META SERIES TODAY, SCALABLE PARALLEL PROCESSING TOMORROW

WHITE PAPER SERIES

BY MIKE TIGHE
CONVEX COMPUTER CORPORATION

ABSTRACT

This paper describes how the Convex Meta Series can be used as the foundation for developing parallel applications for Convex's upcoming Scalable Parallel Processing (SPP) system.

INTRODUCTION

On October 13, 1992, Convex Computer Corporation unveiled the first phase of its Scalable Parallel Processor (SPP) development strategy by announcing the Meta Series, a system that links Convex C Series supercomputers with Hewlett-Packard (HP) high-performance PA-RISC processors.

The Meta Series is comprised of a C Series supercomputer, two or more Hewlett-Packard processors based on the Series 9000/735 PA-RISC workstation, system software for both components, and specialized cluster software.

The Meta Series couples the advantages of a large-memory, vector supercomputer with the advantages of high-performance RISC-based processors. This combination allows Convex customers to solve a wider range of problems than ever before.

The Meta Series provides an ideal platform for developing SPP applications today. To fully understand why this is so, a brief overview of Convex's first SPP system, dubbed SPP-1, is in order. The design of the SPP-1 system makes the advantages of using the Meta Series for application development readily apparent.

CONVEX SPP-1 SYSTEM

The first vector supercomputers were difficult to use and program. As a result, they were commercial failures, being replaced by easy to use vector machines running standard UNIX-based software such as the Convex C Series.

Similarly, many users look forward to moving their applications to a scalable parallel system. However, they have avoided doing so to date because the current offerings are very difficult to use and even more difficult to program.

Convex believes that the most successful scalable parallel system will be the easiest to use and program. In order for a scalable parallel system to be successful, Convex believes it must have the following characteristics:

- *Scalability.* The system must be scalable from small to large configurations. It must scale in multiple dimensions such as price, processors, memory, and I/O.
- *Low Cost of Entry.* There must be an affordable, entry-level system. The system should not require expensive front-end machines in order to function.
- *High-Performance CPUs.* Since few applications are 100% parallel, the system must be comprised of high-performance RISC processors that can efficiently execute scalar code.
- *Ease of Use.* The system must be easy to use and easy to program. It should be a time-shared, general-purpose, shared memory computer system that can run standard UNIX-based software.
- *Third Party Applications.* Since many customers can no longer afford to develop and maintain their own applications, the system must have a large set of ready-to-run third-party applications.
- *User-Developed Applications.* For those customers who do develop their own applications, the system must have a complete set of software development tools such as compilers, debuggers, profilers, and libraries, all of which are easy to use and provide a familiar interface.
- *Protected Investment.* The system must be part of an overall strategy that protects the customer's investment, by providing a painless upgrade path to future machines that will provide greater price/performance and software availability.

These characteristics are not revolutionary. In fact, they are the same characteristics that were required for vector machines to be successful. The Convex SPP-1 system has all of these characteristics.

CONVEX SPP-1 HARDWARE DESIGN

At the center of the SPP-1 hardware design is the Hewlett-Packard PA-7100 processor. The PA-7100 processor is Hewlett-Packard's first superscalar implementation of PA-RISC. By superscalar, we mean that the PA-7100 can execute up to 2 instructions each clock period: 1 integer (including loads and stores) and 1 floating point.

The PA-7100 is able to issue floating point instructions in consecutive clock cycles without delay. By using compound instructions such as `fmpyadd`, the PA-7100 can deliver 2 floating point results each clock period. By interleaving these floating point instructions with load/store instructions to achieve maximum superscalar execution, the PA-7100 can deliver performance approaching its theoretical peak.

The following code segment highlights this programming technique. This example is taken directly from the ConvexMLIB (more on this later) scientific routine DCONV, which is a double-precision (64-bit) convolution.

Cycle	Instruction	Registers
1	`fmpyadd,dbl`	`%fr6,%fr12, %fr22,%fr20,%fr24`
1	`flddx,sm`	`%r25(%sr2,%r21), %fr8`
;		
2	`fmpyadd,dbl`	`%fr7,%fr12, %fr23,%fr21,%fr25`
2	`flddx,sm`	`%r25(%sr2,%r21), %fr9`
;		
3	`fmpyadd,dbl`	`%fr8,%fr12, %fr20,%fr22,%fr26`
3	`flddx,sm`	`%r25(%sr2,%r21), %fr10`
;		
4	`fmpyadd,dbl`	`%fr9,%fr12, %fr21,%fr23,%fr27`
4	`flddx,sm`	`%r25(%sr2,%r21), %fr11`

An SPP-1 node is comprised of four or eight PA-7100 processors. Each PA-7100 processor has its own 1-megabyte instruction cache and 1-megabyte data cache. Access to these caches is made via a 64-bit path and can be completed in 1 cycle. Processors on the SPP-1 are clocked at 100 MHz.

Each SPP-1 node can support from 256 to 2048 megabytes of memory. Memory is arranged in four independent banks, and each bank is further sub-divided into two sections.

Each SPP-1 node also has its own I/O system, which will support a variety of peripherals. The I/O system on each node has a capacity of 250 megabytes per second.

These components are all interconnected via a single, high-performance, non-blocking crossbar. Therefore, a single node of an SPP-1 system can be viewed as a symmetric multiprocessor.

The SPP-1 system is scalable from 1 to 16 nodes. Nodes of the SPP-1 system are interconnected to each other via four high-bandwidth, low-latency CxRing interconnects. Each of these interconnects has a peak bandwidth of 600 megabytes per second.

A fully configured SPP-1 system has 128 processors, a peak theoretical performance of 25.6 billion floating point operations per second (GFLOPS), an I/O capacity of 4 gigabytes per second, and a memory size of 32 gigabytes.

CONVEX SPP-1 SOFTWARE DESIGN

The SPP-1 system will run a UNIX-based operating system in a time-shared, general-purpose mode. The Application Programmer Interface will support ConvexOS as well as HP-UX. There will be source code compatibility with ConvexOS and binary compatibility with HP-UX.

While current parallel systems rely exclusively on message passing in order to run parallel codes, the SPP-1 will support both

message passing and direct generation of parallel code by the compiler, without programmer intervention. The interprocedural optimizing (IPO) Convex Application Compiler will be at the heart of the parallel compilation system. It will support both the C and FORTRAN programming languages. The Application Compiler will generate parallel code much like vector code is generated on vector machines. This will make the system much easier to program.

The Parallel Virtual Machine (PVM) software library will be supported for message sing. Convex PVM will be based on the ConvexPVM version that is already running on the Meta Series, with optimizations for the SPP-1 architecture. Support for other industry-standard message passing libraries is also being considered.

For computationally intensive kernels, Convex will provide MPPLIB. MPPLIB is a parallelized version of ConvexMLIB, which is already available on the Meta Series. MPPLIB is a collection of scientific kernels, such as BLAS, LINPACK, LAPACK, EISPACK, MINPACK, and FFTs, optimized for the SPP-1 architecture.

The SPP-1 system will also provide standard UNIX-based performance analysis tools such as *prof* and *gprof*. The Convex performance analyzer, CXpa, will be available on the SPP-1 system. CXpa for the SPP-1 will support the rently parallel nature of the SPP-1 system, as well as add new features that the SPP-1 hardware will provide.

While standard debugging tools such as *adb* will be provided, debugging parallel applications requires a high-performance debugging tool. Convex CXdb, which already supports parallel debugging of optimized code on the C Series, will be available on the SPP-1 system.

In addition to all of these software tools, every effort is being made to make the SPP-1 system an easy to use, general purpose system. Therefore, all of the standard software tools that are available on the Meta Series, such as editors (*Vi*, *Emacs*), script languages (*awk*, *perl*), and so on will be available on the SPP-1 system.

Convex Meta Series

The Meta Series is comprised of four parts. These are the C Series component, the PA-RISC component, the interconnect, and software to use the Meta Series efficiently.

C Series component

The C Series is an integral component of the Meta Series, providing a large-memory compute node, a high-performance I/O subsystem for file service and connection to data center peripherals, and multiprocessor fine-grained parallelism. The C Series also provides file management services to the cluster with file service software such as CSM and UniTree.

PA-RISC cluster component

The PA-RISC cluster is based on the recently announced Hewlett-Packard Series 9000/735 desktop workstation. The Model 735 employs a PA-RISC uniprocessor with the same PA-7100 architecture as the Convex SPP-1.

Each PA-RISC processor in the cluster can be configured with 32 to 400 megabytes of memory and more than 2 gigabytes of internal disk space. Each processor offers built-in connections for a parallel port, RS-232, SCSI-2, and Ethernet or FDDI. It also has a single expansion slot that adheres to the Enhanced Industry Standard Architecture (EISA).

Meta Series interconnect component

The components of the Meta Series can be interconnected in a variety of ways including Ethernet, FDDI, and a custom Shared Memory Interconnect (SMI).

The Ethernet interconnect is the simplest. Ethernet performance has been measured at 1.1 megabytes per second (processor to processor), and Ethernet latency has been measured at 1 millisecond.

An FDDI interconnect option is also available. FDDI performance has been measured at more than 10 megabytes per second (processor to processor), and FDDI latency has been measured at just under 1 millisecond.

Both the Ethernet and FDDI interconnect options provide simple, cost-effective solutions for Meta Series connectivity. Customers who plan to use their Meta Series as a throughput engine will benefit from either of these cost-effective connectivity options.

Those customers who plan to use their Meta Series as a parallel, time-to-solution engine need a high-performance, low-latency, high-bandwidth interconnect. For these customers, Convex has developed the Shared Memory Interconnect.

The implementation of the SMI is very straightforward. Each processor in the PA-RISC cluster has a high-speed interconnect board that fits into its EISA slot. A similar high-speed interconnect board fits into the VME slot of the Convex C Series system. Each high-speed interconnect can support up to eight PA-RISC processors and one Convex C Series.

Each of these boards is connected by cable to the SMI chassis, which can hold up to three memory controller boards, each of which contains 32 megabytes of memory. In addition, the chassis contains a bus arbitration board that controls access to the memory boards.

Performance of the SMI is impressive. Node-to-node bandwidth has been measured at 14 megabytes per second. Node-to-node latency has been measured at under 500 microseconds.

META SERIES SOFTWARE COMPONENT

The Meta Series comes complete with the standard suite of software normally available on a C Series or PA-RISC system. This includes a complete set of compilers, debuggers, profilers, and libraries.

The Meta Series is a fully integrated cluster solution. Upon delivery, the cluster is immediately usable, unlike home-grown cluster solutions. Software is already loaded on the nodes, and network Ethernet addresses are already configured. More importantly, Convex provides support for all components of the cluster so problem reports, questions and enhancement requests, and in-depth testing are all supported by a single vendor.

System administration of clusters is a major concern of existing cluster sites. Convex has put considerable effort into developing Cluster System Administration Tools that allow the system administrator to view the cluster as one or more logical devices. This software allows the system administrator to perform tasks such as adding/deleting users, system/network configuration, system backups, system performance monitoring, cluster startup/shutdown, and so on much more easily.

To facilitate use of the Meta Series as a time-to-solution computing device, Convex offers ConvexPVM on the Meta Series. PVM is the *de facto* standard message passing library and has been endorsed by a variety of vendors. It is the primary method by which clusters are used for parallel computation

To execute jobs efficiently across all of the nodes, Convex has ported the Network Queuing System to the Meta Series. ConvexNQS+ allows jobs to be automatically distributed to the PA-RISC cluster, the C Series, or both. ConvexNQS+ also allows customers to balance jobs across the nodes of their PA-RISC cluster, or to hold jobs back from the PA-RISC cluster in order to allow a job using PVM software to

run in parallel on all of the nodes. The result is increased job throughput, since jobs run on the architecture most suited for them.

For scientific applications, Convex has developed ConvexMLIB. ConvexMLIB is a collection of scientific routines such as BLAS, LINPACK, LAPACK, EISPACK, MINPACK, and FFTs, specifically optimized for the PA-RISC architecture. The following table highlights performance of some popular ConvexMLIB routines on vectors ranging in length from 100 to 500.

ConvexMLIB Performance for Some Selected Routines (MFLOPS)

Routine	100	200	300	400	500
SCONV	181	188	191	191	191
DCONV	181	188	191	191	191
SGEMM	188	155	155	155	155
DGEMM	187	127	127	127	127

Sample Application

Currently, several customers are developing applications on the Meta Series for the SPP-1 system. One of these applications is an antenna modeling program. It is a real application, used on a daily basis by one of Convex's earliest customers.

This antenna modeling program was originally developed in the 1960s and is written in FORTRAN. Since that time, it has been ported to many different architectures. It was first ported to the Convex architecture in the mid-1980s, when the customer purchased a C1 system. The following table shows execution times for this code on a variety of single-CPU systems.

Execution Times for Single-CPU Systems

System	Elapsed Time
C3810	10 seconds
C3410	25 seconds
C3210	38 seconds
HP 735	67 seconds
HP 730	116 seconds
SUN IPX	368 seconds

The customer is now planning to move the application to a parallel system. To determine how well it will perform on the SPP-1 system, the application was ported to the Meta Series.

The parallelization approach was to first identify where the coarse-grained parallel sections were. Convex software tools such as CXpa and *gprof* were used to identify these coarse-grained parallel sections.

It was then decided that the most efficient approach would be to run identical copies of these coarse-grained sections on the different PA-RISC processors of the Meta Series, and to use ConvexPVM to tell the different processes what data each processor was to work on. Control would be maintained on the C Series component.

Because the design of an antenna involves simulating thousands of different input data sets, the C Series system provided an ideal location for storing these large amounts of data. Since both the C Series and the PA-RISC cluster support the IEEE data format, no data conversion problems existed.

The antenna modeling application now runs on the Meta Series. The following table shows the parallel speedup of the code as it scales from one to four nodes.

Parallel Speedup using ConvexPVM		
Nodes	**Elapsed Time**	**Speedup**
1	123 seconds	0.94
2	63 seconds	1.83
3	42 seconds	2.75
4	32 seconds	3.56

The customer is now able to model scalable parallel performance and estimate future parallel performance. The organization can now use this information to estimate the performance of other applications it is considering moving to the SPP-1 system.

The organization has also realized even greater benefits of the Meta Series. For example, the customer is now using CXpa on the C Series component to optimize the slave instance of the code, which is executed on the PA-RISC component. The purpose of this is to extract fine-grained parallelism from the coarse-grained parallel section of the model.

This is achievable since the SPP-1 compiler will be able to generate parallel code. Code that vectorizes at the DO-loop level with the -O2 option on the C Series system will parallelize across all processors on each node at the DO-loop level with the -O3 option on the SPP-1 system.

Once this task is complete, the customer plans to add calls to MLIB, and to write PA-RISC assembler code for critical portions of the routine, all of which will port without modification to the SPP-1 system.

This is just one successful example of using the Meta Series to develop applications for the SPP-1 system.

CONVEX META SERIES BENEFITS

The previous example highlights some of the many advantages of using the Meta Series to develop applications for the Convex SPP system. Following is a recap of some of the major benefits:

- *Source Compatibility.* The SPP-1 will be source code compatible with the C Series. Source code developed on the C Series component of the Meta Series can be moved to the SPP-1 system with little porting effort.
- *Binary Compatibility.* The SPP-1 system will be binary-compatible with HP-UX. Applications already running on the PA-RISC component of the Meta Series will run on the SPP-1 system.
- *Third-Party Applications.* Since the SPP-1 system is binary-compatible with HP-UX, all of the third-party applications that run on the PA-RISC component of the Meta Series will run on the SPP-1 system.
- *Protected Investment.* Convex is committed to the PA-RISC architecture. Customers can invest in this architecture by making major porting efforts or writing optimized assembler routines, and know that their investment is protected in future generations.
- *Software Environment/Tools.* The Meta Series provides a rich set of software development tools. These tools can be used to port, develop, and model performance on an SPP-1 system.

The following table compares some of the major features of both the Meta Series and the SPP-1 system.

Comparison of Features

Feature	Meta Series	SPP-1
Processor	PA-7100	PA-7100
Clock Speed	99 MHz	100 MHz
Node Size	1 CPU	8 CPU
Max. Nodes	8 (per cabinet)	16
I-Cache Size	256 KB	1 MB
D-Cache Size	256 KB	1 MB
Node-to-Node Latency	500 microsec	<3 microsec
Node-to-Node Bandwidth	14 MB/s	2.4 GB/s
I/O Capacity	132 MB/s	250 MB/s
HP-UX	Yes	Yes
PVM	Yes	Yes
MLIB	Yes	Yes
IPO	Yes	Yes
Tools	Yes	Yes

Summary

The Convex Meta Series is an ideal platform to develop Convex SPP-1 applications. By using the Convex Meta Series to develop applications for the Convex SPP-1 system, users can start moving forward with their migration to scalable parallel computing today.

For more information or for the location of your local Convex sales office, contact:

U.S.A. CORPORATE HEADQUARTERS

Convex Computer Corporation
3000 Waterview Parkway
P.O. Box 833851
Richardson, Texas 75083-3851
Phone: 214/497-4000

EUROPEAN HEADQUARTERS

Convex Computer, Ltd.
Randalls Research Park
Randalls Way
Leatherhead
Surrey, U.K. KT22 7TS
Phone: 44-372-386696

FAR EAST

Convex Computer PTE, Ltd.
1 Scotts Road
#25-06 Shaw Centre
Singapore 0922
Phone: 65-733-4355

AUSTRALIA

Convex Computer PTY, Ltd.
Level 20, Como Office Tower
644 Chapel Street
South Yarra, Victoria 3141
Phone: 61-3-823-6216

Printed in U.S.A.
080-002248-000