

80170NX

ELECTRICALLY TRAINABLE ANALOG NEURAL NETWORK

- Ultrahigh Performance
 - 2 Billion Connections Per Second (CPS)
 - Low-Power CHMOS* III EEPROM Technology
 - 300K Patterns Per Second (64- or 128-Input Patterns)
- Versatile Architecture
 - 64 Dot-Product Neurons with Sigmoid Functions
 - Separate Input and Feedback Arrays
 - Hold and Feedback Capability
 - Supports Hopfield, Multi-Layer and Recurrent Nets
 - 10,240 Nonvolatile Analog Synapse Weights
 - 6- to 7-Bit Typical Resolution
- 208-Pin PGA Package

(See Packaging Specification, Order #231369)

- Compatible with Popular Training Methods
 - Back-Prop and Competitive Learning
 - Single-Neuron Perturb for Madaline III Learning
- Flexible I/O
 - 64 Physical Inputs and Outputs, Configurable for 128 Inputs
 - Supports I/O Cascading
 - I/O Referencing Supports TTL Operation
 - User-Controlled Gain on Neuron Sigmoid Functions
- Many Real-Time Applications
 - Signal Processing
 - Process Optimization
 - Robotic Motion
 - Associative Memory

GENERAL DESCRIPTION

The 80170NX Electrically Trainable Analog Neural Network (ETANN) chip offers unprecedented single-chip bandwidth and storage capacity for pattern-recognition problems. The chip simulates the data processing functions of 64 neurons, each of which is influenced by up to 128 positively or negatively weighted "synapse" inputs. The chip's data inputs and outputs are analog; its control functions for setting and reading synapse weights are primarily digital.

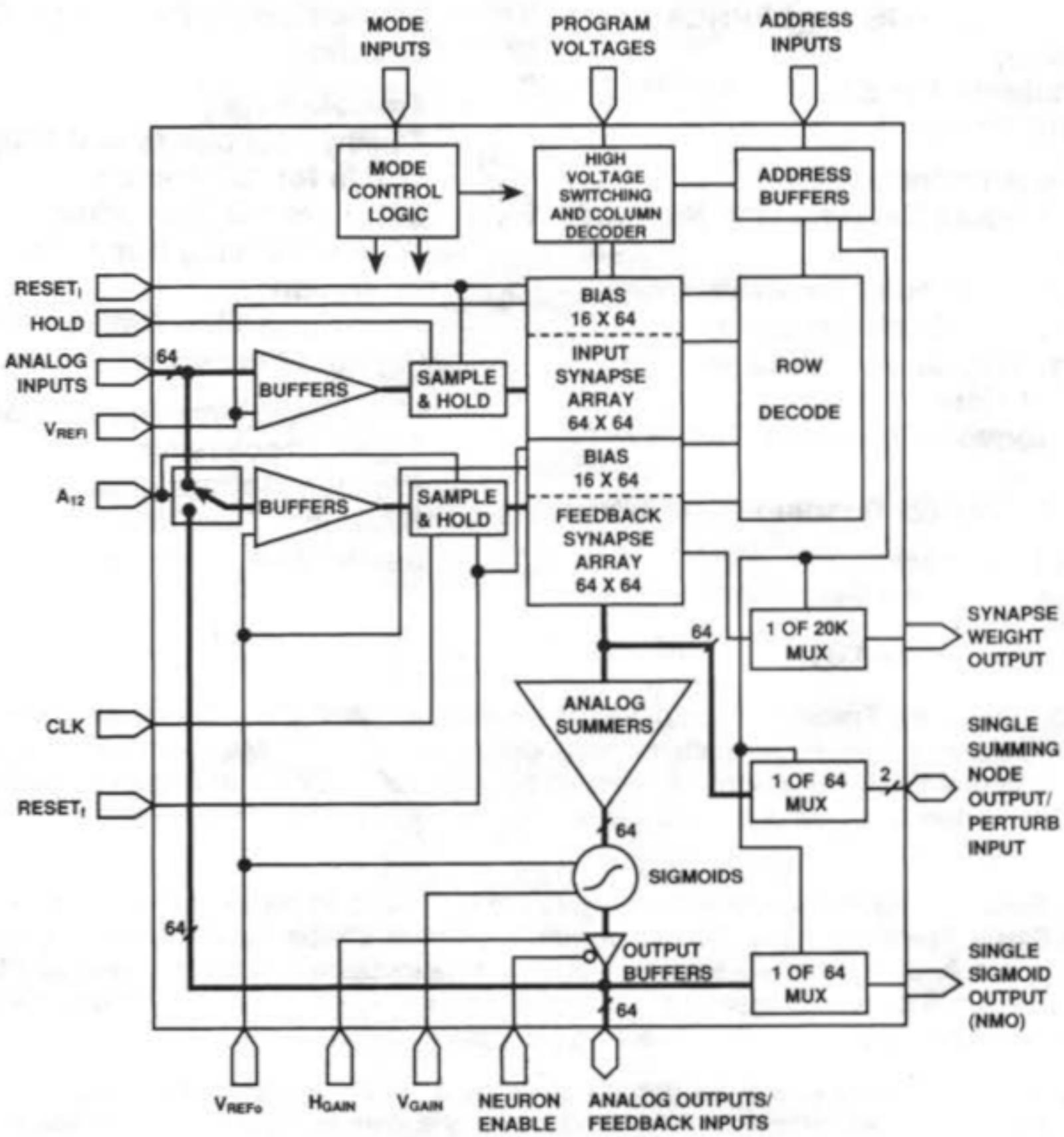
As shown in Figure 1, inputs and synapse weights are organized in two processing arrays—a 64x64 Input Array and a 64x64 Feedback Array. The chip has 64 external analog inputs. These inputs can be routed either to the Input Array or to the Feedback Array. Using input-hold functions, sets of 64 inputs can be multiplexed to the two arrays for mapping 128 inputs into 64 neuron outputs. Using these inputs, clocked feedback functions can support many interesting processing options.

Both the Input and Feedback Arrays are accompanied by a 16x64 bias array, for a total of 10,240 synapses. The chip computes the dot product (inner product) of the analog inputs and the weights stored at the synapses. The results are summed along each of the 64 neurons, generating 64 scalars that are passed through a sigmoid function of each neuron to the outputs.

Fully-parallel processing using both the Input and Feedback Arrays yields performance in excess of 2 billion multiply-accumulate operations (connections) per second (CPS), well beyond the capability of conventional CPUs. By interconnecting eight 80170NX chips, systems can achieve more than 16 billion CPS, a performance level that exceeds most supercomputers.

Multiple neural network layers can be implemented with either a single 80170NX chip or with multiple chips. By segmenting the two 64x64 arrays, networks with more than two layers can be implemented on a single chip. Multiple 80170NX chips can be cascaded by feeding analog outputs of one chip directly into analog inputs of subsequent chips.

*CHMOS is a patented process of Intel Corporation.



290408i1

Figure 1. 80170NX Chip Block Diagram

Interfacing flexibility is provided through voltage-controlled I/O referencing and sigmoid gain controls. A high sigmoid-gain control (H_{GAIN}) supports fully digital 0-to-5-volt interfacing directly to digital CMOS or TTL logic.

The four-quadrant multiplier synapses, implemented with Intel's CHMOS III EEPROM technology, provide truly nonvolatile, reprogrammable weights without refreshing or battery-backup. EEPROM

reprogrammability supports iterative and application-specific learning, avoiding mask-process limitations.

A glossary of neural network terms is included at the back of this data sheet for your convenience.

PIN DESCRIPTIONS

Symbol	Type	Description and Function
I_0-I_{63}	I_A	NEURON INPUTS —fully buffered analog channels for stimulus-vector input, where input voltages range from 0V to V_{IA} (see D.C. Characteristics). Input voltages above V_{REFI} represent positive numbers while those below V_{REFI} represent negative numbers. Inputs and stored synapse weights can be either positive or negative. Four-quadrant multiplications are performed on the inputs and stored synapse weights. External inputs drive the Input Array and, for PDP/128-input operation, can also be multiplexed to drive the Feedback Array, in which case V_{REFO} defines \pm inputs for the Feedback Array.
N_0-N_{63}	I/O_A	NEURON OUTPUTS and FEEDBACK INPUTS —fully buffered analog outputs, representing neuron activity levels, ranging from 0V to $2*V_{REFO}$ (see D.C. Characteristics). For PDP/128-input operation, these same pins can be driven externally for clocked Feedback Array input when neuron outputs are disabled (see NE#).
CE#	I_D	CHIP ENABLE# —an active-low TTL input that enables the chip when CE# = 0. When CE# = 1, the chip is put into a power-down or standby state, and the outputs N_0-N_{63} are disabled.
NE#	I_D	NEURON ENABLE# —an active-low TTL input that enables neuron outputs N_0-N_{63} when NE# = 0. Output drivers are disabled when NE# = 1, allowing several 80170NX chips to share a bus. NE# allows clocking of data into the Feedback Array's sample and hold buffers from an off-chip source.
V_{REFI} , V_{REFO}	I_A	INPUT AND OUTPUT REFERENCE VOLTAGES —voltage-reference inputs that define the 0 or null voltage levels independently for inputs and outputs. Voltages above these references represent positive numbers while those below are negative. The maximum output voltage is $2*V_{REFO}$ (see D.C. Characteristics). Setting V_{REFI} and V_{REFO} to 1.4V allows TTL compatible operation. V_{REFO} is equivalent to the Input Array's V_{REFI} .
V_{GAIN}	I_A	GAIN CONTROL VOLTAGE —a voltage-reference input for controlling analog sigmoid gain. An input voltage from 0V to V_{CC} adjusts the sigmoid gain from the minimum to the maximum. Sigmoid response to V_{GAIN} saturates at about 3.5V. All neurons' gains are set simultaneously.
H_{GAIN}	I_D	HIGH GAIN MODE —a TTL input that turns on the high-gain sigmoid characteristic when brought active-high. This in effect converts the sigmoid into a comparator for 0V or 5V output.
HOLD	I_D	HOLD INPUT —a TTL input that causes neuron inputs to be sampled and held when taken high. Input levels are captured and capacitively stored on the rising edge of HOLD. The inputs are then driven into the Input Array while HOLD is high. I_0-I_{63} resume driving the Input Array directly when HOLD goes low.

PIN DESCRIPTIONS (Contd.)

Symbol	Type	Description and Function
RESET_F	I _D	RESET FEEDBACK —a TTL input that resets the Feedback Array inputs and associated bias inputs to 0V when brought active-high. This effectively disconnects the Feedback Array from the neuron summing columns.
RESET_I	I _D	RESET INPUT —a TTL input that resets the Input Array inputs and associated bias inputs to 0V when brought active-high. This effectively disconnects the Input Array from the neuron summing columns.
CLK	I _D	CLOCK —the TTL input that causes outputs to be sampled and held at inputs to the Feedback Array on the rising edge of CLK. Outputs are never connected directly back to inputs of the Feedback Array; they are only transferred synchronously in successive CLK cycles. When A ₁₂ =1, the sample function in the input buffers for the Feedback Array is disabled. The CLK pin then behaves like the HOLD pin, but with inverted polarity. In addition, the feedback path is disconnected, and the external inputs are multiplexed into the Feedback Array input buffers.
A₁₂	I _D	INPUT MULTIPLEXER —in the forward-propagation mode, address pin A ₁₂ has the function of enabling the multiplexing of input pins into the Feedback Array. When A ₁₂ is brought active-high, the path from output pins to the Feedback Array is disconnected, while a path from the input pins to the Feedback Array is enabled. At the same time, the function of the CLK pin is modified so that it behaves like an inverted-polarity HOLD pin.
A₁₂-A₀	I _D	ADDRESSES —TTL-compatible inputs used to select the individual synapses and neurons. The lower 6 bits (A ₅ -A ₀) are column/neuron addresses and the upper 7 (A ₁₂ -A ₆) are row/input addresses, which together select one synapse. A ₁₂ has a special function as an address pin and input multiplexer pin as described above.
LRN	I _D	LEARN CONTROL —a TTL input for enabling the learning modes used to change the values of the synapse weights.
SSYN	I _D	SYNAPSE SELECT —a TTL input that enables the row-decoding circuitry for applying program, erase, or input voltages to the row selected by A ₁₂ -A ₆ and BIAS pins. SSYN also activates the column-select circuitry for accessing a single synapse (per addresses A ₅ -A ₀) through the neuron and summing-node multiplexers in SYNR (Synapse Read) mode.

PIN DESCRIPTIONS (Contd.)

Symbol	Type	Description and Function
WT	I _D	WEIGHT-CELL SELECT —a TTL input for selecting one or the other of the two EEPROM cells in the selected synapse (referred to as the <i>weight</i> EEPROM cell or the <i>reference</i> EEPROM cell). WT = 1 selects the <i>weight</i> EEPROM cell, WT = 0 selects the <i>reference</i> EEPROM cell.
INC	I _D	INCREASE or DECREASE SYNAPSE WEIGHT —a TTL signal that specifies whether to increase (INC = 1) or decrease (INC = 0) the threshold voltage of the floating-gate EEPROM cell selected by the WT pin. Increasing a synapse weight corresponds to decreasing (erasing) the threshold of the corresponding <i>weight</i> EEPROM cell and/or increasing (programming) the threshold of the corresponding <i>reference</i> EEPROM cell. Similarly, decreasing a synapse weight corresponds to increasing the threshold of the <i>weight</i> EEPROM cell and/or decreasing the threshold of the <i>reference</i> EEPROM cell. The difference in threshold voltages between the <i>weight</i> and <i>reference</i> EEPROM cells is called V _{T_DIFF} . The relationships are: Weight = [V _{T_REF} - V _{T_WT}] and V _{T_DIFF} = - Weight Typically, the weight-setting algorithm maintains a constant common-mode threshold voltage for all EEPROM cells.
BIAS	I _D	BIAS-SYNAPSE SELECT —a TTL input to select the nonvolatile bias synapses. It is used in conjunction with the A ₅ -A ₀ addresses to select one of 64 neurons and with A ₉ -A ₆ addresses to select one of 16 rows of bias synapses. A ₁₂ determines the group of 16 rows selected: A ₁₂ = 0 for Input Array bias, A ₁₂ = 1 for Feedback Array bias. Bias-row addresses 10–15 are reserved for Intel internal use. Users should make sure that these bias synapses are zeroed if using custom software.
V _{PP1} , V _{PP2}	S _A	PROGRAMMING VOLTAGE SUPPLIES —V _{PP1} and V _{PP2} are high-voltage pulse inputs required for training (see A.C. and D.C. Characteristics and Weight Setting). V _{PP1} and V _{PP2} are set to V _{CC} during normal operation. They are connected to the gate and source of the EEPROM device during the WVTR (Weight V _T Read) or RVTR (Reference V _T Read) modes.
QUER	I _D	QUERY —a TTL input that enables the query mode when QUER = 1. QUER activates interrogation of synapse weight and sum measurement.

PIN DESCRIPTIONS (Contd.)

Symbol	Type	Description and Function
SMO+, SMO-	O _A	<p>SUMMING-NODE OUTPUTS—differential analog outputs of the neuron summing nodes in the SMR (Sum Read) mode. SMO+ and SMO- are connected to the selected neuron by the A₅-A₀ addresses for direct measurement of the differential voltage representing:</p> $\text{Net}_j = (\text{SMO}+) - (\text{SMO}-) = \sum \{\text{Weights} * \text{inputs}\} + \text{bias}$ <p>Differential currents can also be read, but only for a very limited range.</p> <p>In the SYNR (Synapse Read) mode, addresses A₁₂-A₀ and BIAS select an individual synapse to obtain its response to a corresponding input. The output is a differential voltage or current across the SMO± pins. SMO+ and SMO- are also used in the SMR/P (Sum Read/Perturb) mode to allow adding or subtracting increments of differential current from a selected neuron's summing nodes for use with the Madaline learning algorithm (see Training).</p>
NMO	O _A	NEURON OUTPUT —a multiplexed analog output connected to the neuron output selected by addresses A ₅ -A ₀ in the NR (Neuron Read) mode. Selective sensing of neuron outputs facilitates implementations of back propagation and other learning algorithms with minimal hardware overhead. NMO is essential for system debugging.
SYNO	O _A	SYNAPSE OUTPUT —an analog output connected to one of the selected synapse's two EEPROM-cell drains for threshold measurement in WVTR, RVTR, or V _T Read modes. The synapse is selected with the A ₁₂ -A ₀ , BIAS and WT inputs (see Weight Setting).

KEY: S = Power Supply A = Analog O = Output
 I = Input D = Digital

PRINCIPLES OF OPERATION

Figure 2 is a simplified model of a typical biological neuron cell. The neuron has several inputs attached to its body at points called "synapses." The variable-strength synapse connections can be either output-exciting or output-inhibiting. Input signals reaching the neuron body through these variable-strength synapses are summed by the neuron body. When the sum exceeds a threshold, the neuron output is activated.

The 80170NX chip is designed around this basic neuron model. The chip normally operates as 64 parallel processors, where each such processor is analogous to this neuron model. Figure 3 illustrates the 80170NX chip architecture as an array of external inputs (μ_i) and neuron outputs (v_j). The intersections of external inputs and neurons are connected through synapses that consist of a multiplier (X) and a weight-storage unit (W_{ij}). W_{ij} is the weight of the connection from input μ_i to output v_j . Figure 3 illustrates only one array of intersections; the chip has

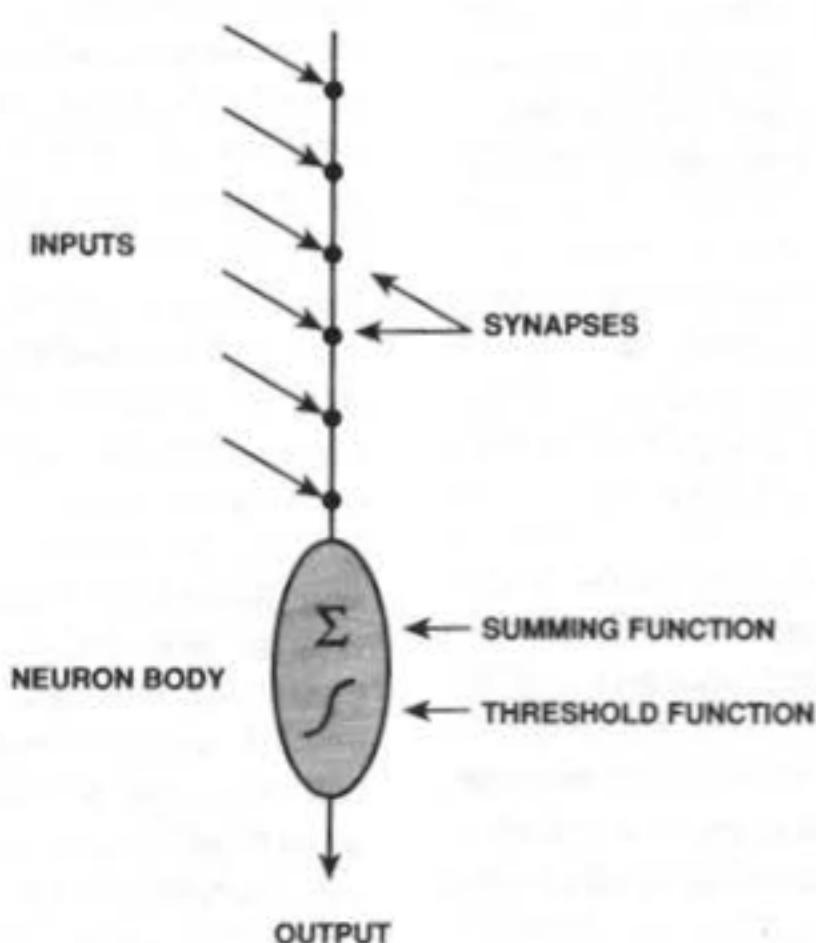
two such arrays—an Input Array and a Feedback Array. In addition to these arrays, each neuron is connected to fixed-input bias synapses (B_{kj}) in each of the two arrays.

This normal mode of operation is called Parallel Distributed Processing (PDP). In PDP operation, each neuron computes the dot (inner) product of the external inputs (μ_i) times their corresponding synapse weights (W_{ij}), and adds the sum of the fixed-input bias synapses (B_{kj}). This dot product is then compressed through a sigmoid function, as follows:

$$v_j = \text{sigmoid} \{ \sum_i (\mu_i * W_{ij}) + \sum_k B_{kj} \}$$

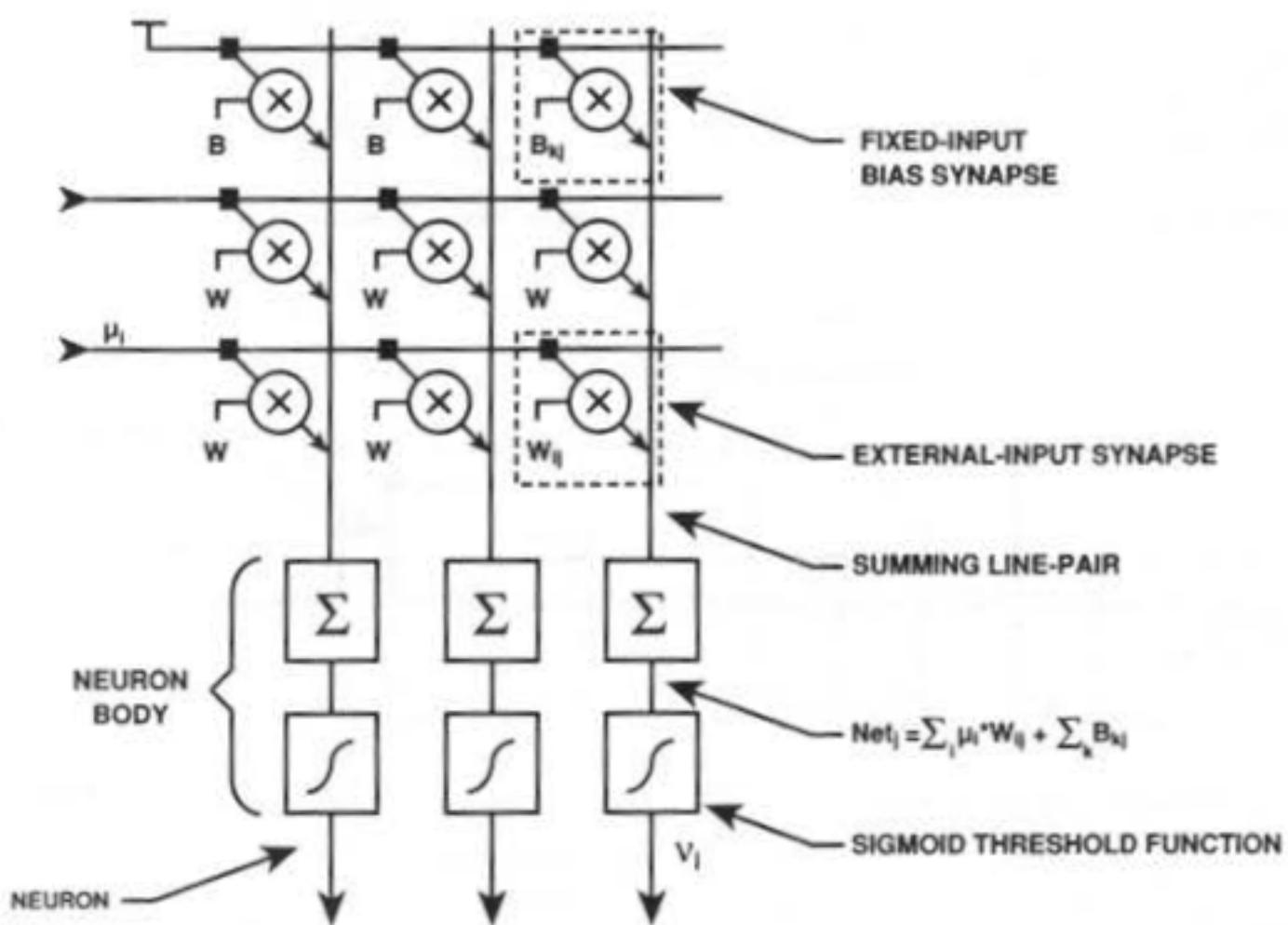
The neuron output (v_j) is a scalar quantity that indicates how closely the input vector matches the stored-weight vector. The sigmoid function through which the result of the dot products pass is illustrated later in Figures 19, 20 and 21.

The Input and Feedback Arrays each have 16 fixed-input bias synapses. Mathematically, all of these bias synapses act as a single bias to provide the



290408i2

Figure 2. Model of Biological Neuron Cell



290408i3

Figure 3. Typical Neural Net

threshold for the neuron's sigmoid function and to cancel offsets due to physical non-linearities in the chip. However, since the dynamic range of a single EEPROM bias synapse is limited, 16 of them are used. All 16 contribute to the dot product performed by the neuron, although seven are set by the training program, and the other nine are set by Intel's chip-initialization process.

The weights of all synapses—external-input synapses and fixed-input bias synapses are programmed by an external training system. Intel provides such a training system, as described in the TRAINING OPERATIONS section on page 22.

The chip uses Intel's fast CHMOS-III EEPROM technology. Synapses are implemented as differential EEPROM cell pairs, as shown in Figure 4. This implementation provides much more compact storage than conventional digital EEPROM memory because synapse weights are stored as analog values on the gates, and because analog multipliers are more compact than digital multipliers. Typical weight resolution of the differential EEPROM cell pairs is 6 to 7 bits. Worst-case resolution over a lifetime of 10 years is at least 4 bits, and greater than 6 bits can be obtained if the Bake-Train-Bake

Method is used. (See the Bake-Train-Bake section on page 29.)

Synapse weights are stored as analog transconductance values, each producing an analog output current from an analog input voltage and a stored-weight voltage. Currents generated by each synapse along a pair of summing lines (shown as a single column in Figure 3) are summed, converted to a voltage, and passed through a sigmoid function with voltage-controlled gain. When both Input Array and Feedback Array are active, the output corresponds to the sum of their two dot products. Identical mechanisms in each of 64 columns correspond to the 64 neuron outputs. The sigmoid function (illustrated in Figure 19) is a threshold function where dot products below the neuron's output threshold indicate a poor match between input and stored weight vectors. Conversely, neuron output will be high for dot products above the neuron's threshold. Each neuron's threshold can be changed independently by the learning algorithm, allowing different match criteria for different vectors. Changing bias-synapse weights alters the sum of currents, shifting the neuron threshold.

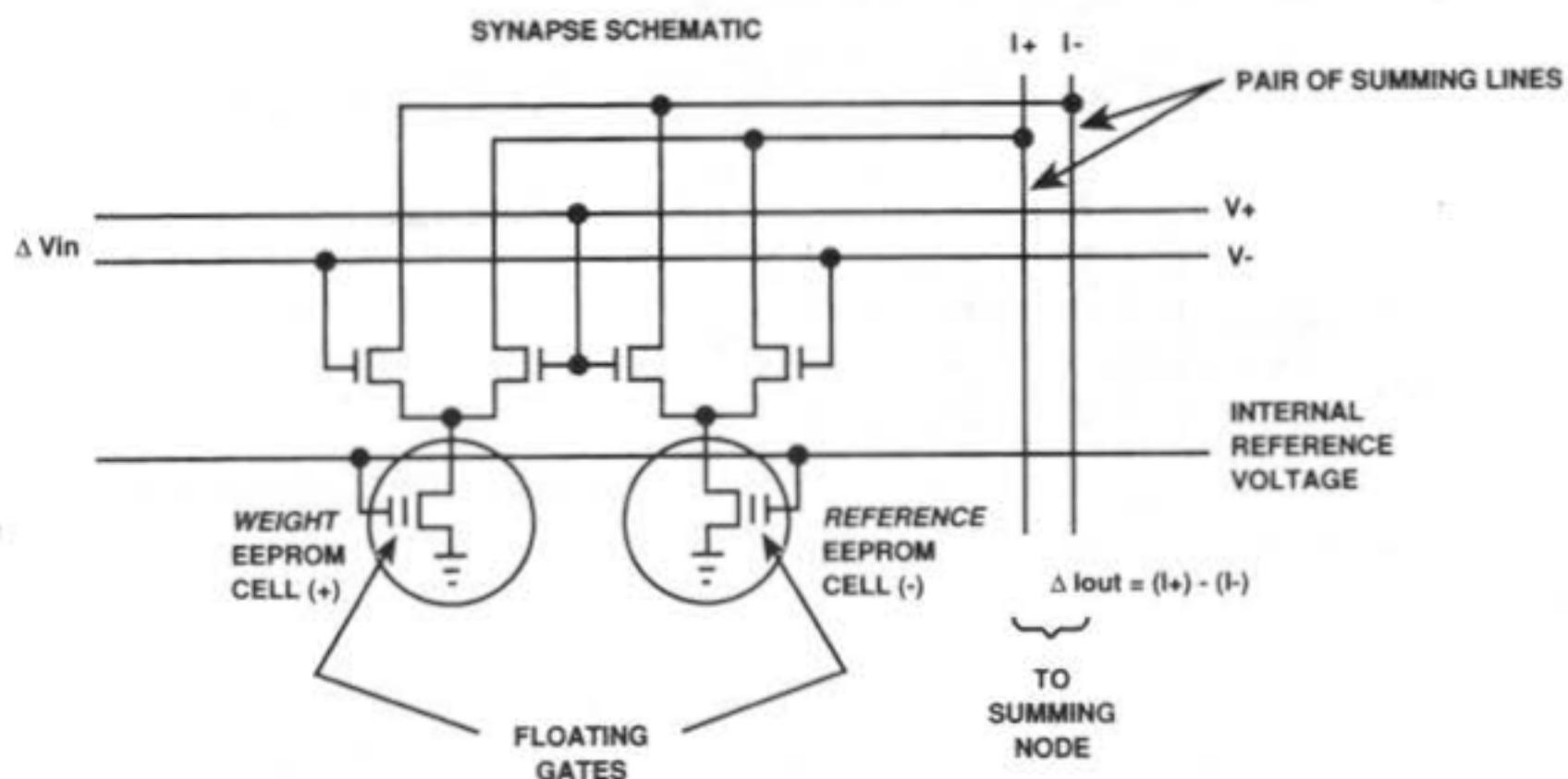
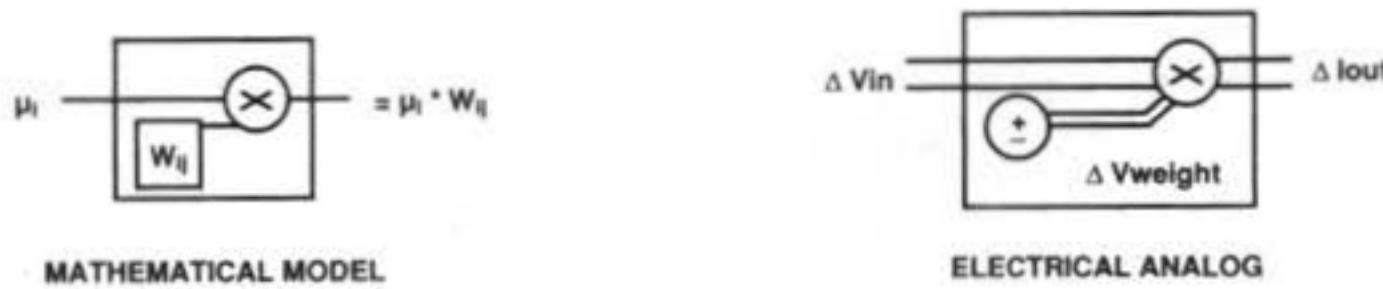


Figure 4. Synapse Implementation

As Figure 4 illustrates, each synapse contains a multiplier and a stored weight, and calculates one product. The synapse product is summed with those from other synapses along the summing lines to produce the dot (inner) product in the form of a differential current, I_+ and I_- . The synapse weights are stored as a difference in floating-gate thresholds of a pair of EEPROM cells. Changing weight values involves altering the charge stored on the floating gates. These modifications represent variable-weight strength as well as polarity. Since the floating gates of an EEPROM cell pair have equal capacitance, unequal charge causes differences in floating-gate voltages.

The synapse's multiplier circuit generates a differential current, I_+ and I_- , proportional to the product of the differential floating-gate voltage, and the differential-input voltage, ΔV_{IN} . The larger the differential voltages, the larger the differential current generated. Thus, the input and weight products are mapped into associated differential currents. Taking advantage of Kirchoff's Law, many such currents can be summed along a pair of bit lines. The sum is converted to a voltage by load devices attached to the bit lines.

PROCESSING CONFIGURATIONS

The chip supports many neural network configurations. For example, several 80170NX chips can share an input bus, capturing different time slices by sequential activation of the **HOLD** function. Taking the **HOLD** pin high capacitively stores 64 analog input levels for each time slice. Many forms of signal processing (such as speech processing) can use this sampling method in which input data is gathered from different points in time.

Similarly, 64 analog voltages can be directed to the lower half of the synapses the—Feedback Array. Figure 5 shows a simplified block diagram of the input multiplexing circuits. The path from the input pins I_0-I_{63} to the Input Array buffers is always connected. In addition these 64 signals can also be multiplexed from the input pins to the Feedback Array buffers. In PDP operation the A_{12} address pin enables input multiplexing by creating a path from the input pins to the Feedback Array buffers, and disables the path from the output pins to the Feedback Array buffers. The **HOLD** and **CLK** signals control the propagation of the input signal to the

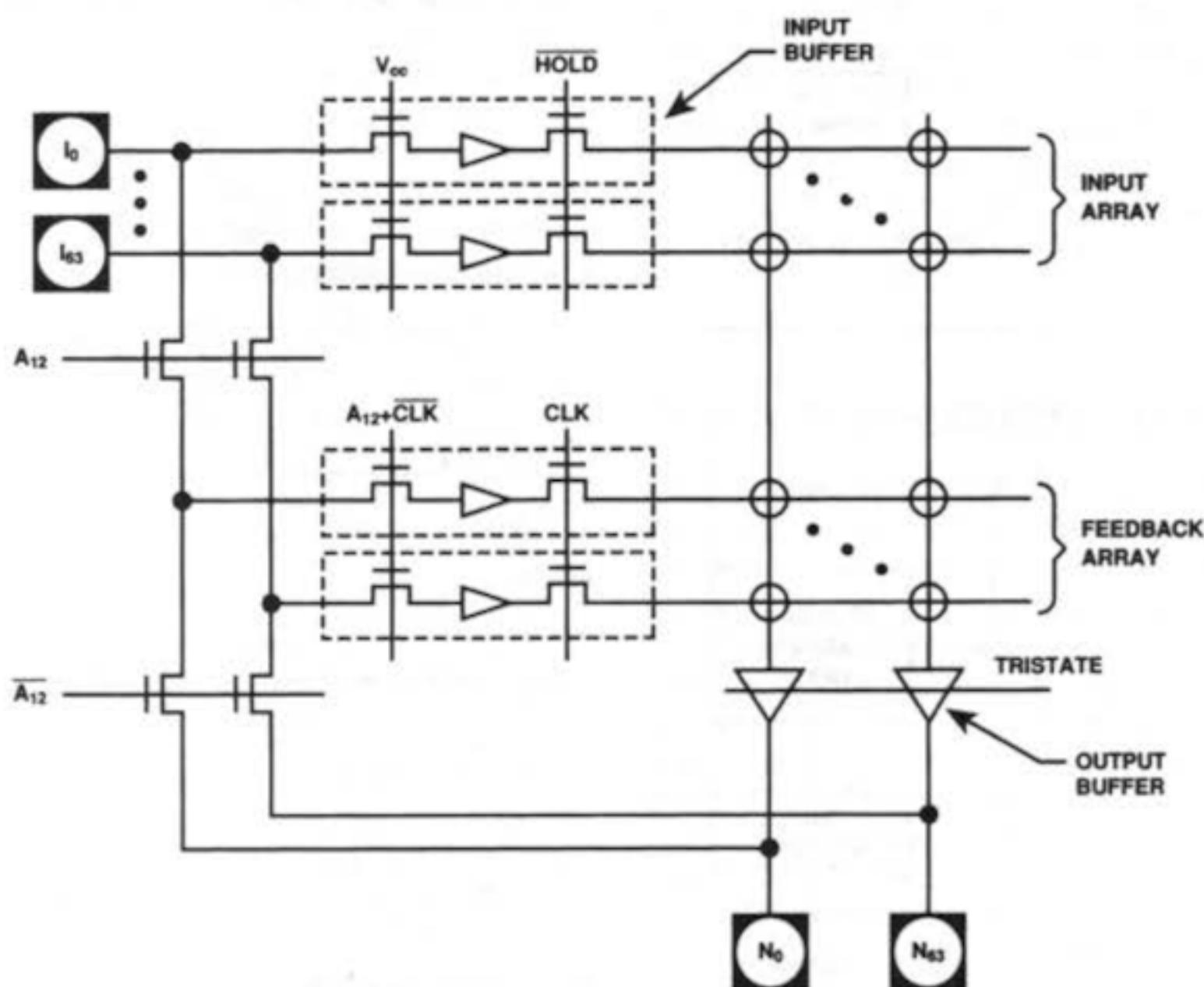


Figure 5. Input Multiplexing Circuits

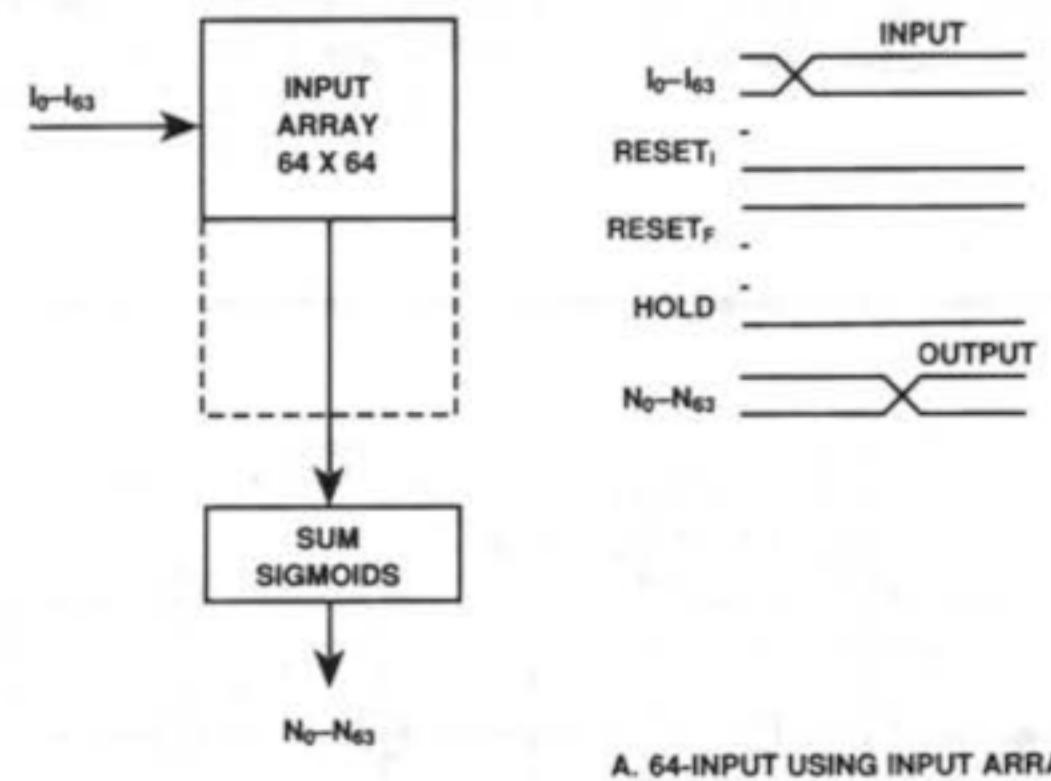
Input Array and Feedback Array, respectively. Similarly the **RESET_I**, and **RESET_F** disable their respective arrays.

Figures 6 through 10 illustrate some basic processing configurations. The simplest example is a 64-input single-layer feedforward network, shown in Figure 6(A). Here, 64-dimension input vectors (or "patterns") are directly mapped into 64-dimension output vectors, as determined by the synapse weights in the Input Array. In this minimal configuration, the Input Array inputs are enabled and the Feedback Array inputs are disabled by setting the **RESET_I** pin low and the **RESET_F** pin high. This is called the 64-input parallel distributed processing mode (PDP/64-input operation).

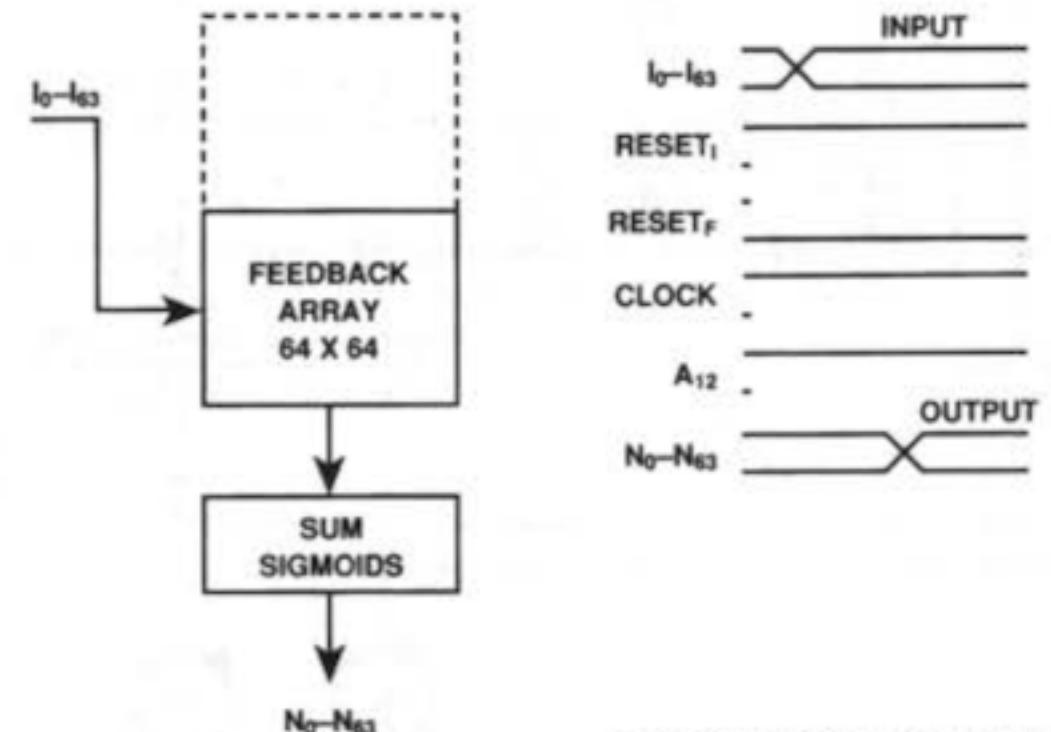
Figure 6(B) shows another variant of the PDP/64-input feedforward configuration. Here, the Feedback Array is driven directly from the external input

pins by taking **A₁₂** and **CLK** high, disabling the Input Array by taking **RESET_I** high, and enabling the Feedback Array by taking **RESET_F** low.

In addition to the chip's 64-input configuration (PDP/64-input operation), a 128-input configuration (PDP/128-input operation) can be implemented. In this configuration, shown in Figure 7, a second set of 64 inputs on input pins **I₀-I₆₃** is multiplexed to the Feedback Array. Each of the 64 neurons then calculates the dot product of both the Input Array and Feedback Array input vectors with their corresponding synapse weights. Pin **A₁₂** enables this multiplex mode. When **A₁₂** and **CLK** are set TTL-high and **RESET_I**, **RESET_F**, and **HOLD** at TTL-low, the first half of the input vector is applied to the input pins and routed to both the Input Array buffers and the Feedback Array buffers. When **HOLD** goes to TTL-high, the Input Array buffers store the analog levels and disconnect from the input pins. The second half



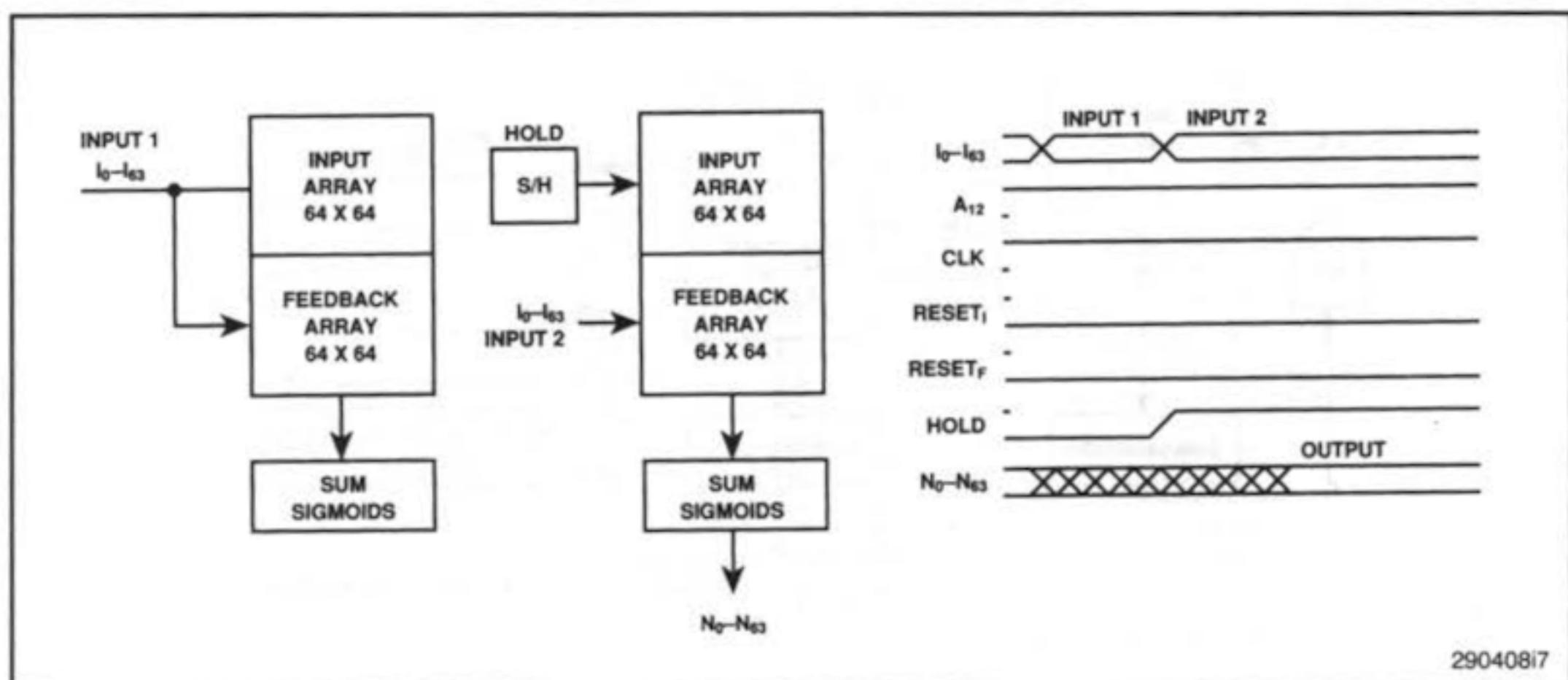
A. 64-INPUT USING INPUT ARRAY



B. 64-INPUT USING FEEDBACK ARRAY

290408i6

Figure 6. 64-Input Feedforward Configurations



290408i7

Figure 7. 128-Input Feedforward Configurations

of the input vector is applied to the input pins which propagates to the Feedback Array, and a valid output will appear at the output pins $N_0 - N_{63}$ one processing-delay later.

A single chip can perform two-layer operation by using the feedback connections and the clocked sample-and-hold features. Alternate activation of the two reset pins, **RESET_I**, and **RESET_F**, disables the corresponding Input Array and Feedback Array inputs, as shown in Figure 8(A). During Input Array processing, Feedback Array inputs are reset by taking **RESET_F** high. The outputs then reflect the dot product of the inputs and the synapse weights stored in the Input Array. After a **CLK** pulse, which initiates processing by the Feedback Array, neuron outputs are driven into the feedback inputs of the Feedback Array. To prevent the external inputs at input pins from interfering with Feedback Array processing, inputs are disconnected from the Input Array by taking **RESET_I** high. The fixed internal inputs to the bias synapses are also reset when **RESET_I** or **RESET_F** are disabled.

Hopfield networks are supported with the clocked feedback capability shown in Figure 8(B). This requires applying alternating **CLK** and **NE#** pulses. Continuous feedback is also supported and can be obtained by driving the **A₁₁** pin to $12V \pm 0.5V$, the **A₁₂** pin to $0V$ and the **CLK** pin to TTL high.

Feedback connections can be used to recognize time-dependent input sequences, as in Figure 9(A). Here, results from the Input Array are fed back for multiple clock cycles to the Feedback Array. Figure 9(B) shows a configuration capable of generating sequences from a fixed input pattern.

Figure 10 combines the PDP/128-input operation with two-layer processing. The Input and Feedback Arrays are combined to allow up to 128 inputs and up to 64 outputs, as shown in Figure 10(B). If a 128-by-j matrix is used for the first layer of processing then a j-by-(64-minus-j) matrix located in the Feedback Array, as shown in Figure 10(B), can be used for the second layer of processing. The number of processing elements in the hidden layer, j, is defined as an integer that is less than 64. The outputs of the first cycle are clocked back into the Feedback Array; only the first j outputs of this cycle are valid. These outputs are multiplied by the j-by-(64-minus-j) matrix in the second cycle. Then, the outputs of the second cycle are read at the output pins $N_j - N_{63}$. Those portions of the synapse array outside the two matrices defined above, marked by "0" in Figure 10(B), must be set to weights of zero.

MULTI-CHIP CONFIGURATIONS

One of two multi-chip interconnection configurations can be used: direct pin-to-pin interconnection, as shown in Figure 11(A), or bus interconnection, as shown in Figure 11(B). Direct pin-to-pin wiring is the more straightforward since no multiplexing is required. The busing example in Figure 11(B) shows a common, multiplexed bus for all communication. Alternatively, separate buses could be used between processing layers and for inputs and outputs. Bus interconnection provides a more flexible development environment; the clocking sequence of the **NE#**, **HOLD**, and **CLK** pins allows changes in data routing. The bus configuration has two drawbacks compared to direct interconnect: (1) data can be moved only in complete 64-neuron blocks, and (2) multiplexing is somewhat slower.

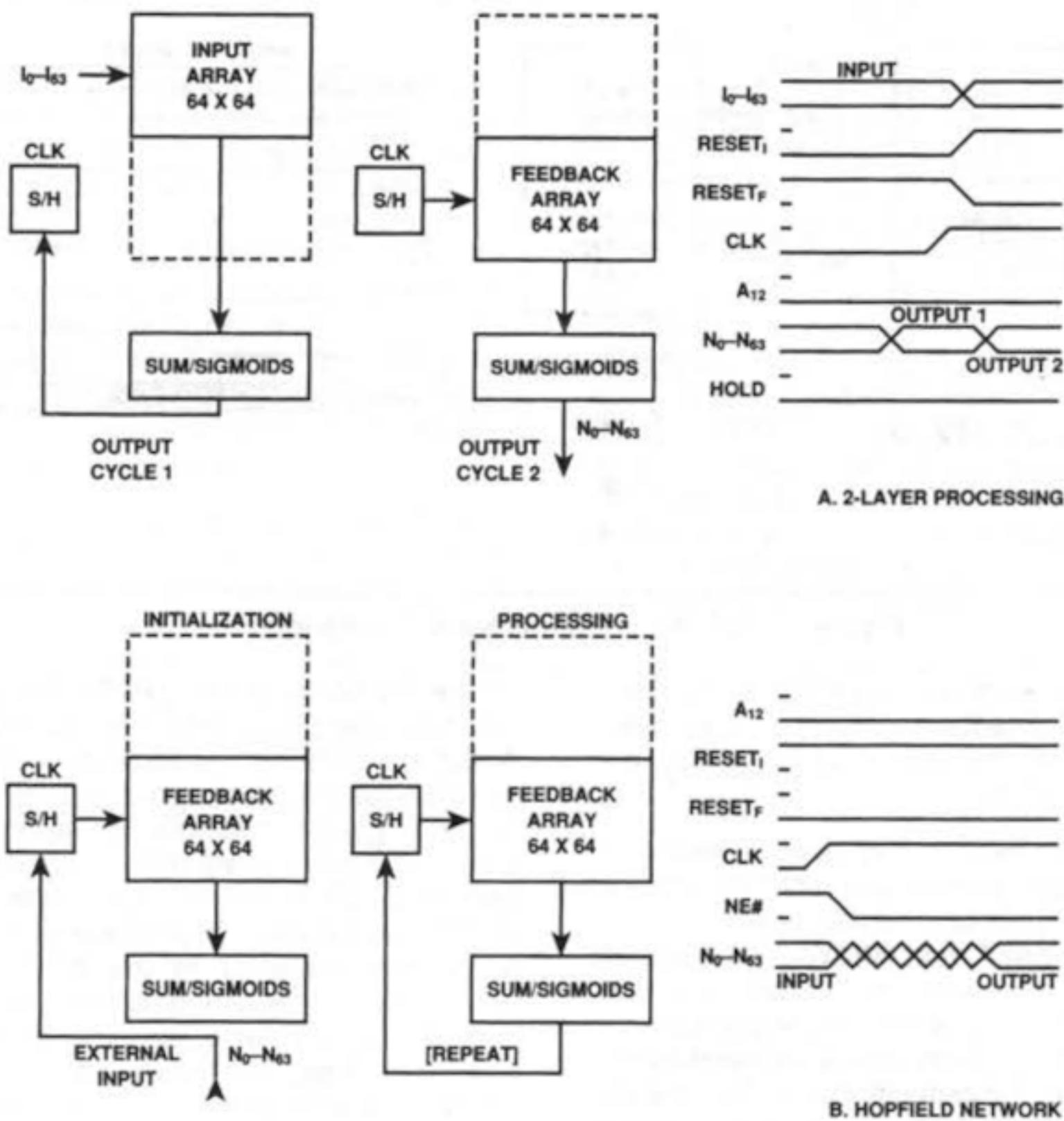


Figure 8. 64-Input Feedback Configurations

Intel offers a multi-chip prototyping board for 80170NX chips—the *ETANN Multi-chip Board (EMB)*. The EMB supports both pin-to-pin and bus interconnection for prototyping up to eight chips. It is an add-on to the iNNTS training system, described further on in the TRAINING section. Multi-chip versions of the required hardware drivers, linking subroutines, and network simulation programs are provided.

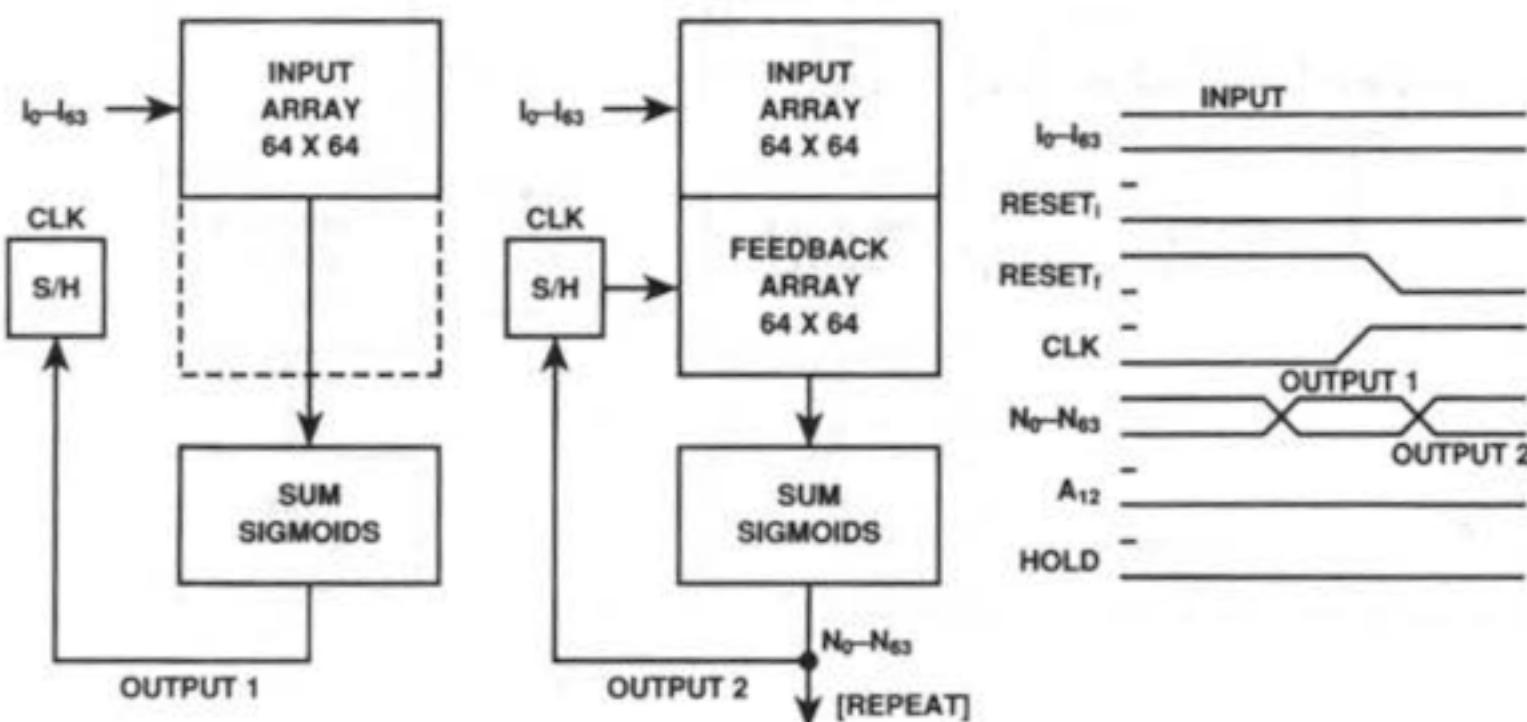
SYNAPSE CHARACTERISTICS

Synapse Weight Operation

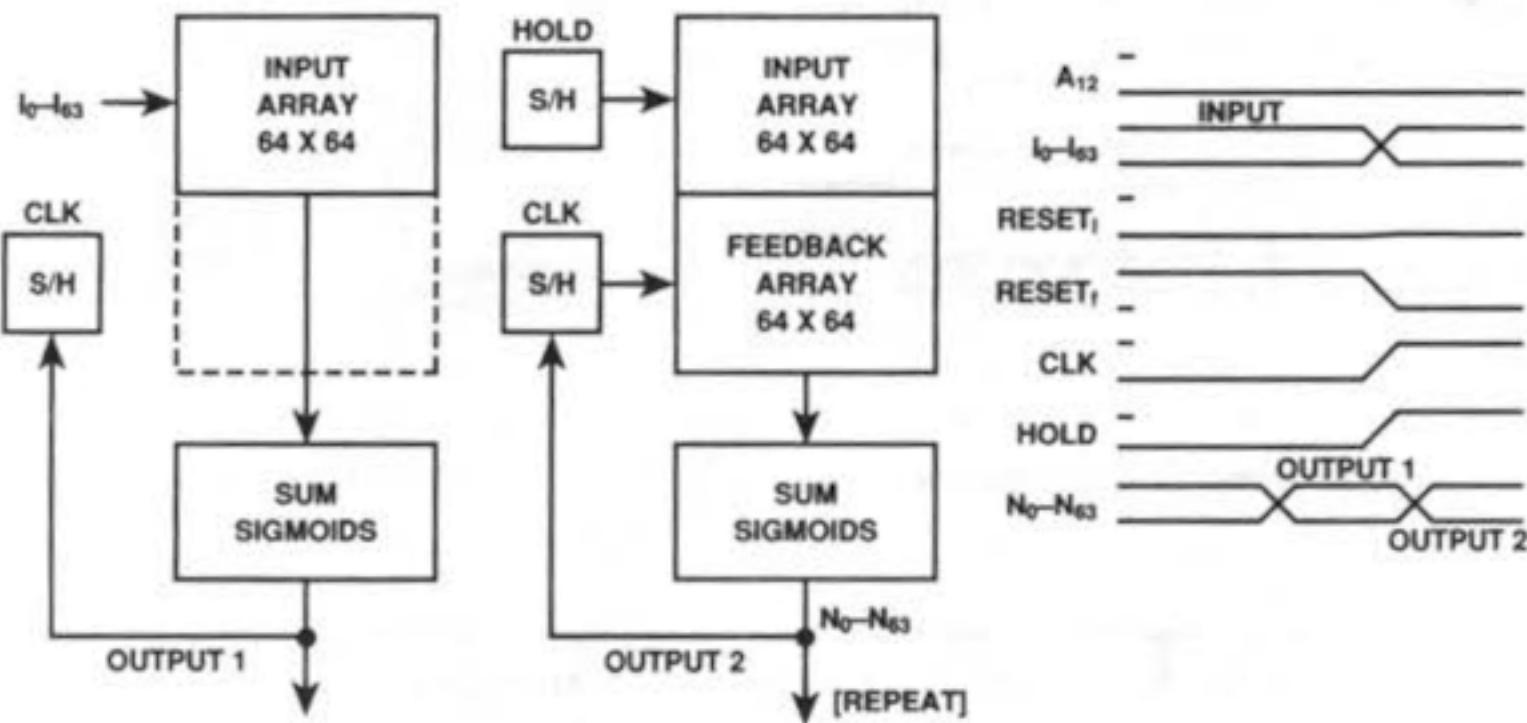
To change a synapse weight, the difference in voltage between the two EEPROM floating-gate cells is changed by adding charge to one cell and removing charge from the other using Fowler-Nordheim tunneling. Tunneling is a physical mechanism commonly used in conventional EEPROM cells; high voltage applied across a thin dielectric causes elec-

trons to tunnel directly through the dielectric, to or from a floating gate. The weight-modification voltage pulses induce Fowler-Nordheim tunneling in a selected synapse (see Figure 12).

A measurable indicator of the floating-gate voltage is the threshold voltage (V_T) of the floating-gate cell itself. Figure 12 shows the floating-gate cell's current-voltage characteristics before and after a weight modification, along with the corresponding V_T 's. The V_T is the control-gate voltage at which the cell begins conducting a given amount of current, typically 1 to 10 μ A. In this example, the floating-gate voltage becomes more negative as electrons are added, corresponding to an increase in the cell V_T , since a more positive control-gate voltage is required to overcome the effect of the added electrons on the floating gate.



A. SEQUENCE RECOGNITION



B. SEQUENCE GENERATION

290408i9

Figure 9. 64-Input Sequence Recognition and Generation Configurations

Synapse weights are represented by threshold-voltage differentials ($V_{T,DIFF}$) between the reference and weight EEPROM cells, where Weight = [$V_{T,REF} - V_{T,WT}$] and $V_{T,DIFF} = -$ Weight. All weight-changing procedures refer to V_T changes, which are measured by using the V_T read modes, WVTR and RVTR, as described in the MODE DESCRIPTIONS section.

Synapse Multiply Characteristics

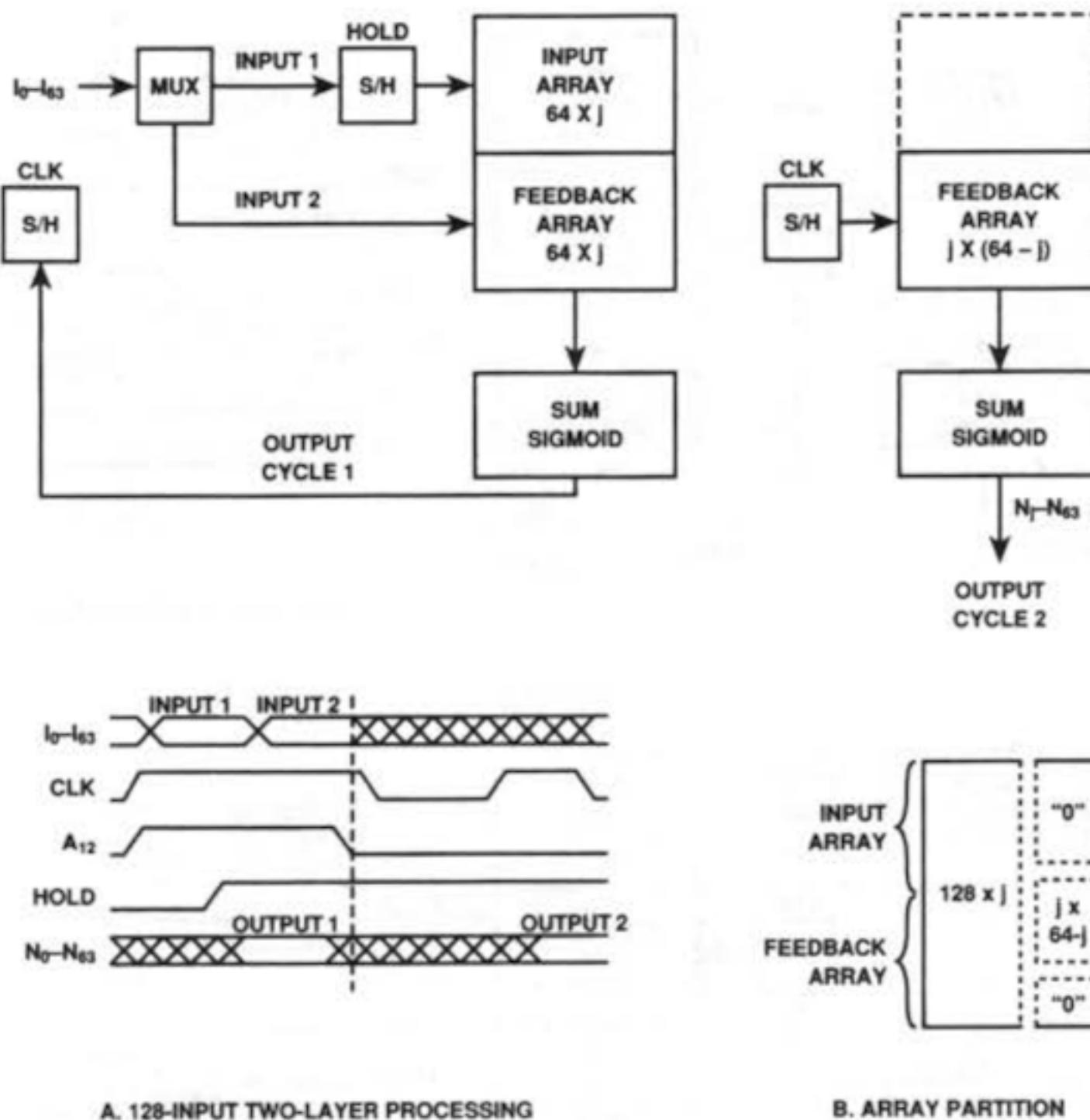
Figure 13 shows a typical four-quadrant multiply characteristic of a synapse. The output differential current is plotted as a function of the input voltage for eleven values of weights. A constant common-mode V_T among the two EEPROM cells is maintained in these results. The response is linear over

most of the weight range. (See the DC CHARACTERISTICS section on page 35 for a definition of $V_{T,DIFF}$, $V_{T,COM}$, $V_{T,WT}$, and $V_{T,REF}$.)

MODE DESCRIPTIONS

Signature Read

The 80170NX chip contains a signature code to differentiate it from other chips. There are two ways to read the code. When A_9 goes to high voltage (12V \pm 0.5V), the device is recognized as an Intel component when 89H is read from pins N_0-N_7 and as an 80170NX chip when CFH is read from pins N_8-N_{15} . The other option is to toggle through 16 bits at the NMO pin, reading 89H followed by CFH.



290408i10

Figure 10. 128-Input Feedback Configuration

PDP – Parallel Distributed Processing Mode

Parallel distributed processing is the normal processing operation, described in the Principles of Operation section on page 6. As shown in the Processing Configurations section on page 8, the input derivation is configuration-dependent, yielding a number of different processing characteristics.

WMI/WMD – Weight-Modify Increase and Weight-Modify Decrease Modes

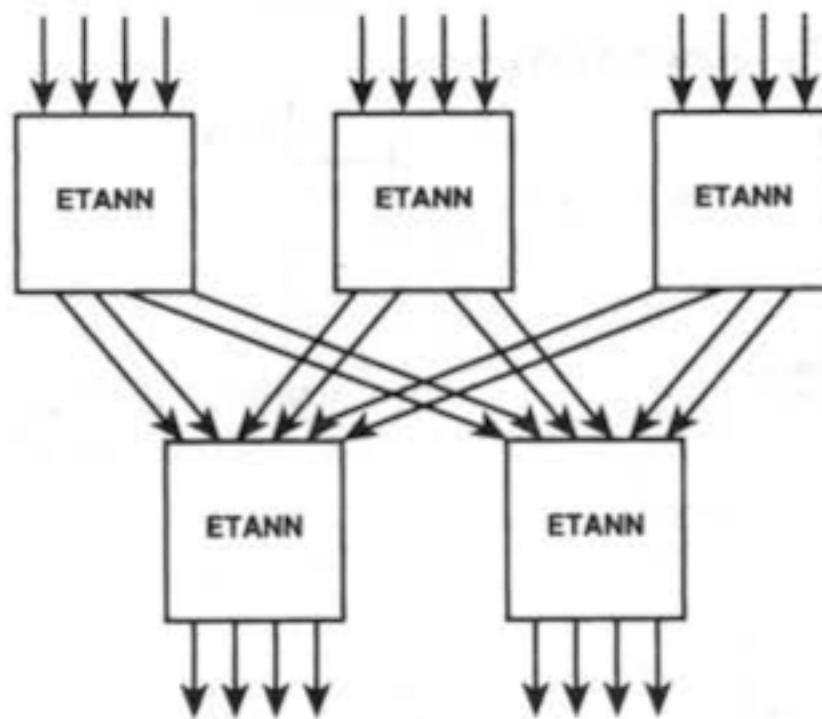
When the LRN pin is set to 1 the chip enters the learning, or training, mode. A synapse is selected by addresses A_{12} – A_0 and **BIAS**. Synapse weights are changed by first measuring the EEPROM-cell threshold (the stored weight) by using the WVTR mode. A calculation is then made of the voltage and pulse width needed for the desired weight modification. By pulsing the calculated high voltages on pins V_{PP1} and V_{PP2} , the weights are modified one at a time. For details, see the next section, TRAINING OPERATIONS on page 22.

The weight-setting algorithm applies multiple pulses in the process of setting a weight by using the error observed after each pulse and then making a better estimate for the next pulse. The algorithm also carries what it has learned in the process of setting one weight to the setting of the next weight. This adaptation is important for reducing the number of pulses and hence the total weight-setting time.

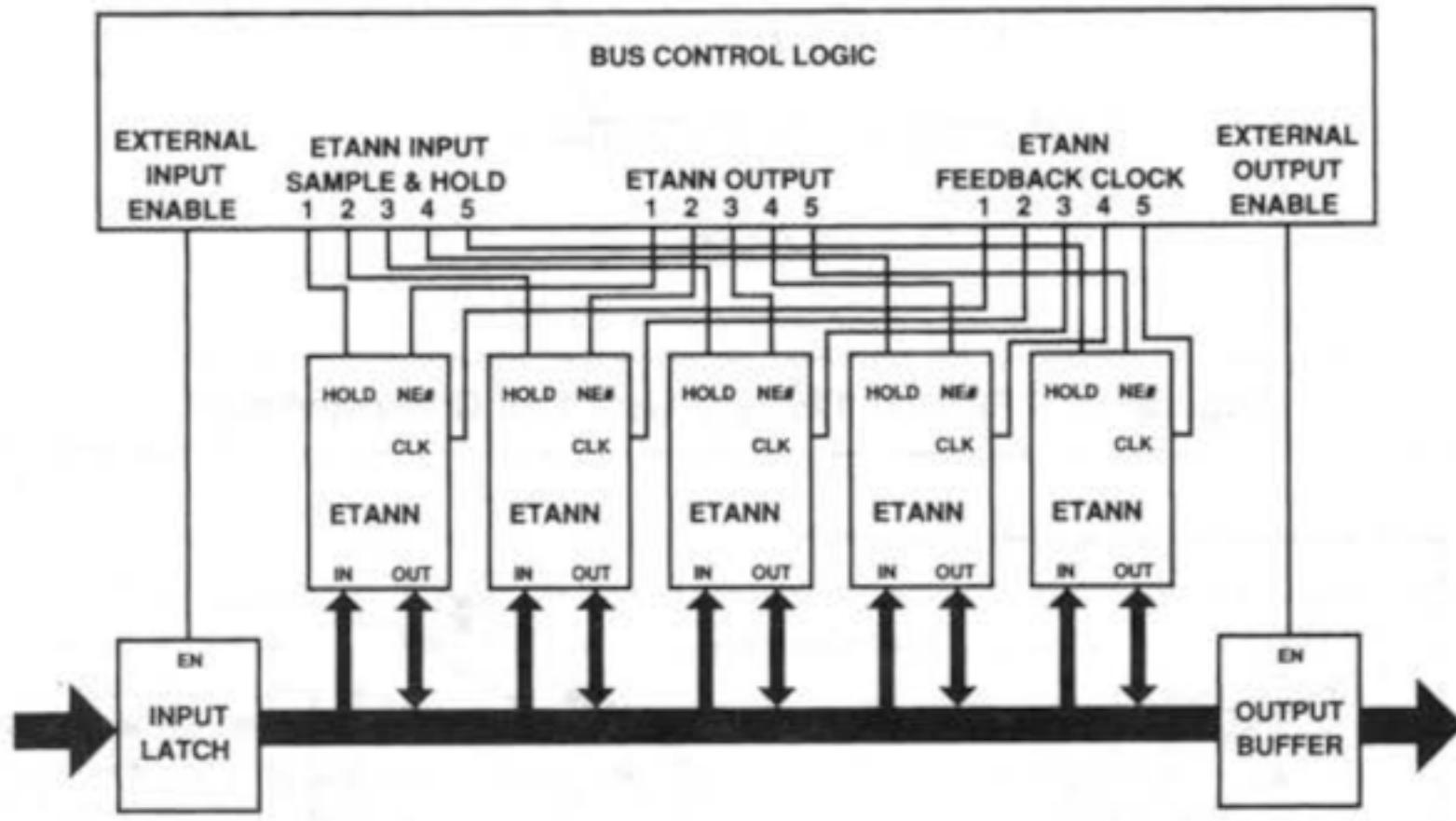
WVTR – Weight V_T Read Mode

A specific floating-gate device storing a desired synapse weight is selected by addresses A_{12} – A_0 , **BIAS**, and **WT** when in WVTR mode. The device's drain terminal is connected to the **SYNO** pin by on-chip decoding circuitry to allow V_T measurement as shown in Figure 14.

As an example, during V_T measurement, 3.0 volts is applied to **SYNO** while monitoring the current into **SYNO**. V_{PP1} and V_{PP2} supply the gate and source voltages, respectively, to the selected floating-gate transistor. V_T is measured by setting V_{PP2} (device source) to 2.0 volts and incrementing V_{PP1} (the



A. DIRECT PIN-TO-PIN INTERCONNECTION



B. BUS INTERCONNECTION

290408i11

Figure 11. Multi-Chip Configurations

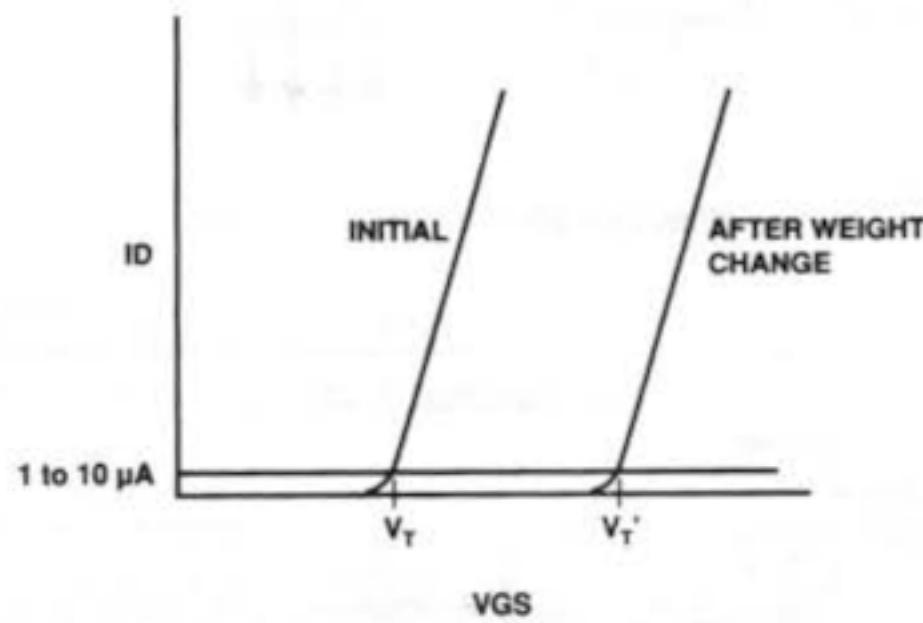
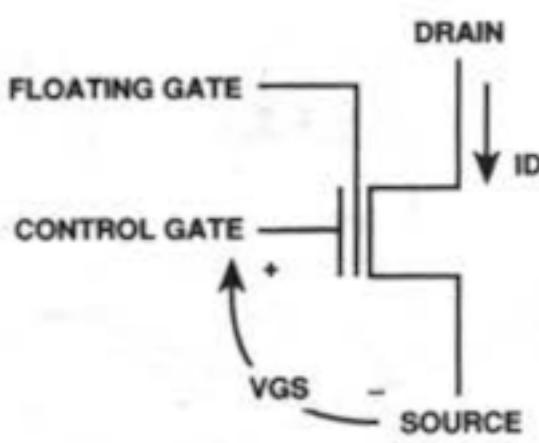
gate) from 0 to 4V until the SYNO (drain) current reaches the specified level (typically 1 to 10 μ A). The gate-to-source voltage, V_{PP1} minus V_{PP2} , that generates this current is the V_T . For example, given a 0.25 volt *reference* EEPROM cell V_T , a 0.25 volt *weight* EEPROM cell V_T corresponds to a synapse weight of zero; *weight* EEPROM cell V_T 's below 0.25 volt correspond to positive weights. Better synapse multiply characteristics are obtained when a constant common mode is maintained. V_T 's ranging from -1.0 to 1.50 volts are attainable corresponding to a weight range of -2.5 to +2.5 volts.

Figure 15 shows synapse multiply characteristics versus floating-gate device weights at different fixed-input voltages (the inverse of Figure 13).

RMI, RMD, and RVTR—Reference Modify and Read Modes

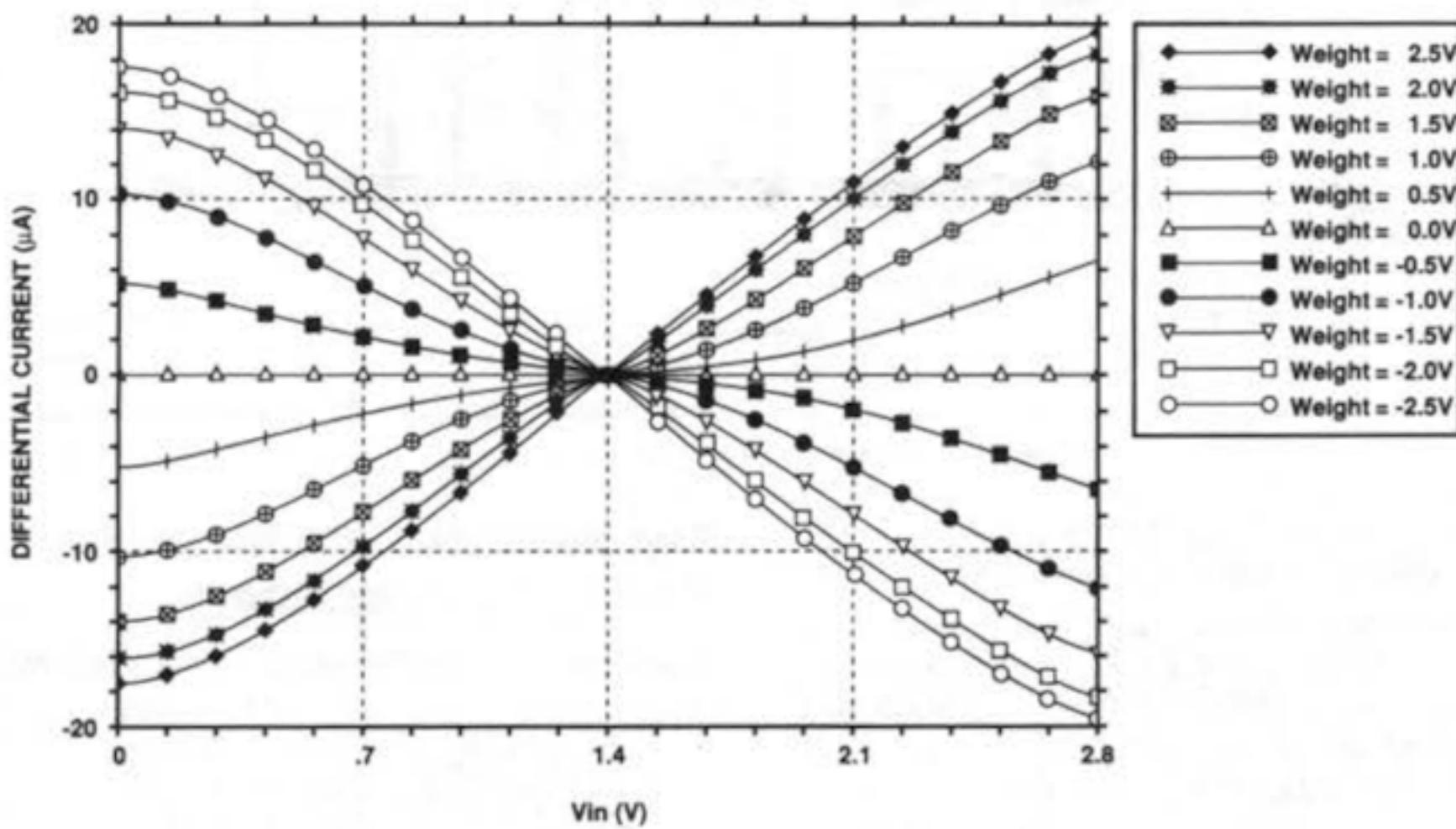
These modes are identical to the corresponding weight-related modes WMI, WMD, and WVTR except that they affect the *reference* EEPROM-cell threshold ($V_{T,REF}$) instead of the *weight* EEPROM-cell threshold ($V_{T,WT}$); see Figure 4.

Zero synapse-weight values occur when $V_{T,WT}$ is set equal to $V_{T,REF}$. Alternatively, unused synapses can be zeroed by programming both *weight* and *reference* cells to maximum (and equal) V_T 's. This is simpler and faster than setting $V_{T,DIFF}$ precisely to zero, and it reduces power consumption by turning the synapse cells off.



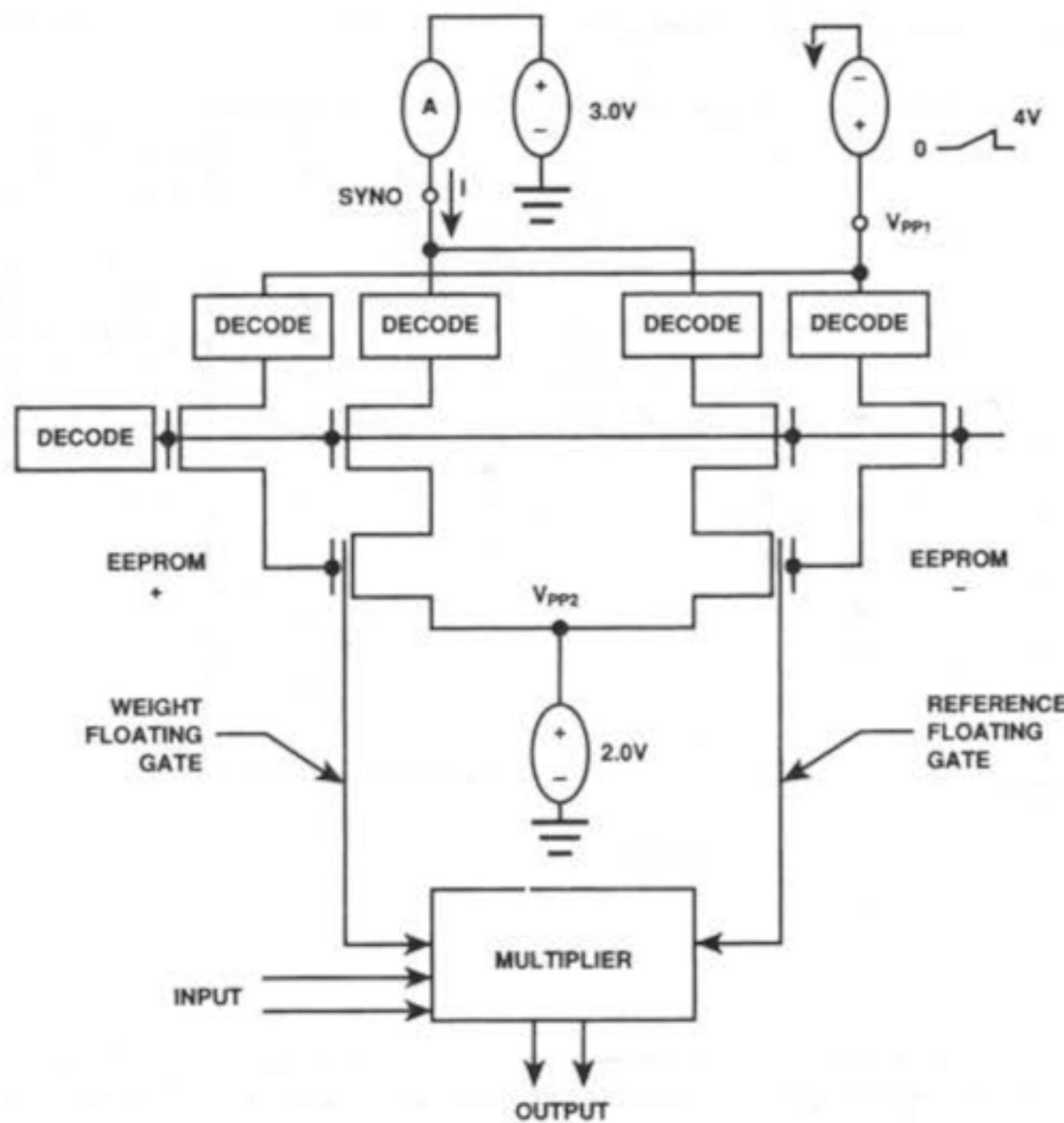
290408i12

Figure 12. EEPROM Floating-Gate Cell Characteristics



290408i13

Figure 13. Four-Quadrant Synapse Multiply Characteristics vs. Input Voltage



290408i14

Figure 14. V_T Measurement System

The *reference EEPROM* cell's floating gate can be set at a fixed value or to a value opposite from that of the *weight EEPROM* cell. For maximum dynamic range of weights, $V_{T,REF}$ should be set opposite to $V_{T,WT}$, and the average of $V_{T,WT}$ and $V_{T,REF}$ ($V_{T,COM}$) should be held constant at 0.25 volt, approximately the intrinsic (UV-erased) threshold level, as follows:

$$V_{T,WT} = V_{T,COM} - \frac{\text{Weight}}{2}$$

and

$$V_{T,REF} = V_{T,COM} + \frac{\text{Weight}}{2}$$

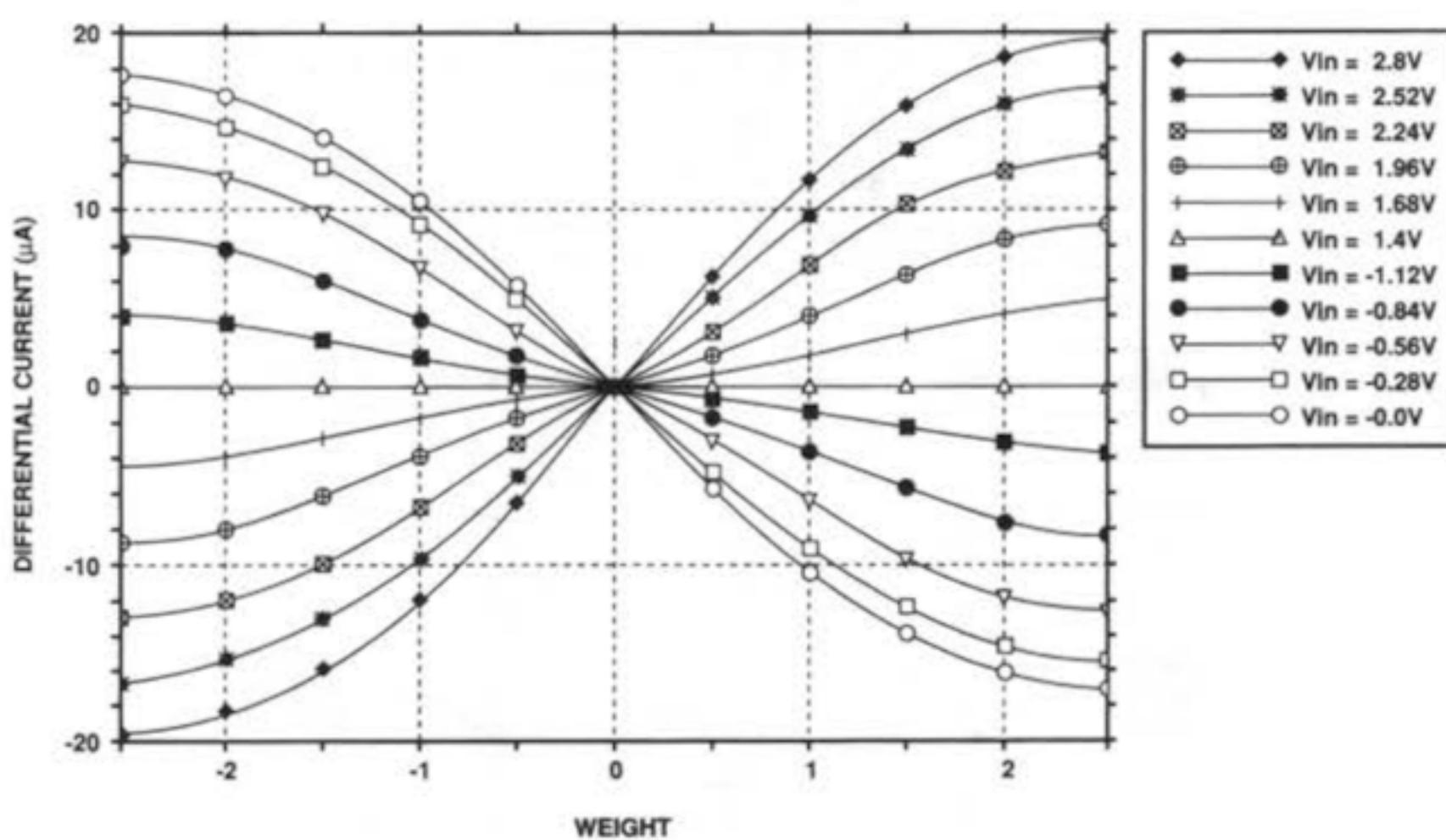
See Weight Setting in the TRAINING OPERATIONS section on page 24 for details.

SMR—Sum Read Mode

The differential voltage generated at a pair of summing nodes by the sum-of-products differential current can be measured directly in the SMR mode.

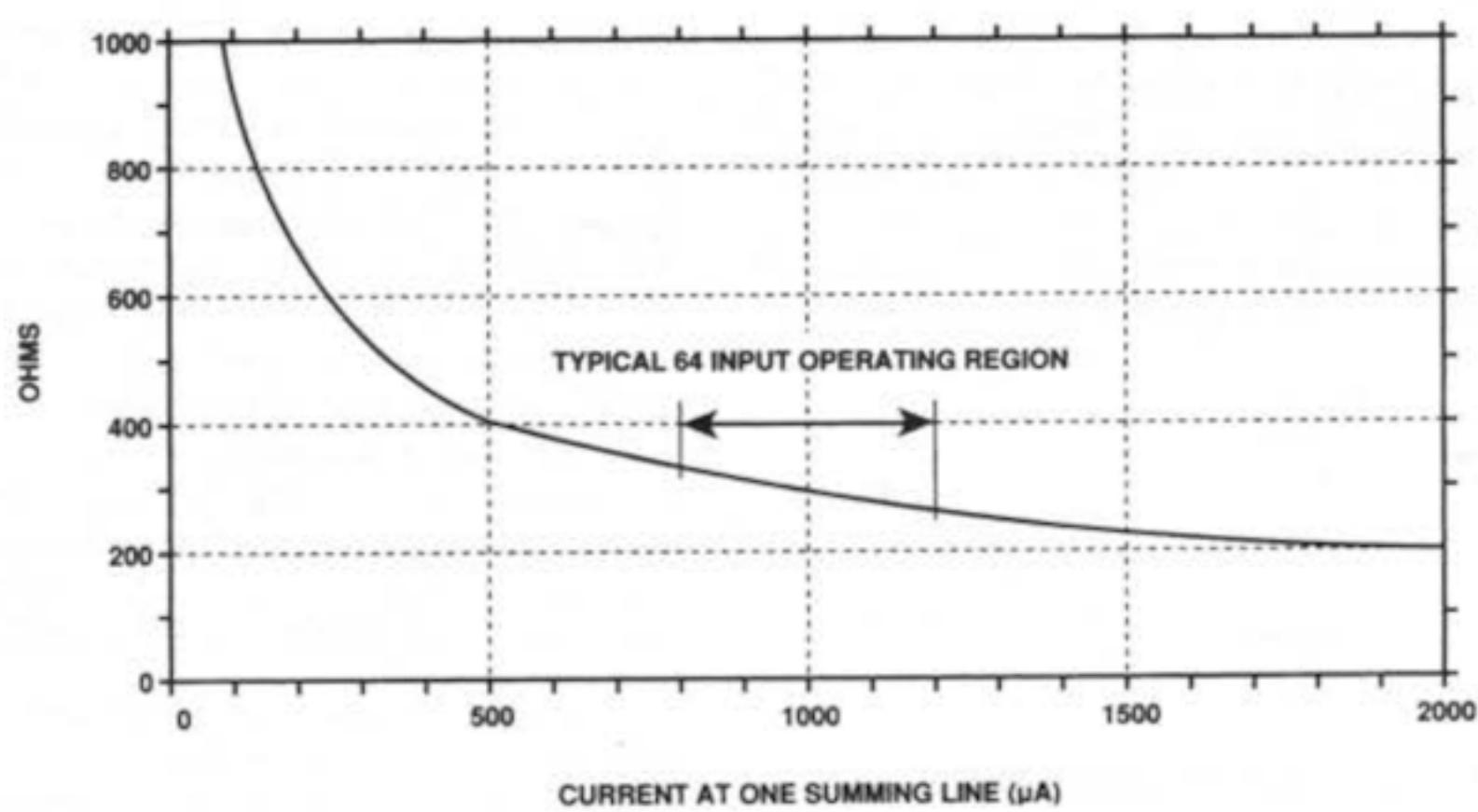
The differential current is converted to a differential voltage through the use of non-linear resistive loads. The effective resistance of these loads vary with the common-mode current through them (see Figure 16). The common-mode current varies with the number of active synapses in the neuron. Addresses A_5-A_0 select the desired pair of summing nodes (see Figure 17). The selected summing nodes are switched to the **SMO+** and **SMO-** pins. This differential voltage is a physical representation of the sum of products, $\Sigma\{\text{Weights} \cdot \text{inputs}\} + \text{bias}$, for the selected neuron. This voltage is very small and unbuffered; it may contain systematic offsets and is not intended for external use.

Figure 18(A) shows how a differential voltage would be measured between **SMO+** and **SMO-**. This mode may also be used to perturb the selected sum by drawing current either from pins **SMO+** or **SMO-**. Drawing current from **SMO+** will increase the sum while drawing current from **SMO-** will decrease the sum, which is useful for implementation of the Madaline III learning algorithm.



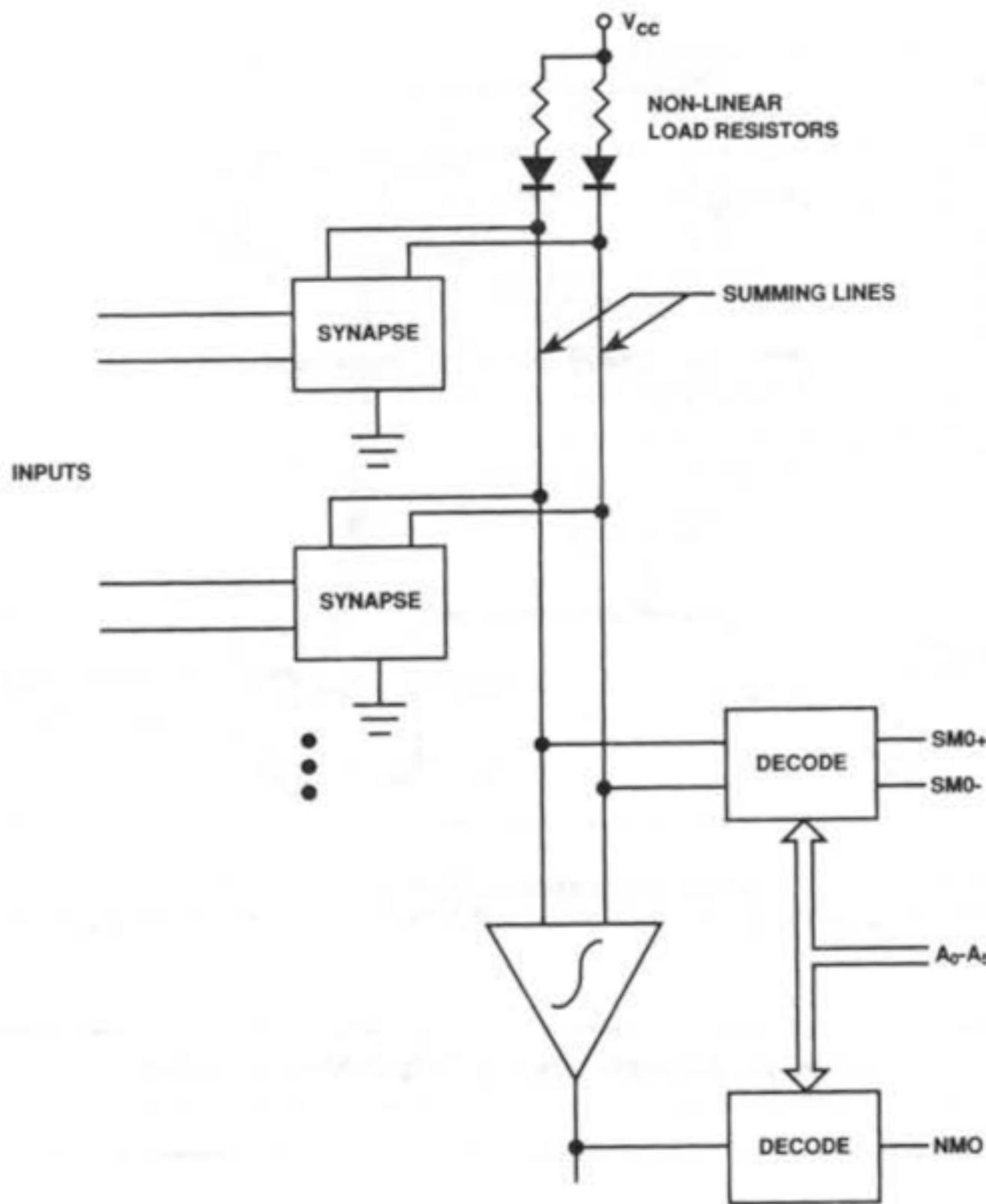
290408i15

Figure 15. Synapse Four-Quadrant Multiply Characteristics vs. Weight



290408i16

Figure 16. Non-Linear Resistive Summing Load



290408i17

Figure 17. Summing Nodes Selection and Output

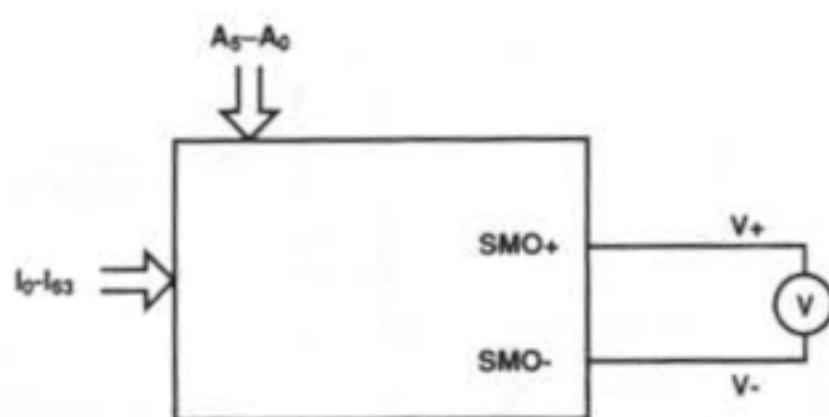
SYNR—Synapse Read Mode

A particular synapse is selected by addresses A_{12} – A_0 and **BIAS** in the SYNR mode, while other synapses connected to the same pair of summing nodes are disabled. This allows measurement of a synapse product contribution at that particular synapse to the total sum through the **SMO** \pm pins as a function of input level. Although a differential voltage can be measured at the pins **SMO** \pm , this is not an accurate representation because of the non-linear load characteristics. It is better to measure actual-cell current by forcing both pins to V_{CC} and measuring the difference in **SMO** $+$ and **SMO** $-$ current levels, as shown in Figure 18(B). Feedback Array synapses can be measured by using the input pads in the input multiplex mode. Rows used for threshold setting (a subset of the bias units) always have

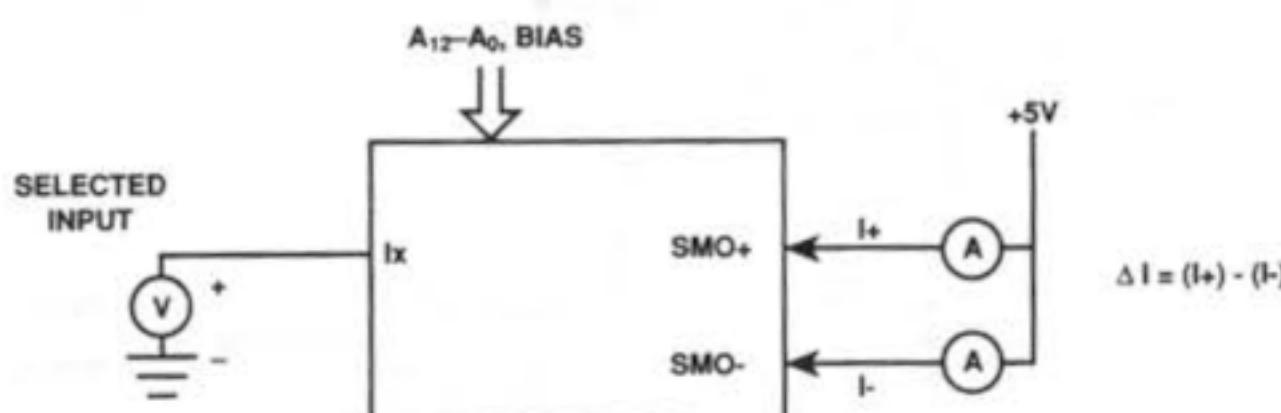
their inputs set to an internal, constant, positive reference. The effective input varies with V_{REFI} for the Input Array and with V_{REFO} for the Feedback Array.

NR—Neuron Read Mode

Addresses A_5 – A_0 select a neuron output to be connected to the **NMO** pin in NR mode. This allows the neuron outputs to be multiplexed out one at a time by the training system, eliminating the need for multiplexers between layers in a multi-layer network. Figures 19 and 20 show the neuron output's sigmoid transfer characteristics as a function of gain and reference control, versus summing-node current input. Figure 20 illustrates the neuron output transfer characteristics versus the summing current at various V_{REFO} 's.



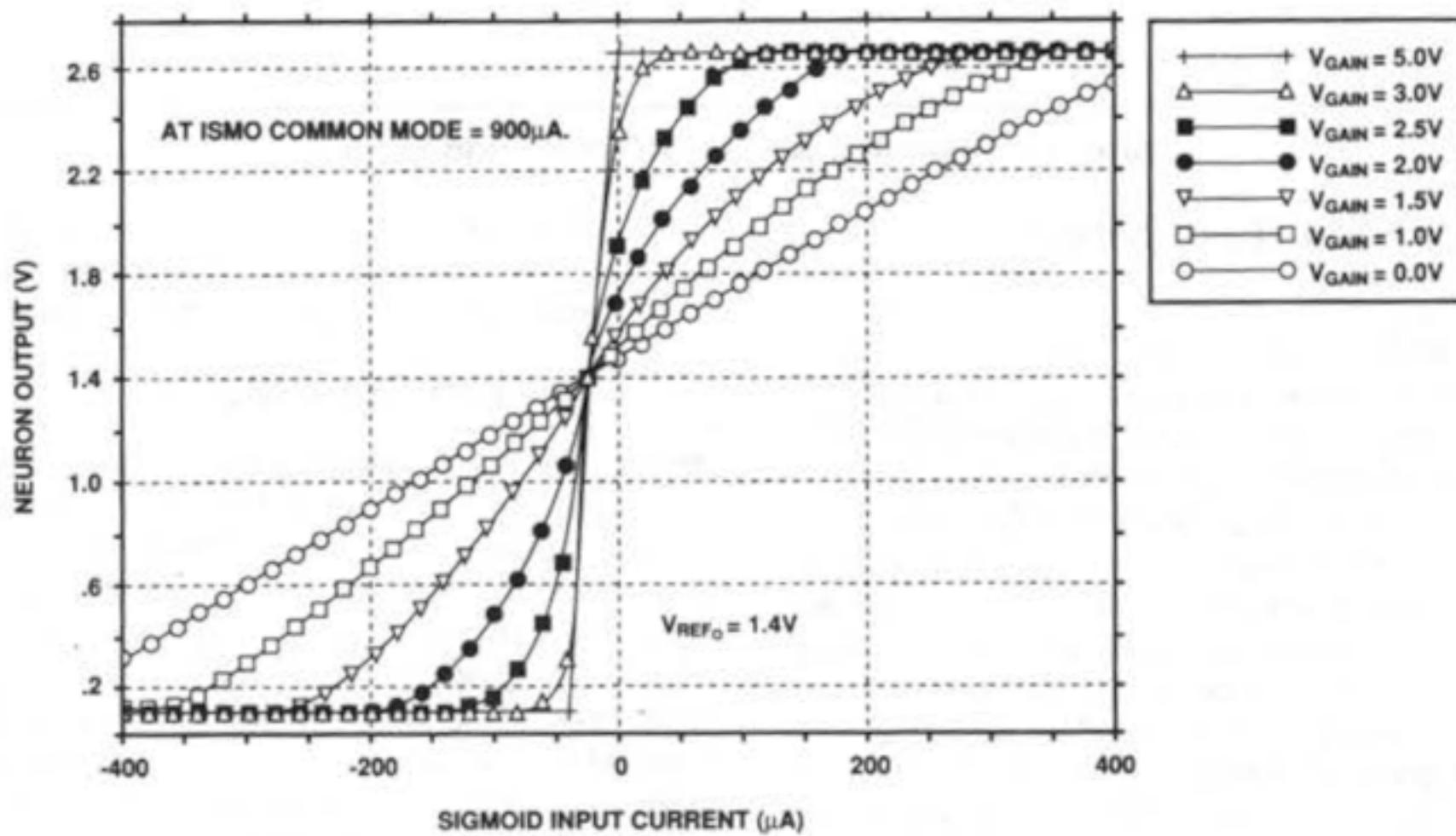
A. SMO +/- DIFFERENTIAL VOLTAGE METHOD



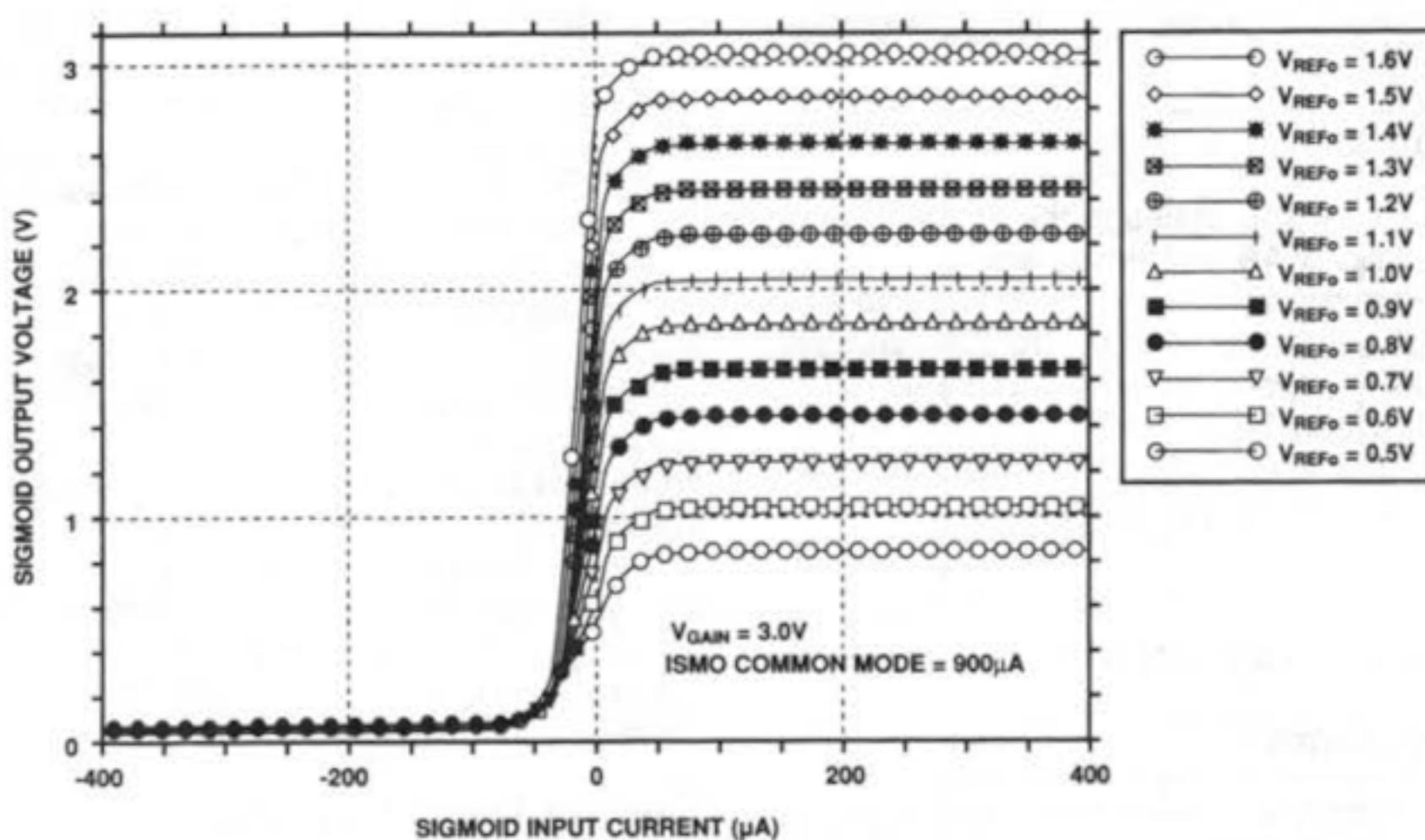
B. SMO +/- DIFFERENTIAL CURRENT METHOD

290408i18

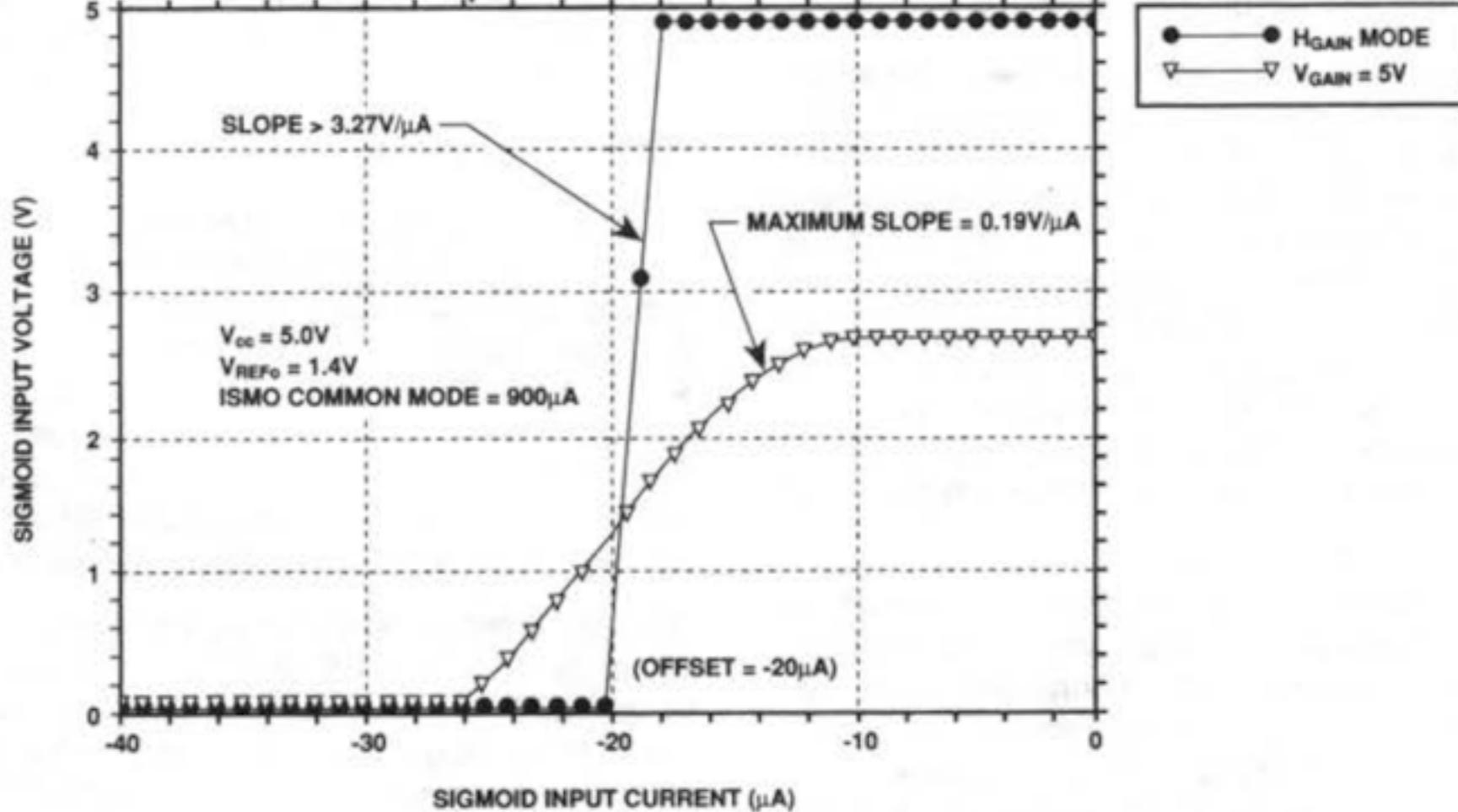
Figure 18. Summing Node Output Measurement

Figure 19. Neuron Output Transfer Characteristic vs. Summing Current and V_{GAIN}

290408i19



290408i20

Figure 20. Neuron Output Transfer Characteristic vs. Summing Current and V_{REF} 

290408i21

Figure 21. Sigmoid Characteristic in V_{GAIN} and H_{GAIN} Modes

The sigmoid has an input offset that varies slightly from chip to chip. The offset can be nulled during chip initialization by using one or more of the bias synapses. Figure 21 shows the H_{GAIN} sigmoid threshold mode.

Mode Summary

Table 1 shows the signal states in the various operating modes. Valid **BIAS** addresses are A_9 through A_0 (A_5-A_0 are neuron addresses), with A_{11} and $A_{10} = 0$ (A_{11} and A_{10} are only used for normal synapse selection). In all synapses (external input and bias), A_{12} selects between Input and Feedback Arrays (input = 0, feedback = 1). For AC waveforms, see Figures 30, 31 and 32 on pg. 35-36.

TRAINING OPERATIONS

Training Systems

Intel offers a complete development and training system for 80170NX chips—the **Intel Neural Network Training System (iNNTS)** and the **ETANN Multi-chip Board (EMB)**. The iNNTS and EMB are compatible with any personal computer based on an Intel 80286 (or later) processor with at least 1 MB of memory in addition to the 640K of conventional memory.

The iNNTS includes a Personal Computer PROM Programmer (PCPP), 80170 Adapter, Generic Programmer Interface (GUPI) base, Confidence Test Module, two 80170NX chips, interface software, and two neural network simulation programs.

The simulation programs are *iBrainMaker* from California Scientific Software, and *iDynaMind* from NeuroDynamX, Inc. Both simulation programs incorporate the 80170NX chip's sigmoid threshold characteristics during simulation. Both programs provide menu-driven functions to download synapse weights to the chip, test through the chip, and perform chip-in-loop optimization to compensate for analog variations. The programs support partitioning of the Feedback Array to allow networks with multiple hidden layers to be implemented on a single chip.

Users who prefer to write their own simulation or training algorithms are provided with a Training System Interface Library (TSIL) containing over two dozen subroutines which can be used to control the operations of the chip. These routines are described in detail in the Programmer's Reference section of the *iNNTS Neural Network Training System User's Guide*.

Multi-Chip Training

The 80170 Adapter provides only one socket with which to program and train one chip. Multi-chip designs require the EMB add-on board, which attaches to the iNNTS 80170 Adapter. An **EMB** includes an eight-socket prototyping board and two 80170NX chips. The iNNTS software is upwardly compatible with the EMB board, and multi-chip versions of the two simulation programs, *iBrainMaker* and *iDynaMind*, are provided with the EMB. The TSIL library contains routines to access each socket of the EMB separately. The board is shipped with jumper blocks attached on the bottom side to configure two analog buses for multiplexing 64 analog inputs and outputs to a target socket. When the chips are trained, the user has the option of removing the jumper blocks and wirewrapping directly from pin to pin. By connecting one chip's output pins directly to another chip's input pins, real-time processing configurations can be prototyped.

Chip-In-Loop Training

Due to fabrication variations, the characteristics of synapses and neurons vary within a single chip and within batches of chips. To compensate for these variations and to optimize chip performance, both the iNNTS and the EMB support chip-in-loop training.

This method replaces the synapse weights that would otherwise be simulated by a simulation program, with actual weights stored in a chip. During training, weights are downloaded to the chip and outputs of the chip are fed back to the training simulator, which evaluates the result and updates the chip's weights accordingly.

Weight Setting

Changes to synapse weights are accomplished by applying high-voltage pulses in the WMI and WMD modes. First, one of the two floating-gate devices in a single synapse is selected by the addresses $A_{12}-A_0$, **BIAS**, and **WT**, then high-voltage pulses are applied on pins V_{PP1} and V_{PP2} (see Figure 32). An adaptive algorithm is used to give a precise target weight with a minimum number of pulses, independent of synapse-to-synapse variations.

The EEPROM cell's floating-gate charge is modified by applying a high-voltage pulse for tens of microseconds. The weight-setting algorithm determines the pulse voltage, V_{PP2} . After the pulse is applied, the algorithm reads the synapse weight. If there are differences from the target weight, a subsequent pulse is applied.

The synapse weight is directly related to the difference in threshold voltages of the two floating-gate

Table 1. Mode Summary

	CE#	NE#	HOLD	CLK	RESET _T	RESET _F	H _{GAIN}	QUER	LRN	SSYN	A ₀ -A ₅	A ₆ -A ₁₁	A ₁₂
Parallel Distributed Processing (PDP):													
PDP/64 Inputs	0	0	X	X	0	1	0	0	0	0	X	X	0
PDP/128 Inputs	0	0	0/1 ⁽⁸⁾	1	0	0	0	0	0	0	X	X	1
Neuron Disable	0	1	X	X	0	1	0	0	0	0	X	X	0
Power Down	1	X	X	X	X	X	X	X	X	X	X	X	X
Weight Modify:													
Wt Mod Increase (WMI)	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	1	V	V	V
Wt Mod Decrease (WMD)	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	1	V	V	V
Wt Vt Read (WVTR)	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	1	1	1	V	V	V
Ref Mod Increase (RMI)	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	1	V	V	V
Ref Mod Decrease (RMD)	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	1	V	V	V
Ref Vt Read (RVTR)	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	1	1	1	V	V	V
Query:													
Sum Read (SMR)	0	X	V	V	V	V	X	1	0	0	V	X	V
Synapse Read (SYNR)	0	X	V	V	V	V	X	1	0	1	V	V	V
Neuron Read (NR)	0	0	V	V	V	V	V	1	0	0	V	X	V
Reset/Perturb:													
Erase Wt Column	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	0	V	X	X
Erase Ref Column	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	0	V	X	X
Prog Wt Column	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	0	V	X	X
Prog Ref Column	0	1	X	X	X ⁽¹⁾	X ⁽¹⁾	X	0	1	0	V	X	X
Sum Read/Perturb (SMR/P)	0	0	V	V	V	V	V	1	0	0	V	X	V
Signature Modes:													
Standard	0	0	X	X	X	X	0	0	0	X	X	X ⁽⁴⁾	X
NW vs NX	0	0	X	X	1	1	1	0	0	X	X	X ⁽⁴⁾	X
NMO Signature	0	0	X	X	X	X	1	1	0	X	V	X ⁽⁴⁾	X
Continuous Feedback:													
With Input Array Disabled	0	0	X	1	1	0	0	0	0	X	X	X ⁽⁶⁾	X

- NOTES:**
- (1) Preferred RESET_T = RESET_F = 1, if RESET_T / RESET_F = 0, the selected synapse will conduct current. This may affect the programming conditions. The state of these pins must be consistent between programming and verifying.
 - (2) Force SMO± to V_{CC} and measure current.
 - (3) Sink differential current from the SMO± pins.
 - (4) A₉ = V_{HH}.
 - (5) V_{REF0} is less than 0.7V.
 - (6) A₁₁ = V_{HH}.
 - (7) Outputs may be active, depending on state of NE#.
 - (8) See Figure 7.

- SYMBOLS:**
- 0 = Logic "0"
 - 0/1 = Transition from 0 to 1 (see Figure 7).
 - 1 = Logic "1"
 - X = Don't Care Inputs, Indeterminate State Outputs
 - Q = Valid Data Output
 - V = Valid Input
 - Z = High Impedance
 - V_{CC} = Logic Supply Voltage
 - V_{HH} = 12 ± 0.5V
 - V_{P1} & V_{P2} = High-Voltage Pulse Inputs
 - V_G & V_S = Gate and Source Voltages—used in SYNR mode
 - I_D = EEPROM Cell Drain Current

Table 1. Mode Summary (Contd.)

	BIAS	WT	INC	V _{GAIN}	V _{REFI}	V _{REFO}	I ₀ -I ₆₃	N ₀ -N ₆₃	SMO±	NMO	SYNO	V _{PP1}	V _{PP2}
Parallel Distributed Processing (PDP):													
PDP/64 Inputs	X	X	X	V	V	V	V	Q	Z	Z	Z	V _{CC}	V _{CC}
PDP/128 Inputs	X	X	X	V	V	V	V	Q	Z	Z	Z	V _{CC}	V _{CC}
Neuron Disable	X	X	X	V	V	V	V	Z	Z	Z	Z	V _{CC}	V _{CC}
Power Down	X	X	X	X	X	X	X	Z	Z	Z	Z	V _{CC}	V _{CC}
Weight Modify:													
Wt Mod Increase (WMI)	V	1	1	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Wt Mod Decrease (WMD)	V	1	0	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Wt Vt Read (WVTR)	V	1	X	X	X	X	X	Z	Z	Z	I _D	V _G	V _S
Ref Mod Increase (RMI)	V	0	1	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Ref Mode Decrease (RMD)	V	0	0	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Ref Vt Read (RVTR)	V	0	X	X	X	X	X	Z	Z	Z	I _D	V _G	V _S
Query:													
Sum Read (SMR)	X	X	X	X	V	V	V	X ⁽⁷⁾	Q	X	Z	V _{CC}	V _{CC}
Synapse Read (SYNR)	V	X	X	X	V	V	V	X ⁽⁷⁾	V _{CC} ⁽²⁾	X	Z	V _{CC}	V _{CC}
Neuron Read (NR)	X	X	X	V	V	V	V	Q	Q	Q	Z	V _{CC}	V _{CC}
Reset/Perturb:													
Erase Wt Column	X	1	1	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Erase Ref Column	X	0	1	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Prog Wt Column	X	1	0	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Prog Ref Column	X	0	0	X	X	X	X	Z	Z	Z	Z	V _{P1}	V _{P2}
Sum Read/Perturb (SMR/P)	X	X	X	V	V	V	V	Q	Q ⁽³⁾	Q	Z	V _{CC}	V _{CC}
Signature Modes:													
Standard	X	X	X	X	X	V	X	Q	Z	Z	Z	V _{CC}	V _{CC}
NW vs NX	X	X	X	X	X	V ⁽⁵⁾	X	Q	Z	Z	Z	V _{CC}	V _{CC}
NMO Signature	X	X	X	X	X	V	X	Q	Q	Q	Z	V _{CC}	V _{CC}
Continuous Feedback:													
With Input Array Disabled	X	X	X	V	X	V	X	Q	Z	Z	Z	V _{CC}	V _{CC}

- NOTES: (1) Preferred RESET_I = RESET_F = 1, if RESET_I/RESET_F = 0, the selected synapse will conduct current. This may affect the programming conditions. The state of these pins must be consistent between programming and verifying.
(2) Force SMO± to V_{CC} and measure current.
(3) Sink differential current from the SMO± pins.
(4) A₉ = V_{HH}.
(5) V_{REFO} is less than 0.7V.
(6) A₁₁ = V_{HH}.
(7) Outputs may be active, depending on state of NE#.
(8) See Figure 7.

- SYMBOLS: 0 = Logic "0"
0/1 = Transition from 0 to 1 (see Figure 7).
1 = Logic "1"
X = Don't Care Inputs, Indeterminate State Outputs
Q = Valid Data Output
V = Valid Input
Z = High Impedance
V_{CC} = Logic Supply Voltage
V_{HH} = 12 ± 0.5V
V_{P1} & V_{P2} = High-Voltage Pulse Inputs
V_G & V_S = Gate and Source Voltages—used in SYNR mode
I_D = EEPROM Cell Drain Current

devices in that synapse (see V_T Read Mode description). Floating-gate threshold voltages are measured repeatedly during the weight-setting algorithm to assess how well the weight-change process is proceeding and to make better estimates of the voltage to be used in subsequent weight settings.

The difference between target V_T ($V_{T,\text{target}}$) and real V_T ($V_{T,\text{eo}}$) is ΔV_T . The floating gate is programmed when ΔV_T is positive and erased when ΔV_T is negative. In other words, programming raises the EEPROM cell threshold, while erasing reduces its threshold:

$$\Delta V_T = V_{T,\text{target}} - V_{T,\text{eo}}$$

If $\Delta V_T > 0$, then program
If $\Delta V_T < 0$, then erase

Figures 22 and 23 illustrate the calculated relationship between a desired synapse-weight change, ΔV_T , and the high voltage which should be applied at V_{PP2} to achieve it. A family of calculated curves is shown for various initial thresholds, $V_{T,\text{eo}}$, for a fixed pulse width of 100 μs to illustrate the relationship. The exact relationship may vary slightly from these values. Table 2 shows the equations from which the curves were generated.

Figure 24 is a flow chart of the weight-setting algorithm. The algorithm begins by initializing the address of the floating-gate device and two coefficients: B_{PGM} and B_{ERS} (B_{PGM} for programming, B_{ERS} for erase). The ΔV_T for the next iteration is determined from the difference between the $V_{T,\text{target}}$ and the most recently measured V_T ($= V_{T,\text{eo}} = V_{T,\text{real}}$). The V_{PP2} for the next pulse is iteratively

solved for by using the functions given in Table 2, $f(V_{PP,\text{old}})$ or $g(V_{PP,\text{old}})$, for programming and erasing, respectively. The functions f and g signify the derivatives of these functions, respectively.

After applying a pulse, the threshold is measured again and subtracted from the original threshold to arrive at the real ΔV_T . If the measured ΔV_T deviates by more than 1% from the desired ΔV_T , a new B_{PGM} or B_{ERS} will be calculated based on the previous V_{PP} and B_{PGM} or B_{ERS} . This procedure is repeated until the measured V_T falls within 1% of the $V_{T,\text{target}}$. The final B_{PGM} and B_{ERS} will be stored for use by the next synapse in the same chip. Because synapses on a chip have very similar program and erase characteristics, the B coefficients do not change significantly after the first few synapse weights are modified.

The difference between the desired ΔV_T and the actual ΔV_T is reduced by each subsequent pulse as the adaptive algorithm learns from experience.

Madaline Learning

Using **SMO+** and **SMO-** in the SMR/P (Sum Read/Perturb) mode allows adding or subtracting incremental differential current from a selected neuron's summing nodes. Sum perturbation supports training by the Madaline learning algorithm. This involves the application of many perturbations followed by grading the improvements produced by each perturbation. Synapse weights are then changed to produce the same effects as the perturbations which caused the most significant improvements in the output responses.

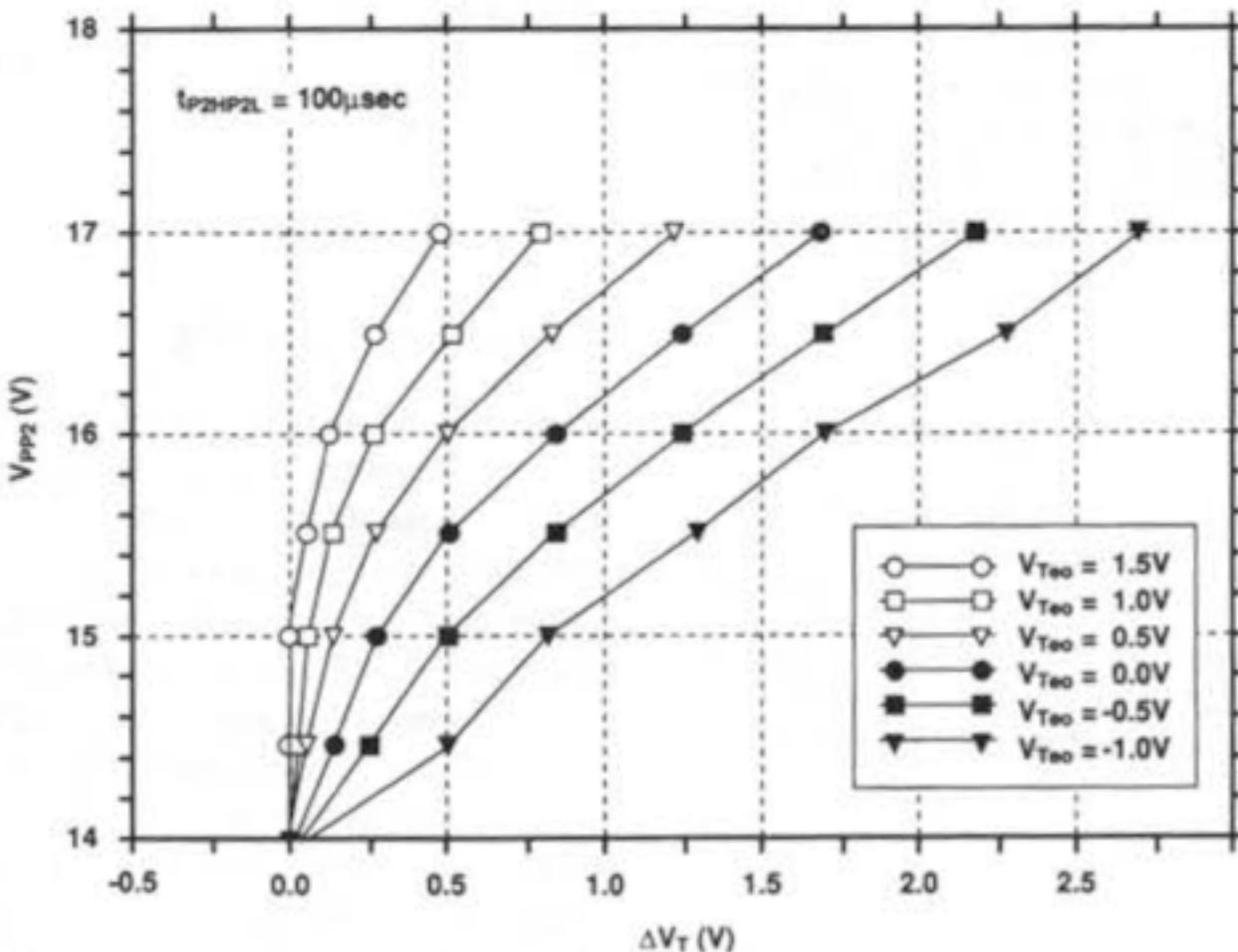
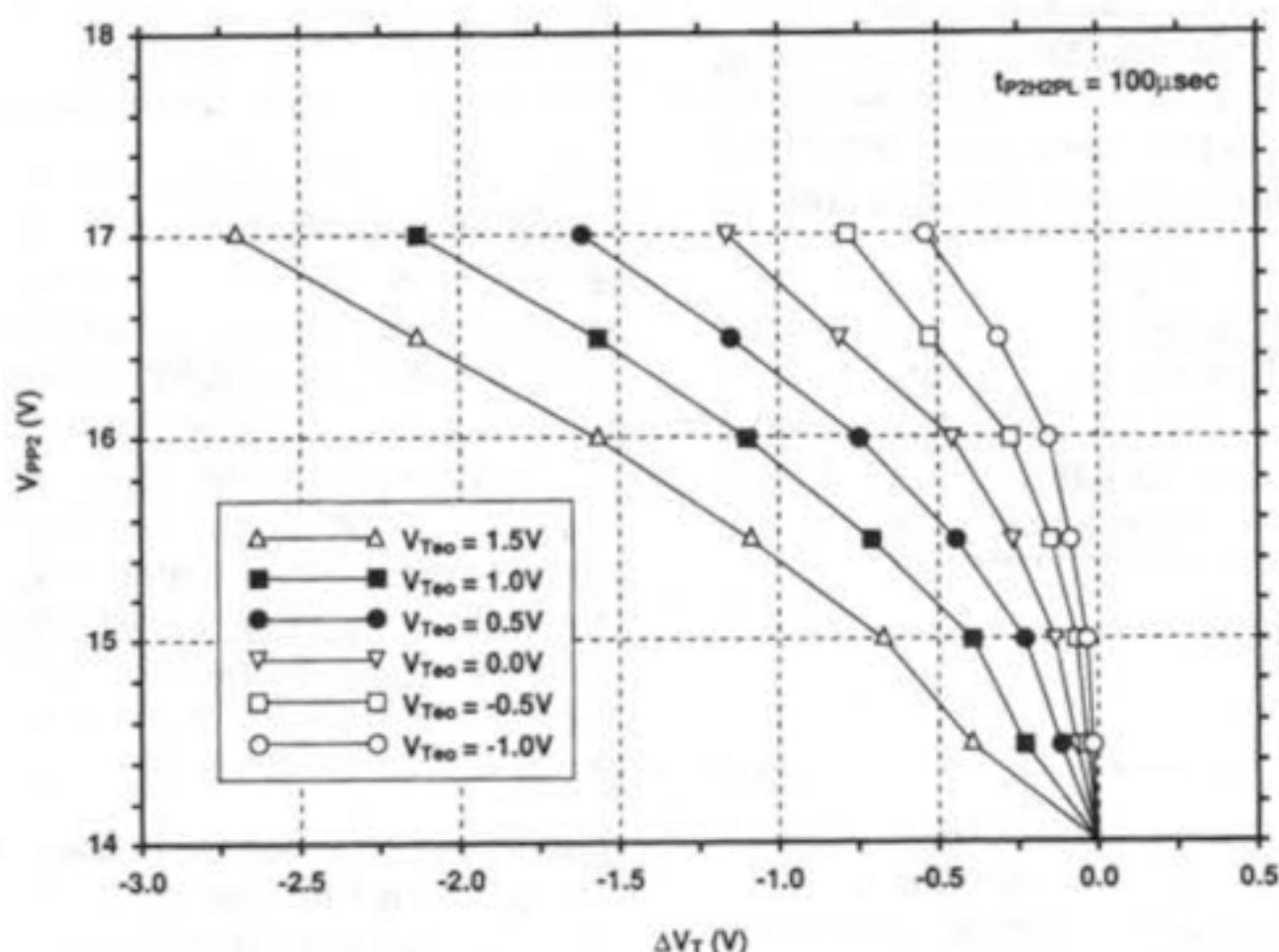


Figure 22. V_{PP2} Voltage Curves for Positive ΔV_T – Program



290408i23

Figure 23. V_{PP2} Voltage Curves for Negative ΔV_T — Erase

Function Simulation

Many standard neural network simulators can model the 80170NX chip's dot product neurons reasonably well. The constraints of the chip's analog circuits can be reflected by specific limits on input and weight-setting ranges and gain values. A first-order emulation allows evaluation of whether a chip or group of chips can be trained to perform the desired pattern-mapping application. Currently there are no effective analytical techniques or algorithms for determining appropriate neural network architectures by simply observing the training information set.

The following transfer characteristic simulates the chip's neuron output at typical operating conditions ($V_{GAIN} = 5V$, $V_{REFo} = 1.4V$, $V_{REFI} = 1.4V$, PDP/64-input operation):

$$v_j = \left\{ \frac{2}{(1 + e^{-8\sum \mu_i * W_{ij}})} \right\} - 1$$

where

- μ_i Inputs, constrained so that $-1 \leq ((V_j - V_{REFI}) / V_{REFo}) \leq +1$.
- W_{ij} Weights, constrained so that $-1 \leq ((weight V_T - reference V_T) / 2.5V) \leq +1$.
- v_j Outputs, constrained so that $-1 \leq ((Output - V_{REFo}) / V_{REFo}) \leq +1$.

These normalized input and synapse-weight values correspond to actual inputs of 0 to 2.8V and weights of -2.5 to 2.5V. The normalized output will also correspond to an actual range of 0 to 2.8V.

A more accurate model of the chip's output under the above conditions and with $V_{GAIN} = 5V$ is:

$$v_j = \left\{ \frac{1.8}{(1 + e^{-8(\sum \mu_i(1.2 - 0.2\mu_i^2)W_{ij}(1.5 - 0.5W_{ij}^2) - \sum k B_{kj})})} \right\} - 0.9$$

This model includes the roll-off in the synapse characteristic at large weight and input magnitudes, as well as the output range limitations. The estimated equation for $V_{GAIN} = 3.3V$ (with $V_{REFI} = V_{REFo} = 1.5V$) is:

$$v_j = \left\{ \frac{1.83}{1 + e^{-1.74 \sum (\text{Synaptic Contributions})}} \right\} - 0.94$$

Bake-Train-Bake

The chip's weight retention over extended periods of time can be significantly improved by a process called Bake-Train-Bake. Charge movement in the nitride layer between the EEPROM cell's control gate and floating gate creates dipoles. The dipoles compensate the electric field created by charges on the floating gates, thus affecting the effective floating-gate potential. This electrical reaction offsets the effective $V_{T DIFF}$'s from their original targets; thus, affecting the chip's processing accuracy. High-temperature bake operations accelerate this phenomenon. When combined with re-training, the

Table 2. Training/Weight Setting Algorithm Equations

Program	$\Delta V_T = V_{TO} + V_{PP} - V_{Teo} - 1.375 \times 10^{-6} B_{PGM} / \{ \ln[e^{(B_{pgm}/E_0)} + 3.242 \times 10^4 * B_{PGM} * T_p] \}$ and $E_o = 7.273 \times 10^5 * V_{PP} + 7.273 \times 10^5 * (V_{TO} - V_{Teo})$
Erase	$\Delta V_T = V_{TO} - 1.125 * V_{PP} - V_{Teo} + 1.375 \times 10^{-6} B_{ERS} / \{ \ln[e^{(B_{ers}/E_0)} + 3.242 \times 10^4 * B_{ERS} * T_e] \}$ and $E_{yo} = 8.182 \times 10^5 * V_{PP} + 7.273 \times 10^5 * (V_{Teo} - V_{TO})$
V _{PP} (Program) Calculations	$f(V_{PP}) = V_{TO} + V_{PP} - \Delta V_T - V_{Teo} - 1.375 \times 10^{-6} B_{PGM} / \{ \ln[e^{(B_{pgm}/E_0)} + 3.242 \times 10^4 * B_{PGM} * T_p] \}$ $f'(V_{PP}) = 1 - \{ B_{PGM}^2 * e^{(B_{pgm}/E_0)} / \{ [\ln[e^{(B_{pgm}/E_0)} + 3.242 \times 10^4 * B_{PGM} * T_p]]^2 * E_o^2 * [e^{(B_{pgm}/E_0)} + 3.242 \times 10^4 * B_{PGM} * T_p] \} \}$ where $V_{PP_new} = V_{PP_old} - f(V_{PP_old}) / f'(V_{PP_old})$ until $V_{PP_new} = > \sim V_{PP_old}$
V _{PP} (Erase) Calculations	$g(V_{PP}) = V_{TO} - \Delta V_T - V_{Teo} - 1.125 * V_{PP}$ $+ 1.375 \times 10^{-6} B_{ERS} / \{ \ln[e^{(B_{ers}/E_0)} + 3.242 \times 10^4 * B_{ERS} * T_e] \}$ $g'(V_{PP}) = -1.125 + \{ 1.125 * B_{ERS}^2 * e^{(B_{ers}/E_0)} / [E_{yo}^2 * [\ln[e^{(B_{ers}/E_0)} + 3.242 \times 10^4 * B_{ERS} * T_e]]^2 * [e^{(B_{ers}/E_0)} + 3.242 \times 10^4 * B_{ERS} * T_e]] \}$ where $V_{PP_new} = V_{PP_old} - g(V_{PP_old}) / g'(V_{PP_old})$ until $V_{PP_new} = > \sim V_{PP_old}$
B _{PGM} Calculations	$h(B_{PGM}) = 7.273 \times 10^5 * (V_{TO} + V_{PP} - \Delta V_T - V_{Teo}) * \{ \ln[e^{(B_{pgm}/E_0)} + 3.242 \times 10^4 * B_{PGM} * T_p] \} - B_{PGM}$ $h'(B_{PGM}) = 7.273 \times 10^5 * (V_{TO} + V_{PP} - \Delta V_T - V_{Teo}) * \{ [e^{(B_{pgm}/E_0)}] / E_0 + 3.242 \times 10^4 * T_p \} / \{ e^{(B_{pgm}/E_0)} + 3.242 \times 10^4 * B_{PGM} * T_p \} - 1$ where $B_{PGM_new} = B_{PGM_old} - h(B_{PGM_old}) / h'(B_{PGM_old})$ until $B_{PGM_new} = > \sim B_{PGM_old}$
B _{ERS} Calculations	$k(B_{ERS}) = (V_{Teo} + \Delta V_T - V_{TO} + 1.125 * V_{PP}) * \{ \ln[e^{(B_{ers}/E_0)} + 3.242 \times 10^4 * B_{ERS} * T_e] \} - 1.375 \times 10^{-6} B_{ERS}$ $k'(B_{ERS}) = (V_{Teo} + \Delta V_T - V_{TO} + 1.125 * V_{PP}) * \{ [e^{(B_{ers}/E_0)}] / [E_{yo} * [e^{(B_{ers}/E_0)} + 3.242 \times 10^4 * T_e]] \} - 1.375 \times 10^{-6}$ where $B_{ERS_new} = B_{ERS_old} - k(B_{ERS_old}) / k'(B_{ERS_old})$ until $B_{ERS_new} = > \sim B_{ERS_old}$

NOTES:

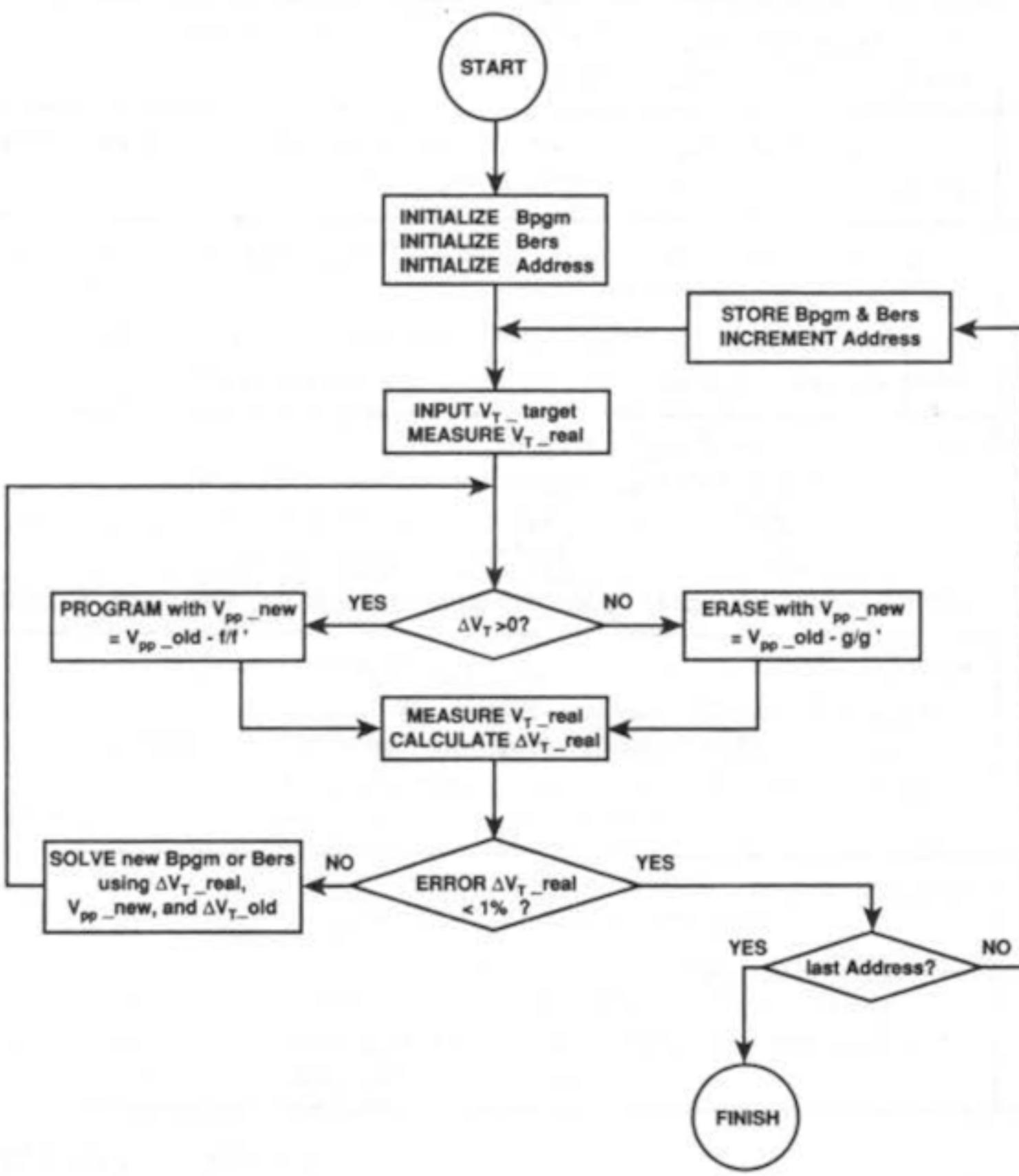
 $V_{TO} \sim 0.5V$ $B_{PGM} = B_{pgm}$ = program mode constant $\Delta V_T = \text{del. } V_T$ = EEPROM cell threshold change ("threshold delta") $V_{TEO} = V_{Teo}$ = present EEPROM threshold $B_{ERS} = B_{ers}$ = erase mode constant

net effect is to reduce this V_T shift, since the electrical field differences between re-trainings is much smaller than between the originally untrained and the first-training conditions. The change in synapse weight versus bake time is shown in Figure 25. The improvement in a single 24-hour 250°C bake-train operation represents 3 to 4 bits of resolution over the chip's minimum data-retention lifetime of 10 years. Two bake-train cycles compensate for lifetime V_T shifts well within the chip's intrinsic processing accuracy. The Acceleration Factor note at the top-left of the chart indicates that baking at 260°C causes the percent change shown on the vertical axis to occur 1000 times faster than it would occur if the chip were stored at 125°C, and 100000 times faster than it would occur if the chip were stored at 75°C.

Temperature and V_{CC} Performance

Figures 26 and 27 show a typical neuron's response to ambient temperature and V_{CC} , respectively. For typical input patterns and synapse weights, four inputs at weight = 0.5V are swept while monitoring the neuron output at $V_{GAIN} = 5V$.

In Figure 26, two temperature effects are visible in the high-gain region of the response: the offset of the neuron shifts and the slope (effective gain) of the neuron decreases. The offset change corresponds to a change in the sum of products of about 2 μA or about 1/5 of one synapse at full weight and full input (0.25% out of 64 synapses) over the commercial temperature range. The offset change is highly dependent on the inputs and weights employed. The output slope decreases with temperature by 0.3% per degree Celsius over the nominal at 25°C (total 20% change over the commercial range).



290408i24

Figure 24. Weight-Setting Algorithm

In Figure 27, the slope in the high-gain region of the sigmoid characteristic does not change significantly with V_{CC} . The V_{CC} effect is characterized as almost exclusively as a change in offset. The offset shifts by about 0.26V which corresponds to less than 2.5 μ A in the total sum over the 4.75V to 5.25V V_{CC} range (± 5).

For both the temperature and V_{CC} cases, the voltage change at the saturated regions of the extremes (of low and high inputs) is less than 10 mV.

PERFORMANCE

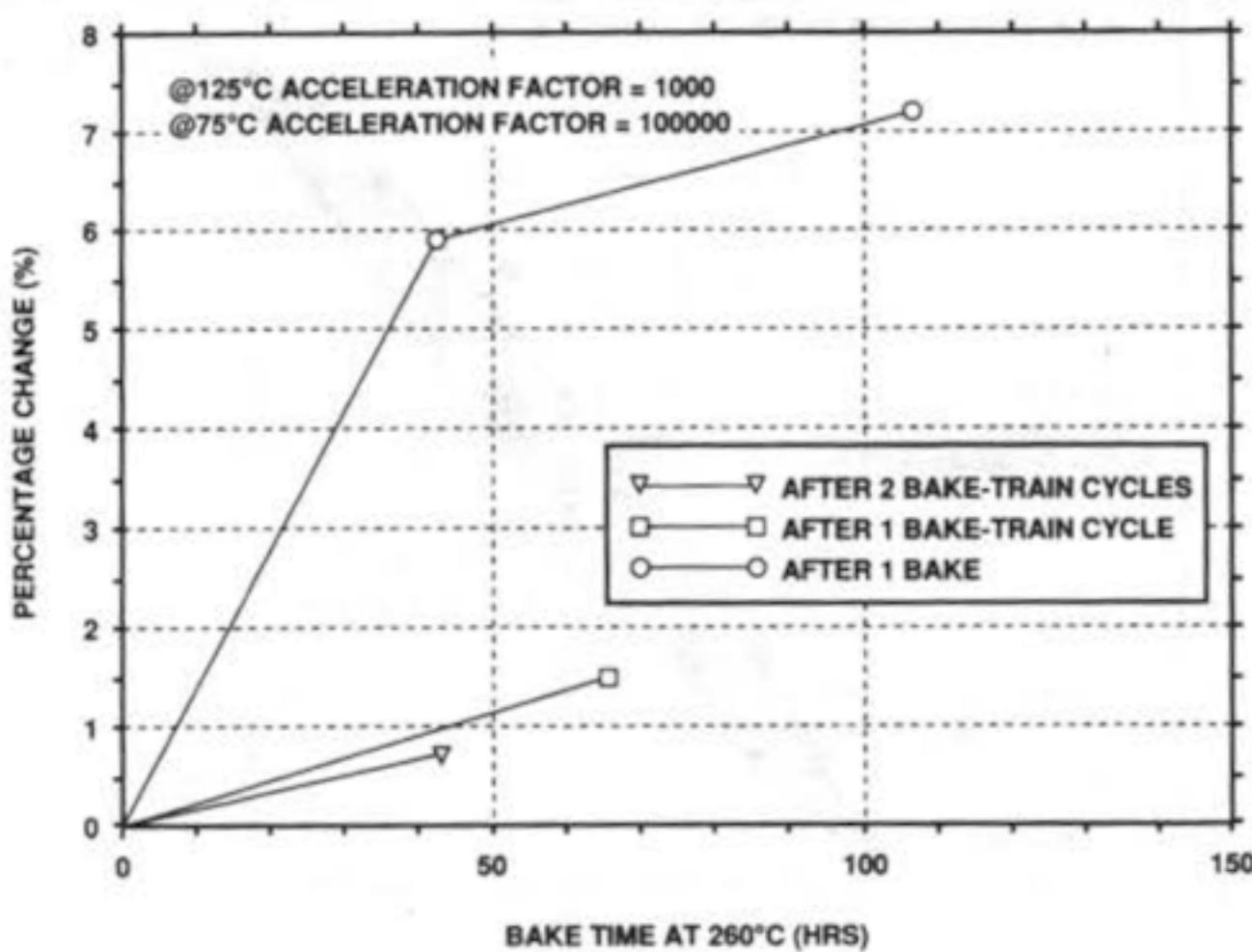
Weight-Cycling Performance

EEPROM cycling failures are a common concern. The strong electric field required by thin oxide

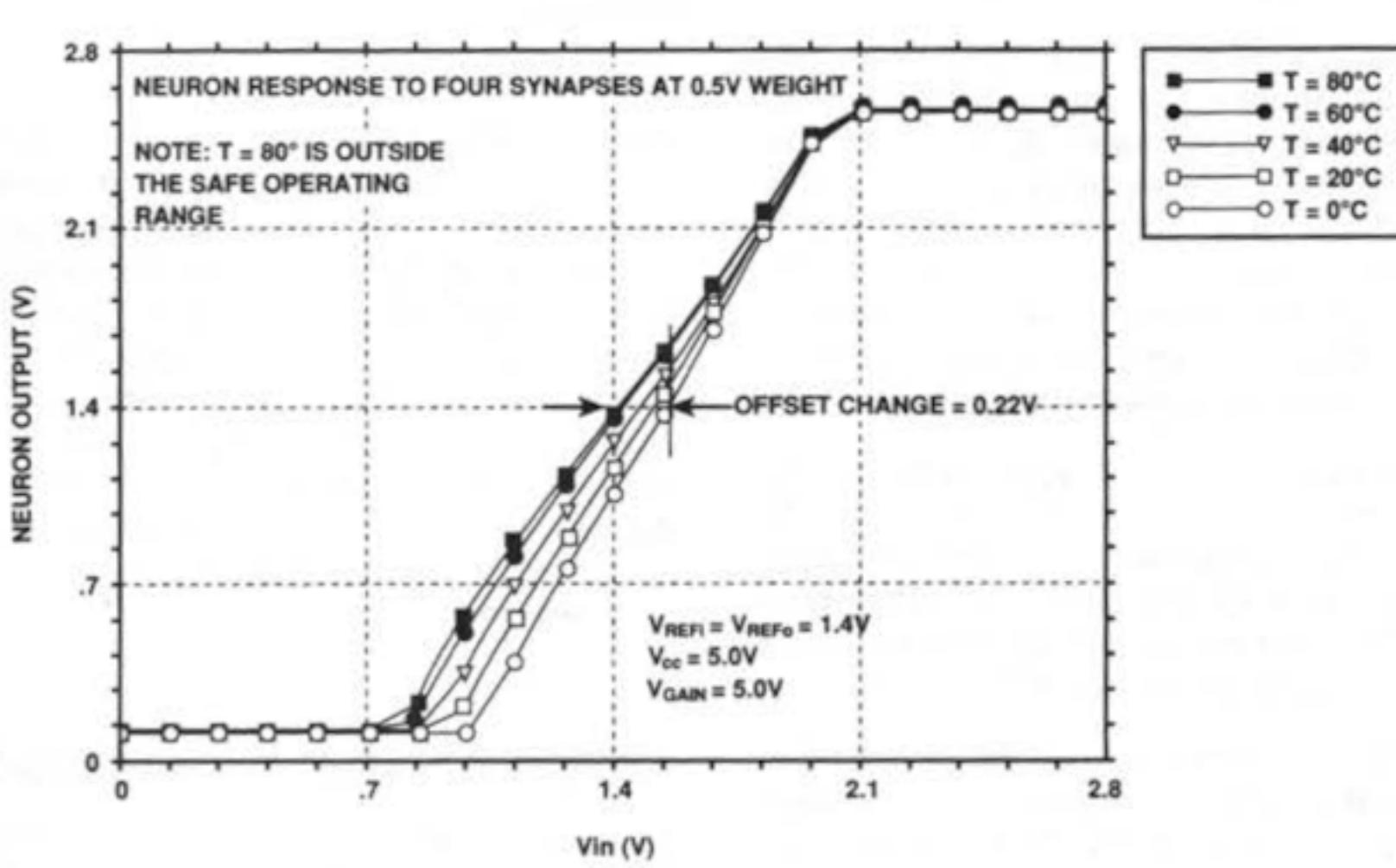
EEPROMs for tunneling can tear the oxide in defect regions. Several process and design improvements have been introduced to combat the problem.

Intel's ETOX-II flash memory technology is designed for extended cycling. An advanced tunnel oxide increases charge carrying capability ten-fold, minimizing the oxide area per synapse subjected to the tunneling electric field, and consequently reducing the probability of encountering an oxide defect.

During normal operation with the iNNTS training system, programming and erasing is performed by the weight-setting algorithm described above (see TRAINING OPERATIONS on page 22). Programming and erasing synapse weights is accomplished by cycling the transistors of the *reference* EEPROM cells and *weight* EEPROM cells. The 80170NX chip is specified for a minimum of 10,000 such cycles.

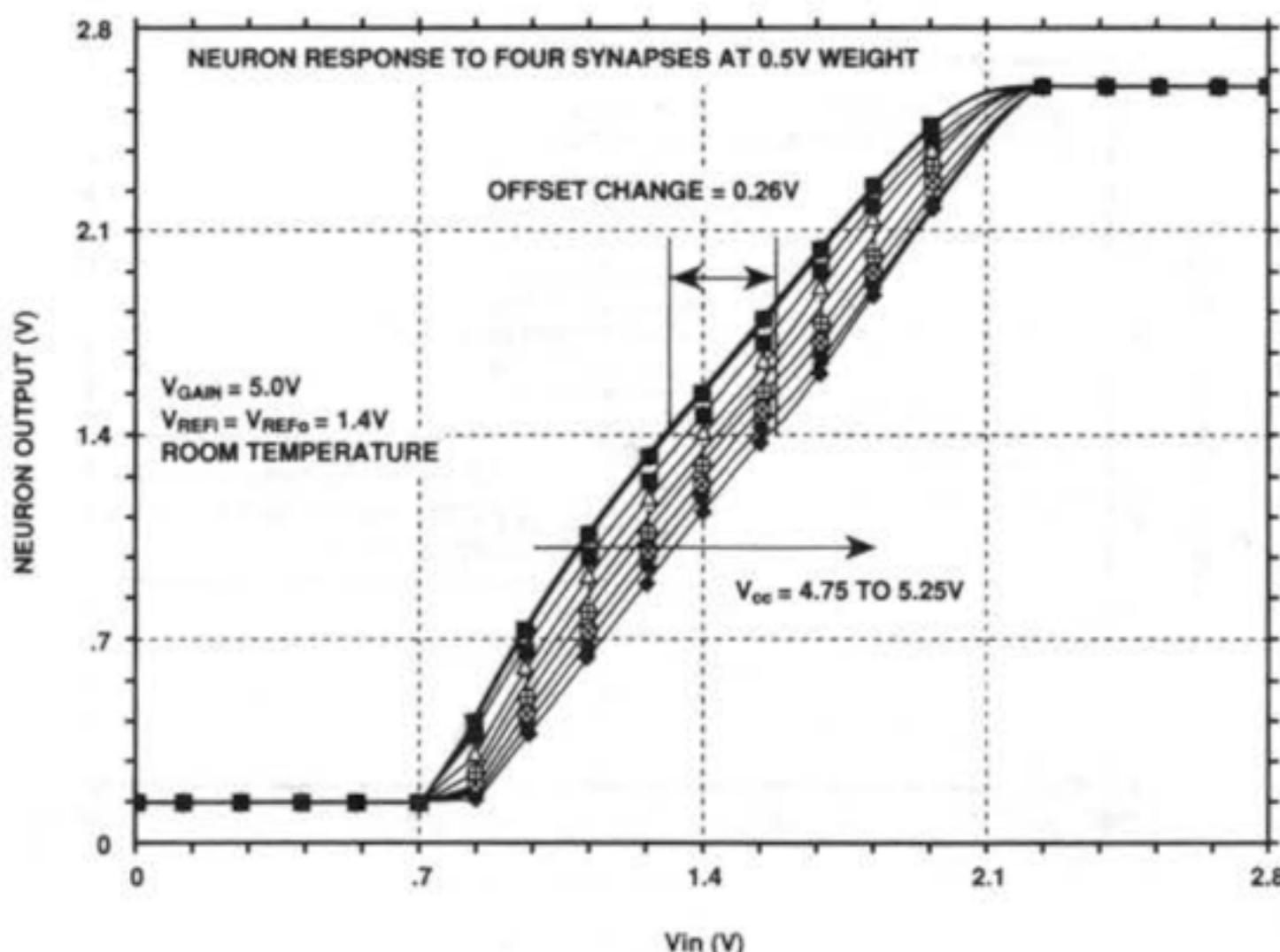


290408i25

Figure 25. Reduction of V_T Shifts Through Bake-Train-Bake

290408i26

Figure 26. Sigmoid Temperature Sensitivity



290408/27

Figure 27. Sigmoid V_{cc} Sensitivity

Chip-cycling performance was verified by performing cycling experiments on 50 chips. A cycle was defined as first setting the weight to $-2.5V$ (or less) and then setting it to $+2.5V$ (or more). All 10,240 synapses of each chip were subjected to 10,000 cycles. After the experiment, the iNNTS still had no difficulty setting all weights to $\pm 2.5V$. Performance After Radiation Exposure

The 80170NX chip's behavior and recognition performance were evaluated after subjecting the chip to ionization radiation (see Castro, et al, 1992). The chips were irradiated with electrons at 10MeV at a flux density between $1.8 \times 10^{10}/\text{cm}^2$ and $4.1 \times 10^{11}/\text{cm}^2$ up to an accumulated dose of about 26krad. Two pattern-recognition networks were involved in the study: a 64-45-52 character-recognition network and an 80-54-10 digit-recognition network.

Significant degradation in recognition performance occurred after about 6krad of irradiation. Figure 28 illustrates the mean-square-error (MSE) and misclassification statistics for the character-recognition network (for 104 patterns) as a function of radiation dose and elapsed time for one chip.

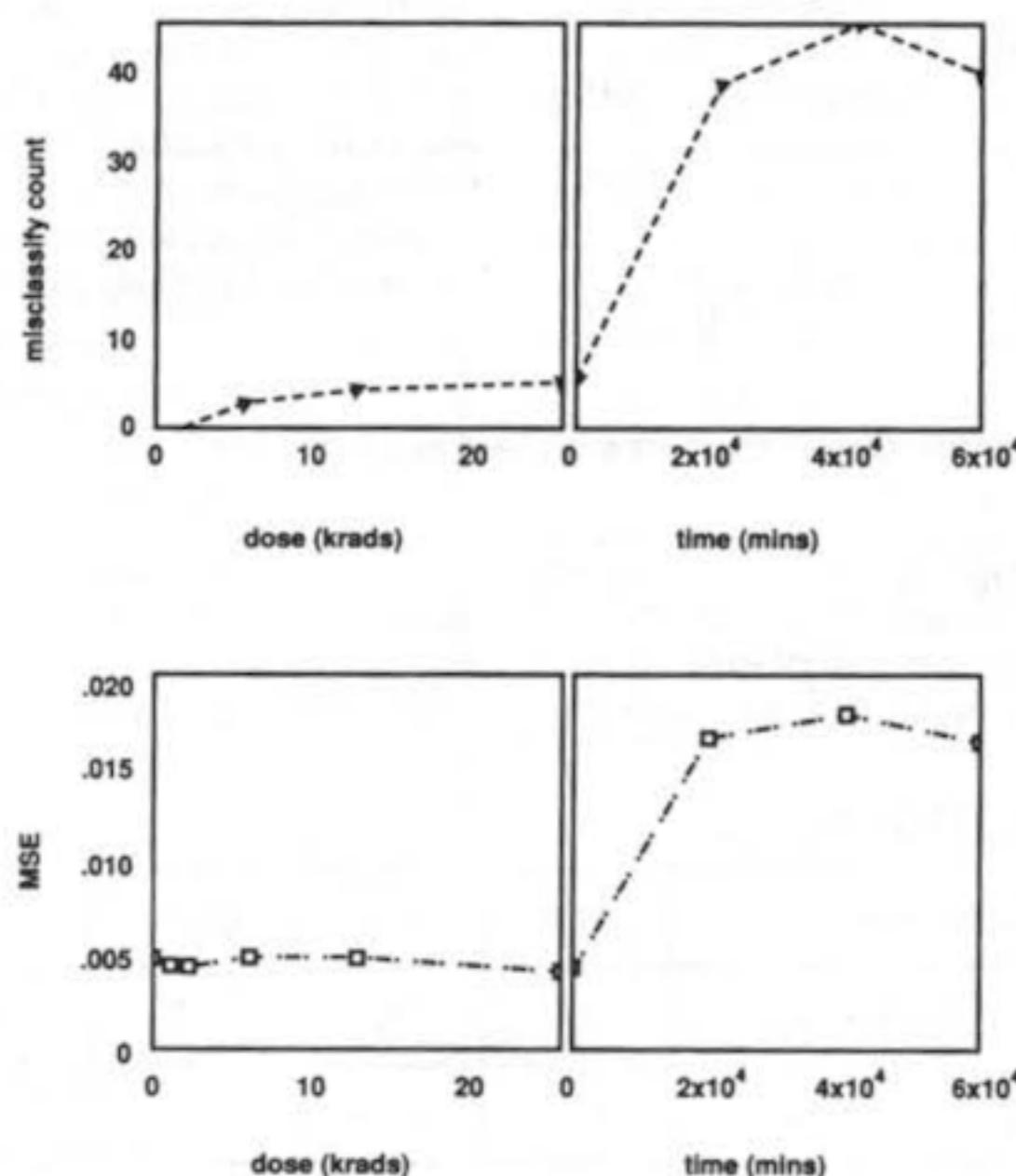
The degradation in performance can be attributed in part to the relaxation of the synapse weights shown for the same chip in Figure 29. The relaxation of synapse weights results from the combined effects of oxide charge generation and trapping, and the

photovoltaic effect caused by Bremsstrahlung radiation. Figures 28 and 29 show that the relaxation of weights continues after irradiation and has the most dominant effect on the degradation of recognition performance.

The chip-in-loop training method was employed on one of the character-recognition chips after each exposure to irradiation. 100% recognition was achieved within two such chip-in-loop training sessions, up until the last exposure at the accumulated dose of 26krad. Some permanent damage to the synapse weights occurred at that dose. Nevertheless, none of the irradiated chips exhibited total malfunction in performing classifications. Despite the failure of some synapses, most irradiated chips can be re-trained with the chip-in-loop method and/or by re-locating the weight array. However, the level of degradation in classification performance and the ability to be re-trained via chip-in-loop training are dependent on the particular network architecture, synapse weights, and reserve capacity of the synapse weights.

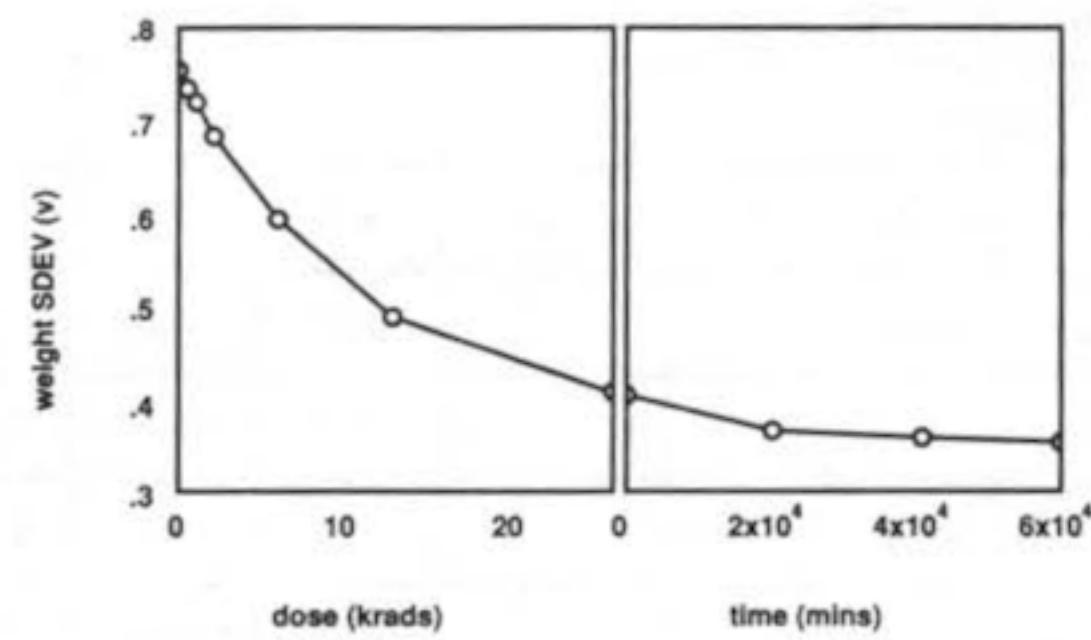
ABSOLUTE MAXIMUM RATINGS

Operating Temperature.....	0°C to +70°C
Storage Temperature	-65°C to +125°C
Bake-Train-Bake Temperature	75°C to +260°C



290408i28

Figure 28. Misclassified Patterns and MSE After Radiation Exposure



290408i29

Figure 29. Weight Distributions After Radiation Exposure

Voltage on Digital Inputs with Respect to V_{SS}	-2.0V to +7.0V ⁽¹⁾
Voltage on Analog Inputs and Outputs with Respect to V_{SS}	-2.0V to +7.0V ⁽¹⁾
V_{CC} Supply Voltage with Respect to V_{SS}	-2.0V to +7.0V ⁽¹⁾
V_{PP1}, V_{PP2} Voltages with Respect to V_{SS}	-2.0V to +21.0V ⁽¹⁾
Output Short Circuit Current	100 mA ⁽²⁾

* **NOTICE:** Stresses above those listed under Absolute Maximum Ratings may cause permanent damage to the device. This is a stress rating only and functional operation of the device at these or any conditions above those indicated in the operational sections of this specification is not implied. Exposure to Absolute Maximum Rating conditions for extended periods may affect device stability.

All specifications contained within the following tables are subject to change.

NOTES:

1. Minimum D.C. input voltage is -0.5V. During transitions: digital and analog inputs may undershoot to -2.0V; V_{PP} supplies may overshoot to +21.0V; and analog inputs and outputs may overshoot to voltages of $V_{CC} + 2.0V$, but for time periods of only 20 ns or less.
2. Output Short Circuit is for no more than 1 second, with no more than 1 output shorted at a time.

OPERATING CONDITIONS

Symbol	Parameter	Min	Max	Units	Conditions
T_A	Operating Temperature	0	70	°C	Processing and Training
V_{CC} ™	V_{CC} Supply Voltage	4.75	5.25	V	

DC CHARACTERISTICS At Room Temperature

Symbol	Parameter	Min	Max	Units	Conditions
I_{LI}	Input Leakage Current		±0.1	µA	$V_{IN} = 3.5V$
I_{LO}	Output Leakage Current		±1.0	µA	$V_{OUT} = 3.5V$
I_{OH}	Output High Source Current		-400	µA	$V_{OUT} \geq 2.4V, V_{REFo} = 1.4V$
I_{OL}	Output Low Sink Current		2.1	mA	$V_{OUT} \leq .45V, V_{REFo} = 1.4V$
V_{IA}	Analog Input Voltage	0	3.5	V	$V_{CC} = 5.0V$
V_{OA}	Analog Output Voltage	0	4.0	V	$V_{CC} = 5.0V, R_L = 10K\Omega$
I_{CC}	V_{CC} Active Supply Current		450	mA	$V_{CC} = 5.0V, CE\# = V_{IL}$
I_{CCS}	V_{CC} Standby Supply Current		1	mA	$V_{CC} = 5.0V, CE\# = V_{IH}$
I_{PI}	V_{PP1} Programming Supply Current		1	mA	$V_{PP1} = 20.0V, V_{CC} = 5.0V$
I_{P2}	V_{PP2} Programming Supply Current		1	mA	$V_{PP2} = 18.0V, V_{CC} = 5.0V$
V_{P1}	V_{PP1} High-Voltage Switch Voltage	18	19	V	see Training Algorithm
V_{P2}	V_{PP2} Weight Modify Pulse Voltage	12.5	18	V	see Training Algorithm

DC CHARACTERISTICS

At Room Temperature

Symbol	Parameter	Min	Max	Units	Conditions
V_{PL}	V_{PP1}, V_{PP2} Inactive/Low Voltage	4.75	5.25	V	
V_{IL}	Digital Input Low Voltage	-0.1	0.8	V	$V_{CC} = 5.0V$
V_{IH}	Digital Input High Voltage	2.0	5.5	V	$V_{CC} = 5.0V$
V_{REFi}	Input Reference Voltage	0	1.7	V	$[V_{REFi}(TTL) = 1.4V]$
V_{REFo}	Output Reference Voltage	0.5	2.0	V	For Symmetric operation maximum $V_{REFo}(TTL) = 1.6V$
V_{GAIN}	Gain Control Voltage	0.0	5.0	V	
V_{T_WT}, V_{T_REF}	Threshold Voltage (V_T) of weight EEPROM cell or reference EEPROM cell	-1.0	1.5	V	$I_D = 10 \mu A @ SYNO$ $V_S = 2.0V @ V_{PP2}$
V_{T_COM}	V_T Common Mode = $(V_{T_WT} + V_{T_REF})/2$	0	0.5	V	V_{T_COM} typical = 0.25V
V_{T_DIFF}	V_T Difference = $(V_{T_WT} - V_{T_REF})$, Weight Dynamic Range	-2.5	2.5	V	
Weight	Weight = $-V_{T_DIFF}$				
See Figure 27 for power supply sensitivity.					

AC CHARACTERISTICS

Symbol	Parameter	Min	Max	Units	Conditions
t_{IVOV}	Processing Delay (T_P) V_{GAIN} H_{GAIN}		3 1.5	μs μs	with 6- to 7-bit resolution at output pin, $Z_i = 100 \text{ pf}$ and $10K\Omega$ to ground. $V_{REFo} = 2V$
t_{ELQV}	Chip Enable to Output Valid (T_{CE})		20	μs	
t_{NLQV}	Neuron Enable to Output Valid (T_{NE})		5	μs	
t_{RHQV}	RESET _{i/f} High/Low to Valid Output		5	μs	
t_{IVHH}	Input Valid to HOLD High Setup	150		ns	
$t_{CL-H/CH-L}$	CLK Input Rise and Fall Time		100	ns	
t_{CHCL}	CLK High Pulse Width	150		ns	

AC CHARACTERISTICS

Symbol	Parameter	Min	Max	Units	Conditions
t_{OVCH}	Output Valid (Feedback Input) to CLK High Setup	150		ns	
t_{IH}	CLK Overlap for Reset	150		ns	
Vf_O	Output Slew Rate	5		V/ μ s	$C_L = 100 \text{ pF}$
t_{HOLD}	Sample and Hold Circuit Hold Time		10 0.8	s s	1% droop @ $T_A = 25^\circ\text{C}$ 1% droop @ $T_A = 70^\circ\text{C}$
t_{IVP1H}	Input Valid to V_{PP1} Setup	200		ns	
t_{P1HP2H}	V_{PP1} Setup Time Before V_{PP2}	100		ns	
$t_{P1L-H/H-L}$	V_{PP2} Pulse Rise/Fall Time	500		ns	
t_{P1HP1L}	V_{PP2} Pulse Width	0.5	1000	μ s	
t_{PLIV}	V_{PP1}, V_{PP2} Recovery Time	10		μ s	

CAPACITANCE

Symbol	Parameter	Min	Max	Units	Conditions
C_{IN}	Input Capacitance		12	pF	
C_{OUT}	Output Capacitance		12	pF	

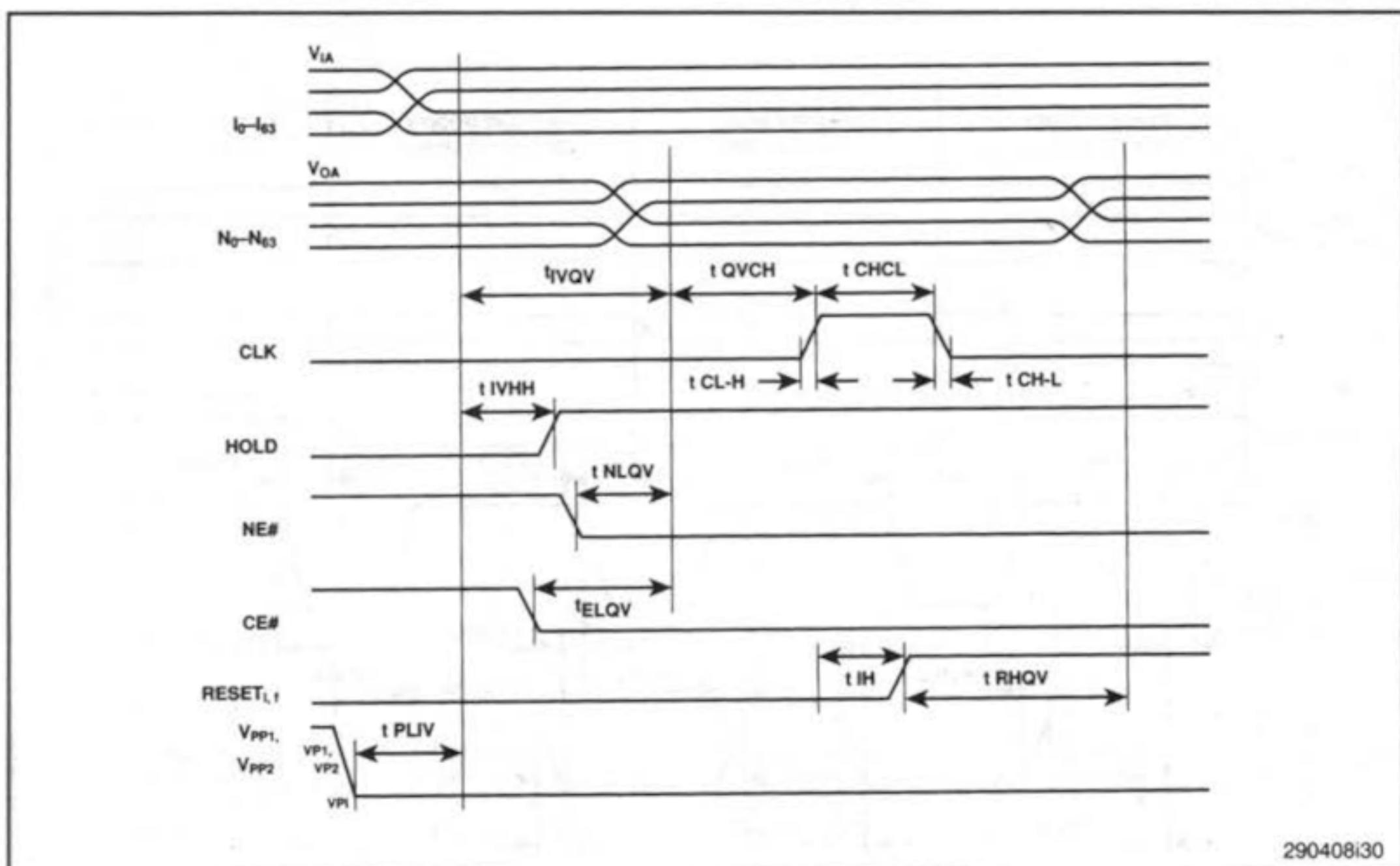


Figure 30. AC Waveforms for Parallel Processing

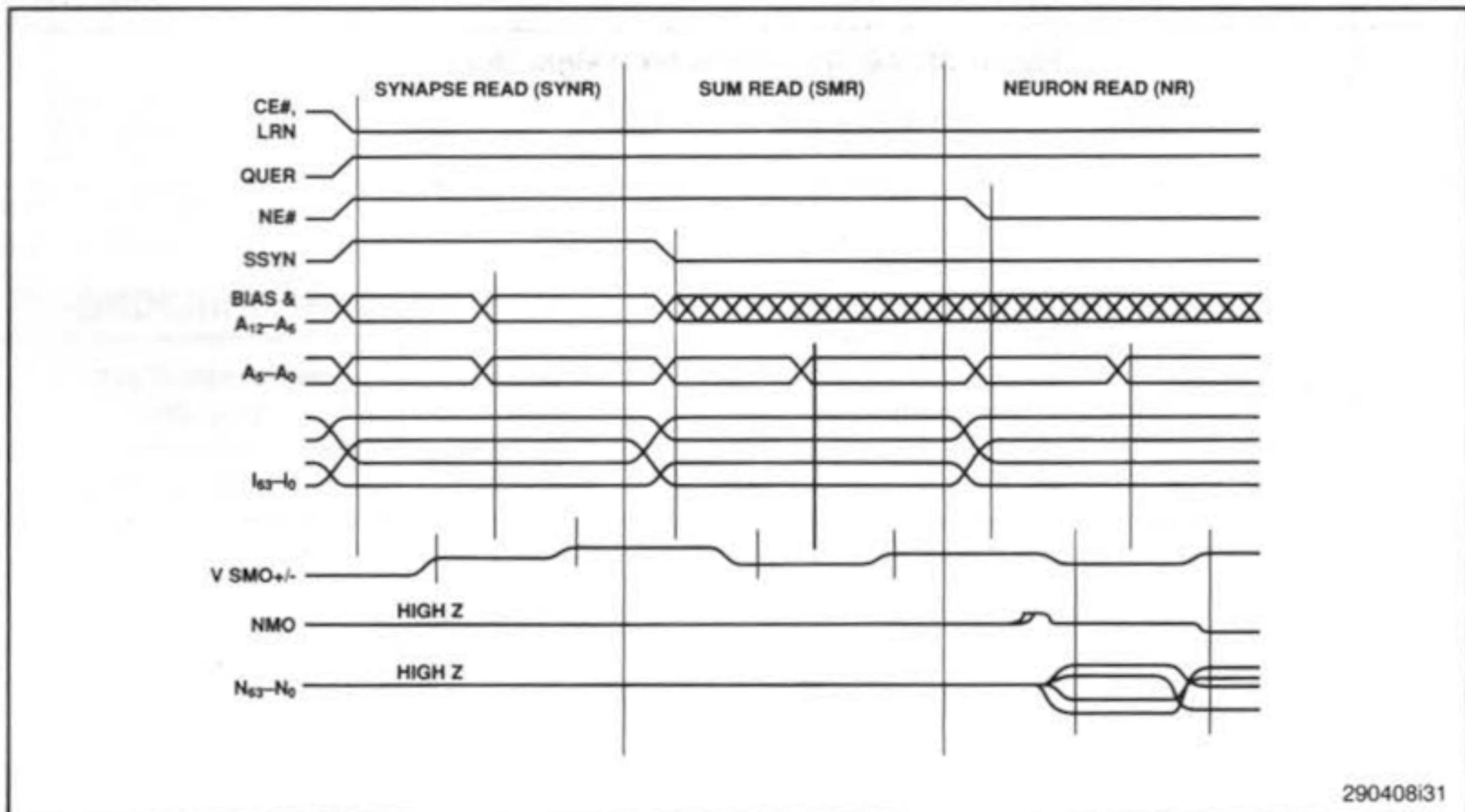


Figure 31. AC Waveforms for QUERY Modes

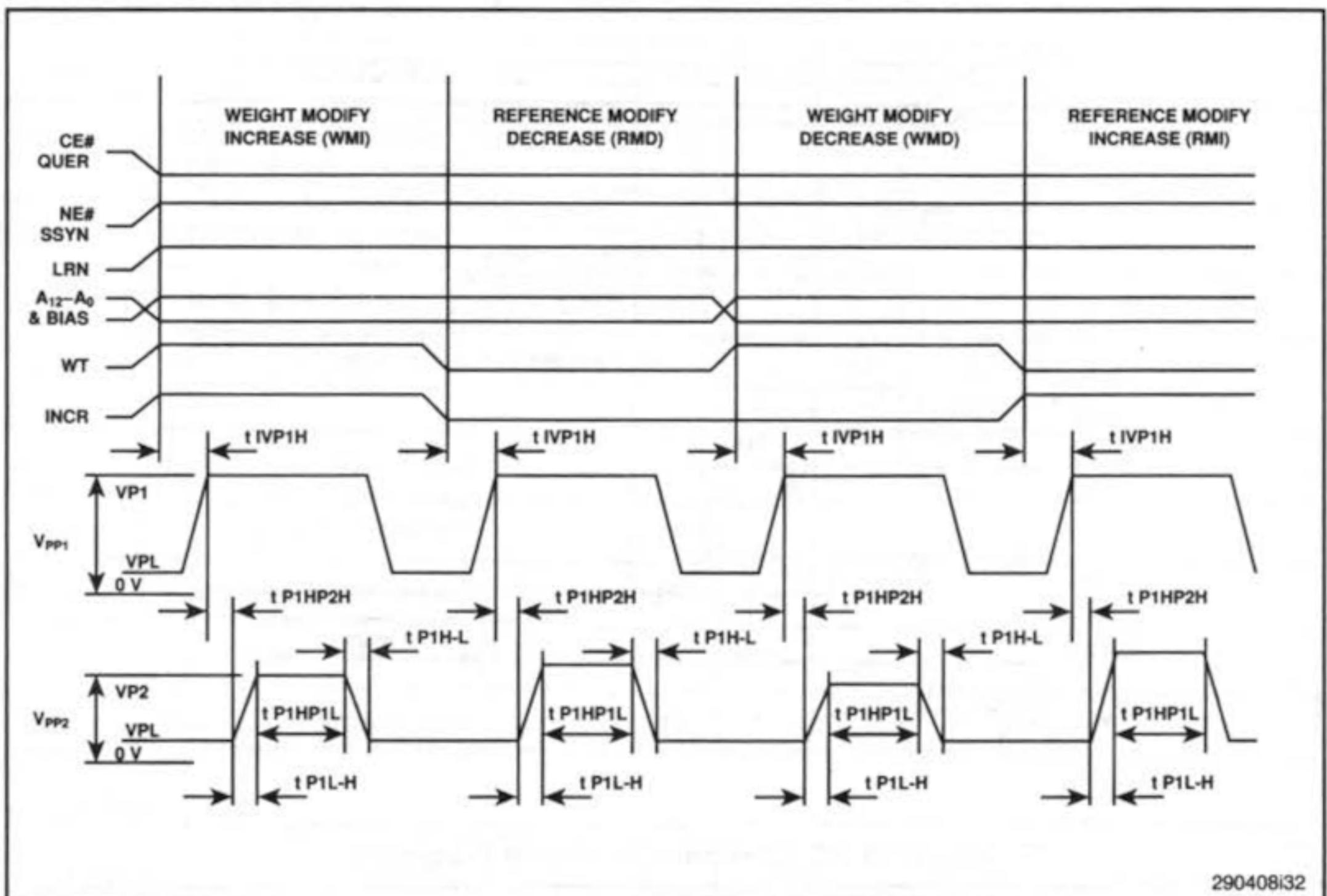


Figure 32. AC Waveforms for Weight Modify

PIN ASSIGNMENTS

A ₁ - N ₁₄	C ₂ - N ₁₃	E ₃ - N.C.	J ₁₅ - A ₃	P ₁ - I ₄₉	R ₂ - I ₄₀
A ₂ - N.C.	C ₃ - N ₁₂	E ₄ - N ₁₁	J ₁₆ - A ₁	P ₂ - I ₄₆	R ₃ - I ₃₉
A ₃ - N ₁₇	C ₄ - N.C.	E ₁₄ - N ₅₂	J ₁₇ - A ₄	P ₃ - N.C.	R ₄ - I ₃₇
A ₄ - N ₂₀	C ₅ - N ₁₉	E ₁₅ - N.C.	K ₁ - I ₆₁	P ₄ - N.C.	R ₅ - N.C.
A ₅ - V _{CC}	C ₆ - V _{SS}	E ₁₆ - N ₅₈	K ₂ - I ₅₉	P ₅ - N.C.	R ₆ - I ₂₉
A ₆ - N ₂₁	C ₇ - N ₂₆	E ₁₇ - N ₅₉	K ₃ - I ₆₂	P ₆ - N.C.	R ₇ - I ₃₀
A ₇ - N ₂₅	C ₈ - N ₃₀	F ₁ - N ₂	K ₄ - I ₆₀	P ₇ - I ₃₁	R ₈ - I ₂₆
A ₈ - N ₂₉	C ₉ - N ₃₂	F ₂ - H _{GAIN}	K ₁₄ - A ₇	P ₈ - I ₂₅	R ₉ - I ₂₂
A ₉ - V _{CC}	C ₁₀ - N ₃₃	F ₃ - N ₃	K ₁₅ - A ₅	P ₉ - V _{SS}	R ₁₀ - I ₁₇
A ₁₀ - N ₃₄	C ₁₁ - N ₃₇	F ₄ - N ₇	K ₁₆ - A ₈	P ₁₀ - I ₁₈	R ₁₁ - I ₁₃
A ₁₁ - N ₃₈	C ₁₂ - V _{SS}	F ₁₄ - N ₅₆	K ₁₇ - A ₆	P ₁₁ - I ₁₂	R ₁₂ - I ₁₄
A ₁₂ - N ₄₂	C ₁₃ - N ₄₄	F ₁₅ - N ₆₀	L ₁ - I ₅₇	P ₁₂ - N.C.	R ₁₃ - N.C.
A ₁₃ - V _{CC}	C ₁₄ - N.C.	F ₁₆ - D.U.	L ₂ - I ₅₅	P ₁₃ - N.C.	R ₁₄ - I ₆
A ₁₄ - N ₄₃	C ₁₅ - N ₅₁	F ₁₇ - N ₆₁	L ₃ - I ₅₈	P ₁₄ - N.C.	R ₁₅ - I ₄
A ₁₅ - N ₄₆	C ₁₆ - N ₅₀	G ₁ - V _{GAIN}	L ₄ - I ₅₄	P ₁₅ - N.C.	R ₁₆ - I ₃
A ₁₆ - N.C.	C ₁₇ - N ₅₅	G ₂ - N ₀	L ₁₄ - BIAS	P ₁₆ - V _{PP2}	R ₁₇ - V _{PP1}
A ₁₇ - N ₄₉	D ₁ - N ₆	G ₃ - V _{REF0}	L ₁₅ - A ₉	P ₁₇ - INC	S ₁ - I ₄₁
B ₁ - N ₁₀	D ₂ - N ₉	G ₄ - N ₁	L ₁₆ - A ₁₂	Q ₁ - I ₄₇	S ₂ - N.C.
B ₂ - N ₁₅	D ₃ - N.C.	G ₁₄ - N ₆₂	L ₁₇ - A ₁₀	Q ₂ - I ₄₂	S ₃ - I ₃₈
B ₃ - N ₁₆	D ₄ - N.C.	G ₁₅ - SMO-	M ₁ - I ₅₃	Q ₃ - I ₄₃	S ₄ - I ₃₅
B ₄ - N ₁₈	D ₅ - N.C.	G ₁₆ - N ₆₃	M ₂ - I ₅₆	Q ₄ - N.C.	S ₅ - I ₃₄
B ₅ - N.C.	D ₆ - N.C.	G ₁₇ - NMO	M ₃ - I ₅₂	Q ₅ - I ₃₆	S ₆ - I ₃₂
B ₆ - N ₂₄	D ₇ - N ₂₂	H ₁ - V _{CC}	M ₄ - I ₄₈	Q ₆ - I ₃₃	S ₇ - I ₂₈
B ₇ - N ₂₃	D ₈ - N ₂₈	H ₂ - V _{REFI}	M ₁₄ - LRN	Q ₇ - I ₂₇	S ₈ - I ₂₄
B ₈ - N ₂₇	D ₉ - V _{CC}	H ₃ - RESET _F	M ₁₅ - NE#	Q ₈ - I ₂₃	S ₉ - V _{SS}
B ₉ - N ₃₁	D ₁₀ - N ₃₅	H ₄ - V _{SS}	M ₁₆ - A ₁₁	Q ₉ - I ₂₁	S ₁₀ - I ₁₉
B ₁₀ - N ₃₆	D ₁₁ - N ₄₁	H ₁₄ - V _{SS}	M ₁₇ - WT	Q ₁₀ - I ₂₀	S ₁₁ - I ₁₅
B ₁₁ - N ₄₀	D ₁₂ - N.C.	H ₁₅ - A ₀	N ₁ - I ₅₁	Q ₁₁ - I ₁₆	S ₁₂ - I ₁₁
B ₁₂ - N ₃₉	D ₁₃ - N.C.	H ₁₆ - SMO+	N ₂ - I ₅₀	Q ₁₂ - I ₁₀	S ₁₃ - I ₉
B ₁₃ - N.C.	D ₁₄ - N.C.	H ₁₇ - V _{CC}	N ₃ - N.C.	Q ₁₃ - I ₇	S ₁₄ - I ₈
B ₁₄ - N ₄₅	D ₁₅ - N.C.	J ₁ - I ₆₃	N ₄ - I ₄₄	Q ₁₄ - N.C.	S ₁₅ - I ₅
B ₁₅ - N ₄₇	D ₁₆ - N ₅₄	J ₂ - CLK	N ₁₄ - SSYN	Q ₁₅ - I ₀	S ₁₆ - N.C.
B ₁₆ - N ₄₈	D ₁₇ - N ₅₇	J ₃ - HOLD	N ₁₅ - N.C.	Q ₁₆ - I ₁	S ₁₇ - I ₂
B ₁₇ - N ₅₃	E ₁ - N ₄	J ₄ - RESET _I	N ₁₆ - QUER	Q ₁₇ - CE#	
C ₁ - N ₈	E ₂ - N ₅	J ₁₄ - A ₂	N ₁₇ - SYNO	R ₁ - I ₄₅	

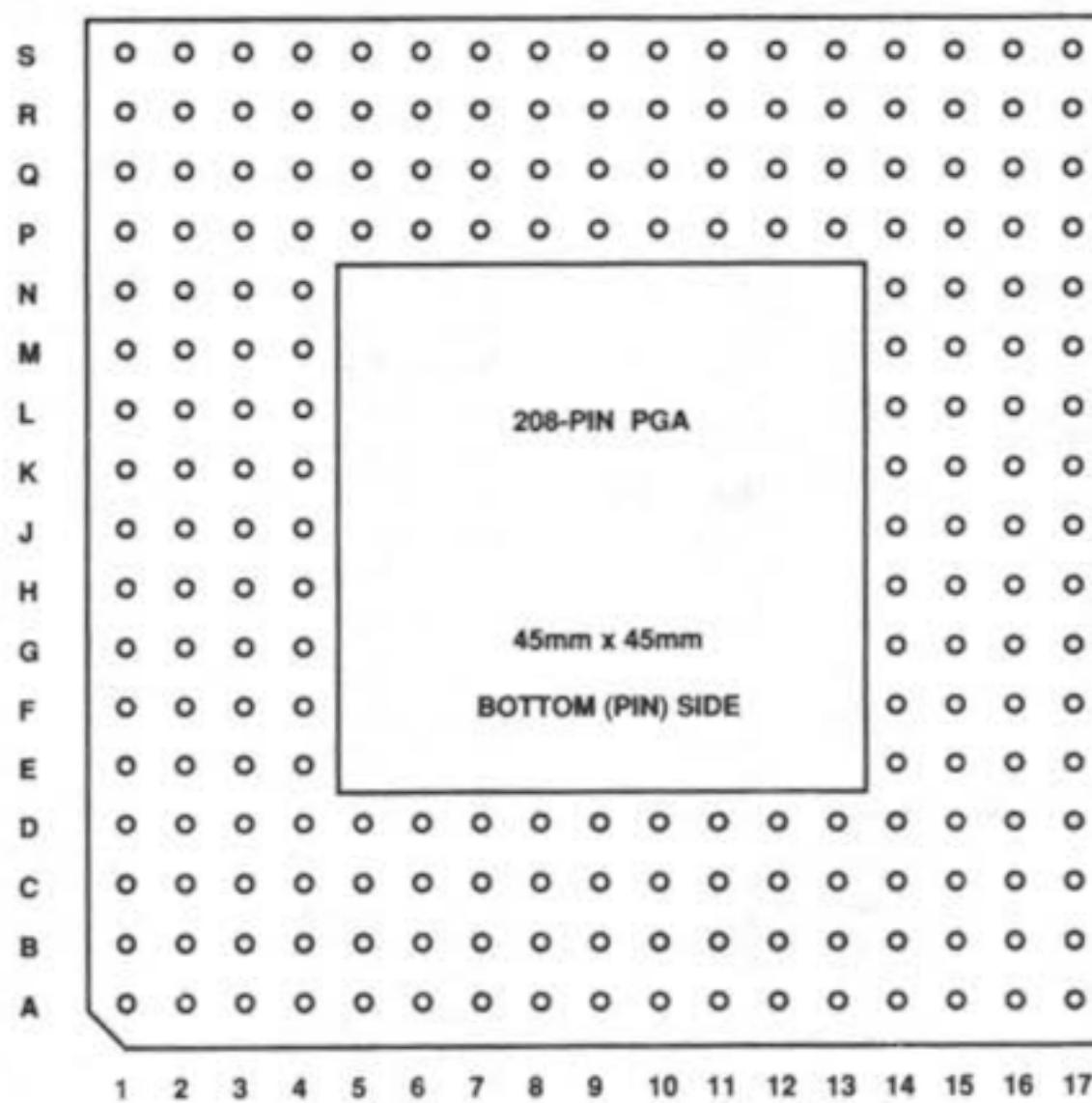
N.C. = NO CONNECT

D.U. = DON'T USE (RESERVED)

ORDERING INFORMATION

Test/Temperature Category	Package Type	Performance Specification	Order Code
Commercial	208-PGA	Standard t _{IVQV} = 3 μS	A80170NX

208-PIN PGA PACKAGE



290408i33

REFERENCES

General

- 80170NX Neural Network Technology and Applications. Intel Corp., 1992.
- Holler, M., S. Tam, H. Castro, and R. Benson, "An Electrically Trainable Artificial Neural Network (ETANN) with 10240 'Floating Gate' Synapses," *International Joint Conference on Neural Networks*, June 1989. Vol. II, 191.
- Roy, M., M. Holler, S. Tam, and H. Castro, "Testing An Analog VLSI Neural Network with 10240 'Floating Gate' Synapses," *1990 IEEE VLSI Test Symposium*.
- Tam, S., M. Holler, J. Brauch, A. Pine, A. Peterson, S. Anderson, and S. Deiss, "Reconfigurable Multi-Chip Analog Neural Network; Recognition and Back Propagation Training," *1992 International Joint Conference on Neural Networks*, Baltimore, Maryland. Vol. 2, 625.
- Tam, S., B. Gupta, H. Castro, and M. Holler, "Learning On An Analog VLSI Neural Network Chip," *Proceedings of the 1990 IEEE International Conference on Systems, Man, and Cybernetics*. November 1990, pp. 701-703.
- Calvin, J., S. Rogers, D. Zahirniak, "Characterization of the 80170NX (ETANN) Chip Sigmoidal Transfer Function for a Device $V_{GAIN} = 3.3V$ " (available from Intel).
- Castro, H., et al, "Radiation Exposure Effects on the Performance of the ETANN (80170NX)," Intel Corporation, 1992.
- Kern, L., D. Andes, "Design and Development of a Real-Time Neural Processor Using the Intel 80170NX Electrically Trainable Artificial Neural Network (ETANN)," *1992 Government Microcircuit Application Conference, Digest of Papers*, pp. 269-272.
- Badgett, W., K. Burkett, M. Campbell, D. Wu, B. Denby, C. Lindsey, R. Blair, S. Kuhlmann, J. Romano, "A Neural Network Calorimeter Trigger Used in CDF," *IEEE 1992 Nuclear Science Symposium and Medical Imaging Conference*, October 25-31, 1992, Orlando, Florida.
- Thorson, M., F. Warthman, M. Holler, "A Neural-Network Audio Synthesizer," *Dr. Dobb's Journal*, February, 1993.
- Brauch, J., S. Tam, M. Holler, A. Shmurun, "Analog VLSI Neural Networks for Impact Signal Processing," *IEEE MICRO Dec.*, 1992, pp. 34-45.
- Akkila, T., T. Francke, T. Lindblad, A. Eide, "An Analog Neural Network Hardware Solution to Cherenkov Ring Imaging Particle Identifier," *Nuclear Instruments and Methods in Physics Research*, North-Holland, March 1993
- Akkila, T., T. Lindblad, B. Lund-Jenson, G. Szekely, A. Eide, "A Hardware Implementation of an Analog Neural Network for Gaussian Peak-fitting," *Nuclear Instruments and Methods in Physics Research*, North-Holland, March 1993.
- Wang, C., C. Wu, "Analog Neural Networks Solve Ambiguity Problems in Medium PRF Radar Systems," *1993 IEEE International Conference on Neural Networks*, San Francisco, CA.
- Brown, J., E. DeRouin, E. Straub, "Integrated Detection and Segmentation for Hyperspectral Imagery Using Neural Networks," *Proc. International Symposium on Spectral Sensing Research*, Maui, Nov. 15-20, 1992.

Cognitive Psychology

- Rumelhart, D.E. and J.L. McClelland, *Parallel Distributed Processing Explorations in the Microstructure of Cognition*.
Volume 1: Foundations
Volume 2: Psychological and Biological Models
MIT Press, Bradford Book, Cambridge, Mass, 1988. Original printing 1986.

Rumelhart, D.E. and J.L. McClelland, *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. (Programs on floppy will run on an IBM PC. C Source code included). MIT Press, Bradford Book, Cambridge, Mass., 1988.

Rumelhart, David, Geoffrey E. Hinton, and Ronald J. Williams, "Learning Representations by Back-Propagating Errors," *Nature* Vol. 323 No. 9, October 1986.

Neurobiology

Kandel, E. R., and J. H. Schwartz, *Principles of Neural Science*. 2nd edition. Elsevier, New York. Medical text with extensive detail on the function of the neuron and the human nervous system.

Burnod, Y., *An Adaptive Neural Network: The Cerebral Cortex*. Prentice Hall, London, 1990.

Neural Network Models and Learning

Kohonen, T., *Self-Organization and Associative Memory*. Springer-Verlag, New York. 2nd edition 1988. First edition 1984. Good general reference.

Lippmann, R. P., "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, April 1987. Overview article of the various networks and learning paradigms for an issue of ASSP dedicated to neural networks.

Koch, C. and I. Jegev, ed. *Methods in Neuronal Modeling*. The MIT Press, Cambridge, Mass., 1989.

Hinton, G., and J. Anderson, ed. *Parallel Models of Associative Memory*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 1989.

Kanerva, P., *Sparse Distributed Memory*. Bradford Book, The MIT Press, Cambridge, Mass., 1988.

Hopfield J. J. and D. W. Tank, "Computing with Neural Circuits: A Model," *Science* Vol. 233, August 1986: 625.

Anderson, J. and E. Rosenfeld, ed. *Neurocomputing Foundations of Research*. The MIT Press, Cambridge, Mass., 1988.

Andes, D., B. Widrow, M. Lehr, E. Wan, "Madaline Rule III: A Robust Algorithm for Training Neural Networks," *International Joint Conference on Neural Networks*, 1990, Vol. I: 533-536.

Speech Recognition

Parsons, T., *Voice and Speech Processing*. McGraw-Hill, New York.

Waibel, A., T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE Trans. on Acoustics, Speech and Signal Processing* Vol. 37, No. 3, March 1989. Work by Japanese Research consortium ATR in conjunction with Carnegie-Mellon University.

Waibel, A. and K. Lee, ed. *Readings in Speech Recognition*. Morgan Kaufmann Publishers Inc., San Mateo, California, 1990.

Temporal Pattern Recognition

Wolf, L., *Recurrent Nets for the Storage of Cyclic Sequences*.

Williams, R. J. and D. Zipser, "A Learning Algorithm For Continually Running Fully Recurrent Neural Networks," *Institute for Cognitive Science Report #8805*, Oct. 1988, U.C. San Diego.

Pearlmutter, B. A, "Learning State Space Trajectories in Recurrent Neural Networks," *Proceedings of the International Joint Conference on Neural Networks*, June 1989, Vol. II: 365.

Implementation

Mead, C., *Analog VLSI and Neural Systems*. Addison Wesley, New York, 1989.

Mead, C. and M. Ismail, ed. *Analog VLSI Implementation of Neural Systems*. Kluwer Academic Press, Boston, 1989.

von Neumann, J., *The Computer and the Brain*. Yale University Press, New Haven, Connecticut, 1958.

Control

Nguyen, D. and B. Widrow, "The Truck Backer Upper: An Example of Self Learning in Neural Networks," *International Joint Conference on Neural Networks*, June 1989, Washington, D.C.

Miller, W., R. Sutton, P. Werbos, "Neural Networks for Control," *The MIT Press*, Cambridge MA, 1991.

Staib, W. and R. Staib, "The Intelligent Arc Furnace Controller: A Neural Network Electrode Position Optimization System for the Electric Arc Furnace," *International Joint Conference on Neural Networks*, June 1992, Vol. III: 1.

GLOSSARY

Bias Synapse—See *fixed-input bias synapse*.

Bake-Train-Bake—An experimental procedure used to enhance the long-term stability of the 80170NX synapse weights. In this procedure, the chip is baked, then trained, then baked again.

Chip-In-Loop Training—A procedure used to train an 80170NX-based neural network, also known as chip-in-loop optimization. In this procedure, actual 80170NX neuron outputs are used in the training algorithm to arrive at the weight updates. The procedure compensates for non-ideal characteristics of the chip and the training algorithm.

Dot Product—The quantity obtained by multiplying the components of two vectors (the *external inputs* and their corresponding *synapse weights*) and adding the products. Also called *inner product* or *scalar product*. The equation is:

$$\text{Dot Product} = \sum_i \text{input}_i * \text{Weight}_i$$

Erase—To decrease the threshold voltage of a *weight EEPROM* cell or *reference EEPROM* cell.

EEPROM—Electrically Erasable Programmable Read-Only Memory.

EEPROM Cell—One of two floating-gate EEPROM cells that, together, store a synapse weight. The two cells are the *weight EEPROM* cell and the *reference EEPROM* cell. The synapse weight is the difference between the threshold voltages at each of these cells.

External Input—One of 64 input pins on the 80170NX chip.

External-Input Synapse—A synapse connected to an external input.

Feedforward Neural Network—A neural network in which inputs are received externally or from a previous layer, are processed, and are passed on to only one subsequent layer. Neurons in a given layer do not connect to one another.

Fixed-Input Bias Synapse—A synapse used to provide the threshold for the neuron's sigmoid function and to cancel offsets due to physical non-linearities in the chip. The Input and Feedback Arrays each have 16 fixed-input bias synapses. Mathematically, all of these bias synapses act as a single bias. However, since the dynamic range of a single EEPROM bias synapse is limited, 16 of them are used. All 16 contribute to the dot product performed by the neuron, although seven are set by the training program and the other nine are set by a chip-initialization process.

Hidden Layer—A group of neurons that is not connected directly to output pins on the chip. Any of the layers in a feedforward neural network that processes information prior to the output layer.

Inner Product—See *dot product*.

Neuron—In biology, a cell in the brain that produces an output signal in response to multiple input signals. In the 80170NX chip, a structure that computes the dot product of input signals and synapse weights, then computes a sigmoid threshold function to produce an output signal.

Output Layer—A group of neurons with outputs that are externally accessible.

Pattern—An input vector. The 80170NX can accept input vectors with up to 128 components.

PDP—Parallel distributed processing. A term for neural network processing. In PDP, a particular piece of information is represented by a distributed pattern of activity on a number of neurons rather than by the activity of a single neuron. This method of distributed representation is the source of a neural network's robustness to faults in individual neurons. The term was coined by Rumelhart, et. al.

Program—To increase the threshold voltage of a *weight* EEPROM cell or *reference* EEPROM cell.

Reference EEPROM Cell—One of two floating-gate EEPROM cells that, together, store a synapse weight. See EEPROM cell and *weight* EEPROM cell.

Sigmoid Function—A monotonically increasing function that is shaped like a stretched-out letter "S." This nonlinear function gives the neuron the ability to make "decisions" and perform nonlinearly separable classification tasks.

Synapse—In biology, a variable-strength connection between neurons that can be either output-exciting or output-inhibiting. In the 80170NX chip, a structure that includes a multiplier and a stored weight. The synapse multiplies an input by the weight stored in the synapse and contributes this product to the dot product calculated by the neuron.

Synapse Weight—A nonvolatile weight that is stored at a synapse. The value of the weight can be between +2.5 and -2.5.

Two-Layer Neural Network—A neural network consisting of one hidden layer and one output layer.

Weight—See *synapse weight*.

Weight EEPROM Cell—One of two floating-gate EEPROM cells that, together, store a synapse weight. See EEPROM cell and *reference* EEPROM cell.

Weight Setting—The process of programming one EEPROM cell and erasing another EEPROM cell, in a synapse cell-pair, so as to set a desired weight value for that synapse.

NEURAL NET CUSTOMER SUPPORT

Intel Neural Net Group

MS: RN3-17

2200 Mission College Blvd.

Santa Clara, CA 95052-8119

Phone: (408) 765-9235

Fax: (408) 765-9797



NORTH AMERICAN SALES OFFICES

ALABAMA

Intel Corp.
600 Boulevard South
Suite 104-I
Huntsville 35802
Tel: (800) 628-8686
FAX: (205) 883-3511

ARIZONA

†Intel Corp.
410 North 44th Street
Suite 500
Phoenix 85008
Tel: (800) 628-8686
FAX: (602) 244-0446

CALIFORNIA

Intel Corp.
1 Sierra Gate Plaza
Suite 280C
Roseville 95678
Tel: (800) 628-8686
FAX: (916) 782-8153

†Intel Corp.
9665 Chesapeake Dr.
Suite 325
San Diego 92123
Tel: (800) 628-8686
FAX: (619) 292-0628

Intel Corp.
1781 Fox Drive
San Jose 95131
Tel: (800) 628-8686
FAX: (408) 441-9540

*†Intel Corp.
400 N. Tustin Avenue
Suite 450
Santa Ana 92705
Tel: (800) 628-8686
TWX: 910-595-1114
FAX: (714) 541-9157

†Intel Corp.
15260 Ventura Boulevard
Suite 360
Sherman Oaks 91403
Tel: (800) 628-8686
FAX: (818) 995-6624

COLORADO

*†Intel Corp.
600 S. Cherry St.
Suite 700
Denver 80222
Tel: (800) 628-8686
TWX: 910-931-2289
FAX: (303) 322-8670

CONNECTICUT

†Intel Corp.
103 Mill Plain Road
Danbury 06811
Tel: (800) 628-8686
FAX: (203) 794-0339

FLORIDA

†Intel Corp.
800 Fairway Drive
Suite 160
Deerfield Beach 33441
Tel: (800) 628-8686
FAX: (305) 421-2444

Intel Corp.
2250 Lucien Way
Suite 100-B, 27
Maitland 32751
Tel: (800) 628-8686
FAX: (407) 660-1283

GEORGIA

†Intel Corp.
20 Technology Parkway
Suite 150
Norcross 30092
Tel: (800) 628-8686
FAX: (404) 605-9762

ILLINOIS

*†Intel Corp.
Woodfield Corp. Center III
300 N. Martingale Road
Suite 400
Schaumburg 60173
Tel: (800) 628-8686
FAX: (708) 706-9762

INDIANA

†Intel Corp.
8910 Purdue Road
Suite 350
Indianapolis 46268
Tel: (800) 628-8686
FAX: (317) 875-8938

MARYLAND

*†Intel Corp.
10010 Junction Dr.
Suite 200
Annapolis Junction 20701
Tel: (800) 628-8686
FAX: (410) 206-3678

MASSACHUSETTS

*†Intel Corp.
Westford Corp. Center
5 Carlisle Road
2nd Floor
Westford 01886
Tel: (800) 628-8686
TWX: 710-343-6333
FAX: (508) 692-7867

MICHIGAN

†Intel Corp.
7071 Orchard Lake Road
Suite 100
West Bloomfield 48322
Tel: (800) 628-8686
FAX: (313) 851-8770

MINNESOTA

†Intel Corp.
3500 W. 80th St.
Suite 360
Bloomington 55431
Tel: (800) 628-8686
TWX: 910-576-2867
FAX: (612) 831-6497

NEW JERSEY

*†Intel Corp.
Lincroft Office Center
125 Half Mile Road
Red Bank 07701
Tel: (800) 628-8686
FAX: (908) 747-0983

NEW YORK

*Intel Corp.
850 Crosskeys Office Park
Fairport 14450
Tel: (800) 628-8686
TWX: 510-253-7391
FAX: (716) 223-2561

†Intel Corp.

300 Westage Business Center
Suite 230
Fishkill 12524
Tel: (800) 628-8686
FAX: (914) 897-3125

*†Intel Corp.
2950 Express Dr., South
Suite 130
Islandia 11722
Tel: (800) 628-8686
TWX: 510-227-6236
FAX: (516) 348-7939

OHIO

*Intel Corp.
Four Commerce Park Square
23200 Chagrin Blvd., Suite 600
Beachwood 44122
Tel: (800) 628-8686
FAX: (216) 464-2270

*†Intel Corp.
3401 Park Center Drive
Suite 220
Dayton 45414
Tel: (800) 628-8686
TWX: 810-450-2528
FAX: (513) 890-8658

OKLAHOMA

Intel Corp.
6801 N. Broadway
Suite 115
Oklahoma City 73162
Tel: (800) 628-8686
FAX: (405) 840-9819

OREGON

†Intel Corp.
15254 N.W. Greenbrier Pkwy.
Building B
Beaverton 97006
Tel: (800) 628-8686
TWX: 910-467-8741
FAX: (503) 645-8181

PENNSYLVANIA

*†Intel Corp.
925 Harvest Drive
Suite 200
Blue Bell 19422
Tel: (800) 628-8686
FAX: (215) 641-0785

SOUTH CAROLINA

Intel Corp.
7403 Parklane Rd., Suite 3
Columbia 29223
Tel: (800) 628-8686
FAX: (803) 788-7999

Intel Corp.
100 Executive Center Drive
Suite 109, B183
Greenville 29615
Tel: (800) 628-8686
FAX: (803) 297-3401

TEXAS

†Intel Corp.
8911 N. Capital of Texas Hwy.
Suite 4230
Austin 78759
Tel: (800) 628-8686
FAX: (512) 338-9335

*†Intel Corp.

5000 Quorum Drive
Suite 750
Dallas 75240
Tel: (800) 628-8686

*†Intel Corp.
20515 SH 249
Suite 401
Houston 77070
Tel: (800) 628-8686
TWX: 910-881-2490
FAX: (713) 988-3660

UTAH

†Intel Corp.
428 East 6400 South
Suite 135
Murray 84107
Tel: (800) 628-8686
FAX: (801) 268-1457

WASHINGTON

†Intel Corp.
2800 158th Avenue S.E.
Suite 105
Bellevue 98007
Tel: (800) 628-8686
FAX: (206) 746-4495

WISCONSIN

Intel Corp.
400 N. Executive Dr.
Suite 401
Brookfield 53005
Tel: (800) 628-8686
FAX: (414) 789-2746

CANADA

BRITISH COLUMBIA
Intel Semiconductor of Canada, Ltd.
999 Canada Place
Suite 404, #11
Vancouver V6C 3E2
Tel: (800) 628-8686
FAX: (604) 844-2813

ONTARIO

†Intel Semiconductor of Canada, Ltd.
2650 Queensview Drive
Suite 250
Ottawa K2B 8H6
Tel: (800) 628-8686
FAX: (613) 820-5936

†Intel Semiconductor of Canada, Ltd.
190 Attwell Drive
Suite 500
Rexdale M9W 6H8
Tel: (800) 628-8686
FAX: (416) 675-2438

QUEBEC

†Intel Semiconductor of Canada, Ltd.
1 Rue Holiday
Suite 115
Tour East
Pt. Claire H9R 5N3
Tel: (800) 628-8686
FAX: 514-694-0064



UNITED STATES
Intel Corporation
2200 Mission College Boulevard
P.O. Box 58119
Santa Clara, CA 95052-8119

JAPAN
Intel Japan K.K.
5-6 Tokodai, Tsukuba-shi
Ibaraki, 300-26

FRANCE
Intel Corporation S.A.R.L.
1, Rue Edison, BP 303
78054 Saint-Quentin-en-Yvelines Cedex

UNITED KINGDOM
Intel Corporation (U.K.) Ltd.
Pipers Way
Swindon
Wiltshire, England SN3 1RJ

GERMANY
Intel GmbH
Dornacher Strasse 1
8016 Feldkirchen bei Muenchen

HONG KONG
Intel Semiconductor Ltd.
32/F Two Pacific Place
88 Queensway
Central

CANADA
Intel Semiconductor of Canada, Ltd.
190 Attwell Drive, Suite 500
Rexdale, Ontario M9W 6H8

Printed in U.S.A./0393/3K/CG/NCG CC
NUERAL NET PRODUCTS