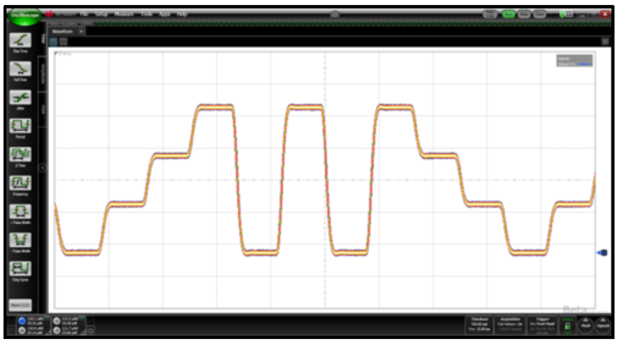
WHITE PAPER

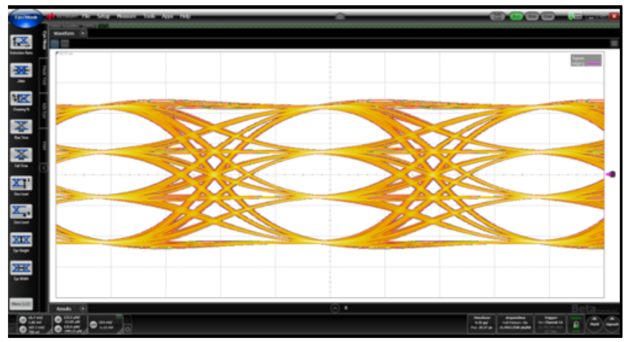
Advanced Modulation and Coding Challenges

Accelerating from 100GE to 400GE in the data center

Demands for a connected world with instant data access continue to drive data center transceiver innovation. 100 gigabit Ethernet (GE) data transmission is in production and will continue to evolve. But 100GE speeds aren't fast enough to support the expected surge in connected devices and the applications they will run, opening the door for 400GE.

Non-return-to-zero (NRZ) and four-level pulse amplitude modulation (PAM4) are two modulation technologies that enable 400GE. Each comes with its own set of challenges.





The basics

The move from 100GE to 400GE in the data center is revolutionary, not evolutionary. Optical transceivers use advanced signal modulation and coding to reach 400GE speeds. These techniques create new test challenges for transceiver manufacturers.

Modulation Schemes Evolve

The most common modulation scheme for 100GE is NRZ modulation. NRZ is a two-state transmission system (also called two-level pulse amplitude modulation, or PAM2) where positive voltage represents a logical 1 and an equivalent (generally) negative voltage represents 0. 100GE uses four lanes of 25 gigabit-per-second (Gb/s) NRZ modulated signals. NRZ has gradually evolved over the last 50 years, with improved speeds from 110 bits per second to 100 Gb/s. Therefore, researchers have addressed many new concepts and challenges. Theoretically, reaching 400GE speeds with NRZ is possible by applying these same concepts using eight lanes of 56 Gb/s signaling. However, as the speed of NRZ designs increases above 28 Gb/s, channel loss of the transmission medium becomes a limiting factor.

According to the Shannon–Hartley theorem, there is a theoretical maximum amount of error-free data over a specified channel bandwidth in the presence of noise. As such, either the channel bandwidth or the number of signal levels must increase to improve the data rate or channel capacity. That requires new multilevel signal modulation techniques.

PAM4 signals use four amplitude levels, with logical bits 00, 01, 10, and 11 representing symbols. The number of symbols transmitted per second (baud rate) is half the number of bits transmitted per second. For example, a data rate of 28 gigabaud (Gbaud) represents 56 gigabits of data transmitted per second. This rate is double the data rate (throughput) in the same bandwidth as 28 Gbaud NRZ, which is essentially 28 Gb/s, since 1 bit represents one symbol.

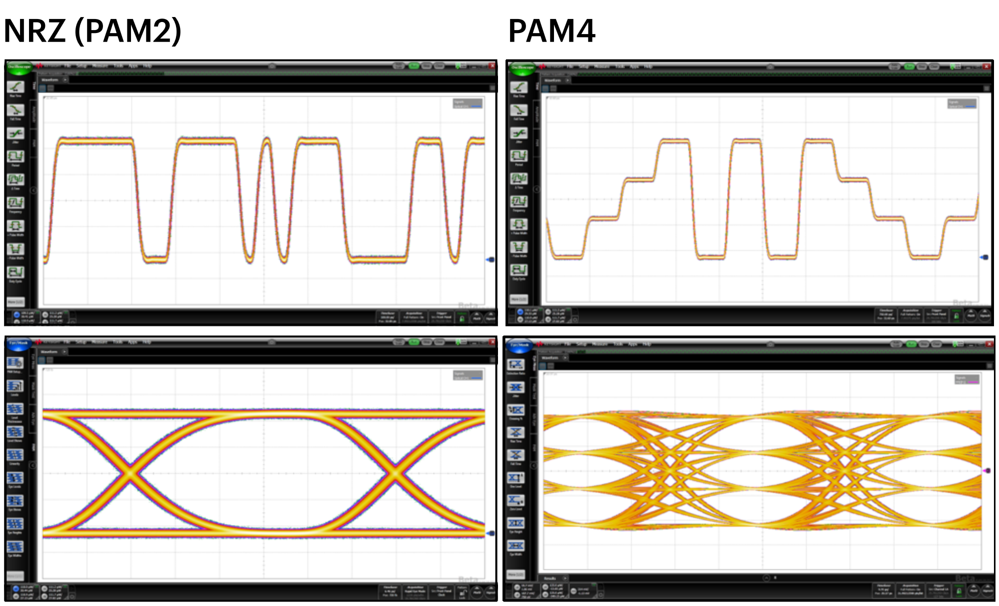


Figure 1.
NRZ versus PAM4 modulated signals

However, PAM4 designs are far more susceptible to noise because they pack four signal levels into an amplitude swing of two, as shown in Figure 1. Therefore, the signal-to-noise ratio is lower. Analyzing noise from transceiver designs needs to account for channel return loss and noise from the test instrumentation. PAM4 uses forward error correction (FEC) to account for this. FEC is an advanced coding technique that sends the information required to correct errors through the link along with the payload data. FEC introduces challenges in physical-layer testing of PAM4 signals. To facilitate error correction, PAM4 supports the use of Gray coding pattern.

Gray code, also called reflected binary code, is a coding pattern in which successive symbols differ by one binary bit. In the case of the PAM4 bit sequences defined above, the Gray code representation for the same symbols would be 00, 01, 11, and 10 for the levels 0, 1, 2, and 3, respectively. The Institute of Electrical and Electronics Engineers and Optical Internetworking Forum standards recommend Gray coding to encode bits onto a PAM4 signal.

FEC introduces challenges in physical-layer testing of PAM4 signals.

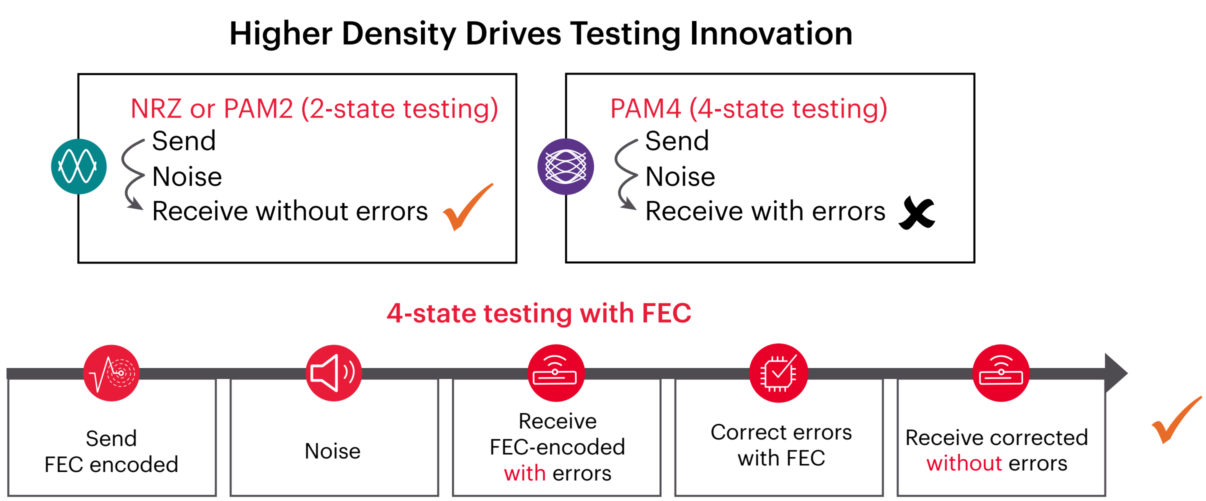


Figure 2. Testing with and without FEC

As shown in Figure 2, FEC sends the information required to correct errors through the link along with the payload data. The decoder uses this information to recover corrupted data without requesting that the transmitter resend it. Both the transmitting and receiving ends of the link must know which coding scheme is in use for the link to operate. Links employing FEC use a variety of coding systems. The more common coding schemes used in data center networking are variants of the Reed–Solomon (RS) system, developed in the 1960s by Irving Reed and Gustave Solomon for use in satellite data links.

RS coding operates on a block of data with a fixed size known as a symbol. Grouped symbols form a frame. It is important to note that the “symbol” and “symbol error rate” terms that appear in data center networking standards using PAM4 refer to the RS symbols and not the PAM4 symbols. RS encoding and decoding work on binary data, before and after the Gray coding conversion from binary to PAM4 and back again.

Test Implications of FEC

With 400GE, naturally occurring errors in the system are acceptable to a certain level and then corrected with FEC, resulting in a nearly error-free environment post-FEC. However, transceiver manufacturers need to consider three things when testing FEC-encoded PAM4 signals: coding gain, burst errors, and striping.

Coding gain

The encoding process converts the payload data to a format that allows decoding and creates the additional data required to correct errors. The resulting encoded data is known as code words. Decoding on the receiving end recovers the data. Coding gain is a figure of the robustness of the error correction code. Higher coding gain allows the correction of a greater number of errors. However, there are trade-offs. RS systems using higher coding gain require more overhead sent in the block of code words to facilitate decoding at the receiving end. Also, increasing the coding gain increases the amount of logic required for coding and decoding, as well as the processing time, or latency, needed to encode and decode the data. Higher-speed serial data links using PAM4 have a higher native error rate than those using NRZ line coding, requiring an FEC with a higher coding gain.

The three main considerations when testing with FEC are as follows:

- coding gain
- burst errors
- striping

Burst errors

A given coding gain in an RS system can correct a defined maximum number of errors in a code word. If the errors exceed this number, the code word cannot be decoded and all the data is lost. This event is known as a frame loss. Frame loss ratio (FLR) is similar to bit error ratio (BER). BER is the measure of the percentage of bits received with errors, because of noise or interference, divided by the number of bits transmitted. Therefore, FLR measures the percentage of frames not delivered divided by the number of frames sent.

Selecting the FEC coding gain avoids frame losses given the target worst-case BER of the native link. This process assumes that the error distribution in the native link is approximately random. A large burst of errors that exceeds the number of correctable errors in the frame will result in frame losses, even if the average error rate in the link is better than the specified native BER. Note that a “burst” in this context is not necessarily consecutive bits. The errors could be interspersed with correct data bits and would still result in a frame loss if the maximum number of correctable bits for the FEC code being used is exceeded.

Error bursts can originate in the receiver end of the link or anywhere in the link where the data is retimed without FEC decoding and re-encoding, such as the pass-through mode in optical modules. Therefore, anticipating error bursts and using striping minimizes their impact.

Striping

Data striping reduces the incidents of frame losses in links employing multiple lanes operating at a subrate of the total link data rate. Striping the data rotates the individual data streams through all the available lanes in the link in a round-robin fashion. By striping, burst errors generated by pass-through retimers in the link will have the length of any error burst effectively divided by the number of lanes used for the striping. For example, in a 100GE link using four lanes of 25.78 Gb/s NRZ data, an error burst of 100 bits generated in an optical module on a single lane would result in only 25 errors on that lane after the striping is deinterleaved at the end of the link. Therefore, while striping does not increase the computed coding gain, which assumes random error distribution, it effectively increases the gain when error bursts occur.

Summary

In addition to RS coding and striping, additional coding takes place in data center networking links. The data first passes through a scrambler to remove any DC conditions and preserve symmetry. Then the data is transcoded before passing through the FEC encoder. After striping, the binary data passes through a PAM4 Gray coder. At the opposite end, logic blocks perform the decoding in the opposite order. The logic required to perform this process is complex. The decoding end of a 400GE link using eight lanes of RS FEC with striping requires several hundred thousand gates. As such, there are several test implications to consider in links using RS FEC. Transceiver manufacturers need to change their testing procedures to account for FEC and use different metrics to test performance characterization and compliance to standards.

For information on how Keysight's solutions can help you address your 400GE data center implementation challenges using advanced modulation and coding techniques, visit the following links:

- For data center infrastructure high-speed digital design and test solutions, check out www.keysight.com/find/DCI.
- For 400GE design and test solutions, including PAM4 simulation and transmit and receiver test, go to www.keysight.com/find/400G.
- To learn more about 800G solutions, check out www.keysight.com/find/800G.