Design Considerations for Next Generation LAN and SAN Gigabit Ethernet Switches

Design considerations for 24-port and higher switches, enabling low cost, pervasive deployment of Gigabit Ethernet over Category 5 cable plants.

DOC00849, V1.0

CONTENT

1 DESIGN CONSIDERATIONS FOR NEXT GENERATION GIGABIT ETHERNET SWITCHES – EXECUTIVE SUMMARY 3

2 A LOOK BACK AT RECENT ETHERNET HISTORY 4

2.1 2001: Ethernet History Will Repeat Itself 5

3 DESIGN CHALLENGES FOR NEXT GENERATION SWITCHES........ 6

3.1 Matching System Configuration Requirements 6

3.1.1 Fixed or Modular Configuration 6
3.1.2 Blocking or Non-blocking Performance........ 6
3.1.3 Configurations for Desktop & Workgroup Switches 7
3.1.4 Configurations for Enterprise Switches 8

3.2 Maximizing Port Density 9

3.2.1 Selecting The MAC-to-PHY Interface 9
3.2.2 Specify Intelligent I/O Pins 10

3.3 Minimizing Overall Systems Cost 11

3.3.1 Target Stacked RJ-45 Modules 11
3.3.2 Look for Advanced Design-for-Testability (DFT) 12
3.3.3 Specify Mainstream CMOS Processes 12

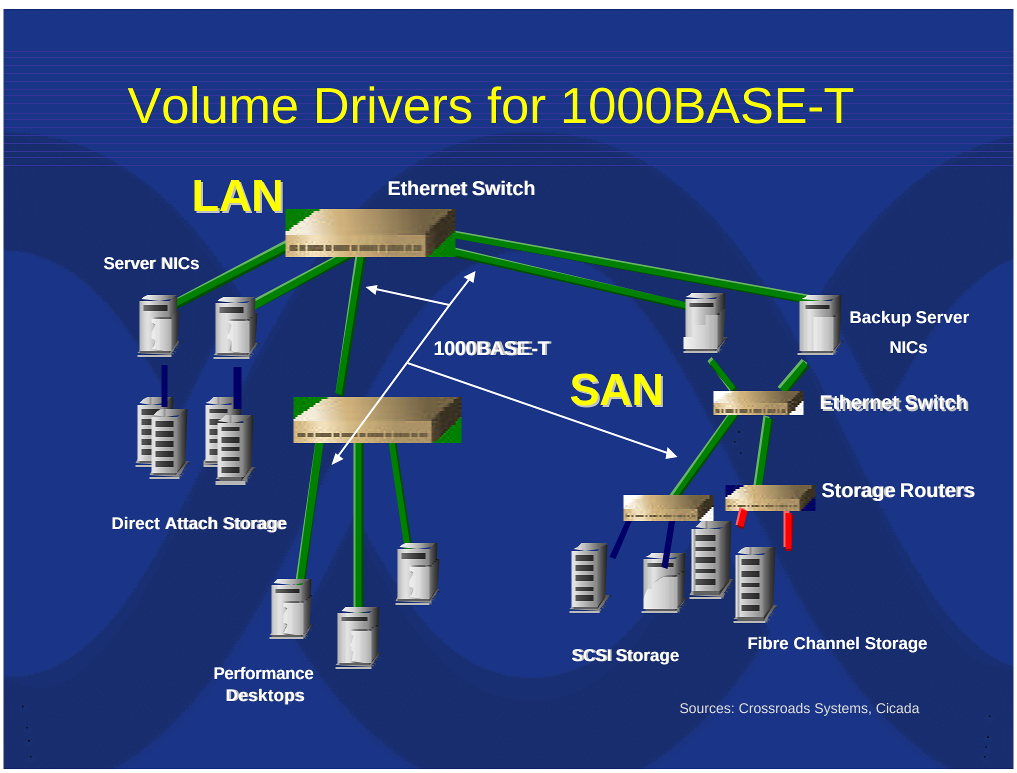
4 CONCLUSIONS 12

1 Design Considerations for Next Generation Gigabit Ethernet Switches – Executive Summary

With the IEEE's formal ratification of the 1000BASE-T standard in June of 1999, the promise of Gigabit Ethernet networks which could take advantage of ubiquitous category 5 copper cabling became a *technical* reality. However, up until the recent availability of cost effective silicon solutions for the switch controller and transceiver, the widespread *commercial* deployment reality has been a very different story.

Although first generation 1000BASE-T Network interface Card (NICs) and low port density[1] switches have been available for almost two years, the power consumption, primary systems component costs and design complexity of these first generation products have been too high compared to their 10/100 Ethernet predecessors. Therefore, unlike what occurred in the market for 10Mbit and 10/100Mbit Ethernet products, these factors have directly delayed widespread deployment of 1000BASE-T by not satisfying the "2X the price for 10X the performance" rule-of-thumb[2].

In 2001, the situation is changing dramatically. OEM designers of high port density 1000BASE-T switches for LANs and SANs will now be able to overcome these design challenges and deliver extremely cost effective solutions to their customers. The availability of a new class of highly integrated Switch Controllers (SwitchCore's CXE-16 and CXE-2000) and low power, multi-port 1000BASE-T Physical Layer (PHY) devices (Cicada's CIS8204 Quad Port PHY) will enable the LAN or SAN switch designer to easily architect and tailor extremely flexible Gigabit Ethernet switch systems for a variety of market segments. For example, single card switches based on these advanced CMOS solutions for 24-port and higher Workgroup LAN switches and Storage Routers will accelerate the rapid deployment of Gigabit Ethernet not only to the desktop, but also throughout the corporate data center and server farm.

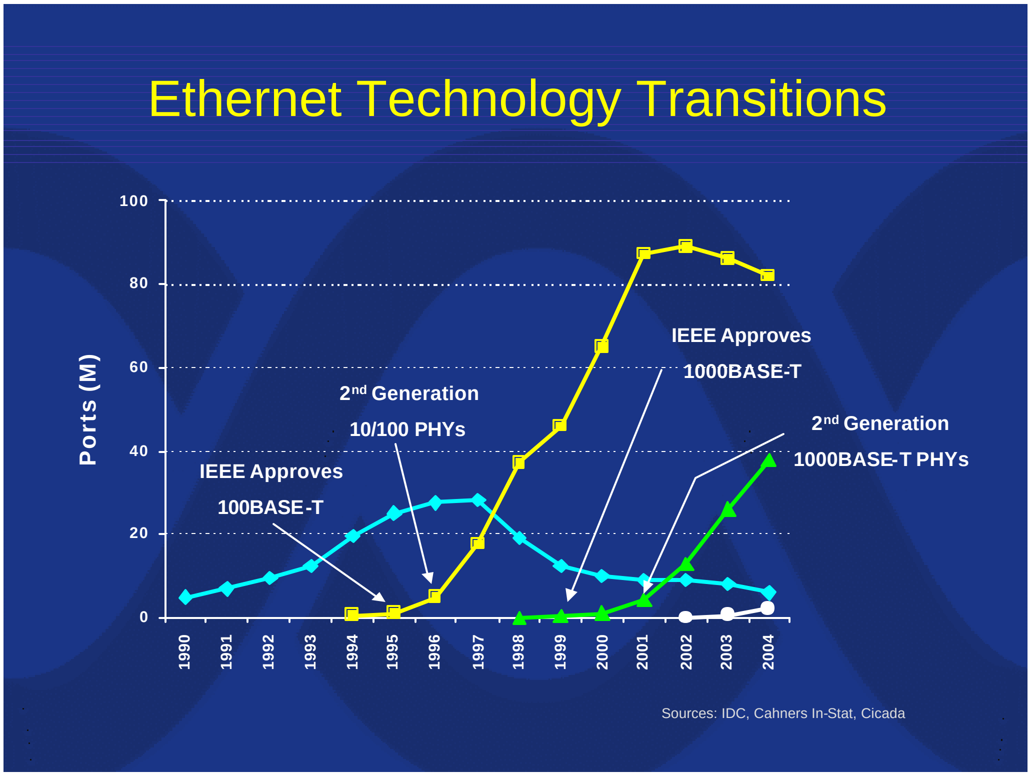


[1] Typically including only one, two or four, 1000BASE-T ports.
[2] Information Technology professionals have often used this metric to justify the upgrade of next generation to Ethernet equipment for the corporate enterprise.

2 A Look Back at Recent Ethernet History

The installed base of Ethernet in the enterprise LAN *alone* now exceeds 300 million ports, the majority of which are presently 10/100Mb ports, as supported by the port shipment graph below. Between 1995 and 2000, overall Ethernet port shipments quickly transitioned from 10Mbit to 100Mbit, and from shared to switched media topologies. The fundamental reason for the displacement of 10Mb ports by 10/100Mb ports was the silicon component cost differential between a "10Mb only" and a 10/100Mb solution was no longer significant. Thus, the well known "2X the price for 10X the performance" rule was easily satisfied.

Beginning in late 1996 and early 1997, one of the key underlying factors which led to this breakneck transition from 10Mb to 100Mb Ethernet switching was the emergence of highly integrated 100BASE-TX PHY and switch controller ICs, key components which together form the heart of the switch system and comprised a large percentage of the system's cost. These new IC solutions offered practical solutions to initial obstacles preventing early availability of affordable, high port density, 100BASE-TX switches: high cost and high power consumption.



Historically, in the particular case of the PHY, some of the first 100BASE-TX PHY ICs arrived in late 1994 and early 1995. Considering the mainstream silicon process technologies available at that time (0.8µm to 0.6µm CMOS, for example), the devices were nevertheless quite sophisticated transceivers; including complex adaptive equalization and baseline wander compensation features. These first generation PHY ICs, however, were priced at almost $50 per port, typically consumed more than two to three watts, and were not available in multi-port variants. These factors did not initially allow 100BASE-TX NIC cards, for example, to compare favorably to cheaper 10BASE-T alternatives until 1997, when increased competition and the use of 0.5µm and 0.35µm CMOS process technologies permitted relatively inexpensive single port 100BASE-TX transceivers to be priced below $10 for the first time. In retrospect, these second and third generation PHYs helped spark the rapid transition to 100BASE-TX switching (as shown in the graph above), since there was no longer any significant system cost difference between a 100BASE-TX switch or NIC, and a 10BASE-T equivalent.

Largely because of the improvements in digital and analog circuit integration of low cost CMOS process technology and increased vendor competition, the new 100BASE-TX PHY IC products introduced in 1997 offered solutions to four critical obstacles facing the 10/100 Ethernet switch designer:

1) Decrease *power consumption*, to enable 100Mbit switches to use similar board designs and components as 10Mbit switches.

2) Increase *port densities* to enable 100Mbit desktop switches to fit into in the same 24 or 48 port (single line card) form factors as 10Mbit switches.

3) Ensure robust *performance* to eliminate any doubt about the viability of 100Mbit Ethernet as a desktop LAN solution.

4) Minimize PHY *costs* to encourage customer transition to higher margin 10/100 capable switches.

As the transition from 10BASE-T to 100BASE-TX accelerated through 1998 and 1999, 100BASE-TX PHYs delivered a 10X reduction in pricing (from $50 to significantly less than $5 per port), *and* more than a 5X reduction improvement in power consumption (from about 3W to lower than 500mW) - in less than 3 years!

A similar story is unfolding for 1000BASE-T silicon for the switch controller and PHY functions.

2.1 2001: Ethernet History Will Repeat Itself

Although the complexity of the analog and digital signal processing challenges required by the 1000BASE-T standard produced many skeptics and only one commercially available PHY product line until approximately mid 2000, the gigabit LAN or SAN switch designer's job in 2001 will become remarkably simpler.

Cost effective gigabit components for 1000BASE-T have already begun to appear in the marketplace, with SwitchCore's 12-port CXE-2000 Workgroup switch-on-a-chip controller, and Cicada's low power, quad port CIS8204 transceivers leading the way.

Compared to fixed configuration, Layer 2, managed 10/100Mb switch port end user pricing of $41 already this year, sub $35[3] per port gigabit component pricing to the system manufacturer can easily support switch manufacturers' cost requirements, satisfying the "2X / 10X" rule for end user switch port pricing and accelerating the transition to widespread Gigabit Ethernet deployment in both the LAN and SAN in 2001. In other words, per port pricing for 1000BASE-T switches should drop dramatically in 2001, to around $80 per port, and then to less than $68 in 2002.

Speed	**Average End User, Per Port Pricing for Layer 2, Fixed, Managed Switches**	
	2001	**2002**
10/100Mb	$41[4]	$34
1000Mb	*$82*[5]	*$68*

[3] Per port "Switch Controller + PHY" pricing to the switch manufacturer
[4] Dell'Oro (April 2001, Gigabit Ethernet Conference 2001)
[5] Projected using the "2X / 10X" rule

3 Design Challenges for Next Generation Switches

From a systems level perspective, let's look at three of the major design challenges facing the LAN or SAN switch designer, along with some of the solutions.

1) Matching system configuration requirements

2) Maximizing port density

3) Reducing system component costs

3.1 Matching System Configuration Requirements

The most basic requirement to consider when designing the next generation switching / routing system is to match the end user's application requirements with the switch system configuration type. In this case, the user's requirements determine whether the switch needs to be modular or fixed, the total number of ports to support, and if the switch must provide non-blocking switching throughout or not.

3.1.1 Fixed or Modular Configuration

Modular configurations or chassis-based systems are most often used where high-port count and system flexibility are important factors, with cost being a secondary consideration. For these reasons, modular configurations are typically centralized in the enterprise network such as in backbones and server farms / data centers.

Fixed configuration systems have been used for distributed switching and they have generally been provided at a considerably lower cost than a modular system. With the emergence of highly integrated switching silicon and PHYs, a stand-alone "pizza box" can now hold the number of ports that is required in backbone applications (typically 24, 48 or more ports), therefore lowering the cost significantly for the enterprise network.

Among other design considerations for the switch engineer is the possibility to use the same components across both fixed and modular configuration product types. Due to the flexible scalability[6] of the CXE product family, systems built on SwitchCore's and Cicada's devices will be similar regardless of whether they are used in line cards for a modular solution or in a single PCB or stand-alone pizza box.

3.1.2 Blocking or Non-blocking Performance

Fixed configuration boxes have traditionally delivered non-blocking or slightly blocking performance. Non-blocking performance implies that in a mesh test (each port receives traffic addressed to one unique port) the switch will not drop packets. In a blocking situation, the blocking rate is calculated as the amount of traffic sent out of a switch divided by the amount of in coming traffic at full load. Generally, a blocking factor higher than 80% is considered as reasonable. Because of the "bursty-ness" of Ethernet traffic, a slight blocking configuration seldom gets oversubscribed.

No matter what the system configuration, non-blocking solutions always increase overall performance, as well as the cost, so a compromise has to be reached based on the application requirements.

[6] Please refer to the document: "Scaling with the CXE Architecture" - SwitchCore Corporation, DOC00332, V1

3.1.3 Configurations for Desktop & Workgroup Switches

As mentioned earlier, a high port density, fixed configuration system typically translates into a lower system cost per port, at the expense of modularity.

For the workgroup or Gigabit-to-the-desktop switching market, a 24-port, Layer 2 Switch can be designed in a blocking configuration with just four, CXE-2000s and six, CIS8204 PHYs (a).

A high performance, non-blocking configuration would require six, CXE-2000 switch chips (b).

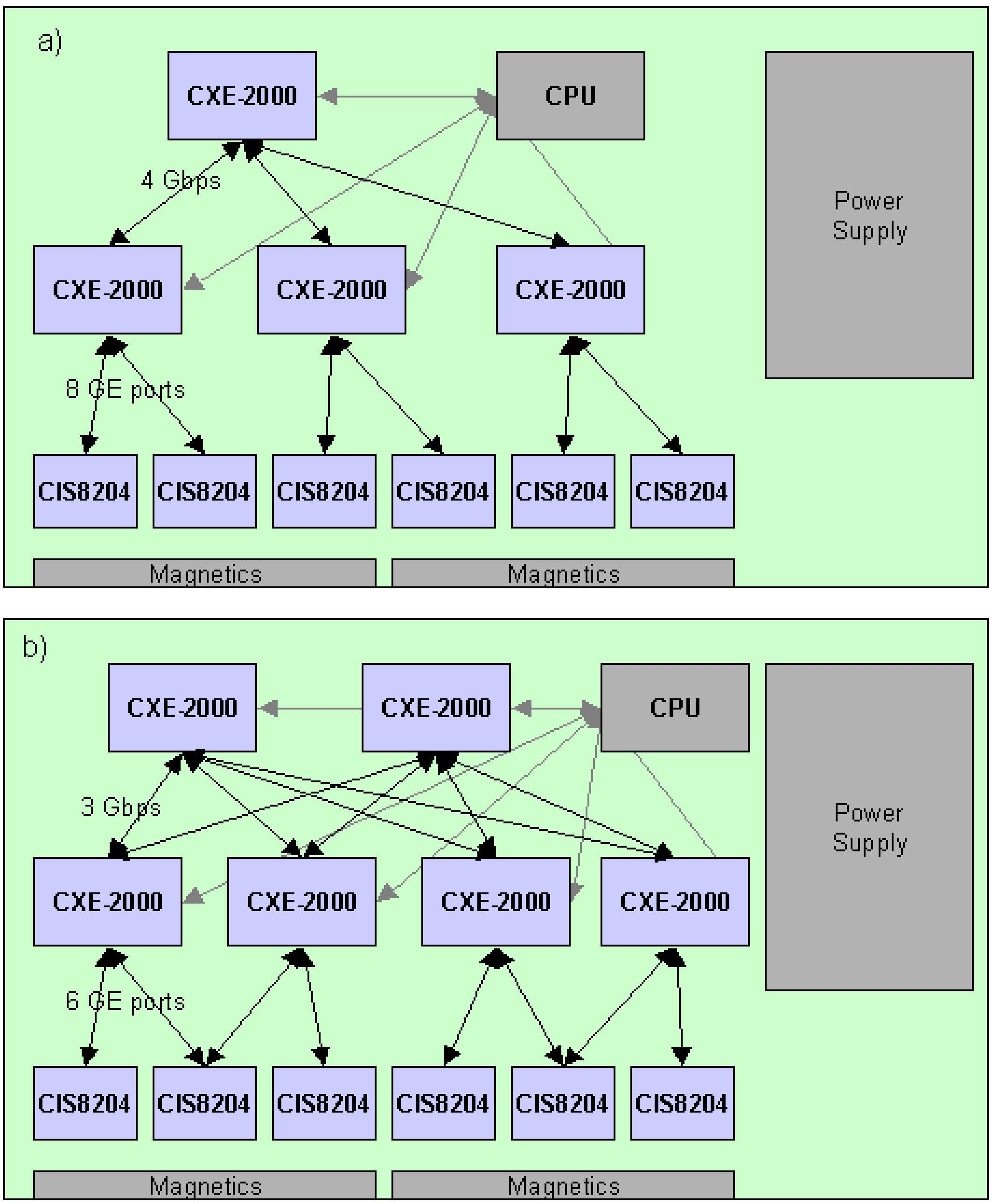


Figure 1 - CXE-2000 Based Configurations for 24-port Workgroup and Desktop Switches

3.1.4 Configurations for Enterprise Switches

For the enterprise switching market, a 32-port Gigabit Ethernet Switch/Router in a 1HU "pizza box" form factor would be considered a dream configuration, but it is far less complicated than it might seem from a controller and PHY component perspective. Such a system (a) can easily be designed with only four, 16-port switch chips (SwitchCore's CXE-16) and eight quad PHYs (Cicada's CIS8204).

For an ultimate performance switch, a non-blocking configuration of six, 16-port switching chips in a mesh topology are required (b). (In the blocking example, the blocking factor is higher for 10 ports than the other 22 ports, making those 10 ports suitable for uplinks or a server farm connection.)

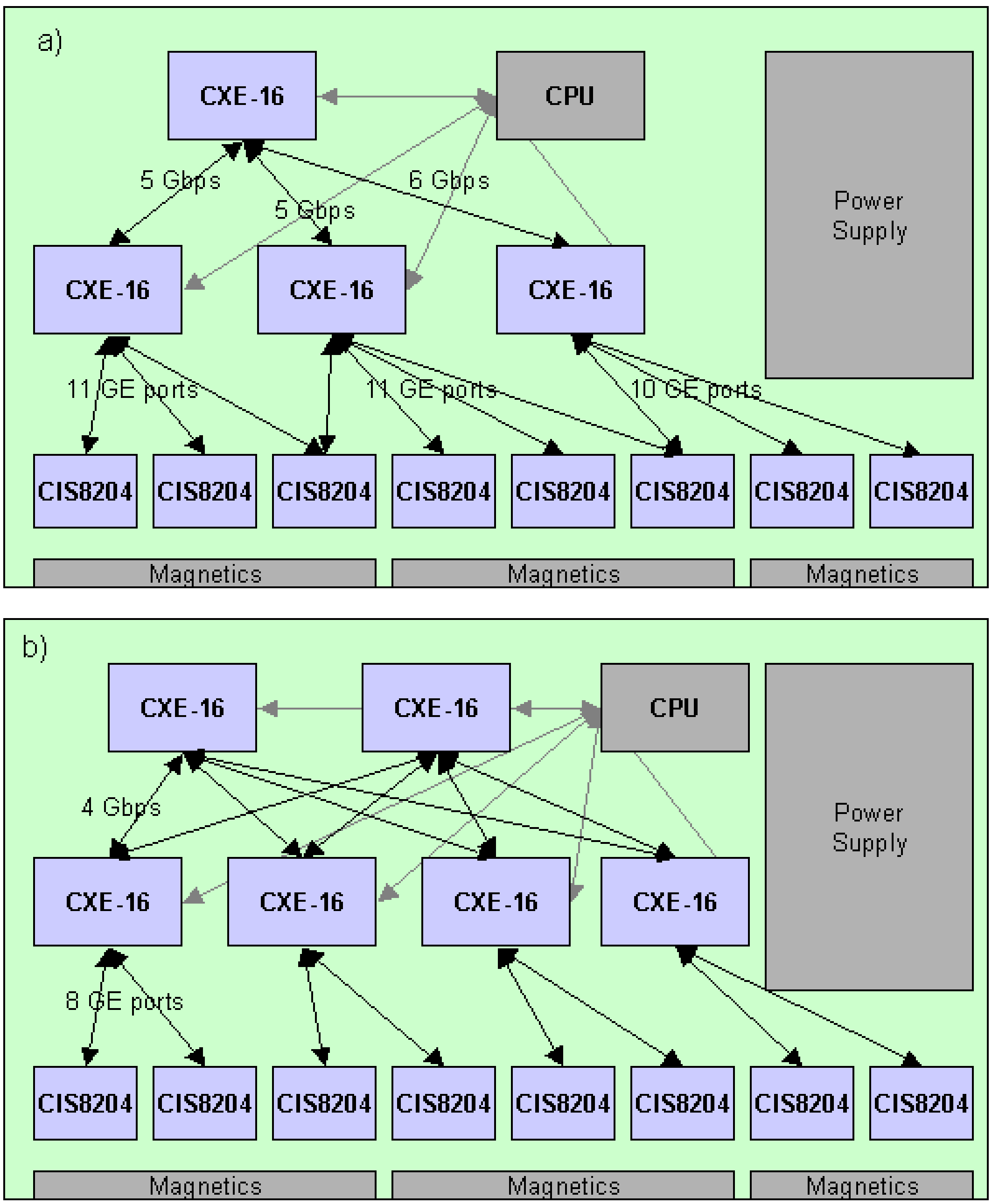


Figure 2 - CXE-16 Based Configurations for 32-port Enterprise Switches

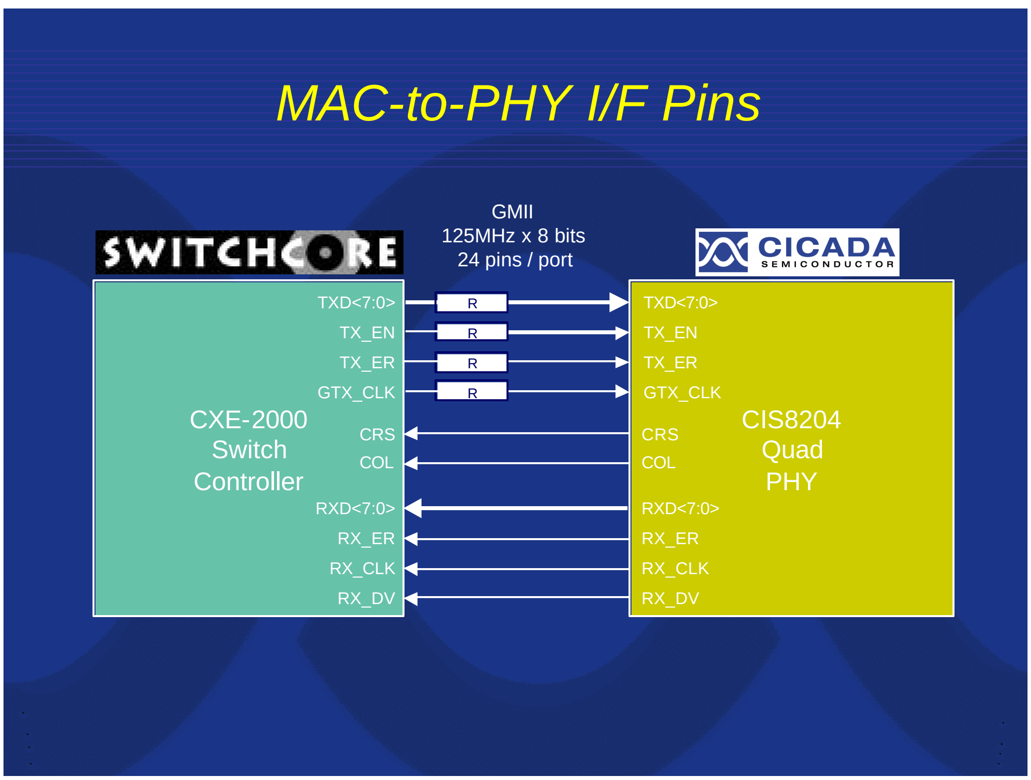
3.2 Maximizing Port Density

High port density card switch designs on a single PCB involves several considerations:

1) Selecting the MAC/PHY interface
2) Minimizing external PCB termination components and timing challenges
3) Targeting high density RJ-45 connectors

3.2.1 Selecting The MAC-to-PHY Interface

The existing IEEE GMII is proven, well behaved from a PCB timing perspective, and is low risk choice for interfacing to proven MACs and controllers. Nonetheless, the switch designer should also be aware of several new system interfaces. These new interfaces have been developed to reduce the signal overhead between the switch controller or MAC ASIC and the PHY, each with their respective merits.



The Reduced Gigabit Media Independent Interface (RGMII), developed by HP, in conjunction with several switch and PHY silicon suppliers, reduces the device pin overhead from 24 (GMII) and 28 pins (TBI) to 12 pins, by multiplexing data and control signals onto common pins and using both edges of the 125MHz reference clock. The pin savings become dramatic for multi-port PHYs and switch controllers: a 24-port switch based upon the RGMII interface eliminates almost 300 pins between the MAC and PHY devices, as compared to the GMII. In addition, fewer pins means there are also fewer series termination components on the board as well.

The Serial Gigabit Media Independent Interface (SGMII), developed by Cisco Systems, reduces the pin overhead even further, to four differential pairs, with a 625MHz double data rate (DDR) clock and a 1.25Gbaud data differential pair for both the transmit and receive data paths.

A close variant of the SGMII interface is also beginning to appear with the clocks embedded in the data streams, making the interface similar to the familiar Serializer-Deserializer (SerDes) interface used to connect with most gigabit fiber optics modules.

The following table summarizes the various MAC-to-PHY system interfaces available to the switch system designer.

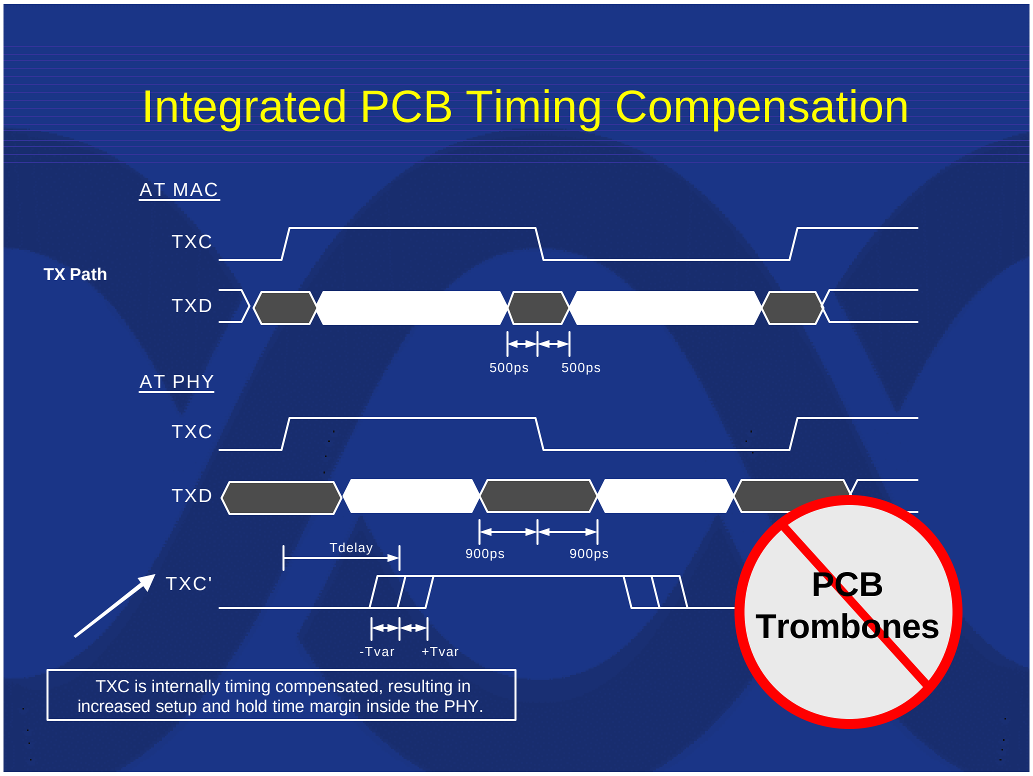
Gigabit MAC-to-PHY Interfaces

Interface	Sponsor	Data Format	Signaling	Key Attributes
GMII	IEEE	125MHz x 8b	Single-ended	IEEE standard
RGMII	HP, Cicada, others	125MHz x 4 bits (DDR)	Single-ended	Saves 12 pins/port, eases signal integrity challenges
SGMII	Cisco	1.25Gbd data 625MHz clock (DDR)	LVDS, Source Synch.	8 Pins (4 diff pairs)
SerDes	Several vendors	1.25Gbd SerDes	LVDS with CDR	4 Pins (2 diff pairs)

3.2.2 Specify Intelligent I/O Pins

To maintain signal and timing fidelity between the MAC and PHY, each output pin is normally source terminated with an external 50? series resistor. This requirement consumes valuable board space and contributes to higher PC board component and assembly costs, not to mention increased board design complexity. Cicada's SimpliPin™ intelligent I/O pin drivers eliminate hundreds of these resistors (>300 for a 24-port switch) on the PC board. More compact PC board routes produce the added benefit of faster time to market, since the effort required for signal integrity engineering is simplified.

Another feature to consider for MAC/PHY I/O pins are drivers that can automatically compensate for sensitive MAC-to-PHY interface timing requirements. For example, the setup and hold time specifications in the current RGMII specification can require the use of serpentine PCB "trombones" to manage the chip-to-chip timing budgets on the PCB. The following diagram depicts one example of the integrated PCB timing compensation capabilities of the CIS8204, which completely eliminates troublesome PCB timing sensitivities, resulting in improved timing margins, increased manufacturability, and lower overall costs of the switch system.



3.3 Minimizing Overall Systems Cost

3.3.1 Target Stacked RJ-45 Modules

To allow 24, 32 and 48 port switches on a single PCB, the PHY component package and pin assignments should support direct placement behind a standard, 14mm-wide RJ-45 connector. For example, quad port PHYs should offer package body sizes less than or equal to two times this width, or 28mm, to allow direct placement behind "stacked" RJ-45 connectors, as shown below. The CIS8204 Quad port PHY features a 27mm wide Plastic BGA which satisfies this form factor, as well as offering optimal pin-out placements to the magnetics and to the RJ-45 connectors, minimizing the number of PCB layers and trace route lengths.

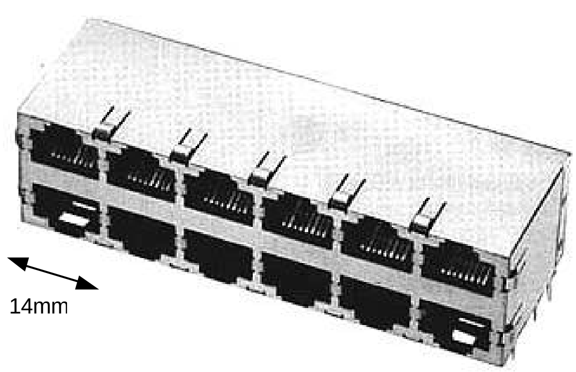


Figure 3 – Example of a 12-port "2 x 6" Stacked RJ-45 Jack Module with Integrated Port Status LEDs

To reduce component count even further, new "stacked RJ-45s" with integrated magnetics and per port status LEDs offer the ultimate level of integration. Transpower Technologies (www.trans-power.com/gigabit.html) is one vendor offering several of these 10/100/1000BASE-T modules.

3.3.2 Look for Advanced Design-for-Testability (DFT)

While minimizing silicon and packaging costs represent two of the obvious ways the switch component supplier is able to deliver low cost products, one of the cost factors systems designers sometimes overlook is what the supplier has included for reducing the increasing cost of IC test.

Thanks to Moore's Law, silicon costs per function continue to drop at an astonishing rate, yet IC test costs are becoming a growing component of an IC's manufacturing costs. To help combat the cost of test, some of the most effective DFT techniques involve Built in Self Test (BIST), scan, and loop-back testing, which can be used for efficiently testing both the switch controller and the PHY during both IC manufacturing and during system level testing.

For example, with essentially no cost penalty to the switch designer, the sophistication of the DSP processing engine in the Cicada CIS8204 1000BASE-T PHY can be leveraged, in conjunction with the CXE-2000 at the systems level, to provide advanced functions such as analyzing cable plant quality and performing in-system Bit Error Rate (BER) analysis. These benefits produce more manageable systems with lower maintenance and support costs, easily supported via the switch's Station Management CPU.

3.3.3 Specify Mainstream CMOS Processes

The switch controller and PHY components should be manufactured in an advanced CMOS process, offering the highest integration capability, lowest cost and the widest availability. However, in order to take advantage of the low cost structure of mainstream digital CMOS process technology, 1000BASE-T PHYs should not require the use of special "analog" process variants. These process variants are not only more expensive, but also take longer to migrate from process to process, slowing the availability of higher port density PHYs.

4 Conclusions

Throughout 2001, we will see the availability of a new breed of low cost, low power, multi-port Gigabit-over-Copper PHY and switch-on-a-chip IC solutions.

Solutions from SwitchCore and Cicada will allow design engineers to simplify their switch design challenges and to develop 1000BASE-T products to enable:

- Switch port densities comparable to 10/100Mb products (24 ports and higher on a single PCB)
- Unmatched deployment simplicity across all real world cable installations
- Per port, end-user prices supporting the "2X-the-Price-for-10X-the-Performance" rule

New, low cost 1000BASE-T products based on SwitchCore's and Cicada's IC products will include cost effective workgroup, desktop and enterprise switches in both fixed and modular configurations for corporate data centers, server farms and in new IP-based SANs.

SwitchCore's 12-port, Layer 2 Workgroup Switch-on-a-Chip, in combination with Cicada's CIS8204 quad port 1000BASE-T PHY, will together serve as the catalysts to the transition from 10/100BASE-TX to 1000BASE-T Ethernet connectivity in LANs and SANs.

Cicada Semiconductor Corporation
811 Barton Springs Rd. Suite 550
Austin, TX 78704
USA
Phone: 512 327 3500
Fax: 512 327 3550

info@cicada-semi.com
www.cicada-semi.com

SwitchCore Corporation
675 N. First Street . Suite PH3
San Jose . CA95112 . USA
Phone: 408 995 3850
Fax: 408 995 3859

www.switchcore.com
info@switchcore.com