Designing with 10GBase-T transceivers

Rich Pell - May 29, 2012

Three recent and major trends are elevating the popularity of 10Gbps Ethernet: the growing importance of cloud computing, the increasing utilization of unified data/storage connectivity, and software permitting server virtualization by enterprise data centers. All of these factors increase networking traffic and they took a technology that not long ago was considered an exotic connectivity option relegated to high-capacity backhaul and brought it into the mainstream. As was the case with three prior generations of Ethernet, the ubiquity, the ready and familiar management tools, and the compelling cost structure are allowing 10G Ethernet to quickly dominate the computer networking scene.

Crehan Research, a leading industry analyst of data center technologies, estimates that by 2014, 10G Ethernet will overtake 1G Ethernet as the preferred network connectivity option in computer servers. And in one of its most reports on the subject, The Linley Group, another leading industry analyst, predicted robust 10GbE growth and estimated that 10GbE NIC/LAN-on-motherboard (LOM) shipments alone will surpass 16 million ports in 2014.

Several standards-based options exist for 10G Ethernet and span the gamut, from single-mode fiber to twin-ax cable. But of all the options available, 10GBase-T, which is also known as IEEE 802.3an, is arguably the most flexible, economical, backwards compatible, and user friendly 10G Ethernet connectivity option available. It was designed to operate with the familiar unshielded twisted-pair cabling technology, which is already pervasive for 1G Ethernet and can interoperate directly with it.

10GBase-T is capable of covering, with a single cable type, any distance up to 100 meters and thereby reaches 99% of the distance requirements in data centers and enterprise environments. This article will explore the basics of 10GBase-T technology, explain the many benefits it brings to the data center, and outline information the designer needs in order to make a 10GBase-T design successful.

10GBase-T Basics
Ratified in June 2006, IEEE 802.3an, also known as 10GBase-T, provided a stable blueprint for chip manufacturers to develop and introduce compliant and interoperable devices allowing for 10Gbps communications over unshielded twisted-pair cabling. 10GBase-T is the fourth generation of so-called Base-T technologies, which all use RJ45 connectors and unshielded twisted-pair cabling to provide 10Mbps, 100Mbps, 1Gbps, or 10Gbps data transmission while being backward-compatible with prior generations. Because Base-T devices have used an auto-negotiation protocol defined by IEEE to determine the capabilities supported by the other end of the link, this backward compatibility has meant that upgrades could be performed one end at a time, allowing quick and easy incremental improvement of network speed without either changing the wiring or performing forklift upgrades of equipment.

The 10GBase-T transceiver uses full duplex transmission with echo cancellation on each of the four
twisted pairs available in standard Ethernet cables, thereby transmitting an effective 2.5Gbps on each pair. These bits are transformed into a bandwidth reducing line code called 128-DSQ (for *double square*), which limits the analog bandwidth utilization of the 10GBase-T modem to 400MHz.

High-performance line equalization countermands the low-pass filter effects of the transmission channel, and additional digital signal processing (DSP) functions cancel the crosstalk and echo impairments present in the cabling. Additionally, powerful Low-Density Parity Check, or LDPC, forward error correction coding rounds out some of the DSP functions and allows nearly error-free detection at close to fundamental limits in signal-to-noise ratio. Figure 1 shows a block diagram of typical 10GBase-T transceiver.

![Figure 1: A 10GBase-T transceiver includes the major DSP blocks responsible for line equalization, LDPC forward error correction, and analog line code data transformation.](image)

**Benefits of 10GBase-T Technology**

10GBase-T technology was designed to provide a backward-compatible, incremental upgrade for 10Gbps Ethernet, and minimize the disruption that a transition to 10Gbps speed could cause to other interconnect technologies. This benefits of such easy transition shows up in everything from support for 100G and 1000Base-T interconnects to architecturally enabling larger, simplified management domains to the reuse of existing cabling practices.

When compared to other 10Gbps connectivity solutions, one of the most important advantages of 10GBase-T is the ability to communicate and interoperate with legacy, often slower Base-T systems. Most commercially available 10GBase-T transceivers are perfectly capable of reversion to both 1000Base-T (1Gbps) and 100Base-TX (100Mbps) protocols. This way, data centers can “future proof” their switching architectures.

A 10GBase-T switch purchased today can communicate effectively with all legacy 1G and 100M servers while providing the infrastructure to upgrade to 10G switching when commensurate speed servers are introduced. This also means that data center expenditures can grow incrementally; rather than a wholesale conversion of all servers and switches to 10G speeds (which would be required with a non-compatible technology such as SFP+ Direct Attach), 10GBase-T switching systems can convert only those links that truly need upgrades to 10G speeds while maintaining 1G speed on legacy servers that don’t require such data rates.
Unlike direct-attach twin-ax cabling systems, which constrain full performance over distances of up to 7 meters (depending on cable thickness), a network designer can use 10GBase-T to enable cable spans to reach the full 100-meter length permitted by structured cabling rules. This extra reach affords data center designers the flexibility of locating switches away from server racks, opening up the system to architectures beyond configurations that rely on more centralized switching.

Heretofore, lack of economical cabling options for 10G Ethernet beyond a single or adjacent rack has led to the popularity of Top-of-Rack (ToR) architectures, in which a stack of rack-mounted servers are connected with short cables to a fixed configuration switch in close proximity -- typically on top of the server rack. However, such architecture has the drawback of increased management domains with each rack switch being a unique control plane instance that must be managed and updated.

A more centralized switching architecture, also known as End-of-Row (EoR), in which server ports are routed to a larger switch servicing several racks of servers, can have the benefit of a singular entity for management with commensurate reduction in maintenance costs. Also, because larger switches amortize the cost of common elements such as power supplies and cooling fans, the per-port cost of a larger EoR switch may be lower than the equivalent number of ports in a collection of ToR switches.

Designers also benefit from 10GBase-T’s transmission media – the alternative being a hodgepodge of cabling types, lengths and connectors: Cat6 for 1000Base-T, twin-ax with SFP+ connectors for short rungs of 10G, optical modules and multimode fiber for longer runs of 10G. By standardizing on 10GBase-T the data center designer can focus on only one cabling system for all speeds and all distances. Furthermore, that cabling system can be inexpensive Cat6A with familiar, cost-effective and easy-to-use-and-install RJ45 connectors.

**10GBase-T design practices**
The design engineer would be well advised to become familiar with methods and guidelines for creating systems utilizing 10GBase-T transceivers. These include parameters for board layout and routing, power distribution and decoupling, and electromagnetic interference (EMI) reduction. Figure 2 illustrates a typical 10GBase-T switching system depicting the interfaces between the PHY and adjacent components.
Component Placement, Routing
10GBase-T transceivers are mixed-signal (analog and digital) components, and as the analog front-end connections are the most critical aspects of the entire design, the designer must ensure that proper routing techniques are followed. Specifically, the component placement should be optimized to reduce the length of the Media Dependent Interface (MDI) traces from the 10GBase-T transceiver to the integrated connector and magnetics (ICMs).

Note that the ICM component can be substituted by a discrete transformer (magnetic) with a RJ45 connector. In general, the ICM should be placed as close to the transceiver device as the design allows. The MDI traces should be 100Ω differential and single ended impedance should not exceed 60Ω. MDI trace length should be three inches or less while pair-to-pair trace length difference should be 0.25 inch or less.

Clock routing is also important as jitter or other anomalies that are present in the reference clock signal affects the quality of the MDI and Media Access Controller (MAC) interface signals. The length of the oscillator traces should not exceed seven inches. These should be 100Ω differential and single ended impedance should not exceed 60Ω. Positive/negative trace length should be matched to 20mils or less.

The interface between the 10GBase-T transceiver and the MAC or switch it is connected to can be in one of four interface types: KR, XFI, KX-4 or XAUI. KR is a superset of XFI and KX-4 is a superset of XAUI. Similar to Base-T MDI, these are also differential pairs and require care in routing. KR and XFI are single differential pairs for transmit and receive. KX-4 and XAUI are four lanes each for transmit and receive. In more modern 10GBase-T transceivers, the XFI interface pins are shared with KR interface and thus have the capability of driving 40 inches with two connectors, assuming the channel characteristic meet the specifications in IEEE 802.3-2008 annex 69B. For these traces, the length between positive and negative within TX and RX pairs should be matched to 10mils or better.

If the XAUI interface is chosen, the length of the XAUI traces can be 20 inches maximum. Positive/negative within each pair should be matched to 20mils or better. Pair-to-pair matching should be limited to 1.0 inch or better.

The general guidelines designers need to follow in routing the differential signals in MDI, KX-4/XAUI, KR/XFI and oscillator are:

- All differential pairs should be 100Ω +/-10% differential impedance between positive/negative pairs.
- Single-ended impedance should not exceed 60 Ω.
- Keep the positive and negative traces symmetric; pairs should be matched at pads, vias and turns. Any mismatch contributes to an impedance mismatch.
- Avoid bends in the signal as much as possible. Where bends are needed, however, use bevel corners of 135 degrees angle or more.
- Avoid serpentine routing method to match differential pair signals lengths.

Power Distribution, Filtering
Power rails and clock supply voltages must be properly filtered and decoupled to reduce noise. The
filter and decoupling circuit should be designed specific to the particular voltage and component serviced. It is critical to have the proper quantity and position of decoupling caps for each supply.

Designers should assure that analog and digital supply planes not cross each other. In many 10GBase-T transceivers, the part of the chip and its pins connecting to the magnetics and RJ45 is analog, and the part of the chip opposite is digital. Ideally, regulators would be kept a large distance away from the proximity of MDI traces and RJ45 connector. If it is not possible to put all the regulators a large distance away, the regulators providing the digital supplies, at least, should be kept away.

It's also preferred that designers separate the regulator for each supply and per port. However, if supplies are shared, there should be an inductor/ferrite placed in series with each supply from a common voltage point generated from the regulator. The common point should be bypassed to ground for broadband frequencies by inserting multiple capacitors. The supplies going to different domains after the ferrite should also be well coupled to ground through several "in parallel" capacitors. Ferrite beads should be used to isolate analog power rails and keep them from being contaminated. Figure 3 shows the required ferrite bead frequency response.

![Figure 3: To prevent analog power rails from being contaminated, ferrite beads are used.](image)

High-frequency (HF), board-level capacitor placement is critical to the proper function of the power supply into the chip. These capacitors must be able to respond to short duration and HF load changes (compared to the low-frequency or bulk capacitors). As a result, it is necessary to place the HF capacitors as close as possible to the final load to minimize the interconnect inductance. The optimum location, leading to minimum interconnect inductance, is the area directly under the 10GBase-T transceiver device on the backside of the board.

Once all of the available area under the transceiver device is filled with capacitors, the next best location to place additional capacitors is on the topside or backside of the board, as close to the transceiver device as possible without interfering with the routing.

Typical HF capacitors are ceramic multi-layered ceramic, surface-mount chips. Due to their construction, they have a low parasitic inductance of 1nH to 2nH. The most critical factor limiting the effectiveness in minimizing the power and ground noise is the inductance associated with the capacitor connection to the board.
The designer's placement strategy for the HF capacitors should consist of locating the lowest-valued capacitors closest to the transceiver device. Since these components would typically have the smallest body size, they are least likely to interfere with the dense routing in the area. Higher-valued capacitors should be placed as close to the transceiver device as possible. However, their larger body size may force one to have to be placed further away from the transceiver.

**EMI Considerations**

ICMs or discrete magnetic transformers isolate the local circuitry from other equipment that the Ethernet port connects to. The center tap of the isolated winding typically has what the industry calls a “Bob Smith” termination. This termination is a patented network invented by its namesake and provides common-mode termination of all of the wires in the cable. This is especially important for those wires not connected to signals, since they can respond as parasitic elements. The termination uses a 75Ω resistor and 1000pF capacitor to a chassis ground connection. The termination capacitor should have a voltage tolerance of 3kV.

ICMs typically embed these resistors and capacitors inside the ICM housing. If discrete magnetic and RJ45 connectors are used, add the Bob Smith termination components.

The 75Ω resistors and the 3kV capacitor must be placed with adequate separation between each other and other circuit components to prevent arcing (static discharge). The separation will depend on the actual static discharge requirements. Practical testing up to 2KV of potential will need a minimum 0.2-inch separation. The discharge path will be through the 1000pF, 3KV capacitor to the chassis ground connection.

Designers are advised to follow these guidelines for EMI test compliance when using discrete magnetic and RJ45:

- The twisted-pair signals from the RJ45 to the magnetic must be routed with 100Ω (+/- 10%) differential impedance.
- Metal shielding of the RJ45 should be connected to the chassis ground to minimize EMI emission.
- A small plane connecting to chassis ground should be placed under the RJ45. This is to be connected to the RJ45 shield pins and the aforementioned 1000pF, 3KV capacitor. The small plane needs to be only directly underneath and in the immediate vicinity of the RJ45 connector. The remainder of the plane on that layer of the board may be used for other purposes.
- Twisted-pair signals must be routed close together and in parallel to reduce EMI emission and reject common mode noise.
- Ideally, most of the plane right underneath the top layer should be chip ground. The chip ground should be cut directly underneath the MDI traces between the RJ45 and transformer to provide a chassis ground island.
- Coupling between the digital power and ground planes and the chassis ground should be avoided. A gap of 0.2 inch or more should be created between the chassis ground and digital ground.
- Three or more pads between the chassis and digital ground void can be placed for EMI testing. These extra pads will facilitate the addition of components if needed for EMI compliance.

**Summary**

The rising prominence of 10G Ethernet in the data center brings new attention to the 10GBase-T connectivity option, as it is the most flexible, economical, backward-compatible, and user-friendly 10G Ethernet connectivity option available. It’s becoming increasingly clear to designers that 10GBase-T transceiver operation and use offer a range of benefits - the ability to interoperate with legacy slower technologies, the use of ubiquitous and inexpensive cabling and connectors, the
flexibility of full structured wiring reach, and the ease of Cat6A cabling deployment. Further, they can take advantage of 10GBase-T board layout and routing guidelines, power distribution and decoupling requirements, and EMI reduction design concepts to employ best practices in network designs.

About the author
Ron Cates, is vice president of marketing, networking products, PLX Technology, Sunnyvale, Calif. (www.plxtech.com). Prior to PLX, he was senior vice president and general manager of Wide Area Networking Products at Mindspeed Technologies. He has over 30 years of experience in the semiconductor industry and holds BSEE and MSEE degrees from the University of California at Los Angeles and an MBA from San Diego State University. He can be reached at rcates@plxtech.com.