# The Vector Coprocessor Unit (VU) for the CM-5

Jon Wade, Zahi Abuhamdeh, Dan Cassiday, Mahesh Ganmukhi,
Jeff Hill, Bob Lordi, Margaret St.Pierre, Scott Smith, Monica Wong,
Shaw Yang, Bob Zak: Thinking Machines Corporation

David Bural, Don Steiss, Bob Farrell, Simon Koval, Maria Gill,
Ken Cyr: Datapath VLSI Products Design, Texas Instruments Inc.

Hot Chips IV Symposium

August 11, 1992



# Presentation Outline

- **Design Objectives**
- **Architecture**
- **Implementation**
- **Performance**

USES SPARC µProc

# Design Objectives

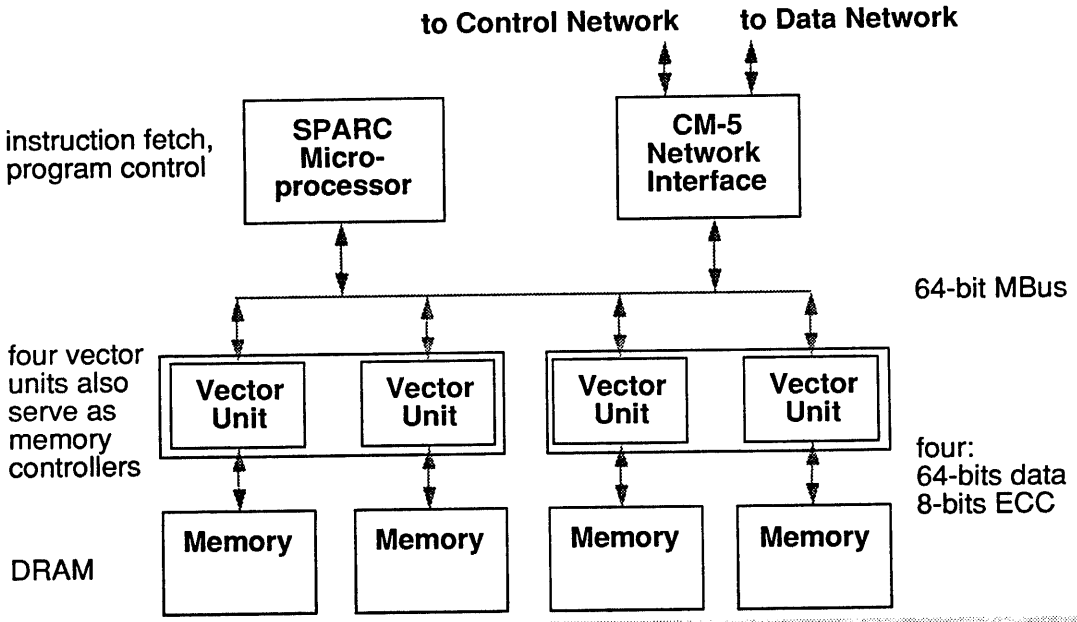
- Maximum sustained MFLOPS/$
  - * memory bandwidth — - direct fast-page DRAM support, strided and indirect memory & register addressing
  - * FLOPS (IEEE) — - high performance FPU
  - * instruction bandwidth — - vectorized instructions
  - * low startup cost — - minimize bubbles in pipeline
- System-level diagnostics
  - * test/debug support — - JTAG, full internal scan



# Design Objectives (cont.)

- System Configurability
  - * support future SPARCs — - use MBus interface
  - * memory size — - 4 to 64 Mbit DRAMs
  - * expandable — - 2 to 8 VUs per node
- Software
  - * compilers & libraries — - support data parallel programming model
- Time to Market
  - * existing technologies — - TI EPIC-2 process, CPGA, DRAMs
  - * existing designs — - TI megacells, gate-array
  - * thorough verification — - transistor, gate, arch & system level

# CM-5 Processing Node with Vector Units






# Execution Model

- memory mapped into MBus address space
- remote accesses
  - * memory
  - * Vector Unit state (register file, control state)
- instructions
  - * may be issued to one, a pair or all VUs on a node (decoded from MBus address)
  - * load/store memory base address is decoded from MBus address
  - * instruction is decoded from MBus write data

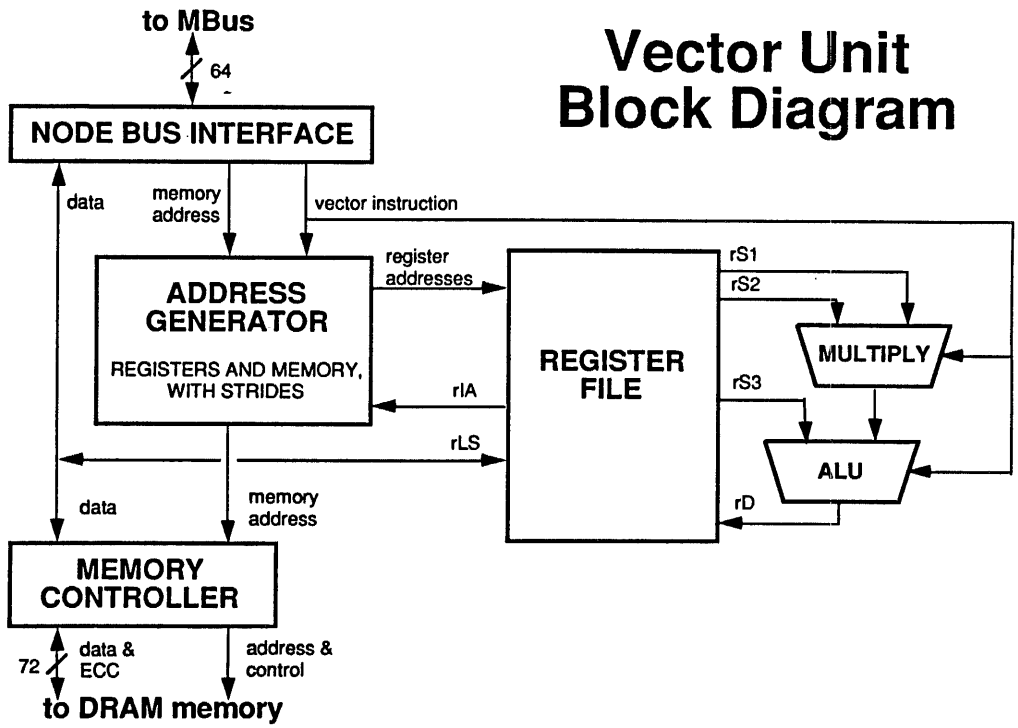
# CM-5 Processing Node Address Mapping

| name | lower | upper |
|---|---|---|
| main memory | N 0000 0000 | N 07FF FFFF |
| data registers | N 4000 0000 | N 4000 01FF |
| instruction register | N 8000 0000 | N 87FF FFFF |
| read only memory | F FN00 0000 | F FN7F FFFF |
| control registers | F FN80 0000 | F FNFF FFFF |
| NI registers | 0 0800 0000 | 0 080F FFFF |

where N is the VU ID number (0 to 7, 8 for common space).



# Vector Unit Block Diagram

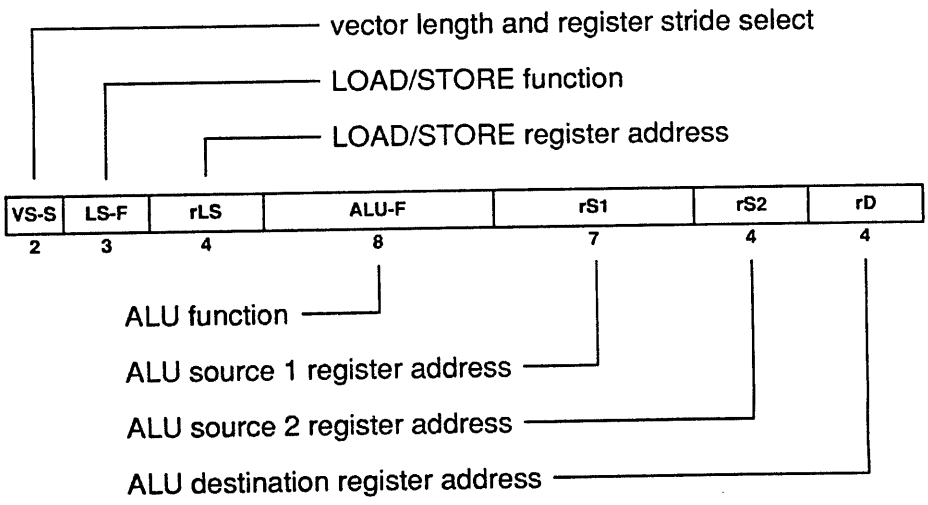


# Vector Unit Instructions

- load/store architecture (load chaining)
- concurrent arithmetic & load/store operations
- single & double-precision FP, 32- & 64-bit integers
- floating-point (div, sqrt), integer & logical operations
- triadic multiply-adds (floating-point & integer)
- vectorized instructions with strided or indirect addressing of memory & register file, vector lengths from 1 to 16
- elemental masking of load/store & arithmetic functions



# Vector Instruction Format: Short



# VU Register File Addressing

- **Vector Registers**
  **e.g. 4 vector registers: 16 x 64-bit, 16 x 32-bit elements**
  **8 vector registers: 8 x 64-bit, 16 x 32-bit elements**
  **16 vector registers: 4 x 64-bit, 8 x 32-bit elements**

- **Arbitrary Base Address & Striding**
  **wrap around ends of register file, negative striding**

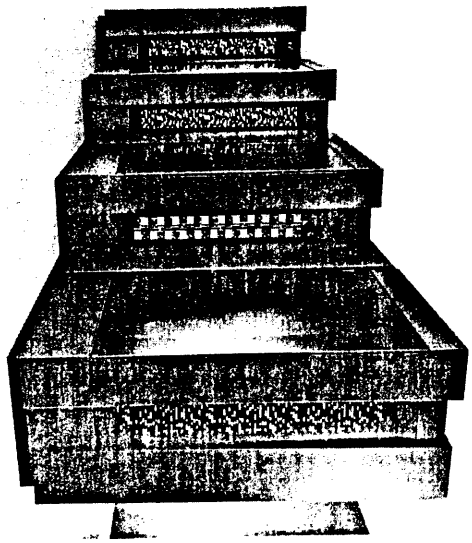
- **Indirect Addressing for Operand-1**
  **arbitrary address sequence**

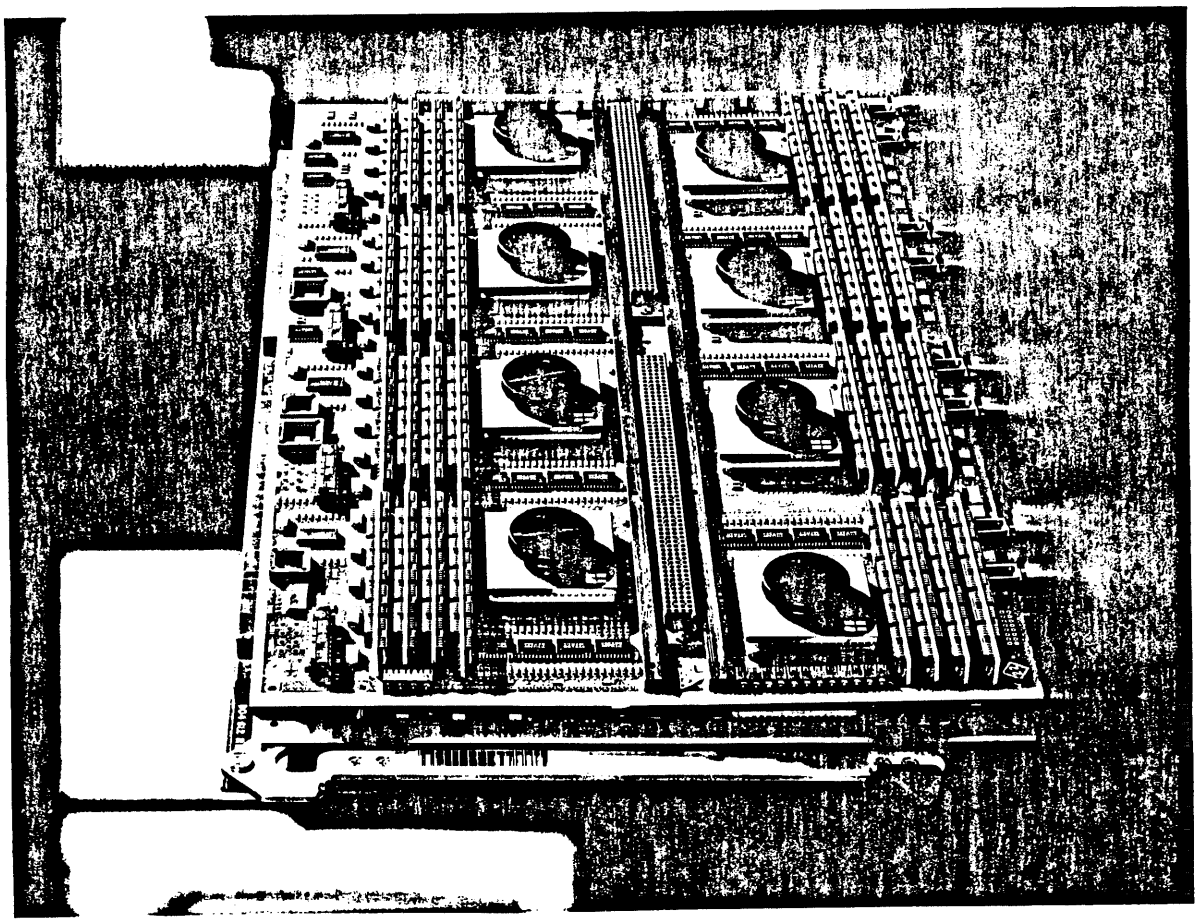


# Vector Instruction Format: Long

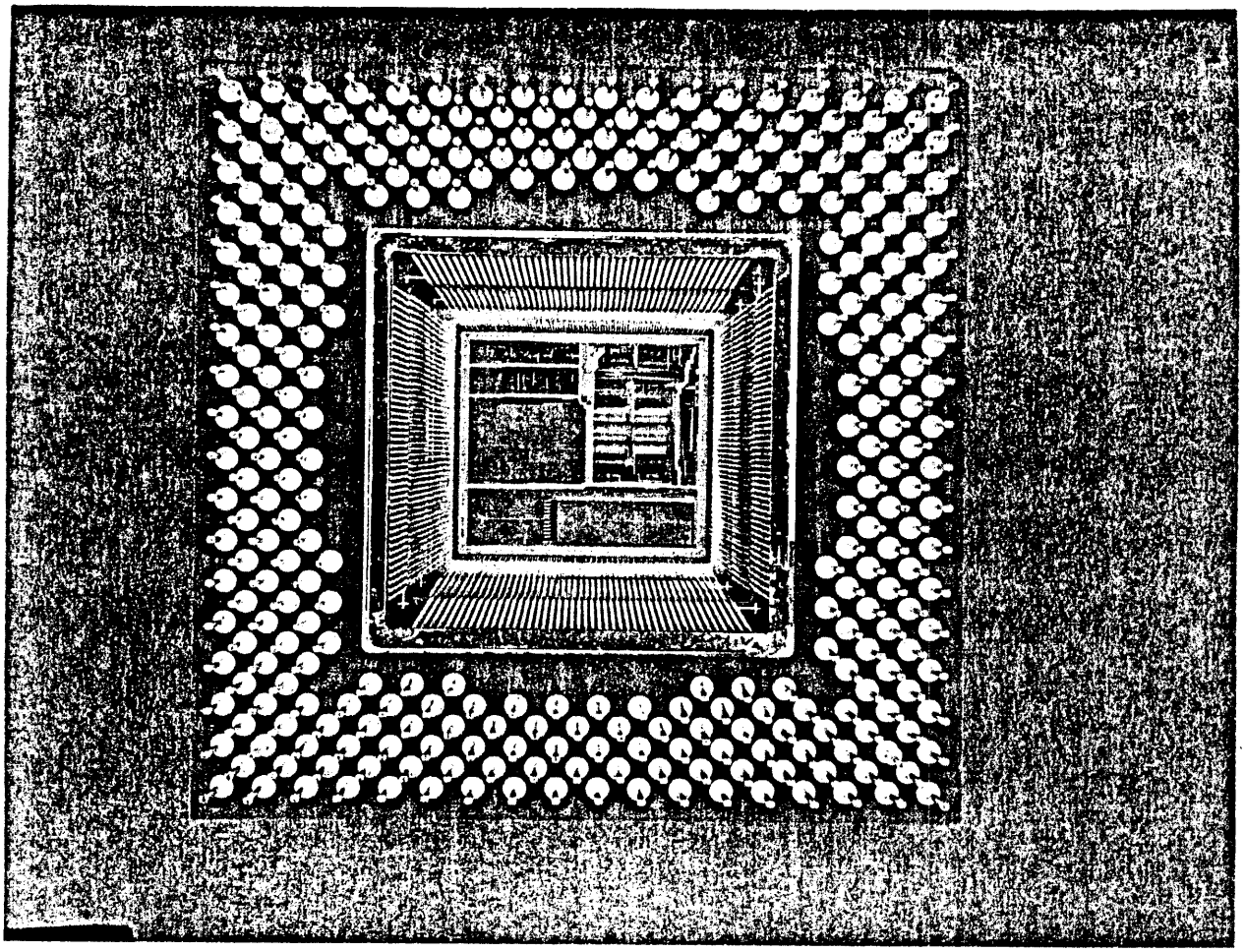
**Long format extends the 32-bit short format with an extra 32 bits. This is further decoded to specify these operations and modes:**

- immediate operands (integer or floating-point)
- arbitrary memory stride
- arbitrary register base addresses and strides
- indirect addressing of memory & register file
- change default vector length
- vector mask conditionalization control
- arithmetic result exchange between paired vector units





6.1.7



# Technology Summary

| | |
|---|---|
| Number of Transistors | 1.0M |
| Die Size | 14.7 x 15.8 $mm^2$ 550 X 620 MILS |
| Package | 319 Pin CPGA |
| Operating Frequency | 40MHz Worst Case |
| Power Dissipation | 5.0W (typical @ 40MHz) |
| Process Technology | 0.8um CMOS, metal pitch<br>1st, 2.0um; 2nd, 2.0um |

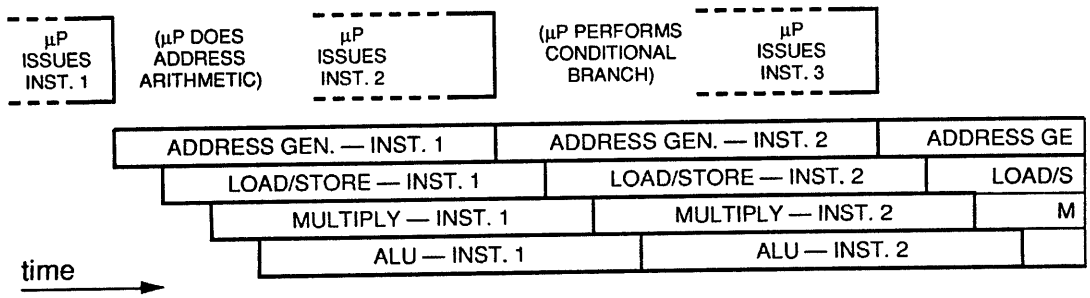
# Chip Features (2 Vector Units)

| | |
|---|---|
| • Peak FP performance | 80 MFLOPs (IEEE SP, DP) |
| • MBus bandwidth | 320 MB/s |
| • DRAM bandwidth | 320 MB/s |
| • Register file | 128 x 64-bit, 256 x 32-bit, 5-read, 3-write ports |
| • DRAM support | 16 MB to 256 MB |



# Vector Instruction Pipelining

The node microprocessor (μP) issues vector instructions. The timing need not be exact. The μP overlaps scalar control computations with vector processing.




Consecutive vector instructions overlap head-to-tail. After the first vector, vector startup time is zero when vector instructions are issued quickly enough.

# Benchmark Test Results

**This slide will list the performance of the vector unit for several numerical benchmarks.**

| | MFlop @ 32MHz | % Peak |
|---|---|---|
| LL #1 | 48 | 75% |
| LL # 7 | 61 | 96 |
| MATMULT | 58 | 90 |
| MATMULT (VOCSZA) | 58 | 90 |
| FFT RADIX-2 (128 COMPLEX) | 40 | 62 |