

# Data timing is critical in distributed systems

Synchronizing data to an accurate clock through time stamping and other methods reduces transmission delay and skewing in distributed systems.

By Fred Mohamadi

A distributed system consists of autonomous computing and processing nodes linked together as a network. It is supported by software and hardware that allow the system to operate as an integrated entity. Moving toward an integrated distributed environment, the ability to determine accurately the timing of the data as it moves between elements of the system is critical.

Several issues arise when considering the timing of data elements as they move from measurement through the network to the processing elements to the human machine interface (HMI). Latency or delay due to transmission of the data between points in a network can cause a significant impact to accurate processing and interpretation of data. For example, even short latency between the measurement of an aircraft navigation sensor and its input and processing by a mapping program can cause large errors in aircraft position displays.

External clock synchronization relates to the problem of using a universal time coordinated (UTC) reference for all clocks in a system. However, for widely distributed systems, inaccuracy in the measurement of the UTC reference and the limited accuracy of that reference (typically several milliseconds) limit the ability to clock the data elements with sufficient resolution.

Clock accuracy has improved due to the advent of global positioning system (GPS) technology. However, limitations to GPS timing occur due to omissions and errors with the GPS receivers. Additionally, accuracy below 1 nanosecond (ns) is currently desired, which requires that GPS technology (down to ~100 ns) be complemented by additional techniques.

## Global clock distribution

Figure 1 indicates the clock distribution management in two and three dimensions. The transmitter sends known patterns as part of the synchronization process to address global clock distribution. The transmitter in this configuration sends the master clock signal generated from a low phase noise VCO or highly stable crystal oscillator. The transmitted master clock propagates through a wired or wireless media to the receiver blocks for time stamp information. The receiver block receives the signal and treats the signal as data. Local clock and

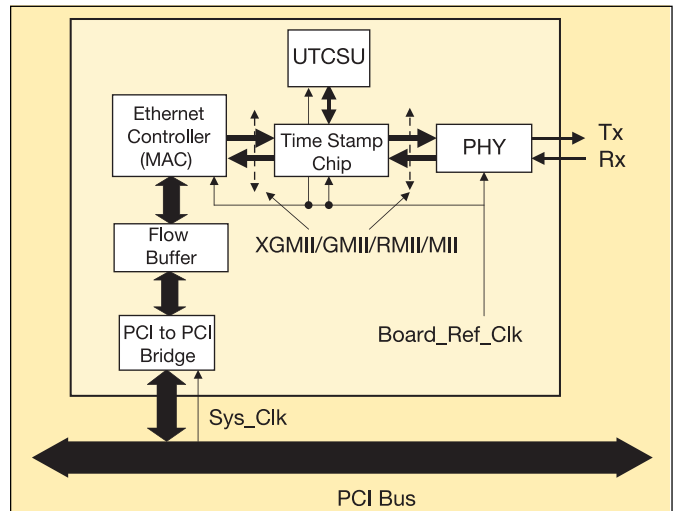


Figure 2. Demonstration of clock synchronization and time stamping.

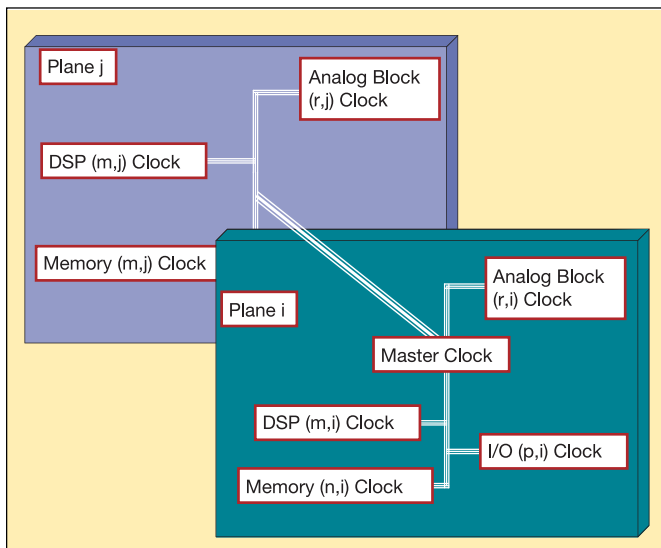


Figure 1. Demonstration of a clock distribution system.

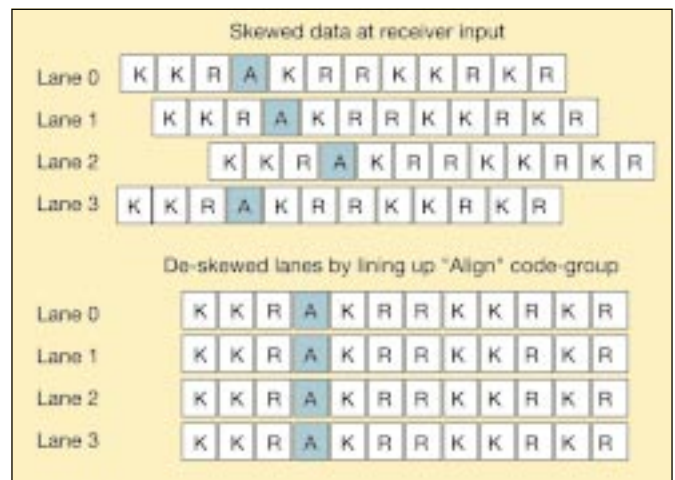


Figure 3. Implementation of XAUI-agnostic de-skewing.

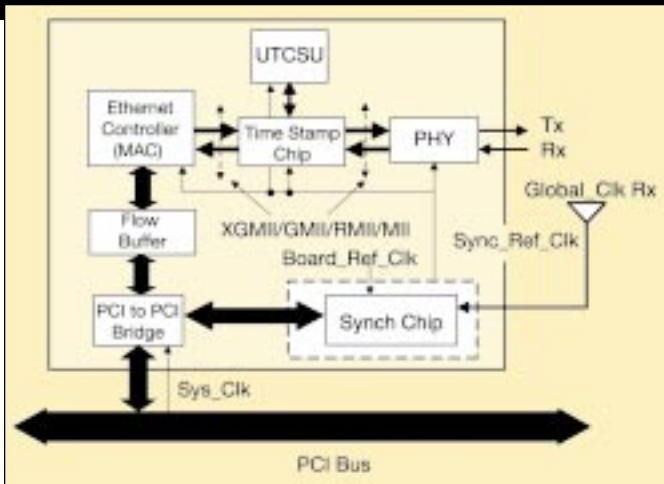


Figure 4. Proposed global clock synchronization and time stamping.

received signals are fed to a phase-locked loop (PLL).

The recovered data from the master clock are then processed further to retrieve the relative and the absolute time elapses with respect to the global clock. The information is used to manage global synchronization within the distributed system.

*Experiments revealed, however, that the original implementation using internal packet FIFOs of the network controller prohibit a further reduction of transmission delay for the memory-based time stamping method.*

### Global clock distribution in a wired environment

Figure 2 is a proposed solution<sup>1,2</sup> using the local clock provided in a custom ASIC-termed universal time coordinated clock synchronization (UTCSU). Time stamps were triggered and inserted in packet memory when the network controller grabbed and deposited a packet there. Experiments revealed, however, that the original implementation using internal packet FIFOs of the network controller prohibit a further reduction of transmission delay for the memory-based time stamping method. A further refinement reflected in Figure 2 was addressed<sup>2,3,4,5</sup> to implement a media-independent interface-based time stamping method. The approach further reduced the transmis-

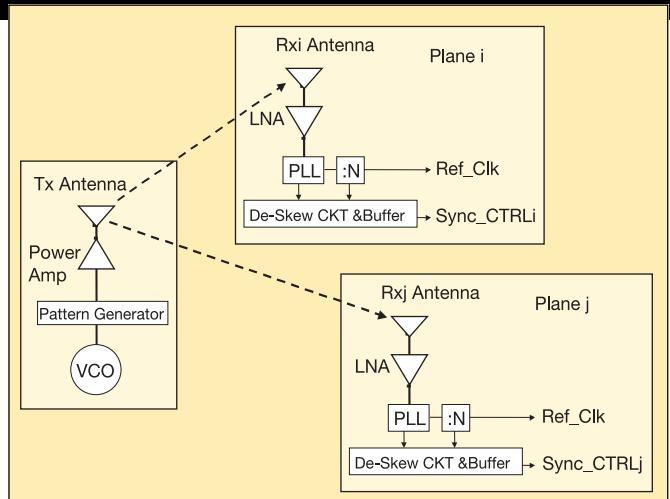


Figure 5. Proposed global wireless blind synchronization.

sion delay uncertainty.

An Ethernet-based burst mode packet is re-shaped by being converted to digital data after a physical layer. After parsing the packet at the IEEE 802.3<sup>6</sup> proposed media-independent interface (MII), time

stamping information is inserted in allocated tagging fields dedicated for quality of service (QOS) using UTCSU. The drawbacks of this approach are complexity of a co-processor (UTCSU), and latency due to the buffering while accessing the header fields. Further latency adds up while adjusting the frame check sequence (FCS), which is a 32-bit CRC calculation of destination/source address, type/length field, data and optional pad bytes. The system clock that controls the interface to a processor (PCI in Figure 2) has a different phase and frequency than the clock supplied from the board. The deskewing of transmit/receive data then requires proper buffering using FIFO or similar design. Further deterioration of latency will occur.

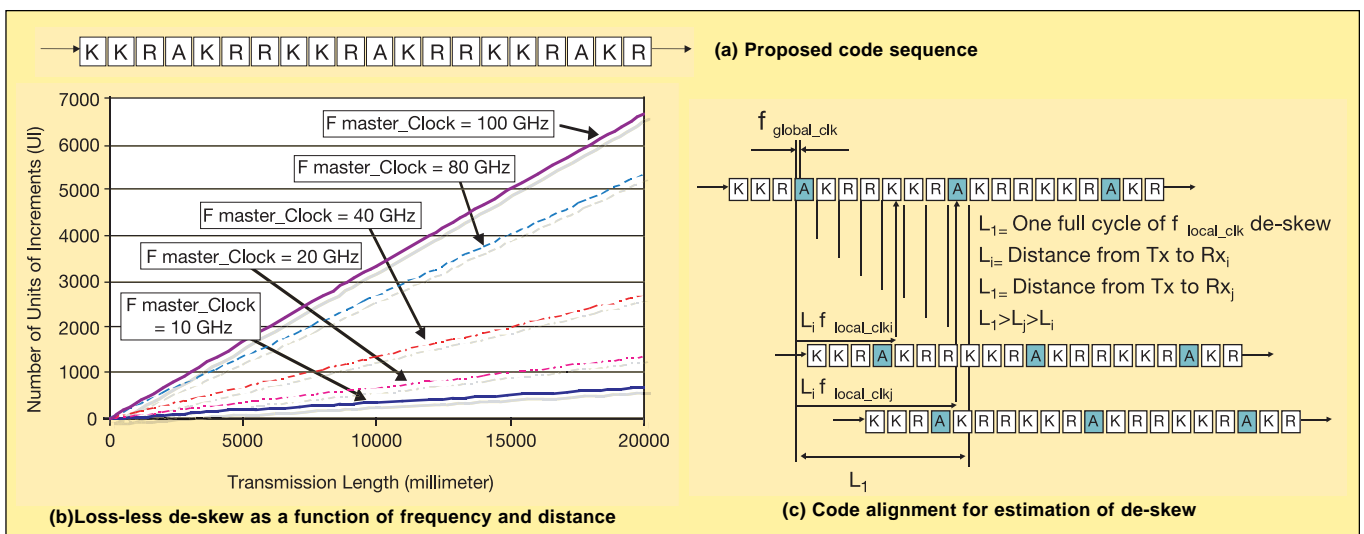
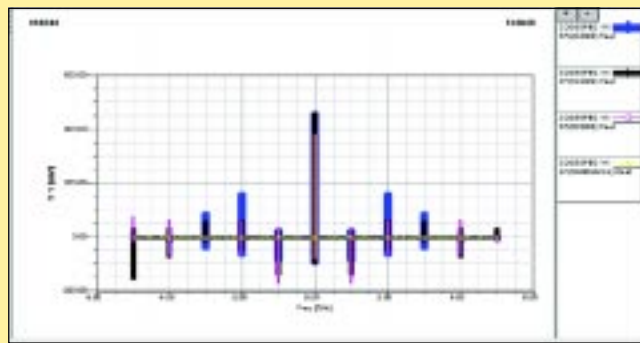
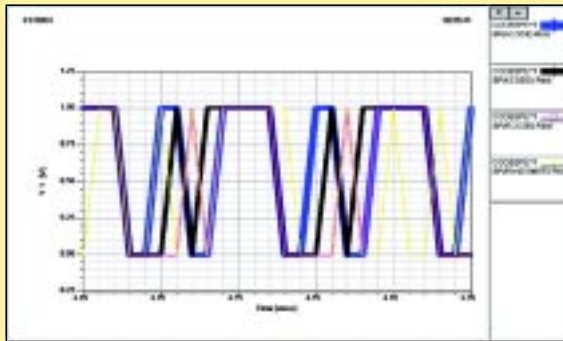


Figure 6. Code repetition and synchronization of clock vs. distance.



(a) Spectra



(b) Waveforms

Figure 7. A, R, K and random codes at 10 GHz.

### IEEE 802.3ae wired synchronization scheme

Extension of the IEEE 802.3<sup>6</sup> code words are used for wired and wireless synchronization. Figure 3 indicates a method commonly used for synchronization of multi-lane signals implemented in a wired environment. According to the IEEE 802.3ae (10 Gbps Ethernet) standards, code words of “K” referred to as idle synchronized character (K28.5); “R” referred to as idle skip character (K28.0); “A”

*An alternative approach is a proposed globally transmitted clock distribution mechanism to implement synchronization spatially to the best of attained accuracy and free of interference*

referred to as idle align (K28.3) are used to address synchronization of the lanes. A code word “E” referred to as error (K30.7) is used to address the occurrence of error or loss of synchronization. Accordingly, it is proposed that “A”, “K”, and “R” codes be used as source of transition to transmit the master clock wirelessly. Additional steps can be taken to scramble (randomize) these codes such that no spectral radiation can be generated to deteriorate the transmitted signal. An optional randomizer can be a  $X^7+X^6+1$  or  $X^7+X^3+1$  type number generator. The “K”, “R”, “A”, and “E” codes are unique such that no true data can be identified with these codes in an 8B/10B coding scheme. Accordingly, the unique order of information embedded in these codes is easily decoded. Another importance of 8B/10B encoding and decoding is the equal number of zeros and ones in a code. This coding property ensures the maximum transition edge and hence sustains the locking function of the local PLL.

### Proposed global clock distribution

An alternative approach is a proposed globally transmitted clock distribution mechanism to implement synchronization spatially to the best of attained accuracy and free of interference (Figure 4).

Various methods of transmitting master clock using differential pair and swapping dual metal layers have been proposed to carry the

clock across distributed systems. The inter-symbol interference (ISI), skin effect and coupling through parasitic capacitance, however, significantly degrade the signal quality and amplitude at these speeds.

A clock signal that is orders of magnitude faster than the processing (DSP, memory and I/Os) clock of the desired distributed system is generated through a master VCO and transmitted via a dipole or microstrip antenna. The oscillation is also further coded with the stamped data, alignment bits and “dummy” bits to sustain a robust and flawless time stamp information. The master clock and time stamping data are received at the receiver (Figure 4). The clock is then recovered and divided by a programmable integer number (clock reduction factor) such that a clock close to the local clock frequency (Board\_Ref\_Clk) is provided to the local functions (DSP, controller, memory, I/Os, etc.). The daughter-board can be easily added to the existing network interface card (NIC). The time stamp is converted to a field or control signal that can be passed to the CPU via a bus such as PCI or PCMCIA. The global master clock transfers the time stamp that is received from a source such as a GPS-based system or a high-precision reference system. This function can eliminate the need for UTCSU or provide an augmentation to the system.

### Proposed wireless synchronization algorithm

Various methods of coding may be used to ensure synchronization to a global clock transmission through wire or wireless systems. A conceptual diagram of such a global clock transmission is illustrated in Figure 5 in which a master VCO couples its output to a pattern generator. For example, if each VCO forms part of a PLL, the coding must ensure sufficient signal transitions to sustain the edges necessary for PLL to achieve lock. The data and clock may be encoded together such that a “global clock” transmission may represent both a global clock and time stamp. After time stamp coding by pattern generator and amplification by a power amplifier, the resulting global clock signal is transmitted through air by feedline/receptors. At the receiver block, each slave feedline/receptor couples to a low-noise amplifier. In turn, each low-noise amplifier couples to a PLL. After de-skewing

using a module in response to the coding provided by pattern generator, divided-down reference clocks and synchronization signals (time stamp) are available for local use.

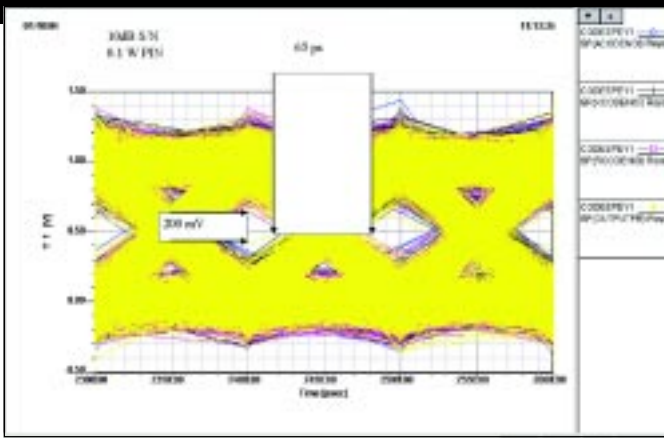
The skew associated with propagation is determined by the actual voltage wave  $v(x)$  that propagates through a wired or wireless media as a function of the propagation distance  $x$ . The voltage wave  $v(x)$  may be expressed as:

$$u = \frac{1}{\sqrt{L_u - C_u}}$$

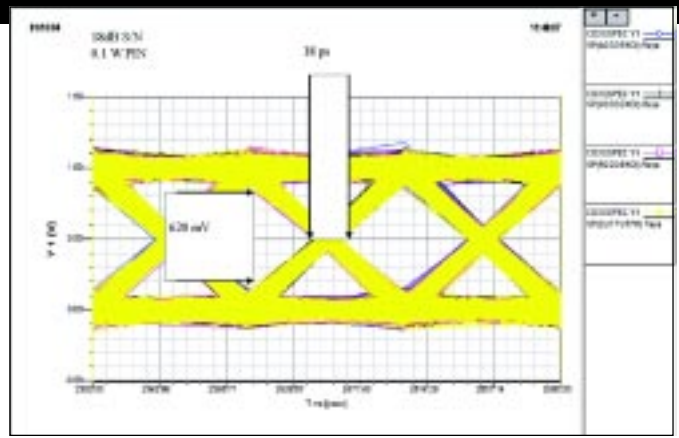
$$v(x) = v_i \cdot e^{-\alpha x + j\beta \cdot x}$$

$v$  = propagation velocity;  $L_u$  = inductance per unit length;  $C_u$  = capacitance per unit length;  $\alpha$  = resistive loss (negligible in the proposed scheme), however, restrictive in a microstrip line used for clock distribution,  $0.05 \text{ mm}^{-1}$  for a 3 micron eighth-metal line;  $\beta = 2 \cdot \pi/\lambda$ ;  $\lambda$  = propagation wavelength.

To address this skew, a pattern generator may generate a sequence of “K”, “R”, and “A” codes as illustrated in Figure 6a. In this code sequence, the “A” code is transmitted at the start of transmission and after a “K”, “R”, “KR”, “KRRK”, “KRRKRR”, or “KRRKRRK” code sequence has been transmitted. Thus, the final transmitted code sequence is AKA, ARA, AKRA, AKRRKA, AKRRKKRA, or



(a) Received with 10dB S/N.



(b) Received with 18dB S/N.

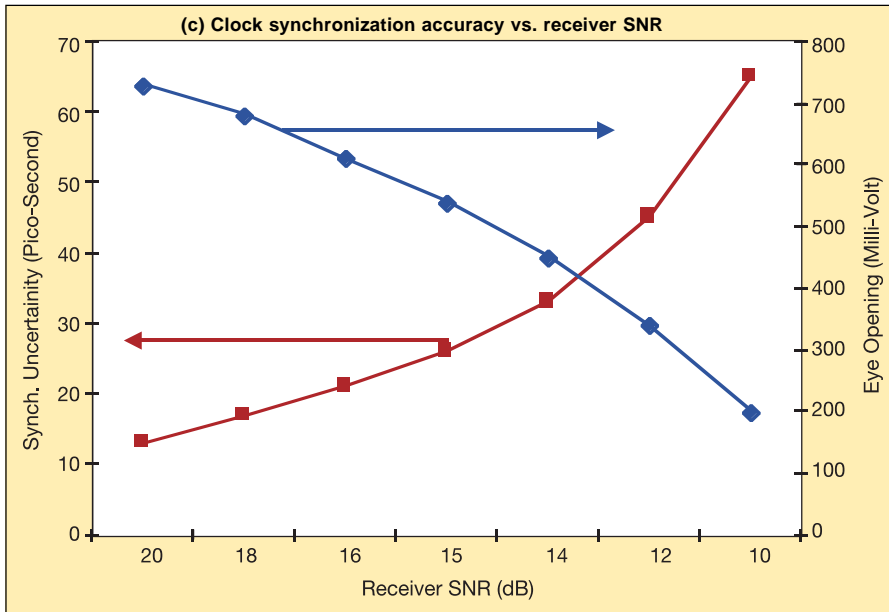


Figure 8. Eye diagram of A, K, R codes at receiver.

AKRRKKRKA.

In this fashion, depending on the transmission frequency and the propagation distance between a transmitting feedline/receptor and a receiving feedline/receptor (Figure 6), a receiving unit may, after receiving an initial “A” code, make an assumption about the number

of a normal distribution.

The transition of the “K”, “R”, and “A” codes guarantees the locking of the receiving PLLs. The 7-bit comma string preceding each symbol in the previously-mentioned K28.5 code may be defined as b’0011111 (comma+) or b’1100000 (comma-). An associated proto-

*With the proper amount of buffering achieved, for example, through the use of innovative high-speed stack or barrel shifters, the de-skewing function between local clocks occurs.*

of transmission cycles that may have expired. An example of suitable A, R, and K codes is:

A = 28.3 = 001111 0011,  
 K = 28.5 = 001111 1010,  
 and R = 28.0 = 001111 0100.

Figure 6b is a graphical representation of the number of cycles generated as a function of propagation distance (in mm) and transmission frequency. Analysis of Figure 6b indicates that a 40 GHz transmission will complete about 3000 UI (unit of increment) while propagating a distance of 20 meters. Accordingly, if the “AKRRKKRA” sequence (80 UI) repeats eight cycles at 40 GHz,

col ensures that “comma+” is transmitted with either equivalent or greater frequency than “comma-” for the duration of the transmission to ensure compatibility with common components.

The comma contained within the /K28.5/ special code group is a singular bit pattern that cannot appear in other locations of a code group and cannot be generated across the boundaries of two adjacent code groups in the absence of transmission errors.

A graphical representation of the propagation delay between a pattern generator generating the K28.5 code and two receiving PLLs is illustrated in Figure 6c. After transmission of an initial “A” code, different amounts of propagation delay is encountered at the recei-

ing PLLs, each receiving a delayed “A” code, respectively.

With the proper amount of buffering achieved, for example, through the use of innovative high-speed stack or barrel shifters, the de-skewing function between local clocks occurs.

Figure 7a reflects the spectra of various codes, indicating its neutral distribution for the “K” and “R” pair used in conjunction with the “A” code. This neutral energy distribution is ideal to minimize the impact of electromagnetic radiation interference. The DC component of the spectra can easily be filtered out with a high-pass filter.

### Time accuracy system implementation

Spectra of various code words were simulated to provide a baseband system that can be used to transmit and receive the global clock (Figure 8). No mixer is used as the signal transmission, and reception is at the speed of the carrier frequency (10 GHz in the simulated example). Code-word generating transceiver chips compliant to IEEE 802.3ae are available as part of the 10 Gb-Ethernet chips from many vendors <sup>7</sup>.

Figure 8 presents uncertainty (jitter) associated with the edge of a received master reference clock due to the environmental noise, frequency and phase drift and signal impairment. The figure also indicates the available voltage swing for signal detection (the larger eye opening is the more accurate synchronization). Figures 8a and 8b are the plots of eye-opening amplitude and total jitter for various strength of received signal (global master clock). Clock synchroniza-

tion accuracy in the order of 100 picoseconds or better can be obtained with 300-400 mV eye-opening (Figure 8c), securing a robust design implementation in CMOS base products<sup>8</sup>. **DE**

### References

1. “Continuous Satellite Two-Way Time Transfer using Commercial Modems,” T. P. Celano, et al, Paper published by the Timing Solutions Corporation.
2. “NTI: A Network Time Interface M-Module for High-Accuracy Clock Synchronization,” M. Horauer, et al, Proceedings of the 6th International Workshop on Parallel and Distributed Real-Time Systems (WPDRTS), Orlando, Fla., March 30-April 3, 1998.
3. “Experimental Evaluation of High-Accuracy Time Distribution in a COTS-based Ethernet LAN,” U.Schmid and H.Nachtnebel, Proceedings of the 24th IFAC/IFIP Workshop on Real-Time Programming (WRTP’99), Germany, June 1999, pp. 59-69.
4. “Integrating External and Internal Clock Synchronization,” C.Fetzer and F.Cristian, Journal of Real-Time Systems,” May 1997, No. 3, vol. 12 (2), pp. 123-172.
5. “NIST Computer Time Services: Internet Time Service (ITS), Automated Computer Time Service (ACTS),” J Levine, et al, time.gov web sites.
6. IEEE Stand-ards 802.3ae, “Media Access Control (MAC) Parameters, Physical Layer, and Management Parameters for 10 Gb/s Operation,” <http://grouper.ieee.org/groups/802/3/ae/public/>
7. “A 10 Gbps (4 X 3.125 GBd) Transceiver,” F. Mohamadi, [http://grouper.ieee.org/groups/802/3/ae/public/may00/mohamadi\\_1\\_0500.pdf](http://grouper.ieee.org/groups/802/3/ae/public/may00/mohamadi_1_0500.pdf).
8. “Si Integration with Millimeter Wave Phased Array Antenna,” F. Mohamadi, *RF Design*, February 2004, pp. 44-48.

#### ABOUT THE AUTHOR

Fred Mohamadi is founder and CEO of TaiLinx based in Irvine, Calif., and holds a Ph.D. in electrical engineering from Stanford University and an MBA from Santa Clara University. He owns three U.S. patents and has published dozens of technical papers.