

C.mmp: THE CMU MULTIMINIPROCESSOR COMPUTER

Requirements, Overview of the Structure  
Performance, Cost and Schedule

C. G. Bell, W. Broadley, W. Wulf, A. Newell

and

C. Pierson

R. Reddy

S. Rege

Department of Computer Science

Carnegie-Mellon University

Pittsburgh, Pennsylvania

August 24, 1971

This work was supported by the Advanced Research Projects Agency of the Office of the Secretary of Defense (F44620-70-C-0107) and is monitored by the Air Force Office of Scientific Research.

This document is still a working document and may not be copied, or referenced without the premission of the authors.

## I. INTRODUCTION

This document proposes a CMU multiprocessor system constructed around a set of PDP11 computers connected through a crosspoint switch to a large sharable primary memory. The system is to operate as a third node in the existing Computer Science computer system, along with two existing PDP10's.

The present proposal constitutes a solution to a specific set of needs existing in our environment. It replaces a prior planned solution, which consisted of a collection of individual stand-alone PDP11 systems and PDP11 I/O processor systems situated in front of the PDP10's. It is not the only system that satisfies our requirements, as the existence of the prior plan indicates. However, it provides a highly effective solution and does so within the cost framework with which we have been working.

The proposed system has research consequences that reach well beyond the particular demands it was designed to satisfy. For although multiprocessors have been much talked about and advocated, there are remarkably few operational systems more complex than dual-processor systems, and even fewer documented scientific investigations into their performance and operating structure. Thus, the proposed multiprocessor system offers substantial opportunities for significant research that is consonant with our research interests and capabilities.

This document is limited to a presentation of the proposed system. It gives enough description of the usage requirements and the research potentials and problems to make clear why we believe the effort to be a sound one. It does not attempt a systematic discussion of the field of multiprocessor research, nor of alternative systems that might be of interest, either to meet our computing demands or as research directions.

Section II discusses the requirements and research potential. Section III lists the design constraints adopted. Section IV lays out the PMS structure

of the system, giving the various components which have to be fabricated, their gross specification and the design decisions remaining. Section V describes the main specifications of the operating system. Section VI provides some details on a performance analysis. Section VII gives schedules and costs.

## II. BACKGROUND ON REQUIREMENTS AND RESEARCH POTENTIAL

The CMU multiprocessor project is directly responsive to two requirements:

1. The need for particular computational facilities.
2. A research interest in computer structures.

The design itself may be viewed as attempting to satisfy the computational needs with a system that is conservative enough to ensure successful construction in an immediate time frame. Within this constraint, the system appears to be a research vehicle of considerable potential, both directly in terms of research on multiprocessor systems and in its ability to support a wide range of important investigations in computer design and systems programming. Beyond this, the system will produce a significant amount of information processing capability, thus satisfying an important side condition that the funds we have available to us for computing facility provide enough general power for the continued growth of the computer science community at CMU.

The co-existence of usage and direct research interests raises the (not unknown) spectre of conflicts between these interests in terms of stability and availability. Consistent with the view that this system is to provide computing facility, such conflicts, should they arise, will be resolved in favor of the users of the system.

All of our interests in the system are computer research interests. The users of the system are engaged in varieties of computer science research not significantly different from the direct research into multiprocessor hardware and software. (Indeed, they are often the same people, wearing different hats). The existence of the system will generate new research not now being done in the environment, some of which we can quite clearly see and only part of which will be direct research on the system. When these additional efforts come into existence they will have as great a claim as present projects to the use of

the system and will make as great contributions to justifying the system as satisfying their computation requirements. The temptation is to lump all these research efforts, direct and indirect, pre-existing and potential, into a single collection for the purposes of justifying the design. Instead we will preserve some fidelity to history, describing the requirements in terms of (1) usage by those research efforts who had already laid a claim for PDP11-style resources; (2) potential research efforts and (3) research into multiprocessor systems.

#### USAGE REQUIREMENTS

The pre-existing research plans that already involved using small computers (PDP11's in fact), and which define the core requirements for the present multiprocessor system, cover almost the full range of computer science research at CMU: artificial intelligence, system programming, and computer structures. We take up each in turn.

##### Artificial Intelligence

The relevant work in artificial intelligence is in speech and vision, the development of a speech-understanding system being a major research venture at CMU, research on vision being a minor one. The requirements are of two kinds. The first, common to speech and vision, is that special high data rate, real time interfaces are required to acquire speech and vision data from the external environment. For reasons of total system reliability these devices should be acquired by an input-output processor (Pio) and then transferred to Mp under standard internal conditions. A major cause of instability at a PDP10 research installation is the specialized hardware added by that installation.\*

---

\* This is extremely difficult to prove, of course. In our own system the addition of any hardware causes some instability for a period of up to several weeks.

Use of a P10 also permits some pre-processing of the incoming high rate data, thus lightening the load on the main system (the PDP10's) which is tuned to a slower interactive rate.

The second requirement, and by far the more stringent one, is to obtain real time processing for the speech-understanding system. The class of schemes under active pursuit involve multiprocessing with several PDP11's plus a PDP10, organized to work on all the levels of speech representation in parallel and with continuous intercommunication (i.e., the acoustic, parametric, phonemic, lexical, syntactic and semantic levels). The extent and shape of the parallel computation and intercommunication is a matter for intensive investigation, but some such scheme seems a fruitful approach to achieve real time speech processing. The total requirements of the system involve very large amounts of Mp, shared between several processors, backed up by extensive secondary storage. In short it is a computational problem that puts demands on the system along most dimensions. The original multiprocessing scheme envisioned was a more limited one, especially in numbers of Pc's, and one to be much more tailored to what seemed justifiable as a large P10 for speech.

#### Systems Programming and Operating Systems

At CMU we are intensively involved in research on operating systems and on understanding how software systems are to be constructed. Three faculty members independently (Habermann, Parnas and Wulf) and many graduate students are interested in these problems. Research in these areas has a strong empirical and experimental component, requiring the design and construction of many systems. We have discarded generally the notion of using virtual or simulated machines for reasons of performance, credibility and

realistic users. Given the unavailability of the PDP10 system, which operates in a user oriented time-sharing mode around the clock, a uniform solution has been to move to small stand-alone computers as the object machines for such experimental systems work, and at least two proposals for using PDP11's had been made prior to the emergence of the present multiprocessor.

The primary requirement of these systems is isolation, so they can be used in a completely idiosyncratic way and be restructured in terms of software from the ground up. On the other hand, the exact configurations needed varies from time to time. More important, work on the various experimental systems does not go on continuously 24 hours a day, but exhibits a variable pattern of usage, depending on how many people are working on the system and how intensively. These systems also require in an essential way access by multiple users and varying amounts of secondary memory.

#### Computer Structures

There is currently intensive activity at CMU in working with register transfer modules, using a system of modules (RTM's) developed here and at DEC (Bell, Grason et al, 1970) and now in production as the PDP16. A dedicated facility is needed for the design testing of experimental systems constructed of these modules. A small computer is appropriate, not only to avoid building specialized test equipment, but to permit the design testing to become fully automatic. When operating in test mode the computer must be directly coupled to the system being tested and isolated from all other systems.

A similar need for a dedicated system arises from an interest in fault tolerant systems. Here, it is necessary that the various processors can have access to one another so that checking can take place. For example, a console of a faulty computer would be attached to a checking computer so that all tests could be carried out under control of the checking computer.

### General Requirements

All of these research efforts require the same type of general purpose service as is now provided by our PDP10 system: files, editing, higher languages, interactive access, etc. That, under our original plans, some of the stand-alone systems might have had to give up some of these facilities only expresses the price that appeared necessary to obtain an isolatable system that could be restructured from the hardware up. Access to these facilities simultaneously with the capability for isolation on demand is certainly to be preferred.

The reasons for requiring general facilities for the speech and vision research appear somewhat different, since these programs are already large integrated software systems with an essential component running on the PDP10.

### POTENTIAL RESEARCH EFFORTS

We include in this category those researches that have not guided the design of the multiprocessor system, in the sense of providing constraints it had to satisfy, but which appear to capitalize on the system so strongly that they have a probability of occurring that ranges from high to certain.

### Network Research

Our original plans in obtaining the PDP10 system involved moving towards a network configuration of several PDP10's. Our motives for so doing were in large part conditioned on our wanting to do research on networks, as an appropriate structure for a within-institution computer system (Bell, Habermann, McCredie, Rutledge, Wulf, 1969). The growth of an actual network has been slow, being affected by funding limitations and by other developments in the environment. Currently we have two PDP10's, the second one just achieving full system status.



The multiprocessor affects the network plans in two ways. The multiprocessor itself provides a field within which one can set up experimental network configurations of assorted sizes and shapes. There are limits to what can be achieved by simulation and artificially constructed abstract networks, just as there are limits to what can be achieved by simulated and virtual operating systems. But they provide an essential tool. The proposed multiprocessor will permit systems with up to perhaps a dozen real nodes to operate in real time; and we expect a substantial amount of network research to evolve in this direction.

The multiprocessor system, taken as a single computer system, provides a third node in a network consisting of itself and the other two PDP10's. This is a real network with real problems of exploitation. The multiprocessor will serve as a highly specialized node with unique functional capabilities. It will be in no way a minor node in the system in terms of processing power, though this power will be realized on relative short work lengths (16 b/w). Thus, the multiprocessor will become the third node in our proposed network, displacing current plans to acquire a third PDP10.

#### Design of Processors

The multiprocessor is designed around a set of identical Pc's, namely PDP11's. However, as long as certain interface design philosophies are maintained, it is quite possible for processors of quite different designs to operate in the system, and to obtain from the rest of the system support with respect to files, input/output, initialization, etc. In fact, various systems constructed with RTM's will undoubtedly take advantage of this.

The recent design exercise on an ARPA list processing system led to a local design (C.ai, see Bell, Freeman, et al, 1971; Barbacci, et al, 1971; and McCracken and Robertson, 1971) that stressed, among other things, the design of processors that went a long ways in specialization to a given list processing system (e.g., a LISP system, an L\* system, etc.). The multiprocessor may offer an opportunity to design and realize some experimental processors. Considerable investigation is required before one can be sure such a venture is worth the costs involved, but with the lowering cost of processors it appears on first analysis to be quite feasible. On the positive side is the principle, now stated several times, that only a small fraction of the necessary research on computer systems can be done via simulations and virtual machines. An actual unconventional processor (say for L\*), which offered substantial processing advantages over programmed versions of the same system, would call forth the programming efforts to explore the system's real worth and determine its ultimate defects as a user system.

#### MULTIPROCESSOR RESEARCH

It is a fact that one cannot put together a multiprocessor system of any complexity today without becoming involved in research on multiprocessors. ~~Multiprocessor systems (other than dual processor structures and nPio-Pc structures)~~ have not become current art. Some of the reasons for this state of affairs appear to be:

1. The absolutely high cost of processors and primary memories so that a complex multiprocessor system was simply beyond the computational realm of all but a few extraordinary users, no matter what the advantages.
2. The relatively high cost of processors in the total system, so that an additional processor did not increase the performance/cost ratio.

3. The unreliability and performance degradation of operating system software, which made going to more complex system structures a venture in futility.
4. The inability of technology to construct the central switches required for such structure, due in part to low component density, in part to high cost.
5. The loss of performance in multiprocessors due to memory access conflicts and switching delays.
6. The unknown problems of dividing tasks into subtasks that can be executed in some parallel, though possibly interactive way.

Thus, the expense was prohibitive, even for discovering what advantages of organization might overcome the obvious decrements of performance. And in any event, it seemed mostly to complicate matters in a world that was already too complicated to function properly.

We appear to have entered into a technological domain when many of the difficulties listed above no longer stand so strongly:

- 1'. Providing we limit ourselves to multiprocessors of minicomputers, the total system costs of Pc and Mp are now within the price range of a reasonable facility. (E.g., the system proposed here with 16 Pc, suitable Mp, Ms and T will still be around \$500,000, yet execute approximately 3~5 million instructions/sec.)
- 2'. The Pc is now generally a smaller part of the total system cost. Hence, a small incremental cost that provides a substantial increase instruction/sec will be cost effective.
- 3'. Software reliability is now somewhat improved, primarily because a large number of operating systems have been constructed. Something is understood about the overhead costs imposed by various forms of organization.

- 4'. Current medium and large scale integrated circuit technology enables the construction of switches that do not have the large losses of the older decentralized switches.  
Centralization of the switch permits fewer and shorter cables.
- 5'. Memory conflict is not high for the right balance of processors, memories and switching system. The trade-offs possible, of course, are characteristic of a given domain of technology. Appropriate balances appear to be attainable currently.
- 6'. There has been work on the problem of task parallelism, not only general studies, but highly specific studies, e.g., around ILLIAC IV and CDC STAR. Other work on modular programming (Krutar, 1971; Wulf, 1971) at CMU suggests how subtasks can be executed in a pipeline.

In short, the price of experimentation appears eminently reasonable, given that there are requirements that appear to be satisfied in a sufficiently direct and obvious way by a proposed multiprocessor structure (i.e., for us, the usage requirements listed at the beginning of the section).

However, the state indicated above does not settle many issues about multiprocessors, nor does it make the development of one routine. We list below the main areas of research interest in the multiprocessor system itself. All of these call for some attention, although the extent of our effort in each is unclear.

1. The multiprocessor design itself, i.e., its PMS structure.  
Few enough multiprocessors have been built, so each one represents an important point in design space.
2. The processor-memory switch design, especially with respect to reliability.

3. The configuration of computations on the multiprocessor.

There are many processing structures and little is known about when they are appropriate and how to exploit them, especially when not treated in the abstract but in the context of an actual processing system:

Parallel processing: a task is broken into a number of subtasks and assigned to separate processors.

Pipeline processing: various independent stages of the task are executed in parallel (e.g., as in a co-routine structure).

Network processing: the computers operate quasi-independently with intercommunication (with various data rates and delay times).

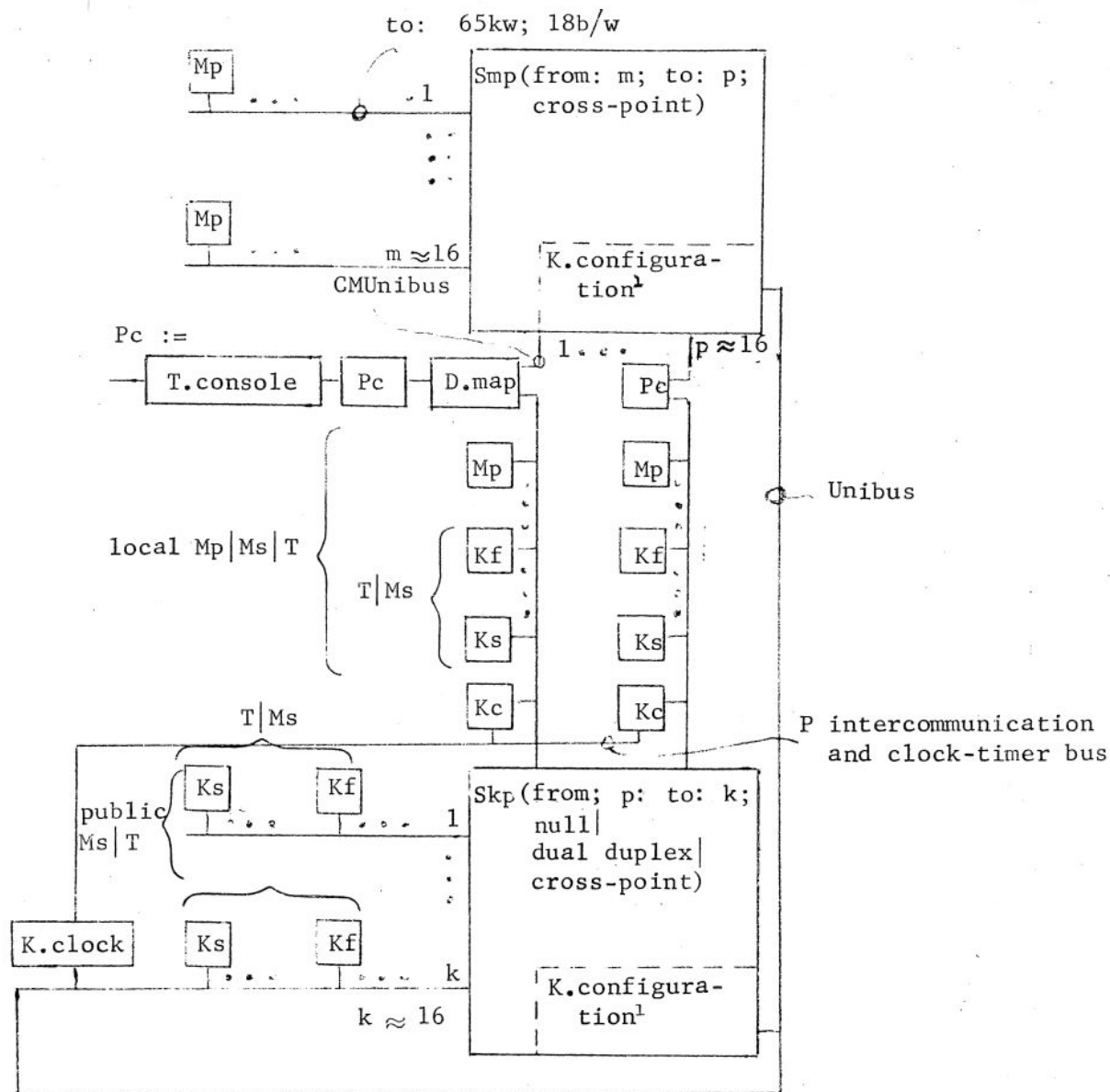
Functional specialization: the processors have either special capabilities or access to special devices; the tasks must be shunted to processors as in a job shop.

Multiprogramming: a task is only executed by a single processor at a given time.

Independent processing: a configurational separation is achieved for varying amounts of time, such that interaction is not possible and thus doesn't have to be processed.

4. The decomposition of tasks for appropriate computation. Detailed analysis and restructuring of the algorithm appear to be required, as in the work with ILLIAC IV. The speech-understanding system is the one major example we know we will undertake. It has research interest from the multiprocessor viewpoint, as well as from the speech recognition viewpoint.
5. The operating system design and performance. While the basic operating system design must be conservative, since the multiprocessor will run as a computation facility, it will have substantial research interest. Variations and alternative operating systems will also be constructable.

6. The measurement and analysis of performance of the total system. One of our complaints about the current field is the high ratio of design studies to performance studies. Where it is documented how well the Burroughs D825 operates as a multiprocessor, or even in what configurations it has operated?
7. The achievement of reliable computation by organizational schemes at higher levels, such as redundant computation.



where: Pc/central processor;  $M_p$ /primary memory; T/terminals;  
 $K_s$ /slow device control (e.g., for Teletype);  
 $K_f$ /fast device control (e.g., for disk);  
 $K_c$ /control for clock, timer, interprocessor communication

<sup>1</sup>Both switches have static configuration control by manual and program control

Fig. 1. Proposed CMU multiminiprocessor computer/C.mmp

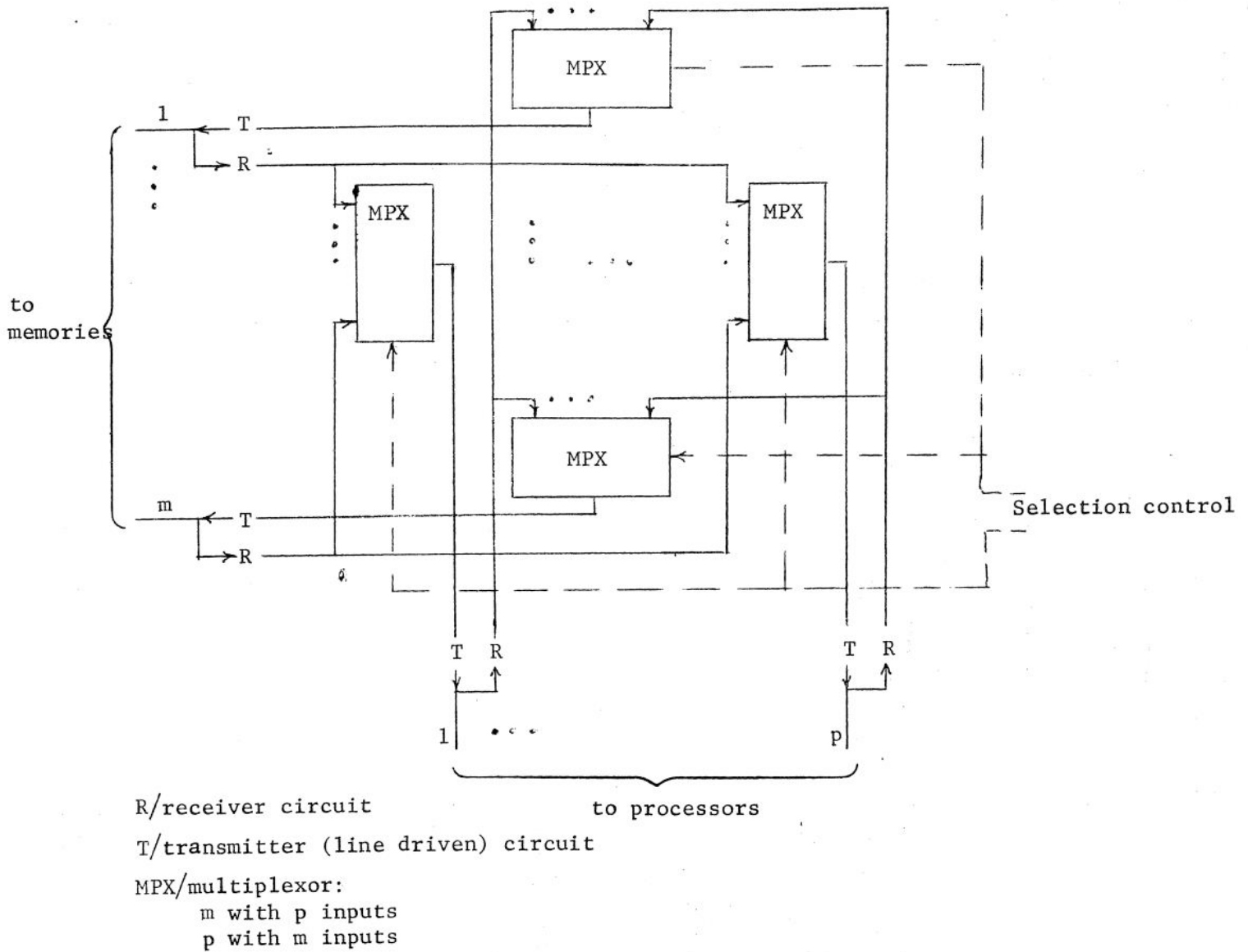


Fig. 2. Bi-directional, cross-point switch of 1 bit for Smp.

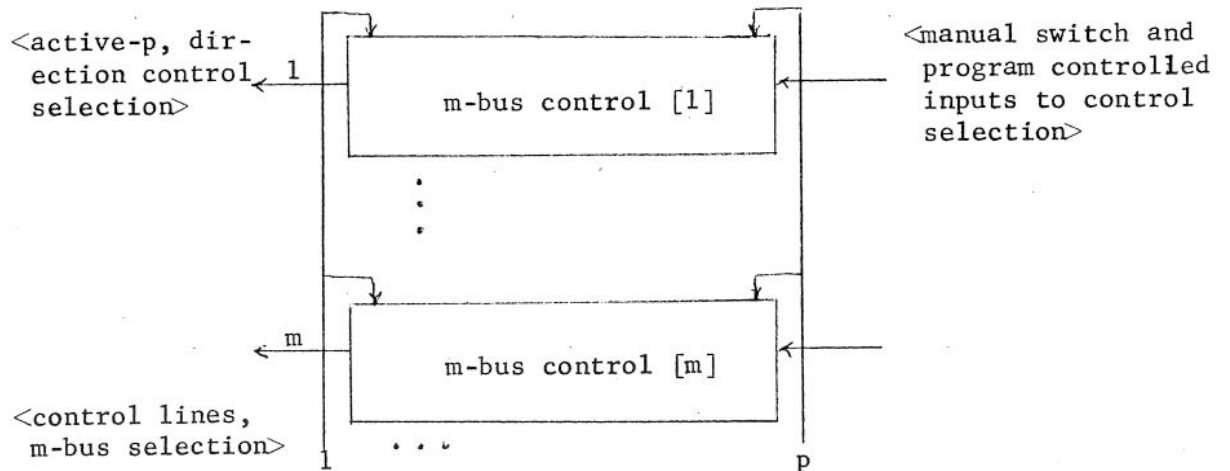


Fig. 3. Control part of Smp.