VMSbus FACILITIES FOR MULTIPROCESSING AND FAULT TOLERANCE

Mira PAUKER, PHILIPS CTI, Fontenay-aux-Roses, France

ABSTRACT

This contribution emphasizes VMSbus as a complementary control and data path to the parallel backplane VMEbus.
Multiple paths between boards provide multiprocessors with shared resource allocation and fault tolerant functionality.
Event message latency times are minimized for tightly and loosely coupled multiprocessor systems, using VMSbus intrabackplane and interbackplane links with self arbitration and resynchronization capabilities.
Redundancy in the fault isolation path, an alternative route to reset or disable failed boards, group addressing for broadcast and broadcall operations are basic means for fault localization and recovery.
Data link compound groups such as semaphores, signature-checking semaphores, tokenpassing, multiaddress talker and listener, locking transaction listener and talker are high reliability tools provided by the VMSbus.

1 INTRODUCTION

The VMSbus provides a serial communication path within a closely-coupled computer system and/or among systems in close proximity. It represents a complementary control and data path to the VMEbus parallel data path.
VMSbus uses two signal lines, a common single sourced clock signal and a data line on which the stations place and sample data. The data is logically OR-ed among active transmitters. Every transmitter continuously monitors and tracks frame progress on the link, solving bus contention problems by a straightforward protocol.

This presentation focuses on VMSbus as a response to multiprocessor system requirements and fault tolerant system needs.

2 VMSbus OVERVIEW

2.1 Layered representation

The VMSbus can be represented with a **layered** organization as in Figure 1 :

- Link Layer modules
  - HEADER SENDER (HS)
  - FRAME MONITOR (FM)
  - HEADER RECEIVER (HR)
  - DATA SENDER (DS)
  - DATA RECEIVER (DR)

They are always organized in groups of functional modules. These groups can be regarded as a higher sublayer within the Link Layer.

- Physical Layer modules (free standing modules) :
  - BUS ACCESS MODULE
  - SERIAL CLOCK MODULE

- Medium :
  - Conductive paths

- Higher layer management controls transactions between **functional groups** of Link Layer modules like the following :
  - A **Controller** group, formed by a HEADER SENDER and a FRAME MONITOR,
  - A **Flag** group, formed by a HEADER RECEIVER and a latch,
  - A **Talker** group, formed by a HEADER RECEIVER and a DATA SENDER,
  - A **Listener** group, formed by a HEADER RECEIVER and a DATA RECEIVER.

The functional groups involved in a transaction are typically situated on different boards, as illustrated in Figure 2.

2.2 Layer interface signals

2.2.1. Physical Layer and Physical Medium Interface

Two signals in the Medium provide communication among Physical Layer modules :
SERCLK - provides clocking information to Bus Access Modules
SERDAT* - bidirectional.

2.2.2 Link Layer and Physical Layer Interface

Six signals represent this interface :
SCLK, RONE, RSTART - from the Physical Layer to the Link Layer
XONE, XSTART, XJAM - from the Link Layer to the Physical Layer

SCLK transmits timing information to the Link Layer. The other signals relate to SERDAT*.

2.2.3 Link Layer module Interface

Five signals describe operation between the Link Layer modules :

SELECT - HEADER SENDER output to the FRAME MONITOR, indicating that HS won the bus arbitration

FRAME IN PROGRESS - FRAME MONITOR output to the HEADER SENDER, indicating a frame is in progress

DSAE - HEADER RECEIVER output to the paired DATA SENDER, indicating DSAE bit state in the Header subframe

S SELECT, R SELECT - HEADER RECEIVER outputs to the paired DATA SENDER, or DATA RECEIVER, upon detection of respective selection code in Header subframe

2.2.4 Higher Layers/Link Layer Interface

Some of the following signals, used to represent the interface of the Link Layer modules **with higher layers serial bus management,** will be detailed by the explanations concerning high reliability built-in features, presented in 6.3 to 6.6.

S STROBE, R STROBE - sourced by HR
S ENABLE, R ENABLE - sent to HR
SENT, CANCELLED - sourced by FM
RESOURCE FREE - sent to HS
Priority Port, S Code, R Code, Data Frame Port, Status Port
- sourced by FM
Code 1 Port, Code 2 Port, SEND12, SEND21
- sent to HS
LOST ARB - sourced by HS
Code Port - sent to HR
Data Port - sent to DS or DR
Data Size Port - sourced by DR
DSENT - sourced by DS
RESET* - sent to all LL modules

2.3 Frame protocol

The serial communication frame defines a transaction protocol surrounding data or controls with synchronization and error detection bits.

A typical frame representation is :

|(——— Header Subframe ———)|

Start bit	PRI	S	R	DSAE	HS VAL	Frame Type	Data	Frame Status	Jam Detect
1	3	10	10	1	1	3	8-256	3	1

Length (bits)

Control frames do not contain the data subframe.
Cancelled frames contain neither data nor status subframes. The Frame Type value is forced to 111 by the entity which cancels a frame.
The Frame Type subframe value 000 characterizes control frames. The values 001 to 110 give the length of the data subframe.
The Frame Status field indicates the status of selected modules. A"1" in the most significant bit indicates an unsuccessful operation due to a data size conflict between DATA SENDERS or RECEIVERS, or incorrect selections during control or data frames. The other two bits indicate the selection of modules by the fields S or R.

Different entities provide the information in the various subframes. It is a **read-write** protocol, checking the active participation of the implied modules.

The Jam detect bit is sampled by the modules involved in the frame transmission to check frame synchronization. If frame synchronization is lost, the FRAME MONITOR drives a sequence of 512 one-bits, which will resynchronize all modules. The presence of a "1" in the Jam detect bit makes the participating modules ignore preceding frame transmission.

3 SUPPORT FOR MULTIPROCESSOR SYSTEMS

The main features of the VMSbus are oriented to support multiprocessor systems.

3.1 Messages between distributed processing elements

VMSbus data frames carry messages :
- in tightly coupled configuration sharing a common memory and a single operating system, as well as
- in loosely coupled configurations with a local memory and local O.S. for each processing element.

3.2 Addressing capabilities

VMSbus generalized addressing is widely useful for :
. broadcast addressing,
. group addressing,
. polling by broadcall operations.

3.3 Resource allocation

VMSbus provides special **mechanisms** like :
. semaphore implementation,
. token implementation,
which are important for reliable multiprocessing.

4 VMSbus MULTIPROCESSING ORIENTATIONS

4.1 Throughput oriented

The VMSbus autonomy allows :
. maximizing the number of independent jobs done in parallel by general purpose computers.
. balancing the workload between CPU-s and I/O-s.

4.2 Availabiliby oriented

The alternative link offered by the VMSbus is suitable :
. for real time on line applications.
. for failsafe operation (or short downtime).
. for I/O intensive applications.
. to maximize the number of interdependent tasks done in parallel.

4.3 Response oriented

Optimized VMSbus allocation and predictable delivery time are essential :
. for dedicated/embedded applications.
. in specialized applications.
. in CPU intensive applications.
. **to maximize** the **number of cooperating processes** done **in parallel.**

The VMSbus functionality that responds to these needs is summarized by the following.

5 VMSbus SUPPORT FOR FAULT TOLERANCE

The VMSbus frame protocols are oriented for short autonomous communications, efficient to failure reply.

5.1 In fault **confinement**

VMSbus can be used to limit the spread of fault effects in a system, in the following ways :

. under Operating System management, VMSbus can be used to effectuate consistency checks of range.

. before performing a function, the VMSbus can be used for requesting and confirmation of resource availability.

5.2 In fault **detection**

VMSbus frames are most suitable :

. to **notify** system elements of a fault occurence,

. to do consistency **checking** concurrently with the useful transfers on the parallel bus,

. to implement **watchdog** timers and time outs.

5.3 **Dynamic** redundancy can be achieved by **switching on** spare components, using VMSbus control frames.

5.4 **Retry** operation

In case of **transmission error** detected by error checking included in the VMSbus protocol, a second operation attempt can be successful (up to 15 are provided by the SCC 68173 controller).

5.5 **Diagnosis** information can be provided about **the location** and its type of a failure, using VMSbus data frames.

5.6 **Reconfiguration**

VMSbus can **isolate** failed components or boards by switching them off, with a possible degradation of performance, or **replace** them by back-up spares.

5.7 **Restart** of the system

Can be achieved in three ways : **hot** - if no information was lost ; **warm** -if some processes can be resumed ; or **cold** if no process is surviving and a complete reload is necessary.

5.8 **On line repair** is possible using procedures equivalent to reconfiguration.

5.9 **Reintegration** of an on-line repaired board can be achieved without interrupting serial bus operation, by using the jam protocol of the VMSbus.

6 VMSbus BUILT-IN HIGH RELIABILITY FEATURES

6.1 Arbitration Mechanism

An arbitration mechanism is built into the procedure for transmitting frames :

- no separate lines are used,
- no dedicated time period is used.

Fair access to the bus can be ensured by the higher-layer logic **arbitration strategy.** The priority value of a HEADER SENDER can be **increased** when the arbitration is lost or can be **decreased** when a frame is cancelled.

6.2 HEADER SENDER VALIDATION

The HSVAL bit guards **against the possibility that all HEADER SENDERS might drop out of the VMSbus arbitration.** If HSVAL is a zero bit, the frame must be ignored.
The protocol allows several HEADER SENDERS to send the same Header subframe together. For applications in which this is not appropriate (e.g., Semaphore Set frames), the higher layers must ensure unique header subframes.

6.3 ENABLE S, ENABLE R validation

When a HEADER RECEIVER matches a selection codes in the S or R field, it will cancel the frame if the corresponding higher layer input ENABLE S or ENABLE R is false. This is useful if a data receiving buffer is not yet ready to accept new data or if some other resource is not ready for the frame operation.

6.4 S, R STROBE validation

The entire frame is **checked for successful transmission before** the S and/or R STROBE signal validates the frame.

No STROBE will be generate if :
- the most significant bit of the frame status field is one (abnormal operation),;
- the jam bit is true (frame desynchronization).

6.5 DSENT validation

The DATA SENDER output DSENT informs higher layers that it has successfully transmitted data. This allows two strategies for Talker or Listener groups : an **On-Demand** Talker/Listener or a **Transaction** Talker/Listener. An On-Demand Talker is always ready to send the most recent data provided by its higher-layer logic. An O.D. listener is always ready to receive new data and present it to its higher layers. Such "always ready" Talkers and Listeners, require double buffering or other means to avoid mixturing of old and new data.

A Transaction Talker sends data only once when it is provided by its higher layer logic. A Transaction Listener requires that its higher layers must read out received data before it will accept more. For such groups, the ENABLE S or ENABLE R signal controls whether the group is ready or whether it cancels the frame.

6.6 SENT/CANCELLED loop signals

The FRAME MONITOR, coupled with a HEADER SENDER, informs higher layers of the serial bus management of the results of frame transmission. The SENT output indicates successful transmission. The CANCELLED output indicates that a selected group was not ready. The LOST ARB output indicates the frame lost the arbitration to a higher priority one.

6.7 JAM protocol

The Jam procedure provides security against frame-level desynchronization among VMSbus modules.

This procedure is also useful in **extended configurations** where the serial bus modules do not share a common Reset signal. The Jam sequence allows **live insertion** of modules, assuring frame resynchronization with other controllers.

7 VMSbus MODULE GROUPS DEDICATED TO HIGH LEVEL FUNCTIONS

7.1 Simple Flag

The simplest group is formed by a HEADER RECEIVER (with its ENABLE S, R inputs set true) and a latch. Among other applications, the latch output can **enable/disable** the system **bus interface** of a board, **or** provide a signal to **Reset** the logic of **a board. Fault area confinement, reconfiguration** after repair, or backup recovery can be managed using Flag groups.

7.2 Multiaddress Flag

A Flag can be set or reset by means of any of several selection codes. Such a group is described as consisting of several HEADER RECEIVERS whose S STROBE outputs and R STROBE outputs are OR-ed logically. Such a Flag can be used to **dynamically disable** Masters on particular boards or all boards of a certain type.
Multiaddress Flag HEADER RECEIVERS are configured with their ENABLE S and R inputs true.

7.3 Multiaddress Talker

In this group, a DATA SENDER can be selected by means of any of several selection codes. The S-SELECT outputs of several HEADER RECEIVERS are OR-ed to produce the DATA SENDER's S SELECT input.

This group can be useful in system **monitoring** or **diagnosis,** reading data from a unique address or from a set of DATA SENDERS. In the latter case, the result can be the largest value among them (due to Data Arbitration Enable feature), or the logical OR of their data.

7.4 Multiaddress Listener

In this group a DATA RECEIVER can be selected by any of several selection codes, the outputs of several HEADER RECEIVERS are OR-ed to produce an R SELECT input.
In multiprocessor systems, this group allows **event notification** or **message sending** to all processors or to a group of processors.

7.5 Semaphore group

A single-bit **mutual exclusion mechanism** is provided by a compound group formed by a controller, a Flag and some additional logic.
The Flag status represents the status of a sharable resource in the multiprocessor system.
The controller provides access to set or clear this Flag, corresponding to the allocation or deallocation of the shared resource.
This group differs from a Simple Flag group in that the set frame is cancelled if the Flag is already set.
It is recommended to implement Flags in parallel with same addresses, one for each controller that can request the resource. These Flags must be in compliance. The serial bus protocols ensure this : a Flag is set by a successful "set frame" and also when a set frame is cancelled.

The semaphore operation with VMSbus includes two selection codes : a semaphore code, assigned to each sharable resource controlled by the Serial bus and a requester code, assigned to each group that can request control of any resource. The different Controllers trying to acquire the same resource MUST use different requester codes, so that there is a unique winner of the VMSbus arbitration. The resource and requester codes are typically assigned at configuration time.

7.6 Signature-checking semaphore

For **high security** operations, an additional requirement is applied to this type of semaphore group. The requester code is checked by the Flag logic, for the right to share the resource. Additional HEADER RECEIVERS, one for each permitted requester code, have their R STROBE outputs OR-ed. The access rights for each requester code are determined at the configuration time.

7.7 Token Passing Group

This is an alternative way to **share common** resources. A set of token passing groups form a **logical "ring"**. Each group knows the selection code of its "successor" group, to which it sends frames representing the token.
The resource can be controlled by a group as long as it has received the token and has not yet sent it to its successor.
Two mechanisms for error checking are considered :

- one RING ERROR logic included in the controller surveys the cancelled frames, on an attempt to set a Token Flag already set or to reset a Token Flag already reset. This ensures that only one token frame is passing through the logical ring ;
- a recommendation is done to maintain a timer controlling the maximum time the resource is used per group.

7.8 Locking transaction Listener/Talker

Two other compound module groups are aimed to allocate a resource to a process, **instead of semaphore** usage.
The Locking Transaction Listener, similar to a Multiaddressed Transaction Listener and a Flag, ensures a Listener receiving several frames of a data unit, in case of a file transmission, from one Talker, before allowing that Listener to be open to any other suitable-configured Talker. So, in case of a printer, a semaphore mechanism can be directly build in the printer serial bus interface.
The Locking Transaction Talker, equivalent to a Multiaddress Transaction Talker and a Flag, ensures the locking of a Talker to a Reading Controller, up to the end of the transmission of successive frames, until it is released by the controller.

These were **some examples** of VMSbus modules possible operations, in order to **satisfy** multiprocessor system requirements and high reliability environment.

8 FRAME TRANSMISSION OPTIMIZATION

Several built-in protocol features ensure minimal occupancy of the VMSbus medium by shortening rejected frames and then, avoiding unecessary one's.

8.1 Frame cancellation

Any addressed module, not ready to participate to the VMSbus transaction, has to cancel that frame by sending 1's in the Type field which follows the Header Subframe. In such a case, the frame is shortened to its minimal length of 30 bits :

Start bit	PRI	S	R	DS AE	HS VAL	1 1 1	JAM det
1	3	10	10	1	1	3	1

But this is not a "productive use" of the serial bus and cancelled frames has to be prevent if possible.

8.2 - Handshaking groups

Handshaking Writing Controller and Handshaking Transaction Listener are built to avoid frame cancellation and improve bus efficiency, controlling the transactions by the rate of the Listener data processing.

8.3 Variable Priority Controller

This module group associates a 3-bit counter to the priority port of the HEADER SENDER. Its initial value is set by the serial bus management, when the Header is configured.
When a frame is cancelled, the FRAME MONITOR of that controller clears the priority counter, avoiding to repeat sending. The same clearing mechanism can be used to ensure bus "fairness", after winning the bus allocation. In case the bus allocation is lost, the counter value must be incremented to increase the chances to win. This method ensures productive traffic on the bus.

9 CONCLUSION

I tried to show how VMSbus contributes to multiprocessing offering a concurrent and independent link to interconnect system boards. This link will facilitate scheduling and synchronization of the Processing Elements.
Fault tolerance capabilities introduced by this complementary bus will reinforce VMEbus usage in high reliable applications.
The Silicon support of the basic modules of this VMSbus protocol is opening its spread usage.
The newcoming SCC68173 VMSbus controller and the SCB68171 VMSbus interface chip are first circuits dedicated to this protocol. The operating systems have to incorporate efficient management of this communication link with all its wide opening features.

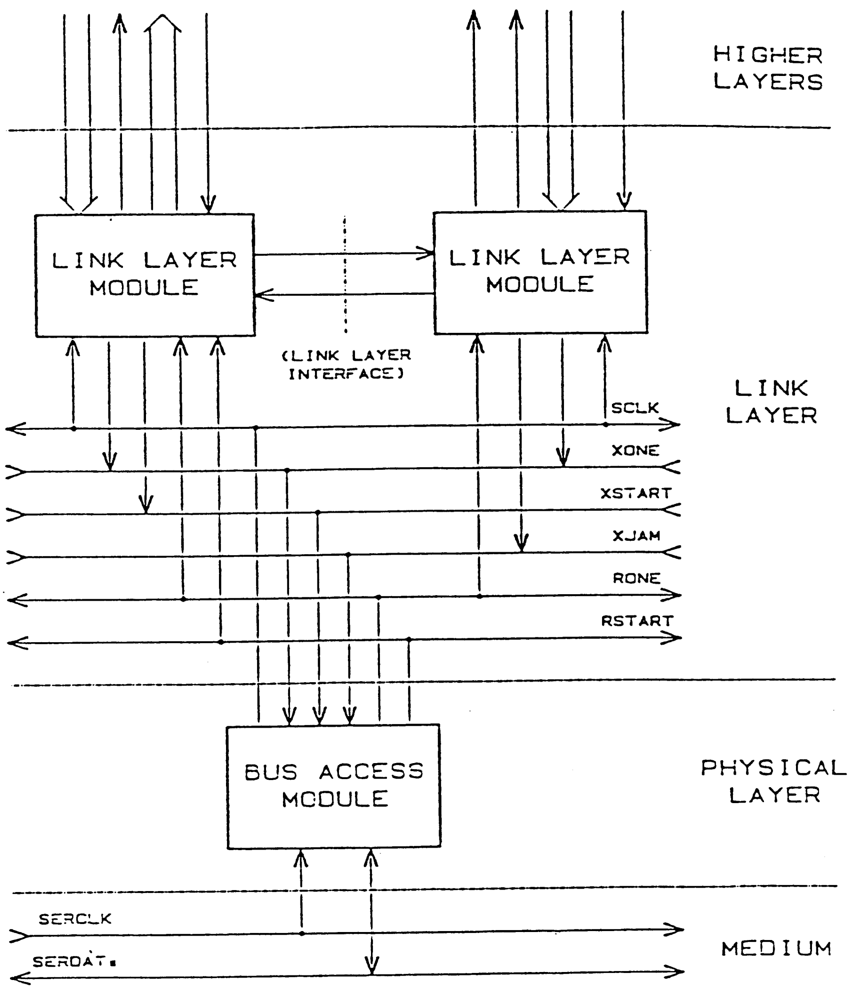

Figure 1 - VMSbus layered organization

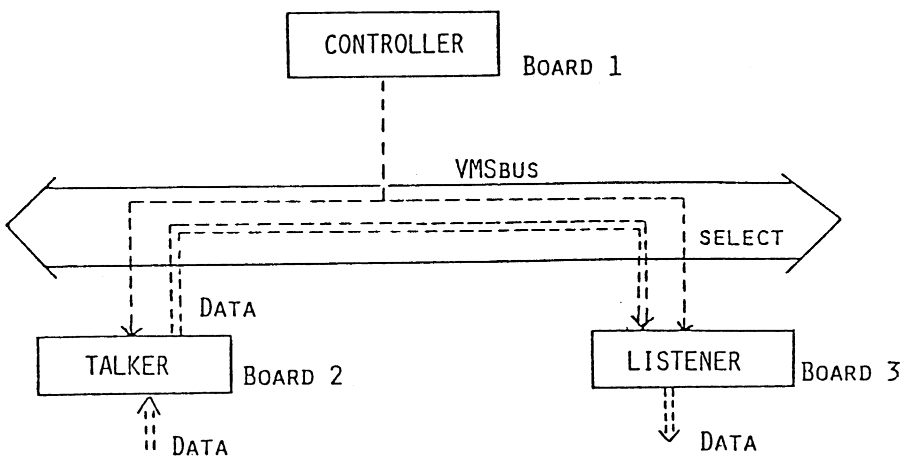

Figure 2 - Groups of modules on different boards