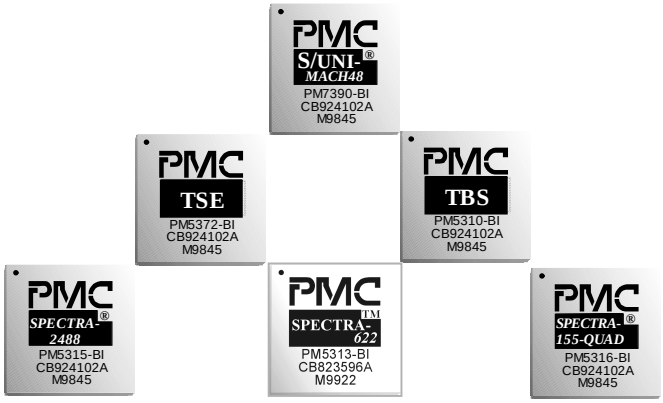
CHESS



TECHNICAL OVERVIEW

PROPRIETARY AND CONFIDENTIAL

PRELIMINARY

ISSUE 2: MAY 2000

PUBLIC REVISION HISTORY

Issue No.	Issue Date	Details of Change
1	Feb. 2000	Document created.
2	May 2000	Update document (power, thermal information, synchronization methodology). Added TSI mappings for CHESS devices. .

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This document introduces the reader to the overall system considerations for implementing components of the CHESS chipset. It also provides design architectures available when using the CHESS chipset. This document is a supplement to the individual CHESS Set device datasheets. Due to the vast number of configurations available with the CHESS chipset, this document may not cover all possible configurations. Please contact a PMC-Sierra Field Applications or Applications Engineer for specific uses not covered in this document.

Although every effort has been taken to ensure that this document is consistent with the datasheet, some errors may occur. Where there are discrepancies with this document, the datasheet and datasheet errata (if any) will take precedence.

1.1 Target Audience

This document has been prepared for Design Architects that are implementing the CHESS set devices and require a quick reference on how to implement the devices in a design. It is assumed the reader is familiar with SONET/SDH and other common telecom/datacom terminology.

1.2 Numbering Conventions

Binary	0001b, 1110b
Decimal	198, 234, 2
Hexadecimal	2H, 200H

2 REGISTER DESCRIPTION

The CHESS Set Device Registers have the following characteristics:

- All values written into unused register bits should be written with logic 0 unless otherwise stated. This action will ensure software compatibility with future versions of the product. Reading back unused bits can produce either logic one or logic zero; hence, unused register bits should be masked off by software during a register read access.

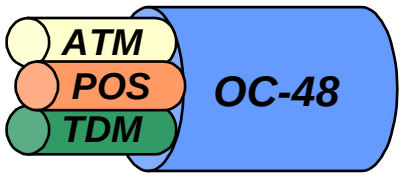
- Certain register bits are reserved. To ensure that the CHESS Set Devices operate as intended reserved register bits must only be written with their default values unless otherwise stated in the datasheet.

The CHESS Set Devices have 2 types of register spaces -- the "normal" registers, which are accessed directly by the microprocessor bus interface of the CHESS Set Device, and the "Indirect" registers which are internal to the CHESS Set Device. Indirect registers are accessed through multiple writes to normal registers and polling of a BUSY bit, ensuring data is visible at the microprocessor bus or written into a normal register stored internal to the CHESS Set Device.

3 THE CHESS ARCHITECTURE

The CHESS (**CH**annelizer **E**ngine for **S**ONET/**S**DH) chipset enables the salient benefits of SONET/SDH to be leveraged in new multi-service switches and routers. Specifically, multi-service switches and routers can be architected to implement SONET ring and protection functions including UPSR, BLSR, or 1+1 protection. The chipset is designed to process channelized (down to a STS-1/STM-0/AU3 granularity) SONET/SDH traffic carrying a variety of services including ATM, POS and TDM, as shown in Figure 1. The CHESS architecture enables like traffic to be groomed to the appropriate processing cards. That is, ATM traffic from the STS-48 data stream can be groomed down to an STS-1 granularity to the ATM processing card and POS traffic to a packet-processing card. Not only is grooming available with the CHESS set, but other features inherent to CHESS make it easy to implement ADM's, provide line and equipment card protection and cross-connect switch fabrics, and even channelize to a further granularity (sub DS3) once the initial grooming is performed.

Figure 1: Representative SONET/SDH data stream of the CHESS Devices



The CHESS set is composed of modular components that allow the customer to pick and choose the system architecture. They are not limited to a single type of design but rather support a variety of flexible Layer 1/ Layer 2 architectures. Some examples of possible system architecture are presented in the coming sections.

3.1 The CHESS Set Devices

The CHESS set consists of six different devices:

3.1.1 SPECTRA-2488

The SPECTRA-2488 is an STS-48 (2.488 Gbps) or 4 x STS-12 (622.08 Mbps) channelized SONET/SDH framer. The chip provides section, line, and optional path termination. The devices provides 48 pointer processors for multiplexing/demultiplexing traffic down to STS-1 granularities and includes a Time Slot Interchange for switching STS-1 level to one of 48 slots. The device can seamlessly interconnect to the TBS for back-plane serialization or protection purposes. Alternatively, the SPECTRA-2488 can be mated directly to any device using the TeleCombus, including the S/UNI-MACH48.

3.1.2 SPECTRA-622

The SPECTRA-622 is an STS-12 (622.08 Mbps) channelized SONET/SDH framer with section, line and optional path termination. The chip is very similar in functionality to the SPECTRA-2488, and includes STS-12 clock/data recovery as well as 12 x DS3 mappers.

3.1.3 SPECTRA-4x155

The SPECTRA-4x155 is a 4xSTS-3 (155 Mbps) channelized SONET/SDH framer with section, line and path termination. The chip is very similar in functionality to the SPECTRA-2488, and includes 4 STS-3 clock/data recovery units.

3.1.4 TBS

The TBS (TeleCombus Serializer) serializes and deserializes each of four 77.76 MHz TeleCombus data-paths to an 8B/10B encoded STS-12 serial link (777.6 MHz), thereby supporting an STS-48 bandwidth equivalent of traffic. There are twelve 777.6 MHz LVDS full duplex ports available (3 sets of 4 links called working, protect and auxiliary). The TBS provides the ability to drive 1.6 metres of back-plane and board trace to enable line, equipment or ring protection and has its own STS-1 Time Slot Interchange (TSI) for grooming traffic. The TSI provides the ability to groom traffic at an STS-1 level to different devices or boards. Most users will use one of the three ports as the main port for grooming to the TSE switch fabric or connecting to a layer 2 processing card. The second port may be used for connecting to a redundant fabric or layer 2 processing card. The third port might be used to pass non-drop traffic to another TBS/SPECTRA-2488 pair. This specific example is shown later in Figure 5 for a Metro/Access Ring ADM.

3.1.5 TSE

The TSE (Transmission Switch Element) is a scaleable 40 Gbps non-blocking STS-1 cross-connect fabric. A single stage fabric can be built to 160 Gbps using only 4 TSE devices. A three-stage fabric can be built to provide fabrics of 320 Gbps to 10.24 Tbps, using 24 to 768 TSE devices respectively.

3.1.6 S/UNI-MACH48

The S/UNI-MACH48 provides ATM mapping and packet mapping for both SONET/SDH and DS3 frames for channelized STS-48/STM-16 streams. The device contains:

- 48 DS3 framers for ATM and packet traffic
- 48 PLCP framers (option for mapping ATM cells into DS3 frames)
- 48 DS3 bit HDLC mappers for packet over DS3
- 48 ATM cell delineation blocks (ATM)

- 48 byte SONET/SDH HDLC processing blocks for Packet Over SONET/SDH

Any mix of the above traffic is workable, with the constraints that traffic must be DS3, STS-1, STS-3c, STS-12c, and STS-48c and must total no more than an STS-48 worth of bandwidth. For example, the S/UNI-MACH48 can simultaneously support 12 x DS3 packet, 12 x DS3 ATM, 4 x STS-3c ATM, 1 x twelve STS-12c POS. The S/UNI-MACH48 has both a standard parallel TeleCombus to mate to SPECTRA devices and a 1+1 serial LVDS TeleCombus interfaces to mate directly to the TSE without the need of serialization by the TBS.

3.1.7 S/UNI-ATLAS 3200

Although the S/UNI-ATLAS-3200 is not a member of the CHESS set, it does interface to the S/UNI-MACH48. The S/UNI-ATLAS-3200 is a monolithic single chip device which handles ATM Layer functions for one direction including VPI/VCI address translation, cell appending, cell rate policing, per-connection counting and I.610 compliant OAM requirements for 64K VCs (virtual connections). It supports an instantaneous transfer rate of 3200Mbit/s equivalent to a cell transfer rate of 5.68×10^6 cells/s (one STS-48c or four STS-12c). Two or more S/UNI-ATLAS-3200 devices can be cascaded to support additional VC's.

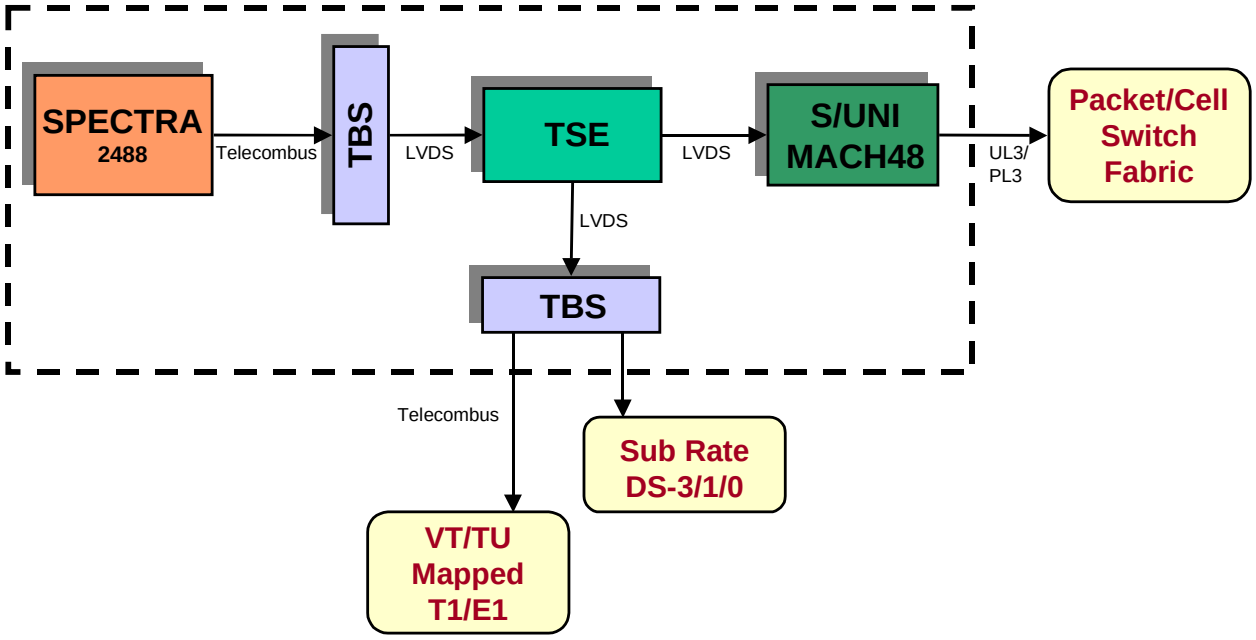
3.2 CHESS SET Example Architectures

There are many possible ways to interconnect this chipset. The following section illustrates some common examples.

3.2.1 The Basic Architecture

The example below demonstrates the power of the CHESS chipset to implement Any Service, at Any Rate on Any Channel. While the following example in Figure 2 does not give exact chip placement, it does demonstrate how a system can be partitioned.

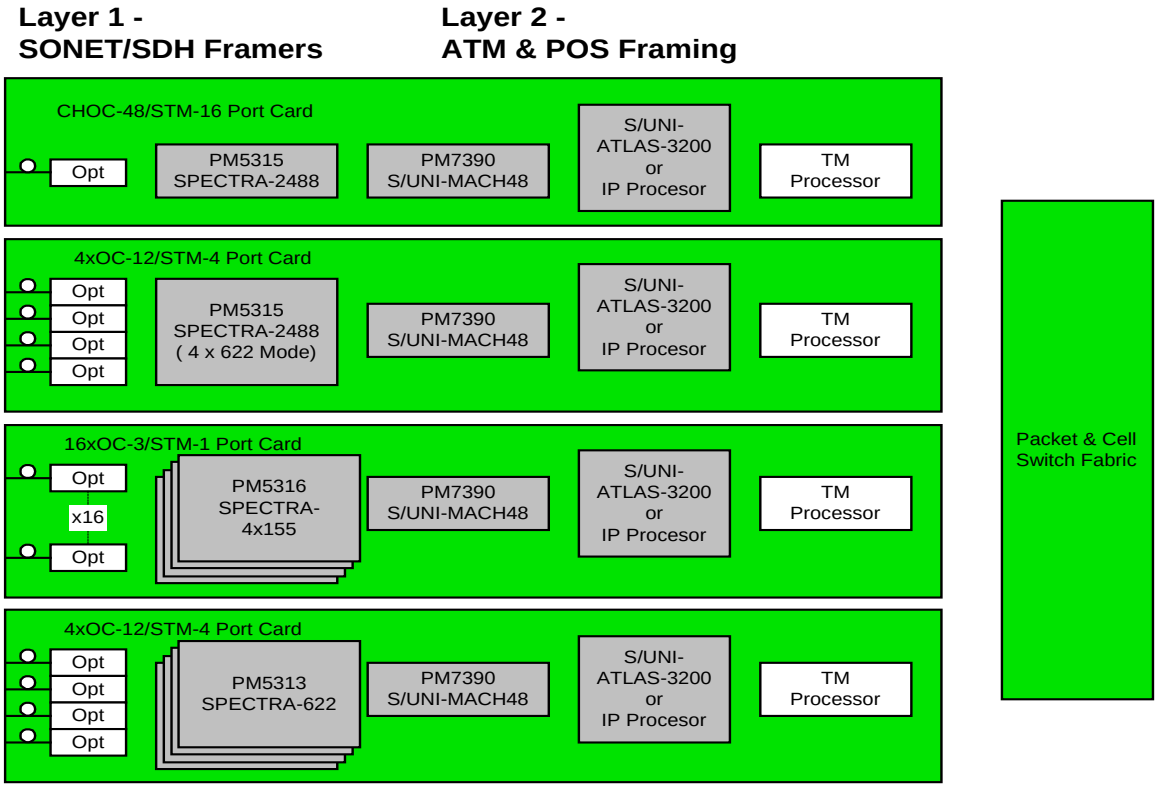
Figure 2: Any Service (at any Rate) on any Channel



3.2.2 One Port - One Service

The following example (Figure 3) illustrates a system in which each port card has its own service. There is no grooming of traffic between cards. One card may be strictly ATM, another TDM, and yet another POS, etc.

Figure 3: One Port One Service (OC-N to DS3)



3.2.3 Complete Cross-Connect Switch

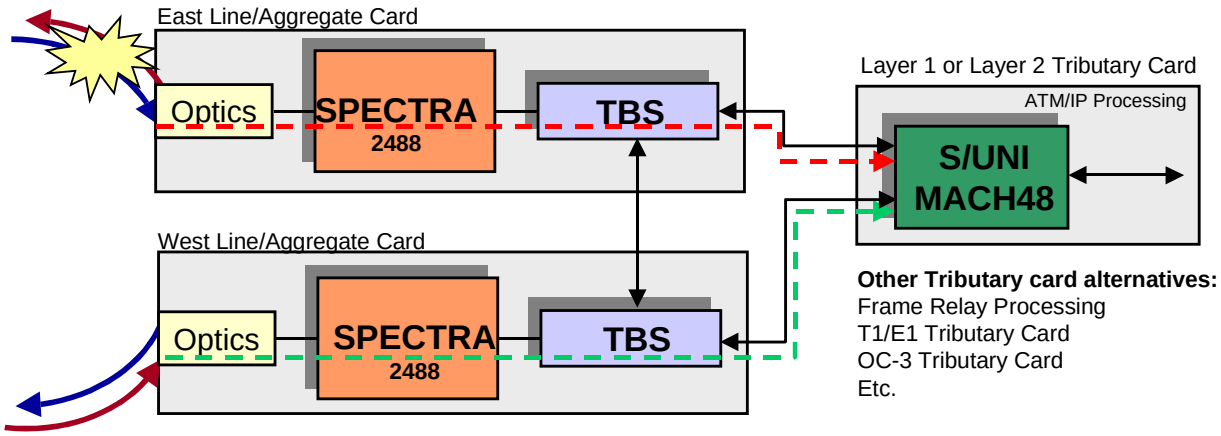
Below is an example (Figure 4) utilizing the TSE fabric to provide switching of STS-1s from any port card to any processing card. In this fashion, any type of traffic can be groomed to its correct processing card regardless of the source. The Layer 2 rates supported include DS-1 through to STS-48. The TSE fabric can be scaled from 40 Gbps – 160 Gbps in a single stage design. 4 TSE devices are used in the 160 Gbps fabric. Note that the TBS devices are used to drive the backplane in this system and can participate in switching (TSI).

Figure 4: Any Port Any Service (OC-N to DS-1)

3.2.4 Building a Small ADM

The following example (Figure 5) demonstrates an implementation of a small North-South (STS-48) Add/Drop Multiplexor. Aside from the serializing and a backplane driving functions, the TBS can provide STS-1 switching and with use of its auxiliary port, drop/ add traffic to the S/UNI-MACH48. For example, in a North-South ADM one SPECTRA 2488 device receives from (for example) the counter-clockwise ring and transmits on the clockwise, while the other SPECTRA receives and transmits the opposite. In this case, pass-through traffic can be passed from one card to the other via the TBS so that it can continue around the ring. In an East-West ADM, one SPECTRA-2488 would both receive and transmit on the same ring. This would require that traffic be passed form the Drop bus to Add bus. This approach is less popular.

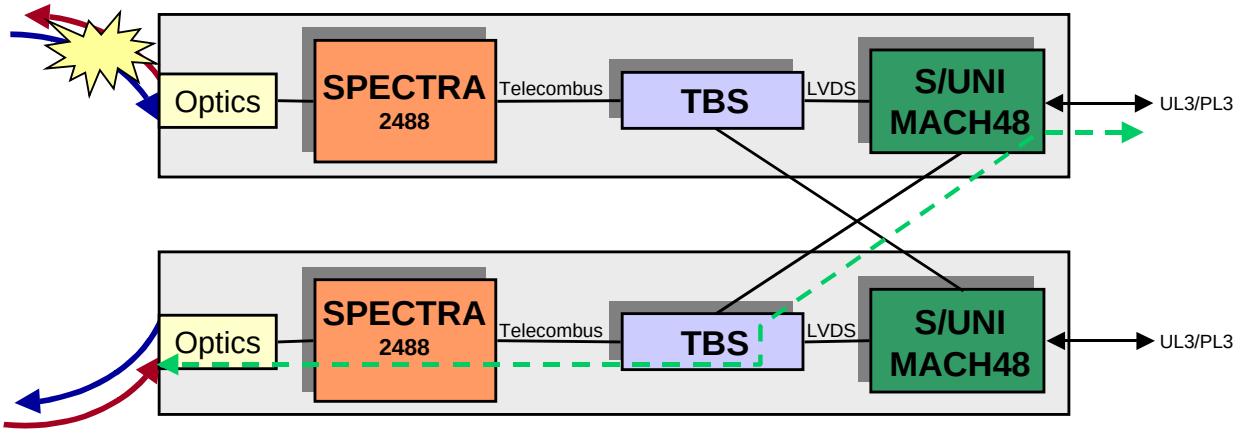
Figure 5: Metro/Access Ring ADM



3.2.5 1+1 Protection on Line Cards

The CHESS chipset provides the ability to perform 1 + 1 line card protection by using the TBS as shown in Figure 6. The secondary (protection) and auxiliary LVDS ports of the TBS give the flexibility to perform many types of protection and redundancy configurations. In this example, the LVDS links would drive the system back plane. Each set of LVDS ports can handle a full 2.488 Gbps, and can be configured independently. In the event that the optics of card 1 failed then traffic can be routed from the S/UNI-MACH48 device of card 1 through TBS, SPECTRA-2488 and optics of card 2. This is illustrated more clearly by the dashed line below.

Figure 6: 1+1 Protection Switching



3.2.6 Building Redundant Switch Fabrics

Again the TBS chips give the ability to create a completely redundant switch fabric without the use of muxes or other external circuitry. Figure 7a illustrates a very simple redundant fabric, but in the same manner, much larger redundant fabrics can be built. This architecture is ideal to implement a switch fabric for slot limited chassis or to leverage on an existing line interface card (SPECTRA-2488, TBS and S/UNI-MACH-48 card) to build a switch fabric. This configuration can be used if the design is slot limited and separate line and port cards cannot be accommodated in the frame. The one limitation is that the top SPECTRA-2488 cannot groom traffic to the bottom S/UNI MACH-48. Figure 7b illustrates a more popular approach that decouples the Layer 1 line card from the Layer 2 service card to provide for full grooming.

Figure 7a: Implementing Redundant Switch Fabrics

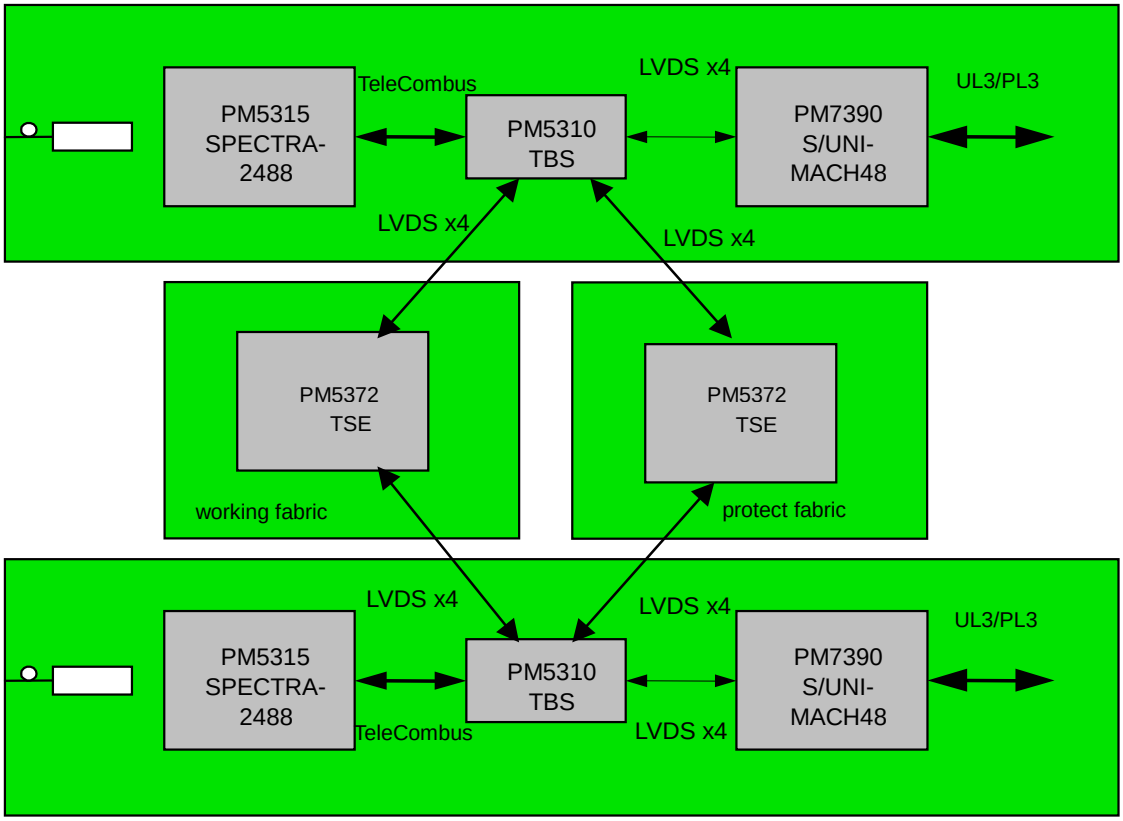
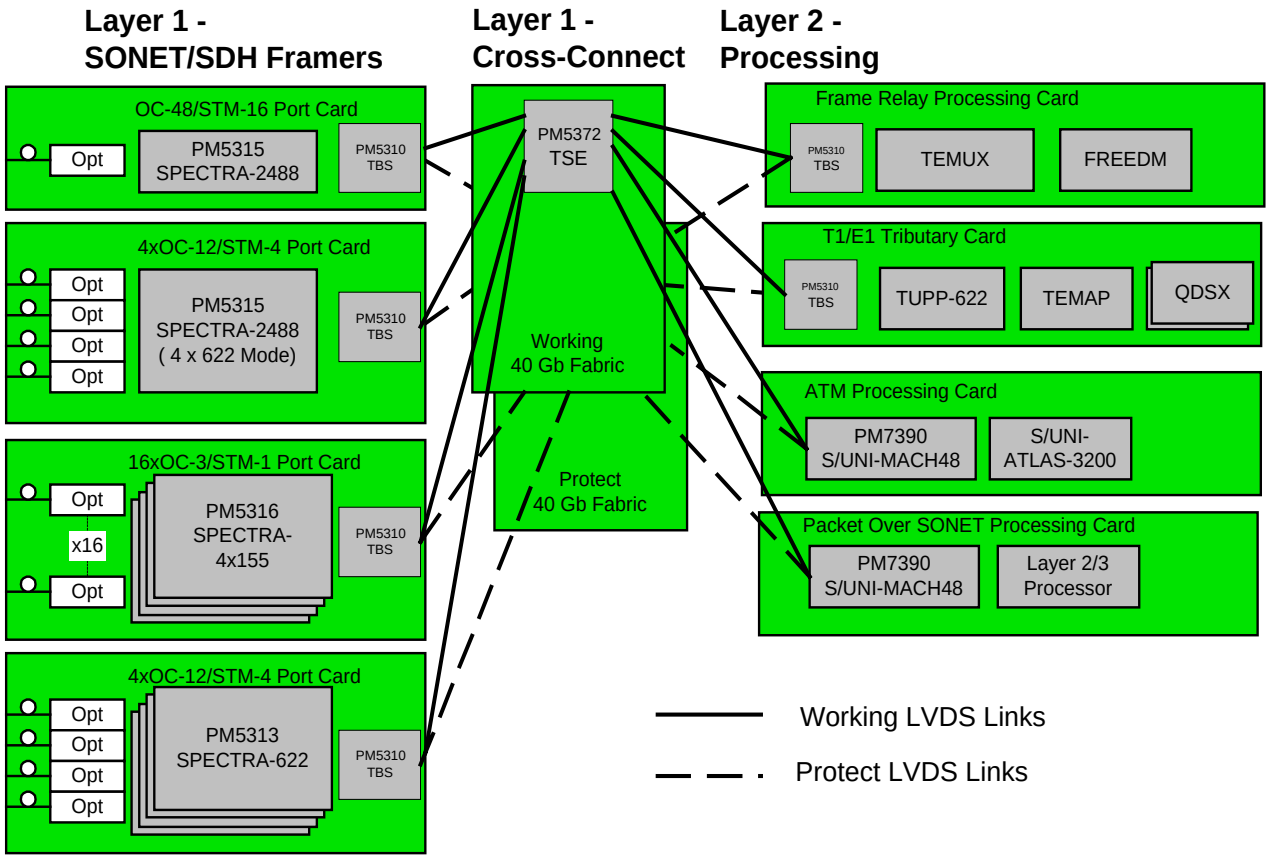


Figure 7b: Redundant Switch Fabrics with Decoupled Line & Service Cards



3.2.7 Implementing DS-1/0 Cards with Protection

The first system (Figure 8) illustrates an implementation for an OC-48 to DS1/0 card. Each TBS feeds one STS-12 link to a service card. Twelve TEMUX devices can terminate an STS-12 worth of traffic. Four such cards can terminate the entire traffic load (STS-48). Note that in this example "a" can have a value of 0 to 3 (up to 4 cards). Service card protection is available by using the protection links. This is better illustrated in the next example. The second example system illustrates an OC-12 to DS1/0 system with equipment protection (Figure 9). Each interface card uses only one LVDS link to interface to the service card providing an STS-12 worth of traffic. A spare link is connected to the protection service card. The protection card TBS has 12 links (4 working, 4 protect and 4 auxiliary), and can thus protect up to 12 such cards. Therefore n can have a value of 12 (12 working cards for one protection card).

Figure 8: OC-48 to DS-1/0 on multiple cards

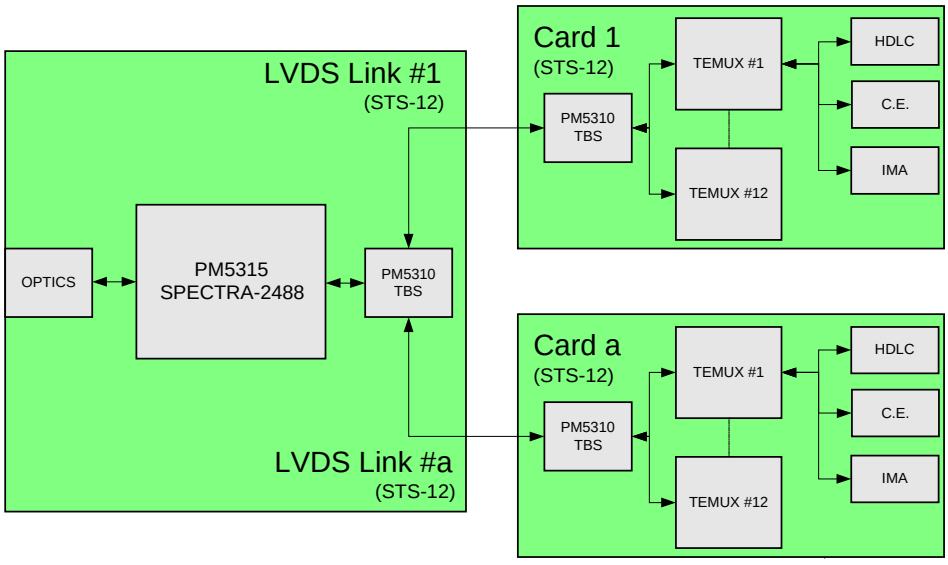
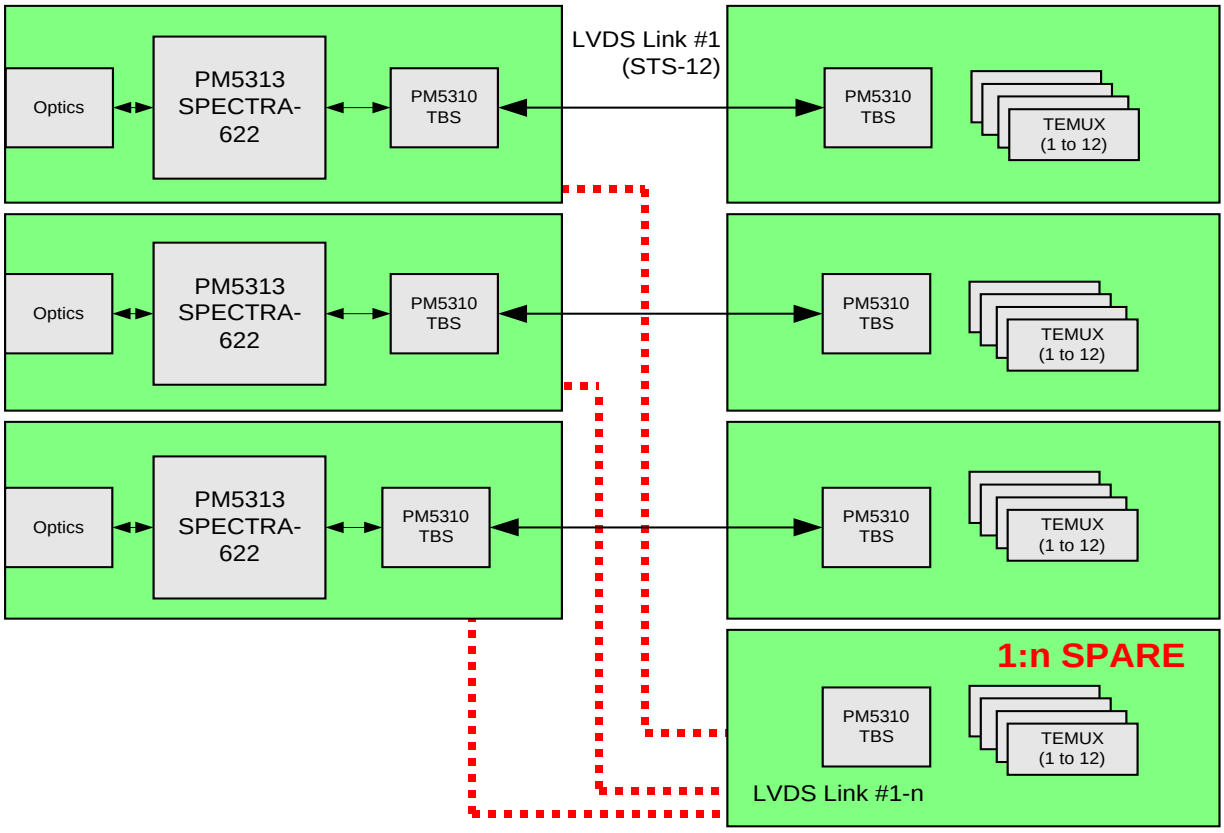


Figure 9: OC-12 to DS-1/0 with 1:n protections



4 FABRIC ARCHITECTURES

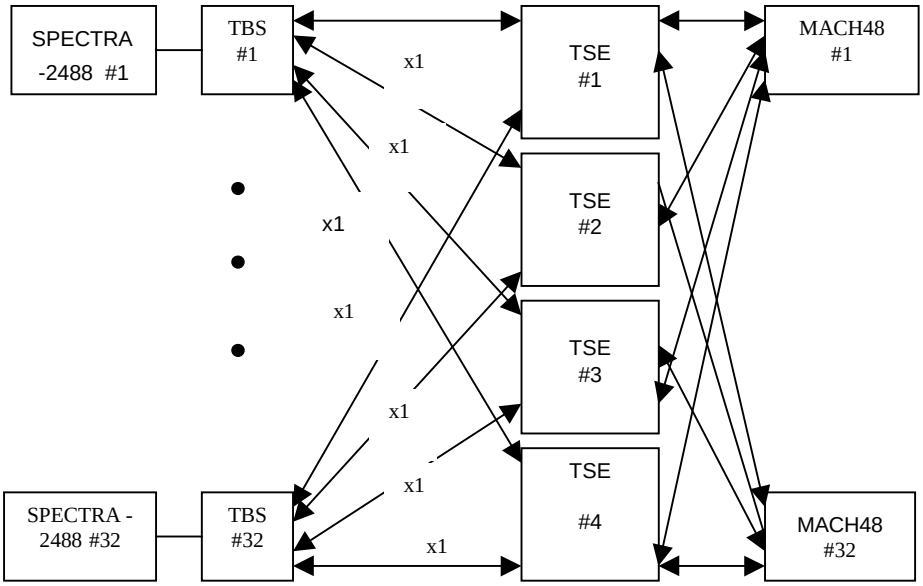
A few common multi-service switch architectures will be shown to assist in understanding how to build the fabric and connect the TBS with the TSE chips. The fabric can also provide for sub-lamda grooming without SONET path termination by replacing the S/UNI-MACH48 with the TBS and SPECTRA-2488.

4.1 Single Stage Fabric: Up to 160 Gbps

Since each TSE can support 40 Gbps of bandwidth, a very large single stage fabric can be built without the requirement for multi-stage architectures. Recall each TBS has four STS-12 LVDS ports (actually 3 sets of 4 – working, protect, auxiliary) which can connect to the TSE. Therefore, a non-blocking single stage fabric can be built with up to 4 TSE's by connecting a single LVDS from the TBS to each of the TSE devices. This gives a 160 Gbps fabric as shown in Figure 10. 32 TBS and 32 S/UNI-MACH48 chips will be connected in this fashion to fully utilize the 160 Gbps fabric (assuming both the TX and RX links of the devices are used).

A single TSE can be used to build a 40 Gbps fabric and two-chips yield a 80 Gbps fabric. . In a 40 Gbps fabric all four ports of the TBS and S/UNI-MACH48 feed a single TSE. Thus 8 total TBS and 8 S/UNI-MACH48 devices would interface to a single TSE. In an 80 Gbps fabric two of the TBS STS-12 LVDS links mate to each of the two TSE devices. A total of 16 TBS and 16 S/UNI-MACH48 devices would interface to the two TSE cross connects. In addition, an equivalent number of SPECTRA-2488 devices could interface to the line side of the TBS. This example is bi-directional in which both the TX and RX links of the respective devices are connected to the TSE.

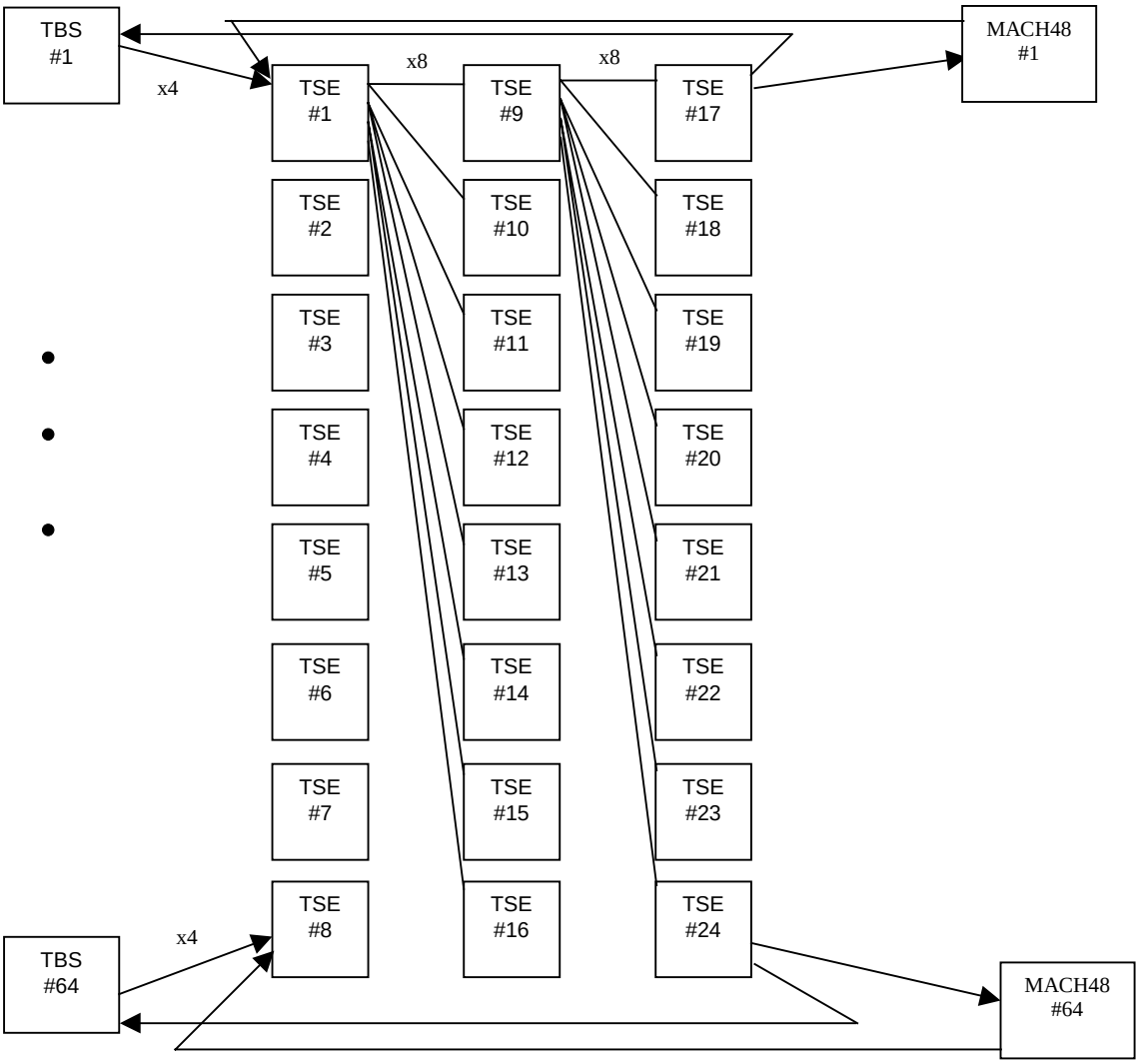
Figure 10: Single Stage 160 Gb Fabric



4.2 Three Stage Fabric: Up to 2560 Gbps

Moving beyond 160 Gbps requires a three-stage architecture to guarantee all possible port mappings and a non-blocking architecture. The example in Figure 11 shows a 320 Gb fabric. Not all connections are shown for simplicity and clarity in this example. The 64 ports on the TSE are sent in groups of eight between TSE chips in each stage. All four links from a TBS are connected to a single TSE. This architecture requires 64 TBS and S/UNI-MACH48 devices and 24 TSE chips. A 2560 Gbps fabric is constructed similarly, with each TSE feeding a single STS-12 LVDS link to one of the 64 other TSE devices in the next stage.Note that in Figure 11 the TBS and S/UNI-MACH48 device sit on both sides of the first and third fabric.

Figure 11: Three Stage 320 Gb Fabric



4.3 Device Count Required For Fully Populated Fabrics

The following table illustrates the number of devices required for a fully populated TSE fabric as shown in Figure 7b. Note that the TSE device count doubles if implementing fabric protection on the TSE (single point of failure). In general, the device count scales directly with the bandwidth (as the bandwidth doubles, the chip count doubles). Note that a three stage TSE fabric is required for bandwidths beyond 160 Gbps. This produces a tripling of the TSE chip count on top of the normal doubling with bandwidth at the 320 Gbps fabric size. Beyond 320 Gbps all devices scale with bandwidth until 20480 Gbps where a 5 stage TSE fabric is required to produce a non-blocking architecture. Planes refer

to the number of different TSEs that a TBS is connected to. If all four TBS links are connected to a single TSE, then this represents a plane of 1. If half (two) of the TBS links are connected to the first TSE, then the other two are mated to a second TSE then this is a plane of two. Finally if each of the four TBS link services an individual TSE, then this represents a plane of 4. Note that for a 10240 Gbps fabric, that each TBS link feeds 1 of four 2560 Gbps three stage TSE fabrics.

Table 1 Device Count For Various Bandwidths

Bandwidth	Fabric Layout		# of Devices Required		
	Stages	Planes	SPECTRA-2488, TBS and S/UNI-MACH48	SPECTRA-2488, TBS and S/UNI-MACH48	TSE (x2 for fabric protection)
			Unidirectional	Bi-directional	
40 Gb	1	1	16	8	1
80 Gb	1	2	32	16	2
160 Gb	1	4	64	32	4
320 Gb	3	1	128	64	24
640 Gb	3	1	256	128	48
1280 Gb	3	1	512	256	96
2560 Gb	3	1	1024	512	192
5120 Gb	3	2	2048	1024	384
10240 Gb	3	4	4096	2048	768
20480 Gb	5	1	8192	4096	2560

5 DEVICE INTERFACES

5.1 LVDS Interface

The LVDS interface implemented on the TBS and TSE follows the IEEE 1596.3-1996 specification with some minor differences. The changes are implemented to customize and optimize the LVDS interface for the system and are described in detail below. Even with these differences the LVDS interface should function with the physical layer of other LVDS interfaces. The differences include:

1. Faster rise/fall times (200 – 400) ps versus (300 - 500) ps. Faster edge rates are commonly used with higher speed LVDS interfaces in the industry to ease the interfacing. The IEEE 1596.3-1996 edge rates are optimized for data rates of 400 Mbps and below.

2. Hysteresis is not implemented on the receive LVDS interface. Hysteresis is used in many implementations to negate the effect of noise that may exist on any of the unused LVDS links. Hysteresis was not implemented in the CHESS set devices to minimize circuit complexity, power and cost. The RX interface and the DRU can be both disabled (powered down) by the appropriate register to prevent any sensitivity to noise.

3. The LVDS transmitter contains an on-chip 100-ohm termination. Most implementations have single 100-ohm termination on the receiver. By implementing a double termination (on both the LVDS receiver and transmitter) a higher signal integrity and matching is ensured.

4. Although not a difference with the layer 1 IEEE 1596.3-1996 specification, the layer 2 8B/10B encoding is presented here for completeness. 8B/10B encoding guarantees transition density as compared to scrambled encoding, which provides only a certain probability of transition density. This guaranteed transition density allows a simpler and more power-effective data recovery unit, provides a more robust serial interface (greater trace or back-plane distance achievable). It also negates the need for complete SONET framing since the A1A2 and J0 bytes can be encoded into special escape characters of the LVDS data stream.

5.2 Hot Swapping and Floating Links on the LVDS Interface

The interface is a LVDS, not a CMOS interface; therefore there is no electrical problem in leaving the LVDS interfaces floating. Note that the LVDS receiver consists of a differential amplifier with a wide common-mode range. The power dissipation is independent of the data transitions (that is, if the input is

connected). There is an internal 100 Ω termination across the positive and negative input. Floating inputs will settle to an arbitrary voltage (between VDD and VSS) determined by leakage paths. Regardless of this arbitrary voltage, the input structure of the receiver will operate in its proper range and the receiver output will be logic 1 or 0 depending on internal offsets. Noise events (power supply noise, crosstalk) may induce the receiver to toggle randomly from 1 to 0 generating "ambiguous" data. This ambiguous data will not result in any problems but is not a desirable condition. Recall that it is the responsibility of the user to configure the appropriate links for calls.

Unused links should be disabled in software. The consumed power for that link will be nearly 0mW. There is no requirement for how quickly the link should be disabled. Disabling the link, simply results in lower power dissipation since the circuitry will be shut down. This action is not mandatory, but is good practice to improve margins. There are no problems with hot-swapping. The “hot-swap” channel can be left enabled and the device will sync up once the far end transmitter is connected. There are no affects on other channels. Hot swapping of cards is still allowed by reprogramming of the links in software.

5.3 No device damage will occur when driving an unpowered input board. (For example, if a board is inserted and power has not come up yet, or a fuse has blown on the board). If a LVDS driver is connected to an unpowered receiver, the LVDS transmitter will force a maximum of 7mA to one of the input pins. This produces a forward bias condition on the ESD protection scheme but not sufficient to elevate the power supply of the device at the receiving end and produce device damage (nor to result in a latch-up condition). Also note that the LVDS interface should not be damaged by interfacing to a 3.3V TTL slot. Calculating the Maximum Difference in Trace Lengths

The TSE utilizes 64 different input and output differential LVDS pairs. It is critical to match the lengths of the positive and negative traces of each differential pair to minimize skew and maximize the eye opening. However, matching one differential pair to another pair is not as important. To accommodate this, the high-speed serial LVDS links have a 24 word (10 bit byte) FIFO. Of this 24 word FIFO, 8 words should be allocated for clock skew and wander between cards or within devices. The remaining 16 words are then available to accommodate differences in trace lengths between LVDS pairs.

The 16 word FIFO yields an allowable delay of 205.8 ns or 41.2m.
⇨ 16 words x 10 bits/word = 160 bits of margin in FIFO
⇨ 1/777.6 Mb/s = 1.29 ns/bit on the serial link
⇨ 160 bits x 1.29 ns/bit = 205.8 ns of margin or 16 clock cycles (at 77.76 MHz)

A transmission speed of 2/3 the speed of light, this corresponds to a trace length difference of 41.2m.
⇨ 205.8 x 10^{-9} s x 2/3 x 3 x 10^{8} m/s = 41.2 m

It is important to note that the LVDS interface itself is designed to drive 1m of backplane plus 30cm of trace length on either side. Therefore this will be the limiting factor. This 1.6m of trace length corresponds to 8ns of delay or 0.6 clock cycles (77.76 MHz) between any LVDS link. Low loss cable or an optical interface can be used to connect to the LVDS interface to realize greater back-plane distances.

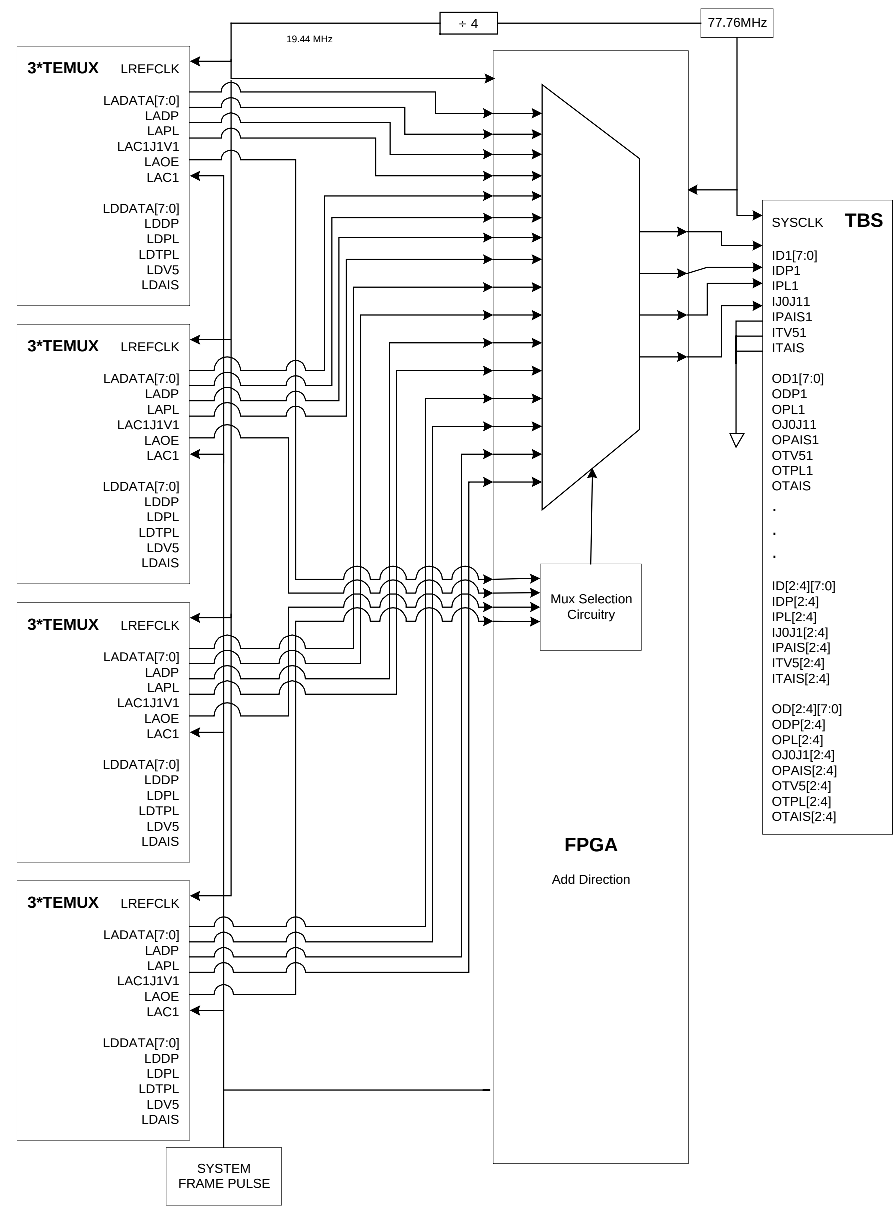
5.4 Interfacing a 19.44 MHz TeleCombus to a 77.76 MHz TeleCombus

The task of interfacing the 19.44 MHz TeleCombus to the 77.76 MHz TeleCombus will be illustrated by means of the TEMUX devices (19.44 MHz) and TBS (77.76 MHz) as shown in Figure 12. Three TEMUXs together drive a 19.44MHz TeleCombus. Four such 19.44MHz buses are muxed together into a 77.76MHz bus, so twelve TEMUXs are required to interface to each of the four 77.76 MHz TBS buses.

Each 19.44 MHz stream is muxed in sequentially- first, second, third, fourth, first, … The exact order doesn't matter, as long as it's consistent and the timing signals are muxed in the same order. The J1 pulses must be muxed together in the same order as the data streams. The same criterion applies for demuxing. Synchronizing the clock will have to be done carefully to avoid timing problems. The only criterion is that set-up and hold times are met throughout the system.

Note the IJ0J1 signal into TBS expects only one J0 pulse, marking the first J0 byte in the incoming stream. But there will be one J0 (C1) pulse coming from each of the four sets of TEMUXs. Only the first J0 (C1) must be sent to TBS; the others are discarded. The converse is true in the opposite direction; one J0 pulse from TBS needs to be replicated into four, one for each of the TEMUX sets.

Figure 12: Interfacing the TEMUX to the TBS



6 SYSTEM SYNCRONIZATION

6.1 Purpose of Synchronization

Synchronization is required to align the serial (STS-12) LVDS links (such as the 64 serial links of the TSE). A global frame pulse (8 kHz, 125µs) is generated and fed to each CHESS device and the J0 byte of the STS-12 data stream is used to synchronize each STS-12 link. Synchronization is accomplished on each link by programming the clock cycle delay from the global frame pulse to the time when the J0 byte is received on that link. Thus the device is programmed to expect the J0 byte a set number of clock cycles following the frame pulse.

There are two primary sources of delay. First, the delay can be introduced by clock skew or trace length differences between different links. However these forms of delay can be ignored in practice because the input receive links compensate for them. A 24-word input FIFO is available to accommodate for any mismatches. As discussed earlier, 1 m differential plus clock skew is easily tolerable from link to link. The second and most significant source of delay is introduced by the device latency.

6.2 Implementing Synchronization in CHESS SET Devices

Consider the implementation shown in Figure 13. All devices receive the global frame pulse simultaneously at time t_0 (ignoring any trace length differentials). The SPECTRA-2488 emits the J0 byte onto the TeleCombus upon receiving the global frame pulse on the DJ0REF input. This action is entirely independent of receiving a J0 byte from the optical line. SPECTRA-2488 pointer adjustments will define the start of the payload envelope (the J1 byte indicates start of payload) and this payload will be outputted over the TeleCombus. The SPECTRA-2488 can be viewed as the master by which the synchronization of the other CHESS devices is determined. The TBS expects the four incoming eight bit 77.76 MHz TeleCombus data paths to be synchronized and upon processing emits the serialized data with J0 character 33 clock cycles after receiving the J0 on the parallel TeleCombus. The J0 byte on each of the twelve independent 777.6 MHz LVDS links are not exactly simultaneous and may have a slight amount of skew relative to each other (because of presence of an 8 word FIFO on the LVDS transmitter output). The LVDS links are then mated to the TSE through a back-plane. The TSE is programmed (via indirect register access of the RJ0DLY[13:0] word) to expect the J0 byte a certain number of clock cycles after it receives the global frame pulse.

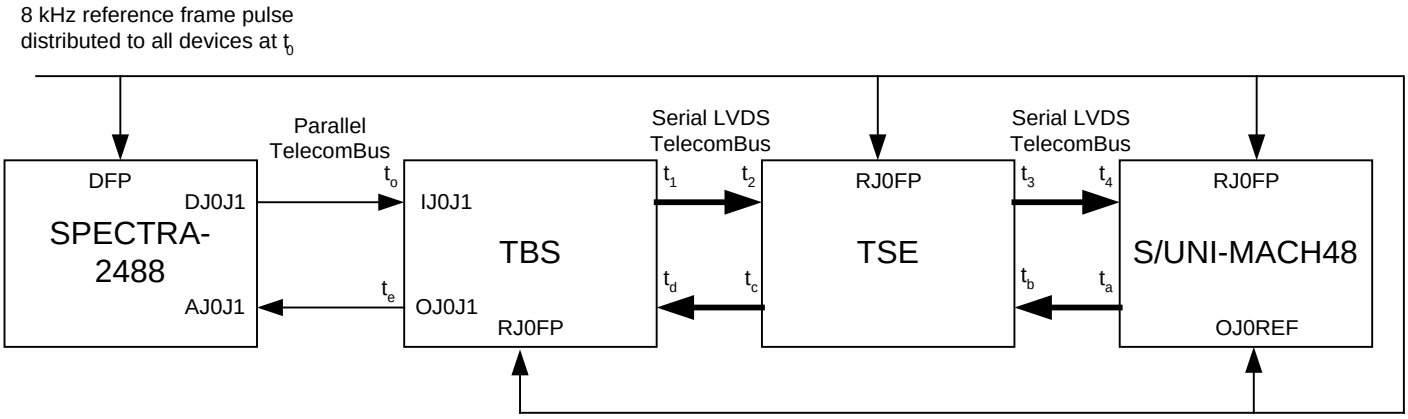
The ingress FIFOs permit a variable latency in J0 arrival of up to 16 clock cycles. That is, the largest tolerable delay between the slowest and fastest LVDS link of

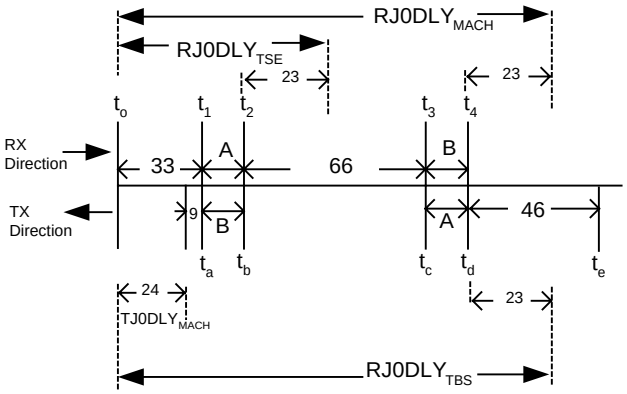
the 64 TSE LVDS links is 16 bytes. Consequently, the external system must ensure that the relative delays between all the 64 receive LVDS links be less than 16 bytes. The minimum value for the internal programmable delay (RJ0DLY[13:0]) is the delay to the last (slowest) J0 character plus 15 bytes. The maximum value is the delay to the first (fastest) J0 character plus 31 bytes. The actual programmed delay should be based on the delay of the "slowest" of the 64 links – the link in which J0 arrives last plus a small safety margin of 1 or 2 words. The magnitude of the clock cycle delay is bounded by two parameters. First, the programmed delay register RJ0DLY is 14 bits. This implies that a clock cycle delay of 2^{14} –1 or 16,383 clock cycles can be programmed. However, the second parameter, the frame rate (125 μs), bounds the delay to one STS-12 frame or 9719 (9720 unique values but 0 is the value for no delay) clock cycles (125 μs x 77.76 MHz), after which the next SONET frame begins. The TSE, upon receiving the global frame pulse, will wait the programmed amount of time (56 clock cycles + cable length delays) before searching each of the 64 links for the location of the J0 pulse to initialize synchronization.

The number of clock cycles can be determined by simply adding the relevant device and cable length latencies. In practice, the programmed delay can be obtained by measuring the clock difference between the global 8kHz frame pulse and the presence of the J0 on the TJ0FP pin. The programmed delay is this clock cycle difference plus a few clock cycles for margin.

This synchronization mechanism is flexible to accommodate differing path lengths. Consider a TSE that is mated to a S/UNI-MACH48 on one link and a SPECTRA-2488 feeding a TBS on the other link. The alternate data paths have different delays; the SPECTRA-2488/TBS link has a greater delay than the S/UNI-MACH48 link delay. In this case, the S/UNI-MACH-48 is programmed to emit the J0 pulse later than SPECTRA-2488 (but aligned with the TBS serial output) such that the J0 from both sources arrive at the TSE within the allowed 16-clock cycle window. The S/UNI-MACH48 programmed delay is 24 clock cycles after the receipt of the frame pulse.

Figure 13: J0 Synchronization Control





In the line side RX direction:

1. An 8kHz frame pulse is received by all devices at time t_0
2. Upon receipt of the 8kHz frame pulse, the SPECTRA-2488 outputs data (and J0) onto the TeleCombus at time t_0
3. The TBS emits the serialized J0 byte and data 33 clock cycles later at time t_1
4. The J0 byte arrives at the input of the TSE A clock cycles later at time t_2. A and B represents the clock cycle delay of the link
5. 23 clock cycles after the receipt of the J0 on the input link, the various J0 bytes are centered in the TSE's 24 word FIFO. This cumulative time from t_0 is the programmed RJ0DLY of the TSE.
6. The TSE emits the J0 66 clock cycles after receipt of the J0 at time t_3.
7. The J0 byte is present at the MACH48 device B clock cycles later at time t_4
8. 23 clock cyles after the receipt of the J0 on the serial link, the four incoming J0 bytes of each LVDS link are centered in the 24 word FIFOs. This cumulative time from t_0 is the programmed RJ0DLY of the TSE.

In the line side TX direction:

1. An 8kHz frame pulse is received by all devices at time t_0
2. Nine clock cycles after receiving the 8kHz frame pulse, the S/UNI-MACH48 outputs data (and J0) to the TSE device. Since t_b and t_2 must be equal (synchronous) for the TSE to function, a 24 clock cycle delay must be programmed as the TJ0DLY of the S/UNI-MACH48. This assumes the link delay A = the link delay B or is negligible. Actually, $t_b - t_a$ is the propogation time of the signal between the devices measured in 77.76 MHz clock cycles.
3. The TSE outputs the J0 byte at time t_c
4. The TBS receives the J0 byte at time t_d and 23 clock cycles later, each of the four J0 bytes are centered in the 24 word FIFOs. This cumulative time from t_0 is the programmed RJ0DLY of the TBS.

6.3 Implementing Synchronization in CHESS SET Devices

Table 2 Synchronization Pins & Registers

Device	Pin Name		Register Bit(s) Name	
	Transmit J0 Frame Pulse	Receive J0 Frame Pulse	Transmit J0 Delay	Receive J0 Delay
TBS	TJ0FP	RJ0FP		RJ0DLY[13:0]
TSE	TJ0FP	RJ0FP		
S/UNI-MACH48	TJ0FP	RJ0FP/ OJ0REF	OJ0DLY[13:0]	RJ0DLY[13:0]
SPECTRA-2488		DFP		
SPECTRA-622		DFP		
SPECTRA-4x155		TBD		

7 PROGRAMMING CONNECTIONS

7.1 Introduction

This section describes the process of physically mapping a call onto a fabric comprised of CHESS devices. The operation of the various blocks within the devices is described, and the programming of the devices in a coordinated fashion to provide hitless rearrangement of the fabric is addressed.

7.2 Problem description

Connections within a switch fabric made from PMC-Sierra's CHESS devices are made at the STS-1 level. Programming multiple parallel STS-1 connections supports higher levels of concatenation. Any level of concatenation can be supported in this way through a TSE fabric. As the devices have aggregate capacity of an STS-48 or higher and physical connections between the devices are made with STS-12 links, a method is needed to calculate how to route a call through the fabric.

A call is defined as a single STS-1 connection from a timeslot on an input port of an input device to a timeslot on an output port of an output device. Depending on the topology of the fabric, the call may traverse one or more core stages comprised of TSEs.

The CHESS architecture is a re-arrangeably non-blocking Time-Space-Time STS-1 switch fabric. The "Space" portion of the fabric takes data from a set of available inputs and moves it to a set of outputs. There are multiple settings for the space switch each corresponding to an STS-1 timeslot. The first "Time" portion of the fabric is responsible for moving the data from the timeslot it is received on to the timeslot that matches the space switch's opportunity to move the data to the desired output. The second "Time" portion of the fabric is responsible for moving the data from the output of the space switch to the desired timeslot on the output port. Depending on existing connections within the fabric, there may not be a time where the space stage has both the desired input and the desired output free. In this case existing connections may need to be rearranged to add the desired connection.

Setting up a call therefore requires the following steps:
1) Determining a set of STS-1 level connections needed for the aggregate bandwidth of the desired call,
2) Determining a path for each of these connections through the fabric,
3) Programming the connections into each of the devices in the path,
4) Committing the new connection definitions in all devices simultaneously.

7.2.1 The Routing Algorithm

A call can be described as a collection of Device-Port-Timeslot (DPS) pair. A single DPS pair describes a single STS-1 call. N DPS pairs therefore describe an STS-N call. Reference software is available from PMC-Sierra to perform the routing of each STS-1 call through a fabric. The output of the algorithm identifies which timeslot through the space portion of the switch (called a wave) can be used to make the desired connection. Based on the topology of the fabric, this information is used to derive the actual device settings required to set up the call.

7.2.2 Multicast considerations for protection and port-mirroring

The routing algorithm provided by PMC handles multicast as well as unicast connections. In general, a mix of multicast connections from any port to any port will result in an indeterminate route calculation time and in some cases the multicast routing may fail to find a path. The documentation for the algorithm provides a set of conditions for which multicast routing is guaranteed.

7.2.3 Device Programming Connections Overview

There are currently three devices in the PMC CHESS Set which together form a complete switch fabric. These devices are the TBS, TSE, and S/UNI-MACH48. The S/UNI-MACH48 and TBS implement the “time” portion of the fabric while the TSE implements both the “space” portion and optionally the ingress and egress “time” portion in some configurations.

The TBS has a parallel STS-48 TeleCombus interface for connection to a SONET framer and twelve STS-12 serial interfaces for connecting to the fabric. Only four of the twelve are required for connecting to a fabric. The other eight are available for redundant configurations and other unique topologies. Any STS-1 can be mapped from the STS-48 TeleCombus to any one or more of the STS-1s on the outgoing serial links. In the reverse direction, any of the STS-1s on any of the twelve serial links can be mapped into any STS-1 on the parallel TeleCombus.

The S/UNI-MACH48 has the same STS-1 interchange functionality as the TBS with the exception that it has eight serial STS-12s instead of twelve. The S/UNI-MACH48 will thus be treated as a subset of the TBS in the following discussions and not treated directly except when discussing individual register addresses.

The TSE implements the “Space” portion of the fabric and optionally part of the “time” portion. It has 64 serial STS-12 interfaces. The time-based portion of the TSE can rearrange the order in which the STS-1s on an incoming STS-12 link are presented to the space fabric. STS-1s from the 64 incoming STS-12s are

presented to the input of the space switch in twelve time periods. Connections from the inputs to the outputs of the space switch within each of these time periods are described as a wave. Each wave can have a unique mapping of inputs to outputs.

7.2.4 Rearrangement coordination

All internal switching blocks of the CHESS Set devices have two configuration tables. One of the tables describes the active switch settings. The second table is used to modify the switch settings off-line. The devices can be instructed by a global signal to switch the definition of the active and inactive tables on the second 125µs STS-12 frame boundary after receiving the global switch over signal. (As STS-1's are transported over STS-12 links, the switch is coordinated on the first STS-1 within an STS-12). This allows a coordinated hitless reconfiguration of a fabric containing multiple devices. Alternatively the user can change which page is active via software control on a per-STS-48 block basis within each device. This is useful mainly in situations where the modification being made does not affect other existing calls or coordinated switchover with other devices is not required. Example configurations of this type could include devices in an inactive redundant switch plane or small fabrics comprised solely of TBS and/or S/UNI-MACH48 devices. In all cases, the selected page bit within each functional block is exclusive-OR'd with the external page-select pin allowing for a mixture of the coordinated switch models.

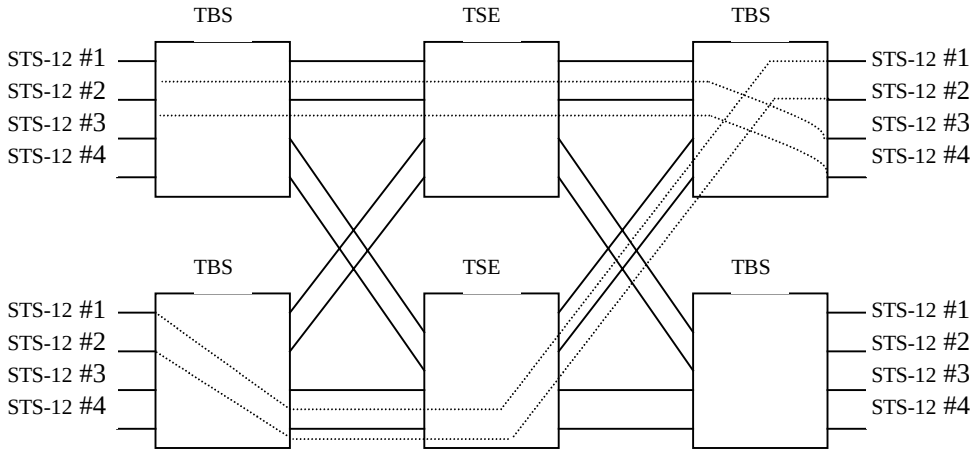
The user may choose one of several methods for updating the switch table settings when adding or deleting calls. One method is to write the entire new switch configuration for each device in all of the inactive tables, then force the switch between active and inactive settings. This is inefficient in terms of managing the hardware when only incremental changes are required for adding a new connection but it may be useful for implementing network connection preplans. The second method is to write only the necessary changes to the inactive tables before forcing the switch. This requires that the software keep track of the existing state of the inactive tables both for configuration purposes and when routing new calls. To ease this burden, it is assumed that the user will perform the following steps when reconfiguring the devices:
1) perform a call routing calculation,
2) write the necessary changes to the inactive switch setting tables,
3) force the switch between active and inactive tables,
4) update the now inactive tables with the same changes made to the now active tables.

This enables the user to keep track of only one set of switch settings, which simplifies the call setup and route calculation steps.

7.2.5 Implications of Topologies

There are many topologies available for interconnecting the devices in the CHESS chipset as discussed earlier. Topologies using the TSE device have multiple time-stages – those in the TBS/S/UNI-MACH48 devices and those in the TSE. The question arises as to whether all of the time stages are necessary, and if not which of these time stages should be used. For a topology other than a 40 Gbps or smaller fabric using one TSE (or two TSE's for redundant fabrics), the time stages in the TBS/S/UNI-MACH48 devices should be used and the TSE time stages left in the "identity mapping" or pass-through mode. If there is only one active TSE in the switch fabric, then the TSE's time stages can be used and the TBS/S/UNI-MACH48 time stages can be left in the identity-mapping mode. The reason for this is illustrated in Figure 14. In this configuration there are two TSE's. Two TBS's with four STS-12 ports each are shown. For simplicity of the diagram, the other 30 TBS's supported by the TSE's are not shown. In order to make the desired connections in this example shown by the dashed lines, the time stages within the input TBSs must be used to direct the STS-12's to the proper TSE for final connection to the desired output TBS. The time stage in the output TBS is used to direct the STS-12's to the desired output ports. This can always be done without making any timeslot rearrangements within the TSE time stages.

Figure 14: Time-Space-Time Switching



In summary, the guidelines for which TSI blocks need to be used in which devices can be broken down by topology as follows:

Single TSE fabrics (<=40Gbps): The time stages in the TSE can be used and the time stages in the TBS/MACH devices can be bypassed.

Single-stage, multi-TSE fabrics (80Gbps to 160Gbps typically): The time stages in the TBS/MACH must be used and the time stages in the TSE are bypassed.

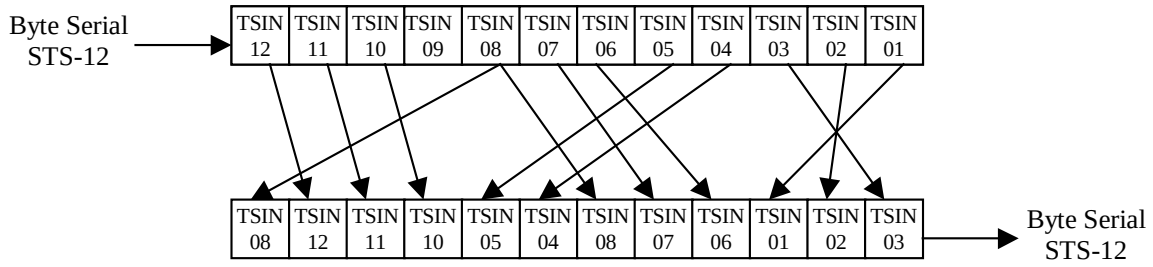
Multi-stage TSE fabrics (320Gbps to 1.2Tbps typically): The time stages in the TSE can be used and the time stages in the TBS/MACH devices can be bypassed.

7.3 Time-Slot-Interchange

7.3.1 Operation of the TSI

The TBS, TSE, and S/UNI-MACH48 all have STS-1 time-slot interchange capabilities. To describe how these work, first consider the basic case of a block with a byte-serial STS-12 stream input and byte-serial STS-12 output. Timeslot interchange capability in such a device is achieved by storing the incoming data in a shift register. Once the entire STS-12 stream has been received in this register, bytes are moved to a second shift register whereby any input register byte can be written to any output register byte. Shifting out the contents of the output register forms the outgoing STS-12 stream. This operation is shown in Figure 15. Note that multicasting in this case is trivial. A single byte in the input register may be written to multiple timeslots in the output register. This is the basic operation of the Ingress Time-Switch Element (ITSE) and Egress Time-Switch Element (ETSE) blocks in the TSE device.

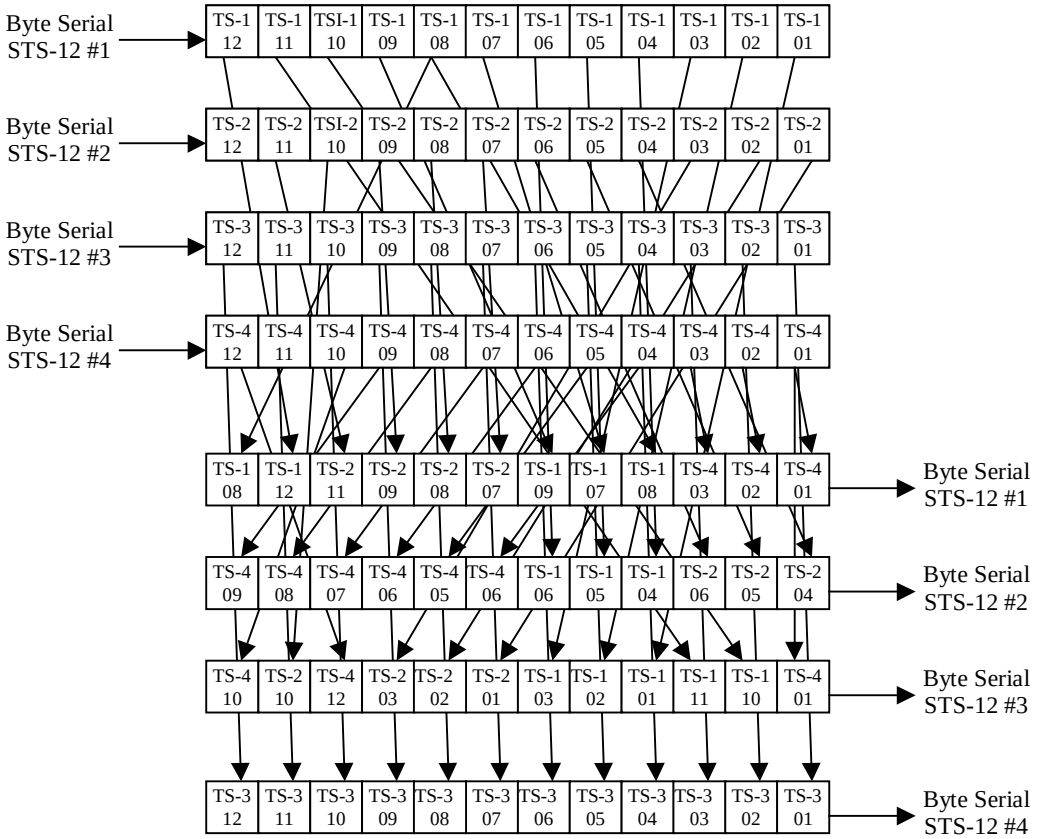
Figure 15: Operation of the Time Slot Interchange



Extend Figure 15 by allowing multiple aligned STS-12 streams at the input and output and extend the switching such that bytes from any of the input registers can be switched to any one of the output registers in any of the multiple locations. For example, the TBS has three copies (working, protect and auxiliary) of a configuration with four inputs and four outputs. In the ingress direction (STS-48 parallel to 4xSTS-12 serial), the STS-48 stream is demultiplexed into four STS-12s and each STS-12 is shifted into three of the input shift registers (working, protect and auxiliary). This operation is illustrated in Figure 16. There are three identical blocks in the TBS that implement the function in this diagram, all sharing the same four Byte Serial input streams

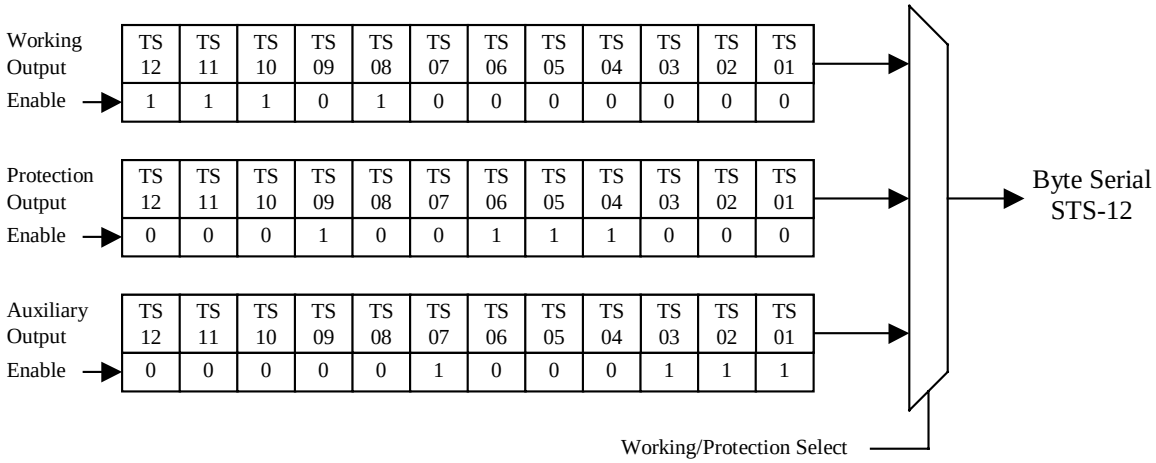
originating from the TeleCombus. These are the Transmit Working Timeslot Interchange (TWTI), the Transmit Protection Timeslot Interchange (TPTI), and the Transmit Auxiliary Timeslot Interchange (TATI).

Figure 16: The TBS Time Slot Interchange Operation



In the egress direction, the TBS must choose from one of three output registers on a time-slot by time-slot basis to form each STS-12 byte-serial output. This selection is achieved by programming an enable into the block (working, protect or auxiliary block) that is to be chosen for the output (see Figure 17). Alternatively, an external pin can be used to force the simultaneous selection of either all of the "working" STS-12s' timeslots or all of the "protection" STS-12s' timeslots (see Figure 17). The blocks in the TBS that implement this function are the Receive Working Timeslot Interchange (RWTI), Receive Protection Timeslot Interchange (RPTI), and Receive Auxiliary Timeslot Interchange (RATI). The operation of selecting a single output from several candidate STS-12 streams is illustrated in Figure 17.

Figure 17: Selecting an Output STS-12 Data-stream



The S/UNI-MACH48 has the same capability as the TBS with the exception that it has eight byte-serial STS-12 links as opposed to twelve. The blocks are named Input Working Timeslot Interchange (IWTI) and Input Protection Timeslot Interchange (IPTI) in the ingress direction toward the system side of the S/UNI-MACH48 and Output Working Timeslot Interchange (OWTI) and Output Protection Timeslot Interchange (OPTI) in the opposite direction.

7.3.2 Programming the TBS

The TBS contains three Timeslot Interchange (TSI) blocks in each direction. The transmit direction is defined as from the parallel TeleCombus to the serial STS-12 interfaces. The receive direction is defined as from the serial STS-12 interfaces to the parallel TeleCombus. For programming purposes, the 32-bit TeleCombus is identified as four individual 8-bit bus segments numbered 1 to 4. The three TSI blocks are the Working, Protection, and Auxiliary respectively labeled TWTI, TPTI, and TATI in the transmit direction and RWTI, RPTI, and RATI in the receive direction. Each block controls four outgoing STS-12 serial streams numbered 1 to 4. Timeslots are numbered 1 to 12. Accesses to these blocks are via indirect reads and writes.

Any time-slot can be over-written with a programmable idle pattern in the transmit direction. This is normally done when a connection to the output timeslot is torn down. The pattern is programmed as a 10-bit code as it is inserted after 8B/10B encoding of data has been performed internal to the device. Care must be taken not to insert 8B/10B control characters or invalid 8B/10B codes. The following steps perform the programming an idle code insertion function:

1) Poll the indirect address register until the BUSY bit is low.

2) Write '11' to the IDLE[1:0] bits and the desired 8B/10B encoded idle pattern in the IC[9:0] bits of the indirect data register (1010110001b for encoded all 1's).
3) Write the desired output stream number in IWDSEL[1:0], the desired timeslot number in IWDTSEL[3:0], and the inactive page number in the PAGE bit in the indirect address register.
4) Perform other system configurations in other devices if desired.
5) Force switchover to the inactive page using the TCMP pin or CMPSEL register bit(s).
6) Update the newly inactive page with the same configuration information.

The change will take effect on the second STS-12 frame boundary after the inactive page is selected as the active page. Consider this example, to send an 8B/10B encoded idle all '1s' pattern in timeslot 3 of the protection LVDS link #2 with page 1 being active. The following writes would be made after polling BUSY inactive in the TPTI Indirect Address register: TPTI Indirect Data register = 0x1xxx, TPTI Indirect Address register = 0x4032. After the active page switch has been performed, the following two writes would occur (simply inverting the polarity of the PAGE bit in the indirect address register): TPTI Indirect Data Register = 0x1xxx, TPTI Indirect Address Register = 0x4432.

A connection can be set up by writing the source timeslot and input TeleCombus segment to the indirect data register, followed by writing the target serial STS-12 stream number and timeslot in the indirect address register. As with the idle code assertion, this should first be written to the inactive configuration page, then a page switch should be performed, followed by the updating of the newly inactive page. Consider this example, to set up a connection from timeslot #10 of STS-12 #1 on the parallel TeleCombus to timeslot #5 of working serial STS-12 link #0 with page 0 active. The following writes would be made after polling BUSY low in the TWTI indirect address register: TWTI Indirect Data register = 0x00a1, TWTI Indirect Address register = 0x4450. After the active page switch has been performed, the following two writes would occur (simply inverting the polarity of the PAGE bit in the indirect address register): TWTI Indirect Data Register = 0x00a1, TWTI Indirect Address Register = 0x4050.

Programming the receive direction of the TBS (from the serial STS-12 streams toward the parallel TeleCombus) follows the same procedure as for the transmit direction. The only differences are that
1) there is no idle code insertion available in the receive direction, and
2) the enable bit for one and only one of the three streams (RWTSEN, RPTSEN, RATSEN for working, protection, or auxiliary respectively) must be set for any given timeslot when setting up a connection or cleared to tear down a connection.

Note that per-timeslot selection between W, P, A links can only be done when RWSEL_EN register bit = 0. Care must be taken that two timeslots are not

programmed to the same destination with both having the enable bit set. Besides programming the enable bits, there is an alternative way to choose between the working and protection serial links in the receive direction. This scheme is available when mixing of timeslots between the three links is not necessary. In this scheme, RWSEL_EN register bit must be set to logic 1 to enable the external RWSEL pin. The external RWSEL pin is used to select between the working and protection links. Switchover based on the RWSEL pin occurs on the J0 boundary.

7.3.3 Programming the TSE

There are 16 Ingress Time Slot Interchange Element (ITSE) and sixteen Egress Time Slot Interchange Element (ETSE) blocks within the TSE device. Each controls four serial STS-12 inputs and four STS-12 outputs. Programming these blocks follows the same procedure as the TSI blocks in the TBS device with the restriction that timeslots may only be interchanged within an STS-12 stream and not between the four STS-12 streams controlled by a given block. This restriction requires that the DINSEL[1:0] and DOUTSEL[1:0] bits in the ITSE blocks be programmed with the same value when setting up a TSI connection. Both the ITSE and ETSE functional blocks support idle code insertion via the same method used in the TBS TWTI, TPTI, and TATI blocks.

7.3.4 Programming the S/UNI-MACH48

The S/UNI-MACH48 contains two Timeslot Interchange (TSI) blocks in each direction. The ingress direction is defined as from the serial TeleCombus towards the UL3/PL3 interface. The egress direction is defined as from the UL3/PL3 interface toward the serial TeleCombus. The two TSI blocks are the Working and Protection blocks labeled IWTI and IPTI in the ingress direction and OWTI and OPTI in the egress direction. These blocks are programmed and operated similarly to the working and protection TSI blocks in the TBS device. Exceptions are that the idle code insertion feature is not available. Also note that the selection between the working and protection timeslots in the ingress direction is defined by the state of the SER_PRT_SEL bit in the IWTI block when RWSEL_EN register bit is logic 0. The external RWSEL pin can be used to select between the working and protection ingress serial links when RWSEL_EN register bit is logic 1. The RWSEL pin on the S/UNI-MACH48 has the same operation as on the TBS.

7.3.5 Programming optimizations

Each of the programming sequences described above includes the polling of a "BUSY" bit that indicates when the device is in the process of implementing a previously requested indirect read or write operation. The setup of a connection

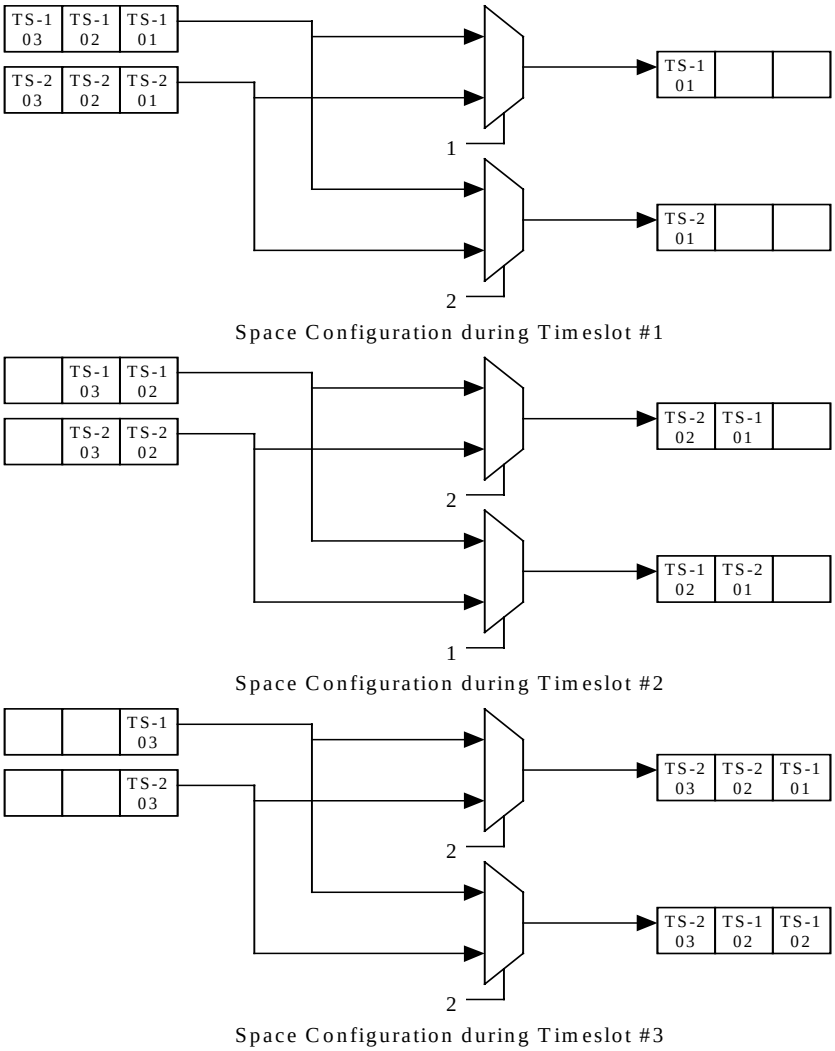
normally involves the programming of several different functional blocks within a device or even multiple devices. , Therefore the time spent waiting for BUSY to clear can be reduced by performing reads and writes in an order such that successive operations are not performed to a given functional block.

7.4 TSE Space Switch

7.4.1 Operation of the TSE Space Switch

The TSE device has what is called a space switch stage as well as the time switch stages. A space switch is able to connect any of the inputs to any of the outputs. Connecting multiple inputs to the same output is not allowed, though multicasting a single input to multiple outputs is allowed. A space switch can be conceptualized as a multiplexer at each output that is able to choose from all the inputs. The TSE has 64 inputs and 64 outputs. It supports twelve settings for the 64 output multiplexers. These settings are used to allow a unique setting of the space switch for each of the twelve time periods during the shift-out of the ingress timeslot output registers. The operation of a simple 2x2-space switch with three ingress timeslots is illustrated in the following figure. During each of the three timeslots there are unique settings programmable for each of the output multiplexers. Figure 18 shows the intermediate and final results after each step of the process.

Figure 18: The TSE Space Switch Operation



7.4.2 Programming the TSE Space Switch

The space-switch configurations are made using pair of registers to perform indirect write accesses. As with other CHESS device functional blocks, there are two pages of configuration memory allowing changes to be made to the inactive page and then committed on the second STS-1 frame boundary after selection of the inactive page in a hitless manner. Writing the source port address (from 1 to 64) to the indirect data register, followed by writing the destination port address and the timeslot in which the connection is to be made in the indirect address register can set up a connection. This connection information should first be written to the inactive configuration page, then a page switch should be performed followed by the updating of the newly inactive page. Consider this

example, to set up a connection during timeslot #10 from the STS-12 input port #28 to the STS-12 output port #5 with page 0 active. The following writes would be made after polling BUSY low in the SSWT Indirect Control Address register: SSWT Indirect Data register = 0x001c, SSWT Indirect Control Address register = 0x2a05. After the active page switch has been performed, the following two writes would occur (simply inverting the polarity of the PAGE bit in the indirect address register): SSWT Indirect Data Register = 0x001c, SSWT Indirect Control Address register = 0x0a05.

7.4.3 STS-1 Timeslot Mappings in the CHESS set

The following tables provide the STS-1 Timeslot mappings from the line to TeleCombus Interface for various CHESS devices assuming the TSI is bypassed. This mapping information is important to properly execute the time portion of the switching before the data gets to the TSE for instance.

Table 3 SPECTRA-2488 Timeslot Mappings

Spectra-2488 (1x2488 mode)												
Line (Time Slots) #1:12	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
x,y: (Byte x of STS-12 # y)	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4
Line (Time Slots) #13:24	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
x,y: (Byte x of STS-12 # y)	4,1	4,2	4,3	4,4	1,5	1,6	1,7	1,8	2,5	2,6	2,7	2,8
Line (Time Slots) #25:36	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
x,y: (Byte x of STS-12 # y)	3,5	3,6	3,7	3,8	4,5	4,6	4,7	4,8	1,9	1,10	1,11	1,12
Line (Time Slots) #37:48	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
x,y: (Byte x of STS-12 # y)	2,9	2,10	2,11	2,12	3,9	3,10	3,11	3,12	4,9	4,10	4,11	4,12
Spectra-2488 (1X2488 or 4 x 622 mode) TeleCombus Mappings												
TeleCombus #1	**1**	**2**	**3**	**4**	**17**	**18**	**19**	**20**	**33**	**34**	**35**	**36**
x,y: (Byte x of STS-12 # y)	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10	1,11	1,12
a, b: (Byte a of STS-3 # b)	1,1	2,1	3,1	4,1	1,2	2,2	3,2	4,2	1,3	2,3	3,3	4,3
TeleCombus #2	**5**	**6**	**7**	**8**	**21**	**22**	**23**	**24**	**37**	**38**	**39**	**40**
x,y: (Byte x of STS-12 # y)	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	2,10	2,11	2,12
TeleCombus #3	**9**	**10**	**11**	**12**	**25**	**26**	**27**	**28**	**41**	**42**	**43**	**44**
x,y: (Byte x of STS-12 # y)	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	3,10	3,11	3,12
TeleCombus #4	**13**	**14**	**15**	**16**	**29**	**30**	**31**	**32**	**45**	**46**	**47**	**48**
x,y: (Byte x of STS-12 # y)	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9	4,10	4,11	4,12
Spectra-2488 (4x622 mode)												
Line 1	1	2	3	4	5	6	7	8	9	10	11	12
Line 2	13	14	15	16	17	18	19	20	21	22	23	24
Line 3	25	26	27	28	29	30	31	32	33	34	35	36
Line 4	37	38	39	40	41	42	43	44	45	46	47	48

Table 4 SPECTRA-4x155 Timeslot Mappings

Spectra-4x155												
Line1	1	2	3									
Line2	4	5	6									
Line3	7	8	9									
Line4	10	11	12									
TeleCombus Mapping (SPECTRA 4 x 155)												
TeleCombus	1	4	7	10	2	5	8	11	3	6	9	12
a, b: (Byte a of STS-3 # b)	1,1	2,1	3,1	4,1	1,2	2,2	3,2	4,2	1,3	2,3	3,3	4,3

Table 5 S/UNI-MACH48 Timeslot Mappings

S/UNI-MACH-48 TeleCombus Mapping (External Mapping)												
TeleCombus #1	1	2	3	4	5	6	7	8	9	10	11	12
x,y: (Byte x of STS-12 # y)	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10	1,11	1,12
a, b: (Byte a of STS-3 # b)	1,1	2,1	3,1	4,1	1,2	2,2	3,2	4,2	1,3	2,3	3,3	4,3
TeleCombus #2	13	14	15	16	17	18	19	20	21	22	23	24
x,y: (Byte x of STS-12 # y)	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	2,10	2,11	2,12
TeleCombus #3	25	26	27	28	29	30	31	32	33	34	35	36
x,y: (Byte x of STS-12 # y)	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	3,10	3,11	3,12
TeleCombus #4	37	38	39	40	41	42	43	44	45	46	47	48
x,y: (Byte x of STS-12 # y)	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9	4,10	4,11	4,12
Internal Mapping of the S/UNI-MACH-48 for OC48c*												
TeleCombus #1	1	13	25	37	5	17	29	41	9	21	33	45
TeleCombus #2	2	14	26	38	6	18	30	42	10	22	34	46
TeleCombus #3	3	15	27	39	7	19	31	43	11	23	35	47
TeleCombus #4	4	16	28	40	8	20	32	44	12	24	36	48
* The S/UNI-MACH-48 uses a different set of mappings internal to the device in OC-48c mode.												
For modes other than OC-48c, the external mapping is used												

8 SYSTEM POWER AND PACKAGE CONSIDERATIONS

8.1 CHESS SET Design Reference

Table 6 DEVICE SUMMARY

Device	Power (max.)*!	Supply	Package Information	Interfaces** Line Side / System Side
TBS PM5310 (TeleCombus Serializer)	4.2W	1.8 V 3.3 V	352 UBGA 27x27mm 1mm pitch θjc = 1.8	4x8x77 MHz TeleCombus (TTL) 3 sets of 4 x 777 MHz (LVDS)
TSE PM5372 (40Gb STS-1 Cross Connect)	13 W	1.8 V 3.3 V	560 UBGA 40x40mm 1mm pitch θjc = 1.2	64 x 777 MHz (LVDS) 64 x 777 MHz (LVDS)
SPECTRA-2488 PM5315 (SONET/SDH terminating device for 1xOC-48 or 4xOC-12)	6.6 W	1.8 V 3.3 V	520 SBGA 40x40mm 1.27mm pitch θjc = 1.0	16 bit x 155 MHz (PECL) in STS-48 mode or 4x8x77 MHz (TTL) in 4 x STS-12 mode 32 bit x 77 MHz TeleCombus (TTL) or 4x8x77 MHz TeleCombus (TTL)
SPECTRA-622 PM5313 (1xOC-12 SONET/SDH terminating device)	4 W	3.3 V	520 SBGA 40x40mm 1.27mm pitch θjc =1.0	8 bit x 77MHz (TTL) or serial 622 MHz (PECL) 8 x77 MHz TeleCombus (TTL) or 32 bit x 19.44 MHz TeleCombus (TTL)
SPECTRA-4x155 PM5316 (4xOC-3 SONET/SDH terminating Device)	3.9 W	3.3 V	520 SBGA 40x40mm 1.27mm pitch θjc =1.0	4x155 MHz (PECL) 8x77 MHz TeleCombus (TTL)
S/UNI-MACH48 PM7390 (OC-48 ATM/POS terminating device)	6.5 W	1.8 V 3.3 V	560 UBGA 40x40mm 1mm pitch θjc = 1.2	2 sets of 4x777 MHz (LVDS) or 4x8x77MHz TeleCombus (TTL) 32 bit x 104 MHz (UL3/PL3)
S/UNI-ATLAS3200 PM7325 (OC-48 S/UNI ATM Layer Solution)	4.4W	1.5V 2.5V 3.3V	576 TBGA 40x40mm 1.27mm θjc =TBD	32 bit x 104 MHz UL3 or 32 bit x 104 MHz PL3 32 bit x 104 MHz UL3 or 32 bit x 104 MHz PL3

* Maximum power (high voltage, full functionality). Typical Power ~ 90% of max.
! Consult datasheet for most up to date power numbers
**Exact rates are: 77 = 77.76, 777=777.6, 155=155.52, 622=622.08

8.2 Analog and Digital Power By Rail

Table 7 Analog and Digital Power By Rail*

Device		TSE[2]	TBS	S/UNI-MACH48[4]	SPECTRA-2488[5]	SPECTRA-4x155
Current	$I(AVDL_{@1.8V})$	965	225	210	505.8	-
Typ.	$I(AVDH_{@3.3V})$	1180	230	160	307.2	124
(mA)	$I(VDDI_{Total@1.8V})$ [3]	3655	1310	2450	1657.7	-
	$I(VDDO_{@3.3V})$	0.001	135	240	366.9	938
Power	$P(AVDL_{@1.8V})$	1737	405	378	910.5	-
Typ.	$P(AVDH_{@3.3V})$	3894	759	528	1013.6	409
(mW)	$P(VDDI_{Total@1.8V})$	6574	2538	4410	2983.1	-
	$P(VDDO_{@3.3V})$	0.005	446	792	1210.9	3096
	Total Power	**12210.0**	**3968**	**6108**	**6188.8**	**3505**
Power	$P(AVDL_{@1.89V})$	1824	425	397	956	-
Max.	$P(AVDH_{@3.63V})$	4283	835	581	1115	450
(mW)	$P(VDDI_{Total@1.89V})$	6908	2476	4631	3133	-
	$P(VDDO_{@3.63V})$	0.005	490	871.2	1332	3405
	Total Power	**13015**	**4226**	**6479**	**6537**	**3855**

* Consult datasheet for most up to date power numbers

1. 1.8V supplies have a ± 5% tolerance; 3.3 V supplies have a ± 10% tolerance
2. TSE I(VDDO) current extremely low - single (non-CBI/non-JTAG) digital output w/ low switching frequency
3. I(VDDITotal@1.8V) = I(VDDIAnalog@1.8V) + I(VDDIDigital@1.8V)
4. Assuming both serial and parallel TeleCombus is active
5. Assuming only 21 RX PECLs, and 18 TX PECLs

8.3 Power Sequencing Rules for 1.8V and 3.3V Supplies:

This section describes the power-sequencing for the device power pins.

Table 8 Power Sequence Rules

Technology	Power Sequencing		Relevant PMC-Sierra - Part
0.35 µm	First	VDDO, QAVDPO	SPECTRA-622 SPECTRA-4x155
	Second	VDD, AVD	
	Third	Data to I/Os	
0.18 µm	First	VDDO no later than AVDH and CSU_AVDH	SPECTRA-2488, S/UNI-MACH48, TBS, TSE
	Second*	VDDI, AVDL	
	Third	Data to I/Os	
0.18 µm	First	VDD33	S/UNI-ATLAS3200
	Second#	VDD25, VDDQ25	
	Third	VDD15, VDDQ15	

* The 3.3 V and 1.8 V supplies can come up at the same time as long as there is never more than 0.3V difference between the 1.8V supply and the 3.3V supply. There are no power up ramp rate restrictions.

During power-up, VDD33 must be brought up before or at the same time as VDD25 and VDDQ25, which must be brought up before or at the same time as VDD15 and VDDQ15. The VDD33 and VDD25 power must be applied before input pins are driven or the input current per pin be limited to less than the maximum DC input current specification (20mA). Power down the device in the reverse sequence.

8.4 Estimating System Power

The following formulas are useful for estimating device power when powering down certain modes of operation.

TBS Power (W) = Core Power + n_{RX} x Receive Power + n_{TX} x Transmit Power

= $3.0 + 0.036n_{RX} + 0.064n_{TX}$

Definitions & Assumptions:

- Core power utilization is independent of the number of link used and hence is constant
- The TBS has 3 groups of 4 RX/TX LVDS 777 MHz links (working, protect, auxiliary)

- n_{RX} and n_{TX} can have values of 1 through 12 dependent on the number of links
- The values of n_{RX} and n_{TX} can be chosen independently

TSE Power (W) = Core Power + α x TX CSU Power + n_{RX} x Receive Power + n_{TX} x Transmit Power

$$= 4.8 + 0.45\alpha + 0.036n_{RX} + 0.064n_{TX}$$

Definitions & Assumptions:

- Core power utilization is independent of the number of link used and hence is constant
- The TSE has 64 transmit and receive LVDS 777 MHz links
  - n_{RX} and n_{TX} can have values of 1 through 64 dependent on the number of links
  - The values of n_{RX} and n_{TX} can be chosen independently
- There is one transmitter CSU (clock synthesis and reference) block for each 16 transmit interfaces
  - The TSE has 4 CSU blocks; therefore α is integer multiple between 1 and 4
- Note that it is possible, by using only 4 TX LVDS links, that all CSUs will be powered if each of the four TX link's clock source is from each of these 4 different CSUs. That is if the LVDS TX links number 1, 17, 33, 49 are used then all 4 CSUs will need to be powered. However if TX links 1 through 16 are used, then only one CSU is powered.

S/UNI-MACH48 Power (W) = Core Power + θ x Parallel Bus + β x DS3 + ϕ (Serial Bus Power1)

$$= 4.7 + 0.35\theta + 0.2\beta + \phi\,(0.45 + 0.036n_{RX} + 0.064n_{TX})$$

Definitions & Assumptions:

- Core power utilization is independent of the number of link used and hence is constant
  - ATM and packet mode are always powered
- The DS3 block can be powered down; thus β can have a value of 1 or 0
- The parallel bus can be powered down; thus θ can have a value of 1 or 0
- The serial LVDS can be powered down; thus ϕ can have a value of 1 or 0
  - Even with the serial bus powered, individuals links can be powered down
  - The serial LVDS bus has 8 TX/RX links (4 working, 4 protect)
    - n_{RX} and n_{TX} can have values of 1 through 8 dependent on the number of links

[1] Serial Bus Power (W) = TX CSU Power + n_{RX} x Receiver Power + n_{TX} x Transmitter Power

- The values of n_{RX} and n_{TX} can be chosen independently

8.5 Maximum CHESS SET System Power

The following power calculations are based on the TSE Fabric Cross Connect as in Figure 7b, using SPECTRA-2488, TBS with Auxiliary Links Powered Down, Spare TSE and S/UNI-MACH48 devices for Layer 2. All devices connected to the TSE are operated bi-directionally. These powers represents maximum power consumption and are based on sections 8.2 and 8.3 above. The shaded portion of the table represents systems that could be more power efficiently produced with smaller fabrics. For example, a 160 Gb fabric with 25% link utilization is better constructed by using a 40 Gb fabric with 100% link utilization. In this case, the 40 Gb fabric power consumption is 275.5 W versus 313.9 W for the larger fabric. However this 14% power tradeoff may be acceptable since the system is future proofed for a four times seamless fabric expansion.

Table 9 System Power versus Link Utilization

System Power (W)	Power (100% Link Utilization)	Power (75% Link Utilization)	Power (50% Link Utilization)	Power (25% Link Utilization)
40 Gbps Cross Connect	291.6	221.9	152.2	82.5
80 Gbps Cross Connect	583.2	443.8	304.4	165.0
160 Gbps Cross Connect	1166.4	887.5	608.7	330.0
320 Gbps Cross Connect	2749.3	2138.2	1527.9	917.5

8.6 CHESS Thermal Information

The following section provides information on the PMC-Sierra CHESS devices to assist in thermal calculations and choosing appropriate heat sinks. The CHESS devices include the SPECTRA-2488, SPECTRA-4x155, S/UNI-MACH48, TBS and TSE. Please refer to the specific datasheets on each of these devices for further information. This section is strictly informative.

8.6.1 Heat Sinking:

The amount of heat that can be removed with a specific temperature difference between the heat sink and the air determines the heat sinks performance. It is most often characterized by thermal resistance, i.e., the lower the thermal resistance, the better the performance. Heat-sink performance can be improved by increasing the physical size of the heat sink (i.e., changing the surface area) or by moving more air across the sink (i.e., changing from natural convection to forced convection). As a general rule, to reduce the thermal resistance by 50 percent, the heat-sink volume must be approximately quadrupled.

The function of a heat-sink is to protect the semiconductor from the heat it produces as a by-product of its normal operation. Depending on environmental conditions, failure to remove this heat may cause the semiconductor device to exceed its safe operating temperature. In this circumstance, the semiconductor's performance, life, and reliability may be reduced. The objective, then, is to hold the junction temperature below the temperature allowed by the device reliability criteria.

The CHESS devices are guaranteed to work at 120°C junction temperature (transient condition). For long-term reliability, it is recommended to keep the junction temperature below 105°C.

The selection of a heat sink requires knowledge of:

(1) The device configuration (package size, orientation);

(2) The power dissipated by the device;

(3) The maximum allowable device junction temperature

(4) The available volume of space to be occupied

(5) Ambient conditions (temperature, air velocity).

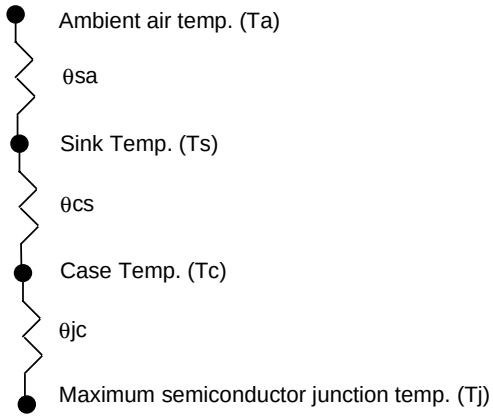
A heat-sink mounted on a semiconductor has three major thermal resistances. The total resistance θtot, is the sum of these individual resistances:

$$\theta tot = \theta jc + \theta cs + \theta sa$$

Thermal resistance from junction to case. θjc. Expressed in °C/W, this resistance is a function of design and manufacturing methods and is specified in the device's datasheet. Thermal resistance from case to sink. θcs. Expressed in °C/W, this is a variable that can be reduced by applying a thermal grease or paste that decreases the high thermal impedance of the air gap between the case and sink.

Thermal resistance from sink to ambient. θsa. Expressed in °C/W, this resistance is used in selecting a heat-sink. The smaller the value is, the more power the device can dissipate without exceeding its junction temperature limit. Heat sinks specifications should provide θsa as a function of airflow. The thermal schematic is represented in Figure 19.

Figure 19: Thermal Schematic of Heat Dissipation



In most applications, values for all these parameters are known, and can be used as the selection criteria for the heat sink (i.e. to find a heat sink, with corresponding air flow, that provides an acceptable θsa). The follow equation can be used to calculate the required θsa, and is applicable for both natural and forced convection cooling.

$$Q * \theta sa = Tsa = (Tj - Ta) - Q (\theta jc + \theta cs)$$

where

Q	=	power dissipated, watts
Tj	=	junction temperature, °C
Ta	=	ambient air temperature, °C
θjc	=	thermal resistance from junction to semiconductor case, °C/W
θcs	=	thermal resistance from case to heat sink, °C/W
θsa	=	thermal resistance from surface of heat sink to ambient air, °C/W
Tsa	=	temperature difference between sink and air

8.6.2 Example, using the S/UNI-MACH48:

Total Power Consumption = 6.5W.

The power consumption assume 3.6V (high voltage on 3.3V rail) and 1.89V (high voltage on 1.8V rail), fully loaded and worst case process. A tighter power supply

(i.e. VDD max = 3.45, 1.9V) will allow you to de-rate the power consumption by the following ratio: V2(new) / V2(VDD max).

Q = 6.5W

Tj = 105°C (fixed; for long term reliability)

Ta = 70°C (variable)

θjc = 1°C/W (fixed, see Thermal Information in Datasheet)

θcs = 0.09°C/W (variable, depending of thermal joint compound)

$Q * \theta sa = Tsa = (Tj - Ta) - Q (\theta jc + \theta cs)$

$\theta sa = [(105°C - 70°C) – 6.5W (1°C/W + .09°C/W)] / 6.5 W$

$= 4.3 °C/W$

This is the largest value of θsa that can be used in this example. Heat sinks providing smaller values of θsa are also acceptable. With this value, you can then proceed to the data presented for various heat sinks and locate one that will provide this performance.

8.6.3 Chart of θja versus Air Flow for CHESS Chipset

Table 10 θ_{ja} versus Forced Air

Package	Part		Forced Air (Linear Feet per Minute)					
		(deg C/Watt)	Conv	100.0	200.0	300.0	400.0	500.0
352 UBGA	PM5310 TBS	Dense Board	24.9	22.9	21.5	20.7	20.3	20.2
		JEDEC Board	16.0	15.2	14.7	14.4	14.2	14.0
560 UBGA	PM5372 TSE	Dense Board	11.1	9.1	7.7	6.9	6.5	6.4
		JEDEC Board	7.2	6.4	6.0	5.7	5.5	5.2
	PM7390 S/UNI-MACH48	Dense Board	12.2	10.2	8.8	7.9	7.5	7.5
		JEDEC Board	7.7	6.9	6.5	6.2	6.0	5.7
520 SBGA	PM5315 SPECTRA-2488	Dense Board	14.4	12.4	11.0	10.2	9.8	9.7
		JEDEC Board	8.1	7.3	6.8	6.5	6.3	6.1
	PM5316 SPECTRA-4x155	Dense Board	14.2	12.2	10.8	9.9	9.5	9.5
		JEDEC Board	8.3	7.5	7.0	6.7	6.5	6.3

560 UBGA package is preliminary. θ_{ja} values subject to change.

PMC-Sierra, Inc.

CONTACTING PMC-SIERRA, INC.

PMC-Sierra, Inc.
105-8555 Baxter Place Burnaby, BC
Canada V5A 4V7

Tel: (604) 415-6000

Fax: (604) 415-6200

Document Information: document@pmc-sierra.com
Corporate Information: info@pmc-sierra.com
Application Information: apps@pmc-sierra.com
(604) 415-4533
Web Site: http://www.pmc-sierra.com

PMC-1991797 (P2)