S-parameters Without Tears

Understand this critical frequency-domain measurement and its interpretations

By Colin Warwick and Fangyi Rao, Agilent Technologies

Richard W. Anderson opens his classic application note on scattering parameters (s-parameters) with the observation:

“Linear networks […] can be completely characterized by parameters measured at the network terminals (ports) without regard to the contents of the networks. Once the parameters of a network have been determined, its behavior in any external environment can be predicted…”

He goes on to observe that at low frequencies it’s trivial to create opens and shorts, and measure impedance (Z) or admittance (Y) parameters, but at high frequencies, opens and shorts become problematic. In that case, s-parameters are easier to measure because the ports are connected to low-loss transmission lines whose characteristic impedance is equal to the reference impedance of the termination. We excite one port at a time with a unit-amplitude cosine wave of a specific frequency, measure the amplitude and phase of the transmitted wave at every other port, the amplitude and phase of the wave reflected from the excited port.

From the above description we can see the s-parameters are intrinsically a frequency-domain concept. For wireless applications, which deal with a modulated carrier wave, the frequency domain is entirely satisfactory. However, in recent years s-parameters have gained favor in characterizing linear components in broadband (DC to gigahertz) applications like signal integrity engineering. For these multigigabit/s applications we need to derive a time-domain representation from s-parameter measurements, for example to create an eye diagram, or to include it in a system that also contains components transient (SPICE-like) models. Like most things, it turns out that there’s a right way and many wrong ways to do this.

For sure we can convert between the time and frequency response functions, \( g(t) \) and \( G(\omega) \) of a linear, time-invariant system like a printed circuit board trace using the Fourier integral and its inverse:

\[
G(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} g(t) dt
\]

\[
g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} G(\omega) d\omega
\]

However, these require we know one or other function analytically at every time (or frequency) between \(-\infty\) and \(\infty\). Not even an Agilent vector network analyzer can give you that! In practice you can measure s-parameters from low frequency to the upper limit of the VNA (tens of gigahertz) and only at a finite number of discrete frequency points. With these band-limited, discrete s-parameters in hand, the temptation is to use the related (but different) inverse discrete Fourier transform (inverse DFT) to produce a time-domain model. And this is where the problems begin.
To see the issue, let’s consider a specific example of an ideal delay line:

\[ g(t) = \delta(t - t_{\text{delay}}) \]

where \( \delta() \) is the Dirac delta function. Its Fourier integral is simply:

\[ G(\omega) = e^{-j\omega t_{\text{delay}}} \]

which has a constant unit amplitude and a radian phase \( \Phi = -\omega t_{\text{delay}} \), which is linear with angular frequency. As expected the group delay \( \frac{d\Phi}{d\omega} = t_{\text{delay}} \), and the inverse Fourier integral reconstructs the original because the integral of \( e^{j\omega(-t_{\text{delay}})} \) is zero everywhere except \( t = t_{\text{delay}} \).

Contrast this with the following band-limited, discrete Fourier transform, here using MATLAB code:

```matlab
npts = 256;
delta_t = 1e-9; % s
t = 0:delta_t:delta_t*(npts-1);
f = linspace(-(npts-1)/(2*npts*delta_t), 1/(2*delta_t), npts); % Hz
amplitude = ones(size(f));
delay = 10.5e-9; %s
phase = -2 * pi * f * delay;
fresp = amplitude .* exp(j * phase);
tresp = ifft(ifftshift(fresp));
plot(t,tresp)
```

Instead of going to infinity, the frequency response is band-limited at the Nyquist frequency \( \frac{1}{2 \times 1\text{ns}} = 0.5 \text{ GHz} \). The resulting time-domain response is far from the Dirac delta function in our previous Fourier integral example.

**Figure 1: Discrete Fourier transform of truncated frequency response of an ideal delay line**

I chose a worst case group delay of 10.5 ns. It lays halfway between the 10 ns and 11 ns points on the discrete 1 ns time grid. There’s no way of representing it because there simply isn’t an output of
the inverse DFT at 10.5 ns. The pulse is spread out so badly that the skirt of the next period leaks into
the end of this one.

Strictly speaking, the output of an inverse discrete Fourier transform isn’t even an impulse response at
all: it’s one period of the repeated pulse train response. The period is the inverse of the spacing of the
frequency points or \( \sim 1/(4 \text{ MHz}) = 256 \text{ ns} \) in our case. But there’s something else. The spreading is
non-causal. The pulse starts to arrive before the 10.5 ns group delay, as well as continuing after it.
Strange? Read on…

One obvious cause of this pickle is that the discrete Fourier transform suddenly truncates the spectrum
at the Nyquist cutoff frequency: 0.5 GHz in our example. This sharp “brick wall” filter in the frequency
domain gives a broad response in the time domain. We might try applying a smooth amplitude roll off
(“windowing”) in the frequency domain. But how about the phase? How to roll that off? Anyhow,
amplitude windowing gets rid of some of the sharp time domain oscillations (“ringing” or “Gibb’s
phenomena”), but it actually makes the time-domain pulse even broader, and so the non-causality
problem grows.

No, to reconstruct the pulse as accurately as possible, we need to take a different approach using
additional information that is missing from the band-limited s-parameter data we have. We know the
measured component is physical and it’s response beyond the measurement cut-off will roll off
smoothly. So we don’t want to window, we actually want extrapolate the amplitude and phase.
Extrapolating the amplitude is a simple matter of fitting a low order polynomial curve to the measured
amplitudes we do have. But again what should we do about the phase?

Let’s go back to what we know about the measured component. Again, it’s a real physical thing, so it’s
time domain response is certainly causal (effect must follow cause) so:

\[
g(t) = 0 \text{ for } t < 0
\]
Knowing this about the time-domain response constrains the set of Fourier integrals that are possible. The constraint is called the Kramers-Kronig relation. A necessary and sufficient condition for a causal system is that the real and imaginary components of the frequency response must be related to each other by a specific mathematical operation called the Hilbert transform (see sidebar below, “Kramers-Kronig in Words”).

Sidebar: Kramers-Kronig in Words (Mostly)

The traditional proof of Kramers-Kronig involves fairly complicated math including contour integration. It’s rigorous but not very intuitive. Inspired by the proof in Hall & Heck’s textbook (see reference below), this sidebar tries to add some insight.

The real and imaginary parts, respectively, of a Fourier integral tell you how “cosine-like” and “sine-like” the time domain waveform is:

\[
G(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} g(t) \, dt = \int_{-\infty}^{\infty} \cos(\omega t) g(t) - j \sin(\omega t) g(t) \, dt
\]

In the time domain, cosines and sines are even and odd functions, respectively. For waveforms in general, there’s no a specific relationship between the sum of the cosines (the even component) and the sum of the sines (the odd component). But for causal waveforms there is: if you take the even component, multiply by the “sign-of-the-argument” step function:

\[\text{signum}(t) = -1 \text{ for } t < 1\]
\[\text{signum}(t) = 1 \text{ for } t > 1\]

...you get a result equal to the odd component. With some elementary algebra and some sketches you can see that this is the necessary and sufficient condition to make the even and odd components exactly cancel to zero for all points with negative time. So we can write a causal waveform as \( h_{\text{causal}}(t) = h_{\text{even}}(t) + \text{signum}(t) h_{\text{even}}(t) \). The \( h_{\text{even}}(t) \) and \( \text{signum}(t) h_{\text{even}}(t) \) terms gives rise the real and imaginary parts of the Fourier integral, respectively.

Now, in the frequency domain, we use two facts:

1) Multiplication in the time domain corresponds to convolution in the frequency domain, and

2) The Fourier integral of \( \text{signum}(t) \) is \( \frac{2}{j\omega} \) which is called the Hilbert kernel.

Convolution with the Hilbert kernel is called the Hilbert transform. Thus the imaginary part is equal to the Hilbert transform of the real part.

For a longer, pictorial version of this argument please also see “Kramers-Kronig in Pictures”

For a rigorous proof see for example Hall and Heck

If you know the real part you can calculate the imaginary and visa versa. Not quite a free lunch, but close! Let’s call the real part \( r \), the Hilbert transform of \( r \) is \( H(r) \), and imaginary part \( i = H(r) \).

Using the relationship between our extrapolated amplitude response and \( r \) we can find an \( r \) such that:

\[ r^2 + H(r)^2 = \text{Amplitude}^2 \]

With \( r \) and \( H(r) \) in hand, then \( \text{Phase} = \tan \left( \frac{H(r)}{r} \right) \)

With this method in hand we can optionally add two further refinements: delay causality and passivity. (Passivity may at first appear unrelated to causality but, as we’ll see, there is a subtle connection.)
Because the ports have a finite physical separation and because no information can pass between them faster than the speed of light, the component must be not only causal but also delay-causal, meaning effect must occur no sooner than light can propagate after the cause. We can create a delay-casual model by subtracting the propagation delay (sometimes called the principle delay) in the frequency domain (where it appears as a phase linear slope $\Phi = -\alpha \omega \Delta t$), applying the algorithm outlined above, and adding back the principle delay to the resulting time domain model.

**Next, the passivity connection.**

Due to measurement noise or numerical error in EM simulation, s-parameter data of a passive component can cross the line and be misrepresented as non-passive at one or more frequency points. Passivity violation could lead to unstable and erroneous waveforms.

If we know the component is in fact passive, we have several options to enforce it. The simplest is to scale the whole spectrum by a constant factor which is less than one. This method has the advantage that the causality is not affected. However, it penalizes not only the incorrect non-passive point but also the correct passive ones. Therefore, the corrected spectrum is overly attenuated when compared to the actual component.

A more sophisticated approach is to correct on point-by-point basis. In this approach, the s-parameter matrix is scaled down by the inverse of the magnitude of the largest eigenvalue only at frequencies where non-passive values occur. Magnitudes at other frequencies are unchanged. Unfortunately this method alone can introduce further causality artifacts. But applying it before our causality method ensures that the artifacts of the first (passivity-enforcing) step are eliminated in the second (causality-enforcing) step. Our second nearly-free lunch!

Before we wrap up, let’s look at some examples using Agilent’s patent pending implementation in the ADS Transient Convolution simulator:
Figure 2a (top) and Figure 2b (bottom)

**Figure 2.** Two comparisons of time-domain waveforms of a bit stream through a power distribution network. Both figures include the directly measured response (in blue). Figure 2a) the red curve is the time domain waveform derived from s-parameters without causality enforcement. The red curve in figure 2b) is the time domain waveform derived from s-parameters with causality enforcement.

In summary, we have shown that we can create accurate, delay-causal, and passive time-domain models by combining band-limited s-parameter data with knowledge about the physical characteristics of our component, namely that all physical components of finite size are delay-causal and components with no external energy input are passive. This method saves you the tears of frustration that would be caused by non-causal and non-passive models.

For more information about Agilent’s signal integrity analysis solutions, please visit [http://www.agilent.com/find/signal-integrity-analysis](http://www.agilent.com/find/signal-integrity-analysis)
References

1) The title is homage to the 1936 comic play “French Without Tears” by Terence Rattigan


3) US Patent Application Number 20080281893


About the authors

Colin Warwick is signal integrity product manager at Agilent EEsof EDA, where he is focused on multigigabit per second signal integrity analysis tools. Prior to joining Agilent, Colin was with Royal Signals and Radar Establishment in Malvern, England, Bell Labs in Holmdel, NJ, and The MathWorks in Natick, MA. He completed his bachelor, masters, and doctorate degrees at the University of Oxford, England. He has published over 50 technical articles and holds thirteen patents.

Fangyi Rao received his Ph.D. degree in physics from Northwestern University in 1997, for research in quantum theory of magnetism and transport. He joined Agilent EEsof in 2006 as a Senior Development Engineer. From 2003 to 2006 he was with Cadence Design Systems, where he made key contributions to the company's Flexible Balance technology and perturbation analysis of nonlinear circuits. Prior to 2003 he worked in the areas of EM simulation, nonlinear device modeling, and optimization.