Via Santa Maria Maddalena 12, 38100 Trento, Italy
tel. +39-0461-260 552 - fax + 39-0461-260 617
e-mail: info@neuricam.com; http: www.neuricam.com

NC3002 TOTEM
Digital Processor for Neural Networks

DATA SHEET
Rel. 12/99

General features

The NC3002 is a digital VLSI parallel processor for fast learning and recognition with artificial neural networks in quality inspection applications where a high throughput of recognition is required. Its architecture (see Fig. 1) is optimised for the implementation of the "Reactive Tabu Search" learning algorithm, a competitive alternative to back-propagation which does not require derivatives of the transfer function and is ideally suited for VLSI implementation.
Its main features are summarised below:

- Pipelined Digital Data Stream, Single Instruction Multiple Data (SIMD) architecture optimised for the execution of the multiply-accumulate operation:

  Acc(n+1) := Acc(n) + DataIn * Weight (n)

  in a single clock cycle in parallel on 32 fast processing units
- 32 fixed-point fully-parallel multiply-and-accumulate processors (MACs) operating in parallel from a common broadcast bus. 2's complement data format
- 64-Kbit internal dynamic random-access memory organised as 32 blocks of 256x8 bits for weight storage with close coupling with processors. Memory can be assigned either to a single neuron or be partitioned among several neurons to implement multi-layer networks with a single chip. Refresh is transparent to the user.
- 32-input, 16-output barrel shifter for scaling of results
- Limited word width for economical layout: 16-bit data, 8-bit weight, 16 or 32-bit results
- Performance of 1500 million multiply-and-accumulate operation per second with a 50 MHz clock. A multi-layer perceptron with a 64-128-64 topology suitable to solve the target application can be evaluated in about 9 µs by two NC3002 chips operating in parallel. Higher performance can be achieved by paralleling up to four chips per network level to implement neurons with up to 256 inputs
- Simple interface with data input, data output, memory address and control buses. Evaluation boards for ISA, PCI, VME, CompactPCI, PC/104 bus and HitachiH8-based embedded board are available, including software drivers and graphic tools.
- Support circuitry for two external 16-input, 16-output look-up tables (LUT) implemented in static RAM to calculate the activation function of the neural network
- Compact chip size of 42 mm^2 and limited pin count for ease of interfacing and low system cost
- Chip implementation in AMS 0.8µm single-polysilicon, double-metal CMOS technology.

Number of processors	32
Total on-chip RAM	64 Kbit
Processing power	1000 million MAC operations per second
Clock frequency	40 MHz
Sustained input rate	1 input/clock cycle
Latency	3 clock cycles
Die size	42 mm^2
No. of transistors	350.000
Power consumption	1.5 W @ 5V and 40 MHz
Package	132-pin CPGA

Table 1. Chip characteristics

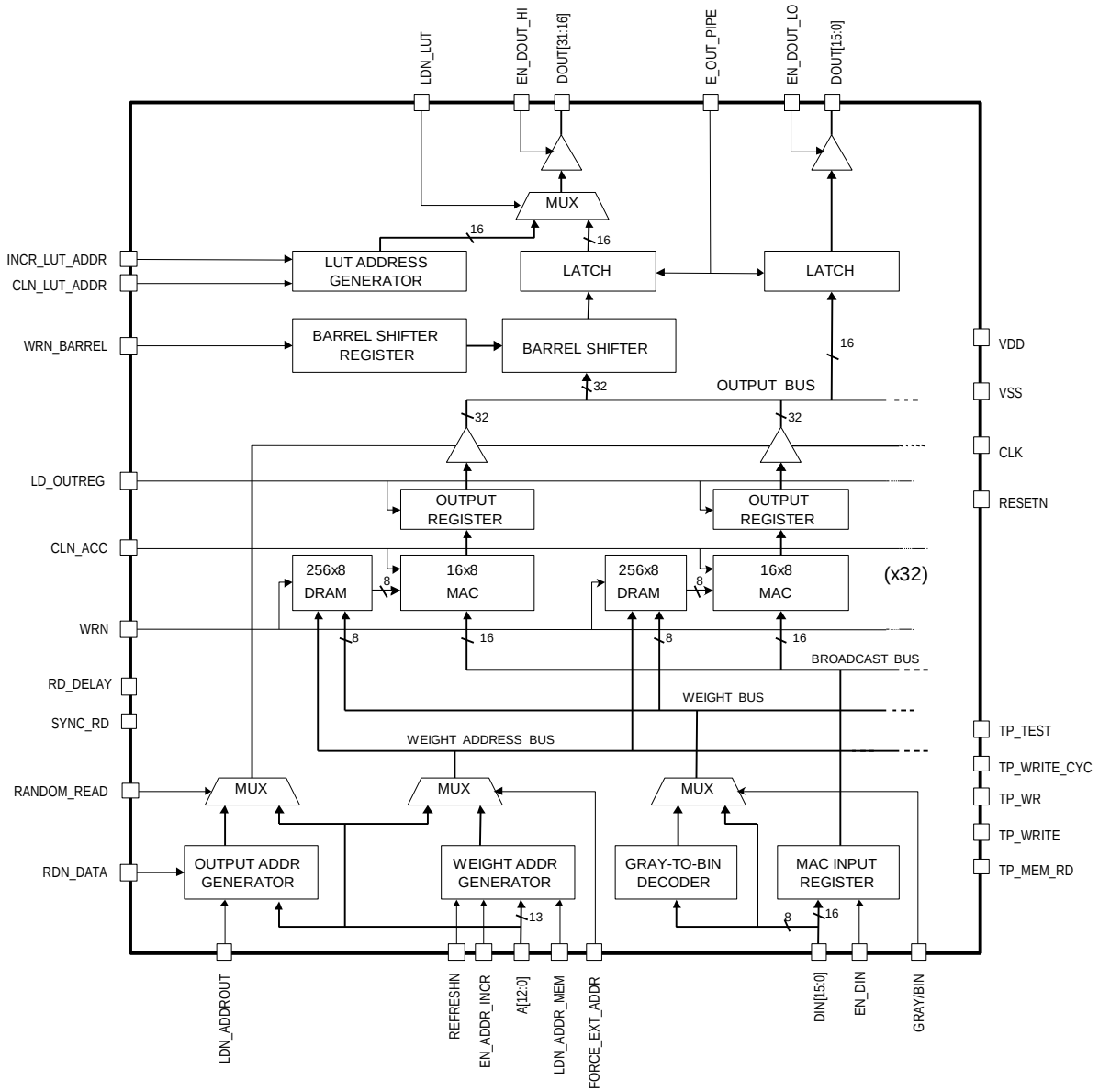


Fig. 1. Block diagram of 32-processor NC3002 chip

Pin descriptions

PIN NAME	PIN TYPE	PIN DESCRIPTION
CLK	CMOS IN	Master clock input. The clock input must run continuously to refresh the internal dynamic pipeline in the MACs and must have a nominal 50% duty cycle. The signals on all input control lines (with the exceptions below) and input buses must be synchronous with CLK as they are sampled by CLK edges. The pins which need not be synchronous with CLK are those related to the output section of the chip: DOUT[31:0], RDN_DATA, CLN_ADDROUT, RANDOM_RD, E_OUT_PIPE, WRN_BARREL, LDN_LUT, INCR_LUT_ADDR, CLN_LUT_ADDR, EN_DOUT_HI, EN_DOUT_LO.
DIN[15:0]	TTL IN PU	Data input bus. In the calculate cycle (EN_DIN active), this input bus carries the 16-bit 2's complement multiplicand data to all MACs via the MAC input register and the broadcast bus. In the weight write cycle, the bus carries on its 8 LSBs DIN[7:0] the 8-bit 2's complement weight to be stored (the 8 MSBs are ignored). Data is latched into the chip synchronously with the master clock when the WRN line is active. The data can be presented in binary or Gray format according to the status of the GRAY/BIN pin. In the barrel_write cycle (WRN_BARREL active), the bus carries on its 5 LSBs DIN[4:0] the value to be stored in the barrel shifter register.
DOUT[31:0]	TS OUT PU	32-bit 2's complement data output bus. This three-state output bus carries the results of the computations from the output register being addressed. The values of the output registers are transferred to the DOUT bus when RDN_DATA is brought "low". According to the status of the RANDOM_RD line, the output register can be addressed either sequentially or randomly. The output bus is sub-divided in two sections. The 16 MSBs DOUT[31:16] can be connected either to the output registers thorough the barrel shifter or to the LUT address generator according to the state of LDN_LUT. Their three-state drivers are activated by EN_DOUT_HI. The 16 LSBs DOUT[15:0] are connected directly to the output registers and their three-state drivers are activated by EN_DOUT_LO.
A[12:0]	TTL IN PU	13-input address bus for weight memory and output register. During the weight write cycle, this input bus carries the addresses of the weight memory locations. During the random read cycle (RANDOM_RD line active), the lines A[4:0] carry the address of the output registers to be read. See the memory map.
CLN_ACC	TTL IN PU	Clear accumulator. Active low. This input pin clears to zero all 32 accumulators in the MACs.
EN_DIN	TTL IN PU	Enable input data. Active low. When this input pin is sampled "low" on a rising CLK edge, the data on the DIN bus is latched in the MAC input register and a multiply-and-accumulate operation is triggered on all 32 MACs.
EN_ADDR_INCR	TTL IN PU	Enable address increment. Active low. When this input pin is sampled "low" on a rising CLK edge, the weight memory pointer is post-incremented (after the present calculation). The range of the addresses is 0 to 255. Automatic wraparound is provided after address 255. Activating EN_DIN and EN_ADDR_INCR together gives rise to a multiply-and-accumulate operation with post-increment of the weight memory pointer.

LDN_ADDR_MEM	TTL IN PU	Load self-incrementing memory pointer. When this input pin is sampled low on a rising CLK edge, the weight address generator is loadednwith the value contained in DINBUS[7:0].
FORCE_EXT_ADDR	TTL IN PU	Test pin: force external address. This input pin determines the addressing mode of the weight memory. FORCE_EXT_ADDR "low" selects sequential access mode in which the weight memory is sequentially accessed. The range of the pointer is 0 to 255 with automatic wraparound. FORCE_EXT_ADDR "high" forces the bits A[7:0] of the address bus to override the self-incrementing memory address pointer thus implementing random addressing of the weight memory.
LD_OUTREG	TTL IN PU	Load MAC result into output register. When this input pin is sampled "high" on a rising CLK edge, the results of all 32 MACS are transferred into their respective output registers. The signal can be activated only after the pipelines in the MACS are full.
WRN_BARREL	TTL IN PU	Write barrel shifter register. Active-low. A low-to-high transition on this input pin stores the data present on DIN[4:0] into the barrel shifter register. The valid range of values for the barrel shifter is 0 - 16.
RDN_DATA	TTL IN PU	Data read strobe. Active low. Bringing this input pin "low" transfers the results from the output register of one neuron to the DOUT bus. A low-to-high transition on this pin increments the output register pointer. Due to internal dynamic precharge circuitry, the pin cannot be held "low" indefinitely.
EN_DOUT_HI	TTL IN PU	Enable three-state drivers of 16 MSBs of output bus DOUT[31:16]. Active-low. When this input pin is "low", the bus is driven. When the pin is "high", the bus is tri-stated with a weak pullup.
EN_DOUT_LO	TTL IN PU	Enable three-state drivers of 16 LSBs of output bus DOUT[15:0]. Active-low. When this input pin is "low", the bus is driven. When the pin is "high", the bus is tri-stated with a weak pullup.
E_OUT_PIPE	TTL IN PU	Enable output pipeline. A "high" value on this input pin puts the latch on the DOUT[31:16] path into transparent mode. When the pin is "low", the data are latched. This pin is useful in implementing a one-stage pipeline on the output of the chip, which relaxes access time requirements on the external LUT.
LDN_ADDROUT	TTL IN PU	Load self-incrementing output address pointer. Active low. This input pin loads the output address generator with the value contained in DINBUS[4:0].
SYNC_RD	TTL IN PU	Test pin: synchronous read. Active low. A "low" value on this input pin forces the chip into the synchronous read mode whereby the RDN_DATA line is sampled on the rising edge of CLK.
RD_DELAY	TTL IN PU	Test pin: read delay. When this input pin is forced "high", DOUT is activated with one half clock cycle delay. Only active when SYNC-READ is "low".
RANDOM_RD	TTL IN PU	Test pin: force random read of results. This input pin determines the addressing mode of the output register to be read onto the DOUT bus. RANDOM_RD "low" selects the sequential read mode: one of the 32 output registers is addressed by the output register pointer which self-increments after the read cycle. The range of the pointer is 0 to 31 with automatic wraparound after address 31. RANDOM_RD "high" selects the random read mode in which the output register is selected by bits A[4:0] of the address bus.
WRN	TTL IN PU	Write weight memory. Active low. This input pin triggers a write cycle of the contents of the 8 LSBs of the data input bus DIN[7:0] into the weight memory location pointed to by the address bus A[12:0]. Gray-to-binary conversion is performed according to the status of the GRAY/BIN line.

GRAY/BIN	TTL IN PU	Gray/binary encoding of input data. This input pin is valid in the weight memory write phase. When the pin is "high", the data presented on the DIN bus are converted from Gray to binary coding before being stored in the weight memory. When the pin is "low", no conversion occurs.
LDN_LUT	TTL IN PU	Load LUT / normal operation. Active-low. When this pin is "high", the 16 MSB of the output bus DOUT[31:16] are connected to the output registers. When the pin is "low", DOUT[31:16] are connected to the LUT address generator. This mode is used to generate the addresses for the LUT RAMs during the RAM write operation.
INCR_LUT_ADDR	TTL IN PU	Increment LUT address. A low-to-high transition on this input pin increments the LUT address generator.
CLN_LUT_ADDR	TTL IN PU	Clear LUT address. Active-low. A "low" level on this input pin resets to zero the contents of the LUT pointer register.
REFRESHN	TTL IN PU	Refresh weight memory. Active-low. When this input pin is sampled "low" on a rising CLK edge, a 2-clock cycle refresh operation is initiated. The full memory is refreshed every 64 refresh operations.
RESETN	TTL IN PU	Reset chip. Active-low. This input pin resets the refresh and the barrel shifter registers. It should always be activated after power-up.
TP_TEST	TTL IN PU	Test pin: test on / normal operation. When this input pin is brought "high", the chip enters test mode in which the test lines TP_WRITE_CYC, TP_WR, TP_WRITE, and TP_MEM_RD are used to override some of the internally generated timing signals. When the pin is "low", normal chip operation occurs and the test lines are ignored.
TP_WRITE_CYC	TTL IN PU	Test pin: force write cycle. This input pin is used to force a write cycle.
TP_WR	TTL IN PU	Test pin: write. This input pin is used to override the internal write line.
TP_WRITE	TTL IN PU	Test pin: write. This input pin is used to override the setting of the multiplexers which select the data on the weight memory side of the MACs.
TP_MEM_RD	TTL IN PU	Test pin: read memory. This input pin is used to override the internal memory read line in the calculate cycle.
V_{DD}	POWER	Digital (5V) power supply. There are 8 V_{DD} pins, all of which must be connected.
V_{SS}	GROUND	Digital ground (0 V). There are 8 V_{SS} pins, all of which must be connected.

LEGEND: TS OUT = three-state output; TTL IN = TTL-level input; CMOS IN = CMOS-level input; PU = weak pullup to V_{DD}.

NC3002 pinout - 132-pin ceramic pin grid array (CPGA)

P	CLN_LUT_ADDR	A12	A11	A9	A6	A3	A2	CLN_ACC	RDN_DATA	EN_ADDR_INCR	LDN_ADDR_MEM	EN_DIN	REFRESHN	n.c.
N	VDD	n.c.	EN_DOUT_HI	A10	A7	A4	A1	RANDOM_RD	LD_OUTREG	LDN_ADDROUT	WRN	n.c.	n.c.	E_OUT_PIPE
M	DOUT2	LDN_LUT	n.c.	EN_DOUT_LO	A8	A5	A0	FORCE_EXT_ADDR	CLK	GRAY	n.c.	n.c.	WRN_BARREL	VSS
L	VDD	DOUT1	INCR_LUT_ADDR									RESETN	SYNC_RD	RD_DELAY
K	VSS	DOUT3	DOUT0									TP_TEST	n.c.	n.c.
J	DOUT6	DOUT5	DOUT4									VDD	VSS	DIN 15
H	DOUT7	VSS	VDD									DIN 12	DIN 13	DIN 14
G	DOUT10	DOUT9	DOUT8									DIN 11	DIN 10	DIN 9
F	DOUT11	VDD	VSS									DIN 6	DIN 7	DIN 8
E	DOUT12	DOUT13	DOUT14									DIN 2	DIN 5	VDD
D	DOUT15	VSS	n.c.									n.c.	DIN 3	VSS
C	n.c.	n.c	n.c.	n.c.	n.c.	DOUT19	VSS	VDD	DOUT28	DOUT31	TP_WRITE_CYC	n.c.	DIN 0	DIN 4
B	n.c.	n.c.	n.c.	n.c.	DOUT16	DOUT20	DOUT23	DOUT24	DOUT27	DOUT30	TP_WRITE	n.c.	n.c.	DIN 1
A	n.c.	n.c.	VDD	DOUT17	DOUT18	DOUT21	DOUT22	DOUT25	DOUT26	DOUT29	TP_WR	TP_MEM_RD	n.c.	n.c.
•	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**

^ Orientation dot

N.B. pins facing the viewer

NC3002 - Pin cross-reference list

DATAIN		DATAOUT		ADDRESS		CONTROL / TEST		VSS	VDD	N.C.
DIN0	C13	DOUT0	K3	A0	M7	CLK	M9	D2	F2	C3
DIN1	B14	DOUT1	L2	A1	N7	RDN_DATA	P9	F3	H3	B1
DIN2	E12	DOUT2	M1	A2	P7	LD_OUTREG	N9	H2	L1	C2
DIN3	D13	DOUT3	K2	A3	P6	CLN_ACC	P8	K1	N1	D3
DIN4	C14	DOUT4	J3	A4	N6	WRN	N11	M14	J12	C1
DIN5	E13	DOUT5	J2	A5	M6	TP_WRITE_CYC	C11	J13	E14	N2
DIN6	F12	DOUT6	J1	A6	P5	TP_WR	A11	D14	C8	M3
DIN7	F13	DOUT7	H1	A7	N5	TP_WRITE	B11	C7	A3	N12
DIN8	F14	DOUT8	G3	A8	M5	TP_MEM_RD	A12			M11
DIN9	G14	DOUT9	G2	A9	P4	TP_TEST	K12			P14
DIN10	G13	DOUT10	G1	A10	N4	LDN_ADDROUT	N10			N13
DIN11	G12	DOUT11	F1	A11	P3	RANDOM_RD	N8			M12
DIN12	H12	DOUT12	E1	A12	P2	LDN_ADDR_MEM	P11			D12
DIN13	H13	DOUT13	E2			FORCE_EXT_ADDR	M8			A14
DIN14	H14	DOUT14	E3			EN_ADDR_INCR	P10			B13
DIN15	J14	DOUT15	D1			EN_DIN	P12			C12
		DOUT16	B5			GRAY	M10			A13
		DOUT17	A4			REFRESHN	P13			B12
		DOUT18	A5			E_OUT_PIPE	N14			B4
		DOUT19	C6			WRN_BARREL	M13			C5
		DOUT20	B6			RESETN	L12			A2
		DOUT21	A6			LDN_LUT	M2			B3
		DOUT22	A7			INCR_LUT_ADDR	L3			C4
		DOUT23	B7			CLN_LUT_ADDR	P1			A1
		DOUT24	B8			EN_DOUT_HI	N3			B2
		DOUT25	A8			EN_DOUT_LO	M4			K13
		DOUT26	A9			SYNC_RD	L13			K14
		DOUT27	B9			RD_DELAY	L14			
		DOUT28	C9							
		DOUT29	A10							
		DOUT30	B10							
		DOUT31	C10							

Principle of operation

The NC3002 chip is a pipelined digital parallel processor for the implementation of Multi-Layer Perceptrons (MLP). It employs high-speed limited-precision integer arithmetic and allows good recognition performance in combination with a novel training algorithm. Internal dynamic RAM is provided for storage of the weights.

Training algorithm
The implementation of biological models in highly-concurrent VLSI systems requires training algorithms that admit low-precision weights, low-accuracy signal processing and only the evaluation phase of the network when an input pattern is given (the forward pass). The Reactive Tabu Search (RTS) learning algorithm proposed in [1] satisfies the above requirements. It is characterised by global memory-based optimisation (no "trapping" at local minima), no need for derivatives (feed-forward network evaluations only), tuneable number of bits per weight and input, robustness to low computational accuracy, and no critical sensitivity to initial conditions. This contrasts with derivative-based training algorithms such as backpropagation which tend to require high-precision computations [2], [3], [4]. The fixed-point, short-word digital implementation made possible by RTS leads to more economical VLSI architectures than its floating-point, long-word format counterparts in terms of silicon estate, power dissipation and speed.

Processor architecture
The digital data stream SIMD computational structure was used as the paradigm for the development of an architecture specially tailored for the RTS algorithm, based on the concepts of simplicity in the basic processor structures, high parallelism, operation on integer numbers, low number of bits in the representation of the inputs and of the weights, deep pipelining and balanced processing versus I/O throughputs [5]. The digital implementation is particularly attractive in the present case because the word widths of the operands are limited. Architectural simplicity is achieved by optimising the execution of multiply-and-accumulate on long data streams, the most compute-intensive operation for the forward pass of the Multi-Layer Perceptron. The calculation of the changes in the weights during the learning phase is delegated to other circuits such as standard microprocessors. The activation function is implemented off-chip in a RAM-based look-up table.

The bit-parallel architecture is used to optimise processing performance. An array of 32 parallel processors implement a single layer of an MLP. A broadcast scheme is used to feed the features to the neurons implemented with multipliers-accumulators (MACs), thus providing the desired full connectivity in the MLP network while requiring the lowest possible bandwidth at the input port of one I/O transfer per accumulation cycle. Efficient processing is achieved by a fully-parallel multiplier architecture, which provides optimum speed of one operation per clock cycle, while the time required for a full multiplication-accumulation is on the order of N_{inp} clock cycles, where N_{inp} is the number of input features. Access to the output of individually addressable neurons is provided.

The global bandwidth required to feed the weights to the MACs during training is on the order of 32 input/output operations per accumulation cycle, which potentially causes a serious I/O bottleneck in the case of external memory and highly parallel architectures. Internal storage of the weights, however, enables a high bandwidth between MAC and memory at the expense of increased silicon area.

The chip can operate, singly or in an array, as a co-processor in a host system. Word widths are optimised for learning with RTS: the 16-bit width of the broadcast bus is adequate to represent signals from transducers and intermediate results between layers, the memory word width of 8 bits is sufficient for many classification tasks [1] and allows economical chip layout, while the 32-word width of the output channel permits high accumulation capacity. The key design constraints of very compact memory and multiplier cells is satisfied by full-custom design of these elements, and by careful physical matching of the form factors of the memory and MAC blocks.

Multiplier-accumulator
The MACs operate on two's complement signed integers. The 16 x 8 parallel multiplier uses a Baugh-Wooley algorithm [6], which provides a more regular layout than the widely used Booth multiplier. The basic multiplier cell employs fully static 28-transistor CMOS adders. For high-speed operation, one level of pipelining is included in the multiplier array. Two additional levels of pipelining are included in the accumulator based on a ripple-carry adder. The larger delay with respect to carry-lookahead or carry-save configurations is counterbalanced by the pipelining. A 32-bit static storage register on the output of the MAC permits an output transfer from a neuron to be performed concurrently with a parallel input-multiply-accumulate operation on all processors for optimal implementation of MLP networks.

Dynamic memory

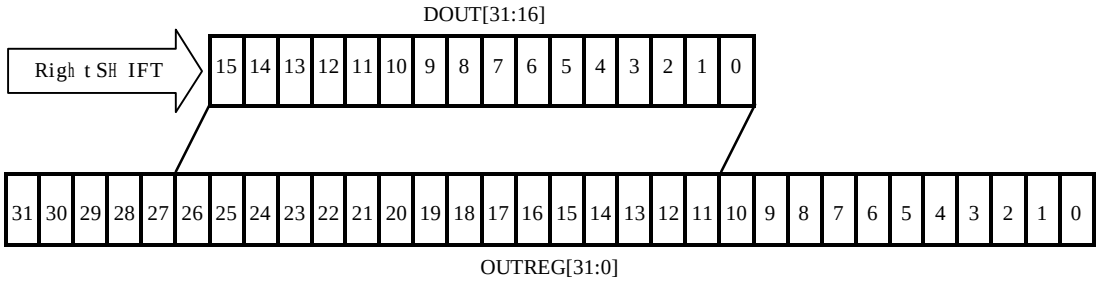
To optimise speed, the weight memory is partitioned into 2-Kbit blocks closely coupled to their related processors. Limited RAM area is achieved by the use of three-transistor n-MOS dynamic cells, with separate data input, precharged data output, read and write strobe lines. Decoder area is minimised by sharing the row decoders among a stacked array of memory blocks. Refreshing is provided by a write-after-read scheme requiring two clock cycles. A one-level pipeline register between RAM and MAC allows the data to be presented to the MAC during the entire clock cycle for optimum speed performance. The memory depth of 256 8-bit words allows neurons with up to 256 inputs to be implemented. Because of the sequential access to the weights, the chip can implement different MLP topologies with a high degree of flexibility: the memory bank can either be assigned to a single neuron or be partitioned among neurons on different layers.

Barrel shifter

A barrel shifter is provided to adapt the 32-bit bus width of the output of the NC3002 chip to the 16-input data bus in the implementation of multi-layer networks. The barrel shifter connects 16 out of 32 lines from the output registers of the processors to the 16 MSBs of the output bus DOUT[31:16] while the 16 LSBs from the output registers are always connected to DOUT[15:0]. The barrel shifter register contains the right-shift factor S for the barrel shifter.

Output register		bit 31.....bit 0

DOUT[31:16] bus	S = 0	bit 31.....bit 16
	S = i	bit (31-i).....bit (16-i)
	S = 16	bit 15.....bit 0



Registers and memory

Register / memory	R/W	Characteristics	Status after reset
MAC input register	W	16-bit 2' complement	zero
MAC accumulators	-	32-bit 2's complement. Not accessible by user. The 32-bit width of the output bus is sufficient to contain the result of 128 multiplication-accumulations with no overflow handling required	undefined
output registers	R	32-bit 2's complement	undefined
weight memory	W	32 ´ 256 ´ 8 bits (8 LSBs of DIN bus) 2' complement or Gray coding	undefined (When not refreshed, the weight memory drifts to FF Hex in a time dependent on temperature etc.)
output register pointer	W	Valid in sequential read mode. Loadable and self-incrementing. Range: 0 - 31 with automatic wraparound	undefined
weight memory pointer	W	Valid during the calculate cycle. Self-incrementing. Can be loaded and incremented by one. Range: 0 - 255 with automatic wraparound	undefined
barrel shifter register	W	Contains the right-shift factor for the barrel shifter. Range 0 - 16	zero
LUT address register	W	Can only be reset to zero. Self-incrementing. Range 0 - 65535 with automatic wraparound	undefined
refresh register	-	not accessible by user. Increments by one after each refresh cycle. Range 0 - 63 with automatic wraparound	zero

Memory map

Output registers

A[12:5]	A[4:0]	address A_{MAC} of output register and MAC
xxx	00 Hex	0
	...	...
	1F Hex	31

The table above is valid for random read mode. When sequential read mode is selected, addresses lines A[12:0] are ignored.

Weight memory

A[12:8]	A[7:0]	address A_{MAC} of memory bank and MAC	address within memory bank A_{MAC}
00 Hex	00 Hex	0	0
	...		...
	FF Hex		255
...	00 Hex	...	0
	...		...
	FF Hex		255
1F Hex	00 Hex	31	0
	...		...
	FF Hex		255

The table is valid for the weight memory write operation.

Instruction set

The instruction set of the processor consists of the following operations:

Instruction	Options	Length (CLK cycles)
calculate pattern for all processors: ACC := ACC + DIN * WEIGHT (WEIGHT_POINTER)	- sequential/random weight pointer - post-increment weight pointer	1
clear accumulators to zero		1
transfer accumulators to output registers		1
read output register	- sequential/random register pointer - post-increment outreg pointer - shift result right by s	(asynch.)
clear weight pointer to zero		1
clear output pointer to zero		1
write weight memory	- convert from Gray to binary	2
refresh weight memory		2
load barrel register		1
increment LUT pointer		1
clear LUT pointer to zero		1

The instructions are invoked by appropriately driving the 25 control lines of the processor as detailed in the AC characteristics chapter below.
Note that calculate and read result instructions can be executed simultaneously to achieve maximum speed in multi-layer networks. The weight memory can be read out through the MACs with a dummy calculate ´1 operation.

ELECTRICAL CHARACTERISTICS (expected data are given)

Absolute maximum ratings

Operating temperature	0 °C to 70 °C
Storage temperature	- 55 °C to 150 °C
Voltage on any pin to V_{SS}	- 0.5 to 7.0 V
Power dissipation	2 W

Operating conditions

Symbol	Parameter	Min	Nom	Max	Units
T_A	Ambient temperature under bias	0		70	°C
V_{DD}	Digital supply voltage	4.75	5.0	5.25	V
f_{CLK}	Master clock frequency	0.1		40	MHz

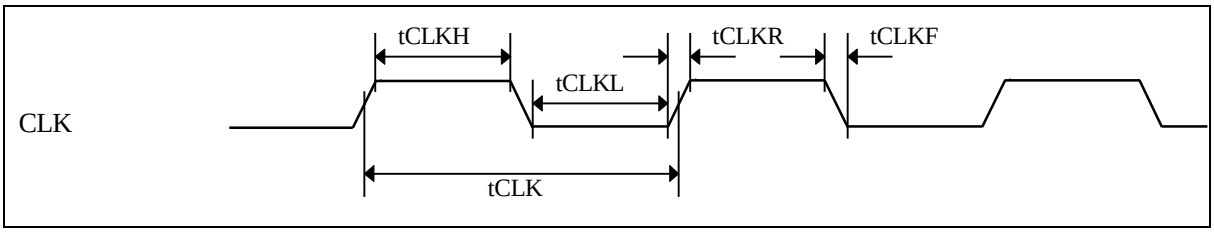
DC characteristics (over specified operating conditions)

Symbol	Parameter	Conditions	Min	Typ	Max	Units	Notes
V_{IH}	Input “high” voltage	TTL inputs	2.0		V_{DD} +0.5	V	
V_{IL}	Input “low” voltage	TTL inputs	-0.5		0.8	V	
V_{IH}	Input “high” voltage	CLK input	3.5		V_{DD} +0.5	V	
V_{IL}	Input “low” voltage	CLK input	-0.5		1.5	V	
I_{IH}	Input “high” current				±10	µA	
I_{IL}	Input “low” current	CLK input			±10	µA	
I_{IL}	Input “low” current	TTL inputs			±200	µA	
I_{OT}	Output three-state leakage				±10	µA	
V_{OH}	Output “high” voltage	IOH=-4mA	V_{DD} - 0.4			V	
V_{OL}	Output “low” voltage	I_{OL}=4mA			0.4	V	
FSLH	Output slew factor (low-to-high transition)				0.08	ns/pF	
FSLL	Output slew factor (high-to-low transition)				0.05	ns/pF	
R_{PU}	Weak pullup on TTL input and output pins			70		Kohm	
CPIN	pin capacitance (any pin to V_{SS})			5.0		pF	
t_{REF}	memory refresh time				2	ms	
I_{DD}	Supply current (active mode)	f_{CLK}=40MHz		250		mA	

AC parameters (over specified operating conditions)

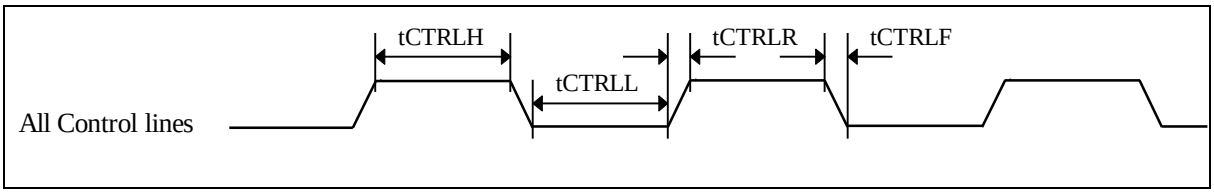
External clock drive

Symbol	Parameter	Min	Max	Units
$1/t_{CLK}$	Clock frequency	0.1	40	MHz
t_{CLK}	Clock period	25	10000	ns
t_{CLKH}	Clock high time	10		ns
tCLKL	Clock low time	10		ns
t_{CLKR}	Clock rise time		10	ns
tCLKF	Clock fall time		10	ns



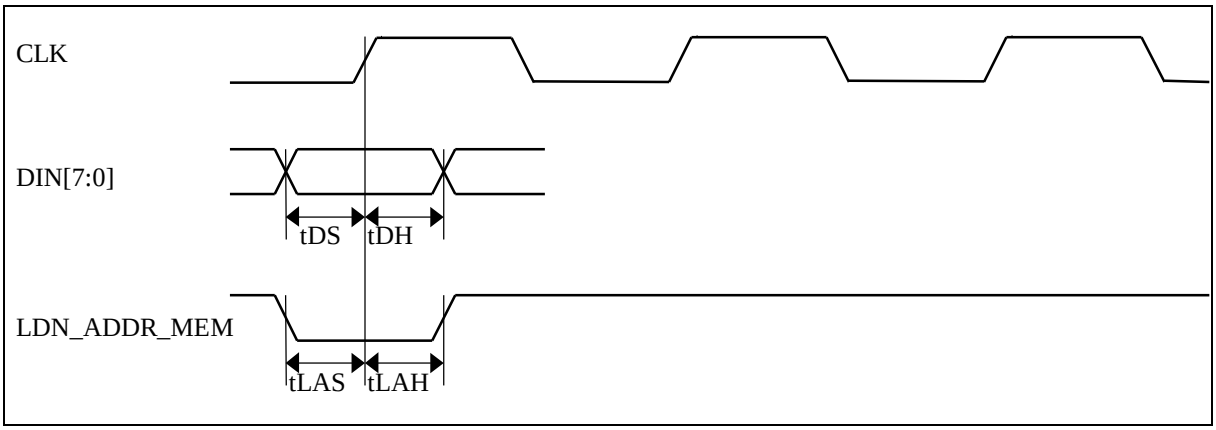
External control line drive (timings apply to all input control lines)

Symbol	Parameter	Min	Max	Units
t_{CTRLH}	Control high time	8		ns
tCTRLL	Control low time	8		ns
t_{CTRLR}	Control rise time		10	ns
tCTRLF	Control fall time		10	ns



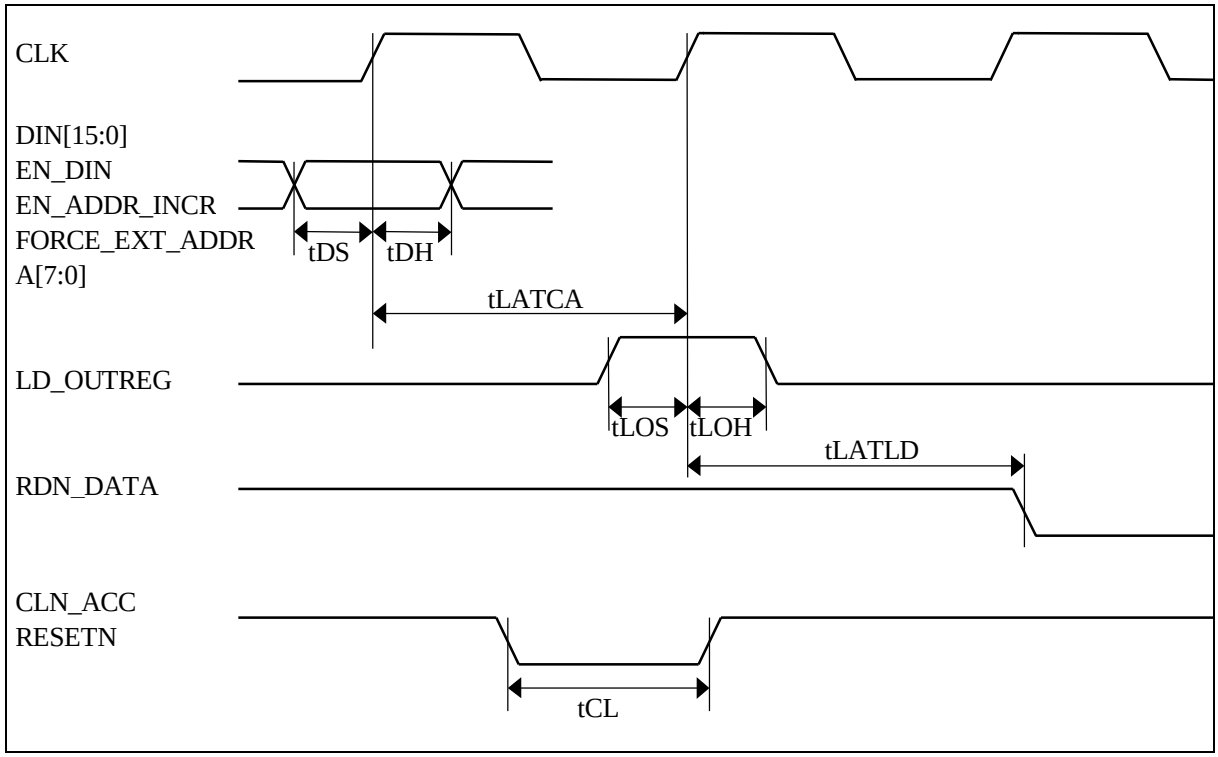
Load address memory cycle timings

Symbol	Parameter	Min	Max	Units
t_{DS}	Input data setup time	2		ns
t_{DH}	Input data hold time	2		ns
t_{LAS}	LDN_ADDR_MEM setup time	2		ns
t_{LAH}	LDN_ADDR_MEM hold time	2		ns



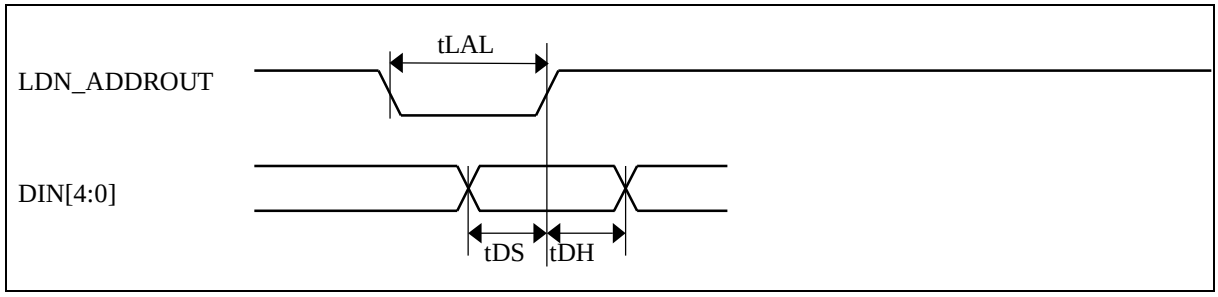
Calculate cycle timings

Symbol	Parameter	Min	Max	Units
t_{DS}	Input data setup time	2		ns
t_{DH}	Input data hold time	2		ns
t_{LATCA}	Data in to LOAD_OUTREG latency	1.0		t_{CLK}
t_{LATLD}	LOAD_OUTREG to output register full latency	1.5		t_{CLK}
tLOS	LD_OUTREG setup time	2		ns
tLOH	LD_OUTREG hold time	2		ns
t_{CL}	CLN_ACC, RESETN duration	2		t_{CLK}



Load address output register cycle timings

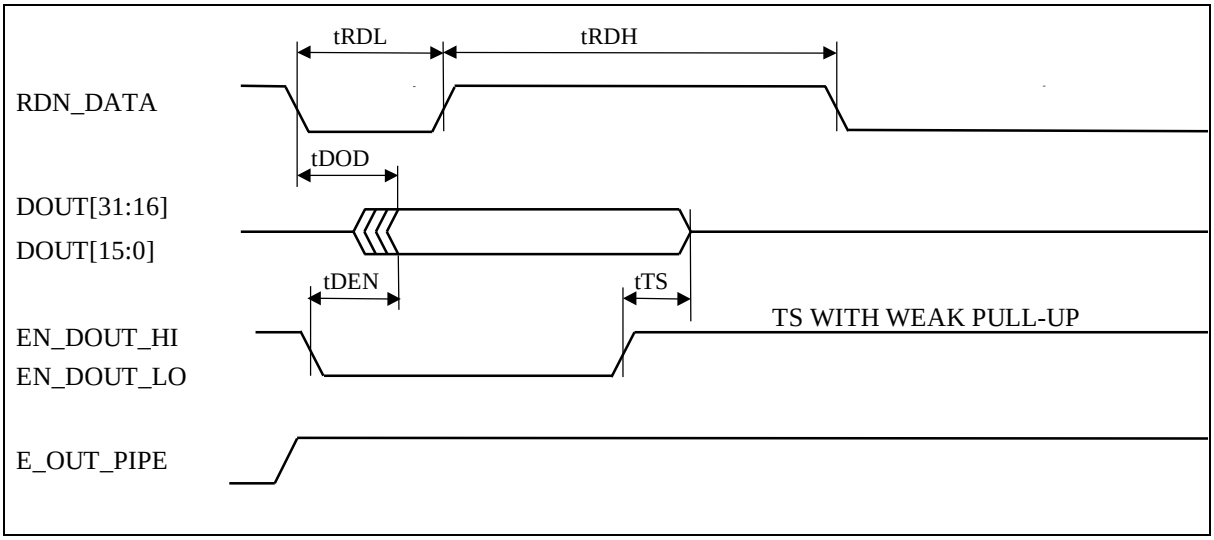
Symbol	Parameter	Min	Max	Units
t_{DS}	Input data setup time	2		ns
t_{DH}	Input data hold time	2		ns
tLAL	LDN_ADDROUT low time	8		ns



Note: the load address output register cycle is asynchronous with respect to the master clock.

Read cycle timings

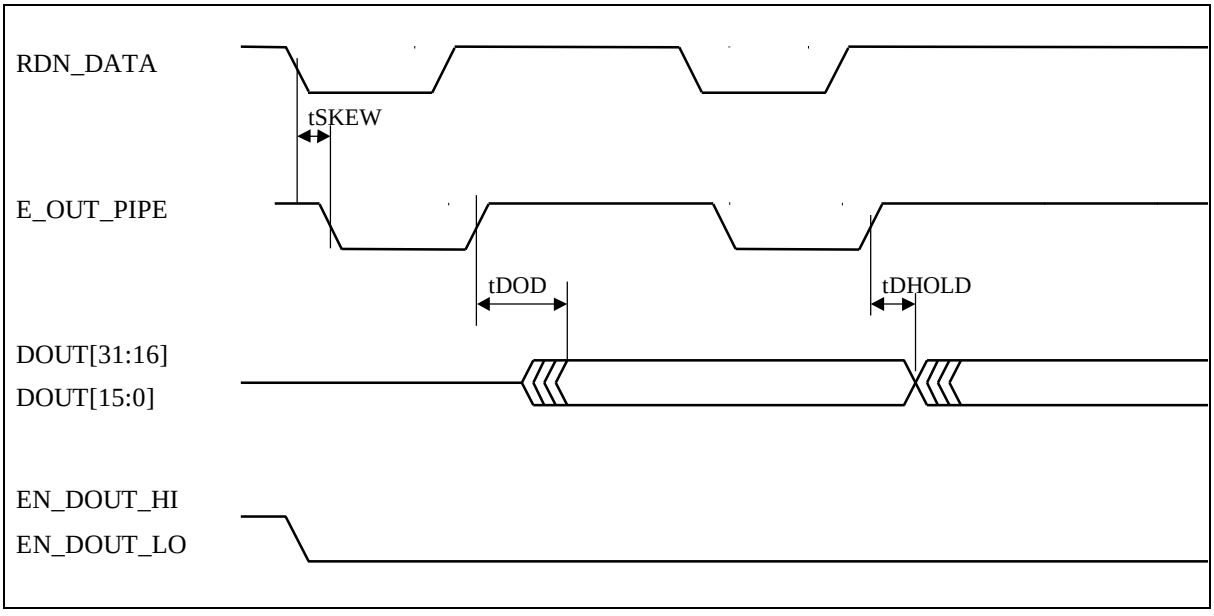
Symbol	Parameter	Min	Max	Units
t_{RDL}	Read low time	12	1000	ns
t_{RDH}	Read high time	2		ns
t_{DOD}	Output data access time		25	ns
t_{DEN}	Output enable to data valid		12	ns
t_{TS}	Output data valid to three-state		5	ns



Note: the read cycle is asynchronous with respect to the master clock.

Read cycle timings with E_OUT_PIPE

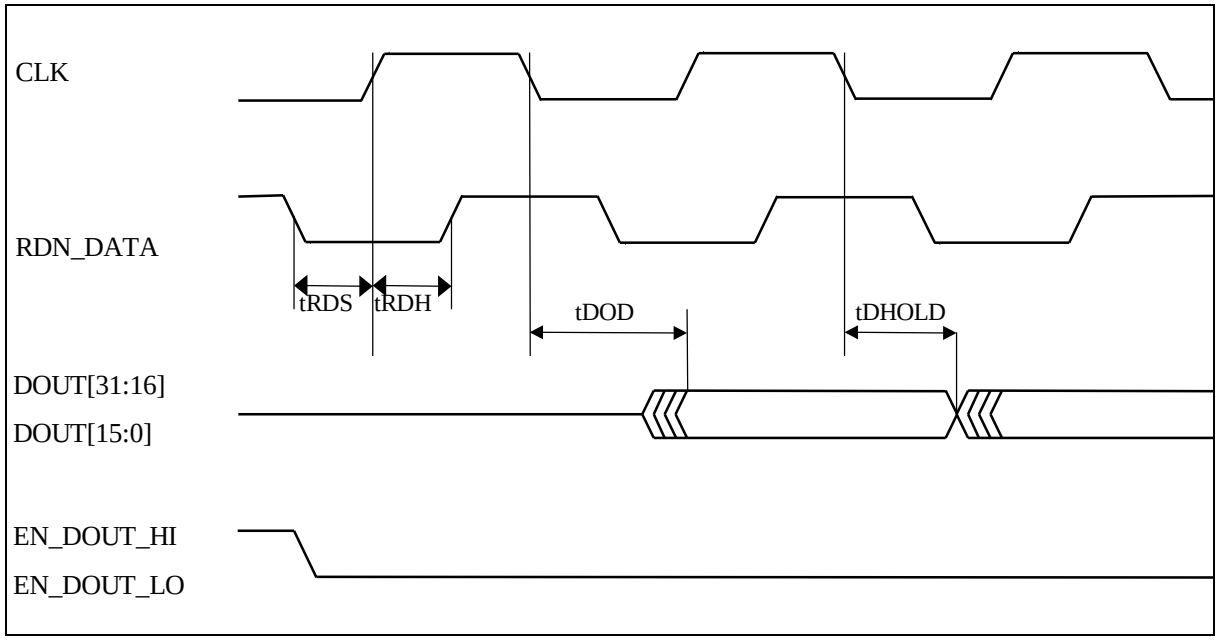
Symbol	Parameter	Min	Max	Units
t_{SKEW}	Skew between E_OUT_PIPE and RDN_DATA	-1	1	ns
t_{DOD}	Output data access time		20	ns
t_{DHOLD}	Data hold time	5	8	ns



Note: the read cycle with E_OUT_PIPE is asynchronous with respect to the master clock.

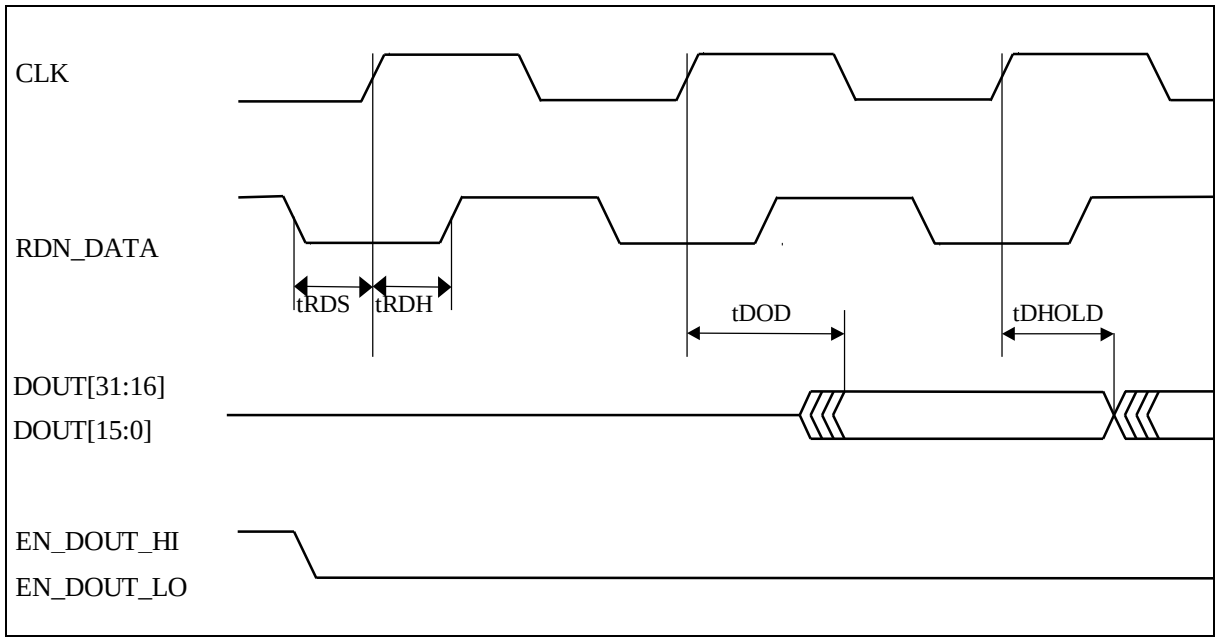
Sinchronouse read cycle (SYNC_RD low ; RD_DELAY high)

Symbol	Parameter	Min	Max	Units
t_{RDS}	Read setup time	2		ns
t_{RDH}	Read hold time	2		ns
t_{DOD}	Output data access time		20	ns
t_{DHOLD}	Data hold time	8	12	ns



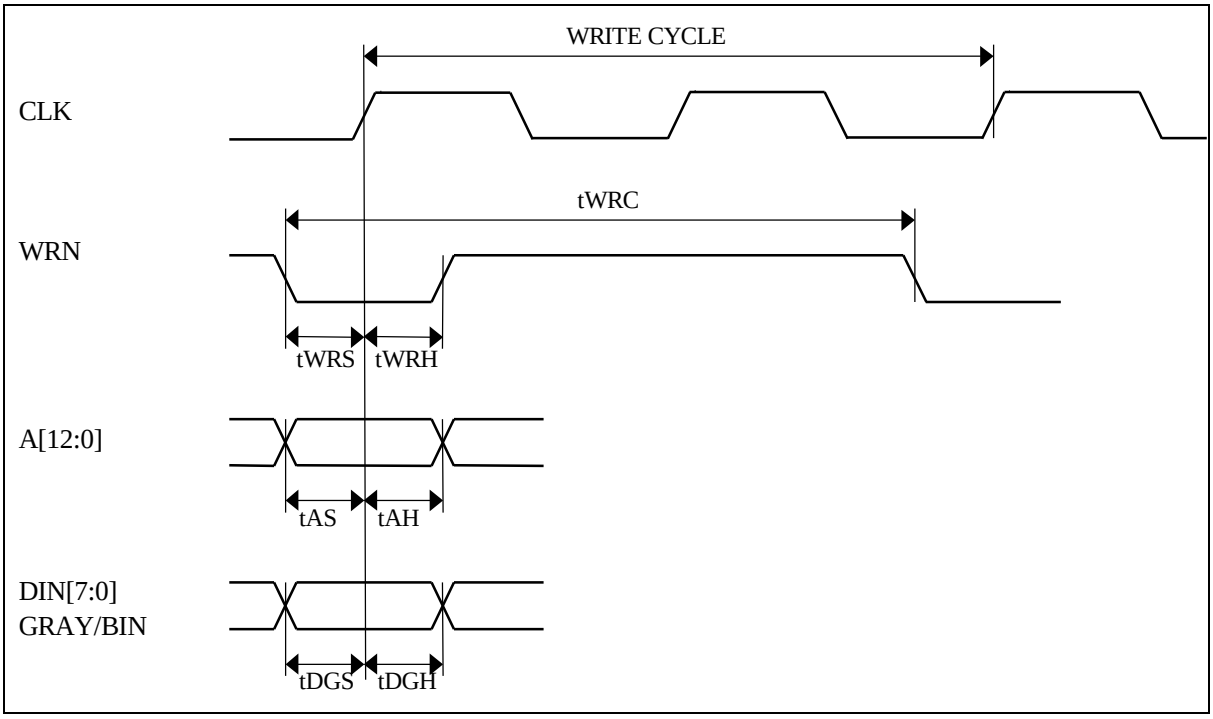
Sinchronouse read cycle with RD_DELAY (SYNC_RD low ; RD_DELAY low)

Symbol	Parameter	Min	Max	Units
t_{RDS}	Read setup time	2		ns
t_{RDH}	Read hold time	2		ns
t_{DOD}	Output data access time		20	ns
t_{DHOLD}	Data hold time	8	12	ns



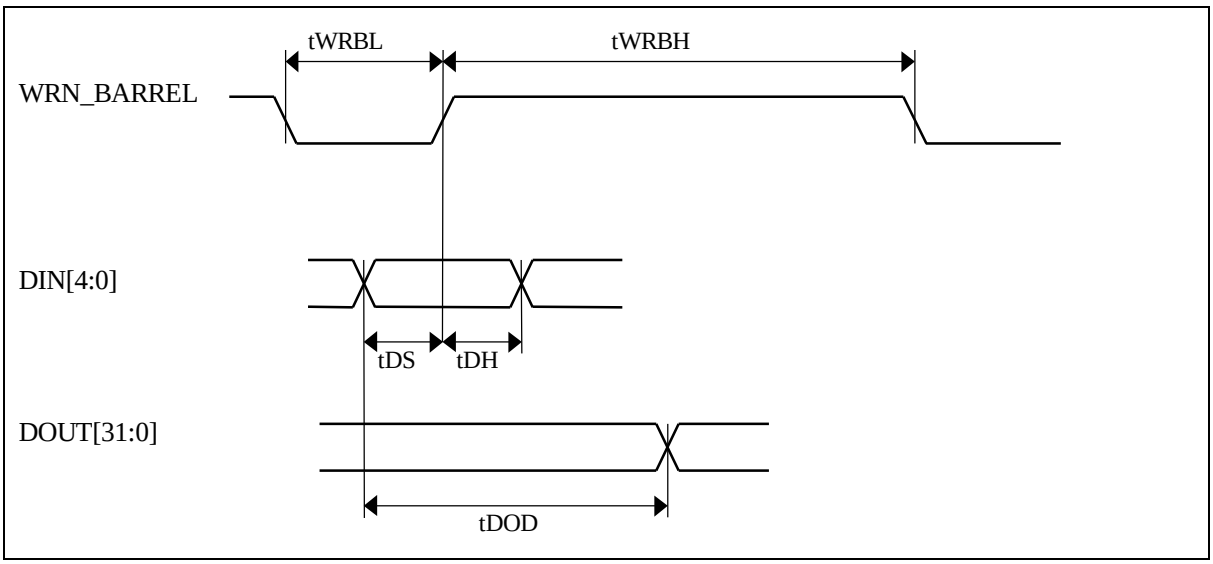
Weight write cycle timings

Symbol	Parameter	Min	Max	Units
t_{WRC}	Write strobe repetition rate	2		tCLK
t_{WRS}	Write setup time	2		ns
t_{WRH}	Write hold time	2		ns
t_{AS}	Address setup time	2		ns
t_{AH}	Address hold time	2		ns
t_{DGS}	Input data setup time	2		ns
t_{DGH}	Input data hold time	2		ns



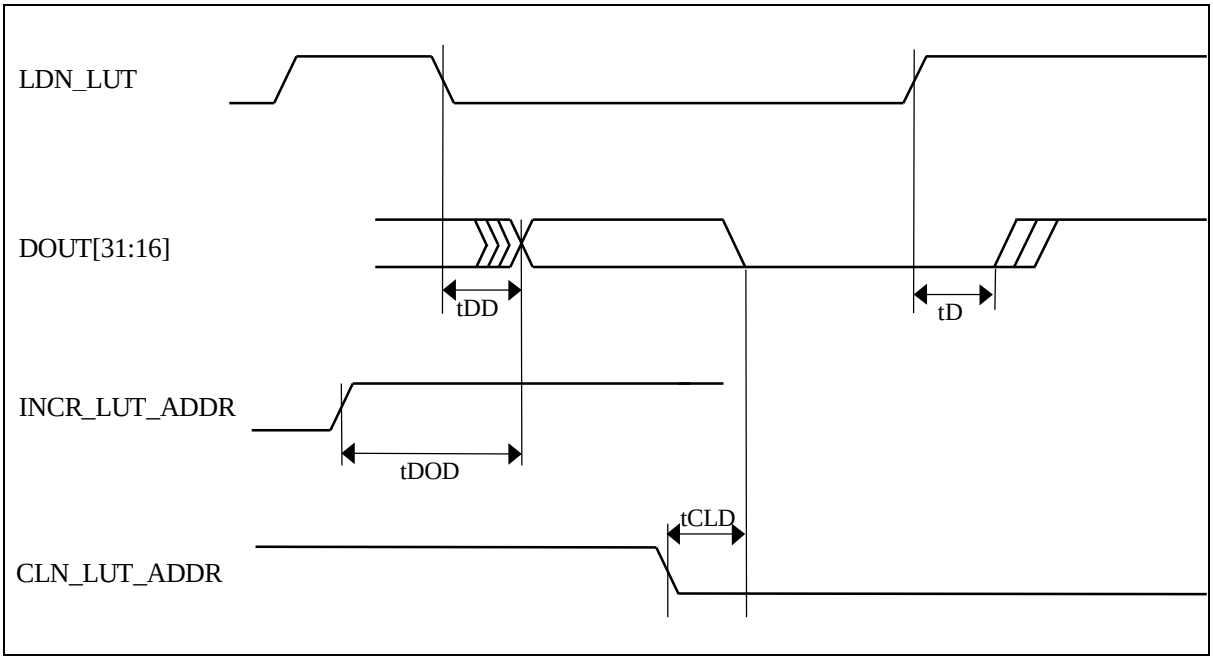
Barrel register write cycle timings

Symbol	Parameter	Min	Max	Units
t_{WRBL}	Barrel write strobe low time	8		ns
t_{WRBH}	Barrel write strobe high time	15		ns
t_{DS}	Data setup time	2		ns
t_{DH}	Data hold time	2		ns
t_{DOD}	data in to data out valid delay		25	ns



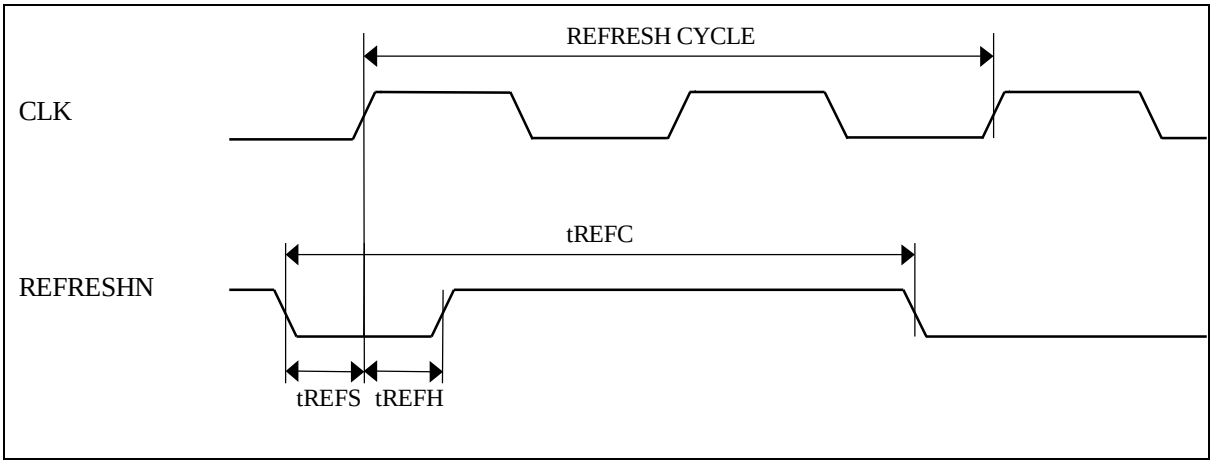
LUT write cycle timings

Symbol	Parameter	Min	Max	Units
t_{DD}	LDN_LUT transition to output data valid		12	ns
tD	Output data valid to LDN_LUT transition	2		
t_{DOD}	INCR_LUT_ADDR rising edge to output data valid		14	ns
t_{CLD}	CLN_LUT_ADDR low to output data valid		14	ns



Refresh cycle timings

Symbol	Parameter	Min	Max	Units
tREFC	Refresh strobe repetition rate	2	30	tCLK ; ms
t_{REFS}	REFRESHN setup time	2		ns
t_{REFH}	REFRESHN hold time	2		ns

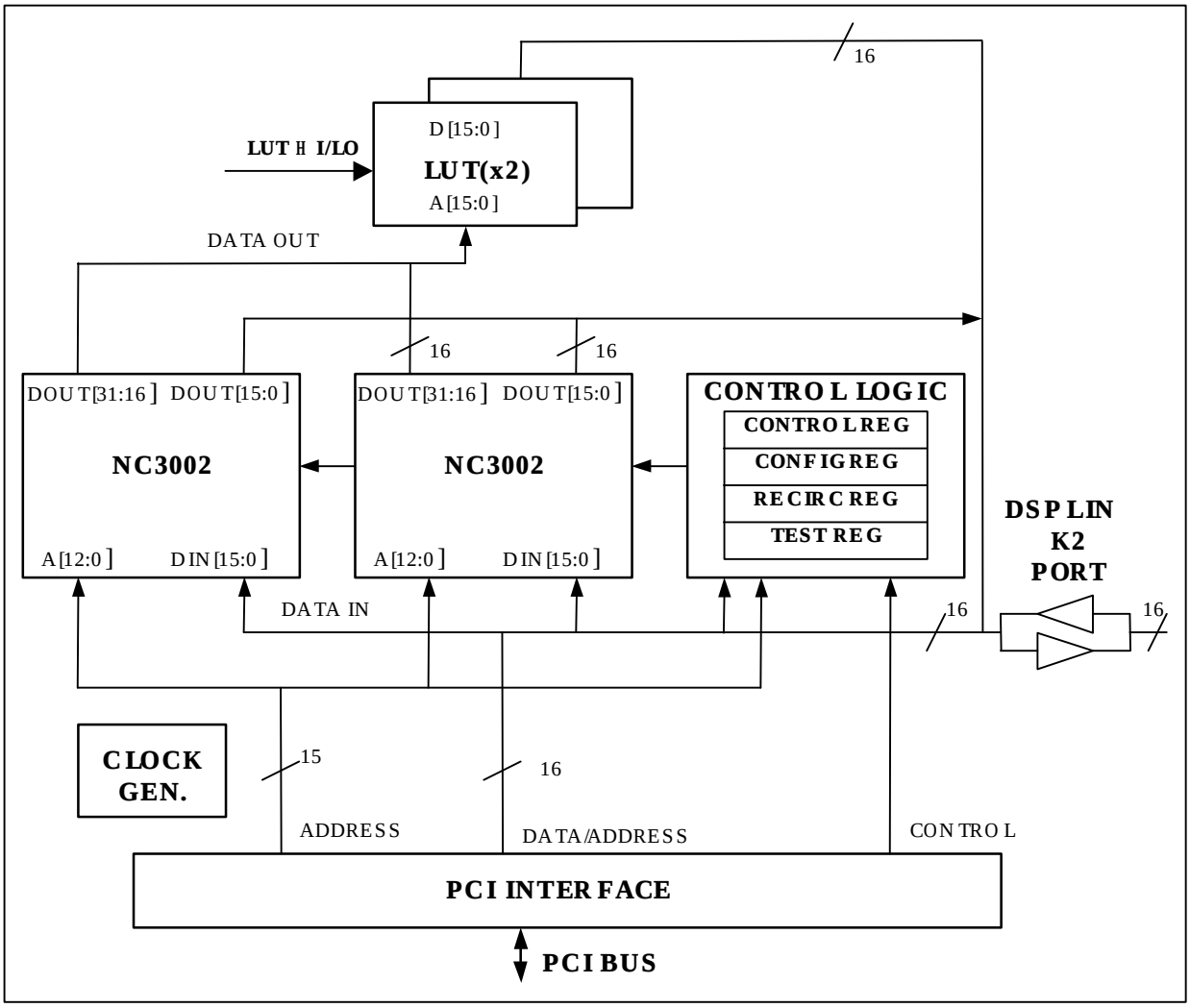


Note: One refresh cycle refreshes one column of RAM cells. The whole memory is refreshed every 64 refresh cycles.

The NC3002 system board

The NC3002 chip is a fast parallel number-crunching engine intended for operation in conjunction with a standard CPU. It is designed for maximum application flexibility and minimum redundancy of on-chip control circuits in single-chip or multiple-chip configurations. The present application requires two NC3002 chips, an external LUT, implemented with two 128K´8 static RAMs with 15-ns access time, a clock generator and a limited amount of external control and interface logic.
The proposed architecture allows a target multi-layer perceptron with a 64-128-64 topology to be evaluated in about 9 µs. Higher performance can be achieved by paralleling up to four chips per network level to implement neurons with up to 256 inputs.
The target bus, which is optimal in terms of throughput and applicability in different systems is the Peripheral Component Interconnect (PCI) bus. A DSPLINK2 port will be provided to allow future integration with third-party hardware such as standard DSP boards [7]. Xilinx FPGAs will be used to provide all the required control functions.
A preliminary block diagram of the board is shown in Fig. 2.

Fig. 2 Block diagram of NC3002 PC PCI board



References

[1] R. Battiti and G. Tecchiolli, "Training neural nets with the reactive tabu search", *IEEE Transactions on Neural Networks*, vol. 6, No. 5, 1995, pp. 1185-1200.

[2] T. Nordström and B. Svensson, "Using and designing massively parallel computers for artificial neural networks", *Journal of Parallel and Distributed Computing*, vol. 14, No. 3, 1992, pp. 260-285.

[3] U. Ramacher, "SYNAPSE - A neurocomputer that synthesizes neural algorithms on a parallel systolic engine", *Journal of Parallel and Distributed Computing*, vol. 14, No. 3, 1992, p. 306-318.

[4] D. Hammerstrom, "A VLSI architecture for high-performance, low-cost, on-chip learning", *Proc. Int. Joint Conf. on Neural Networks 90*, San Diego, Vol. 2, 1990, pp. 537-543.

[5] R. Battiti, P. Lee, A. Sartori and G. Tecchiolli, "TOTEM: A Digital Processor for Neural Networks and Reactive Tabu Search", MICRONEURO'94, Turin, September 1994.

[6] C. R. Baugh and B. A. Wooley, "A two's complement parallel array multiplication algorithm", *IEEE Transactions on Computers*, vol. C-22, No. 12, 1973, pp. 1045-1047.

[7] Loughborough Sound Images plc, "DSPLINK2 Generic Interface Specification", Loughborough, England, 1993.