

Ni1000 Recognition Accelerator

High-Speed Classification Engine for Pattern Recognition

- **Accurate Recognition**
 - Up To 256 x 5-Bit Input Pattern Size
 - Estimates Up To 64 Class Probabilities Using Weighted Sum of Up To 1000 Radial-Basis Functions
 - Estimates Class Probabilities Even When Class Distributions Overlap
 - Parzen Windows Technique
 - Emulates Local Receptive Field Neural Network
- **Fast Recognition**
 - 200x Speed-Up Over Typical Host PC
 - Scalable Acceleration With Multiple Ni1000 Chips
- **Learning**
 - RCE Learning On-Chip
 - Fast Learning: <10 Epoch
 - Incremental Learning Capability
 - Off-Line Learning Supported: LVQ, k-Means + LMS, RBF Back-Prop
- **Memory**
 - 1000 Usable Prototypes of Up To 64 Classes Stored On-Chip
 - 256 x 5-Bit Prototype Size
 - Non-Volatile Flash Memory
 - Scalable Prototype Storage With Multiple Ni1000 Chips
- **Applications**
 - Hand- or Machine-Printed Character Recognition
 - Machine Vision
 - Medical Imaging
 - Biometric Identification
 - Fingerprint Recognition
 - Speech Recognition
- **Pipelined Parallel Processing**
 - 512 Distance Calculation Units
 - Three-Stage Pipeline
 - RAM-Buffered I/O
 - 16-Bit Floating-Point Math Unit
- **Microcontroller**
 - 16-Bit Microcontroller
 - 4K x 16-Bit Program Flash Memory
 - RCE Learning Code Supplied
 - Also Supports PNN Learning
- **Multichip Support**
 - Bus-Oriented Data Interface
- **x86-Compatible Interfacing**
 - 32- or 64-Bit Data Bus
 - Max. Data Transfer Rate Over 200MB/sec
 - Digital CMOS/TTL Technology
- **State-Of-The-Art Technology**
 - 0.8m CHMOS-IV
- **Compact Package**
 - 168-Pin PGA

Operating Mode	Clock Speed	Processing Time	Connections Per Second	Worst-Case Classification Rate
Unpipelined	33MHz	60 μ sec	3.9 Billion	16,000 patterns/second
Pipelined	33MHz	30 μ sec	8.2 Billion	33,000 patterns/second

Contents

1. GENERAL DESCRIPTION	4
2. PINOUT	8
2.1 Pin List	8
2.2 Pin Configuration	9
3. SIGNAL DESCRIPTIONS	10
4. ARCHITECTURE	13
4.2 Bus Interface	16
4.3 The Classifier	34
4.4 Microcontroller	50
4.5 System-Level Architecture	60
4.6 Classification Timing	61
4.7 Computational Precision	62
4.8 Software Modes and Configurations	64
5. BUS OPERATIONS	67
5.1 Hardware-Controlled Access Modes	69
5.2 Bus Cycles	72
6. ELECTRICAL CHARACTERISTICS	81
6.1 Absolute Maximum Ratings	81
6.2 D.C. Characteristics	82
6.3 A.C. Characteristics	83
7. MECHANICAL AND THERMAL CHARACTERISTICS	86
8. ORDERING INFORMATION	88
9. GLOSSARY	89
10. BIBLIOGRAPHY	93

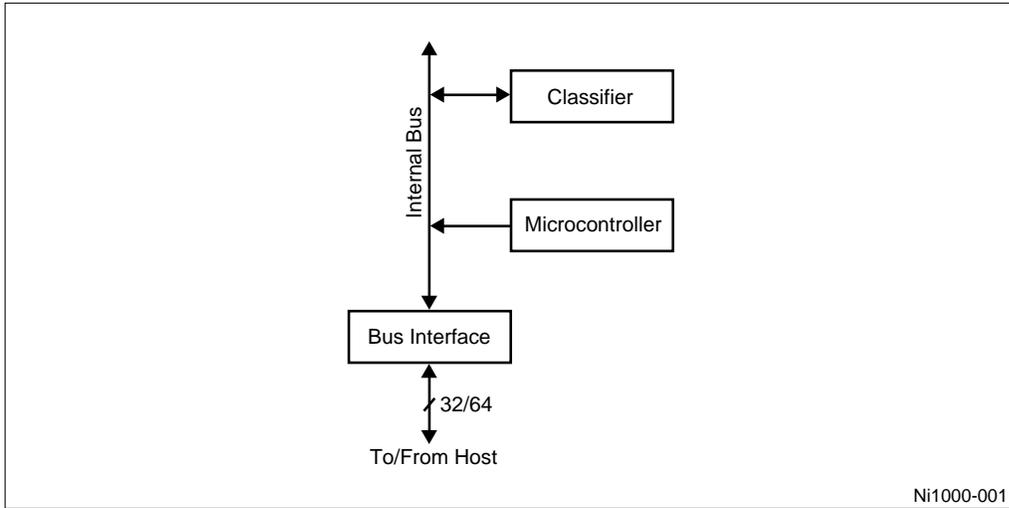


Figure 1-1. Block Diagram

1. General Description

The Ni1000 Recognition Accelerator supports classification speeds of up to 33,000 patterns per second, with real-time adaptation. The chip is compatible with commonly used Radial Basis Function (RBF) algorithms like the Restricted Coulomb Energy (RCE), Probabilistic RCE (PRCE), Probabilistic Neural Networks (PNN), and custom algorithms. The flexible, on-chip microcontroller and its 4-Kbyte, on-chip, non-volatile microcode memory accommodate custom algorithms.

The Accelerator accepts input vectors with a maximum of 256 feature dimensions, each with 32 levels of resolution, and it outputs up to 64 classes or probabilities. High-speed parallel processing units compute the city-block distance between an input vector and up to 1024 stored prototypical examples. The Accelerator's high speed is suitable for computationally intensive applications like optical character recognition, fingerprint identification, and industrial inspection.

Pattern recognition is the process of sorting input data into categories or classes that are significant to the user. The differences or underlying traits of each class must first be loaded into the chip's memory. The contents of the chip's memory can be developed manually or extracted from examples of data typical to the problem, using a learning algorithm. Input data is problem-specific and may consist partially or completely of stored data, such as historical records, or of direct sensor inputs. Once learning is complete, the system is ready to classify input data. The Ni1000 Recognition Accelerator supports incremental learning in the field. This is often necessary to further adapt the recognition system to its environment.

The architecture of the Accelerator consists of two main parts: a classifier and a general-purpose 16-bit microcontroller. The classifier performs distance calculations between the input vector and a set of up to 1000 stored *prototypes* vectors (or 8000, for problems with 32 dimensions or less), using an array of 500 dedicated processors. Its outputs are firing class IDs and probability density functions, which are calculated in six phases handled by a six-stage pipelined processor. The microcontroller directs the passage of data through the classifier and interacts with software running on the host.

Both the input and output ends of the classifier are equipped with alternating double buffers, so that the classifier does not stall during I/O. The host interface can be selected for 32- or 64-bit data bus width, and it supports a single-transfer-per-clock burst mode.

Figure 1-2 is a block diagram of the internal hardware architecture. The upper part of Figure 1-2 shows the classifier, the bottom part shows the microcontroller, and the middle part shows the interface to the host.

In an application environment, the classifier receives data from the host system through the bus interface, processes it, and sends the classification results back through the bus interface to the host. The classifier exploits both array and pipeline parallelism to provide the host with up to 33,000 classifications per second. The parallel hardware of the Distance Calculation Units and their tight coupling to the Prototype Array (PA) are responsible for much of this processing power. The Prototype Array holds 1000 (prototypes) x 256 (dimensions) x 5 (bits per dimension), for a total of 1.3 million non-volatile Flash storage bits. It can also hold prototypes for multiple problems with varying dimensionality—up to 8000 prototypes with 32 dimensions. Problems that need additional prototype storage can be solved with systems using multiple Ni1000 Accelerators. It is also possible to trade lower input-vector dimensionality for higher input-feature bit resolution.

Each of the 500 parallel Distance Calculation Units calculates the city-block distance by summing the differences between each component (dimension) of the input vector and that of a prototype vector stored in the Prototype Array. The DCUs are multiplexed twice in time at a sustainable processing rate of 16.5 billion operations per second and a bandwidth of up to 2.5 x 1010 bits per second. Parallel, absolute-value subtractors sequentially process the dimensions.

The classifier's Math Unit (MU) calculates probability density functions and class results concurrently. It processes floating-point data and computes the exponential and other mathematical functions. The MU uses a six-stage pipeline with a resolution of 16-bits for floating-point computations (10-bit mantissa and 6-bit exponent). It places results in one of two static RAMs. This double-buffering scheme allows the Math Unit to continue processing a second vector without interrupting the classification pipeline. The Prototype Parameter RAMs (PPRAMs) hold parameters needed to generate the classification results.

The bottom part of Figure 1-2 shows the microcontroller. It is a fully custom, 16-bit, Harvard-architecture microcontroller that supervises learning, performs chip maintenance tasks, and maintains communication with the host. It can also exchange interrupts with the host. The 4k x 16-bit PGFLASH Flash memory stores the microcontroller programs. All memory devices are memory-mapped to the microcontroller's address space, with the exception of the PGFLASH, which is the microcontroller's program Flash memory. Other facilities available to the microcontroller include 256 words of general-purpose static RAM (GRAM) and a free-running 32-bit timer. Although not shown in this diagram, the microcontroller has access to virtually all of the memories in the classifier, although classification must stop while the microcontroller accesses these memories.

The microcontroller can enable an automatic classification mode in which a series of logic blocks arranged as a pipeline process data and output the results to the host. The performance of 33,000 classifications per second is made possible by the Ni1000 parallel architecture. A Von Neumann machine would need to execute more than 16.5 billion instructions per second to approach the processing rate achieved by the Ni1000 Recognition Accelerator.

The middle part of Figure 1-2 shows interface to the host, which consists of input buffers (IRAM), an output buffer (ORAM), and sixteen I/O control registers. The external data bus can be either 32 or 64 bits wide and will perform single-clock burst transfers. The input stage buffers two full-sized vectors. The outputs can be either in IEEE standard 32-bit floating-point format or the internal 16-bit floating-point format. Both the host and the Accelerator's on-chip microcontroller can access the sixteen 16-bit I/O control registers. The registers contain various control parameters for the Accelerator and provide a general channel for communication between the microcontroller and the host.

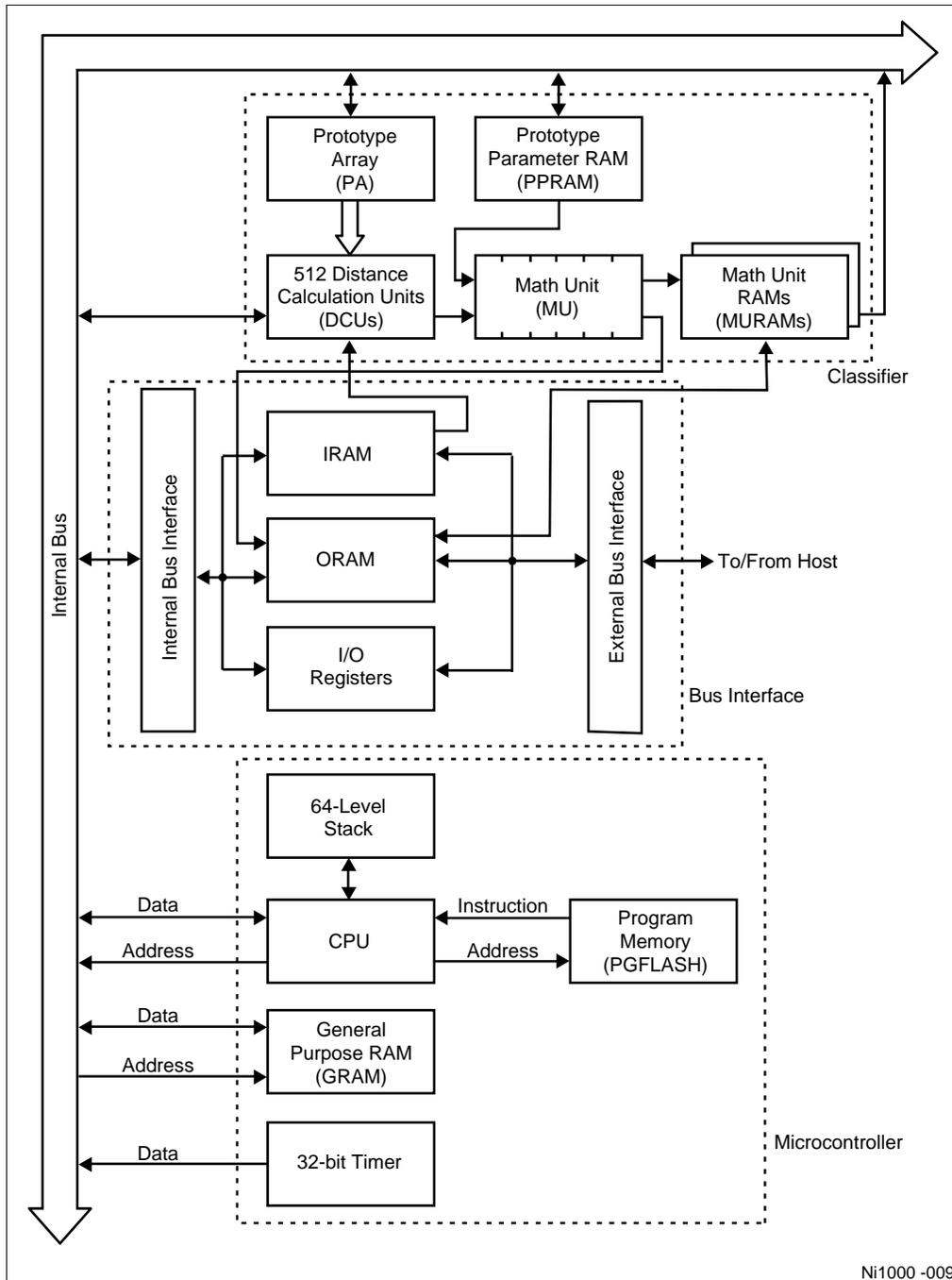


Figure 1-2. Internal Architecture

In most applications, the Accelerator will reside on a bus that is shared with a host CPU and perhaps other Ni1000 Accelerators, as shown in Figure 1-3. The Accelerator will typically be a slave device on the host bus. It will not initiate transactions on the host bus. Both the host and the on-chip microcontroller have the ability to interrupt each other.

Figure 1-3 shows a local host CPU on an add-in card for personal computers and workstations. The CPU manages the flow of data through the Accelerator(s). The CPU may also have other functions, such as preprocessing data, interpreting classification results, or coordinating the operation of multiple chips. Other implementations may rely on the CPU of the system board for these functions.

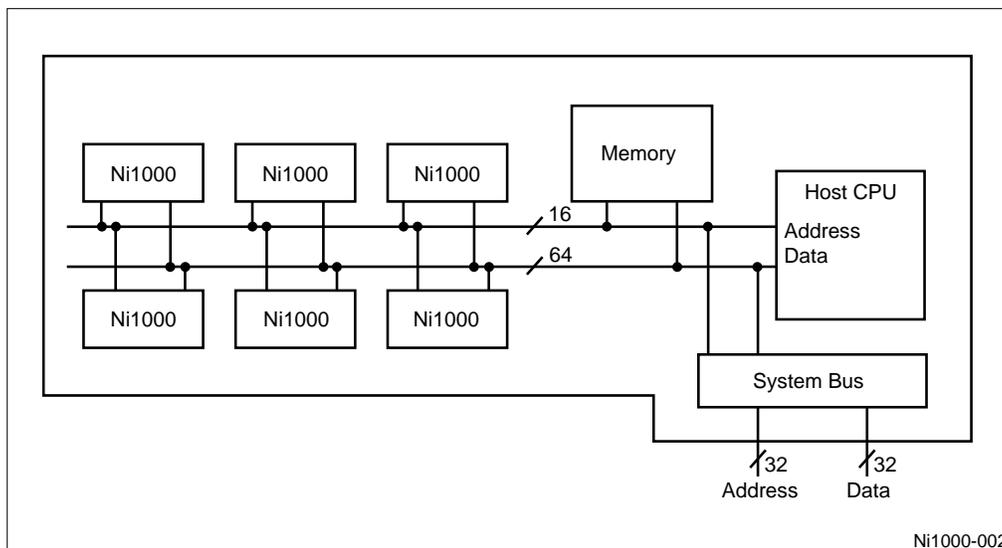


Figure 1-3. Typical Multichip Add-In Board