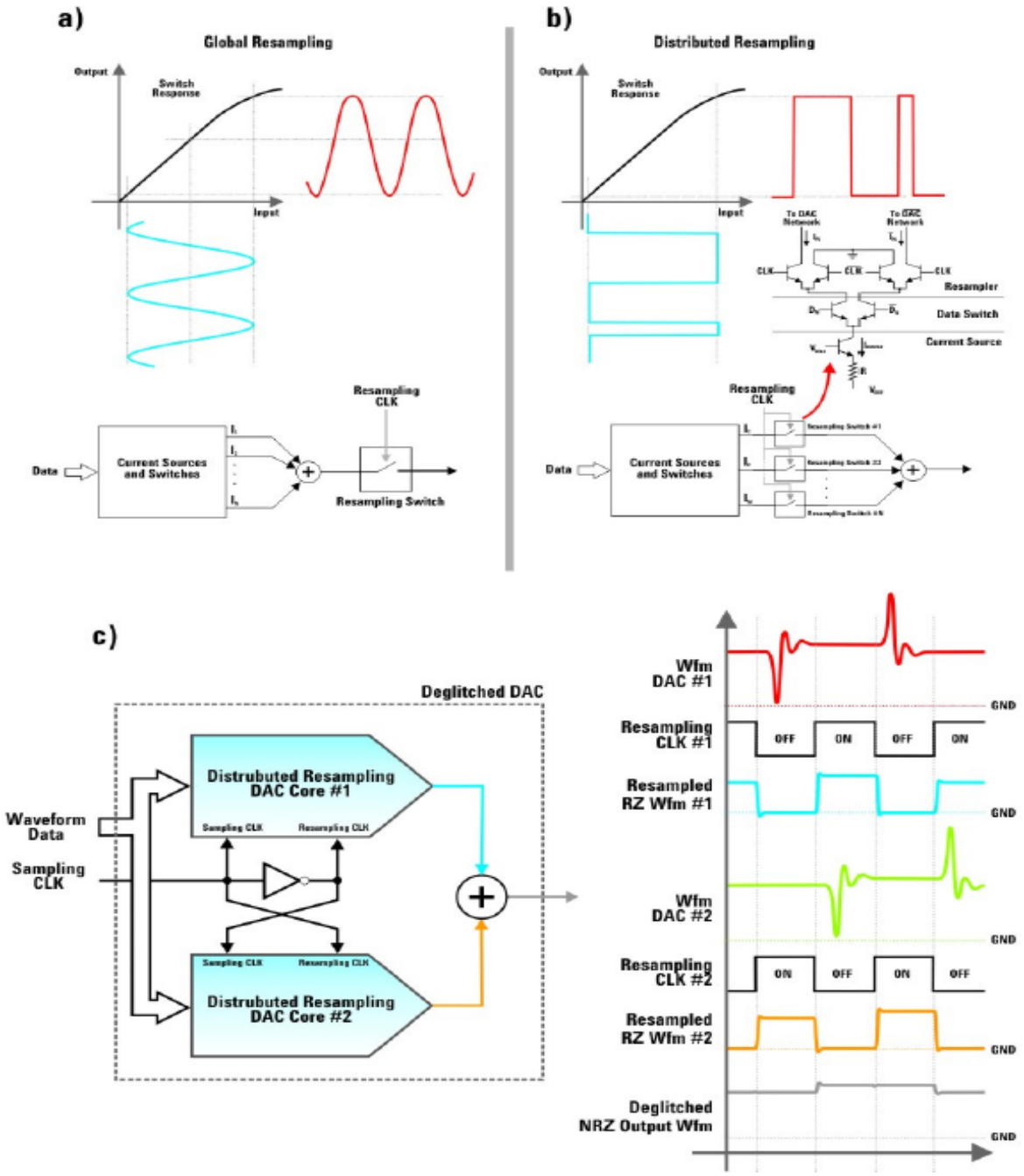


Figure 34: Keysight's exclusive distributed resampling architecture removes the influence of the resampling switch linearity (a) by replacing it with multiple switches associated to each current element (b). The result is that every switch handles only two input levels, so linearity is no longer an issue. As this approach results in a RZ signal, a practical implementation consists of two DAC cores working in parallel with a 50% duty cycle and delayed by half a sample period so that the combined output results in a continuous, smooth signal. The depicted schematic shows a complete current element supporting distributed resampling and complementary outputs. Note that current through the source is never interrupted.

Basically, the deglitching system consists of a track and hold (T&H) stage that disconnects the DAC from the output while the noisy transition is taking place. The overall effect is such that the glitch amplitude is greatly reduced and the sample is delayed by the time the T&H switch is closed. This process is called resampling since the T&H effectively samples the signal again at a different sampling instant but at the same sample rate.

The problem with this approach lies within the linearity of the switch in the track and hold block. The switch must be an ultra-fast solid-state device capable of handling a wide amplitude range, and any non-linearity in its response will impact the quality of the output signal. A proprietary technique developed by Keysight, known as distributed resampling, can overcome this situation by resampling each current source instead of the overall signal (Figure 34). Through this technique, linearity is no longer an issue since resampling switches only have to handle two fixed levels. The output will consist in two levels as well since it occurs with 1-bit DACs. Distributing the resampling clock is simpler and more accurate than distributing the bits in the sample word since it is just one signal, and as with any clock, it is not affected by some impairments such as intersymbol interference. With distributed switches, the DAC output results in a RZ (Return-to-zero) signal since the output voltage is kept only while the resampling switches are in the “on” state. The RZ signal has some advantages (Figure 37) over the NRZ alternative since the sinc(f) zeroth-order hold response has its first null at twice the sampling rate so the DAC response is much flatter over the full Nyquist band. However, for a 50% duty cycle the amplitude at low frequencies is reduced by 6 dB while the unwanted signal images are boosted.

2.11 Dual-Core DAC Architectures

Switching glitches are a very important source of noise;Figure 34c shows an improved DAC architecture used in Keysight AWGs where the return-to–zero issue is solved by combining the output of two DAC cores working in parallel with a 50% duty cycle and delaying one of them by half a sample period so that the return-to-zero section is effectively removed from the output. This arrangement solves the 6dB amplitude loss problem and makes it easier to filter out images in the 2nd Nyquist band. The removal of switching glitches combined with the improved linearity of distributed resampling and segmented DAC architectures are key in the extremely high SFDR available in the M8190A AWG and other high-performance AWGs from Keysight.

The dual-core architecture implementation of the M8190A AWG also makes it possible to control both cores independently in order to improve AWG performance for some applications and/or some frequency... Figure 35 shows four different modes the cores can be controlled to get some application-specific advantages:

- NRZ (Non-Return-to-Zero): For some signals where time-domain behavior is paramount, resampling may be switched off and used in only one of the cores. In this way, switching glitches will be visible, but the FS ripple introduced by switching between cores will be removed. This will slightly improve the appearance of the signal in the time domain, especially for pulsed signals.
- DNRZ (Double Non-Return-to-Zero): This is the basic dual-core DAC working mode. It removes the switching glitches while preserving the amplitude of the output signal. The absence of switching glitches and the improved linearity obtained by averaging the levels of the two DACs in a given sample will result in an improved SFDR performance. This is especially useful for signals where the performance in the frequency domain is important as is the case with digitally modulated signals, whether they are generated in the IQ baseband or the IF/RF domains. Although some FS ripple (caused by the 2FS switching between cores with slightly different output levels) will be present, this is not important for such signals since it will be located beyond the first Nyquist band and it will be removed by applying the appropriate low or band-pass filters.
- RZ (Return-to-Zero): Using just one of the cores while keeping distributed resampling on has some advantages as well. The shorter pulse (50% of the sampling period) will result in reduced average amplitude so SNR will degrade. However, a shorter pulse changes the zeroth-order hold response dramatically since the first zero of the sinc(f) response is located at 2FS. This means that the frequency response of the DAC in the first Nyquist band will be much flatter improving the step response of the AWG. The new response will result in around 0.5dB attenuation at FS/2 compared to NRZ which results in almost 4dB of attenuation.
- Doublet mode: This mode works like the DNRZ mode except for the polarity of one of the cores. The inverted polarity can be accomplished by previously inverting the sample word applied to that core. Every sample will then consist of a “doublet”, where each half of the sample time will be an RZ pulse with the same amplitude but inverse polarity. This response increases the amplitude of some high frequency bands so it is especially useful when dealing with the generation of bandwidth-limited, high-frequency signals.

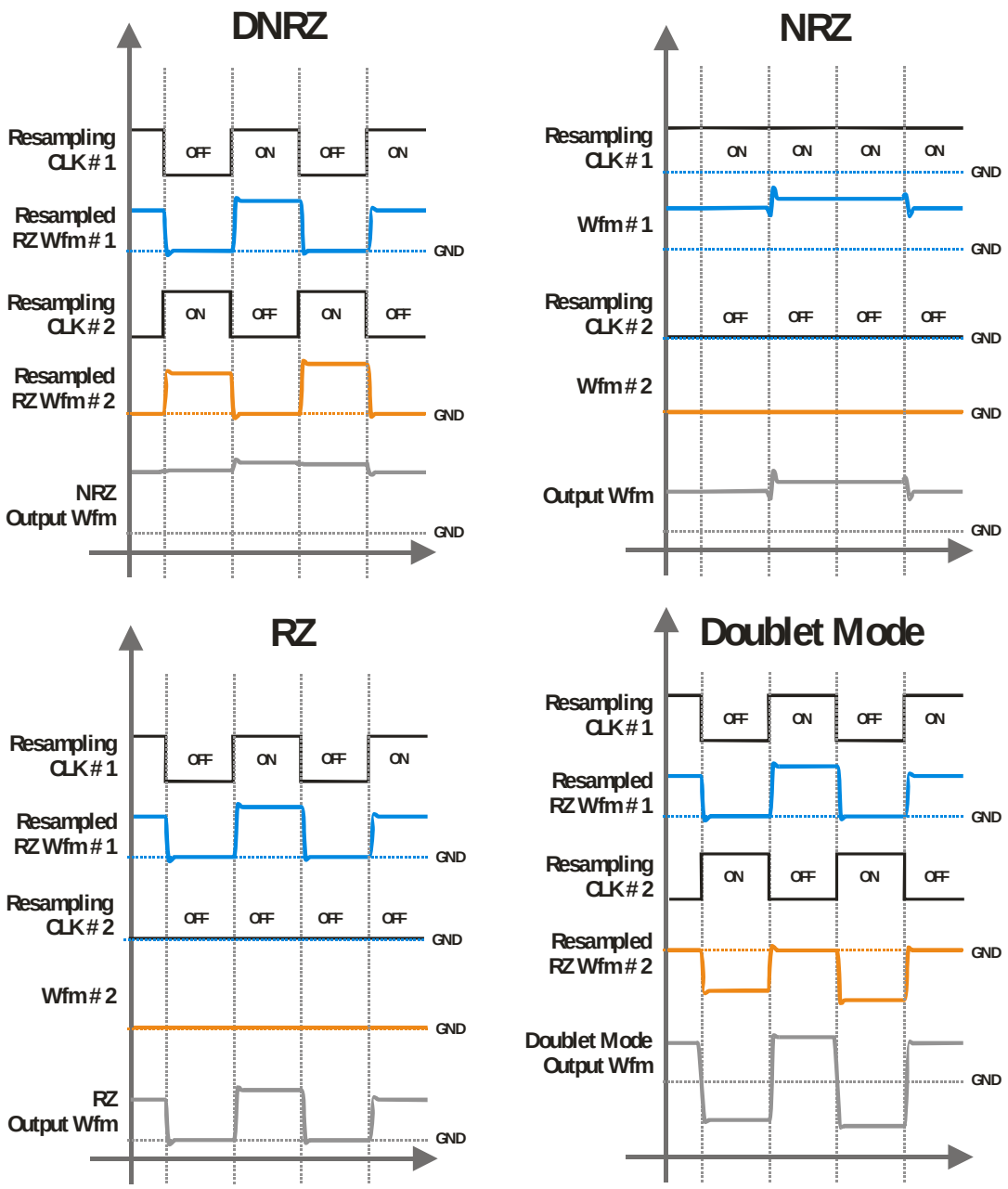


Figure 35: DAC cores in a dual-core architecture can be controlled independently so several working modes can be implemented. NRZ uses only one of the cores without using the resampling switches so better time-domain performance is obtained at the expense of glitch-generated noise and the resulting poor SFDR performance. DNRZ is the highest performance working mode as switching glitches are virtually eliminated. RZ can be used to obtain a flatter frequency response in the first Nyquist band at the expense of higher amplitude images and reduced output amplitude. Doublet mode is unique to dual-core DACs. In it, the second core signal is inverted, and as a result, the response in the second Nyquist band (FS/2-FS) is boosted.

2.12 Doublet Mode

The doublet mode is unique in the sense that it has been designed to generate signals in the second Nyquist band, the one covering the FS/2-FS range where the frequency response of the DAC will reach its maximum (Figure 36). This means that the first signal image will be used instead, and that antialiasing filtering requires a band-pass filter rather than a low-pass one. The response may be calculated by the superposition of the responses two Ts/2 RZ pulses with inverted polarity, one delayed by Ts/4 and the other by –Ts/4:

$$S_{RZ}(f) = 1/2\ \mathrm{sinc}(\pi \times f/2FS)$$

$$S_{Doublet}(f) = 1/2\ \mathrm{sinc}(\pi \times f/2FS) \times e^{i\omega Ts/4} - 1/2\ \mathrm{sinc}(\pi \times f/2FS) \times e^{-i\ \omega Ts/4}$$

Given that Ts=1/FS and $\omega = 2\pi f$

$$S_{Doublet}(f) = 1/2\ \mathrm{sinc}(\pi \times f/2FS) \times e^{i\pi f/2FS} - 1/2\ \mathrm{sinc}(\pi \times f/2FS) \times e^{-i\pi f/2Fs}$$

After developing the above expression the following one is obtained:

$$|S_{Doublet}(f)| = |\ \mathrm{sinc}(\pi \times f/2Fs) \times \sin(\pi \times f/2Fs)\ |$$

This frequency response can be seen in Figure 36. It is interesting to note that there are nulls in the response at DC and 2FS while the maximum response is obtained at 0.74FS, practically in the middle of the second Nyquist band. It is clear that signals with DC or low frequency components cannot be implemented through this working mode. For a particular bandwidth-limited signal, sampling rate should be adjusted so that the center of the band occupied by the signal is located close to the optimal location (0.74 Fs) while its total bandwidth is limited to FS/2.

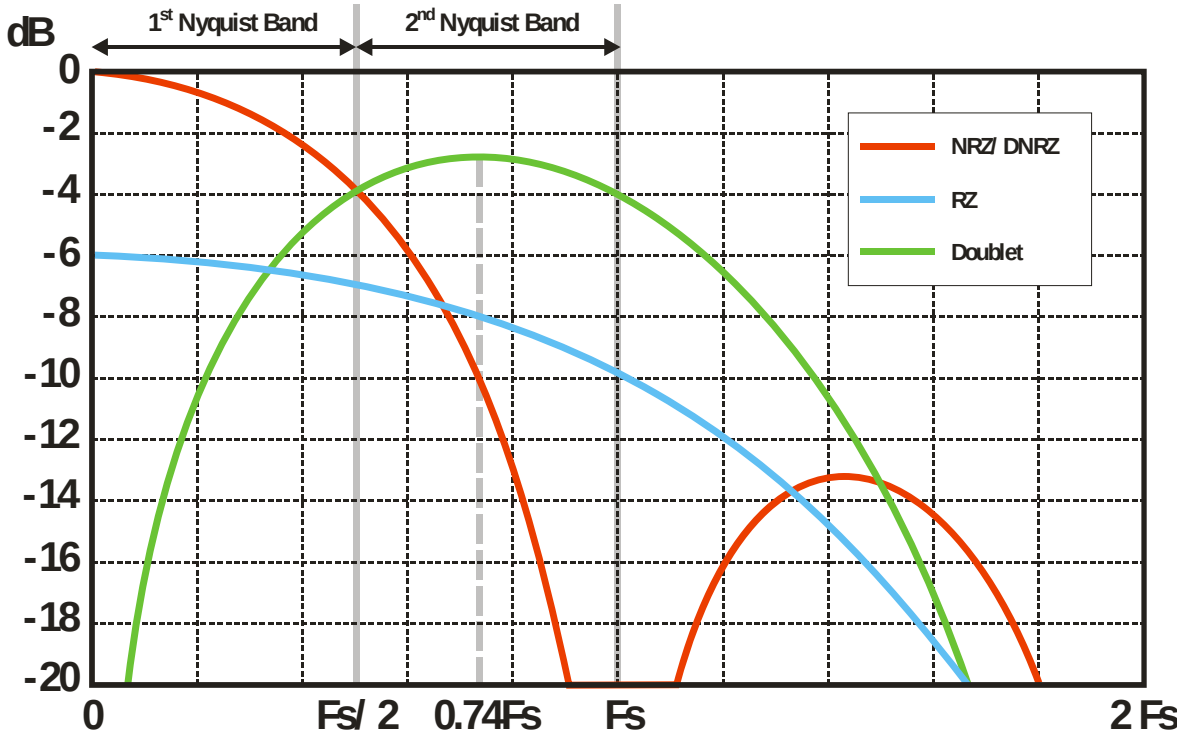


Figure 36: Frequency response for the four working modes in a dual-core architecture DAC. While the NRZ and RZ modes are mainly used to generate signals in the first Nyquist band, the Doublet mode enhances the response in the second Nyquist band making possible the generation of higher quality bandwidth-limited signals at higher frequencies.

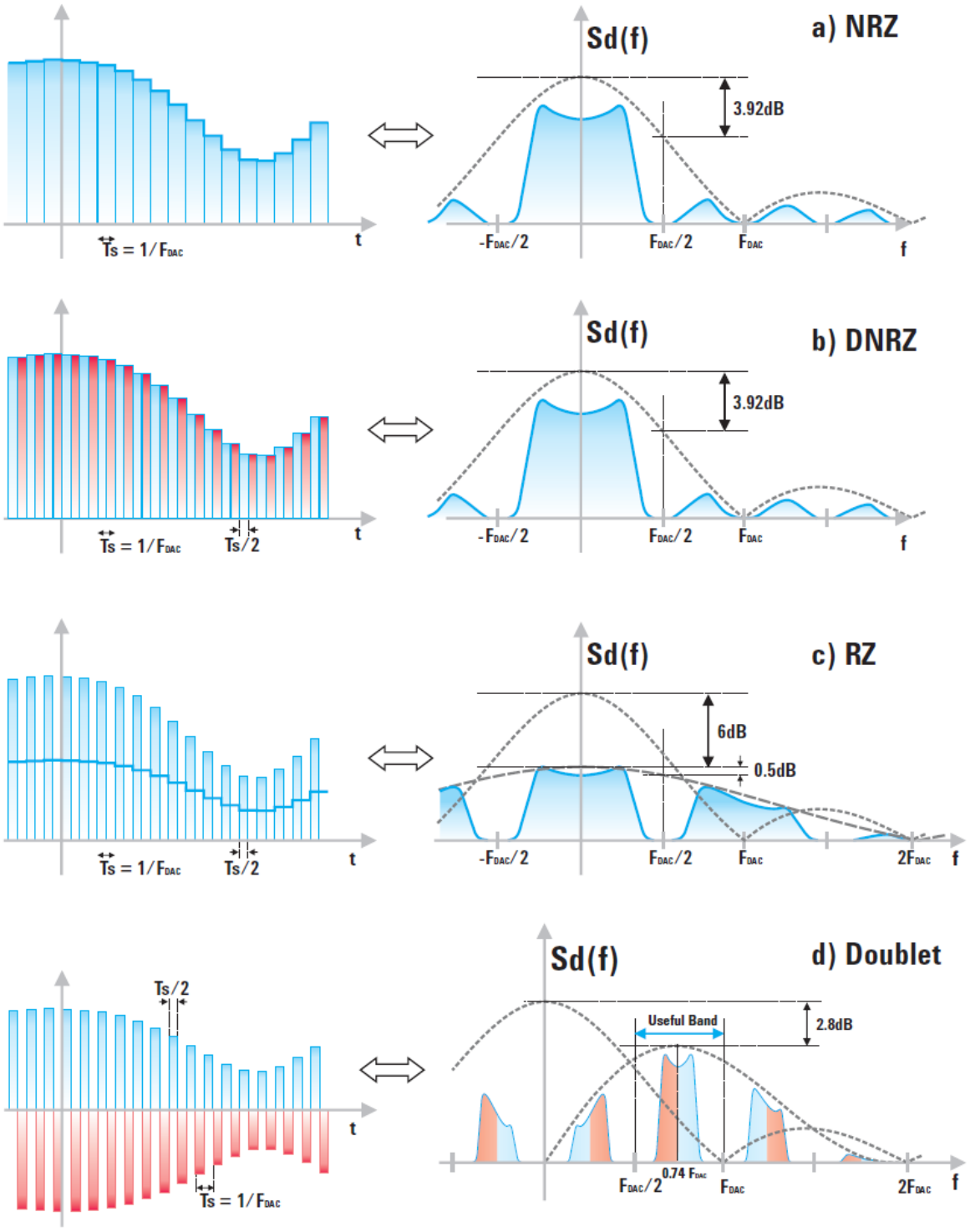


Figure 37: Signal generation using the dual-core DAC working modes. NRZ and DNRZ (a and b) result in the typical zeroth-order, sinc (f/FS) response. The RZ (c) mode will show a flatter, sinc (f/2Fs) response at the expense of a 6 dB attenuation of the signal and higher amplitude images. Finally, the doublet mode (d) boosts the second Nyquist band, with amplitude gains of more than 7dB and 5dB with respect to the NRZ and RZ modes respectively. This mode can effectively extend the frequency reach of any AWG beyond the Nyquist frequency. For signals without a perfectly symmetrical spectrum, the signal stored in the waveform memory must be calculated with its spectrum reversed (complex conjugate in the time domain) so that the right shape is obtained in the second Nyquist band.

When compared to the other modes, the doublet mode shows a much higher amplitude response within the second Nyquist band, up to 5 dBs when compared to the RZ mode, and up to 7dB if compared to the NRZ and DNRZ modes. Additionally, the response in that band is much flatter and close to symmetrical (excluding the effects of the AWG's analog response roll-off).

The doublet mode is designed to use the first signal image located in the second Nyquist band. Similar to any other even Nyquist band, the spectrum of the signal is reversed in relation to the first band spectrum. In other words, the spectrum of the output signal will be that corresponding to the complex conjugate of the signal stored in the waveform memory (Figure 37d). For signals with a symmetric spectrum (i.e. AM or ASK modulated carriers), spectrum inversion has no consequences, but for non-symmetric spectrums (i.e. QAM or a radar chirp) the signal stored in the waveform memory must be calculated with the spectrum reversed (i.e. by inverting the Q component in a QAM signal or by reversing the direction of the frequency sweep in the time domain for a chirped pulse) so that the output spectrum shows the expected shape.

Since the doublet mode uses the second Nyquist band, it is clear that it can also be used to generate bandwidth-limited signals beyond the Nyquist frequency (Fs/2). In this situation, the antialiasing filter becomes a pass-band filter rather than a low-pass one. Nevertheless, the doublet mode may be useful even when the AWG sampling rate allows for the generation of the bandwidth-limited signal in the first Nyquist band. In this case, the sampling rate must be at least twice the maximum frequency component of the signal (Figure 38). If the upper boundary of the signal is close to the Nyquist frequency, it will be attenuated and affected by the generator roll-off. In RZ mode (Figure 38a), roll-off will be smoother than in the NRZ mode, although amplitude will be lower. If the sampling rate is reduced by one half and the mode is switched to "Doublet" (Figure 38b), the amplitude will be higher than that of the NRZ mode and the flatness and symmetry will be better. Better flatness and symmetry will result in an even better dynamic range. Additionally, halving the sampling rate will also halve waveform memory to produce the same time window, saving precious AWG resources and reducing the time required to calculate and transfer the waveform to the generator.

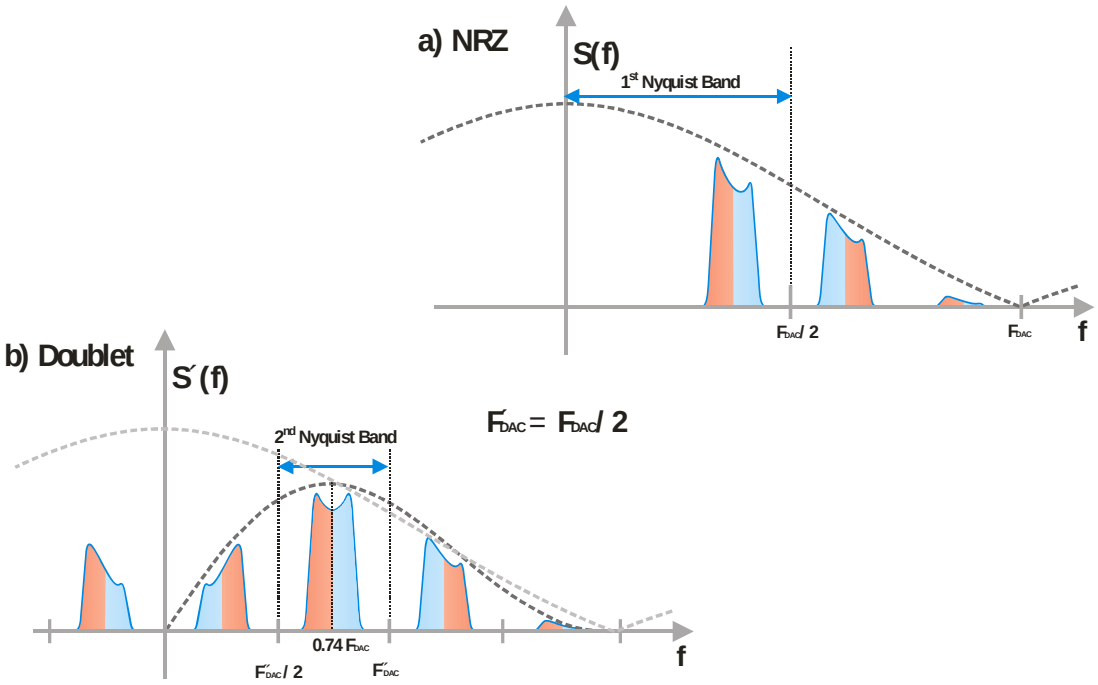


Figure 38: Doublet mode is also useful to generate bandwidth-limited signals even when they are located in the first Nyquist band. In this example, the sampling rate may be halved by using this mode so higher amplitudes and flatter frequency responses can be obtained. One of the consequences of reducing the sampling rate by 50% is that only half of the memory is required to obtain the same time window with the corresponding waveform memory savings.

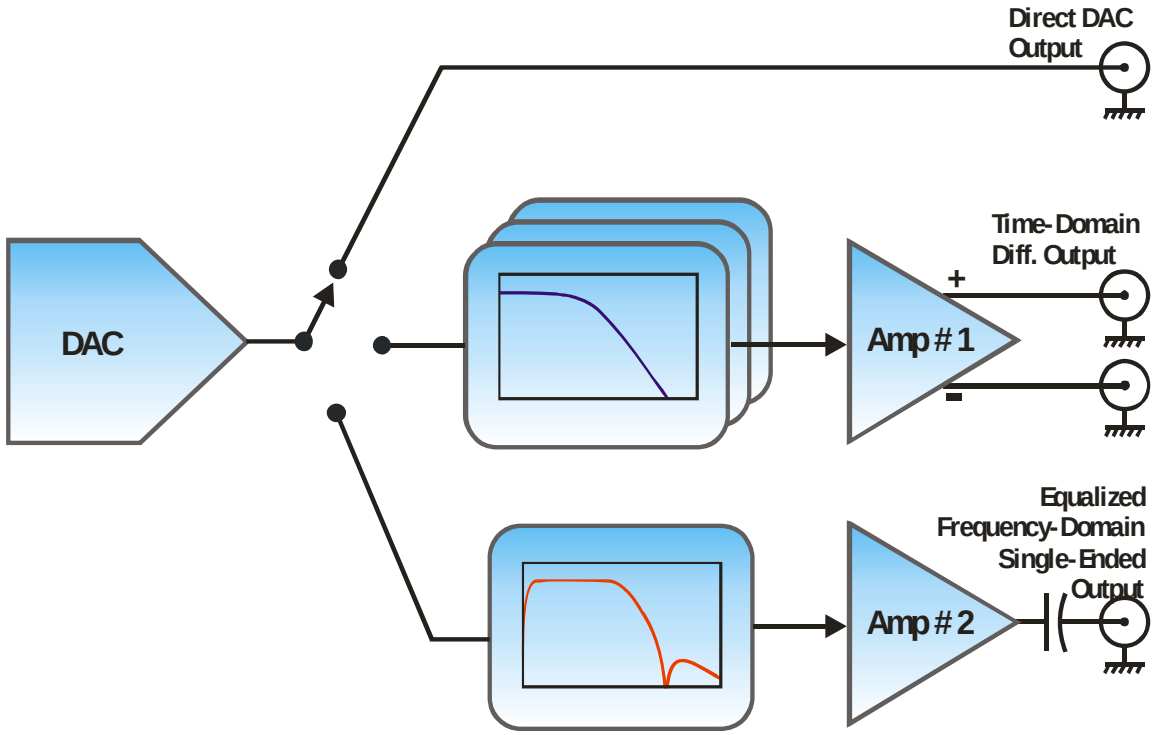


Figure 39: The output stage of an AWG may be composed of several signal paths. High-performance AWGs usually incorporate a Direct DAC Output, offering the highest analog bandwidth and linearity although it may require external signal processing blocks to adapt the signal to the application requirements. Additionally, an alternative bank of filters associated to one or more output amplifiers may be available. Filters and amplifiers are designed to optimize performance in the time or the frequency domains. While some may consist in low-pass filters with a gentle roll-off optimized for the generation of low-ringing differential signals in the time-domain, others may show a bandpass, equalized response to generate high-quality, single-ended signals optimized for their frequency domain behavior.

2.13 Reconstruction Filters

The ideal reconstruction filter response was described in the sampling theory section of this primer. In an actual AWG, the signal is reconstructed by the combined analog response of the whole signal path. The signal processing chain includes the DAC itself, a bank of low-pass (and sometimes bandpass) filters, and an amplifier (Figure 39). Some AWGs may incorporate output modes optimized for signal performance either in the time-domain or in the frequency domain. Requirements may be very different:

- Direct DAC output: In this mode, bandwidth, dynamic range and linearity are maximized. It connects the DAC output to the AWG output, and since there is no further signal processing, noise is minimized and linearity is that of the DAC. The control on the signal characteristics (amplitude, DC offset, etc.) may be very limited and many applications may require external signal conditioning.
- Time-domain optimization: Signal processing for high quality time-domain performance usually implies an analog response with the fastest possible rise/fall times, no additional ringing and good enough image removal. In a high-speed serial signal the previous characteristics result in a low jitter, low intersymbol interference signal with an open eye diagram.
- Frequency-domain optimization: Good signals in the frequency domain must show a good frequency coverage, high-dynamic range, good flatness, and very good image removal.

Implementing a near ideal reconstruction filter in an actual “true arb” architecture AWG is not possible for all of the possible sampling rates since each one would require a different cut-off frequency. Tuneable filters are expensive and difficult to implement, while the filter characteristics greatly depend on frequency. Most instruments incorporate a bank of fixed filters with different cut-off frequencies and responses optimized for several usage models.

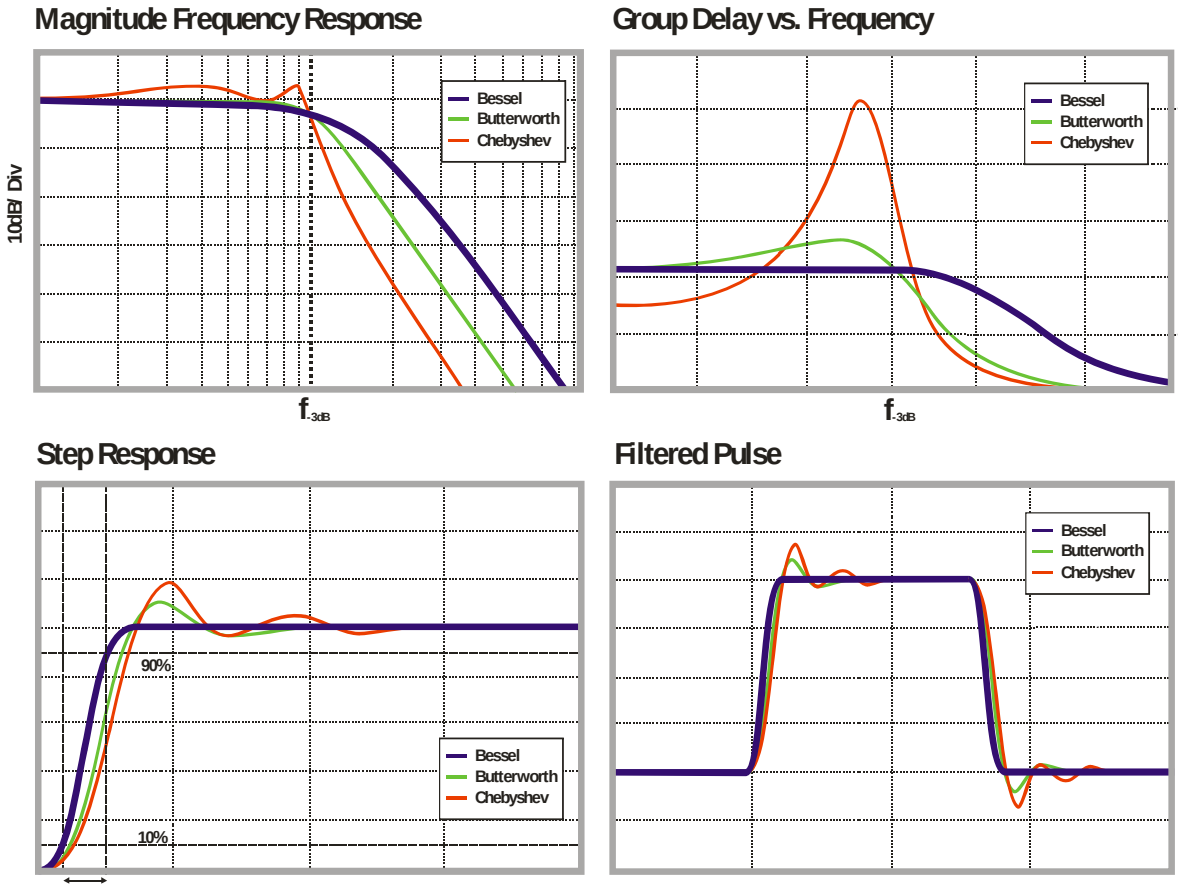


Figure 40: Bessel (Thompson) filters are very popular in AWGs for time-domain signal optimization. When compared to alternative filter types such as the well know Butterworth or Chebyshev, they show a ringingless step response as a consequence of their excellent group-delay response. However, their relatively slow roll-off will result in lower image attenuation. Here the relative performance of 4th order filters can be seen.

Ideal reconstruction filters are not adequate for time-domain signal optimization since any fast transition (i.e. “all zeros” to “all ones”) in the input signal will show ringing and intersymbol interference. Flatness in the pass-band and attenuation in the stop band must be sacrificed in order to obtain a clean step response. Bessel (also called Thompson) filters are very popular in AWGs since they show a constant group delay throughout the passband
(Figure 40). As a consequence, Bessel’s filter step response offers good rise-time performance and does not show any ringing at all. Other more selective (for the same order) and popular filter types such as Butterworth and Chebyshev have better image rejection, but their group delay response results in ringing. Typically, users should select the filter with a fast enough rise-time for the target signal and then create the signal with a sampling rate within the instrument reach, high enough to place unwanted images beyond the desired attenuation level for that filter.

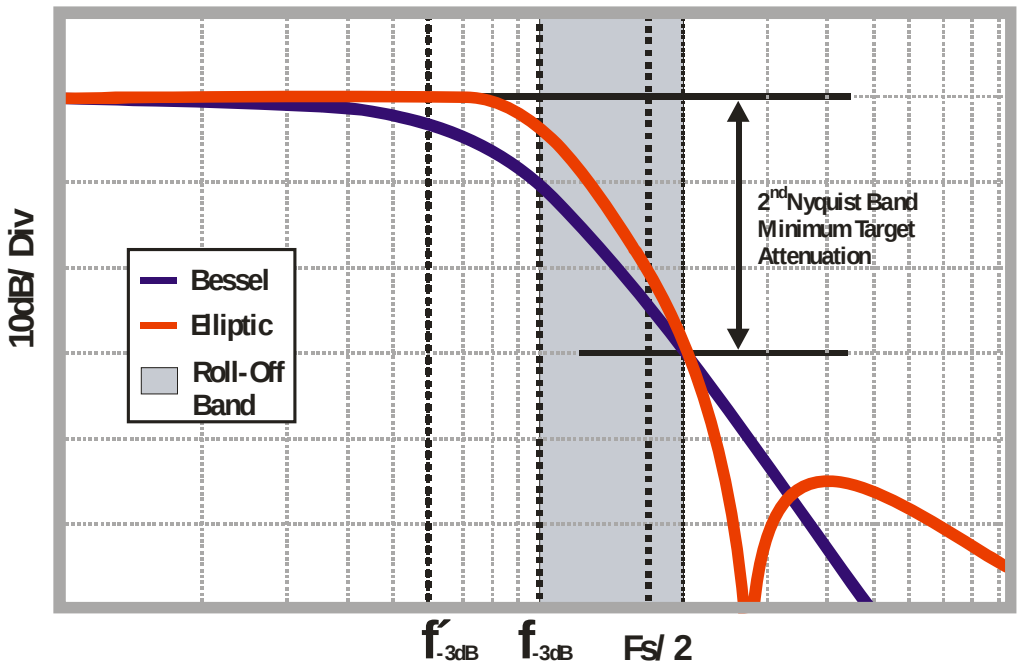


Figure 41: Elliptic (Cauer) filters are more appropriate to reconstruct signals with bandwidths closer to the Nyquist frequency. A Bessel filter with a comparable attenuation in the 2nd Nyquist band would show a much lower 3dB cutoff frequency.

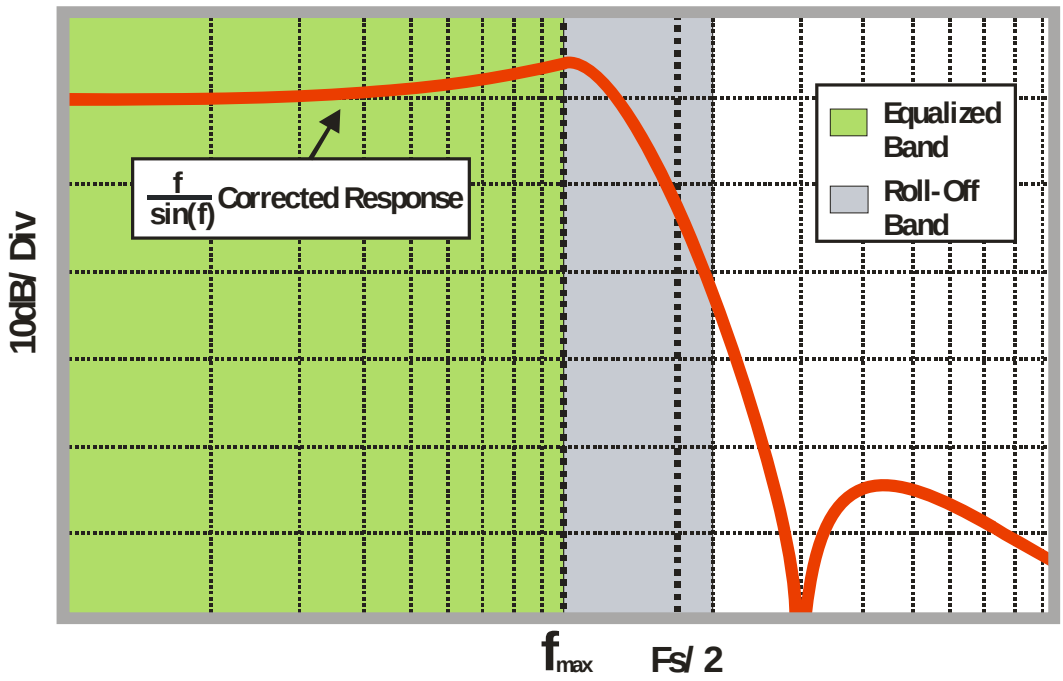


Figure 42: Equalization filters must compensate for the sinc(f) zeroth order hold frequency response within its passband.

In terms of bandwidth, Bessel filters require at least 4 times oversampling relative to the signal bandwidth instead of the 2 times stated by the Nyquist sampling theorem.

For signals with components close to the Nyquist frequency for the highest sampling rates of an AWG, Bessel filters will not attenuate images sufficiently so the DAC steps will be visible in the output signal. In this case, the only solution is to use a filter with a fast roll-off between the pass-band (within the first Nyquist band) and the rejected band that should start in the second Nyquist band. Here, group delay response and step response will not be as critical, so some ringing must be tolerated. The most efficient filters for that purpose are the elliptic filters (also named Cauer) shown in Figure 41. These may be designed with specific parameters regarding ripple in the passband and/or the stopband and selectivity. Some AWGs incorporate elliptic filters for the high frequency cut-off low pass filters, while Bessel filters are kept for the lower cut-off frequency filters. Elliptic filters can be used with oversampling factors as low as 2.5, quite close to the Nyquist limit.

Frequency response flatness is not directly an issue for time-domain optimized signal generation, as transition times and step responses are the primary concern and good performance can only be accomplished through the usage of filters with a gentle roll-off. The overall AWG response will be the combination of the sinc(f) zeroth order-hold response and the filter response where it is not flat at all. For frequency-domain optimized signals (i.e. IF/RF wireless signals) where flatness is important, optimum filters should equalize the sinc(f) response so the overall response is flat (Figure 42). Sometimes a bandpass response may be better as it will remove the DC component of the signal that may be an issue for some devices under test. These equalizing filters are typically calibrated for the highest sampling rate. They will boost the high frequency signals and also any high frequency noise (i.e. quantization noise) present in the DAC output. Beyond the filter, time-domain and frequency-domain optimized signals may use completely different signal paths and even output types since time-domain signals may require differential output while IF/RF signals are generally single-ended.

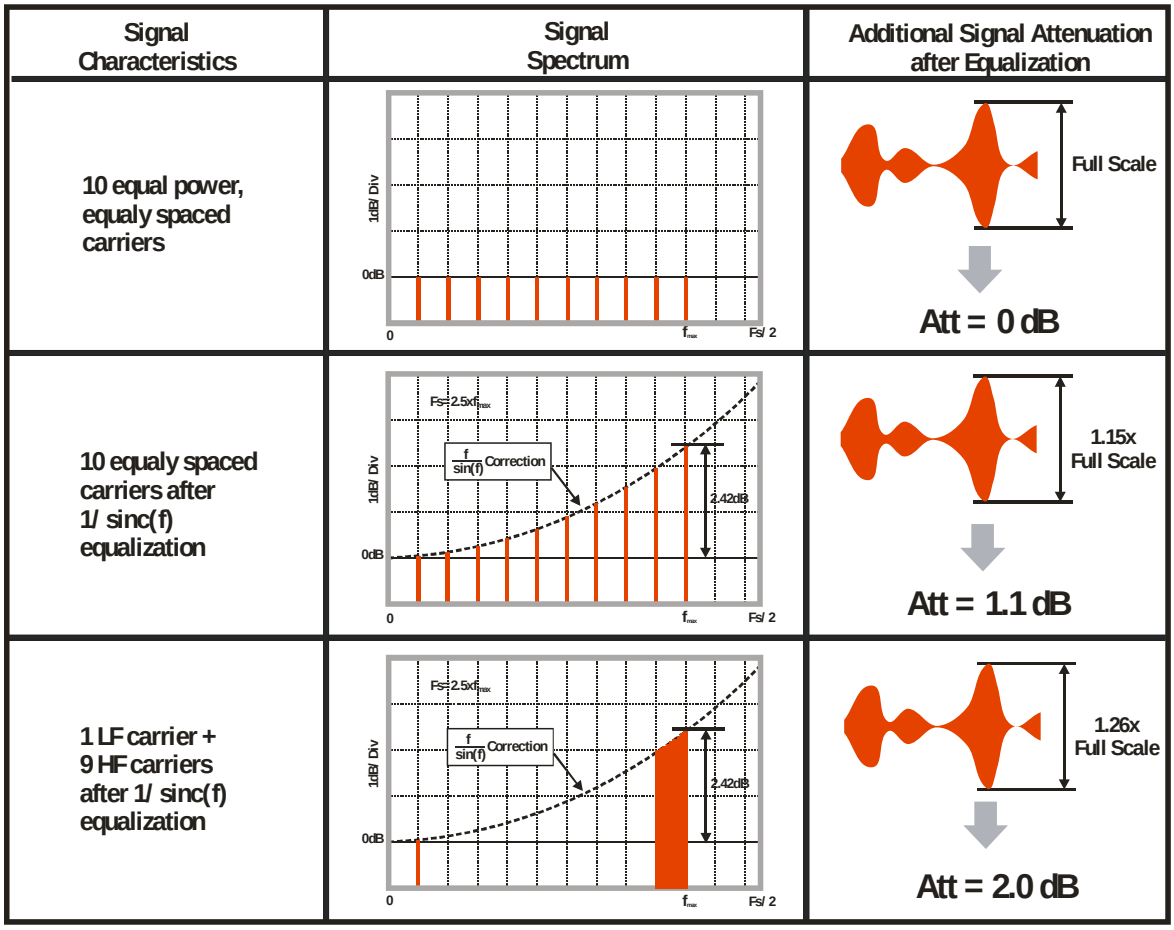


Figure 43: Digital pre-distortion may be used to compensate for the combination of the DAC and signal path frequency response, but it causes a reduction in the dynamic range. In this example, ten evenly distributed carriers (top) are predistorted (center) to compensate for the sinc(f) attenuation introduced by a DAC with a sampling rate 2.5 times the maximum signal frequency component. The resulting signal must be attenuated by 1.1 dBs to fit in the DAC range. A similar signal with one LF carrier and 9 HF carriers will require more attenuation (2 dB) since most signal power is located in a high attenuation band. The Keysight M8195A AWG goes through an in-factory calibration process where the overall frequency response (amplitude and phase) is established for each channel and the corresponding factors are stored in NVRAM. Users can upload the data and apply specific linear pre-distortions filters to any waveform.

2.14 Digital Equalization

One of the advantages of AWGs is their ability to generate any signal, regardless of whether it is distorted or not. Pre-distorting AWG signals on purpose is a widely used technique for different reasons:

- Signal equalization: Any linear distortion introduced by the whole signal path may be compensated for by equalization. Obtaining the inverse impulse response of the signal path (DAC, filters, amplifiers, cabling) through system characterization is a prerequisite.
- Channel-to-channel and instrument-to-instrument matching: Some applications require multiple matched channels (i.e. IQ signals for quadrature modulation). Depending on the level of performance required, the only way to properly match the channels may be by applying differential equalization to all of them.
- Embedding/de-embedding: Sometimes signals must be applied through some intermediate signal path (i.e. a fixturing in an ATE system) or they must emulate a given signal path or processing block. In these cases the original signal must be pre-distorted to emulate (embedding) or compensate (de-embedding) those effects.
- Bandwidth extension: Boosting the high-frequency components affected by the signal path response roll-off may extend usable bandwidth. This strategy can only be applied for moderate attenuation levels (< 10dB), otherwise noise and effective bit performance will suffer.

Some AWGs, such as the Keysight 33500A series or the Keysight M8195A, can even apply linear distortion in real-time by applying a DSP block to the samples coming from the waveform memory. The M8195A AWG incorporates the frequency response for each channel in NVRAM established by a calibration/characterization process performed during the manufacturing process. However, for AWGs without that DSP block, signals must be pre-distorted before downloading them to the waveform memory. In some Keysight AWGs, mainly designed for the generation of IQ signal pairs, pre-distortion is applied transparently to all of the signals downloaded by users by applying equalization and channel-matching FIR filters whose coefficients are stored in non-volatile memory after calibration in the factory. In this way, an excellent flatness and quadrature imbalance and error performance may be obtained.

Linear pre-distortion of AWG signals does not come without consequences since this process often results in a reduction in the dynamic range (Figure 43). In most situations, pre-distortion basically consists in boosting the amplitude of the frequency components of the signal attenuated by the generator and/or any external device. Boosting those components will increase peak-to-peak amplitude of the lineally distorted signal. For instance, if the original non-distorted signal was normalized to use the full-scale of the DAC, the new higher peak-to-peak amplitude of the distorted signal must be attenuated to fit in the DAC range to avoid clipping. Since the overall power of the signal will be reduced while all of the sources of noise will stay the same, the dynamic range will be reduced accordingly. The loss of the dynamic range and amplitude depends on the overall system response and the signal's power distribution over frequency. The range goes from 0dB up to the maximum in-band attenuation.

3 Memory Management and Synchronization in AWGs

3.1 Waveform Memory Access

Conceptually, the connection between the waveform memory and an N-bit DAC may be very simple: a synchronous, N-bit wide bus. The sample address would be provided by an address counter running at FS rate (Figure 44a). The counter would be reset either at the end of the memory or when a predefined record length limit is reached. For continuous generation, the counter will be seamlessly restarted immediately after reset, resulting in a continuous, repetitive signal. Waveform memory becomes then a circular buffer. The only way to implement such an architecture is by using static RAM (SRAM) capable of running at a sufficient read speed. Dynamic memory (DRAM) is banned from this scheme, as refresh cycles would break the flow of synchronous transfers. Static RAM can be much faster than its dynamic counterpart so, traditionally, it has been preferred for faster AWGs although data density performance is poor and achievable record lengths are limited to some tens of mega samples. In this simple architecture, SRAM memory access speed effectively limits the maximum achievable conversion rate as it lacks behind the conversion speed of high-performance DACs.

A direct way to break this limitation is by widening the data bus to transfer more than one sample in one memory read cycle thereby increasing the overall transfer rate accordingly (Figure 44b). A multiplexer attached to the DAC converts the sample flow to the required conversion speed.

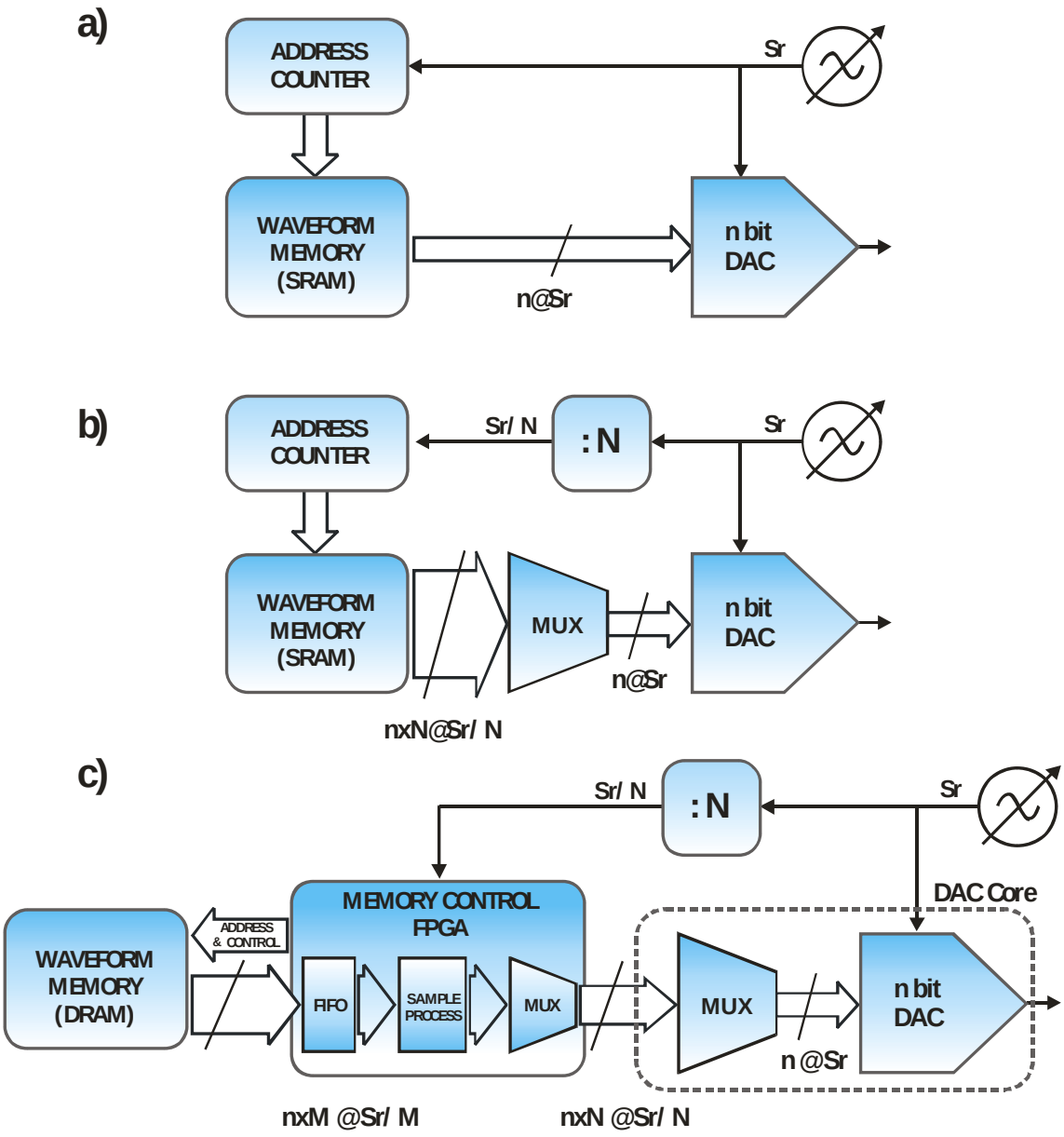


Figure 44: Waveform memory access architectures are influenced by memory technology and access speed. Fast enough SRAM can be directly connected to the DAC (a). If conversion speed is too high then it is possible to transfer more than one sample in a memory access cycle by widening the bus and using a multiplexer close to the DAC (b). Usage of DRAM requires a more complex memory access architecture but it pays off as record lengths may be huge (c).

This method permits the usage of slower, cheaper memory chips and simply adding more lanes is enough to support faster DAC speeds. However, this approach has two drawbacks: waveform length granularity and bus complexity. As data is transferred to the DAC in chunks of N samples, any achievable record length must be a multiple of N. This has been the traditional memory access approach for state-of-the-art high-speed AWGs but resulted in record lengths below 100 MSamples (or 5ms @12GSa/s for a 64MSamples record length).

To go beyond the cost and performance limitations of synchronous memory access using SDRAM memory, a different approach must be taken. The only practical way to reach the long record lengths (>1GSample) at high conversion speeds required by some applications is by switching to DRAM memory technology. DRAM memory features high data density and low cost making it feasible to reach multi-Gigasample memory sizes. However, the need to access the memory asynchronously results in a more complex waveform data access architecture Figure 44c). As an example, it is necessary to add buffering (FIFO structure) in order to accommodate the asynchronous data flow due to processes such as dynamic memory cell refreshing. For the same reason, the average transfer rate must be at least equal to that required by the DAC. The width of the data bus and the characteristics and performance of the memory control system will again resulting record length granularity, and in most cases, a minimum record length typically a multiple of the granularity.

Parameters such as maximum (and minimum) record length, their dependence, if any, on the sampling rate, and the record length granularity are some of the specifications related to waveform memory (Table 3). Granularity limits a user's freedom in signal design, but its effects in most situations are negligible, especially for sufficiently long waveforms. However, special care must be taken in order to avoid unwanted wrap-around glitches since most AWGs just truncate record length to the nearest valid value in the case that the waveform download does not meet the instrument's requirements regarding record length.

Transferring waveform data between the waveform memory and the DACs has become one of the most challenging design problems faced by AWG designers, especially for very high speed devices. In some cases, it is virtually impossible to transfer all of the data required for all of the channels running at full speed. For example, in Keysight's M8195A, , there are two banks of waveform memory: one 4x256KSa implemented inside the 4-channel DAC ASIC capable of feeding all of the DACs at full speed simultaneously and another 16GSample DRAM block connected to the DAC ASIC through a waveform memory handling and synchronization/trigger FPGA. This massive DRAM memory can be accessed at full speed for one channel, at half of the speed for up two channels and at ¼ of the speed for 3 or 4 channels. The internal DSP block in the DAC ASIC can take care of oversampling the incoming waveforms if necessary so that the DACs are fed with waveform data at full-speed, increasing signal quality while keeping images well beyond the waveform bandwidth. The internal DSP is also used to reduce the amount of data stored in either the internal or external memory banks or even to generate real-time waveforms such as PRB sequences.

Table 3: Waveform memory specification for some of Keysight AWGs

	Sampling Rate	Minimum Record Length	Maximum Optical Record Length	Record Length Granularity
33522A	250 MSa/s	8 Samples	16 MSamples	1 Sample
N8241A	1.25 GSa/s	128 Samples	16 MSamples	32 Samples
81180B	4.6 GSa/s	320 Samples	64 MSamples	32 Samples
M8190A Option-14B	8 GSa/s	240 Samples	2048 MSamples	48 Samples
M8190A Option-12G	12 GSa/s	320 Samples	2048 Msamples	64 Samples
M8195A Rev. 1	65 GSa/s (4 Ch)	128 Samples	256 Ksamples (Internal to the ASIC)	128 Samples
M8195A Rev. 2	65 GSa/s (1 Ch) 32.5 GSa/s (2 Ch) 16.25 GSa/s (4 Ch)	128 Samples	16384 Msamples (External to the ASIC, shared among channels)	128 Samples

3.2 Memory Segmentation

The most simplified model for AWG memory is just a single bank of memory where the user downloads the signal to be generated (Figure 45 left). If more than one waveform is required over time, each one must be loaded into the memory whenever needed. Typically, downloading a new waveform requires either a transfer from some mass storage device (internal or external) or just transferring it as some software, internal or external, computes all of its samples. A different model allows users to store multiple waveforms in separate memory banks (Figure 45 center). Switching the output waveform is as simple as selecting the bank. An advantage of this method is that new waveforms can be downloaded to any inactive bank while the instrument is generating a signal. Banks can be actual waveform memory, fast memory directly connected to the DAC, or just fast access memory that can be quickly downloaded to the actual waveform memory. These two models are common in lower performance AWGs. However, most high-performance AWGs, allow for memory segmentation where massive amounts of waveform memory can be dynamically assigned to several waveforms called segments (Figure 45 right).

Through segmentation, multiple waveforms can be stored simultaneously, ready to use, in the generator's waveform memory. Storing multiple waveforms is useful because users can switch from one to another in a very short period of time just by selecting them from a list of pointers instead of having to download them from a computer or from some shadow memory.

Segmentation is much more memory efficient than using memory banks since users have much more flexibility in the way waveform memory is allocated, and many more waveforms may reside in a given amount of memory. Additionally, switching from one segment to another may be extremely fast, potentially at sampling speed, since there is no waveform transfer or bank switching processes associated to it. This is especially useful in ATE systems where download time of waveforms may be an important component of the time required to set-up a test step.

Record length and granularity limitations also apply to segments. In addition, total amount of segments may be limited as well, reflecting the characteristics of the internal look-up table supporting them within the AWG.

In some cases, it may be difficult to obtain full transfer speed for all of the channels when very long record lengths are required. In the Keysight M8195A AWG, there are two dissimilar waveform memory banks. There is a limited size bank (1 MSample) in the same ASIC that the four DACs are located and capable of feeding all the converters at full speed. For applications requiring longer record lengths, an external, DRAM-based 16GSample memory can supply samples at full speed for one channel, half this rate for two channels, and ¼ this rate for 4 channels.

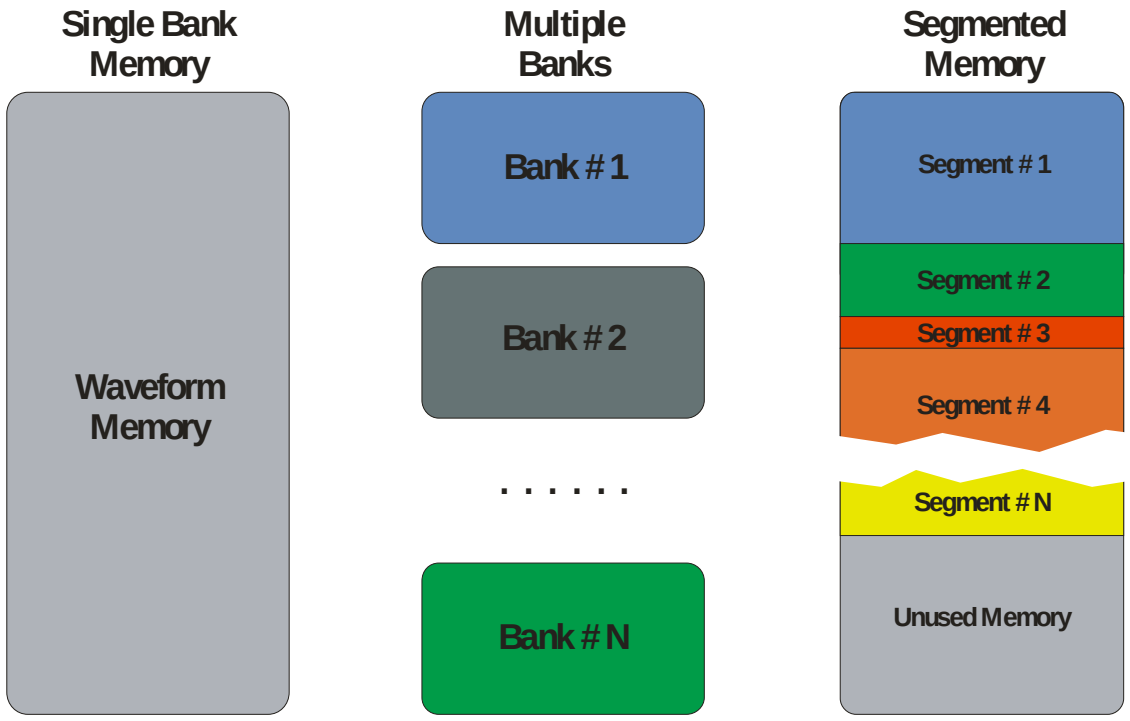


Figure 45: Waveform memory can be arranged in different manners. Single bank memory allows loading just one waveform at a time. Multiple banks can store one waveform per bank, and quick but not seamless, signal switching. Segmented memory allows sharing a unique waveform memory among multiple, variable length, waveforms. Signals can be switched seamlessly so waveform sequencing is made possible.

3.3 Waveform Sequencing

Probably, the most important outcome of memory segmentation is the ability to perform seamless waveform sequencing under the control of a sequencer. In this context, jumping from one segment to the next occurs seamlessly without any gap or additional jitter between the last and first samples of consecutive segments. To be useful, sequencing requires users to be able to define the sequence list and some generation parameters for each entry, such as the number of times that every specific step must be repeated during execution (Figure 46) or the branching conditions. Sequencing has multiple applications and advantages that often can be exploited simultaneously:

- Waveform memory preservation and decreased download time: Often, long, complex waveforms can be split into shorter segments that are repeated several times within the waveform. Using sequencing, each of these pieces can be stored as segments and then played-back in the right order under the control of a user-defined list handled by the sequencer. Breaking up the waveform into these segments can require less memory (Figure 46) to store the waveform and allow it to be downloaded faster than a single un-segmented waveform. Good examples of such waveforms are radar or video signals (Figure 47). In radar signals, short pulses are transmitted with a relatively slow PRF (Pulse Repetition Frequency). With sequencing, the interval between pulses may be synthesized by repeating multiple times a “quiet” segment, while another one will contain the pulse itself.
- Complex signal creation and edition tool: Sequencing may be a valuable waveform creation and modification tool. Video signals may be a good example of this. In Figure 47, a compliant SMPTE colour bars signal is synthesized using sequencing. As an example, modifying the signal to add some grey scale area simply requires the definition of a new segment with one grey stair line and adding a new entry in the sequence list to finally modify the repetition number of the SMPTE bars so that the resulting signal is made of 240 active lines per field.

- Test Executive for AWGs: Many sequencers allow branching between segments under programmatic control. This capability may be used in ATE systems to control the flow of signals very quickly and releasing the control software from clumsy and slow waveform downloads, saving time and simplifying software. A complete test procedure may be divided into sub-sequences, each one involved in a specific step of the test. The sequencer can wait for some command or signal (i.e. the activation of a trigger line in a PXI bus) at the end of each subsequence before jumping to the next step.

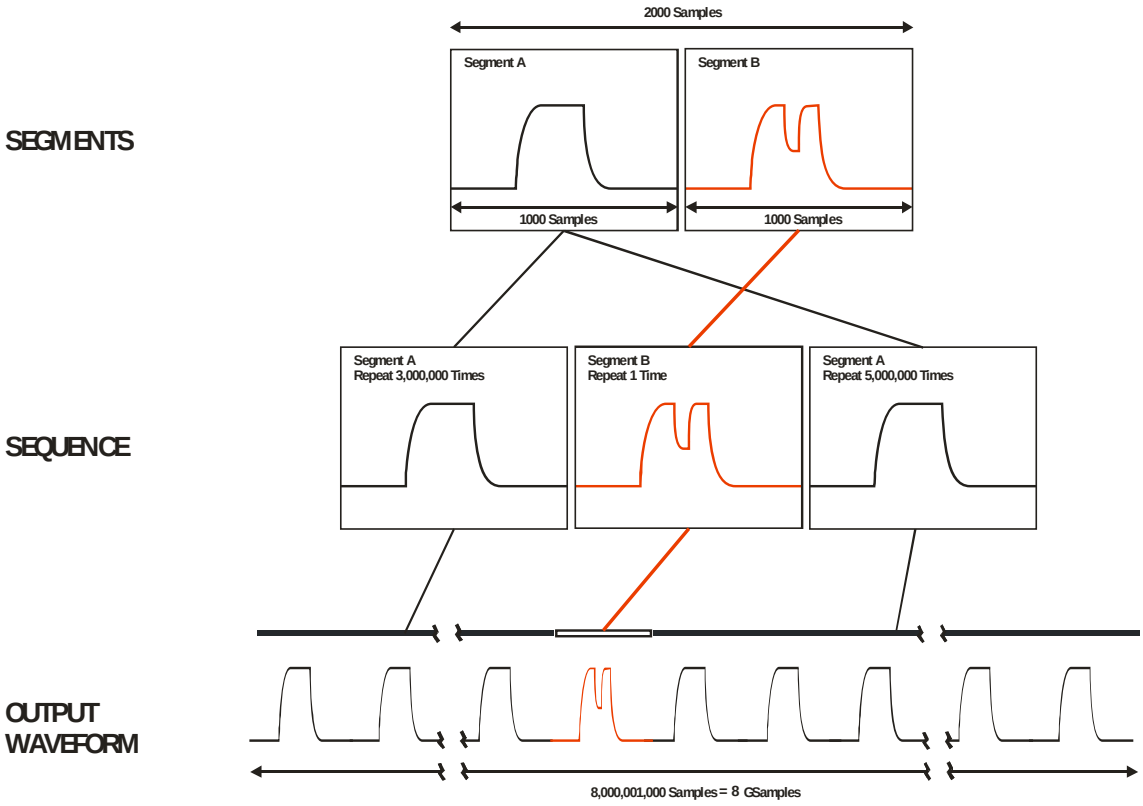


Figure 46: Sequencing can be used to synthesize very long waveforms, even beyond memory limitations, by exploiting regularities in some waveforms. Memory savings may be huge depending on the signal. In this example an 8 GSample waveform is created through sequencing from just two segments of 1000 Samples each.

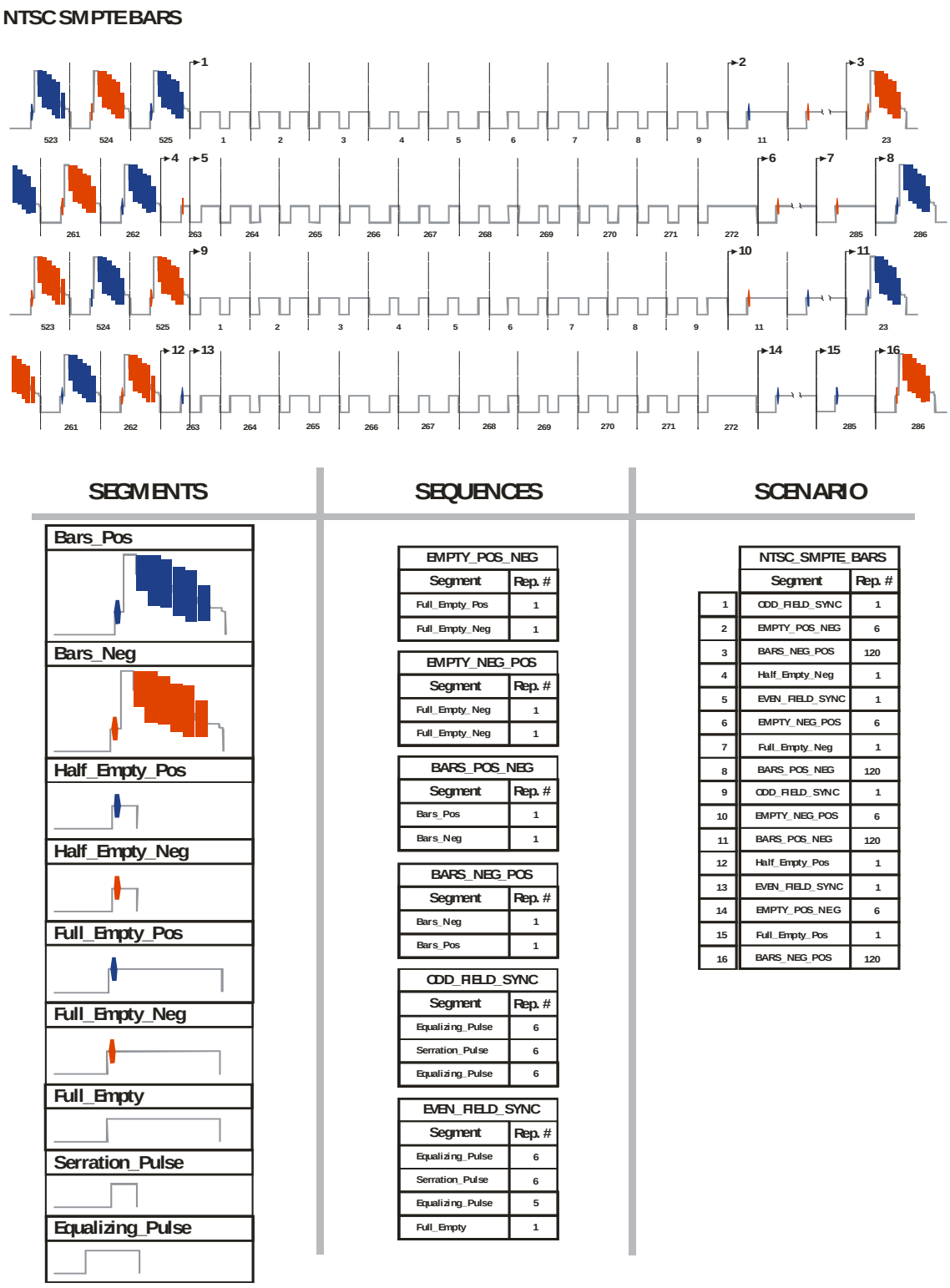


Figure 47: example of segments, sequences and scenarios. Sequencing may be used to generate very complex signals. In this case, a fully compliant NTSC SMPTE bars signal is synthesized using just nine segments. Static NTSC signals require 4 full fields to repeat exactly in the same way. The difference between the segments colored in red and blue is the burst' phase. The usage of scenarios (sequences of sequences) results in an even simpler definition of the signal.

Table 4: Waveform sequencing specification for some of Keysight's AWGs

	# of Segments in a sequence	# of Sequences	Scenario Support	Dynamic Scenario/ Sequence Control	Segment Length Range
33522A	512	32	No	No	8 Samples- 16 MSamples
N8241A	32K	32K	Yes	Yes (13bits)	128 Samples- 16 MSamples
81180B	16K	1000	Yes	Yes(8 bits)	320 Samples- 64 MSamples
M8190A Option-14B	512K	512K	Yes	Yes (19 bits)	240 Samples- 2048 MSamples
M8190A Option-12G	512K	512K	Yes	Yes (19 bits)	320 Samples 2048 MSamples

- Synchronization: Some AWGs can use external event or trigger signals to control the way the sequencer behaves. This capability allows for easy synchronization with real-world events (i.e. trigger events from other instruments) in real-time. The sequencer itself may control some digital outputs so external devices can be synchronized as well (i.e. an oscilloscope could be triggered at a very specific instant in the output waveform). Latency, the lapse of time between the trigger signal and the sequencer action, may be an issue in some applications so this is a critical specification.
- Block and system emulation: The usage of external signals capable of controlling the sequencer permits the generation of waveforms at specific moments in time. In some instruments, external signals can even control the specific segment or sequence to be executed. This strategy can be used to emulate the behaviour of complete functional blocks and systems as signals show up at the output depending on externally provided stimuli.

Sequencing performance depends on several characteristics, both qualitative and quantitative:

- Number of segments.
- Number of segment entries or steps in a given sequence list supported by the sequencer.
- Number of sequences.
- Execution and jump control for each step: Typically, the number of times the segment must be repeated for each step can be defined. This number may be set to “infinity” so the step will be repeated indefinitely until the generator is stopped or some jump condition is met.
- Sequence of sequences (Scenario) support: Adding another level to sequencing, so complete sequences may be sequenced as well, results in improved flexibility and additional memory savings (Figure 49). A sequence of sequences is also known as a “scenario”. It can also help to greatly simplify the creation or edition of complex sequences.

The sequencer is in fact a state machine so every time it jumps from one step to the next, the state changes. As any other state machine it requires some time to transition from one state to another. It is a similar situation to the advanced triggering systems found in logic analyzers or oscilloscopes. The sequencer transition time adds another requirement to segments that are part of a sequence, as its duration must be equal to or longer than the sequencer transition time. For this reason, segments must be longer than a given number of samples to make sure that the sequencer will be capable of handling branching actions before the end of the first iteration of any segment, even at the fastest sampling rate. Otherwise, seamless, glitch-free generation cannot be guaranteed. When segments shorter than the minimum allowable length are required, users can simply create new link segments consisting in the union of the short segment and the next segment in the sequence and/or by repeating the same short segment until the total length exceeds the minimum segment length, depending on the specific signal generation requirements (Figure 48).

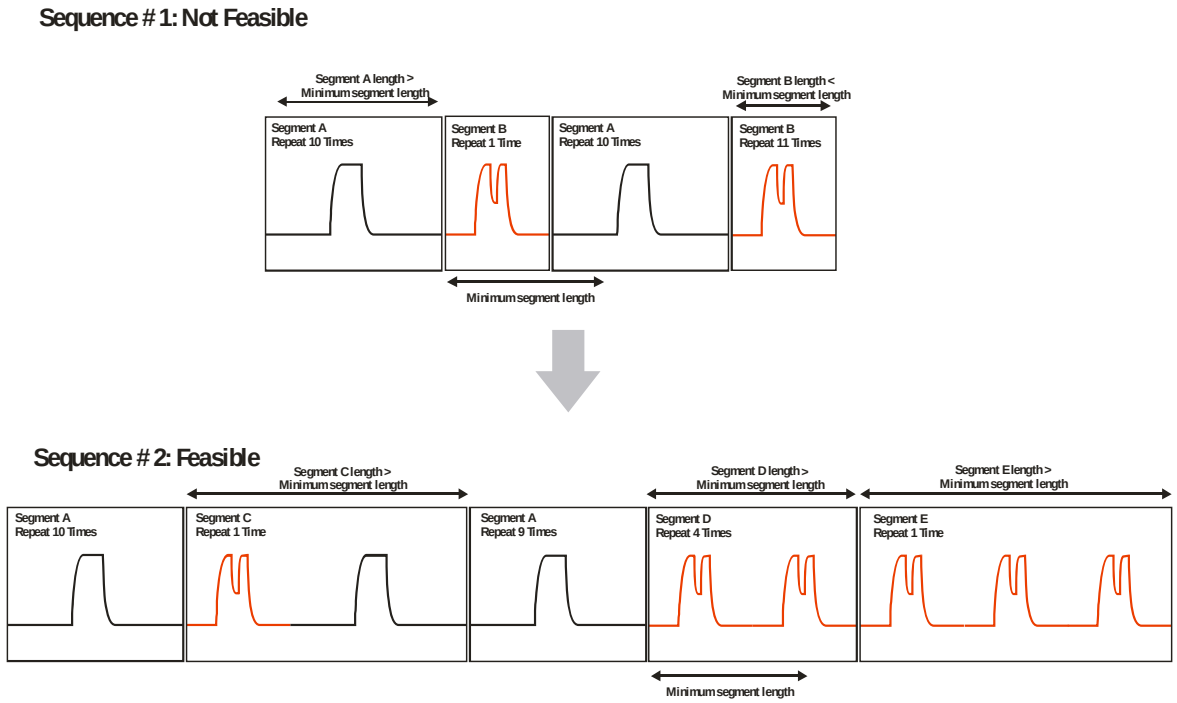


Figure 48: Minimum segment length may be a problem in some situations. However, most times it is possible to solve the issue by generating some additional link segments. Here, three new segments replace a segment too short to be generated: one built by concatenation of two consecutive waveforms and the other two by repeating the short segment several times. The resulting new segments are now longer than the minimum segment length, and they can be combined to obtain the desired sequence.

Scenario (sequence of sequences) support may be implemented in different ways. The most direct way is to use separate lists for sequences and scenarios. In such an arrangement, a scenario memory includes one or several lists of pointers to the sequence memory along with looping and branching information for each sequence in the list. In this way, changing one of the sequences will have automatic consequences in all of the scenarios using that sequence. Using an additional sequence control layer may impact the maximum sequencing speed so this solution may not be desirable for very-high speed AWGs. A different approach is to integrate the scenario related information in a unique sequencing memory. In this mode, sections of the sequencing memory are assigned to different scenarios. Scenarios are then defined by a series of sequences and sequences by a series of segments.

Looping information and branching information for each sequence is also stored in the list. This approach results in simpler and faster handling since all of the information is obtained from the same source.

The only drawback of this method is that repeating sequences will show up repeated times in the sequencing memory so more sequencing memory is used. Changing or replacing that sequence will require tampering with all of its occurrences in the sequencing memory. Both limitations can be overcome with large sequencing memory and a good sequence design strategy.

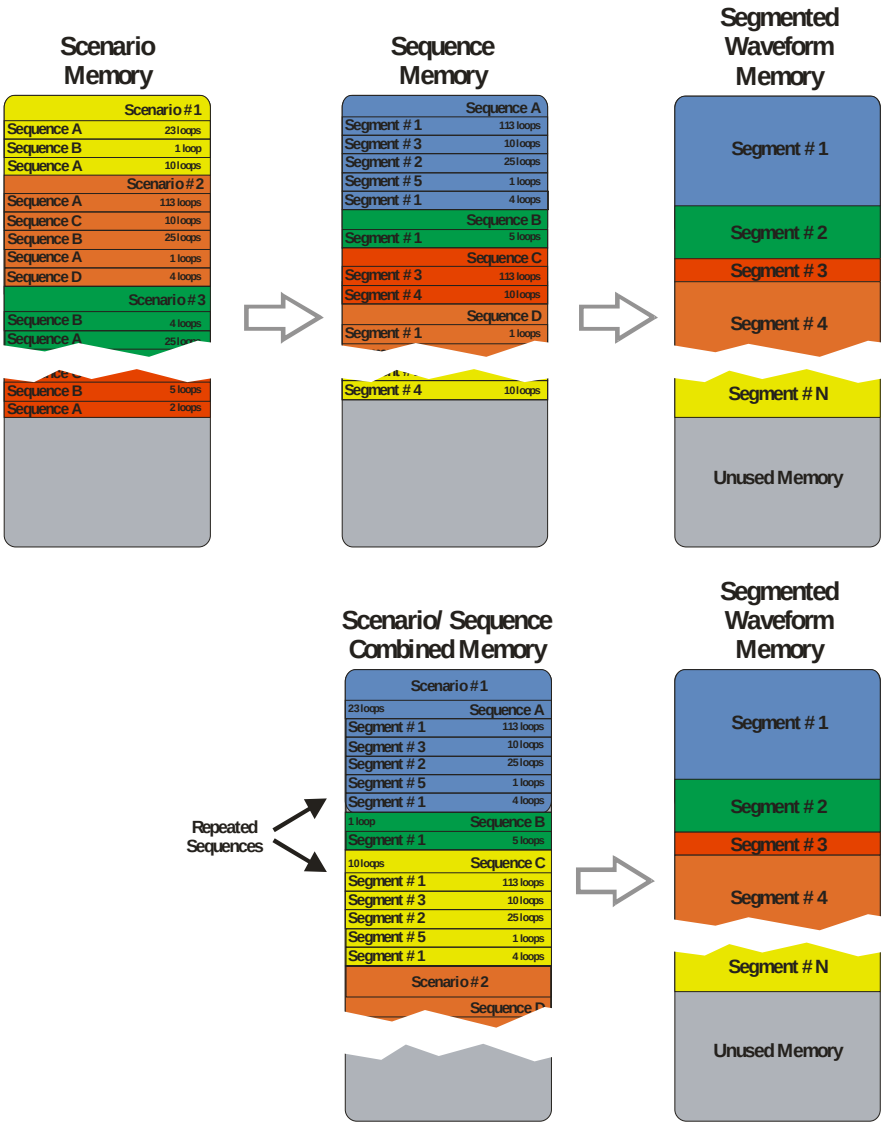


Figure 49: Scenario (sequence of sequences) support adds another degree of freedom to sequence design. It greatly simplifies the set-up and modification of some complex signals. Scenarios may be implemented in a specific memory pointing to the sequence memory or integrated in it. The latest solution is potentially faster and simpler to implement, but repeating sequences must also be repeated in the sequencing memory.

3.4 Sequence Control

Beyond saving valuable waveform memory in an AWG and simplifying signal design, sequencing can be used to dynamically control the output signal according to external events. In order to accomplish that goal, steps in a sequence, sequences in a scenario, and multiple scenarios must respond to predefined conditions through predefined responses. As a consequence, every element in a sequence or scenario must convey information about what to do at the end of the current state and how to advance to the next step. High performance AWGs typically support different advancement modes:

- Automatic: The sequencer jumps seamlessly to the next segment or sequence in the list immediately after completion of all of the loops defined for the current step.
- Conditional: The current step is looped until a predefined event occurs, and then the sequence jumps to the next step seamlessly. In some AWGs, the jump may be immediate, without completing the current iteration, or synchronous, so that the current iteration is completed.
- Repeat: The sequencer waits for an event after completion of all of the loops for the current step in order to jump to the next step. While waiting, the generator's output is set to an "idle" state, typically the DC value of the last sample in the current step or a user-defined level. When this advancement mode is set for all of the steps, the sequence will be executed in "Stepped Mode".

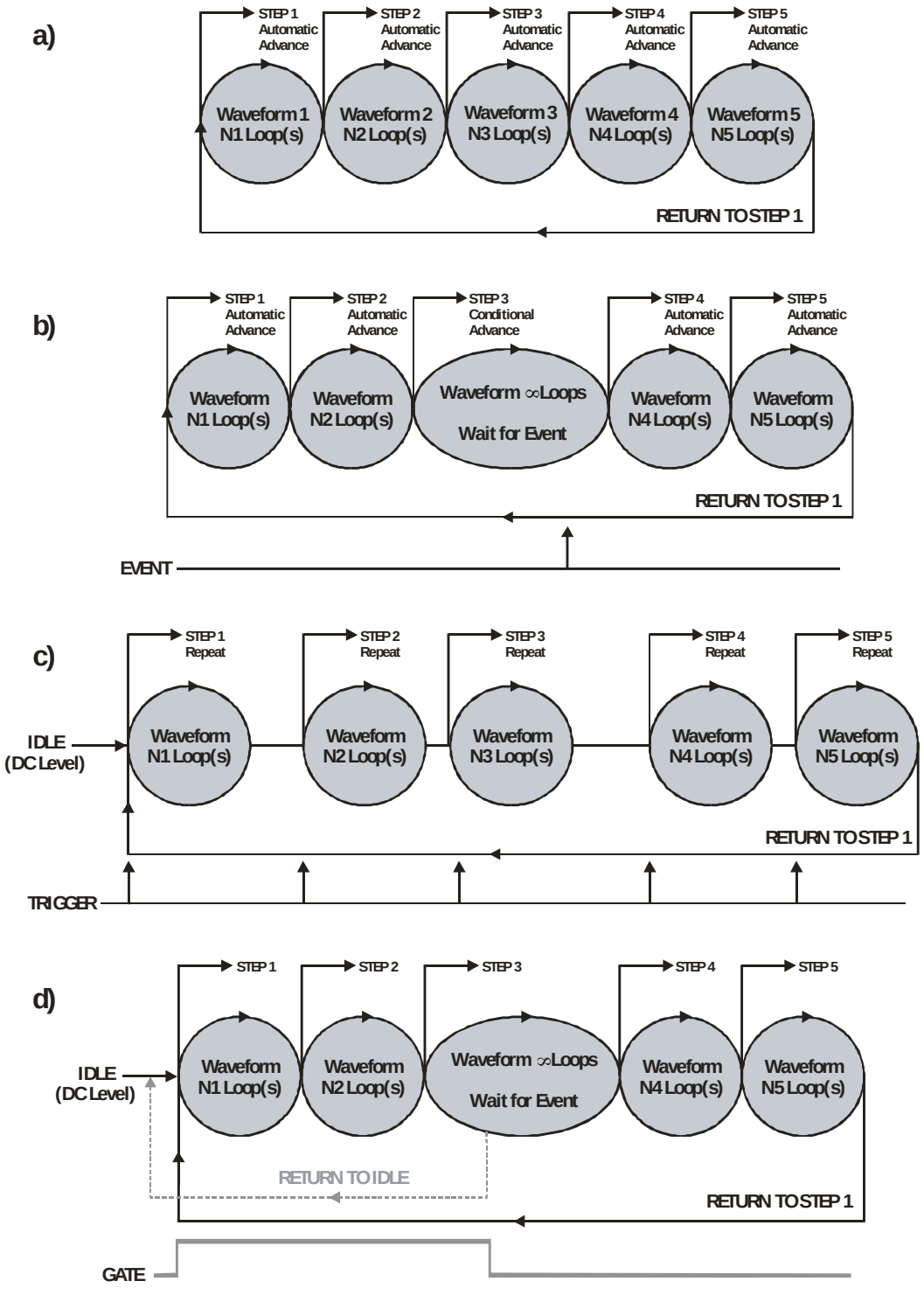


Figure 50: Advance modes for sequences and scenarios. In automatic advance mode (a) the sequence jumps from one step to the next without waiting for any event. Conditional advance mode (b) can be applied to specific steps that will be looped until an event shows up. In repeat mode (c) after completion of all loops for a step the generator enters an “idle” state (generally a DC level) while it waits for a trigger event to advance to the next step. Gated mode (shown in d) applies the repeat mode to all steps in a sequence. In gated mode (d) a gate signal controls the sequencer so it only runs when that signal is active.

Events may come from a variety of sources such as Trigger/Event inputs or commands sent programmatically by some control software. Event latency and maximum frequency are some parameters that affect the performance of event-based sequencing. Latency is the time required by the sequencer to take action after some event shows up. It can greatly influence the usability of the AWG in some applications, since in many occasions, responses to stimuli must happen within a limited time window. All of these events are associated to an “edge” rather than to a level. A transition in the event signal in the right direction will cause the action taken by the sequencer.

The same or other inputs may be used as a gate signal (gated mode). Gate signals are active while a given level is kept. The behaviour of gated modes may be different from instrument to instrument, but it basically enables the sequencer so it only runs while the gate signal is active. In some instruments, the sequencer just stops when the gate signal is off and continues when it is active again. In other ones the transition of the gate signal to “off” may reset the sequencer in addition to stop it, so when the gate signal is active again the sequence will be restarted.

3.5 Dynamic Sequence Control

Sequencing may solve many signal generation requirements saving valuable waveform memory, reducing the time required to transfer them, and easing complex signal creation and edition. However, the sequencer toolbox available to users may sometimes be limited when dealing with some real-world situations. Users can basically setup the order of generation of the different segments in a sequence and sequences in a scenario, the number of loops for each step, and control the sequence advances to the next step through a few hardware or software signals. This scheme is very powerful to control the timing of a predefined waveform according to external events that show up in an expected sequence. Some tests may require generating a different waveform or a different sequence of waveforms depending on the status of the DUT/SUT.

Dynamic sequencing solves this need by transferring total or partial real-time control of the sequencer to an external device.

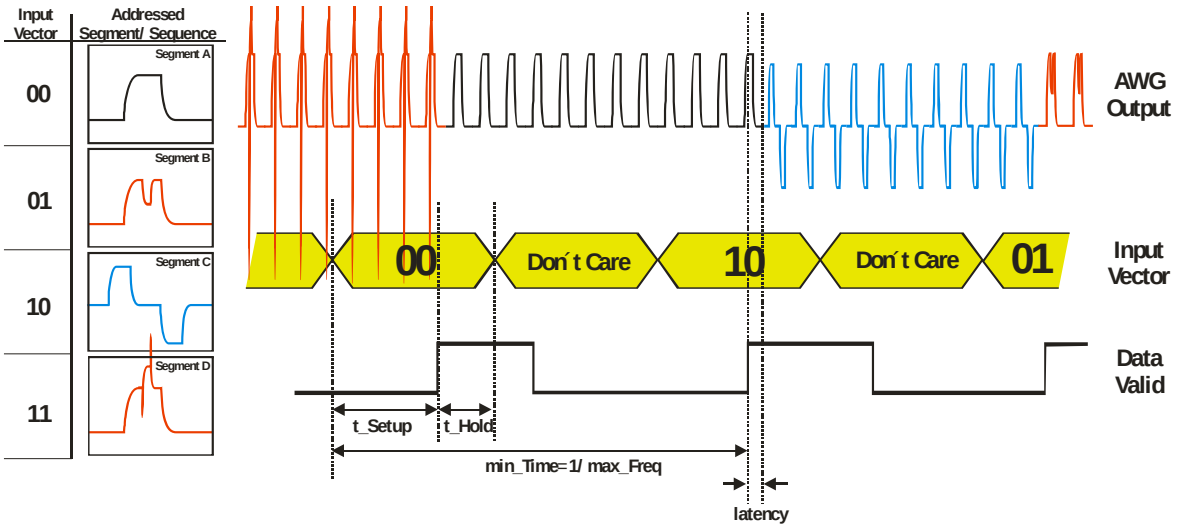


Figure 51: In dynamic sequencing, an input vector points to the next segment in a sequence, sequence in a scenario, or scenario in a sequencing memory to be generated. This allows for random signal access and emulation of complex behavior through external control. In this example a total of four different waveforms may be randomly selected through a 2-bit input vector so the order of occurrence does not have to be previously set up by the user. Maximum vector speed and latency time may limit the usefulness of dynamic sequencing.

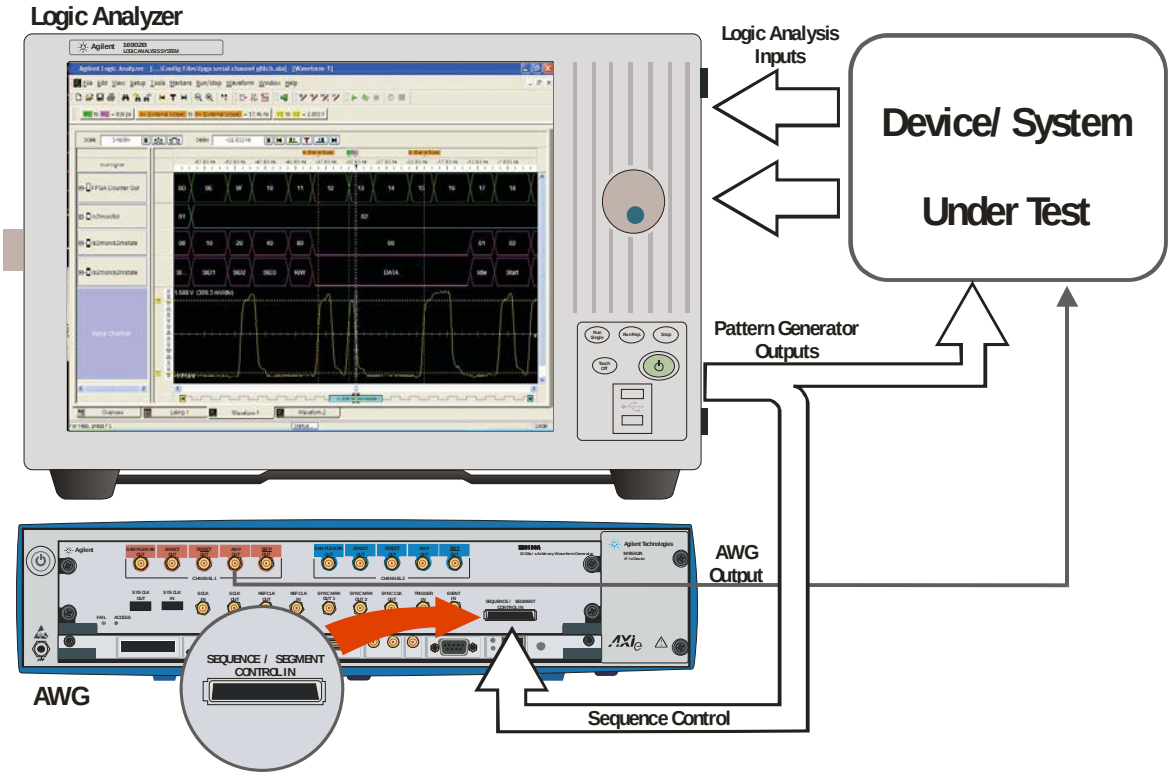


Figure 52: Dynamic sequencing provides appropriate responses in front of conditions coming from the device under test or the real-world. Conditions may be identified by the DUT itself, custom circuitry, or, like in this example, advanced instrumentation such as a high-performance logic analyzer. These can detect complex sequences of events through their trigger system and supply dynamic sequence control through built-in pattern generators.

Dynamic sequencing may be handled by software from an external controller capable of detecting external conditions. In this case, once the condition is detected, the software can activate the required output waveform through a software command sent to the AWG. Software latency may be an issue in many applications so some high-performance instruments incorporate an additional hardware-based dynamic sequence control. In this scheme, the signal to be generated is selected through a digital word, or vector, applied to a parallel input and validated by some data-valid signal. The user still must predefine what segment, sequence, or scenario is assigned to each word, but some external device will take the highest-level sequence control. As segments and sequences can be accessed in a random way, their order in time will not be predefined or fixed as it happens in traditional sequencing. Dynamic sequencing control signals may come from a variety of sources:

- DUT: Signals coming from the DUT. Sometimes it may be necessary to modify the software/firmware of the DUT to supply the right signals at the right moment in time.
- Specific hardware: Some user-supplied circuit capable of detecting external events and supplying the right vector to the AWG.
- Other instruments: Some instruments can detect complex events though sophisticated triggering systems. Oscilloscopes are a good example. However, most oscilloscopes are limited to output a single signal when a trigger event is detected, therefore its usage may be limited to traditional event-based sequencers in AWGs. Probably the most flexible instrument that can be used to dynamically control AWG sequences is an advanced logic analyzer. First of all, logic analyzers incorporate extremely powerful sequential triggering systems capable of detecting a complex sequence of events, validate their timing and/or number of occurrences, and take multiple actions based on them. Secondly, some logic analyzers have standard or optional embedded high-speed digital pattern generators. These pattern generators can be fully defined by the user in a similar way to segmented memory AWGs. Users can create complex sequences of pattern segments, and best of all, these can interact with the logic analyzer triggering system.

Although the power and flexibility of the dynamic sequencing scheme enables the emulation of extremely complex systems since the AWG will provide adequate waveforms depending on the device under test requirements, it is also limited by some specific parameters:

- Addressing space: The width of the vector pointing to specific segments/sequences within the waveform memory imposes a limit to the possible unique responses the AWG can supply to the device under test. Instruments such as Keysight's M8190A AWG provide a 13-bit input, so 2^{13} (8192) segments sequences are directly addressable. If more are needed, the instrument can handle up to 19 bits in two consecutive address read cycles (13+6) so a total of 2^{19} (512K) unique responses can be provided.
- Switching time (latency): This is the time required to jump to the segment/sequence pointed to by the input vector after validation. This parameter may sometimes depend on the sample frequency, or the way the AWG handles the transition since some complete the current iteration of the current segment/sequence before jumping to the next one. It may also depend on the options available on the instrument, as sometimes fast switching is not a standard feature of some instruments.
- Maximum switching frequency: This is the maximum speed that the AWG can accept new vectors.

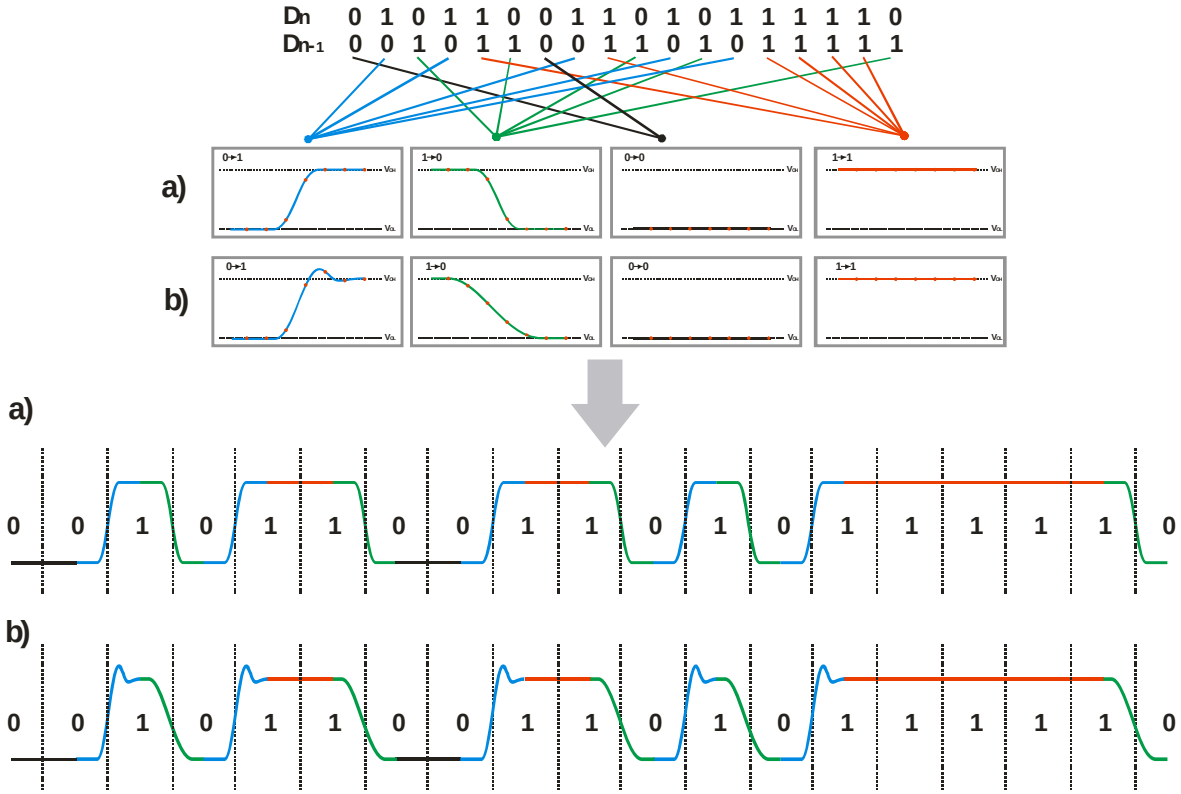


Figure 53: Some Pulse Arbitrary Function Generators such as Keysight's 81160A use dynamic sequencing and real-time data (internal or external) to synthesize long sequences such as PRBS31 that would not fit in their internal waveform memory. For binary signals just four, 1-bit time long segments must be defined: 0-0, 0-1, 1-1, 1-0. This methodology also speeds-up testing as creating a new binary signal with or without distortion only requires the definition of four segments according to the desired parameters. Here a clean binary signal (a) and another one showing duty cycle distortion and overshoot (b) are generated by editing the four segments.

Some instruments such as the Keysight 81160A Pulse Function Arbitrary Noise Generator make clever usage of dynamic sequencing to generate very long binary or PAM sequences by applying an internally generated data stream in real-time or by using an external data input that can be synchronized with the AWG. The solution is based on storing a series of one-symbol long segments with all of the possible transitions and steady levels. These basic building blocks are rather short in terms of samples so creating and loading new ones to the instrument is as fast as it is simple. Those segments can be designed with distortion added to them so realistic impairments can be simulated

While binary signals require only four segments (0->0, 0->1, 1->0, 1->1), PAM signal generation requires more depending on the modulation level (i.e. 16 segments for PAM4). In Figure 53 the same binary sequence results in either a clean or a distorted signal depending on the analog waveforms loaded into the instrument.

Some popular PRBS sequences require extremely long record lengths (PRBS31 is 2^{31}-1 bits, 2 Gbits long) if they would be generated using a traditional single-segment, true-arb approach. Even shorter sequences may take an unacceptable long time to be calculated and downloaded.

3.6 Synch Signals and Markers in AWGs

Event and trigger signals are inputs to synchronize an AWG with the DUT and other instruments. Sometimes it is necessary to synchronize the DUT or other instruments with the AWG. Some specific signals for this purpose are available in most high-performance AWGs: Synch and marker outputs.

Most signal generators incorporate synch signals that are typically used to trigger additional instrumentation such as oscilloscopes. Often, synch signals are associated to a single, important event such as the start of a sweep or a burst. In AWGs, synch signals may be associated with a segment although the user can often establish their position and duration as well.

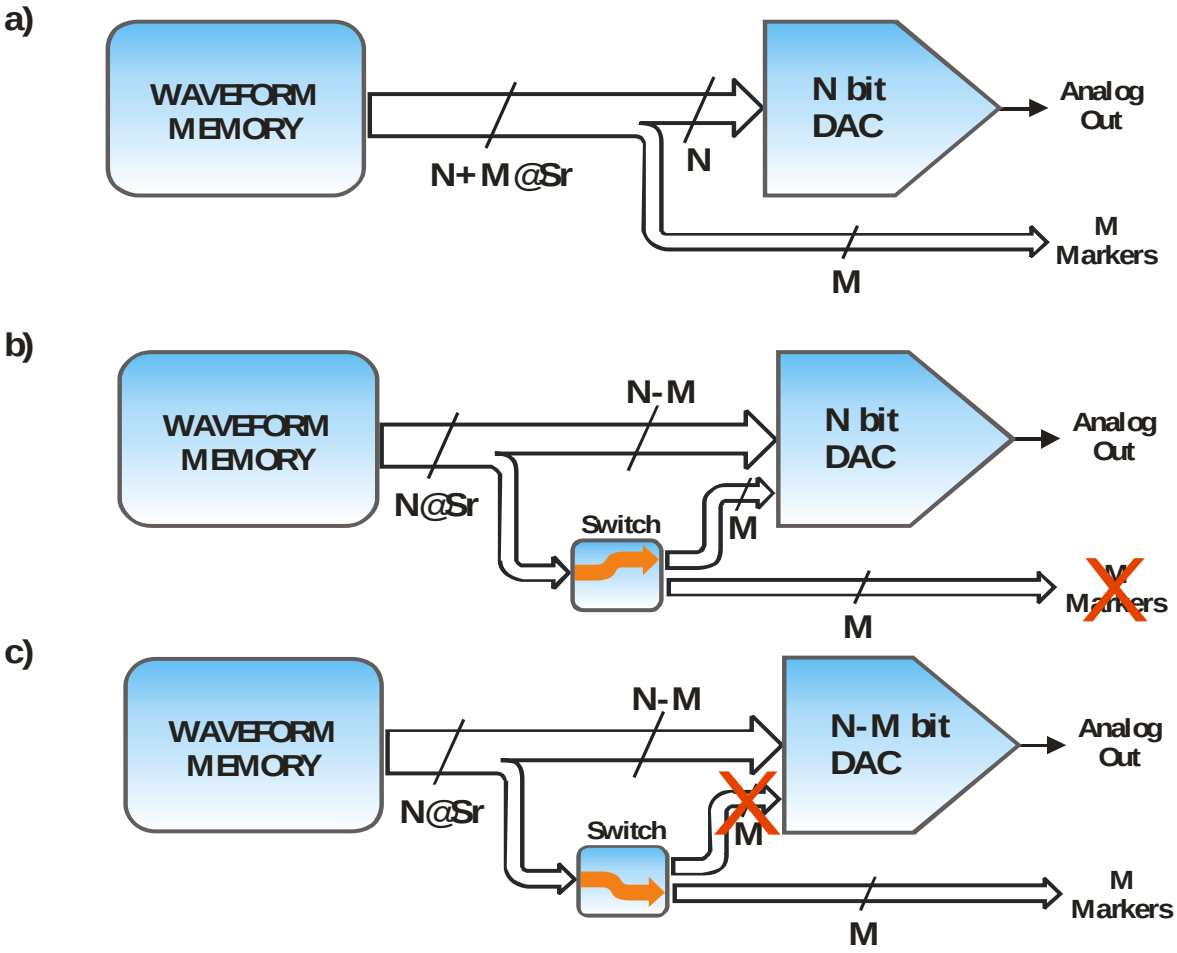


Figure 54: Markers are associated to waveform samples and their evolution depends on information stored in the waveform memory. Marker information should be accommodated by widening the waveform data transfer bus (a). However, some high-speed arbs cannot operate in this way as there is no remaining bandwidth available. In this case, a multiplexed approach may be implemented (b). Some LSBs can be shared with the markers if necessary but at the price of reducing the vertical resolution of the AWG when markers are activated (c).

Marker signals are specific to AWGs and they are somehow connected to the samples in the waveform memory. Users can mark specific moments in the waveform by setting markers accordingly. Markers are two-level, digital signals and users can set the low and high voltage levels according to their needs. There may be one or more marker outputs assigned to every analog channel or shared among them. Given the difference in signal path (including cabling) and bandwidth for the analog and marker outputs, it is highly desirable to be able to adjust delay either programmatically or from the instrument front panel. As an example, markers could be used to signal the position of symbols for a baseband QAM IQ signal pair, so an eye diagram may be obtained by using the marker as the trigger for an oscilloscope
(Figure 55), or they could signal the location of error bits in a binary sequence. Markers may greatly simplify some tests as they can be used to control any measurement device to obtain data at very specific moments in time so they are extremely useful in automated test equipment (ATE). Finally, they can be used as additional stimuli as they can generate serial binary signals or as supporting control signals (i.e. to switch off the RF output in a vector signal generator as seen if Figure 56). The most direct way to implement markers is by widening the waveform data memory to accommodate the extra bits required to control the marker signals. Typically, high-end arbs incorporate 2 markers per channel so two extra bits are needed (Figure 54a). This may be a problem for very high-speed arbs as the extra bits increase the overall transfer speed from the waveform memory. One way to solve this problem is sharing some of the LSB bits for each sample with the marker information (Figure 54b/c). This means that markers can only be activated at the price of reducing the DAC vertical resolution. Another way to reduce the marker data bandwidth is by limiting the speed of the transitions for these outputs and their positioning resolution. This is not a real limitation in most applications, as synchronization signals tend to be significantly slower than the sampling rate for the arb.

Markers are time-aligned with specific samples in a waveform while samples are not necessarily aligned with events in the waveform.

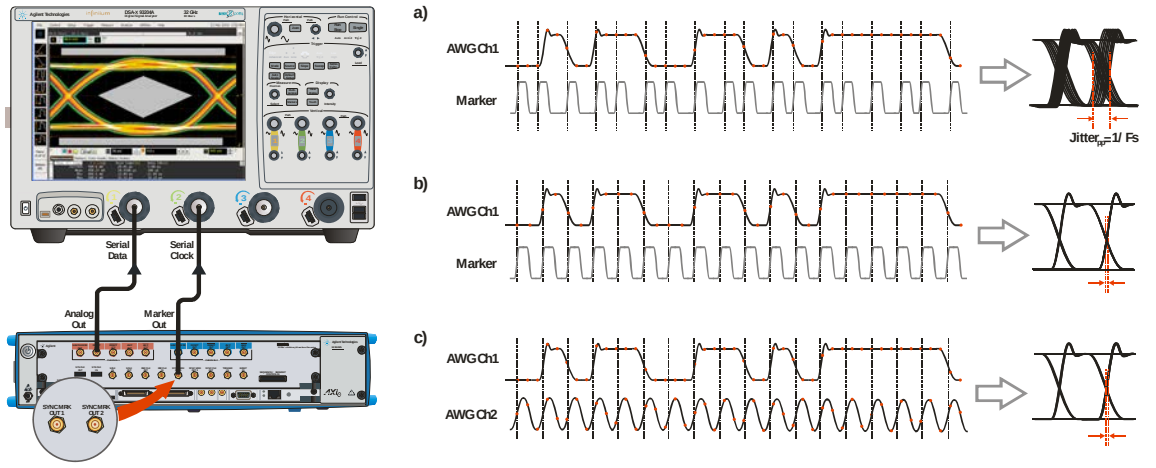


Figure 55: One of the most popular applications of markers is synchronization with other instruments and/or the DUT. In this example an AWG generating a high-speed serial data signal use the marker output to provide a trigger signal to a digital oscilloscope in order to build an eye diagram. Edges in the marker signal should point to the exact location of each symbol in the stream. Special care must be taken to avoid undesired jitter in synchronization signals as markers are not directly connected to signal features but to specific samples (a). Marker generated jitter may be removed if data rate is constant and sampling rate is a multiple of it (b). If this is not the case and a very low jitter synchronization signal is required, it is better to generate a clock signal with a second analog channel in the arb (c).

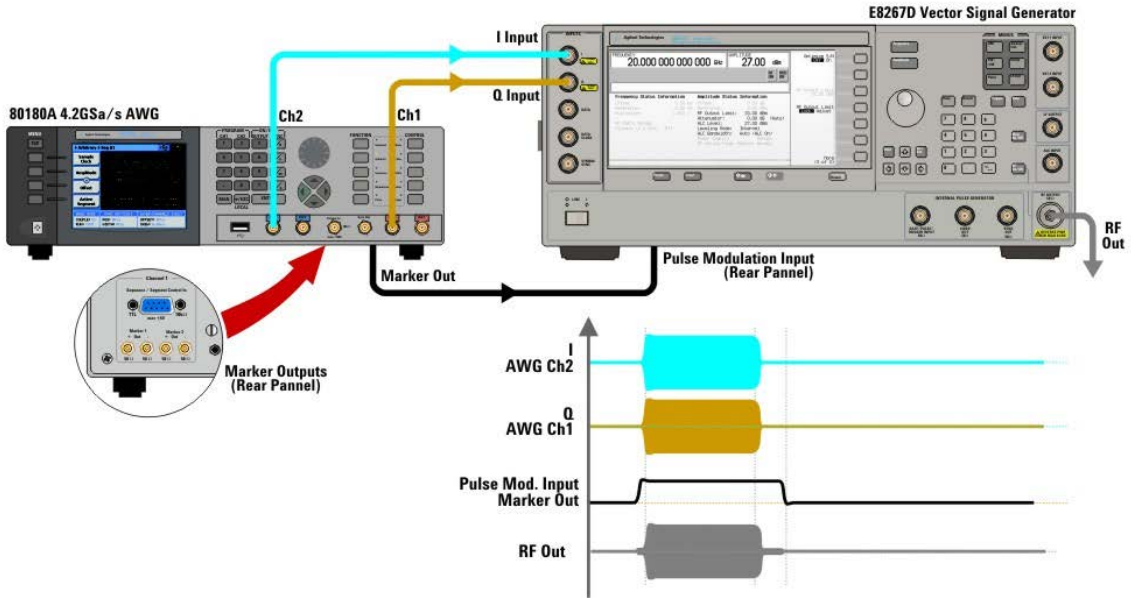


Figure 56: Controlling other instruments and/or the DUT is another application of markers. Here, one of the markers of an AWG is used to switch- off the RF output of a vector signal generator used to generate a pulsed RF signal while the same arb is generating the I and Q baseband signals through its two analog channels. Notice that even when the I and Q signals are not active, a remaining RF carrier is still present at the RF generator output. The Pulse modulation input controlled by the marker switches off the RF output improving the dynamic range of the RF signal.

This may result in undesired jitter between the synchronization signals transported by markers and the waveform being generated by the AWG (Figure 55a). Generally speaking, peak-to-peak jitter will be equal to the marker positioning granularity. In a typical test situation, Figure 55 shows a high-speed serial data signal generated by one of the channels of an AWG to build an eye diagram using a digital oscilloscope. A data clock is necessary to build an eye diagram. A possible test strategy consists in recovering the clock from the data stream. Clock recovery systems may be expensive and are not jitter-free anyway. Alternatively, one of the markers could be set up to transport the data clock so it can feed the scope trigger input to directly build the eye diagram. Figure 55a shows the test results when the AWG's sampling rate is not harmonically related to the data rate. As edges in the marker output are aligned with some specific samples and these do not occupy specific location in one bit time, the eye diagram will show an additional jitter. If sampling rate is a convenient multiple of the data rate, then this jitter component will go away (Figure 55b).

Unfortunately, it is not always possible to adjust sampling rate to be a multiple of the data rate. In

some cases, sampling rate may be not high enough; in some others data rate may be not constant throughout the complete waveform so there is no way to choose a sampling rate valid for the complete signal. In this particular case, the best way to obtain a low jitter clock signal is to use a second AWG synchronized channel where a clock signal (i.e. a sine wave) running at the data rate (or a submultiple) has been previously loaded (Figure 55c). As this synchronization signal will not have to be aligned with sampling instants and it can change over time according to the evolution of the data signal, jitter may be greatly reduced and limited to the channel-to-channel jitter component typically much lower than the sampling period.

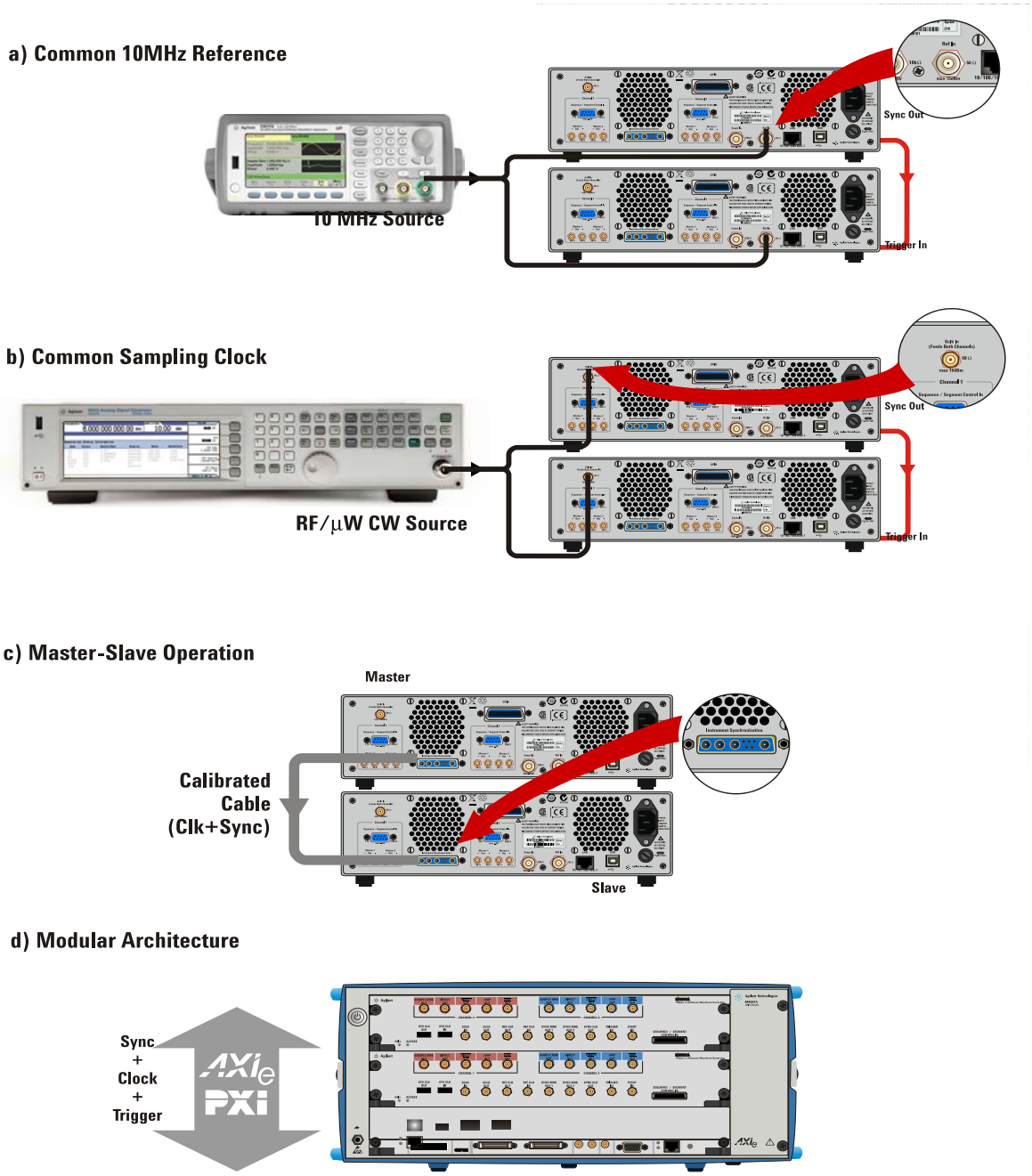


Figure 57: There are several methods to synchronize multiple AWGs to generate multi-channel signals beyond the capacity of one single instrument, generally limited to two analog channels. Using an external 10MHz reference (a) allows multiple AWGs to generate the same sampling frequency while the usage of trigger inputs can be used to align the waveforms coming out from all the generators. This method does not result in a repetitive sampling instant time alignment and it may be insufficient for wideband signals. A much better solution when it comes to sample time alignment is using a common external sampling clock (b). In this case, sample clock edges come from the same source and any differential delay may be adjusted after a simple calibration process. Some AWGs are designed from scratch to support multiple instrument synchronization, some, as the Keysight 81180B, incorporate specific hardware and software to implement the so-called Master-Slave Operation (c). Hardware includes a series of inputs and outputs to share the same sampling clock and initialization event as well as a calibrated cable that allows for a simple connection and predictable, repetitive time alignment between instruments. The same approach is possible using modular instruments where multiple AWG modules can be attached to the same standard bus (d).

3.7 Synchronization of Multiple AWGs

Many applications require more than one stimulus signal and those often must be synchronized. Many AWGs in the market incorporate two or even more analog channels. In some cases, these multiple channels are designed to operate together, so multiple settings must be common to all:

- Sampling clock
- Record length
- Sequence related setting: Waveform memory segment layout, sequences and scenarios, etc.
- Trigger and event signals effects.

Some instruments, such as the Keysight 81180B AWG, can alternatively operate both channels as virtually independent AWGs so each channel may be independently set-up if desired. For these instruments, even when sampling rate is set up differently for each channel, all are referred to the same internal reference so phase relationship is kept. As all the channels are internally controlled, timing specs (delay, channel-to-channel jitter) are specified and kept under reasonable boundaries. Typically, a fine channel skew control allows for precise time alignment between all the channels including the effect of cabling and any external device. When more channels are needed, the solution is to use more than one AWG and to synchronize them appropriately. There are several ways to hook together several AWGs to operate as a single instrument:

- Common frequency reference: Many time or frequency domain instruments incorporate an input for an external frequency reference, typically a 10 MHz signal. Through it, instruments can operate synchronously and/or with a better timing reference in terms of stability and phase noise than the internal one. Sharing the same reference among multiple identical AWGs may result in exactly the same sampling rate for all the instruments if it is set-up in the same way. When AWGs are of different make or model, it is possible that the same nominal sampling rate will result in different actual values depending on the way each instrument synthesizes it from the common reference. Even if differences are tiny they will accumulate over time. Even when all the instruments are identical, initial phase for each sampling clock will be random (and unpredictable) but constant during any given generation session.

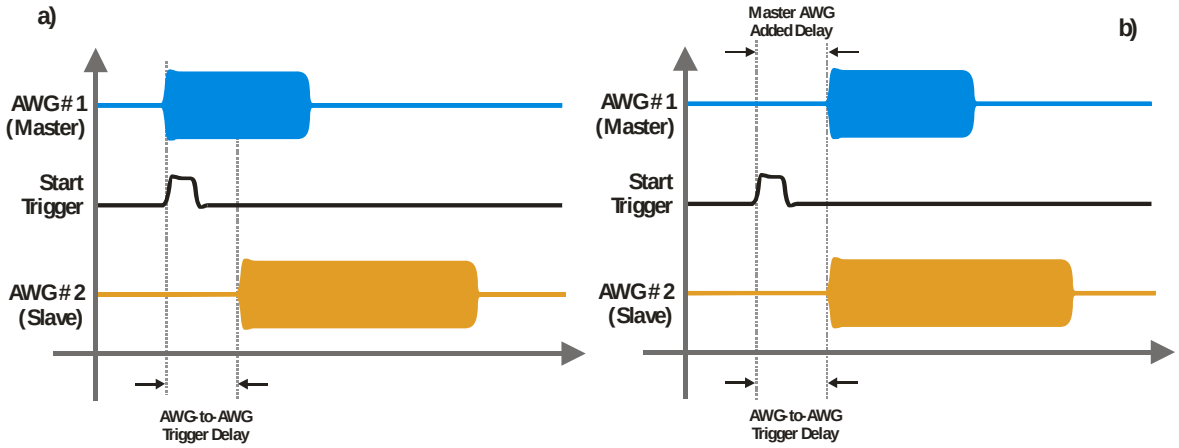


Figure 58: Beyond sampling clock synchronization, it is necessary to make sure waveforms from different AWGs are properly time-aligned. In any master-slave arrangement the reference generator will source the trigger event to start signal generation in all the slave arbs. In order to remove the effects of master-to-slave delay, it is necessary to delay signals coming from the Master generator. This can be accomplished through signal extension in the Master device (adding and idle section to the waveform or an idle segment to a sequence), usage of cables with appropriate differential propagation delays, or by adding some delay to the Master unit. AWGs specifically designed for Master-Slave operation such as the Keysight 81180B ease time-alignment between generators as they incorporate all the hardware and software functions required to do it quickly and consistently.

- Another issue is time-alignment of all the channels. One effective method is using the armed generation mode, where the AWG will start generating the signal after a trigger event is detected. Feeding the same trigger signal to all the instruments will align the waveforms within some time boundaries influenced by the trigger jitter for each instrument, cabling caused skews, and the previously mentioned sampling clock random phase issue. It is possible to use a synch or marker signal from one of the arbs to trigger the rest of them. In this way starting this instrument (master) will cause the others to start (slaves). There will be an additional delay between the master and the slave instruments that sometimes must be compensated to obtain a useful signal set. A direct way to do it it is to insert an idle group of samples at the beginning of the master's waveform to wait for the other instruments to start. Sequencing may be used instead if the signal must be looped in order to avoid the idle section showing up in its waveform occurrence. For short delays, the internal fine delay control can be an alternative.

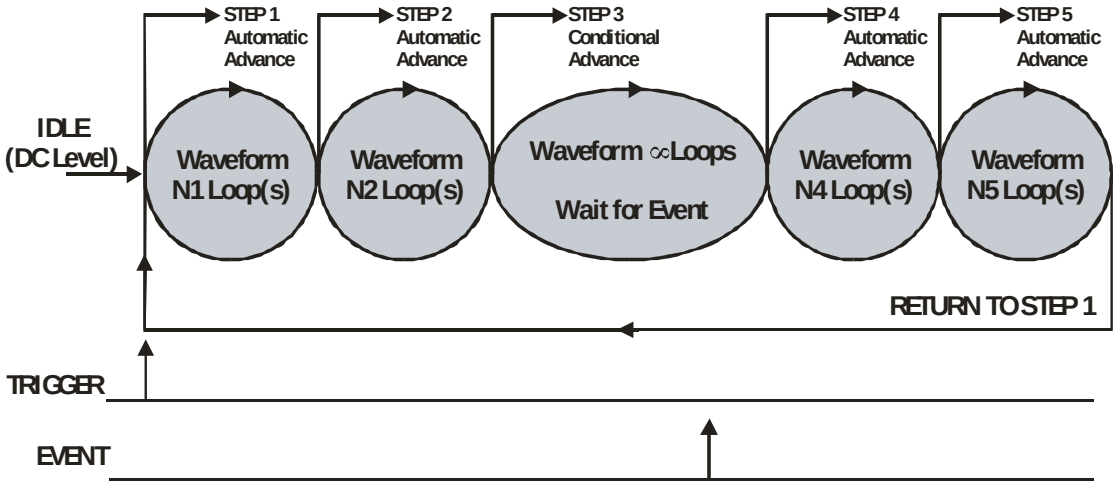


Figure 59: In armed mode, AWGs start generating after reception of a trigger signal. This mode can be used to synchronize multiple generators to start generating signals in the very same instant.

- Common sampling clock: Some arbs incorporate an external sampling clock input. Again, this feature can be used to improve the timing characteristics of the generator if a better quality external clock than the internal one is available. External sampling clocks can be also useful for non-uniform sampling rate signal generation (i.e. for jitter generation). When applied to multiple instruments, it will make sure that all the generators will share the same sampling clock and, as a consequence, that they will keep the same phase. Any phase difference can be removed after calibration through the internal fine delay control. Waveform time alignment can be handled in the same way than in the common frequency reference case. Typical clock signal sources for high-speed AWGs are RF and microwave CW generators.
- Integrated Master-Slave Operation: Some AWGs, such as the Keysight 81180B, have been designed to operate in an integrated master-slave mode. In this case, two instruments, one designated to be the “Master” and the other to be the “Slave”, can be operated as a single instrument. In order to accomplish that goal, specific hardware and software is used. The hardware side consists in a multiple signal connector to make sure the same sampling clock and starting instant is shared along with a calibrated cable so the delay between both instruments is bounded and repeatable.The software part makes possible to operate both instruments from the Master unit and apply a programmable delay (in samples) to the Master waveform to account for the unavoidable instrument-to-instrument delay.
- Modular architectures: Modular AWGs based on standard instrumentation buses such as PXI, AXIe, VXI, etc. can share clocks and trigger signals through the bus backplane so the above strategies are easily implemented in a controlled, repeatable environment with no or limited extra cabling.

4 Waveform Development for AWGs

4.1 Introduction

AWGs can generate virtually any signal if the right set of samples is stored in a big enough waveform memory and they are converted to analog with a sufficient sampling rate, vertical resolution, and bandwidth. However, signal samples must be calculated having the AWG requirements in mind. Otherwise signals will not match the expectations in terms of quality or applicability. AWGs characteristics such as maximum sampling rate, vertical resolution, and waveform memory size will influence some decisions about the way waveforms are designed. Waveform calculation and transfer times may also be an important factor, especially when complex waveforms requiring huge amount of memory must be created in an interactive test environment, where characteristics of new waveforms may depend on previous results or parameters from the device under test. In most situations signal quality and calculation time may be contradictory so some trade-off must be accepted.

4.2 Record Length and Sampling Rate Optimization

Any AWG incorporates a limited amount of waveform memory and it is not possible to seamlessly generate signals requiring more memory than the available in the instrument. Usage of segmented memory and sequencing also motivate users to develop signals as short as possible in terms of their time window and number of samples. Once the time window is minimized, the only way to further reduce the size of the waveform is by reducing the sampling rate as much as possible. Memory size (or record length, RL), sampling rate (Fs), and time window (TW) are connected by the following formulae:

$$TW=RL/Fs,\quad RL=TWxFs,\quad Fs=RL/TW$$

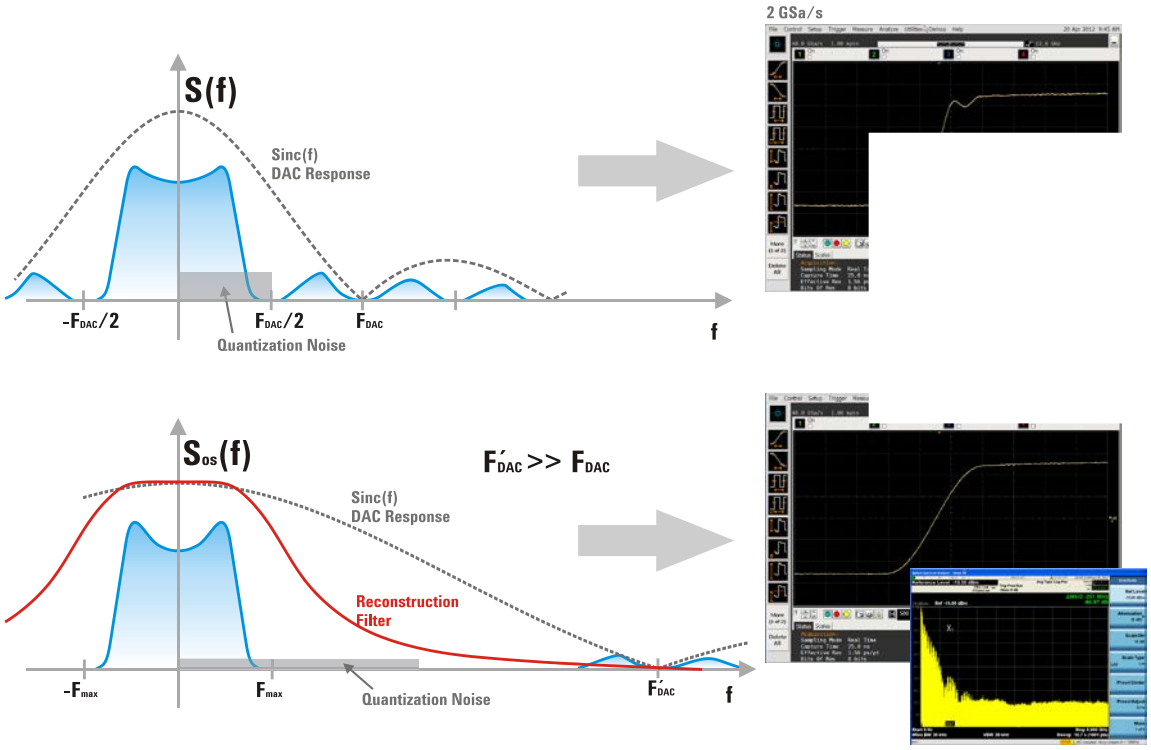


Figure 60: Keeping the sampling rate close to the Nyquist theorem requirements (above) saves memory although signal quality suffers due to the influence of the sinc(f) DAC response, the highest quantization noise spectral power density, and the poor image attenuation. A higher sampling rate (oversampling, below) results in a better looking signal as image rejection is higher and the sinc(f) response is flatter while the same quantization noise power is spread over a larger band and reconstruction filter can remove most of the images. Waveforms in the right show the same waveform generated at 2GSa/s (above) with no additional filtering and 8 GSa/s (bottom) and an appropriate reconstruction filter.

Reducing the sampling rate to a minimum, close to the Nyquist sampling theorem requirements, seems to be the first thing to do when optimizing memory usage. However, reducing the sampling rate leads to some signal quality issues:

- AWG's Sinc(f) response will impact more the signal as most of its energy will be located near the Nyquist frequency. Equalizing the signal through digital processing will consume more dynamic range from the DAC and, as a consequence, it will reduce the SFDR performance.

- More energy of the images will show up in the output signal as those will be closer to the wanted signal, where the reconstruction filters are not as effective. Boosting the high frequencies to equalize the wanted signal will also result in higher amplitude of the images' frequency components closer to the first Nyquist band.
- Quantization noise power density will grow as sampling rate is reduced compromising the SQNR performance.

Setting the sampling rate well over the strict signal requirements is an operation known as oversampling. It solves or attenuates the effects of most of the above issues. Oversampling can be equivalent to use a higher ENOB DAC at a lower sampling rate:

$$\Delta bits = 10\log_{10}(\text{Oversampling Factor})/6.02$$

Oversampling also eases the application of reconstruction filters with a gentle roll-off such as Bessel filters, resulting in a better looking, undistorted signal. Ideally, sampling rate should be selected taking these considerations into account. An alternative strategy is to choose always a high sampling rate (i.e. the maximum) and play with the record length to obtain the desired time window. In this way, oversampling factor will be always the maximum but memory needs will be maximized as well.

Record length granularity (RLG) impacts the ability to establish a precise time window for a given sampling rate. For single shot signals this is not an issue as record length may be extended to the nearest multiple of RLG greater than the original record length:

$$RL' = TW \times Fs$$

$$RL = \text{ceiling}(RL'/RLG) \times RLG$$

"Zero padding" (or any other convenient stuffing) should be applied to the last RL'-RL samples. When such a signal is looped for continuous generation (or used in a sequence) the extra samples will create a signal inconsistency: a gap between consecutive occurrences.

When the signal must be looped or sequenced, timing inconsistencies must be removed. The most obvious way to do it is by readjusting the sampling rate, so the target time window results in a record length which it is a multiple of the instrument's RLG. To do so it is necessary to first establish an approximate value for the final sampling rate (F's), then calculate the closest record length RL being a multiple of RLG that results in the closest time window to the one required by the signal, and then readjust the sampling rate (Fs) to obtain an accurate time window (TW) value:

RL = round((TW x F's)/RLG) x RLG

s = RL/TW

To obtain sampling rates as close to the maximum (MFs) as possible, record length should be

RL = floor((TW x MFs)/RLG) x RLG

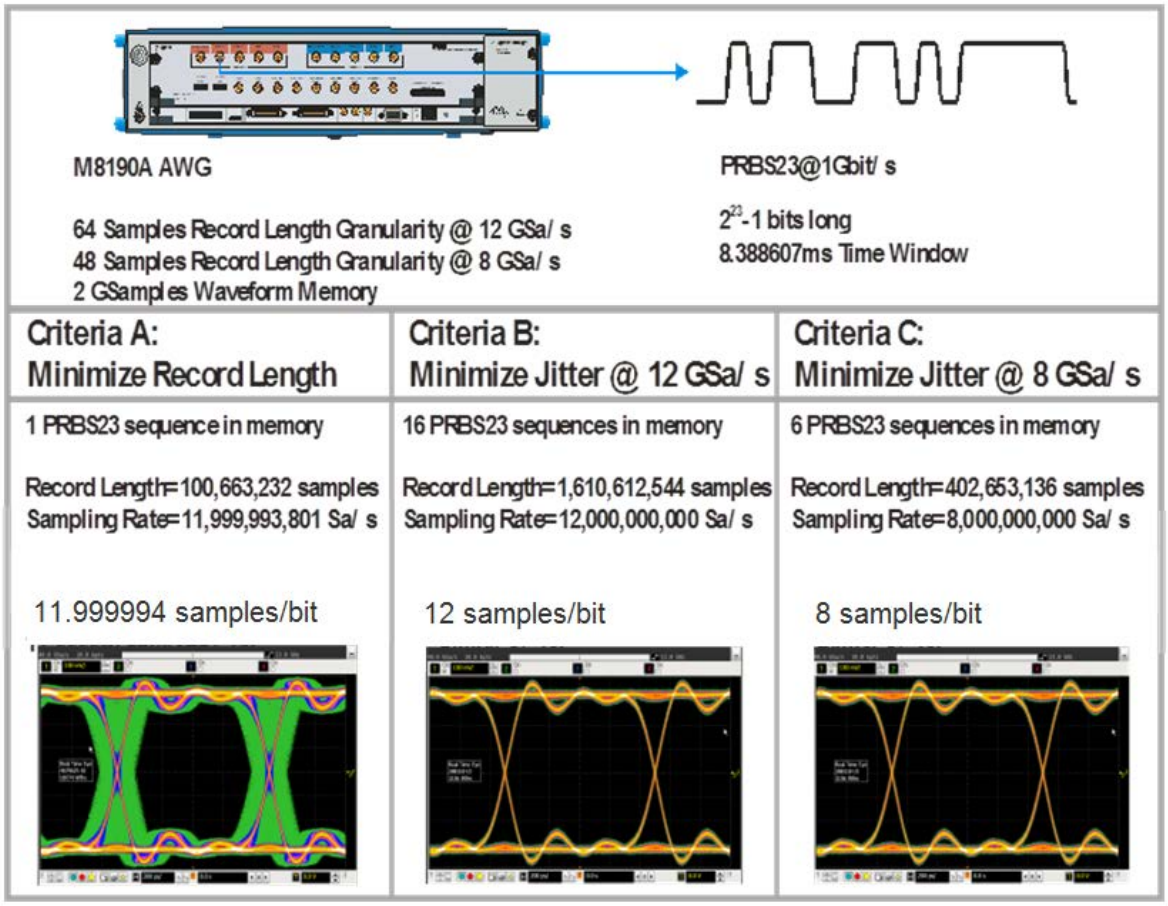


Figure 61: There may be several criteria to select the record length for a waveform. Record length granularity is an important constraint. Sometimes it may be better to repeat the signal several times instead of adjusting the sampling rate carefully to obtain the target time window for a valid record length. In this way, any convenient record length can be chosen and record length granularity requirement may be met by replicating the same samples an appropriate number of times.

$$Fs = RL/TW$$

The floor (integer part) operator in the above expression makes sure that sampling rate will be equal or lower than the absolute maximum. Time window accuracy then will depend on the sampling rate accuracy and resolution.

To minimize the record length for a given sampling rate may not always be the optimal criterion. When sampling rate must be set precisely (i.e. to be a multiple of some frequency component present in the signal or a bit rate) the resulting record length will be defined as always by the expression

$$RL' = TW \times Fs$$

If the resulting Record Length is not a multiple of the granularity parameter, then the waveform must be repeated an integer number of times until the total number of samples is a multiple of RLG. The new record length will be:

$$RL = LCM (RL', RLG), LCM = \text{Least Common Multiple}$$

The number of repetitions (NR) in the memory will then be:

$$NR = RL/RL'$$

In the example shown in Figure 61 another criterion can be seen. In this example, 1 Gb/s PRBS23 sequence must be generated. Minimum record length calculations for 12 GSa/s generators result in memory size larger than 100Msamples (criterion A). Given the constraints regarding record length (64 samples granularity at 12GSa/s, sampling rate must be slightly reduced in order to obtain 1Gb/s nominal data rate. Anyway, running the AWG at the maximum rate of 12 GSa/s gives a bit rate error of 0.5ppm, lower than the accuracy of internal sampling clocks in most arbs. Given the length of the binary sequence (223-1 bits), it is not possible to obtain an integer number of samples per bit time. It may be convenient an integer number of samples per bit time to simplify calculations, as sample positions are the same for all the edges in the waveform and the marker to bit alignment will be always perfect. Both factors will reduce jitter and calculation time while signal usability will improve. In this particular example, 12 samples per bit would maximize quality but the only way to meet all the requirements is by repeating the same PRBS23 sequence 16 times so the new time window will contain a number of samples multiple of 12 and 64 and an integer multiple of PRBS23 sequences. In fact the new optimum record length (1,610,612,544 samples) must be exactly the least common multiple (LCM) of, 12, 64, and 2^{23}-1 (criterion B). Sometimes waveform memory size is the limiting factor. The larger the waveform memory, the more freedom users will experience when developing waveforms and the better signal quality will be obtained. If the same exercise is performed in a 8GSa/s AWG with 48 samples record length granularity (criterion C), the same low jitter results can be obtained by repeating the PRBS23 sequence 6 times with exactly 8 samples per bit. Total record length is roughly 25% of that for 12 GSa/s.

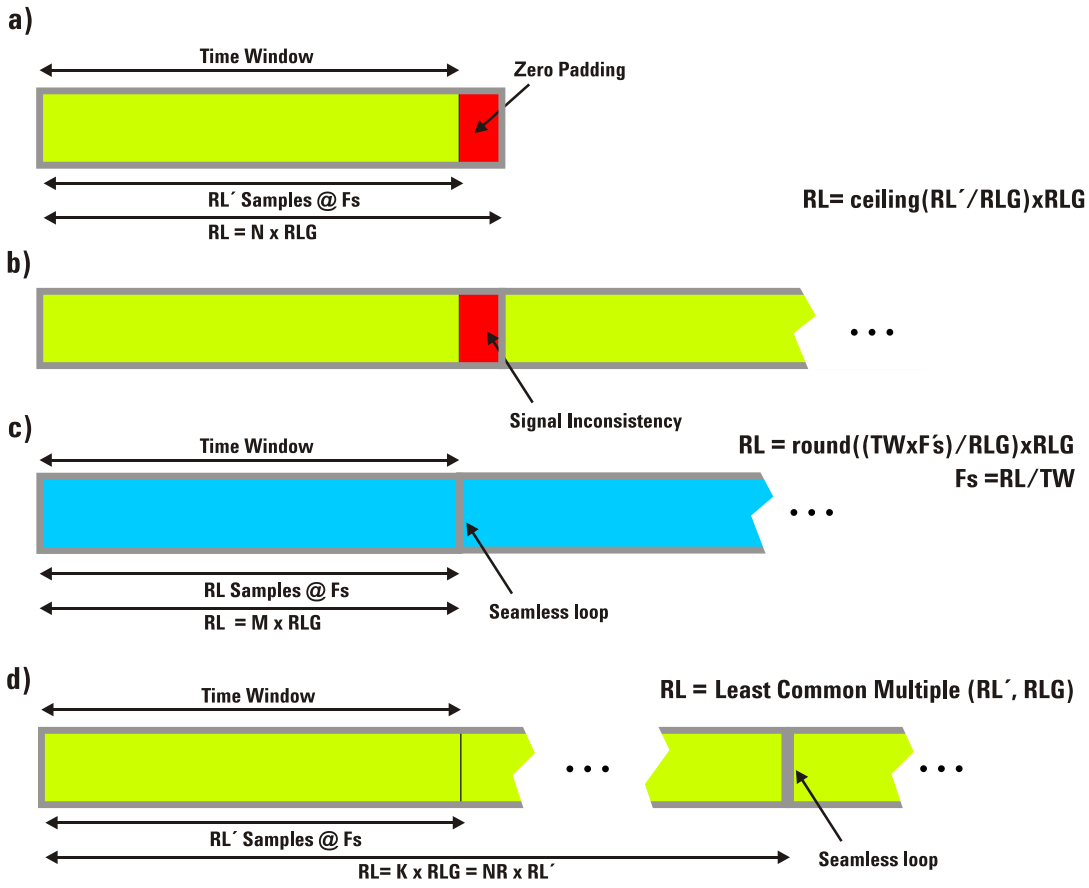


Figure 62: Time window selection typically depends on the application. Sampling rate selection depends on a series of factors such as signal bandwidth, image rejection, and signal quality. Both combined result in a required record length. As it must be a multiple of the record length granularity (RLG), most times it must be readjusted. In single-shot signals, extra samples may be filled with any desired DC level or padding waveform (a). However, if this signal must be looped for continuous generation, it will show inconsistencies at the end of every loop (b). A different approach consists in selecting the closest record length which is a multiple of RLG and adjust sampling rate to obtain the desired time window (c). Another solution is repeating the same waveform section several times until the total record length is a multiple of RLG (d). The above formulae can be used to calculate the optimum record length for each case.

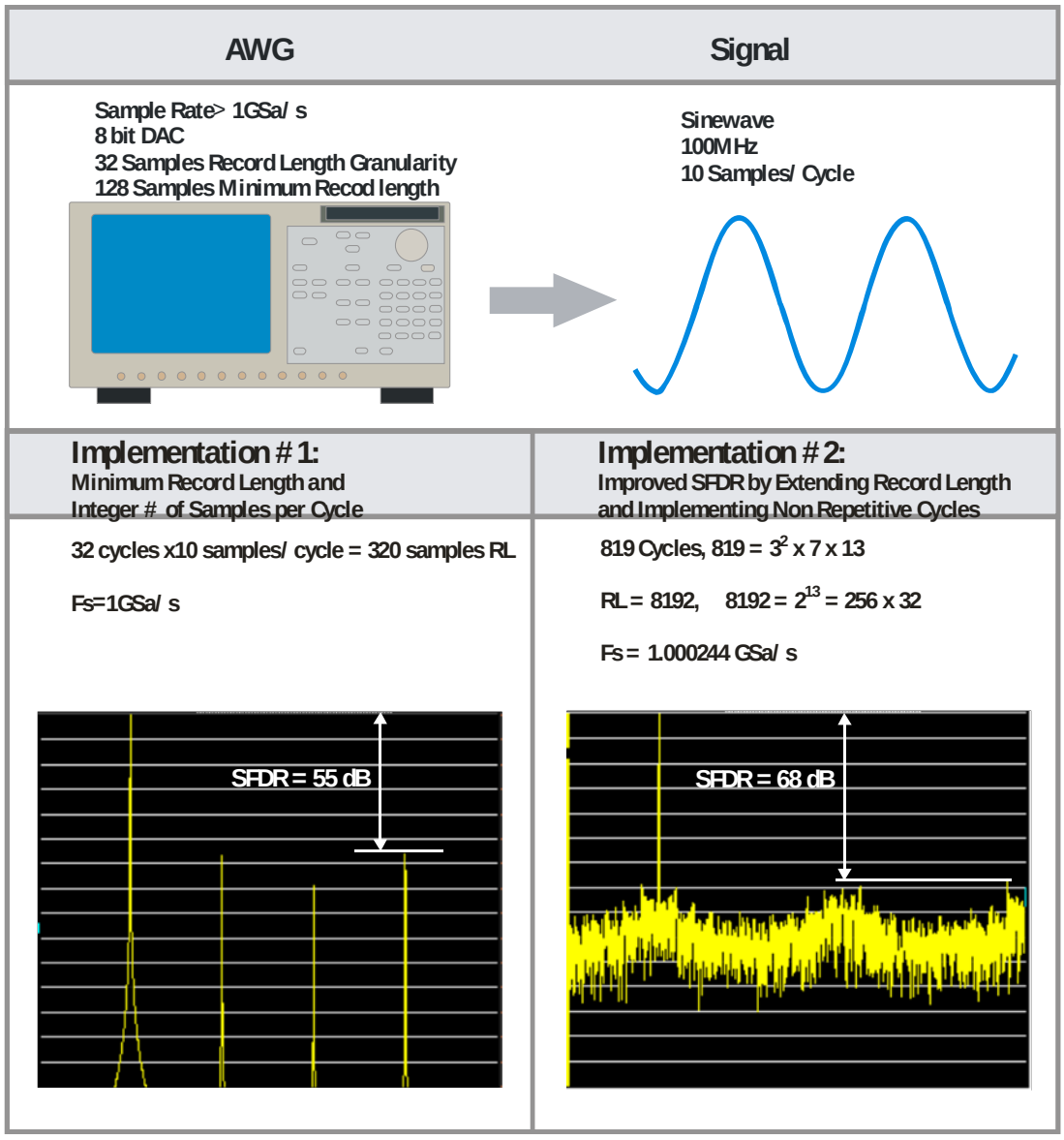


Figure 63: When an AWG generated signal repeats using exactly the same samples, quantization noise becomes periodical and it shows up at specific frequencies, multiples of the repetition rate. In this example the same sinewave is generated using the same AWG with around 10 samples/cycle. Implementation #1 repeats 32 times the same 10 samples in order to fulfill the minimum record length and record length granularity criteria simultaneously. The spectrum of the signal clearly shows some harmonics caused by the repeating quantization noise as its period is the same than the carrier being generated. In the second implementation, a longer waveform with 8192 samples (a multiple of 32) with 819 cycles in it destroys the periodicity of the quantization noise, which now spreads over a much larger number of harmonics of a much lower repetition frequency (around 819 times). The factorizations of the record length and the number of cycles do not show any common prime factor making sure that there are not repeated signals within the waveform memory. The resulting spectrum shows a much better SFDR.

Some test signals happen just once (“single shot”) while other must be generated continuously over long periods of time. Real-time sequencing apart, AWGs can only supply continuous signals by repeating the same segment, sequence, or scenario over and over again. Some applications require signals to start and/or end in a gentle, specific way (ramping up and ramping down sequences). Sequencing allows to do that by adding adequate portions at the beginning and at the end of the sequence list. The length of the repeating section must be chosen so it can contain a functionally consistent waveform. For example, to generate a continuous sinewave, there must be at least one cycle of it in the generator’s memory, while some serial data or wireless applications may require a minimum number of symbols to build a self-consistent, meaningful frame. Definition of “Self-consistent” and “meaningful” depends on the application and the type of tests being performed.

4.3 Signal Looping

One interesting outcome of repeating waveforms is that their spectrum is made of discrete spectral components (narrowband spectrum), harmonics of the signal repetition frequency (1/TW), instead of a continuous dense spectrum (wideband spectrum). Another consequence is that everything in the signal will be periodic, including quantization noise. This means that quantization noise will show up as discrete lines in the spectrum as well. The total quantization noise power will be the same than in a non-repeating, everlasting signal, but now noise will concentrate in specific in-band harmonics and these can impact the SFDR performance. The average power for each harmonic depends on the total number of in-band harmonics, so the more harmonics the better SFDR. The number of harmonics grows with sampling rate for a given time window and with record length (if the signal is not repeated exactly the same within the waveform memory) for a given sampling rate:

$$N_{Harmonics} = (Fs/2)/(1/TW) = Fs \times TW/2$$

$$N_{Harmonics} = (Fs/2)/(Fs/RL) = RL/2$$

This means that increasing the sampling rate and/or the time window will improve signal quality and minimize the effects of quantization noise in the frequency domain. Both actions will result in an increment of the record length required to build the waveform.

For periodic signals it is better to build a waveform made of multiple repetitions of the basic waveform, as quantization noise will be spread over a large number of harmonics. Additionally, minimum record length considerations make sometimes storing multiple cycles of the same basic signal mandatory. However, as stated above, it is important that periodic signals do not repeat exactly in the same way within the generators memory. When this happens is just like using a shorter record length. For example, a 10,000 samples waveform memory containing exactly 10 sinewave cycles will be equivalent to a shorter 1000 samples memory containing just one cycle as the same exact cycle will repeat 10 times within the complete memory. Signal repetitions happen when the same sampling instants are applied to successive occurrences of the basic waveform or some multiple of it. To make sure this does not happen, greatest common divisor (GCD) of record length and number of repetitions of the basic signal must be 1:

GCD (RL, Number of Repetitions) = 1

Basic arithmetic shows that this can be accomplished if prime factorisations for both parameters do not share any factor. Record length granularity makes this condition more difficult to meet as usually the granularity parameter is a power of two or a simple product of low prime numbers including 2 (i.e. $64=2^6$, $48=3x2^4$). Playing with prime numbers usually leads to a quick solution.

A possible strategy may be to select an approximate sampling rate and a time window that will result in the desired number of harmonics for quantization noise. Sampling rate must be greater than twice the signal bandwidth (Nyquist sampling theorem) and the time window should be much longer than one period of the basic signal. Using this data, the nearest valid multiple of RLG can be obtained. Time window can be used to obtain a preliminary value for the number of repetitions. Once obtained the prime factorisation of record length, number of repetitions (in practise it must be always an odd number as RLG is always a multiple of 2) must be adjusted so its factorisation does not include any of the prime factors found for record length. Once the final number of repetitions is obtained, then the final sampling rate can be calculated to precisely match the required time window. Figure 63 shows how this strategy may be used to generate a better quality sinewave. The effects of quantization noise periodicity greatly depend on vertical resolution. While they may be noticeable for a 6 or 8 bit vertical resolution AWGs, they can be negligible for >12 bit vertical resolution AWGs.

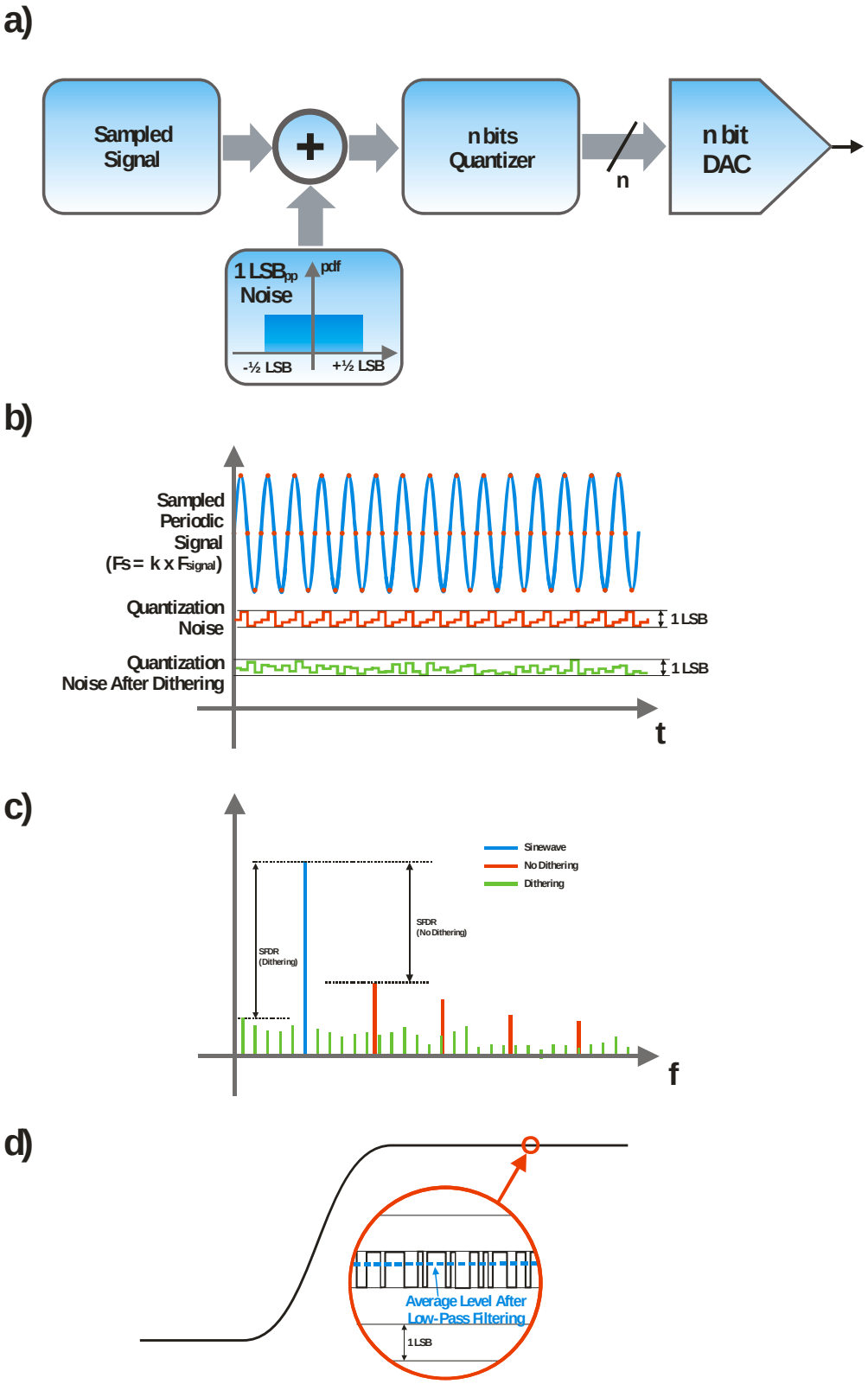


Figure 64: Dithering consist in adding a low level noise (1 LSBpp) to the waveform samples before quantization (a). In this way, any periodicity in the quantization noise will be removed (b) so the quantization noise will be spread out over a much larger number of

harmonics, thus improving the SFDR performance (c). Additionally, dithering can increase the DAC resolution for sections with constant amplitude (d).

4.4 Dithering

Sometimes the exact repetition of the same waveform portion cannot be avoided. For example, waveform expansion required by record length granularity for continuous playback (Figure 62d) results in the repetition of multiple copies of exactly the same samples. In some other cases, signals may improve if samples are always located in the same relative position of the repeating cycle. When not enough samples can be placed in the edges of the high frequency square signals (i.e. for a clock signal or a serial data stream), to precisely control their timing, they are synthesized with a lower edge jitter if one cycle requires an integer number of samples. Additionally, repeating exactly the same waveform may speed-up signal calculations. In these cases the number of harmonics will depend on the number of samples in a cycle and not on the total record length. As the problem comes from the repeating quantization noise pattern cycle after cycle, a simple solution consists in adding a small non-repeating, random noise over the whole record length (Figure 64). Although this additional noise will reduce slightly the overall SNR parameter, quantization noise will not be periodic anymore within the complete waveform. As a result, the noise will be more evenly distributed over the whole Nyquist band and the high amplitude noise harmonics will disappear, improving then the SFDR performance (Figure 64c). This noise addition operation is known as dithering and it has been often used in DAC and ADC signal processing and image enhancement.

Noise should be as low as possible to obtain the expected results (removal of quantization noise harmonics) without significantly degrading the signal to noise ratio. Ideally, noise should be white, so power is evenly distributed over the whole Nyquist band. AWGN (Additive White Gaussian Noise) seems a good candidate but its high peak to average power ratio (PAPR) will reduce the available dynamic range for the useful signal, as the overall peak-to-peak amplitude will increase accordingly. It is better to use a uniform distribution noise (as the distribution of numbers returned by the C function *rnd*) so its peak-to-peak amplitude is ±½ LSB. This amplitude will make sure that the quantization noise periodicity will be removed at the same time noise power is minimized. The peak-to-peak amplitude of the overall signal will increase by 1 LSB, so especial care must be taken in order to avoid signal clipping.

Dithering can also improve the effective DAC resolution in sections of a waveform with low frequency contents (i.e. a DC level) if it is applied to the waveform before being quantized to the DAC resolution (Figure 64d). The reason for this is that dithering will randomly assign two consecutive quantization levels to consecutive identical samples. Statistically the ratio of samples assigned to each level will depend on the actual intermediate level of the original signal. Consequently, the average level for that section will be an intermediate value closer to the original value. This effect can make dithering useful in some situations where the signal does not repeat within the waveform memory.

4.5 Wrap-Around Artifacts

When waveform segments are looped, the first sample follows the last sample at the end of each loop iteration. Wrap-around artifacts are caused by the discontinuity between both ends of a given waveform being repeated in a loop or, in a more general case, between the end and the beginning of two consecutive waveform segments in a sequence (Figure 65).

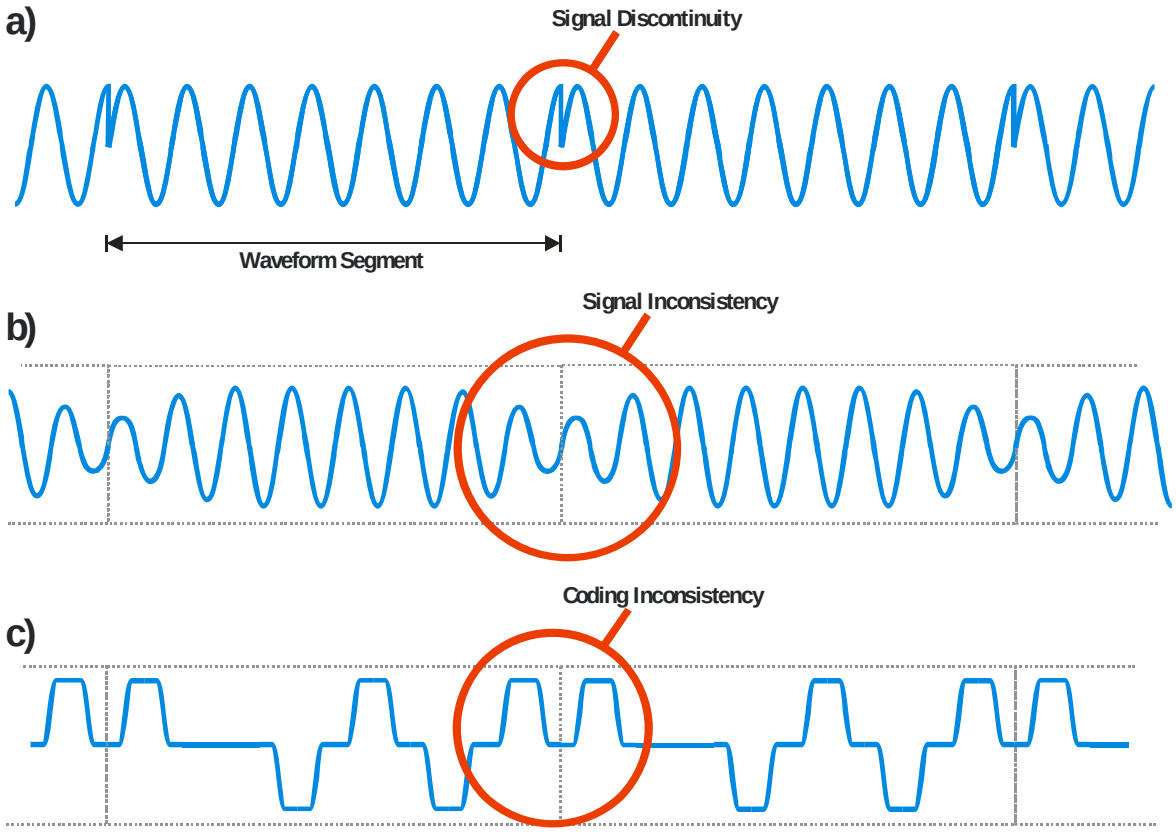


Figure 65: Looping a waveform segment or linking different segments in a sequence may result in periodic signal discontinuities and/or inconsistencies. Those can be caused by incomplete symbols or cycles (a), border effects of signal filtering (b, a band-pass filter is applied to a carrier segment after applying zero padding to the isolated segment), or illegal coding sequences (c, a AMI coded serial data signal shows two consecutive marks with the same polarity)

It is important to understand that AWGs can link signal segments without any gap or glitch related to the sampling period as samples are converted seamlessly regardless of their location. The effect of wrap-around artefacts may be negligible in many time-domain tests as users can restrict their measurements made on the device under test (DUT) to areas away from the transition between signal segments. The usage of synchronization signals (i.e. markers) provided by the AWGs makes this approach easy to implement in most test situations.

Even if measurements are performed at the right moments in time (i.e. using properly located markers as a trigger), the inter-segment transients can affect the behaviour of the DUT. For example, in a serial-data transmission test, the clock recovery circuits in a DUT can lose lock if a transient results in truncated symbols or illegal line coding. Depending on the length of the data and the time taken by the DUT to lock the signal clock again, test results will not be valid anymore as they will not reflect the DUT steady-state behaviour. Wireless test can be even more demanding as they occur in many domains consecutively or simultaneously: time, frequency, modulation, channel coding, and protocol. In many wireless test situations, measurements cannot be easily restricted to signal areas away from the wrap-around transients.

Discontinuities may happen at many different layers in a signal. Signals must be designed to avoid all the discontinuities at all levels that are relevant to the test to be performed. Physical layer is the most evident source of discontinuities, as they can be directly observed. There are several sources of problems that must be handled separately:

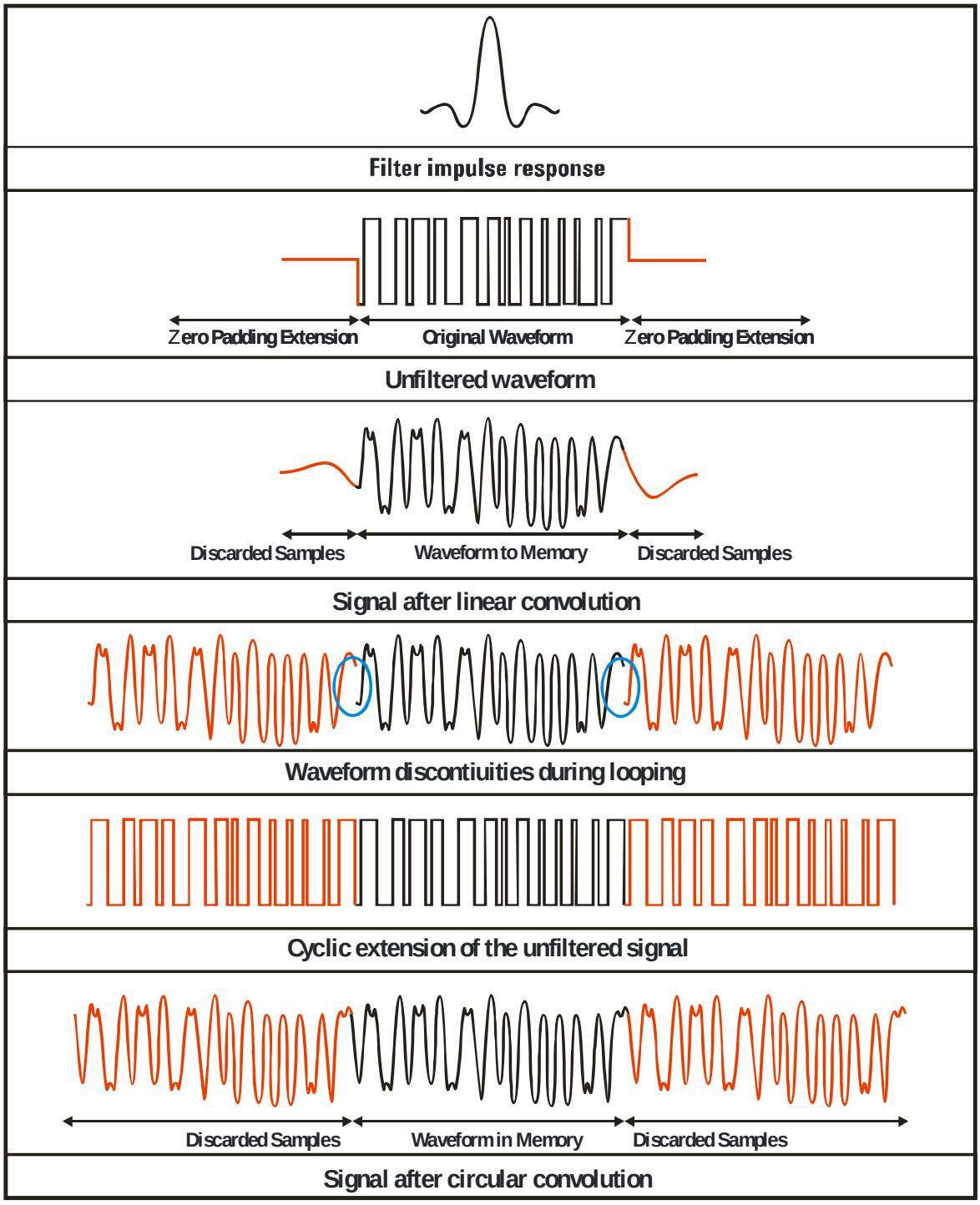


Figure 66: Circular convolution solves the wrap-around problems caused by inconsistent filtering. Isolated segments can only be filtered after adding zero-padding sections to each segment end. Removing the extra samples and looping that segment will result in signal discontinuities. Circular convolution extends the segment by repeating it and applying the filter to the extended signal. As the filtered signal is repetitive and consistent, just looping a cycle of it will result in a seamless, self-consistent signal.

- When signals are made by a series of symbols or cycles, there will be a fractional symbol or cycle at the end of every loop if a non-integer number of them are stored in the segment.
- If a filter is applied to a waveform segment there will be some border effects caused by applying the filter to a non-continuous signal. Even if a longer signal segment is truncated to avoid those effects, there will be a discontinuity, as convolution will not be, in general, applied to the same waveforms sections.
- Channel coding (i.e. line coding) is also an issue as a fully compliant waveform may not be compliant at all when looped.

Solutions to the above problems are sometimes quite straightforward and sometimes quite difficult. First of all, any looping waveform segment must accommodate an integer number of symbols or cycles for all its signal components. This condition may influence the choice for record length and/or it may limit the choices for some of the signal parameters. As an example, any multi-carrier signal has to be made of carriers located at multiples of the segment repetition rate (1/TW= Fs/RL). In this particular case, carrier spacing gives a minimum viable time window while the maximum frequency component requires a minimum sampling rate. Both requirements combined result in a minimum acceptable record length.

Signal filtering continuity may be obtained by applying circular convolution. This operation consists in filtering the looped signal instead of an isolated segment of it (Figure 66). Once the filter is applied, a segment with the original time window and number of samples can be extracted, stored in the waveform memory, and seamlessly looped. As the looped signal will be exactly the same than the original, a perfectly continuous signal will result. Circular convolution with an N-taps FIR filter is as simple as appending the N/2 samples from the end at the beginning of the waveform and the N/2 samples from the beginning at the end and apply the filter. The signal to be looped will be obtained after discarding the first and the last N/2 samples. If the number of taps is larger than the length of the unfiltered signal, then multiple copies of it must be appended in both ends until the appended section length is equal or larger than N/2. IIR (Infinite Impulse Response) filters require a more complex approach as it may require cycling the unfiltered signal several times until the filtered signal converges into its steady state. Convergence criteria may be based on comparing the samples of consecutive occurrences of the filtered signal until the worst-case difference is lower than 1 LSB.

For long enough signals and typical filter time-domain impulse responses, cycling the signal twice may be enough. Another practical approach is converting the IIR filter to a FIR one by truncating its impulse response when all its later samples are lower than a given small value (typically it should be much lower than 1 LSB relative to the impulse response' peak-to-peak amplitude).

For sequences, the circular convolution method remains valid except for transitions from one segment to the next. To solve this issue, it is necessary to add "link segments" obtained by merging two unfiltered consecutive segments, extending the signal asymmetrically and then applying the filter (Figure 67). The sequencing list must be modified to accommodate the linking segments without altering the overall signal timing. This approach also means that different linking segments must be calculated for all possible inter-segment transitions.

Solutions to channel coding or line coding wrap-around problems greatly depend on the specific situation. Most times, a careful selection of the symbol sequence may be enough to obtain a consistent signal. Difficulty to find a consistent sequence mainly depends on the memory of the coder. For example, a consistent AMI (Alternate Mark Inversion, a very simple coding scheme) coded data stream may have the first and the last '1's coded with the same polarity. When looped, two consecutive '1's will have the same polarity what it is illegal in that coding scheme (Figure 65c). This happens because there is an odd number of ''1's in the complete stream. Just using a stream with an even number of '1's will solve the problem. Some more complex codes may be more difficult to fix. A possible general solution consists in creating a sequence of coded symbols (typically a pseudo-random sequence) of length M, much longer than the number of symbols N for the target waveform. Applying a "sliding window" of length N on that sequence until an N-symbols legal sequence is found will generate a consistent set of symbols that can be looped seamlessly. Another viable solution may be truncating the coded sequence at any point and then tweak the initial and/or the final symbols to artificially obtain a consistent sequence.

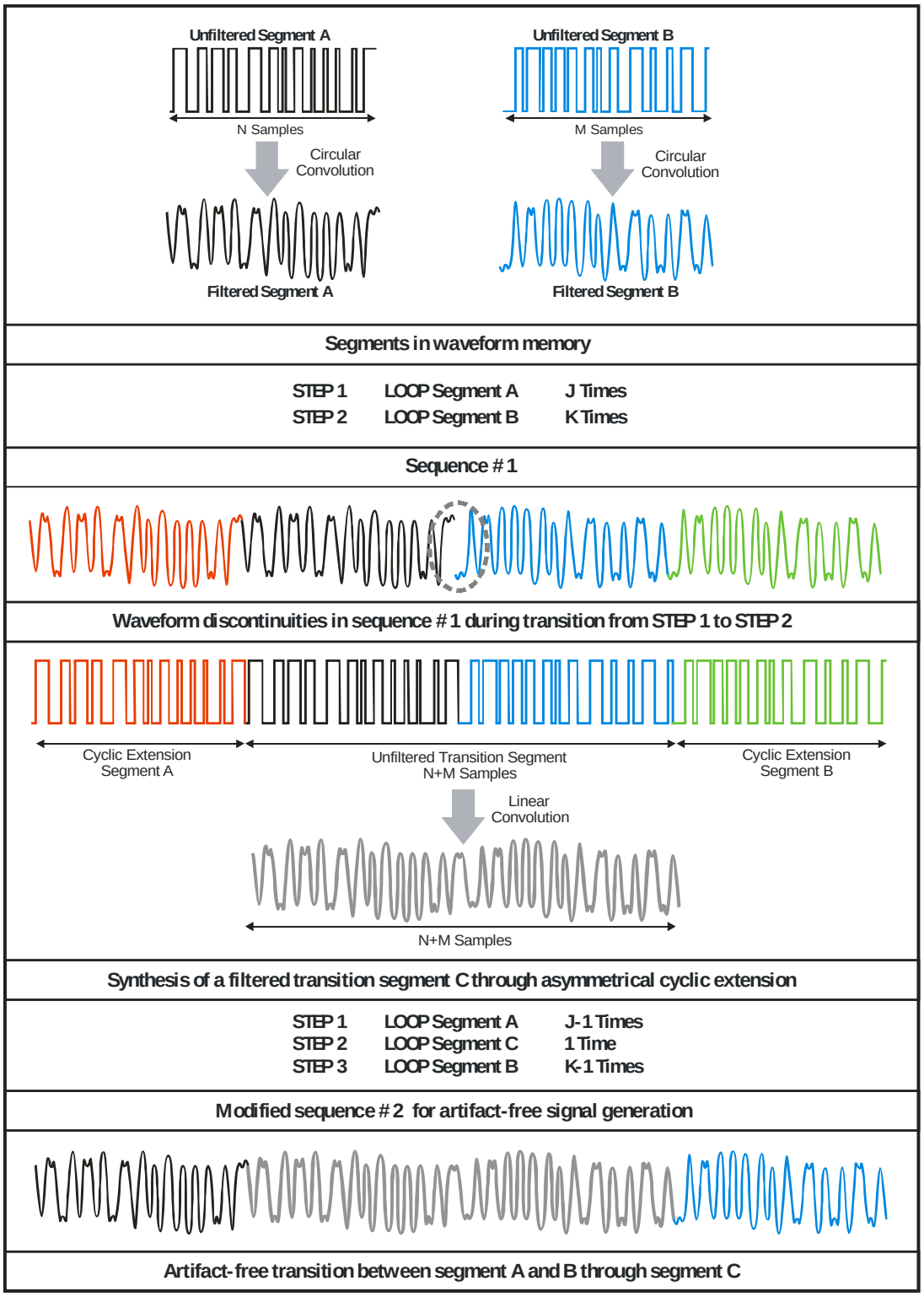


Figure 67: Circular convolution is not enough when different segments must be sequenced, as inconsistencies will show up when jumping from one segment to the next. Additional linking segments must be synthesized to solve this problem. This new segments can be calculated by merging two unfiltered segments and applying cyclic extension asymmetrically. Once the filter is applied, the linking segment may be extracted and added to the sequencing list.

4.6 Importing Signals from Other Instruments

AWGs can be used to playback real-world signals once they have been captured using some analysis instrumentation such as a DSO (Digital Storage Oscilloscope), a VSA (Vector Signal Analyzer), or even a Logic Analyzer. Being able to replicate real signals and conditions is very useful to troubleshoot problems difficult or impossible to synthesize realistically through a purely mathematical description. Those instruments typically supply waveforms uniformly sampled with their native vertical resolution. DSOs are the most popular source of waveforms and, given the similarity between their block diagram and that of AWGs, it is quite straightforward to adapt them.

Modern state-of-the-art high-speed DSOs can typically sample signals at higher sampling rates than the fastest AWGs although their vertical resolution (8 bits) is limited compared to that of the AWGs (>10 bits). Sample rate and record length flexibility use to be much lower than those in AWGs as they can be set only to a limited set of values. In the past, maximum record lengths for high-performance scopes were bigger than that of the AWGs but the latest generation of instruments, such as the Keysight M8190A (2 GSamples record length), has changed the panorama. It is important to maximize the quality of the captured signal in order to obtain optimal playback results. Often, signals must be processed to further improve their quality or just to make them playable by a specific generator.

The most direct way to capture a signal to be played back is setting up the same sample rate that will be applied in the target AWG. In order to maximize the signal to noise ratio, it is important to adjust the scope's vertical controls (volts/div and offset) so most of the ADC range is used to sample the signal. It is also important to make sure that the bandwidth of the signal is limited so the Nyquist sampling theorem is met. Otherwise, aliasing present in the captured signal will be also part of the generated signal. Ultra-high speed oscilloscopes (Sample rate >20GSa/s) show a quite flat frequency response, far away from the more traditional gaussian-like responses implemented in lower bandwidth instruments.

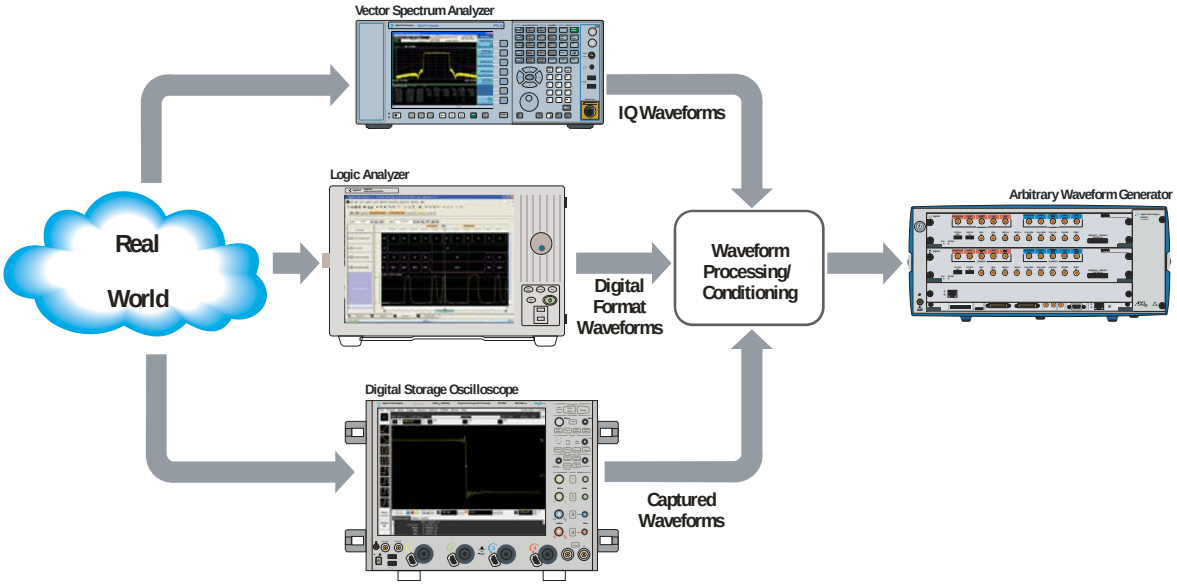


Figure 68: Many instruments can capture waveforms from analog or digital sources. AWGs can play-back those signals properly processed. Sometimes, real-world signals may be the best way to emulate unique test conditions. Additionally, AWGs can be also used in block substitution during system development.

High-enough sampling rates (2.5 times the 3dB bandwidth) combined with steep roll-off characteristics results in captures with low signal distortion and free of aliasing. Sampling at the highest available speed may be, then, highly advisable if the available record length in the scope allows the capture of the intended time window. Once the signal has been captured, applying a low-pass or band-pass filter may improve it, as any unwanted signal or noise will be removed and some additional bits of vertical resolution will be gained. Filter characteristics are influenced by the intended sampling rate for the AWG, as the high cut-off frequency must be set to avoid any aliasing component. A good practice is selecting the AWG sample rate to be an integer fraction of the scope's, or the other way around. In that way, the resampling process is reduced to a simple decimation.

Sometimes, though, record length granularity, sampling rate resolution and settability, and time window accuracy result in a non-integer sampling rate ratio. In that case, new samples must be obtained by interpolation of the captured ones after applying a filter. Resampling through sample interpolation consists basically in calculating new samples located somewhere in the middle of two actual filtered samples. An ideal interpolator consists in applying the sinc interpolator.

Any given sample will be made by the weighted sum of the nearby samples after applying the sinc impulse response. Ideally, calculation of any sample would require an elevate number of terms. Fast calculations may require limiting the number of samples involved what it is equivalent to truncate the sinc impulse response. There is then a trade-off between the interpolator accuracy (and its frequency response) and the calculation time for each interpolated sample. A simpler and faster method may consist in interpolating new samples just using the two nearby actual samples. The simplest method to do so is by using a linear interpolator. In order to accelerate calculations, interpolating functions may be stored in a high-resolution back-up table.

Often, vertical resolution in AWGs is greater than in DSOs so it may be interesting to increase this through some additional signal processing. As described previously, oversampling results in an increased vertical resolution, even for portions of the signal with a constant voltage level as analog noise acts as dithering. Many modern oscilloscopes provide some additional methods to increase vertical resolution:

- Waveform Averaging: Repetitive signals may be averaged by accumulating multiple acquisitions. Averaging also removes any non-repetitive component from the signal such as noise. Most oscilloscopes implement exponential averaging where each new averaged waveform is obtained from the previous one and the new captured waveform through the following formula:

  $$Avg(m) = (N-1)/N \times Avg(m-1) + 1/N \times Acq(m)$$

  Although the practical effects of linear and exponential averaging are similar, statistically speaking, exponential averaging requires twice as many averages to obtain the same results in term of signal-to-noise-ratio or vertical resolution improvements. Waveform averaging requires the signal and the oscilloscope to be very stable over multiple acquisitions. Any timing change will affect the final result and its influence will grow with the distance to the trigger sample what makes it difficult to use with very long acquisitions.

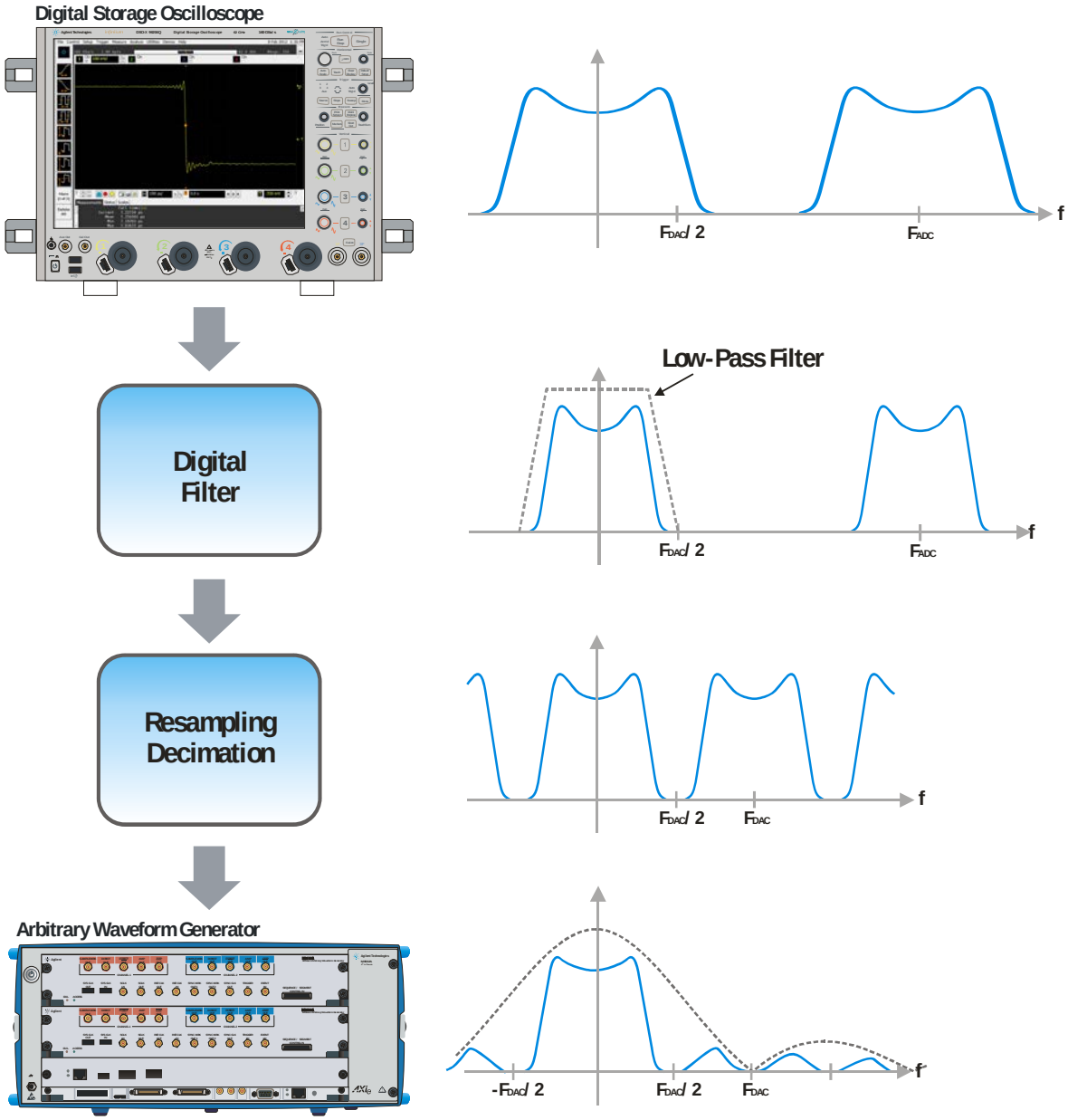


Figure 69: Signals captured with an oscilloscope must be properly processed for AWG usage. Basically, the signal must be bandwidth limited and resampled so it can be generated by an AWG at a different sampling rate. The resampling process may result in an increased vertical resolution when the oscilloscope sampling rate is greater than that of the target AWG.

- Signal Smoothing: This process replaces each captured sample by the average of N samples around it (Boxcar Filter). It is equivalent to a zeroth-order hold low-pass filter. The advantage of smoothing in front of other filters is its simplicity and processing speed. Theoretical improvement of vertical resolution (Δn) can be calculated through the following expression:

  $$\Delta n = ½ \log_2(N)$$

- High Resolution acquisition mode: Some DSOs can apply a boxcar filter in real-time. Samples are averaged in real-time and the result is stored into the acquisition memory. The number of averaged samples (N) depends on the ratio between the maximum sampling rate and the one set-up for a given time base and record length (decimation factor). The difference with signal smoothing is that the “hi resolution” mode does not require capturing all the samples before applying the filter so, for a given record length, the captured time window is much longer as the signal is automatically decimated. It also saves processing time, as there is no additional processing required after acquisition.

5 High-Speed Serial Data Signal Generation Using AWGs

5.1 Introduction

Historically, high-speed serial data has dominated the electrical and optical telecom landscape. The meaning of “high-speed” has evolved from 1.5Mb/s up to 40Gb/s and more of today’s optical networks. Most short distance digital connections found in computers, backplanes, and PCBs transport high-speed serial data that share some common characteristics:

- The clock is transported by the signal and it is reconstructed at the receiver through a clock recovery system. The ability of the signal to successfully carry information critically depends on the timing characteristics, specifically jitter, of the signal and the clock recovery system performance.
- The signal is noticeable distorted by the transmission path. In some cases, the connection length is the main problem (i.e. optical fibre long distance links) as attenuation, dispersion, and noise affect the signal so it must be regenerated before it becomes unusable. In PCBs and backplanes, bandwidth, reflections, electrical noise, and crosstalk have similar effects. Differential signalling is often used as it helps to minimize the impact of some of the impairments although it may add some of their own.

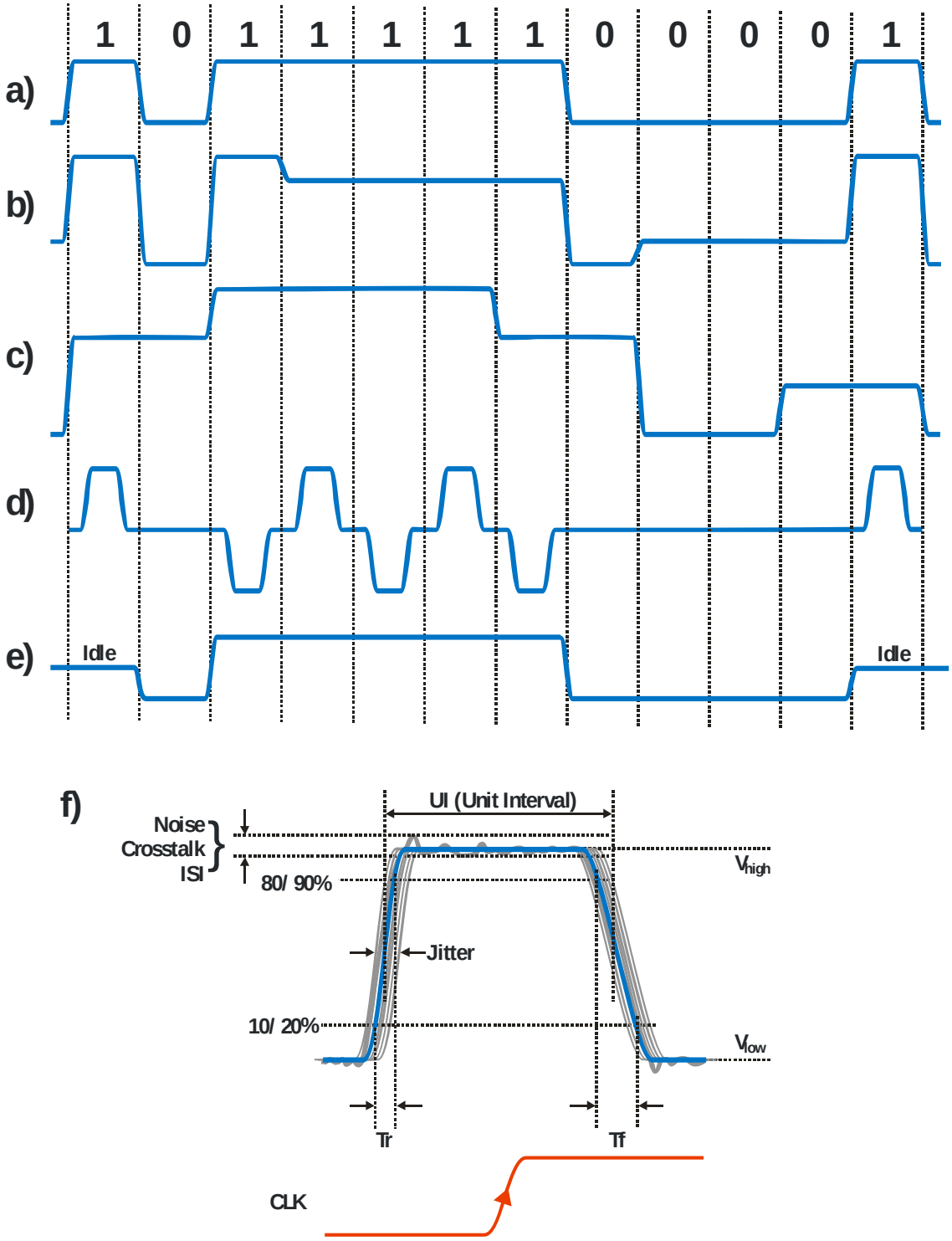


Figure **70**: Basic serial data signals have just two levels: "low" and "high" (a). However, many real-world signals need more than two levels. De-emphasis (b) compensates lossy interconnections while PAM (c) can pack more than one bit per symbol. Line coding can be another reason. AMI (d), or Alternate Mark Inversion, alternatively inverts the polarity of "1s" (Marks) to make sure the DC component of the signal is 0. Some serial standards define "idle state" levels when data in not flowing **(e).** Additionally, receiver test requires complete signal shape and timing control (f). AWGs can directly supply these signals, with or without distortions

Most high-speed digital signals consist in a sequence of “low” and “high” states signals representing some binary stream (Figure 70a). The binary stream may be properly coded in order to reduce the transmission errors and/or ease the receiver work. Some transmission schemes require more than two analog levels. In PAM (Pulse Amplitude Modulation) signals, multiple levels are used so more than 1 bit per symbol may be transmitted (Figure 70c). De-emphasis is a very popular technique to transmit fast transitions through bandwidth-limited connections (Figure 70b). Bipolar encoding requires two opposite polarity marks so the resulting signal has no DC component (Figure 70d). Error correction techniques can be also applied so that a better bit error rate may be obtained in the presence of noise and dispersion.

Analog characteristics of high-speed serial signals (Figure 70f) are extremely important and they must be characterized by the appropriate test equipment (i.e. oscilloscopes). Receiver designers must also check the ability of their designs to cope with distorted signals and establish their operating range. Pattern generators, either stand-alone or integrated in BERTs (Bit Error Rate Tester) or Logic Analyzers, are the traditional stimuli devices for this task. Pattern generators can supply two level signals following a predefined binary sequence. Users can set-up many analog and timing characteristics of the signal such as “high” and “low” voltage levels, rise and fall times, and jitter addition. In some cases, some more analog characteristics may be emulated through the usage of external devices and/or clever combination of multiple synchronized channels (Figure 71).

AWGs can supply any analog signal so any binary stream using any coding scheme can be generated if its bandwith falls within the instrument capabilities. Signals may be created without any distortion or they may be distorted on purpose to emulate real-world conditions regarding noise, cross-talk, jitter, or transmission path characteristics. AWGs flexibility greatly simplify receiver margin testing as analog and timing distortions may be accurately set-up.

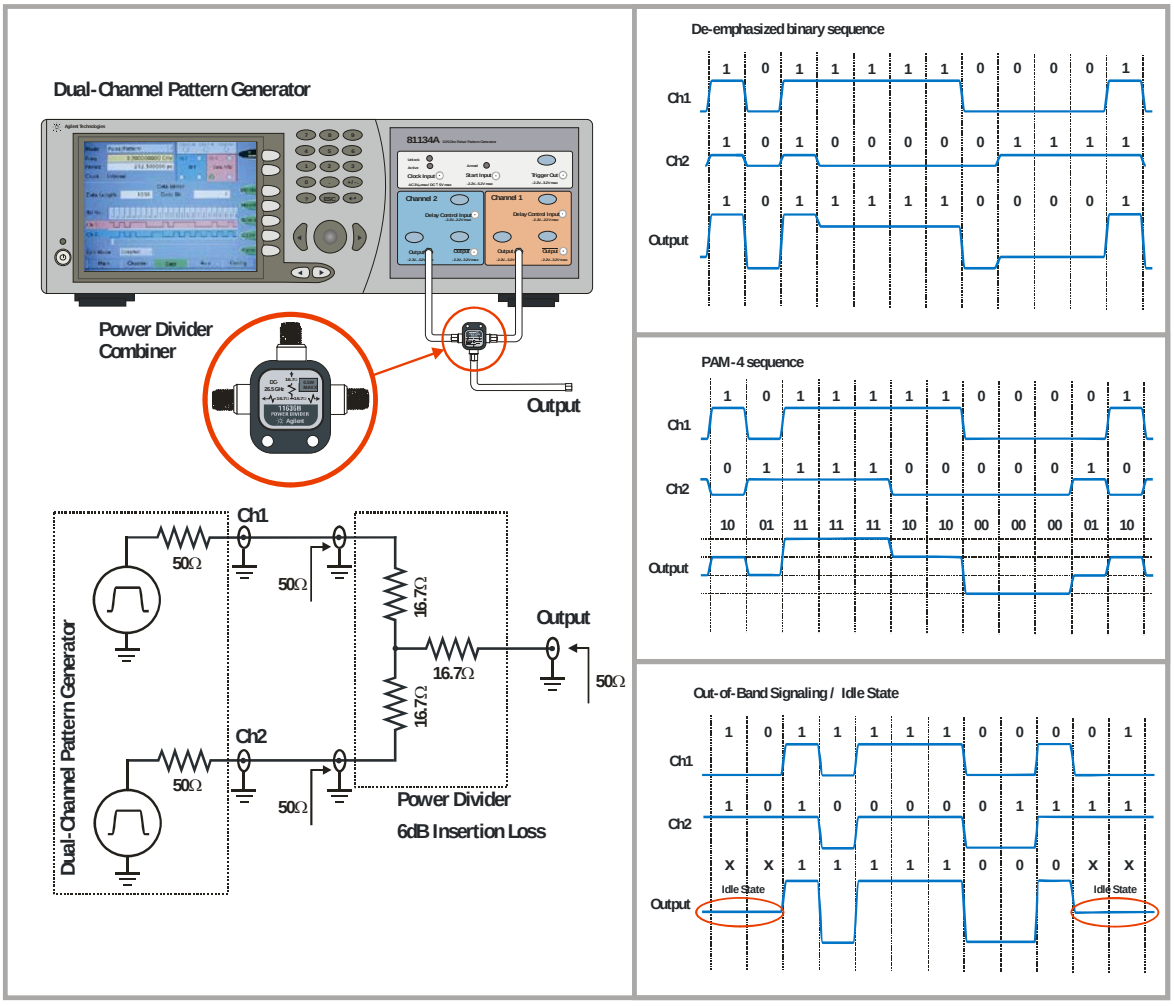


Figure 71: Pattern generators supply binary level signals. Combining two synchronous channels, some multilevel signals may be synthesized by playing with the data sequence, amplitude, and delay. The lack of flexibility and the limitations of these arrangements make AWGs much more convenient. However, pattern generator capability to generate very long data sequences can only be partially matched by most AWGs.

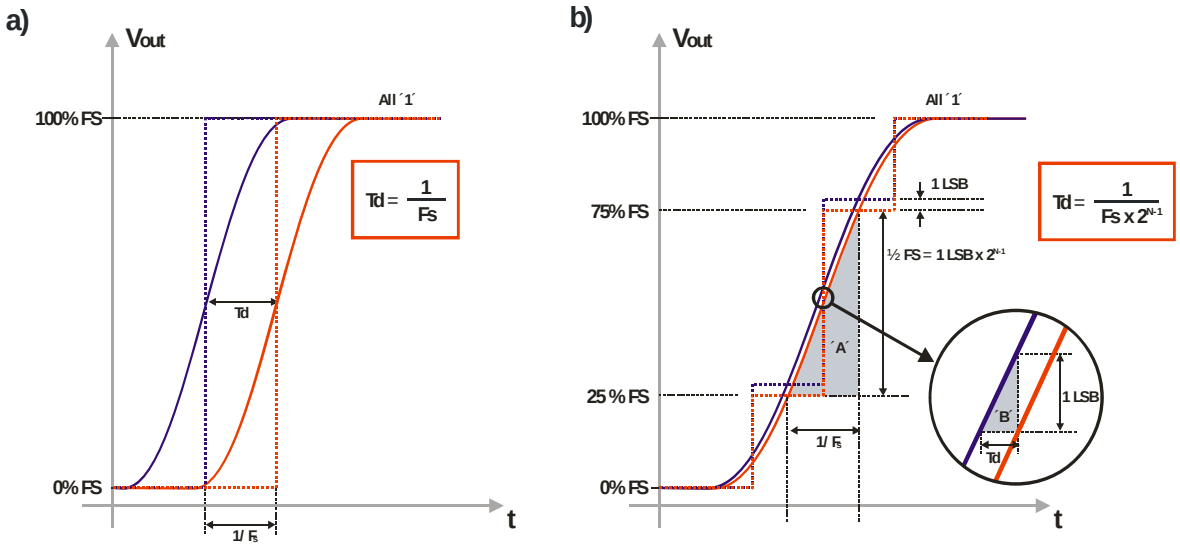


Figure 72: Serial data signal timing control is made through edge positioning. If edges are made by 0 samples (a), sampling period (1/Fs) is the timing resolution. Slower edges where some samples may be position allow for very fine edge positioning (b) by manipulating the vertical position of the samples in the edge so timing control resolution improves with vertical resolution.

5.2 Timing Control in AWGs

Timing control is paramount when generating any serial data signal. The capability to position all the edges in the desired moment in time influences the usability of the signal as clock (and therefore signal) recovery may depend on the accuracy and stability of it. At first sight, it looks like sampling period limits the timing resolution for any AWG. This is true when transition from one level to another is implemented by changing the output voltage from one sample to the next. For a given baud rate, sampling rate should be then the same or a multiple of it. Otherwise, a visible amount of jitter will show up in the output signal. This jitter is bounded by the sampling period so its peak-to-peak amplitude is equal to the sampling period (1/Fs). It may seem that setting up sampling rate to be a multiple of the baud rate of the signal is not an important limitation. However, many applications, such as a jitter insertion for jitter tolerance or jitter transfer test, may require a careful and accurate control of the position for all the edges in a serial stream. Additionally, it may be important to control other timing parameters such as rise and fall times or duty cycle distortion. Under these conditions it is not possible to guarantee an integer number of samples for all the symbols.

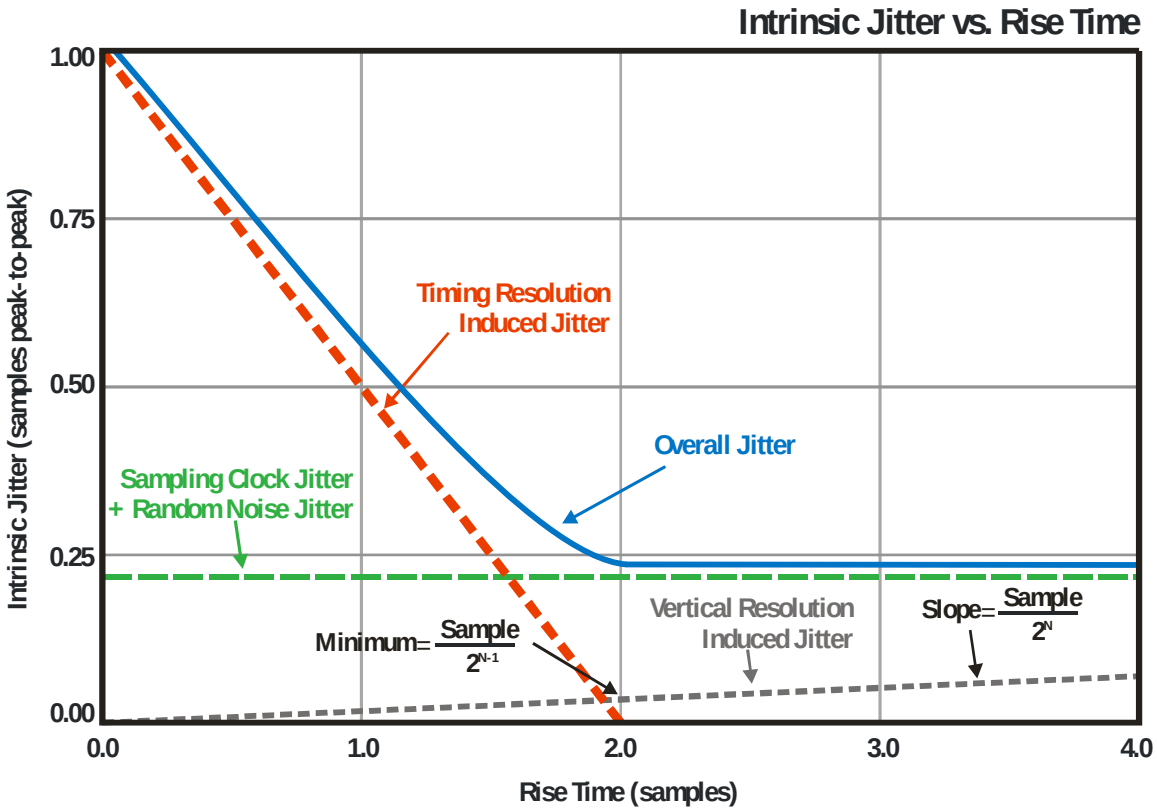


Figure 73: The number of sample periods in an edge also impacts signal jitter. Combining all the intrinsic jitter sources, the best results are obtained for around 2 sample periods in one edge. Beyond that figure, other sources of jitter dominate, especially sampling clock jitter and noise.

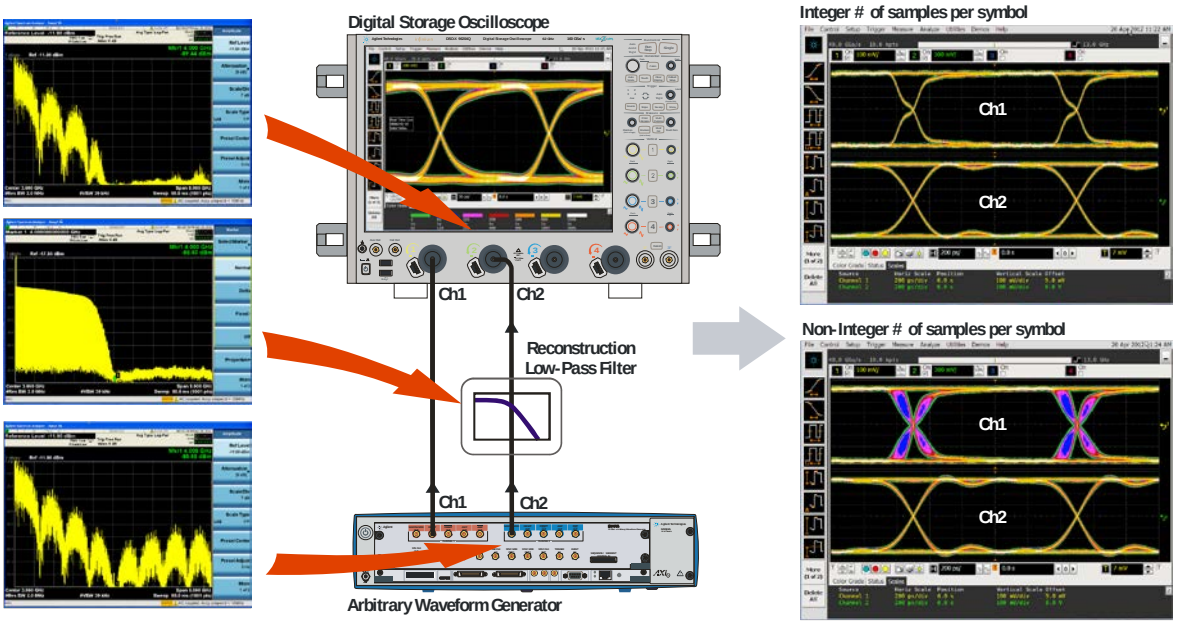


Figure 74: Using a reconstruction filter is important to obtain good jitter performance. Here two different synthesized 1Gbps signals are being generated using an 8GSa/s AWG. Rise and fall times have been selected so two sample times fit in every edge. The filtered version of the signal is fed into ch1 in an oscilloscope while the filtered version is captured using ch2. Intrinsic jitter is very sensitive to the number of samples per symbol for the unfiltered signal. The filtered signal looks the same no matter the ratio between the sampling and symbol rates.

Edge shape and positioning can be properly controlled in an AWG if transition times are equal or greater than two sample periods. Such situation can be seen in Figure 72b. Any edge can be slightly advanced or delayed by adding or subtracting one or more LSBs to the level of each sample in the edge. This results in a much finer timing resolution than the sampling period. Given that quantization error is also1 LSB peak-to-peak for any two levels in the edge, the same amount of peak-to-peak timing jitter will be generated. This component of jitter is generally very small when compared with other sources for actual devices. As an example, in a 12 bit vertical resolution, 12 GSa/s AWG the vertical resolution induced jitter is less than 50fs (but it is over 2.5 ps in a 6 bit vertical resolution AWG running at the same speed). Actual numbers depend on some additional factors such as AWG output linearity. Figure 73 shows the dependence between the intrinsic jitter and the number of samples periods in one edge. It improves as the number of samples in the edge grows to 2, but it does not fall to zero as there are other components of jitter such as sample clock jitter (including sampling instant uncertainty) and noise. The overall frequency response, including any reconstruction filter and cabling, will also add another component of deterministic jitter caused by ISI (Inter Symbolic Interference).

As a rule of thumb, two sample periods per edge are required to minimize intrinsic jitter while allowing full timing control. This requirement limits the effective baud rate that can be fully supported by a given AWG to about 1/3 to 1/5 of the instrument' maximum sampling rate. In fact, the majority of signal integrity literature states that a complete timing characterization of a serial data signal is only possible when the signal bandwidth includes the 3rd or, even better, the 5th harmonic of the serial signal highest fundamental frequency (baud rate/2). Following these criteria, bandwidth required for the signal matches the Nyquist bandwidth for the AWG' sampling rate. Bessel-type frequency responses are ideal as they avoid ringing in the output pulses while ISI generated jitter is minimized, as group delay is constant throughout the pass band. Given the limitations of Bessel-type frequency responses in terms of roll-off and, therefore, image rejection, maximum supported baud rates must be further reduced in order to obtain optimum results (Figure 74).

5.3 Jitter Generation with AWGs

For proper receiver testing and SSC (Spread Spectrum Clock) emulation, it is necessary to generate signals with a controlled amount of jitter. Although in AWGs, serial data streams exist only as analog waveforms, it is possible to control the signal timing by manipulating the sampling clock timing as done in pattern generators. Sampling clock may be controlled in two different ways: indirectly through the reference clock (Figure 75a) or directly through the DAC clock (Figure 75b). The principle is simple: if sampling clock is jittered then the output signal will be jittered as well. Reference clock control (i.e. through FM or PM modulation) may be only useful for relatively low frequency jitter and moderate amplitude as typically the external reference input is connected through a PLL circuit with a low-pass type of response and a limited locking range. This method may be the only practical way to generate wander (jitter under 10Hz) or other low frequency jitter transients (i.e. jitter hits). Given the indirect nature of timing control in this particular method, a proper jitter calibration procedure may be necessary to apply the desired amount of jitter to the signal. Some AWGs incorporate an external sampling clock input.

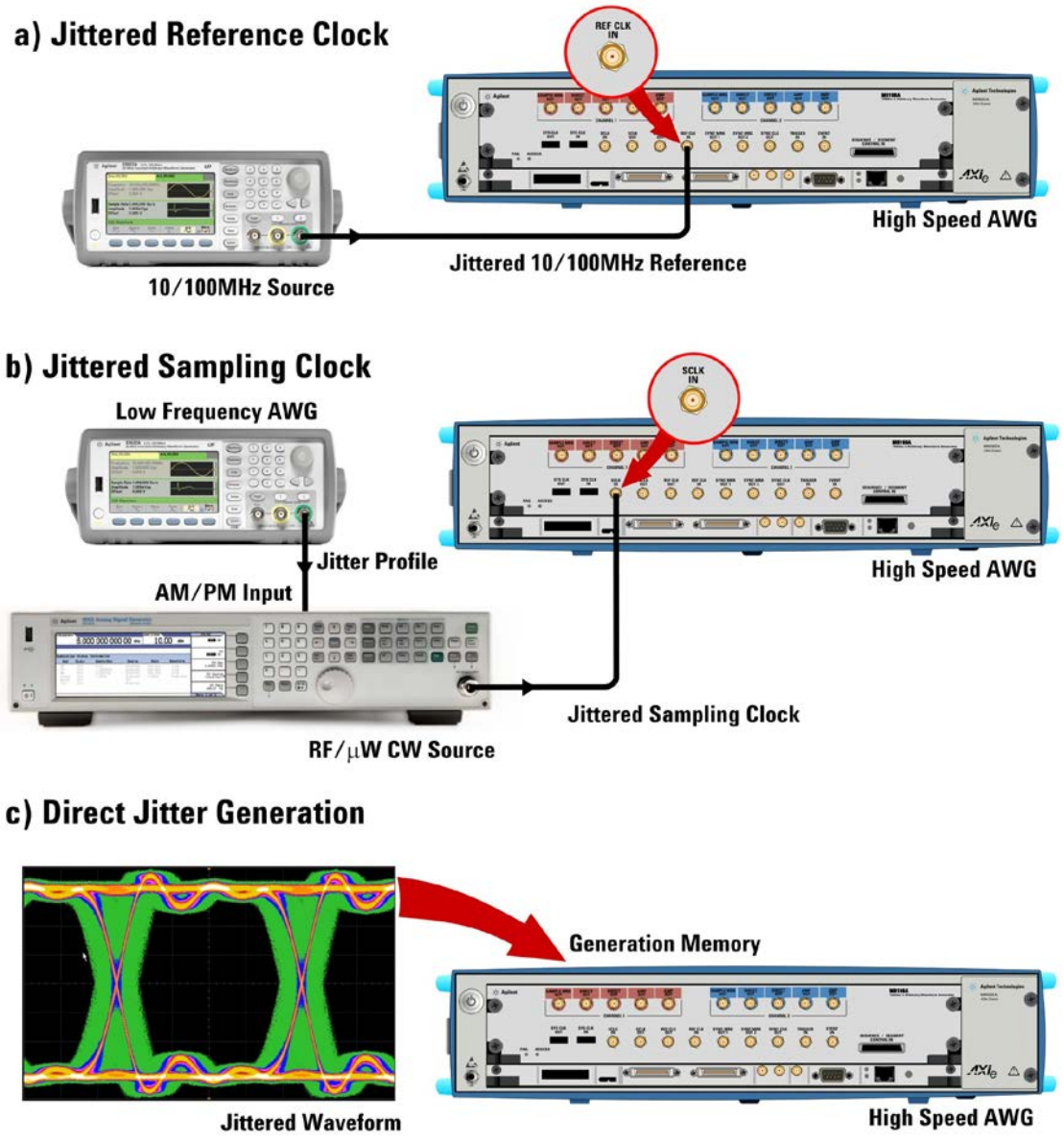


Figure 75: There are several ways to generated jittered waveforms. For relatively low-frequency jitter, a jittered reference clock may be used (a). Jitter will be transferred to the output signal through a PLL oscillator so its jitter transfer response must be calibrated. Some AWGs incorporate external sampling clock inputs. A jittered sampling clock will result in a jittered waveform (b). The most flexible way to generate jitter is by calculating an already jittered waveform and generate it directly at a fixed sampling rate (c).

Again, jitter may be generated by using a PM or FM modulated clock signal. Direct control of the sampling clock allows almost unlimited control of the signal jitter and avoids the need of careful calibration as the inserted jitter depends only on the characteristics of the signal. Any generator capable of generating FM and PM modulated signals, including another AWG, may be used as a jittered sampling clock source in any of the above methods. There two important advantages for these two methods: 1) jitter can be modified without changing the waveform in the AWG generation memory and 2) there is no correlation between the symbol stream and jitter in the signal. As most multi-channel AWGs share the same external reference and external sampling clock inputs, the same jitter will be applied to all the channels, what it is an important limitation of this methodology for some applications. The need of additional equipment is also a handicap.

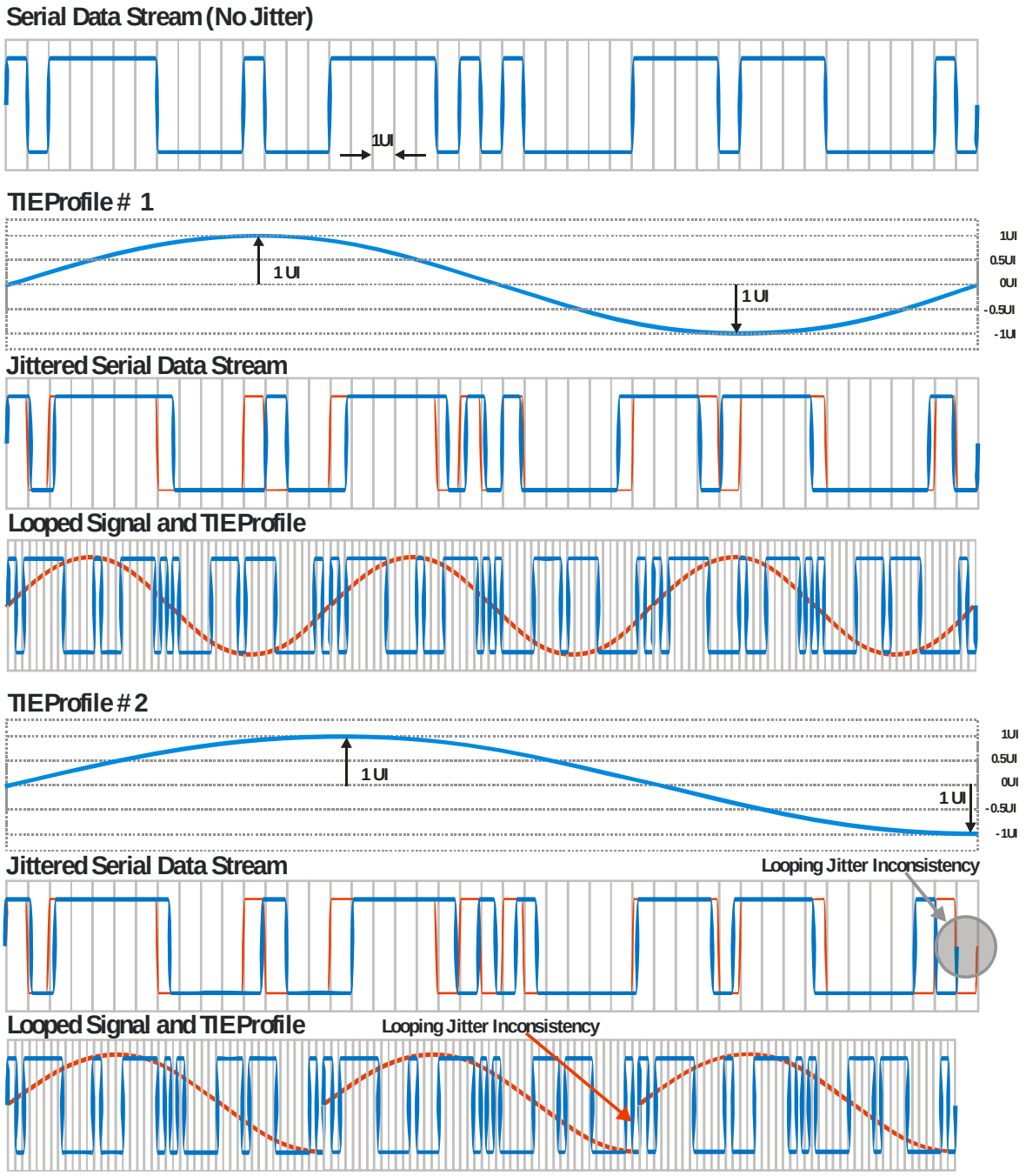


Figure 76: Looping jittered signals can result in inconsistencies even if the basic waveform has been properly built through wrap-around artifact-free techniques. Jitter TIE (Time Interval Error) profiles must be also designed to handle wrap-around. For sinusoidal jitter, record length must accommodate an integer number of cycles. If not, there will be jitter hits for every repetition of the signal, leaving it unusable for proper receiver testing.

A simpler jitter generation methodology consists in creating an already jittered waveform and generating it with an AWG (Figure 75c). This can be accomplished by placing the edges at the right timing locations following a time-domain jitter profile. The basic jitter parameter is TIE (Time Interval Error), which measures the time shift between the location of an edge and the ideal position according to a reference symbol clock. The jitter profile consists then in a TIE vs. time waveform. This method allows users to dynamically define all the timing and amplitude parameters for the waveform so any jitter profile can be created if it fits in the available time window. As seen in the previous section, edge timing can be controlled with high accuracy, resolution, and repeatability if the edges accommodate two or more samples. When this condition is not met, the shape of the jitter profile suffers a non-linear distortion, as achievable TIE values are quantized (Figure 79). Non-continuous jitter profiles result in unwanted jitter frequency components and they are very stressful for clock recovery circuits at the receiver under test taking to errored bits, bit slips, or PLL loss of lock that should not show up with a continuous jitter profile with the same amplitude.

5.4 Jitter Profiles and Waveform Memory

Standard jitter profiles include sinusoidal, square, triangular, and random. Other profiles can emulate real-world jitter behaviour such as “jitter hits” where a slow jitter transient shows up as the result of mapping and demapping tributaries to/from higher order transport signals in some long haul telecom networks. As any other signal characteristic, jitter profiles must be accommodated within the available waveform memory. As any other signal characteristic, it must be properly set-up to avoid unwanted wrap-around artefacts if the signal is going to be looped. In Figure 76, a sinusoidal TIE profile is applied to a binary serial sequence. First, an integer number of TIE cycles are applied (one in this particular case). The edges will shift according to the profile but as the TIE value is the same at both ends of the binary sequence, the overall duration of the signal will be exactly the same than the same sequence before applying any jitter to it.

Table 5: Required bandwidths and sample rates for some popular high-speed serial data standards over 2Gb/s depending on the type of test. Signal path emulation may reduce bandwidth needs while slow edges may limit sampling rate requirements.

		Basic Data Generation No Timing Control		Functional Testing Partial Timing Control		Compliance Testing Full Timing Control	
Standard	Bit Rate (MHz)	Fundamental (MHz)	Sample Rate (MSa/s)	3rd Harmonic (MHz)	Sample Rate (MSa/s)	5th Harmonic (MHz)	Sample Rate (MSa/s)
RapidIO	2,000	1,000	2,000	3,000	6,000	5,000	10,000
PCI-Express	2,500	1,250	2,500	3,750	7,500	6,250	12,500
SATA 3Gb/s	3,000	1,500	3,000	4,500	9,000	7,500	15,000
XAUI	3,125	1,563	3,125	4,688	9,375	7,500	15,000
HDMI 1.3	3,400	1,700	3,400	5,100	10,200	8,500	17,000
Fibre Channel (electrical)	4,250	2,125	4,250	6,375	12,750	10,625	21,250
Display Port	5,400	2,700	5,400	8,100	16,200	13,500	27,000
SATA 6Gb/s	6,000	3,000	6,000	9,000	18,000	15,000	30,000
PCI-Express 3.0	8,000	4,000	8,000	12,000	24,000	20,000	40,000
Thunderbolt	10,000	5,000	10,000	15,000	30,000	25,000	50,000

If a non-integer number of cycles are applied (0.75 cycles in the example) the duration of the signal will be either longer or shorter than the unjittered version. Changing the waveform record length, the AWG sampling rate, or a combination of both, always respecting the record length granularity requirements, will accommodate the new time window requirements. If the signal is looped, it will not show any apparent wrap-around artefact but the jitter profile will, making the signal unusable for tests such as jitter transfer or jitter tolerance. Looping that signal will result in a non-sinusoidal jitter profile with a very noticeable discontinuity in the transition. To solve this issue, the same wrap-around handling procedures followed with waveforms must be applied to TIE profiles.

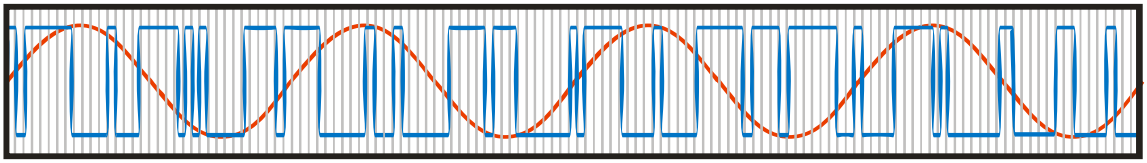
If sinusoidal jitter is taken as a reference, waveform record length (and time window, TW) fixes the minimum jitter frequency that can be generated: 1/TW. Another consequence is that frequency for any sinusoidal jitter will be a multiple of the minimum frequency as this condition results in an integer number of cycles. Reversing the argument, a given sinusoidal jitter frequency results in a specific time window requirement and, as a consequence, a specific record length. If the application requires the generation of a specific data sequence then the record length must accommodate an integer multiple of data sequences and jitter profile cycles. For any AWG, the longest time window achievable at the fastest sampling rate gives a good indication of the flexibility of the instrument to generate a wide range of jitter profiles at high symbol rates. As an example, the Keysight M8190A, with 2 GSamples record length at 12 GSa/s, can generate jitter with frequency components as low as 6Hz, well within the wander domain.

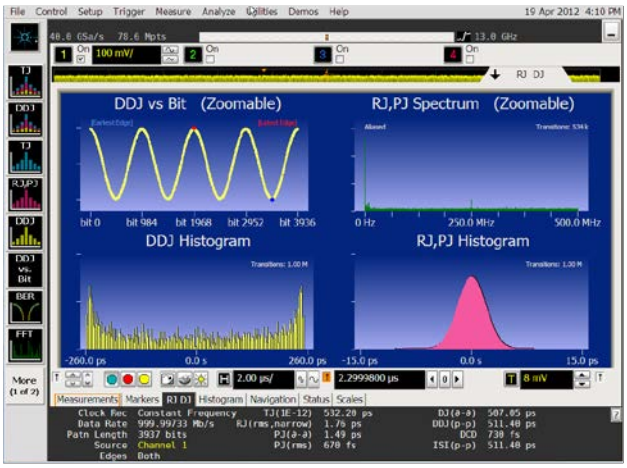
5.5 Data to Jitter Correlation in Direct Jitter Generation

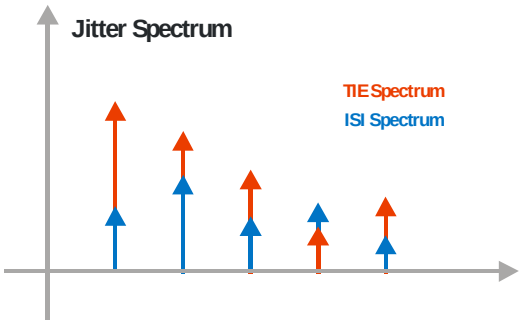
There are many jitter components in a serial signal. Some are deterministic and some are random. Some components may be bounded while others are unbounded. Finally, some are correlated with the data stream while some others are not. Advanced jitter analysis can separate the different components of jitter so designers can track the sources and improve link performance. Jitter analysis' algorithms look for patterns in jitter and correlations between the data stream and jitter in the time or frequency domain. ISI (Inter Symbolic Interference) is an example of jitter connected to the data stream (or DDJ, Data Dependent Jitter). Depending on the frequency response of the transmission system, previous symbols interfere with the current symbol and modify the shape and time position of its edges resulting in an increased jitter. In a typical jitter tolerance test, an AWG-generated data stream with some specific jitter profile should feed a data receiver under test.

The output stream should be then analysed by a DSO equipped with some advanced jitter analysis software. Comparing jitter at the input and the output and analyzing its components will provide some insight into the way the receiver handles the incoming jitter and adds some of its own. Input jitter not correlated with data is necessary to distinguish between periodic jitter and ISI. Continuous direct jitter generation always results in periodic jitter as the signal repeats exactly in the same way every time the same waveform is looped and all the components of the jitter spectrum will show up at multiples of the waveform repetition frequency. For a relatively short, non-repeating data stream, all the data correlated jitter sources, such as ISI, will show the same correlation and jitter analysis algorithms will not be capable of telling the difference between periodic jitter, random jitter, or ISI as all of them will be connected to the signal (Figure 77a).

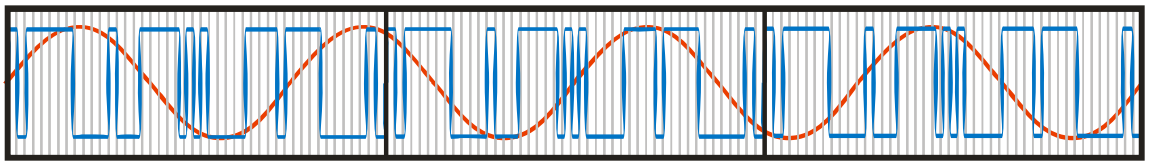
a) Long Binary Pattern and Correlated TIE Profile

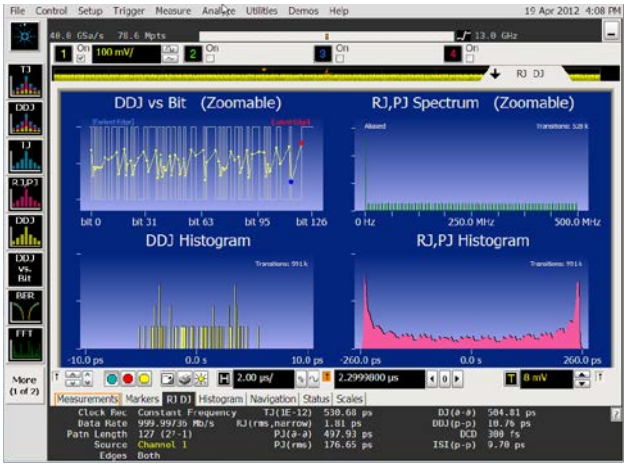






b) Repetitive Short Pattern and Uncorrelated TIE Profile





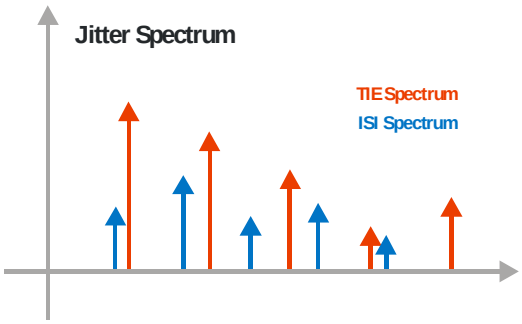


Figure 77: Direct Jitter generation may result in unexpected jitter behavior. For looping signals, all synthesized jitter components will be periodic. In (a) periodic jitter (sinusoidal in this example) is mistaken as data dependent jitter because there is a perfect correlation between data and jitter in both time and frequency domains. In (b) sinusoidal jitter is reported as periodic jitter while the data dependent jitter profile shows a small amount of ISI caused jitter. In this case, the same data sequence is repeated within the waveform memory while the sinusoidal jitter frequency has no direct harmonic relationship with it.

Although jitter will always be deterministic and repetitive, decoupling as much as possible the repetition rate for the data sequence and the jitter profile can minimize the correlation between data and jitter.

This goal can be accomplished by extending the data sequence within the generator memory. Some jitter analysis algorithms require the data sequence to be repetitive so, in this case, extension may be obtained by repeating the same sequence and making sure the jitter profile length is not an integer multiple of the basic sequence; so, frequency components of the jitter profile and the data correlated jitter will not overlap except for a few, high-frequency spectral lines (Figure 77b).

5.6 Jitter Calibration

The overall signal jitter at the AWG output will consist of the combination between the synthesized jitter and all the intrinsic sources. The latest include random and periodic jitter originated in the sampling clock, random jitter caused by thermal noise, ISI (Inter-Symbolic Interference) resulting from the system frequency response. As a result, total jitter in the generated signal will be higher than expected. ISI and random jitter may be compensated after calibrating the generation system (AWG, cabling, fixturing, etc). Proper calibration requires some jitter analysis tool capable of separating jitter components with enough accuracy. Modern high-performance real-time DSOs, such as the Keysight Infiniium 90000Q series (Figure 78), equipped with advanced jitter analysis software, like the Keysight EZJIT+, can offer very accurate characterization (jitter noise floor under 500 fs_{rms}). A flat oscilloscope frequency response, low noise, and the capability to de-embed the response of probes and fixturing will minimize the contribution of the characterization system to the measured jitter.

Overall random jitter may be typically characterized by its rms value. One way to characterize the intrinsic random jitter is by generating a jitter-less signal and analyze it. An alternative method consists in generating a signal with the desired level of synthesized random jitter. In both cases the excess random jitter will come from the generation system (plus some small contribution from the measurement system). Overall random jitter will be the result of combining all the sources of random jitter. For uncorrelated random jitter components the following expression may be used:

$$(R_j^t)^2 = (R_j^s)^2 + (R_j^i)^2 + (R_j^m)^2$$

R_j^t Total random jitter

R_j^s Synthesized random jitter

R_j^i AWG intrinsic jitter

R_j^m Measurement system random jitter

Once actual random jitter is obtained, synthesized jitter can be corrected so the overall jitter matches test requirements.

ISI correction is more elaborate. Generally speaking, ISI (an important component of DDJ or Data Dependent Jitter) results from non-flat, non-constant-group-delay frequency responses. A way to add a given level of ISI to a serial data waveform is by applying a low-pass filter properly characterized.

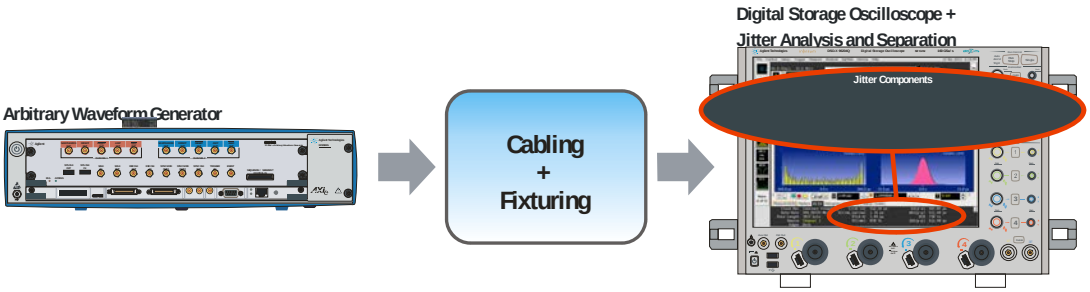


Figure 78: Jitter calibration must be used to compensate the effects of the AWG's intrinsic jitter and other sources such as cabling and fixturing. Without calibration random and data dependent jitter (especially ISI) would be higher than expected affecting the accuracy of any jitter tolerance test. Proper jitter calibration requires a low jitter noise floor analysis instrument and some jitter analysis software capable of doing jitter component separation.

However, the combined frequency response of the AWG and the connection to the DUT will modify the combined response resulting in different than expected ISI jitter. Proper frequency response characterization may help to compensate the filter applied to the waveform so the combined response is closer to the one expected. Again, calibration may provide the overall ISI jitter so the right correction may be applied to the synthesized signal.

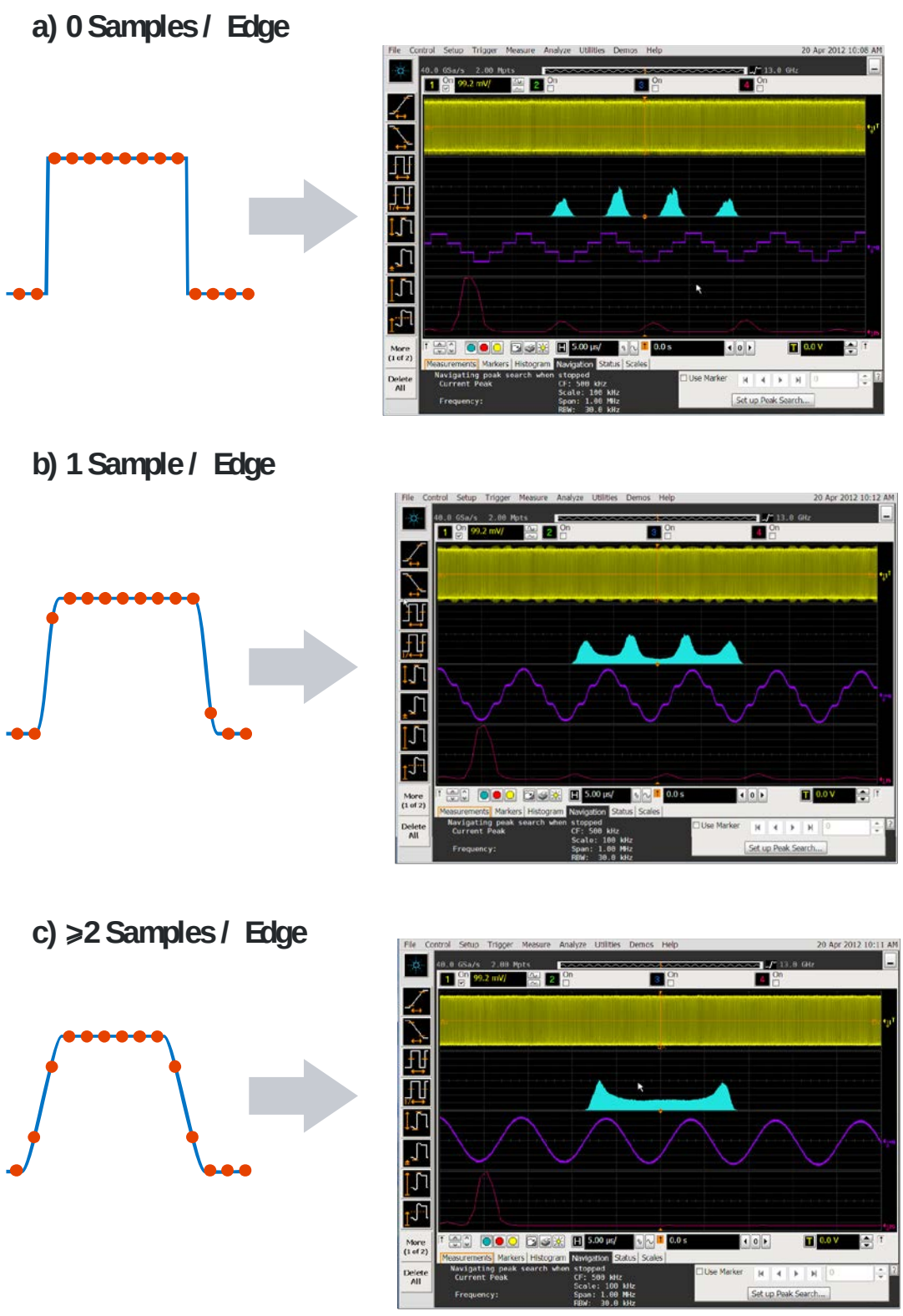


Figure 79: Then number of samples in an edge defines the timing control granularity. Here, the same sinusoidal jitter is applied to a binary sequence. With 0 samples in an edge (a) sinusoidal jitter is quantized (also visible in the jitter histogram) and its spectrum shows unwanted harmonics. One sample in an edge (b) results in an smoothed TIE waveform but it is still distorted. Setting up two samples per edge solves the distortion problem (c) as timing granularity is only limited by the vertical resolution.

5.7 Differential Signal Generation

Most high-speed serial (and many other) signals transmitted through cables, backplanes, and tracks in a PCB, are differential (i.e. LVDS, Low Voltage Differential Signalling). Differential signals require two independent transmission paths where the information is transported as a voltage difference between them. Typically, the signals sent through each path, called true and complement, are made of a common mode and differential mode parts. Common mode part may be described as the average voltage ($V^+/2 + V^-/2$) while the differential part results from subtraction ($V^+ - V^-$). Well-behaved differential signals use to have a constant DC component as the common mode signal. This implies that the varying part of the signals in each differential part have the same shape but inverted. Some new signalling schemes even propose the usage of superimposed common mode and differential mode signals to transport different information (i.e. Data and Clock), sometimes at different rates (i.e. data and interface control information) and in different directions. Generally speaking, differential receivers must be sensitive only to the differential mode signal and must reject any common mode signal. Testing differential receivers requires generating near-ideal and impaired signals. There are some impairments that are unique to differential signals. Most of them are one way or another related with the common mode signal. Any shape, amplitude, or timing asymmetry between the positive and negative component will result in an unwanted common mode component showing up and the distortion of the differential mode signal. Any unwanted common mode signal (i.e. crosstalk) may also show up as a differential component if both signal paths and/or the receiver are not properly balanced.

Most high-speed AWGs incorporate complementary output DACs that directly offer well-balanced, time-aligned differential signals as the same current sources and switches are used for both outputs (Figure 80a). Additionally, an offset control can shift the DC component for both outputs simultaneously. Latest generation products, such as the Keysight M8190A, even provide a fine differential offset control and a VTT (Termination Voltage) control to improve the accuracy and usability of the supplied differential signal. To generate a valid differential signal, users just have to load the differential signal waveform into the generation memory, set the desired amplitude and common-mode DC level, and play it back. For 2-channel AWGs with only single ended outputs, polarity reversed signals may be downloaded to each channel' generation memory and play-it back synchronously (Figure 80b). In this way, one of the channels will supply the S^+ signal while the other will do the same with the S^- signal. Though, in this scheme, any channel-to-channel jitter will show up as additional jitter in the differential signal. Any imbalance or lack of accuracy between both channels will result in an unwanted common mode signal.

Complete differential receiver characterization requires signal pairs with both differential and common-mode signals. In this way, the sensitivity of the device under test to common-mode signals can be properly established. Common-mode signals may come from different sources such as output or signal path imbalance, noise, and crosstalk. In the other side, common-mode signals result in a differential component in case there is some imbalance between the positive and negative signals in the differential pair. Generating both signal components simultaneously requires two arbitrary generation channels per differential pair. Figure 80 shows different AWG arrangements for differential signal generation.

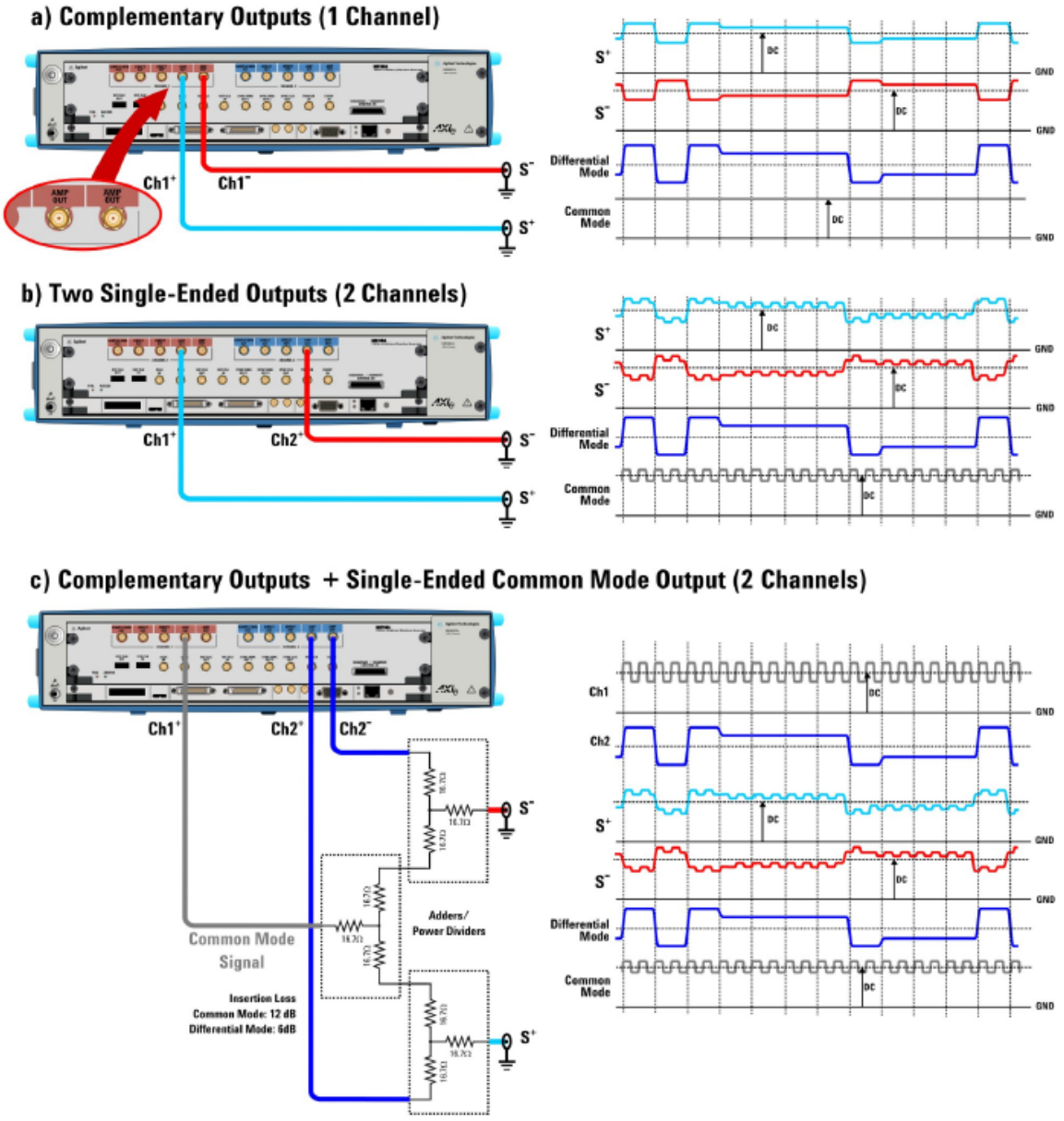


Figure 80: There are several ways to generate differential signals with a AWG other than a single-ended to differential converter. Most high-performance arbs incorporate a complementary output where the DC level can be set-up independently (a). Using two independent channels allows for full differential connection emulation (b) as any combination of common mode and differential mode signals may be implemented. Combining complementary output and single-ended channels can offer the same results. Combining the signals, though, may result in lower amplitude differential pairs.

When only the complementary outputs from one channel are used, a DC level is the only common-mode component that can be generated (Figure 80a). Using two synchronous channels generating the signals being transferred through each line in a differential pair (Figure 80b) allows for the simultaneous generation of differential and common-mode signals. Waveform in Channel 1 will be composed by the combination of the common-mode signal for both channels and the corresponding differential-mode component for each channel. Finally, two channels may be used to independently generate the common-mode component (single-ended by nature) and the differential component (through a complementary output). This arrangement requires the usage of a series of power splitters and combiners (Figure 80c) to combine both components into a differential pair. In this case, both channels do not have to be synchronous unless common-mode and differential components are.

The emulation of a differential pair using two different synchronous AWG channels is, perhaps, the most flexible way to generate realistic differential signals as it allows for the generation of virtually any combination of differential-mode and common-mode signals. However, especial care must be taken as combining the common-mode signal and one of the differential signals into one single channel may be problematic if the available voltage range is not sufficient. As the DC component may be handled by the offset control, users must fit the complete signal swing in the channel's full scale and make sure the maximum and minimum voltage do not go beyond the output amplifier limits. Differential jitter is also an issue as both signals come from different DACs while complementary outputs are generated by the same components in a single DAC.

5.8 Signal Path Response Embedding and De-Embedding

Most receiver tests check the capability of any device to withstand the degradation of serial data signal so the error level is kept under an acceptable level. The feasibility of errors may be evaluated directly (i.e. through a bit error rate tester, or BERT) or indirectly (i.e. through the statistical analysis of the eye diagram and/or jitter using a real-time or sampling digital oscilloscope). Generators must be capable of generating undistorted, jitterless signals that become distorted after being transmitted through a given signal path or interconnection. Sometimes, tests may be more flexible if a generator supplying an equally distorted signal directly to the receiver replaces the actual signal path. Worst-case scenarios and operating ranges are easier to characterize in this way while cost and time of testing may be reduced. The process of adding a controlled level of linear distortion to a signal in order to emulate the effects of the transmission path is known as signal embedding. On the other hand, sometimes the connection between the AWG and the system under test (i.e. cabling and fixturing in an ATE system or a section or a PCB track from a test point to the IC under test) may add unwanted distortion to the signal. Linear compensation of the signal to obtain an undistorted signal at a given point in the signal chain is known as signal de-embedding. Embedding and de-embedding techniques are also becoming very popular in high-bandwidth digital scopes.

Embedding and de-embedding are quite straightforward to apply for two port systems. First of all, it is necessary to obtain an accurate model of the signal path to emulate or compensate. These models may come from a simulation tool integrated in an EDA system or modeling software or they may be obtained through direct measurements using frequency-domain (i.e. a VNA or Vector Network Analyzer) or time-domain (i.e. through TDR/T or Time Domain Reflectometry / Transmission) techniques. The outcome of these characterization procedures is a 2x2 array with 4 S-parameters (Figure 81) and phase) function of frequency. The S21 parameter specifies the transmission response between the input and output ports of the interconnection so it can be used as its frequency response. The embedding procedure implies obtaining the impulse response corresponding to the S21 parameter, resample it to the target AWG sampling rate, and convolve it with the undistorted waveform. Circular convolution must be used if the signal is intended for looping. The resulting waveform coming out from the generator will be equivalent to that at the output of the interconnection. De-embedding follows the same steps except that the frequency response should be inverted before obtaining the impulse response. Applying the correction will result in an undistorted signal at the output of the interconnection.

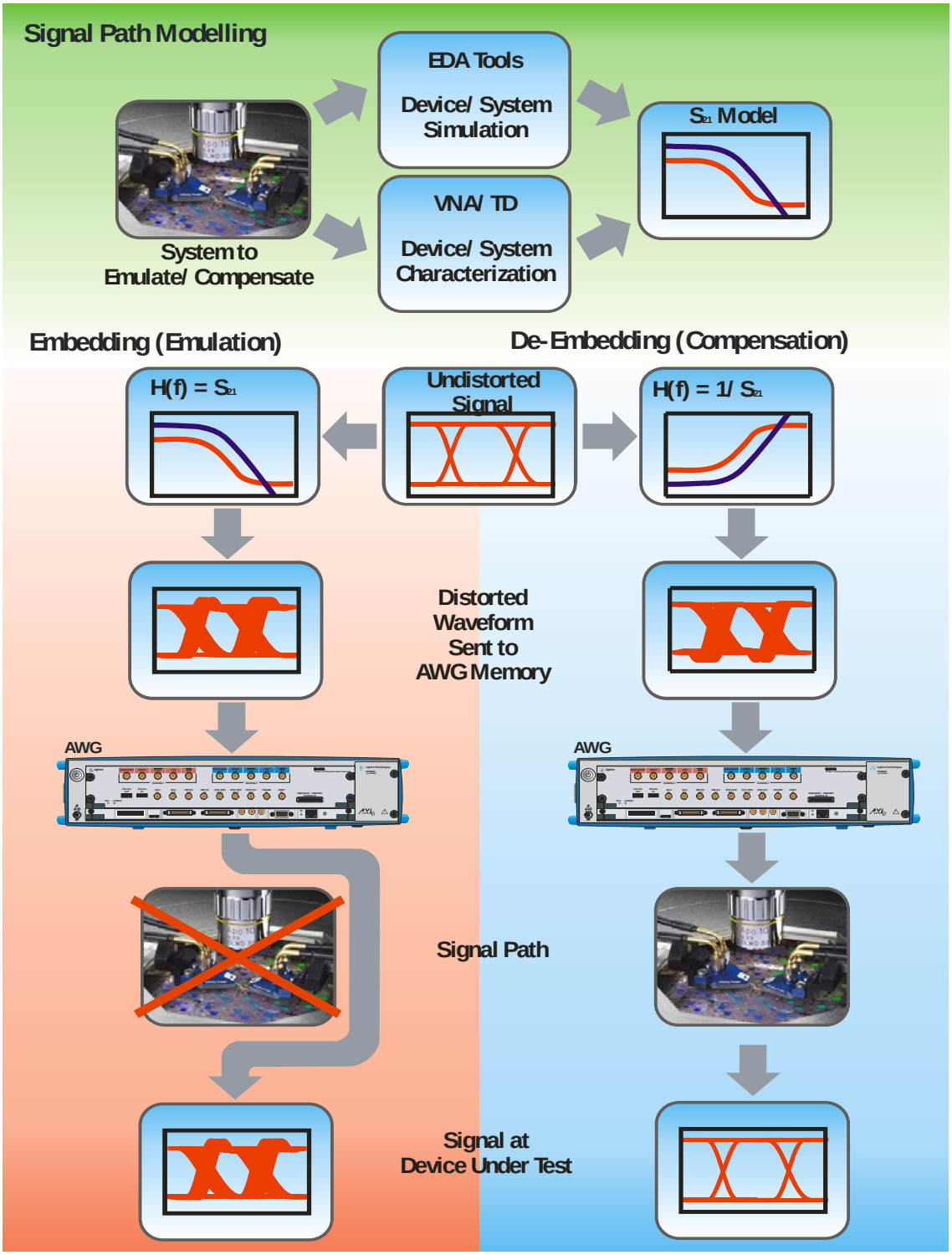


Figure 81: Embedding and de-embedding follow the same basic procedure. However, their goals are the opposite. Before application, an accurate model of the signal path (i.e. an interconnection in a PCB) must be obtained. Embedding applies the model to the undistorted signal so the effects of the modelled interconnection may be emulated. De-embedding applies a reversed response derived from the interconnection model so the signal at the end of the interconnection will be undistorted.

Differential interconnects are far more complex to deal with as they involve 4 ports referred to ground. The resulting S-parameters consist of a 4x4 matrix. This matrix can be applied directly to obtain the output voltages (V_o^+ and V_o–) if the input voltages in the differential pair (V_i^+ and V_i^-) are known:

$$V_o^+ = S_{31} \times V_i^+ + S_{32} \times V_i^-$$

$$V_o^- = S_{41} \times V_i^+ + S_{42} \times V_i^-$$

Most times, it is more convenient to deal with the differential and common modes of a differential signal instead of doing it with the corresponding four single-ended signals, as this approach is better aligned with the way signals are transmitted. There are several methods to directly obtain mixed-mode parameters using standard instrumentation. Differential (V_D) and common-mode (V_C) signal components can be easily obtained from the single ended voltages (V^+ and V^-):

$$V_D = V^+ - V^-$$

$$V_C = ½ (V^+ + V^-)$$

$$V^+ = V_C + V_D /2$$

$$V^- = V_C - V_D /2$$

Following this convention, a new set of S-parameters, which relate the common and differential modes at the input and the output of the connection, can be defined. These are known as the Mixed-Mode S-parameters and they can be easily derived from the 4-port model (Figure 82). Common and differential mode at the output (V_O^C and V_O^D) can be calculated from those at the input (V_I^C and V_I^D) through the following expressions:

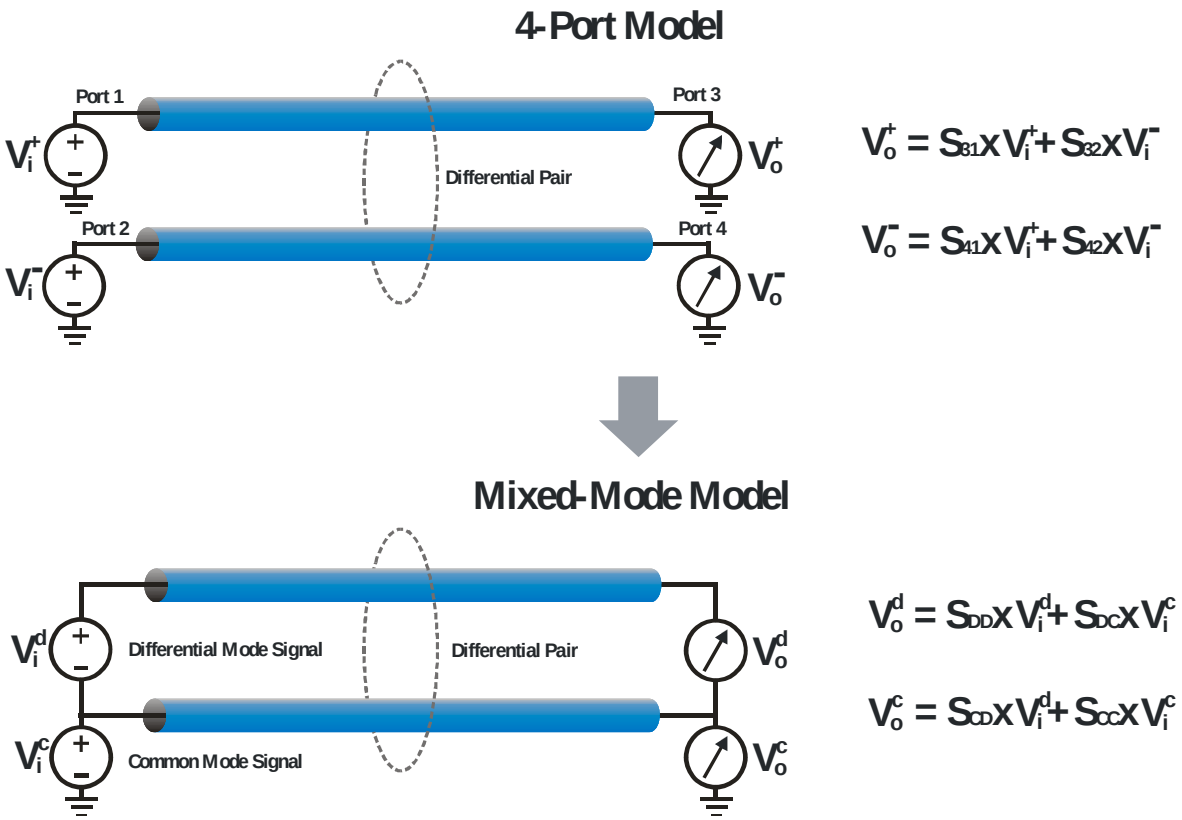


Figure 82: Differential interconnections must be properly characterized using a 4-port model. In this way, the influence of the interconnection on both the differential and common-mode signal will be correctly applied or compensated. Mixed mode models are more intuitive and easier to apply as they take differential mode and single-ended mode signals as inputs and outputs. Mixed-mode S-parameters may be derived from 4-port models or characterized directly using adequate instrumentation.

$$V_O^D = S_{DD} \times V_I^D + S_{DC} \times V_I^C$$

$$V_O^C = S_{CD} \times V_I^D + S_{CC} \times V_I^C$$

The above expression reflects the fact that imbalanced connections results in a fraction of the differential component at the input becoming common mode component at the output, while part of the common mode component at the input becomes differential mode component at the output.

In the previous paragraph, different strategies to generate differential signals with AWGs were discussed (Figure 80). For single channel, complementary output generation (Figure 80a) it is not possible to generate a common mode component other than a DC level. However, it is still possible to partially emulate the influence of a common mode (typically unwanted) signal through an imbalanced interconnection over the differential mode signal at the output. Complete embedding and de-embedding is only possible with implementations based in the usage of two independent (but synchronized) AWG channels. In particular, differential signal de-embedding can compensate for imbalance in the differential interconnection by applying properly compensated signals at the input. Compensation requires the inversion of the mixed-mode S parameter matrix so the following expression must be used:

$$V_I^D = (S_{CC} \times V_O^D - S_{DC} \times V_O^C) \times (S_{DD} \times S_{CC} - S_{DC} \times S_{CD})^{-1}$$

$$V_I^C = (-S_{CD} \times V_O^D + S_{DD} \times V_O^C) \times (S_{DD} \times S_{CC} - S_{DC} \times S_{CD})^{-1}$$

In case a differential-only signal is required at the test point, there will not be any common mode signal at the output (V_O^C = 0 + DC component) so the above expressions may be simplified to the following ones:

$$V_I^D = (S_{CC} \times V_O^D) \times (S_{DD} \times S_{CC} - S_{DC} \times S_{CD})^{-1}$$

$$V_I^C = (-S_{CD} \times V_O^D) \times (S_{DD} \times S_{CC} - S_{DC} \times S_{CD})^{-1}$$

In practice, embedding and de-embedding are equivalent to apply a filter. De-embedding is an especially sensitive operation as most times it involves the compensation of lossy interconnections, especially at high frequencies. This implies a reduction in the maximum signal amplitude. Although device and interconnect characterizations made using VNAs or TDR/T systems are capable of covering a very wide frequency range, models applied should be limited to the signal useful bandwidth, typically 2.5 times the baud rate (fifth harmonic of the signal's fundamental frequency).

5.9 Crosstalk and Noise Insertion

Another unavoidable signal component is noise. In high-speed serial interconnections crosstalk is a very important contributor to noise. Thermal noise is always present and, unlike other noise sources, its amplitude is unbounded. Generating a realistic test signal requires to add noise and crosstalk in a repeatable, controlled way. For a given interconnection (i.e. a track in a backplane or a PCB), most of the noise comes from other nearby interconnections. One way to test in realistic conditions is to inject additional uncorrelated (“aggressor”) signals into these interconnections. This can be done with additional channels from the same or other AWG generators, pattern generators, or even actual traffic. Noise may result in errored bits coming out of the receiver as the signal can trespass some decision threshold.

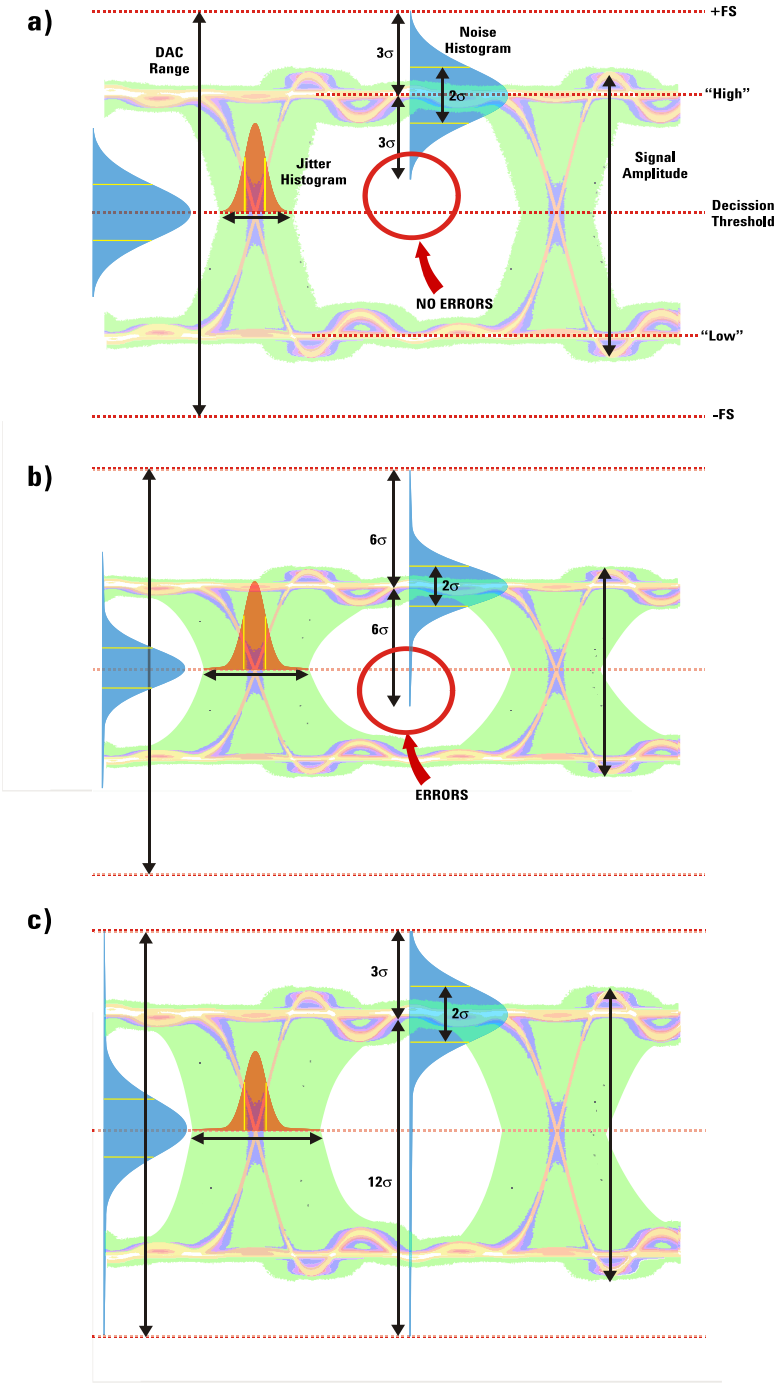
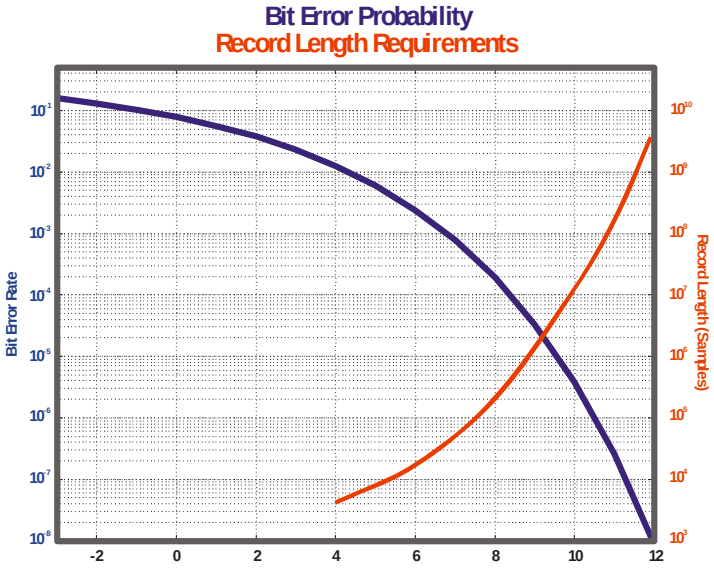


Figure 83: Different AWGN (Additive White Gaussian Noise) addition scenarios. Gaussian noise must be mandatorily clipped to fit in the available AWG output range. In (a) noise have been clipped so much that noise never crosses the decision threshold and, as a result, there will be no errors. In (b) the same noise power has been kept but the noise crest factor has been increased. Errors will happen but an increased combined peak-to-peak amplitude results in a noticeable reduction of the signal's amplitude. In (c) it is the combined signal + noise waveform what is clipped. Signal's amplitude is kept and the right error level can be reached. Random jitter caused by AWGN is also affected by the way Gaussian noise is added.

Again, these interconnections may be modelled through simulation or direct measurements. Once the model is obtained, it is possible to synthesize additional uncorrelated, synchronized or not, data streams and to compute a new differential or single-ended signal including the different contributions of crosstalk noise according to that model.

The overlapping of multiple bounded, uncorrelated noise sources results in a bounded Gaussian-like amplitude statistical distribution. Ideal thermal noise is fully Gaussian so its amplitude is unbounded. AWGs can only generate bounded signals so Gaussian noise cannot be fully implemented (Figure 83). Additionally, maximum peak-to-peak amplitude is also limited in AWGs and, as a consequence, as noise peak-to-peak amplitude grows, the available amplitude range for the serial data signal decrease. One way to deal with this issue is by artificially limiting the amplitude of Gaussian noise to an acceptable peak-to-peak value. This may be done by clipping the noise when its current value goes beyond or below the limits. A comfortable way to define the clipping level is by defining the noise' crest factor (or PAPR, Peak-to-Average Power Ratio). This number is the ratio between the maximum acceptable peak and the root-mean square (or σ) of the original unbounded Gaussian distribution. The actual rms amplitude of the noise will be lower as the highest peaks will be removed. Using this approach, the statistical amplitude distribution will be symmetrical and it will show two Dirac δ at the extremes of the distribution. The problem with this approach is that, depending on the peak-to-peak amplitude of the noise (a combination of its rms amplitude and its crest factor), there may be no errors caused by it
(Figure 83a). One solution is to increase the crest factor so a significant portion of the histogram goes beyond the decision threshold (Figure 83b). In this case, the overall peak-to-peak amplitude of the signal + noise will be higher and, given the available voltage range in the DAC, the signal amplitude will be reduced and potentially it may be lower than some desired value, leaving the signal unusable. A third approach is to synthesize an unbounded Gaussian noise, adding it to the signal, and clipping the overall signal when it goes beyond the DAC boundaries (Figure 83c).



Record Length vs. SNR

SNR (dB)	Uncoded	
	BER	RL
4	1.25×10^{-2}	4 KSa
6	2.39×10^{-3}	17 KSa
8	1.91×10^{-4}	262 KSa
10	3.87×10^{-6}	12.9 MSa
11	2.61×10^{-7}	192 MSa

Figure 84: Adding computed gaussian noise is not enough to make sure the signal will behave correctly from the statistical standpoint. Given a target bit error rate, the signal must accommodate at least 10 times the average error period (1/BER). Serial data signals using FEC (Forward Error Correction) results in lower error rates so required record length will significantly grow respect to that of uncoded signals with the same SNR.

The clipping levels must be enough to accommodate the highest and lowest signal peak plus a significant portion of the Gaussian histogram. 3σ may be a good choice as the signal is under that level 99% of the time. This ensures that the average noise level is very close to zero. This approach will result in near ideal error levels caused by noise without taking too much dynamic range out of the AWG.

Generating realistic noise may seem enough to properly emulating a noisy channel. Unfortunately, there are some other requirements. The most important one is related to the amount of time that non-repeating noise must be generated so that results in a statistically valid error behaviour. Figure 84 shows the relationship between SNR and BER (Bit Error Rate) for a binary digital sequence with AWGN. For 1E-6 error level there will be only one error in 1,000,000 bits. In order to generate a well-behaved signal, the total record length must accommodate at least 10 times that value (10,000,000 bits or 50 MSamples @ 5 samples/symbol). If not, errors may be significantly lower or higher than expected. The usage of FEC (Forward Error Correction) techniques, so popular in long distance optical links and radio transmissions, makes things more difficult given that even poor SNR, resulting in very high number of errors before correction, will cause a very low error level after correction, depending on the coding gain of the algorithm in place. Generating a short signal will probably result in the right level of errors before correction but those few errors per waveform segment may be either corrected, so there will be no errors after correction, or uncorrected, and error level after correction will be much higher than expected. It is not uncommon that the error level after correction reaches 10^{-10} or 10^{-12} (known as QEF, Quasi Error-Free, and conditions). Generating a data sequence made out of 10^{13} bits, or about 10^{14} samples is out of the reach of today's general-purpose AWGs. Some AWG generators, such as the Keysight 33500 series, can generate Gaussian noise algorithmically (repetition period is over 5 years) so this noise does not suffer of any unwanted statistical effect.

6 Wireless and Modulated Signal Generation Using AWGs

KEYSIGHT TECHNOLOGIES

6.1 Wireless Signal Generation Architectures

Complex modulated signals are the basis for today's wireless revolution. They can be found in mobile phones, military radars, contactless credit cards, WiFi networks, GPS navigation systems, RFID tags, or deep space communication, just to name a few. The same techniques can be found in wired communication systems such as xDSL or CATV and they are being successfully applied to boost the capacity of long-haul optical networks by applying complex modulation to optical carriers. The rapid evolution of modulation schemes and coding systems, the need to support multiple wireless standards with the same hardware, the continuous performance improvement required for the successful deployment of faster services for a growing number of users, and the more demanding transmission environments make wireless testing a challenge. Generators must be capable of supplying multiple standard signals, sometimes simultaneously, while simulating complex linear and non-linear distortions so receiver designs and components can be properly validated. Arbitrary waveform generation techniques have been the foundation of most RF/Wireless vector signal generators (or VSG). Two-channel external or internal AWGs typically supply high precision baseband signals to feed a quadrature modulator (Figure 85a). This working scheme allows for the complete control of both the amplitude and phase of a carrier so any analog or digital modulation scheme, distorted or not, are feasible. AWGs are ideal as no special hardware is required for neither a given modulation scheme nor a modulation standard, and adding new ones only depends on the availability of adequate external software. As sampling rate and SFDR performance in AWGs have dramatically improved over time, direct generation of multiple, dissimilar, modulated IF/RF carriers is now feasible (Figure 85b). New AWG architectures, specifically designed for modulated signal generation, are being introduced in state-of-the-art high performance, ultra-high-speed arbs such as the Keysight M8190A.

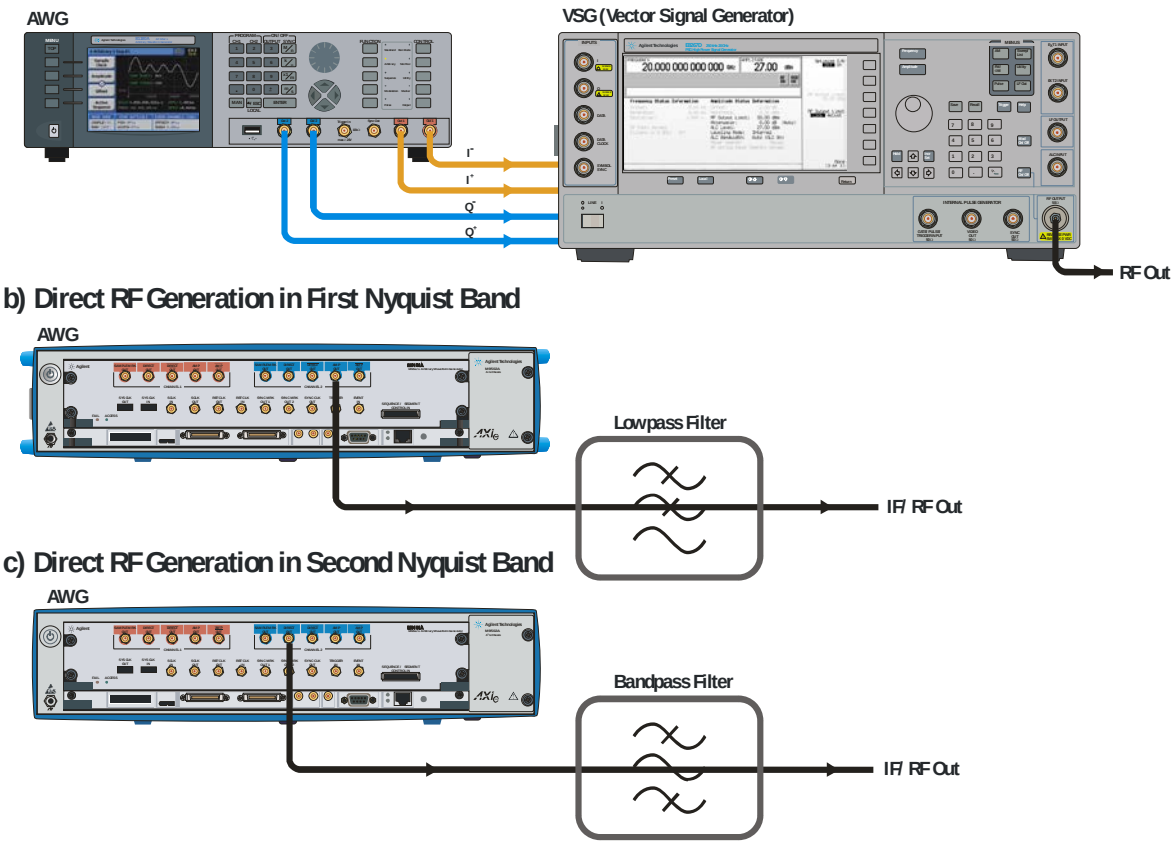


Figure 85: AWGs can generate RF signals in different ways. Traditional vector signal generation (a) relies on two-channel AWGs to generate quadrature baseband signals (I and Q). Modern high-speed AWGs can directly generate one or multiple dissimilar modulated carriers (b) right at the target RF frequency or at an useful intermediate frequency (IF). It is possible to reach even higher carrier frequencies by using the signal image in the second Nyquist band (c). The doublet mode introduced by the M8190A 12 GSa/s is especially adequate for this kind of operation.

These architectures include the efficient usage of the second Nyquist band through the introduction of the “doublet” DAC mode (Figure 85c), and the implementation of full-digital real-time quadrature modulators and up-converters within the AWGs.

6.2 External Quadrature Modulation

Quadrature modulators require two external baseband signals (Figure 86). The In-Phase (or I) and quadrature (or Q) components modulate two carriers of the same frequency with a 90º nominal phase difference between them. The baseband signal pair can be also seen as a complex signal (I being the real part and Q the imaginary part) modulating a single carrier. The spectrum of a complex signal is not symmetrical around the zero frequency so the bandwidth of a quadrature modulated signal doubles the bandwidth of the baseband signal and upper and lower sidebands must be transmitted undistorted as there is not redundancy as it happens with AM signals. In other words, given a modulation bandwidth B, the baseband signal bandwidth must be B/2. Modulation bandwidth depends on the modulation scheme and the characteristics of the filter applied to the baseband signals in order to keep the power of the signal within the boundaries of a channel. Any unfiltered DAC images coming from the arb supplying the I and Q signals will show up as unwanted signals in the RF spectrum. These images must be kept under a minimum acceptable level so an aggressive reconstruction filter, a high oversampling factor or a combination of both must be applied. Good quality signals must meet a series of requirements:

- Good modulation accuracy
- Low noise and spurious level
- Low phase noise
- Low residual carrier
- Good ACPR

Any difference between the I and the Q signals will impact the modulation accuracy. Differences may arise from amplitude, phase,

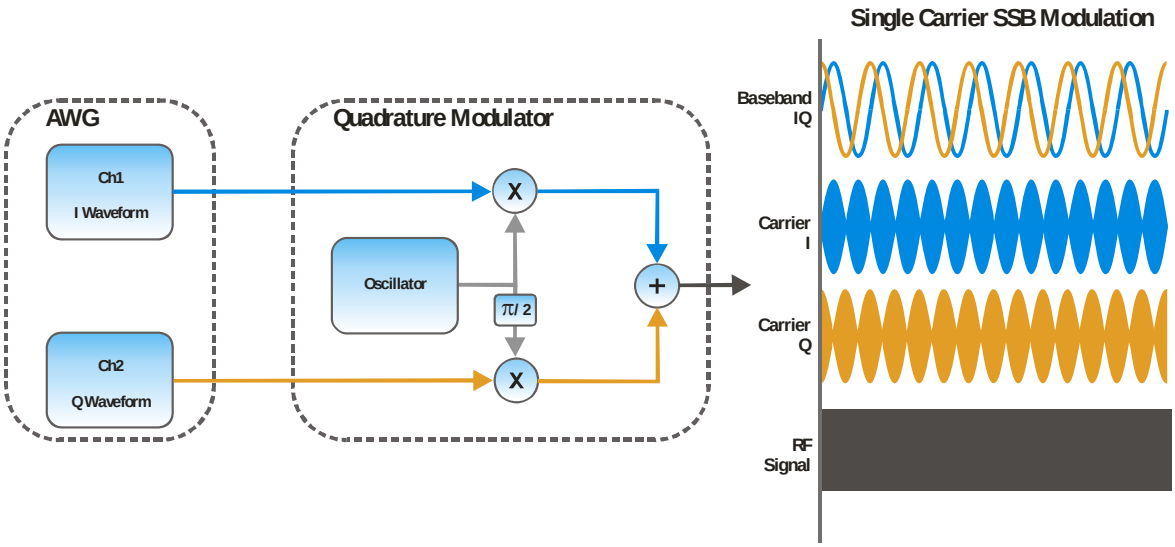


Figure 86: Quadrature modulation apply two basebands signals to carrier-suppressed AM modulators fed by two carriers of the same frequency and 90° phase difference. The orthogonality of the two carriers makes possible the transmission of two independent signals in the same frequency band. A perfect quadrature modulator produces a single carrier of fc-fm or fc+fm when two equal power sinusoidal signals of frequency fm and 90° relative phase are applied. The resulting RF signal will be a single side band (SSB) carrier with constant amplitude envelope.

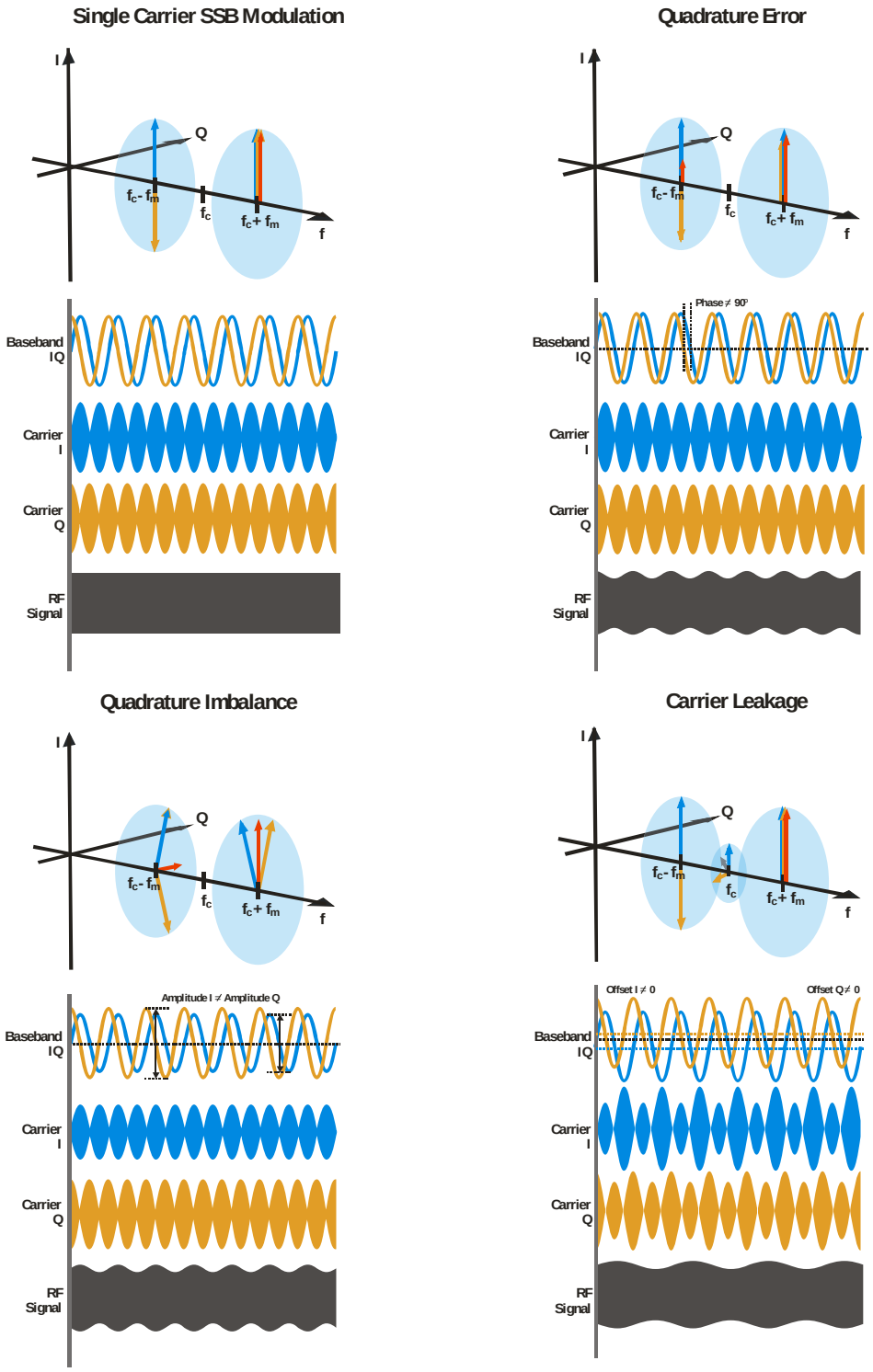


Figure 87: In an SSB sinusoidal quadrature modulation the output of each modulator is a dual side band signal as expected from any amplitude modulator. Adding them together, though, results in the cancellation of one of the sidebands. Actual quadrature modulators are not perfect and unwanted sidebands or signal components can be produced by amplitude and/or phase errors between the I and Q components.

delay, and frequency (magnitude and phase) response (Figure 87). A good way to look at the consequences of unequal I and Q components is by analyzing the behavior of the spectrum of an ideal single-side-band sinusoidal modulation. Baseband signals are just sinusoidals with the same amplitude and 90 (or -90) degrees phase difference. The ideal modulated signal will result in a single carrier with a frequency f=fc+fm or f=fc-fm depending on the relative phase. Modulation for each carrier is in fact a suppressed carrier AM modulation so their spectrum will consists of two carriers at a distance fm from the carrier frequency fc. When both signals are added together in a perfect quadrature modulator, two of the side carriers cancel each other while the others get boosted. Effects of different quadrature impairments may be analysed separately. These effects behave in a linear way as they are superimposed independently in the final signal. When the amplitude of the I and Q components are not equal, an impairment known as quadrature imbalance, the cancellation of the unwanted side carrier is not complete so a residual image will show up. The phase of the unwanted carrier with respect to the wanted carrier is 0 or 180 degrees. If the phase between the orthogonal carriers is not 90 degrees apart, an impairment known as quadrature error, the unwanted side carriers will not be in the same plane so, again, a residual, unwanted image will show up but, now, the phase of this carrier will be orthogonal to the wanted carrier. Wanted carriers are also affected by the impairments but for small differences the effects on those will be almost negligible. The relative amplitude of the images can be easily calculated:

1. Quadrature Imbalance

   $$Image_{QI}\ (dBc) = 20\ log((A_I-A_Q)/(A_I+A_Q)$$

   For small differences, $A_I=A_Q=A_{BB}$

   $$Image_{QI}\ (dBc) = 20\ log\ (\square A/2A_{BB})$$

   Quadrature imbalance (QI) is expressed as a percentage of the amplitude difference between the I and Q baseband components and their average amplitude. As a consequence

   $$Image_{QI}\ (dBc) = 20\ log\ (QI/100) - 6.02$$

2. Quadrature Error

$$Image_{QE}\ (dBc) = 20 \log ((2\ A_{BB} \sin (\phi/2)/\ 2\ A_{BB} \cos (\phi/2))$$

Where ϕ is expressed in radians. For small ϕ,

$$\sin (\phi/2) = \phi/2$$

$$\cos (\phi/2) = 1$$

As a consequence

$$Image_{QE}\ (dBc) = 20 \log (\phi/2)$$

Quadrature Error (QE) is usually expressed in degrees so

$$Image_{QE}\ (dBc) = 20 \log (QE \times \pi/360^{o})$$

I versus Q delay can be seen as a linear phase shift depending on the frequency. As a result, the amplitude of the image will change with fm.

The previous discussion λεαδσ to the conclusion that good frequency response and channel match is critical to generate good quality signals. AWGs are just one of the factors influencing the final result. The response of the cabling and the quadrature modulator itself are important too. Some two-channel AWGs specifically designed for IQ baseband signal generation, such as the Keysight N8241A, follow an in-factory calibration procedure to match the response of both channels through DSP. Its combination with high-quality, wideband quadrature modulators such as the Keysight PSG, allows for the generation of highly accurate digitally modulated signals with over 1 GHz of modulation bandwidth without further frequency-dependent corrections.

DC components in either the I and Q signal paths will also result in an unwanted residual carrier component. This impairment, known as carrier leakage, can be minimized by carefully adjusting the I and Q offsets in the AWG although some modulators, such as the Keysight PSG, also incorporate fine DC offset adjustment controls.

For relatively low modulation bandwidths, adjusting the overall baseband signals amplitude and offset and the 90 degrees phase between the I and Q carriers and using good-quality, matched cabling may be enough to get an acceptable level of quadrature impairments. High modulation bandwidths, though, may require a careful frequency dependent system calibration and signal correction to get good results.

6.3 Direct IF/RF Generation

AWG's role in the generation of modulated signals using an external quadrature modulator is limited to the baseband component. As a result, sampling speed requirements are basically set by half of the modulation bandwidth. Aggressive oversampling and/or reconstruction filtering may be necessary, though, to avoid DAC images showing up as unwanted RF signals that may interfere with adjacent channels, limiting the overall ACPR (Adjacent Channel Power Ratio) and spur performance. Today's high speed, high-bandwidth AWGs can support the direct generation of modulated carriers so there is no need for an external quadrature modulator (Figure 88).

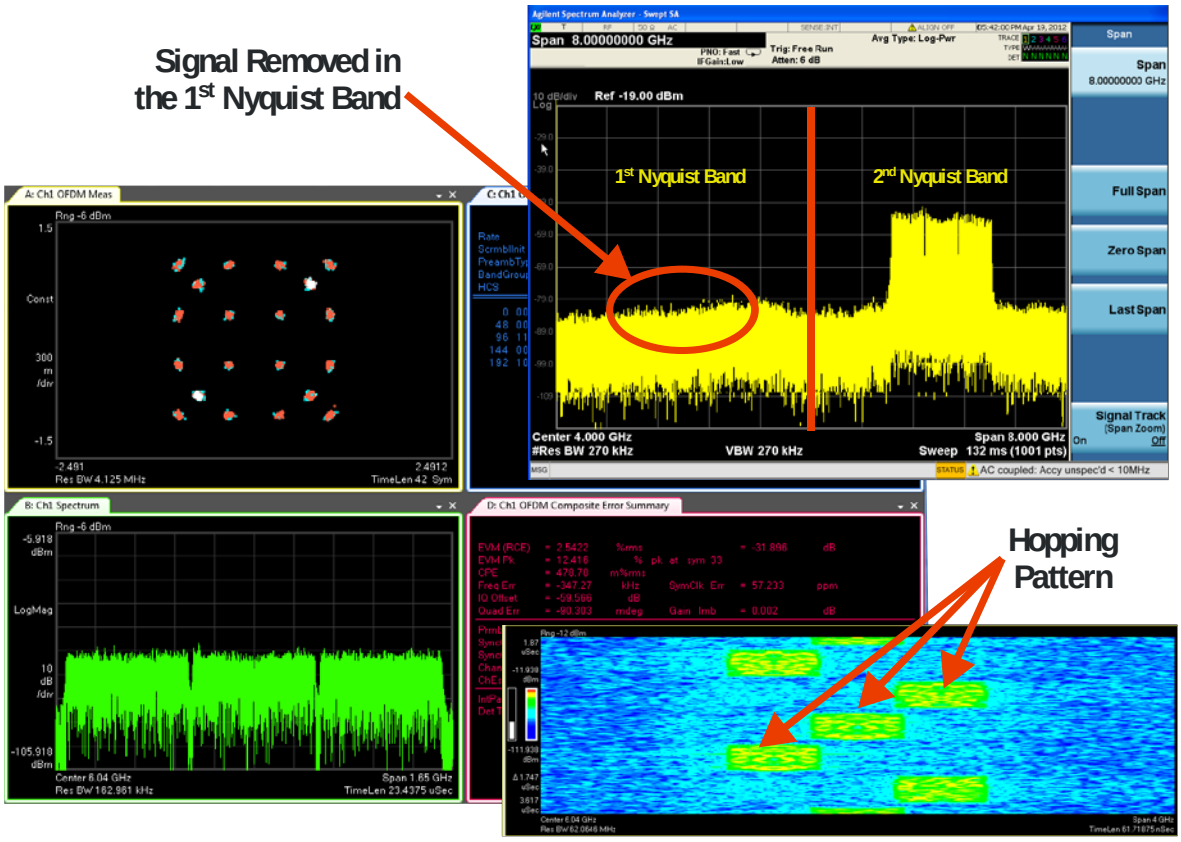


Figure 88: A fully compliant UWB (WiMedia) signal centered at 6GHz directly generated in the second Nyquist band by A Keysight M8190A generator running at 8 GSa/s. Signal amplitude and modulation quality is excellent thanks to the usage of the "Doublet" mode. An adequate band-pass filter has been applied to fully remove the signal image in the first Nyquist band. Notice the modulation bandwidth required for this signal (@1.5 GHz) and the hopping frequency characteristic of this signal, visible in a spectrogram (in the lower right corner).

Direct signal generation shows some important advantages:

- Simpler generation set-up as there is no need for external equipment.
- Only one channel is required.
- No quadrature impairments as quadrature modulation is performed mathematically.
- Modulation bandwidth only limited by the AWG sampling rate and bandwidth.
- Generation of multiple modulated carriers over a large bandwidth.

Direct RF signal generation has also some drawbacks:

- Sampling rate requirements depend on the highest frequency component. Carrier frequency and modulation bandwidth, especially for wideband signals, influence the sampling rate requirements.
- As sampling rate is high, waveform memory requirements may be much higher than the ones for baseband generation. The same time window results in a larger waveform memory requirement and a longer signal calculation. Baseband signals must be properly interpolated to the DAC sampling rate for good quality signals.
- High sensitivity to DAC resolution and non-linear impairments in the AWG.
- Unwanted signal images will show up beyond the Nyquist frequency. Those signals may be eliminated through proper low-pass or band-pass filtering.

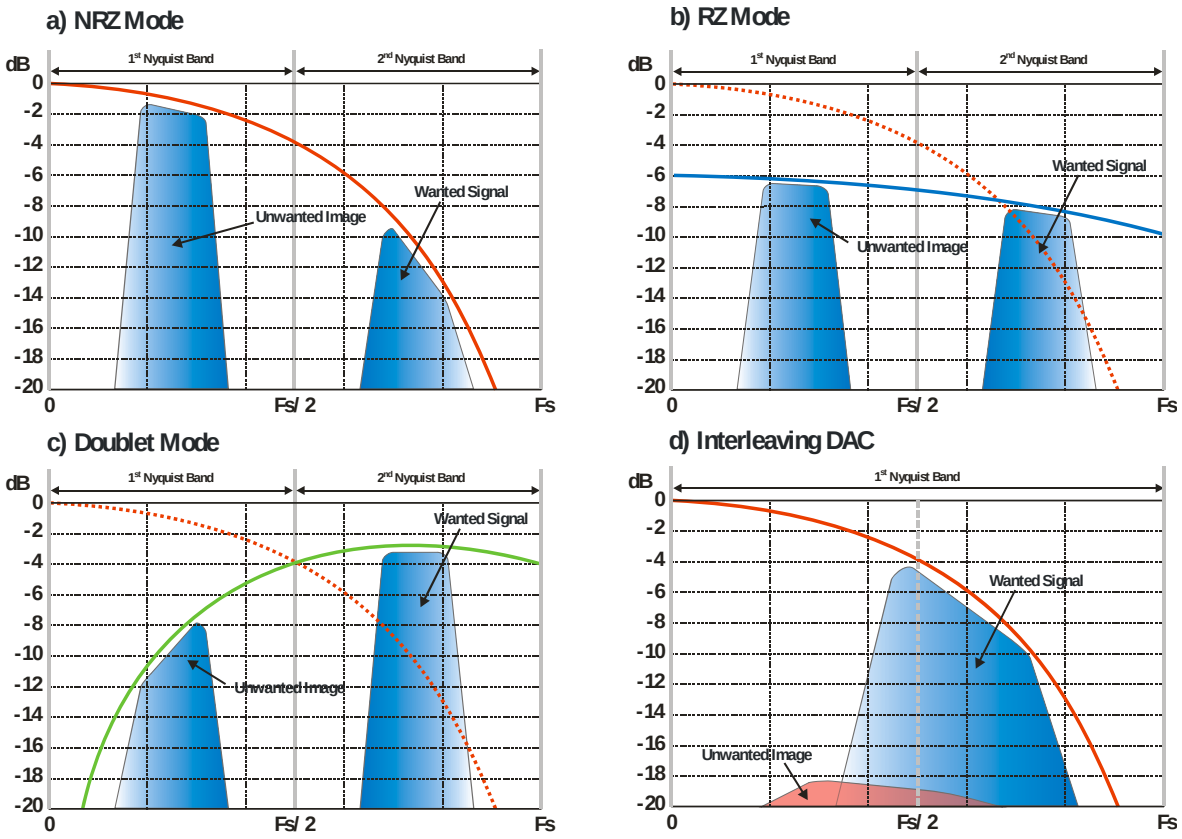


Figure 89: Traditionally, generation of signals in the second Nyquist band results in a lower quality signal (a). The RZ mode available in some generators slightly improves the amplitude and flatness of the wanted signal (b). Doublet mode (c) is the only conversion architecture specifically designed for 2nd Nyquist band operation, showing a much larger signal amplitude, a much better flatness as it eliminates the zero located at Fs, and a lower amplitude of the unwanted image. Interleaving DAC (d) does not improve usability of frequencies located the second Nyquist band but it extends the 1st Nyquist band instead. Although this strategy can improve usable bandwidth, it does not provide good quality signals in the frequency domain due to the in-band images and the low dynamic range available at frequencies close to Fs.

Some modern high-speed arb architectures incorporate some characteristics specifically designed for direct RF signal generation. Some AWGs incorporate reconstruction filters to flatten the frequency response up to near the Nyquist frequency at the maximum sampling rate. In that way, the combination of the AWG's analog roll-off and the sin f/f response neither results in a linearly distorted signal nor in a corrected signal with a lower amplitude and, as a result, in a poorer dynamic range and SNR.

Using images beyond the first Nyquist band allows the generation of higher frequency RF signals (Figure 89). Regular AWGs are designed so their frequency response falls rather quickly beyond the Nyquist frequency. Additionally, the sinc(f) DAC response falls to zero for f=Fs. The Keysight M8190A 12 GSa/s AWG incorporate two DAC modes specifically designed to generate signals in the second Nyquist band (6-12 GHz). The RZ mode boost the sinc(f) response by doubling the frequency for its first zero. However, amplitude of the signal is reduced by 6 dB so gains in the amplitude of the images in the second Nyquist band are limited. The higher amplitude of the signals (now unwanted) in the first Nyquist band makes necessary the usage of very selective band-pass filters to get rid of them. The M8190A' "Doublet Mode" improves the second Nyquist band amplitude while reducing the amplitude of signals in the first Nyquist band as it adds a zero at DC. It does not show the 6dB amplitude loss, as the doublet pulse is active over the whole sampling period. The double mode also removes the zero located at Fs resulting in a flatter frequency response. A band-pass filter is still required to remove images out of the second Nyquist band but overall signal quality is greatly improved. Sampling rate requirements depend now on the signal in the first Nyquist band. Spectrum in the second Nyquist band will be mirrored so signals must be spectrum reversed. This is easy to accomplish by inverting one of the baseband components during the signal calculations. The carrier frequency that must be used for calculations can be obtained using the following formula:

$$Fc' = Fs - Fc$$

Fc' = Carrier frequency used in calculations

Fc = Desired carrier frequency

Fs = Sampling Frequency

The interleaving DAC architecture is also a method to generate signals beyond the Nyquist frequency. However, it does not improve the usability of the second Nyquist band. The main goal of this architecture is to extend the first Nyquist band to allow for the generation of higher bandwidth signals. In fact, current implementations, based on the addition of two delayed channels, do not even remove the first zero located at Fs. Interleaving DAC architecture may be useful to generate very wide bandwidth signals (i.e. high-speed serial signals) but signal quality may be extremely limited for RF signal generation as any imbalance between the two channels involved results in unwanted signal images that would interfere with the signal itself for any modulated signal sitting in the Fs/2 frequency.

6.4 Full-Digital Quadrature Modulation and Up-Conversion

External quadrature modulators and direct RF signal generation both have pros and cons. The Keysight M8190A incorporates a new generation method that combines the advantages of both methods and eliminates the drawbacks (Figure 90). Each DAC incorporates two independent digital oscillator blocks along with two baseband signal processing blocks. Up to two Baseband (i.e. I and Q) signals are fed into the processing block at a fraction of the sampling speed where they are resampled (through real-time interpolation) to match the sampling rate for the DAC. Each baseband signal goes to a digital multiplier where a digitally synthesized carrier is also applied. Finally, the output of both multipliers is added and supplied to the DAC. Amplitude, frequency, and phase can be set independently for each oscillator. In a typical configuration, both digital oscillators will share the same frequency but with a 90º phase difference to effectively implement an all-digital quadrature modulator. As all the signal processing is digital there will not be any misalignment caused by differences in the signal path or analog imperfections, just as it happens in the direct RF generation method, while baseband samples are supplied at a much slower speed, thus optimising memory usage and opening the door at extremely high play-back times. Digital oscillators allow changing generation parameters without stopping operation. As an example, carrier frequency can change in real-time under programmatic control to implement complex frequency hopping signals (i.e. Bluetooth). Some impairments, such as quadrature error and imbalance or phase noise, can be applied without the need to recalculate the baseband signals as it is the case with regular direct RF generation. As digital oscillators are independent, two different ASK signals at different carriers frequencies can be also generated with a single DAC. This generation mode can be used together with the “doublet” DAC conversion mode and inverting the phase of one of the oscillators can be used to reverse the spectrum as required by second Nyquist band operation. Digital oscillators do not have to be synchronous with the modulating signals (unlike carriers in direct RF generation) so the repetition of the same signal segment will not result in a repetitive quantization noise pattern.

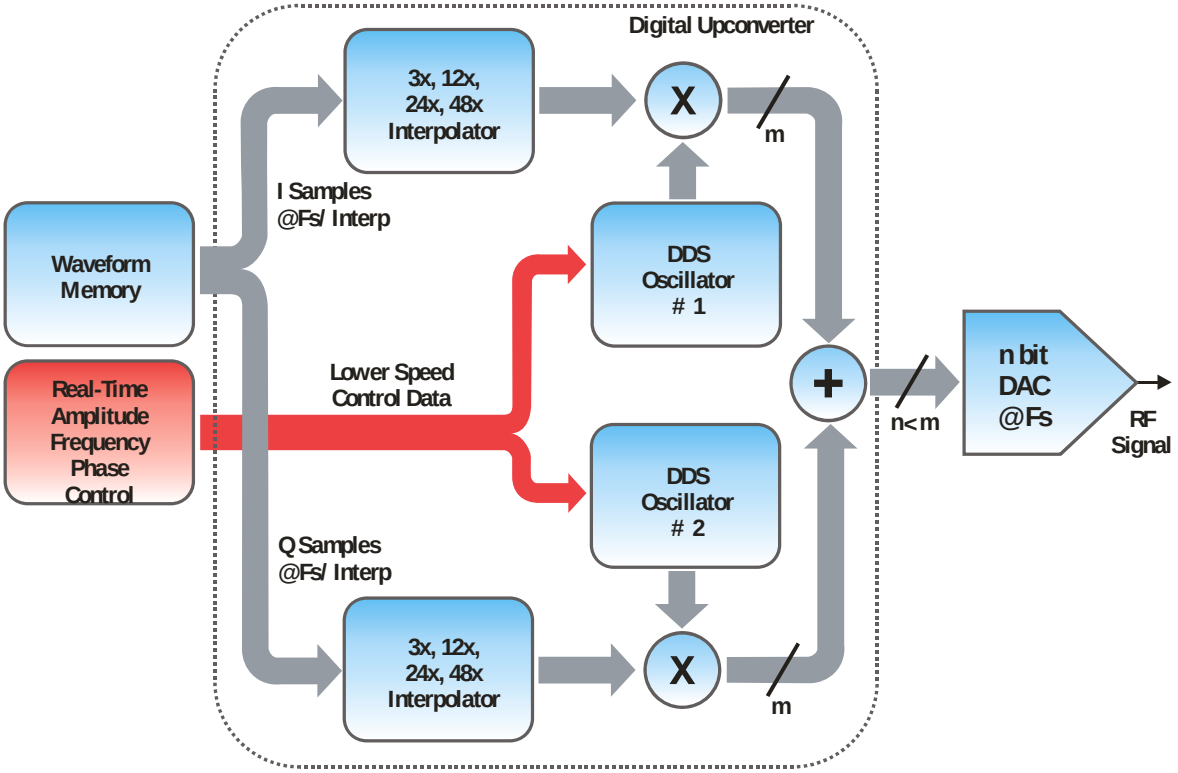


Figure 90: Block diagram for the full-digital up-converter and modulator architecture as implemented in the M8190A' DACs. It saves waveform memory as baseband signals samples are supplied at a much lower speed. The real-time control of the internal digital oscillators permits the implementation of complex signals behaviors such as hopping frequency, phase noise, etc. Integer arithmetic is implemented in a higher resolution than that of the DAC to avoid losing dynamic range.

6.5 Wideband Signal Generation, Calibration, and Correction

Modulation accuracy depends on a variety of factors. A very important component is the accuracy of the signal's spectral shape (magnitude and phase). Numerical synthesis of baseband and IF/RF signals allow for an almost perfect definition of the signals. However, the digital-to-analog conversion process and the filtering, amplification, transmission, and, eventually, modulation or up-conversion of the analog signal will introduce linear and non-linear distortions. Even ideal DACs result in the application of a sinc(f) low-pass filter. Reconstruction filters and amplifiers also act as low-pass filters.

For very small bandwidths, frequency response may be considered almost constant but, in general, the combined frequency response of the AWG and all the associated devices (cabling, external filters, quadrature modulators) will lineally distort the signal worsening modulation quality (typically expressed through the EVM, or Error Vector Magnitude, parameter). It is evident that the effects of the lack of flatness in the frequency response grow with the signal bandwidth and the signal proximity to the Nyquist frequency. Fortunately, AWGs can generate undistorted or distorted signals. If the overall frequency response within the signal's band is known then a reversed filter can be applied to the signal during calculations so a corrected, undistorted signal can be finally obtained.

Proper signal correction requires system characterization through a calibration process (Figure 91). Calibration is made possible through the usage of accurate analysis devices. These include spectrum analysers and oscilloscopes. There are several ways to characterize the generator system frequency response although there is an important difference between generating a modulated carrier (or carriers) using a two-channel arb along with an external quadrature modulator and using a single channel AWG for direct RF signal generation. Direct RF generation does not result in any unwanted quadrature impairment while the usage of an external quadrature modulator will suffer of any baseband signal imbalance, delay, o carrier phase inaccuracies by generating unwanted signal images and distorting constellations. Calibration of wideband quadrature modulation systems where the AWG supplies two baseband signals may require two steps: one for quadrature impairment characterization and another for frequency response characterization. Direct RF generation systems only require a frequency response characterization process.

Quadrature impairment characterization for a given frequency (positive or negative) may be performed by generating a single SSB carrier. This can be accomplished by generating two sinusoidal waveforms at the desired Fm frequency with either 90 or –90 degrees phase (for positive and negative frequency characterization) between them. It is possible to establish the amplitude of the unwanted image using a spectrum analyser. Although the amplitude of the wanted carrier is also affected, the effects are insignificant for small quadrature impairments. Complete characterization requires the separation of the effects caused by quadrature imbalance (amplitude difference between the I and Q signals) and those of quadrature error (those caused by relative carrier phase errors). One way to establish those two components is to generate the same level of impairment separately on both components (Figure 92).

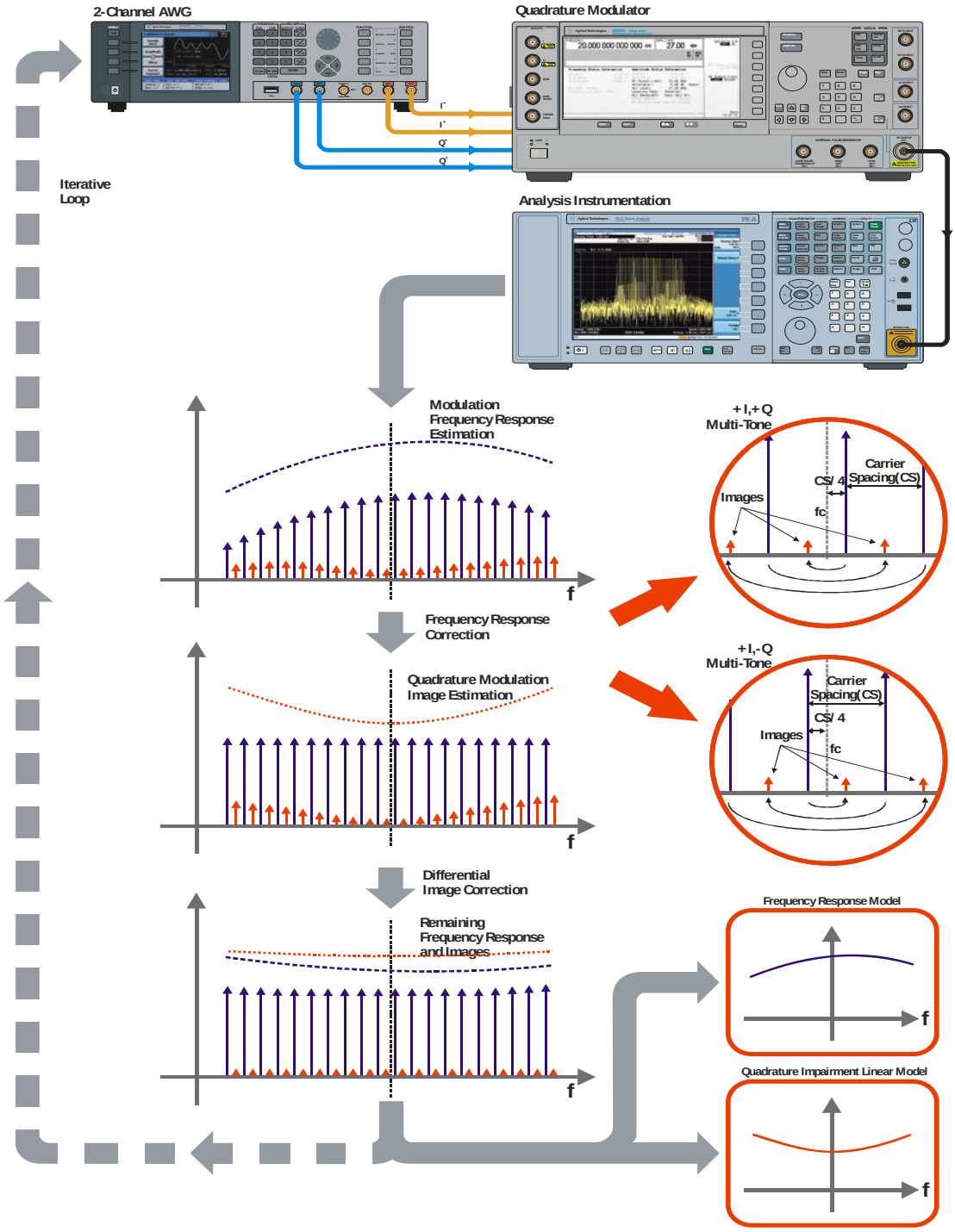


Figure 91: Highest quality wideband signal generation using an external quadrature modulator is only possible through careful generation system calibration. This process must account for both the overall system frequency response and any quadrature impairment. A shifted multi-tone signal may be used for calibration. Wanted modulated tones and unwanted images caused by quadrature impairments will be located in different locations. Inverting one of the baseband will swap both types of carriers increasing the frequency resolution of the analysis. Iterating the process will further improve results.

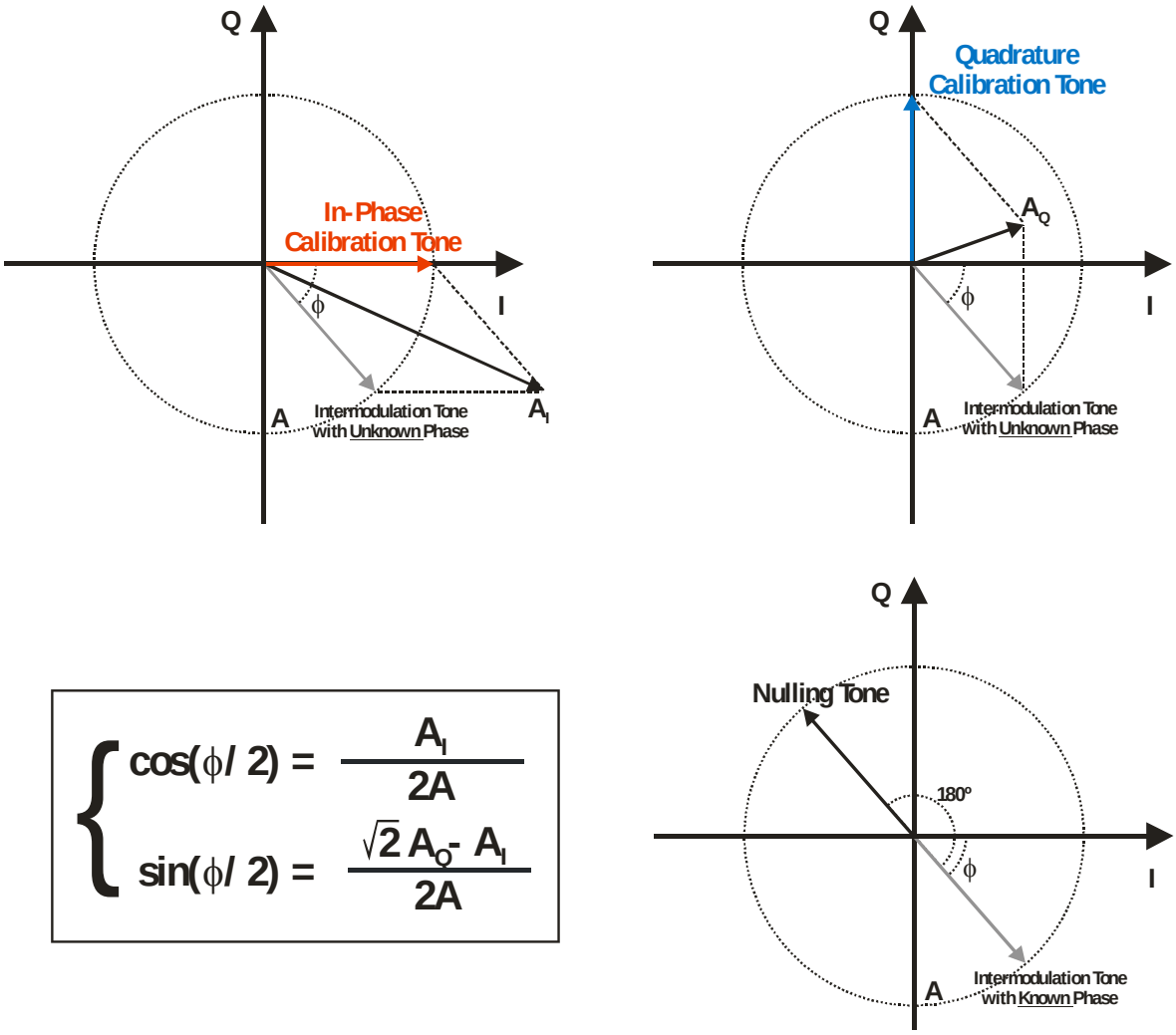


Figure 92: The nulling-tone technique can be used to find out the phase of any unwanted synchronous tone produced by a generation system. In a generation system using an external quadrature modulators, image tones caused by quadrature impairments in the system are caused by a combination of I/Q amplitude imbalance and carrier phase error. By injecting sequentially a known level of each impairment and analysing the resulting amplitudes, it is possible to obtain the phase of the original tone. Calculations are simplified and accuracy optimized if the amplitude of the calibration impairment is the same than the original tone. This procedure can be applied simultaneously to multiple frequencies when applied to a multi-tone calibration signal.

First, the same quadrature imbalance level is added and a new image amplitude reading is obtained; then, the same quadrature error impairment level is added and a different image amplitude reading is obtained. The relative phase of the image signal can then be obtained by solving a simple trigonometrically problem (Figure 92).

Once the amplitude and phase of the image has been established, a differential correction in amplitude and phase can be applied to the I and Q SSB carriers to get rid of the image. It is possible to characterize the quadrature impairments over the complete band of interest by repeating this process for a series of frequencies. I to Q signal delays will show up as linear phase component in the overall phase response. A differential correction filter must be applied to any I and Q signals to get rid of any impairment due to quadrature error and imbalance.

Frequency response can be characterized following a similar process. In this case only the wanted carrier is analysed. However, some equipment, such as swept spectrum analysers, are not capable of characterizing the relative phase for each frequency so they can only supply partial correction information. Even with this limited information, modulation accuracy can be improved by applying an amplitude-only correction to the signal. Both quadrature impairment and frequency response characterization can be performed simultaneously. It is also possible to do it over multiple frequencies all over the band of interest at once. To do so, a frequency-shifted comb of carriers can be used. Given a carrier distance Fd, all the carriers must be shifted in either direction by Fd/4. In this way, images will show-up centered in the gap between consecutive carriers. Reversing one of the components (either I or Q) will swap the location of the direct carriers and the images so, in fact, the analysis resolution will be, Fd/2 and correction data for quadrature impairments and frequency response will be obtained for all frequencies in two steps. Analysis results can be improved by iterating the same calibration procedure until the desired levels of image suppression and response flatness have been reached.

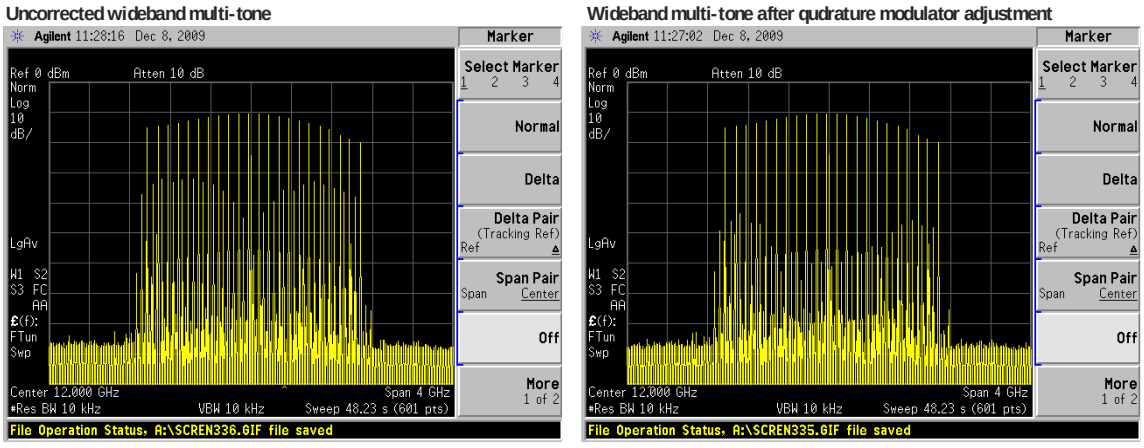


Figure 93: Adjusting the relative amplitude of the I and Q baseband signals and the relative carrier phase in a quadrature modulator results in some image level reduction. In this 2.5 GHz bandwidth multi-tone, some images (the smaller tones) have been reduced by more than 30 dB but others have been reduced by a mere 10 dB. Wideband signal correction requires obtaining correction factors as a function of frequency. The unchanged line in the center is caused by carrier leakage and can be minimized by adjusting the DC level of both baseband components.

An alternative method for frequency response characterization is based on the application of an adaptive equalizer to a calibration (Figure 94). High performance VSA software, such as the Keysight 89601B software, can automatically apply a complex equalization filter so that it includes information about magnitude and phase, which optimises modulation quality at the output. VSA software can be applied to spectrum analysers, logic analysers, and scopes. In a linear system, the equalization filter can be used directly to correct any signal, no matter its modulation scheme. To use this method, it is enough to apply any modulated signal using a modulation scheme supported by the analysis software and modulation parameters (i.e. baud rate) so the signal completely covers the band of interest. Adaptive equalizers, though, may be very sensitive to some linear and non-linear impairments in the signal, such as quadrature errors and imbalance, and sometimes they do not converge to a valid solution. Again, applying this calibration method iteratively (where the equalization filter is applied to an already corrected signal) can improve results.

Non-linear impairments can show-up in many test situations, especially when RF power amplifiers are involved. Pre-distortion techniques are widely used in real transmitters to improve modulation quality and ACPR performance. AWGs provide a very convenient environment to apply these techniques as non-linear corrections can be applied to the signal mathematically and no especial hardware or real-time processing software is required. The most convenient way to characterize the non-linear behaviour of some system is by obtaining the AM-AM and the AM-PM responses. These can be obtained through direct measurements (i.e. using a VNA, or Vector Network Analyzer, with built-in X-parameter characterization) or through system simulation. Some VSAs can also provide the AM-AM and AM-PM responses by analysing some known signals before and after the non-linear distorting block. Non-linear response data (typically an adjusted polynomial) can be used directly to emulate a non-linear device.

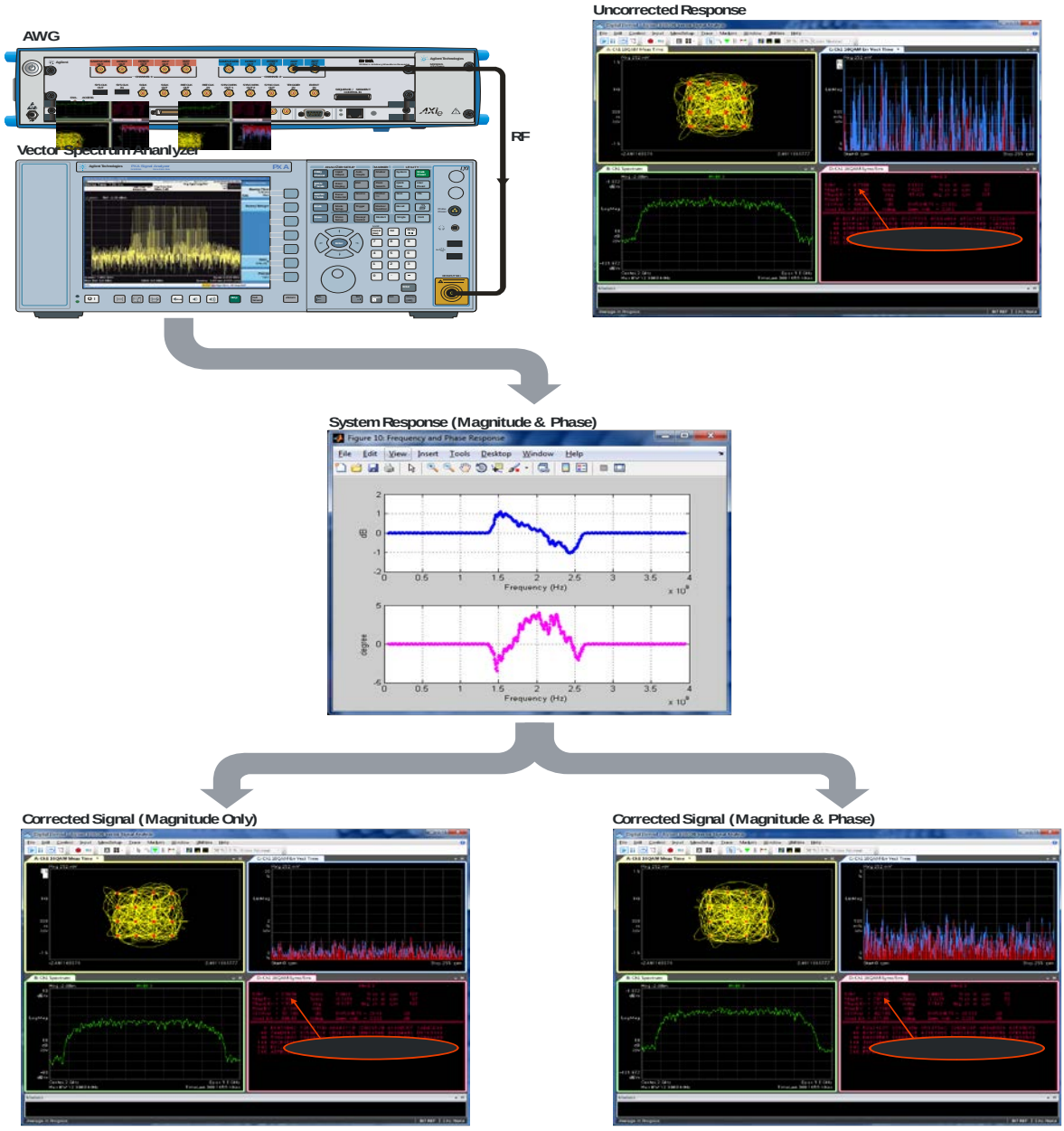


Figure 94: Direct RF signal generation does not suffer of quadrature impairments. Nevertheless, correcting the signal frequency response is very important to obtain a good modulation quality. Here a 1 GBaud 16QAM signal is being generated. Without any correction EVM for this signal is close to 5%. Applying just amplitude response correction, EVM falls to 2.5%. Applying both amplitude and phase correction, EVM is further reduced to about 1%. Here, correction data is obtained from the adaptive equalizer set by the analysis instrument. This is a very straightforward method to obtain magnitude and phase response information and the correction data obtained can be applied to any signal, no matter its modulation scheme. However, adaptive equalization does not typically correct quadrature impairments so it may be difficult to apply to RF generation systems using an external quadrature modulator.

6.6 Generation of Multi-Carrier Signals

Generating multiple carriers over a wide frequency range is one of the unique capabilities of AWGs when compared to traditional vector signal generators. Multiple carrier signals, modulated or not, are useful in a variety of test situations:

- Multi-tone signals are traditionally used to test non-linearity or to establish the frequency response of RF systems and components. NPR (Noise Power Ratio) testing is a particular case where one or more bands are free of tones so intermodulation products in that band can be characterized.
- Interference testing. Generation of additional signals in adjacent channels or interference within the signal bandwidth without the need for additional equipment.
- Bandwidth-limited, time-dependent, and/or spectrum shaped noise generation.
- Realistic traffic load generation. Power amplifiers in cell towers often handle more than one channel while CATV networks support tens of channels simultaneously. Realistic test of these components requires feeding all the required signals simultaneously. Direct RF generation with a generator opens the door to generate all of them with one single instrument.

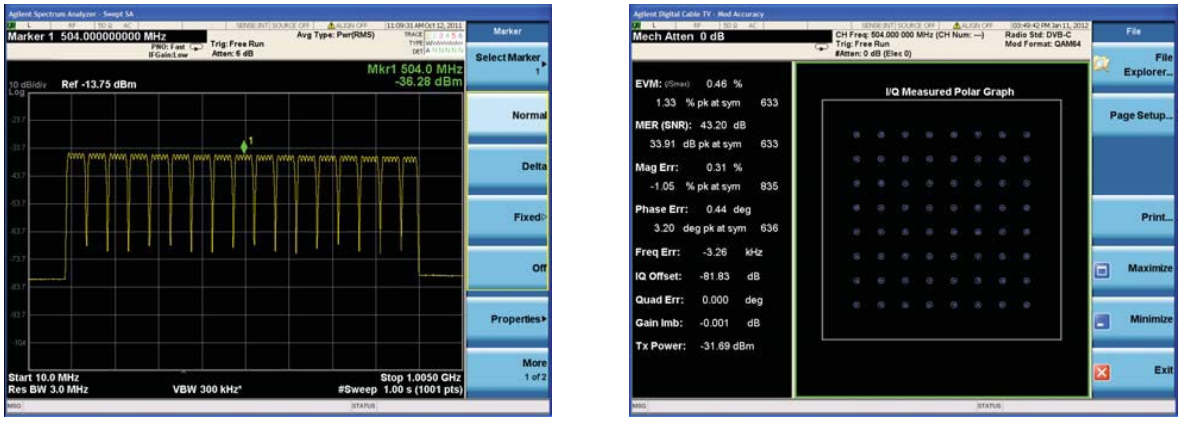


Figure 95: 84 64QAM DVB-C signals being generated simultaneously by an AWG (A Keysight M8190A). Notice the excellent available dynamic range and the lack of spurs (left) and the modulation quality obtained (right) with EVM lower than 0.5%.

Often the different signals are similar (i.e. CATV, see Figure 95) but sometimes may be very different (i.e. a Bluetooth signal interfering a WiFi network). An important concern when generating multiple signals is that the available dynamic range must be shared by all of them. It is, then, very important preserving such a precious resource. A good indication on how well dynamic range is used is by computing the peak-to-average power ratio (or PAPR). This parameter can be either directly computed or it may be extracted from a CCDF (Complementary Cumulative Distribution Function) analysis. Preserving dynamic range requires that all carriers are as uncorrelated as possible. In that way, at any moment in time each carrier will have a different amplitude and high peaks will only occur when many (or all of them) have high amplitude and similar phases simultaneously. When multiple uncorrelated signals are added together the combined amplitude probability distribution functions is close to Gaussian. When generating multiple similar modulated carriers (i.e. to emulate a CATV head-end) it is important to minimize signal correlation as well. This can be accomplished by using different data streams for each carrier. One way to obtain uncorrelated data is by re-using sufficiently delayed versions of the same data. As an example, 10 QAM16 modulated signals can improve power by more than 6 dB just by using uncorrelated data in all the carriers when compared to use the data in all of them (Figure 96).

Multi-tone (or Frequency Comb) signals consist of a set of equally spaced, equal amplitude non-modulated carriers. They are very useful to obtain the frequency response of a system under test. Initial phase is the only available degree of freedom to make them uncorrelated. For a large number of carriers, randomizing the phase for each carrier offers good results in terms of PAPR. As an example, a 100 tones signal can reduce PAPR by more than 13 dB just by randomising the phase (Figure 96). In this context random phase does not mean “unknown” phase. Relative phase for each carrier must be known in order to obtain the phase vs. frequency response. Multi-tone flatness is influenced by the DAC’s zeroth order hold response, the AWG’s analog roll-off, and any external processing block (cabling, amplifiers, filters, modulators) between the generator and the DUT. Calibration is required to generate constant amplitude tones.

OFDM (Orthogonal Frequency Division Multiplex) signals are a special case of multi-carrier signals. Modulation for each carrier is defined in the frequency domain for each symbol unlike single-carrier modulated signals (i.e. QPSK or QAM) where modulation is defined in the time-domain. Some standards, such as WiMedia or WiGig, extend this strategy to signal bandwidths in the GHz range. Correction data is often obtained in the frequency domain as a complex (magnitude-phase) filter response. Applying the correction to time-domain data requires obtaining the corresponding complex impulse response (through the inverse Fourier transform) and applying convolution to the signal to be corrected (Figure 97a). Although this method can also be applied to OFDM signals, it is much simpler and efficient to apply the correction directly to the frequency domain data before obtaining the time-domain signal through IFFT (Figure 97b).

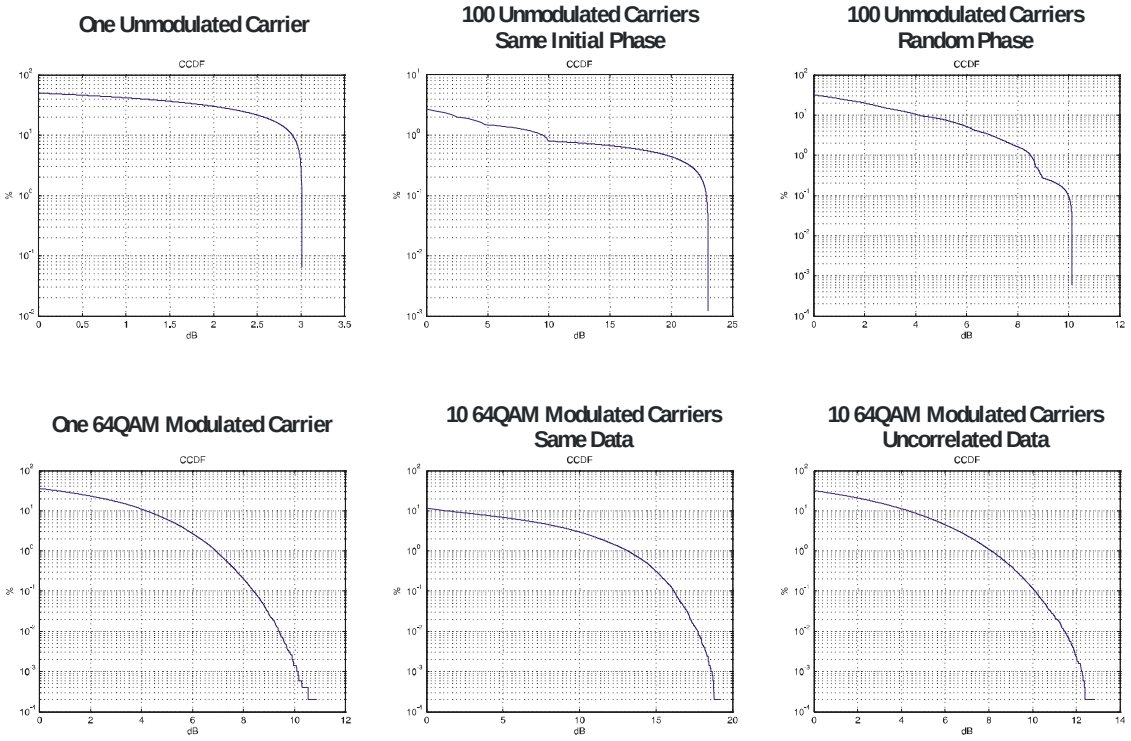


Figure 96: These CCDF curves are calculated for a different number of similar carriers, modulated or not. The 0 dB reference corresponds to the combined average power for all the carriers. Available power, signal usability, and dynamic range dramatically improve by generating uncorrelated carriers. For multi-tone signals, phase is the only variable. For modulated signals, uncorrelated data results in uncorrelated carriers.

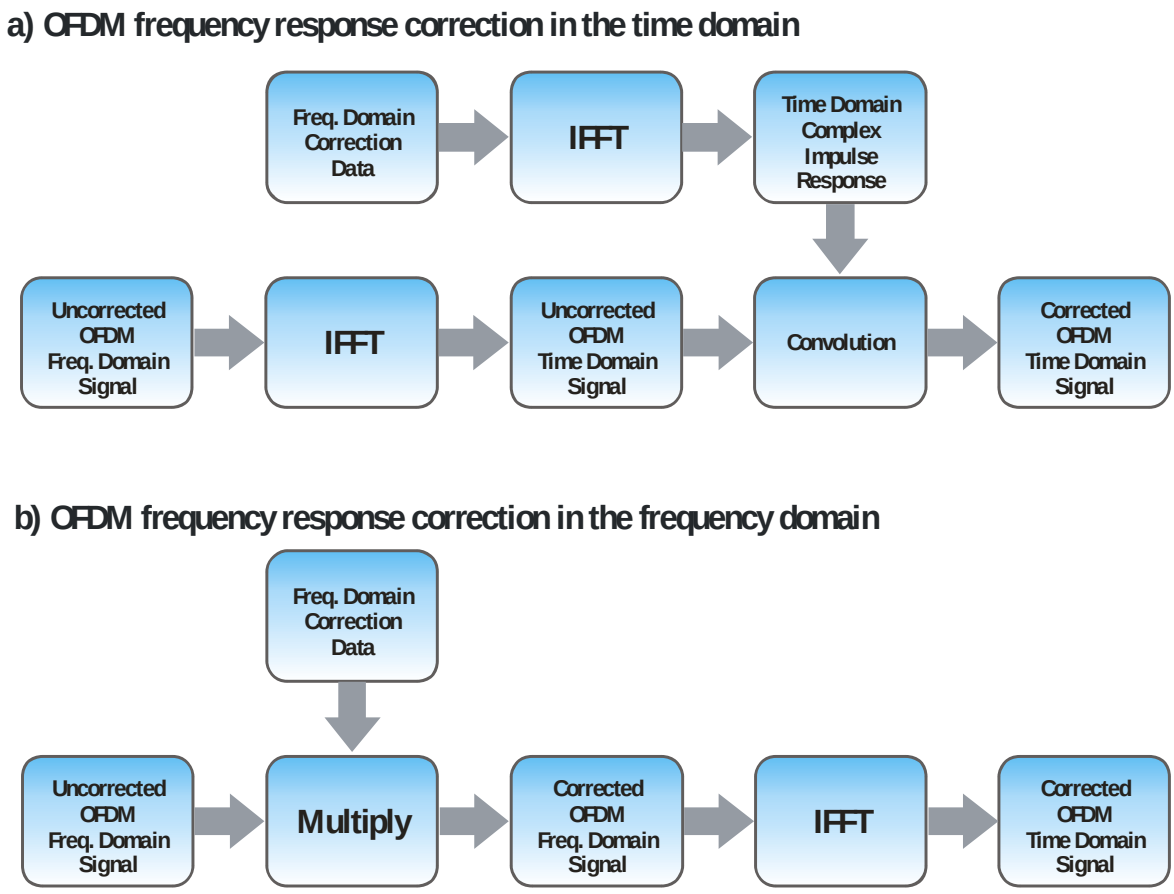


Figure 97: OFDM is a particular case of multi-carrier signal. Correction of OFDM signals may be done using the same time-domain techniques applied to single-carrier wideband signals (a) but doing it in the frequency domain (b) is more efficient and accurate. The same method works better with multi-tone signals and for the generation of multiple narrow-band signals where sometimes just amplitude correction is necessary

The correction data obtained will be more accurate if the calibration signal contains carriers exactly at the same frequencies than the OFDM carriers. OFDM analysis software packages also provide equalization data that can be used for correction purposes. However, equalization for many OFDM standards depends on data for just a subset of all the carriers, known as pilots, and correction for non-pilot carriers is interpolated in the time and/or frequency domains, so accuracy is limited.

NPR (Noise Power Ratio) testing is based in the usage of an especially configured multi-tone signal with an empty band (or notch) in it (Figure 98). Feeding this signal to a non-linear device will generate multiple intermodulation products that will be visible as a lower SFDR in the notch band. The sensitivity of this test depends on the initial SFDR performance in the original test signal. AWGs can easily generate such signals. All the tones will be synchronous as they are derived from the same sampling clock. However, any non-linearity in the AWG or attached devices can result in unwanted intermodulation products. As all the carriers generating intermodulation products are synchronous and spacing between all of them is exactly the same, all the intermodulation products will be located at known frequency locations and their relative phase will be constant over time (in other words, IMD products are synchronous too). Correcting the non-linear behaviour of the generation system after characterizing can reduce the unwanted intermodulation tones within the notch band. This characterization may be cumbersome and inaccurate. An alternative methodology is generating “anti-tones” (or nulling tones). Once the location and amplitude of the in-band IMD products are found using a spectrum analyzer, their respective phase may be established by adding calibration tones of the same amplitude and orthogonal phase sequentially. Measuring the combined amplitude for both calibration signal phases allows obtaining a unique phase for each IMD tone. A corrected signal can be re-calculated by adding an inverted phase tone for each IMD tone so it will be partially or totally cancelled. This process may be iterated to further reduce the unwanted tone amplitude. Given the non-linear nature of the process, calibration must be applied to all the IMD tones simultaneously to improve convergence. Generating this kind of signal with an external quadrature modulator adds some issues as a symmetrical location of carriers will result in IMD and image tones sharing the same exact location in the frequency domain. Carrier feed-through can be also confused with an IMD product. Shifting the whole test signal in the frequency domain beyond the resolution bandwidth of the analyser will separate IMD products from carrier feed-through and images generated by quadrature modulation impairments. The previously described quadrature impairment and flatness calibration process may be used to remove images while nulling-tones may be used to remove unwanted IMD tones.

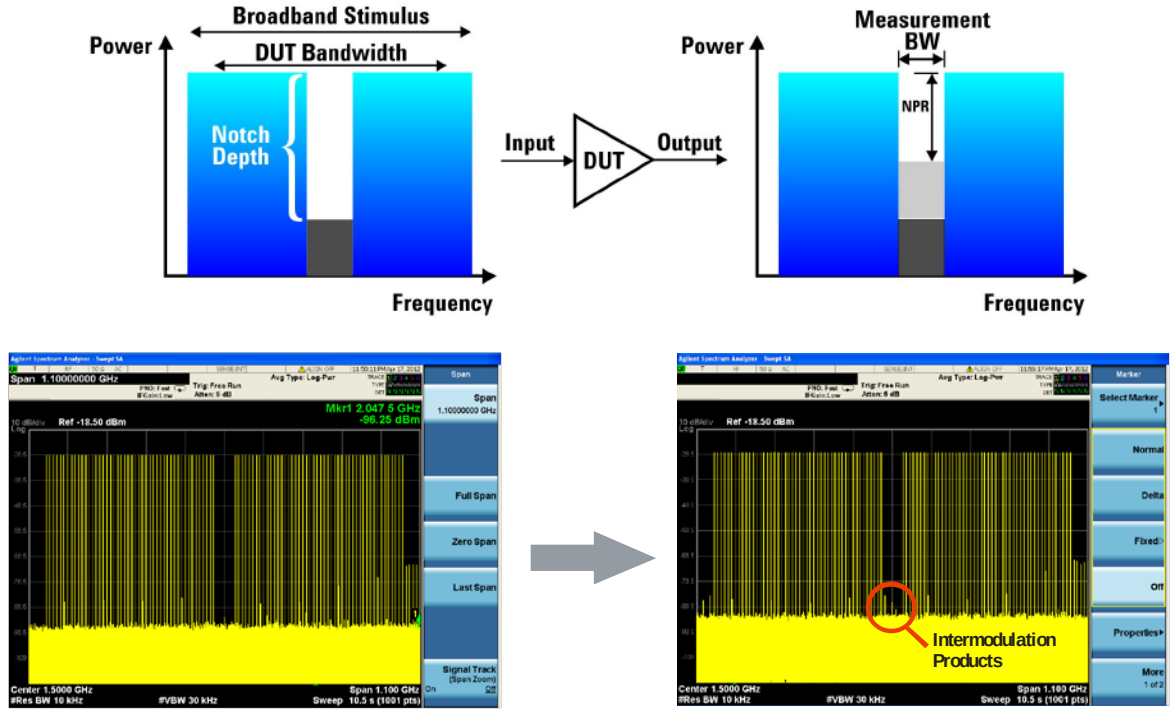


Figure 98: Noise Power Ratio (NPR) test relies in the spurious-free dynamic range in a user-defined notch. Intermodulation and noise produced by the DUT will show up as a reduction in the depth of the notch. Accuracy and dynamic range of the NPR measurements depend on the SFDR in the notch in the input signal. Proper signal correction may be necessary to reach an acceptable performance though the application of nulling tones to cancel the AWG generated intermodulation products.

Limited bandwidth Gaussian noise may be used to emulate noisy transmission channels. Most considerations on noise generation made in chapter 5 are also valid for RF signal generation. However, improving the available dynamic range in the AWG through noise or overall signal clipping is unacceptable as this is a highly non-linear process that creates unwanted frequency components in the form of spectral growth. As clipping distortions are confined in time, the statistical effects will greatly differ from those caused by genuine noise and they will affect more to high power sections of the signal. Typically, bandwidth limited noise may be generated by calculating noise samples for the two orthogonal components (I and Q) at the noise BW rate. The final noise waveforms are calculated by applying an ideal interpolation function with BW=NBW/2 and adapting the sampling rate to the AWG sampling rate. This process results in a limited bandwidth signal no matter the values or statistical distribution of the input noise samples. It is possible then to clip the input samples to some bounded range.

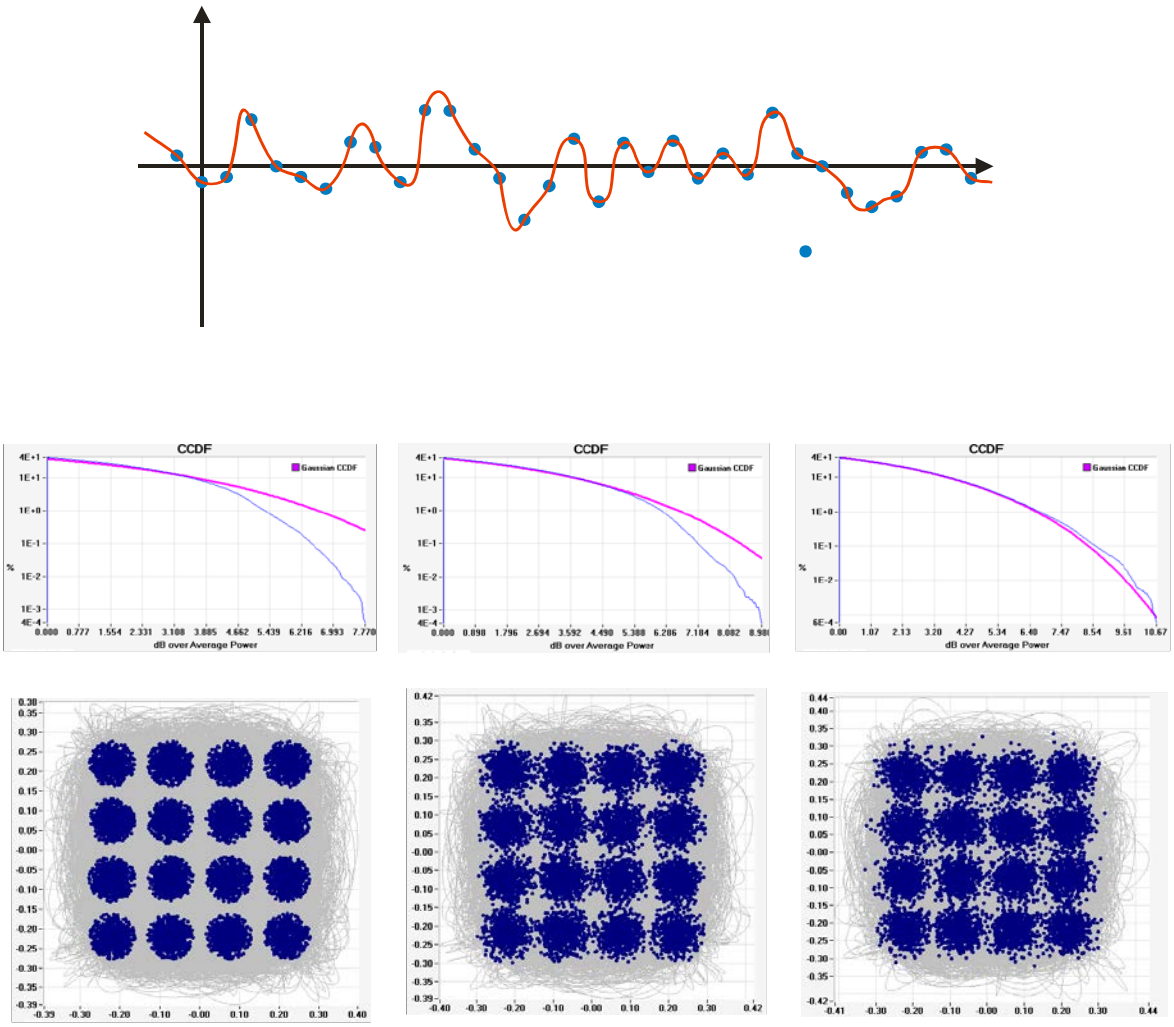
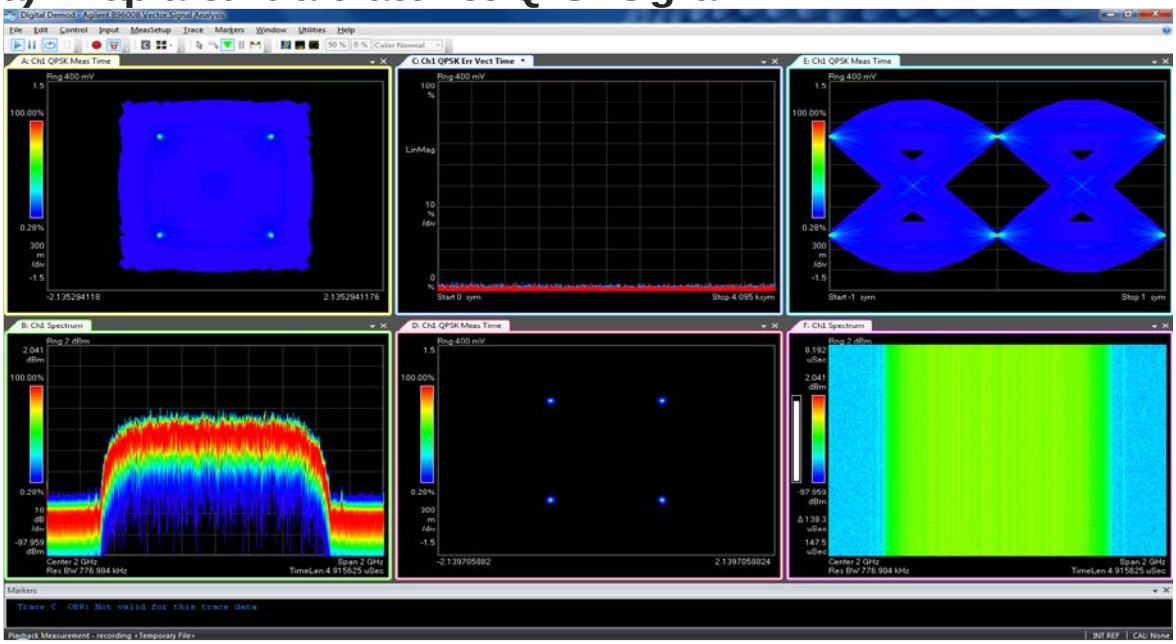


Figure 99: When synthesizing a signal with added Gaussian bandwidth-limited noise, clipping the signal to limit the effects of the noise crest factor is not an option. Clipping is a highly non-linear operation that results in signal distortion and spectral growth. Baseband noise samples, though, may be limited to some extreme values but, after applying a near-ideal bandwidth limited interpolator, the actual PAPR will grow significantly. Here, three different PAPR have been applied to the same baseband samples so the SNR is approximately the same for the 16QAM target signal. In some cases, there will be no errors caused by noise as the signal never goes beyond the decision thresholds. In this example, selecting the 12 dB PAPR will result in a very realistic noise behavior with a very little overall (signal + noise) peak-to-peak amplitude penalty.

This process will modify the statistical distribution of the signal but the bandwidth limited interpolated version of it will have peaks beyond this limit. Figure 99 shows results for three different crest factors (PAPR) applied to the input noise samples and the corresponding CCDF (and crest factor) for the bandwidth-limited interpolated noise. As it can be seen, the resulting crest factor is far bigger than that of the original noise samples. Additionally, the shape of the probability distribution function (and as a consequence, of the CCDF function) changes dramatically beyond the clipping level so error statistics will also change if the decision threshold is close or beyond this level. For direct RF generation, one AWG channel may be used to generate the RF signal while a second AWG channel may generate the bandwidth-limited noise. The noisy signal can then be obtained by adding the signals coming from both channels. In this way, there is no connection between the signal and the noise levels and there will be more dynamic range (and power) available for each signal. An alternative may be to use an external noise source and an appropriate band-pass filter. For baseband signal generation, the same noise source cannot be used for both channels as noise for the I and Q components must be uncorrelated. A way to obtain uncorrelated noise with a single source is to use a sufficiently delayed version of the noise signal for one of the channels.

a) Wrap-around artifact free QPSK signal



b) Wrap-around related artifacts in a QPSK signal

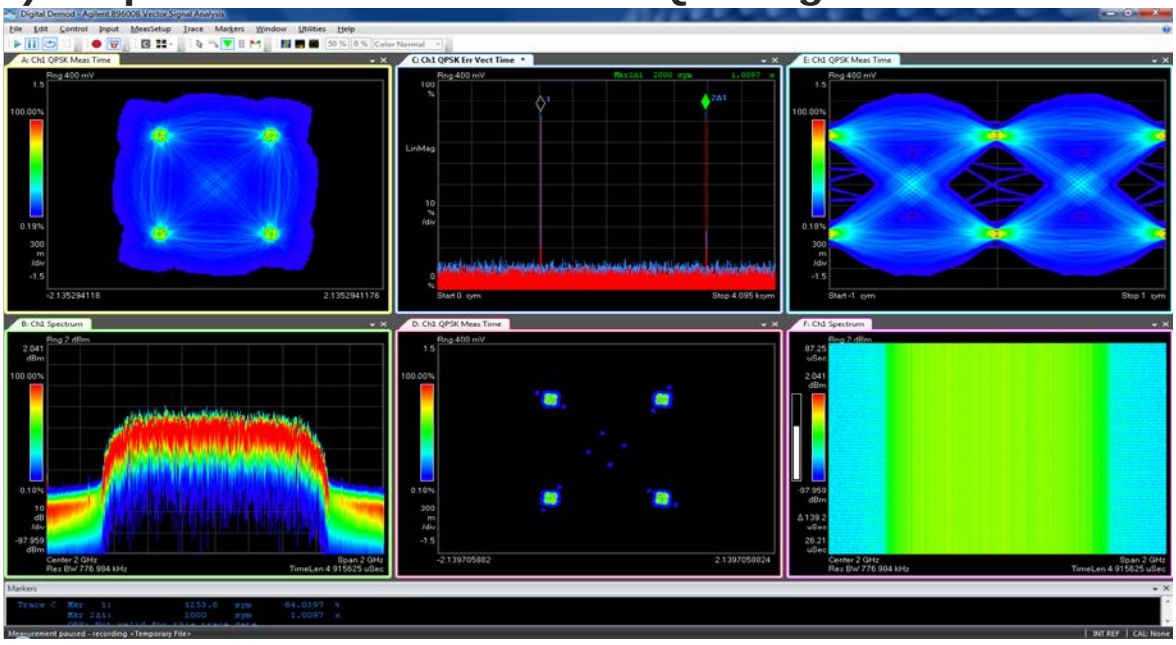


Figure 100: Incorrect wrap-around handling results in highly distorted signals when the signal is looped for continuous generation (b). Distortions show-up at all domains: time (see eye diagram), frequency (see spectral growth in the signal spectrum), and modulation (see constellation diagram). EVM spikes show up every occurrence of the signal. Properly synthesized signals (a), designed for continuous play-back, do not show any of these impairments and allow accurate measurements and glitch free operation of the DUT.

6.7 Wrap-Around Handling for RF/Wireless Signals

Wrap-around artefacts are especially damaging in the RF/Wireless signal generation context (Figure 100). Like in serial digital signals, baseband signal discontinuity results in signal inconsistencies and limits the signal usability. Direct RF signal generation adds another layer of wrap-around established by any carrier. A discontinuous carrier also results in an inconsistent signal as phase changes abruptly when any segment is repeated and phase is carrying information. Additionally, either baseband signal or carrier discontinuities also result in power being spread beyond the boundaries of the signal BW, a phenomenon known as spectral growth. This effect is specially disrupting as it compromises the capability of performing some frequency-domain measurements (i.e. the Adjacent Channel Power Ratio, or ACPR, parameter). It will also result in unwanted interferences on adjacent channels that will degrade modulation quality performance for signals located on those channels. Direct RF generation signals must meet the wrap-around handling requirements applied to any serial-digital signal (see chapter 5) and, additionally, there must be an integer number of carrier cycles within any waveform segment. For QAM kind of signals, the latest condition is met if any carrier frequency is a multiple of the signal repetition rate. As this parameter is a function of record length and sampling rate, especial care must be taken when selecting those parameters to obtain the required symbol rate, number of symbols, and carrier frequency for all the carriers in a segment with good enough accuracy while meeting the requirements regarding sample rate set ability, maximum record length, and record length granularity.

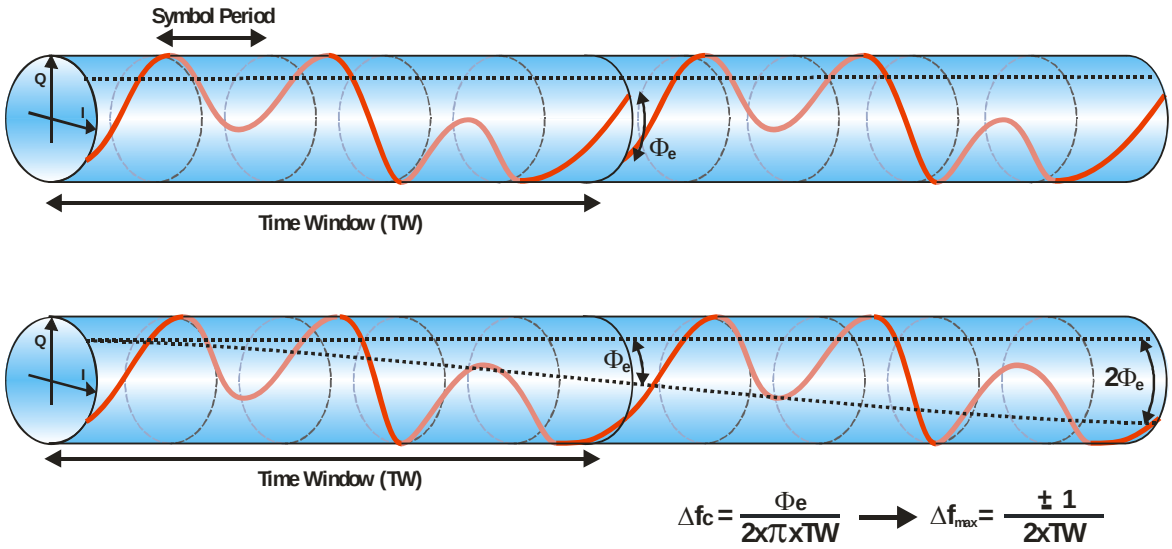


Figure 101: PSK/FSK/MSK signals require additional wrap-around handling processing as properly designed baseband signals may have different initial and final phases resulting in a discontinuous play-back when the signal is looped. A general solution for this problem consists in twisting linearly the phase over the whole waveform segment. Twisting may be implemented by changing the frequency by Δf or by rotating the corresponding I and Q components.

For PSK/FSK/MSK kind of signals, it is not enough to adjust the baseband modulating signal for wrap-around operation and setting the carrier frequency to a multiple of the signal's repetition rate. Figure 101 shows that, if the initial carrier phase is taken as a reference, the final phase can be anywhere, so there will be a phase jump of up to 180 degrees at every loop. A way to solve this issue is by applying a linear phase twist to the signal so the final and initial phase are the same. Phase twisting is equivalent to change the carrier frequency:

$$\Delta f = \Delta\phi \; / \; (2 \times TW)$$

$$\Delta f|_{max} = 1 \; / \; (2 \times TW)$$

The maximum carrier error will depend then on the time window length. PSK/FSK/MSK signals for direct RF generation can be calculated in two ways: by direct modulation of the carrier or by quadrature modulation of two orthogonal carriers. For direct carrier modulation, phase twisting must be performed by applying a correction (Δf) to the original target frequency.
Δf may be obtained by calculating the signal with the target frequency first, and then comparing the initial and the final phase. For quadrature modulation, the I and Q components are obtained using the following expressions:

$$I(t) = A \times \cos(\phi(t))$$

$$Q(t) = A \times \sin(\phi(t))$$

I(t) and Q(t) are the projections of $\phi(t)$ over the I and Q axis. Again, comparing the initial and final phases will result in a $\Delta\phi$ error. The corrected signals will be the following:

$$I'(t) = A \times \cos(\phi(t) - \Delta\phi \times t/TW)$$

$$Q'(t) = A \times \sin(\phi(t) - \Delta\phi \times t/TW)$$

Applying these formulae is equivalent to linearly rotate the signal by $-\Delta\phi$ over the whole waveform segment. When using an external quadrature modulator, re-adjusting the carrier frequency of the modulator's local oscillator can compensate the Δf error. Some differential and rotating constellation based modulation schemes can also benefit of phase twisting although applying circular convolution to the baseband signals may be much more difficult.

6.8 Coherent Optical Signal Generation Using AWGs

Digital modulation techniques are also used in coherent optical communication systems (Figure 102). They are based on the same principles but with much higher signalling speeds. Current systems use fairly simple modulation schemes (DQPSK) along with polarization division multiplex (PDM) to work up to 40 and 100Gb/s. Future links based in more complex modulation schemes such as QAM16 (and higher) or OFDM will reach speeds of hundreds of Gb/s with one single optical carrier in a 50 GHz DWDM ITU grid. Figure 102 shows a typical optical quadrature modulator. Polarization division multiplexed systems use two of these modulators, each one working with one orthogonal polarization. Although the basic modulation principles are similar, there are some important differences:

- The basic modulator component is typically a Mach-Zehnder modulator. This device can symmetrically modulate the amplitude of an optical carrier so the output will have two possible phases: 0 or 180º. However, as it works through an interferometry principle, the transfer function between the input and the output is far from linear.
- Modulation bandwidths go well beyond 10 GHz compared to less than 1-2 GHz for most wireless systems.

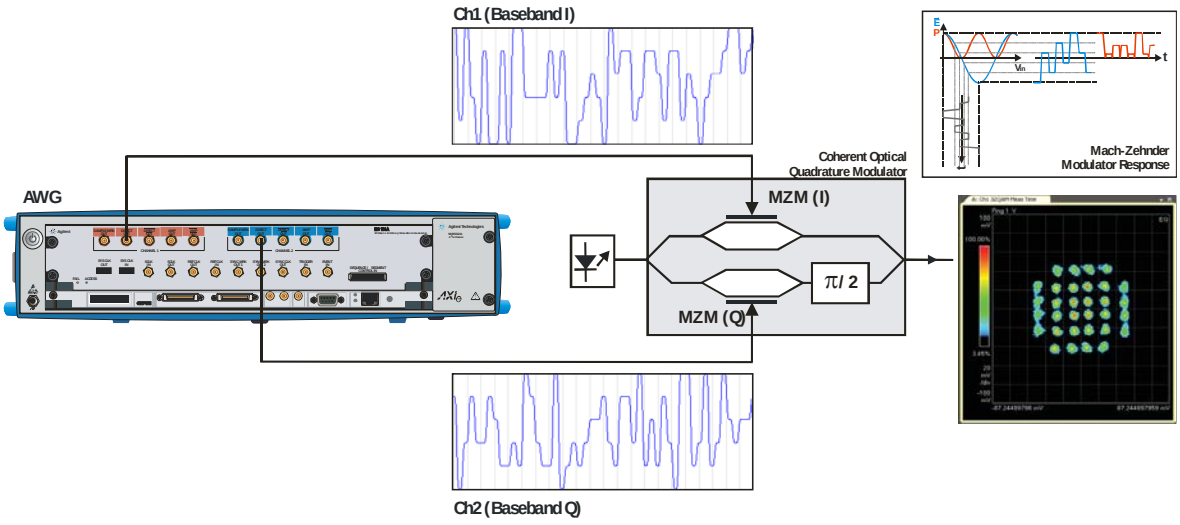


Figure 102: AWGs can be connected to coherent optical quadrature modulators. In this example a 32QAM signal is generated through a Mach- Zehnder interferometer based modulator. The modulator transfer function is not-linear so the electrical signal has been distorted previously to obtain a good quality constellation. A coherent receiver is required to demodulate this signal as traditional intensity detection cannot detect the phase of the optical carrier, just the power (P). Polarization division multiplexed (PDM) systems duplicate the number of channels and modulators as each orthogonal polarization is modulated independently.

- Usage of polarization division multiplex is very rare in wireless systems although the basic principle (transmitting two or more messages in the same band though the same path) is similar to the widely spread MIMO (Multiple-Input, Multiple Output) wireless transmission schemes. As long haul fiber optics links do not keep polarization over time, receivers must be capable of reconstructing the original SOP (State of Polarization) through adaptive algorithms, so the original signals may be recovered.

AWGs are a very convenient baseband test signal source as any modulation scheme may be implemented without the need of especial hardware, they can emulate linear and non-linear distortions or signals may be pre-corrected for modulation device linearization. Bandwidth and the number of synchronous channels are the main challenges. A complete generation system would require 4 AWG channels (Ix, Qx, Iy, Qy) running, at least, at the signalling rate of the link to be emulated. A significantly higher sampling rate may be necessary to implement proper baseband filtering or a good-enough device linearization. The timing accuracy (skew, jitter) requirements are much higher than in typical wireless systems. Four independent channels also allow for quadrature impairment handling and SOP emulation so coherent receivers and algorithms can be properly tested over a wide range of controlled, repeatable conditions. SOP changes depend on mechanical and thermal conditions of the fiber so they change quite slowly. Realistic testing of SOP tracking algorithms requires stimuli capable of changing at a sufficiently low speed. Maximum record length is the key specification as it fixes the largest available time window and, consequently, the lowest frequency at what any signal parameter can change. As an example, the 2 GSamples waveform memory available in the M8190A results in SPO changes as low as 6 Hz at the maximum sample rate (12 GSa/s).

Cost is another factor when considering 4-channel ultra-high speed arbitrary generation systems. Although full polarization division multiplex emulation requires 4 independent AWG channels, basic testing may be performed using only one channel. To do so, the same channel is reused after delaying it by an integer number of symbol periods (Figure 103).

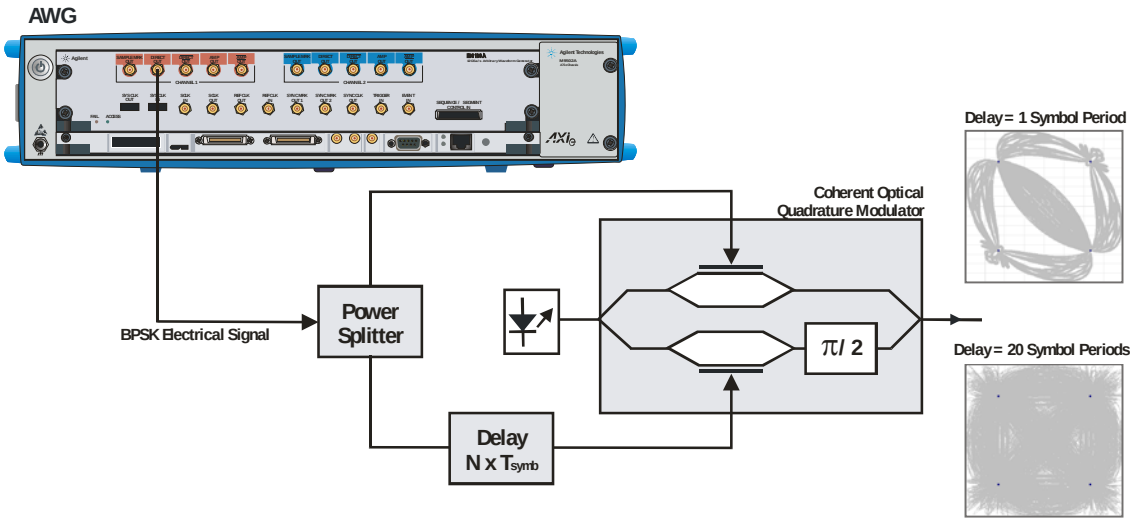


Figure 103: It is possible to generate QPSK/QAM signals using a single channel AWG by re-using the in-phase signal for the quadrature input after delaying it by an integer number of symbol times. Even with random data, delay must be long enough to avoid the memory effects that result in anomalous signal statistics. Here, a QPSK constellation made of combining signals with just one symbol delay shows a distorted phase diagram. The same BPSK sequence delayed by 20 symbols results in a statistically valid behavior.

The original and delayed signals must be as uncorrelated as possible for proper emulation. Pseudo-Random Binary Sequences show very low self-correlation so they are ideal as data sequences. Another source of correlation is inter-symbolic interference so delay should be long enough to make sure that any remaining influence of a given symbol with the same symbol delayed is negligible.

7 References

Chapter 1

Hewlett-Packard Journal. April 1988

Chapter 2

- Steven W. Smith. The Scientist and Engineer's Guide to Digital Processing. www.dspguide.com.1997
- Ken Poulton et al. Bandwidth and Bits: New AWG Design Achieves Both. Autotescon 2004
- Bob Jewett et al. A 1.2GS/s 15b DAC for Precision Signal Generation. ISSCC2005
- Bob Jewett et al. A 7.2 GSa/s or 12 GSa/s, 12-bit DAC in a 165GHz ft BiCMOS Process
- Keysight Application Note 5990-7460EN. Comparing Function Generator Performance: Direct Digital Synthesis Versus Point-by-Point Technology. 2011
- Joan Mercade. DAC Interleaving in Ultra-High-Speed Arbs. Evaluation Engineering, December 2009
- US Patent 6,812,878 B1, Jewett et al. Per-Element Resampling for a Digital-to-Analog Converter

Chapter 3

- Joan Mercade. Maximize a Waveform Generator's Memory. Test & Measurement World. May 2011
- Keysight Application Note 5989-2920EN. Trigger and Marker Interfacing on the N6030A. 2005
- Keysight Application Note 5989-2921EN. Triggering the N6030A Arbitrary Waveform Generator. 2005
- Keysight Application Note 5989-3173EN. Using Generate Scenario Mode in the N7509A. 2005
- Keysight Application Note 5990-3467EN. Synchronizong Multiple Function Generators to Produce Phase-Related Signals. 2009
- Keysight Application Note 5990-5965EN. Understanding Sequence Run and Sequence Advance Modes. 2010
- Keysight Application Note 5990-9360EN. Multi-Instrument Synchronization with the 81180B. 2011
- US Patent Application 2011/0109349, Dippon et al. Waveform Generation From an Input Data Stream

Chapter 4

- Joan Mercade. Unwrapping Wireless Signals. Test & Measurement World. December 2006

Chapter 5

- Keysight Application Note 5988-4797EN. Low-Voltage Differential Signaling. 2002
- Keysight Application Note 5988-5634EN. An Introduction to Multiport and Balanced Device Measurements. 2002
- Keysight Application Note 5988-5635EN. Concepts in Balanced Device Measurements. 2002
- Keysight Application Note 5988-9109EN. Measuring Jitter in Digital Systems. 2008
- Keysight Application Note 5988-9740EN. Finding Sources of Jitter with Real-Time Jitter Analysis. 2008
- Keysight Application Note 5989-1650EN. Advanced Jitter Generation and Analysis up to 3.35 Gbps with the Keysight 81134A Pulse Pattern Generator & 86100C Infiniium Wide-Bandwidth DCA-J
  Oscilloscope. 2004
- Keysight Application Note 5989-3776EN. Analyzing Jitter Using the Keysight EZJIT Plus Software. 2008
- Keysight White Paper 5989-4518EN. Advanced Measurement and Modeling of Differential Devices. 2006
- Keysight Application Note 5989-5765EN. Keysight Signal Integrity Analysis Series. Part 3: The ABCs of De-Embedding. 2007
- Keysight Application Note 5989-8094EN. Flexible Signal Conditioning with the Help of the Keysight 81134A Pulse Pattern. 2008
- Keysight Application Note 5990-3585EN. Using Equalization Techniques on Your Infiniium 90000A Series Oscilloscope. 2009
- Keysight Application Note 5990-5897EN. Creating Differential Signals with a Two-Channel Arbitrary Waveform Generator. 2010
- Keysight Application Note 5991-0168EN. Crossing the Digital-Analog Divide. 2008

Chapter 6

- Joan Mercade. Ruling the Waves. Evaluation Engineering. July 2004
- Keysight Application Note 5965-7160E. Digital Modulation in Communications Systems-An Introduction. 2001
- Keysight Application Note 5952-8898E. The Fundamentals of Signal Analysis. 2000
- Keysight Application Note 5968-6875E. Characterizing Digitally Modulated Signals with CCDF Curves. 2000
- Keysight Application Note 5989-4138EN. The ABC's of Arbitrary Waveform Generation. 2005
- Keysight Application Note 5989-5280EN. Keysight Technologies Solutions for MB-OFDM Ultra-wideband. 2008
- Keysight Application Note 5990-7451EN. Keysight Vector Signal Analysis Basics. 2011
- Keysight Application Note 5990-7725EN. Creating a Complete and Flexible Solution for WiGig Testing. 2011
- Keysight Application Note 5990-8349EN. Ultra-Wideband Radar System Design. 2011
- Keysight Application Note 5990-8365EN. Testing Transmitters and Receivers with Pulse Ultra-Wideband Signals. 2011
- Keysight Application Note 5990-8430EN. Creating a Radar Threat Simulator and Receiver Calibrator with Precise Angle of Arrival. 2011
- Keysight Application Note 5990-8883EN. Wideband Digital Pre-Distortion with Keysight SystemVue and PXI Modular Instruments. 2011
- Keysight Application Note 5990-9558EN. Creating a Solution for Testing Frequency-Hopping Spread Spectrum Devices. 2011
- Keysight Application Note 5990-9871EN. Testing Broadband CATV Amplifiers with True Wideband Channel Rasters. 2012
- Keysight White Paper 5990-3748EN. Metrology of Advanced Optical Modulation Formats. 2009
- Keysight Application Note 5990-9972EN. Testing Next-Generation Optical Systems with Simulated OFDM Signals. 2012

Edition 4.0, June 2015
M8190-91050
www.keysight.com