

JOÃO ANTONIO ZUFFO - LSI/USP

I - INTRODUÇÃO

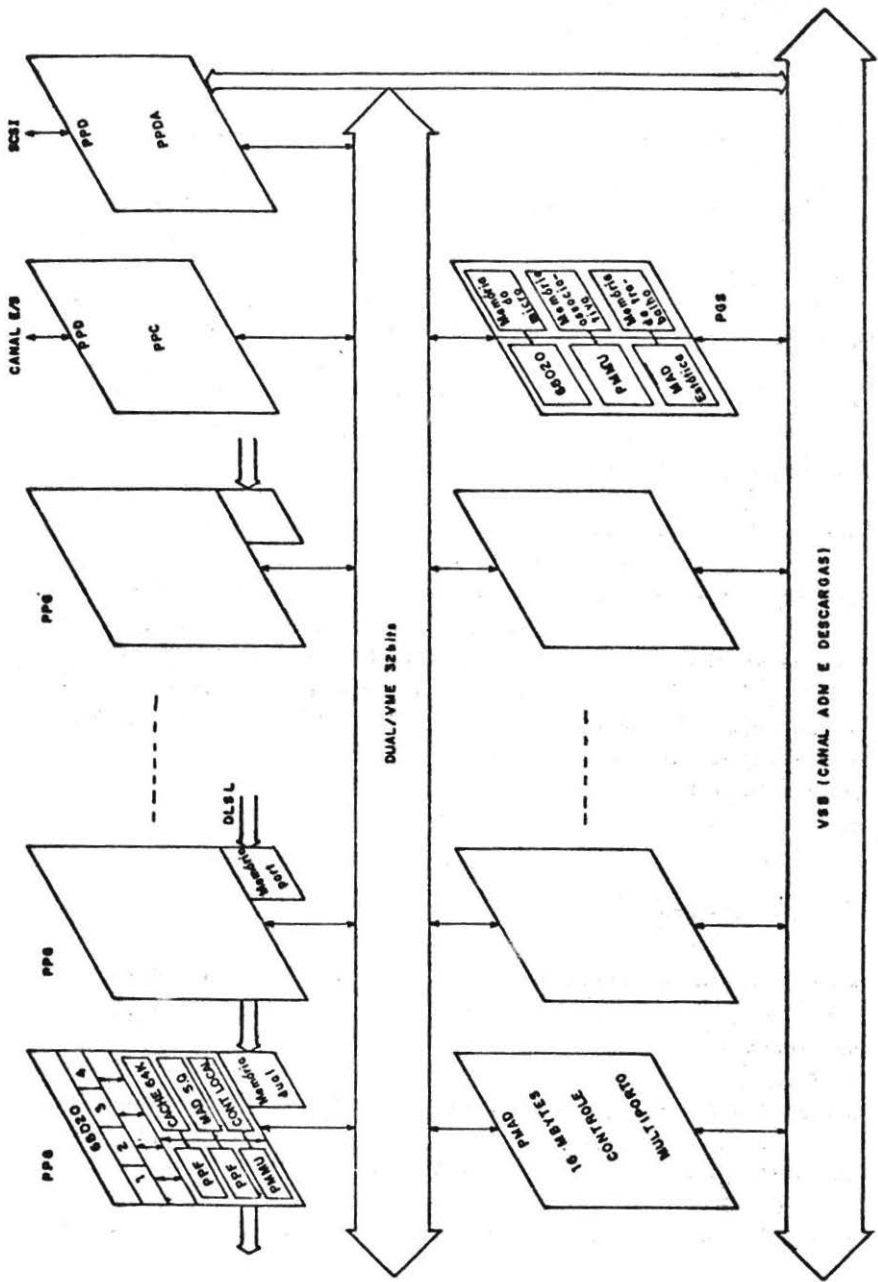
A universidade brasileira ressen-te-se sem duvida de um equipamento de maior porte que atenua não sô suas necessidades de processamento científico, mas também que seja uma máquina aberta que permita aos grupos de informática o acesso a circuitaria e a logicionaria de baixo nível, possibilitando deste modo uma ampla gama de facilidades em pesquisas e desenvolvimentos, não facilmente disponíveis em outros sistemas.

Através do MP-LSI 8701 é proposta uma organização e a arquitetura de sistema que além de versátil apresente um baixo custo de implementação, acessível portanto a grande parte das universidades brasileiras.

II - ORGANIZAÇÃO

A organização tentativa do sistema MP-LSI 8701 esta mostrada na figura 1.

A comunicação interprocessadores e a memória e a distribuição de tarefas tem sido o principal gargalo no desenvolvimento do multiprocessamento. O LSI partindo de sua experiência prévia procurando manter-se dentro de arquiteturas convencionais, de-



lineou tal organização. Como o leitor pode observar, o sistema consiste em uma série de dutos e tipos de placas. As placas são as seguintes:

PGS - Placa de gerenciamento de sistema

PPG - Placa de processamento geral

PMAD - Placa de memória de acesso direto

PPD-PPDA - Placa de processamento de dispositivos de Armazenamento

PPD-PPC - Placa de processamento de comunicações

Embora ainda não delineadas são possíveis outras placas de processamento especializado de acordo com a particular destinação do sistema.

A comunicação entre as placas dar-se-a através dos seguintes dutos:

DUAL VME - Duto VME duplo em termos de endereços e dados e em termos de todas as variáveis de controle e status que se tornem necessárias. Estes dutos interligam-se com outras placas do sistema. Velocidade total prevista 128 M bytes/s.

VSB - Duto utilizado na comunicação E/S do sistema. Destina-se à transmissão de dados individualmente ou em blocos. Faixa prevista 64 M bytes/s.

DLSI - Duto de comunicação entre as placas PPG, através de memórias portas duais. Faixa prevista 64 M bytes/s.

No delineamento da organização e arquitetura do MP-LSI 8701 foi levada em conta o desenvolvimento futuro da microeletrônica e a possibilidade da existência de novas pastilhas no mercado. É exatamente a queda de custos da circuitaria e a integração em escala cada vez maior que possibilitaram a colocação de todo este sistema em um único armário, sendo que seu núcleo central seja colocado em um único bastidor, VME compatível expandido.

III - DESCRIÇÃO DAS PLACAS

Apresentamos aqui uma descrição funcional rápida das placas de impresso do MP-LSI 8701, mostrando a organização interna em termos de blocos:

a) PGS - Placa de gerenciamento do sistema

Detalhes gerais: Como seu nome indica a função da placa de gerenciamento do sistema é gerenciar em baixo nível todas características e tarefas do multiprocessamento. A PGS completará portanto as funções do núcleo dos sistemas operacionais, sem contudo interferir com as características individuais de cada particular S.O.. É claro que isto só será possível dentro do contexto de cada particular S.O..

Caberá a PGS gerenciar também o processamento tolerante a falhas, como uma forma particular de multiprocessamento. Esta placa deverá contar e gerenciar lista de circuitaria em falha, correndo periodicamente programas de diagnose. A própria placa PSG poderá ser duplicada em sistemas de alta confiabilidade, comparando-se resultados de diagnose. Não se pretende na primeira fase, que estamos desenvolvendo, implementar facilidades tolerantes a falhas.

A placa de gerenciamento PGS poderá, em função do comando do usuário, determinar a granulação do paralelismo em multiprocessamento, selecionando parte ou a totalidade dos processadores para a execução de tarefas. Sob o comando da PGS a totalidade ou parte dos microprocessadores poderão operar em rede, executando as mesmas instruções e operações, sobre diferentes dados (Array Processors), ou então poderão operar em linha realizando diferentes operações em cascata, em um esquema de bombeamento (Pipelining). É claro que pretende-se implementar todas estas facilidades paulatinamente e de acordo com a especialização do sistema.

A placa de gerencia PGS providenciará também status para as páginas e instruções nos caches das PPGs para a execução de instruções vetoriais e matriciais em nível de linguagem de máquina.

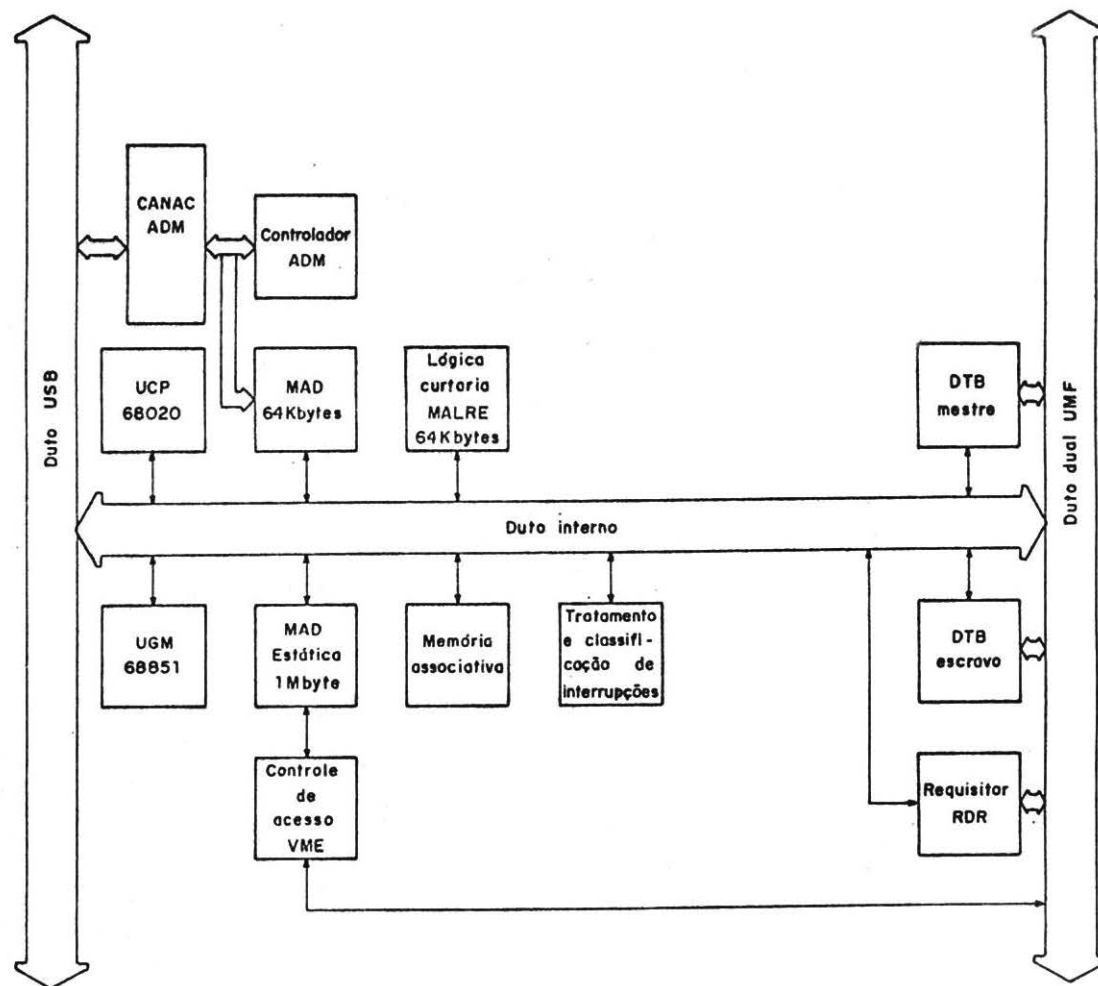


Fig.2 PGS- Placa de gerenciamento do sistema

Em princípio a PGS deverá conter:

- Microprocessador local 32 bits
- Memória apenas de leitura (ROM) com logicircuitaria de comando de placas (firmware)
- Memória associativa na qual se estabelece relação entre processos e tarefas de um lado e facilidades de circuitaria colocadas à disposição, no outro.
- Memória de trabalho com semáforos, status dos múltiplos microprocessadores e facilidades de circuitaria, incluindo presença de operação errática.
- Memória de Acesso Direto estática, utilizável por todo o sistema.
- Circuitaria complementar as placas PPG, PPDA e PPC.

Aprioristicamente preve-se as seguintes funções para a PGS:

- 1) Listagem das placas PPG e dos microprocessadores ociosos e designação a eles de acordo com as respectivas prioridades.
- 2) Correr programas de diagnose e estabelecer listagem de microprocessadores e facilidades de circuitaria com operação deficiente. Levantar setores de memória com deficiência.
- 3) Controle dos segmentos de tempo e dos recursos circuitais destinados a execução de cada tarefa. Este controle deve-se dar em nível de cada microprocessador individual permitindo a sua operação em tempo partilhado e em multiprogramação.

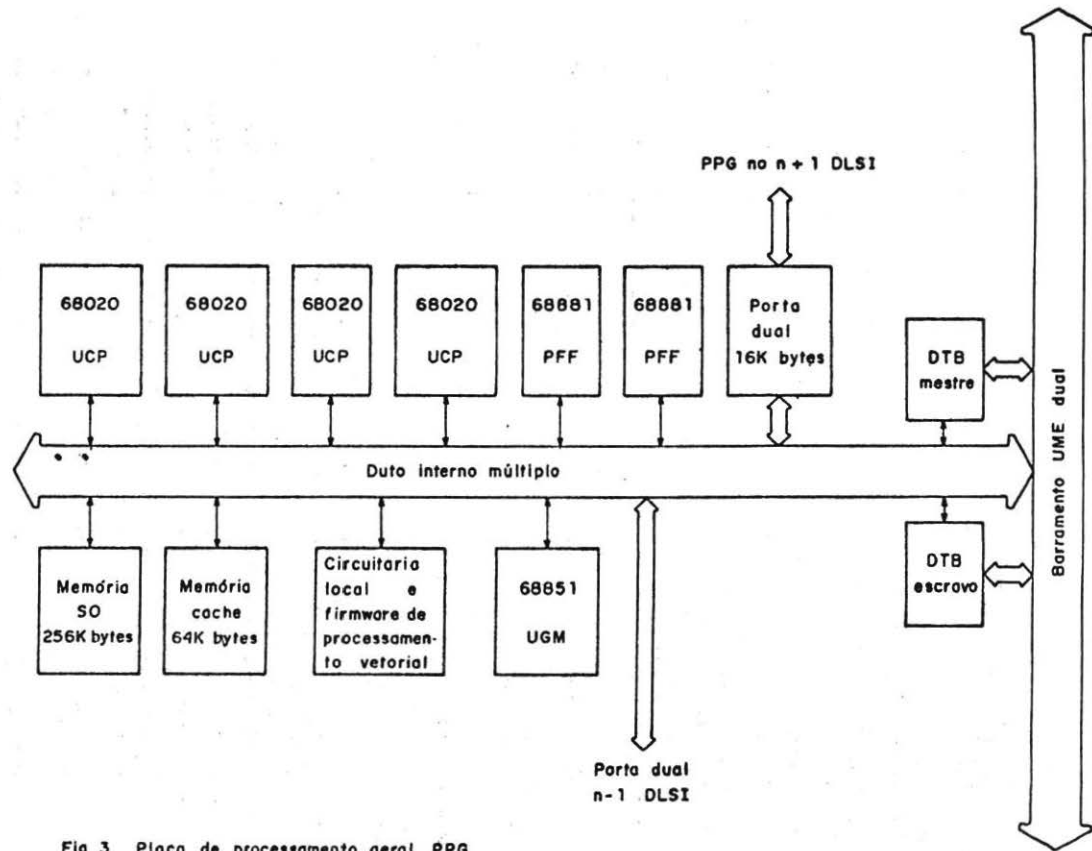


Fig 3 Placa de processamento geral PPG

- 4) De acordo com a configuração do sistema e disponibilidade de facilidades, estabelecer as prioridades das tarefas e os tempos de atraso nas respostas em tempo real. Reagir e analisar as especificações estabelecidas por usuários externos.
- 5) Supervisionar as trocas de informações com o meio externo colocando-as no estado de espera. O estado de espera de uma tarefa pode também ser originado, por necessidade de dados a serem obtidos pela execução de outra tarefa.
- 6) Avaliar "a priori" o tempo de execução das tarefas em função dos recursos e facilidades disponíveis. Se este tempo de execução for longo as facilidades disponíveis podem variar. Em função da prioridade da tarefa aumentar ou reduzir os recursos disponíveis.
- 7) Estabelecer canais de comunicação entre tarefas em execução que necessitem comunicação com outros processos ou tarefas ociosos ou não, ou então com o meio externo.
- 8) Permutar processos e tarefas com segmento de tempo esgotado por novas tarefas e processos. Dependendo do status e prioridade dessas tarefas e processos estes podem residir na memória principal ou no armazenamento secundário ou mesmo ter páginas na memória da PGS ou nos caches das próprias PPG.

- 9) Atender as solicitações das PPG e de seus microprocessadores que determinam suas tarefas, redistribuindo a eles novas tarefas de novo a otimizar o uso dos recursos e facilidades evitando que estes permaneçam ociosos. Esta estratégia visa a minimizar o tempo de execução dos processos e tarefas.
- 10) A gerência PSG está subordinada a obediência das condições dos sistemas operacionais, S.O. colocando sobre estas novas facilidades.
- 11) Interagindo com os S.O. sem neles interferir, estabelecer status de processos e tarefas interruptíveis ou não e estabelecer tipos de interrupção, que possam interagir com tarefas.
- 12) Supervisão de alocação dinâmica de paginas e segmentos aos microprocessadores dentro do processo de supervisão de distribuição de tarefas. Deve ser possível a alocação de instruções individuais a múltiplos microprocessadores, fazendo-os operar em rede.
- 13) Supervisionar o sistema de memória virtual controlando os status das tarefas e processos. Interagir com as PMMU com este fim e estabelecer a estratégia a ser adotada em cada processo ou tarefa.
- 14) Estabelecer comunicação com as Placas de Processamento Dedicado supervisionando status e canais de comunicação.

- 15) Receber as solicitações de interrupções externas, classificá-las, detectar se são dependentes de tarefas ou não, analisar sua prioridade e a partir daí destinar as facilidades e recursos circuitais, tudo de acordo com as características particulares da interrupção.
- 16) Esquema de ADM e de transferência em blocos com a MP, PPG de um lado e as placas de processamento dedicado de outro.
- 17) Gerenciar a falta de energia e armazenar todas as informações essenciais em periféricos apropriados. Restabelecer as condições anteriores à falta de energia.
- 18) Operar em esquemas tolerante a falhas com microprocessadores de diferentes PPG operando em paralelo com acesso a diferentes módulos de memória. Em sistemas altamente confiáveis prever mais de uma PPG de modo que estas de auto-testem.
- 19) Gerenciar a memória associativa local, ligando tarefas a recursos de logicionária, circuitaria e microprocessadores, a função do tipo e prioridade da execução. Estabelecer estas prioridades em função do tempo de execução exigido e dos recursos e facilidades disponíveis.
- 20) Supervisionar a memória MAD estática local colocando-a à disposição de tarefas como meio de comunicação inter-tarefa e tarefa-meio externo.

- 21) Alocar parte ou a totalidade dos recursos de circuitaria para o processamento em rede. Nesse caso os microprocessadores executarão simultâneamente a mesma instrução, operando dados colocados préviamente no cache local. Esta execução pode ser vetorial (sequencial) ou não.

Observe-se aqui que dependendo da necessidade de comunicação entre os microprocessadores estem podem executar micro-tarefas (ou grupo de instruções) sob o comando da PGS ao invés de uma instrução de cada vez.

- 22) A PGS pode supervisionar a distribuição de instruções ou páginas de instruções a partir da MP e estas instruções podem ser colocadas nas memórias portas duais das PPG.

- 23) A PGS pode alocar parte ou a totalidade de recursos circuitais e microprocessadores para as operações de bombeamento de dado duto (Pipeline). Nesse caso, o gerenciador pode destinar diferentes segmentos de programas ou diferentes micro-tarefas aos diferentes microprocessadores. Cada processador executará sua tarefa em determinado tempo e as memórias portais-duais operam como acopladores.

- 24) A operação dado duto (pipeline) para a comunicação interprocessadores em diferentes PPG utilizam as memórias portais-duais que, em princípio terão 16 K bytes. Estas memórias interligan-se diretamente com as PPG vizinhas adjacentes. Com

isso é possível destinar no dadoduto tarefas encadeadas com diferentes tempos de execução aos microprocessadores das PPG. Estes operando em cascatas em diferentes tarefas com diferentes tempos de execução vão acumulando dados intermediários nas memórias portais duais inter PPG.

Os tempos ociosos dos microprocessadores que executam suas tarefas mais rapidamente podem ser ocupados por outras tarefas não pertencentes ao encadeamento que forma o dadoduto sobre a gerência da PSG. Com isso é possível casar as velocidades dos diferentes microprocessadores (estágios) sem a introdução de ineficiências.

OPGS supervisiona as memórias portal dual de propagação PPG bem como os resultados intermediários intermicroprocessadores colocados nas memórias cache locais. Em função dos dados e dos resultados na fila é realizada a comutação para novos processos e tarefas

- 25) O gerenciador estabelece aos usuários níveis de operacionalidade e níveis de acesso a recursos e facilidades, através do controle das tarefas e do controle de acesso às placas PPG.
- 26) O PGS possivelmente supervisionará senhas e palavras chaves para o acesso a recursos computacionais e logiciais de todo o sistema.

b) PPG - Placa de Processamento Geral

Detalhes Gerais: As placas PPG conterão provavelmente quatro microprocessadores locais de 32 bits. Os microprocessadores locais com partilham a memória cache de 64 K bytes locais, bem como os processadores de ponto flutuante PFE, a memória de sistema operacional e trabalho, MSOT e a circuitaria que impõem processamento vetorial local aos processadores. O duto interno da PPG opera em alta velocidade, 20 ns de tempo de comutação, comunicando-se com o exterior através do VME dual ou através de memória portal dual. Estão no acesso direto deste duto, tanto a memória portal dual local, como a memória portal dual da PPG n-1, sendo que a PPG n-1 fecha o ciclo com a PPG n.

Os microprocessadores locais da PPG são totalmente autônomos podendo passar do modo usuário ao modo supervisor de forma independente do gerenciador. O PGS tem apenas a função de distribuir, ativar e desativar tarefas segundo os respectivos status e prioridades além de atender funções a elas ligadas, não interferindo com a operação de S.O. que esteja correndo em microcomputadores locais.

Os microprocessadores locais podem gerar interrupções e pedidos de comunicação com tarefas e processos correndo em outros microprocessadores, ou em estado ocioso através de DM ou de interrupções. Estas comunicações ocorrerão através do VME ou através do du-

to DLSI, que engloba as memórias portais duais. Isto ocorre principalmente quando a comunicação é com seus vizinhos adjacentes. De modo geral, estas comunicações serão supervisionadas pelo gerenciador PGS. Na figura 2 temos mostrado um diagrama em blocos das placas PPG.

Aprioristicamente prevê-se as seguintes características e funções para as placas PPG.

- 1) As placas PPG tem um subespaço de endereçamento próprio e pertencem ao espaço de endereçamento geral do sistema e do gerenciador.
- 2) O subespaço de endereçamento de cada PPG permite endereçar suas memórias e seus microprocessadores para comunicações com o sistema.
- 3) São armazenados em memória apenas de leitura MAL (ROM) os programas de inicialização de cada PPG.
- 4) Os microprocessadores locais de cada PPG enviam periodicamente seu status ao gerenciador e quando estão ociosos solicitam tarefas.
- 5) Cada microprocessador dentro de uma PPG possui uma chave além de seu endereço. Esta chave permite o acesso a suas facilidades (Esta chave possivelmente existirá complementando as chaves existentes nos processos)

- 6) As placas PPG são ligadas em cascata através das memórias portal-dual, formando um dadoçuto . Com isso os elementos na memória portal-dual podem eventualmente formar uma fila.
- 7) As memórias portais-dual tem acesso rápido (15 a 20 ns) permitindo dois acessos simultâneos a seus dados. Em princípio cada PPG conterà 16 K bytes de memória portal dual podendo esta capacidade eventualemnte ser ampliada para 64 K bytes.
- 8) O gerenciador pode em situações especiais intervir nas memórias de uma PPG e em particular na memória portal dual.
- 9) Através das memórias portais duais temos um duto extra que denominamos de DLSI. O DLSI permite a comunicação entre PPG adjacentes. Estas PPG tem as memórias portais duais como sub-espaco comum de endereçamento de memória.
- 10) A ultima PPG possivelmente será interligada a primeira formando um anel de comunicações através do DLSI.
- 11) Com a palavra chave conveniente e através dos dutos VME um endereço de uma PPG poderá percorrer sucessivas placas até atingir a memória dual portal desejada.
- 12) O sub-espaco de endereçamento de uma PPG está apresentado do modo tentativo na fig. 4.

- 13) A PPG conterá circuitaria local, cujo status é imposto pela PGS. Esta circuitaria pode interferir de modo independente em cada microprocessador, desenvolvendo operações vetoriais e matriciais, na mesma instrução a UCP, quando esta tentar acessos à memória, a partir de um dado ponto do programa de máquina. Neste ponto a UCP deve saltar incondicionalmente para um dado ponto P_i e aí executar a instrução vetorial ou matricial imposta. Feito isto é imposto um novo salto de retorno à sequência original. Tudo opera como se fosse imposto um salto à subrotina sem salvamentos.
- 14) A circuitaria externa pode também impor à UCP certas subrotinas padrões em casos especiais. Nesses casos especiais englobamos subrotinas de diagnose, operação em rede, e operação em bombeamento.
- 15) O gerenciador supervisiona o estabelecimento de palavras-chaves nas PPG para a comunicação em cada sentido.
- 16) A prioridade de comunicação de uma tarefa com a outra depende da prioridade da primeira tarefa e o estado de execução da segunda. Em caso de necessidade poderá ser enviada mensagem ao gerenciador para priorizar a execução da segunda tarefa em função do atendimento e prioridade da primeira tarefa.

c) PMAD² - Placas de Memória de Acesso Direto Dinâmicas

Detalhes Gerais: As placas de memória de acesso dinâmicas, PMAD² possuem 16 M Bytes, com a utilização de pastilhas de 1 M bits. Estas memórias possuem filas de endereços e um acoplador de memória MAD estática a fim de não reterem o duto de comunicações e permitirem a transferência rápida de dados.

As operações externas de ADM podem ser executadas ciclo a ciclo de relógio, ou então através de transferência maciça de dados.

Além de um acoplador rápido de 16 K bytes (ou mais) as PMAD² possuem circuitaria necessária para as transferências ADM simples e em blocos.

Como organização possível da matriz de memória estamos pensando num arranjo de 1 M x 128 com um byte a mais de redundância. Com isso temos distância Hamming 4, correção automática de um erro e detecção de presença de dois. A fig. 5. mostra-nos um diagrama em blocos das placas PMAD².

Aprioristicamente preve-se as seguintes características e funções:

- 1) Correção automática de um erro e detecção de dois.
- 2) Cada placa PMAD² conterà 16 M bytes de memória dinâmica de acesso direto, 16 K bytes de acoplador rápido e suportará transferências ADM monociclo e em bloco.
- 3) As placas PMAD² sinalizarão o PGS no caso de detecções de erros.

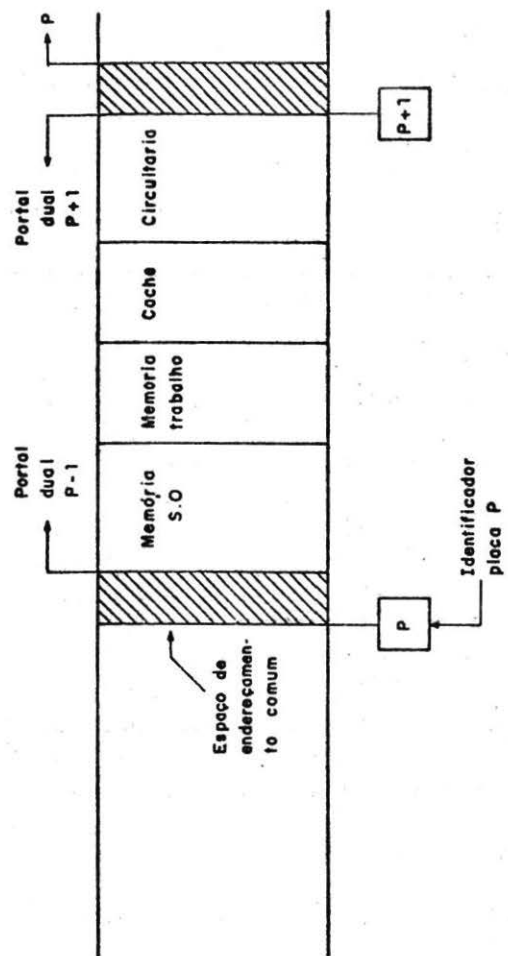


Fig 4 Espaço de endereçamento PPG

- 4) As placas PMAD² constituem a memória principal do sistema que poderá atingir 128 M bytes (9 placas de 16 M bytes com pastilhas de 1 M bit).
- 5) Cada placa PMAD² poderá ser acessada de maneira independente num esquema multiporto.
- 6) Cada PMAD² poderá permutar dados em um ciclo ou envia-los em descarga tanto através de dutos VME como através do duto VSB.

d) PDD - Placas de Processamento Dedicado

As placas de processamento dedicado tem a função de realizar todas as operações de E/S do sistema de multiprocessamento. Possuem uma UCP própria dedicada ao estabelecimento de protocolos de comunicação com o meio externo. Prevemos dois tipos de placas:

d.1. Placa de Processamento de Comunicação - PCC

Descrição Geral: Esta placa é responsável pelo processamento da entrada e saída de caracteres, permitindo em cada seção o controle de até 64 terminais assíncronos.

Possuem de cada seção um microprocessador 68010 operando em 10 MHz. A comunicação entre o sistema e a placa é efetuada através de uma memória porta-dual. Os processadores locais podem também acessar a memória principal. A interface com os terminais é efetuada através de placas padrão EURO CARD duplo instaladas num barramento Canal E/S Motorola. Na fig. 6 temos representado um diagrama em blocos de tal placa:

Fig. 6 - Diagrama em blocos da placa de processamento de comunicações PCC.

d.2. Placa de Controle de Dispositivos de Armazenamento em Massa - PCDAM

Descrição Geral: Esta placa efetua o gerenciamento de dispositivos de armazenamento em massa e operações de entrada e saída bloqueadas. A placa possui até 4 seções sendo baseada em processadores 68010 de 10 MHz, responsáveis pelo atendimento de E/S em bloco utilizando algoritmos de escalonamento e "caching" de discos. A comunicação com os dispositivos periféricos de armazenamento em massa é feita através de interface SCSI (Small Computer Systems Interface Ansi X3 T9). Os processadores podem acessar a memória principal do principal através de interface VME e podem realizar transferências

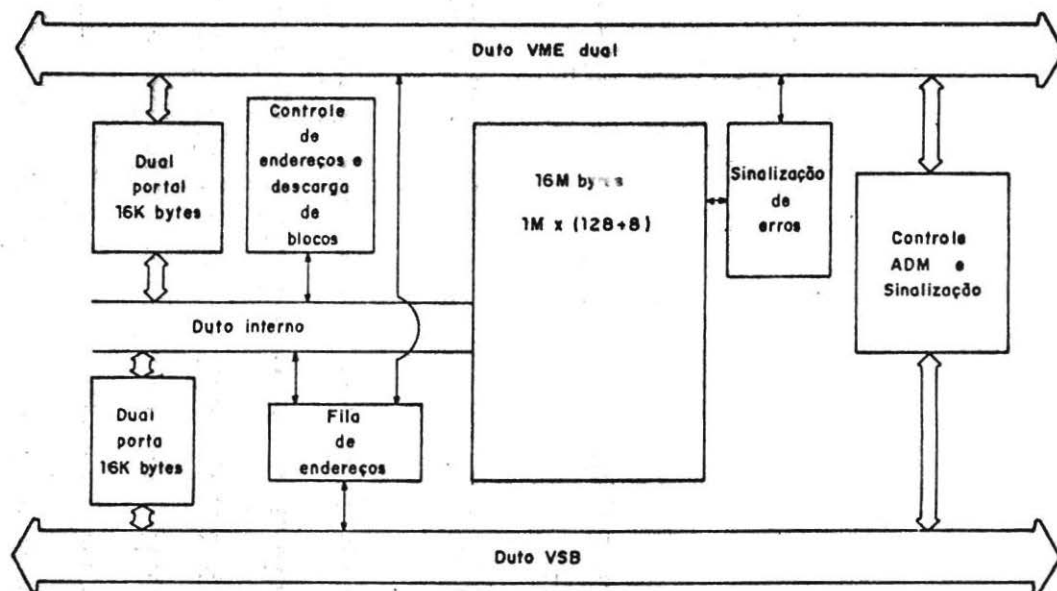


Fig. 5 Placa de Memória de Acesso Direto Dinâmica PMAD². Organização possível de $1M \times (128+8)$ sendo um byte de redundância por endereço. Com isso temos distância Hamming 4, Correção automática de 1º erro e detecção de 2

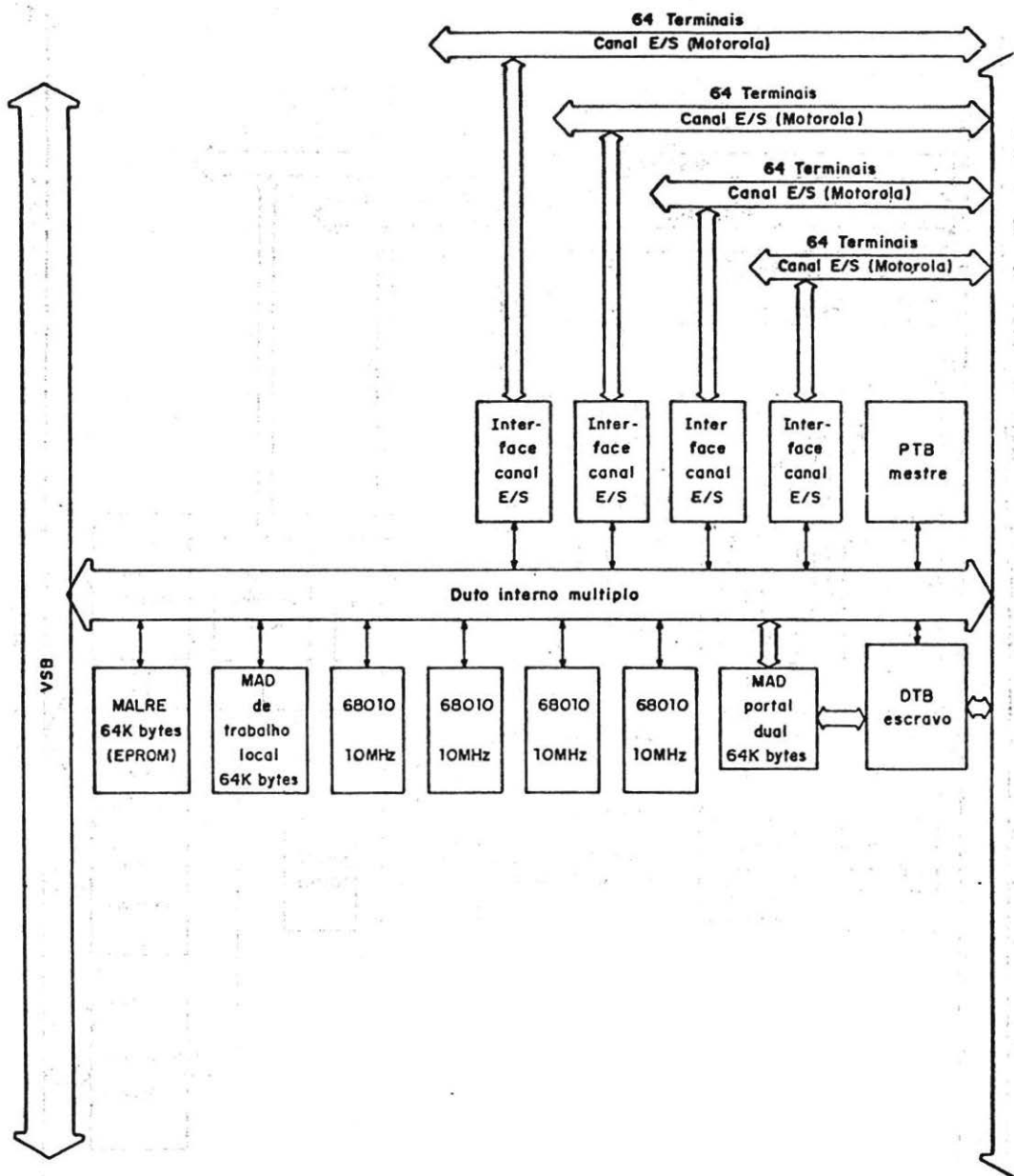


Fig. 6 Placa de controle de comunicações

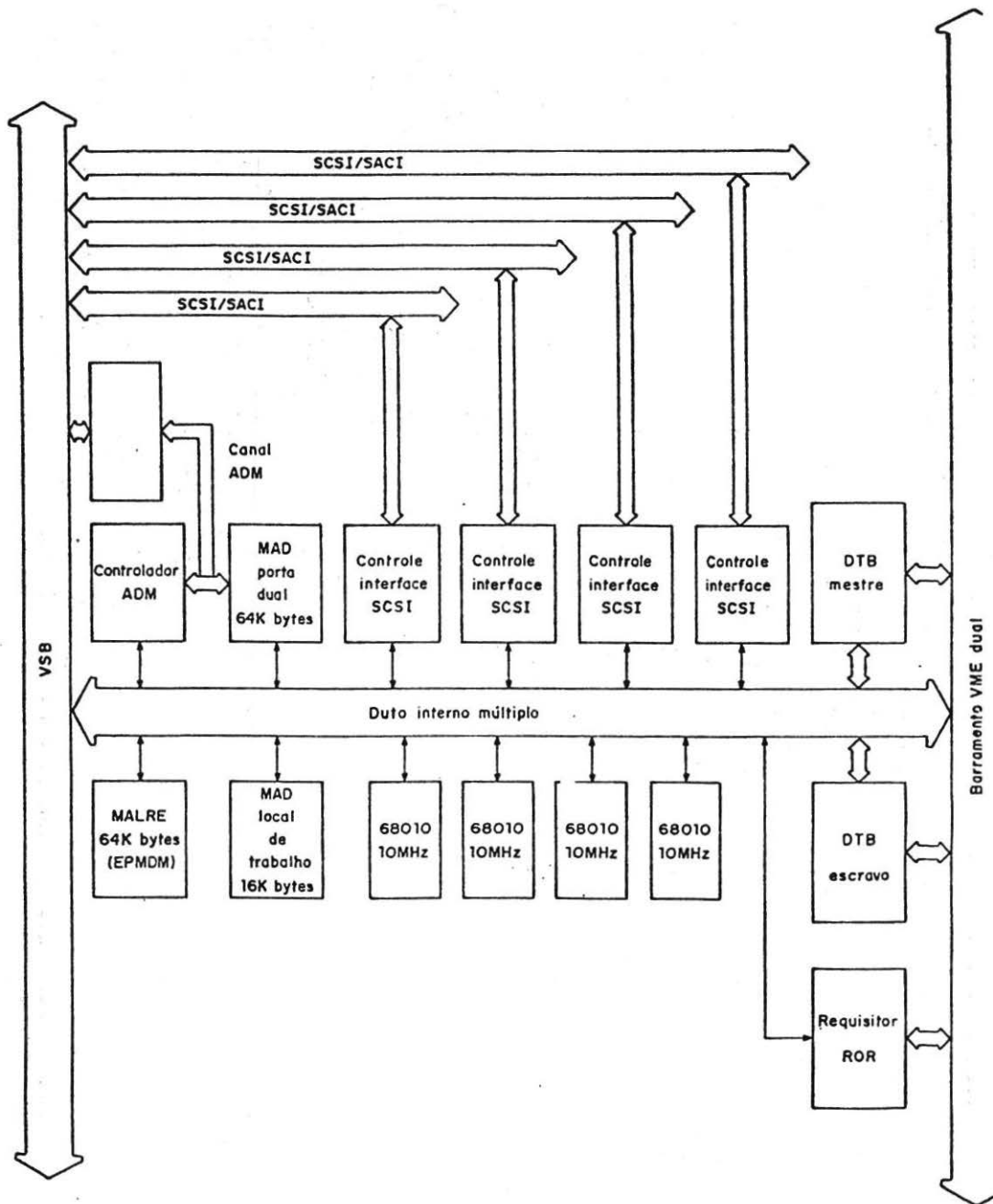


Fig. 7 Placa de controle de dispositivos de armazenamento

ADM bloqueadas ou não através de interface VSB. na fig. 6 temos mostrado um diagrama em blocos de tal placa.

Fig. 7 - Diagrama em blocos de placa de controle de dispositivos de armazenamento em massa.

IV - CONSIDERAÇÕES GERAIS SOBRE CIRCUITARIA

O diagrama adotado utiliza grupos de microprocessadores colocados em placas de processamento geral PPG ou em placas de processamento dedicado. Espera-se que a utilização de memórias de S.O. e memória cache reduza sensivelmente a necessidade de dutos externos às placas. De qualquer forma foi previsto um duto VME dual para comunicação com a memória principal e um duto DLSI para a comunicação interplacas adjacentes. Na comunicação com o meio externo preve-se um duto VSB com capacidade potencial de comunicação de 64 M bytes/s. Toda a circuitada

ria apresentada é tentativa, podendo sofrer alterações profundas no decorrer do projeto.

Todavia a experiência passada do LSI indica que a organização apresentada tem grande chance de vir a ser implementada e apresentar um desempenho que possa ser considerado satisfatório.

V - LOGICIONARIA

O sistema é um sistema projetado para vir operar em multiprocessamento, sendo o gerenciador PGS encarregado de realizar todas as tarefas de multiprocessamento supervisionando o acesso privilegiado e a proteção de dados. Embora a organização seja voltada para sistemas S.O. UNIX compatíveis, a logicionária UNIX deve fornecer apenas as tarefas a serem executadas em termos de prioridades relativas. A PGS deverá etiquetar associativamente estas tarefas e as encaminhar ao microprocessador apropriado.

Agradecemos ao Professor Sérgio Takeo Kofugi pelas inúmeras discussões sobre o texto.

JAZ/n

22 de abril de 1987