1991

MULTICHIP MODULE WORKSHOP

EXTENDED ABSTRACT VOLUME

March 28th - 29th, 1991

Dream Inn
Santa Cruz, California

Sponsored by IEEE Computer Society

Supported in part by the National Science Foundation

Impact of MCM's on System Performance

**Ayman I. Kayssi, Karem A. Sakallah, Richard B. Brown,
Ronald J. Lomax, Trevor N. Mudge, and Thomas R. Huff**

Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109

1 Introduction

As the performance goals of computers increase, packaging techniques become critical for reducing the delays that result from chip crossings and interconnect among chips. Packaging technology has developed to the point where it is now possible to integrate a processor and cache onto a single Multi-Chip Module (MCM) of a few inches on a side. MCM's offer the following advantages over Printed Circuit Boards (PCBs). First, they eliminate one level of packaging, which also eliminates the parasitic delays, and improves overall reliability. Second, they allow chips to be placed closer together, improving the area utilization, and decreasing the length of the interconnect lines between the chips. Finally, denser lines can be used on the MCM, because the line width, pitch, and thickness on MCMs are much smaller than those on PCBs. All of these factors contribute to an improvement in signal delay times on the MCM substrate as compared to a PCB, which directly affects system performance. Simulations of a GaAs microcomputer currently under development at the University of Michigan [1] show that using an MCM for packaging improves system performance by approximately 50% as compared to a PCB, due to a factor of three decrease in the signal propagation delay from CPU to memory.

This paper describes the effect of using different MCM technologies on the optimized performance of the GaAs Microcomputer. This high-performance microcomputer, which implements the MIPS RISC architecture [2], is designed using GaAs direct-coupled FET logic (DCFL) technology form Vitesse Semiconductor Corporation. The system components mounted on the MCM are the CPU and floating-point accelerator, the primary instruction- and data-caches, and a memory management unit (MMU).

A multilevel optimization procedure similar to that described in [3] is used in Sec. 2 with actual MCM layouts. The effects of the dielectric constant of the insulator and of the die attachement method are described in Sec. 3. Concluding remarks are made in the last section.

2 Cache Size Optimization

The optimization of the size of the instruction cache (I-cache) is done at multiple levels to maximize the *MIPS* rating of the GaAs microcomputer. The *MIPS* figure (millions of instructions per second) is a suitable measure of system performance, and is defined as

$$MIPS = \frac{1000}{T_c \times CPI}$$

where T_c is the processor clock period in nanoseconds, and *CPI* is clock cycles per instruction for a representative set of application programs. Both of these quantities vary with, among other things, the size of the I-cache.

This variation of T_c and *CPI* with I-cache size was obtained after the MCM layouts were generated for 2 K, 4 K, 8 K, and 16 K I-cache sizes. For each cache size, circuit simulation was used to obtain the

2.1 MCM Layout Generation

The Floorplanner place and route tool, from Seattle Silicon Corporation (SSC), was used with NetEd and SymEd, from Mentor Graphics, to perform MCM placement and routing. This procedure starts by creating a ruleset for the MCM process using the SSC makerule utility, then a footprint for each chip, containing pad and port information, is created with the SSC Compiler Development System. SymEd and NetEd are used at this point to define symbols and connectivity of the chips; the resulting NetEd schematic is then converted to the SSC database by running SSC routexpand, schem2net, and net2geo. Floorplanner can now automatically place and route the design, a process which should be repeated after sizing power and ground using SSC pdabs.

MCM layouts obtained in this fashion are used to extract the line lengths that are used for circuit simulation. The layouts for the 4 K I-cache and 8 K I-cache cases are shown in Figs. 1 and 2.

2.2 Circuit Simulation

As detailed in [3], the clock period T_c is

$$T_c = \frac{1}{2}(T_{MCM} + T_{CPU} + T_{MEM})$$

where T_{CPU} is the critical delay through the processor chip, T_{MEM} is the access time of the cache, and T_{MCM} is the roundtrip signal propagation delay on the MCM, including input and output buffers on the CPU and cache chips. The worst-case delay through the processor chip, as simulated in SPICE, is T_{CPU} = 3.0 ns, while the access time for the memory, made of custom GaAs 1 K × 32 bit SRAM chips, is expected to be T_{MEM} = 3 ns.

T_{MCM} was obtained from the actual MCM layouts; the interconnect lines on the MCM are modeled by lossy transmission lines. The values of R, L, and C per unit length were calculated for a typical Polycon process (assuming a benzocyclobutene dielectric) to be R=1070 Ohm/m, C=108 pF/m, and L=272 nH/m. Output pad transmitters include active pull-up and pull-down with no line terminating resistors, as the resistive loss along the line was high enough to adequately suppress line reflections. The input pad receivers consist of two voltage limiting diodes and one pass transistor. The die-to-substrate bonding parasitics, assuming flip-chip attachement, are modeled by a low-pass pi-section with two capacitances and one inductance equal to 0.548 pF, 0.353 pF, and 0.1nH, respectively.

The resulting circuits for a one bit line going from the CPU to memory and back were simulated in HSPICE (from Meta-Software), using their built-in lossy transmission line model, for each I-cache size considered. Time delays were calculated at the $V_{in} = V_{out}$ point for a GaAs DCFL inverter (410 millivolts for a 2-Volt power supply.) The resulting MCM roundtrip delays are plotted in Fig. 3, and the corresponding clock periods are plotted in Fig. 4. These plots clearly show how the MCM delay, and therefore, processor clock period grow with larger cache sizes.

2.3 Architectural Simulation

The clock cycles per instruction (*CPI*) metric was obtained for various I-cache sizes from the cache simulator `cacheUM`, developed at the University of Michigan. Figure 5 shows the variation of *CPI* with cache size and clock period. The cache simulations were performed for sixteen benchmarks reflecting the typical workload of an engineering workstation [3].

2.4 Optimization Results

For each I-cache size, the *MIPS* figure can now be obtained from T_c and *CPI* . Figure 6 predicts that the maximum performance, a *MIPS* rating of 155.9, will be achieved with an I-cache size of 8 K. Results for the 4 K cache size are virtually the same, having a *MIPS* rating of 155.2.

In the rest of the paper, we investigate the effect of the MCM on this small difference in performance between 4 K and 8 K I-caches.

3 Effects of MCM Parameters

We studied the effects of two parameters on the *MIPS* rating of the 4 K and 8 K I-cache sizes. The first one is the dielectric constant of the insulator. The relative dielectric constant is varied from 2.56 to 4.5, to include the range of all polymers available as insulators. This is electrically equivalent to varying the capacitance per unit length of the transmission line from 108 pF to 188 pF. The second variation was the die attachement method, which was assumed to be flip-chip, TAB, or wire-bonding. This can also be modeled by changing the values of the low-pass pi-section components at the transmitters and receivers, to reflect the different parasitics associated with TAB and wire-bonding as compared to flip-chip. The 0.353 pF capacitance and 0.1 nH inductance values were changed to 0.83 pF and 1.0 nH for TAB, and to 1.4 pF and 1.2 nH for wire-bonding.

Using these ranges of values for circuit elements, HSPICE simulations gave a new set of results for T_{MCM} and T_c. Figure 5 was still used for the *CPI* numbers. The resulting *MIPS* ratings for all six combinations of I-cache sizes and die attachement method versus relative dielectric constant are plotted in Fig. 7. As seen in the plot, the *MIPS* rating in the case of 8 K is always superior to that of 4 K when using flip-chips; however, this is not true for TAB or wire-bonding, where the relative 4 K performance improves as the dielectric constant increases. The break-even point between 4 K and 8 K with TAB is at $\epsilon_r = 3.9$, while for wire-bonding, it occurs at $\epsilon_r = 2.8$.

4 Remarks and Conclusions

The difference in performance between a 4 K I-cache and an 8 K I-cache is small for advanced flip-chip technology with a low dielectric constant insulator. Therefore, other criteria for choosing the I-cache size are used. First of all, the number of memory chips for an 8 K I-cache is twice that of a 4 K. This not only occupies more of the MCM area, but also implies that the power dissipated in the I-cache will be doubled for the 8 K cache. With more power dissipated, the substrate temperature increases, thereby increasing the temperature of all MCM-mounted integrated circuits and the bonds between ICs and substrate. Higher temperatures increase the thermal shear stresses due to the thermal coefficient of expansion mismatch, decrease noise margin, and decrease system reliability and MTBF. Since power supply current increases with the number of memory chips, larger power and ground buses should be used in the 8 K cache, again occupying more area on the MCM. Signal wave shape becomes more distorted as the line length and the number of loads on the line increase.

In conclusion, a combined architectural-electrical analysis showed that an 8 K I-cache offers only marginally better performance than a 4 K I-cache. Further analysis showed that for certain MCM technologies, a 4 K cache has the best performance. This led to an arbitration at the physical level of the MCM, where smaller caches have several advantages.

References

[1] T. N. Mudge, R. B. Brown, W. P. Birmingham, J. A. Dykstra, A. I. Kayssi, R. J. Lomax, O. A. Olukotun, and K. A. Sakallah, "The Design of a Microsupercomputer," *IEEE Computer*, pp. 57–64, January 1991.

[2] G. Kane, *MIPS RISC Architecture*, Prentice-Hall, 1989.

[3] O. A. Olukotun, R. B. Brown, R. J. Lomax, T. N. Mudge, and K. A. Sakallah, "Multilevel Optimization in the Design of a High-Performance GaAs Microcomputer," *IEEE JSSC*, May 1991.

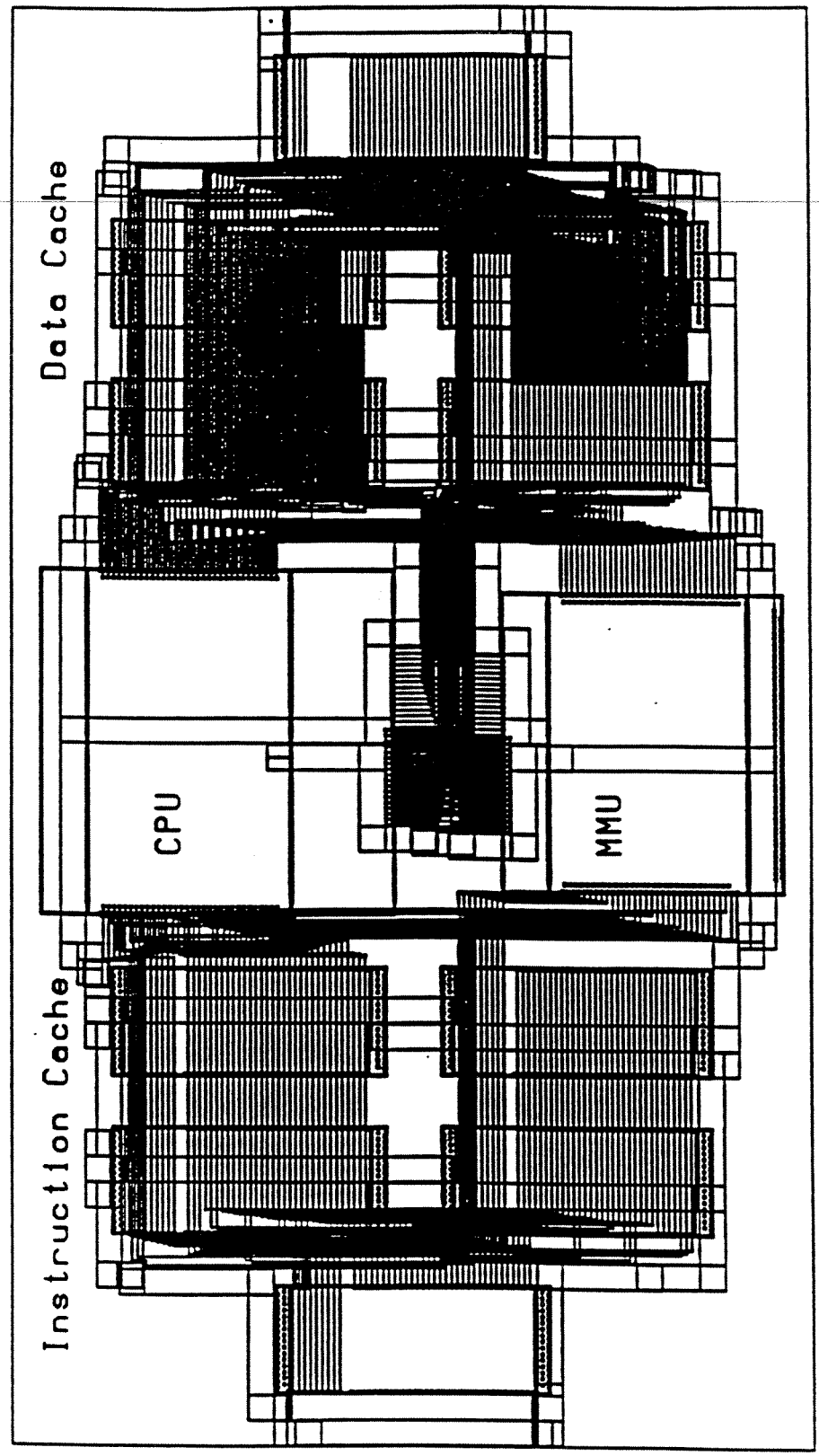

Figure 1: MCM layout with a 4 K I-cache.

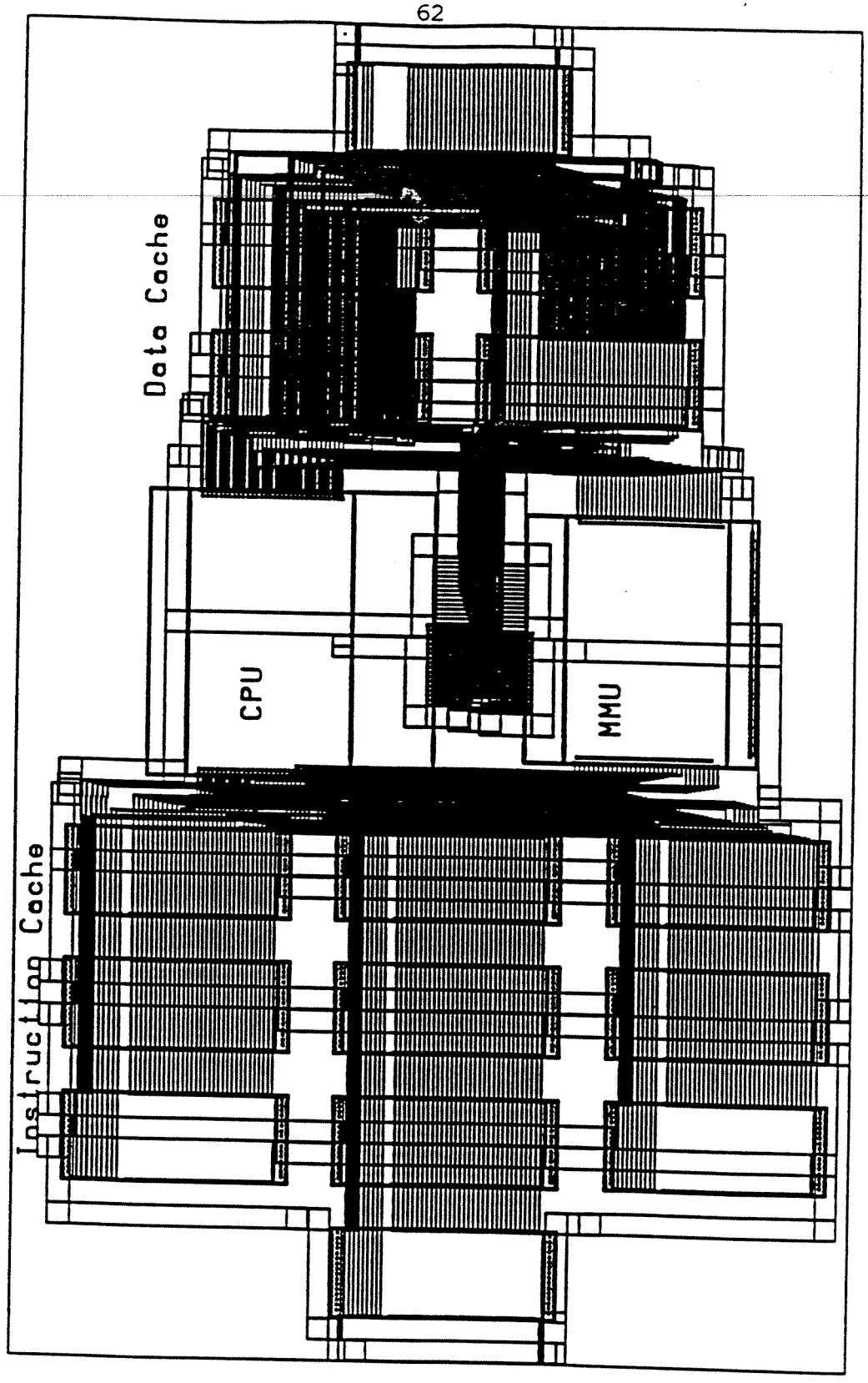

Figure 2: MCM layout with an 8 K I-cache.

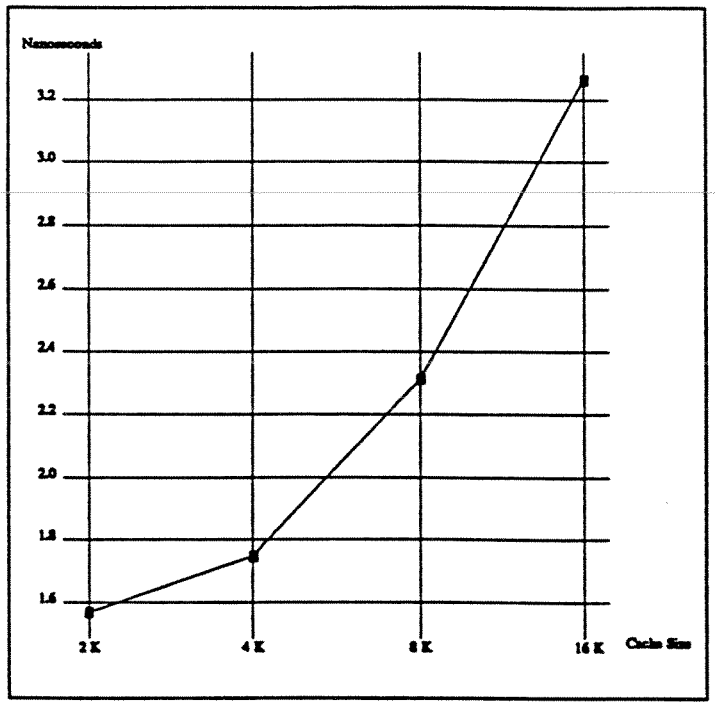

Figure 3: MCM delay variation with I-cache size.

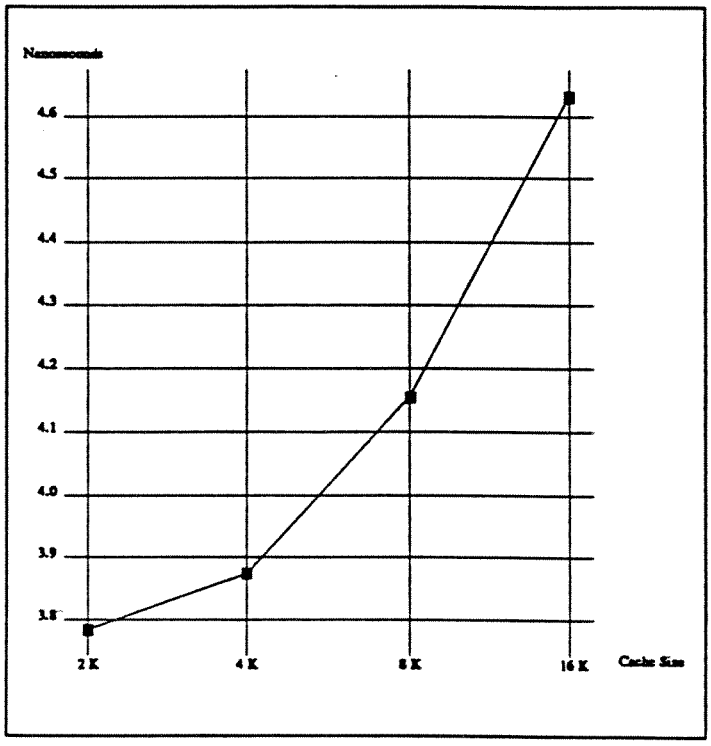

Figure 4: Clock period variation with I-cache size.

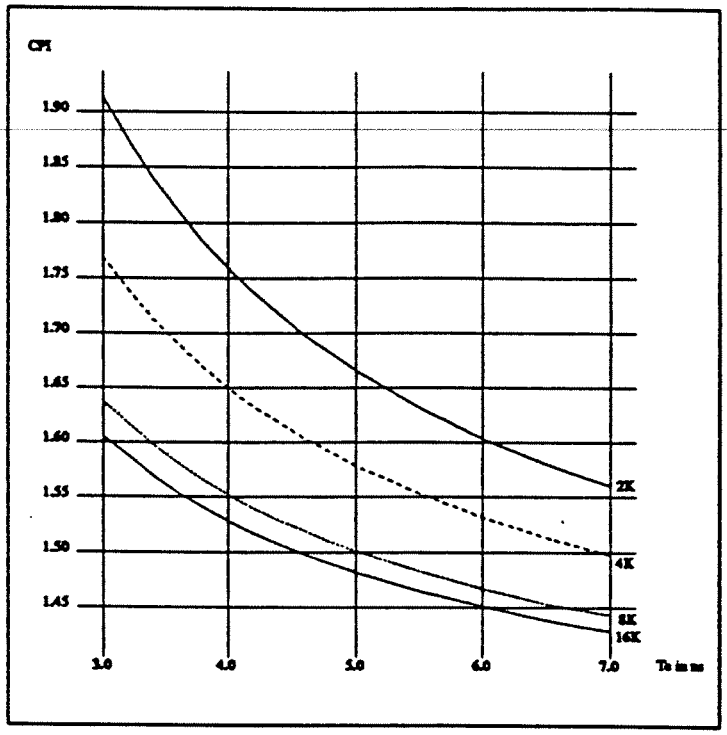

Figure 5: Variation of *CPI* with I-cache size and clock period.

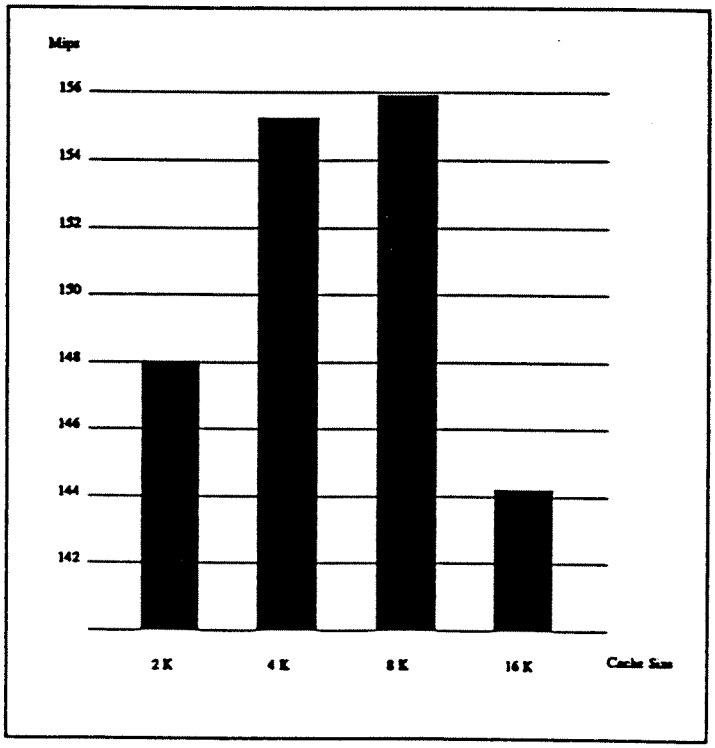

Figure 6: *MIPS* variation with I-cache size.

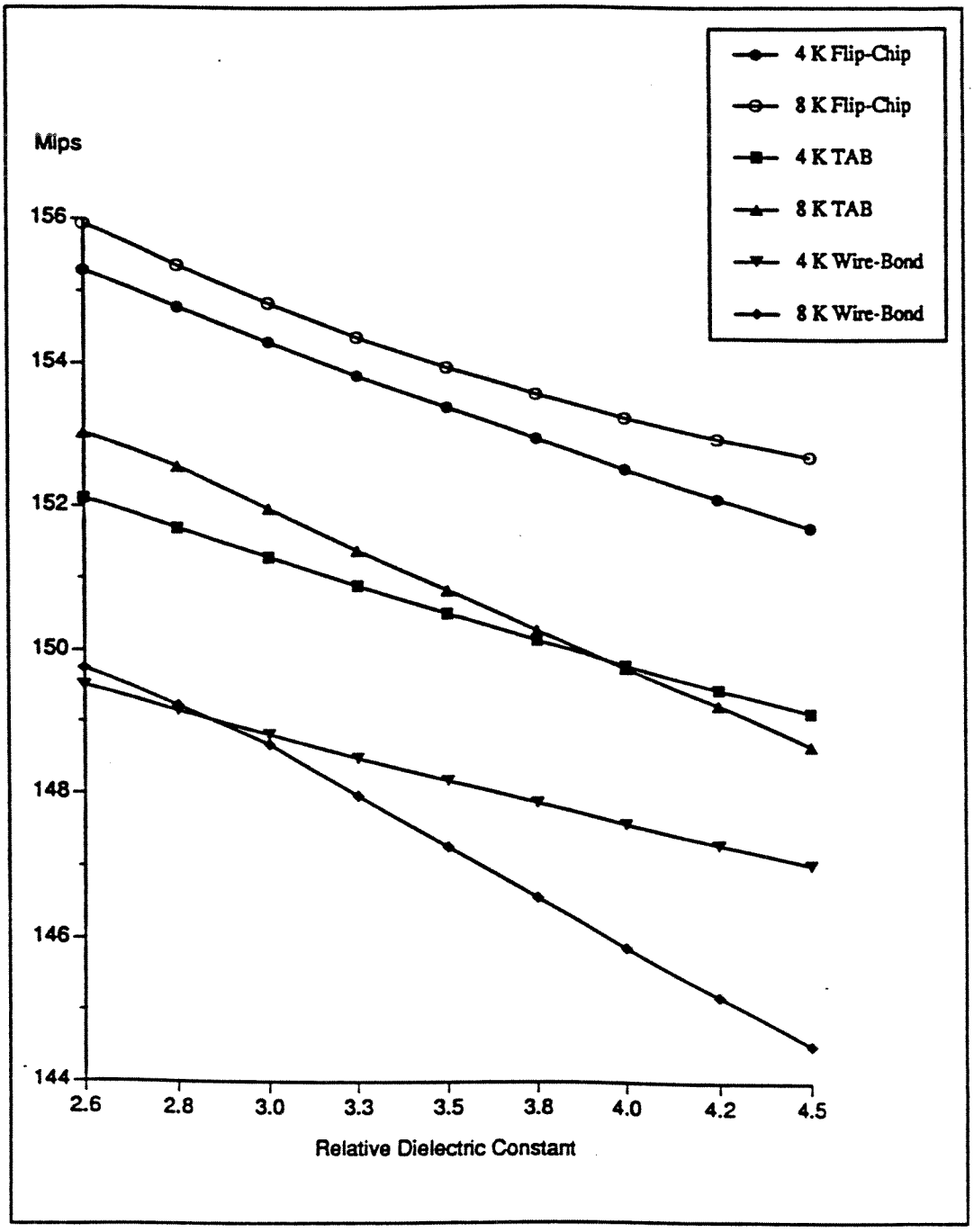

Figure 7: Impact of different MCM technologies on system performance.