SAE Aerospace
An SAE International Group

AEROSPACE INFORMATION REPORT

SAE AIR4288 REV. A

Issued 1992-12
Reaffirmed 2001-10
Stabilized 2012-05

Superseding AIR4288

Linear Token Passing Multiplex Data Bus User's Handbook

RATIONALE

This document has been determined to contain basic and stable technology which is not dynamic in nature.

STABILIZED NOTICE

This document has been declared "Stabilized" by the SAE AS-1A Avionic Networks Committee and will no longer be subjected to periodic reviews for currency. Users are responsible for verifying references and continued suitability of technical requirements. Newer technology may exist.

TO PLACE A DOCUMENT ORDER: **Tel: 877-606-7323 (inside USA and Canada)**
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: **http://www.sae.org**

SAE values your input. To provide feedback on this Technical Report, please visit
http://www.sae.org/technical/standards/AIR4288A

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

1. SCOPE:

This document is intended to explain, in detail, the rationale behind the features and functions of the AS4074, Linear, Token-passing, Bus (LTPB). The discussions also address the considerations which a system designer should take into account when designing a system using this bus. Other information can be found in these related documents:

AIR4271 - Handbook of System Data Communication
AS4290 - Validation Test Plan for AS4074

Introduction:

The AS4074 LTPB is the result of work by the Linear Implementation Task Group (LIT) of the Society of Automotive Engineers (SAE) Avionics Systems Division (ASD).

The standard was developed to address the peculiar requirements of advanced systems projected for fielding during the late 1980s and beyond. These applications may be in the areas of avionics systems as well as ground based vehicles (vetronics) and stationary systems. Among the requirements for which this bus was developed as a solution are:

- Low Data Latency
- Highly Fault Tolerant
- Distributed Control (data driven systems)
- Quick, independent reconfiguration in the event of a failure
- High throughputs in excess of 50 Mbps

In order to achieve these goals, the task group investigated existing standards, pinpointed potential problems with these protocols in the areas of real-time system requirements, and proceeded to address the problem areas with protocol solutions. In some cases, these characteristics of the AS4074 standard seem innocuous. In others, they stand out as areas which require an understanding of the performance characteristics in order to properly utilize the standard to the full capability in a system.

This handbook is intended to introduce the reader to the AS4074 standard and explain, in detail, the operation of all facets of the protocol. It seeks to explain the rationale for the specified operation of the state machine which controls the protocol and offers the user an understanding of how to apply this data bus as a solution to a systems data communications problem.

Standard Overview:

The AS4074 LTPB standard is organized with a section to provide the user with an understanding of the operation of the protocol followed by the detailed requirements which result from that operation. Following a brief discussion of the overall protocol, the Media Interface Unit (MIU) is discussed for both fiber optic and wire (coax) implementations. Next, the actual frame formats and data field definitions are discussed, followed by a discussion of station management functions. Detailed protocol operational characteristics are found in Section 5 of the document, which includes a high level state transition diagram and state transition descriptions.

1. (Continued):

This document discusses these topics in much the same order as they are presented in the standard. In addition, there is a hypothetical system design section in which the various aspects of data bus design are covered. Some of the topics addressed are: token rotation and token holding timer settings, station address assignment, and bus test considerations for design activities.

Appendix B allows the user to quickly reference between the AS4074 standard and the handbook and, conversely, between the handbook and the standard.

1.1 History of the SAE AS4074 Standard:

As early as 1979, members of the SAE who were involved in advanced aerospace systems development, recognized the need for high speed system data transfer capabilities. The advent of MIL-STD-1553 had given the designer a flexible tool for overall system integration; however, it lacked the speed and flexibility which is necessary for future systems.

SAE A-2K, the predecessor subcommittee for AS-2, was asked by representatives of the Department of Defense, to develop a standard which would address data communications requirements for the time frame of 1990 and beyond. Initial calls to industry and government for inputs to the committee led to the formation of two task groups. These two groups, the High Speed Data Bus Applications and Requirements Task Group (HART) and the Topology and Protocol Task Group (TAP) were responsible for developing the requirements for a data bus to integrate these high performance systems and select an appropriate protocol for development.

As a result of these activities, the subcommittee elected to pursue two related standards. The first, and the subject of this handbook, is the AS4074 LTPB. The other, discussed in a companion handbook, is the AS4075 high-speed ringbus (HSRB). Each of these standards offers specific benefits in areas such as throughput, fault tolerance, and fault isolation. The designer is encouraged to review the standards in light of his particular requirements and utilize the handbooks as a guide in the selection and utilization of these standards.

2. PHYSICAL LAYER CONSIDERATIONS:

2.1 Fiber Optic Implementation of AS4074:

2.1.1 Fiber Optic Topologies:

2.1.1.1 Single Central Topologies: Single central networks are characterized by a single point at which all fiber optic signals are concentrated and are arguably the simplest network for a fiber optic data bus. As can be seen from Figure 2.1.1.1-1, all stations have fiber connections to this point which is usually referred to as a "star". Single central networks can be classified as passive transmissive star networks, passive reflective star networks, and active networks.

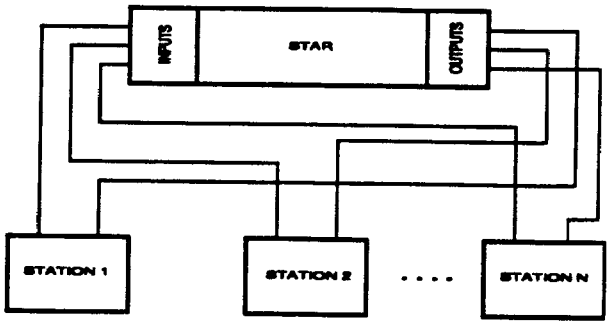


FIGURE 2.1.1.1-1 - Star-Coupled Central Network Topology

2.1.1.1 (Continued):

The main concern in passive transmissive networks is received power. The large splitting (furcation) losses in the passive star limit its usefulness to networks with a small number of stations. This is because signal power input to the star is divided among the stations; in a 64 station network, for example, each station can expect 1/64th of the power input to the star (that is, the power out of the star is down 18.1 dB) due to splitting losses. This, coupled with other system losses, quickly consumes link budget. Figure 2.1.3-4 illustrates an optimistic case for a 64 station system. Notice that, with assumed connector losses of 0.5 dB and a minimum transmitter output power of 1.75 dBm, only 4.75 dB of margin is achieved. This system would cease to function with connector losses on the order of 1 dB.

Active star systems can, to an extent, overcome the reduction in station received power due to splitting (and other) losses. This can be accomplished by the addition of an active device which amplifies the signal prior to transmission to the receiving station. The active star may simply be a repeater; in this case the signal is received by the star and amplified before being sent to the receiving stations. The active star could be a regenerator, in which case the signal is restored to its original (electrical) form and then retransmitted to the receiving stations. The latter case is considerably more complicated since the star is operating as a store-and-forward device, Note, however, that an active star can only recover the loss between the source and the star. Losses between the star and the receiving station, which may be quite large, must still be considered.

Since active stars must acquire the signal prior to passing it along to the receiving stations, any change in the signal which occurs during the acquisition process must be restored if the receiver requires that the received signal be a bit-for-bit duplicate of the transmitted signal. For example, if the receiver is AC-coupled, some number of preamble bits will be lost during signal acquisition. If the loss of these preamble bits cannot be tolerated by the system, it will be necessary to restore these bits prior to retransmission.

2.1.1.1 (Continued):

Cabling can be a drawback in single central systems since it is necessary to run an individual cable (containing one or more fibers) from each station to the star. In large networks this may be difficult to install and maintain. These cable runs may also require many bulkhead penetrations resulting in a large number of bulkhead connectors and their associated losses. These features may also make it difficult to add stations to the network.

Another consideration with single central networks is reliability. The single concentrated point for all cables (signals) means protection for a single central network is some form of redundancy.

Despite these drawbacks, single central networks have proven to be the network-of-choice for a variety of avionics programs. They feature a well understood initialization (claim token) process, and offer cost, weight, reliability, and power advantages for applications with few connectors in the link and relatively short cable lengths. Applications requiring long cable runs and perhaps numerous bulkhead penetrations, may require a distributed network in which active circuitry must be used to overcome the enormous losses experienced.

2.1.1.2 Distributed Topologies: We begin by asking ourselves how many degrees of freedom we wish to have in our generalized topology development process. Here, we will only consider nonredundant configurations. Also, we are limiting the scope to two-dimensional topologies (from a mathematical perspective). Obviously two-dimensional constructs can be put into three-dimensional space (an aircraft). Anything more complex is not likely to be attractive in an application where size, weight, power, and cost are the name of the game. So, in our two-dimensional framework, we can readily identify three parameters over which we wish to have control: the degree of activity (passive [P], hybrid, active [A]), the degree of distributivity (centralized [C] and distributed [D]), and the level of multiplicity (single [S], dual [D], triple, multiple [M]). This mentally suggests a building block approach with each block carrying a unique three-letter identifier, one for each variable. For illustration purposes, we stack these blocks to make a "high rise" whose height is limited only by the level of multiplicity we choose. Figure 2.1.1.2-1 shows this idea. For simplicity, the hybrid block has been left out and all levels above dual are grouped in the multiple category. The cornerstone is the single central passive (SCP) star-coupled network, the most elementary topology for a data bus and the baseline for all HSDB programs and standardization activities. Clearly there must be missing blocks in this building because "central" (i.e. not distributed) with higher levels of multiplicity is uninteresting at best, and "single" and "distributed" are contradictory and therefore unrealizable, at least within the traditional concept of distributed topologies. Physically a SDP topology based on an exotic star coupler can be conceptualized, but we wish to stay within existing technological bounds. The DDP and MDP topologies are generally called distributed star-coupled or DISCO buses and that terminology will be used here and will be expanded to include DDA and MDA topologies.

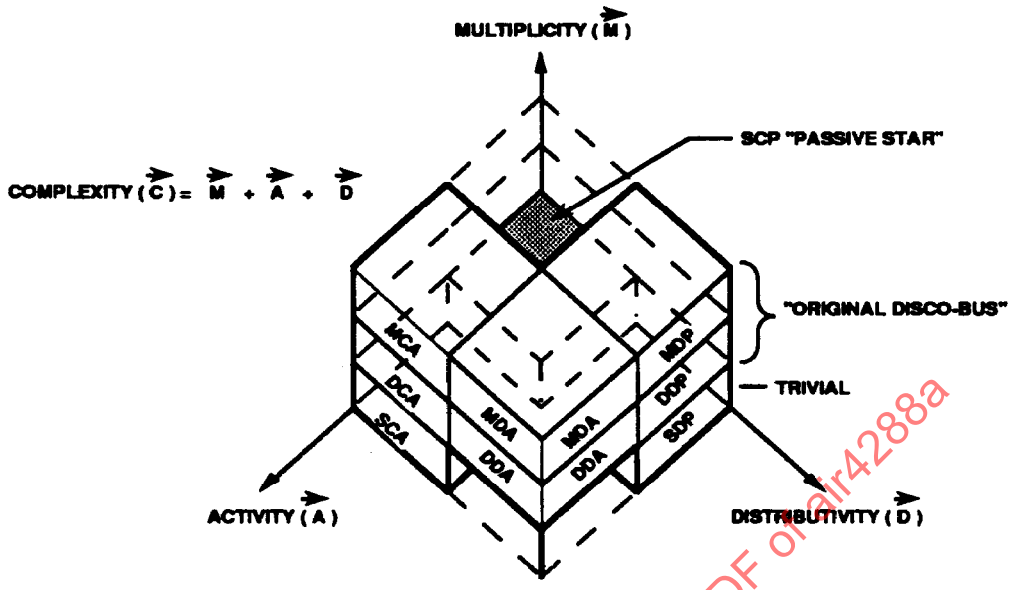


FIGURE 2.1.1.2-1 - Distributed Topology "High Rise"

2.1.1.2.1 Distributed Passive Topologies: Before passing into the world of distributed active topologies, let us look at Figure 2.1.1.2.1-1 to see how the elementary single central passive (SCP) network expands in multiplicity and distributivity. In the middle box, the star coupler has been subdivided in such a way that full connectivity is maintained, but the split stars are distributed so they coexist with their local cluster of remote terminals (RTs). Note that in the limit, as the "coupler nodes" get closer to the terminals, and the space between the coupler nodes gets larger, the amount of optical fiber needed to connect all terminals is reduced by one half. This is one of the attractive features of this distributed star-coupled (DISCO) bus. Note also that each coupler node houses two 2x2 (2 input and 2 output port) transmissive star couplers instead of one 4x4 coupler, and any RT-to-RT signal passes through exactly two 2x2 couplers. Thus the ideal (splitting only) losses are identical in both cases and there is no theoretical optical power loss budget penalty. (In practice there is a small penalty because two couplers will have more excess loss than one regardless of size.) Finally, if the coupler nodes are together in one box, this is the uninteresting DCP case.

2.1.1.2.1 (Continued):

The DISCO bus concept of distributivity can now be expanded indefinitely. In the lower box of Figure 2.1.1.2.1-1, there are four coupler nodes (M=4) so this is a 4DP topology. If symmetrical (NxN) star couplers are used, the number of terminals which can be served by the network is N^2 (16 here) and the ideal link losses are the same as those for the SCP topology with a 16x16 star coupler. This is also called a fully-connected network. There are no passive redundant paths, but if one takes advantage of the intelligence resident in the RTs, a fault-tolerant network can be constructed by virtue of multiplicity of paths available via the remote terminals. The lower illustration is called "ring-configured" because it can be laid out as a physical ring as is illustrated in more detail in Figure 2.1.1.2.1-2. From a practical standpoint, this fully-connected network is not as awkward as some believe. The number of inputs and outputs at each coupler node is less than one half that for the SCP network, and the harnesses connecting coupler nodes in the ring can be configured so they are identical (except for length perhaps) making their fabrication quite straightforward. These principles apply equally well to the active configurations and should be kept in mind when trading options.

2.1.1.2.2 Distributed Active Topologies: Now, the same methodology will be used to develop basic distributed active topologies. Figure 2.1.1.2.2-1 is analogous to Figure 2.1.1.2.1-1. The top box shows the single central active (SCA) star coupler serving four terminals. In effect, the passive star coupler has been split in the middle and a receiver/transmitter "gain block" added. The optical combiner (Σ) and distributor (Δ) functions could be 4x1 and 1x4 star couplers, respectively, or some other device or scheme which provides the same function. The R triangle is the optical receiver; the T box is the optical transmitter. This is the most elementary fiber-optic active star coupler possible and may have to be more complex after all system issues are considered.

The dual central active (DCA) topology is shown in the middle box of Figure 2.1.1.2.2-1. The number of transmitters and receivers has doubled, and the OR gates are included to illustrate the practical aspects of combining logic level signals. The two OR gates are actually redundant, but are shown symmetrically as they are to develop the distributive possibilities. It is worth noting that the optical combiner and distributor functions could be outside of the dashed lines if they are discrete devices such as 1xN and Nx1 star couplers.

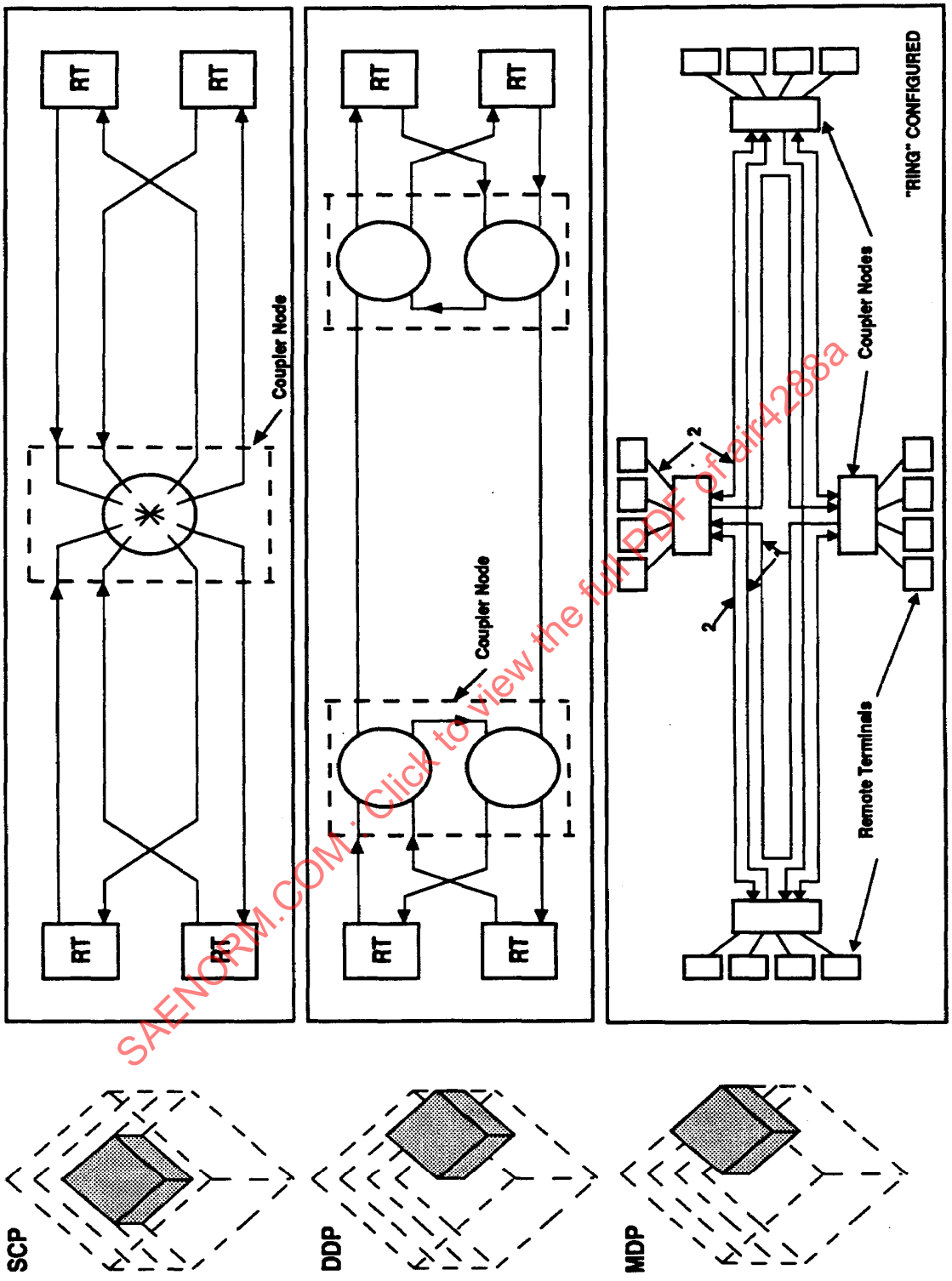


FIGURE 2.1.1.2.1-1 - Development of Distributed Passive Technologies

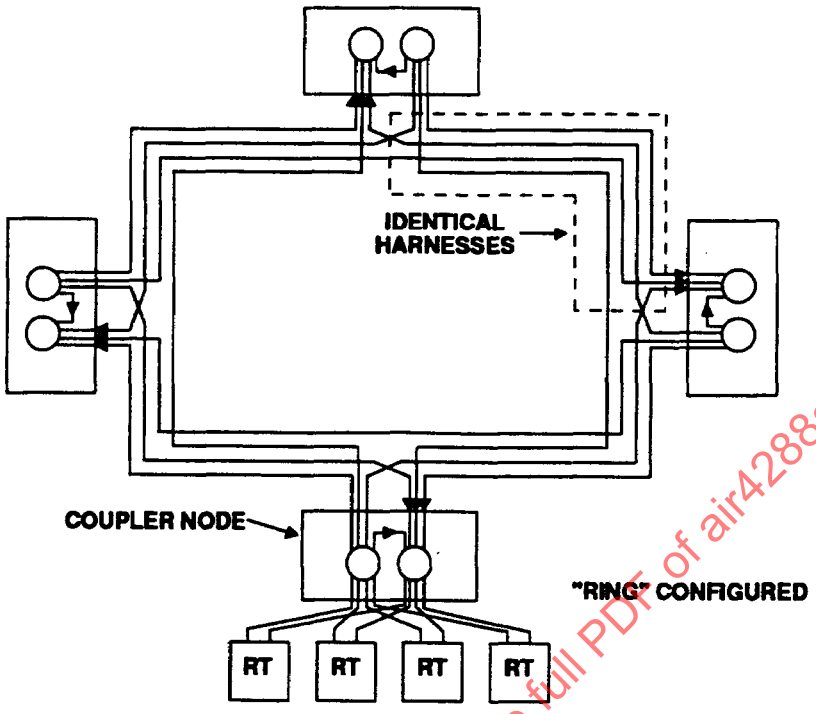


FIGURE 2.1.1.2.1-2 - A Four-Coupler-Node Passive DISCO Bus

2.1.1.2.2 (Continued):

In the lower box of Figure 2.1.1.2.2-1, the dual centralized active star coupler is now distributed to become a DDA topology. The coupler in the middle box is split vertically along the axis of symmetry and distributed. In so doing, electrical wires were cut, not optical fibers, so the two active star couplers must be linked by a transmitter/receiver pair, one for each direction. This configuration is analogous to the middle box in Figure 2.1.1.2.2-1. We must now continue to expand this concept, as we did for the passive case, to achieve multiple distributed active (MDA) topologies. In so doing, we will find it convenient to call each of the structures within the DDA dashed boxes of Figure 2.1.1.2.2-1 a "unit cell" because it will continue to show up in most distributed active configurations. The illustration at the top of Figure 2.1.1.2.2-2 is just that. The unit cell (UC) parameters listed will be defined shortly, after the unit cell has been completely generalized.

2.1.1.2.2 (Continued):

We find that we can proceed along a number of paths to further develop a family of distributed active configurations. In a manner like that done above with the passive networks, we begin by defining the active DISCO bus shown in the lower left of Figure 2.1.1.2.2-2. The basic unit cell now replaces the passive coupler node, and depending on the size of the network, the unit cell must expand beyond the basic cell shown in Figure 2.1.1.2.2-1. Specifically, the active DISCO unit cell in Figure 2.1.1.2.2-2 has been purposely shown "star-configured" for reasons that will soon become obvious. We can exploit the star-like topologies by first adding a third unit cell between the two shown at the top of Figure 2.1.1.2.2-2, and connected to the others in the same way. The center unit cell will be different than the "concentrators" at the remote terminal clusters depending on the number of terminals in a cluster. But now we can keep adding concentrators and feeding them into the central unit cell "star." We will call this configuration the "star-concentrator" topology. The illustration at the lower right of Figure 2.1.1.2.2-2 shows four concentrators and one star. The single central star can be passive or active, and if passive this is actually a hybrid configuration since there are both passive and active distribution elements in the network. Note the similarity between DISCO and star-concentrator topologies. The star-concentrator trades off fiber-optic interconnection simplicity for the added risk of one additional passive or active star coupler at the center which becomes a classic single point of failure we are trying to avoid by going to distributed configurations. Additional study of this issue is clearly warranted.

Next, we are tempted to keep cascading unit cells of some kind with those shown in Figure 2.1.1.2.2-2. When we added one, we used the middle one in a single central star coupler to create a two-dimensional structure. If, instead, we expand one-dimensionally, we create a new class of configurations that are not strictly star-like. Therefore, we refer to this class as "trunk" topologies, as shown in Figure 2.1.1.2.2-3. At the left is the active-active trunk, named because a signal must keep passing through the active cells to get to the destination. A unique characteristic of this topology is the theoretical unlimited extent to which the trunk can be expanded. (There are practical limits, of course, which must be explored.) Finally, this one-dimensional distributivity can be carried another step. As shown in Figure 2.1.1.2.2-3, the unit cells in the active-active trunk are replaced with passive trunk taps, while the unit cells are moved out of the way to act as concentrators for the clusters served by each passive trunk tap. Because of the mix of passive and active distribution elements, this is called an active-passive trunk configuration. From the point-of-view of optical fiber technology, this configuration must be viewed skeptically because of the difficulty implementing passive taps with optical fiber. However, the topology is obviously attractive for aircraft applications, and several manufacturers have developed passive in-line tap technology in the laboratory.

All possibilities have not been exhausted, but nearly anything else will be a variation on one of the themes presented such as some exotic hybrid configuration. The next step is to look into the mysterious unit cell to see from what it is made and how many unit cells are needed for all of the topologies presented.

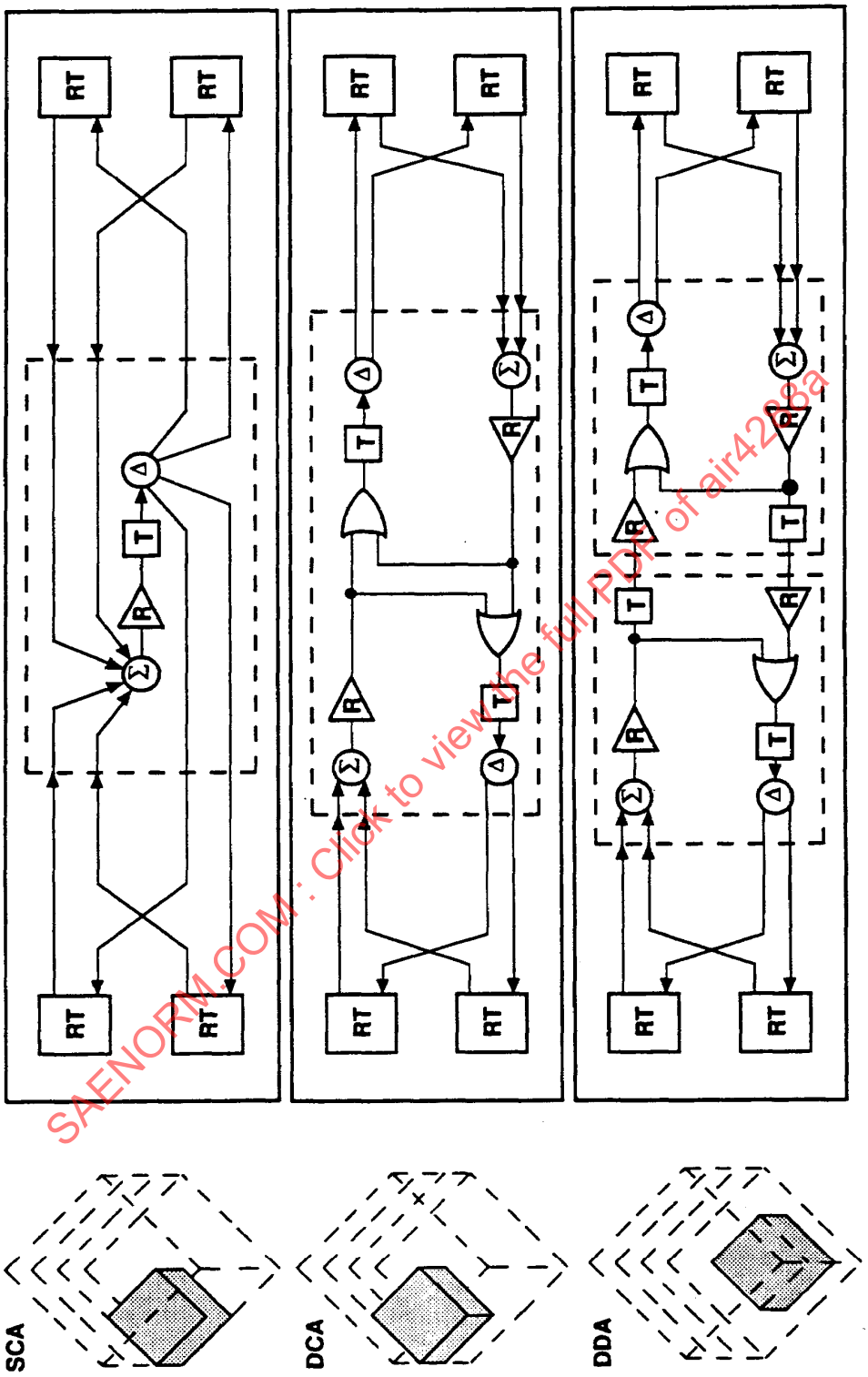


FIGURE 2.1.1.2.2-1 - Development of Distributed Active Technologies

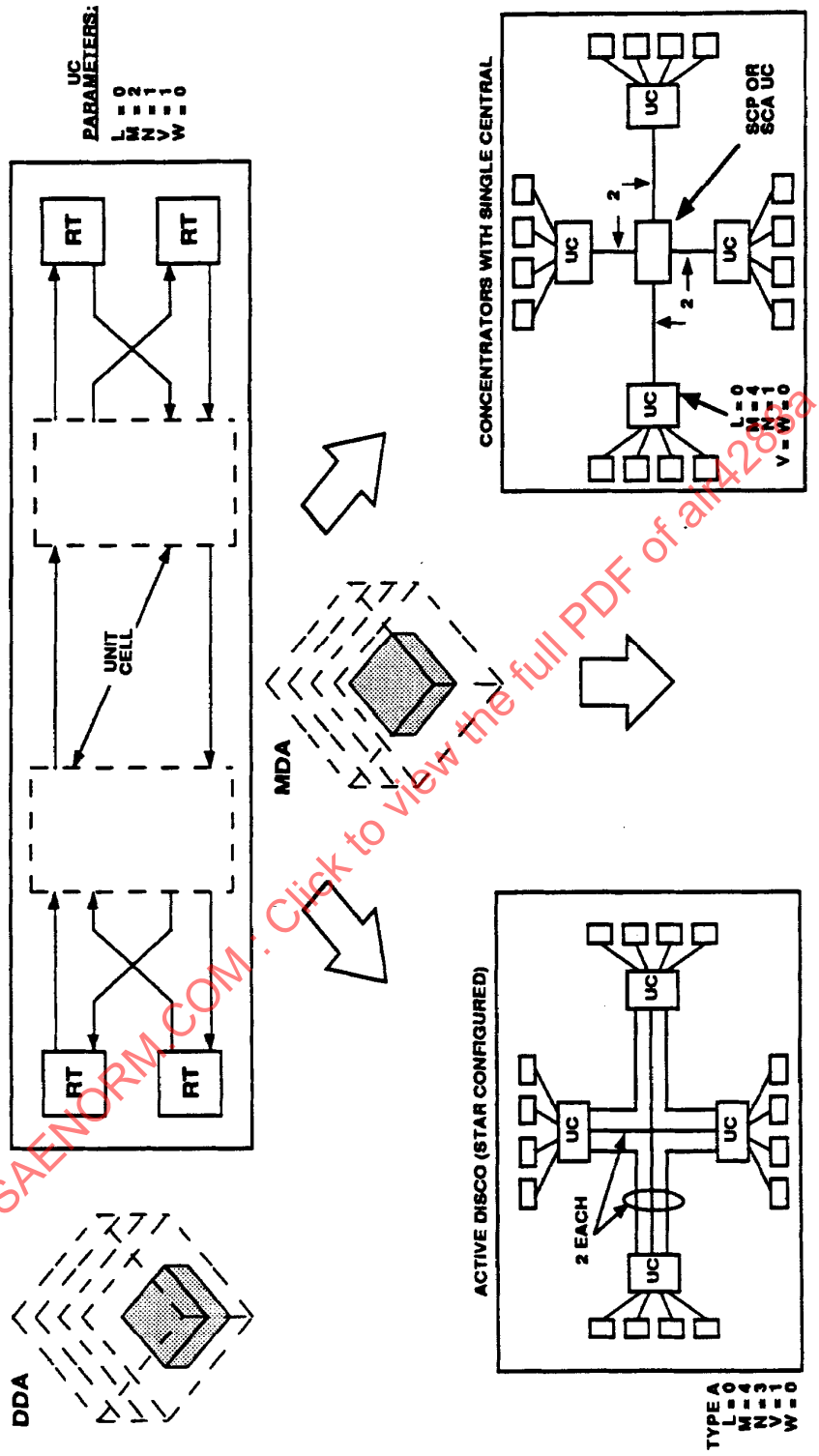
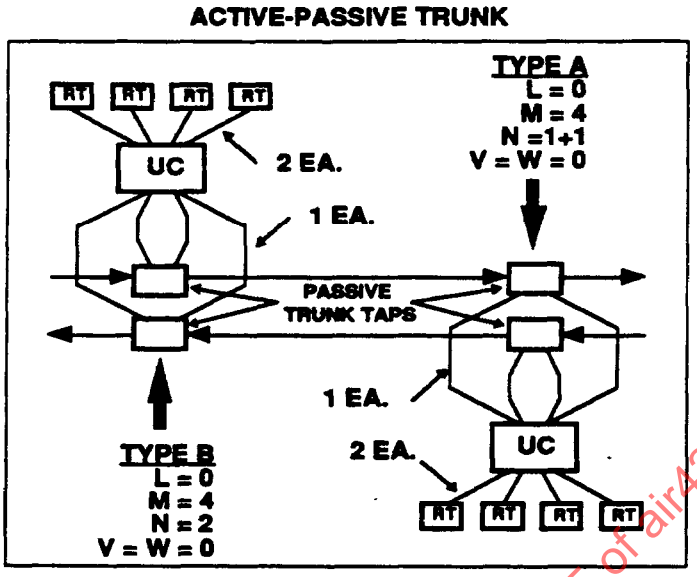


FIGURE 2.1.1.2.2-2 - Development of Multiple Distributed Active Topologies



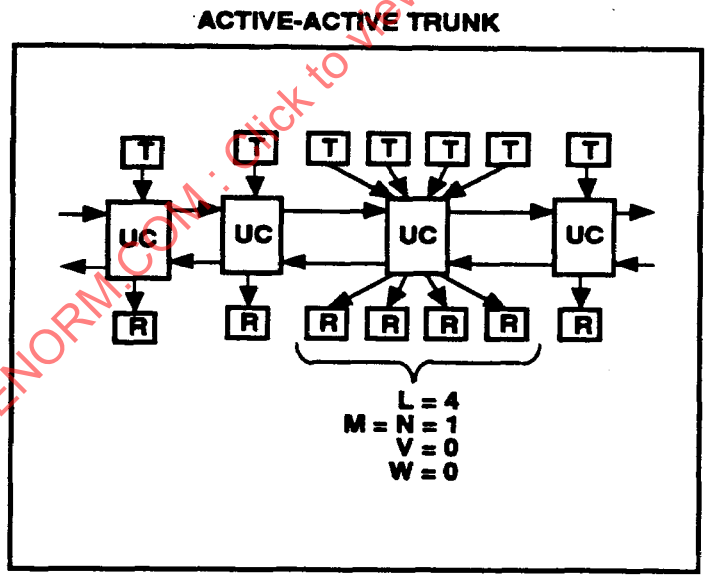


FIGURE 2.1.1.2.2-3 - Trunk-like Distributed Active Topologies

2.1.1.2.3 Universal Unit Cell: By studying the circuit requirements for all the active unit cells described, one finds that all configurations can be implemented with one unit cell logic block that is remarkably simple, highly symmetric, and requiring only two mode bits to select three states for all topologies. Figure 2.1.1.2.3-1 illustrates the "universal" unit cell. As simple as it is, the active-active trunk topology actually makes the logic block as complicated as it is because there must be three transmitters and receivers as a minimum. For universal use, the number of optical inputs and outputs required varies depending on configuration. The integers L, M, and N in the figure are those values. If the number of terminals in a cluster served by a unit cell exceeds the capacity of an individual receiver and/or transmitter, additional devices are added and their logic level (electrical) inputs and/or outputs are appropriately combined before connecting to the logic block shown in Figure 2.1.1.2.3-1. Generally receiver outputs are ORed and transmitter inputs are tied directly together only requiring an adequate fan out from the logic block. Figure 2.1.1.2.3-1 was drawn specifically with the active-active trunk (Figure 2.1.1.2.2-3) in mind. Here L is as required, and M=N=1. Figure 2.1.1.2.3-2 is an alternative representation giving it a more "universal" look. V and W are the binary mode select control lines. To use the universal unit cell, select decimal values of L, M, and N, and binary values V and W according to Table 2.1.1.2.3-1 for the configuration desired. If the dual central configuration is excluded, only one mode select line is needed. The "L" channel is only used for the active-active trunk and one version (C2) of the active-passive trunk. With these there must be at least three transmitter and receiver channels, although two of the three only have one optical input and output on the trunk channels. The two types of active DISCO buses (A and B) and six types of active-passive trunks (A through E) are described in the following sections.

TABLE 2.1.1.2.3-1 - Universal Unit Cell Implementation Table

CLASS	TOPOLOGY	TYPE	NO. OF REPEATERS	TX TO RX LINKS
SCA	SINGLE CENTRAL	-	1	2
DCA	DUAL CENTRAL	-	1	2
DDA	DUAL DISTRIBUTED	A	1 OR 2	2 OR 3
		B	2	3
MDA	DISCO	A	1 OR 2	2 OR 3
		B	2	3
	STAR-CONCENTRATOR	PASSIVE STAR	2	3
		ACTIVE STAR	3	4
	ACTIVE-ACTIVE TRUNKS	-	1,2,..., K	2,3,...,K+1
	ACTIVE-PASSIVE TRUNKS	A, B	2	3
		C1, C2, D	1 OR 2	2 OR 3
		E*	2,3,..., K	3,4,..., K+1

K = NO. OF UNIT CELLS USED
* = FOR TWO OR MORE NODES

2.1.1.2.3 (Continued):

Finally, the expandability of the unit cell, touched on earlier, and additional circuit requirements need to be discussed. Any channel that serves more terminals than can be handled by a single receiver or transmitter must have additional units per channel (L, M, or N) added as required. The way they are connected is shown in Figure 2.1.1.2.4-1. Note that if clock recovery and retiming (a clock recovery unit or CRU) is required in the active star coupler unit cell, and multiple receivers are needed per channel, there must be a CRU for each receiver in that channel, as shown. Also additional circuits may be needed to prevent a receiver from locking-up in the "one" state during an intertransmission gap thereby preventing any other receiver's output from being retransmitted. This is serious if there is no CRU because some CRU circuits can tolerate this situation. A more serious and likely case is the chattering receiver. The gain is so high on these devices that, in the absence of an input signal, such as during the intertransmission gap, the data detector (comparator) may trigger on noise and chatter. That may spoil a transmission even with clock recovery thereby disabling the entire data bus. A similar requirement is imposed on the transmitters to prevent "blabbing" in the event of a component failure. In distributed configurations, this is especially important since a component failure in a terminal cluster served by a unit cell may crash that cluster, but must not crash the entire bus, that is, the remaining clusters and active star coupler unit cells.

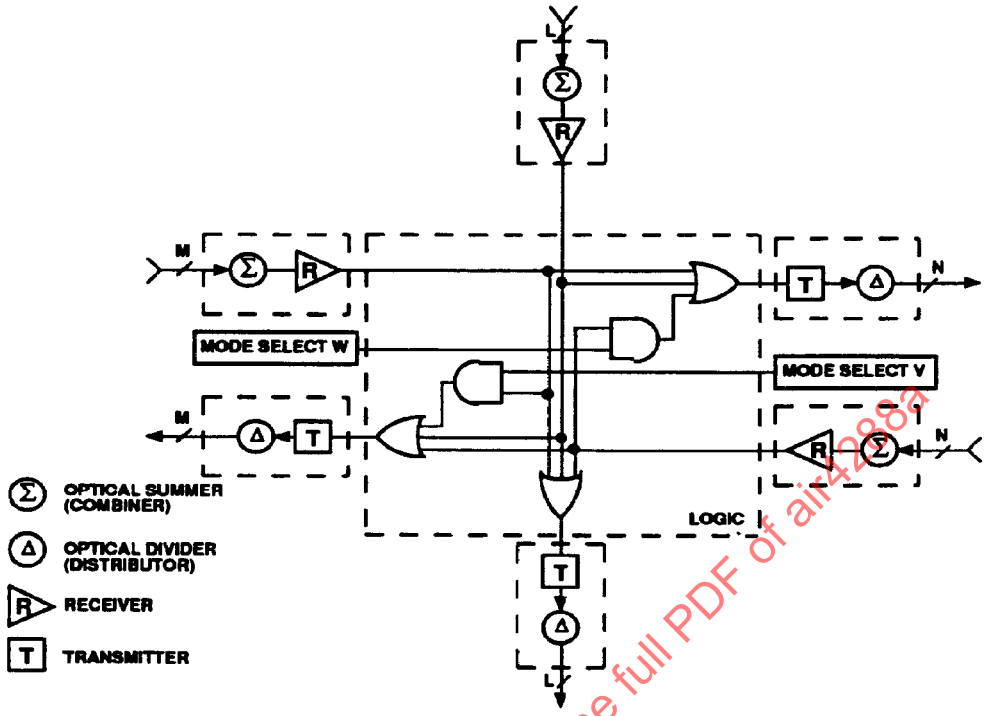


FIGURE 2.1.1.2.3-1 - Universal Cell Set

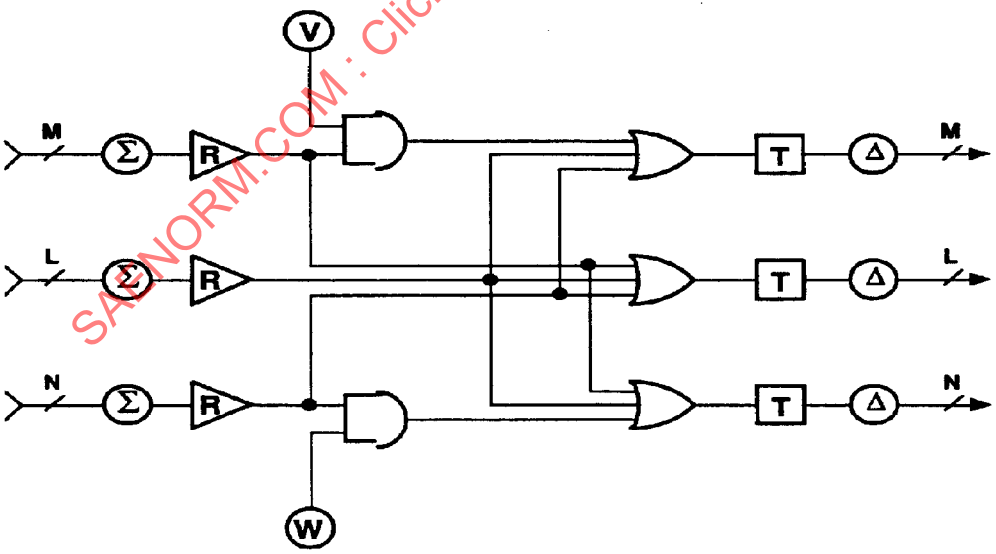


FIGURE 2.1.1.2.3-2 - Alternate Universal Unit Cell Representation

2.1.1.2.4 Single Central Active (SCA) Star Coupler: Figure 2.1.1.2.4-1 illustrates the SCA unit cell with and without expansion provisions for large numbers of terminals (>M). It is the direct replacement for the passive star coupler and the least complex of active implementations. Since all messages from all terminals sum to a single electrical path internal to the unit cell, initialization is expected to be as predictable as it is for the passive star-coupled bus. Of all the active star coupler implementations, this one is the least likely to require clock recovery and retiming circuitry. With a single receiver implementation, antilockup circuits are not required and some low-rate receiver chatter is allowed. With multiple receivers, antilockup is required and no chatter can be tolerated. Depending on the number of terminals served by the single central active star coupler, multiple transmitters/distributors may be required to satisfy link budget requirements. Finally, the SCA star coupler is a single point of failure to the network, as is the passive star coupler.

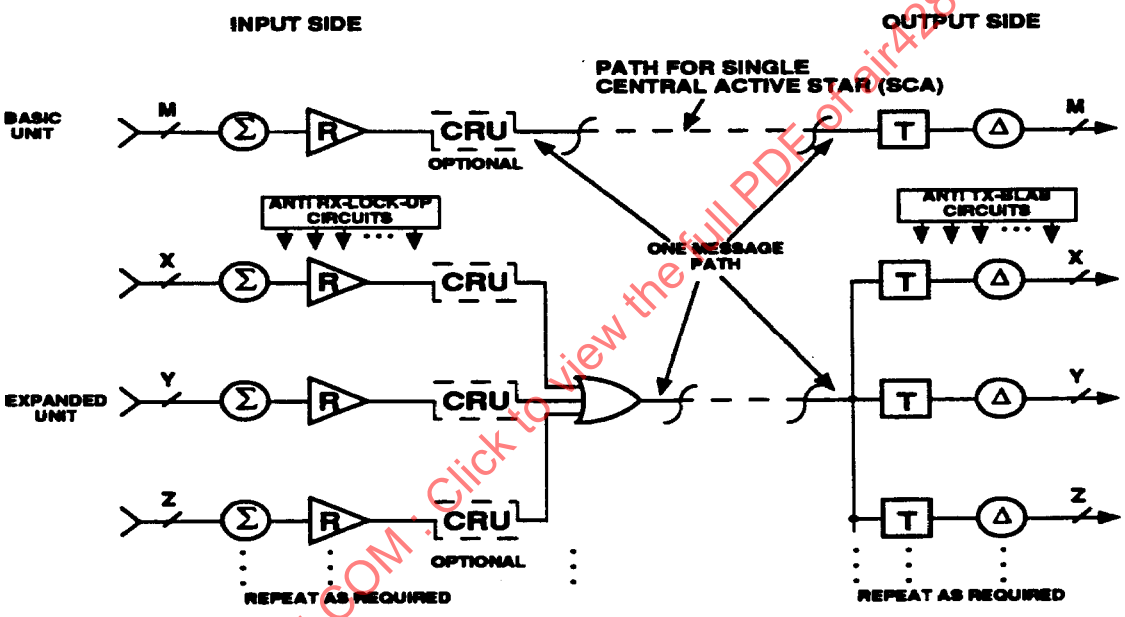


FIGURE 2.1.1.2.4-1 - Expansion and Completion of the Active Universal Cell Set

2.1.1.2.5 Distributed Star-Coupled (DISCO) Unit Cell: Figure 2.1.1.2.5-1 illustrates two implementations of a DISCO bus unit cell. Type A (see Table 2.1.1.2.3-1) conforms to the original concept introduced in 2.1.1.2.2. Type B utilizes passive star couplers to perform the local terminal loopback function optically rather than electrically as in Type A. There does not seem to be any advantage to the Type B implementation. However, Type A must have an antiblab circuit on the transmitter serving the network. Generally, networks of this type have increased fault tolerance although reliability is a sensitive function of the degree of fault tolerance employed. There are always two unit cells between any two terminals, so the requirement for retiming in the unit cell must be carefully considered. All messages follow different paths with different delay and optical attenuation, so receiver performance must be factored into bus design and the differential delay must be compatible with restrictions imposed by the protocol. Initialization may be more difficult to analyze due to the multiple paths, delays, and attenuations. Finally, fiber-optic interconnections may be more complex resulting in difficulties in maintenance in a large installation. As was pointed out in 2.1.1.2.2, this may be mitigated by incorporating a "standardized" harness to interconnect unit cells.

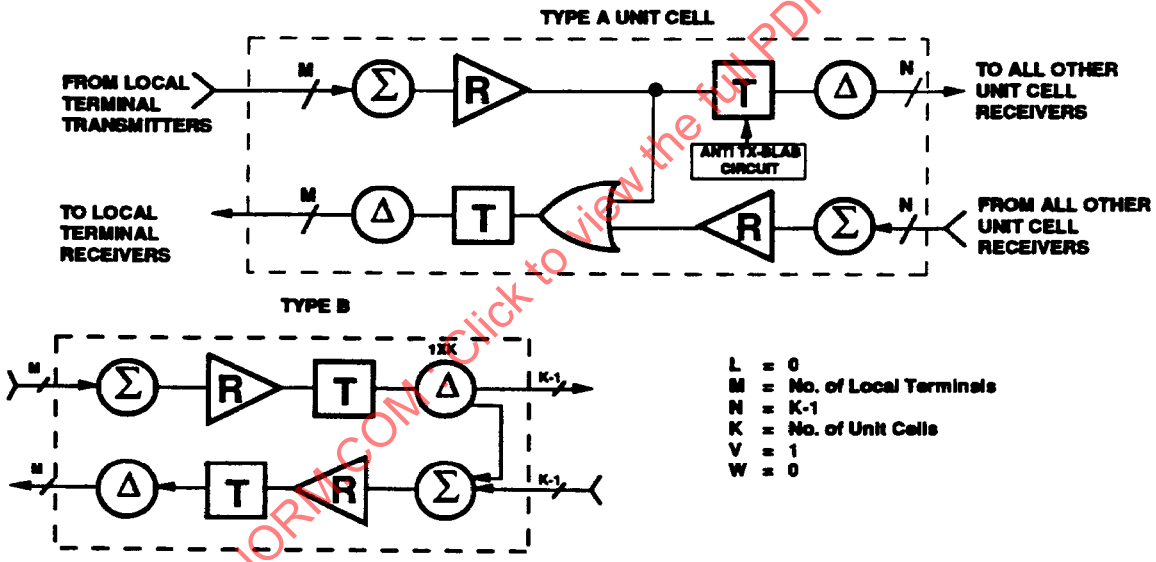


FIGURE 2.1.1.2.5-1 - Active DISCO Bus Unit Cell Implementation

2.1.1.2.6 Star-Concentrator Unit Cell: The star-concentrator configuration consists of as many concentrators as needed to serve all terminal clusters plus a single central active or passive star coupler tying the concentrators together. Figure 2.1.1.2.6-1 illustrates the concentrator unit cell and its connections to local terminals and a passive star coupler. All concentrators have only one optical input and output on the network side, and as many as are required on the local terminal side. Receiver antilockup and transmitter antiblab circuits are required to prevent crashing the entire network if a problem arises in the local terminals or its concentrator. Whether an active or passive star is used depends on the number of concentrators employed and the network optical power budget. Since all messages pass through the single central star, it is a single point of failure and thus has the same fault tolerance liability as does the SCA configuration. For the same number of terminals, the star coupler in the star-concentrator topology is smaller than a single central star coupler. This has practical benefits in terms of the number of inputs and outputs, and the size of the connector(s) needed to implement the star coupler.

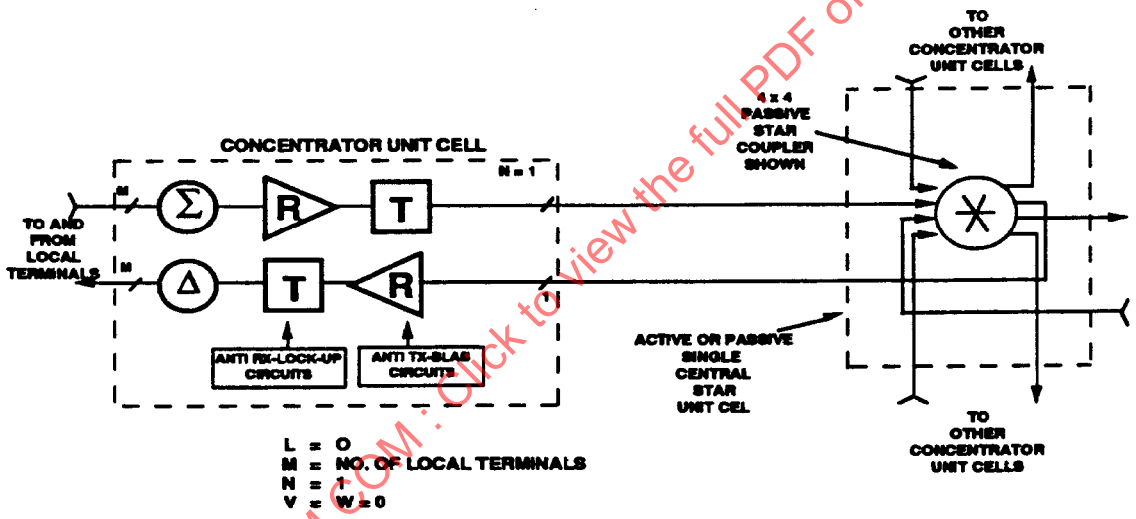


FIGURE 2.1.1.2.6-1 - Star-Concentrator Cell Unit Implementation

For the SCA topology, the initialization process is relatively easy to analyze here because all signals pass through a single "thread," as it were. With an active central star coupler, all signals pass through three cascaded stages of amplification, thus the retiming issue must be carefully considered. With a passive central star coupler, only two repeater stages are used, but the optical power budget may not be satisfied. An analysis of the network in a specific application must be conducted to determine the best technical approach.

2.1.1.2.7 Active-Active Trunk Unit Cell: This is the most complex of the unit cells and drives the complexity of the universal unit cell (2.1.1.2.3). This is because there must be at least three receivers and transmitters as shown in Figure 2.1.1.2.7-1. The number could increase if more are needed to serve a local terminal cluster. Note that two of the three receivers and transmitters have only one input or output; they serve the trunk.

The obvious advantage of this network is the minimal amount of cabling needed. If the clusters of local terminals are closely spaced and adjacent to the unit cell, the only interconnect of consequence is the cable pair connecting unit cells. This could be a significant advantage for large installations. Furthermore, the network is easy to expand off the ends. Only two cables need to be connected to an existing unit cell. Remember, though, that this is not a ring network and the trunk cannot be looped in a ring and closed. It will simply oscillate at a frequency related to the length of the ring. The large number of repeaters that are possible, and the way they simply connect, means very large optical power link margins (allowed loss budget minus all losses) can be achieved. Care must be exercised, however, to ensure that in the best case, receiver saturation or overload does not occur. This topology may only be practical if high link losses between unit cells can be guaranteed.

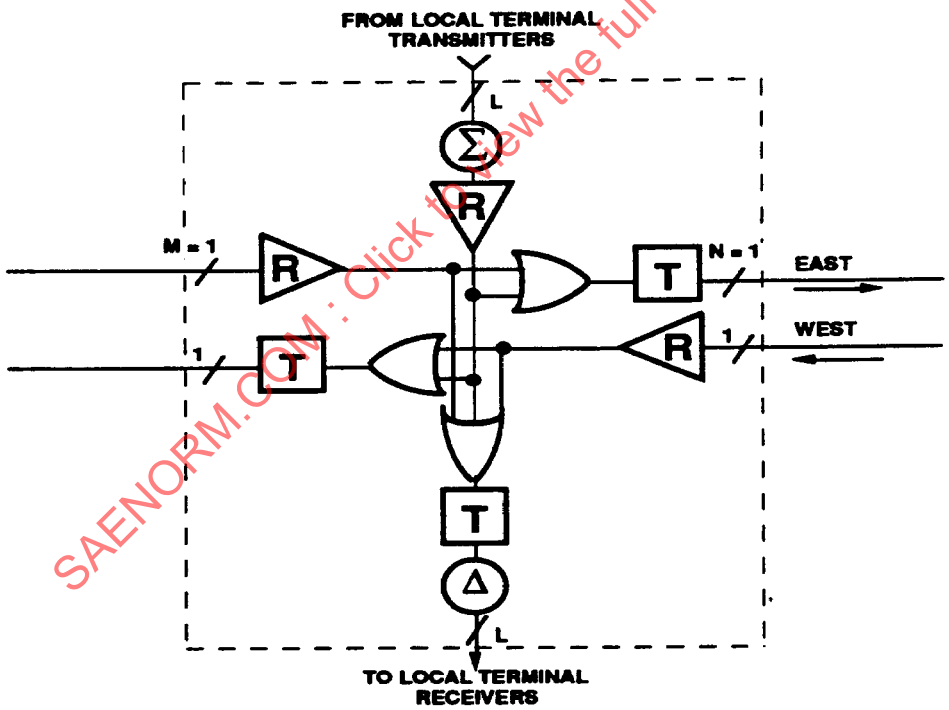


FIGURE 2.1.1.2.7-1 - Active-Active Trunk Unit Cell Implementation

2.1.1.2.7 (Continued):

Note that there is no single point of failure in this network, but the failure of a unit cell node is likely to create two separate networks which cannot communicate. This can be an important fault tolerance feature, or it may be disastrous depending on specific configuration features. Another potential disadvantage lies in the ability to expand the network so readily. If the number of repeaters is too large, the signal distortion buildup may force the designer to retime the signal in each unit cell. This will add significant throughput delay for the longest paths and create large differential delays for all signals. This may have a significant impact on overall bus performance including the possibility of failing to fall within the protocol timing boundaries. This effect must be carefully considered.

Finally, the initialization behavior of this network is complex and exceedingly difficult to analyze because of the wide range of values of the parameters that are important in this process.

2.1.1.2.8 Active-Passive Trunk Options: An entire family of active-passive trunk configurations can be constructed depending on the technology used to implement the passive trunk taps. Three tap options will be presented and some of their features discussed. Six unit cells are needed to cover all cases, but they can be implemented with the universal unit cell described in 2.1.1.2.3 and itemized in Table 2.1.1.2.3-1.

Some general features of the three variations of active-passive trunk topologies are given next. The trunk configuration is attractive for large aircraft. (This includes the active-active trunk, too.) Because the trunk coupler is passive, a failure in a terminal cluster or its unit cell does not mean the network will be subdivided. Compare this result with the active-active trunk case. This is an important plus for the topology. Of course, low-loss tap or access couplers must be used for this network to work. The technology is relatively undeveloped so risk is much higher here than it is with any other topology presented so far. At least two companies are actively developing this technology for the commercial aircraft industry. Whether their development is suitable for the military aircraft environment is yet to be determined. In any event, the technology just does not allow moderate to large link margins to exist. This may jeopardize the utility of the technology for high-speed data bus applications because of the lower optical power budget available at high data rates, and low link margins.

Again, initialization is difficult to analyze with several of the implementations because of the large optical signal range the receivers must handle. The size of the network determines the degree of difficulty in analyzing this problem.

2.1.1.2.8.1 Type A and B Active-Passive Trunk Unit Cells: The tap used in this implementation is the kind currently being developed for commercial applications and may be the most likely to be usable here. It uses optical fibers of different sizes which are specially prepared and permanently attached to the trunk fiber at predetermined locations. The result is a very low trunk throughput loss which is absolutely essential if more than a few taps are used. However, the tap-on losses are moderate and tap-off losses quite high, and this is where the problem lies.

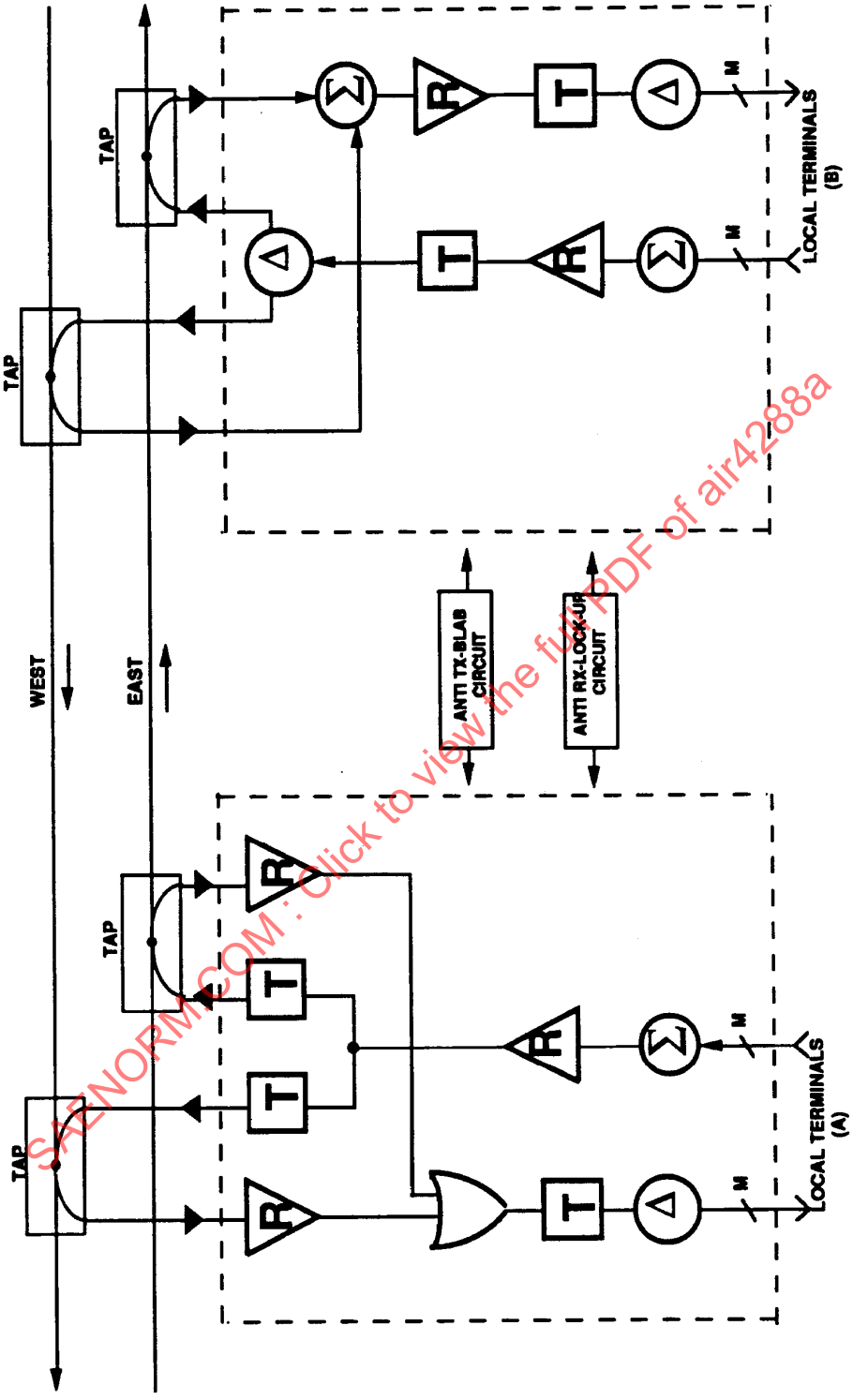


FIGURE 2.1.1.2.8.1-1 - Type A and B Unit Cell Implementation

2.1.1.2.8.1 (Continued):

Figure 2.1.1.2.8.1-1 illustrates this bus with two applicable types of unit cells we call Type A and B. Both an eastbound and westbound trunk are needed to serve all terminals on the bus so there are two passive access couplers serving each unit cell. Loopback for local terminal service must occur on the bus at the taps or there will be difficulties. There are two unit cell options available with the choice depending on optical power budget considerations mostly. Type A uses separate transmitters and receivers for eastbound and westbound traffic, whereas Type B uses only one of each to get on and off the bus and relies on a simple 2:1 optical combiner and divider to make two-way traffic possible. The number of local terminal receivers and transmitters is determined by the number of terminals served in the cluster as it is for all other active star coupler unit cells. Antilockup and antiblab circuits are needed as before to keep the network up if there is a local terminal or unit cell difficulty.

2.1.1.2.8.2 Type C and D Active-Passive Trunk Unit Cells: The tap used with these unit cells is hypothetical although the technology described in the previous section is applicable. Effectively, each tap consists of two single taps spliced back-to-back which insert and extract signals from the bus in the reverse order from that discussed above. Again an eastbound and westbound trunk is required which is never closed (no ring must be created). The greatest difference is local terminal loopback cannot occur on the bus and must be done inside the unit cell. This is an advantage in that local service can be preserved in the event of a bus or tap failure. Of course, one must trade the risk of tap failure against unit cell failure to determine the value of this approach. The Type C unit cells use multiple transmitters and receivers to access the bus; Type C1 is illustrated in Figure 2.1.1.2.8.2-1. Type C2 is identical to the active-active trunk unit cell shown in Figure 2.1.1.2.7-1, but is obviously applied differently. As before, antilockup and antiblab circuits are needed. Type D uses a single transmitter and receiver, and a 2:1 combiner and divider, like Type B, to access eastbound and westbound traffic.

2.1.1.2.8.3 Type E Active-Passive Trunk Unit Cell: Figure 2.1.1.2.8.3-1 shows the last variation on the active-passive trunk configuration. A simple tap is used to access the bus. There are eastbound and westbound "lanes" as before, but the loopback must be provided by connecting the appropriate end of the eastbound and westbound lanes together. An optional repeater may be needed at the union making the loopback an electrical process. Optional repeater unit cells can be used along the trunk to maintain a specified power level everywhere. This applies to the other tap trunk methods as well.

The unit cell serving local terminal clusters is identical to the star-concentrator unit cell and is the simplest. Only one transmitter and receiver is needed to access the bus, and no internal loopback is necessary since it is provided by the bus. As usual, antilockup and antibabble circuits are required.

2.1.1.2.9 Unusable Implementations: There are several subtle variations on the configurations discussed above that appear to have merit, enough so that all appeared in the open literature at one time, but for reasons which will be discussed, can get the data bus designer into deep trouble.

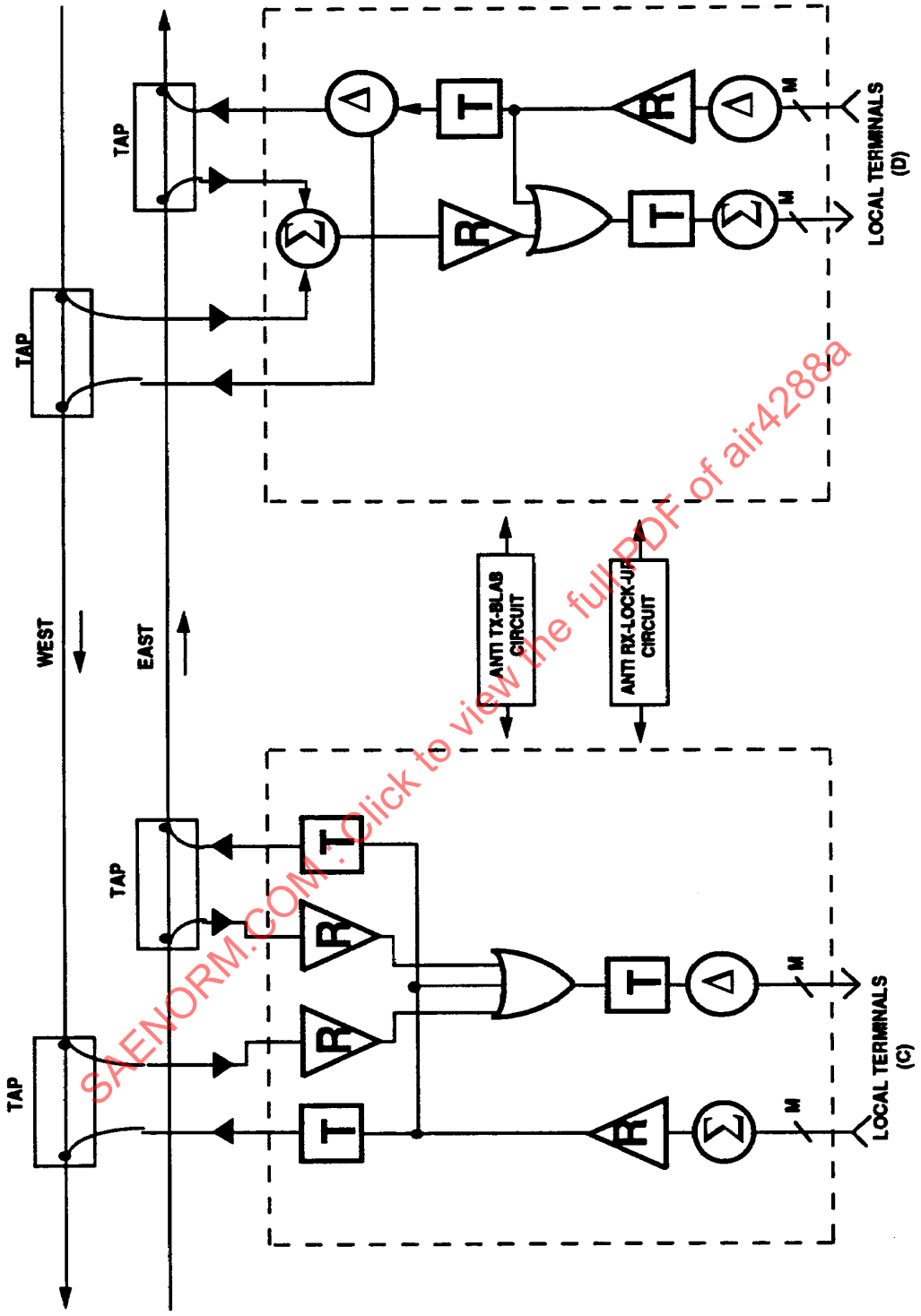


FIGURE 2.1.1.2.8.2-1 - Type C and D Unit Cell Implementation

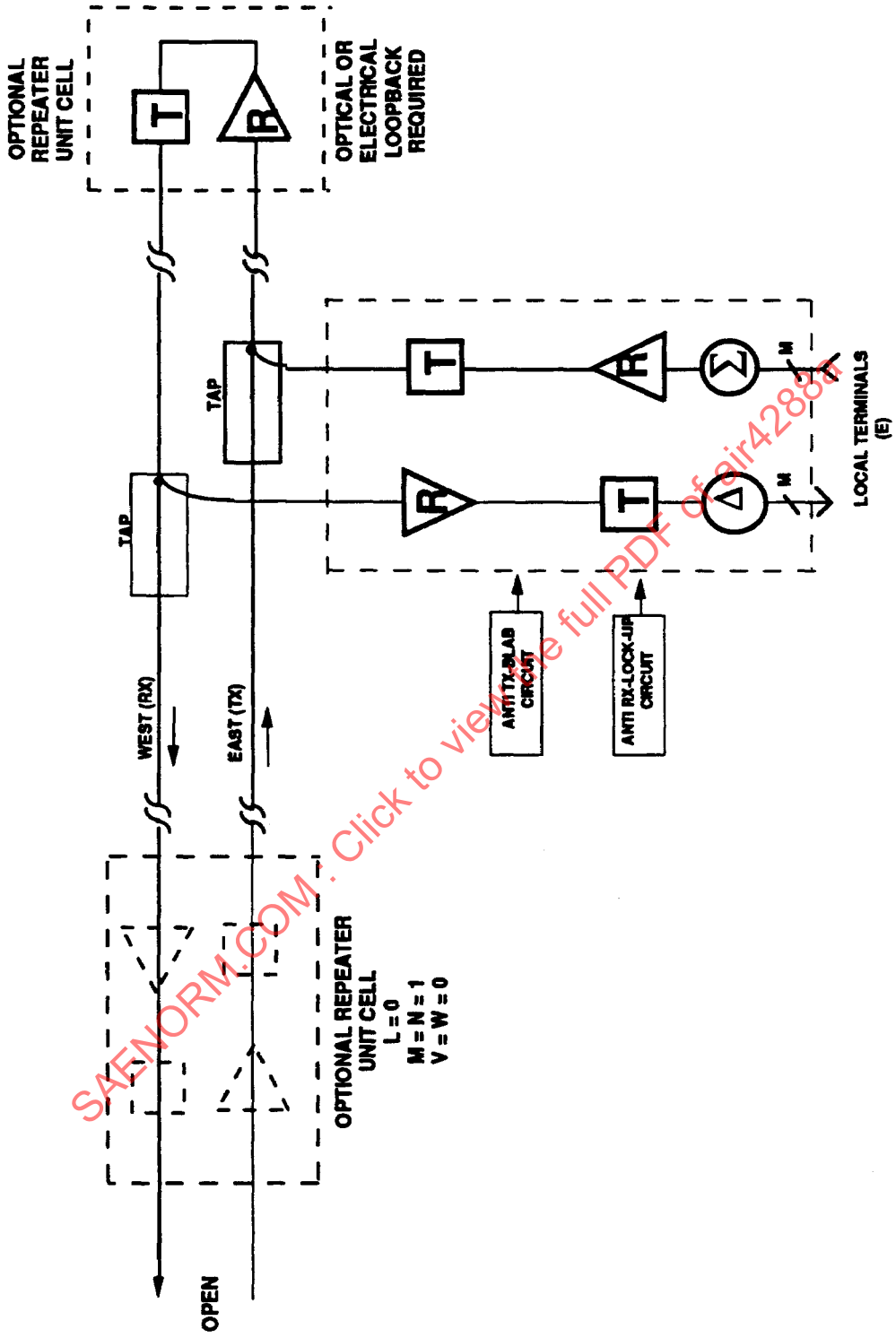


FIGURE 2.1.1.2.8.3-1 - Type E Unit Cell Implementation

2.1.1.2.9 (Continued):

Figure 2.1.1.2.9-1 is the first of these. The unit cell is the elementary active star coupler shown in the middle of Figure 2.1.1.2.2-1 for the DCA topology with only one input and output per receiver and transmitter. The network is created by tying the unit cell to conveniently located passive star couplers, as shown, where the number of input and output ports is one greater than the number of terminals. Loopback for each set of terminals (left and right) is provided by the passive star coupler, thus the left and right terminals do not have an active star coupler in their part of the network. The DCA unit cell is intended to provide gain for signals passing from left to right and right to left. However, there is a serious flaw in this design.

There is a feedback path through both passive star couplers that connect the active star inputs to the outputs. The result will be an oscillator with frequency determined by the loop path length. Removing the OR gates and replacing them with straight wires does not help. A smart lock-out logic circuit that only lets traffic flow in one direction could be added, but is more complex than can be justified for the purpose intended. This circuit should be avoided!

Note that the problem goes away if the M+1 and N+1 passive star couplers are each replaced by two passive stars acting as Mx1 or Nx1 combiners and distributors. However this is just a variation on the DCA topology of Figure 2.1.1.2.2-1 with the combiner and distributor outside of the unit cell.

Figure 2.1.1.2.9-2 shows three active-passive trunk variations seen in the literature that have implementation problems. On the left, the M+1 passive star coupler is employed to simplify the "unit cell" and provide passive loopback for local terminals, like before. However, the U-shaped tap provides loopback too, and so this link will oscillate at a quite predictable frequency. If a Y-shaped tap is used, like the one on the eastbound lane of traffic, there will be an echo signal produced on the eastbound traffic which will increase intersymbol interference and possibly prevent bus operation. A similar situation exists with the "unit cell" shown on the right in Figure 2.1.1.2.9-2. Loopback is provided twice, internal to the unit cell and in both traffic lanes on the bus. This will create echo at the local terminals and result in intersymbol interference which may shut down the bus or at least the local terminals. Again, these configurations must be avoided.

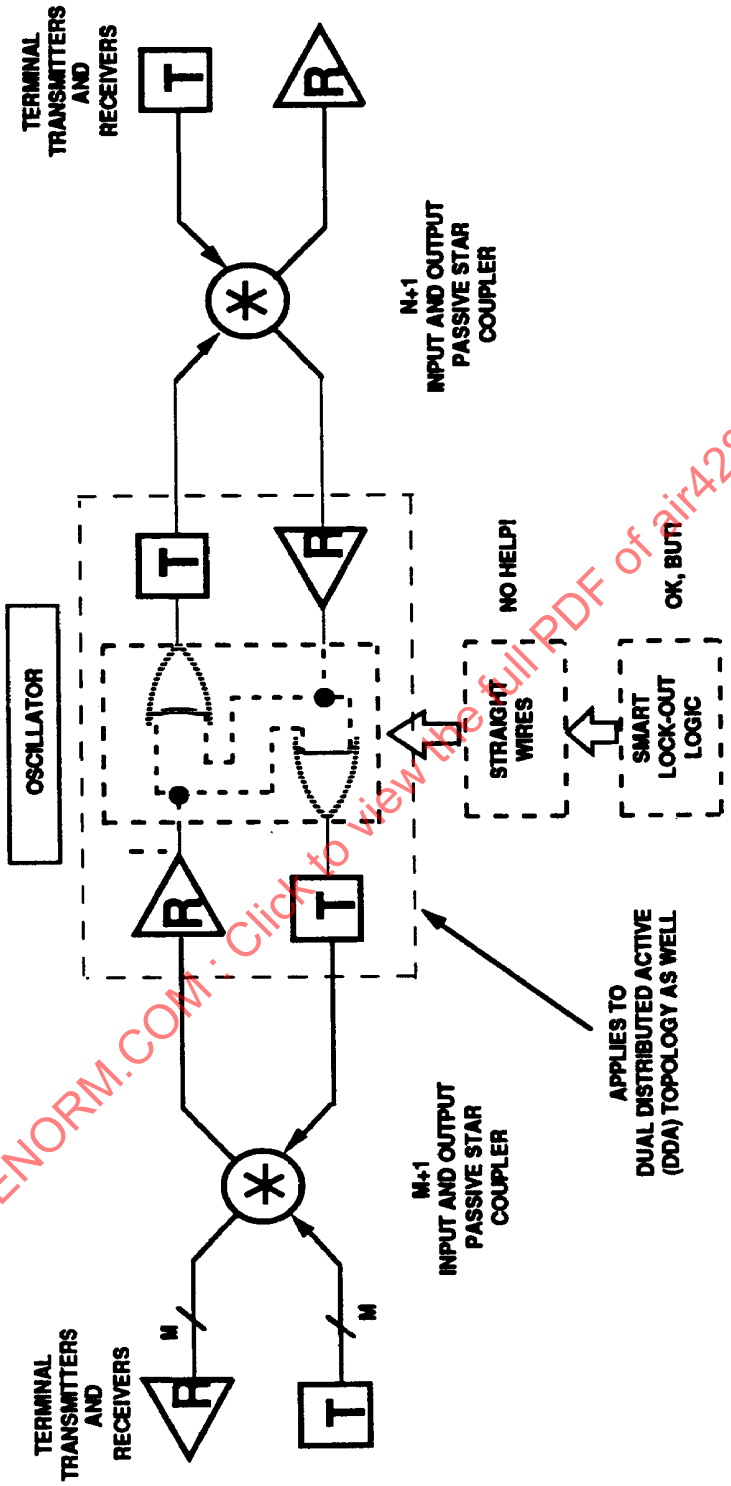


FIGURE 2.1.1.2.9-1 - Hybrid DCA Star Coupler Topology

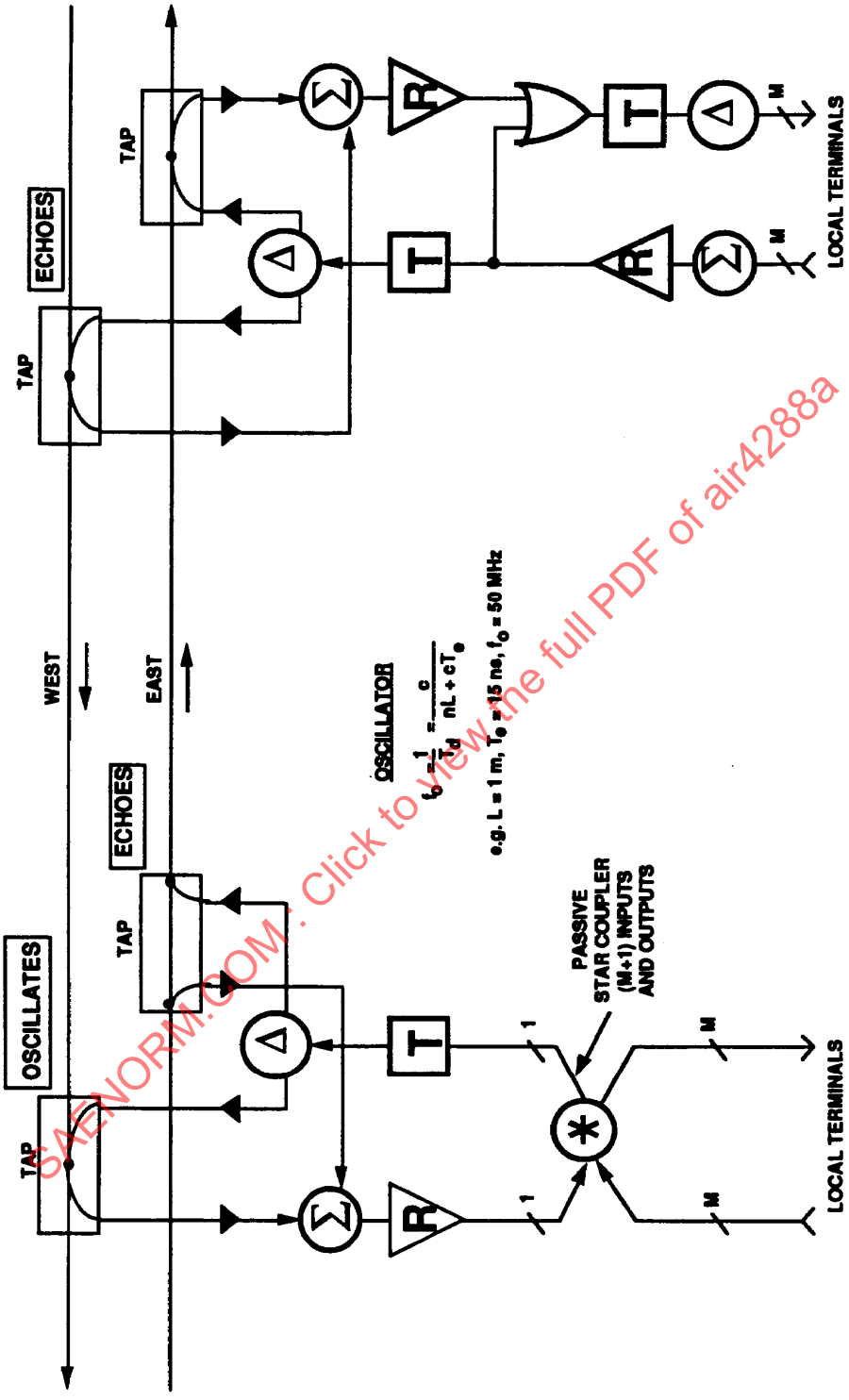


FIGURE 2.1.1.2.9-2 - An Unusable Active-Passive Trunk Implementation

2.1.1.2.10 Summary: Based only on their merits from the viewpoint of utility in a fiber-optic serial high-speed data bus, the single central, DISCO, star-concentrator, and active-active trunk are the most viable. This view may change when they are "installed" in an aircraft. The active-active trunk, though, has the liability of potentially having too many cascaded repeater stages since there is no maximum. Care must be exercised with this configuration.

Table 2.1.1.2.10-1 summarizes the active topologies introduced here in terms of the number of repeaters in a network, and the number of links between any transmitter and receiver. From the retiming and initialization issues standpoint, this should be as low as possible, so the active star-concentrator and active-active trunk may be pushing limits of practicability. Clearly, a more thorough analysis is required before any conclusions are reached.

TABLE 2.1.1.2.10-1 - Active Topology Repeater and Link Count

CLASS	TOPOLOGY	TYPE	NO. OF REPEATERS	TX TO RX LINKS
SCA	SINGLE CENTRAL	-	1	2
DCA	DUAL CENTRAL	-	1	2
DDA	DUAL DISTRIBUTED	A	1 OR 2	2 OR 3
		B	2	3
MDA	DISCO	A	1 OR 2	2 OR 3
		B	2	3
	STAR-CONCENTRATOR	PASSIVE STAR	2	3
		ACTIVE STAR	3	4
	ACTIVE-ACTIVE TRUNKS	-	1,2,..., K	2,3,...,K+1
	ACTIVE-PASSIVE TRUNKS	A, B	2	3
		C1, C2, D	1 OR 2	2 OR 3
		E*	2,3,..., K	3,4,..., K+1

K = NO. OF UNIT CELLS USED
* = FOR TWO OR MORE NODES

2.1.2 Fiber Optic Technology: This section is an overview of fiber-optic component technology from aspects other than behavior in a radiation environment. The components considered are optical fiber and cable, optical connectors, sources and detectors, and those which have special application to this data bus.

2.1.2.1 Optical Fiber and Cable: A fiber-optic cable is defined as an optical fiber which has been suitably buffered, strengthened, and jacketed to be useful in an application. A typical optical fiber is .006 inch in diameter (140 micrometers); a cable ready for installation, about 0.1 inch (2.5 mm). When terminated with optical connectors, the structure is referred to as a terminated cable, cable segment, or interconnect segment. A cable run consists of concatenated cable segments which provide a communication channel between other fiber-optic components such as a transmitter and access coupler or receiver. A complete communication channel from transmitter to receiver is usually called a fiber-optic data link. These terms will be used often in the remainder of this discussion.

An ideal fiber-optic cable would have low loss and adequate bandwidth at the wavelength of interest, and retain these properties throughout the operating environment, and over the lifetime of the system. Obviously, a bare optical fiber will have to have considerable protection to achieve that goal. In this section, the characteristics of a suitable aerospace cable will be discussed.

Characteristics - The required characteristics can be subdivided into optical and mechanical ones. The optical properties are usually stated for room temperature operation without forces applied. Since there is an interaction between the optical properties and the mechanical forces, limits have to be applied to variations expected under all operating conditions. The optical performance of the bare fiber and the jacketed fiber (the fiber-optic cable) is not necessarily the same. Since only the fiber-optic cable will be used by a manufacturer for installation, the optical properties must be specified for that configuration.

Earlier it was stated that the ideal cable had low and fixed loss and sufficient bandwidth at the wavelength(s) of interest. A minimum specification thus calls out loss or attenuation per unit length, the fiber bandwidth, and the wavelength at which these values are valid. Other optical parameters which are usually stated include the fiber's numerical aperture or angle of acceptance of light, and the index of refraction profile of the fiber's core (the waveguide portion of the structure).

The mechanical properties generally consist of the dimensions and materials for all components of the cable, and the strength properties of the bare fiber, the cable's strength member, and the cable's outer jacket. This is usually sufficient to characterize the cable for procurement purposes provided there is sufficient test data to justify the selection of component materials.

Some additional comments regarding fiber bandwidth, attenuation as a function of mechanical stress, static fatigue, and high temperature operation follow.

2.1.2.1 (Continued):

Bandwidth - Fiber "bandwidth" is usually specified in units of MHz·km, that is as a product of bandwidth and length. This means that the longer the fiber, the narrower the bandwidth of that run. For example, a 100 MHz·km fiber has, in principle, a bandwidth of 100 MHz for 1km, 1000 MHz for 100 meters, and 10 GHz for 10 meters. For a data bus installation of less than 10 meters long for a complete link (transmitter to receiver), it does not appear necessary to be overly concerned over the bandwidth specification of an optical fiber for this application. That is not necessarily the case, however.

The factor which determines a fiber's (frequency domain) bandwidth is called dispersion in the time domain and has units of time per length such as nanoseconds per kilometer. There are two kinds of dispersion: modal and chromatic. Modal dispersion arises because different light rays follow different paths in multimode fiber. Chromatic dispersion is a result of different wavelengths of light (different colors) traveling at different speeds in the fiber. Manufacturers of multimode optical fiber generally specify bandwidth for a zero-linewidth optical source thus avoiding any reference to the effect chromatic dispersion may have on the net bandwidth. Since most data bus and network applications use light-emitting diodes which have relatively wide spectral outputs (typically 50 to 150 nanometers), chromatic dispersion may not be negligible. It is recommended that total fiber dispersion be used to specify an optical fiber for any application. For multimode fiber which behaves as a gaussian low-pass filter, the total dispersion is the root-sum-squared of the modal and chromatic dispersion factors:

$$\Delta t_f = \sqrt{(\Delta t_c)^2 + (\Delta t_m)^2}$$

where Δt_f is the total fiber dispersion, Δt_c is the chromatic dispersion and Δt_m is the modal dispersion, with all terms having units of seconds. Furthermore, the total dispersion which is actually found in a fiber is not linear with length of the fiber and is sensitive to the manner in which the light from a source is coupled to the fiber (the launch conditions). Therefore, the total fiber dispersion should be specified for the maximum length expected in the application. Here is an example for a data bus application. For a 100 micrometer core fiber at 850 nanometers, the modal dispersion is about 1 nanosecond after 100 meters. The chromatic dispersion is about 0.1 nanoseconds per nanometer per kilometer; so for 100 meters and the 50 nanometer spectral width source, this term contributes 0.5 nanosecond. The total dispersion is:

$$\Delta t_f = \sqrt{(1)^2 + (0.5)^2} = 1.1 \text{ ns}$$

This quantity can be thought of as the rise time of the fiber to a step input of light, analogous to the response of a single pole filter or amplifier. It is related to bandwidth by the approximate relation:

$$\text{BW } 0.44/t_f = 0.44/1.1 \text{ ns} = 400 \text{ MHz}$$

2.1.2.1 (Continued):

This is the bandwidth of 100 meters. If the bandwidth scaled linearly with length, this would be a 40 MHz·km fiber. However, the manufacturer measured modal dispersion for a one kilometer length and may have obtained a 100 MHz·km bandwidth. This corresponds to 1000 MHz for 100 meters. The reason for the difference is that dispersion is linear with length only up to a certain point called the coupling length, l_c, and beyond that dispersion varies approximately with the square root of length. The coupling length differs with type of fiber. It can vary from a few meters for step-index core fibers to hundreds of meters for some graded-index core fiber. In summary, to avoid pitfalls associated with manufacturer's specified bandwidth, it is preferable to specify total dispersion for a length not exceeding the maximum in the system.

Although not apparently a large factor, dispersion can have a critical effect on overall system performance, especially in a worst case situation. This is because the total dispersion of typical multimode fibers is of the same order of magnitude as the minimum rise and fall times of high-radiance LEDs and the step response of high sensitivity receivers in data buses with link lengths near the maximum. A good approximation to the rise time (or fall time) of a signal at the analog output of a receiver is:

$$\Delta t = 1.1\sqrt{(\Delta t_f)^2 + (\Delta t_t)^2 + (\Delta t_r)^2}$$

where t_t is the rise (fall) time of the transmitter optical output and t_r is the step response of the receiver. If the receiver output waveform is badly distorted due to the finite response time of all the link components, the intersymbol interference will be too great to achieve the desired measure of performance, usually specified by bit error rate. For example, if a 100 Mbaud data bus has a transmitter with a 3 ns rise and fall time optical output, a 2 ns modal dispersion, and a high sensitivity band-limited receiver step response of 4 ns, the actual rise time of the receiver's output will be 6.0 ns. With a minimum signaling bit time of 10 ns, there may be a considerable amount of intersymbol interference even without pulse width distortion. By making the modal dispersion negligible, the receiver output is improved to 5.5 ns.

Bending Losses - The loss in a fiber due to bending can be subdivided into microbending and macrobending loss components. The cabling and jacketing process is responsible for increasing microbending loss, but can be controlled through careful design. Microbends can be thought of as continuous, minute changes in the radius of curvature of the fiber axis as shown in exaggerated form in Figure 2.1.2.1-1(a). The loss results from the coupling of energy between the guided modes and leaky or unguided modes of the fiber. A large diameter fiber of given numerical aperture resists microbending better than a small one. The aspect ratio of the fiber (the ratio of core to cladding diameter) should be kept small to resist microbending loss unless large numerical apertures (NAs) are involved. But the use of high NA fiber may be precluded due to accompanying high modal dispersion. Aspect ratios less than 70% are recommended.

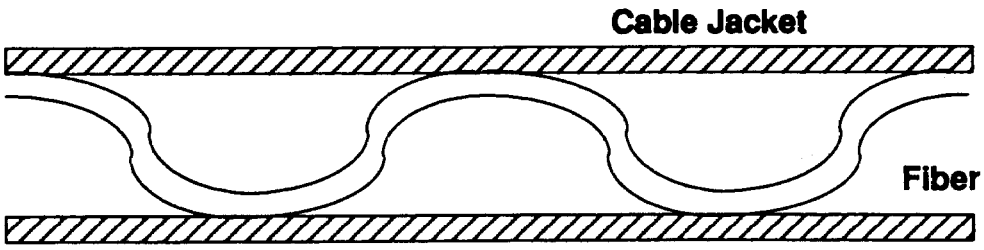


FIGURE 2.1.2.1-1(a) - Microbends in Loose-Tube Constructed Cables

2.1.2.1 (Continued):

Another way to minimize microbending loss is to extrude a compressible jacket over the fiber as shown in Figure 2.1.2.1-1(b). External forces deform the jacket, but the fiber remains relatively straight. It has been shown that the loss of a jacketed fiber can be reduced from that of an unjacketed fiber by a factor given by the following equation:

$$F = \left\{ 1 + \pi \Delta^2 \frac{b^4}{a} \frac{E_f}{E_j} \right\}^{-2}$$

where Δ = index of refraction difference between core (on axis) and cladding, b = cladding radius, a = core radius, and E_f and E_j are the Young's modulus of the fiber and jacket material. Typical values are E_f = 64 GPa (glass) and E_j = 58 MPa (Hytrel 4056). For a 100/140 micrometer core/cladding ratio fiber with Δ = 0.01, F = 0.18 implying that microbending loss can be reduced to 18% of the original value just by properly jacketing the bare fiber. Note that as the aspect ratio (a/b) increases, the factor F increases. A 200/230 micrometer core/cladding ratio fiber with identical properties has F = 0.4.

Bends with radii large compared to the fiber diameter cause loss increases also. This is the macrobending loss. Slight bends of the cable have so little excess loss that it is unobservable. As the bend tightens, the loss increases exponentially until a critical radius of curvature is reached and the losses become large. The loss occurs because energy in the higher modes is coupled out of the fiber core along the curved fiber thus reducing the total energy available at the fiber output. An expression for the number of modes which are guided by a curved graded-index fiber with fiber radius a is

$$N = N_\infty \left\{ 1 - \frac{\alpha + 2}{2\alpha\Delta} - \left[\frac{2a}{R} + \frac{3}{2n_2 kR} \right]^{2/3} \right\}$$

2.1.2.1 (Continued):

where a defines the fiber index of refraction profile (a = 2 typically for graded-index fiber; a = ∞ for step index fiber), Δ = index of refraction difference between the core (on axis) and cladding, n_2 is the cladding index of refraction, k is the wavenumber, R is the radius of curvature of the bent fiber, and:

$$N_\infty = (a/a+2)\,(n_1 ka)^2 \Delta$$

is the total number of modes in a straight fiber, where n_1 is the core index of refraction. At 850 nm wavelength and a = 2, k = 7.4 X $10^6 m^{-1}$. Choose n_2 = 1.47, n_1 = 1.46, Δ = 0.01, a = 50 micrometers. N_∞ = 2860 and

$$\frac{N}{N_\infty} = 1 - 10^{-4}\left[\frac{100}{R}\,\frac{26.7}{R^{2/3}}\right]$$

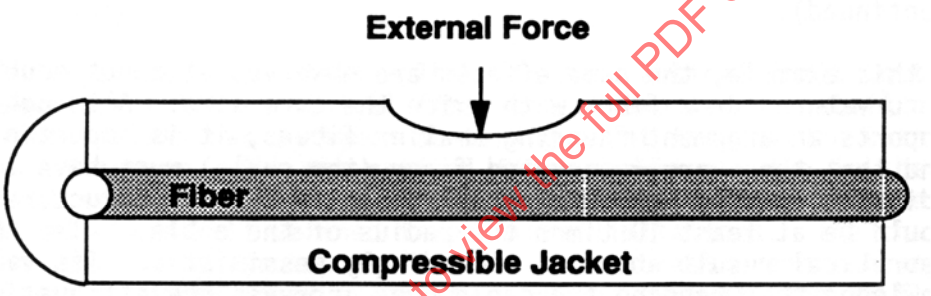


FIGURE 2.1.2.1-1(b) - Microbending Reduced with a Compressible Buffer Jacket Material

This expression is plotted in Figure 2.1.2.1-2 for values of a = 50 and 100 micrometers, where R is in centimeters. For a = 50, half of the modes are lost at R = 2.0 cm and there is no energy left at R = 1.0 cm. Since the energy is not distributed uniformly among all the modes, loss of the higher-order half of the modes does not represent loss of half of the energy. Nevertheless, at R = 2.0 cm, one can expect a measurable increase in loss which rapidly increases as R decreases. In practice microbending loss from mode coupling actually keeps macrobending loss under control, and R has to be decreased past the elastic limit of the glass (it has to break) for all the light to be attenuated.

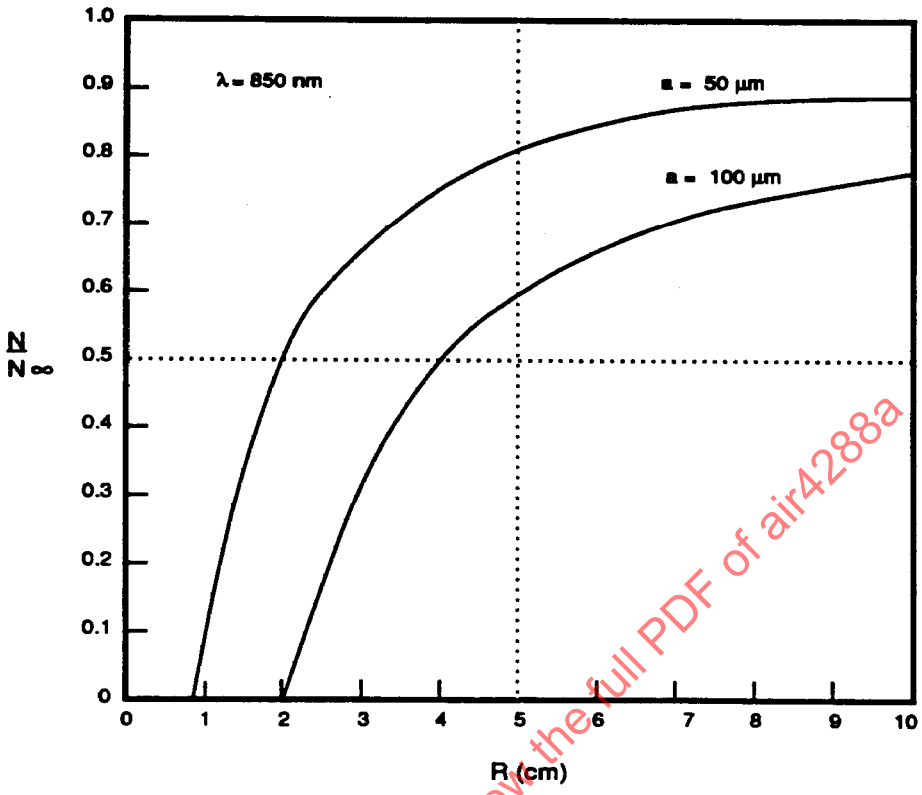


FIGURE 2.1.2.1-2 - Guided Mode Ratio as a Function of Bend Radius, R

2.1.2.1 (Continued):

In this example, the same effects are observed at about double the radius of curvature for a fiber with twice the core size. Although this result supports an argument favoring smaller fibers, it is important to keep in mind that the overall jacketed fiber (the cable) must have a minimum bend radius to preserve the mechanical integrity of its structure and that should be at least 10 times the radius of the cable. Also the theoretical result above is exceedingly pessimistic. Its value is in the knowledge that bending the cable does increase the attenuation, and that minimum bend radii for fiber-optic cables must be observed. This requires training in the manufacturing and installation world which is accustomed to folding copper wire as tightly as required despite the fact that there are minimum bend radii specified for it also. Actually, fiber-optic cable is less susceptible to folding and bending problems than is coaxial cable. It simply will have to be handled more carefully. A passive solution may be to make the fiber-optic cable so stiff that it cannot be bent beyond a minimum radius. Most agree that that is a drastic solution. A meaningful specification for minimum bend radius must be established based on experimental results.

Static Fatigue - Strength and static fatigue are the two basic mechanical characteristics of glass optical fiber. Strength specifies the failure level of a fiber under an applied load; static fatigue deals with the slow growth of preexisting flaws in the fiber under humidity conditions and tensile stress. The concern is that a fiber may fail before its 20 to 30 year required lifetime is reached due to static fatigue.

Gradual flaw growth causes the fiber to fail at a lower stress level than that which would be reached under a strength test. A flaw propagates through the fiber because of chemical erosion of the fiber material at the flaw tip. The primary cause of this erosion is the presence of water in the environment which reduces the strength of the SiO_2 bonds in the glass. The speed of the growth reaction is increased when the fiber is put under stress. However, based on experimental investigations, it is generally believed (but not yet fully substantiated) that static fatigue does not occur if the stress level is less than approximately 20% of the inert strength (in a dry environment, such as a vacuum). Certain fiber materials are more resistant to static fatigue than others, with fused silica being the most resistant of the glasses. In general, coatings which are immediately applied to the fiber during the manufacturing process afford a good degree of protection. All fibers currently available for avionics applications have polymeric coatings on the fiber. These are permeable to water vapor and therefore do not afford the optimum protection against static fatigue. The ideal coating is a thin (0.03 nanometer) dielectric which hermetically protects the fiber. A metallic coating provides the needed hermeticity, but it also makes the fiber an electrical conductor and antenna defeating one of the benefits fiber-optic technology offers. Some dielectric coating materials being considered are Si_3N_4, SiON, SiC, and SnO. Other techniques for reducing flaws which contribute to static fatigue such as improving preform preparation, and modifying furnaces and draw conditions are also being funded.

2.1.2.1 (Continued):

High Temperature Operation - There is no practical limit to the highest temperature which an optical fiber can be exposed and still provide service. Glass has a higher softening point temperature than the melting point of copper. But the optical fiber cannot survive in the real world without suitable jacketing for mechanical protection, and the jacket materials have well defined temperature limitations. The characteristics of an ideal outer jacket include adequate abrasion and cutting resistance, dimensional stability, freedom from kinking in flex (which causes excessively high losses), flame retardancy, and operational suitability over a wide operating temperature range. Several materials come close to meeting these requirements and have been used for jacketing cables in the past. Below is a list of a few of the higher performance outer jacket materials and their properties. Teflon and cross-linked Tefzel allow the fiber to be used continuously over a temperature range of -65 °C to +200 °C. A principal supplier of this fiber forbids the reproduction of the specifications and drawings for his high temperature fiber-optic cables. Data obtained from advertisement for it states its weight at 3.7 pounds per thousand feet and a survival life of 4000 hours at 200 °C continuous. It is a "loose-tube" constructed cable, a type which is usually used for long distance applications where attenuation is a major concern. Tight tube construction is generally preferred for applications in which high mechanical integrity is more important than low attenuation. These various outer jacket materials for fiber-optic cables and their characteristics are:

Kynar® (polyvinylidene fluoride)	A tough (abrasion and cut-through resistant), thermally stable and self-extinguishing material. It has low smoke emission and is resistant to most chemicals. Its inherent stiffness limits its use as jacket material. It has been approved for low-smoke applications.
Teflon® FEP	Specified in fire alarm signal system cables. It will not emit smoke even when exposed to direct flame, is suitable for use at continuous temperatures of 200 °C, and is chemically inert.
Tefzel®	Like Teflon FEP, it is a fluorocarbon and has many of its properties. Rated for 150 °C, it is a tough, self-extinguishing material.
Irradiated Cross-Linked Polyolefin (XLPE)	Rated 150 °C operation. Cross-linking changes thermoplastic polyethylene to a thermosetting material with greater resistance to environmental-stress cracking, cut-through, ozone, solvents and soldering than either low or high density polyethylene.
Zero Halogen Thermoplastic	A thermoplastic material with excellent flame retardancy properties. Does not emit toxic fumes when it burns. Originally designed for shipboard fiber-optic applications, it can be used for any enclosed environment.

Kynar is a registered trademark of Pennwalt, Inc.
Teflon and Tefzel are registered trademarks of E.I. DuPont de Numours & Company.

2.1.2.2 Optical Connectors: A variety of fiber-optic connector styles are usually required to fulfill all data bus and network applications on aircraft and spacecraft. The use of line replaceable modules (LRMs) forces the use of individual contacts to provide the backplane interface up to about 8 single channels. Beyond that, too much space is occupied by the single channel contacts, and some form of multichannel single interface, such as an imaging connector, better serves the needs of the module. At rack connections, either single-channel connectors or a multicontact single-connector-shell interface looks best depending on the total number of modules served by the rack. At stand-alone LRUs (WRAs) with mainly input and output channels, one or more multichannel imaging connectors are particularly attractive. Their higher loss per channel (at the present time) limits the number which can exist along any one path from transmitter to receiver. The limitation is relaxed, however, in data buses incorporating active star couplers or in point-to-point links.

Finally, bulkhead disconnects may be required at certain penetration boundaries especially where a pressure difference exists. All of the connector approaches mentioned above are candidates here depending on the system power budget. Therefore, we limit the connector discussion which follows to single-channel contact technology for single and multiple-channel connector shells, and multichannel interfaces with a common optical aperture.

Characteristics - The single most important connector characteristics for all applications is its insertion loss. Repeatability of the loss over time with and without remating is also important. Since connector loss is a major system design driver, it is always the case that the lower the loss, the better. So the question becomes, how low is low, and how low is low enough? Then one must ask how the losses will be kept low.

The connector loss picture is not as bleak as it is sometimes painted. Single-channel connectors with losses consistently below 1 dB for hundreds of matings and repeatability on the 0.1 dB range are now available from multiple suppliers. Also consider that a situation in which a group of concatenated connectors in a transmitter to receiver path all have high losses simultaneously is not reasonable, especially when consistency in connector performance is taken into account. Once the connection is made, dirt and other foreign objects will be virtually unable to enter the connector space in any reasonable design.

There are connector terminii designs now being fielded which use lenses and still retain small diameters. They have the advantage of being much less sensitive to small obstructions in the space between connector halves because of their relatively large optical cross-sectional area. This property offers an advantage over connector terminii in which the ends of the fibers face each other directly with a small air gap between them.

2.1.2.2 (Continued):

A final consideration is the size of the optical fiber accommodated by the connector terminus or contact. Because of the lack of firm standards, connector design must be flexible enough to handle several sizes. The defacto standard avionics fiber has a 100/140 micrometer core/cladding ratio and most connectors for avionics-like applications are designed with that fiber in mind. There are efforts underway to promulgate other standards. Specifically, the telecommunications industry is now competing for the military and space market and believes their 125 micrometer cladding standard is best. Unfortunately there are several core sizes competing for survival inside of 125 micrometer claddings and all are smaller than desired to get the most power out of multimode optical sources. At the other extreme, designers of the data buses incorporating passive star-coupled topologies truly need all the power possible and many support the use of 200 micrometer core fibers. A standard cladding dimension, which is important to the connector manufacturer, has not yet been established for this core size; nearly every fiber maker has a unique cladding diameter. Until that issue is resolved, the prospects for connectors for this fiber are much lower than they ought to be.

The remainder of this section presents examples and characteristics of connectors and contacts (terminii) being developed or considered for avionics and space applications.

Single-Channel Connectors - Here, a single-channel connector refers to a terminus for an optical fiber and the connector shell which houses it. The shell provides the apparatus to mount it on a cabinet panel or "splice" it to another connector half with the addition of an adapter. The best known example of this connector type is the so-called "SMA-style" because of its resemblance to the RF SMA connector. Actually, the only part in common is the coupling nut which characterizes the overall diameter.

This connector has been available in various versions from different suppliers for 10 or more years and has been refined about as much as it can. By modifying it with a better alignment sleeve (for connector-to-connector interfaces), and providing a coupling nut with holes for safety wire, this connector found its way onto the Navy/Marines AV-8B Harrier II, the first production military aircraft to incorporate fiber-optic interconnections. It is not a low-loss connector, having a typical loss of 1.0 to 1.5 dB. Although it works well in the specific application, it is not a state-of-the-art connector and would not be used in a new aircraft design. Today, connectors with a typical loss of 0.5 dB and a maximum loss of less than 1 dB are available at lower cost than that of the SMA connector. The prototype for the modern single-channel connector is the AT&T "ST" style. It is now available from numerous suppliers with essentially identical specifications. The key to its higher performance (low loss and high repeatability) is the use of a ceramic tube rather than a metal one to contain the optical fiber and provide the reference (outside) surface for alignment. The performance improvement results from the high precision achieved in placing the hole for the fiber in the true center of the ferrule and because the sleeve which aligns two ferrules in an adapter is designed in such a way that its cross-section remains circular when the cylindrical ferrules are inserted. Keying the connector shell assures repeatability from mating to mating.

2.1.2.2 (Continued):

The suitability of this connector for aerospace applications from the standpoint of environmental and mechanical stress has yet to be established. As it stands today, it is an excellent laboratory connector and will now and forever displace the SMA-style for that purpose. All indications suggest that the ceramic ferrule is entirely suitable for military aircraft and space applications having better dimensional stability and a much closer material match with glass optical fibers. Another property which appears to put ceramics at a disadvantage actually helps. The ceramic material is brittle compared to steel. Thus a large deflection force on the ferrule will cause steel to bend while the same force breaks ceramic. Since a bent steel ferrule is not repairable and obviously more difficult to diagnose as a culprit, the ceramic ferrule approach offers a true maintainability advantage in avionics use. With a slightly different coupling ring or nut, this connector could meet requirements imposed by aircraft and spacecraft launch-vibration levels. A version of this is now available. Its suitability for the avionics environment is unknown; any tests which may have been conducted have not yet been made generally available.

Single-Channel Contacts - The term single-channel contact or terminus refers to the structure which supports an optical fiber at its end and is designed to be incorporated into a connector housing more than one terminus. An example is a fiber-optic contact which can be retained by the insert of a standard multichannel electrical connector shell. The best known type is described by the MIL-C-38999 specification for military avionics. A similar (or the same) type of contact is desired for the backplane connector for the Standard Electronics Module. Several connector manufacturers are currently designing and fabricating fiber-optic contacts which are intended to be an integral part of the electrical connector for that application.

For years there has been a demand for a small fiber-optic contact which could be interchanged with an electrical contact in a standard military connector. For obvious reasons, the MIL-C-38999 connector was targeted from the beginning. Primarily due to the stimulus provided by the aircraft industry, a number of manufacturers began developing size 16 or smaller contacts for MIL-C-38999 connectors. A number of products resulted and have been evaluated for avionics applications. Testing of each supplier's product is still required for full qualification.

One of the most interesting of the new single-channel contact concepts is based on the lensed-connector method and is called "fiber-lens" by its manufacturer, ITT/Cannon. It uses an epoxy preform within the contact to hold the fiber in place rather than an externally-introduced liquid epoxy. Additional support is provided by crimping, thus the assembly process is simpler than that of many other connectors. Its unique feature is the way a lens is formed at the end of the fiber. Rather than using a discrete or graded-index lens, a spherical lens is formed from the end of the fiber itself. This method was demonstrated many years ago, but is only showing up now as a product. The difficulty in producability comes from the need for a special instrument, the "fiber-lens fuser," to form the lens.

2.1.2.2 (Continued):

Another single-channel contact system which is attractive because of the easy termination process is produced by Raychem and requires no polishing, epoxy, lenses, or liquids. It is currently made a nonstandard size but the same contact can be used as a single-channel connector with its own coupling nut or as a component of that manufacturer's own multichannel connector designs. Its losses are not claimed to be as low as those using ceramic ferrules, and it differs from the "fiber-lens" concept above in that a mated pair of contacts have fiber ends looking at each other through a small air gap much like the SMA style connector.

Multichannel Connectors - Multichannel fiber-optic interfaces are required in some systems. A modification to some existing single-channel connector concept, or a contact set designed for use with a multichannel connector housing, is the approach being most actively pursued by connector manufacturers. Circular connector shells (such as MIL-C-38999 style) and rectangular housings for SEM-E line replaceable modules are the chief targets for the fiber-optic contacts. There are at least five manufacturers currently developing an LRM-type connector with a fiber-optic insert to take contacts of their own making or those of another vendor.

There is yet another class of multichannel fiber-optic connector to be considered for avionics applications: one which is a dedicated optical interface (fiber-optics only) and has multiple channels which are integral and nonseparable from its housing. Two examples are worth mentioning. The first, from AT&T, takes a large number of optical fibers and places them in a precision one-dimensional array. Two arrays are then butted fiber-to-fiber to create a connection for as many channels as there are fibers in each array. The advantage of this approach is that a large number of fibers can have a demountable interface in a space which is small compared to what would be required if individual contacts like those described above were used. The disadvantage is that the faces of the fiber arrays are placed in contact to keep losses minimized, and in a high vibration environment, the mating halves of the connector are vulnerable to rapid wear. There are also undesirable alignment and contamination sensitivities which single-channel butted connectors and contacts have. It seems to be most suitable as a permanent splicing concept for multiple channels.

Figure 2.1.2.2-1 shows a two-dimensional array implementation of the above concept. The second class of multichannel single aperture connector is a variation on that shown in the figure, but is carried out in such a way that it is much more tolerant of the environment and attractive for avionics. In the simplest terms, the two butted two-dimensional arrays shown in Figure 2.1.2.2-1 are first separated by a relatively large distance (the spacing is greater than the diagonal measure of the array). Then the space between the arrays filled with a symmetrical optical relay (a lens) so that one array (the object) is mapped onto the other (the image). If the lens is made of at least two components and is split in half, and each half is permanently associated with its array and put to a suitable housing and a connector mated pair is formed. This connector concept has proven to have significant environmental advantages over many other approaches and has very high channel density. It is currently being developed by the Harris Corporation and has been built in bulkhead versions with up to 84 channels in a 1 inch diameter shell, and with 12 channels in a 0.58 inch wide housing for use on single-width, 3/4-ATR, SEM-E line replaceable modules.

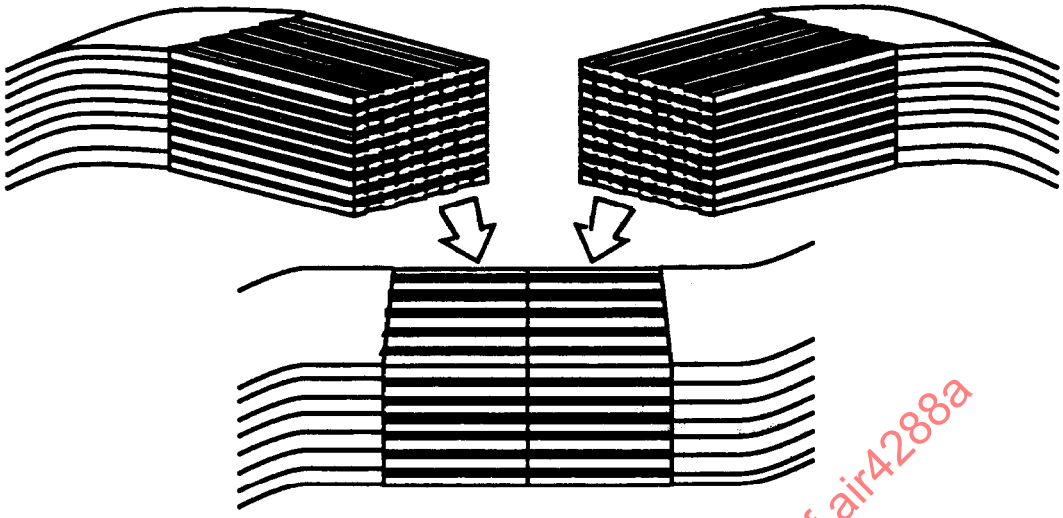


FIGURE 2.1.2.2-1 - A Demountable, Butted, Array Pair

2.1.2.3 Fiber-Optic Cable Segments Fabrication and Installation: Once a fiber-optic cable and connector are selected based on their apparent suitability, the two components must be joined and the utility of the resulting cable segment or segments must be determined. This is done in a mechanical and environmental testing program with emphasis on the cable to connector joint, and the properties of connector-to-connector, and connector-to-source and sink interfaces. After acceptability on the basis of laboratory testing, it can be introduced into the development and production programs. There are two facets which must be addressed: its fabrication and its installation. Fabrication deals with the process of putting a connector or contact onto a fiber and cable; installation with putting the cable segments onto the platform. It is important that both be considered as early in a program as possible to ensure a smooth transition of the technology into production. Some of the early issues for each are discussed here.

Fabrication - Factors to consider in the evaluation of concepts for attaching connectors to cables include the initial difficulty, the slope of the learning curve, the yield, the cost of the components and capital equipment to perform the fabrication steps, including checkout, and the potential of the system for easy field maintenance. In this paragraph, a description of some of the standard and newer termination procedures is given which will guide the reader in making knowledgeable decisions about the best fabrication procedure for their application. The factors mentioned should be kept in mind to assure a best technical approach over the life cycle of the installation.

2.1.2.3 (Continued):

The processes used today to terminate fiber-optic cables can be subdivided into those which require polishing of the fiber end, and those which do not. The latter group is further divided into simple cleave, and specially processed terminations such as the new "fiber-lens" method described. Another distinction can be made about the way the fiber, its jacketing, and the cable's strength member are retained in the contact or ferrule. In almost all cases, a crimp is used with or without an adhesive, usually epoxy, to retain the strength members. Some form of epoxy is also used to hold the fiber to the ferrule. Two variations are found here; either a liquid, which is injected into the contact during fabrication, is used or the contact is supplied with an integral adhesive preform. In both cases, some sort of oven is required to cure the epoxy, or melt-then-cure the preform. Some manufacturers consider their preform approach to be epoxyless, but it really only means the steps required to prepare and administer a two-part adhesive are eliminated. It is obviously more convenient, but not all of the steps associated with epoxy use are omitted in this process.

Thus, the two ways of subdividing contact types leads to the common names associated with termination processes such as "pot (epoxy) and polish," or "cleave and crimp." In general, there has been a decided trend away from the epoxied and polished ferrule for the simple reason that combination of processes is the most labor intensive. Nevertheless, for avionics and spacecraft applications, it still remains as the most reliable and lowest risk approach, primarily because of the experience base. As time goes on, it will surely be supplanted by faster and cheaper methods without sacrificing reliability. Selected termination procedures are summarized in Table 2.1.2.3-1. This is only a representative list of popular methods; there are others, and other suppliers of products which are interchangeable or intermateable with those listed in the table. Based on experience in the termination process gained by many, it is difficult to determine a best way from written instructions. Hands-on experience is essential to determine which method best meets the needs of each user. For avionics applications, it has already been well established that the fabrication process for fiber-optics is no more difficult than that for coaxial cable, the learning curve is steeper (fabricators learn faster), and the yield is higher (as long as production shortcuts are not taken). Newer processes are likely to be competitive with single-conductor wire termination performance.

As more extensive use of fiber-optics begins in avionic and spaceborne systems, further advantages will be found because nearly all applications can be served by one fiber-optic cable and a few connector types.

TABLE 2.1.2.3-1 - Selected Termination Procedures for Fiber Optic Connectors

Kind	Representative Supplier	Product Name	Polish/ Cleave/ Special	Epoxy/Crimp/Preform	
				Jacket Strength Members	Fiber
Single Channel Connector	Amphenol (Metal)	905, 906 (SMA)	P	E, C	E
	Amphenol (Ceramic)	905, 906 (SMART)	P	E, C	E
	AMP, Inc. (Ceramic)	Optimate 2.5 mm bayonet	P	E, C	E
	AT&T, Inc. (Ceramic)	ST (P2020A-C)	P	E, C	E
Single Channel Connector or Contact	Raychem (Metal)	PELL (SSMA)	C	E, C	P
Single Channel Contacts	ITT/Cannon (Metal)	Size 16 for MIL-C-38999 Series I & III	P	E	E
	ITT/Cannon (Ceramic Tip)	"Fiber-lens" Size 16	C & S	C	P

2.1.2.3 (Continued):

Installation - After a cable segment is fabricated and its continuity verified with an optical power meter, it is ready for integration with other fiber-optic cable segments and conventional wire into harnesses which will then be installed in the spacecraft. At this stage of the process, it is only necessary that the fiber-optic cable segments be rugged enough to survive handling during the storing, lay-up, harnessing, overbraiding (if applicable), more storing, and delivery to the installation site. This may be a greater challenge than it seems at first.

Throughout this process, the contacts and connectors must be adequately protected by covering the vulnerable optical fiber end or lens surface. Both the cable and connector as components are durable. The weak link is where the cable meets the connector in the segment. Attention to this region of the cable is essential.

During design, care must be taken to ensure that both wire and fiber-optic cables are routed in such a way that they receive reasonable protection from personnel who might step on cables during installation; from debris which is created during fabrication; from any objects which might move during operations (actuators, cables); from wear due to the effects of vibration on the harness or cable outer jacket where it contacts structure; and from any gases or fluids which either belong where they are, or show up where they do not.

2.1.2.3 (Continued):

In general, the same steps used to install conventional wire cable segments and harnesses should be applicable to fiber-optics. However, fiber-optic cables will not be able to stand kinking to the extent copper wire can. Although there are minimum bend radii specifications for wire, they are often ignored. This can cause major performance problems for coaxial cable, and eventually will require a maintenance action. Optical fiber which is folded over and crushed during installation may break. It will probably be necessary to enforce a minimum bend radius of 10 times the outside diameter of the largest cable segment in a harness.

Some special connector considerations which must be quantified prior to installation of fiber-optic segments and harnesses include the requirements for cleaning the connector prior to mating; a specification for torquing any threaded connectors; requirements for securing couplings such as the use of safety wire and heat shrink tubing; and the manner in which unmated connectors must be stowed.

A specification is required for the material to be used and methods employed to attach a cable or harness to the aircraft or spacecraft structure. The requirement for service loops must be established. Criteria for repairing a cable (splicing) and replacing may also be needed. This is usually based on some visible evidence of damage or by a measurement of cable attenuation showing out-of-limits performance. It is also possible to identify the location of faults in cable runs using special instrumentation.

In summary, before fiber-optics can be installed in aerospace vehicles, a specification outlining all procedures in detail is required including quality assurance provisions and control methodology for the process and acceptability of an installation.

2.1.2.4 Optical Sources and Detectors: An obvious requirement is that the source must give off light and the detector must be sensitive to light at the wavelengths that optical fiber is highly transmissive. The kinds of sources and detectors which are effective with optical fiber are:

(1) Small devices
(2) Bright and highly directive sources
(3) Responsive detectors generating little noise
(4) Useable in the range 600 to 1600 nm

Two types of optical sources for fiber-optics have emerged. Both are referred to as electroluminescent devices. They are light-emitting diodes (LEDs) and injection laser diodes (ILDs). In the most elemental form, they are semiconductor chips with two adjacent (p- and n-type) regions whose interface is called a pn junction. When forward biased current flows through it and light is emitted in the junction region.

ILDs are generally edge-emitters; LEDs can be edge or surface-emitting devices. Figure 2.1.2.4-1 illustrates this difference and also provides a concise summary of the principal characteristics of state-of-the-art devices for fiber-optic communications applications.

2.1.2.4 (Continued):

Figure 2.1.2.4-2 lists a number of semiconductor materials used to make fiber-optic sources, but for most high speed data bus applications, only GaAlAs and InGaAsP are viable. Their characteristics are emphasized in the Figure 2.1.2.4-1 data. Data bus applications in general require high radiance (high power) sources. ILDs can provide the greatest power into an optical fiber but their peculiar characteristics, instability, and relatively low lifetime currently preclude their use for this application. The best choice is a GaAlAs surface-emitting LED at 850 nm followed by a InGaAsP surface-emitter at 1300 nm. The latter has somewhat better total dose radiation resistance and is preferred if sufficient optical power is available for the system design.

Detector technology is more mature than source technology. Two semiconductor photodiode types are suitable for fiber-optic applications. They are pin photodiodes and avalanche photodiodes (APDs). Both types are basically pn junction diodes (as are the LEDs and ILDs) with an undoped "intrinsic" region in the middle. Both types are usually reverse-biased to full depletion for use; light striking the junction creates electrical current flow which is then amplified. The avalanche photodiode provides gain within the diode reducing the external amplification requirement, but it is more temperature sensitive.

Figure 2.1.2.4-3 provides a concise summary of the characteristics of typical silicon pin-type and avalanche photodiodes for short wavelength use, and InGaAs pin-type photodiodes for short or long wavelength fiber-optic systems. Except for higher dark current and higher cost, the InGaAs photodiode is generally preferred. It also exhibits somewhat greater resistance to the undesirable effects of large total doses of ionizing radiation. The APD has poorer radiation resistance, high temperature sensitivity, need for a high voltage reverse bias power supply and stabilization circuitry, and unneeded gain advantage. If the intrinsic dark current can be tolerated and does not dominate the noise output of the receiver, it is the preferred device. When the photodiode is suitably connected to a receiver carefully designed for the specific application, the net performance of the combination typically varies with the NRZ bit rate, that is with the receiver bandwidth required to support a given bit error rate (BER), as shown in Figure 2.1.2.4-4. A BER of 1 error in 10^9 bits is frequently used in digital fiber-optic system specifications. The average (or peak) optical power level required at the detector input for a given BER is called the receiver's sensitivity. It depends on detector responsivity (notice the difference between Si and InGaAs performance), receiver type and characteristics, clock recovery requirements (if any), received waveform characteristics (pulse width distortion, jitter, etc.), interconnect component losses and distortion contributions, and transmitter characteristics.

FIGURE 2.1.2.4-1 - Comparison of Typical LEDs and ILD

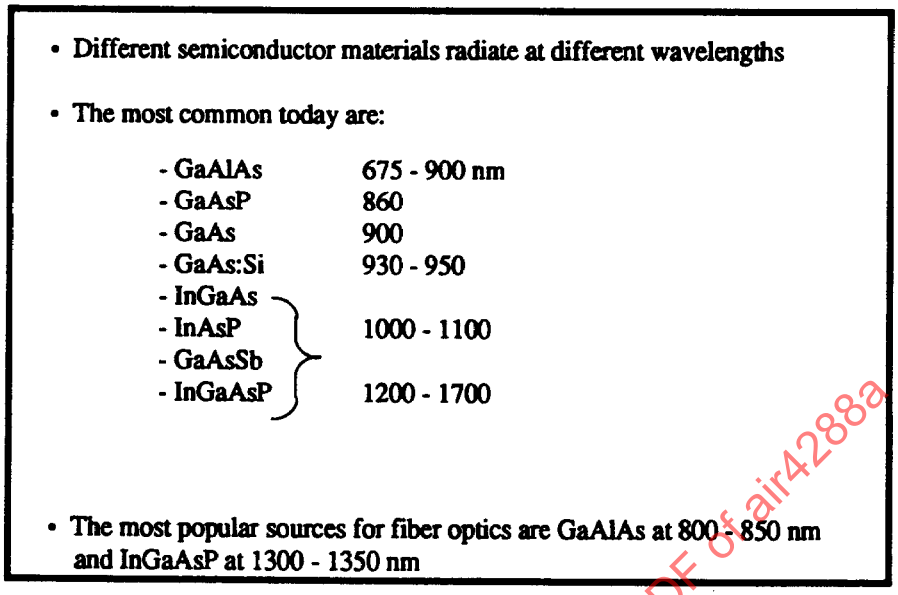
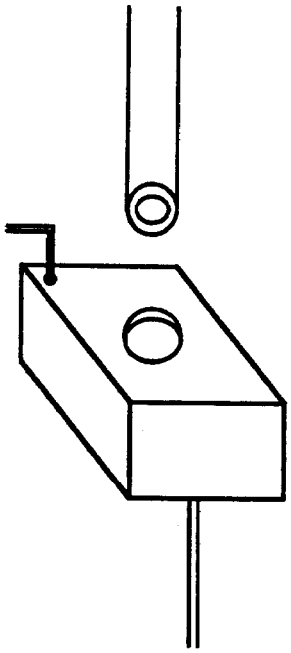


FIGURE 2.1.2.4-2 - Optical Source Wavelength Selection



	SI PIN	SI APD	InGaAs PIN
ACTIVE DIAMETER (µm):	300 - 1000	500	5 - 150
DARK CURRENT (nA):	1 - 10	15	0.5 - 4.0
CAPACITANCE (pf):	1 - 3	1 - 2	0.5 - 1.0
RISE/FALL TIME (ns):	1 - 3	0.5	0.5 - 1.0
NEP (W/√Hz):	$1x10^{-13}$	$2x10^{-13}$	$3x10^{-13}$
RESPONSIVITY (A/W)			
At 850 nm:	0.3 to 0.6	5 - 80	0.3
At 1300 nm:	Not Usable	Not Usable	0.7
At 1550 nm:	Not Usable	Not Usable	0.8
TEMPERATURE - RESPONSIVITY (%A/W/°C):	0 - 0.1	5 - 60	0 - 0.2

FIGURE 2.1.2.4-3 - Detector Characteristics

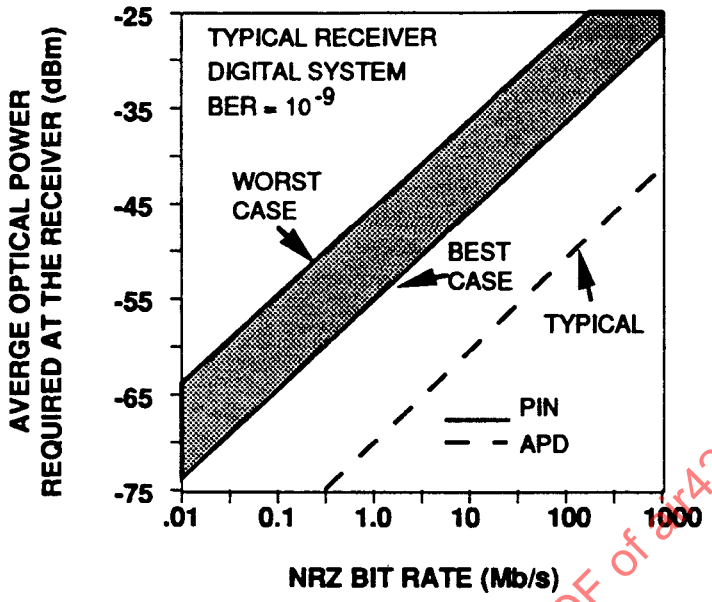


FIGURE 2.1.2.4-4 - Average Optical Power Required at Detector for Digital Systems

2.1.2.5 Summary: Except for a few issues over which there is disagreement, such as static fatigue, the use of antibuckling elements, excessive fiber strain, and the maximum practical fiber size, it is generally agreed that a tight tube construction is preferable since fiber loss in the fractions of decibels per kilometer is not necessary. This type of fiber has been successfully used in military aircraft including a current application in a product setting. At least three European aircraft manufacturers are confident in its suitability and are using fiber-optics on military vehicles. Although not in production yet, their applications are more aggressive than any undertaken in the United States to date in the sense that the fiber-optic systems operate at high data rates and employ extensive use of the technology. With a much greater likelihood than ever that fiber-optics will be used on the next air vehicle and on the next generation of spacecraft, manufacturers of all components are picking up their development activity for specialized parts. This is most obvious in the connector world. Much lower loss single-channel connectors are now being introduced and the militarization of them is underway. More suppliers of single-channel contacts for connectors such as the MIL-C-38999 type are emerging and some innovative approaches for contacts, such as the "fiber-lens" concept, are being applied. All of the latest generation of connectors and contacts have greatly reduced losses and repeatability compared to the standard connector of just a few years ago. A low cost, dependable fiber-optic contact for aircraft and spacecraft applications is on the horizon.

For interfaces with large numbers of data channels such as at bulkheads, the imaging connector is extremely attractive. It is a robust interface being tolerant of most sources of contamination and able to be mated and remated thousands of times with no change in performance since there is only a single wear surface which is relatively large. Some improvements in this connector's losses is still required for certain applications, and its low cost producibility needs to be exploited.

2.1.2.5 (Continued):

Optical source and detector technology has reached a high level of maturity. Detector technology is especially well positioned. Long wavelength systems lag short wavelength systems in certain respects, but are rapidly catching up. Integrated circuit technology needed to support the optoelectronic interface: the transmitter, receiver, and clock recovery function is available. The specific circuit designs can be especially sensitive to overall system requirements. For example, a small and seemly insignificant change in the higher protocol layers of data bus architecture can have a profound effect on the physical layer which includes the transmitter and receiver. It is now possible to proceed with further specific design and integration of these components. When taken to completion, a high reliability product will result.

2.1.3 Fiber Optic Receiver Considerations: The fiber-optic receiver detects light and converts it into an electrical signal. For a data bus such as AS4074, the receiver design is complicated by requirements for detection of low level optical power, wide dynamic range, and rapid acquisition time. The system requirement is for a receiver which meets or exceeds a specified BER for a given minimum received optical power. This minimum optical power (defined as sensitivity) is usually stated in dBm (peak), where 0 dBm equals 1 milliwatt. While optimum sensitivity is the basic goal in receiver design other considerations require trade-offs to achieve the best balance in performance while meeting the requirements of the standard.

A block diagram illustrating a star-coupled LTPB bus is shown in Figure 2.1.3-1. Figure 2.1.3-2 illustrates the physical configuration of a representative bus, and a block diagram of a receiver (including the clock recovery unit) is shown in Figure 2.1.3-2. For purposes of an example, this discussion considers a passive star coupled fiber-optic system. As can be seen from Figure 2.1.3-1, this typical application consists of bus interface units interconnected via a star coupler, connectors, and fiber-optic cables. It is the receiver which must overcome the losses introduced by these components. While other bus configurations are possible Figure 2.1.3-1 can be used to describe the properties of the bus which affect the optical receiver and as an aid in understanding terms used in the standard which apply to receiver design. The following paragraphs are intended to define terms unique to the standard, provide some insight into the derivation of receiver parameters, and present several design concepts.

Detailed receiver design information can be found in several excellent sources (Appendix A, Reference 1).

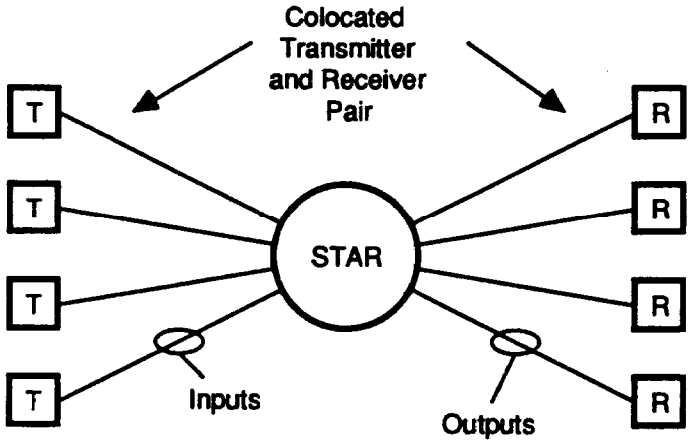


FIGURE 2.1.3-1 - A Star-Coupled Linear Token-Passing Bus

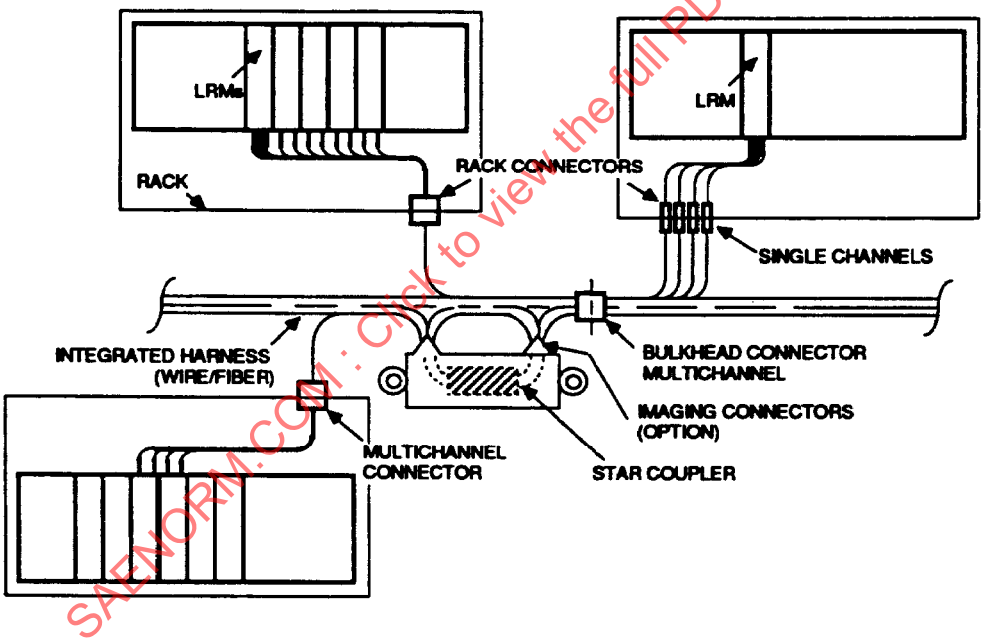


FIGURE 2.1.3-2 - Typical Aircraft Installation of a Star-Coupled LTPB

2.1.3 (Continued):

The bus operates by successive transmitters sending bursts of information of varying length. A minimum length of time between successive receptions (system minimum intertransmission gap) is specified to ensure the quality of the transmission (that is, guarantee that there are no cases of a particular receiver receiving simultaneous transmissions from two or more sources). Thus, any receiver gets successive receptions via different paths with different losses. The maximum range of peak optical power which can occur for this situation in a particular design is called the intertransmission dynamic range (IDR). In addition if a single receiver design is to be used at all terminals it must handle an even greater range to accommodate all transmissions (not just successive ones) over all operating conditions, from the best case to the worst case. This greater range is defined as the receiver operating range or ROR. The ROR must be greater than the IDR since the effects of environmental conditions, aging, etc. have the effect of shifting the IDR within the ROR (see Figure 2.1.3-3). The IDR then, is the difference between the paths with the highest and the lowest losses; while the ROR is the required operating range of the receiver under all conditions.

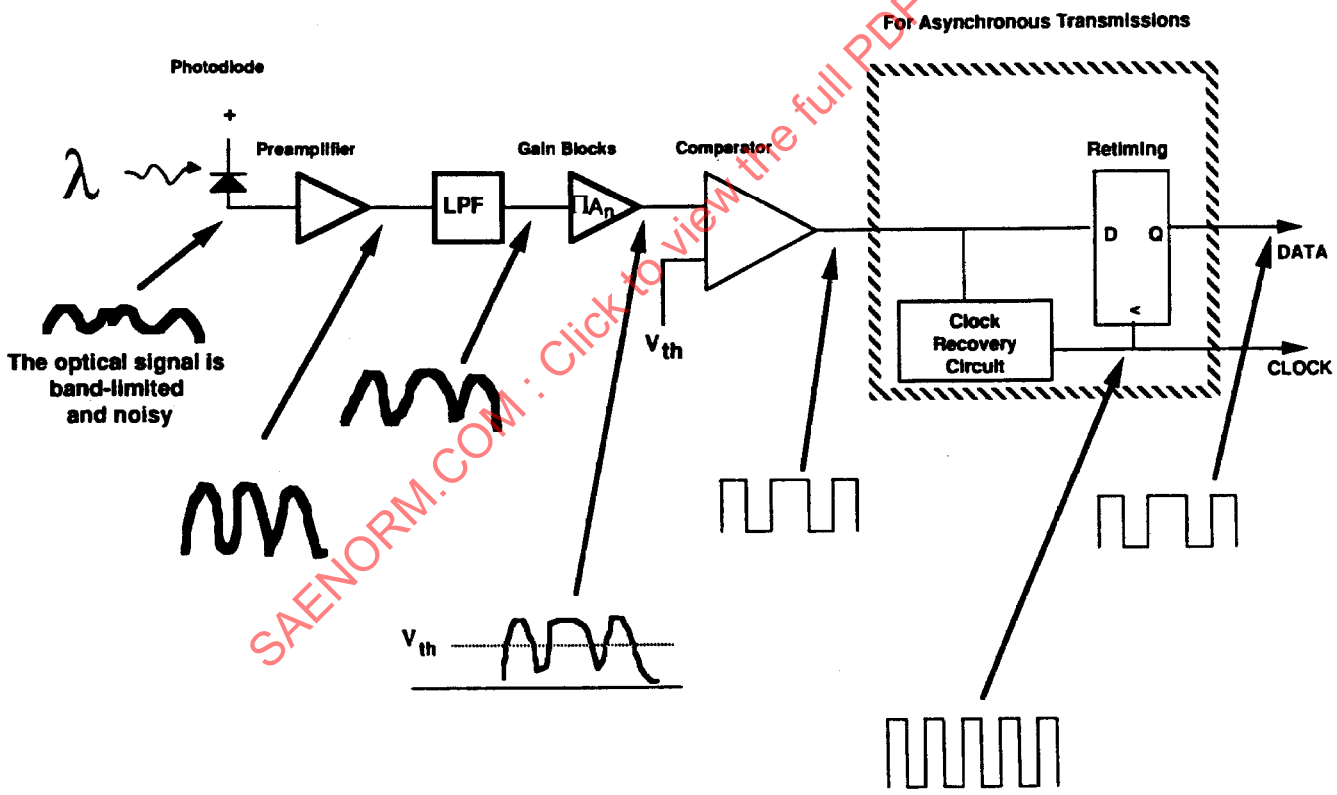


FIGURE 2.1.3-3 - Block Diagram of a Typical Fiber Optic Receiver

2.1.3 (Continued):

In order to minimize the impact on receiver design the standard specifies a minimum loss (A_{min}) of 10 dB (for type F-1 fiber-optic media characteristics). This dictates that the minimum optical attenuation in any path be at least 10 dB and the following data suggests that for a bus configuration of less than 8 ports consideration should be given to ensuring at least 10 dB in every path. Similarly the standard specifies a maximum optical attenuation (A_{max}) in any path of 28 dB. Therefore, the sum of all losses in any path cannot exceed 28 dB and should be less than this if the designer chooses to have margin in his system to account for increased link loss due to changing conditions (e.g., aging). Losses in the bus consist of those losses encountered in connectors, the fiber-optic cable, and (for the bus shown in the example) the splitting device. In a given bus configuration the BIUs are optically separated by the fiber-optic cable with cable loss a function of the length of the cable. Any number of connectors may be in the path with nominal losses per connector ranging from 0.5 dB to (say) 2 dB. The fiber optical splitting device introduces a splitting loss of:

$$\text{Splitting Loss} = 10 \log N$$

where: N is the number of ports.

A 64 port star coupler (which would support to a 64 BIU bus) has a splitting loss of about 18 dB, while an 8 port coupler has a splitting loss of 9 dB. The link margin can be calculated by the equation:

$$\text{Margin} = ((\text{transmitter power-link loss}) - (\text{receiver sensitivity}))$$

Thus a link with 8 connectors at 0.5 dB each, a 64 port coupler with an insertion loss of 3.5 dB, and 100 meters of cable at 0.005 dB/meter has a total loss of 26 dB. With a transmitter output power of -1.75 dBm this link would have a margin of:

$$\text{Margin} = -1.75 - 26 - (-32.5) = 4.75 \text{ dB}$$

Figure 2.1.3-4 details these values.

The data bus requires a receiver with high sensitivity and rapid acquisition. These parameters are difficult to obtain simultaneously due to the optical bus characteristics of unipolar signaling, wide dynamic range, and burst mode operation. With unipolar signaling a reference which will permit bit decisions to be made must be established. This threshold voltage must be a function of signal amplitude and, therefore, some method of adjusting this threshold to various signals of greatly different amplitude must be included in the design. Accurately and dynamically selecting thresholds over a wide dynamic range is a major design challenge. Figure 2.1.3-5 shows three possible concepts for the preamplifier.

Transmitted Power	**-1.75 dBm**
LINK LOSS Coupler Splitting Loss Coupler Insertion Loss Connector Loss (8 ea.) Cable Loss (100m)	 18.0 dB 3.5 dB 4.0 dB . 5 dB
Total Loss	**26.0 dB**
Receiver Sensitivity	**32.5 dBm**
Margin	**4.75 dB**

Margin =(Transmitted power - Link Loss - Receiver Sensitivity)

FIGURE 2.1.3-4 - Power, Loss, Margin Calculations

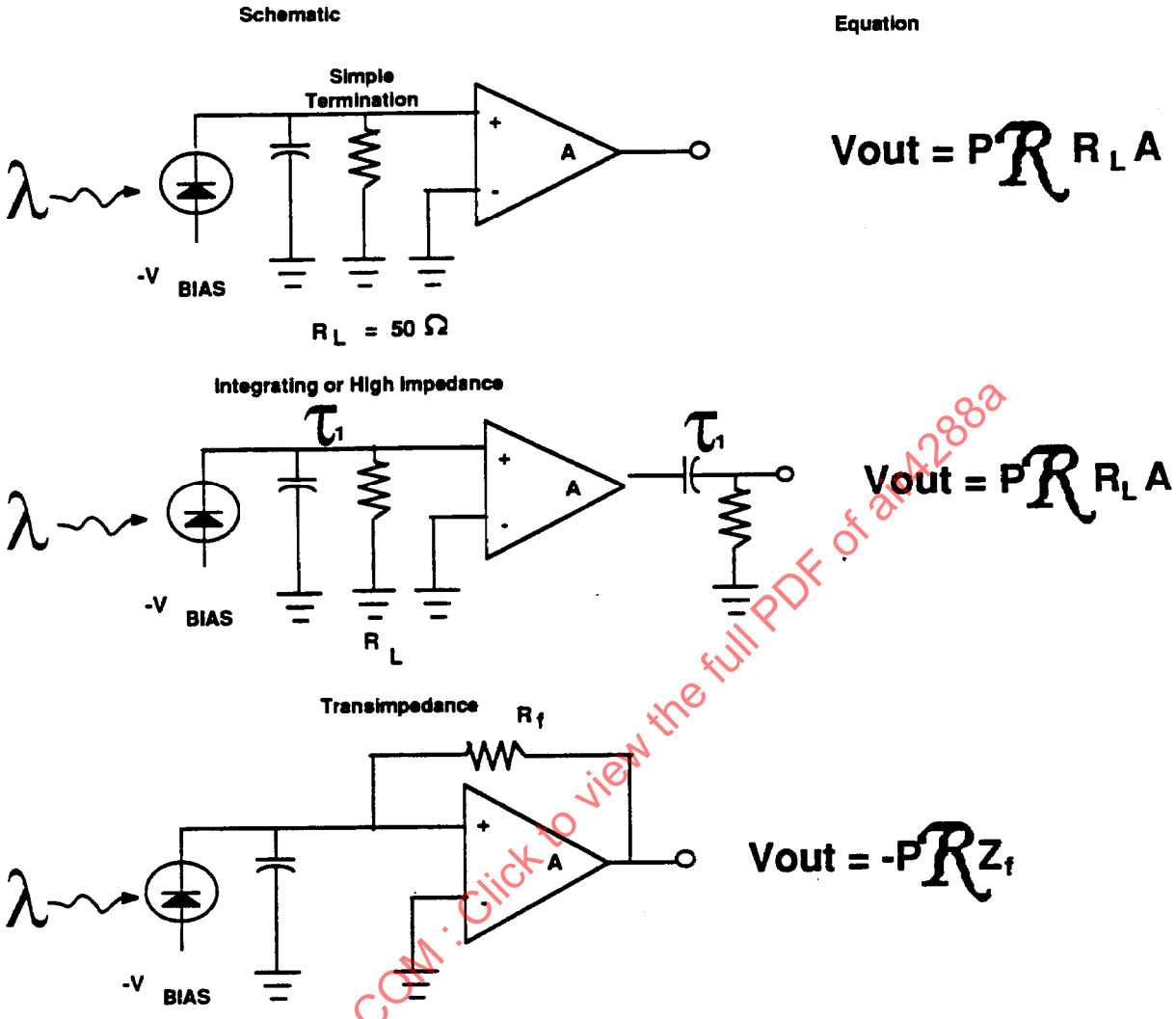


FIGURE 2.1.3-5 - Three Approaches to Design of a Fiber Optic Preamplifier

2.1.3 (Continued):

The DC-coupled receiver is designed such that the output signal to the bit detection circuitry is zero volts for zero light input and, with a light pulse present, the signal is proportional to the light intensity. Since the negative peak value of the signal is fixed at zero volts, the threshold voltage may be derived as one-half of the peak positive signal voltage. For systems operating at 100 MBaud, this receiver works well with moderate to large light levels, but sensitivity is limited. The combination of detector leakage current (which is indistinguishable from signal current), bias current, and amplifier input offset voltage combine to require input light levels much higher than those available under worst case conditions. Circuitry can be added to improve one parameter (e.g., sensitivity) at the expense of another parameter (e.g., initialization). However, if direct-coupled amplifiers are to be used in this application, input devices with very low leakage currents are required.

Another approach is the high impedance or integrating amplifier. In this design, noise is minimized, especially thermal noise contributed by resistors. Thus, R_L is made as large as possible (see Figure 2.1.3-5), and the result is an amplifier with high sensitivity (the highest of any of the designs) and minimum noise. Unfortunately, this design also has the worst dynamic range of those considered for this application.

A third approach to the design of the preamplifier is the transimpedance amplifier. As shown in Figure 2.1.3-5, Z_f is the effective feedback impedance from the output to the input of the amplifier. This amplifier is often used because it is capable of wide bandwidth, is almost as sensitive as the integrating amplifier, and provides a wide dynamic range. The drawbacks of the transimpedance amplifier are that it is noisier than the integrating amplifier (but only slightly) and propagation delay limits high frequency response.

An example of a high speed fiber-optic receiver designed for maximum sensitivity and usable over a wide temperature range is the AC-coupled design shown in Figure 2.1.3-6. It removes the large leakage currents likely to be encountered in the photodetector and input transistor. Compensating components (such as another FET) are not stable enough to remove large offsets to the required level. Active circuitry requires an interval of zero light to develop a zero current reference. This interval is impossible to detect with an uncalibrated circuit and it leads to difficulties in initialization of the network. Thus, the AC-coupled receiver offers distinct advantages for this standard applied to systems with severe operational requirements.

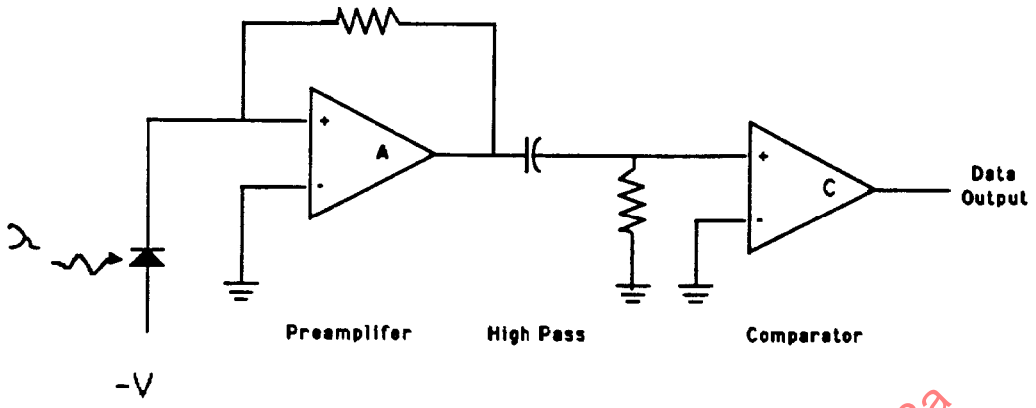


FIGURE 2.1.3-6 - Simplified Schematic of an AC-Coupled Receiver

2.1.3 (Continued):

In addition to solving the leakage current problem, AC-coupling also converts the unipolar optical signal into a bipolar signal which is simple to detect with a zero referenced comparator. For an AC-coupled amplifier to work several conditions must be met:

a. The optical signal must have nearly (but not necessarily exactly) an equal number of ones and zeros.

b. Significant time must be allowed for the receiver time constants to decay to their final steady-state value and that is dependent on the amplitude of the received optical signal.

c. The optical signal must not go without a data transition for a time which is long compared to the receiver time constant.

d. A preamble is required for the receiver time constant to decay to a value appropriate to the incoming data.

These conditions are all considered in the standard but they place constraints on the receiver design. The receiver must acquire the signal with the specified intertransmission gap and preamble while limiting pulse droop (Figure 2.1.3-7) to a value consistent with the coding technique employed. The pulse droop can be expressed by equation 2.1.3.1:

$$D = 1 - \exp(-NT/RC) \qquad (2.1.3.1)$$

where: T is the time per bit, N is the run length in bits, and R and C are as shown in Figure 2.1.3-6.

Solving equation 2.1.3.1 for RC yields:

$$RC = -NT/\ln(1 - D) \qquad (2.1.3.2)$$

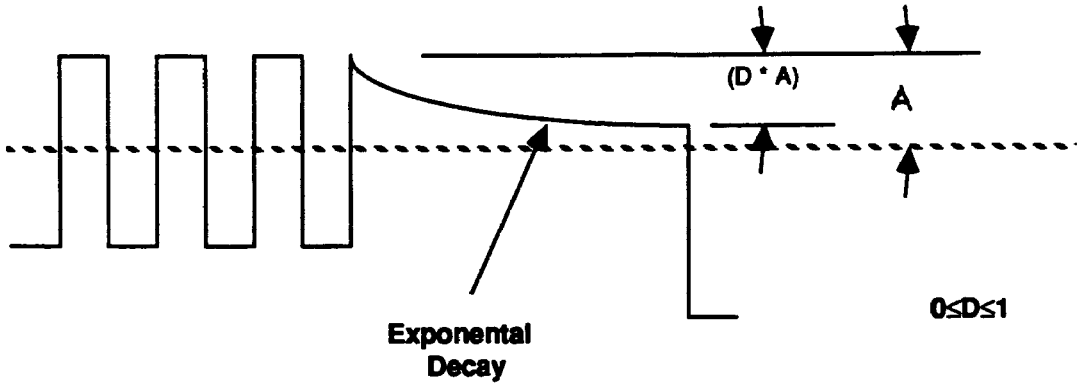


FIGURE 2.1.3-7 - Illustration of Pulse Droop

2.1.3 (Continued):

Acquisition time is illustrated in Figure 2.1.3-8. A short acquisition time requires a short receiver time constant so pulse droop and acquisition time have to be traded accordingly. Acquisition time can be expressed by equation 2.1.3.1:

$$A = (-RC/T)\ln(E/(K-1)) \qquad (2.1.3.3)$$

where: A is the acquisition time in bit times, K is the instantaneous dynamic range, and E is the threshold error expressed as a decimal percentage of peak signal amplitude.

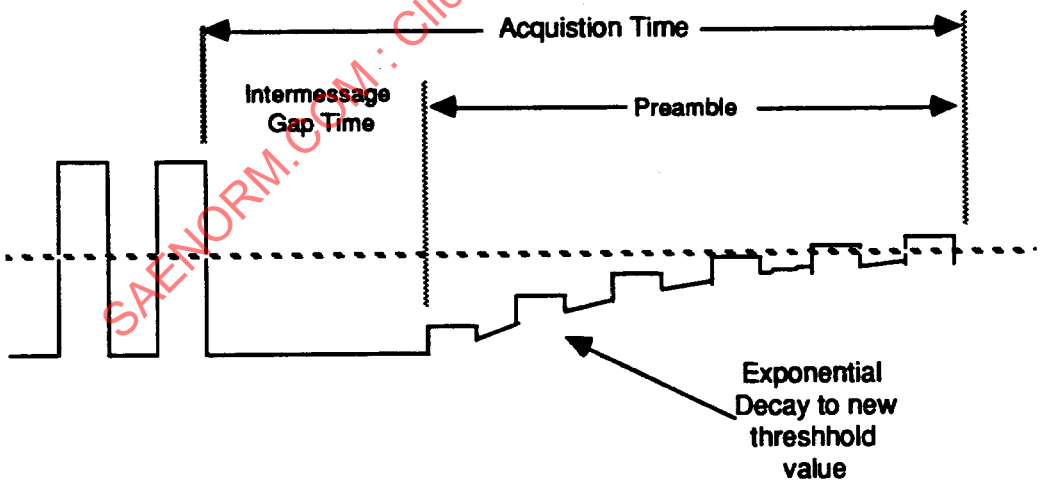


FIGURE 2.1.3-8 - Illustration of Signal Acquisition Time

2.1.3 (Continued):

Combining these two equations and setting E equal to D (equal threshold errors for droop and acquisition):

$$A = N\{\ln[D/(K-1)]/\ln(1-D)\} \quad (2.1.3.4)$$

Since acquisition time (intertransmission gap and preamble) and run length are fixed by the standard, equation 2.1.3.4 can be used to select the receiver time constants once acceptable limits on droop are determined.

2.1.3.1 Preamble Requirements: As discussed in 2.1.3, the unique means of detection used in fiber-optic receivers places special demands on the preamble specification. Because active star couplers may not transmit the same number of "good quality" preamble bits as they receive, the standard specifies both the transmitted preamble length (P_t) and the minimum length of good quality preamble bits that the receiver should expect (P_r). Because of distortion in some active star couplers, the beginning of the preamble may be of poor quality (i.e., may not meet the waveform requirements of 3.2.2.1.3.4 of the AS4074 standard). It is likely, however, that there would be sufficient transitions at the optical power level of the incoming frame to adjust the receiver thresholds and sufficient valid preamble bits in P_r to allow clock synchronization. Because the standard is implementation independent (i.e., does not specify the use or nonuse of any particular active star coupler design, nor any particular optical receiver design) these limits must be specified so that designs which meet the standard will work with any other design that meets the standard.

2.2 Wire-Based (Coaxial) System Implementation:

The AS4074 standard has been designed with implementation flexibility as a consideration. The discussion in this section addresses two potential topologies and the engineering trade-offs which must be made when designing a system using this medium as a design solution.

2.2.1 Topology Options: When utilizing coaxial cable as a media for the LTPB system, there are two basic topologies which may be used. The first is quite familiar to most designers and is the basis of the MIL-STD-1553B bus design - this is the bidirectional trunk bus design. It is a single main bus - or trunk line, to which all stations are coupled. The coupler, acts to match the impedance of the terminal stub (including the terminal itself) to the main trunk line in order to introduce as little loss and impedance mismatch (reflections) to the signal path as possible. Figure 2.2.1-1 details just such a system.

2.2.1 (Continued):

Another approach is to use directional couplers to implement a trunk bus system. Figure 2.2.1-2 shows a system designed using this approach. In this approach, directional couplers are used to match the terminal stubs to one of the main trunks. One coupler is used for the transmitter interface and another is used for the receiver interface. At first glance, this seems to be a substantial amount of additional hardware. The couplers are, however, simpler and easier to design than bidirectional couplers required by the previously described topology. When analyzing the behavior of the system, however, you discover that the use of this type of coupler introduces a high level of fault tolerance at the physical layer. The directional coupler allows the signal on the media to only go one direction. Another advantage of the directional coupler approach is that with the use of a central repeater a higher signal level can be maintained on the bus. This will result in a higher signal level at the input of the receiver which reduces the effects RFI and EMI will have on signal-to-noise ratio.

Both topologies discussed suffer from the same problems - primarily, the electrical characteristics of coax cable at higher frequencies, especially on a long main trunk line. Most notable of the problems are the direct relationship between frequency and loss, and the distortion of the signal (zero crossings) as the signal propagates down the media. The distortion is due to the effects of capacitance and inductance inherent in the media. The directionally coupled system seeks to overcome this problem by installation of repeater amplifiers which decode the incoming signal and then regenerate it at full transmitter power on the receive portion of the coax. This approach imposes additional hardware, power, and weight demands and increases propagation delay in the transmissions.

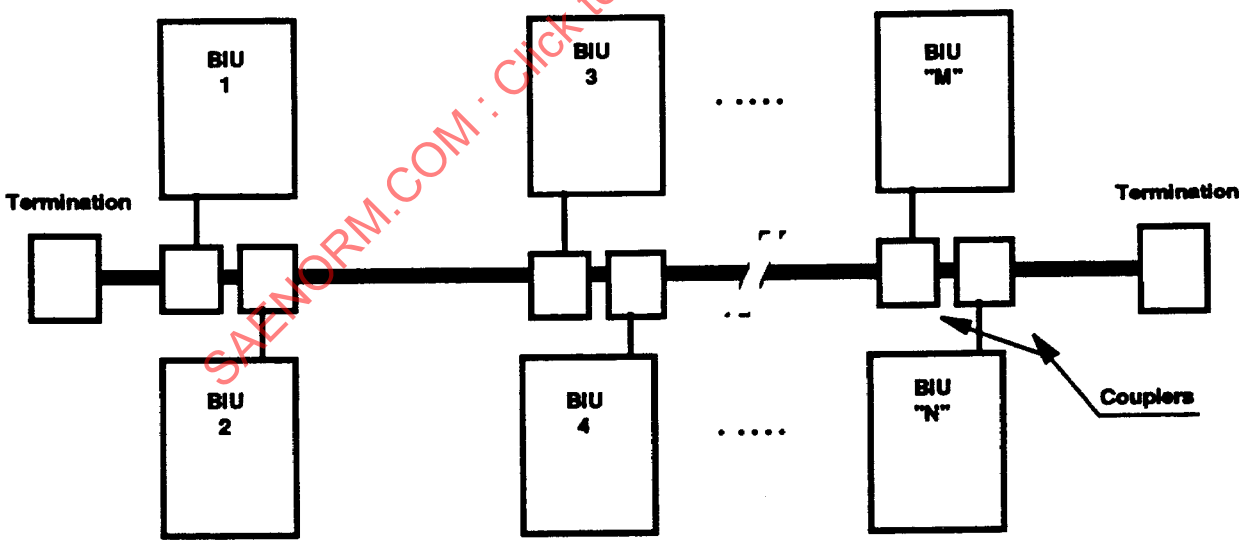


FIGURE 2.2.1-1 - Bidirectional Trunk Bus Topology

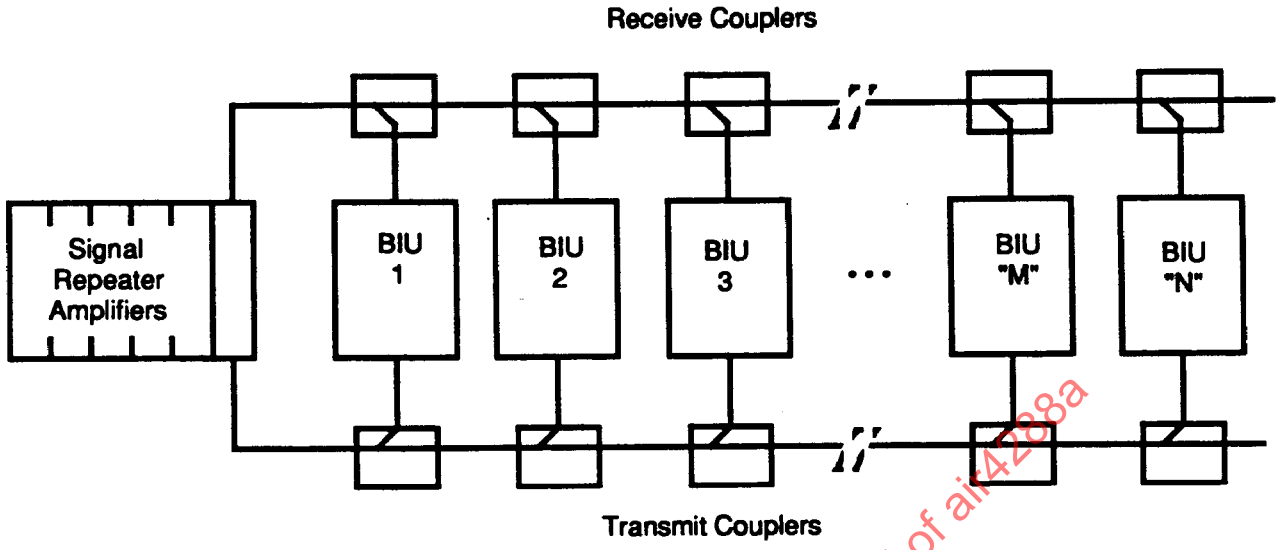


FIGURE 2.2.1-2 - Directionally Coupled Trunk Bus Topology

2.2.1 (Continued):

With either topology, design attention must be given to ensure that reflections on the bus (such as those created if battle damage cuts a coax link) are severely attenuated and cause very little, if any, disturbance to the received signal upstream. As a design requirement, should the downstream portion of the bus be lost due to the damage, the upstream portion must continue to function.

2.2.2 Power Budget: In much the same way as a fiber optic implementation, the designer must carefully plan the design of his system to make certain that adequate power levels are available throughout. This is less of a technical problem than for fiber optic networks, however, because of the more flexible technology involved. For example, transmitter power can be much greater, receiver sensitivity and dynamic range can be much greater, and couplers can be designed with a wide range of coupling factors. Figure 2.2.2-1 shows a typical system design for a trunk bus LTPB, utilizing bidirectional coupling techniques.

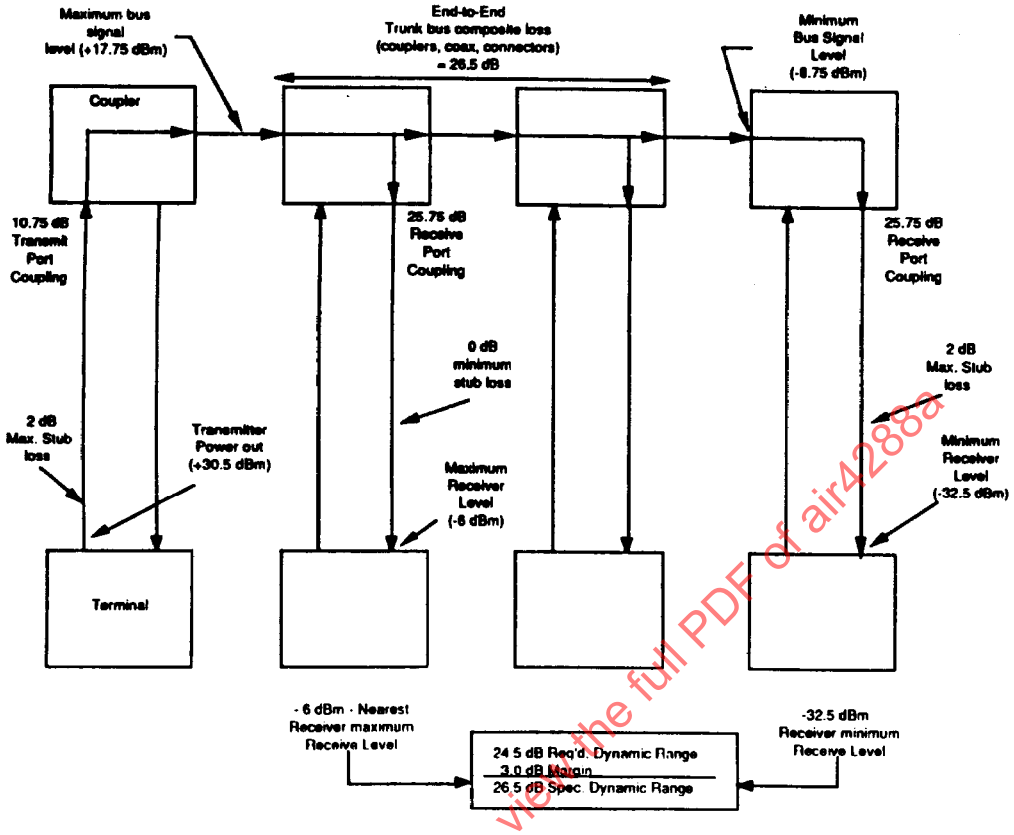


FIGURE 2.2.2-1 - System Power Budget Calculation

2.2.2 (Continued):

Notice the issues that must be considered when designing the system. First and foremost, the system designer must determine the power budget of the system. This is a function of the number of network taps required, the characteristics of the couplers, and the loss of the interconnecting coaxial cable. There will be three principal components to this loss:

a. Loss on the main trunk bus
b. Coupling loss from main trunk bus onto the stub to the receiver
c. Coupling loss from transmitter onto the main trunk bus

Normally the network design process begins by defining the worst-case receiver input signal level. This is the receiver sensitivity requirement. In the example case, the receiver sensitivity has been set as -32.5 dBm. From there the power budget sets a minimum signal level on the bus. This allows us to determine the minimum power level at which the transmitter must launch the signal onto the bus.

2.2.2 (Continued):

The first component of the budget occurs when the signal is tapped off the main trunk and fed into the receiver stub. The loss in the system detailed above for this component equals 25.75 dB. Next the stub attenuation must be taken into consideration. The sum of these losses determines the minimum signal level which must be present on the trunk bus in order for the system to meet the minimum input level requirements of the receiver. In our example, the receiver design can not handle signals greater than -32.5 dBm so the minimum bus signal level is -8.75 dBm.

Next we determine the loss requirements of the trunk bus. In the example given, we show a coupler throughput loss (loss of signal amplitude) on the main trunk (as the signal traverses the coupler) of 0.1 dB. For each coupler in the system then, we will experience a 0.1 dB loss in signal amplitude as it propagates through the trunk from one end to the other. Assuming that we are designing for a maximum of 64 couplers the total loss is 6.4 dB. If each coupler requires a pair of connectors of 0.05 dB loss, this is an additional 6.4 dB. Accumulated loss of signal amplitude due to losses in the coaxial cable itself will vary depending on the media being used. In our example we have allocated 13.7 dB for coax loss measured at the highest frequency of the signal waveform. This can be met using any one of the different types of coaxial cable. The system engineer must select coax compatible with the size, weight, and reliability of the host platform.

We are now prepared to determine the minimum signal level required on the trunk bus. It can be determined knowing the minimum signal level at the receiver trunk and the maximum trunk loss. In our example this becomes +17.75 dBm. This computation is important for two reasons: it determines the required receiver dynamic range and it determines the transmitter output power requirement.

A receiver stub coupled from the trunk at the maximum signal level point will present a level of -8 dBm to the receiver stub. We have postulated a situation in which we have no loss in the stub (ideal case). The entire loss, in our case, comes from the coupler insertion loss on the receiver side of the power budget. In this example, the required receiver dynamic range is the difference between the minimum receive level and the maximum receive level of 24.5 dBm. We have also added a design margin of 3 dB to arrive at our specification value of 26.5 dB dynamic range.

The requirement for transmitter output power is determined knowing the worst case loss between the transmitter output port and the trunk bus. It consists of the coupling factor of the transmitter port of the coupler summed with the worst case attenuation introduced by the stub connecting the transmitter to the coupler. In our example, the coupler factor for the transmitter port is -10.75 dB and the loss due to the stub connecting the BIU transmitter to the coupler port is set at worst case 2 dB. This means that the transmitter in our example must deliver a minimum of +30.5 dBm at its output port to satisfy the power budget requirements for this system.

2.2.2 (Continued):

The example is, at the very least, overly optimistic. Mismatch due to coaxial cable parameter variations, connector losses and mismatches, coupler variations, transmitter power variations, aging of components in the system (due to temperature extremes, vibration, exposures to chemicals and other agents) must be planned for during design of a real system and requires that a design margin be introduced into the design. In the case of our sample system, we have chosen a 3 dB margin for receiver dynamic range. Other margins are appropriate, as well. The individual designer should evaluate the worst case conditions under which his system will be operated and select a design margins which is compatible with those conditions.

2.2.3 Waveform Distortion: Signals on a long coaxial medium will, by the nature of the media, incur some distortion as they traverse the length. This distortion is composed of several elements:

a. Distortion due to reflections from mismatch
b. Distortion due to differential propagation delay
c. Distortion due to R-L-C characteristics
d. Distortion due to RFI/EMI susceptibility

2.2.3.1 Distortion due to Reflections: Discontinuities on coaxial media, even those which produce only a slight impedance mismatch, will result in reflection of signal voltage back into the line from which the signal is propagating. The polarity and amplitude of the reflected signal depends upon the magnitude of the mismatch. The resulting waveform distortion will be the sum of the reflected and incident signals based on their time relationships along the media. If the magnitude of the mismatch is large enough, significant distortion of the waveform, in the form of intersymbol interference, can occur. This directly impacts the bit error rate and the proper operation of the bus. The bus couplers, terminations, and impedance characteristics of the coaxial media being used are, therefore, critical to the success of implementing this type of bus. This factor is especially critical in the design of bidirectional trunk bus networks since reflections from both ends of the bus are superimposed at each network node.

2.2.3.2 Distortion Due to Differential Propagation Delay: If one were to use traditional Fourier Transform techniques on the pulse train which results from the transmission of a serial digital data signal on a data bus, it would be discovered that each pulse consists of a summation of sinusoids of varying harmonics of the fundamental signal frequency, each of varying amplitude and phase. A fundamental characteristic of coaxial media is that each of these different frequencies propagate in the media at a different speed. The difference in propagation delay (differential propagation delay) can cause distortion of the signal, especially when it must travel along a long length of line. This can result in increases in rise and fall times and zero crossing shifts. These characteristics should be taken into account when planning the system, especially in receiver design.

2.2.3.3 Distortion Due to R-L-C Characteristics: A pulse signal applied to the input of a coaxial cable will suffer distortion as it travels along length of the media. This distortion is due to the R-L-C characteristics of the media. The lumped element equivalent circuit of a coaxial cable transmission line is depicted in Figure 2.2.3.3-1. Note that the signal will incur losses due to the resistive component of the model as well as a frequency component degradation due to the low-pass filter characteristics of the inductance and capacitance. The complex R-L equivalent network represents the skin effect, a phenomenon which must be modeled to completely understand how waveform distortion results. The waveform takes on smooth transitions (which correspond to the actual transitions in the input signal) and, depending on the duration and amplitude of the pulse train, may or may not make a zero crossing at the appropriate time relative to the input signal. Some portions of the pulse train may not cross zero at all. Therefore, the use of zero crossings in the receive circuitry of stations on relatively long coaxial trunk lines may be impossible.

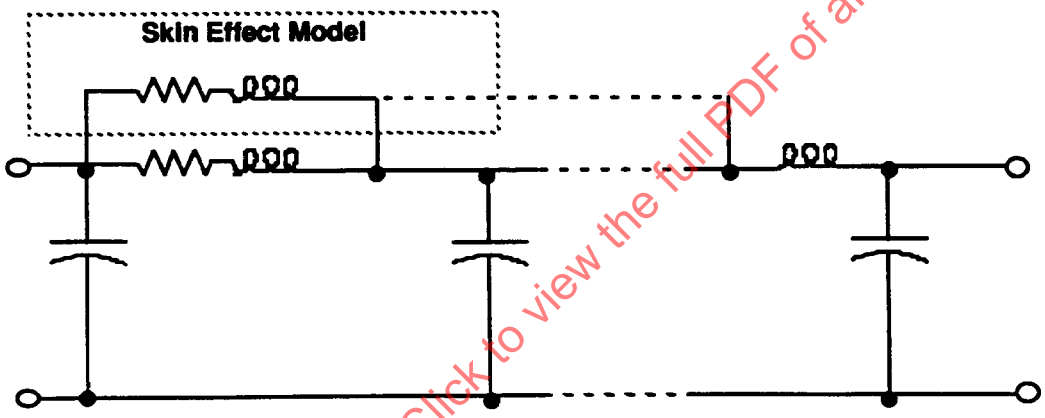


FIGURE 2.2.3.3-1 - Lumped Element Equivalent Circuit of a Coaxial Transmission Line

2.2.3.4 Distortion Due to RFI/EMI Susceptibility: A network designed using coaxial cable interconnect is susceptible to distortion effects caused by pickup of radiated noise from co-located electrical equipment and from external sources. This noise component acts to reduce the signal-to-noise (S/N) ratio of the signal applied to the receiver and, therefore, increases the network error rate. The severity of this effect is a function of the number and radiating flux of the other equipments and the shielding characteristics of the network interconnect. Assuming a minimum receiver signal level of -32 dBm and a required 25 dB S/N ratio in order to meet the bit error rate specification, the maximum noise allowed in the frequency band of interest (2 MHz to 73 MHz) would be -57 dBm. Traditional EMI/RFI design techniques can be used to select coaxial cable type and network routing to satisfy this requirement.

2.2.4 Hardware Design Considerations:

2.2.4.1 Transmitter Design Considerations: Design of a transmitter for a coaxial based system requires that the designer investigate two main areas:

a. Output drive capabilities
b. Signal rise and fall time requirements

2.2.4.1.1 Output Driver Characteristics: After planning the system and generating a plan for the power budget, the transmitter output stage may be designed. The design must be capable of delivering the required signal level into the coupler and have a frequency response capable of meeting the bandwidth requirements of the system.

In addition, the design must be capable of meeting the required circuitry protection specifications. That is, if a short circuit occurs on the media, between the transmitter and coupler, or in the coupler itself, the transmitter output circuitry should suffer no damage. Some type of current limiting circuitry should be incorporated into the output stage to protect the transmitter when current through the power driver circuitry exceed design limitations. The transmitter must also be protected such that if an electrical surge occurs on the media, the transmitter circuitry is not damaged.

2.2.4.1.2 Transmitter Quiescent Characteristics: A transmitter should be designed such that when it is not activated to transmit a message or token, it should be off-line, that is, it should not contribute significant energy to the bus. The transmitter should not create any noise which might decrease the signal-to-noise ratio on the trunk. But another consideration is that the transmitter should remain as close to the zero voltage level as possible (no DC value). This is important, especially in transformer coupled circuits, where DC offsets are to be avoided.

2.2.4.2 Receiver Characteristics: The receiver typically consists of a linear gain preamplifier which drives clock recovery and data recovery circuits. The preamplifier stage requires technologies quite similar to those required in the design of any VHF receiver. Important characteristic are linearity, sensitivity, dynamic range, and input impedance, The stages in the following paragraphs are somewhat nontraditional and are described in more detail.

2.2.4.2.1 Clock Recovery: The most critical circuit in the BIU receiver circuit is the clock recovery circuit. This circuit must detect the presence of transitions on the bus (usually corresponding to the preamble of the message) and cause the generation of a local clock signal. This local clock will be used to sample the incoming pattern of transitions from the bus and cause detection on start and end delimiters, frame control information, source and destination addresses, and message information. All of this information must be decoded properly, that is, it must be detected and translated back into NRZ information matching that which was transmitted. This requires the clock to lock to the exact frequency and phase of the clock used by the transmitting station. This lock must be achieved rapidly (before the end of the preamble) and must be immune to the lack of transitions inherent in the start and end delimiters as well as any noise which might be present in the system.

2.2.4.2.1 (Continued):

The designer should note that there are a number of clock recovery schemes which may be utilized in the design and implementation of a BIU clock recovery circuit. Phase lock loops are a typical choice. They are simple to design and can normally be constructed from standard integrated circuits. The principal negative aspect of the phase locked loop approach is the difficulty in balancing the requirement for rapid acquisition against the ability to maintain adequate noise immunity. Use of a ringing tank is another approach which has been successfully used. This usually requires more board space due to the lack of a standard integrated circuit device, but can be designed to provide superior performance. The principal technical challenge with this approach is to design a tank circuit with adequate Q and adequate temperature stability. This should be one area of the design which is thoroughly investigated during the design phase. Problems with clock recovery can lead to a highly unreliable system.

2.2.4.2.2 Decoding Circuitry: The receiver should contain circuitry which will utilize the regenerated clock from the clock recovery circuit to sample the incoming signal stream and generate an output (data and control outputs) which correspond to the information originally transmitted in the serial stream. This should include the appropriate detection of start and end delimiters, bus activity, frame type, and any other control data the designer needs to properly receive and route the messages and maintain the proper statistics on bus activity within the station.

Problems may occur in the design of this circuit due to the nature of the media being used in the system. The designer should remember that various signal distortions which will be encountered in the reception of messages on the bus will all create problems in sampling and decision making by the receiver. Development of circuitry which is capable of accepting the distorted signal and providing adequate sampling and decision making must be a high priority in this design activity. In many systems, it will be found that traditional approaches using threshold detectors do not provide adequate performance due to the unreliability of zero crossings and the resultant pulse width distortions of the decoded signals. The most reliable approach has been to detect changes in waveform polarity using a type of rate detection circuit.

2.2.4.3 Coupler Design: Design of the coupler represents the most critical design activity involved in building a coax based LTPB. Proper management of coupler losses allows the designer to maintain a healthy signal level on the main trunk, while protecting the receiver from an overload condition. The coupler design must adequately match the receiver impedance with that of the bus in order to avoid impedance discontinuities and resulting reflections. If the directionally coupled trunk bus topology is selected, the design of the coupler is straightforward. Suitable couplers can be procured to specification from a number of well known RF component suppliers. If the bidirectional trunk bus topology is selected, the choice of couplers is more complex since commercial equivalents are not available. This may require that they be designed and produced as a part of the network development program. Couplers have been designed and built under government programs which have demonstrated insertion loss of 0.1 dBm, VSWR of 1.05:1, and fault tolerance in which no single component failure will disable the network.

3. AS4074 MEDIA ACCESS CHARACTERISTICS:

3.1 Data Bus Redundancy:

The purpose of this section is to present the rationale for data bus redundancy and describe the method chosen for the AS4074 implementation.

The system designer must determine the level of redundancy for the data bus based on reliability goals for the entire system. The type of redundancy will also be dictated by the goals to be fulfilled in adding redundant capacity. For example, the level of busing in a vehicle management system with four wholly independent processing racks might consist of four completely independent buses to provide the same level and type of redundancy implemented in the remainder of the system. By contrast, an avionics system where each bus terminal has only limited redundancy and contains some single points of failure would generally require protection only against total loss of the data bus. This protection would consist of detecting single failures within a terminal that would disrupt the operation of other terminals, using the detected failure to passively shutdown the terminal, and providing a method for implementing media redundancy only.

3.1.1 Media Redundancy Approaches:

Media redundancy methods are used to prevent faults in the physical media path and transceivers from causing loss of terminal function. One example is the loss of the star coupler which interconnects all terminals in a fiber optic implementation. Media redundancy provides an alternate path through a redundant star coupler in this case. The two couplers should be physically separated to reduce the possibility that they could both be lost to battle damage.

Three different approaches for providing media redundancy have been studied. The first approach, shown in Figure 3.1.1-1, is actually full dual redundancy. In this approach not only the media but all bus interface hardware is replicated.

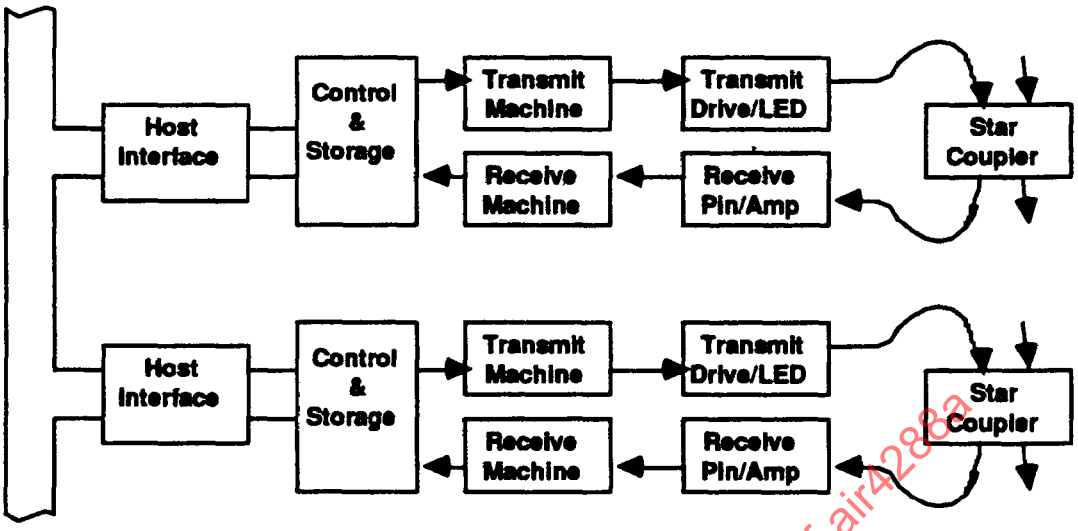


FIGURE 3.1.1-1 - Full Dual Data Bus Redundancy

3.1.1 (Continued):

Messages can be launched and received independently on either of the channels. Clearly this design can provide two completely independent bus channels when redundancy goals demand that capability. In the comparisons that follow, this approach is evaluated as a method of providing media redundancy only. The approach is particularly applicable when higher levels of system redundancy is required, and is recommended for providing system level redundancy.

Another media redundancy approach, shown in Figure 3.1.1-2, is referred to as Synchronous Redundancy since a single message is transmitted on both bus media at the same time. Within the transmitter section, the same signal is used to drive both optical transmitters to minimize the time skew between optical signals.

Time skew will generally be present at the receiving terminal due to the difference in optical path length. Because of this time skew, the receiver section uses independent receive hardware for decoding data from each optical media. Combine/select logic selects which set of received data will be passed to the remainder of the receiver.

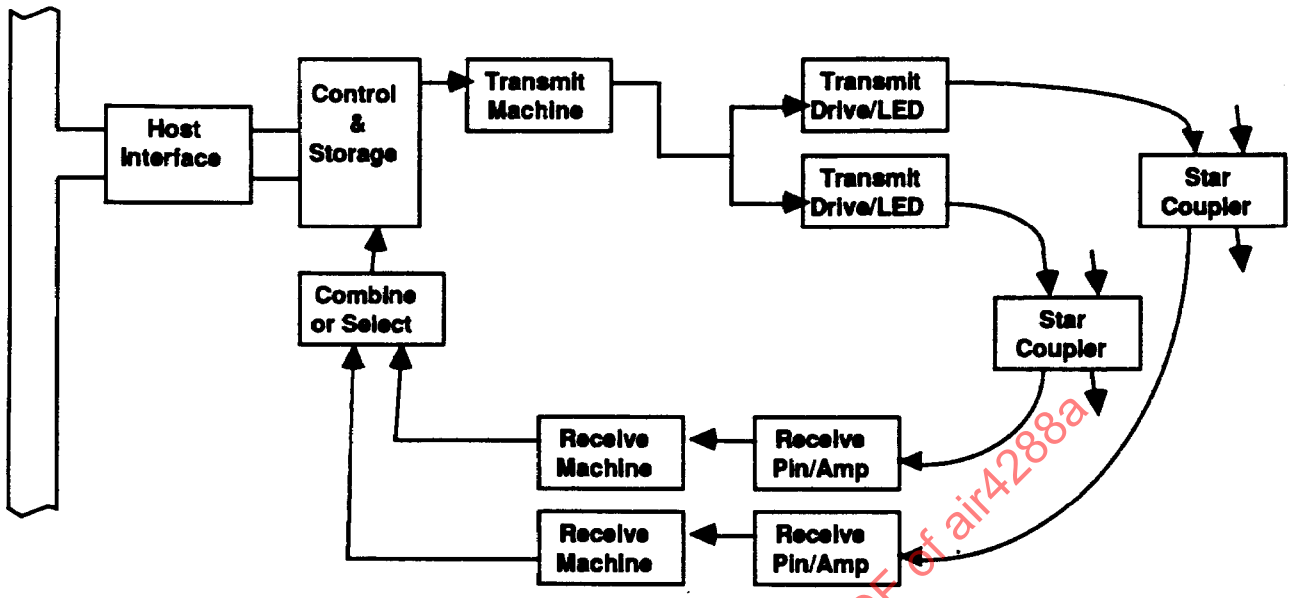


FIGURE 3.1.1-2 - Synchronous Redundancy

3.1.1 (Continued):

The final method studied is Standby Redundancy, shown in Figure 3.1.1-3. In this scheme, only a portion of the receiver and transmitter circuitry is replicated. For example, by duplicating the receiver front end, the decoder and the clock recovery circuitry (in the receiver) and the encoder and transmitter driver (in the transmitter), we may now use bus selection logic to determine which of the two bus media we desire to transmit on. The other bus would be held in standby so, in the event of a failure on the one medium, we could use the control logic to activate the other. In this method we could bypass the failed medium and restore bus operation by using the other medium.

In a typical implementation, the bus selection logic is driven by a topology memory which contains information on the health of all terminals in the system. The data on media health within the topology map is used to determine which channel a message is launched on. Station management messages are used to maintain information in the topology memory within all the active terminals in the system.

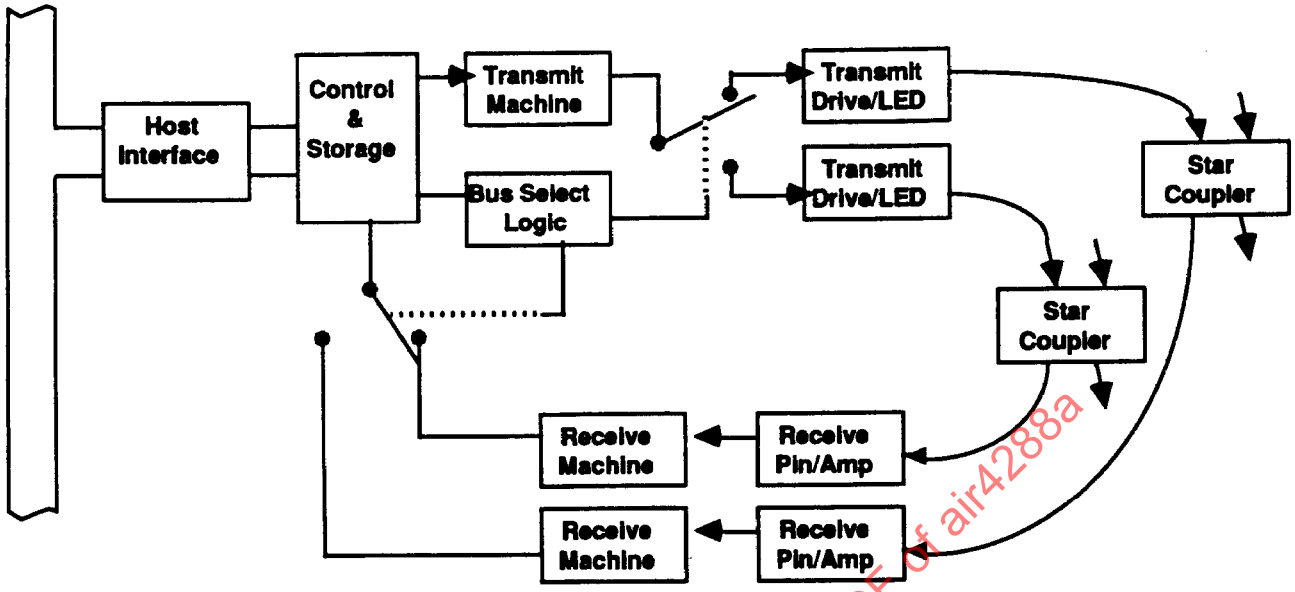


FIGURE 3.1.1-3 - Standby Redundancy

3.1.2 Selection of a Media Redundancy Approach:

3.1.2.1 Full Dual Redundancy: The principal advantage of the full dual redundant method of implementing media redundancy is retention of module compatibility with systems that implement more or less than dual redundancy. The alternate media channel in a module implementing synchronous redundancy would be superfluous in a system that requires a more independent type of redundancy and would probably remain unused.

The primary disadvantage of the full dual redundant approach is that substantially more hardware is needed than is required by other redundancy schemes. A second disadvantage is that supervisor processor intervention is required whenever a media fault occurs. Since the control logic within the channel controls only a single media path, software must be used to reroute messages to another bus channel. A potential solution to this problem is to launch the same message through both independent channels to assure its successful passage through the media. A serious drawback to the dual-launch method is the need for complex software to differentiate between a new version of a message and the second copy of a previously received message.

3.1.2.2 Synchronous Redundancy: The synchronous redundancy method has some important advantages which are summarized below:

a. Recovery from media errors is accomplished at the receiving terminal in a manner transparent to any processor software.

b. Recovery from media errors does not require retries or immediate acknowledgment, which saves bandwidth.

c. Media error recovery can be accomplished on all messages (including logical messages and messages containing sequential data).

d. Maintenance of topology maps within every bus interface module is not required.

e. As an added benefit, the possibility of a bus collision caused by the erroneous broadcast of an initialization frame by a terminal with a failed receiver is greatly reduced. The front end of the receiver is redundant and the remainder of the receiver would be tested by electrical loop back tests at power up.

The primary disadvantage in the synchronous redundancy approach is the increased complexity of the receiver logic due to the error recovery circuitry required and the need for time skew correction introduced by path length variations.

It should be remembered that when absolute resistance to single failures is required, only full replication of bus channels (full dual redundancy) can provide that resistance.

3.1.2.3 Standby Redundancy: The primary advantage of the standby redundancy method is that it allows a retry on alternate media to be controlled by logic local to the BIU. This allows messages to be retried on alternate media without intervention by launching processor software.

The primary disadvantage of this method is in the use and maintenance of a topology map for determining which of two media paths to use. In a system that makes wide use of logical messages, where the physical address is unspecified, topology map data would be of limited use in selecting which media channel should be used for a message. The implementation of automatic retries generally requires an immediate acknowledgment to determine when a retry is necessary. Automatic retries can also cause problems with sequential or time critical messages since the receive terminal could potentially attempt to process two copies of the same message. The proper operation of the topology map requires that all operational terminals in the system maintain an accurate copy of the map. This requires that the copies of the map be transferred to terminals as they enter the network. This, in turn, requires the use of bus bandwidth consuming station management messages that may cause the propagation of a failure in one station's topology memory or associated logic to other terminals in the system.

3.1.2.4 Selection of a Redundancy Scheme: Because of the many advantages of synchronous redundancy, this method was chosen by the SAE Linear Bus Task Group as the method for implementation of media redundancy. It is recommended that any design implementing less than full redundancy make use of babble timers and self-monitoring (3.7) to increase overall bus resistance to single-point failures. Finally, it should be remembered that none of the above techniques offers absolute resistance to single failures. Only replication of the entire BIU (full dual redundancy) can provide that resistance. Use of full dual redundancy is not precluded by the standard.

3.1.3 Implementation of Synchronous Redundancy: Synchronous redundancy refers to hardware methods of implementing redundant physical paths where the transmitter, receiver, decoder, and physical media are replicated with no further effect on any higher level protocol operations. Transmissions occur on both media simultaneously and are received by either or both, receivers with the first valid data stream available being accepted.

Implementation of the transmit function is as simple as routing a single transmitted serial data stream simultaneously to two separate transmitters. The combining of the two received data streams is more complicated and may be accomplished in several ways. Two approaches to implementation of a dual synchronous receiver are presented. One involves the combining of the raw serial bit stream at the receiver with no further impact of higher level hardware. The other approach continues the dual reception through the serial to parallel conversion level, performs double buffering of the dual data, and selects one copy on a message by message basis. The two approaches will be referred to as the data bit oriented and data block oriented implementations.

3.1.3.1 Data Bit Oriented Implementation: The data bit oriented implementation of a dual synchronous receiver combines the two received serial data streams before parallel conversion is applied. The operation of the combine/select logic may be better understood by examining the sequence of operations that occur when a message is received.

a. A message start delimiter is detected within the decoder for a particular media channel. The delimiter is received on this channel (designated the first channel) before the same delimiter is received on the other channel (designated the second channel) because of the path length induced time skew. Figure 3.1.3.1-1 details an implementation of synchronous redundancy in a fiber optic system. In this example, the decoder sends a start delimiter indication to the combine/select logic. This causes the combine/select logic to prepare for reception of the new message by clearing the bit counters and designating the first channel as the primary channel for the current incoming message.

b. The receiver for the first media channel begins decoding data from the message and placing the data in a clock synchronizing FIFO.

c. The combine/select logic begins extracting data bits from the FIFO for the first channel and passing data to the remainder of the receiver logic. The bit counter for the first channel is incremented every time that a bit is extracted from the FIFO.

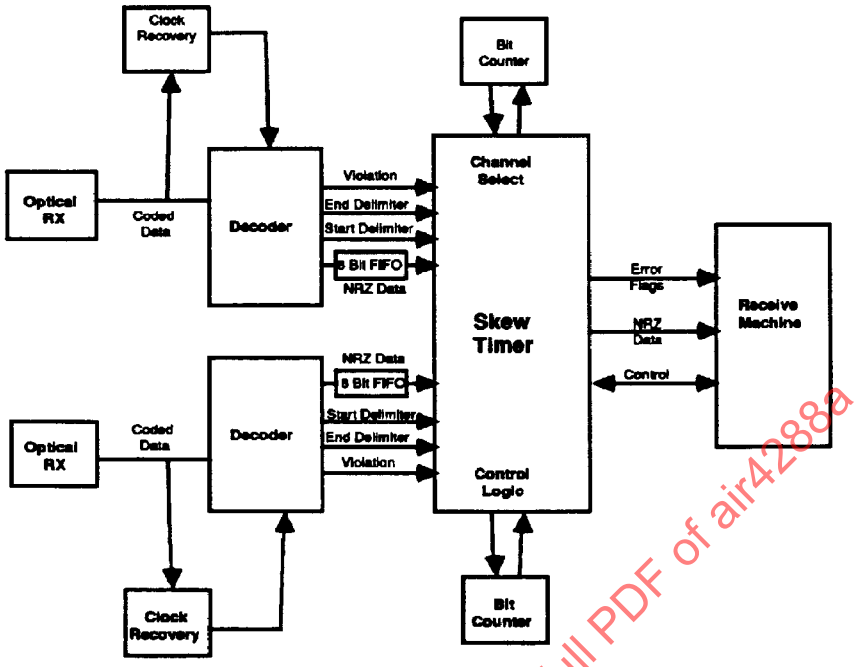


FIGURE 3.1.3.1-1 - Example of a Data Bit Oriented Implementation of Synchronous Redundancy

3.1.3.1 (Continued):

d. After the time skew (3.1.4) has elapsed, a start delimiter is received on the second channel. The decoder for the second channel sends a start delimiter indication to the combine/select logic. This causes the combine/select logic to set a flag indicating activity on the second channel.

e. The decoder for the second media channel begins decoding data from the message and placing the data in the FIFO for the second channel.

f. The combine/select logic begins extracting data bits from the FIFO for the second channel. Each bit extracted causes the bit counter for the second channel to be incremented but the data bits for this channel are discarded unused.

g. The decoder for the first channel continues to process data bits until an end delimiter is detected on that channel. At that time an end delimiter indication is sent to the combine/ select logic.

h. Receipt of an end delimiter for the second channel would normally terminate the combine/ select message processing. If a media fault has occurred prior to this message, no start delimiter will be received for the second channel. In that case, the skew time out counter will time out indicating a media fault.

3.1.3.1 (Continued):

i. If an error occurs in the first media channel during the message, the decoder for the first channel sends a violation flag to the combine/select logic. The combine/select logic stops extracting data from the FIFO and freezes the bit counter for the first channel. Data continues to be extracted and discarded from the second media channel and the bit counter for that channel is incremented. When the value of the second channel bit counter equals the frozen value in the first bit counter, data from the second channel is passed to the remainder of the receiver. In this way, data from the good media channel is substituted for the data on the first channel.

Advantages of the data bit oriented approach include a simple implementation, minimal hardware, and transparency to all layers above the physical media interface. The main disadvantage of this approach is the addition of logic to the high speed section of the LTPB circuitry. Another problem is that the circuit only provides redundant protection from detected coding errors, and does not validate the CRC on both frames. It must also be noted that the variations between clocks in to LTPB nodes must also be managed by the FIFO function and could add to the station response time of the receiving station if the FIFO becomes full.

3.1.3.2 Data Block Oriented Implementation: The data block oriented implementation (Figure 3.1.3.2-1) of a dual synchronous receiver operates similarly to the data bit oriented implementation, except in this case the receiver buffers both data streams on a message (or token) basis then selects the data in a buffer that has been received without error and discards the other copy. The full receive protocol machine must be replicated and storage provided for both copies of an individual frame. The operation of the selection process is described:

a. A frame is received on one channel and is stored temporarily to a frame buffer for selection if error free.

b. The redundant frame is received on the other channel and is also stored temporarily to a frame buffer for selection if error free.

c. Upon complete storage of both frames, an error free frame is selected and forwarded based on some arbitration criteria. The other copy is discarded. If both frames are in error, both are discarded.

d. Some method for double buffering alternate messages must be provided to account for multiple message frames.

e. It is likely that to meet station response times, special handling and arbitration must be provided for token, initialization, and some station management frames.

f. The requirement for a skew timeout counter (3.1.4) still applies to detect media faults on a specific channel.

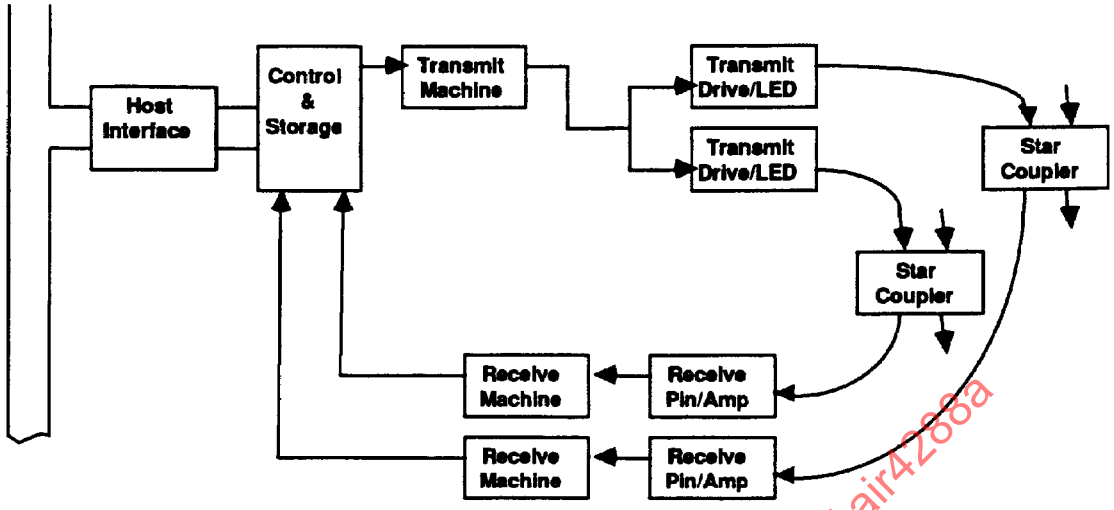


FIGURE 3.1.3.2-1 - Example of a Data Block Oriented Implementation of Synchronous Redundancy

3.1.3.2 (Continued):

The main advantage of the data block oriented implementation is that the CRC may be validated on each message (or token) individually, reducing the likelihood of lost messages. This implementation, however, requires much more hardware than the data bit oriented approach. The designer should pay particular attention to the maximum message length required by the application for which he is designing his BIU.

3.1.4 Accounting for Time Skew Between Redundant Bus Media: Successful implementation of the synchronous redundancy scheme requires that time skew between the two media be properly accounted for during system design and managed to within a worst case value.

Time skew on the redundant media of the LTPB occurs when delay elements are introduced into one medium without a corresponding delay being introduced into the other. These delays may be due to propagation delay from additional media length, signal processing time due to an active element (active coupler or repeater), or other such system design related items. Figure 3.1.4-1 shows the ideal relationship between signals at the receivers on both media (buses) of a synchronous redundant system. The signals are required to be transmitted simultaneously (within 1 bit time) by the standard. In the case shown, there would be no problem receiving the message from either bus. In the event of an error on the primary bus, the message could easily be recovered from the secondary bus. Note that the designation "primary bus" and "secondary bus" are determined strictly by which bus a message start delimiter is detected and validated on first. This becomes the primary bus. The other is then designated the secondary bus. Notice the use of an intertransmission gap time (t_{itg}) of at least 280 nS between frames in the figure. This bears some explanation in order to clarify the terms used.

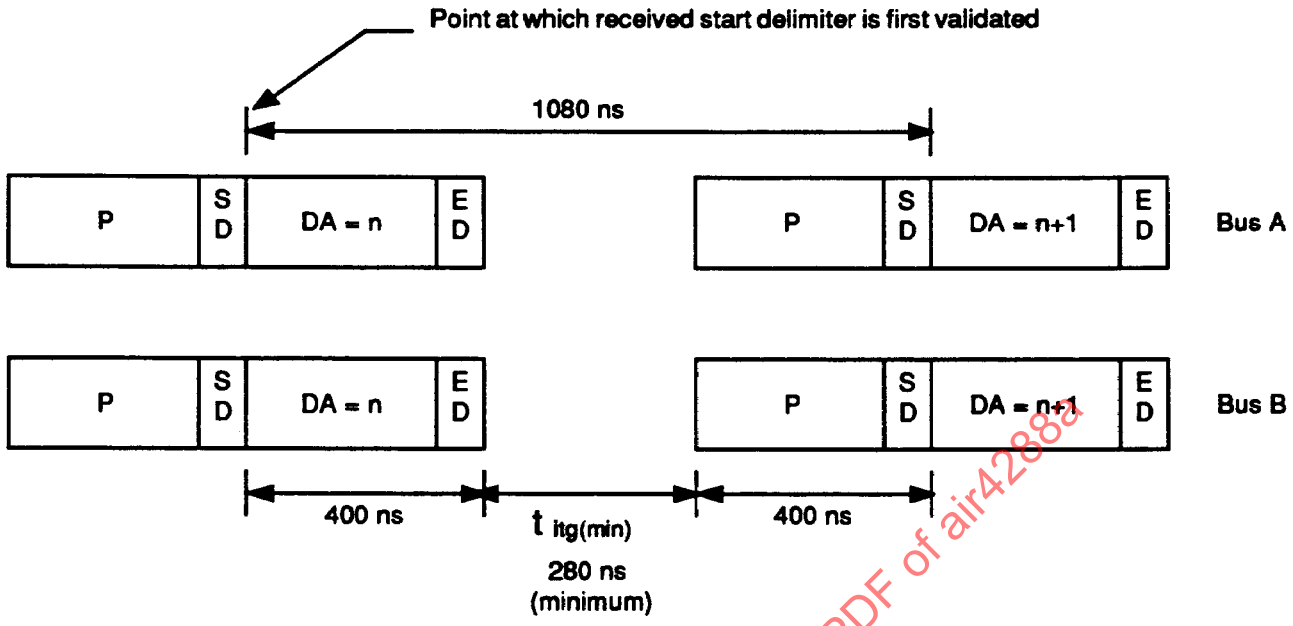


FIGURE 3.1.4-1 - Ideal Relationship Between Transmissions in a Synchronously Redundant LTPB

3.1.4 (Continued):

The gap which occurs between the transmissions of different stations on the LTPB is bounded by an upper and a lower limit. The upper limit, called the station response time (t_{sr}) puts an upper limit on the amount of time a station may take to begin transmitting a message after having received a valid token for his address. This value is assigned 500 nS. The lower limit, called the intertransmission gap time (t_{itg}), is called out in the slash sheet and is the minimum amount of time a station must wait to begin transmission after receipt of a valid token for its address. The t_{itg} is provided to compensate for any signal level changes which might occur at the receiving station due to the relative physical locations of the station which just finished transmitting and the station which is to begin transmitting. This time allows the automatic gain adjustment circuit of the receiver to recover from a potentially saturated state to one which is sensitive enough to see traffic from the new station, which might be at a substantially lower power level. Since 280 ns is the lower of the two values, it is used for our worst case scenario.

3.1.4 (Continued):

Figure 3.1.4-2 illustrates the concept of time skew. Time skew is defined as the time from reception and validation of a start delimiter on the primary bus (bus A in this example) to the reception and validation of a start delimiter on the secondary bus (bus B). Note that reception and validation of a start delimiter is the first point in message reception which defines what the bus activity being detected means. Since we cannot be certain when the clock recovery circuit will be properly phase locked to the incoming encoded clock information (we hope that it occurs well before the end of the preamble) we must assume that everything prior to a start delimiter is noise or some other activity. The time between reception and validation of a start delimiter on the primary bus and the reception and validation of a start delimiter on the secondary bus is the skew (t_{skew}). This time must be accounted for during the design of the BIU and the LTPB system in order to assure proper operation and minimization of latencies.

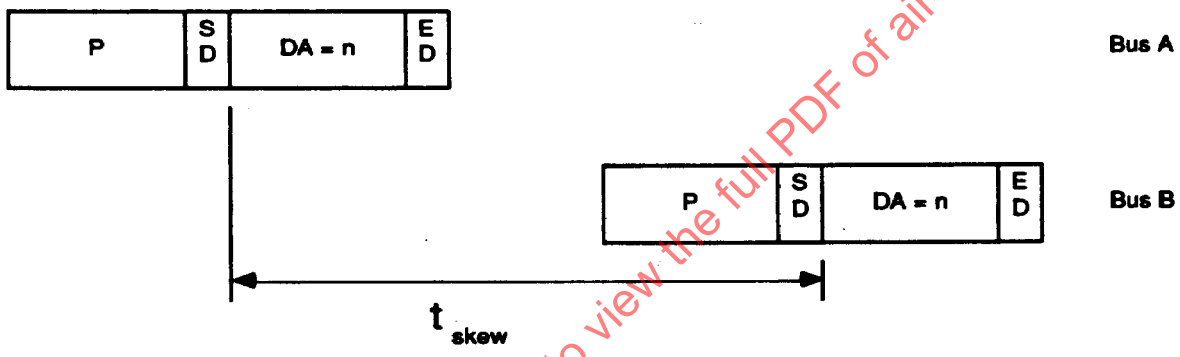


FIGURE 3.1.4-2 - Illustration of Time Skew in a Synchronously Redundant LTPB

A determination must now be made as to the worst acceptable time skew for the LTPB system. Figure 3.1.4-3 depicts the worst case situation and, hence, the worst time skew acceptable for t_{skew}. The token frame is used as an example since it is the shortest frame which will be transmitted on the bus and, therefore, gives us the worst case timing. Time skew must be minimized such that the end delimiter on the message present on the secondary bus is received and validated prior to the reception and validation of the next message start delimiter of the token present on the primary bus. The reason for this requirement becomes obvious if you allow the time skew to increase such that this rule is violated:

a. We require that the bit counter (3.1.3.1) or other incoming data correlation counter to be larger, since we will receive more bits on the primary bus before we begin receiving the corresponding data bits on the secondary bus. This delay would also increase message latency in the event an error is detected on the primary bus. This delay would be due to the need for the message being received on the secondary bus to reach the same point so that reception could be switched over. The message latency problem also occurs in a block reception oriented scheme.

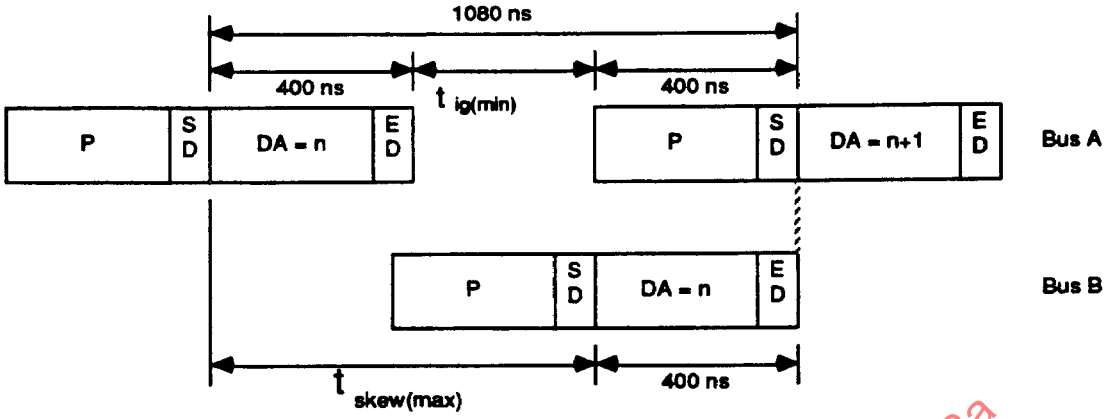


FIGURE 3.1.4-3 - Illustration of Worst Allowable Time Skew

3.1.4 (Continued):

b. Circuitry would be required to keep track of message reception on each of the two buses. That is, if the time skew is sufficiently large the primary bus could be starting a message frame which is, say, two or three frames ahead of the token currently being received on the secondary bus. Decision making hardware would be required to differentiate between messages so that the error recovery function would be able to pick up the right message from the incoming data stream and recover a message in error. Additional data buffering space would be required.

This points to the fact that it is to the system designer's advantage to limit the time skew in his system. As a matter of fact, this maximum skew will be specified by the BIU hardware designer. BIU hardware should be designed to handle the specified amount of skew and still operate properly. The AS4074 standard sets 150 nS as the maximum allowable skew in a system using this standard.

3.1.5 Additional Redundancy: The LTPB has been defined to operate with dual synchronous redundancy. Dual redundancy was chosen because of the experience with MIL-STD-1553. Synchronous redundancy was chosen because of the difficulty of deciding to change from one bus to another within a distributed access control scheme and the need to get the message transfer right the first time because of the lack of transfer status response.

Many applications of MIL-STD-1553 utilize dual redundancy, although greater or less redundancy can be handled by the system. Standby redundancy was defined, where the bus controller (BC) would choose the bus on which to command a remote terminal (RT) and the RT would respond on the same bus. As many different buses as are required may be used, with the bus controller choosing on which bus a particular message should be transmitted.

The disadvantages of synchronous redundancy are the need to have both transmitters active all the time and the need to take different path lengths on the two buses into account. The latter is necessary to enable tokens and messages on each bus to be synchronized with the other bus. Tokens and messages are transmitted on the two buses at exactly the same time.

3.1.5 (Continued):

The problem of trying to use synchronous redundancy on more than two buses is that of propagation delays. Since multiple paths are likely to be routed differently, paths between stations on different buses will also be different, although the total delay for a token rotation on each bus might be the same. Since each station must wait for the token to be received on each bus, or wait for a timeout before transmitting, the time taken for a token rotation will increase for greater synchronous redundancy since there is a greater probability of there being a long path present.

Greater redundancy must be achieved through the replication of the LTPB station. This means that the different buses, connected to different stations, will operate independently of one another so that tokens, messages or even the order of messages will not be synchronized between the different bus systems. This should not be thought of as a disadvantage. The LTPB is defined such that a message will be passed by the system within a determined time. The systems are not designed to require a specific data order, as MIL-STD-1553 systems were, because of the increased performance gained by not doing so. Since most systems will only require dual redundancy, it does not seem too great a hardship on those systems requiring greater redundancy to also endure redundant stations. This may be required anyway to prevent single point failures.

If data order is important, it can be implemented by time tagging the data when it is queued and sorting it by time when it is received. This is not necessary for data from the same station, since this will always be transmitted in the same order for the same priority, but different orders could occur when data is merged from more than one transmitting station.

There are a number of ways in which multiple redundancy may be performed. Taking a requirement for quadruple media this could be implemented as two separate synchronous redundant stations both of which are always active, two separate synchronous redundant stations operating as a standby pair: four single bus stations, all always active, or four single bus stations acting in some standby redundant manner.

For the first of these, the token and message passing would be synchronized between the pair of buses connected to a single station, but each station would be operating independently of the other (Figure 3.1.5-1). Both systems would transfer the same data, albeit at slightly different times, and the user would have to determine which of the two versions of the data to choose.

The second method would only have one synchronous redundant bus active at a time, the other being activated if a fault occurs on the first (Figure 3.1.5-2). The token and message passing would be synchronized on the synchronous redundant pair as if there were no other buses present. The User would always be presented with a single copy of the data. Since a single point failure should not be able to prevent both of the synchronous redundant pair from operating, it will be possible for a system command to be issued to indicate the switch to the standby system.

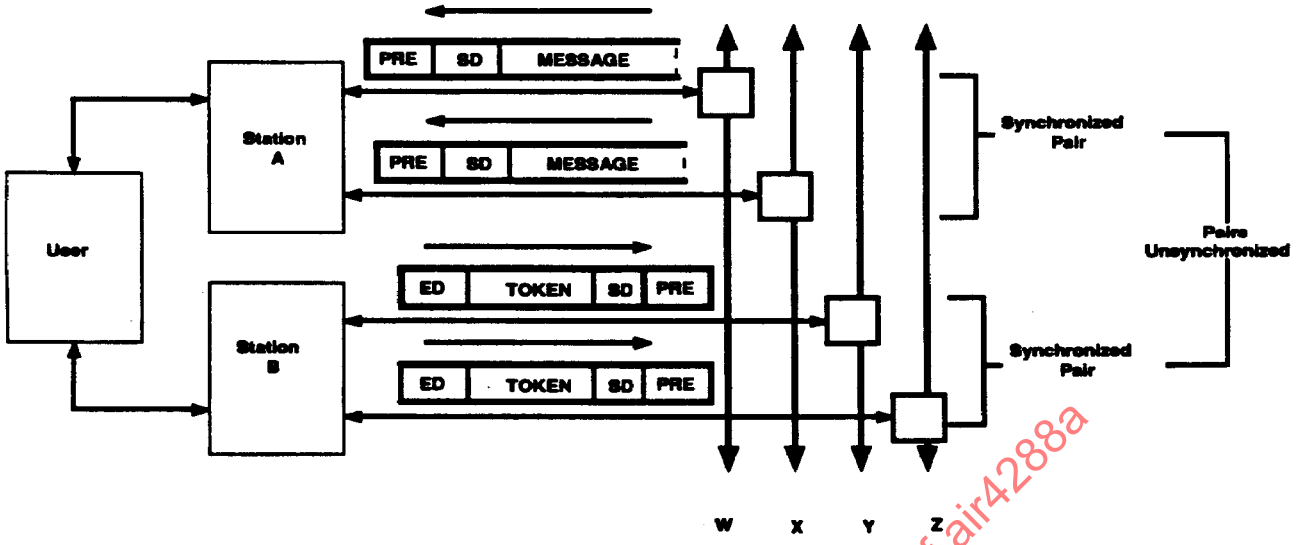


FIGURE 3.1.5-1 - Two Active, Synchronous Redundant Stations

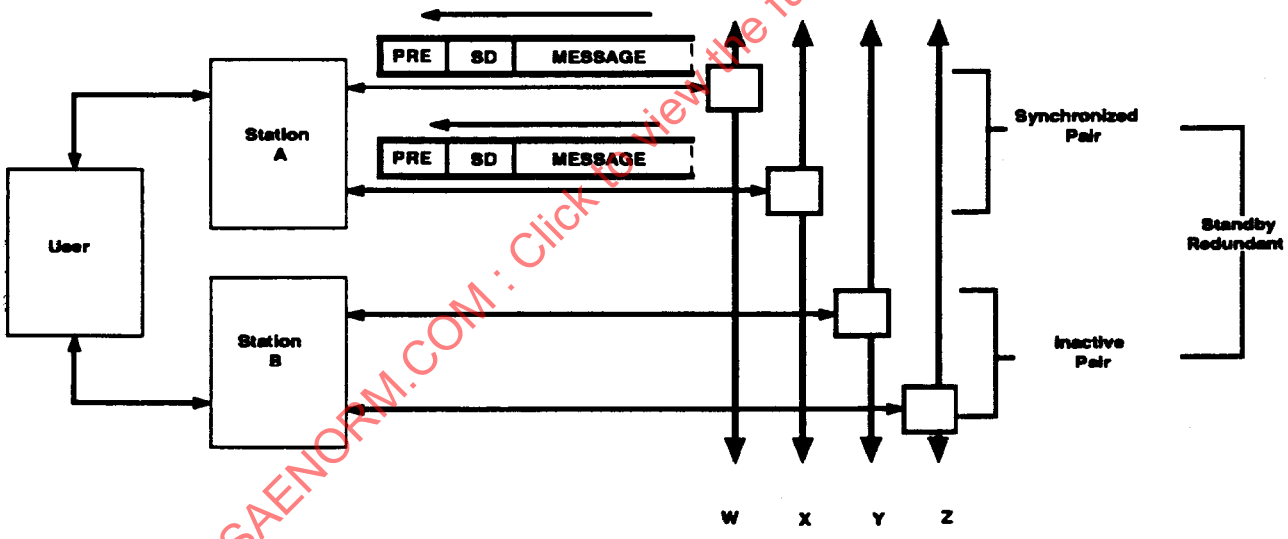


FIGURE 3.1.5-2 - Two Standby, Redundant, Synchronous Redundant Stations

3.1.5 (Continued):

The third method again has all four buses always active, except now each bus operates independently such that token and message passing is not synchronized (Figure 3.1.5-3). Each bus would transfer the same data, albeit at different times, and the User would have to determine which of the four versions of the data to choose.

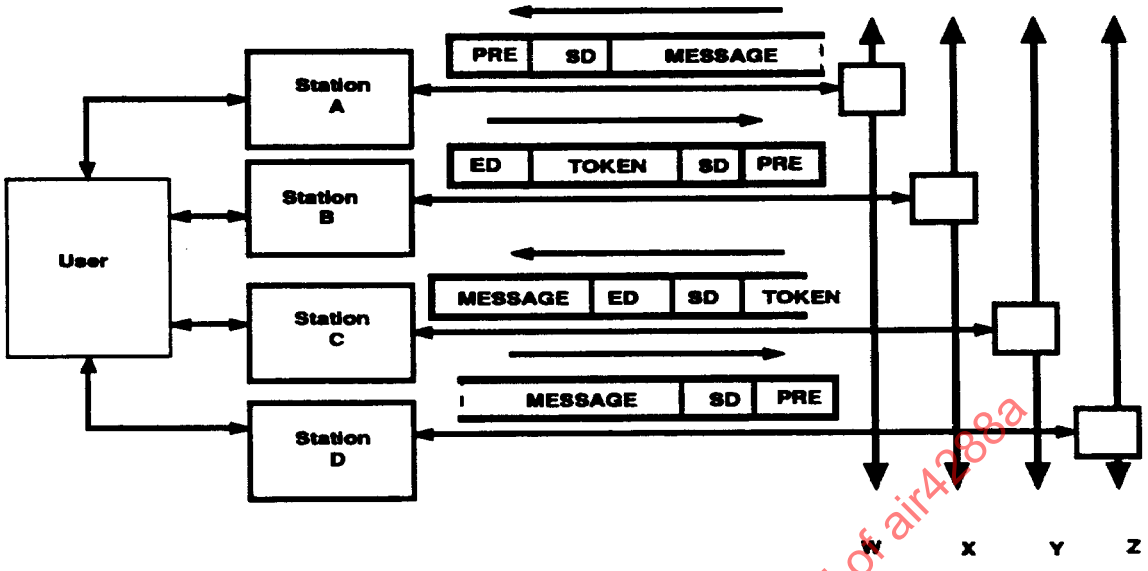


FIGURE 3.1.5-3 - Four Active, Independent Stations

3.1.5 (Continued):

The final method could take pairs of buses active together. Each of these buses would operate independently of the other active bus with token and message passing not synchronized (Figure 3.1.5-4). Each bus would provide the user with its own version of the data and the user must choose which to use. In the case of a fault on one of the buses, a system command is issued over the other active bus to activate one of the standby buses.

In summary, the limitation of bus redundancy handling by a station to dual synchronous redundant does not prevent additional redundancy being added at a system level by replication of LTPB interfaces. While this does impose additional restrictions on the users connected to the bus, these are preferable to the loss of performance encountered by the addition of extra synchronous redundant buses. A variety of ways may be used to implement the additional redundancy, the choice of which will depend on required availability, power dissipation, user complexity and volume.

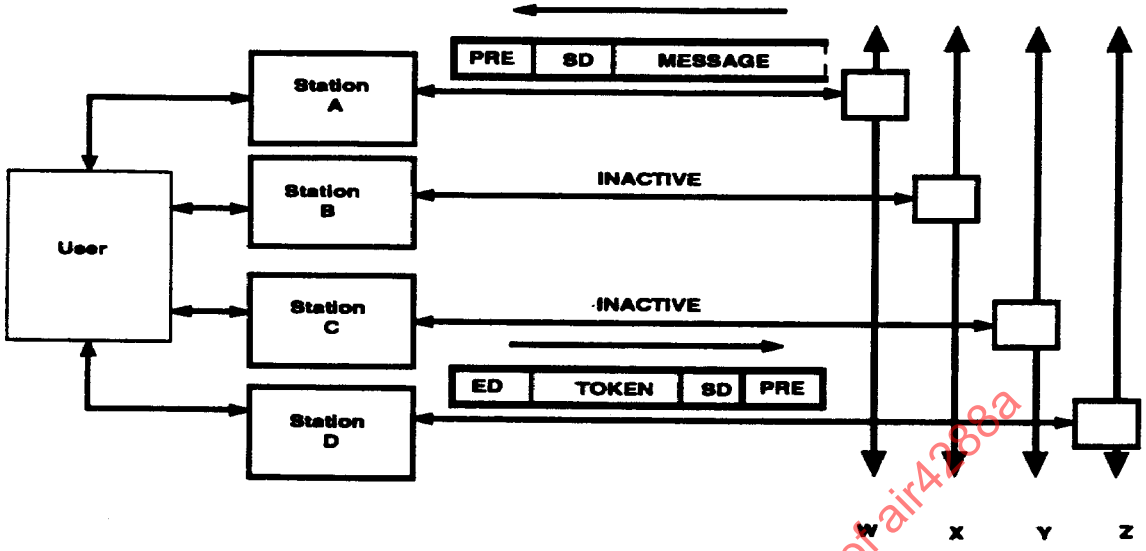


FIGURE 3.1.5-4 - Four Standby Redundant, Independent Stations

3.2 Claim Token Process:

3.2.1 Operation of the Claim Token: The claim token is utilized as a method of initializing the normal operation of the protocol. This initialization occurs under two conditions:

a. Cold Start: (bus has not been running prior to this point for example, system power-up)
b. Warm Start: (bus has been running but a fault has occurred which has caused the bus to go "dead" for example, a "lost token")

Claim token activity is initiated by the expiration of the bus activity timer (BAT). This timer is explained in more detail in 3.9.3. The BAT determines the maximum amount of time that the bus may be dead before a failure is declared. The value is set by the user and generally varies with the station physical address (i.e., lower numbered stations have the shorter BAT maximum values and the higher numbered stations have longer BAT values). Failure of the station to resolve the claim token after four attempts causes the station to go into the IDLE state.

A cold start occurs when stations on an inactive LTPB are simultaneously powered up. All stations on the LTPB receive their operating voltage(s) from their power supplies and begin built-in test (BIT) procedures. Since all power supplies do not come up at the exact same time and different amounts of time will be required for BIUs to complete BIT (due to different oscillator frequencies, different vendor design, etc.) the exact time that the station initializes his BAT and recognizes that the bus is dead varies. For this reason, more than one BIU may try to vie for control of the LTPB during a cold start.

3.2.1 (Continued):

In Figure 3.2.1-1 the stations have powered up such that the BATs of Stations 1 and 3 expire at the same time. Stations 2 through "N" still have time remaining on their BATs at this time. Since both stations 1 and 3 consider the bus to be dead, both try to initialize it simultaneously. This causes the claim token message frames from both stations to collide. This activity causes all the other stations on the bus to detect activity. Since they consider the bus to be active, they go to IDLE to await the token pass. Stations 1 and 3, each see the collision as a coding error in their data stream and cease transmission, resetting their BATs and monitoring the bus. Since station 1 has the shorter BAT, it will expire first and begin transmitting a second claim token frame. This time, station 3 will see the claim token frame and will go to IDLE to await a token pass to its address. Station 1 will complete transmission of the claim token frame, with no errors, and will win the right to initialize the logical ring. Normal token passing must now be established under control of Station 1, the claim token winner. Since these stations have been powered down, the register which would normally contain the address of their successor would contain a hardware initialized value, not one which represents a valid successor station on the active network. Station 1, as well as every other station on the LTPB, will have to go through the "hunt" procedure (3.3.1) to determine which station will be his successor on the active LTPB as the logical ring is built up from scratch.

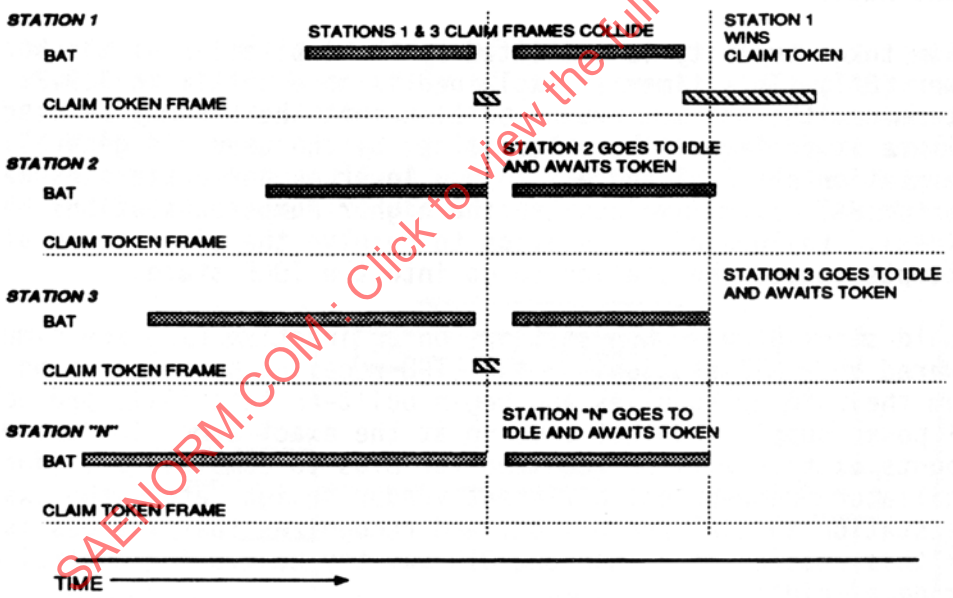


FIGURE 3.2.1-1 - Normal System Power-Up Scenario (Cold Start)

3.2.1 (Continued):

Figure 3.2.1-2 details a "warm start". A warm start occurs when the bus has been in operation for some period of time (a time great enough to at least have stable token passing between all active stations) and a failure occurs such that the bus goes dead (e.g., lost token). All stations will recognize the dead bus condition at the same time since they are all active and their BATs will start timing the absence of traffic at the same time. The station with the shortest BAT will begin the claim token activity and win control of the token.

Logical ring build up is where the difference between cold start and warm start occurs. In the warm start condition, the station will still have the address of its last valid successor in the "next station register". That means that once token passing starts, tokens will immediately be passed to the active stations in the station's "next station register" and a "hunt" will be unnecessary. This is in contrast to the required hunt for a successor in the case of the cold start.

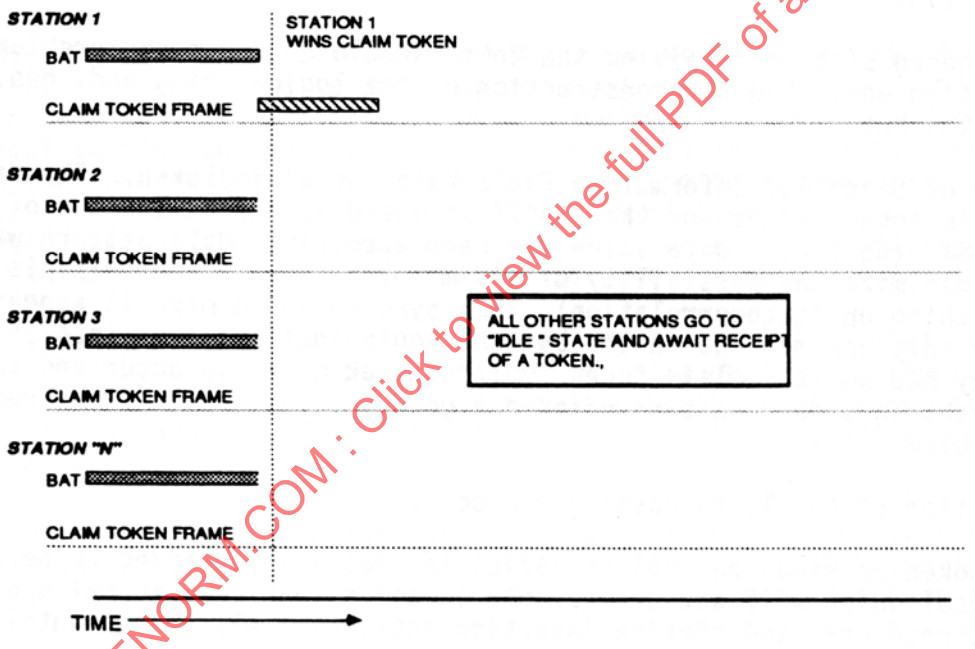


FIGURE 3.2.1-2 - Bus Reinitialization (Warm Start)

3.2.2 Variable Length Claim Token Frame: The use of a variable length claim token frame, with length related to the physical address of the station using the following equation:

$$\text{Number of words in claim token frame} = (\text{station address} + 1) * (T_{pd} * R_d / 16)$$

3.2.2 (Continued):

This difference in claim token word count between stations allows the claim token process to avoid problems in specific situations. For example, in a fiber optic implementation, assume two stations begin transmitting the claim token frame simultaneously (as depicted in Figure 3.2.1-1). In a situation where they are separated by a long distance and have a large amount of loss in couplers and connectors, each station's own signal may swamp-out the received signal from the other station. Each station would then see it's own claim token frame as error-free and assume control of the bus. The contention would not be resolved and both stations would assume a win of the claim token and begin bus initialization. The bus would then crash due to multiple transmissions. If however, the claim token frame length is based on station address, one of the stations would be transmitting a longer frame. This would be seen by the distant station after its claim token frame was complete it would go to IDLE, since bus activity existed. The right to initialize the bus would then go to the station having the longer claim token frame and a potential bus crash will be avoided.

Notice that this method of resolution of bus control also would allow a station other than the lowest numbered station in the system to get initial control of the bus. While this is a possibility, it should be remembered that reconstruction of the logical ring (i.e., recovery of a failed bus) is the prime concern. The potential of a station other than the lowest numbered station receiving the token should not pose any problem since that station would begin reconstruction of the logical ring and, hence, recovery of the bus.

3.2.3 Use of Specified Information Field Data for Claim Token: The data field of claim token frames for the AS4074 standard specifies the use of the word "4884" (hex) as a data value for each word. This data pattern was derived to minimize the possibility of claim token frames from multiple stations synching-up (autocorrelation). This sync-up would make it appear that there was only one message on the bus and would fool stations into believing that they had won the claim token. Multiple tokens would occur and the bus would crash. This data pattern offered a good level of protection from this problem.

3.3 Operation of the Token Passing Protocol:

The token passing protocol specified in AS4074 is intended to be a robust protocol which will act predictably in all situations. Normal operation, fault recovery, and station insertion activities all operate utilizing the same protocol mechanism (token passing procedure).

3.3.1 Logical Ring Buildup: This paragraph discusses logical ring buildup from a cold start. For the difference between cold start claim token and warm start claim token, please refer to 3.2.1. Having won the right to initialize the bus through the claim token process, the winning station proceeds to hunt for the next active station (Figure 3.3.1-1), starting at the station with a physical address one greater than his own (This Station (TS) + 1). On each token pass, the station waits for the token passing timer (TPT) to expire prior to declaring a failure of the token pass. If the bus remains dead for a period longer than the TPT, the station retransmits the token to the same address and again waits, looking for bus activity.

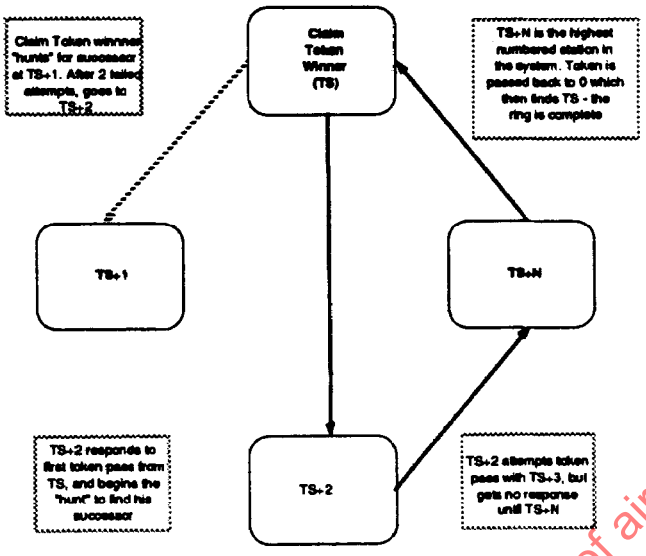


FIGURE 3.3.1-1 - Logical Ring Buildup

3.3.1 (Continued):

Should the station detect bus activity after the first or second token passes, the token pass is declared good and the station which passed the token goes to IDLE. The station which received the token then proceeds to attempt to find another active station by utilizing the same procedure. This continues until all stations in the system have been offered the token. The token then is routed back to the station which originally started the token and the logical ring has been built up. Message traffic may now be transmitted by the station holding the token.

Should two token passes to a particular address fail (TPT expires twice), the station is declared not active and the station holding the token increments the successor address by one and begins the hunting sequence over at that address. This activity continues until a station responds to the token pass.

3.3.2 Station Insertion: On a regular basis, programmed by the user, each station will attempt to pass the token to all stations between its address and that of its current successor, unless its current successor is the station with a physical address one greater than his own. This allows any new stations or those which might have failed or momentarily dropped out due to power bumps, to re-enter the active network. This ring insertion activity is controlled by the ring admittance timer (RAT) described in 3.9.1.

3.3.2 (Continued):

Figure 3.3.2-1 details the operation of the ring admittance mechanism. When a station's RAT expires and message traffic on the bus is relatively light (as indicated by an unexpired TRT3 timer), the station attempts to pass the token to the station with the physical address one greater than its own. This hunting is identical to that accomplished following claim token to build up the active network. Each address is attempted twice. If there is no response, the successor address in the token frame is incremented by one and the token pass attempted at the next station. This continues until a station accepts the token, as indicated by bus activity prior to expiration of the TPT. The new station then proceeds to use the same hunting mechanism to find its successor. This continues until all addresses have been checked and successors established.

3.3.3 Failed Station Deletion: In the event a station fails or drops out of the network due to an unplanned event (such as emergency power shedding), the removal of the station is done rapidly utilizing the same token passing methods discussed earlier.

Figure 3.3.3-1 shows the deletion of a failed station from the network. Note that the station passing the token to the failed station will attempt the token pass twice. After that time, the station goes into the standard token passing hunt to find a new successor. Once a new successor is located, the new successor becomes the permanent successor for that station until such time as ring admittance is performed (the ring admittance procedure is the method by which a failed station, if and when recovered, can re-enter the network).

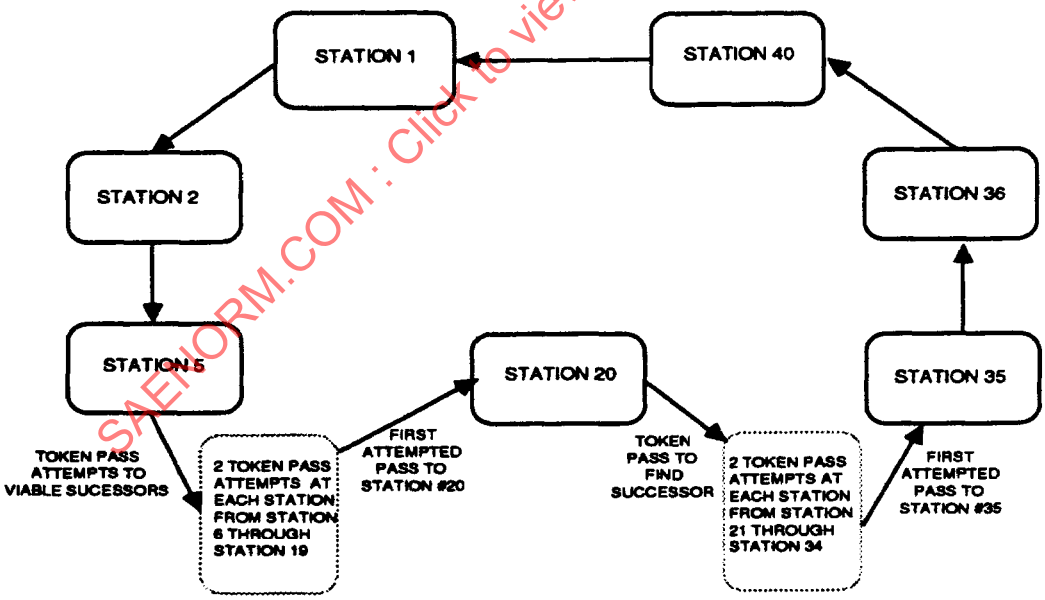


FIGURE 3.3.2-1 - Ring Admittance Procedure

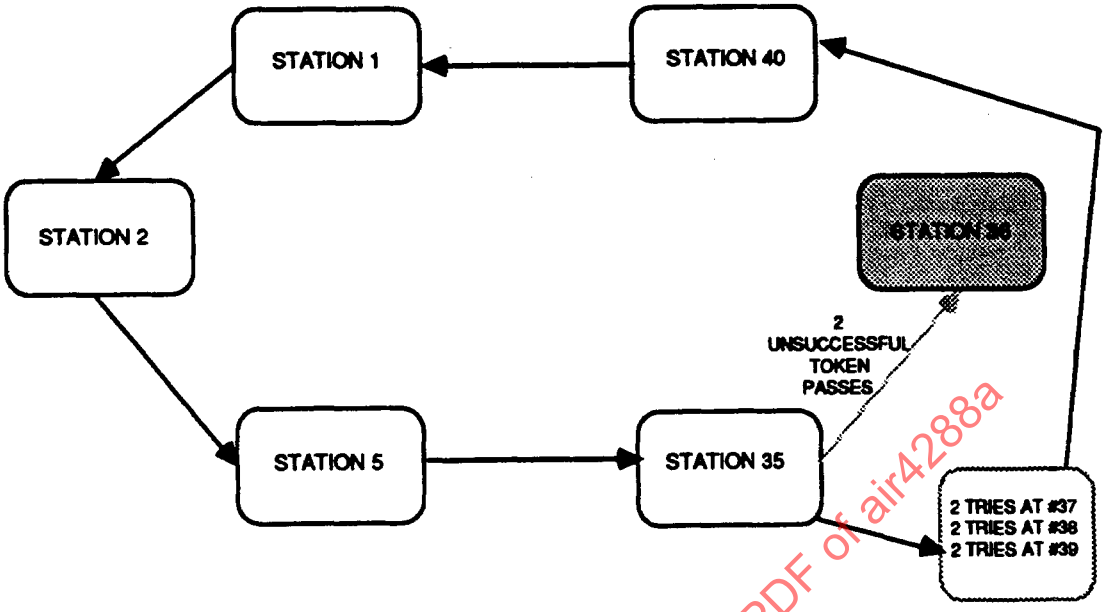


FIGURE 3.3.3-1 - LTPB Recovery from Failed Station

3.4 Send Message Procedure:

Transmission of a message by the BIU requires that a number of factors be taken into account. This section gives a brief overview of the procedure used by the BIU to determine when a message is pending for transmission and when that message is actually sent on the media. A more detailed description of the process can be found in the actual standard document itself (Section 5). Another explanation which outlines specific considerations is found in the sample system design section (Section 4) of this document.

The BIU will normally just generate a token for the successor address when no message traffic is pending. When the host generates a message for transmission, it must notify the BIU that a message is pending after that message is queued in a mutually agreed upon location. The priority level of the message must also be given. Some mechanism must present this information to the BIU transmitter. The host must make certain that the complete message is accessible to the BIU prior to flagging it for transmission. This is due to the fact that the exact arrival time of the token is unknown to the BIU and host interfaces (it is based on a dynamic message traffic load) and once a message transmission has been started, it can not be stopped or paused for the host interface to catch-up with the BIU.

Once a message has been placed in the queue, the BIU waits for a token pass to its address. When the token arrives, the destination address field is checked for it's address. It then verifies that no error was detected in the frame by completing a check of the token frame check sequence.

3.4 (Continued):

Once satisfied that it has received a good token, the BIU initializes the token holding timer (THT) to the maximum value and checks for any P0 (highest priority) messages. If any are pending they are transmitted. The BIU will transmit P0 messages until they are exhausted (no more P0 messages are pending) or until the THT expires. In all cases, the THT determines the maximum amount of time the BIU will use the token for message transmission. This, in turn, determines the worst case token rotation time for the system.

If P0 messages are exhausted and time still remains on the THT, the BIU checks the token rotation timer (TRT) for the P1 (second highest) priority level. If there is still time on the TRT1, and this amount of time is less than the time remaining in the THT, the TRT1 value is loaded into the THT and this value is used to limit message transmission at this priority level. TRT1 is initialized to the maximum value and restarted.

NOTE: If TRT1 is greater than THT, the THT is used as the limiting time for message transmission at this priority level. The station then proceeds to transmit messages at the P1 level until all messages are exhausted or until the current value of the THT expires. This procedure is repeated for P2 and P3 level messages.

It should be stressed that the THT maximum value is the maximum amount of time that a BIU in the system may use the bus to transmit any messages. When servicing P0 messages only, they may consume up to the maximum value of the THT, if required. The worst case token holding time would be the THT maximum value plus one maximum length message, if the station began transmitting a maximum length message just before the THT expired. If there are insufficient P0 messages pending, then the lower priority messages (P1,P2,P3) may be transmitted, if there is time remaining on the TRT associated with that particular priority level. That remaining value is loaded into the THT if, and only if, the value is less than the remaining value on the THT at the time it is checked. If the value is greater (which would cause the station to hold the token longer than $THT_{(max)}$) then the remaining value in the THT must be used to bound message transmission at that level to guarantee the token rotation time.

When a station has a number of messages of any priority ready for transmission, it may send them on a single token hold subject to the restrictions of the priority scheme (THT and TRTs) detailed above. The messages are to be concantenated by transmitting the start delimiter of the following message immediately after the end delimiter of the previous message. There shall be no gap (intermessage gap) between the end and start delimiters of consecutive messages.

When all message traffic has been exhausted, when a TRT is checked and found to have expired (prior to sending any messages at that priority level), or when the THT expires the station will pass a token to the current successor. In times of heavy message traffic, some messages will remain at the station and be deferred to a later token hold.

3.4.1 Transmission Monitoring: Each station within the system is required to monitor the data which it transmits on the LTPB. This monitoring is accomplished by the receiver associated with that channel. Monitoring is accomplished by decoding the incoming data stream and computing and checking the frame check sequence at the end of the frame (TFCS for tokens and MFCS for message frames). Failures also include encoding errors which are detected. Should a failure be detected, the higher level protocols must be notified and specific action must be taken. The notification and action to be taken vary depending on the type of failure detected.

3.4.1.1 Failure to Detect Bus Activity: Should the transmitting station fail to detect its own bus activity within two propagation delays, a failure condition will be generated. This failure indicates that there is a problem somewhere on the physical medium (couplers, interconnection, receiver, etc). The station is required to cease transmission immediately upon detection of the fault and to notify the TPIU of the failure. This allows the station to remove itself from that physical path to avoid causing a failure on the path should the failure be due to his station, as opposed to the external media.

3.4.1.2 Failure Detected During Claim Token Activity: During a claim token frame transmission, errors which occur in the monitored data stream are generally an indication of other stations vying for control of the bus, not actual system or hardware errors. These must be handled in a manner unlike transmission errors which occur during other types of frames.

When an error is detected during claim token transmission, the station must immediately cease transmission and reset the BAT (3.9.3). The station then follows the protocol procedure outlined for claim token activity as detailed in 3.2. No transmission monitoring error is generated following a failure of this type.

3.4.1.3 Failure Detected During Token Frame Transmission: Failures detected during token transmission are potentially the most catastrophic failures which can occur. The token frame is the key to proper LTPB operation. Problems which occur during this can cause a total failure of the bus resulting in a need to reinitialize the protocol. This can also result in message traffic being delayed beyond acceptable boundaries. Should a transmission monitoring error occur during token transmission the first error is logged by the station, but no action is taken. A second consecutive failure detected on the same token hold requires that the station generate a fault indication to the higher levels of the protocol and disable itself. This has the effect of "isolating" a potential troublemaker from the bus and avoids the possibility of causing a bus crash. It is then up to the station's host to decide by wrap-around tests and BIT whether a problem exists and whether the station can be enabled onto the active bus again.

3.4.1.4 Failure Detected During Message Frame Transmission: Should an error be detected during transmission monitoring of a message frame, the protocol operation is not affected. It is assumed that the message will be discarded by any receiving station. The problem is assumed to be due to transient problems in the media or the transmitter. Stations are not to re-transmit messages which are deemed as failed by this transmission monitoring process. The message retry function is assigned to the host.

Messages which are received by the station during transmission monitoring are not allowed to be sent to the host as a received message. These are messages which this station has transmitted. Sending the recovered message frames to the host only consumes bandwidth and processor time from the host to handle messages which it already has sent.

3.5 Receive Message Procedure:

The BIU receiver is always active for receiving messages. When enabled (3.8), the BIU acts upon those tokens addressed to it and receives message frames addressed to the physical address of the station or labeled with a logical address which is recognized by the station.

When receiving a token frame, the receiver checks the destination address field for the BIU physical address. If this address matches and if the token frame check sequence (TFCS) in the 8 bits following the destination address is correct (indicating no error in the frame), the station accepts the token and performs the proper message transmission/token transmission activities for the current station situation.

When receiving a message frame, the station must check the header frame for the following information to determine whether the message is being transmitted to it or whether it is a broadcast message desired by the host:

a. Determine frame type (claim token, token, station management message, data message)
b. Determine destination address or capture logical address
c. Capture word count

Once the frame type is determined, the station can determine the exact type of message (claim token, station management, or data) and activate the appropriate data handling circuits. The destination address can then be determined to be physical or logical. If the flag bit is a "0", then the next 7-bits can be checked for the BIU physical address. If the flag bit is a "1", the message filter circuit (3.6.5.2) must be activated to determine if the host wants to receive this message. After it is determined that the message is desired to be received, the word count field is captured and is used to indicate to the receive machine the total number of words which are to be received in the information field. This can be used for a overall error check on the message reception process. After the total word count has been received, a final 16-bit word is received which constitutes the message frame check sequence (MFCS), generated by the transmitting station. If the result in the frame check generator of the receiving station matches the frame check word transmitted, the message is without error and can be passed to the host. A nonmatching result in the frame check generator (a residual value of zero) indicates that an error has occurred somewhere in the header or data fields and the message must be discarded.

3.6 Frame Convention:

This section explains the meanings of the various components of the frames transmitted on the data bus. In some cases, where not immediately obvious, some explanation of the rationale accompanies the description. Figure 3.6-1 details the components of the three frame types.

3.6.1 Preamble: The preamble is generated by the transmitting BIU as a clock reference for the receiving stations. The receiving BIU uses this clock reference to synchronize the local receiver clock to the frequency and phase of the transmitting clock to assure error free data decoding. The standard allows the system designer to program the length of the preamble to minimize the overhead required in certain systems where the clock recovery scheme is sufficiently fast to capture proper synchronization in fewer bit times.

The exact characteristics of the preamble will vary according to the data encoding scheme utilized. This is documented in the slash sheet detailing the particular implementation. The preamble should always be chosen so that it contains the highest edge density possible.

Note that for multiple messages transmitted on a single token hold, the preamble is only transmitted before the first message of the multiple message frames contained in the transmission.

3.6.2 Start and End Delimiters: The start and end delimiters frame the PDU to provide synchronization of the start and end of the frame at the receiving BIU. Both the start and end delimiters are unique symbols which depend on the particular data encoding scheme utilized. The exact symbols used for each are documented in the slash sheet detailing the particular implementation.

3.6.3 Frame Control Field:

3.6.3.1 Frame Type Field: The frame type field consists of the first 3 bits of the word following the start delimiter. This field determines the exact content of the frame currently being received. This information allows the receiver to prepare the proper reception circuitry for the type of information which will be found in the rest of the received data. The definitions of the bit patterns are:

a. 0XX - Token
b. 100 - Claim Token Frame (3.2)
c. 101 - ILLEGAL FRAME
d. 110 - Station Management Message (3.8)
e. 111 - Data Message

3.6.3.2 Priority Field: The priority field is a two bit field immediately following the frame type field. This field indicates the particular priority which the message was transmitted with. It is, however, important when it is associated with certain station management messages which require a replay to an originating station. In this case, the priority field is used to indicate the priority at which the response to the station management message is to be queued for transmission (3.8).

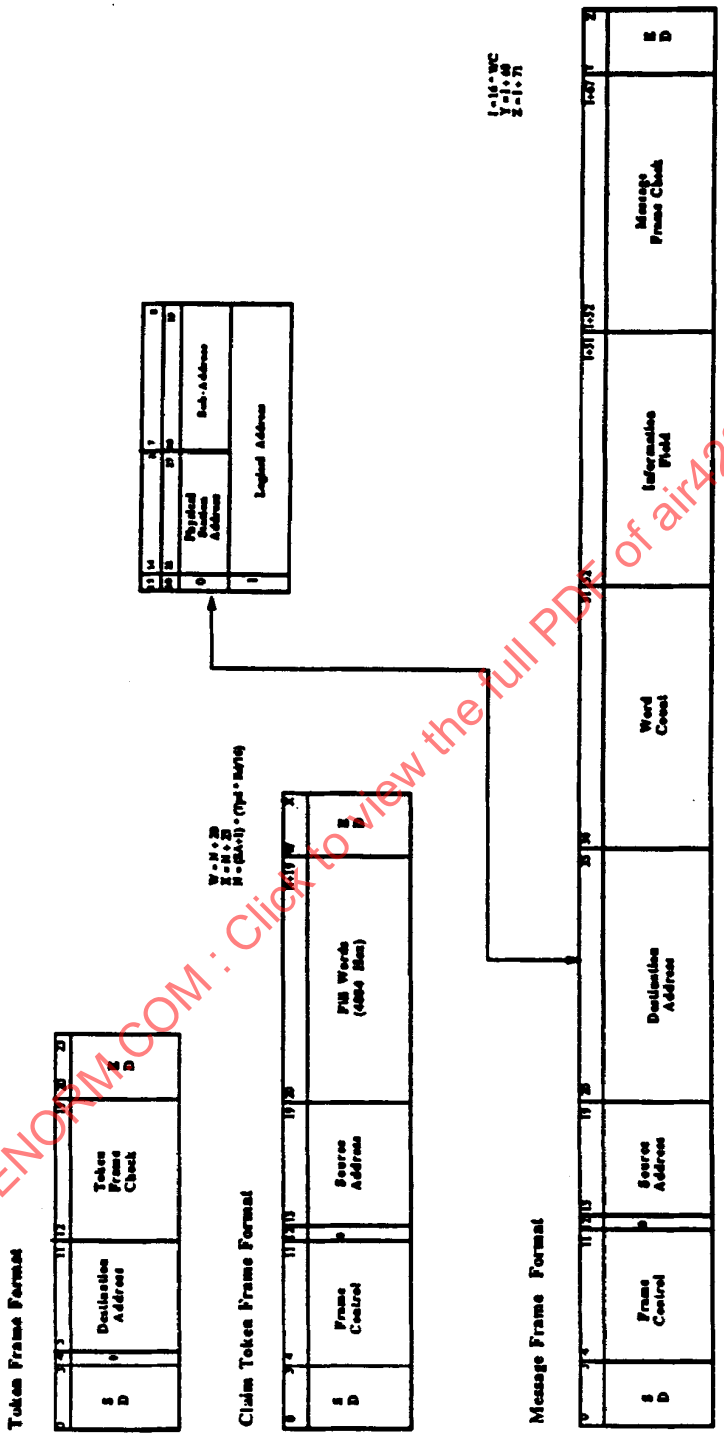


FIGURE 3.6-1 - LTPB Frame Formats

3.6.3.3 Station Management Code: The station management code field consists of the 3 bits immediately following the priority field. This field is meaningful to the BIU when the frame type indicates that the message being received is a station management message. Station Management messages, as well as the definition of the station management codes are discussed in detail in 3.8.

3.6.4 Source Address Field: The source address field is located in the 8 bits immediately following the frame control field. It indicates which station on the network transmitted a particular message frame. Normally, the information is of no consequence to the receiving station or host. However, during bus monitoring activities (such as fault analysis) having this information available can be beneficial.

3.6.5 Destination Address Field: The destination address (DA) field is the second 16-bit word following the start delimiter and is used to "point" the message toward the station or stations which use the information contained in it. The DA can be one of two types:

a. Physical address (PA)
b. Logical address

3.6.5.1 Physical Address: The PA field is active when the address type flag (most significant bit of the 16-bit field) is a zero. The remaining bits are divided into two fields:

a. The 7 bits immediately following the address type flag contain the physical address of the station to which the message is being sent.

b. The least significant 8 bits of the field contain a subaddress which routes the message to the particular entity in the host.

When a message is physically addressed, the BIU decodes the 7 bits constituting the physical address to determine if the message is directed to it. The remaining 8 bits are passed to the host with the message to indicate the intended routing of the message from the transmitting station.

3.6.5.2 Logical Address: When the address type flag (most significant bit of the field) is set to a "1", the field contains a logically addressed message. The remaining 15 bits in the field constitute an address for one of 32K addresses in a "Message Filter" located in the BIU (3.1.3.2). The incoming logical address is decoded to provide a pointer to the bit. The bit is then checked to see if the BIU has been "told" either by the host or an external source to receive messages with that logical address, using the Load/Report Configuration Command detailed in 3.82. A logical address field containing an address type flag of "1" and "1"s in the remaining lower order 15 bit positions (address field value of FFFF (hex)) is the required address for broadcast messages. These messages are required to be received by all stations.

3.6.6 Word Count Field: The word count (WC) field is the third 16-bit field following the start delimiter of a message frame. This field contains a 13-bit number representing the number of words contained in the information field which follows. This field covers the maximum allowed message sizes from 1 to 4096 words. All other WCs outside this range are invalid. The WC may be used by the receiver to determine the expected end of information field, for queuing purposes and also to determine the expected number of words to be read from the queue. The receiver must use this number to detect short frame errors, should a transmitting station or other bus fault cause too few information words to be transmitted.

3.6.7 Frame Check Sequences: Frame check sequences are used as a method of detecting errors in transmissions on the LTPB. There are two different frame check sequences implemented. One is used on the token frame, the other on message frames. There is no frame check sequence on the claim token frame. During the claim token process, the claim token frame is used to detect the presence of other transmitting station on the bus and the concern is the validity of the encoded waveform at the receiving station.

As mentioned in the following paragraphs, the FCS generators are initialized to zeros prior to beginning computation. The decision to use an initialization value of zeros instead of all ones is a rather arbitrary choice since there are no outstanding technical reasons why it should be done one way or the other. It is important that the designer heed the requirement for initialization of the generators otherwise noninteroperability will result between different implementations.

3.6.7.1 Token Frame Check Sequence (TFCS): The token is the most important frame transmitted on the LTPB. It controls access to the bus by the various stations. In order to prevent multiple stations accepting a corrupted token that looks like it is addressed to them, some type of error detection mechanism was needed. An 8-bit frame check sequence (FCS) was devised to provide this protection. An 8-bit FCS was used to maintain the token on a 16-bit word boundary.

The TFCS provides coverage of the leading zero bit, which identifies the frame as a token, as well as the destination address field. Each station must check the TFCS attached to the token to make certain that it is the proper value for the address contained in the destination field before accepting the token. In the event that the TFCS is invalid, the station shall not accept the token.

The TFCS generator is initialized with all zeros prior to each transmission and the 8-bit value computed is transmitted immediately following the destination address in the token. The value computed is not inverted prior to transmission.

3.6.7.2 Message Frame Check Sequence (MFCS): Error detection on message frame is provided by a 16-bit frame check sequence which is appended to the message frame. This 16-bit word is computed by the transmitting station and covers the entire message frame header (3.6.3) and all words contained in the information field. Each receiving station must generate an MFCS on the received data and compare it to the MFCS attached to the message. They should be identical if no errors have occurred. If an error is detected, the message must be flushed from the buffers. The host is not notified of the message arrival or of the detected error.

3.6.7.2 (Continued):

The MFCS generator is initialized with all zeros prior to each message frame transmission and the 16-bit value computed is transmitted immediately following the last message word in the information field. The value computed is not inverted prior to transmission.

3.7 Error Handling:

This section provides that context for explanation of the basics of fault recovery. The following subjects will be briefly addressed:

a. terminology
b. pathology
c. implementation (of dependability characteristics)

Subsequently, a fault recovery strategy will be developed for the linear LTPB, based on three ground rules:

a. on-line recovery
b. off-line recovery
c. harmony recovery

3.7.1 Terminology: This section provides an overview of the basic concepts of "system reliability" or, more appropriately, "system dependability". This is concerned with that dependability which applies to data processing (DP) and data communications (DC) (i.e., several pieces of equipment connected by one or more LTPBs).

The goal of DP and DC system dependability is to design, to implement, and to use DP and DC systems in which faults are considered to be natural, foreseen, forecasted and tolerable events.

The dependability of a system can be defined as the quality of service that this system is capable of delivering to its users such that all users can place a justified confidence in the systems ability to accomplish the required service.

A system failure occurs when the accomplished service differs from the required service in given circumstances. An error is that part of the system state which deviates from what it should be in order for the system to have behaved properly (i.e., to be able to accomplish the required service). Finally a fault is the phenomenological cause of an error:

a. a failure is an event,
b. an error is a state,
c. a fault is a cause.

3.7.2 Pathology: A fault can be the result of human or physical activity. Physical faults are disorders which impact the system behavior, either internal or external to the system.

Human faults can occur when the user is interacting with the system. These can occur during the design phase (from requirements through physical implementation), during subsequent modification phases, or as a result of improper definition of operational and maintenance procedures. Interaction faults occur because of voluntary or involuntary violations of properly defined operational and/or maintenance procedures.

Internal faults are physical failures of the system due to usage, such as short circuits, damage, and breakdowns. External faults are environmental disturbances such as EMP, pressure, temperature, and vibration, which adversely impact the system behavior.

As soon as a fault occurs, it creates an error. Most often, this error stays latent (hidden or masked) until becoming effective (unmasked or visible) when activated by specific operational circumstances.

An error can cycle between "latent" and "effective" states. It can propagate from its origin (chip, module, or architectural layer) through other system parts (other chips, other modules, or other architectural layers).

A failure results from the activation of a latent error by specific operational input conditions (unmasking). It is detected by the means of differences between the expected service and the actually accomplished service. After a failure occurrence, the delivered service is said to be unexpected. A system life is a continuous alternation of two modes of operation: expected/ unexpected delivered service.

Three quantities give insightful measurements of a system's dependability (the so called RAS parameters):

a. Reliability,
b. Availability,
c. Serviceability.

Reliability is a measure of the system's mean time between failures (MTBF). It is the period during which a system delivers continuously, the expected (required) service.

Availability is a measure of system's ability to deliver the expected service in the presence of faults. It is the service delivery rate relative to the system alternation of expected/unexpected service deliveries.

Serviceability is a measure of the duration of the system failures. It is the continuous time of unexpected service delivered by the system after a failure occurrence.

3.7.3 Implementing a Dependable System: Implementation of a dependable system demands that its designer use the methods listed below:

a. Specify, design, implement and describe system utilization rules to minimize all possibility of fault occurrence by system construction and fault avoidance methods to ensure continuous accomplishment of the required service, in spite of foreseeable fault occurrences.

b. Allow users as well as designers to place confidence in the system's ability to provide the required service by minimizing the number of existing latent errors through use of system verification methods and by forecasting all possible consequences of errors due to foreseeable faults.

It is important to point out that the validation concept requires system designers and builders to use subjective, as well as objective, methods during the system design-implementation-realization process. An example of this is the basic notion of a validation-prone design. Everyone must realize the importance of this key issue. The results of proper application of the concept will be a natural willingness of the human operators and users to place their confidence in the system's ability to provide the required service. Each designer must realize that every DP and DC system design must be devoted to robustness and simplicity.

After having defined the concepts and the terminology, we will now turn ourselves towards the recommended fault recovery strategy for the linear LTPB. This strategy will be developed on three ground rules:

a. on-line recovery,
b. off-line recovery,
c. harmony recovery.

3.7.4 On-Line Recovery: Many of the BIU functions, if they fail, can have a severe impact on the performance or on the integrity of the LTPB system. The following is a list of the key hardware based functions which can cause such failures:

TABLE 3.7.4 - Key Hardware Based Functions

BA (Bus Activity)	Detector	(TP mechanism)
NS (Next Station)	Register	(TP mechanism)
NS + 1	Generator	(TP mechanism)
TS (This Station)	Register	(Successor Hunting mechanism)
TS + 1	Generator	(Successor Hunting mechanism)
THT/TRT		(Token Holding Timer & Token Rotation Timers)
TPT		(Token Passing Timer)
RAT		(Ring Admittance Timer)
BAT		(Bus Activity Timer)
CTS		(Claim Token State behavior)
MSA	Register	(Maximum Station Address)

3.7.4 (Continued):

All such mechanisms must be thoroughly identified since they require constant monitoring by each BIU. In case of any detected fault in any of them requires immediate corrective action, and notification to the host (user). The combination of all of these individual monitoring functions is termed the self-monitor function (SMF).

3.7.5 Off-Line Recovery: Monitoring for failure mechanisms which only effect the faulty station and its host but do not impact the transmission system as a whole, should not be included in the SMF. These failure mechanisms must be detected by other appropriate test actions such as power-on self-test or periodic self-test. In addition the SMF itself must be subject to all such off-line tests just as is the rest of the BIU.

3.7.6 Harmony Recovery: Both previous sections (ON-LINE and OFF-LINE RECOVERY) dealt with the concepts of BIU self-checking and self-testing activities. These concepts are needed to validate the individual operational behavior of each BIU (i.e., the consistency of each BIU's parameters relative to the required BIU's behavior).

Bus wide consistency is another subject. It is concerned with the consistency of the parameters of all of the BIUs in a LTPB system (i.e., relative to one another's interactions). This task requires the use of a system monitor. The task of that system monitor should be to ensure that all BIU parameters will result in an overall cooperative operational behavior of all BIUs, in harmony with one another.

3.8 Station Management Concept:

AS4074 was designed with Station Management functions as a means to allow the system designer to implement a highly testable system. The commands available are meant to allow maximum visibility and control of the BIU, either from the BIU-to-Host interface or from an external control source, across the bus network itself. Note that when a station management command for a BIU is received from an external source, that the command is reported to the user. This allows the user to verify that the command is proper and, if required, countermand the command. The user also has the capability to disable reception of and response to station management commands received across the bus. This allows the user to prohibit a malicious outside station from using station management commands to cause unwanted or unnecessary changes.

This section discusses the various station management command functions and fields. Use of station management functions for testability is discussed in 3.15. Notice that in many cases, the user determines the priority at which station management messages are handled. This allows the user to determine the operational impact of the situation he is working with and tailor the system impact (is it serious enough to use top priority bandwidth to get it through fast?) of transmitting the message.

3.8.1 Mode Control Command/Status Report: This command allows the system designer a means to control the various modes of operation of a BIU, as well as collect various types of information about BIU and system operation. The command allows the user to command a complete RESET and/or BIT on a BIU. The command also allows station condition information to be ascertained by use of the mode code which requests a two or ten word status report. Functional control over the Global Time Reference (GTR) (3.12) is also provided in this station management command.

3.8.2 Load/Report Configuration Command: Each BIU contains several programmable timers, control values, and tables which the system designer may use to optimize the performance of the system. Among these programmable values are:

a. Token holding timer (THT) (3.11.3)
b. Token rotation timers (TRT) (3.11.3)
c. Ring admittance timer (RAT) (3.9.1)
d. Token passing timer (TPT) (3.9.2)
e. Bus activity timer (BAT) (3.9.3)
f. Maximum number of stations (MNS) (3.13)
g. Message filter table (MFT) (3.13)

The load/report configuration command is utilized to perform the programming of these functions. The various fields within the command are used to load system specific values to replace the default values which are automatically loaded into the registers at power-up and during any system/host commanded RESET.

The command can also be used to read back the values loaded into the various programmable registers to allow the host/system (or designer, during system integration) to verify that the data has been loaded correctly or that the data has not been corrupted by some system occurrence which might be considered anomalous. Note that a report configuration command will cause the entire message filter table to be reported. There is no provision for partial message filter table report.

3.8.3 Test Messages: The ability to verify proper operation of the various data paths throughout the LTPB system is vital to testability of the system - during both operational, system integration, and system/BIU level testing. In order to provide visibility to the data paths, test messages, or wrap-around test capability has been provided in the station management command complement.

Two methods of test are available at the BIU. The first tests the ability of the BIU to receive messages across the LTPB, properly decode the message, store it in the transmitter queue, and transmit it back to the source station during the next token hold of the station being tested. The host interface of the originating station is also tested. The second test allows the host to test the Host/BIU interface. The method of handling is similar to that of the first test message. The host, utilizing the proper interface protocol, places a message into the BIU transmit queue. The station then takes the message, transfers it to the receive queue, and notifies the host that a message is present. Both messages are described in more detail in the following paragraphs.

3.8.3.1 Loopback Operation:

3.8.3.1.1 HOST/BIU Message Loopback: Issuance of this command by a station, to another station, requires that the sink station take the message, change the SMC from 101 to 100, exchange the source address and destination address field information within the message header, and return it to the source station as a BIU Loopback Message (3.8.3.2) when the sink station next receives the token. Upon receipt of the returned loopback message, the source station will inspect the information field contents and verify that the information matches that which was transmitted. Notice that the information field length and contents is user determined. It may contain any number of words from 1 to 4096 (or the system determined maximum length) utilizing any data pattern which fulfills the user's test goals.

3.8.3.1.2 BIU Loopback Mode: Issuance of this command by a host to the associated BIU, requires that the BIU disable its transmitters and receivers, form a connection between the local transmitter and receiver circuits, for test purposes, accept the message, place it in the proper transmit queue, route the message from the transmit queue to the receive queue (since they are connected), and notify the host of a message being present. During the transfer, the contents of the information field must remain intact. The host then reads the message out of the receive queue and verifies the contents of the information field. Again, notice that the information field contents and length are user determined, as is the priority level at which the message is to be handled.

3.8.3.2 Bus Loopback Message Echo: This message is transmitted in response to a receipt of a Bus Loopback Test Message from a source station. The Bus Loopback Message is received from the bus (source), the SMC changed from 101 to 100, the source address and destination address fields are exchanged, and the message frame is stored in the transmit queue for transmission on the next token hold, based on the message priority level contained in the priority field of the message header.

The information field from the Wrap-Around Test Message is stored and returned in the Bus Loopback Message Echo unaltered. This allows the source station to verify that the sink station (station under test) is properly receiving, storing, and transmitting data from the bus.

3.8.4 Time Synchronization Message: The global time reference (GTR) function (3.12) is addressed and controlled through the use of this station management command. This command allows the user or the time master station on the bus to control, update, and/or read the value currently stored in the registers. Note that this command is specified with a latency requirement. This requirement specifies the maximum amount of time that the BIU has to process a received time update command as well as the amount of time the station has to process a time synchronization message (when acting as time master).

3.8.4 (Continued):

The final field present in the time synchronization message is the Message Update Rate (MUR) field. This field allows the user to tell the time master how often to transmit the time synchronization message. By making this rate a user selectable value, this allows the user to achieve a needed level of clock accuracy while utilizing a less accurate (less expensive) clock source for maintaining the clock count. Explanation of this scheme is contained in 3.12. The update values available are 1 through 8, corresponding to 8, 16, 33, 66, 131, 262, 524, and 1049 millisecond update periods.

3.8.5 BIU Diagnostics: Station management is the area of BIU functionality which controls all aspects of BIU diagnostics and station status reporting. Among the functions included in this area are:

a. Full (interruptive) power-up BIU test
b. Loopback tests
c. Background (noninterruptive) BIU test

Testing of data paths has been covered in the discussion on loopback message modes (3.8.1 and 3.8.2) earlier in this section.

3.8.5.1 Full (interruptive) Power-Up BIU Test: The BIT capability of the LTPB BIU should be designed to perform an exhaustive test if all logic and memory (and associated addressing/control circuitry) which comprise the BIU. This BIT should be configured to be performed immediately upon power up as well as on command (from Mode Control command) from a source on the LTPB or the Host. This test should be performed within a 2-second window after power has been stabilized. The BIU should still respond to requests for status from the host across the Host/BIU interface. Among the tests required to be performed are:

a. Verify all internal data paths on bus and host interface paths
b. Verify parity on hardwired BIU physical address pins
c. Verify functionality of message filter circuits
d. Check functionality of Token Passing circuitry
e. Test all BIU memory/storage (including control ROM, if any)

The BIT should be capable of disabling the transmit function of the BIU such that a detected failure does not accidently propagate onto the active bus. The BIU module should have a visual indicator which indicates a no-go condition.

This test should also be activated when the station receives a PERFORM BIT command (see 3.8.1).

3.8.5.2 Background (Noninterruptive) Diagnostics: During normal BIU operation, the BIU will perform a background diagnostics test to assure the functionality of the unit. This test is performed in a noninterruptive manner - that is, the test should not affect the state of any registers which are currently being utilized in normal operation of the bus. Normally, this type of test interrogates the status of various circuits in the hardware to determine their activity status. Inactive blocks of circuitry are then quickly tested and the results (if a failure is detected) are stored in a status word for reporting. The host may access this status word via a status update request across the Host-BIU interface. The same information may be requested by an external source by means of the LTPB.

3.8.5.3 Use of External Loopback Test Messages:

3.8.5.3.1 Equipment State Determination: In the event of a failure in the user, the BIU is no longer useful, even if operative. In such a case, the equipment is said to be in an error state. In order to detect an error state, a wraparound test must test the following:

a. The BIU
b. The BIU/user interface
c. The user

Usually, a predefined data pattern is sent to the equipment being tested. The data pattern is then returned to the sending entity and compared for accuracy. This is the same type of test available in the AS4074 LTPB standard. They are detailed in the station management section of the standard. The two tests available are: the bus loopback test message and the bus loopback test message echo.

The action to be taken in the event a fault is detected depends on the type of response:

a. If the response contains an error in the data pattern, the tested equipment should be disabled. (This reaction could be modified to take place only after a failure of two or more attempts, to avoid disabling on transient problems.)

b. If there is no response to the test at all, the unit is said to be in a dialog error state. The unit should not be disabled since it may be unpowered.

From a general point of view, disabling of faulty equipment is not useful to the overall network scheme since the error impacts data only, not general network performance characteristics.

3.8.5.3.2 Equipment Point of View: It is important to the user to know if data is being correctly received from the bus by the BIU or if the bus is operating normally. For example, a frozen display in an aircraft could have disastrous results. This could be avoided by use of loopback test. By periodically sending a loopback message a user can be assured that the network is functioning properly.

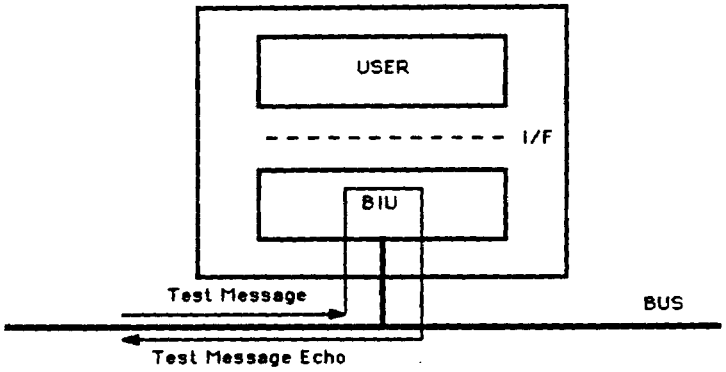


FIGURE 3.8.5.3.2-1 - Path of Loopback Test Message and Test Message Echo

3.8.5.3.2.1 Equipment Start/Restart:

3.8.5.3.2.2 Start: At powerup, the BIU and the user are initialized (generally simultaneously). When the user has initialized, and determines that the BIU has completed initialization (or vice versa). At this time, message transfer can start. During user initialization, messages may not be sent. Here an "initialization flag" would be desirable. If messages are sent during the initialization phase, then data invalidation mechanism must be defined.

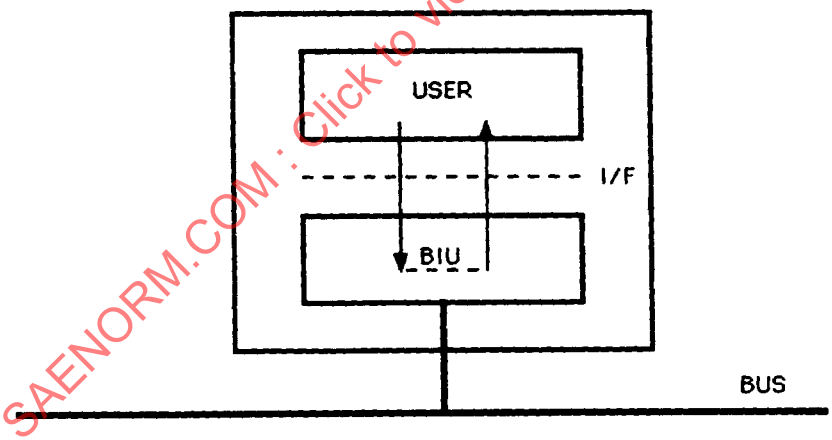


FIGURE 3.8.5.3.2.2-1 - BIU/User Verification Utilizing Self-Test Command

3.8.5.3.2.3 BIU Restart: A general order which would be followed is:

a. First: Try to restart the user. If unsuccessful, nothing else is done. If successful, go to step two.

b. Second: The user commands the reinitialization of the BIU.

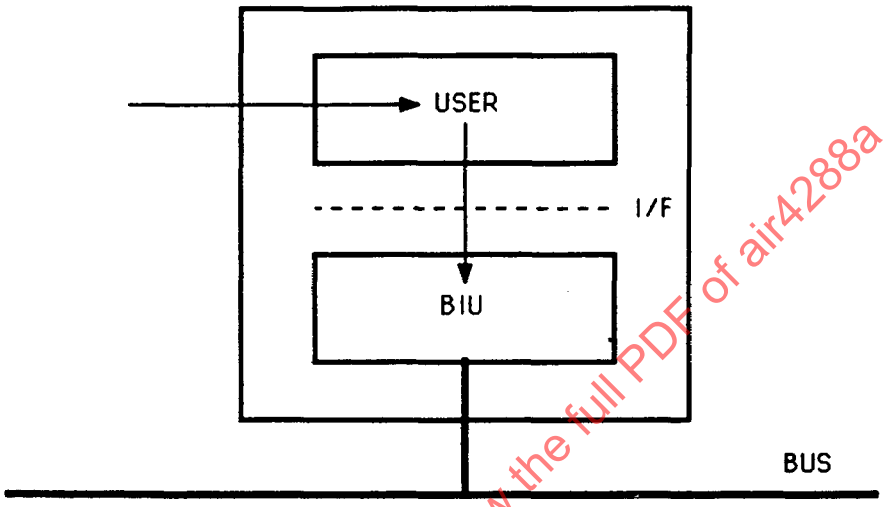


FIGURE 3.8.5.3.2.3-1 - User Commanding Reinitialization of the BIU

Meanwhile, the wraparound test is completed. As soon as the equipment has been reinitialized a good wraparound test will be observed.

3.8.5.3.2.4 Bus Restart: The same procedure as the one described in the case of a discrete reset is followed. A resident mechanism is required to initialize the user tasks through communications functions not yet initialized.

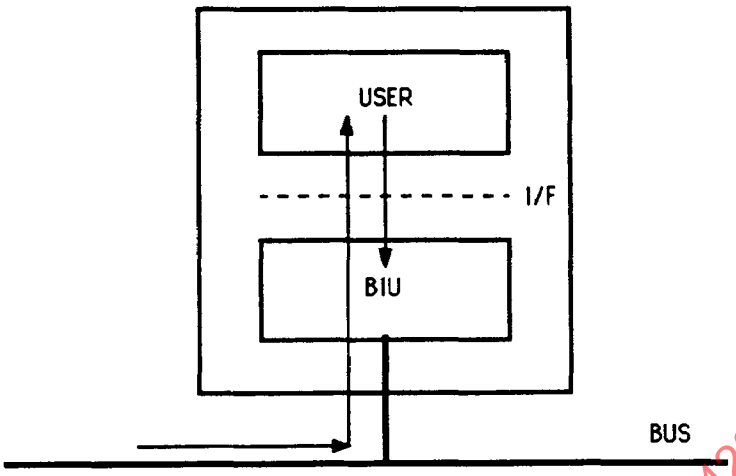


FIGURE 3.8.5.3.2.4-1 - Restart of the BIU from the Bus

3.8.5.3.3 Test Management: The management of system equipment tests and function seems to be easier when the control is centralized. However, distributed control of these activities will be implemented in many cases. This requires that certain problems be addressed including equipment state tables in each user, definition of who tests who, and test sequence integrity.

3.8.6 Traffic Summary Counters: The AS4074 standard specifies a number of 16-bit counters which record events which occur at the station. Some of the events recorded are events which are seen on the bus (traffic from other stations). Others are events which the station generates (messages transmitted by this station). These traffic summaries can be a useful system design/integration/fault isolation tool. For example, a bus system which is crashing for unknown reasons can be polled by a test device from a single point on a regular basis during operation and the make-up of the traffic can be determined from the summaries in each station. The potential "bad station" can then be weeded out from the statistics reports. The traffic events which are maintained in the station are:

a. Valid Messages Transmitted: Counts the number of messages frames that the station transmits with no errors.

b. Claim Tokens Transmitted: Counts the number of claim tokens transmitted (that is, how many times did the station attempt to initialize the bus).

c. Transmissions Aborted: Counts the number of times that a station aborted transmission of a frame due to a perceived error in the transmission.

d. Frame Validity Errors: Counts the number of messages that a station transmits which fails validity test during transmission monitoring. There is a separate counter for each bus medium implemented.

e. Frame Received Errors: Counts the number of frames not transmitted by this station, which contain validity errors. There is a separate counter for each bus medium implemented.

3.8.6 (Continued):

f. Valid Messages Received: Counts the number of frames not transmitted by this station which are received with no detected validity errors.

g. Receive Queue Overflow Errors: Counts the number of times that a message could not be passed to the user because the receive queue overflowed, indicating a problem between the BIU and the user.

These counters operate on a continuous basis - that is, the counters will rollover to zero when the next event is detected on a full count. This requires that the user software be designed to poll the counters on a regular basis when the traffic summary information is important to the user. The software must then perform the simple mathematics to extract the number of events which occurred since the last polling. This information can then be used to generate information on individual station operation and to isolate problem stations. The counters can only be reset by direct command using the station management message.

3.9 Timer Concept:

The following timers are considered to be "protocol related" timers. That is, their job is to control the overall higher level functioning of the token passing protocol (e.g., error recovery, station admittance, initialization). For a discussion of the message handling timers (message priority related) see 3.11 and Section 4.

3.9.1 RAT: The RAT is used as the interval timer required for the ring admittance procedure. The ring admittance feature allows an inactive station, which is presently not part of the logical ring, to be admitted to the logical ring on a periodical basis. The duration of the period is determined by the value set into the RAT. The period of the RAT is a 16-bit, user programmable value which can be between 0 and 6,553.5 milliseconds. A default value of 100 milliseconds will be automatically set at initialization. This value can be changed by the use to a different value, if desired.

An active station will also initialize its RAT to the maximum value upon completion of the admittance process. This will be explained in the following text which is a description of the total ring admittance procedure.

Any station which is a part of the active ring, upon the expiration of its RAT, is now conditionally allowed (only during periods of low bus traffic) to admit a presently inactive station to the active ring during its next token hold period (or present token hold period if the RAT expired during its token hold period). When an active station that has an expired RAT (which will be called this station (TS)) receives and accepts the token, TS will initiate and perform its normal message transmission process. In accordance with the message priority system implemented in all stations, lower priority messages may be deferred during periods of high bus traffic. Also, since the ring admittance process is deferred during periods of high bus traffic, TS will only initiate the ring admittance process if, after transmitting all messages in the various active queues associated with all priority levels, time still remains on the TRT.

3.9.1 (Continued):

If a check of TRT3 shows it has not expired, and there is a gap between the address of TS and the current successor station, TS will set the next station address (NSA) to TS+1 and attempt to pass the token to station TS+1. If TS fails on two successive attempts to pass the token to station TS+1, TS will increment the NSA to TS+2 and repeat the process until a station is found that accepts the token. If all inactive stations contained between TS and the original successor station (SS) do not want to be admitted to the active ring, when the NSA becomes SS, the ring admittance process will stop since the SS accepts the token and begins to transmit (TS will hunt the gap between the present token holder and its SS). Once the token has been successfully passed by TS, TS will initialize its RAT to the maximum value and set its successor station address to the value of the station that accepted the token.

During periods of high bus traffic, TS (with an expired RAT) will not initiate a ring admittance process since TRT3 will have expired. In this case, TS will pass the token to its established SS in the normal manner. TS will not initialize the RAT to the maximum value and upon receiving the token on the next rotation, if bus traffic is low, (TRT has not expired) will attempt to admit an inactive station into the active ring. TS will not initialize the RAT to the maximum value until it has performed and completed a ring admittance process.

3.9.2 TPT: The TPT is an interval timer that is used to determine if the token has been successfully passed from a station (which will be called TS) to the successor station. TS, upon passing the token to its SS, monitors the bus media for a period of time determined by the TPT. If the token has been successfully passed, the SS immediately starts transmitting (within the allowable station response time (T_{sr})) which will result in TS detecting bus activity prior to the expiration of the TPT. However, if no bus activity is detected prior to the expiration of the TPT, TS will conclude that a token pass failure has occurred. In this event, TS will enter its recovery procedure (3.3) which consist of retransmitting the token to the SS. If a token pass failure to the SS occurs for the second time, TS will increment its SS address by one and continue with the recovery procedure.

The AS4074 standard specifies that the TPT be an 8-bit counter which results in a TPT time range from 0 to 10.20 microseconds. The value set into the TPT is determined by three factors: (1) the worst case round trip propagation delay between the two furthest stations on the bus, (2) time for the SS to respond, and (3) time for TS to detect the response. To determine the value which should be set into the TPT, a station response time of 500 nS is equivalent to the maximum station response time. Also, a value of 20 bit times has been recommended for TS to detect a response. Twenty bit times (consisting of the start delimiter and preamble) was determined to be the maximum number of bit times it should take for the clock to phase lock and for the BIU to detect bus activity. The system designer should take these worst case numbers into account when computing a value to program into the TPT for his system.

3.9.3 BAT: The setting for the BAT present in the standard is based on a scheme which allows the network to quickly initialize in a distributed, yet predictable manner. At the same time, this scheme allows compatibility with both wire and fiber optic systems. The BAT is a user programmable 11-bit timer with a range of 0 to 2047 microseconds. It is initialized, at power-up with a default value of 15*(station physical address + 1).

3.9.3 (Continued):

The BAT is set based upon the station physical address. The following equation describes the relationship:

$$BAT = (\text{station physical address} + 1) * (T_{tp} + T_{sr} + T_{pd} + T_{ba}) \text{ microseconds}$$

which can be reduced to:

$$BAT = (\text{station physical address} + 1) * (2T_{sr} + 3T_{pd} + 2T_{ba}) \text{ microseconds}$$

The resultant timer value assures that a dead bus situation will always be resolved such that the lowest physically addressed station active on the bus will win the claim and begin circulation of a token to rebuild the logical ring. Figure 3.9.3-1 illustrates a system in initialization.

Looking at a system power-up scenario, it is recognized that all stations will not enter the same BIU states at the same time. This is due to the different amounts of time required for power supplies to settle, internal BIT to be completed, and other related factors. Therefore, there will be contention for claim token in a normally operating system during this stage. In the figure, an "N" station system has just been powered up. Stations 1 and 3 happen to power up in a way which causes their BATs to expire simultaneously. Stations 2 and "N", still have time remaining on their BATs. Stations 1 and 3, believing that there are no other stations active on the bus (correct assumption) each begin to transmit a claim token frame. This activity causes two events to happen. First, the activity is detected by Stations 2 and "N" which initialize their BATs to the maximum value and go to IDLE. They assume that since activity has occurred on the bus that the bus is now active and that they are to avoid claim token state and, instead, wait for a valid token pass. The second occurrence is that Stations 1 and 3 each detect a coding error in their receiver which indicates that their claim token frames have collided with other activity on the bus. Stations 1 and 3 immediately abort their transmissions and initialize their BATs to the maximum value. The next event which will occur, will be the expiration of the Station 1 BAT (since time is based in station address, the Station 1 BAT will be the smallest value on the active bus) which will cause it to transmit another claim token frame. This time, the station will see no error in the claim token transmission, since Station 3 has not yet had a BAT expiration. Station 3, upon seeing the claim token frame, will assume another station is active on the bus, will initialize the BAT to the maximum value and will go to IDLE. Station 1 , having completed the transmission of the claim token frame, without error and waits Ttp without detecting bus activity, will have won the token claim and will proceed to pass the token and begin construction of the logical ring.

Bus recovery works in a manner identical to power up, except that all stations will have the same reference point in time (in other words, they all see the bus go silent at the same time) therefore, the lowest physically addressed station in the system will end up with the token on the first try, since the BATs will begin counting down simultaneously, with the lowest addressed station BAT expiring first.

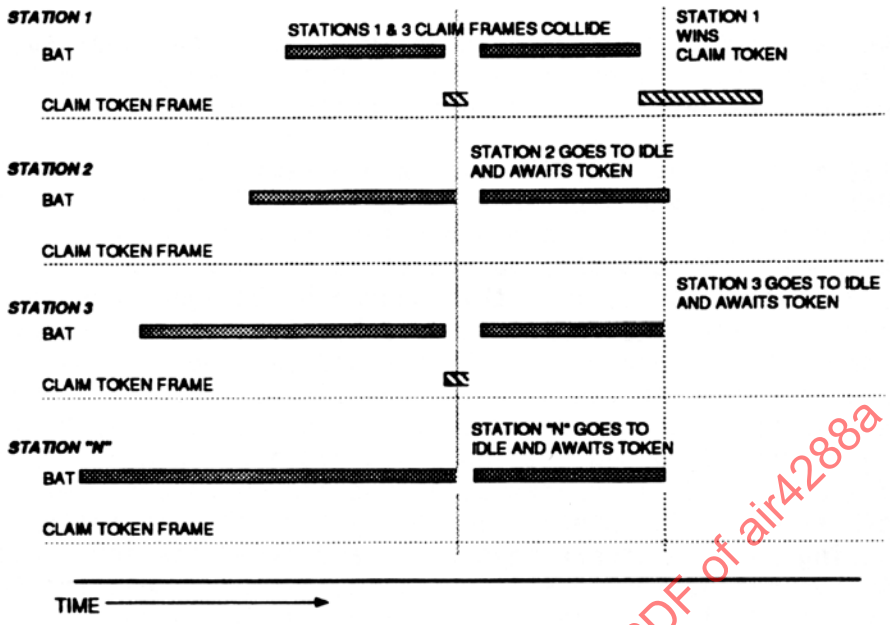


FIGURE 3.9.3-1 - System Initialization Diagram

3.10 Station Address Assignment and Physical Placement:

The AS4074 standard imposes no restrictions on the assignment of station addresses or physical placement of the stations relative to each other. There are, however, several guidelines which may improve system performance when the bus is used to interconnect subsystems over a longer distance (length of media connecting the stations at the furthest points on the bus greater than 400 m).

3.10.1 Station Address Assignment:

3.10.1.1 Initialization: During initialization, the station with the lowest physical address will normally win the claim token process. This is due to the method used to determine the setting of the bus activity timer (BAT). The setting of the BAT is discussed, in depth, in the claim token and BAT sections of the handbook (3.2.1 and 3.9.3) and is not repeated here.

Despite the fact that AS4074 is designed to be a distributed control bus, some users may wish to designate a station or stations which will be responsible for certain "housekeeping" functions on the bus. These functions may occur at power-up and reconfiguration due to station faults or bus failures. The user, in this case, should consider assigning those stations in the lowest addresses in the system. This would allow those stations to be the most probable winner(s) of the claim token. Having them assigned as consecutive physical addresses has the effect of making them redundant in that, if the main station (lowest address) fails, the next logical winner of the claim token will be the backup station, having the next lowest physical address. The logical ring will then be built up using normal token passing from the lowest physical address to the highest physical address.

3.10.1.2 Station Addition: AS4074 uses a programmable ring admittance timer (RAT) to create a user variable periodic function which allows stations not currently active on the bus (due to mission requirements or station failure) to be admitted to the active network. The function of the RAT is discussed in 3.9.1 of this handbook. Designer consideration in the area of address assignment should be given to making the ring admittance time as small as possible. This could be done by grouping all stations in as small an address space as possible. This eliminates interstation address gaps and would reduce the number of times a hunt would be required to find new stations. In addition, use of the maximum number of stations (MNS) programmable function which declares the highest addressed station present in the system will eliminate system level hunting for the higher physically addressed stations which do not exist. The highest addressed station would, instead, send the token directly back to the lowest addressed station in the system.

3.10.2 Physical Placement: While it is not necessary to assign station addresses on the basis of where the station is physically located, some performance benefits may be obtained in larger systems. In particular, if stations with adjacent physical addresses are physically located next to each other, then the time required for the predecessor station and the successor station to exchange tokens and detect each other's bus activity could be reduced by:

$$2 * (\text{bus length} * \text{propagation delay of the media})$$

In longer bus trunks, this minimization of media could provide a desirable performance improvement.

3.11 Message Handling:

3.11.1 Transmit and Receive Queue Sizing Considerations: For purposes of this discussion the term "queue" will be used to describe the transmitter and receiver message buffering space. This term should not be construed as limiting the designer to a FIFO type of memory. Other implementations, such as RAM with pointers or other memory configurations are acceptable implementations.

3.11.1.1 Transmitter Queue Sizing Considerations: The BIU will most likely require a transmitter queue which will provide for the temporary storage of all data messages that are to be transmitted on the LTPB. This temporary storage will dissociate the requirements of the serial, token-passing bus from the various parallel, internal host-user interfaces. The word size of this transmitter queue is most likely to be 16 bits; this is the size of the words on the LTPB. The access bandwidth of the queue would have to allow for providing words to the LTPB at full speed without introducing gaps between words. In a 50 Mbps implementation this translates into one 16 bit word every 320 nS. It is also desirable to have the transmitter queue servicing the host user simultaneously, which would require additional queue access bandwidth. If 16 bit words could be either enqueued or dequeued at a rate of 160 nS cycle time, then the transmitter output queue would be capable of providing for the full speed of the network to pass through the BIU as long as the host user can support these high data rates on its internal bus. There is of course a trade-off between transmitter output queue speed and size. The expected operation of the output message queue is to first completely enqueue a message that is to be transmitted on the LTPB before that message is offered to the LTPB. However, if both the output queue and the host user's internal bus are of sufficient speed, then output messages can be offered to the LTPB before they are entirely contained within the output queue. That is to say that the first word of the message can be transmitted on the LTPB before the last word of the message has been stored into the transmitter output queue. This type of implementation is strongly discouraged because of the requirement for contiguous data to be contained within the message frame that is transmitted over the LTPB.

If the operation of the output queue for a particular BIU is designed with the data generation and transmission requirements of the host user in mind, then an optimum size and speed can easily be determined. If the data that is generated by the particular host is widely varying in size, and its creation rate is entirely asynchronous, then a large transmitter output queue would be recommended to alleviate overflow conditions associated with those periods of time when the buffers are created at the maximum rate and of the maximum size. If, on the other hand, the data created by the BIU's host is synchronous and its message size is constant, then a transmitter output queue large enough to accommodate two buffers (the message being output and the next to be output) would suffice. This would be even more appropriate if the data were refresh data. Then the primary concern would be to maintain the consistency of the data sample (ensure that a message containing half old and half new data is never transmitted).

Attention must be paid to the access of the transmitter output queue from the two distinct users, the host and the BIU. Since the BIU will only possess the token for a fraction of the total time, the output queue can be refilled by the host user during those times that the token is not in the possession of the BIU. However, the token rotation rate is not entirely predictable and for this reason priority must be given to the BIU whenever a conflict in queue accessing exists. It is recommended that the enqueuing and dequeuing of data be implemented entirely asynchronous to each other so that simultaneous operation will not produce undesirable effects.

3.11.1.1 (Continued):

The size of the transmitter output queue must also consider the four classes of priority that are associated with the output messages. If we assume for a more general purpose BIU that all four classes of priority will be supported, then enough storage capacity in the transmitter output queue to accommodate at least two maximum message sizes for each of the priorities would consume 32 768 16-bit words (the message being output and the next message to be output).

There may also be some transmitter output queue storage requirements derived from the station management messages depending upon how these features are implemented into a particular BIU.

3.11.1.2 Receiver Queue Sizing Considerations: The designer should consider the worst case scenario for bus message reception and host data transfer when designing the receive queues for the BIU. Again, one should consider the asynchronous characteristics of the interface. That is, a high speed serial interface on the bus side and the user interface (which could be any one of many available bus protocols) on the other. The key design consideration is providing enough queue memory to allow the queuing of the worst case number of message words based on the projected access and transfer time of the host interface. In other words, how fast can the host vie for and get the host interface and transfer words out compared to how fast the receiver puts them in.

It should be noted that data integrity should be considered here also. There is a temptation to limit the amount of receive message queue space by immediately passing the words to the host upon reception. This scheme poses a distinct danger of passing a corrupted message to the host. In addition, systems which require implementation of data security features require that a message be validated (that is received with no errors) prior to notifying the host that a message has been received (Section 3.14). This means that the designer should provide a queue which will hold the entire message so that the CRC can be validated prior to notifying the host of the message reception in the BIU. It is recommended that the design provide enough space in his BIU receive queue to provide queuing for a worst case number of messages, based on the data transfer capabilities of the host interface. Provisions must be made in queue memory to provide the host with not only the message information, but any control information from the header which might be important in the recognition and/or utilization of the information (e.g., Logical address/message content type).

3.11.2 Transmitter Queue Management (Priority Implementation): Management of the transmitter output queue will consist of accepting messages from the host, storing these messages temporarily, and providing these messages to the transmit circuitry when the appropriate conditions exist in the BIU (i.e., possession of the token, token holding timer and token rotation timer not expired).

3.11.2 (Continued):

A simple implementation would provide four separate transmitter output queues, one for each priority. If any data structure besides this four queue structure is implemented, there is most likely to be some "garbage collection" algorithm that will be needed to be executed periodically to reassemble fragmented storage areas. As an example of how these fragmentations will occur, consider the case where a single random access memory packs all output messages into the memory in contiguous fashion. Memory fragmentation will occur if any of the messages in a particular priority are required to wait for the next token possession while adjacent messages of a different priority are all sent out during this token possession.

Some mechanism must be conceived to mark the beginning and end of each message along with its priority and its destination. There will also be a need for a transmit queue not empty flag. It is advisable to completely move the message into the output queue before this flag is set to not empty. This avoids the problem of not having the entire message present at the transmitter when the token arrives and running out of data because the host interface can not provide the remaining data fast enough for the transmitter.

3.11.3 Priority System Concept: The AS4074 standard uses a number of timers to control message access to the bus. This timer based control is the basis of the priority scheme. A token holding timer (THT) controls the maximum amount of time a station may hold the token and bounds the time that the highest priority messages may be transmitted. Three other timers, called token rotation timers (TRT), control transmission of messages at three lower priorities. All of these timers are required to be 16-bit counters which allow for a range of 0 to 65 535 microseconds. The user has direct access for programming the values using the appropriate station management message (3.8).

3.11.3.1 Terms and Definitions:

PRIORITY: Priority refers to a system for defining latency classes for messages. Messages which require minimal latencies are assigned to a higher priority than messages which can tolerate longer latencies.

When utilizing the standard, the system designer must establish the priority of all messages relative to each other, regardless of which station sends them.

LATENCY: Latency is the time a message takes to go from the source to the destination. The bus system designer is concerned with that portion of the latency time during which a message can't be transferred because the station doesn't posses the token or is prevented by other traffic on the bus.

3.11.3.2 Priority System Operation: Priorities are used to determine latency classes. Priorities generally establish the order in which messages will be transmitted during the time the station is allowed to hold the token. However, this is not true when using the network. This is because what typically makes a message a high priority is how little latency it can tolerate. In our system every message must, and will get through (except in a heavy message load situation). Priorities are established to assure that the messages which are classed as vital to the operation of the system get transmitted, within the maximum allowable latency, utilizing deferral of lower priority messages to free up bandwidth on the bus.

Figure 3.11.3.2-1 shows proper operation of a four priority level implementation under normal bus loading conditions. Note that THT and TRT settings are such that there is adequate time at each priority level to complete transmission of all pending messages. Also notice the characteristics of the priority scheme. In particular, notice the time at which the THT and TRTs are reset. The THT is always reset to the maximum value upon receipt of a valid token addressed to the station. The TRTs are reset to the maximum value as permission to transmit message traffic at that priority level is granted during the token hold. Note that prior to resetting the TRT, the amount of time remaining in the timer is stored in the THT and becomes the new "bound" time for token hold unless the time remaining on TRT is greater than THT in which case THT remains at its current value. Hence, in this case the TRT does exactly what it's name implies - it "measures" the time the token takes to complete the logical ring. The THT acts as the bound for transmission of the highest priority message from a station (in a worst case scenario, a station may have enough messages to "fill-up" the THT and hence, this will bound the total amount of time the station may hold the token). Another consideration is the fact that the station may begin to transmit a message just before the THT expires. Since the station will complete transmission of this message before passing the token, this amount of time should be taken into account during calculation of the worst case times.

Now, assume that a couple of high priority message sources elsewhere on the bus have caused an urgent need to transmit large blocks of emergency data back and forth. This causes the message traffic on the bus to become quite heavy. All of a sudden there is not enough available bus bandwidth to accommodate all pending messages of all priorities. Here is where proper determination of individual message priority level and TRT setting becomes important. Figure 3.11.3.2-2 shows just such as system, during heavy loading. Note that priority level three messages are required to be deferred, since TRT3 has expired. In this case, the TRT3 is reset to maximum value and the token is passed while the message remain in queue. In a properly designed system, the deferred message will be passed within a specified time frame.

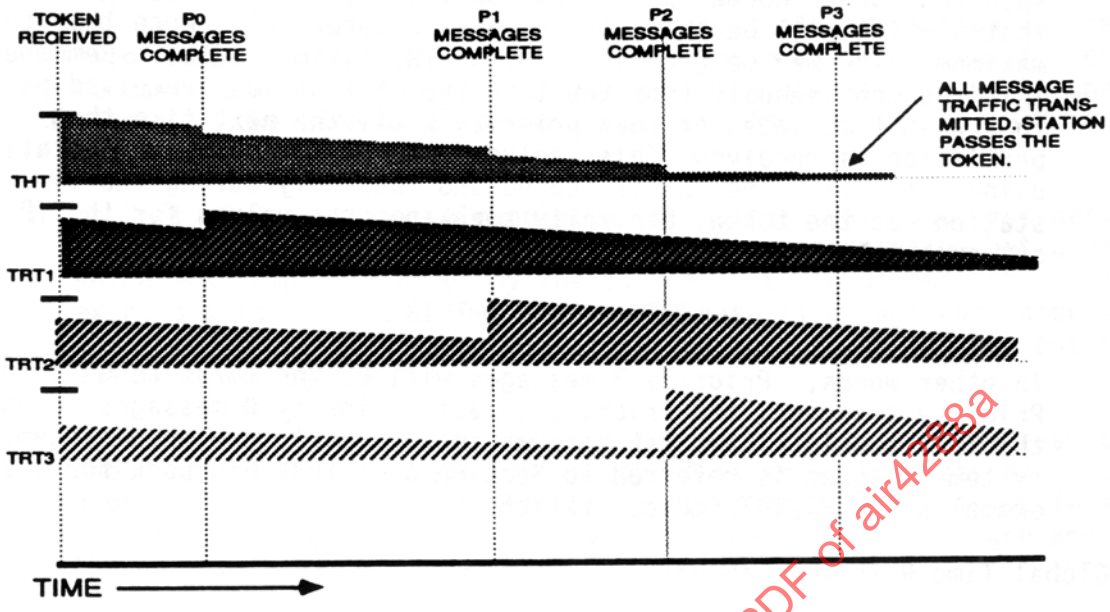


FIGURE 3.11.3.2-1 - THT/TRT Operation on a Normally Loaded Bus

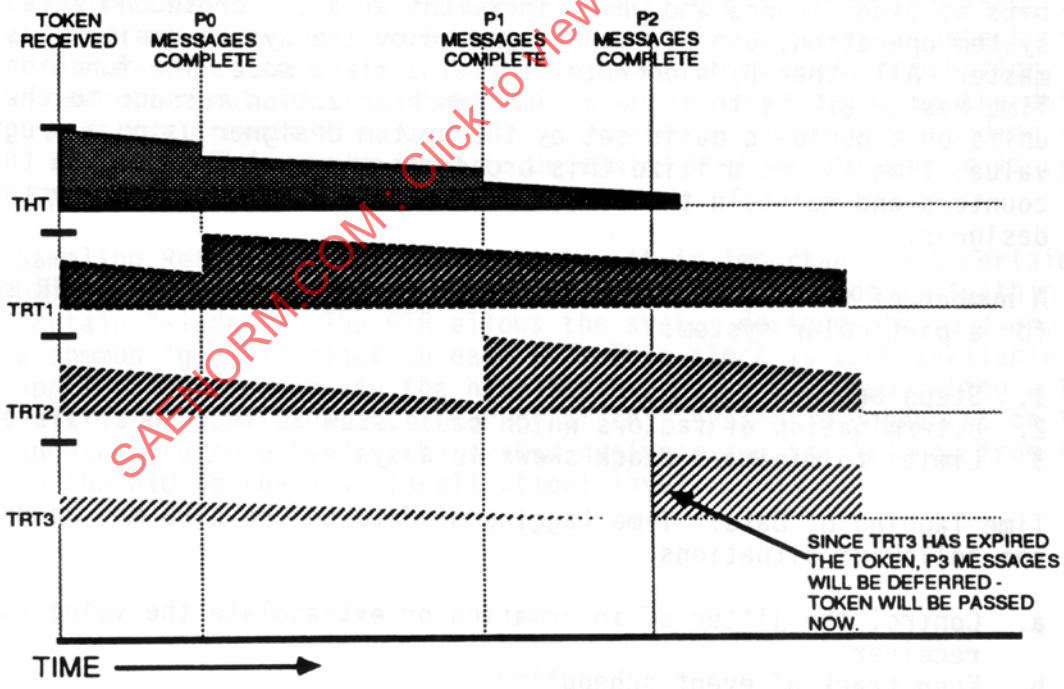


FIGURE 3.11.3.2-2 - THT/TRT Operation on a Heavily Loaded Bus

3.11.3.2 (Continued):

Again, refer to Figure 3.11.3.2-1. The value for the TRTs must be set such that under normal operating conditions, all message traffic from that station will be transmitted on each token hold. Note that the TRT maximum value may be greater than the THT maximum value. Remember - the TRTs run continuously from the time the station last received permission to transmit messages of that priority until the next time that permission is received. This includes the time that other stations are using the token. The THT is loaded and runs only during the time the station has the token. Generally speaking, the values for the TRTs are set such that the following relationship occurs:

$$TRT1 \geq TRT2 \geq TRT3$$

In other words, Priority 3 messages will be the first to be deferred, Priority 2 second, and Priority 1 last. Priority 0 messages (bounded by the THT) should always get through in a properly designed system. The system designer is referred to Section 4 of this handbook for several examples of THT/TRT implementation.

3.12 Global Time Reference (GTR):

3.12.1 Global Time Reference Overview: The GTR concept required by AS4074 has been formulated to provide synchronization for the system served by the bus. Use of this feature is optional, depending on the system designer's needs for data sampling/transmission update error limits.

The requirements state that each BIU shall maintain a local clock with 48 bits of significance and which increment at a 1 microsecond rate. In system operation, one BIU is designated by the system designer as the time master. All other BIUs operate in a time slave mode. The function of the Time Master BIU is to issue a Time Synchronization message to the slave units on a periodic basis set by the system designer using a programmable value. Time slaves utilize this broadcast time value to update the 48-bit counters and maintain the time error rate determined by the system designer.

A number of factors must be considered when formulating the GTR settings for a particular system:

1. Steps between time correlations
2. Determination of factors which cause skew in individual slave clocks
3. Limits to minimize clock skews in a system

3.12.2 Time Tagging of Data: Time tagging of data on the LTPB may be useful in the following situations:

a. Control the jitter of information or extrapolate the value at the receiver
b. Keep track of event scheduling
c. Control information integrity

3.12.2 (Continued):

Generally there is only one time tag associated with each message since it is assumed that all of the data in that message was created at the same time. If this is not the case, certain pieces of information in the message may be time tagged separately by the host. The precision of the time tagging will depend upon the application and will be up to the system designer at the time of LTPB implementation for his system. GTR accuracy is described in 3.12.4.

3.12.2.1 Control of Jitter and Extrapolation of Value: When the destination host receives information, the value is not the present value but a value which was sampled earlier by the source host. To accomplish extrapolation of the value, the source host time tags the information at the time of creation and the destination host reads the current time when processing the data.

Time tagging of information allows the user to measure the real time end-to-end (HOST-to-HOST) LTPB delay. This knowledge does not solve all problems which occur when the data arrives too late or is stale and unusable. it may not be useful even if end-to-end delays are known (in the case of a stores release).

3.12.2.2 Event Scheduling: An example of the use of time tagging to keep track of events can be during maintenance activity. It is helpful to know when certain failures occur within a system. By time tagging the data, the failure can be captured and the relationship to other events in the system analyzed.

3.12.2.3 Control of Information Integrity: A system designer who knows the behavior of various subsystems can use time tagging to detect problems. For example, assume a sensor usually exhibits certain behavior in normal operation but fails by freezing values. Using time tags, the destination host can compare when data was generated and the corresponding value reported. This can be used to determine whether the sensor is operating properly.

3.12.3 Sampling Reference: In some systems, it is important that certain events, being sampled to generate data, be sampled simultaneously (within a certain "window"). The GTR allows the system designer to load and maintain a common "global" clock in each BIU. This clock is then available to all applications running in the host for the purpose of generating a time "tick" for sampling. The accuracy of this system is dependent on the update rate selected by the system designer and the latency from the GTR on the BIU to the host (application) level.

3.12.4 Timer Accuracy Considerations: A number of factors contribute to the absolute error of the clock time and the skew between all clocks in the system. These factors include:

a. Long and short term clock frequency drift
b. Master clock read time
c. Latch time of slave clock
d. Transmission time
e. Clock update periodicity

3.12.4 (Continued):

Consider a drift rate as depicted in Figure 3.12.4-1. If our oscillator had a drift rate of plus or minus one-half unit of time per unit of time, the value of our clock could be anywhere within the shaded area. If we wanted an accurate time reference, we could be in trouble. The difference between the master and any slave would be as shown in Figure 3.12.4-2 if the update rate were to be 2 milliseconds and a drift rate of minus one-half unit per unit time. A higher update rate would allow us to control the instantaneous value of the clock to a much smaller error. Remember that long term aging of the oscillator should also be taken into account when looking at system needs.

The reading of the master clock and the amount of time consumed by the queing and media access, followed by the reception, checking, and latching of data into the slave clocks has been set at a worst case time of 6 microseconds. This time was derived as shown in Figure 3.12.4-3 and detailed in Tables 3.12.4-1 and 3.12.4-2. Propagation or transmission time is computed (worst case) to be the time between the master and the most distant slave. This number is based strictly on the physical separation of the two stations. It is obvious that the system designer who is concerned with a high level of GTR synchronization should select the master station such that the physical distance to all other stations (or to those stations in which time synchronization is critical) in the network is as consistent as possible. This helps minimize skews between the clocks.

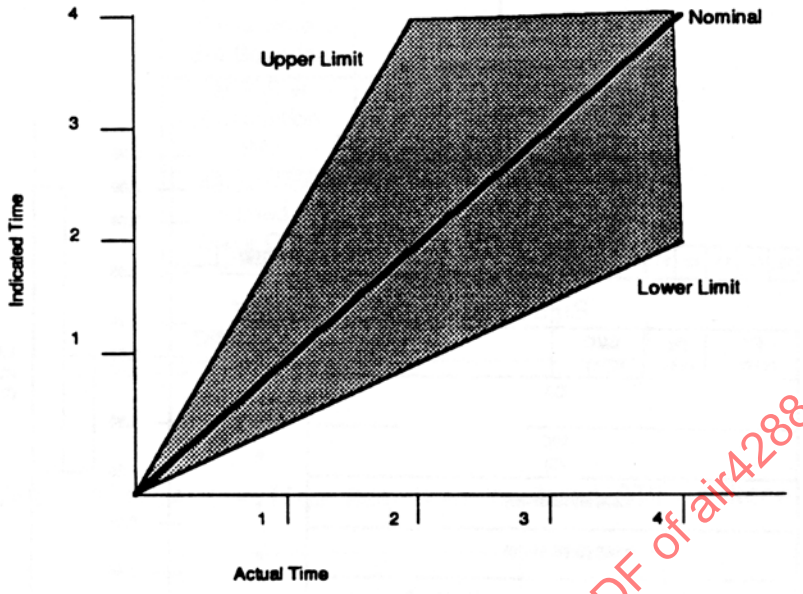


FIGURE 3.12.4-1 - Clock Frequency Drift

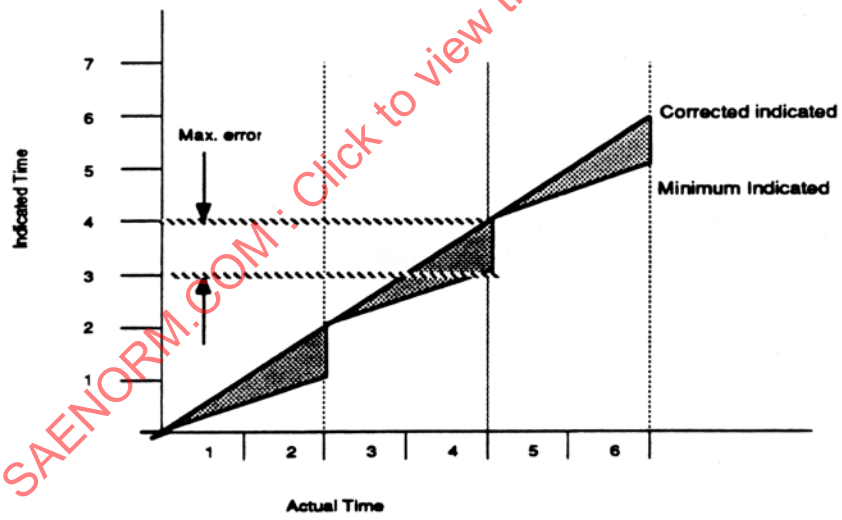


FIGURE 3.12.4-2 - Clock Frequency Drift for all Clocks

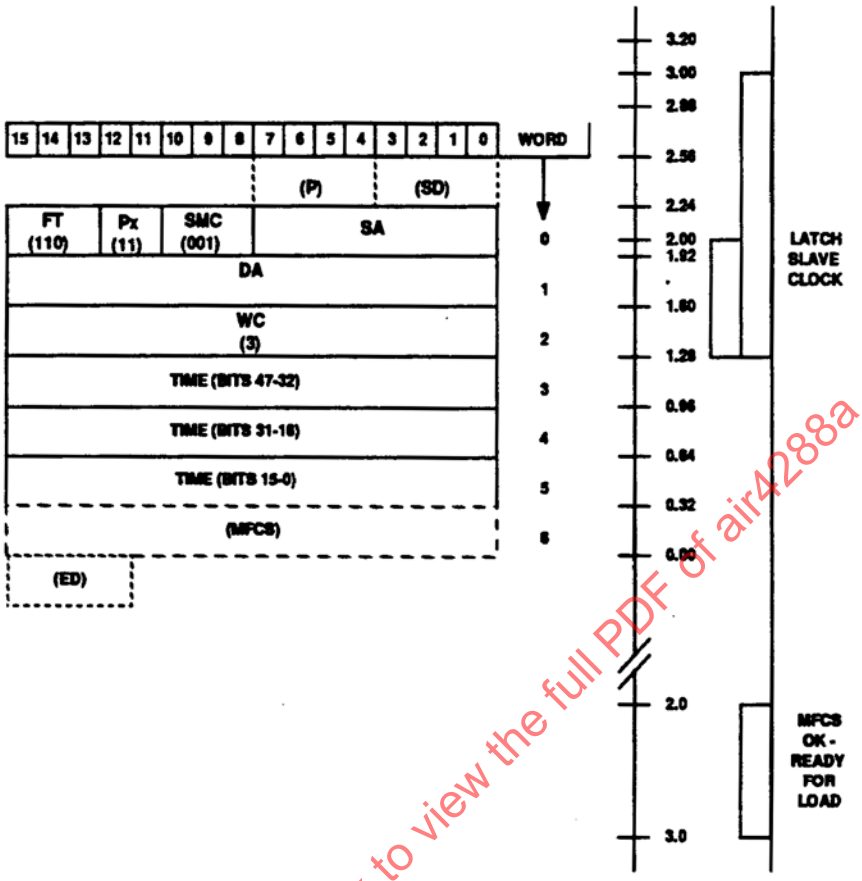


FIGURE 3.12.4-3 - System Related Delays Encountered in Clock Update at Time Master and Time Slaves

TABLE 3.12.4-1 - Time Required to Load GTR at Transmitter

Function	Actual Time	Indicated Time
Latch 48-Bit Value word at 3-4 Boundary	0	0+/- 0.5
MFCS at Destination BIU	1.28	0+/- 0.5
48-Bit Word Latched in Slave Clock	1.28 - 2.28	0+/- 0.5
Time Difference	0.78 to 2.78 µS	

TABLE 3.12.4-2 - Time Required to Load GTR at Receiver

Function	Actual Time	Indicated Time
Latch 48-Bit Value at word 1-2 Boundary	0	0 +/- 0.5
MFCS/ED on Bus	1.92	0 +/- 0.5
MFCS/ED at Destination BIU	2.22	0 +/- 0.5
Time at which MFCS is Validated	3.22	0 +/- 0.5
48-Bit Word Latched into Slave Clock	3.54 to 4.54	0 +/- 0.5
Time Difference	3.04 to 5.04 µS	

3.12.5 System Issues: The designer has control over the overall clock skew in the system by being able to specify the update rate and the location of the time master relative to the time slave units. The standard specifies an equation which can be used by the system designer to optimize the GTR for his application. That equation is:

$$\text{Global Clock Difference (GCD)} < 6 + T_{ms} + (1000 * T_u * T_d * 2 \text{ microseconds})$$

As can be seen from the formula, the 6 microsecond figure developed in Figure 3.12.4-3 and considered in Tables 3.12.4-1 and 3.12.4-2 represents the maximum difference between the actual time and the time reported in the TSM (3 microseconds) added to the maximum amount of time required to load and use the value at the slave (3 microseconds). The propagation time provides another additive error. Finally the last factor combines the drift rate for the 1 microsecond reference clock with the update rate and is listed in Figure 3.12.5-1 for various drift and update rates.

Update Period	Errors Due to Drift			
	$\pm 10^{-3}$	$\pm 10^{-4}$	$\pm 10^{-5}$	$\pm 10^{-6}$
2S	4mS	0.4mS	40µS	4µS
1S	2mS	0.2mS	20µS	2µS
500mS	1mS	0.1mS	10µS	1µS
250mS	0.5mS	50µS	5µS	0.5µS
125mS	0.25mS	25µS	2.5µS	0.25µS
50mS	0.1mS	10µS	1µS	0.1µS
25mS	50µS	5µS	0.5µS	50nS

Shaded area shows how a reasonable clock error can be maintained either by use of a more accurate oscillator or a more frequent update period.

FIGURE 3.12.5-1 - Error Rate Versus Update Rate

3.12.5 (Continued):

The system designer, therefore, on specification of the maximum GCD limit can specify the Time Master update rate for each unit and, to a lesser extent the propagation time by selection of a location of the time master with respect to the slaves.

A value judgement will be made trading off clock drift maximum limits allowed against update rates. The system designer has some freedom to specify the limits of the various determining factors of the GCD and can do so to achieve the most economical hardware.

3.13 Software Design Considerations:

Software design, without trying to create a standard high level protocol, is almost completely dependent on the specific interfaces and overall system architecture used, and is above the level intended by this handbook. There are, however, some standard software interfaces that must be implemented to allow the LTPB to operate. These include:

a. LTPB interface initialization
b. LTPB message filter functions
c. Message input/output requirements
d. LTPB status monitoring
e. LTPB system control

Generic requirements for each of these five major functions are presented in the following paragraphs.

3.13.1 LTPB Interface Initialization: For the most part, the LTPB interface itself is self-initializing, and most of the initialization requirements come from the queueing mechanisms and message screening functions implemented alongside the LTPB interface. The only requirements imposed by the LTPB interface itself include the assignment of static (program pin) configuration data and optional loading of programmable timer values. The following programmable functions must be handled by the LTPB interface:

a. LTPB Physical Station Address - the LTPB Station Address is typically hardwired "program pin" data and requires no loading from software. It is recommended that the software be capable of reading and verifying this data for correctness.

b. LTPB Maximum Number of Stations (MNS) - the MNS is implemented to limit the highest number station on a particular system implementation so that ring admittance and deletion functions do not require token passing to large numbers of unused stations on smaller systems. The MNS is typically "program pin" data and requires no loading from software. It is recommended that the software be capable of reading and verifying this data for correctness.

c. Token Passing Timer - the appropriate value of the token passing timer is dependent on the physical characteristics of the bus and is computed by the system designer (paragraph 3.9.2). This value is a user programmable 8 bit value.

3.13.1 (Continued):

d. Bus Activity Timer - the appropriate value of the bus activity timer is dependent on the physical characteristics of the bus and the hardwired source address of the station and is computed by the system designer (paragraph 3.9.3). The BAT is a user programmable 11 bit value.

e. Ring Admittance Timer - the ring admittance timer should allow full programmability by the software interface under control of the system designer (3.9.1). The RAT is a user programmable 16 bit value.

f. Token Rotation Timers - the token rotation timers should allow full programmability by the software interface under control of the system designer (3.11.3). The three TRTs are user programmable 16 bit values.

g. Token Holding Timer - the token holding timer should allow full programmability by the software interface under control of the system designer (3.11.3). The THT is a user programmable 16 bit value.

h. Global Time Reference (GTR) - the GTR (3.12) should allow full programmability by the software interface. Additionally, programmable functions to control the GTR "Time Master Mode", update rate and message priority must be provided. The system must designate one node as "Time Master" by placing it in the Time Master Mode (through a command register provided by the BIU or other application dependant means). For the designated Time Master station, the host must also be able to set both the priority and frequency at which the Time Synchronization messages are sent. The priority may be any of the four priority levels provided. The update rate is limited to the following values: 8, 16, 33, 66, 131, 262, 524, and 1049 milliseconds, approximately (based on incrementing bits 13 through 20, respectively, of the 48 bit GTR).

i. Preamble Length - a preamble length factor "Sp" is provided for the user to program in the desired number of preamble symbols required by the implementation. This is typically "program pin" data.

3.13.2 LTPB Message Filter Function: The LTPB interface must have the capability to screen logically addressed messages under control of an external source or the host system so that the host only gets the messages required. Message filter functions may be simple one-bit filters or address translation maps, and may be provided for both transmit and receive. In order to support the broadcast mode of message transmission, it is required that location 7FFF (hex) in the message filter be permanently enabled. Detailed description of these functions is application specific and is beyond the scope of this handbook.

3.13.3 Message Input/Output Requirements: The host system must define the requirements for transfer of data to, and reception of data from the LTPB interface, as needed by the specific application involved. The format for the information transfer is application specific but should contain the following information:

a. Start of data
b. Destination address
c. Priority level
d. Word count
e. End of data
f. Other specialized information

It should be noted that the responsibility for insuring successful message transmission and reception across the bus belongs to the host and particular application, since the media access protocol itself provides no means for message acknowledge. Message acknowledgment services, if required, must be provided by the host software.

3.13.4 LTPB Status Monitoring: The LTPB software interface should provide a supervisory processor function to the LTPB as required by the specific application to continuously monitor the health of the LTPB interface. This supervisor interface should establish a certain level of control over the LTPB interface to assist in the rapid detection and recovery from errors. Typical information provided includes:

a. LTPB current status
b. Error counts
c. Bus traffic counts

The specification and format of this information is application specific and is beyond the scope of this handbook.

3.13.5 LTPB System Control: The use of the token passing distributed access control of the LTPB requires a radical change in the way systems are designed. This change reflects the increasing autonomy of subsystems and will result in more efficient systems since data will only be transferred when necessary, rather than periodically. In spite of these advantages, situations still exist where the use of centralized control is preferred; whether for compatibility with existing system methods or simply the warm feeling of one subsystem in control. While it is preferable that the system be designed to use the distributed control in the way intended, such centralized control can be implemented using the LTPB.

One way is to implement the centralized command/response protocol on top of the token passing. Tokens are passed round the logical ring in the normal way, but only one station may unilaterally request a message transmission. These messages are used to initiate data transfers from other stations. The other stations must wait for these messages before requesting a message to be transmitted. The transmission requests are then handled as usual when the station receives the token.

3.13.5 (Continued):

During one token hold, the controlling station may issue a number of commands to different stations before releasing the token and the remote stations transmitting their messages on subsequent token round(s). Status responses may be provided with each transmission or, since the stations continually receive tokens, could be issued periodically without the need for the controlling station command.

Although the remote stations receive many tokens, no messages are transmitted until requested by the controller, so that, as far as the system is concerned, messages are only transmitted from remote stations when commanded.

3.14 Data Security Issues:

3.14.1 General Considerations: Within an aircraft weapon system, signals fall into two levels of classification, unclassified (black) and classified (red) signals. Hence, it is necessary for the system designer to provide the appropriate protection for these signals. Red signals require a high degree of EMI, EMC, and TEMPEST protection while black signals require a lesser degree of protection. The protection of signals within a weapon system can vary between different areas of the avionics suite. Guidelines within DOD and the National Security Agency (NSA) are in development.

3.14.2 Protection on the LTPB: The LTPB will be required to provide protection for classified data. Some designers are considering multilevels of protection while others are providing only two levels of protection - unclassified (Black Data) and classified (Red Data). Three approaches being investigated to date are:

a. Physical Isolation
b. Encryption
c. Time Share

In physical isolation approach, all black data is resident on one bus while the red data is resident on a separate bus. The designer must assure that proper Red/Black separation techniques are employed. With encryption, the red data is uniquely processed prior to distribution on the bus. In a time sharing system, the bus handles both red and black data. When the red data is being distributed, the system allows only subsystems with the proper access on the bus. This technique will need to be demonstrated with a high degree of probability that the red data can be protected properly. The policy organization at NSA should be consulted on each of the above mentioned approaches prior to actual implementation.

3.14.3 Encryption Tradeoffs: If encryption is selected as part of the protection process, a decision must be made as to where the encryption is done. It can be done in the host or in the BIU. Placing the encryption in the BIU leads to a more complex terminal but aids commonality. If encryption is done in the host, the potential for many unique coders exists.

3.14.4 Cautions: In utilizing the LTPB with data requiring protection, there are several known pitfalls or cautions worth noting. Guidelines are not always consistent and vary within DOD. Also, some protocol systems employ a tag for identifying secure data. Care must be taken when receiving data using this technique. The error detection code should be checked before sending data to the subsystem.

3.15 Test Concepts:

Test of the BIU utilizes some concepts familiar from MIL-STD-1553B testing. However, it is obvious from a review of the protocol operation that it also presents some challenges not previously encountered in avionics system test and integration. There are three primary areas which require test and analysis in order to provide a properly operating system. These areas are:

a. Media (coax or fiber optic) Verification
b. Protocol verification
c. Host/BIU interface verification

Each of these areas require different treatment in test planning for the various stages of system development (e.g., BIU test, box or module level test, and system level test). The protocol, however, has been designed to support testing of all interface functions at any level of integration by means of Station Management functions which may be accessed either across the bus or across the Host/BIU Interface. The discussion which follows is primarily a test concept which results from task group discussions on the topic of AS4074 BIU testing - that is, how do we make sure what we designed can be tested?

3.15.1 Test of Media Interface: Test of the individual station media interface is addressed in the AS4074 standard document. The body of the document calls out specific parameters for both the coax and fiber optic implementations of the BIU. Actual values are specified in a "slash sheet" arrangement, found in an appendix to the standard. Table 3.15.1-1 details a sample slash sheet for a fiber optic implementation of the LTPB. Table 3.15.1-2 provides parameters for implementation of a coax system. The slash sheets currently a part of the standard are representative implementations and are not proposed as the only implementations allowable under the standard. As new systems come on-line, it is envisioned that the slash sheet arrangement will allow those systems to take advantage of a common protocol, yet operate at different data rates, etc by appending a new slash sheet.

Note that each slash sheet, in conjunction with the body of the text, explicitly specifies the tests which must be accomplished as well as the test value boundaries. Values for the transmitter and receiver of the BIU interface offer not only quantitative values for test and verification of the interface, but also for test and verification of the media interconnection for the final system installation.

TABLE 3.15.1-1 - Fiber Optic Slash Sheet from AS4074

AS4074.1 SAE Page 72

TYPE 1-2 FIBER OPTIC MEDIA INTERFACE CHARACTERISTICS

Parameter	Description	Units	Requirement
COMMON CHARACTERISTICS			
-	Encoding Method		Manchester II
R_d	Data Rate	MBPS	50 +/- 0.01%
R_s	Signaling Rate	MBaud	100 +/- 0.01%
T_o	Nominal Bit Time	nS	20
T_m	Minimum Duration Between Transitions	nS	10
S_{ite}	System Minimum Intertransmission Gap	nS	280
S_p	Preamble Minimum Size	Bit Times	16
-	Transmit Optical Connector		TBD
-	Receive Optical Connector		TBD
W1	Optical Wavelength Lower	nM	800
W2	Optical Wavelength Upper	nM	860
BW	Spectral Bandwidth	nM	60
TNT	Token Holding Timer Initializ. Value	Bit Times	800
TRT_1	Token Rot. Timer 1 Initializ. Value	Bit Times	1280
TRT_2	Token Rot. Timer 2 Initializ. Value	Bit Times	640
TRT_3	Token Rot. Timer 3 Initializ. Value	Bit Times	320
BAT	Ring Admit. Timer Initializ. Value	Bit Times	$1.6*10^6$
TRANSMITTER CHARACTERISTICS			
T_{po}	Transmitter Optical Power (Signal High)	dBm	-2.0 +/- 2.0
T_{pr}	Transmitter Residual Power (Signal Low)	dBm	-15
T_{pl}	Transmitter Leakage Power (Tx Off)	dBm	-40
T_r	Transmitter Maximum Rise time	nS	4
T_f	Transmitter Maximum Fall Time	nS	6
T_{dpw}	Transmitter Maximum Pulse Width Distortion	nS	2.4

Page 73 SAE AS4074.1

TYPE 1-2 FIBER OPTIC MEDIA INTERFACE CHARACTERISTICS (Continued)

Parameter	Description	Units	Requirement
T_{cos}	Transmitter Combined Over/Under-Shoot	%	5
T_{to}	Transmitter Nominal Bit Time	nS	20 +/- 10%
T_{ts}	Transmitter minimum signaling duration	nS	10 +/- 5%
T_{dt}	Data Streaming Timer	uS	1640 +/- 10%
RECEIVER CHARACTERISTICS			
R_{po}	Receiver Maximum Optical Power Input	dBm	0
R_{pr}	Receiver Operating Range	dB	21
R_{idr}	Receiver Intertransmission Dynamic Range	dB	15
R_{pn}	Receiver Minimum Optical Power Input	dBm	-31.5
R_{pos}	Receiver Combined Over/Under Shoot	%	5
R_r	Receiver Input Maximum Rise Time	nS	5
R_f	Receiver Input Maximum Fall Time	nS	7
R_{dpw}	Receiver Input Maximum Pulse Width Distortion	nS	2.4
R_{to}	Receiver Nominal Bit Time	nS	20 +/- 10%
R_{ts}	Receiver Maximum Signaling Rate	nS	10 +/- 10%
R_{ber}	Receiver Maximum Bit Error Rate	-	$10^{exp}(-10)$
MEDIA CHARACTERISTICS			
B_{core}	Fiber Core Diameter	µM	200
B_{clad}	Fiber Cladding Diameter	µM	240
NA	Fiber Minimum Numerical Aperture	-	0.2
B_{to}	Maximum Total Dispersion	nS	3
A_{min}	Minimum Optical Attenuation	dB	11.5
A_{max}	Maximum Optical Attenuation	dB	28.5

TABLE 3.15.1-2 - Electrical Media Slash Sheet from AS4074

AS4074.1 SAE Page 74

TYPE 6-1 ELECTRICAL MEDIA INTERFACE CHARACTERISTICS

Parameter	Description	Units	Requirement
COMMON CHARACTERISTICS			
-	Encoding Method		Manchester II
R_d	Data Rate	MBPS	50 +/- 0.01%
R_s	Signaling Rate	Mband	100 +/- 0.01%
T_b	Nominal Bit Time	nS	20
T_d	Minimum Duration Between Transitions	nS	10
$Sitg$	System Minimum Intertransmission Gap	nS	200
S_p	Preamble Minimum Size	Bit Times	4
S_v	Surge Voltage Level	V	+/- 1000
S_d	Surge Voltage Duration	nS	10
-	Transmit/Receive Connectors		TBD
TRT	Token Holding Timer Initialize. Value	Bit Times	800
TRT_1	Token Rot. Timer 1 Initialize. Value	Bit Times	1200
TRT_2	Token Rot. Timer 2 Initialize. Value	Bit Times	640
TRT_3	Token Rot. Timer 3 Initialize. Value	Bit Times	320
RAT	Ring Admit. Timer Initialize. Value	Bit Times	$1.6*10^6$
TRANSMITTER CHARACTERISTICS			
T_{vo}	Transmitter Output Voltage Level (On)	V	+/-(7.5+/-20%)
T_{vr}	Transmitter Peak Output Voltage Level (Off)	V	+/- 0.05
T_r	Transmitter Maximum Rise Time	nS	4 +/- 50%
T_f	Transmitter Maximum Fall Time	nS	4 +/- 50%
T_{cmt}	Transmitter Combined Over/Under-Shoot	%	5
T_{to}	Transmitter Nominal Bit Time	nS	20 +/- 10%
T_{tm}	Transmitter minimum signaling duration	nS	10 +/- 5%

Page 75 SAE AS4074.1

TYPE 6-1 ELECTRICAL MEDIA INTERFACE CHARACTERISTICS (Continued)

Parameter	Description	Units	Requirement
T_{12}	Transmitter Drive Impedance	Ohms	50 +/- 20%
T_{dt}	Data Streaming Timer	µS	1640 +/- 10%
RECEIVER CHARACTERISTICS			
R_{vo}	Receiver Maximum Input Voltage	V	+/- 12
R_{vm}	Receiver Minimum Input Voltage	V	+/- 0.1
R_{dr}	Receiver Dynamic Range	dB	26
R_{cmt}	Receiver Combined Over/Under-Shoot	%	10
R_r	Receiver Input Maximum Rise Time	nS	10
R_f	Receiver Input Maximum Fall Time	nS	10
R_{to}	Receiver Nominal Bit Time	nS	20 +/- 10%
R_{tm}	Receiver Maximum Signaling Rate	nS	10 +/- 10%
R_{12}	Receiver Input Impedance	Ohms	50 +/- 20%
R_{cmr}	Receiver Common Mode Rejection Ratio		0.5
R_{ber}	Receiver Bit Error Rate	-	10exp(-12)
R_{nv}	Receiver Noise Voltage Input	mV	2.5
MEDIA CHARACTERISTICS			
A_r	Attenuation of Reflected Signals	dB	34
A_{max}	Maximum End-to-End Attenuation	dB	40
A_{min}	Minimum End-to-End Attenuation	dB	14
$m1$	Transmission System Lower Bandpass Freq.	MHz	0.5
$m2$	Transmission System Upper Bandpass Freq.	MHz	100
D_{pp}	Group Propagation Delay Difference	nS	1
D_j	Jitter	nS	1

3.15.2 Test and Verification of Protocol: The distributed nature of the bus access method of this bus complicates overall test and verification of operation. During the design of the proposed standard one of the key goals was to provide "hooks" to allow the user to implement testing of all functions through use of a "test" interface attached at a single point on the bus. The result was a set of Station Management functions which, while useful in the management of an operating bus system, can also be used successfully in the performance of a complete interface test at both the BIU and system levels. Detailed test and validation procedures may be found in the AS 4290 Validation Test Plan for AS4074 LTPB.

At the BIU level, test could be accomplished by utilizing an automatic test set which simulates two or more "stations" on the bus (Figure 3.15.1.2-1). This test set would contain an appropriate Host/BIU interface as well as a fiber optic or electrical connection to the transmitter and receiver sections of the BIU. The unit would be capable of testing the ability of an interface to perform protocol activities in the areas of initialization, logical path buildup, fault recovery, ring admittance. Test of receiver and transmitter functions, message queuing, message framing could be performed by use of the station management function of "Test Message Wrap-around". As in MIL-STD-1553B, this command causes the station to accept a specially coded message from the bus and, upon the next receipt of the token, return the message, with the same data, to the originating station. A wrap-around test is also provided for the Host/BIU interface. All of the timers can also be programmed and verified either from the Host/BIU interface or the bus, as can BIU status. Tests of timer operation can be accomplished by setting the timers to certain determined values, loading the transmit queues with known message traffic and causing predetermined traffic conditions on the bus utilizing the automated test-set.

System level test concepts center on the use of an automated test set connected to the bus media (Figure 3.15.1.2-2). This test set, utilizing the station management functions could perform tests of the individual station and overall system functions as detailed in Table 3.15.2-1.

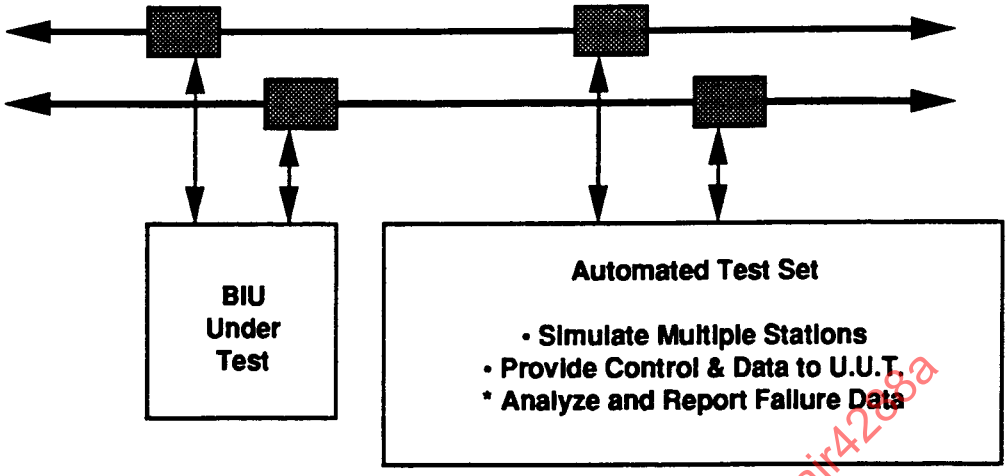


FIGURE 3.15.2-1 - BIU Automatic Test-Set Concept

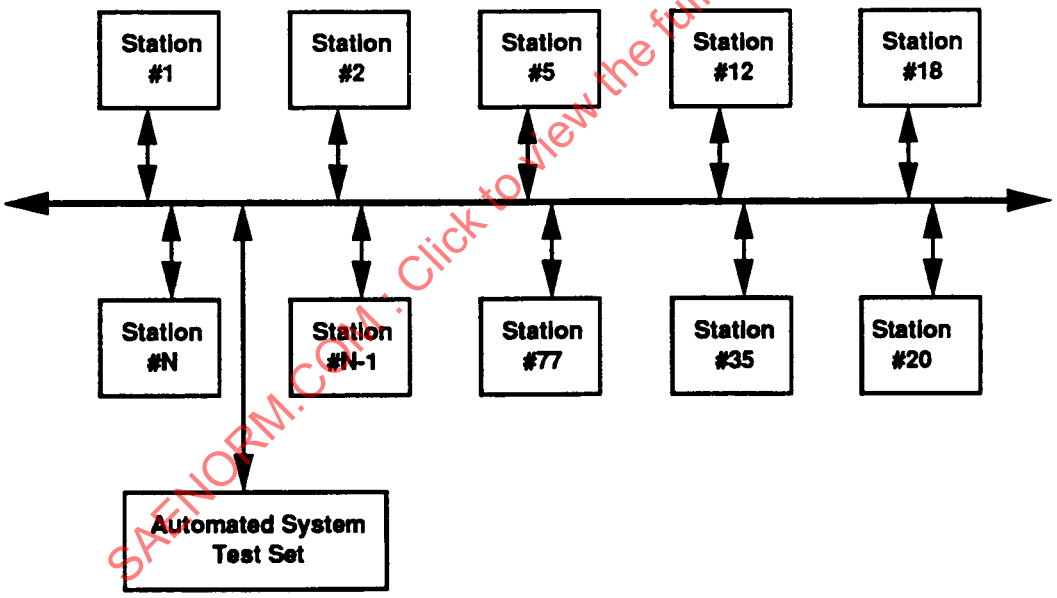


FIGURE 3.15.2-2 - System Test Concept

TABLE 3.15.2-1 - System Test Using Station Management Functions

TABLE 3 - SYSTEM TEST USING STATION MANAGEMENT COMMANDS	
COMMAND	FUNCTIONS TESTED
1. RESET	ALLOWS USER TO RESET STATION AND VERIFY THAT ALL INTERNAL VALUES HAVE BEEN RESET TO DEFAULT VALUES AND STATION HAS DISABLED ITSELF.
2. ENABLE	PERMITS CONTROL OF THE BIU TO ALLOW TESTER TO TEST STATION ADMITTANCE FUNCTION, BUS ACTIVITY SENSING FUNCTION, STATION INITIALIZATION FUNCTION.
3. DISABLE	PERMITS CONTROL OF BIU TO ALLOW TESTER TO TEST FAILED STATION BRIDGING FUNCTIONABILITY TO CONTROL BIU WHEN DISABLED ("LISTEN" FUNCTION).
4. REPORT STATISTICS	PERMITS TESTER TO RETRIEVE DATA ON BIU PERCEIVED BUS FUNCTIONS, SUCH AS TOKENS/MESSAGES SEEN, TOKENS/MESSAGES SENT, FAILED TOKENS/MESSAGES. COULD BE USED IN A COMPARISON BETWEEN STATIONS TO DETECT WEAK TRANSMITTER, IMPROPER RECEIVER SENSITIVITY OR INTERMITTANT CIRCUIT OPERATION.
5. LOAD/REPORT CONFIG.	PERMITS TESTER TO VARY VALUES IN PROGRAMMABLE FUNCTIONS FOR TEST PURPOSES, AND TO VERIFY THAT BIU POWERS UP WITH PROPER DEFAULT VALUES
6. TEST MESSAGE WRAP	PERMITS TESTER TO SEND TEST MESSAGE TO BIU UNDER TEST AND RECEIVE A REPLAY OF THE MESSAGE TO VERIFY STATION DATA PATHS. IDENTICAL TEST EXISTS FOR HOST/BIU INTERFACE
7. TIME REFERENCE COMMANDS	PERMITS TESTER TO PERFORM SETTING AND VERIFICATION OF THE 48 BIT REAL TIME CLOCK IN THE BIU ACROSS BOTH THE BUS AND THE HOST/BIU INTERFACE

3.15.2 (Continued):

Several functions supported by the Station Management command set are noteworthy. The Enable and Disable commands allow the test person to alternately enable and/or disable the transmit function of the BIU. This makes it relatively easy to verify the functionality of the fault recovery mechanism in a system and then to provide a check of the Ring Admittance function. By forcing a station off of the bus (creating a "failed" station), the predecessor station is forced to bridge the bad station. By then enabling it, one can watch for the ring admittance opportunity (a predecessor function) and verify that the "failed" station rejoins the bus by accepting the token (proper behavior for a new station).

Another useful tool is the ability to set the various programmable functions to user determined values. This is accomplished with the Load/Report Config(uration) command. The test person could, utilizing this command, set the THT, TRTs, RAT, and other timers to certain values in order to force deferral of message traffic, increase the frequency of Ring Admittance, or verify that the station powers up to the appropriate default values.

A final test tool provided in the Station Management Command set is useful, not only during system level test, but also during the integration of a new system. Statistics on all bus activity, as perceived by a particular station, may be retrieved from each BIU by use of the "Report Statistics" command. This command gives the user access to information relative to the number of valid messages seen on the bus by the BIU, valid messages transmitted by this BIU, and the number of aborted frames caused by or validity errors seen by the BIU. By comparing data from all stations in a system, an intermittent BIU could be isolated.

4. SAMPLE SYSTEM DESIGN APPROACH:

There are many equally viable approaches to system design. The purpose of this section is to introduce the system designer to the characteristics of the LTPB protocol and to demonstrate approaches to determining the appropriate values for the station timers. This should allow the system designer to adjust his or her approach to support designs utilizing the LTPB protocol. Three approaches are presented in this section. The first approach begins with a detailed development on the relationship of the timers. Equations are developed which can be used to set and verify the timers. A step-by-step design example of a simple system is included. The second method is closely related to the discussion in 3.11.3.2 and introduces a BASIC analysis program in Appendix C. The final approach is based on the determination of classes of methods and develops a set of equations which represents time to transmit messages at selected priorities at selected stations. It is suggested that the system designer review the three approaches and determine the best one for his application.

4.1 System Design Approach I:

The LTPB protocol is a token passing protocol. That is, a station in the system gains the right to use the media for transmission of messages when it receives the token. When a predetermined set of circumstances no longer exists, the station may no longer send messages and must pass the token to a successor station on the logical ring.

As a part of this predetermination of how long a station may use the token, each station has a set of associated timers:

a. One token holding timer (THT)
b. Three token rotation timers (TRT)

The THT is a local timer set to a value, THT[max], as soon as the station receives the token. A station is allowed to send messages of any priority until such time as the THT becomes zero. Hence, THT[max] is the maximum amount of time that any one station may hold the token. If many P0 messages are pending, THT[max] acts as a bound for transmission of P0 messages, since they are sent first.

The TRTs are timers which are used to authorize and bound the transmission of lower priority messages. At the station level, lower priority message transmissions are still bounded by the THT. At the system level, however, time must be left for succeeding stations to transmit their highest priority messages. The TRTs determine the amount of remaining bandwidth available on the LTPB for transmission of lower priority message traffic before the token must be passed to enable the transmission of higher priority message traffic. There is one TRT associated with each of the three lower priority message categories (TRT1, TRT2, TRT3) in each station.

A TRT is reset to its maximum value whenever the station begins to process messages at that priority level. That is TRT_j is set to TRT_j[max] when a station begins to handle priority j messages. The station is allowed to sent priority j messages if, just prior to TRT_j reinitialization, the residual value of THT and TRT_j were not equal to zero. Since TRT_j has been running since the last time the station used the token to send messages at this priority level, this timer has, in effect, measured the amount of bus traffic, by measuring the time taken by the token to return, and this is used to determine the available time for transmission of messages of this priority on the current token hold.

4.1 (Continued):

From a system point of view, if each TRT in N stations on a LTPB is set to a different maximum value there would be 3N+1 levels of priority available (i.e., for a 128 station network, there would be a possibility of 385 different levels of priority). More discussion on this method of TRT/THT setting is discussed in 4.8. More than likely, however, the system designer will opt to utilize a smaller number of message priorities by setting the TRTs in all stations to the same maximum values. Note that the THT, set to a specific value for each and every station bounds the time that any one station may hold the token. The maximum amount of time around the logical ring may be bounded by the relationship:

$$\sum_{i=o}^{N-1} (THT_i[max]) + T \qquad [s]$$

where: T represents the time taken for a token to circulate the logical ring in the absence of messages (empty token round)

The proper operation of the priority scheme is guaranteed by setting the TRTs to the following relationship:

$$TRT1[max] \geq TRT2[max] \geq TRT3[max]$$

4.2 Data Transfer Types:

The system designer will recognize two different types of data transfers which will occur on the LTPB:

a. Periodic data transfers
b. Aperiodic data transfer

4.2.1 Periodic Data Transfers: Periodic data transfers are characterized by a given data generation rate (or update rate) which remains constant. This data transfer may also have a maximum and minimum allowable latency defined, normally less than the period of data generation.

4.2.2 Aperiodic Data Transfers: Aperiodic data transfers are characterized by a transmission probability density function and a maximum allowable latency. The transmission probability density function is analogous to the data generation rate of the periodic data transfer. The main difference between the two is the theoretical nature of aperiodic transfers (i.e., no definite figure can easily be determined).

4.3 Analysis of Message Traffic:

For purposes of discussion, the following classes of messages are defined:

- $m_{p,i}(F,L_j)$: This identifies periodic data transfers of frequency F and Latency L_j.
- $m_{a,i}(P_d,L_j)$: This identifies aperiodic data transfers with a transmission probability density P_d and a latency L_j.

4.3 (Continued):

A data flow for each station i may be computed (for periodic data transfers) and estimated (for aperiodic transfers). This can be expressed as:

$$\phi_i = \phi_{p,i} + \phi_{a,i} \qquad \text{[bit/s]}$$

where:

$\phi_{p,i}$ is the periodic message data rate sent by station i
$\phi_{a,i}$ is the aperiodic message data rate sent by station i

The equation which governs mixed message transfers is expressed as:

$$\phi_{i,j} = \phi_{p,ij} + \phi_{a,ij}$$

or

$$\phi_i = \sum_{j=0}^{3} (\phi_{p,ij}) + \sum_{\varphi=0}^{3} (\phi_{a,ij}) \qquad \text{[bit/sec]}$$

where:

$\phi_{p,ij}$ is the periodic message data rate sent by station i with priority j
$\phi_{a,ij}$ is the aperiodic message data rate sent by station i with priority j

When viewed from the standpoint of transmission time, the following values may be defined for station i:

d_i = transmission time per token round for station i
d_{ij} = priority j transmission time per token round for station i

Generally, only the data rate is specified. Nevertheless, data rate and transmission time per token round are interrelated as:

$$\phi_i = \frac{d_i * R}{TR}$$

where:

R = Protocol throughput
TR = Token Round time

4.3.1 System Level Analysis: The parameters defined in this paragraph are a synthesis of those defined previously.

$$\phi_j = \sum_{i=0}^{N-1} \phi_{ij}$$

where:

ϕ_j is the whole system data rate for priority j

The corresponding transmission times may also be defined:

From the above:

$$D_j = \sum_{i=0}^{N-1} d_{ij} \qquad [s]$$

where:

D_j is the transmission time per token round for all priority j messages

$$D_j = n_j * D_j$$

where:

D_j is the transmission time for all priority j messages, taking n_j token rounds

$$\phi_j = \sum_{i=0}^{N-1} \frac{d_{ij}*R}{TR} = \frac{D_j*R}{TR} \qquad (4.3.1) \qquad [bits/s]$$

$$= \frac{D_j*R}{TR*n_j} \qquad (4.3.2) \qquad [bits/s]$$

4.3.2 THT and TRT Settings:

4.3.2.1 THT Setting: Computation of the value for THT[max] is based on the following relationships:

a. The THT guarantees the minimum amount of time that a station may use the media for transmission of periodic messages and leave enough time for transmission of aperiodic messages.

b. The THT must guarantee the worst case token round (i.e., limit the amount of time a station may use the medium).

4.3.2.1 (Continued):

The token round time is expressed as:

$$TR = \sum_{i=0}^{N-1} (d_i) + T$$

where:

$\sum_{i=0}^{N-1} (d_i)$ is the transmission time for all stations and T is the time taken for an empty token round

Due to the message transmission algorithm, a message may be sent provided the instantaneous value of the THT has not reached zero when the message is presented for transmission. It may happen that the THT is close to zero, but has not reached it, such that the station may overrun the value of THT[max] by the transmission time for the maximum length message, since the message transmission will be completed prior to passing the token.

The maximum transmission time for a given station is thus:

$$THT_i[max] + M_i$$

where:

M_i is the maximum message length (time) sent by station i

If M is the duration of the system level maximum length message, and this value is the same for all stations, this results in the expression:

$$d_i \leq THT_i[max] + M_i$$

So that the maximum time of one token round is given by:

$$TR[max] = \sum_{i=0}^{N-1} (THT_i[max]) + (N*M) + T \qquad (4.3.3)$$

The value given to T is the value under the normal token passing conditions. The failure modes, which increase the time taken to pass the token per round, must be included in the remaining system bandwidth as a safety margin.

$$n_0 * TR[max] \leq L_0 \qquad (4.3.4) \quad [s]$$

4.3.2.1 (Continued):

Low latency is achieved in a worst case system if the following relationship is used:

where:

n_0 is the number of token rounds used for transmitting all priority 0 messages
L_0 is the maximum latency for priority 0 messages

The maximum length token round (Equation 4.3.3) now must be computed. This involves determining values for M, n_0, and the summation of the maximum token holding times for all stations.

During the period equal to L_0 (n_0 * TR[max]) all messages of priority 0 must be sent.

If D_0 is the transmission time for all priority 0 messages, the worst case can be expressed similarly to equation 4.3.4:

To guarantee the message latency, the amount of message data to be transmitted within the latency delay L_0 must be limited.

$$D_0 + n_0 * T \leq L_0$$

If the expression above does not meet system requirements, another type of bus should be investigated.

Rearranging this, the maximum number of token rounds is:

$$n_0 = \frac{L_0 - D_0}{T} \qquad \text{[no units]}$$

Substituting this and equation 4.3.3 in equation 4.3.4:

$$\frac{L_0 - D_0}{T}\left(\sum_{i=0}^{N-1} THT_i[max] + (N*M) + T\right) \leq L_0$$

Which allows the solution for M and the summation of the maximum THTs in each station.

The next step consists of dividing up the guaranteed bandwidth among all stations in the system to determine the THT[max] value for each.

4.3.2.2 Setting of the TRT Maximum Values: Proper operation of the priority scheme requires the following expression to be true:

$$TRT3[max] \le TRT2[max] \le TRT1[max]$$

In order to guarantee L_0 latency, the following inequality must be true:

$$\sum_{i=0}^{N-1} (THT_i[max]) + (N*M) + T \le L_0 \qquad [s]$$

Also, $TRT1_{[max]}$ must be limited in order to guarantee L_0 latency. Therefore, the following relationship must be enforced:

$$TRT1[max] + M < \sum_{i=0}^{N-1} (THT_i[max]) + (N*M) + T \qquad (4.3.5) \qquad [s]$$

WARNING: This inequality is required, but is not sufficient, to guarantee L_0 In the event of the previous token round being empty. For this situation, it is necessary to use the THT as a transmission time "watchdog" since the instantaneous time remaining on a TRT ("residue") may exceed the value of the instantaneous value of the THT. When the situation occurs that the residue in the TRT exceeds the value in the THT, the value remaining in the THT should be the limiting factor in determining station transmission time.

4.3.2.2.1 TRT1 Computation: The maximum transmission time for priority 1 messages in station i (d_{i1}) is represented by:

$$d_{i1} = THT_i[max] + M - d_{i0}$$

where:

d_{i0} is the priority 0 message transmission time for station i

The maximum transmission time for all stations, during n_1 token rounds is thus:

$$D_1 = n_1 \left(\sum_{i=0}^{N-1} THT_i[max] + M - d_{i0} \right)$$

or:

$$D_1 = n_1 \left(\sum_{i=0}^{N-1} THT_i[max] + (N*M) - D_0 \right) \qquad (4.3.6) \qquad [s]$$

4.3.2.2.1 (Continued):

In order to guarantee the L_1 latency delay, the message transmission will have to be accomplished during a number of maximum length token rounds in the worst case (n_1). Therefore, the following must be true:

$$n_1 * TR[max] \le L_1 \qquad [s]$$

Thus, rearranging and substituting from equation 4.3.3:

$$n_1 <= \frac{L_1}{TR[max]} = \frac{L_1}{\sum_{i=0}^{n-1} (THT_i[max]) + (N*M) + T} \qquad [\text{no units}]$$

In order to guarantee L_1, D_1 must fulfill equation 4.3.6 when n_1 takes its maximum value.

$$D_1 \le n_1 \left(\sum_{i=0}^{N-1} (THT_i[max]) + (N*M) - D_0 \right) \qquad (4.3.7) \qquad [s]$$

Notice that the transmission time for the priority 0 messages enters the equation (D_0). The highest latency transmission time must be subtracted from the token round time. Substituting from equation 4.3.5 into equation 4.3.7:

$$D_1 \le n_1 (TRT1[max] + M - T - D_0) \qquad [s]$$

Rearranging, the minimum value for TRT1[max] can be calculated:

$$TRT1[max] \ge \frac{D_1}{n_1} + D_0 + T - M \qquad (4.3.8) \qquad [s]$$

Substituting from equation 4.3.1 and equation 4.3.2 in equation 4.3.8:

$$TRT1[max] = \frac{\phi_1 + \phi_0}{R} * TR + T - M \qquad [s]$$

Since:

$$TRT1[max] + M \le TR[max] \qquad [s]$$

then:

$$\phi_0 + \phi_1 \le R \qquad [\text{bits/s}]$$

4.3.2.2.2 Other TRT Computations: Generally speaking, the equations developed for TRT1, can be extended to the determination of settings for TRT2 and TRT3. These are derived from the following equations:

$$n_j = \frac{L_j}{TR[max]} \qquad (4.3.9)$$

$$D_j \le n_j \left(TRT_j[max] + M - T - \sum_{q=0}^{j-1} D_q \right) \qquad (4.3.10) \qquad [s]$$

$$TRT_j[max] = \sum_{q=0}^{j} (\phi_q) * \frac{TR[max]}{R} + T - M \qquad (4.3.11) \qquad [s]$$

4.4 Notes on the Equations:

a. In order for the priority scheme to function such that four levels of priority are provided for, and messages throughout the system are deferred, starting with the lowest priority, the following relationship must hold true:

$$TRT_j[max] \le TRT_{j-1}[max]$$

This involves:

$$D_j \le D_{j-1} \qquad [s]$$

but does not imply:

$$\phi_j \ge \phi_{j-1} \qquad [bit/s]$$

b. To prevent potential throughput limitations:

$$\sum_{j=0}^{3} (\phi_j) \le R \qquad [bit/s]$$

In order to provide a predefined safety margin for failure mode and system growth, the following relationship should be observed throughout the computations for system design:

$$\sum_{j=0}^{3} \phi_j = (1 - \alpha_R)R \qquad [bit/s]$$

where:

α_R is the portion of protocol throughput reserved for the safety margin

4.4 (Continued):

c. The latency for priority 3 messages (L_3) is also related to the station addition (ring admittance) function (section 3.3.2). Therefore, the insertion mechanism is handled as an priority 3 message.

4.5 System Requirements:

In the development of any system, the suitability of the candidate approaches must be determined. One of the criteria involved in that determination for a communication network is whether or not the candidate approach can support the required message traffic with the necessary data latencies. The example presented in this section describes an approach to setting the timers which, although nonoptimal, will help the system designer to make that determination.

4.5.1 Network Physical Parameters Determination: The physical characteristics of the planned network have an impact on the determination of the station timers. In this section the characteristics of the example system are presented and the value for T, the time for the token to circulate the logical ring in the absence of messages, is determined. Table 4.5.1-1 lists the physical characteristics of the example system.

TABLE 4.5.1-1 - System Physical Characteristics

Number of Stations	**10**
Distance Between Stations	**100 [m]**
Preamble Size	**16 [bits]**
Bit Rate	**50 [Mbits/sec]**
Propagation Speed	**21.4 [cm/nsec]***

*** Assumes 1.4 index of refraction multimode fiber.**

T represents the time for token rotation, it includes components which account for propagation delay, the time needed to receive the token, and the transmit time of the token. The worst case approximation is given by:

$$T = \{N * \text{Bus Length} / \text{Propagation Speed}\} + N * \text{Station Response Time} + N * \text{Time to Transmit Token [sec]} \quad (4.5.1)$$

where N is the number of stations on the network. The bus length represents the distance from the station's transmitter to the next station's receiver, in a passive star configuration, the Bus Length will be twice the distance of a station to the star. In an active star configuration additional delay may need to be added to allow for the operation of the active star coupler. The times to receive and transmit the token are based on the number of preamble bits, the token length, and the bit time as shown in the following equation.

$$\begin{aligned} \text{Time to Transmit a Token} &= (\text{Preamble Size} + \text{Token Length}) * \text{Bit Time} \\ &= (16\ [\text{bits}] + 24\ [\text{bits}]) * 0.02\ [\mu\text{sec/bit}] \\ &= 0.80\ [\mu\text{sec}] \end{aligned} \quad (4.5.2)$$

4.5.1 (Continued):

The 0.02 μs/bit is based on the 50 Mhz transmission frequency. Now, calculating T for the example system,

$$T = \{10*100[m]/0.214\ [m/ns]\} + 10 * 0.50\ [\mu sec] + 10 * 0.80\ [\mu sec] = 17.7\ [\mu sec]$$

This value for T will be used in the remainder of the example system design.

4.5.2 Data Parameter Determination: In order to proceed with the system design, the bus traffic must be categorized. There are three important parameters for each message: data size, frequency of transmission, and required latency. The data size should be in 16-bit words. The frequency should be specified as both a nominal frequency and a peak frequency. The latency represents the allowed transit delay for the message. This is the time between the end of computation producing the data at the transmitting host and the reception of the data by the receiving host. Table 4.5.2-1 presents a set of hypothetical data which will be used in the remainder of this example. In the table, a set of 10 messages are described. For this example it will be assumed that all stations on the bus have identical message traffic and that there are 10 stations.

TABLE 4.5.2-1 - Example Message Traffic

Message Name	Data Size Words	Frequency (Hz) Nominal	Frequency (Hz) Peak	Latency (ms)
A	20	100	100	10
B	50	50	75	10
C	50	50	50	20
D	150	25	50	20
E	20	25	25	40
F	225	25	25	40
G	1025	as required	15	66
H	1000	12.5	12.5	80
I	150	as required	12.5	80
J	2000	10	10	100

It is important to note at this time that the messages are ordered according to their required latency. This is because the data latency requirement is the parameter which must be used to determine the messages' priority.

The next step is to determine the appropriate target latencies for the four priorities. The data given in Table 4.5.2-1 lends itself to a breakdown in which each high latency message is a multiple of the lowest latency message. This is a good feature to shoot for because it can simplify the design process. In this example, the priorities will be established to support 10 [msec], 20 [msec], 40 [msec], and 80 [msec] latencies. With this assumption, messages A and B are priority 0, C and D are priority 1, E, F and G are priority 2, and H, I, and J are priority 3.

4.5.2 (Continued):

With the messages assigned to priorities, it is time to calculate the transmission time for all priority j messages for each priority D_j. This is done on a station by station basis according to the following equation:

$$d_{ij} = \{(\#\text{ of priority i words}) * 16\text{ bits/word} + (\text{overhead bits such as address and MFCS}) * (\#\text{ of priority i messages})\} * \text{Bit Time} \quad [\text{sec}] \tag{4.5.3}$$

With the data given in table 4.5.2-1,

$$d_{i0} = \{(70\ [\text{words}]) * 16\ [\text{bits/word}] + (27\ [\text{bits/message}]) * (2\ [\text{messages}])\} * 0.02\ [\mu\text{s/bit}] = 23.48\ [\mu\text{sec}]$$

Since all stations are assumed identical in this example this yields 23.48 [μsec/station] * 10 [stations] = 234.8 [μsec] for a value of D_0. Repeating the calculations for the remaining three priority levels yields the values listed below:

TABLE 4.5.2-2 - Example Transmission Times

Priority	d_{ij} [μsec]	D_j [μsec]	n_j	D_j [μsec]
0	23.48	234.8	1	234.8
1	65.08	650.8	2	325.4
2	408.02	4,080.2	3	1,360.07
3	1,009.62	10,096.2	4	2,524.05

4.5.3 Bus Loading Estimation: The next step in the example is to determine the expected amount of bus traffic on each token round. By setting the priority latencies to be multiples of the priority 0 latency the determination of the number of token rotations necessary to transmit all of the messages of each priority was simplified. In this phase of the design, it will be assumed that one token rotation is capable of transmitting all of the priority 0 messages of each station. This provides the greatest amount of media bandwidth for message transmission because the token rotation overhead T is minimized. Proceeding logically it follows that 2 rotations are required for priority 1 messages, 3 for priority 2 messages, and 4 for priority 3 messages. Dividing the D_js by the appropriate n_js yields the D_j values needed for the next step. For convenience they were also listed in Table 4.5.2-2. Now, the sum of the D_j values and the token rotation overhead T must be less then the desired priority 0 latency for the physical implementation to be suitable for the system. This is stated in the following equation:

$$\sum_{j=0}^{3} D_j + T < L_0 \tag{4.5.4}$$

4.5.3 (Continued):

For this example,

234.8 [μsec] + 325.4 [μsec] + 1360.07 [μsec] + 2524.05 [μsec] + 17.7 [μsec] = 4437.19 [μsec]
(Less than 10 000 [μsec])

which indicates that the planed implementation of the LTPB protocol has sufficient bandwidth to support the necessary system traffic. From equation 4.5.4, the system designer can see that more stringent latency requirements for priority 0 messages directly effect the bandwidth of the designed network. It is important to point out that the D_j values for the lower priorities can be reduced if their respective latency requirements are lowered. Thus the determination of the required latencies for all of the priorities has a direct effect on the bandwidth capabilities of the designed system.

It would also be expected that any desired growth margin would need to be considered at this time. For example, it was determined that a 100% growth margin was desired for all priorities. The values in Table 4.5.2-2 would be modified to those given in Table 4.5.3-1.

TABLE 4.5.3-1 - Example Transmission Times With 100% Growth

Priority	d_{ij} [μsec]		D_j [μsec]		n_j	D_j [μsec]	
0	23.48	46.96	234.8	469.6	1	234.8	469.6
1	65.08	130.16	650.8	1,301.6	2	325.4	650.8
2	408.02	816.04	4,080.2	8,160.4	3	1,360.07	2720.13
3	1,009.62	2,019.24	10,096.2	20,192.4	4	2,524.05	5048.1

The above data demonstrates that to increase the margin equally for all priorities it is only necessary to increase the respective D_j values. There is also the option of having a different growth budget for each priority.

The designer must also note that the above calculation assumes proper operation of all of the stations on the network. It would be advisable to allow some additional margin for station failures and multiple token pass attempts.

4.6 Simple Design Example:

This section presents a simplified example by providing one method for timer determination. This approach is simplified by working from the estimated message traffic up to the timer settings instead of working from the latencies to the maximum allowed timer settings. The latter approach is based more closely on the equations of 4.2 through 4.4. This approach is also simplified by making certain predeterminations of variables. Most notably, the n_j values are set as described in the preceding section. It is also assumed that the required latencies for the priorities have been set and that the various transmissions times, D_j, D_j, d_{ij},and d_{ij} are known. Recall from 4.3 that $D_j = n_j * D_j$. In this example, the data from Table 4.5.3-1 including the growth budget is utilized.

4.6.1 Setting the THT: The THT for each station represents the amount of time that the station may control the network. As such it must be long enough to allow the transmission of the necessary messages of all the priorities. This is basically the sum of the d_{ij} values. This can be stated as the equation,

$$\sum_{j=0}^{3} d_{ij} < THT_i$$

where d_{ij} is simply D_j from Table 4.5.3-1 divided by the number of identical stations N = 10. Thus for this example,

$$THT_i > 46.96\ [\mu sec] + 65.08\ [\mu sec] + 272.013\ [\mu sec] + 504.81\ [\mu sec] > 888.863\ [\mu sec]$$

This will ensure the ability to transmit the required message traffic under normal operating conditions. In a realistic system, each station would be expected to have different amounts and types of traffic and so would be expected to have different values for its THT.

4.6.2 Setting the TRTs: The TRTs are slightly more involved. As system-wide timers, their value must represent the overall system communication at each of the remaining priority levels. This calculation yields the minimum value needed for each TRT_J as follows,

$$TRT_J > \sum_{j=0}^{J} D_j + T \qquad (4.6.2)$$

However, equation 4.6.2 does not consider the need to transmit messages of lower priorities than the J used in the equation. Because the THT is actually replaced by the residual TRT_J value when priority J-1 messages are all sent, it is imperative that the higher priority TRTs allow time for the transmission of lower priority messages in the same manner that the THT needs to be set to allow for the transmission of all of the lower priority data traffic for each token rotation. The equations below show the TRT values determined if the need to transmit lower priority messages is not considered.

$$\begin{aligned} TRT_1 &> D_0 + D_1 + T \\ &> 469.6\ [\mu sec] + 650.8\ [\mu sec] + 17.7\ [\mu sec] \\ &> 1138.10\ [\mu sec]\ , \end{aligned}$$

$$\begin{aligned} TRT_2 &> D_0 + D_1 + D_2 + T \\ &> 469.6\ [\mu sec] + 650.8\ [\mu sec] + 2720.13\ [\mu sec] + 17.7\ [\mu sec] \\ &> 3858.23\ [\mu sec]\ , \end{aligned}$$

4.6.2 (Continued):

and

$$TRT_3 > D_0 + D_1 + D_2 + D_3 + T$$
$$> 469.6\ [\mu sec] + 650.8\ [\mu sec] + 2720.13\ [\mu sec] + 5048.1\ [\mu sec] + 17.7\ [\mu sec]$$
$$> 8906.26\ [\mu sec]\ .$$

Note that the value determined for TRT_3 represents the total expected bus traffic as calculated in equation 4.5.4 with the growth budget included. As explained above, these values do not agree with the relation given in 4.4,

$$TRT_j[max] \leq TRT_{j-1}[max] \qquad (4.6.3)$$

In order to support the proper operation of the priority scheme, each TRT must exceed the traffic requirements based on the TRT_3 setting. Moreover, it is suggested that each timer be set to allow for the transmission of at least one message of a lower priority in a worst case situation. This means that

$$TRT_j[max] \geq TRT_{j+1}[max] + M_j \qquad (4.6.4)$$

Where M_j represents the longest message of priority i. This attempts to guarantee the transmission of at least one message of a lower priority for each token rotation. For equation 4.6.4 to function properly, the TRT for priority 2 must be calculated first based on the calculation given in equation 4.6.2 for TRT_3. This should be followed with a calculation of TRT_1 based on the new value of TRT_2 as follows,

$$TRT_2[max] > 8906\ [\mu sec] + 1025\ [words] * 16\ [bits/word] * 0.02\ [\mu sec/bit] > 9234\ [\mu sec],$$

and

$$TRT_1[max] > 9234\ [\mu sec] + 150\ [words] * 16\ [bits/word] * 0.02\ [\mu sec/bit] > 9282\ [\mu sec].$$

These values represent the minimum values necessary to ensure proper operation of the priority scheme with the data given in Tables 4.5.2.1 and 4.5.3.1. In 4.6.4 these will be compared with the theoretically allowed maximum values for the THT and TRTs to verify if this design is suitable.

4.6.3 Comments on System Design with the LTPB: There are several comments which need to be made at this point. They are presented in the following paragraphs.

The information presented in 4.5 is important regardless of the chosen design approach or system philosophy. The one factor which may be changed is the setting of the number of token rotations n_j for each priority. By allowing more rotations then are necessary it is possible to decrease the average latency for messages of all priorities. This is achieved at the cost of additional system overhead in the form of token transmit and reception times and propagation delays and does not decrease the guaranteed latency.

4.6.3 (Continued):

The effects of station failures will have the greatest impact on the transmission of lower priority messages. Because the TRTs are system level parameters, their time will be used during searches for the next available station during the station admittance process or in the event of a failed station. It would be advisable to include some extra time in these timers to allow for system problems. The amount of time allowed will depend on the size of the network, the probability of the station failures, and the importance of the latency requirement on the data. Even priority 0 messages may not achieve their target latencies in the event of failures such as the loss of the token. In this event, all of the stations would again need to vie for admittance to the logical ring and the new logical ring would need to be established. These factors must be considered in the overall system design process.

The design philosophy of the LTPB expects the protocol to be able to properly impose a priority scheme on messages received from the host system regardless of the order or frequency of reception. In a system of this nature, the designer must resort to the use of probabilities to set the timer values. An approach of this nature is described in reference 4. Still, at some point the frequency of message generation must itself be limited or the lower priority messages of the system will never be transmitted. This can create a more complicated system design problem then the more traditional message scheduling approach of a centralized control oriented system. However, these two approaches are not necessarily mutually exclusive. If each of the hosts implements a relatively simple scheduler, the benefits of the distributed token passing protocol can be achieved while still supporting the structured data traffic of some avionics applications. The implementation of this scheduler can either be a part of the host or a part of each process contained in the host. Again, this is a design specific problem which is left up to the system designer for the particular system.

4.6.4 Verification of Timer Values: This section uses the equations developed in 4.2 through 4.4 to verify that the values determined in section 4.6 satisfy the requirements. As stated in 4.6, the equations developed in 4.2 through 4.4 approach the problem from the direction of finding the maximum allowed timer values to meet the given latency requirements. These can be used to set the upper bound on the timer settings and to verify that the values determined are suitable.

4.6.4.1 The THT Maximum Value: Starting with the final equation of 4.3.2.1,

$$n_0 \left\{ \sum_{i=0}^{N-1} (THT_i[max]) + (N*M) + T \right\} \le L_0 \qquad (4.6.5)$$

4.6.4.1 (Continued):

Now, substituting into this equation, with M, the maximum message length = 2000 [words] * 16 [bits/word] * 20 [nsec/bit] = 0.64 [msec], and solving for the sum of the THTs,

$$1\left\{\sum_{i=0}^{N-1}(THT_i[max]) + (10 * .64\ [msec]) + .0177\ [msec]\right\} \le 10\ [msec]$$

$$\sum_{i=0}^{N-1} THT_i[max] < 3.5545\ [msec]$$

The value of 888.863 [μsec] * 10 for the 10 stations = 8.888 [msec] exceeds the value allowed according to equation 4.6.5. The primary reason for this is the size of the largest message, 2000 words. There are at least two different approaches to solve this problem. The first is to solve it at a higher level. More specifically, the system must be designed with the scheduled message concept to guarantee that all of the stations will not try to transmit their longest message on the same token rotation. This would be difficult to do and is not recommended because it is not consistent with the philosophy behind the LTPB protocol. The more appropriate solution is to limit the system maximum message size. A message size of 2000 words represents 6.4% of the transmission time allowed by the priority 0 latency requirement. In order to determine the maximum allowed message size, return to equation 4.6.5 and solve for M.

$$M \le \left\{(L_0/n_0) - \sum_{i=0}^{N-1}(THT_i[max]) - T\right\}/N \qquad (4.6.6)$$

Which yields,

$$M \le \{(10\ [msec]/1) - 8.888\ [msec] - 0.0177\ [msec]\}/10 \le 109.43\ [\mu sec]$$

Which is equivalent to 335 words per message. Therefore, in order to satisfy the inequality of equation 4.6.5, the maximum message length for the system must be limited to 335 words. This limit will increase the number of messages sent and so will effect the overhead of the system by 27 bits per message. For a real system design, it would now be necessary to repeat the calculations of 4.5 and 4.6 to reevaluate the capabilities of this implementation of the LTPB to meet the system requirements. Instead of repeating the above work, the discussion will move on to the development of the TRTs.

4.6.4.2 The TRT Maximum Values: As presented in 4.3.2.2, for the proper operation of the network, the token rotation time must be less than the allowed latency. This basically provides an upper bound on the value of any f the TRTs given by the following equation.

$$TRT_j[max] + M < TR[max] < L_0 \quad (4.6.7)$$

Thus, for the example system,

$$\begin{aligned} TRT_j[max] &< L_o - M \\ &< 10\ [msec] - 109.43\ [\mu sec] \\ &< 9.891\ [msec] \end{aligned}$$

The values calculated in 4.6.2 meet this requirement. In fact, with the new maximum message size they would actually be lower then the values presented.

4.6.5 System Design Approach Summary: This section presents a summary of some of the important points for designing a system using an implementation of the LTPB protocol.

In order to begin a design of a LTPB protocol implementation, certain parameters must first be determined. These are the system physical characteristics which yields the value for T, the required latencies for the four priority levels, L_j, the planned bus traffic, and the desired or required growth budget. After these have been established it is possible to determine whether the implementation meets the requirements of the system. Once this hurdle has been crossed, the designer can proceed with an initial timer determination by using equations 4.6.1, 4.6.2, and 4.6.4. As was demonstrated in the example it is important to verify the solutions through the use of equation 4.6.5 and to modify the design if necessary. A method of modifying the design through the limiting of the maximum message length is given in equation 4.6.6. This method requires another iteration of the transmission time determinations and the timer values but it can meet the requirements without effecting the desired latencies or growth budget.

Many other concerns face the system designer, some of those mentioned in this section included the budgeting of time to allow for system failures or exceptions, the possible need for message scheduling, the effects of allowing more then the minimum required number of token rotations, and the differing impacts of system failures on the different priority levels. Hopefully, this section has at least alerted the system designer to some of the issues involved with design of implementations of the LTPB protocol.

As a final note, a different approach to timer determination is presented in reference 4. It is based more on probabilities than 100% guarantees and so is more in keeping with the pure LTPB protocol philosophy. In addition to describing a method for setting the system timers, it describes an approach to message segmentation which is worth the attention of the system designer. One of the major points of the paper is that the probabilities for meeting the required latencies need only be on the order of magnitude of the bit error probability for the planned implementation. One relaxation that is suggested is that the maximum message length considered in the equations should be the value of M_j for the associated priority and not the worst case for the entire system. The designer is referred to that work for further detail.

4.7 System Design Approach II:

The bus system designer works with the priority system to bound the latency of each class of message. To do this, classes of messages for the network must be established. The protocol allows this to be done in several ways. One approach is to construct a system in which, under normal loading conditions, four levels of priority are implemented system-wide. This assumption is the basis for this design approach. The designer must, knowing the characteristics of bus traffic which will occur, decide which messages will be placed at each of the four levels of priority. It is assumed, for this approach, that the messages transmitted at the high priorities are such that it is critical that they get through within a guaranteed latency period, while messages of a lower priority are allowed to be deferred for a short period during the brief periods of high bus loading.

After having decided the issue of priority assignment, the designer must decide how to make the system work with normal traffic, such that system messages are delivered within the system latency constraints. First, he must guarantee that the token will complete each logical ring within a certain amount of time. This time is based on the worst case message latency which the system requires and is determined by the token holding timer (THT). The THT acts as the bound for transmission of the highest priority message from a station (in a worst case scenario, a station may have enough messages to "fill-up" the THT and hence, this will bound the total amount of time the station may hold the token). Another consideration is the fact that the station may begin to transmit a message just before the THT expires. Since the station will complete transmission of this message before passing the token, this amount of time should be taken into account during calculation of the worst case times.

Next, the system designer must look at the remaining three levels of priority and decide how to set the token rotation timers (TRTs) for these levels of messages. The values for the TRTs must be set such that under normal operating conditions, all message traffic from each station will be transmitted on each token hold. Note that the TRT maximum value may be greater than the THT maximum value. Remember, the TRTs run continuously from the time the station last received permission to transmit messages of that priority until the next time that permission is received. This includes the time while other stations are using the token. The THT is loaded and runs only during the time the station has the token. Generally speaking, the values for the TRTs are set such that the following relationship occurs:

$$TRT1 \geq TRT2 \geq TRT3$$

In other words, Priority 3 messages will be the first to be deferred, Priority 2 second, and Priority 1 last. Priority 0 messages (bounded by the THT) should always get through in a properly designed system.

Assume, for sake of developing a general equation describing the worst case latency on the bus, that we have a three station system (N_1,N_2,N_3). In this system, the THT maximum value and the TRT_x maximum values in each station are identical. We will use a value of 7 for the THT. The length of messages pending in the stations is 1 unit of time.

4.7 (Continued):

Now assume that the TRTs in each of the three stations are set to the following values:

$$TRT1 = T_{base} + 17$$
$$TRT2 = T_{base} + 11$$
$$TRT3 = T_{base} + 07$$

(T_{base} is simply a variable representing a value of time. In this case, that value of time is a "base" value which is present in all of the TRTs and is used to simplify calculations.)

Now assume that:

$$T_{base}+(N_x*THT_{max}) \geq TRT1 \geq TRT2 \geq TRT3$$

Since:

$$(T_{base}+21) > (T_{base}+17) > (T_{base}+11) > (T_{base}+07)$$

In the case of an empty token round (no messages are passed by any station, only a token) (see Figure 4.7):

	Station 1				Station 2				Station 3			
	P0	P1	P2	P3	P0	P1	P2	P3	P0	P1	P2	P3
Pass 1	0	0	0	0	0	0	0	0	0	0	0	0
Pass 2	0	17	0	0	7	0	0	0	7	0	0	0
Pass 3	7	0	0	0	AND SO ON.....							

FIGURE 4.7

After having finished the empty token round, we set up a round in which message traffic is handled, we can build a worst case scenario which will result in an equation which will allow us to describe the setting of the THT and TRT1 (they are related, as will be demonstrated).

In Pass 2, station 1 has 17 messages at P1 and stations 2 and 3 have 7 messages each at P0.

4.7 (Continued):

Now assume that station 1 has just finished checking the P0 message queue and proceeds to move on to handle P1 messages when the host places a P0 message in his queue. The worst case time for that P0 message to get on the bus is:

$$TRT1 + (N_x * THT_{max}) \leq L$$

where:

L is the maximum latency for the system

The relationship TRT1≥TRT2≥TRT3 guarantees that this latency delay will hold for all P0 messages in all possible situations. Also note that the variable "T_{base}" has disappeared from the computation.

By using the BASIC program included in appendix 3 of this handbook, the user can explore the characteristics of the priority scheme for various THT/TRT settings with a simple system. Also note that the scheme detailed above is not the only method which can be implemented within the standard. There will, undoubtedly, be systems in which certain peculiar message traffic characteristics are encountered. In these cases, it may be desired that certain stations have different TRT or THT settings than other stations. This is allowable within the scope of the standard. The standard defines a hardware approach to the priority system. Ultimately the user (system designer) is responsible for making the system function properly and may use the available hardware functions in any given system to implement his particular needs.

4.8 System Design Approach III:

We now look at a second approach to the problem. Let's suppose that the designer decides to establish six classes of messages. These might correspond in descending priority to 128, 64, 32, 16, 8, and 4 Hz data.

Bear in mind that while these classes intentionally correspond to data rates, they are really classes of messages, so far as the system design is concerned. For example during the reload of a mission computer, the reload messages containing the program may compete with 64 Hz data by virtue of the priority level selection. This, in spite of the fact that the entire package of messages which contain the program is already in the queue at the time the transmitting station first gets the token (arrival rate much greater than 64 Hz).

For the designer to be able to guarantee that the 128 Hz data have an opportunity to get out at that rate, the system design must assure that a station that has 128 Hz data sees the token at at least the 128 Hz rate or at least once every 7.8 milliseconds and is able to transmit the messages. To guarantee that, the designer must be able to establish the maximum amount of time each station around the logical ring can hold the token.

4.8 (Continued):

We assume that the designer has been able to establish that the stations with 128 Hz messages will see the token at least once every 7.8 milliseconds. The designer is now ready to make sure that those stations with 64 Hz data see the token at least once every 15.6 milliseconds.

What the designer needs to be able to guarantee is that, while still getting out all the messages which must go out at 128 Hz, there is enough bandwidth left to get out at least half the messages that must go out at 64 Hz so that the other half can get out the next time around the logical ring. In this example we picked rates that are multiples of 2 apart. That way 1/2 of the 64 Hz messages are transmitted in 7.8 ms time and the other half in the next 7.8 ms that's all of it in 15.6 ms which of course is the period of 64 Hz.

The logical extension of this is that each trip around the worst case logical ring would have enough bandwidth to handle all of the 128; 1/2 of the 64; 1/4 of the 32; 1/8 of the 16; 1/16 of the 8 and 1/32 of the 4 Hz data messages. If not, the system is improperly designed.

In order to properly assign bus bandwidth to the system message traffic, we must determine the traffic mix. Assume the designer has determined that 10% of the total time that messages occupy the bus they will be 128 Hz messages. This takes into account not only the number of messages, but the length of each message as well. For our example the rest of the messages are distributed as 20% are 64; 30% are 32; 20% are 16; 10% are 8, and 10% are 4 Hz message classes. Figure 4.8-1 shows this mix.

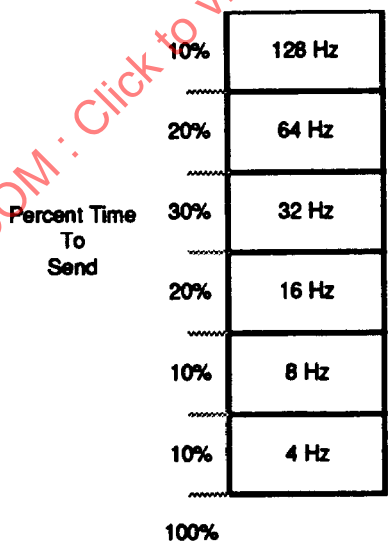


FIGURE 4.8-1 - Percent Time Required to Send All Messages

4.8 (Continued):

The worst case trip around the logical ring must take no more than 7.8 milliseconds (128 Hz) and must allow all of the 128 messages to be sent. That is, 10% of the time (0.78 ms) must be devoted (or able to be devoted) to 128 Hz messages. One half of the 64 Hz or 10% (1/2 of 20% x 7.8 ms = 0.78 ms) must be devoted to 64 Hz messages and so on. The system must be designed such that your worst case message load will be able to be transmitted within the time boundaries assigned.

The token rotation rate necessary is as fast or faster than the minimum rotation rate necessary to guarantee the highest priority update rate. In this case 7.8 ms. However, at those high-speed, worst case rotation rates the same message distribution as established by the priority levels must be maintained. Otherwise, under heavy loading, some priorities will be starved. Remember that heavy loading and long rotation times are synonymous.

Assume that there are 7 stations on the network. Each station has its own mix of traffic. Station 1 transmits 1/2 of the 128 Hz class messages. The other 1/2 are transmitted by station 2. Station 1 also sends all the 4 Hz traffic and station 2 sends all the 8 Hz messages on the network. Stations 2 and 4 share equally in the transmission of the 64 Hz messages. Stations 2, 3, 4, and 5 send 1/6 of the 32 Hz messages each. Station 6 sends 1/3 of the 32 Hz messages. In addition, stations 6 and 7 each send 1/2 of the 16 Hz messages. See Table 4.8-1.

TABLE 4.8-1 - Allocation of Bus Bandwidth to Message Traffic

Station #	128	64	32	16	8	4	Total Time (%)
1	5%					10%	15%
2	5%	10%	5%		10%		30%
3			5%				5%
4		10%	5%				15%
5			5%				5%
6			10%	10%			20%
7				10%			10%
Total	10%	20%	30%	20%	10%	10%	100%

Since the total time that all the stations hold the token at a particular level is proportional to the total time to rotate around the ring, then Table 4.8-1 can be rewritten in terms of the 7.8 ms maximum time to rotate around the ring (see Table 4.8-2). In this table, the total time spent at each priority level appears across the bottom. The total time the station transmits each message appears in the right-most column.

TABLE 4.8-2 - Message Traffic Compared to Transmit Time

Station #	P0	P1	P2	P3	P4	P5	Total Station Xmit Time
1	.39					.78	1.17
2	.39	.78	.39		.78		2.34
3			.39				.39
4		.78	.39				1.17
5			.39				.39
6			.78	.78			1.56
7				.78			.78
Total Time Each Px	.78	1.56	2.34	1.56	.78	.78	7.80

4.8 (Continued):

We now derive an equation which relates the station message traffic to the amount of time required to make sure that all messages pending get onto the bus during a token hold. We will call this value the time to transmit (TTT). For the purposes of the equation, this will be noted as:

TTT_{PN}=Time To Transmit at Priority "P", Station "N"

$TTT_{11}=TRT_{11}-[TX_{11}+TX_{21}+TX_{31}+TX_{02}+TX_{12}+...+TX_{0N}+TX_{1N}+TX_{2N}+TX_{3N}+TX_{10}]$

$TTT_{21}=TRT_{21}-[TX_{21}+TX_{31}+TX_{01}+TX_{12}+...+TX_{0N}+TX_{1N}+TX_{2N}+TX_{3N}+TX_{10}+TX_{11}]$

•
•

$TTT_{2N}=TRT_{2N}-[TX_{2N}+TX_{3N}+TX_{01}+TX_{11}+...+TX_{0(N-1)}+TX_{1(N-1)}+TX_{2(N-1)}+TX_{0N}+TX_{1N}]$

$TTT_{3N}=TRT_{3N}-[TX_{3N}+TX_{01}+TX_{11}+...+TX_{0(N-1)}+TX_{1(N-1)}+TX_{2(N-1)}+TX_{3(N-1)}+TX_{0N}+TX_{1N}+TX_{2N}]$

The resultant general equation for this method of setting the timers is:

$$TTT_{PN} = TRT_{PN} - \sum TX_{XY}$$

$$TRT_{PN} = TTT_{PN} + \sum TX_{XY}$$

We now look at an example system having the message traffic depicted in Table 4.8-3.

TABLE 4.8-3 - Example of System Message Traffic Analysis

	P0	P1	P2	P3	Total
Station 1	**27**	**9**	**1**	**0**	**37**
Station 2	**27**	**9**	**1**	**0**	**37**
Station 3	**27**	**9**	**1**	**0**	**37**
Totals	**81**	**27**	**3**	**0**	**111**

$$\text{TRT2} - \text{TTT} = \text{TRT2} - \sum(TX_{xy})$$

$$1 = \text{TRT2} - 111$$

$$\text{TRT2} = 112$$

$$\text{TRT1} - \text{TTT} = \text{TRT1} - \sum(TX_{xy})$$

$$9 = \text{TRT1} - 111$$

$$\text{TRT1} = 120$$

The calculated TRT values represent the exact times necessary to just complete all message traffic pending in this station after all other stations in the system have been allowed to do the same.

NOTE: When using this scheme, if a station takes longer than the allotted TRT to complete his message, all priority levels will share in the penalty. This is the penalty paid for using this scheme in a system in which message traffic may be dynamic in nature, unless the times are determined for the absolute worst case message traffic, in which case the bus will be inefficient for light loadings.

5. BIU STATE MACHINE DEFINITION:

This section describes the operation of the BIU in terms of a state machine with well defined criteria for transitions from state to state. The state machine defined in Section 5 of the standard only covers the TPIU operation once it is able to join the token passing logical ring. This is included in this section as state 3.0, but additional transfers have been added (identified by *) to integrate it with the rest of the BIU state machine. The additional states define the action on power-up and upon reception of station management subcommands.

5. (Continued):

This complete BIU state machine definition was not included in the standard since many aspects of a state diagram such as this are implementation specific. It is not the purpose of this section to dictate the design and function of an LTPB BIU. This section is meant only to point out that the overall design requires considerations of many more states than the ones detailed in the standard. In fact, during actual design, there will be other states or substates which may occur in individual implementations which are not represented in this section. The designer is urged to use the information in this section only as a starting point for design. The ultimate design goal is compliance with the AS4074 standard and those specific requirements imposed by the end user of the BIU.

The overall state diagram is shown in Figure 5-1. Five states are defined as follows:

State	Name
0	Offline
1	Quiescent
2	Disabled
3	Enabled
4	Fault