Signal Integrity

Determining the Operating Bit Rate of a Socket

Interpreting Spring Probe S-Parameters

Background

The performance for serial IO is now operating single level encoded streams with bit rates ranging from 8Gbs to 25 Gb/s (protocol dependent, application specific) as well as multi-level Pulse Amplitude Modulation (PAM) at double and quadruple these bit rates. This increase in IO speeds and additional modulation modes are needed to compress more data per channel to accommodate the emerging data center and cloud explosion in big data analytics, AI systems and the fiber optic expansion out past 100 Gb/s and network backbone support for 5G wireless.

No surprise that with this evolution in IO speeds, test engineers will need to deal with higher frequencies, faster bit rates and denser packaging. This can leave the socket designer who does not have deep experience in RF, microwave wave engineering and 3D EM field solvers struggling to specify the proper spring probe to meet their application's IO specifications.

This application note will provide some basic information to use for interpreting the specifications for the IO performance of a DUT socket or contact probe. We will also add some context to the S-parameter plots and the frequency and timing specifications provided in a spring probe or contactor spec sheet.

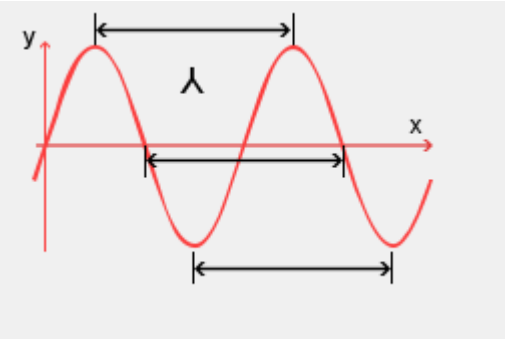
Wavelength of a Probe

Let's establish some basic terms and definitions.

Wavelength:

This is an important concept for understanding the operation of a passive element in a test channel. When we convert the frequency of a signal into the distance it travels over the time for one cycle, we obtain the signal's wavelength. The formula:

Wavelength (ʎ) = wave velocity (m/s) / frequency



A passive component of a test channel has a wavelength, dependent on the signal propagation delay and the material surrounding the conductors.

For example: A 1 GHz sine wave signal has a 1 ns period. It travels 299792458 m/s in air, so the signal is 299.79 mm. The signal slows down when the signal medium is copper surrounded by a dielectric material as it would be in a socket or PCB. Assume a conductor in a dielectric medium with a dK of 4. In this case the wave travels 149860000 m/s. Wavelength is now 150mm. A 5mm long probe is 1/30 of the 1 GHz wavelength in dielectric

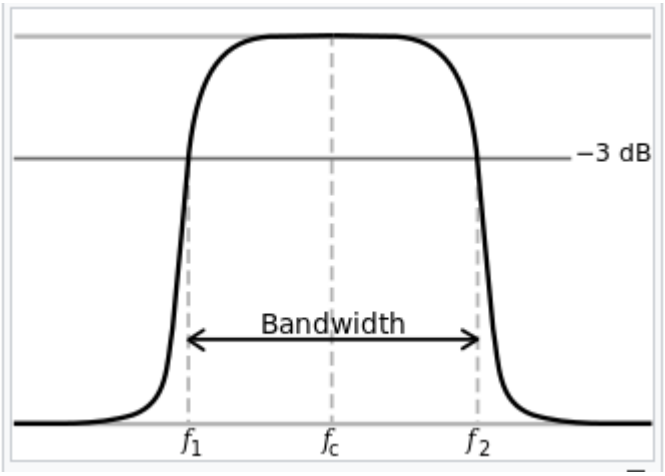
environment. The 5mm length probe has a very small impact on the 1 GHz signal integrity, it is simply a small parasitic lumped element.

At 40 GHz things get a little more interesting. The wave length of a 40 GHz signal in an environment with a 4.0 dK value is now 3.74 mm (150mm/40), less than the length of the 5mm probe. Now the signal integrity is affected significantly since the probe is no longer a lumped element but a transmission line with a wavelength longer than the incident signal. The 3.74mm wavelength signal creates reflections within the 5mm probe wavelength on the incident signal as the wave front reflects off the terminations at the ends of the probe. These can cancel or add to the pin output power and returned power, and at some frequency almost completely cancel the output power of the signal (resonance).

Bandwidth of a Probe

Bandwidth:

Bandwidth is a term primarily used in communications and signal processing. It refers to the magnitude of a channel or filter bandpass function with an upper and lower cutoff frequency.



Source: Wikipedia

Bandwidth in this context, refers to the distance between the upper and lower cutoff frequencies where at least −3dB of the initial signal power or voltage transfers from the input to the output of the channel. The **dB** is a logarithmic unit that describes a ratio of two measurements, in this case input to output voltage. The equation to calculate the dB level for voltage is:

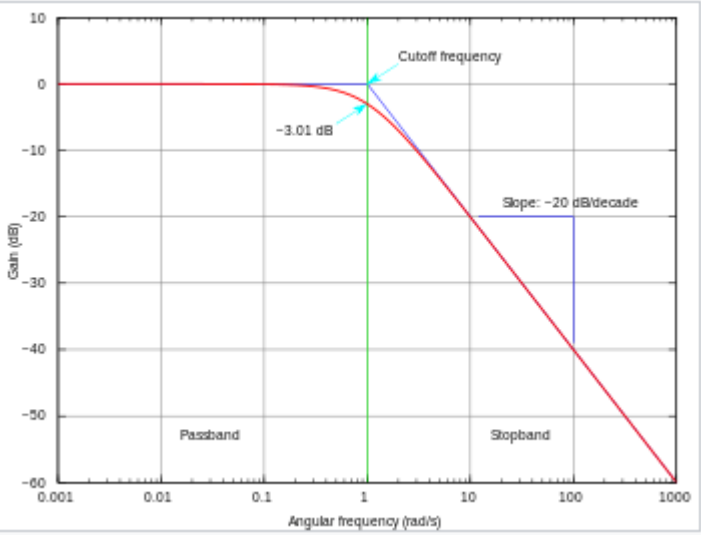
$$L_V = 20 \log_{10}\left(\frac{V}{V_0}\right) \text{ dBV}$$

An output voltage value of −3dB is equal to 0.707 of the input voltage. For spring probes we almost always refer to the voltage calculation rather than a power calculation where the equation is:

$$L_P = 10 \log_{10}\left(\frac{P}{P_0}\right) \text{ dB}$$

In this case −3dB would be equal to 0.5 time the input power.

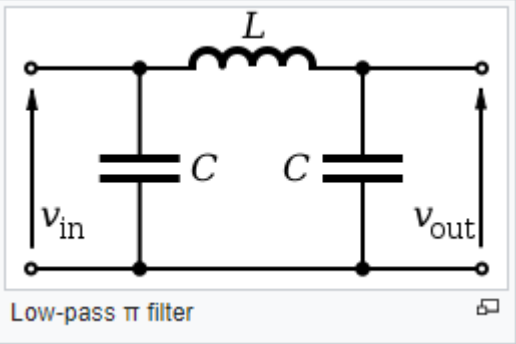
Since the contact probe is a passive element, it operates to DC with just a tiny loss from the small contact resistance (<50mΩ). We now have a low pass filter function:



Power gain is shown in decibels (i.e., a 3dB decline reflects an additional half-power attenuation). Angular frequency is shown on a logarithmic scale in units of radians per second.

Source - Wikipedia

We only concern ourselves with an upper cutoff frequency. Shown here is a single pole filter (RC circuit). Not to make it too complicated but a spring probe model is usually modeled as a Pi filter circuit, so we would have three poles. The circuit abstraction for a spring probe:



Low-pass π filter

This circuit is quite intuitive. If you think about the spring probe structure, it forms a capacitor at the DUT board pad, a capacitor at the DUT contact and is an inductor for series path. There is also a small series resistance is in parallel with the inductor to model the contact resistance. The low-pass π filter (spring probe), is connected within the transmission line, transmitting low frequencies and reflecting high frequencies. The impedance and the frequency pass band are dependent on the component values. This lumped model works well if the wavelengths of the signal are at least *6x* the length of the probe. Once the signal wavelength is less than the 6:1 ratio, the probe starts to act more like a transmission line and should be modeled as a transmission line transfer function for accurate upper band width cutoff frequency.

Contactor Bandwidth Specification

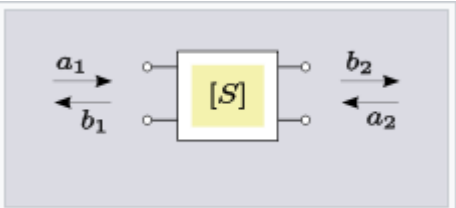
The contactor/probe industry has set an arbitrary upper frequency cutoff boundary of −1dB. This is the standard most vendor specifications for probes will define. This -1dB is not set for any specific application, but it attempts to indicate, mainly as a figure of merit, how much impact the probe will have, on the signal in the test channel. At −1dB, a 1V input produces an 890mV output at the frequency where the -1dB cutoff level occurs.

Since the goal of a probe/socket is to provide as transparent a connection from the DUT to the test channel as possible, the −1dB ***cutoff frequency*** is critical and needs to be applied properly to the device application IO specification.

At this point, we need to add another set of terms to the bandwidth definition, insertion loss and return loss. Signals in a passive network like a spring probe, are always band limited. As the frequency of the signal increases, more signal power is lost to the channel. This loss is both insertion loss, i.e. how much of the signal amplitude is lost as it passes through to the output, and return loss, how much of the signal amplitude is reflected to the source. Bandwidth is sometime interpreted as equivalent to insertion loss, but the socket and probe signal integrity is affected by both types of loss.

Connecting Bandwidth to Insertion Loss and Return Loss- S-Parameters

Since a socket probe is a passive linear network it fits the definition of a two-port network. The ports constitute interfaces where the network connects to other networks, the points where signals are applied, or outputs are taken. In a two-port network, often port 1 is considered the input port and port 2 is considered the output port.



Many electrical properties of networks of components (inductors, capacitors, resistors, filter networks) may be expressed using S-parameters. These include gain or insertion loss, return loss, Voltage standing Wave Ratio (VSWR), and reflection coefficient. In the context of S-parameters for an electrical network, scattering refers to the way in which the traveling currents and voltages in a transmission line are affected when they meet a discontinuity caused by the insertion of a network into the transmission line. This is equivalent to the wave meeting an impedance differing from the line's characteristic impedance.

Although applicable at any frequency, S-parameters are mostly used for networks operating at RF, radio frequency. For probes we are usually working above 500 MHz through upper microwave and emerging mm-wave frequencies. Here signal power and energy considerations are more easily quantified than currents and voltages. S-parameters change with the measurement frequency, so frequency must be specified for any S-parameter measurements stated, in addition to the characteristic impedance or system impedance.

S-parameters are readily represented in matrix form and obey the rules of matrix algebra.

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

For the transfer matrix [S], expanding the matrices into equations gives:

$$b_1 = S_{11}a_1 + S_{12}a_2$$

and

$$b_2 = S_{21}a_1 + S_{22}a_2 .$$

Signal Integrity

.

Each equation gives the relationship between the reflected and incident power waves at each of the network ports, 1 and 2, in terms of the network's individual S-parameters, S_{11}, S_{12}, S_{21}, S_{22}.

The 2-port S-parameters have the following generic descriptions:

S_{11} is the input port voltage reflection coefficient (Return Loss)
S_{12} is the reverse voltage gain (Insertion Loss)
S_{21} is the forward voltage gain (Insertion Loss)
S_{22} is the output port voltage reflection coefficient. (Return Loss)

In case the two measurement ports use the same reference impedance, the insertion loss (*IL*) is the magnitude of the transmission coefficient $|S_{21}|$ expressed in decibels. It is thus given by

$IL = -20\log_{10}[S_{21}]$ dB.

Input return loss (RL_{in}) is the measure of how close the actual input impedance of the network is to the nominal system impedance value. Input return loss expressed in decibels is given by

$$RL_{\text{in}} = 10\log_{10}\left|\frac{1}{S_{11}^2}\right| = -20\log_{10}|S_{11}| \text{ dB.}$$

The output return loss (RL_{out}) has a similar definition to the input return loss but applies to the output port (port 2) instead of the input port. It is given by

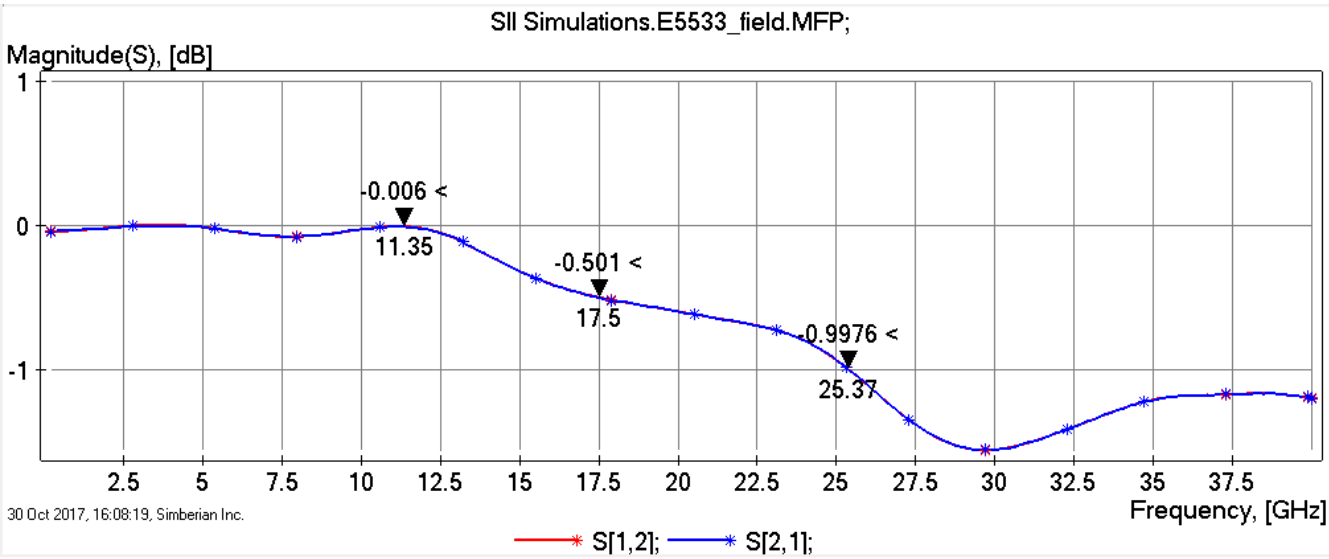
$RL_{out} = -20\log_{10}[S_{22}]$ dB.

It is the extra loss produced by the introduction of the device under test (DUT) between the 2 reference planes of the measurement. Notice that the extra loss can be introduced by intrinsic loss in the DUT and/or mismatch with the termination and intrinsic channel impedance. We have circled back to see that Insertion Loss is equivalent to the bandwidth cutoff frequency, if we match the frequency of the insertion loss where the S-parameter crosses −1dB loss level.

Example Plot of Insertion Loss

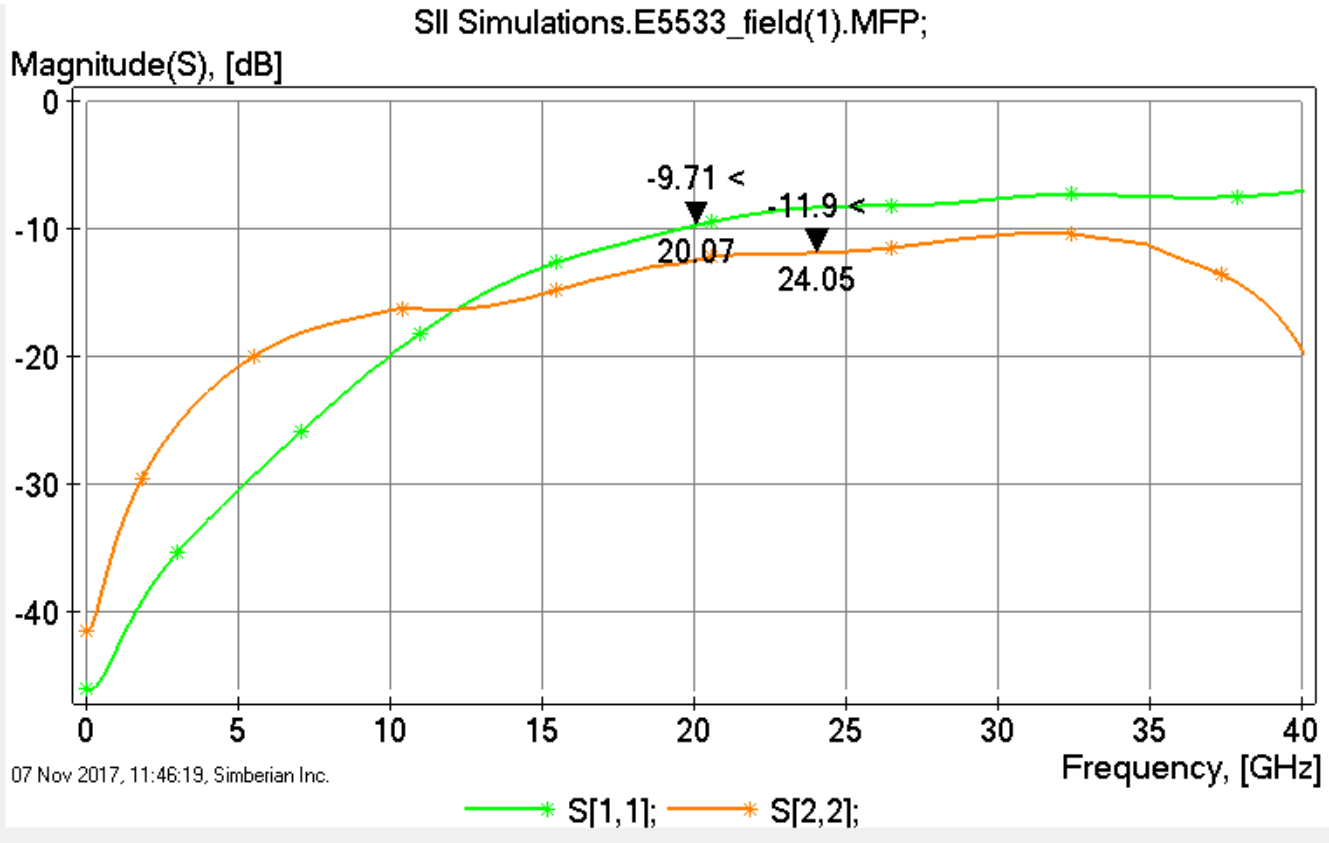
Here is a plot of insertion loss on an SII E5533 pin. First thing to note is any measurement needs a reference, so we are plotting the transfer function insertion loss s-parameter S_{21} for an E5533 with two adjacent ground pins at 1.0mm distance between pin centers (1.0mm pitch). This plot has been measured on a 40 GHz VNA (Vector Network Analyzer).

It is important to note that since we need to add 2 GND return paths at 1.0mm pitch, we have created a transmission line with an impedance that is an important factor in the quality of the signal we can deliver.



The −1 dB cutoff frequency measures 25 GHz and the pin has very low insertion loss to 11 GHz. Beyond 11 GHz the effects of the pin's component capacitance and inductance along with the GND return loop inductance and mutual capacitance from the pin socket configuration takes effect.

Example Plot Return Loss



The Return Loss from this probe measures around -10dB at 20 GHz. -10dB means that 0.31 (31%) of the input

voltage is reflected to the source at 20 GHz. For this example, pin, the return loss varies slightly depending on the direction of the signal. The pin used in this example is asymmetrical with a capture ring surface bulge close to S11 port. Along with the Insertion Loss, Return Loss combines to distort the output waveform, across the pin wavelength frequencies, the reflection can add or cancel part of the output signal. To determine the bandwidth of the probe, the Return loss should be considered carefully.

RF Specifications

If the device under test is a pure RF pin we have enough data from the bandwidth specification and Insertion and Return loss plots to calculate almost exactly how the socket/probe will attenuate the signal. An RF signal is closely modeled as a sinusoid with modulation of some type, usually frequency, and encoding inside the modulation band.

A RF IO signal will perform very closely to the frequency plotted on a VNA. If we have a 2.4 GHz frequency IO and place it on a pin that operates to 10 GHz at -1dB, the loss at 2.4 GHz may be less than -0.2dB or it will reduce the magnitude of the voltage peaks on the signal less than 3% (0.977 of input magnitude).

If we apply a pin that has a bandwidth of 3x the specified bandwidth of the RF signal, we should have no issues with the socket/probe design. It will essentially be transparent. For signals from 1 to 5 GHz it is usually not difficult to find a probe to work at 3x the IO bandwidth. Once we get to device frequencies above 10 GHz, things get more detailed and difficult but there are still many practical socket solutions to 40 GHz.

For auto radar and other mm-wave applications it is not possible to design an interconnect with large margins over the carrier frequencies 20 GHz to 77 GHz and above. The socket and interface board designer is hard pressed to reach the operating speed of the carrier frequency in the mm Wave band without excessive loss and at a reasonable cost. In the case of very high frequencies we look at specifications like VSWR (voltage standing wave ratio) and compromise the maximum output attenuation, Often the test circuits on the part or the equipment will down convert such high frequency outputs to a more practical measurement range.

In the case of RF, microwave and mm wave signals, we want to supply as a pin with a characterized bandwidth at least 4 or 5 GHz above the center frequency of the carrier. The flatter lower loss part of the insertion loss curve and the flatter parts of a VSWR plot before the resonance

Digital SerDes IO Pins

When we move to encoded digital signals in serial IO channels, we need to consider more issues than just the channel bandwidth. The drive signal and interpretation of a one/zero in an advanced digital device with SerDes IO is enhanced by circuits designed into in the transmitter and receiver macros that employ real time analog and digital signal processing methods. As designers seek to improve the amount of information that can be transferred in a band limited channel, the device IO can interpret the bit stream in higher and higher signal to noise environments, and less than ideal board layout with higher material losses than the bit rate would seem to demand.

Loopback Testing and Production Screens

In a production or high-volume characterization test application one common test methodology is to verify the IO operates properly in a channel that has a fixed loss through a frequency range. Specific IO pin pairs have a maximum insertion and return loss modeled for the IO at a given frequency. A typical specification may state that a channel will operate at -18dB at 10 GHz. This type of testing is usually referred to as Loopback Testing and employs a transmitter port looped back on a receiver port with the device's DFT used to determine if the part functions properly. The test board traces act as the IO port's load and fixed filter for the test and must be designed carefully to match the operating test limits. In this type of test, the socket port impedance and S-parameters must be included with the parameters of the package and board traces to insure the socket path does not create a discontinuity or resonance that would prevent the test channel from meeting the loss specifications

A 3D simulation is the best way to determine if the socket interconnect will be an issue. The simulation also allows for the tuning of the trace to the socket package to make the loss point vs frequency more accurate. Ideally a socket should be designed in conjunction with the test board and knowledge of the package losses so that the proper length channel, dielectric material and via designs are applied to the test board to achieve the total loss for the test. The socket does not have to provide a specific maximum bandwidth, it just needs to be added to (or de-embedded) to the total loss of the channel.

In the more typical case, the board is locked down before the socket is designed, the simulation helps us understand if the socket IO channel impedance and losses around the frequency targets. If there is a test issue, we can use the simulation to find ways to modify the socket to squeeze more margin out of the test

Engineering Characterization

The other type of SerDes IO testing uses high accuracy external instruments to measure some component of the output transmitter or input receiver module. Some examples of these types of tests: eye opening, total jitter, input return loss, transmitter rise time, receiver sensitivity, Bit Error Rate. For engineering characterization, the approach on socket design is to be as transparent as possible to provide the instrument with good data on the analog performance of the IO. And a stable calibration offset. Here is where the analog bandwidth figure is useful.

In the case where the test goal is to establish as transparent a channel as possible, the bandwidth of the connector pin socket structure will indicate the amount of signal information that can pass through the socket. The bandwidth will dictate the minimum rise time the transmitter can produce, it will be one of the factors in the duty cycle distortion and inter symbol interference components of the jitter measurement, the receiver will need to function under the signal attenuation of the bandwidth limit and impedance match of the socket channel.

Bit Rate

Serial Data speed is measured in bit rate. Bit Rate is specified in Gb/s (Gigabits per sec) which is time based as opposed to a frequency in which is specified in GHz. We are referring to Gross Bit Rate, $\boldsymbol{R_b}$, or the un-coded raw transmission of bits. In non-formatted or NRZ (non-Return to Zero), $\boldsymbol{R_b}$ is equivalent to the Unit Interval (UI).

$$R_b = 1/T_b$$

For 1 Gb/s, $\boldsymbol{R_b}$ = 1ns or $1/(10^{-9})$ sec.

A serial bit stream can consist of any arbitrary sequence of ones and zeros. There are many serial protocols (PCI express, SATA, USB, Infiniband, SPI…) and discussion of the details of these are far beyond the scope of this paper

The designer needs to understand the conversion of the IO information in the bit stream. The unit interval or smallest time slice information is transmitted can be used as the target frequency for the socket design.

In broadband communication, the higher the channel bandwidth the more harmonic power can be passed through the channel. Ideally power at the third and fifth harmonic is needed to adequately produce an open eye. Reproducing the signal rise time and minimizing inter-symbol interference is best achieved the higher the socket bandwidth. If the goal is raw measurement of the output signal, more bandwidth is better.

Operating bit rate for modern transceiver receiver pairs are functions of the circuits used in the device to operate in the channel. There are many signal processing circuits on the device to achieve operation at high bit rates in very lossy channels. In the case of modern broadband signaling with the extensive use of DSP to operate, the socket channel becomes another channel discontinuity like a board via. It needs to be understood and its contribution to overall channel loss and compensation are the approaches to take for successful socket designs.

Signal Integrity

Serial Data Harmonics

Some examples of bit rates and harmonics at clock pattern rates:

Protocol	Bit Rate	Harmonics 1st 3rd 5th	Cutoff
FBDIMM2	9.6 GB/s	4.8 GHz, 14.4GHz, 24GHz	24 GHz
PCIe3	8.0 GB/s	4.0GHz, 12.0GHz, 20GHz	20 GHz
FBDIMM3	16.0GB/s	8.0GHz,24 GHz, 40 GHz	40 GHz
Coherency Protocol	20 Gb/s	10 GHz, 30 GHz, 50 GHz	50 GHz

This table gives us a projected frequency needed for the probe pin to provide minimal attenuation to the signals 5th harmonics and is transparent in the time and voltage domains. Probe pins for socket channels that have enough bandwidth to pass the target signal's 5th harmonics will exhibit very low channel ISI and DCD. This frees the engineer to concentrate on less forgiving components in the signal interface.

Serial Data Compensation

Modern SERDES transceiver pairs have implemented many architectural features to process the data in a lossy channel. A transmitter/receiver pair in a SERDES lane can usually be designed to operate in loss environments that approach and exceed -20dB. Using real time DSP, de-emphasis, Finite Impulse Response circuitry, advanced CDRs, and pulse shaping, a SERDES lane can often train and operate when the eye observed in the test channel is completely closed. What does this mean for the probe pin response?

The probe/interface should provide a smooth loss curve to ensure that the probe adds minimal lumped distortion points. Large frequency notches on an S21 curve may cause too much variability in pattern discernment functions and inhibit the lane performance. A lossy probe pin at high frequency may be perfectly adequate to test a device IO if the loss remains linear. A resonance spot can be highly detrimental and probe - board via interfaces can produce these resonance points as a function of the probe-via T_{pd}.

Signal Integrity Spring Probes - Projected Bit Rates

0.4mm Pins

Probe	-1dB Cutoff	Clocked Bit Rate
AA909	12.0 GHz	24 Gb/s
A1512	20.3 GHz	40.6 Gb/s
A1520	24.1 GHz	48.2 Gb/s
A1540	16.1 GHz	32.2 Gb/s
A1550	18.7 GHz	37.4 Gb/s
A1561	7.4 GHz	15.4 Gb/s
A1562	11.6 GHz	23.2 Gb/s
A1580	7.4 GHz	15.4 Gb/s
A1582	9.6 GHz	19.2 Gb/s

Clocked Bit Rate = 80% Eye opening, loss less channel estimate

0.5mm Pins

Probe	-1dB Cutoff	Clocked Bit Rate
B2500	6.4 GHz	12.8 Gb/s
B2501	11.2 GHz	22.8 Gb/s
B2502	17.0 GHz	34 Gb/s
B2503	13.0 GHz	26 Gb/s

B2504	8.8 GHz	17.6 Gb/s
B2509	16.1 GHz	32.2 Gb/s
B2514	12.2 GHz	24.4 Gb/s

Clocked Bit Rate = 80% Eye opening, loss less channel estimate

0.8mm Pins

Probe	-1dB Cutoff Frequency	Clocked Bit Rate
D4525	19.5	39 Gb/s
D4565	17.0 GHz	34 Gb/s
D4595	19.5 GHz	39 Gb/s
D4601	12.8 GHz	25.6 Gb/s
D4603	12.4 GHz	24.8 Gb/s
D4613	14.3 GHz	28.6 Gb/s
D4623	8.2 GHz	16.4 Gb/s
D4693	18.3 GHz	36.6 Gb/s

Clocked Bit Rate = 80% Eye opening, loss less channel estimate. Full harmonic Bit Rate is the inside marginal spec for the probe where no observable distortion is present

1.0 mm pitch pins

E5566	>40 GHz	80 Gb/s

E5533	25.3 GHz	50.6 Gb/s
E5548	14.5 GHz	29 GB/s
E5585	31.5 GHz	63 Gb/s
E5593	31.5 GHz	63 Gb/s
E5656	13.9 GHz	27.8 Gb/s

Clocked Bit Rate = 80% Eye opening, loss less channel estimate.