WHITE **PAPER**

cadence®

Pushing the Envelope with PCIe 6.0: Bringing PAM4 to PCIe

By Tony Chen, Cadence

The evolution of new artificial intelligence/machine learning (AI/ML) applications and the accelerating shift of enterprise workload to the cloud are shaping modern data center transformations. These concurrent data-intensive developments are fueling explosive data traffic growth in data centers. To address the insatiable demand for more data bandwidth, silicon providers are racing to adopt the latest high-speed interface technologies, such as the Cadence® PHY IP for PCI Express® (PCIe®) 6.0, in their next-generation silicon.

Contents

Introduction

Over the past two decades, the PCIe interface has gained wide industry support and has become the de facto interface standard for high-speed data transfers between processing/computing nodes due to its high-speed, low-latency, low-power, and low-cost attributes. In fact, within the past decade, PCI-SIG has released a new PCIe generation every three years that doubled the data rate over the previous generation while maintaining full backward compatibility (see Table 1).

PCIe Spec	Data Rates (GT/s)	Encoding	X16 B/W Per Dir*	Year
1.0	2.5	8b/10b	32Gbps	2003
2.0	5.0	8b/10b	64Gbps	2007
3.0	8.0	128b/130b	126Gbps	2010
4.0	16.0	128b/130b	252Gbps	2017
5.0	32.0	128b/130b	504Gbps	2019
6.0	64.0 (PAM4)	FLIT	1024Gbps	2021**

** bandwidth after encoding overhead, **projected*

Table 1: PCIe protocol evolution[1]

PCI-SIG is set to finalize PCIe 6.0 specifications in 2021, expanding yet again both speed and bandwidth to address new application challenges and to enable new innovations in data centers. Given the daunting challenges in achieving blazingly fast 64GT/s speed, PCI-SIG moved to adopt a new signaling technology for the latest PCIe generation: 4-Level Pulse Amplitude Modulation (PAM4).

PAM4 Overview

PAM4 is a multi-level signaling technology that transmits two bits per unit interval (UI) as opposed to the conventional non-return-to-zero (NRZ), which transmits only one bit per UI (see Figure 1).

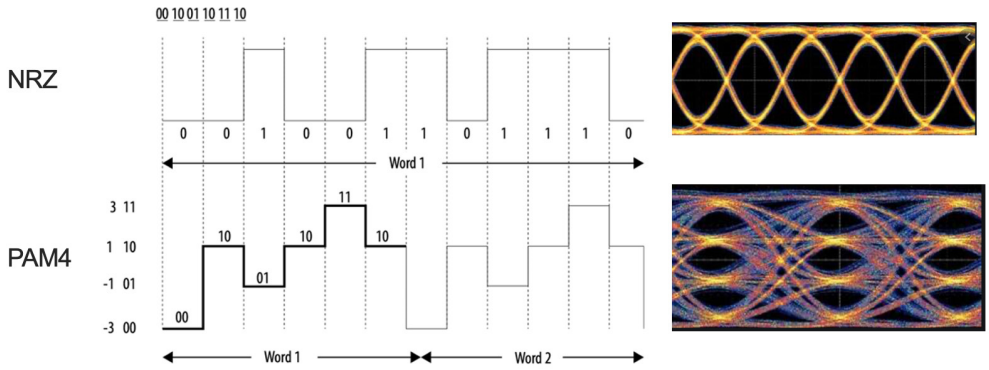

Figure 1: PAM4 encoding

Doubling the I/O bandwidth beyond 32GT/s (PCIe 5.0 speed) to 64GT/s poses significant signal integrity (SI) challenges. The backward-compatibility requirement in PCIe mandates support for legacy channels (PCB + connectors + add-in card, etc.). In NRZ signaling, the channel insertion loss for these legacy channels can be greater than 36dB at the Nyquist frequency (16GHz) for PCIe 5.0 speed (32 GT/s). At 64GT/s, the Nyquist frequency doubles to 32GHz and the channel's frequency dependent loss increases to 70dB (Figure 2)!

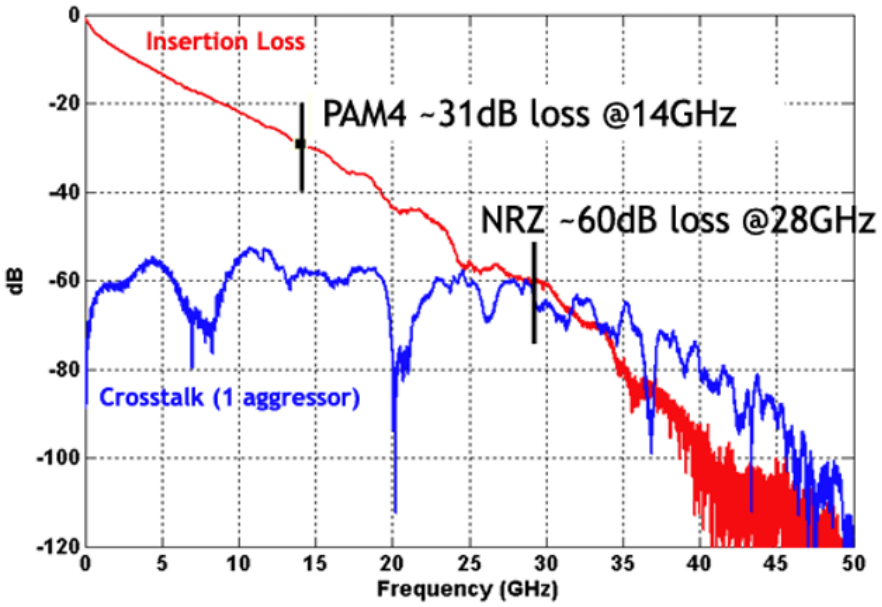

Figure 2: Channel insertion loss vs Nyquist frequency[2]

Adopting PAM4 signaling for PCIe 6.0 is advantageous because by transmitting two data bits per UI, the data rate is effectively doubled without doubling the Nyquist frequency. Channel insertion loss, therefore, is kept at the same manageable level as PCIe 5.0. However, PAM4's advantage over NRZ in halving the Nyquist frequency does come at a cost: The noise margin for PAM4 is reduced by 9.5dB (33%). This reduction in noise margin exacerbates the adverse impact from cross-talk interferences, signal reflections, and power supply noise.

To mitigate the increased noise sensitivity, PCIe 6.0 also adopted Gray code and forward-error-correction (FEC) to minimize the probability of error. Studies have shown that excellent channel operating margin can be achieved by applying these new coding techniques to PAM4.

Another design challenge for PAM4 is its strict linearity requirement. Signal linearity must be carefully considered in the design to avoid performance degradation. Specifically, equalizations and amplifications performed by the receiver and the transmitter must preserve the signal's linearity. Figure 3 shows two PAM4 eyes: One with perfect linearity (left), the other with poor linearity (right). It's easy to see how poor signal linearity translates to reduced noise margin, which can lead to unrecoverable bit errors.

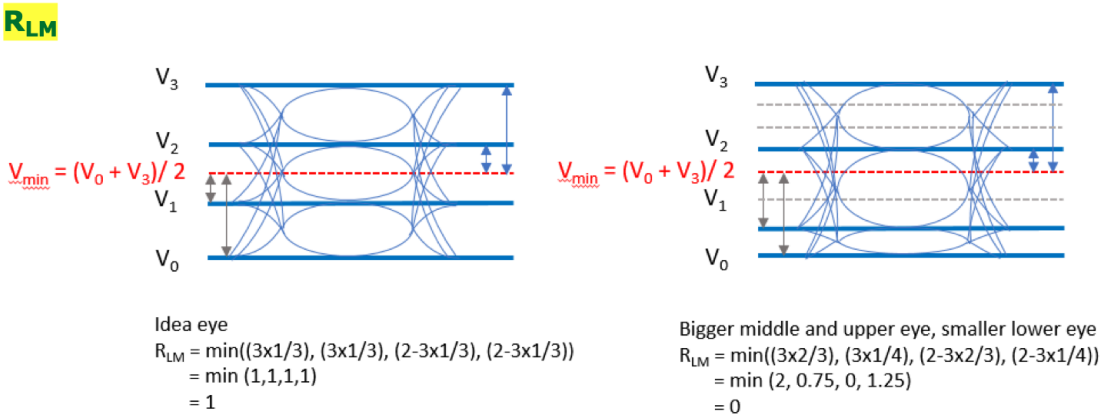

Figure 3: Linearity considerations in PAM4[3]

PAM4 Transmitter

The high-speed transmitter provides two main functions: First, it must be able to generate the correct signal levels to encode the corresponding data bits. Second, it must pre-distort the signal waveforms to compensate for the frequency dependent loss in the channel.

In NRZ signaling, such output drivers have typically been implemented as a finite impulse response (FIR) filter. By summing the weighted contributions from cascaded delay elements, a flexible FIR filter can be constructed (Figure 4).

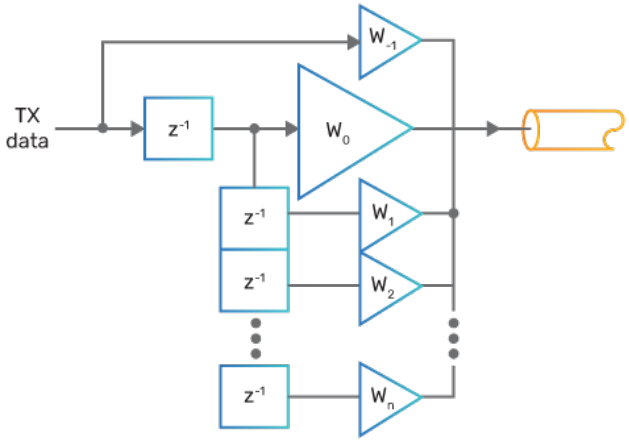

Figure 4: FIR Filter-based output driver implementation

However, this implementation leads to significant capacitive overhead since a dedicated driver is needed for each delayed tap. Although various implementations have been proposed to overcome such shortcomings, these implementations all add complexity and cost overhead.

In recent years, digital-to-analog converter (DAC)-based transmitters (Figure 5) have gained popularity for high-speed signaling due to low parasitic capacitance overhead. In addition, the DAC-based approach provides a highly flexible FIR filtering capability that is limited only by its resolution and linearity. Finally, the area for a high-speed DAC can be smaller compared to the conventional analog approach of the past.

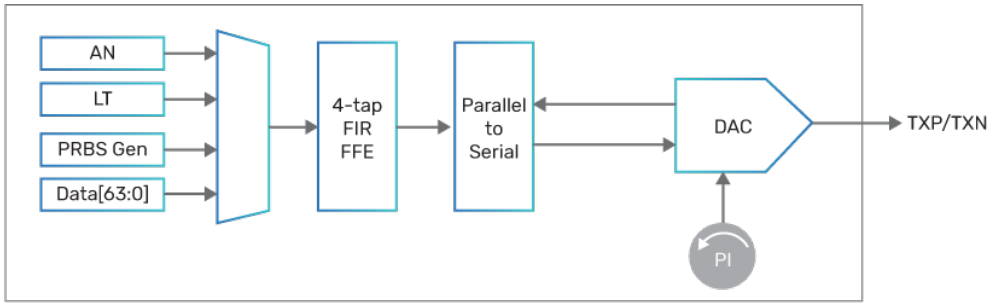

Figure 5: DAC-based transmitter

A TX FIR filter suffers two main drawbacks: 1) TX FIR coefficients are normalized to maximum output swing, which is limited by the power supply. Due to this limitation, TX FIR equalization is performed at the expense of attenuations in the signal's low-frequency contents. 2) A TX FIR filter requires backchannel training to arrive at the optimal coefficient, which requires additional protocol overhead. For the above reasons, a TX FIR filter is usually limited to less than five taps.

PAM4 Receiver

In recent years, digital signal processor (DSP)-based receivers (Figure 6) have gained popularity in wireline communications. The CMOS technology scaling trend is enabling powerful and efficient DSP-based implementations for performing data recovery and channel equalization.

The DSP-based receiver uses an analog-to-digital converter (ADC) to quantize and digitize the incoming data. Once the incoming data is digitized, the DSP processor can then use the digitized data to perform the following tasks:

1. ADC offset and gain correction to compensate the non-ideality in the ADC and its environment
2. Feed-forward equalization, typically more than 20 taps for high-speed PAM4
3. Decision feedback equalization, typically limited to 1 tap for high-speed PAM4
4. Clock recovery and centering
5. Data recovery

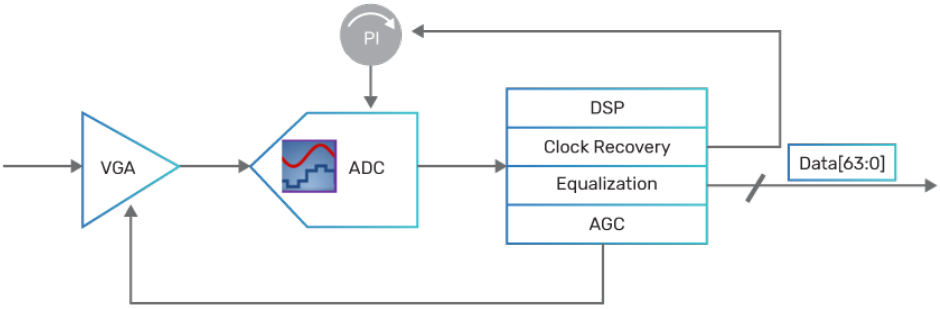

Figure 6: ADC-DSP-based receiver

Equalization and data recovery in the digital domain result in robust tolerance to power, voltage, and temperature (PVT) variations and improve reliability in a data link compared to the conventional analog approach, which is more susceptible to PVT variations. In addition, channel equalization can be done adaptively and continuously in the background to compensate and track voltage, temperature, and humidity drift over time. Moreover, a DSP can be built using a purely digital flow (automated place and route), reducing design cycle time.

FEC and Gray Coding

Although PCIe 6.0 achieved doubling of the data rate by adopting PAM4, the reduced signal-to-noise-ratio (SNR) makes PAM4 signaling far more susceptible to noise in the system compared to its NRZ counterpart. This susceptibility to noise is conducive to a high bit error rate that could lead to system malfunctions or performance loss. To address this vulnerability, PCIe 6.0 features both FEC and cyclic redundancy check (CRC) mechanisms to detect and correct bit errors. FEC is a coding technique that sends redundant data together with the payload data. The FEC decoder in the receiving end can then use the redundant information to correct corrupted data bits, provided that the error rate is below a certain threshold (e.g., 1e-6). This self-correcting mechanism minimizes inefficient data retransmissions. If the CRC detects errors following FEC, the link layer retry mechanism is then initiated to retransmit the data.

In addition, Gray coding is specified for PAM4 signal transmission in PCIe 6.0. Gray coding maps most significant bit (MSB) and least significant bit (LSB) in such a way that symbol errors induced by voltage noise results in a single-bit error at most.

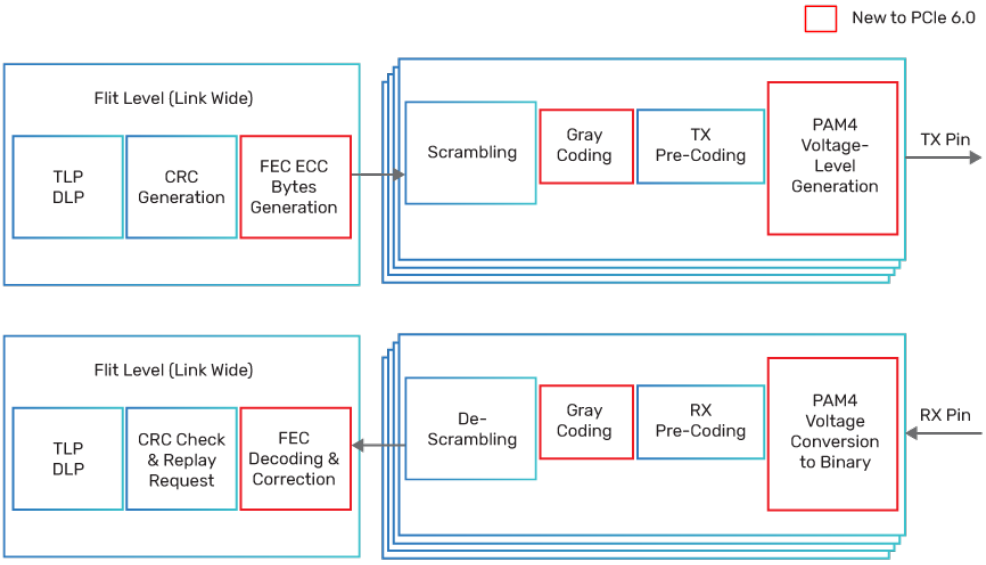

Figure 7: Addition of FEC and Gray coding to PAM4

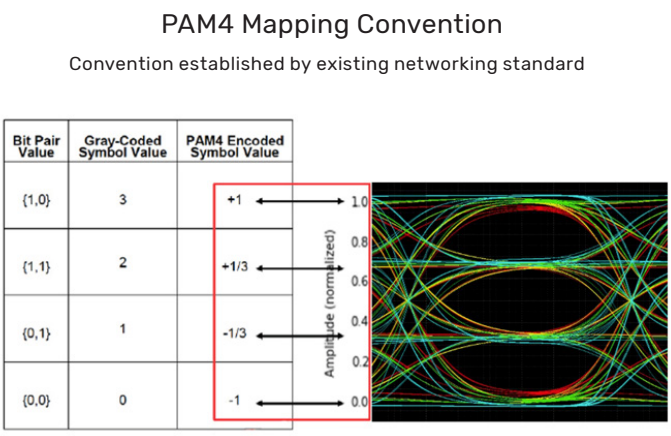

Figure 8: PAM4 to Gray code mapping

Finally, to maintain the low-latency performance of PCIe, the addition of FEC and CRC retry mechanisms in PCIe 6.0 must not increase the latency substantially compared to PCIe 5.0. PCIe 6.0 targets <10ns of latency adder for the inclusion of the FEC function. The PCI-SIG Working Group has shown that for most PCIe use cases, the latency target for PCIe 6.0 could be met (Figure 9).

(X1 Link)

Data Size (DW)	TLP Size (DW)	Latency in ns for 128b/130b @ 32.0GT/s	Latency in ns in Flit Mode @ 64.0 GT/s	Latency Increase due to accumulation (ns)
0	4	6.09375	18	11.90625
4	8	10.15625	20	9.84375
8	12	14.21875	22	7.78125
16	20	22.34375	26	3.65625
32	36	38.59375	34	-4.59375
64	68	71.09375	50	-21.09375
128	132	136.09375	82	-54.09375
256	260	266.09375	146	-120.09375
512	516	526.09375	274	-252.09375
1024	1028	1046.09375	530	-516.09375

(X16 Link)

Data Size (DW)	TLP Size (DW)	Latency in ns for 128b/130b @ 32.0GT/s	Latency in ns in Flit Mode @ 64.0 GT/s	Latency Increase due to accumulation (ns)
0	4	0.380859375	1.125	0.744140625
4	8	0.634765625	1.25	0.615234375
8	12	0.888671875	1.375	0.486328125
16	20	1.396484375	1.625	0.228515625
32	36	2.412109375	2.125	-0.287109375
64	68	4.443359375	3.125	-1.318359375
128	132	8.505859375	5.125	-3.380859375
256	260	16.63085938	9.125	-7.505859375
512	516	32.88085938	17.125	-15.75585938
1024	1028	65.38085938	33.125	-32.25585938

Figure 9: Latency comparison for X1 Link and X16 Link

Looking Ahead

As of this writing, PCIe 6.0 is at version 0.7. Although it is challenging on multiple fronts to achieve doubling of the data rate in PCIe 6.0, much of the technical barriers related to the speed increase and PAM4 adoption have been overcome. The arrival of PCIe 6.0 is expected to enable the next generation of innovations in data centers, AI/ML, and cloud computing.

Cadence is leading the way to bring PHY and controller solutions for PCIe 6.0 to the mass market. Leveraging state-of-the-art PAM4 technologies from Cadence's extensive portfolio of production-proven 112G/56G PAM4 Ethernet PHY, the company is well-positioned to deliver the latest PCIe evolution that will push the frontier of Intelligent System Design™.

The Cadence PHY IP for PCIe 6.0 will be available in selected mainstream FinFET process nodes from leading foundries.

References

1. D. Das Sharma, "PCIe 6.0 Specification: The Interconnect for I/O Needs of the Future." Part of the PCI-SIG Educational Webinar Series, presented June 4, 2020: https://pcisig.com/sites/default/files/files/PCIe 6.0 Webinar_Final_.pdf
2. Y. Frans et al., "A 56-Gb/s PAM4 wireline transceiver using a 32-way time-interleaved SAR ADC in 16-nm FinFET," IEEE J. Solid-State Circuits, vol. 52, no. 4, pp. 1101–1110, Apr. 2017: https://ieeexplore.ieee.org/document/7811205
3. Intel, "AN 835: PAM4 Signaling Fundamentals", March 12, 2019: https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/an/an835.pdf