

The NON-VON Supercomputer Project:
Current Ideology and Three-Year Plan¹

David Elliot Shaw
Department of Computer Science
Columbia University
June, 1983

1 Introduction

While we have learned a great deal during the first two years of the NON-VON Supercomputer Project, I am reluctant to commit myself at this point to anything that might be called a "position" regarding the direction and ultimate outcome of current research in the field of parallel architectures. In part, my hesitation reflects an appreciation for the difficulty of objectively assessing the state of the field as a whole while enmeshed in the "cult of personality" surrounding a particular machine. Fortunately, our local dogma has not yet become so rigid as to preclude the possibility of significant revisions of our beliefs in response to the experiences and ideas of our colleagues. At the same time, it is clear that our understanding of the essential issues of parallel machine design in general is colored by the particular challenges we have faced in the context of the NON-VON Project.

In the following discussion, I will thus try to avoid any claims regarding the ideological correctness or historical inevitability of any of the architectural principles to which I now subscribe. In their place, I will

¹Despite the title, this research was in fact supported in part by the Defense Advanced Research Projects Agency under contract N00039-80-G-0132.

attempt to list a few of our current architectural objectives, and to outline our tentative hardware implementation plans for the next three years. Software considerations will not be discussed in this document, despite the fact that they have occupied a large fraction of our time.

2 Current Architectural Objectives

If pressed to identify the three most important objectives of our current research, I would list the following:

1. The extensive intermingling of processing and memory resources, supporting massive "fine granularity" parallelism.
2. The construction of machines based on heterogeneous interconnection topologies, and incorporating both "large" and "small" processing elements.
3. The provision of hardware support for both SIMD and MIMD control regimes, to support a wide range of parallel algorithms involving different modes of inter-processor communication.

Our vehicle for the pursuit of these objectives is a family of closely related machines that we have come to call NON-VON.

3 The NON-VON Project

NON-VON is a massively parallel non-von Neumann supercomputer architecture that has been under investigation at Columbia since 1980. The machine was originally designed to provide highly efficient support for the kinds of symbolic information processing tasks that seem to arise frequently in the context of large-scale artificial intelligence and database management applications. While such tasks remain our primary focus, we have since come to suspect that the NON-VON architecture may prove applicable to such diverse application areas as signal processing, physical simulation, and low-level

computer vision as well.

The following goals are central to the NON-VON Project:

1. The experimental construction of working prototypes of the NON-VON family of machines, in an attempt to validate certain innovative architectural principles that could have important practical implications.
2. The development of languages, translators, and operating systems capable of effectively exploiting the potential parallelism of such machines without the introduction of prohibitive software complexity.
3. The implementation of a modest corpus of working applications software that demonstrates NON-VON's potential advantages in the context of different kinds of computational tasks.

4 Staged Development of the NON-VON Machine

During the coming three-year period, we plan to proceed in several stages toward the satisfaction of our long-range goals. Our (partially overlapped) three-stage development strategy is designed to minimize the risk involved in developing a highly unconventional supercomputer. We plan to begin by implementing and testing a relatively simple machine which nonetheless incorporates what we regard as the most essential elements of a full-scale NON-VON supercomputer. Architectural enhancements would be added in stages, yielding incremental increases in power and generality without the introduction of an unmanageable increase in conceptual or engineering complexity at any single stage.

4.1 NON-VON 1

The first version we intend to actually implement, which we now call NON-VON 1, is based on a chip we have recently completed and are now in the process of testing and debugging. Each chip will contain a single small processing

4

element (SPE), including its own small local RAM. These single-SPE chips are to be interconnected to form the primary processing subsystem, which is configured as a binary tree, with a control processor (CP) attached to the root.

Because only a single CP will be incorporated in the NON-VON 1 prototype, the machine will be limited to single instruction stream, multiple data stream (SIMD) applications, in which a the CP sends instructions to be executed concurrently by all processing elements. Although a complete system would also include a secondary processing system (SPS) based on a number of "intelligent" disk drives, we do not propose to develop a working SPS within the scope of this contract. In short, NON-VON 1 will be limited to the execution of SIMD algorithms in which the argument and result data does not exceed the capacity of the PPS.

Unlike more recent versions of the architecture, NON-VON 1 performs all arithmetic and logical operations in a bit-serial fashion and is rather limited its choice of operands for most instructions. Because only one SPE is embedded on each chip, a relatively low priority was placed on the minimization of silicon area; detailed measurements of the NON-VON 1 layout have, however, formed the basis for the highly efficient floor plans now under development for use in later versions.

For the sake of completeness, it is probably worth mentioning at this point that the name NON-VON 2 was assigned to an interesting architectural exercise that we do not currently plan to carry beyond the "paper-and-pencil" stage, although its essential ideas may well influence future NON-VON designs.

4.2 NON-VON 3

The machine we now call NON-VON 3 forms the basis for much of the work we plan to do during the next three years. Like NON-VON 1, our NON-VON 3 prototype will include no disk drives and only a single control processor, and will thus be capable of executing only SIMD algorithms in which the data does not exceed the capacity of the PPS. The machine will be similar in most respects to the original NON-VON 1 design, but will incorporate a number of improvements suggested by the results of our initial experiments in chip design and software development. In particular, the NON-VON 3 SPE will feature:

1. An area-efficient eight-bit ALU to replace the one-bit ALU incorporated in the prototype NON-VON 1 SPE chip.
2. Fewer local registers, based on NON-VON 1 area measurements and software simulation results.
3. A far better floor plan, formulated using precise measurements taken from the prototype chip.
4. A generalization of certain NON-VON 1 instructions to support the more efficient execution of many common instruction sequences.

The NON-VON 3 instruction set is nearly identical to, and with few exceptions, more general than the one employed in NON-VON 1. Some of the additions in fact correspond to commonly used macros in our existing NON-VON 1 software. Before adopting this instruction set, however, we were careful to insure that all existing NON-VON 1 software could be simply and mechanically translated into NON-VON 3 instructions, so that none of our work to date would be lost. Such a translator should be completed shortly. Translated programs will take advantage of some, but not all of NON-VON 3's enhancements. In the future, of course, NON-VON 3 software will be written using NON-VON 3 instructions, allowing the exploitation of all of these features.

4.3 NON-VON 4

The NON-VON 1 and 3 machines should serve to validate many of our most important architectural ideas, yielding major performance improvements on a number of problems amenable to SIMD execution. The more sophisticated NON-VON 4 architecture, though, is intended to provide for the highly efficient execution of a much wider range of computational tasks than NON-VON 1 and 3. The most significant enhancements we expect to incorporate in NON-VON 4 involve the addition of a few thousand large processing elements (LPE's) within the top portion of the PPS tree, all interconnected in a high-bandwidth interconnection network, and each capable of serving as a control processor for an independent PPS subtree. This should give NON-VON 4 the capacity for multiple instruction stream, multiple data stream (MIMD) and multiple SIMD (MSIMD) operations, multi-tasking and multi-user applications, and such problems as physical simulation for which the top of the NON-VON 3 tree would otherwise represent a significant communication bottleneck.

We hope to realize an additional multiplicative factor in total throughput by reducing the effective instruction cycle time (which is equal to the time required for parallel inter-SPE communication) far below the estimated two microseconds projected for NON-VON 1 and 3. Among the techniques we plan to employ to achieve such an improvement are a separation of instruction broadcast and inter-SPE data communication functions, the provision of a wider instruction broadcast data path, local caching of instructions, and tree pipelining of blocks of instructions during transfer to the local caches. Rough initial estimates suggest that these techniques might reduce average instruction cycle time by as much as a factor of four or five.

Another important feature of the NON-VON 4 design is the incorporation of a large number of standard, commercially available dynamic RAM chips, which we expect to couple tightly to the individual PPS chips. While we expect this

RAM to be used in several different ways within the NON-VON 4 machine, one of its most important functions would be as a high bandwidth ``swapping memory'', allowing data to be very rapidly transferred to and from the many local RAM's embedded within the PPS.

5 Further Reading

I have enclosed a copy of a technical report entitled "The NON-VON Supercomputer", which contains a rather detailed description of the NON-VON 1 machine, although with little discussion of the Secondary Processing Subsystem or of NON-VON's operation in other than a strictly SIMD mode. Unfortunately, the only papers describing NON-VON 3 are the detailed system design documents we use for internal purposes, which would probably not be of interest to outside readers (but which we would be happy to provide upon request). Even less is available on NON-VON 4. We would be pleased, however, to add anyone expressing an interest in the NON-VON 3 or 4 architectures to our mailing list, and to send out any relevant documents as soon as they become available.

Table of Contents

1 Introduction	1
2 Current Architectural Objectives	2
3 The NON-VON Project	2
4 Staged Development of the NON-VON Machine	3
4.1 NON-VON 1	3
4.2 NON-VON 3	5
4.3 NON-VON 4	6
5 Further Reading	7