Implementing Synchronous Ethernet in Telecommunication Systems

James Aweya, *Senior Member, IEEE*

Abstract—Network infrastructures are gradually migrating from time-division multiplexing (TDM) based onto packet-based architectures. In spite of this convergence, there are a significant number of synchronous applications that require accurate timing to be distributed over the packet networks. Examples of precision timing sensitive applications that need the transport of synchronization over packet networks include interconnection and transport of TDM services over packet networks (TDM switches, TDM PBXs, voice, video-conferencing and broadband video), and connections to 2G, 3G, and 4G wireless base stations. TDM networks, unlike packet networks (e.g., Ethernet, IP, MPLS), have timing transfer inherently built into them. Native Ethernet (IEEE 802.3) is inherently asynchronous and was not designed with timing transfer in mind. Synchronous Ethernet (Sync-E), defined by the ITU-T, has emerged as a powerful, yet simple technology, for accurate timing transfer over Ethernet networks using "TDM-like" (precisely, SDH/SONET) timing techniques. This discussion explains what Sync-E is, followed by a detailed discussion on which flavors of Ethernet can support Sync-E and which cannot. The discussion includes how Sync-E can be implemented in the popular Ethernet versions. We then describe example Sync-E node timing architectures, and some network timing applications and related issues.

Index Terms—Synchronous Ethernet, clock synchronization, timing and synchronization, timing distribution, clock recovery, phase-locked loop, clock and data recovery, mobile backhaul.

I. INTRODUCTION

TELECOMMUNICATION (telecom) carriers and network service providers are designing the next-generation IP based networks to meet the demand for more bandwidth, faster and more reliable services. Ethernet (IEEE 802.3) [1] has emerged as the transport medium of choice for the packet based networking. In general, Ethernet, and complementary technologies like IP and MPLS have gained industry-wide acceptance for deployment in provider network infrastructures due to enhancements in Quality of Service (QoS), Operations, Administration and Maintenance (OA&M), congestion management, and resiliency. Ethernet also gives more flexibility for integrating voice, video and data services.

However, with the transition from TDM to packet technologies, telecom network timing and synchronization will need to evolve to support the emerging packet based infrastructure. A critical piece missing from a total convergence to Ethernet is the ability to provide timing natively within the packet network. This would provide Ethernet with the capability to transport timing-sensitive applications (such as transport of

TDM and circuit emulation services (CES) over packet) as well as distribute precise frequency references.

For example, synchronization plays a crucial role in mobile backhaul networks [30], [31]. Mobile wireless base stations (both Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD) technologies) derive their carrier radio frequencies from a highly accurate reference clock to a quality level within 50 parts-per billion (ppb). In case of FDD operation, there are two carrier frequencies, one for uplink transmission and one for downlink transmission. In case of TDD operation, there is only one single carrier frequency and uplink and downlink transmissions in the cell are always separated in time. To avoid severe interference between uplink and downlink transmissions despite the fact that the two links use the same frequency, the cells in a TDD network typically use the same uplink downlink configuration together with inter-cell synchronization to a common time reference to align the switch-points among all the cells. This avoids interference between the two links as uplink and downlink transmissions do not occur at the same time.

The reference clock is typically derived from TDM (E1/T1) interfaces linking the base station to the switching centers (in FDD networks) or from expensive GPS receivers located at the base station (in TDD networks). GPS receivers used for telecom network synchronization have a much higher specification (high quality oscillators, high holdover capability accuracies, etc.) than those in the average portable satellite navigation system, plus they need all the right interfaces and cabling to communicate with the telecoms equipment.

The Primary Reference Clock (PRC) is the master clock from which all other clocks in the synchronous network directly or indirectly derive their timing. This hierarchy of time synchronization is essential for the proper functioning of the network as a whole. A clock that ultimately derives its rate from the PRC is said to be traceable to that PRC. Accurate synchronization (of not more than 50 ppb) is crucial for mobile networks because the radios used in these networks operate in very strict frequency bands that need separation to avoid channel interference which reduces the call quality and network capacity. Accurate radio carrier frequencies at the transmitter antenna outputs guarantee that radio channels of the neighboring cells do not overlap in the spectrum used. Spectral channel overlap causes channel cross-interference, which in turn causes noticeable and times degradation of voice quality (i.e., signal-to-noise ratio). Without timing information traceable to a highly accurate PRC, local interference between channel frequencies, as well as mutual interference with neighboring base stations occur, ultimately causing dropped calls and degrading the overall user experience.

Accurate synchronization is also required for successful handover processing (handover between base stations). Mobile units derive the accurate frequencies that they use for transmission and reception from the base stations. If the transmission frequencies are not very closely matched between adjacent cell sites (base stations), then annoying "clicks" can occur when the call is being handed over between base stations. In the worst case, the call would drop because the mobile unit would not be able to immediately lock onto and acquire the new signal. An accurate timing reference in each base station is required in order to prevent call drops during handoffs. Supported with results from experimental trials in a real mobile network, reference [43] shows that poor frequency synchronization of references in GSM Base Transceiver Stations (BTS) can lead to dropped calls, slow handover between cells and co-channel interference. These observations are equally applicable to other mobile technologies like UMTS, CDMA, LTE, and LTE-Advanced since they all share the similar synchronization structures and requirements.

Other examples of synchronization applications can be drawn from industrial automation, power industry, and home audio/video networking (IEEE 802.1AS [2]). In all these application domains, providing accurate timing information natively within an Ethernet framework affords an extremely cost-effective means of distributing timing to each subsystem.

The telecom industry already has clearly defined standards governing TDM network synchronization, from to plesiochronous digital hierarchy (PDH) T1/E1 rates up to synchronous digital hierarchy/synchronous optical network (SDH/SONET) rates. The timing requirements for TDM networks are specified in ITU-T, ETSI and ANSI documents. These standards include well defined guidelines and requirements to deploy network synchronization. Some examples of these standards are as follows. G.781 [42] specifies the basic synchronization distribution building blocks and a set of rules by which they are combined in order to describe the synchronization functionality of a digital transmission equipment. Clause 8 in G.803 [3] describes the distribution of timing reference in an SDH network. G.810 [32] provides definitions and abbreviations used in ITU-T timing and synchronization recommendations in addition to background information explaining why there is the need to limit phase variation and the impairments on digital systems.

G.811 [33] defines the requirements for PRCs suitable for synchronization supply to digital networks. G.812 [34] defines a slave clock commonly known as a synchronization supply unit (SSU). G.813 [10] defines a slave clock known as an SDH Equipment Clock (SEC). A SEC is required to have fairly good, but not excellent, timing accuracy while in holdover mode (thus allowing the use of relatively inexpensive crystal local oscillators), but stringent jitter and wander generation (in order to enable chaining of multiple SECs). An SSU can be employed after a chain of SECs to counteract the accumulation of timing inaccuracies. The SSU's master may be a PRC, another SSU or an SEC.

ANSI T1.TR.81 [4] specifies synchronization network architecture for the SONET hierarchy. ITU-T specifies jitter and wander requirements are in G.823 (E1) [5], G.824 (T1) [6] and G.825 (SDH) [7]. ANSI specifies jitter and wander at network and equipment interfaces in T1.105.03 [8]. The Telcordia GR-1244-CORE for North American T1/DS1 equipment [9], and ITU-T G.813 Option 1 for E1 equipment outside North America [10] are the most common timing standards for T1/E1 equipment. The above reference are not meant to be exhaustive, other sources such as references [35] and [36] and the references therein, provide comprehensive discussions on TDM network synchronization. Readers can also refer to these references for further information on evolution of Telecom network synchronization practices.

The current move to packet technologies offers Telecom network operators the increased backhaul capacity they require for the deployment of high bandwidth data services and, also the cost advantage and design flexibility of packet transport. However, this move has also created the need for synchronization technologies that can replace the current TDM and GPS based ones. Synchronization is a key requirement in Telecom networks and the Telecom industry has become a driving force in the development and evolution of synchronization solutions and standards. Telecom network operators are currently evaluating physical layer and packet-based timing solutions suitable for packet-based transport. Current developments in this area are described in a number of articles some which are the following references [30], [31], [37], [38], [39], [40].

Sync-E is defined in ITU-T standards G.8261 [11], G.8262 [12], and G.8264 [13]. Sync-E provides Layer 1 level (physical layer) synchronization similar to PDH and SDH/SONET networks. IEEE 1588 Precision Time Protocol (PTP) [14] provides synchronization at the Layer 2 and higher over a variety of networking technologies, while Network Time Protocol (NTP) [15] provides synchronization over Layer 3 level with IP routed technologies. It is worth highlighting here the key contributions of three related papers on Sync-E that are tailored for the general reader. Reference [30] describes how the Sync-E standards were developed to allow for the transportation of synchronization information over Ethernet and also allow for interworking with and migration from existing SONET/SDH-based transport infrastructure. The paper [41] describes how Sync-E uses Ethernet physical layer frequency transfer together with the ESMC protocol standardized by the ITU-T for frequency distribution. This article also describes how Sync-E could be extended in order to transport phase/time in addition to frequency. Beyond the applications of Sync-E for mobile backhaul, Aweya in [31] describes emerging and potential applications of Sync-E in Telecom networks such as Sync-E for differential timing transfer, Sync-E as a packet backplane for TDM modules, TDM backplane extension over Sync-E, and TDM backplane expansion over Sync-E.

This paper presents a discussion on the basics of Sync-E and explains how Sync-E can be supported on the existing Ethernet flavors. Ethernet provides interoperability/ interworking between all flavors but is also a technology that has a wide range of physical layer types. These physical layers have evolved through time following advances in electronics and optical technologies, and also industry network practices and needs. The differences in all these physical layer types surely will have implications on how Sync-E is implemented since Sync-E passes timing via the physical layer of Ethernet. This paper discusses Sync-E while recognizing these different physical
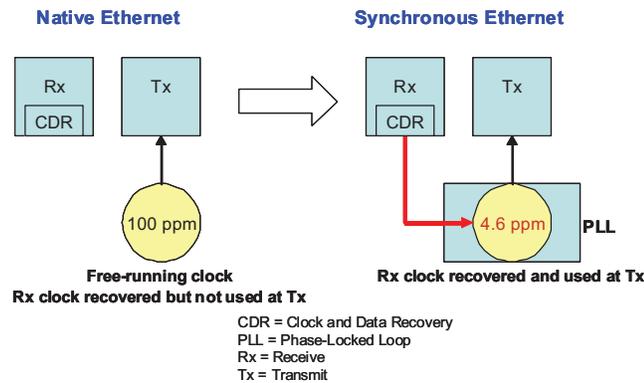
Fig. 1. Timing in native Ethernet vs. Sync-E.

Fig. 2. Timing in synchronous Ethernet.

layer types. The intent of the paper is to explain this emerging technology and link the discussion to common industry practices in timing and synchronization. Digging through the relevant standards and understanding vendor implementations and industry practices can be a nightmare. This intent of the paper is to present in a tutorial style the relevant material that leads to the understanding and implementation of Sync-E by tying in all the relevant material from the standards and industry practices.

In the paper we focus only on the implementation of Sync-E over the most common and widely used 100 Mb/s (100BASE-TX/FX), 1000 Mb/s (1000BASE-CX/SX/LX and 1000BASE-T), and 10Gb/s (10GBASE-R/W) Ethernet technologies. 1000BASE-T and 10GBASE-T, also common technologies, need special handling for them to be used in Sync-E. In addition, we describe how Sync-E timing transfer is handled in the Ethernet media independent interface types for each Ethernet technology. We explain why Sync-E cannot be implemented over 10 Mb/s Ethernet. We describe some generic Sync-E architectures and show how these architectures can be used in example networking timing applications.

II. NATIVE ETHERNET (IEEE 802.3) VERSUS SYNCHRONOUS ETHERNET (ITU-T REC. 826X): TIMING FEATURES

This section describes the timing related features in native Ethernet versus Sync-E. The next section describes the main features of Sync-E according to the ITU-T standards. In native Ethernet the physical layer (PHY) transmit clock is typically derived from an inexpensive ±100 ppm free-running local oscillator. A receive PHY (i.e., the PHY at the receive end of an Ethernet link) locks onto this transmit frequency using clock recovery methods (a clock and data recovery (CDR) unit) to sample the received data (Fig. 1). The Preamble in the Ethernet frame allows the receive PHY to lock onto the incoming bit stream and the Start-of-Frame Delimiter (SFD) indicates the beginning of the encapsulated data [1]. This simple arrangement and appropriate buffering allows a receive PHY to lock onto the bit stream from the upstream PHY and extract the relevant transmitted data without the need for frequency synchronization (as in the TDM case) as packets are buffered at each hop.
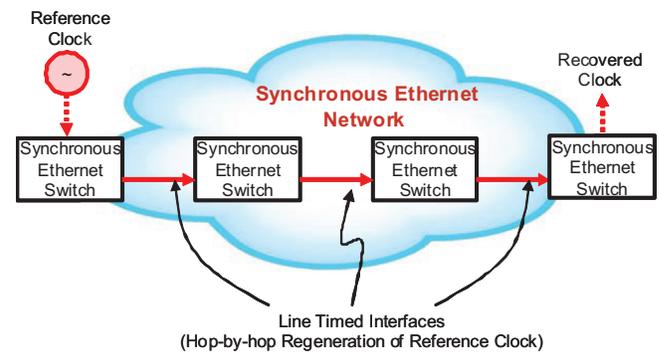
Although Ethernet operates perfectly well with the above arrangement, Ethernet does not preclude the use of high-accuracy frequency references at each node. One could opt to derive the Ethernet transmit clocks from a clock traceable to a PRC which is exactly what timing in Sync-E defines (Fig. 1). By allowing a PHY to recover a PRC traceable clock using low-jitter circuitry and then using this signal as the local transmit clock, a timing path can be built at the physical layer traceable back to the PRC (Fig. 2). PRC timing is embedded into the physical layer path by driving the local transmit clocks with the high-accuracy frequency reference. The ±4.6 ppm (parts per million) indicated in Fig. 2 refers to the frequency accuracy of a timing device in a Sync-E network in free-running state (see G.8262). This means if a device or interface loses synchronization to the PRC, the standards specify that it should still be able to deliver ±4.6 ppm after a specified free-run period. Thus, based on the ITU-T standards, Sync-E has a clock accuracy of ±4.6 ppm - similar to the range found on SDH/SONET, a great improvement on the native (asynchronous) Ethernet limit with a free-running clock of ±100 ppm accuracy. Note that jitter and wander can accumulate at each hop in the Sync-E network, degrading the quality of the recovered clock. However, this is no different from the timing distribution issues already encountered in PDH and SONET/SDH networks and, so with carefully engineered clock recovery methods and proper selection of the type of clocks used in the network, signal degradation can be clearly characterized and controlled. ITU-T G.8262 specifies the minimum requirements for timing devices used in synchronizing network equipment that supports Sync-E.

Synchronization in native Ethernet is only between two adjacent nodes, and is not passed from one hop to another. Sync-E, however, is capable of passing timing from one hop to another. In Sync-E, the recovered signal is cleaned with a phase-locked loop (PLL) to remove incoming network generated jitter and wander in addition to jitter generated from the clock recovery circuit, before being fed to the transmitting device. It should be noted here that, clock recovery is only possible if the Ethernet PHY sends continuous transmissions (pulse transitions) even in idle period where user data is not sent. The recovered clock from the node receiving synchronization feeds all nodes that are transmitting synchronization. Sync-E employs Ethernet's physical layer to forward timing signals

from node to node, much like the timing distribution technique used in SONET/SDH. The benefit of this approach is that the Ethernet data path functions normally, completely unaffected by the timing path built at the physical layer. This timing path remains isolated from Ethernet packetization issues such as drops and packet delay variations. Using physical layer timing transfer techniques as in SONET/SDH, Sync-E solves the problem of synchronization distribution in a packet switched network context where traffic switching or routing is done in an asynchronous way. It is important to note that Sync-E only provides frequency transfer, and does not provide phase or time-of-day transfer which can be done using IEEE 1588 PTP [14] or NTP [15].

III. SUMMARY OF SYNCHRONOUS ETHERNET SPECIFICATIONS AND REQUIREMENTS

Sync-E is a method of distributing frequency over Ethernet Physical Layer as defined in a *suite* of ITU-T Recommendations (G.8261, G.8262, G.8264). This method of frequency transfer is based on the well-established SONET/SDH synchronization principles widely adopted in the telecom industry. Sync-E enables the migration of carrier network from SONET/SDH-centric to Ethernet-centric infrastructure and also the interworking of SONET/SDH synchronization and Ethernet synchronization. Sync-E reuses SONET/SDH principles to allow Sync-E to interwork with SONET/SDH synchronization networks. The Sync-E specifications also ensure interworking with native Ethernet equipment.

Now we list below some of the main features of Sync-E as a synchronization solution as gleaned from the Sync-E standards. We also highlight the similarities to SONET/SDH synchronization:

A. ITU-T Recommendation G.8261

The key elements of G.8261 are as follows:
- G.8261 discusses the challenges in TDM-packet interworking, circuit emulation services (CES), and synchronization transport over packet networks as well as related issues like wander budget definition and performance characterization in the presence of packet delay variation (PDV).
- G.8261 also describes functions that are applicable to the different modes of CES (network-synchronous solutions, and differential and adaptive methods) and provides guidance on the deployment of synchronization solutions.
- G.8261 defines the network limits at the output of the Synchronous Ethernet Equipment Clock (EEC) in a synchronization chain. The EEC is conceptually similar to the slave clock known as SDH Equipment Clock (SEC) defined in G.813 for SONET/SDH networks.
  - Just like in SONET/SDH, frequency transfer in Sync-E is done hierarchically using an interconnection of clocks. At the root of the tree is the PRC. Each and every Sync-E Network Element contains an internal clock called the EEC. Normally the EEC is locked in frequency to one of its incoming traffic line signals, while all outgoing line signals are timed by the EEC. This way a master-slave tree is built

where synchronization travels from the PRC down to all the clock chains formed by the branches of EECs. The same transfer technique has been used in telecommunication for a long time, notably in synchronization based on E1/T1 and SDH/SONET.
  - An EEC (or SEC) is required to have fairly good timing accuracy (but not excellent as in an SSU (see below)) while in holdover mode, but stringent jitter and wander generation (in order to enable chaining of multiple EECs). This lower accuracy allows the use of relatively inexpensive crystal local oscillators in an EEC. An SSU can be employed after a chain of EECs to control the accumulation of timing inaccuracies.
- Annex A of G.8261 defines the network architecture for Sync-E. This work extends the scope of the G.803 reference synchronization chain to Sync-E equipment.

B. ITU-T Recommendation G.8262

The key elements of this specification are as follows:
- G.8262 defines the performance characteristics of EECs. G.8262 specifies the clocks for Synchronous Ethernet equipment to ensure that Synchronous Ethernet clocks are compatible with SONET/SDH clocks as defined in G.813 and G.812.
  - Two options are defined for EEC: 1) EEC Option 1 to support G.813 Option 1 which applies to Sync-E equipment designed to interwork with networks that operate in the 2048 kb/s hierarchy (SDH networks). 2) EEC Option 2 to support G.812 Type IV which applies to Sync-E equipment designed to interwork with networks that operate in the 1544 kb/s hierarchy (SONET networks)
- G.8262 specifies clock parameters compatible with G.813: The recommendation defines requirements for clock accuracy, noise transfer, holdover performance, noise tolerance, and noise generation.
- G.8262 ensures full compatibility with the G.803 SDH reference chain. A mix of SEC and EEC can be done in the G.803 SDH reference chain.
  - Current telecom and service provider networks are implemented as a mix of SDH/SONET and Ethernet technologies. It is not uncommon to see SDH/SONET in the core network and Ethernet in the aggregation and access networks. The network node interfacing the SDH/SONET and Ethernet parts have multi-service ports and functions (e.g., VC cross-connecting function, packet (frame) switching, etc.) to allow for the interworking of these technologies. With this interworking arrangement and the fact that Sync-E and SDH/SONET share similar, in fact, the same clock synchronization attributes, it makes perfect sense to interwork Sync-E and SDH/SONET for clock distribution purposes (see Fig. 3).
  - The synchronization chain consists of EECs and clocks called Synchronization Supply Unit (SSU) (Fig. 3). Like in SDH/SONET, SSUs are needed to reduce the accumulated jitter and wander to levels
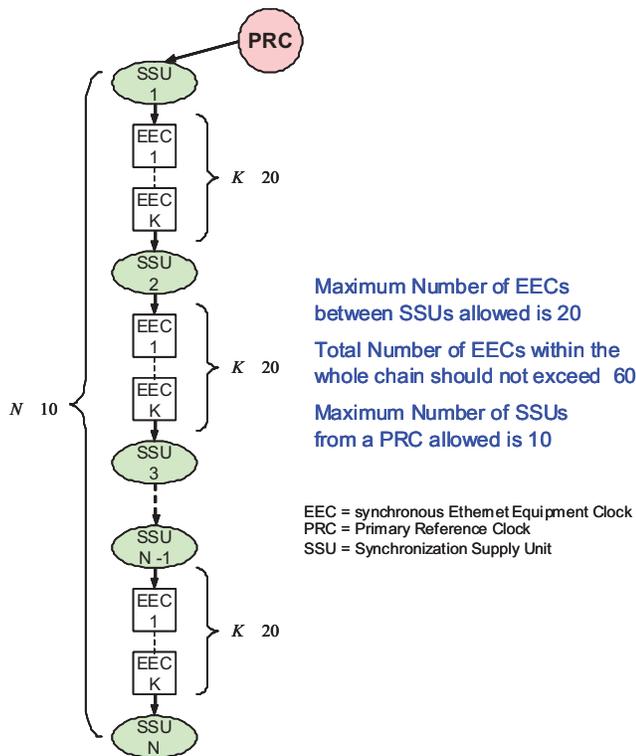
Fig. 3. Sync-E and SDH synchronization network reference chain.

which are compliant with the relevant limits and to act as a node clock, i.e. the central clock of a telecom node or building. This node clock distributes synchronization to other equipment in the building. The SSU also provides holdover protection which kicks in when the traceability to the PRC is lost because of some failure. With this feature, the SSU becomes an autonomous frequency source running on its internal high-quality oscillator (e.g., ovenized quartz or a Rubidium type oscillator). In holdover mode synchronization supply continues with a stability which, while not as good as in locked mode, allows the site to continue operating until the failure is repaired.

- G.8262 specifies Sync-E, STM-N and PDH as interfaces for EEC

To illustrate the compatibility between Sync-E and SDH/SONET, Fig. 4 shows a Sync-E and SDH/SONET interworking example where Ethernet frames (over a Sync-E network) are sent on a SDH/SONET link/network. Ethernet frames are sent through an "encapsulation" block (e.g., Generic Framing Procedure (GFP)) to create a synchronous stream of data from the Ethernet frames. The synchronous stream of encapsulated data is then passed through a mapping block which typically uses virtual concatenation (VCAT) to route the stream of bits over one or more SDH/SONET paths. The SDH paths may be VC-4, VC-3, VC-12 or VC-11 paths.

C. ITU-T Recommendation G.8264

This recommendation has the following major elements:

- G.8264 defines frequency transfer using Sync-E: It provides general information in addition to the operational modes (synchronous and non-synchronous).
- G.8264 defines the Synchronization Status Message (SSM) protocol and formats for Sync-E. G.8264 also contains the specification of a synchronization signaling channel called the Ethernet Synchronization Messaging Channel (ESMC). The ESMC protocol carries the SSM messages used by Sync-E to exchange timing signal traceability information across a link.

  ○ The synchronization signaling system in SDH/SONET is known as the Synchronization Status Message (SSM). The SSM which is a message contained in the multiplex layer overhead carries an information about the quality level of the source clock from clock to clock along the branches of the synchronization distribution tree. This signaling system is used for controlling protection switching in case of link or clock failures. Sync-E uses a similar system like SSM which works in exactly the same way as in SDH/SONET with exception to the communication channel used for transferring the SSM from clock to clock. In SDH/SONET the SSM is contained in the SSM Byte (SSMB) of the STM-n or OC-n frame overhead. Sync-E uses the ESMC which consists of special Ethernet frames defined in G.8264 [13]. The SSM in Sync-E is carried by an Ethernet protocol based on the IEEE Organization Specific Slow Protocol (OSSP). For its use, the ITU-T is assigned a Slow Protocol subtype and an Organizational Unique Identifier Identifier (OUI) by the IEEE. The ITU-T has in turn assigned an ITU-T subtype to identify the ESMC Protocol Data Unit (PDU). Annex A of G.8264 defines the reference source selection mechanism that can be used by Sync-E equipment to provide timing signal traceability to upstream elements and ultimately the PRC.

  ○ The SSM carries the Quality Level (QL) of the clock (source) which is used for the synchronization selection process at a node. The clock QL indicates the holdover performance of a particular clock type in the synchronization network. The latest revision of G.781 [42] includes the definition of clock QL values used for EEC-option 1 and EEC-option2 as well as the synchronization functions needed for Sync-E equipment. G.781 defines the QL values for EEC-option 1 and EEC-option 2 to be the same as G.813 Option 1 and G.812 Type IV (stratum 3), respectively. G.781 also describes the process of selecting a synchronization source. For Sync-E, G.8264 defines two types of ESMC message: heartbeat and event. Heart-beat messages (transmitted once per second) are used to continuously signal the clock quality level. Event messages (generated and transmitted immediately upon detection of a clock quality level change) are used to signal a change in the SSM clock quality level. If an ESMC message
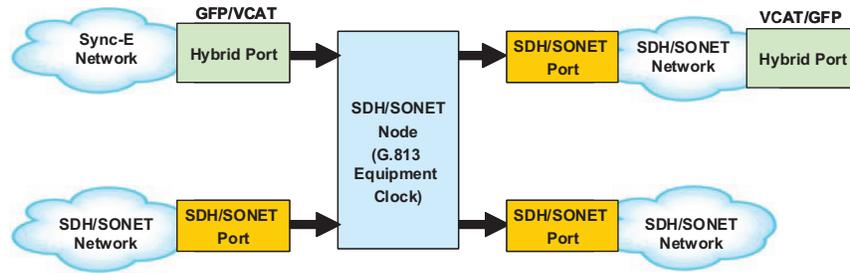
Fig. 4. Interworking of SONET/SDH and Sync-E.

is not received after five seconds the QL value is to be considered as DNU (do not use).

The discussion above highlights why Sync-E has become a key enabler for interworking synchronization capabilities with SDH/SONET networks. This is because the EEC specification is identical with the SEC specification, and because ESMC conveys the same information as the SSM. The interworking allows service providers to mix Sync-E with SDH/SONET network elements to form one unified synchronization network.

IV. ATTRIBUTES OF 100 MB/S, 1000 MB/S, AND 10 GIGABIT ETHERNET THAT FAVOR THEIR USE IN SYNC-E

We discuss in this section the main features of the 100 Mb/s, 1000 Mb/s, and 10 Gb/s Ethernet technologies that make them suitable for use in Sync-E. As we will see below, 10 Mb/s Ethernet PHYs cannot be used in Sync-E, and 1000BASE-T and 10GBASE-T PHYs will require special architectural handling when used in Sync-E. Our discussion focuses, in particular, on the relevant Physical Layer (PHY) features of the most commonly used Ethernet versions that are important to timing distribution.

A. 100BASE-TX/FX

Fig. 5 shows the 100 Mb/s Ethernet reference model. The term PHY (Fig. 5) is a generic industry term used to describe a transceiver for 100 Mb/s, 1 Gb/s, and 10 Gb/s operation. A PHY would connect between a MII/GMII/10GMII and an MDI and consists of PCS, PMA, and PMD functionality. In many other cases, PHY is used as an abbreviation for the Physical Layer of the OSI model. In Fig. 5 (and all related figures in the paper), the MDI defines the connector between the PMD sublayer and the media. The Auto-Negotiation function provides a mechanism for exchanging configuration information between devices at the two ends of a link segment and automatically selecting the highest performance mode of operation supported by both devices. The PMD sublayer takes the serial bit stream provided by the PMA sublayer and converts it to/from optical (or electrical) signals across a medium (fiber or copper). The PMA sublayer provides the PCS with a media-independent interface that a variety of serial physical media can connect to. The PCS provides the functions of data symbol encoding and decoding, carrier sense, collision detect, and synchronization services to the MAC, which are usually independent of the physical medium used.
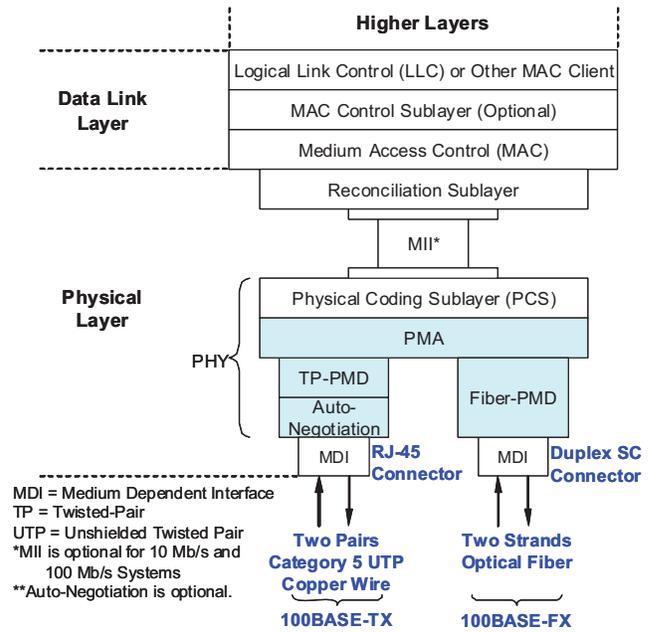

Fig. 5. 100BASE-X Ethernet reference model [1].

1) 100BASE-TX/FX PHY Overview: 100BASE-TX operates over two pairs of Category 5 (Cat 5) cable (one pair for each direction) while 100BASE-FX operates over two individual (multimode) fiber optic cables. 100BASE-TX has a maximum distance of 100 meters (328 ft). 100BASE-FX uses a 1300 nm near-infrared (NIR) light wavelength transmitted via two strands of optical fiber, one for receive and the other for transmit. Its maximum distance is 400 meters (1,310 ft) for half-duplex connections (to ensure collisions are detected), and 2 kilometers (6,600 ft) for full-duplex over multi-mode optical fiber. 100BASE-TX and 100BASE-FX are based on a common set of PCS (Physical Coding Sublayer) and PMA (Physical Medium Attachment) specifications and are collectively designated as 100BASE-X.

The 100BASE-X MAC (as in all IEEE 802.3 MACs) handles the high level portions of the Ethernet protocol (framing, error detection, when to transmit, etc) and the PHY handles the low level logic (4B/5B encoding/decoding, SERDES (serialization/deserialization), and NRZI encoding/decoding) and analog front-ends. 100BASE-X uses 4B/5B block encoding
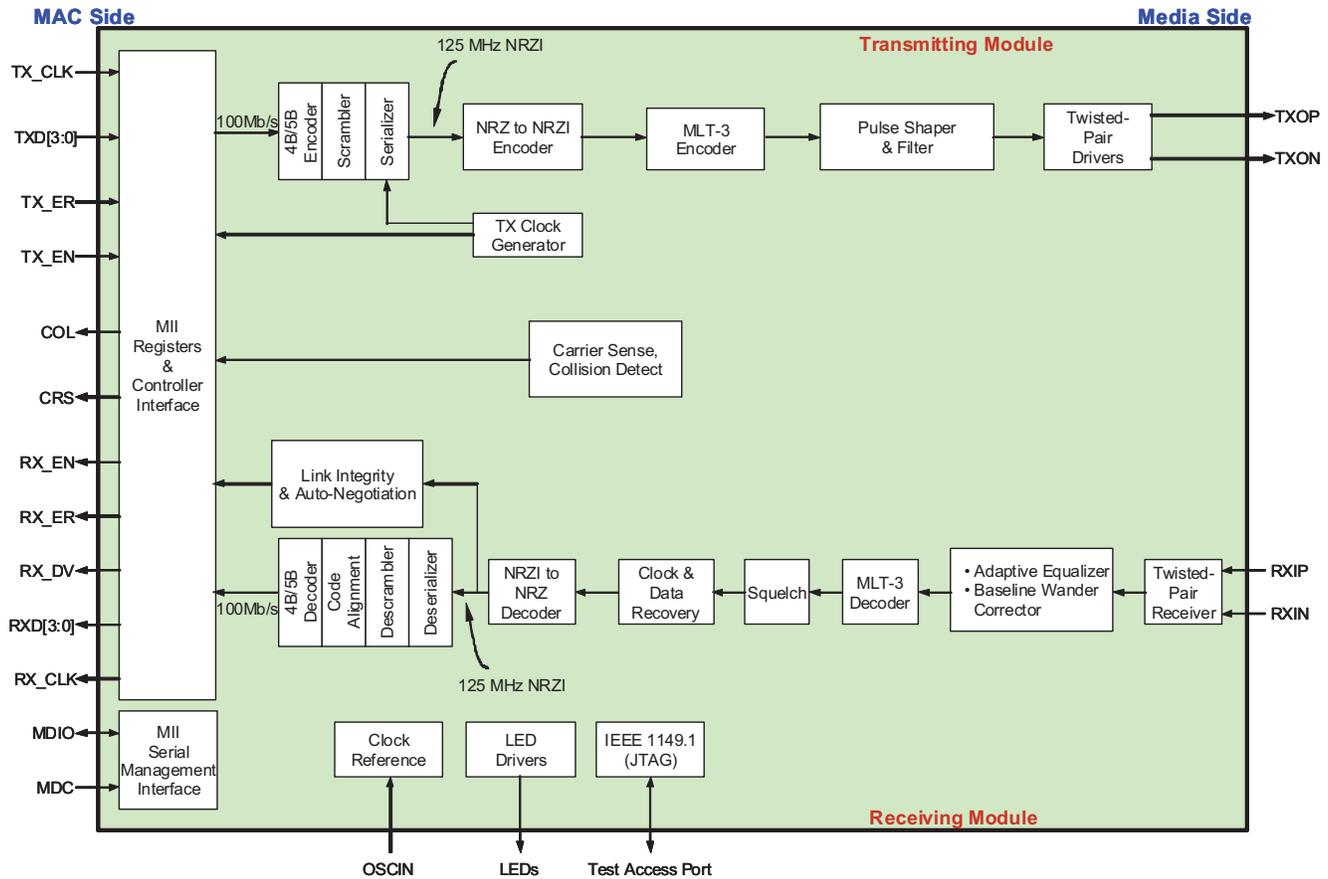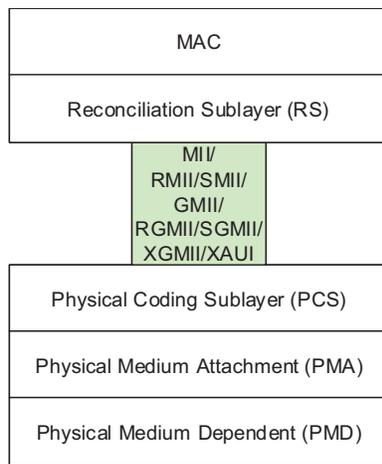
Fig. 6. 100BASE-X transceiver block diagram.

where the PHY takes a nibble (4-bits) from the MII (Medium Independent Interface) and converts this to a 5-bit binary symbol [1]. For 100BASE-FX, this 5-bit symbol is transmitted serially on the medium. Since all data is transmitted over a single fiber, the full data rate of 100 Mb/s is present, with a frequency of 125 MHz due to the 4B/5B code conversion. However, for 100BASE-TX, RFI/EMI effects from the higher 100 Mb/s data rates on the UTP (unshielded twisted pair) are unacceptably high, so additional processing steps are required to reduce the spectral content of the transmission.

In 100BASE-TX, first, the transmitted 5-bit code is scrambled. The transmitter and receiver are kept in lock-step to ensure that the data can be descrambled at the receiver. The scrambling process helps smooth the spectral content of the resulting transmitted waveform. Subsequently, an MLT-3 code (multilevel transmit) is applied to the transmitted serial bit stream. This converts the binary 5-bit symbol to a ternary code and further reduces the spectral content. The data rate on the single pair is 100 Mb/s but the scrambling and MLT-3 coding reduce the frequency on the line to only 31.25 MHz. Fig. 6 shows the block diagram of a typical 100BASE-TX transceiver. The additional scrambling and MLT-3 encoding stages required to minimize EMI/RFI over the UTP cable are not necessary in the case of fiber optic cables (i.e., 100BASE-FX).

The 4B/5B encoding maps the 16 possible 4B data nibble values to a subset of the 5B binary code groups available within the total of 32. During the encoding, normal MII data nibble values are mapped to data codes [1]. An "IDLE" (/I/) code group is used to continuously signal between packets during IPG (inter-packet gap). The remaining 15 code groups are used for special control signaling functions (the /J/, /K/, /T/, /R/, and /H/ code groups), or are reserved (considered invalid). After data packet transmission by an Ethernet node, the signal on the transmit cable pair transitions to the IDLE fill code – the transmit pair *does not become physically idle. The IDLE fill codes sustain continuous clock recovery at the receiver and also serve as keepalive messages to the receiver*. The feature of IDLE fill codes is present in all higher speed flavors of Ethernet but not in 10 Mb/s Ethernet, particularly, 10BASE-T Ethernet which uses special pulses (link test pulses) as keepalive signals (see 10BASE-T discussion below).

2) Physical Layer Timing and 100 Mb/s Ethernet Media Independent Interface Types: Ethernet defines a clear interface between the MAC block (common to all devices) and a PHY, which is media dependent. The MII (as well as its higher speed counterparts) is an interface that decouples the MAC layer from the different requirements of the various PHY layers (see Fig. 7). The MII is capable of operating at both 10 Mb/s and 100 Mb/s. The MII is an 18-pin signal interface consisting of 7 transmit interface signals, 7 receive interface signals, 2 network status signals, and 2 management interface signals (Fig. 8). In many cases the MII is simply referred to as a 16-pin signal interface by excluding the 2 management pins.

MAC

Reconciliation Sublayer (RS)

MII/
RMII/SMII/
GMII/
RGMII/SGMII/
XGMII/XAUI

Physical Coding Sublayer (PCS)

Physical Medium Attachment (PMA)

Physical Medium Dependent (PMD)

MII = Media Independent Interface
RMII = Reduced Media Independent Interface
SMII = Serial Media Independent Interface
GMII = Gigabit Media Independent Interface
RGMII = Reduced Gigabit Media Independent Interface
SGMII = Reduced Gigabit Media Independent Interface
XGMII = 10 Gigabit Media Independent Interface
XAUI = 10 Gigabit Attachment Unit Interface

Fig. 7. xMII as PCS-to-MAC interface.

The nibble-wide data transmit and receive paths, as well as their associated control signals, operate synchronously to their respectively clocks. The MII was developed to provide 4-bit (nibble) interface in both transmit and receive directions, lowering the required speed of operation (from a 100 MHz requirement if a bit-serial implementation were retained) to 25 MHz. The data on each pin in both transmit and receive directions is synchronous with the clock, and the clock rate is one-fourth of the data rate, i.e., 25 MHz for 100 Mb/s and 2.5 MHz for 10 Mb/s. From a clocking perspective, the following clock signals (pins) are provided (Fig. 8):

- **Transmit Clock (TX_CLK)**: TX_CLK is a continuous clock which provides a timing reference for the TX_EN, TX_ER, and TXD[3:0] signals from the MAC. The clock frequency is 25 MHz in 100 Mb/s operation and 2.5 MHz in 10 Mb/s operation.
- **Receive Clock (RX_CLK)**: RX_CLK is a continuous clock which provides a timing reference to the MAC for the RX_DV, RX_ER, and RXD[3:0] signals. The clock frequency is 25 MHz in 100 Mb/s operation and 2.5 MHz in 10 Mb/s operation. The clock is recovered by the PHY from the received line data stream.

The MII can be used as an interconnect at the chip, board, or physical device level. In considering an interface for a Sync-E application, the most important issue is how a reference signal is supplied to and recovered from the interface. It is important to note that Sync-E timing transfer is done link-by-link and thus will require clock signals to be supplied to an upstream interface, recovered from a downstream interface, clean-up and re-injected into adjoining downstream links.

In a Sync-E implementation, the TX_CLK in a transmitting device is traceable to a reference clock (PRC) while the RX_CLK in a receiving device constitutes the recovered

clock of the upstream TX_CLK. The RX_CLK can then be cleaned up and in turn used to source the TX_CLK of the receiving device. In non-Sync-E implementation, TX_CLK in the receiving device is driven by a free-running clock with no more than ±100 ppm accuracy.

One of the biggest issues with the MII, from an implementation point of view, is the number of signal pins required between the MAC and the PHY. The Reduced MII (RMII) [25] resulted from the realization that ever-reducing silicon geometries and ever-increasing port densities for switches and repeaters were bound to make pin-count reduction for MII function a pressing issue. Fig. 9 shows the RMII architecture and signals. The RMII provides the same characteristics as the existing IEEE 802.3 MII while providing an interface that is independent of PHY port densities and saves a total of 9 data and control pins per port, that is, 7 pins out of 16 pins (ignoring the management pins) are required (based on the architecture in Fig. 9). This means a 24-port switch based on RMII would require a total of 168 pins, not including the management pins. Depending on RMII implementation, generally the number of signals/pins required for connecting to the PHY can be reduced from 16 (for an MII-compliant interface not counting the management pins) to between 6 and 10.

The RMII supports both 10 Mb/s and 100 Mb/s data rates, and full-duplex operation. Unlike the original MII, the RMII was designed to be only as *chip-to-chip interconnect*. Unlike the MII, the RMII operates synchronously. This allows the transmit paths and the receive paths to operate based upon the same clock as the switch or repeater, saving both clock pins. However, in order to provide a substantial pin reduction, the data and control paths were reduced by specifying that the RMII operate at 50 MHz, twice the data rate as the MII. This allowed the transmit and receive data paths to be reduced from nibble-wide to two bits wide in each direction, saving four pins.

The Reference Clock (REF_CLK) is a 50 MHz signal and is the synchronous clock reference for receive, transmit, and control interfaces. REF_CLK provides the timing for CRS_DV, RXD[1:0], TX_EN, TXD[1:0], and RX_ER. REF_CLK is provided by the MAC or an external source (see Figs. 9 and 10). The 2 bits RMII Receive Data signals RXD[1:0] are driven synchronously to the 50 MHz reference clock REF_CLK (Fig. 9). When the PHY device uses an external (e.g., 25 MHz) crystal oscillator connected to one of its pins, it internally generates the 50 MHz RMII reference clock for use by the RMII logic (Figs. 9 and 10). The 50 MHz clock is output on RX_CLK and TX_CLK for use as the reference clock for an attached MAC.

Note that in a non-Sync-E implementation, REF-CLK is the synchronous clock reference for receive, transmit, and control interfaces. In a Sync-E implementation using RMII as a chip-to-chip interconnect, the TX_CLK in a transmitting PHY device is driven by the REF_CLK which in turn can be traceable to a reference clock (PRC). The RX_CLK (MII signal) in a receiving PHY device constitutes the recovered clock of the upstream TX_CLK (MII signal). Using the RX_CLK signal as a reference in a receiving RMII PHY for downstream links depends on how accessible the RX_CLK
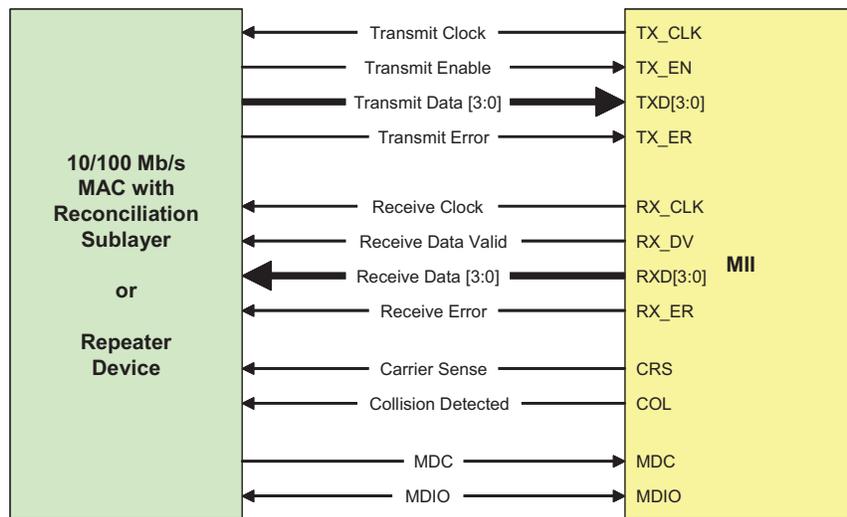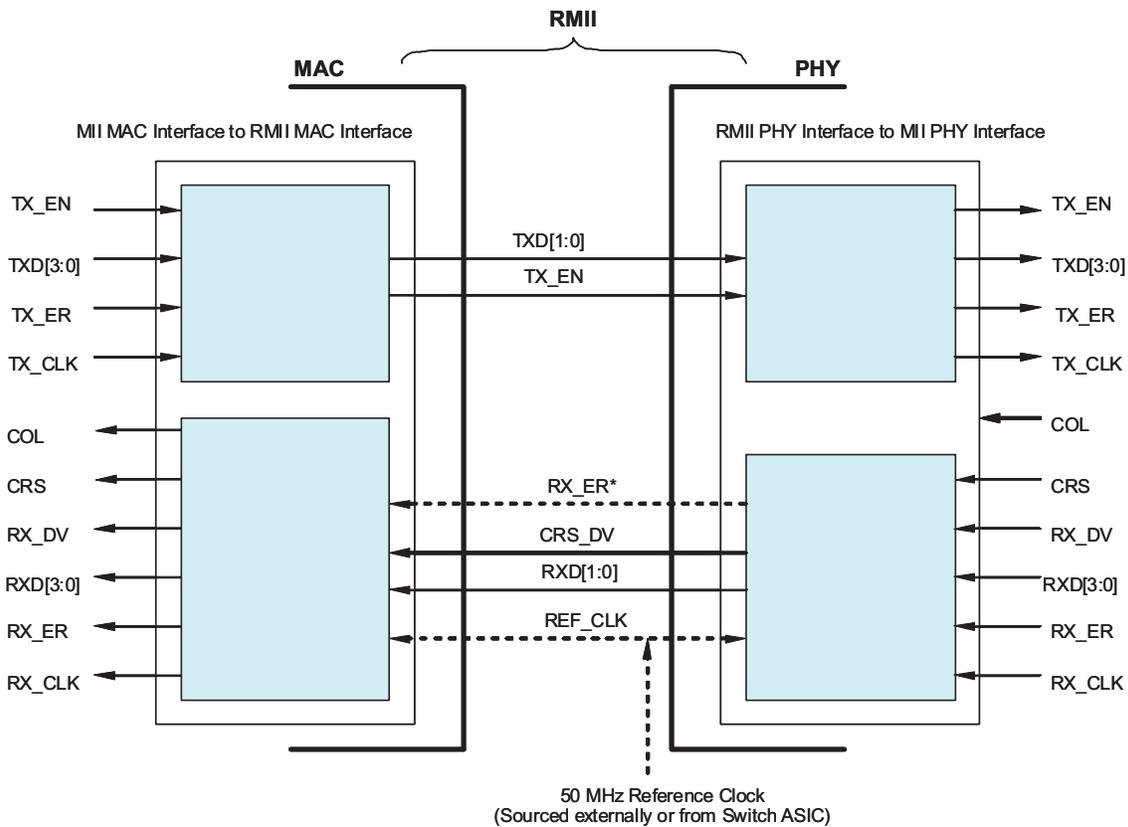
Fig. 8.   MII pin designations.

Fig. 9.   Reduced MII signals and architecture.

signal is in the chip design. If RX_CLK is accessible on the chip, it can be cleaned up and scaled up to 50 MHz for use by the RMII logic and also used to source the TX_CLK of the receiving device. TX_CLK in a non-Sync-E receiving device is driven by a separate REF_CLK clock reference.

The Serial-MII (SMII) [26] is another standard that addresses the connection of 10/100 Mb/s Ethernet PHYs to Ethernet switches or the MAC portion of an end-device's Eth-

ernet interface. The SMII is designed to support the following: carry complete MII information between a 10/100 PHY and MAC with two pins per port, allow multi port MAC/PHY communications with one system clock, operate in both half and full duplex, support per packet switching between 10 Mb/s and 100 Mb/s data rates, and allow direct MAC to MAC communication. SMII provides a Serial MII Interface for a 10/100 Mb/s Ethernet MAC to be used with PHYs that offer
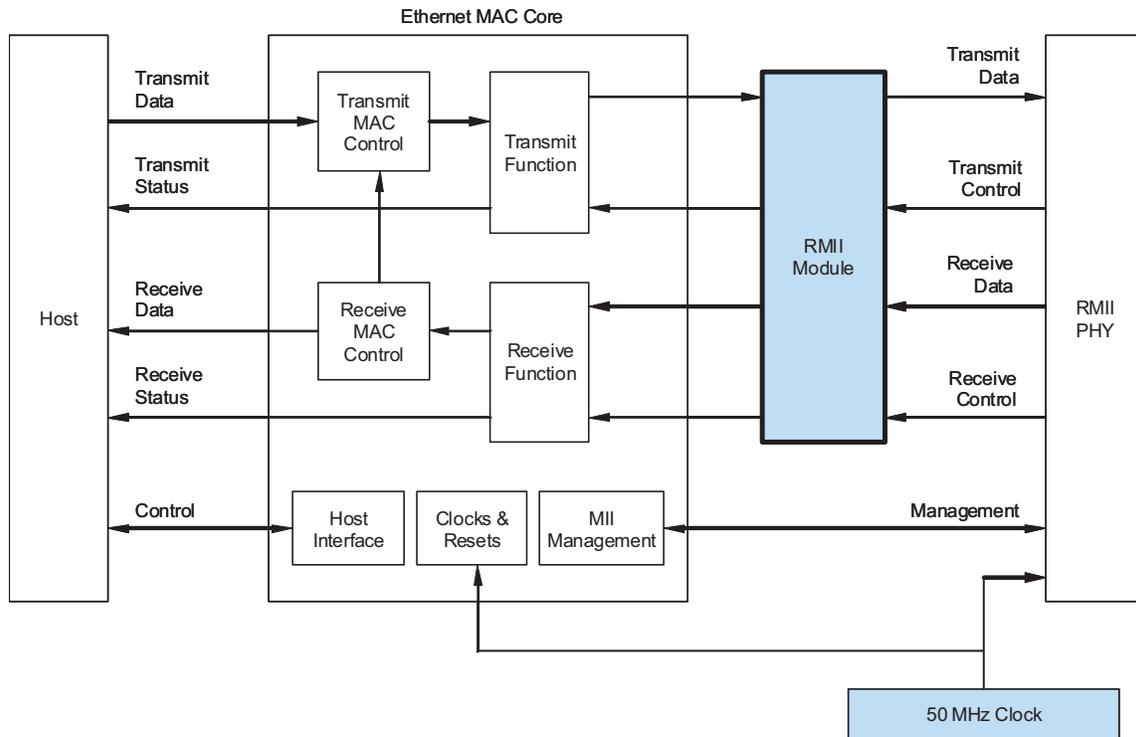
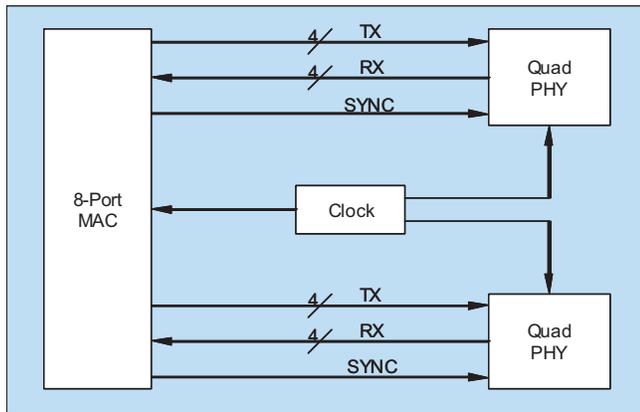Fig. 10. Ethernet MAC core with RMII.

Fig. 11. Typical SMII system.

two pins per port substantially reducing the pin count between the MAC and the PHY. The SMII can be used in multi-port applications allowing multiple MACs to be connected to quad or octal PHYs with minimal routing. The reduction in pin count is achieved by multiplexing data and control information to a single transmit signal and a single receive signal. Figs. 11, 12, and 13 show the SMII signals and typical applications.

SMII is composed of two signals per port (Fig. 11) – a global synchronization signal, and a global 125 MHz reference clock. All signals are synchronous to the clock (Fig. 11). The TX signal (MAC to PHY) is used for transmit data and control. RX signal (PHY to MAC) is used for receive data and control. The SYNC signal (MAC to PHY) is used for synchronization. The system clock (which drives both MAC and PHY) is also used for synchronization.
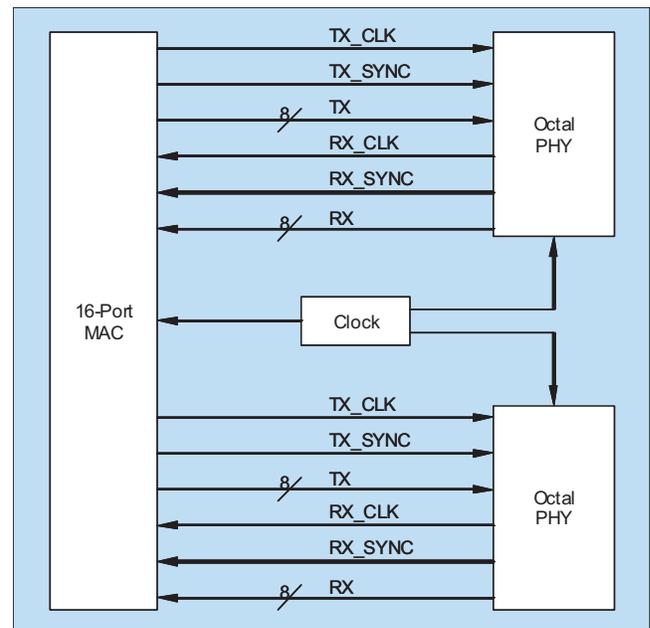
Fig. 12. Source synchronous SMII system.

Fig. 12 shows a source synchronous SMII system that uses four optional signals specified in the SMII specification [26] that may be added in place of SYNC. These signals are for applications requiring longer trace delay's (more than 1ns). RX_CLK and RX_SYNC synchronize RX for one or more ports. Similarly TX_CLK and TX_SYNC synchronize TX for one or more ports. RX_CLK and TX_CLK are frequency locked, but not necessarily phase locked to CLOCK.
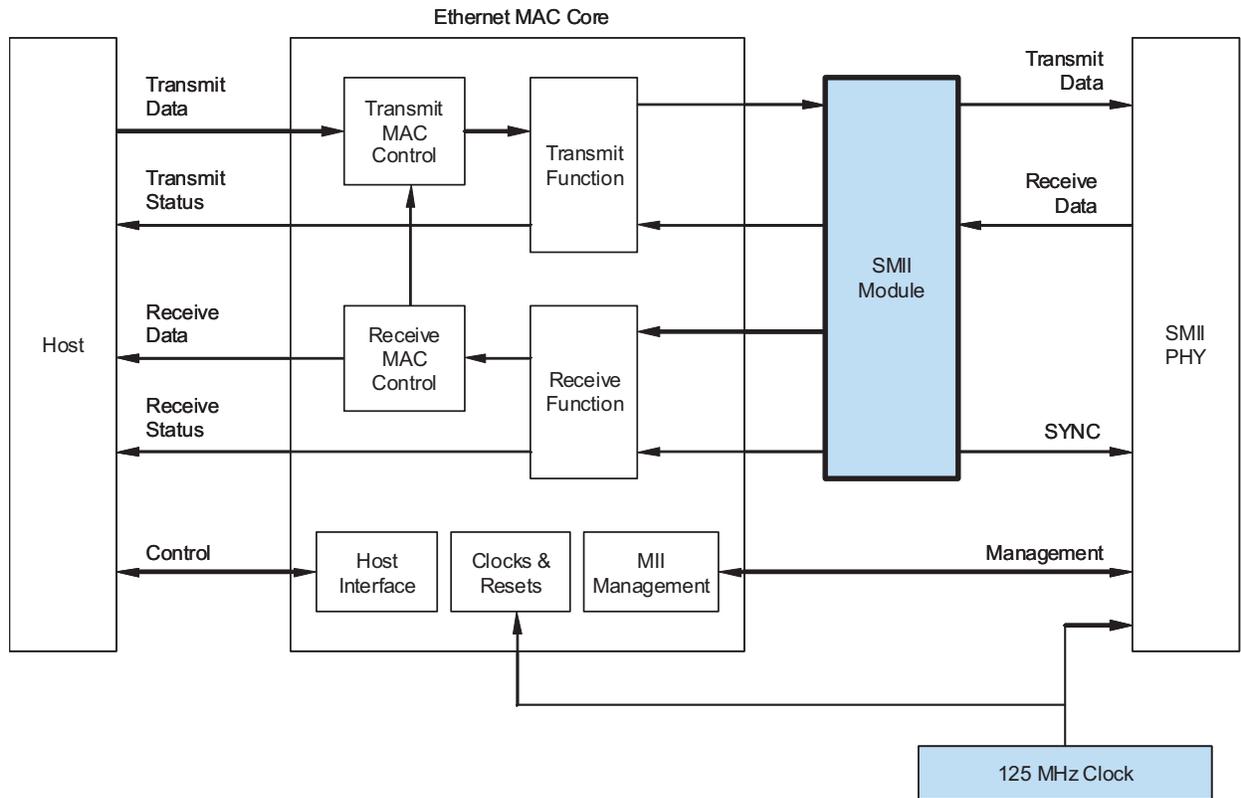
Fig. 13. Ethernet MAC core with SMII.

In the sample application in Fig. 13, the interface between the 16-signal MII interface and the two-signal Serial-MII interface is provided by the SMII module. The required 2-signal plus clock and synchronization interface to the PHY is provided by the SMII module. This module performs the conversion of the 4-bit MII data handled by the basic MAC design to/from the serial data format of the Serial-MII interface in both the transmit and the receive directions. The registers within the PHY are read and written through the MAC's Host Interface.

In the transmit direction, the MAC core generates MII nibble data streams in response to frame byte streams supplied by the host system. The SMII takes the nibble data, which is clocked at 25 MHz (2.5 MHz in 10 Mb/s mode), and outputs serial data at the SMII Reference Clock frequency of 125 MHz. A synchronizing pulse is asserted every 10th clock to signal the start of a new 10-bit segment comprising a Transmit Enable bit, a Transmit Error bit followed by a byte of data. In 100 Mb/s mode, each segment conveys one byte of data. In 10 Mb/s mode, each segment is repeated ten times, thus modulating the bandwidth down to 10 Mb/s.

In the receive direction, the SMII PHY provides serial data synchronous to the reference clock. Again 10-bit segments are transferred with the start marked by a SYNC pulse. The first two bits represent Carrier Sense and Receive Data Valid. The remaining eight bits represent one byte of data. In 100 Mb/s mode, a new nibble is output each RX_CLK. In 10 Mb//s mode, the SMII PHY replicates each received byte ten times and the SMII only needs to pass a nibble to the MAC every tenth RX_CLK.

As described above, a synchronization signal going from the MAC to the PHY and a 125 MHz clock reference are required in SMII, but these can be shared by multiple ports. In a Sync-E implementation using SMII as a MAC-to-PHY interconnect, the 125 MHz clock (driving both MAC and PHY) in a transmitting device can be made traceable to a reference clock (PRC) and the downstream signals in connected modules can be made traceable to this reference clock. The RX_CLK (MII signal) in a receiving PHY device constitutes the recovered clock of the upstream TX_CLK (MII signal). Using the RX_CLK signal as a reference in a receiving SMII PHY for downstream links depends on how accessible the RX_CLK signal in the chip design. If RX_CLK is accessible on the chip, it can be cleaned up and scaled up to 125 MHz for use by the SMII logic and also used to source the TX_CLK of the receiving device. TX_CLK in a non-Sync-E receiving device is driven by a separate 125 MHz clock reference.

B. 1000BASE-CX/SX/LX

The Gigabit Ethernet technologies 1000BASE-CX, 1000BASE-SX, and 1000BASE-LX are collectively designated as 1000BASE-X. The reference model of 1000BASE-X Ethernet is shown in Fig. 14 [1]. The PCS provides the functions of data coding and decoding, which are usually independent of the physical medium used. In Gigabit Ethernet, the encoding scheme used for fiber (1000BASE-SX/LX) and short shielded balanced copper cable (1000BASE-CX), that is, 8B/10B encoding, is different from that of UTP copper media (that is, 1000BASE-T, discussed below).
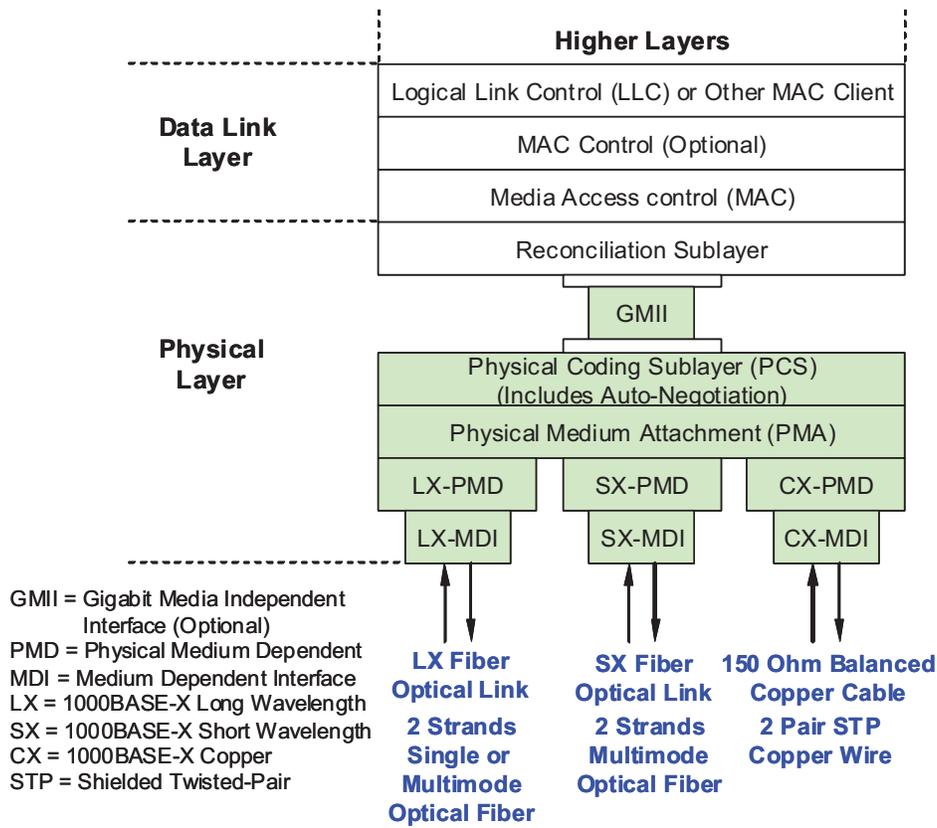
Ethernet Reference Model

Higher Layers

Data Link Layer

Logical Link Control (LLC) or Other MAC Client

MAC Control (Optional)

Media Access control (MAC)

Reconciliation Sublayer

GMII

Physical Layer

Physical Coding Sublayer (PCS)
(Includes Auto-Negotiation)

Physical Medium Attachment (PMA)

| LX-PMD | SX-PMD | CX-PMD |
| LX-MDI | SX-MDI | CX-MDI |

GMII = Gigabit Media Independent
    Interface (Optional)
PMD = Physical Medium Dependent
MDI = Medium Dependent Interface
LX = 1000BASE-X Long Wavelength
SX = 1000BASE-X Short Wavelength
CX = 1000BASE-X Copper
STP = Shielded Twisted-Pair

LX Fiber Optical Link
2 Strands Single or Multimode Optical Fiber

SX Fiber Optical Link
2 Strands Multimode Optical Fiber

150 Ohm Balanced Copper Cable
2 Pair STP Copper Wire

Fig. 14. 1000BASE-CX/SX/LX (1000BASE-X) Ethernet reference model [1].

1) 1000BASE-CX/SX/LX PHY Overview: The 8B/10B coding scheme maps groups of 8 binary bits to a code-group of 10 binary bits. The code has good transition density, is run-length limited, and is DC balanced. This facilitates clock recovery and error detection of most single- and multiple-bit error conditions, making it an excellent choice for optical and copper media types. Packet data presented by the GMII is encoded by the 8B/10B encoder in the PCS and passed to the PMA (with the exception that the first byte of the Preamble is replaced with the Start-of-Packet Delimiter (SPD) code-group).

When no packet data is present on the GMII, such as between frames, an IDLE (/I/) code-group indication (called ordered-set) is issued [1]. Transmission of IDLE ordered-sets also keeps the receive electronics and optics "alive" between packets. The idle ordered-sets /I1/ and /I2/ are continuously transmitted to maintain clock synchronization and are required to delimit packet data. The clock recovery unit in the receiver needs periodic transitions to maintain synchronization of the receive clock. An ordered-set is either a single special code-group or a sequence of code-groups consisting of an initial special code-group followed by additional special or data code-groups [1]. Ordered-sets are always 1, 2, or 4 code-groups in length.

The PMA performs symbol serialization and deserialization (SERDES). The encoded stream of 10-bit symbols (from the

PCS) is serialized before transmission, and the received stream (from the PMD) is deserialized and passed as 10-bit symbols to the PCS layer. In addition, the receive section of the PMA sublayer extracts the symbol-timing clock from the received bit stream for correct symbol alignment (framing) of the received data. The PMA can be implemented as a Ten Bit Interface (TBI) described in the IEEE standard [1].

The PMA transmit clock (PMA_TX_CLK) operates at 125 MHz, similar to the GMII transmit (GTX_CLK) and receive (RX_CLK) clocks [1]. PMA_TX_CLK is used to latch 10-bit symbols in the PMA transmission. A clock multiplier unit in the PMA uses the PMA_TX_CLK to generate the 1250 MHz clock required for transmission of serialized 10-bit symbols on the media. The PMA accepts 1250 Mb/s serial data from the PMD. The received data is deserialized and delivered by the PMA to the PCS in TBI format.

The PMA has two receive clocks PMA_RX_CLK<0> and PMA_RX_CLK<1>. PMA_RX_CLK<0> is a 62.5 MHz recovered receive clock that the PMA uses to latch odd-numbered symbols reconstructed from the incoming serial bit stream. PMA_RX_CLK<1> is the 62.5 MHz recovered receive clock that the PMA uses to latch even-numbered symbols reconstructed from the incoming serial bit stream. PMA_RX_CLK<1> is 180 degrees out of phase with PMA_RX_CLK<0>. The combined effect of the PMA_RX_CLK<0> and PMA_RX_CLK<1> is a 125 MHz
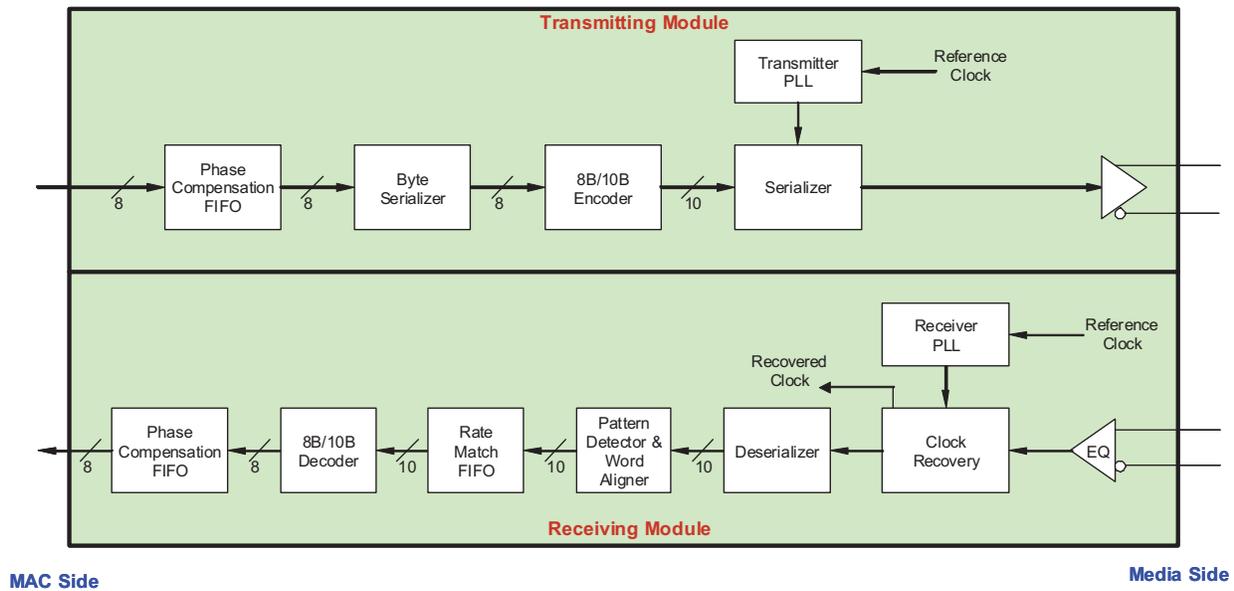
Fig. 15.   1000BASE-X transceiver block diagram.

PMA receive clock of 125 MHz that is used to latch out the 10-bit symbols required by the PCS.

The PMD layer performs the function of converting signals from the PMA into signals appropriate for the specific media. If the medium is fiber, the electrical signals are converted to optical signals and vice versa. The MDI defines the connector between the PMD layer and the media. The block diagram of a typical 1000BASE-X transceiver is given in Fig. 15.

The 1000BASE-X PMA supports three different media types with each of the media types requiring an appropriate corresponding PMD:

- 1000BASE-CX specifies operations over two pairs of 150 ohm balanced copper cable. The "C" in 1000BASE-CX stands for "copper" as it uses specially balanced shielded twisted pair copper jumper cables also called "twinax" or "short haul copper." Segment lengths are limited to only 25 meters which restricts 1000BASE-CX to connecting equipment in small areas like wiring closets. The application for this type of cabling will be short-haul data-center interconnections and inter-rack connections.
- 1000BASE-SX specifies operation over a pair of optical fibers using short-wave optics. The "S" in 1000BASE-SX stands for "short." Only multi-mode optical fiber is supported. 1000BASE-SX uses two strands of multi-mode fiber-optic cable per link – short wavelength optics (850nm) over: 62.5 μm multi-mode fiber; 50 μm multi-mode fiber. The short wavelength optics specified by the standard operate in the wavelength range of 770 to 860 nanometers. The standard specifies a distance capability between 220 meters (62.5/125 μm fiber with low modal bandwidth) and 550 meters (50/100 μm fiber with high modal bandwidth). Short wavelength optics have the advantage of being less expensive than long wavelength optics. To guarantee proper signal recovery, a 1000BASE-SX link cannot exceed 550 meters in length.

- 1000BASE-LX specifies operation over a pair of optical fibers using long-wavelength optics. The "L" in 1000BASE-LX stands for "long." Both single mode and multi-mode optical fibers are supported. 1000BASE-LX uses two strands of multimode or single mode fiber-optic cable per link – long wavelength lasers (1300nm) over: 62.5 μm multi-mode fiber; 50 μm multi-mode fiber; 10 μm single mode fiber. The long wavelength optics specified by the standard operate in the wavelength range of 1270 to 1355 nanometers. Long wavelength optics are more expensive than short wavelength, but have the advantage of being able to drive longer distances. To guarantee proper signal recovery, a 1000BASE-LX link cannot exceed 550 meters in length over multimode fiber or 10 km in length over 10 μm single mode fiber.

2) Physical Layer Timing and 1000 Mb/s Ethernet Media Independent Interface Types: The GMII (Fig. 16) provides an interface between a MAC and any transceiver (1000BASE-X and 1000BASE-T). The GMII is not an exposed interface – it can be used to connect on a chip-to-chip (IC to IC) level, which is typically implemented as printed circuit board traces, or motherboard-to-motherboard connection between two or more printed circuit boards. Excluding the management pins, GMII has 24 pins (Fig. 16). The data path in GMII is specified to be byte-wide instead of nibble-wide (4-bit wide). During each cycle one byte of data is moved across the GMII, requiring the transmit and receive clock to operate at 125 MHz to provide a total data rate of 1000 Mb/s. The GMII can also operate at 10 Mb/s and 100 Mb/s data rates with the clocks operating at 2.5 or 25 MHz, respectively. The operation of the GMII at 10 Mb/s and 100 Mb/s is identical to MII.

GMII has the following clocking signals (pins):

- **Gigabit Transmit Clock (GTX_CLK)**: There are two clocks (GTX_CLK and TX_CLK), depending on whether the PHY is operating at 1000 Mb/s or 10/100 Mb/s speeds. GTX_CLK is a 125 MHz (nominal) clock for
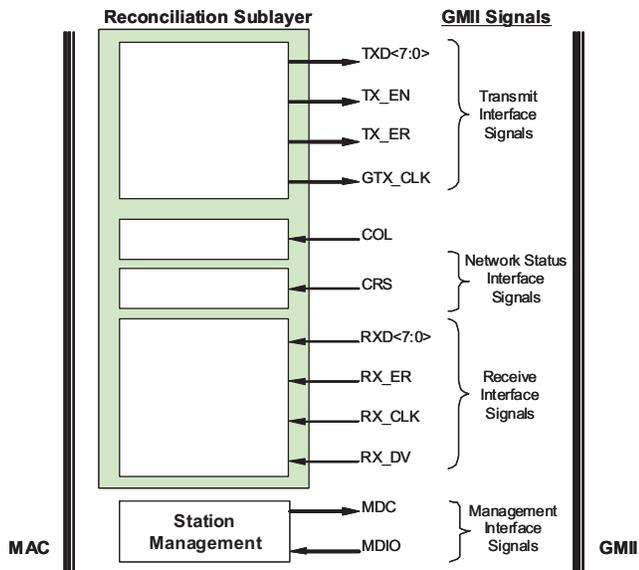
Fig. 16. GMII signals.

Fig. 17. Reduced Gigabit media independent interface (RGMII).

transmission of 1000 Mb/s Ethernet data. GTX_CLK is a signal *supplied* to the PHY and is used as a timing reference to drive TXD, TX_EN, and TX_ER signals synchronously with clock, in order to transmit data and/or issue status. TX_CLK (*supplied by* the PHY) is used when the GMII operates at 10 Mb/s or 100 Mb/s data rate. TX_CLK operates at either 25 MHz for 100 Mb/s or 2.5 MHz for 10 Mb/s connections. TX_CLK is not used in 1000 Mb/s operation.

- **Receive Clock (RX_CLK)**: This is the recovered clock (125 MHz, nominal) from the received data stream.

In a Sync-E implementation, the GTX_CLK in a transmitting device is traceable to a reference clock (PRC) while the RX_CLK in a receiving device constitutes the recovered clock of the upstream TX_CLK. The RX_CLK can then be cleaned up and in turn used to source the GTX_CLK of the receiving device. In non-Sync-E implementation (i.e., IEEE 802.3), GTX_CLK in the receiving device is driven by a free-running 125 MHz clock with no more than ±100 ppm accuracy.

The Reduced GMII (RGMII) [27], [28] specifies a reduced interface between an Ethernet MAC and PHY and can be used as an alternative to the MII, the GMII and the Ten Bit Interface (TBI). This reduction is achieved by clocking data on both the rising and falling edges of the clock in 1000 Mbit/s operation, and by eliminating non-essential signals like the carrier-sense and collision-indication. As shown in Figs. 17 and 18, RGMII consists only of 12 pins (RX_CTL, RXC, RXD[3:0], TX_CTL, TXC, TXD[3:0]) excluding the management pins as opposed to 24 for GMII and 28 for TBI. RGMII has the following clocking signals (pins):

- **TXC (generated by the MAC in both RGMII and RTBI mode)**: TXC is the transmit reference clock. TXC operates at 125 MHz for 1000 Mbit/s operation, and for 10/100 operation, the clocks operate at 2.5 MHz or 25 MHz, respectively. TXC has ±50 ppm accuracy.
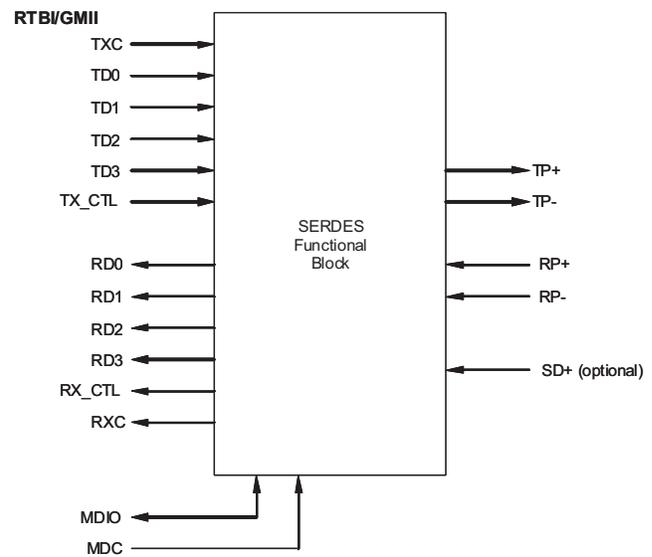- **RXC (generated by the PHY in both RGMII and**

RTBI mode): This is the continuous receive reference clock derived from the received data stream. RXC operate at 125 MHz for 1000 Mbit/s operation, and for 10/100 operation, the clocks operate at 2.5 MHz or 25 MHz, respectively. RXC has ±50 ppm accuracy.

Note that in GMII 10/100 Mbit/s operation, the TX_CLK is provided by the PHY (and supplied to the MAC) and in GMII 1000 Mbit/s operation GTX_CLK is provided by the MAC (and supplied to the PHY).

In a Sync-E implementation using RGMII as a MAC-to-PHY interconnect, TXC generated in a transmitting MAC device can be traceable to a reference clock (PRC). The RXC in a receiving PHY device constitutes the recovered clock of the upstream TXC. Using the RXC signal as a reference in a receiving RGMII PHY for downstream links depends on how accessible the RXC signal in the chip design. If RXC is accessible on the chip, it can be cleaned up and scaled up to 125 MHz for use by the RGMII logic and also used as the reference clock of the receiving device.

The Serial GMII (SGMII) [29] is another Gigabit Ethernet interface used to connect an Ethernet MAC-block to a PHY. To carry frame data and link rate information between an Ethernet MAC and a 10/100/1000 PHY, SGMII uses a differential pair for data signals and for clocking signals, with both being present in each direction (i.e., transmit and receive), resulting in 8 signal lines in total. As shown in Fig. 19, SGMII uses differential pairs at 625 MHz clock frequency dual data rate (DDR) for TX and RX data and the TXCLK and RXCLK clocks. The data signals operate at 1.25 Gbaud and the clocks operate at 625 MHz (a DDR interface). Compared to GMII, SGMII has low-power and low pin count with serial 8B/10B encoded interface (commonly referred to as a SerDes).

The transmit and receive data paths leverage the 1000BASE-SX PCS defined in the IEEE 802.3 specification. The traditional GMII data signals (TXD/RXD), data valid signals (TX_EN/RX_DV), and error signals (TX_ER/RX_ER) are encoded, serialized and output with the appropriate DDR
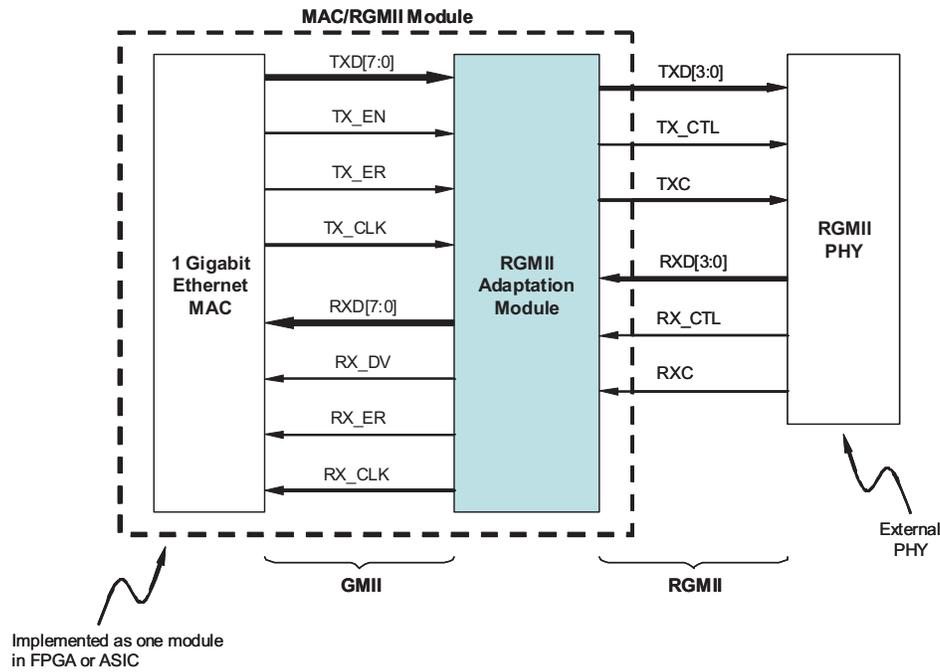
Fig. 18. An application of the RGMII – RGMII to GMII bridge.
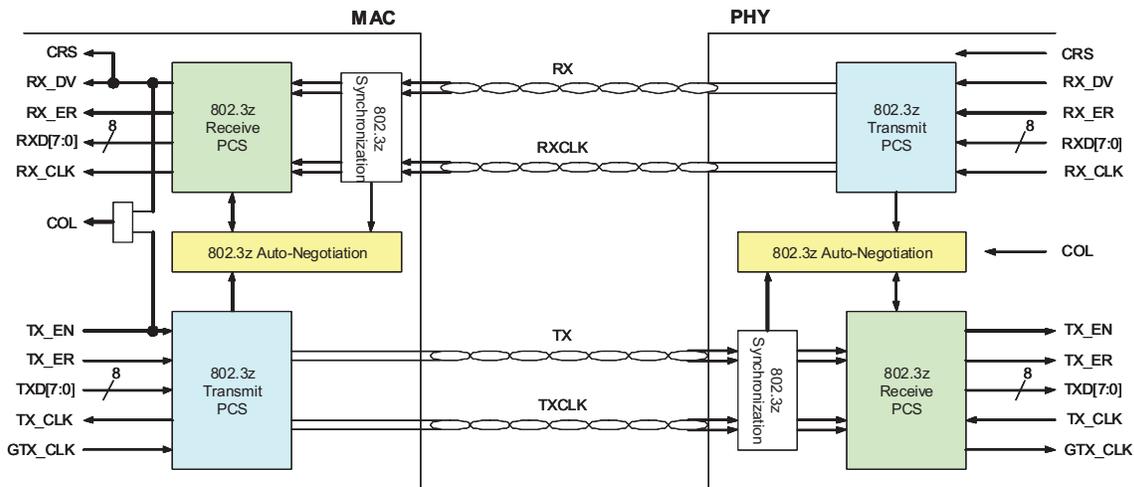

Fig. 19. SGMII connectivity.

clocking – thus giving SGMII a 1.25 Gbaud interface with a 625 MHz clock. SGMII specifications details source synchronous clocking; however, specific implementations may recover clock from the data rather than use the supplied clock. This operation is allowed; however, all sources of data must generate the appropriate clock regardless of how they clock receive data.

At the receive side (Fig. 20), GMII signals come in at 10/100/1000 Mbps clocked at 2.5/25/125 MHz. The PHY passes these signals through the PHY Receive Rate Adaptation to output the 8-bit data RXD[7:0] in 125MHz clock domain. RXD is sent to the PCS Transmit State Machine to generate an encoded 10-bit segment ENC_RXD[0:9]. The PHY serializes ENC_RXD[0:9] to create RX and sends it to the MAC at 1.25 Gbit/s data rate along with the 625 MHz DDR RXCLK.

At the transmit side, the PHY deserializes TX to recover encoded ENC_TXD[0:9]. The PHY passes ENC_TXD[0:9] through the PCS Receive State Machine to recover the GMII signals. In the mean time, the Synchronization block checks ENC_TXD[0:9] to determine the synchronization status between links, and to realign if it detects the loss of synchronization. The decoded GMII signals have to pass through the PHY Transmit Rate Adaptation block to output data segments according to the PHY port speed.

In a Sync-E implementation using SGMII as a MAC-to-PHY interconnect, TXCLK generated in a transmitting MAC device can be traceable to a reference clock (PRC). The RXCLK in a receiving PHY device constitutes the recovered clock of the upstream TXC. Using the RXCLK signal as a reference in a receiving RGMII PHY for downstream links
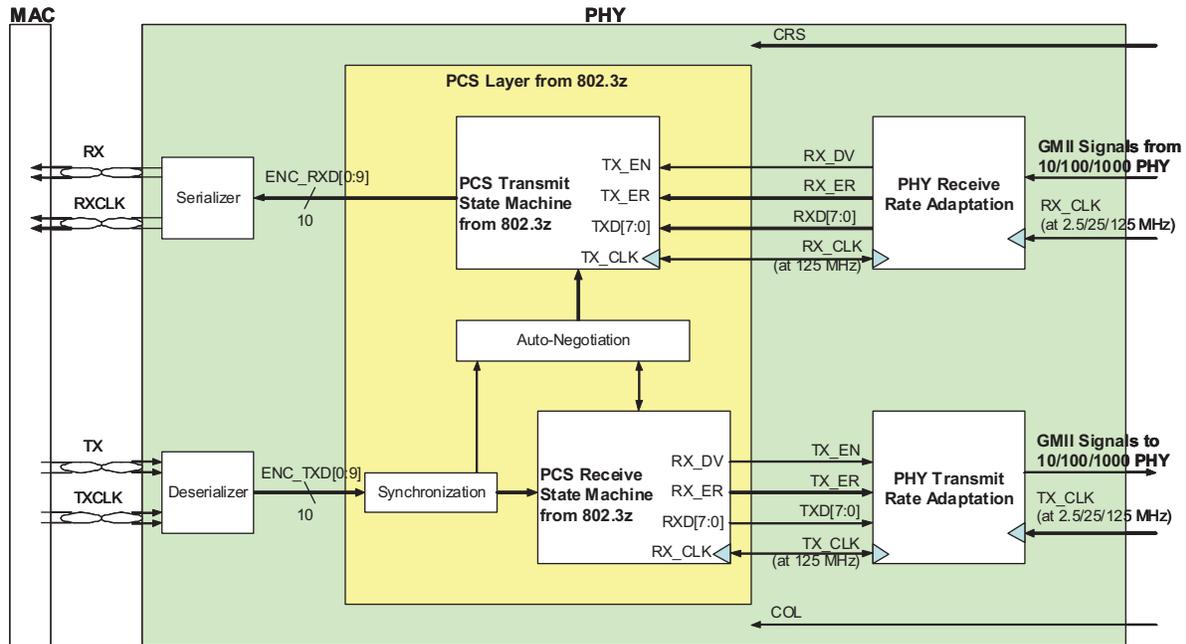
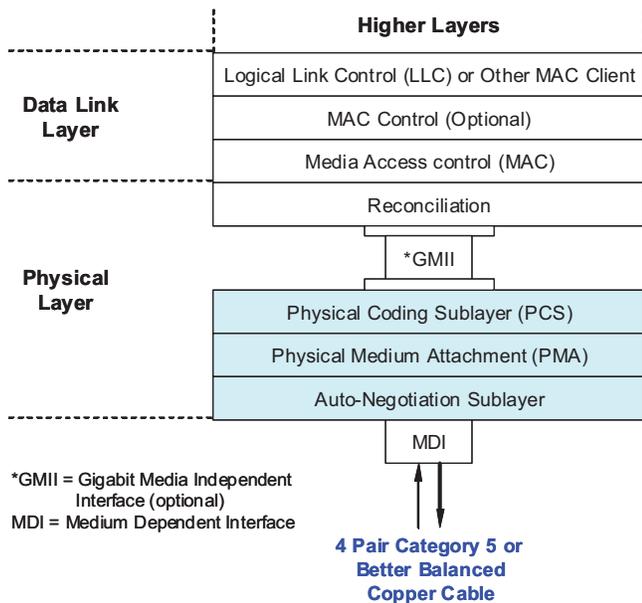Fig. 20. PHY functional block in SGMII specification.

Fig. 21. 1000BASE-T Ethernet reference model [1].

depends on how accessible the RXCLK signal in the chip design. If RXCLK is accessible on the chip, it can be cleaned up and scaled up to 625 MHz for use by the SGMII (receive) logic and also used as the reference clock of the receiving device.

C. Implementing Sync-E with 1000BASE-T Ethernet – Special Considerations

Note that the 1000BASE-T reference model (Fig. 21) is different than that of 1000BASE-X (Fig. 14). Unlike 1000BASE-

X, 1000BASE-T operates over at least Category (Cat) 5 copper cable (with Cat 5e strongly recommended) with four twisted pairs. 1000BASE-T uses PAM-5 (five-level Pulse Amplitude Modulation) line coding as well as transmission over all four pairs of Cat 5 cable (to compensate for limited bandwidth of twisted pairs used in Cat 5 cables). Each pair (that is, each of the two wires) is used in both directions simultaneously – full duplex baseband transmission. Precisely, a wire in a pair supports full duplex transmission. Each of the 4 pairs in the 1000BASE-T medium runs at 125 Mbaud (symbols per second) to achieve the aggregate data rate of 1000 Mb/s (with each wire pair utilizing PAM-5) as illustrated in Fig. 22. 1000BASE-T supports operations over 100 m over the Cat 5 balanced twisted-pair cabling systems.

The use of hybrids and cancellers (Fig. 22) enables full duplex transmission by allowing symbols to be transmitted and received on the same wire pairs at the same time. 1000BASE-T sends encoded signals simultaneously in both directions on the same wire pair as shown in the 4-channel link topology of Fig. 22. This means that the signal at the MDI is the sum of the just-transmitted signal and the signal being received from the entity at the other end of the link. It also means that the receiver must be more than an amplifier and signal detector. The transmitted signal that is added to the received signal at the MDI must be removed (cancelled) before any amplification and detection of the received signal can take place – essentially an echo cancellation problem.

This echo cancellation problem becomes much simplified if the symbol rate, that is, the frequency at which encoded data is transmitted, is identical in both directions. Using a master/slave timing concept (discussed below), 1000BASE-T is able to achieve identical symbol rates in both directions. 1000BASE-T transmission is similar to ISDN and xDSL where digital signal processing algorithms have to be used for echo cancellation.
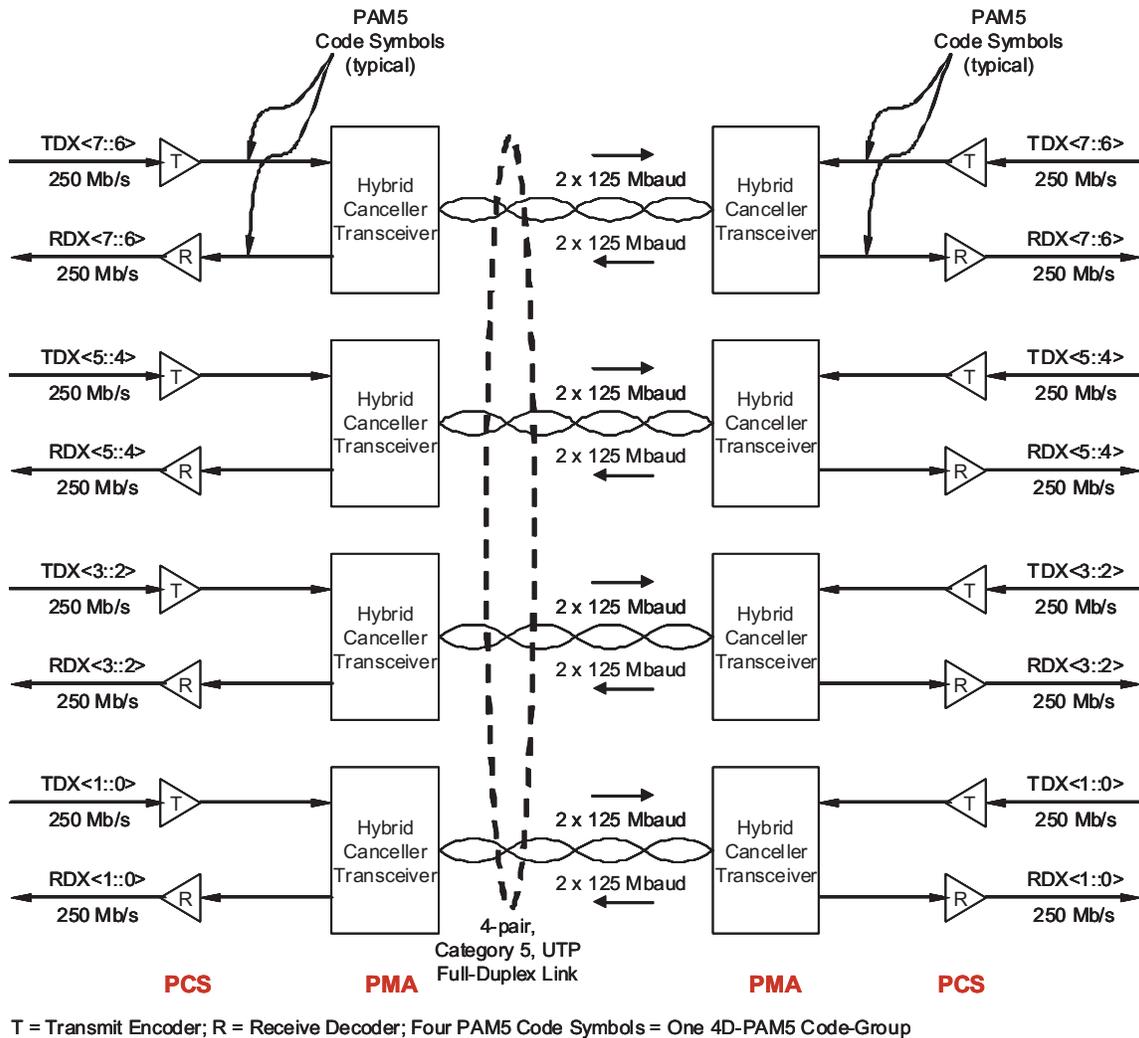
Fig. 22.   1000BASE-T link topology.

Fig. 23 shows a high level diagram of the PCS/PMA data and clock subsystems of the 1000BASE-T PHY. The PMA hybrid transceiver consists of the following:

- A normal Transmit function with four independent transmitters
- A special Receive function, with four independent receivers where the inverse of the transmitted signal is added to the composite signal from the MDI to cancel the transmitted signal
- A clock recovery unit to recover the clock from the received signal sequence

The actual transmitted signal on each wire pair is a 5-level $\{2, 1, 0, -1, -2\}$ pulse amplitude modulated symbol (PAM-5). The four symbols, transmitted simultaneously on the four wire pairs, form a code group (4-D-PAM5) that represents an 8-bit frame octet. The transmitted symbols are selected from a four-dimensional 5-level symbol constellation. Each four-dimensional symbol can be viewed as a 4-tuple (An, Bn, Cn, Dn) of one-dimensional quinary symbols taken from the set $\{2, 1, 0, -1, -2\}$. The 4D-PAM5 coding allows an aggregate 1000 Mb/s data rate to be achieved with a transmission rate of

only 125 Mbaud per wire pairs. The modulation rate of 125 Mbaud matches the GMII clock rate of 125 MHz and results in a symbol period of 8 ns. Encoding of the outgoing frame octets and decoding of the received 4D-PAM5 code groups occurs in the PCS.

1000BASE-T uses a continuous signaling system; in the absence of data, Idle symbols are transmitted. Idle mode is a subset of code-groups in that each symbol is restricted to the set $\{2, 0, -2\}$. A *continuous stream of Idle symbols is sent whenever regular transmission is not in progress to maintain continuous clock synchronization between the two link partners.*

A 1000BASE-T PHY can be configured either as a Master PHY or as a Slave PHY. The Master-Slave relationship between two stations sharing a link segment is established during Auto-Negotiation. The Master can also be set manually. The Master PHY uses a local clock to determine the timing of transmitter operations. The Slave PHY recovers the clock from the received signal and uses it to determine the timing of transmitter operations, i.e., it performs loop timing, as illustrated in Fig. 24. This is required because the echo cancellation problem in 1000BASE-T is simplified if the symbol rate
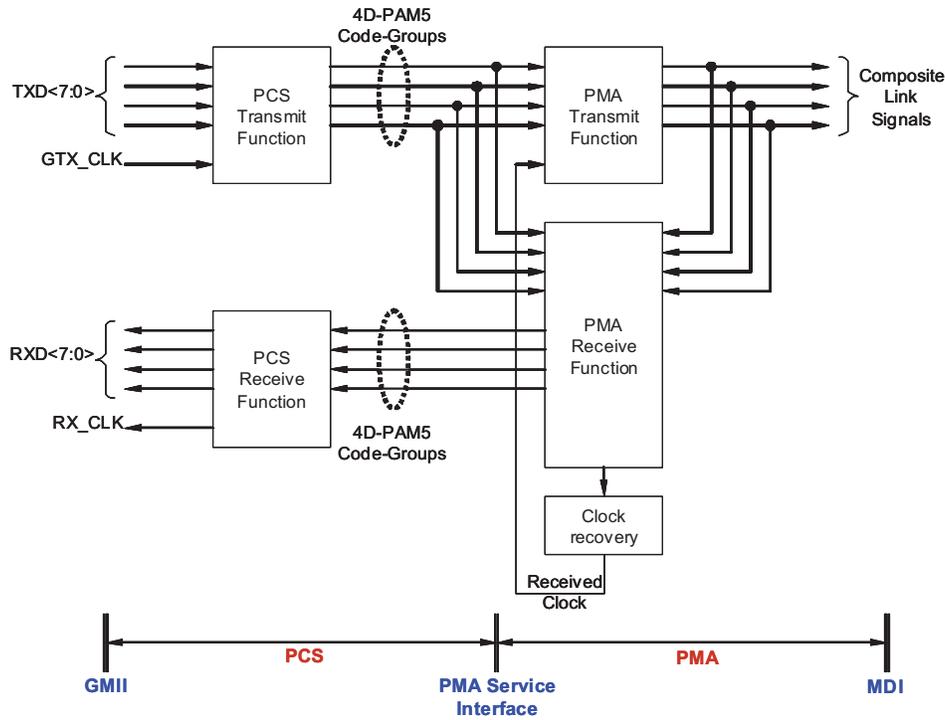
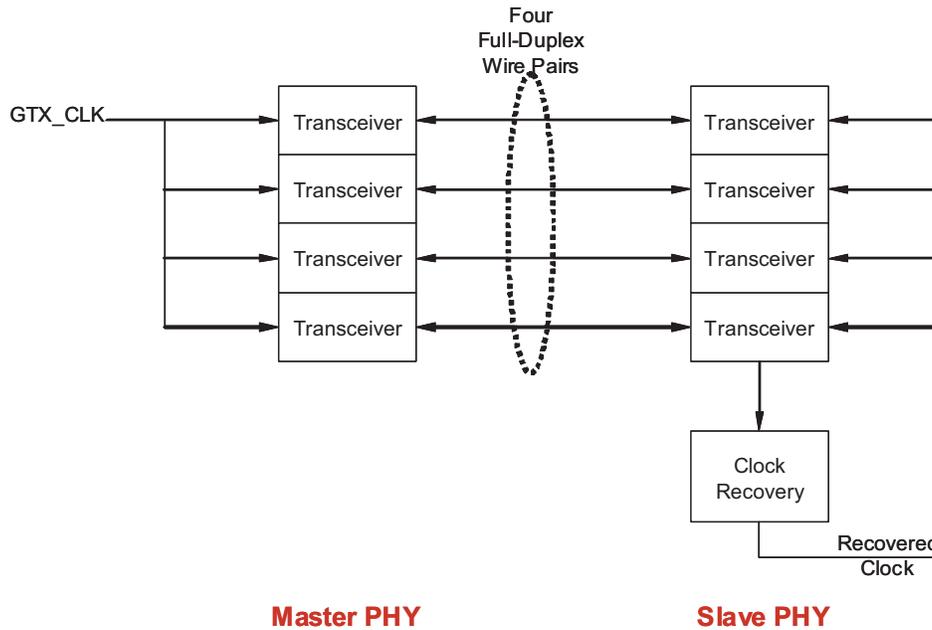Fig. 23.   1000BASE-T PHY data and clock circuits.

Fig. 24.   1000BASE-T master/slave loop timing configuration.

(frequency at which encoded data is transmitted) is identical in both directions. Loop timing is a synchronization procedure in which the Slave is synchronized by the clock recovered from the Master's transmitted symbol stream. In a multiport to single-port connection, the multiport device is typically set to be Master and the single-port device is set to be Slave [1]. *100BASE-X, 1000BASE-X and 10GBASE-R/W do not require master and slave functions because one fiber is always used for transmission and the other for reception (no bi-directional*

transmission on a single fiber or wire) as illustrated in Fig. 25.

Unlike Sync-E with 1000BASE-X (Fig. 26), 1000BASE-T requires special handling when used in Sync-E implementation (see Amendment 1 of G.8262). The master generates a transmit clock that is traceable to a PRC (instead of locally from a free-running crystal oscillator), and the slave recovers the master clock from the received data and uses this recovered clock to transmit its own data. The recovered signal then needs to be cleaned with a PLL to remove noise generated from the
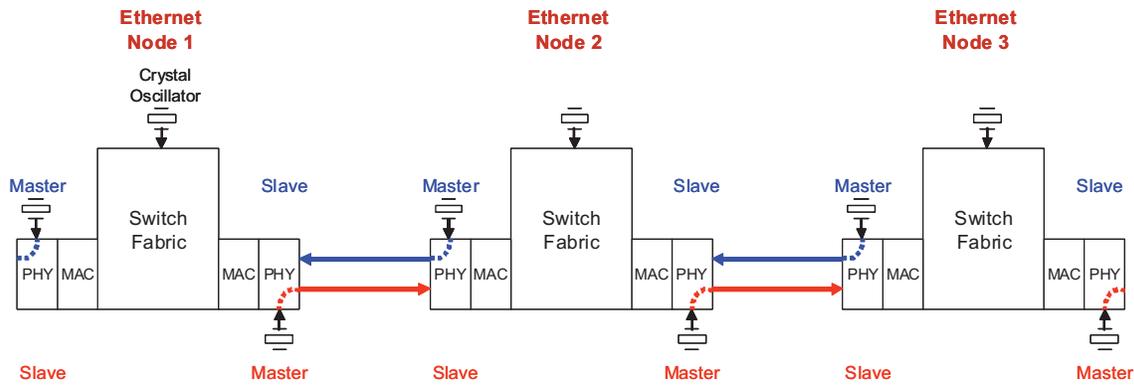
Fig. 25. Timing in native Ethernet versions 100BASE-X, 1000BASE-X and 10GBASE-R/W.
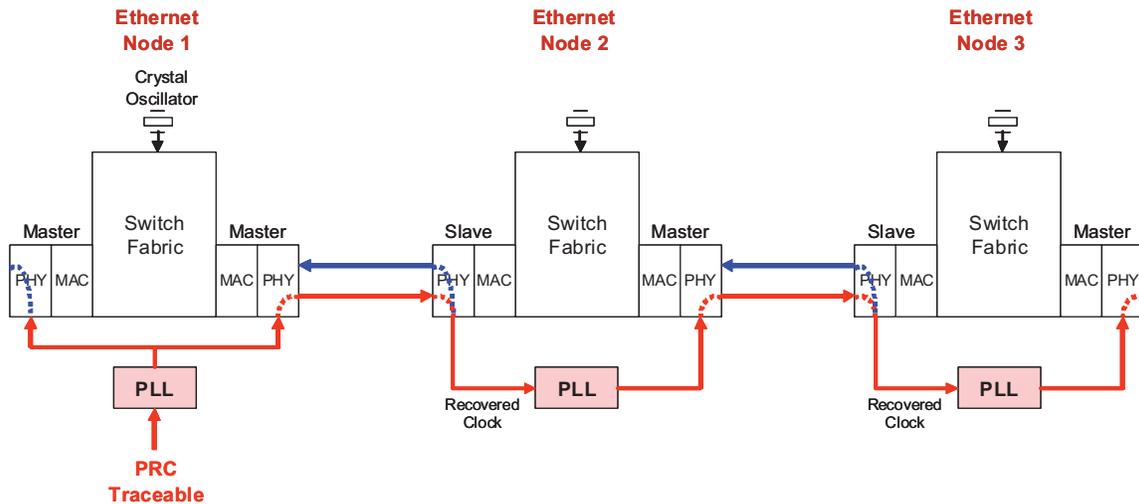

Fig. 26. Timing in Sync-E versions 100BASE-X, 1000BASE-X and 10GBASE-R/W.

clock recovery circuit before being fed to the next transmitting device as shown in Fig. 27.

The clocking arrangement (Fig. 27) is still a Master/Slave concept. The Master and Slave can be configured manually or Master and Slave can still be determined during the auto-negotiation process, but the Master has to be selected while taking into account which of the PHYs has a higher quality clock traceable to the PRC (Stratum 2 preferred over Stratum 3E, Stratum 3E preferred over Stratum 3, etc.).

D. 10 Gigabit Ethernet

The IEEE 802.3 standard [1] defines two main PHY types: the LAN PHY (10GBASE-X (WWDW LAN PHY) and 10GBASE-R (Serial LAN PHY)) and the WAN PHY (10GBASE-W (Serial WAN PHY)). The Serial LAN PHY (10GBASE-R) and Serial WAN PHY (10GBASE-W) provide the same functionality, except the WAN PHY has an extended feature set in the PCS that enables connectivity with SONET STS-192c/SDH VC-4-64c networks. The WAN PHY uses essentially a light-weight SDH/SONET frame running at 9.953 Gbit/s. It is mostly favored for the transport of 10 Gb/s Ethernet across a WAN (a telecom carrier backbone) deploying SDH/SONET or previously installed WDM systems without

having to directly map the Ethernet frames into SDH/SONET [16], [17], [18], [19], [20].

Fig. 28 shows the reference model for the LAN PHYs (10GBASE-X and 10GBASE-R) and the WAN PHY (10GBASE-W). The sublayer types with their corresponding encoding schemes are given in Fig. 29. The purpose of the WIS is to allow 10GBASE-W equipment to generate Ethernet data streams that may be mapped directly to STS-192c or VC-4-64c streams at the PHY level, without requiring MAC or higher-layer processing. The WIS therefore specifies a subset of the logical frame formats in the SONET and SDH standards. In addition, the WIS constrains the effective data throughput at its service interface to the payload capac-ity of STS-192c/VC-4-64c, i.e., 9.58464 Gb/s. Multiplexed SONET/SDH formats are not supported in 10GBASE-W. 10GBASE-W does not provide full SONET/SDH management – provides lightweight SONET/SDH functionality only [1]. Fig. 28 also lists the PHY variants under each type of PHY. Note that 10GBASE-T has a different reference model, a model similar to the 1000BASE-T model (Fig. 21). A majority of the real world 10 Gb/s Ethernet PHYs are of the 10GBASE-R, 10GBASE-W and 10GBASE-T type. The main features of these PHYs are summarized in Table 1. A comparison of the
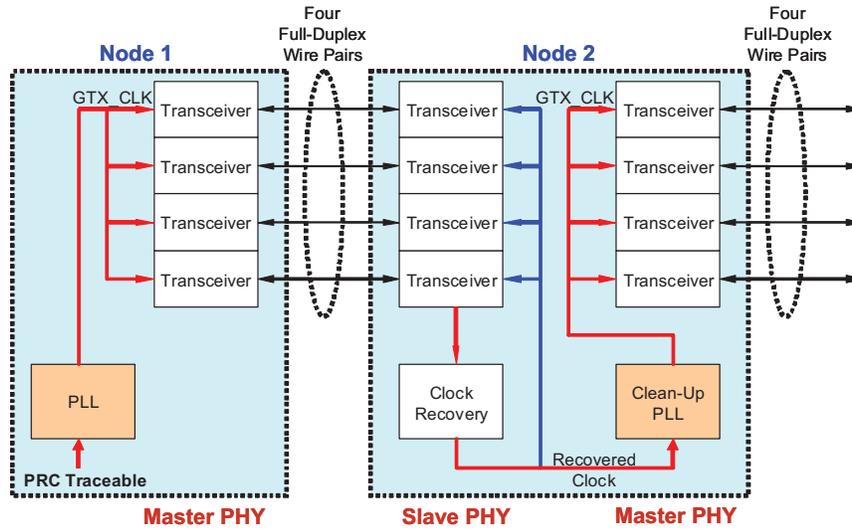
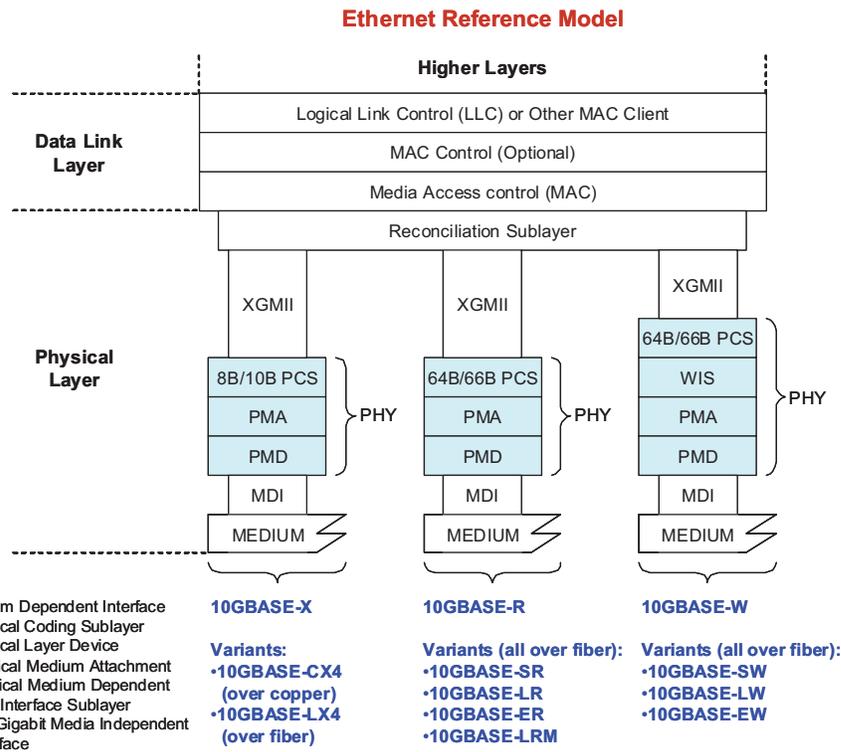Fig. 27. Timing transfer with Sync-E employing 1000BASE-T.

Fig. 28. 10 Gb/s Ethernet reference model [1].

100 Mb/s, Gigabit Ethernet, and 10 Gigabit Ethernet copper-based PHYs are given in Table II.

1) Timing Transfer over Sync-E Employing 10GBASE-R: When Sync-E is implemented over the LAN PHY (10GBASE-R), timing transfer follows the same principles as in 100BASE-X and 1000BASE-X PHYs. Idle period (where data packets are not transmitted) are continuously filled with IDLE symbols and allows for sustained clock recovery at the receiver. High quality timing can be derived from the PHY symbol stream if the transmitter sources it clock from a high quality clock like a PRC.

2) Timing Transfer over Sync-E Employing 10GBASE-W: As noted in Table 1, 10GBASE-W (WAN) defines WAN encoding for 10 Gigabit Ethernet where Ethernet frames are encoded so that they can be carried over SONET STS-192c data rates and SDH VC-4-64 transmission standards allowing for 10 Gbit/s transmission across a WAN. 10GBASE-W does this by wrapping the 64/66B encoded payload into a SONET frame, making the effective rate 9.95 Gbit/s. Thus, unlike 10GBASE-R where timing transfer follows the same principles used in 100BASE-X and 1000BASE-X (as discussed above), Sync-E over 10GBASE-W can adopt the

	PCS		
	10GBASE-X	**10GBASE-R**	**10GBASE-W**
Coding Schemes	8B/10B	64B/66B	64B/66B + SONET Framing
Framing	Ethernet	Ethernet	SONET
Interface Speed (Gb/s)	4x3.125	10.3125	9.95328
	LAN PHY		**WAN PHY**

Fig. 29. PCS sublayer types and encoding schemes.

TABLE I
MAIN FEATURES OF THE 10GBASE-R, 10GBASE-W AND 10GBASE-T PHYS

PCS Specification	Types of PMD Specification	Description
10GBASE-R	10GBASE-ER	10GBASE-ER ("extended range") uses 64B/66B coding at PCS and 1550 nm lasers. It delivers serialized data over single-mode fiber (SMF) at a line rate of 10.3125 Gb/s. 10GBASE-ER has a reach of 40 km.
	10GBASE-LR	10GBASE-LR ("long range") uses 64B/66B coding at the PCS and 1310 nm lasers. It delivers serialized data over SMF at a line rate of 10.3125 Gb/s. 10GBASE-LR has a specified reach of 10 km (but 10GBASE-LR optical modules can often manage distances of up to 25 km with no data loss).
	10GBASE-SR	10GBASE-SR ("short range") uses 64B/66B coding at the PCS and 850 nm lasers. It delivers serialized data over multimode fiber (MMF) at a line rate of 10.3125 Gb/s. Over MMF cabling, it has a range of between 26 m and 82 m depending on cable type. Over new 50 µm, 2000 MHz·km OM3 MMF, it has a range of 300 m.
	10GBASE-LRM	10GBASE-LRM, (Long Reach Multimode) uses 64B/66B coding at the PCS and 1310 nm lasers. It delivers serialized data over MMF at a line rate of 10.3125 Gb/s. 10GBASE-LRM supports distances up to 220 m on FDDI-grade 62.5 µm MMF originally installed in the early 1990s for FDDI and 100BaseFX networks, and 260 m on OM3. 10GBASE-LRM uses electronic dispersion control (EDC) for receive equalization.
10GBASE-W	10GBASE-SW	The 10 Gigabit Ethernet signal encoding types include a WIS which allows 10 Gigabit Ethernet equipment to be compatible with the SONET STS-192c transmission format. SONET equipment is commonly used to carry data communications over long distances, and the STS-192c format provides a payload capacity of 9.58464 Gb/s. 10GBASE-SW is designed for use over short wavelength (850 nm) MMF. The design goal is from 2 m to 300 m of fiber distance, depending on the type and quality of the multimode fiber. The 10GBASE-SW media type is designed to connect to SONET equipment, which is typically used to provide long distance data communications.
	10GBASE-LW	10GBASE-LW is designed for use over long wavelength (1310 nm) SMF. The design goal is from 2 m to 10 km of fiber distance, depending on cable type and quality (longer distances are possible). 10GBASE-LW media type is designed to connect to SONET equipment.
	10GBASE-EW	10GBASE-EW is designed for use over extra long wavelength (1550 nm) SMF. The design goal is from 2 m up to 40 km, depending on cable type and quality (longer distances are possible). 10GBASE-EW media type is designed to connect to SONET equipment.
10GBASE-T	—	10GBASE-T is specified to work up to 55 m with existing Cat 6 cabling. In order to allow deployment at the usual 100 m, the standard uses a new partitioned Category 6a cable specification, designed to reduce crosstalk between UTP cables, known as alien crosstalk. 10GBASE-T is designed to run distances of up to 100 m over four-connector, four-pair structured copper cabling systems. The wire-level modulation in 10GBASE-T is Tomlinson-Harashima precoded (THP) which is a version of pulse-amplitude modulation with 16 discrete levels (PAM-16), encoded in a two-dimensional pattern known as DSQ128.

TABLE II
COMPARISON OF THE COPPER-BASED ETHERNET PHYS

	100BASE-TX	1000BASE-T	10GBASE-T
Line Rate	100 Mb/s	1000 Mb/s	10 Gb/s
Modulation	MLT-3 (PAM-3)	PAM-5	PAM-16
Number of Cable Pairs Used	2	4	4
Modulation Rate/Pair	125 Mbaud	125 Mbaud	800 Mbaud
Key Noise Source	NEXT	FEXT, echo	Alien Crosstalk (ANEXT, AFEXT)
Cabling	Cat5	Cat5 and better	Cat6 and better link segment. Cat5 allowed meeting link segment or alien crosstalk margin computation
Coding	MLT-3	8-state 4D Trellis code across pairs	Tomlinson-Harashima precoding (THP) LPDC + DSQ128

NEXT = Near-End Crosstalk; FEXT = Far-End Crosstalk; ANEXT = Alien Near-End Crosstalk; AFEXT = Alien Far-End Crosstalk

existing optical level timing approaches already employed in SONET/SDH. Alternatively, Sync-E over 10GBASE-W can pass timing via the 64/66B encoded stream since this encoding scheme also guarantees transitions so that the clock can be recovered. However, the logic of using the existing optical carrier timing capabilities of SDH/SONET is very compelling since there is no need to develop another timing transfer functionality for the 64/66B stream.

All timing transfer with SDH/SONET is done by deriving timing from the optical carrier. The optical carrier is synchronous and free from pointer adjustments. SONET/SDH use a pointer mechanism to identify the beginning of frames in their payload. If the payload timing differs from the SONET/SDH network element, or if the timing of two SONET/SDH network elements differs, adjustments are made in the pointers. Pointer adjustments cause significant amounts of jitter and wander. Therefore ANSI [21] recommends that timing is never passed as payload through a SONET/SDH facility. It recommends instead that timing is derived from the previous SONET/SDH network element in the facility.

As noted above, in an SONET/SDH network, all OC-N/STM-N signals are supposed to be synchronous. They are generated by SONET/SDH equipment synchronized by the synchronization network. The latter uses the OC-N/STM-N signals themselves in order to distribute synchronization to all SONET/SDH equipment that generate OC-N/STM-N signals. A SONET/SDH element can be clocked by an external reference signal over its Synchronization Interface Port (T3) (Fig. 30(a)). It can also be clocked by one of the incoming OC-N/STM-N signals. Furthermore, a SONET/SDH element can extract synchronization from one of the incoming OC-N/STM-N signals and deliver it to its Synchronization Interface Port (T4). Using these features, synchronization can be carried from the PRC to the BITS/SSU over OC-N/STM-N signals: at the PRC end, the reference frequency is fed to the SONET/SDH equipment over its Synchronization Interface Port (T3) and synchronizes the OC-N/STM-N signal; at the BITS/SSU end, the OC-N/STM-N signal rate is extracted and fed to the BITS/SSU over the Synchronization Interface Port (T4) of the SONET/SDH equipment.

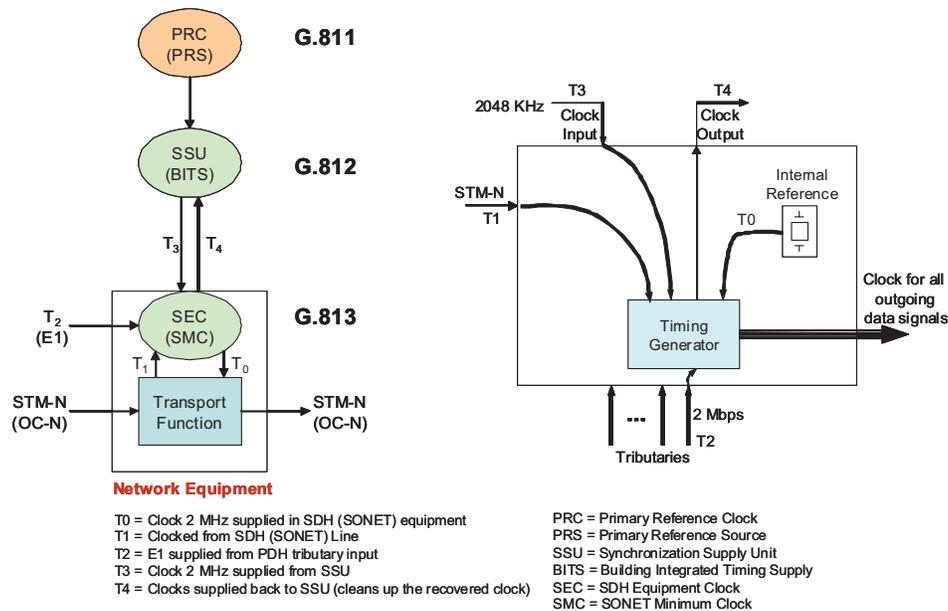The introduction of SONET/SDH technology in transmission networks allow for more effective timing transfer between offices, because SONET/SDH equipment incorporates specific functions of timing generation, filtering, and extraction. The quality of the timing recovered from the optical signals is of high quality, affected only by transmission line jitter (e.g., jitter due to thermal noise and environmental conditions on the optical line), not by bit justification or any other mapping issue. Fig. 30(b) illustrates an example scenario of timing transfer via the optical carrier of Sync-E employing 10GBASE-W. Thus, with Sync-E over 10GBASE-W, the Ethernet frames can be 64/66B encoded as payload and timing transferred via the SDH/SONET optical carrier as illustrated in Fig. 30(b).

3) Timing Transfer over Sync-E Employing 10GBASE-T: In 10GBASE-T data and control symbols are embedded in a framing scheme that runs continuously after startup of the link. Idle periods are continuously filled with Idle symbols as in the other higher speed PHY types. Idle control characters (/I/) are transmitted when idle control characters are received from the XGMII. Similar to a 1000BASE-T PHY, A 10GBASE-T PHY can be configured either as a Master PHY or as a Slave PHY. The Master-Slave relationship between two stations sharing a link segment is established during Auto-Negotiation. The Master PHY uses a local clock to determine the timing of transmitter operations. The Master-Slave relationship may include loop timing. If loop timing is implemented, the Slave PHY recovers the clock from the received signal and uses it to determine the timing of transmitter operations, i.e., it performs loop timing. If loop timing is not implemented, the Slave PHY transmit clocking is identical to the Master PHY transmit clocking. Timing transfer in Sync-E over 10GBASE-T is similar to Sync-E over 1000BASE-T. Instead of a free-running local clock, the Master PHY has a clock that is traceable to a PRC and this clock is transferred as PHY symbol clock to the receiver (Slave PHY).
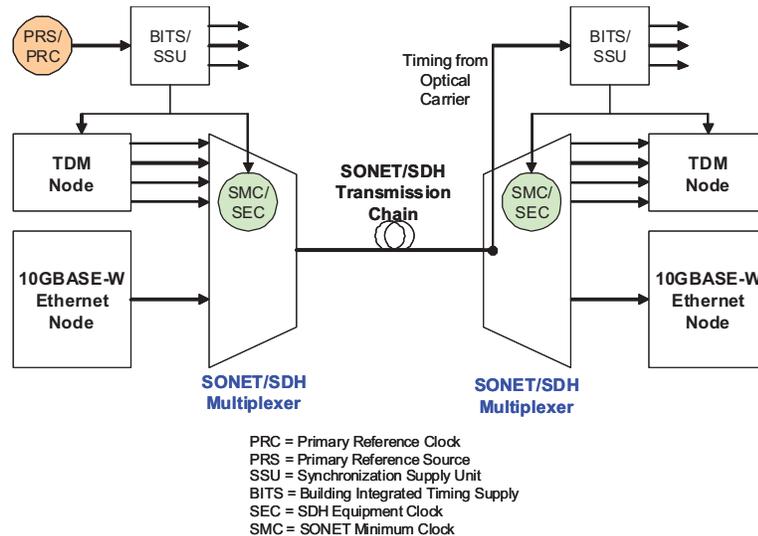
4) Physical Layer Timing and 10 Gb/s Ethernet Media Independent Interface Types: The 10 Gigabit Media Independent Interface (XGMII) has been standardized in IEEE Std. 802.3 and is the interface between the MAC and PHY component. However, this interface can optionally be extended by a so called XGMII Extender Sublayer (XGXS), thus introducing an intermediate interface called 10 Gigabit Attachment Unit Interface (XAUI) (Figs. 31, 32, and 33). Fig. 33 shows the relationships between the XGMII and XAUI layers with corresponding signal rates.

The XGMII (Fig. 34) consists of 74 signal lines and is suitable for short range on-board component interconnection. XGMII is composed of 32 transmit channels (4 lanes of 8 bits each, TXD<31:0>), 32 receive channels (RXD<31:0>), 1 transmit clock (TX_CLK), 1 receive clock (RX_CLK), 4 transmitter control characters (TXC<3:0>), and 4 receive control characters (RXC<3:0>) for a 74-pin wide interface in total. XGMII is typically used for on-chip connections but in chip-to-chip usage it is mostly replaced by the XAUI. XAUI as an optional XGMII Extender is primarily intended as a chip-to-chip (IC to IC) interface implemented with traces on a printed circuit board.

As discussed above, the XAUI is an optional, self-managed interface that can be inserted between the reconciliation sublayer and the PHY layer to transparently extend the physical reach of the XGMII. The XAUI as an optionally inserted

T0 = Clock 2 MHz supplied in SDH (SONET) equipment
T1 = Clocked from SDH (SONET) Line
T2 = E1 supplied from PDH tributary input
T3 = Clock 2 MHz supplied from SSU
T4 = Clocks supplied back to SSU (cleans up the recovered clock)

PRC = Primary Reference Clock
PRS = Primary Reference Source
SSU = Synchronization Supply Unit
BITS = Building Integrated Timing Supply
SEC = SDH Equipment Clock
SMC = SONET Minimum Clock

(a) Synchronization architecture for SDH/SONET and the different clock inputs of a network element

PRC = Primary Reference Clock
PRS = Primary Reference Source
SSU = Synchronization Supply Unit
BITS = Building Integrated Timing Supply
SEC = SDH Equipment Clock
SMC = SONET Minimum Clock

(b) Timing transfer over Sync-E employing 10GBASE-W link

Fig. 30.   SONET/SDH based synchronization.

transparent interface multiplexes some of the XGMII signals resulting in a smaller line count (16) and increased operational distance. XAUI is a self-timed interface allows jitter control to the PCS. Since both ends of this interface will have to deliver XGMII signals, there is a certain glue logic required that does the serialization/ deserialization (including 8B/10B encoding) to match the 16 XAUI signal lines. The overall speed remains at 10 Gb/s as required by the XGMII.

XAUI addresses several physical limitations of the XGMII. When implemented as a chip-to-chip interface, XGMII signaling is based on the High Speed Transceiver Logic (HSTL) class 1 single-ended I/O standard, which has an electrical distance limitation of approximately 7 cm. Because XAUI uses low voltage differential signaling method, the electrical limitation is increased to approximately 50 cm. Another advantage of XAUI is simplification of backplane and board

trace routing. XAUI (Figs. 34 and 35) only consists of 4 differential transmitter channels and 4 differential receiver channels for a 16-pin wide interface in total. This reduction in pin count significantly simplifies the routing process in the layout design.

At the transmit side of the XAUI interface, the data and control characters are converted within the XGXS into an 8B/10B encoded data stream. Each data stream is then transmitted across a single differential pair running at 3.125 Gb/s. At the XAUI receiver, the incoming data is decoded and mapped back to the 32 bit XGMII format. This process provides a transparent extension of the physical reach of the XGMII and also reduces the interface pin count.

XAUI functions as a self-managed (self-clocked) interface because code-group synchronization, channel deskew, and clock domain decoupling are handled with no upper layer
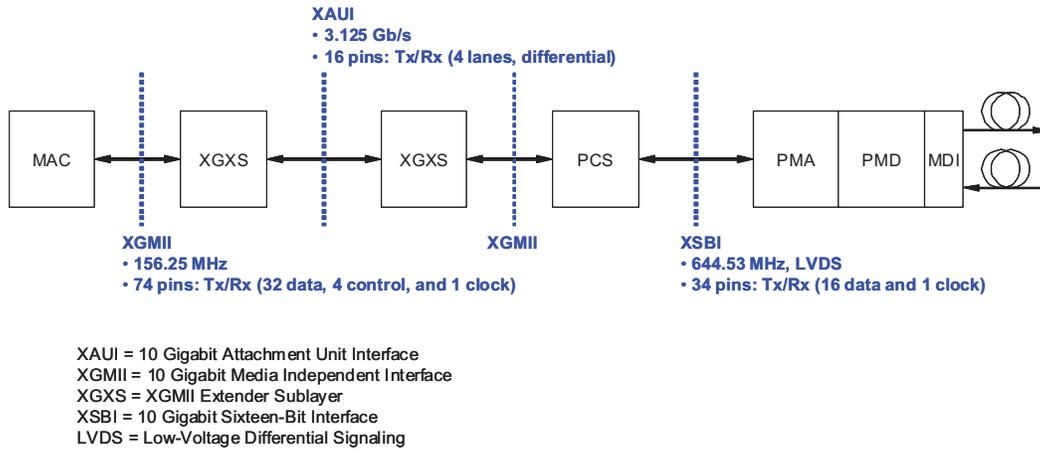
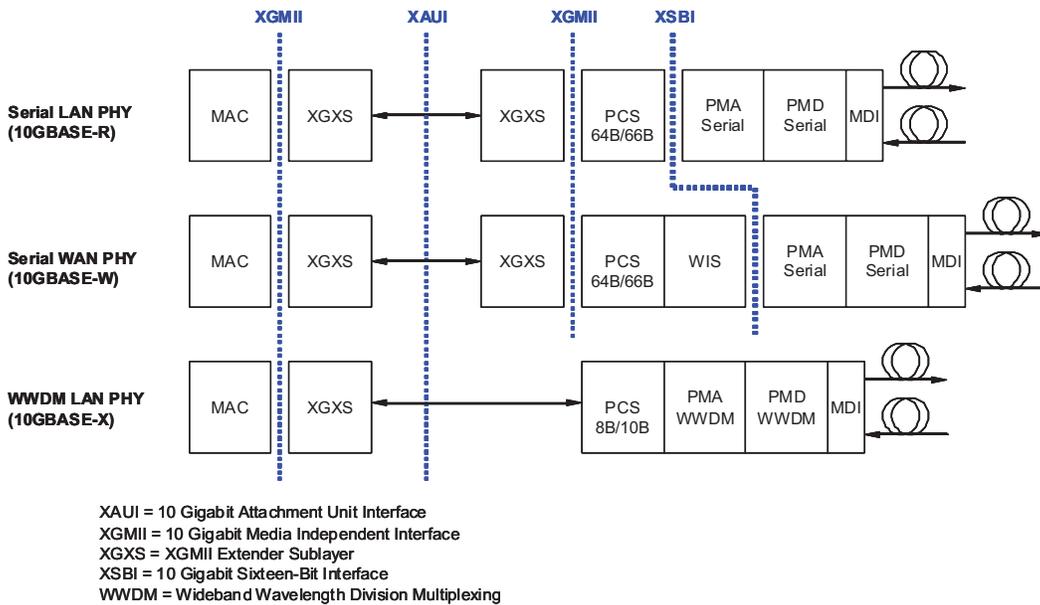Fig. 31. 10 Gigabit Ethernet PHY sublayer electrical boundaries.

Fig. 32. 10 Gigabit Ethernet PHY architectures.

support requirements. These features are accomplished based on PCS code-groups that are used during the Inter-Packet Gap (IPG) time and during idle periods. PCS code-groups are mapped by the XGXS to XGMII characters. Using Figs. 33–36, the following points can be highlighted for XGMII and XAUI:

- XGMII is organized into four lanes with each lane conveying a data octet or control character on each edge of the associated clock. The source XGXS converts bytes on an XGMII lane into a self clocked, serial, 8B/10B encoded data stream. Each of the four XGMII lanes is transmitted across one of the four XAUI lanes.
- The source XGXS converts XGMII Idle control characters (interframe) into an 8B/10B code sequence. The destination XGXS recovers clock and data from each XAUI lane and deskews the four XAUI lanes into the single-clock XGMII.

- The destination XGXS adds to or deletes from the interframe as needed for clock rate disparity compensation prior to converting the interframe code sequence back into XGMII Idle control characters.
- XAUI has built-in capabilities to overcome the inter-lane signal-skewing problems using a type of automatic deskewing.
- The XGXS at the PHY end of the XGMII Extender (PHY XGXS) and the XGXS at the Reconciliation Sublayer end (DTE XGXS) may operate on independent clocks.

The 10 Gigabit Ethernet Sixteen-Bit Interface (XSBI) connects the PMA to its client which can be either the PCS or the WIS (Figs. 31, 32, and 33). XSBI is intended to be an *electrical interface* between 300-pin Multi-Source Agreement (MSA) optical transponders and the ICs implementing the PCS or WIS. The data and clock rates for the various interfaces discussed so far are given in Table III. XSBI supports full duplex
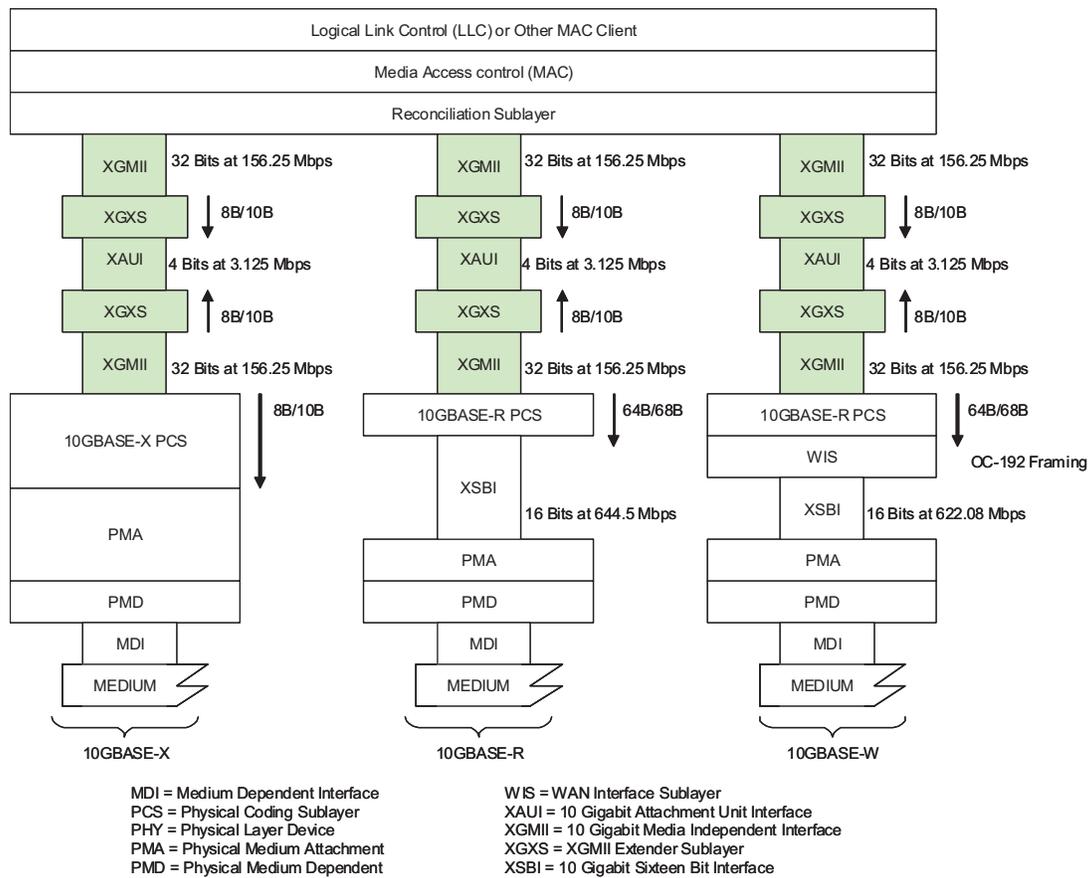
Fig. 33. 10 Gb/s Ethernet reference model showing XGMII, XAUI and XGXS and data rates at each sublayer.

transmission of data-groups with associated data-group clock signals. As shown in Fig. 37, the PMA client provides data-groups on TX_D[15:0] to the XSBI transmit function which latches the data on the rising edge of the PMA_TX_CLK. The PMA_TX_CLK signal is derived from PMA_TXCLK_SRC as provided by the PMA. An internal Transmit Clock Generation Unit (TXCGU) uses REFCLK to generate the internal bit clock that is used to serialize the latched data out of the PMA outputs.

The PMA Receive function accepts serial data from the PMD and extracts a bit clock and recovered data from the serial inputs in the Receive Clock Recovery Unit (RXCRU). The recovered data is deserialized and conveyed to the PMA client on RX_D[15:0]. The rising edge of the recovered clock, PMA_RX_CLK, which is at 1/16 the bit rate, is used by the PMA to send the received 16-bit data-groups to the PMA client. The key XSBI signals are as follows:

- TX_D[15:0]: This is the 16-bit parallel low-voltage differential signaling (LVDS) data presented to the XSBI for serialization and transmission onto the media.
- RX_D[15:0]: This is the 16-bit parallel LVDS data transmitted by the PMA and presented to the PMA client for further processing. The PMA deserializes the incoming data to 16-bit wide data groups, RX_D[15:0].
- PMA_TXCLK_SRC: This is the LVDS transmit data-group clock source used by the PMA client to derive PMA_TX_CLK. The PMA derives the

PMA_TXCLK_SRC from the local REFCLK.

- PMA_TX_CLK: This is the LVDS transmit data-group clock used to latch data into the PMA for transmission. The PMA_TX_CLK is derived from PMA_TXCLK_SRC. The rising edge of PMA_TX_CLK is used to latch data into the PMA for transmission.
- PMA_RX_CLK: This is the LVDS receive data-group clock derived from the serial input data whenever possible. The PMA derives the PMA_RX_CLK from the line rate when the serial data from the PMD is present, or the REFCLK when the serial data from the PMD is absent.
- Sync_Err: This is a signal is used to indicate the inability of the PMA to recover the clock from the serial data stream.

Table IV shows the XSBI clock and data rates for WAN (10GBASE-W) and LAN (10GBASE-R) PHYs. In XGMII, the four lanes in each direction share a common clock—TX_CLK for transmit and RX_CLK for receive. TX_CLK provides the timing reference for the transfer of the TXC<3:0> and TXD<31:0> signals from the Reconciliation Sublayer to the PHY. RX_CLK is a continuous clock that provides the timing reference for the transfer of the RXC<3:0> and RXD<31:0> signals from the PHY to the Reconciliation Sublayer. The RX_CLK clock signal may be derived from the received data or it may be that of a nominal clock (e.g., TX_CLK). RX_CLK has frequency of 156.25 MHz ±100 ppm.
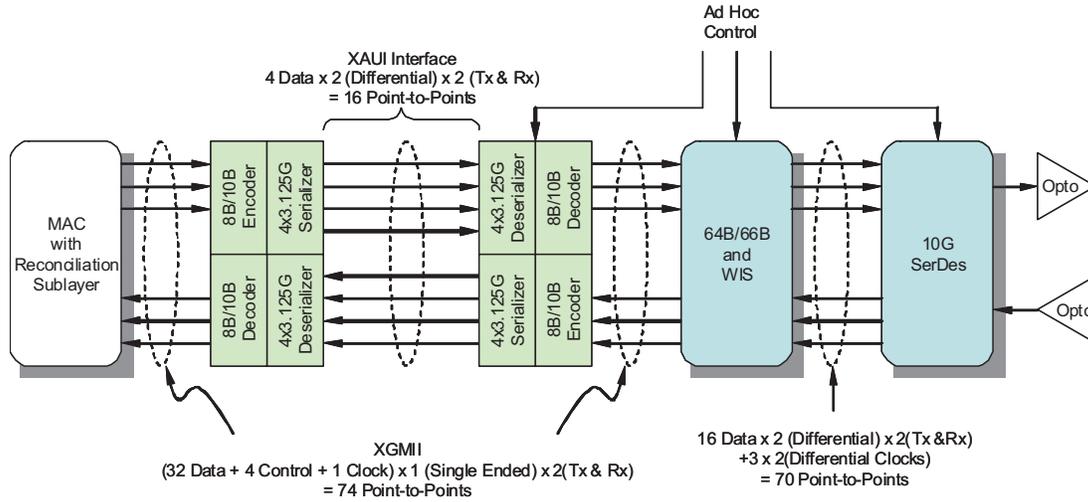
Fig. 34. Building blocks of 10 Gigabit Ethernet rransceiver – Serial WAN PHY.
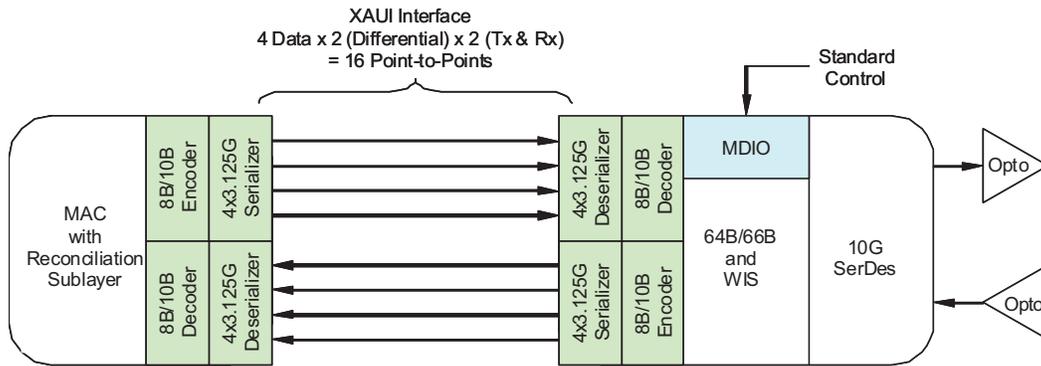

Fig. 35. Selecting XAUI as interface between system electronics and optical transceiver – Serial WAN PHY.

In a pure XGMII Sync-E implementation (no XAUI), the TX_CLK (Fig. 36) in a transmitting device is traceable to a reference clock (PRC) while the RX_CLK in a receiving device constitutes the recovered clock of the upstream TX_CLK. The RX_CLK can then be cleaned up and in turn used to source the TX_CLK of the receiving device. In non-Sync-E implementation (i.e., IEEE 802.3), TX_CLK in the receiving device is driven by a free-running clock. In using XAUI in a Sync-E implementation, the XGXS at the PHY end of the XGMII Extender (PHY XGXS) can derive timing from the XGMII at the Reconciliation Sublayer. With this the RX_CLK of the receiving XGMII (PHY end) can be made traceable to the TX_CLK of the transmitting XGMII (Reconciliation Sublayer end) which in turn is traceable to the PRC.

XSBI can also be used in a Sync-E implementation depending on the client (PCS or WIS chip) and optical transponder integration and pin exposure. In a Sync-E implementation, RE-FCLK can represent a clock traceable to the PRC. The PMA then derives the PMA_TXCLK_SRC from REFCLK and the PMA_TX_CLK, which is then used to latch data into the PMA for transmission, is in turn derived from PMA_TXCLK_SRC. The PMA derives the PMA_RX_CLK from the line rate when the serial data from the PMD is present.

TABLE III
10-GIGABIT ETHERNET INTERFACES CHARACTERISTICS

Interface	Width	Clock Rate (MHz)	Data Rate Per Channel
XGMII	32	156.25	312.5 Mb/s
XAUI	4	None	3.125 Gb/s
XSBI	16	644.5 (LAN PHY) 622.08 (WAN PHY)	644.5 Mb/s (LAN PHY) 622.08 MB/s (WAN PHY)

V. UNDERSTANDING WHY SYNC-E CANNOT BE IMPLEMENTED OVER 10 MB/S ETHERNET

Fig. 38 shows the relationship between the ISO reference model and the 10BASE-T model (which also corresponds to the implementation in an Ethernet node (DTE)). The MAC is responsible for transmit and receive message data formatting, and the enforcement of the CSMA/CD (Carrier Sense Multiple Access with Collision Detection) protocol. In half-duplex mode, an Ethernet node uses CSMA/CD to determine when it is permitted to transmit over the shared medium.
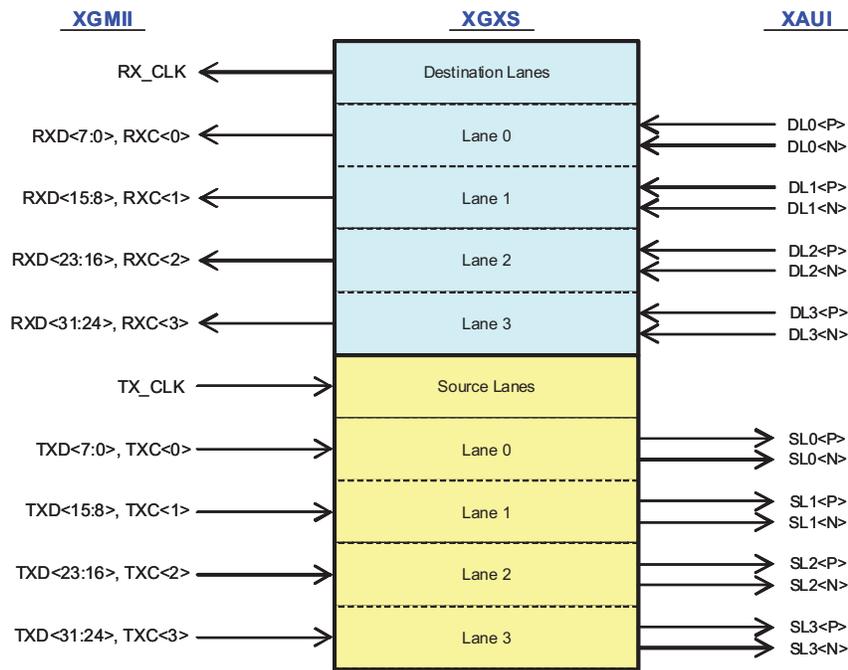
Fig. 36.   XGMII and XAUI signals and inputs and outputs of XGXS [1].

TABLE IV
XSBI CLOCK AND DATA RATES FOR WAN (10GBASE-W) AND LAN
(10GBASE-R) PHYs

Parameter	WAN PHY	LAN PHY	Unit
TX_D[15..0]	622.08	644.53125	Mb/s
PMA_TXCLK	622.08 ± 20 ppm	644.53125 ± 100 ppm	MHz
PMA_TXCLK_SRC	622.08 ± 20 ppm	644.53125 ± 100 ppm	MHz
RX_D[15..0]	622.08	644.53125	Mb/s
PMA_RXCLK	622.08 ± 100 ppm	644.53125 ± 100 ppm	MHz

A. 10BASE-T Physical Layer Architecture

The PLS provides the interface between the MAC and the AUI. The PLS and the AUI subsystems support the signaling scheme between the MAC layer and the MAU. Although considered part of the Physical (PHY) layer, the PLS function is normally implemented locally to the MAC function. The PLS is responsible for four principal functions which are transmit Data Output, receive Data Input, Carrier Sense, and Error Sense [1]. The three input functions – Data Input, Carrier Sense, and Error Sense – are provided by the PLS to the MAC, to allow it to perform the CSMA/CD protocol. The MAC transmits serial data which is output in *Non-Return-to-Zero* (NRZ) format. The data is then *Manchester* encoded by the PLS and transmitted over the AUI to the MAU, using *differential signaling* technique. Manchester encoding allows

both clock and data information to be combined into a single "bit-symbol." Each bit-symbol consists of two halves, with the first half being the logical inverse of the data bit to be encoded, and the second half of the symbol always being the logical value of the data bit itself (or the inverse of the first half). This guarantees that there is a signal transition in the center of each bit-symbol (or bit cell center (BCC)). Fig. 39 shows the relationship of the clock, NRZ, and Manchester encoded data. Note that a sequence where each successive data bit differs from the previous bit (such as in the Ethernet preamble) exhibits a 5 MHz frequency, whereas a sequence containing identical data (all 0s or 1s) exhibits a 10 MHz frequency (Fig. 40).

The AUI provides the signaling path between the PLS function and the MAU. The AUI provides a defined interface to allow a special cable and connector assembly to be used to connect the PLS function to the MAU. This allows the MAC/PLS to be located remotely from the MAU, and hence the network medium. In many practical implementations, the MAC and PHY are fully integrated, with the AUI embedded – no provision is made for an exposed AUI connection (AUI exists in a virtual sense). This is important since the provision of an explicit (exposed) AUI directly results in a significant cost penalty. The AUI (cable) typically consists of three differential signal pairs (Data Out (DO), Data In (DI), Control In (CI), with an optional fourth which is virtually never implemented), plus power and ground connections. Differential signaling is employed on each of the interface signals to allow the signals to pass through a transformer isolation barrier, which provides the Ethernet station protection from severe network medium faults.

The MAU is responsible for the actual physical and electrical interfaces to/from the medium. The MAU provides the functional, electrical and mechanical interface between the
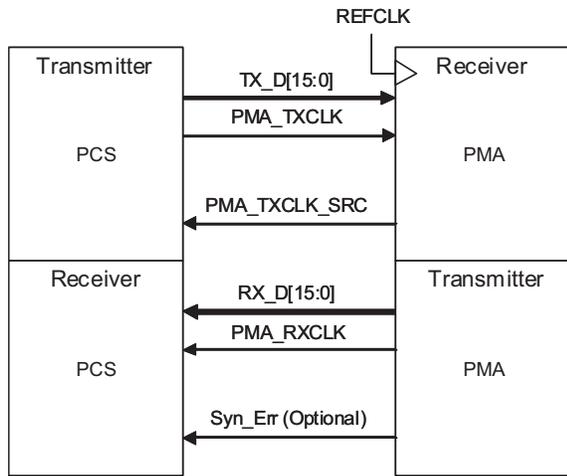
Fig. 37. XSBI functional block diagram.

Ethernet station and the network medium. It is important to note that for 10 Mb/s Ethernet, the MAC, PLS and AUI functions are preserved regardless, of medium, only the MAU is required to change. The MAU performs the following primary functions – Transmit Data, Receive Data, Loopback, Collision Detection, SQE (Signal Quality Error) Test, Jabber Protection, and Link Integrity (used in 10BASE-T Ethernet) [1]. *Among these functions, understanding the workings of Link Integrity function is paramount to understanding why Sync-E cannot be implemented over 10 Mb/s Ethernet.*

The 10BASE-T cable is inexpensive and is the commonly available telephone grade (Cat 3), 100 ohms UTP. For the MDI (see Fig. 38), simple, inexpensive RJ-45 type telephone jack connectors are used to connect the MAU to the medium. The MAC and the PHY layer are primarily supported directly in silicon, with implementations available from multiple semiconductor companies. The LLC sublayer, with which the MAC forms the Data Link Layer (DDL), is normally implemented in software, as are the layers above.

B. Understanding the 10 Mb/s Link Integrity Function

Coax-based 10 Mb/s Ethernet systems (10BASE-2 and 10BASE-5) [1] are designed such that both the transmitter and receiver are connected to the coax center tap. So when the Ethernet node transmits, the MAU simultaneously receives the transmission and returns it to the MAC controller as an indication of transmit-to-receive path integrity. However, the separation of the transmit and receive paths in 10BASE-T leads to two problems. The first is the Ethernet station cannot sense its own transmission (the transceiver will not see activity on the receive pair unless a collision occurs), and the second is it cannot detect a failed link. A broken transmit path is a problem (the node will not be able to transmit data), but a broken receive path is a far serious problem – the node loses its ability to monitor the network for activity and collisions, meaning a node would transmit regardless of the current conditions.

To address the first problem, a loopback path to the MAC controller is implemented internally to the MAU. To address

the second problem (in MAUs that use separate transmit and receive signaling paths, such as twisted pair and fiber), a special mechanism, Link Integrity, is required to detect whether a valid communications path exists over the network. This facility is important in ensuring correct network operation in these systems, since a break in the receive signal path will render the Carrier Sense facility inoperable.

Since separate transmit and receive signal paths are employed in the Unshielded Twisted Pair (UTP) cable, the 10BASE-T MAU at each end of the link employs a Link Integrity test to monitor for end-to-end continuity [1]. This is performed by each MAU transmitting a "link test" pulse when the Ethernet station has no packets to transmit (Fig. 41). In the absence of network traffic, a simple heartbeat pulse is sent periodically (every 16 ms $\pm$ 8 ms) by the transmitter of all 10BASE-T MAUs. The link test pulse is a unipolar pulse (positive only), unlike the normal 10BASE-T differential signaling used for packet data. The receiver in each MAU must detect packet data or link test pulses (Fig. 42) to remain in the "Link Pass" state. If the MAU detects neither, it will enter the "Link Fail" state, and effectively prevent the Ethernet station from transmitting on the network.

If the receiver of a MAU does not see either a packet data or a link test pulse within a defined time window (50 to 150 ms), it will enter the Link Fail state, which disables the data transmit, data receive, and loopback functions. The MAU disables the transmit function to prevent the disturbance of existing network traffic, thereby preventing false late-collision scenarios. The loopback path is disabled to warn the Ethernet station that a failure has occurred, since the transmit (DO) to receive (DI) loopback path is interrupted. When the receive pair is disconnected, the MAU enters the Link Fail and further transmission ceases. However, even during the Link Fail state, the transmission and reception of link test pulses still continues. The link test pulse is passed only between 10BASE-T MAUs and is not observed on the AUI signaling between the MAU and the Ethernet station (Fig. 42).

C. Reasons Why Sync-E Cannot be Implemented with 10 Mb/s Ethernet

From the above discussion, we see that 10BASE-T Ethernet is not capable of timing distribution because signal transmission (Manchester encoding) over the PHY interface stops when there are no data packets for transmission – the 10BASE-T transmitter stops sending normal pulses during idle periods. Instead, the 10BASE-T transmitter simply sends a link test pulse (keepalive pulse) every 16 ms $\pm$ 8 ms to notify its presence to the receiving end. *Even with this, the unpredictability, variability of generation instant, and frequency of these pulses make them unsuitable for clock recovery at the receiver.*

In 100BASE-X and higher speed flavors of Ethernet, in the absence of data packet transmission, IDLE symbols are continuously transmitted unlike in 10BASE-T where link test pulses are transmitted between data packets. The IDLE code also serves as a keepalive signal to the receiver. *By continuously filling idle periods in the faster Ethernet PHYs (100 Mb/s, 1 Gb/s and 10 Gb/s) with normal pulse transition which are in the Idle codes (no unipolar pulses), continuous*
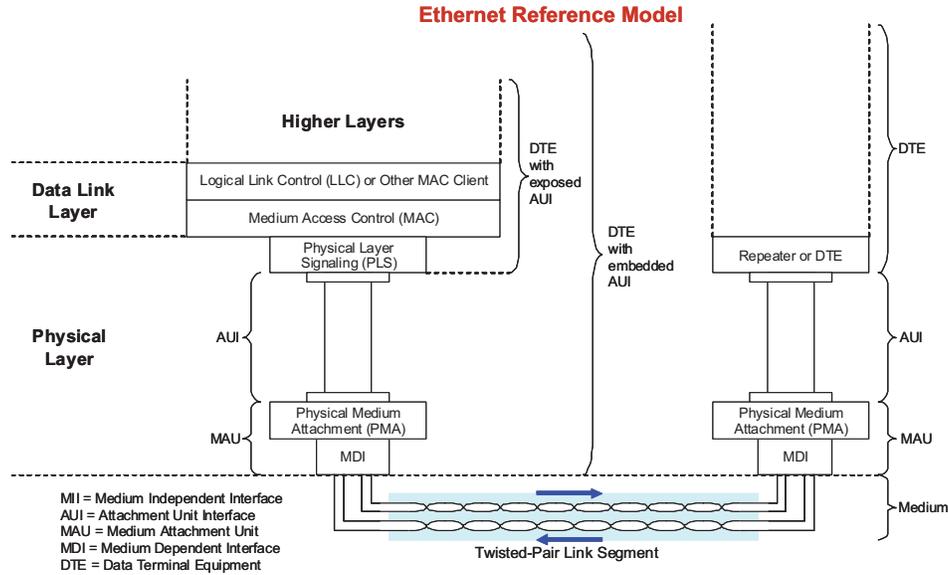
Ethernet Reference Model

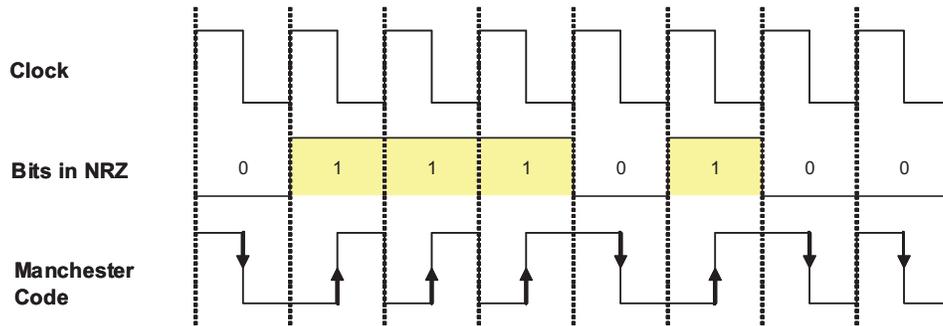Fig. 38.   10BASE-T Ethernet reference model [1].

Fig. 39.   Manchester encoding.

high-quality clock recovery at the receiver becomes possible. This important feature of the higher speed PHYs make them capable of distributing high quality timing from a standard reference clock. The concept of timing distribution over the higher speed Ethernet PHYs is illustrated in Fig. 43.

VI. SYNC-E NODE TIMING ARCHITECTURES

In this section we describe a generic Sync-E node architecture and how timing transfer is achieved through the Ethernet ports. We assume that somewhere in the network a PRC feeds an accurate timing reference into the network.

A. Node Architectures

Fig. 44(a) shows the main timing elements of a generic Sync-E node. At each node a timing recovery unit (clock and data recovery (CDR)) will recover the clock from an upstream synchronized data stream, clean it up, and use it as the transmit clock to the next node which could be a downstream Sync-E or an end node requiring accurate timing reference. As illustrated in greater detail in Fig. 44(b), in Node 1 the CDR recovers the clock and data, passes the recovered clock, R_CLK, to the Clean-up PLL in the timing module. The Clean-up PLL cleans

(improves) the clock and passes it to the clock multiplier unit (CMU) which in turn generates the necessary output clock(s). The output clock remains synchronized to the original R_CLK. The downstream CDR in Node 2 receives the output clock and data from Node 1 and repeats the above process. The receive and transmit of clock and data continues through multiple nodes in the network until it reaches its final destination. As shown in Fig. 44b, the cleaned up clock is also used to transmit data on the transmit port in the upstream direction (loop timing).

Revisiting Fig. 44(a), the PHY will pass the R_CLK to the timing module which in turn will generate all the necessary clocks for all the different elements and protocols (Ethernet, E1, T1, etc.) the node interfaces with. The timing module is also capable of supporting multiple selectable input reference frequencies, and generate multiple output clocks (copies). In this example system, the system timing reference is selectable from either the Building Integrated Timing Supply/Synchronized Supply Unit (BITS/SSU) clock reference or the Sync-E uplink ports. The BITS/SSU timing reference provides a working and protection clock signal to a clock selection multiplexer. The two recovered line clocks from the Sync-E ports are also fed into the clock selection multiplexer.
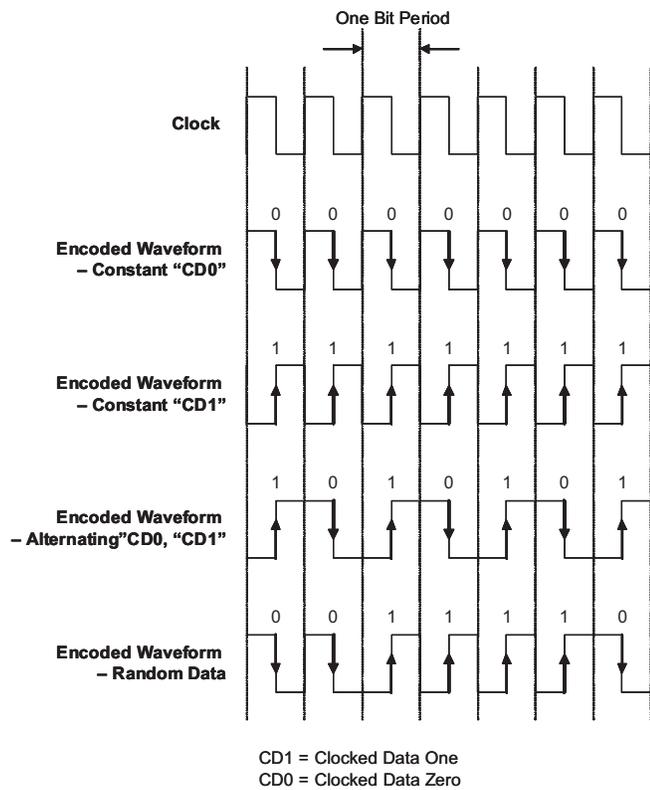
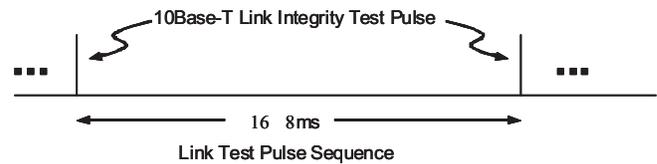Fig. 40. Examples of Manchester encoded waveforms.

Fig. 41. 10 Mb/s Ethernet link integrity test signal.

network, the recovered clock is used as an accurate frequency source for use externally (PDH/SDH clock, radio frequency generation, etc).

Jitter is the short-term phase variations of the significant instants of a digital signal from their ideal positions in time. Wander refers to long-term variations in the significant instants. ITU-T G.810 classifies phase variation frequencies below 10 Hz as wander and frequencies at or above 10 Hz as jitter. As in PDH and SDH/SONET networks, three main timing related concerns in a Sync-E node are ***jitter (wander) transfer (***or ***jitter (wander) attenuation)***, ***jitter (wander) generation (***or ***intrinsic jitter (wander))***, and ***jitter (wander) tolerance***. The key requirements for these timing related issues are defined in ITU-T G.8262 which also describes requirements related to bandwidth, frequency accuracy, holdover of Sync-E clocks. Jitter (wander) transfer is a measure of how much jitter (wander) is transferred between input and output of network equipment. The transfer is a function of the bandwidth of the PLL and the input jitter (wander). For the PLL, the internal low pass loop filter determines the jitter (wander) attenuation. The PLL can be viewed as a jitter and wander filter. The narrower the loop bandwidth, the better the jitter and wander attenuation.

As shown in Fig. 44(c), the physical layer timing path traverses three PLLs – the CDR, Clean-up, and CMU PLLs. The figure shows the relative jitter (wander) transfer characteristics of the three PLLs in the timing path [24]. The Clean-up PLL has the narrowest loop bandwidth. In Fig. 44(b), jitter (wander) transfer is the amount of jitter (wander) that passes through the node from R_CLK to TX_CLK. The R_CLK from the PHY is expected to have a large amount of jitter (wander) in it since it is coming directly from the network. The jitter (wander) transfer gain of multiple PLLs in series is the arithmetic sum of the individual PLL jitter (wander) transfer gains. Since the Clean-up PLL has the narrowest loop bandwidth and attenuates jitter (wander) at a wide range of frequencies, the overall system jitter (wander) transfer gain is dominated by the jitter (wander) transfer gain of the Clean-up PLL. The Clean-up PLL should therefore be selected to meet the exact application requirements.

Jitter (wander) generation is the jitter (wander) created internally by the node itself. Since intrinsic jitter (wander) is always present. The majority of this jitter (wander) is generated by the PLL, power supplies and other components in the individual node. Jitter (wander) tolerance is a measure of the resilience of equipment to input jitter (wander). As the timing signal traverses a network, if the jitter (wander) is amplified as it passes through the network, then it could exceed the jitter (wander) tolerance of subsequent equipment. The designer must ensure that each node limits the amount

The multiplexer in Fig. 44(a) drives a set of clock signals into the Clean-UP PLL that is used to lock the local oscillator to the timing reference. The PLL generates a locked 25 MHz clock signal to drive the Sync-E PHY devices. This synchronizes the data on all of the transmit ports of the device to the timing reference and embeds the locked clock signal into the data stream. If an external clock signal is required to provide a timing reference to the rest of the system (such as chassis line cards), the Clean-up PLL must synthesize the desired frequency.

The timing module could provide additional features such as holdover function, reference monitoring, and reference switching, on top of jitter/wander attenuation it supports. With the holdover function, when a slave clock loses the master clock signal it is expected to go into holdover mode. In this mode it attempts to maintain clock accuracy, although it no longer has access to its reference. Carrier-grade telecom timing modules support redundant PLLs and components to allow for uninterrupted and reliable timing supply [22], [23].

Timing slaves receive synchronized data from the master system and derive PRC timing by performing clock data recovery on upstream ports. In the slave system, the recovered clock signals from the Sync-E PHY feed into a Clean-up PLL to attenuate accumulated jitter/wander. The PLL could also provide clock monitoring and holdover functions to maintain an output reference clock signal in the event that the recovered clock signal is lost or degraded. Using the reference clock signal to drive its transmit interfaces, the timing slave can then act as a timing master for further downstream devices. If the timing slave is a leaf node in the clock distribution
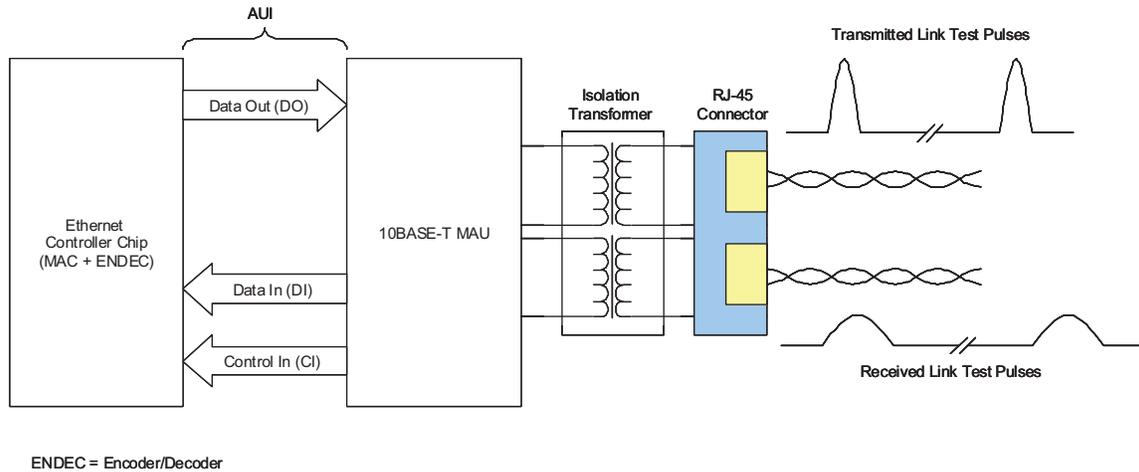
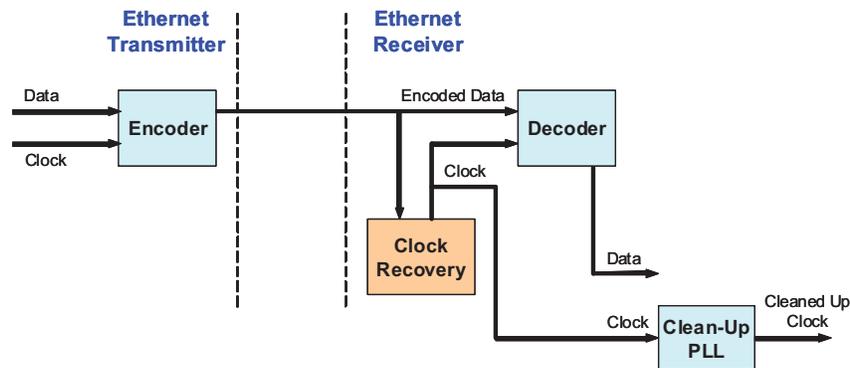Fig. 42. 10 Mb/s Ethernet link test transmission/reception.

Fig. 43. Point-to-point clock synchronization using the higher speed Ethernet PHYs.

of jitter (wander) it generates internally otherwise the jitter (wander) from each node will add to the noise in the network and eventually the output clock from the node will not meet the requirements of the intended application. The PLL should tolerate large jitter and wander at its input and still maintain synchronization without raising any alarms.

B. Sync-E Network Applications

Sync-E can be used in wireless backhaul to connect the wireless base stations to the mobile switching centers. For GSM systems, the timing reference to the base stations is traditional supplied via E1/T1 links which have timing transfer capabilities built into them. However, in current networks, the bandwidth of these links is saturated due to the deployment of new wireless services like UMTS, HSPA, and HSPA+. In addition, next-generation wireless technologies like LTE, Mobile WiMAX and LTE-Advanced have been defined with higher capacity air interfaces which require correspondingly higher capacity packet-based interfaces (e.g., 1 Gigabit Ethernet). Fig. 45 shows the application of Sync-E interfaces with timing support in wireless backhaul. In the example network, the Base Station Controller (BSC) with Sync-E functionalities has access to a high accuracy PRC which it uses to time its transmit Ethernet interfaces. Each intermediate Sync-E switch

in the (Sync-E) network then recovers the PRC-traceable clock at its receive interfaces and subsequently times its transmit interfaces with the recovered clock. The base stations at the ends of the timing path ultimately use the recovered clock timing to derive their carrier radio frequencies.

Ethernet has been a highly understood and widely deployed technology for over a decade. It is now deployed from the desktop to the wide area and has come to compete and is replacing SDH/SONET as the technology of choice in telecom networks. Ethernet interfaces and devices are much less expensive than PDH or SONET/SDH interfaces and devices of the same bandwidth. Ethernet also supports high bandwidths with the emergence of 10, 40 and 100 Gigabit Ethernet technologies. The driving force behind Ethernet as a metro networking technology is the prevalent use of Ethernet in residential, corporate and, more recently, telecom and service provider core networks. With this trend, it therefore, makes sense to bring Ethernet into the metro networking domain as it introduces a lot of advantages to both the service provider and the customer (residential and corporate). The added value of Sync-E in the metro networking is it brings timing readily to timing-dependent interfaces and nodes like TDM multiplexers, PBXs, and wireless base stations which are connected to a packet technology like Ethernet. Fig. 46 illustrates a network node architecture with mixed Sync-E
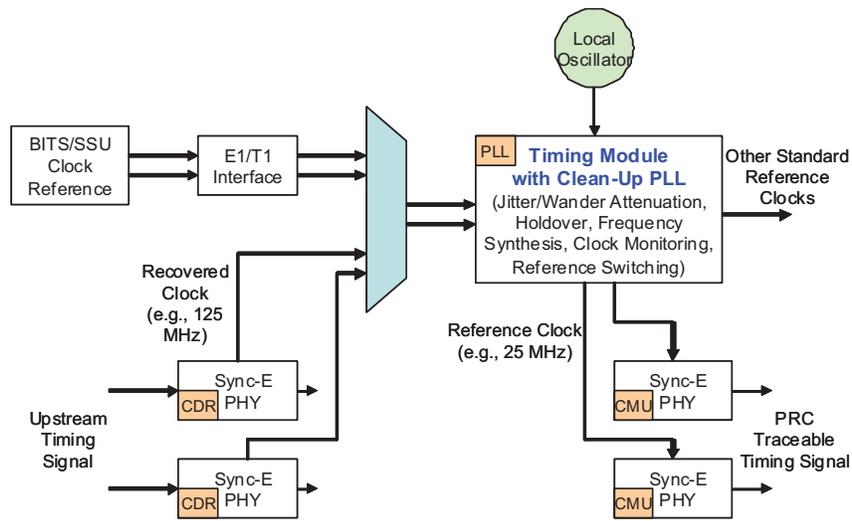
(a) Sync-E Node timing elements
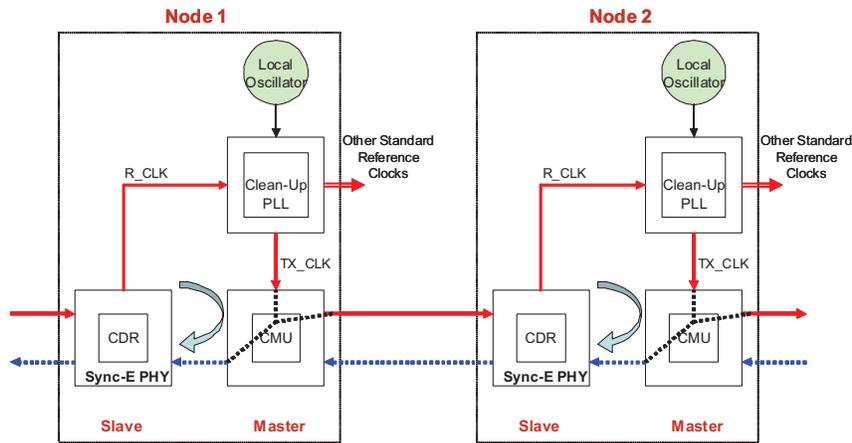
(b) PLLs in timing path
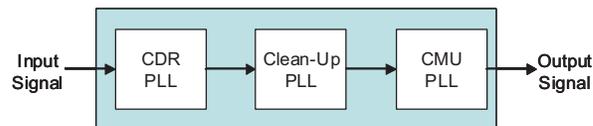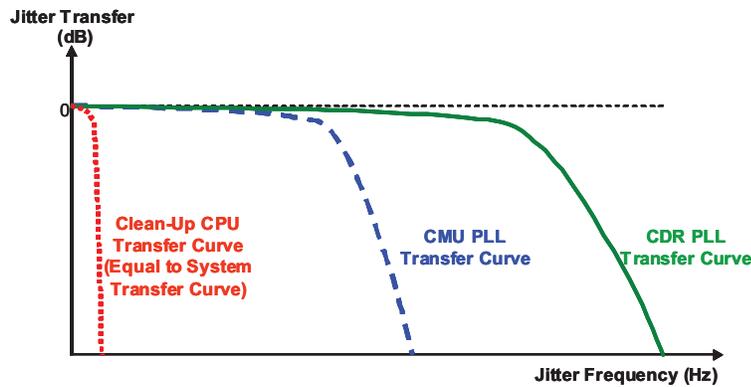
(c) Illustrating desired jitter (wander) transfer of the PLLs in the timing path

Fig. 44. Sync-E Node timing architectures.

and SDH/SONET interfaces as would be found in a Metro Carrier Ethernet networking scenario. There are many more application areas of Sync-E, those listed here are just a few examples to show the reach of Sync-E in networking.
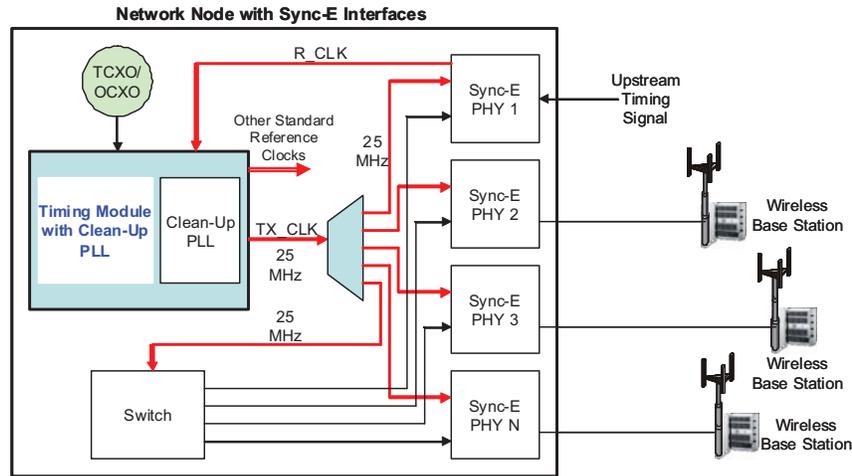
Fig. 45. Network node with Sync-E interfaces applied to wireless backhaul.

VII. CONCLUSION

In this paper we explain that Sync-E passes timing directly via the physical layer interface from node to node similar to timing transfer in SONET/SDH or T1/E1. Sync-E can be implemented easily over the faster versions of Ethernet, (100 Mb/s, 1 Gb/s and 10 Gb/s). Idle periods in Ethernet flavors are continuously filed with Idle symbols thus providing continuous pulse transitions, which allows high-quality clock recovery at the receiver. Sync-E implementation over 1000BASE-X Ethernet is very much like all the other higher speed Ethernet versions. Gigabit Ethernet over copper (1000BASE-T) which uses transmission over 4 copper pairs, however, requires special handling when used for Sync-E implementation. The same applies to 10GBASE-T Ethernet. A proper master and slave arrangement has to be configured between two 1000BASE-T nodes (usually determined during the auto-negotiation process or can be set up manually) for them to participate correctly in timing transfer. 10BASE-T Ethernet cannot be used for timing transfer over the physical layer interface because a 10BASE-T transmitter stops sending data pulses during idle periods – there is no equivalent of Idle symbols as in the faster flavors of Ethernet. Instead, using the link integrity test feature, a 10BASE-T transceiver simply sends a link test pulse (keepalive pulse) every 16ms $\pm$ 8 ms to notify its presence to the receiving end. These link test pulses certainly have poor timing transfer characteristics are not suitable for clock recovery at the receiver. Thus, out of the many Ethernet physical signals standardized by the IEEE only those with continuous transmission can be used in for Sync-E.

Sync-E is based on the well-established SONET/SDH synchronization principles widely adopted in the telecom industry. As a result the architecture of a synchronization network based on Sync-E is the same as with SDH/SONET. Synchronization networks can be built as mix of Sync-E and SDH/SONET elements. Also, the same types of synchronization network elements, i.e. PRCs and SSUs, are used both in Sync-E and in SDH/SONET. Sync-E uses the same clock specifications as SDH/SONET. Sync-E and SDH/SONET share Synchronization Status Message principle and allows interworking.
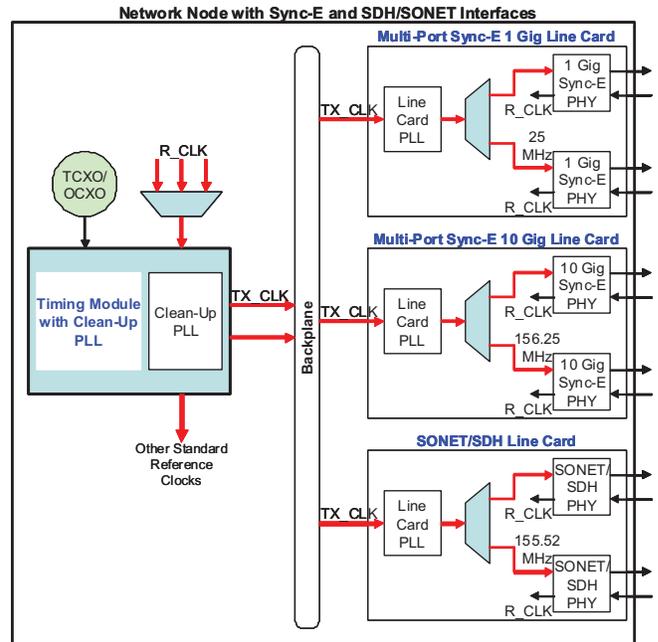

Fig. 46. Network node with mixed Sync-E and SDH/SONET interfaces as could be found in a metro carrier Ethernet application.

REFERENCES

[1] *IEEE Standard for Information Technology-Specific Requirements - Part 3: Carrier Sense Multiple Access with Collision Detection (CMSA/CD) Access Method and Physical Layer Specifications*, IEEE 802.3-2008.

[2] *Draft Standard for Local and Metropolitan Area Networks - Timing and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks*, IEEE P802.1AS/D7.6, Nov. 2010.

[3] ITU-T. Recommendation G.803: "Architecture of transport networks based on the synchronous digital hierarchy (SDH)."

[4] American National Standard for Telecommunications, *Synchronization Network Architecture*, ANSI T1.TR.81.

[5] ITU-T. Recommendation G.823: "The control of jitter and wander within digital networks which are based on the 2048 kbit/s hierarchy."

[6] ITU-T. Recommendation G.824: "The control of jitter and wander within digital networks which are based on the 1544 kbit/s hierarchy."

[7] ITU-T. Recommendation G.825: "The control of jitter and wander within digital networks which are based on the synchronous digital hierarchy (SDH)."

[8] *American National Standard for Telecommunications, Synchronous Optical Network (SONET) – Jitter at Network Interfaces*, ANSI T1.105.03-1994.

[9] Telecordia GR-1244-CORE, "Clocks for the synchronized network: Common generic criteria," Issue 4, Oct. 2009.

[10] ITU-T. Recommendation G.813: "Timing characteristics of SDH equipment slave clocks (SEC)."

[11] ITU-T. Recommendation G.8261: "Timing and synchronization aspects in packet networks."

[12] ITU-T. Recommendation G.8262: "Timing characteristics of synchronous Ethernet equipment slave clock (EEC)."

[13] ITU-T. Recommendation G.8264: "Distribution of timing information through packet networks."

[14] *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, IEEE 1588-2008.

[15] D. Mills, "Network time protocol (version 3) specification, Implementation and analysis," IETF RFC 1305, Mar. 1992.

[16] ITU-T Recommendation G.707/Y.1322: "Network node interface for the synchronous digital hierarchy ([G707])," Oct. 2000.

[17] ITU-T Recommendation G.783: "Characteristics of SDH equipment functional blocks ([G783])," Oct. 2000.

[18] ITU-T Recommendation G.7041/Y1303: "Generic framing procedure ([G7041])," Jan. 2002

[19] ITU-T Recommendation G.7042/Y1305: "LCAS for virtually concatenated signals ([G7042])," Nov. 2001.

[20] ITU-T Recommendation X.86: "Ethernet over LAPS," Feb. 2001.

[21] *American National Standard for Telecommunications, Synchronization Interface Standards for Digital Networks*, ANSI T1.101-1994.

[22] S. Milijevic, "Adding timing redundancy to communication equipment designs," *EE Times*, Dec. 2004.

[23] S. Milijevic, "Timing redundancy in telecommunication systems," Zarlink Semiconductor, White Paper, Jan. 2005.

[24] "Ethernet time synchronization: Providing native timing within the network," Broadcom, White Paper, Oct. 2008.

[25] RMIITM Specification, RMII ConsortiumTM, Mar. 1998.

[26] Serial-MII Specification, Version 2.1, Feb. 2000.

[27] Reduced Gigabit Media Independent Interface (RGMII) Specification, Version 2.0, Apr. 2002.

[28] Reduced Gigabit Media Independent Interface (RGMII) Specification, Version 1.3, Dec. 2000.

[29] Serial-GMII Specification, Version 1.7, July 2001.

[30] J.-L. Ferrant, M. Gilson, S. Jobert, M. Mayer, M. Ouellette, L. Montini, S. Rodrigues, and S. Ruffini "Synchronous Ethernet: A method to transport synchronization," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 126–134, Sep. 2008.

[31] J. Aweya, "Emerging applications of synchronous Ethernet in telecommunication networks," *IEEE Circuits Syst. Mag.*, vol. 12, no. 2012, pp. 56–72.

[32] ITU-T. Recommendation G.810: "Definitions and terminology for synchronization networks."

[33] ITU-T. Recommendation G.811: "Timing requirements at the outputs of primary reference clocks suitable for plesiochronous operation of international digital links."

[34] ITU-T. Recommendation G.812: "Timing requirements at the outputs of slave clocks suitable for plesiochronous operation of international digital links."

[35] S. Bregni, *Synchronization of Digital Telecommunications Networks*. John Wiley & Sons, 2002.

[36] S. Bregni, "A historical perspective on network synchronization," *IEEE Commun. Mag.*, vol. 36, no. 6, June 1998.

[37] J.-L. Ferrant, and S. Ruffini, "Evolution of the standards for packet network synchronization," *IEEE Commun. Mag.*, vol. 49, no. 2, Feb. 2011.

[38] M. Ouellette, Kuiwen Ji, Song Liu, and Han Li, "Using IEEE 1588 and boundary clocks for clock synchronization in telecom networks," *IEEE Commun. Mag.*, pp. 164–171, Feb. 2011.

[39] L. Cosart, "NGN packet network synchronization measurement and analysis," *IEEE Commun. Mag.*, pp. 148–154, Feb. 2011.

[40] K. Shenoi, "Performance aspects of timing in next-generation networks," *IEEE Commun. Mag.*, pp. 156–162, Feb. 2011.

[41] K. Hann, S. Jobert, and S. Rodrigues, "Synchronous Ethernet to transport frequency and phase/time," IEEE Commun. Mag., vol. 50, no. 8, pp. 152–160, Aug. 2012.

[42] ITU-T. Recommendation G.781: "Synchronization layer functions."

[43] S. Bregni and L.Barbieri, "Experimental evaluation of the impact of network frequency synchronization on GSM quality of service during handover," in *Proc. IEEE GLOBECOM*, Dec. 2003, pp. 3457–3461.

James Aweya was a Senior Systems Architect with the global telecom company Nortel, in Ottawa, Canada, from March 1996 to January 2009. He spent 13 years with Nortel working on communication networks, protocols and algorithms, router and switch design, and other telecom and IT equipment design. He received his B.Sc. (Hon) degree in electrical and electronics engineering from the University of Science & Technology, Kumasi, Ghana, an M.Sc in electrical engineering from the University of Saskatchewan, Saskatoon, Canada, and a Ph.D. in computer engineering from the University of Ottawa, Canada. Dr. Aweya has authored approximately 50 international journal papers, 33 conference papers, 43 technical reports, and has been awarded 44 US patents, with a number of patents pending. He was awarded the 2007 Nortel Technology Award of Excellence (TAE), Nortel's highest technology award, for his pioneering and innovative research on "Timing and Synchronization across Packet and TDM Networks." He was also recognized in 2007 as one of Nortel's top 15 innovators. Dr. Aweya is a Senior Member of the IEEE. Currently he is Chief Research Scientist at EBTIC (Etisalat British Telecom Innovation Center) in Abu Dhabi, UAE, working on next-generation mobile wireless backhaul architectures, circuit-to-packet migration strategies, timing and synchronization over packet networks, data center bridging technologies, and other areas of networking of interest to EBTIC stakeholders and partners.