Forgetful Logic Circuits for Pulse-Mode Neural Networks[1]

Richard B. Wells, Anindya Bhattacharya, Ben Sharon, Priyank Gupta, Sam Young, Sanjeev Giri, Terseer Ityavyar, and Dave Cox
MRC Institute
University of Idaho
Moscow, ID 83844-1024, USA
wellsrb@ieee.org

***Abstract* – We introduce a new class of pulse-mode circuit, called forgetful logic. Forgetful logic circuits can be used to implement more complex waveform signaling in pulse-mode artificial neural network circuits. The basic operation of forgetful logic is first explained. Its application is then illustrated by numerous examples.**

I. INTRODUCTION

There is very strong biological evidence that signal-dependent elastic modulation of synaptic weights and neuronal excitability plays a key role in information processing in the brain. Relatively rapid, short-term variations in synaptic efficacy is now believed to be responsible for a transient and reconfigurable 'functional column' organization in the visual cortex [1]-[2]. Dynamical recruitment of neurons into functional units by various selection processes have been theoretically studied by Edelman [3], Pearson et al. [4], Linsker [5], and Anderson and Van Essen [6]. Transient elastic modulation of synaptic efficacy is a central feature in the dynamic link architecture paradigm of neural computing [7]-[8]. One well-known example of the use of elastic modulation is provided by the vigilance parameter in ARTMAP networks[9]. It has long been accepted that firing rate encoding is one method by which information can be presented in a pulse-mode neural network, and it is likewise known that rate-dependent mechanisms exist in biological neural networks that filter information based on both pulse rate and the duration of a signaling tetanus [10]. Similarly, information may also be encoded through synchrony of firing patterns [10]-[12], and it is obvious that synchrony and rate/duration encoding can be combined in determining elastic modulations of synaptic efficacy. Many biological synapses, for instance, show selectivity to both pulse repetition rate and tetanus duration [10].

In this paper we present simple circuits for implementing these sorts of elastic modulation features in pulse-mode artificial neural networks and illustrate its use with some examples of rate- and tetanus-duration selectivity in networks comprised of previously reported mixed-signal VLSI pulse-mode neurons [13]-[16]. The circuit is selective for ranges of input firing rates and number of pulses received. If the firing rate is below the selection range the circuits do not activate, and within the designed frequency range the circuits require a minimum number of incoming pulses before activation. They are based on a logic circuit consisting of a pass element, one or two inverters, and a biasing element that sets its dynamic characteristics. We call circuits based on this design "forgetful logic" circuits (FLCs).

II. BASIC FORGETFUL LATCH CIRCUIT

The basic logic element is the non-inverting forgetful latch (FL) depicted in Figure 1. M2 – M5 comprise a biasing stick, which can be common to several FLCs in a VLSI implementation. M6 and M7 bias a storage node at the gates of inverter M8-M9. A single high-level input pulse applied to M1 charges the storage node and results in a HIGH level output from inverter M10-M11. When the input pulse goes LOW, M1 opens and current source M7 slowly discharges the gate capacitance at the storage node. The output pulse remains high for a brief time determined by the gate capacitance and the value of the drain current of M7. Thus, the input pulse is briefly 'stretched' at the output, typically for about 2.89 μsec for a 1 μsec input pulse in our designs, beyond the end of the input pulse. The FL then 'forgets' and the output goes LOW again.

Figure 2 illustrates the response of the FL to isolated input pulses and to a high-frequency tetanus. Note that for high-rate input pulse trains the FL maintains a constant HIGH output level. This behavior signals the on-going presence of

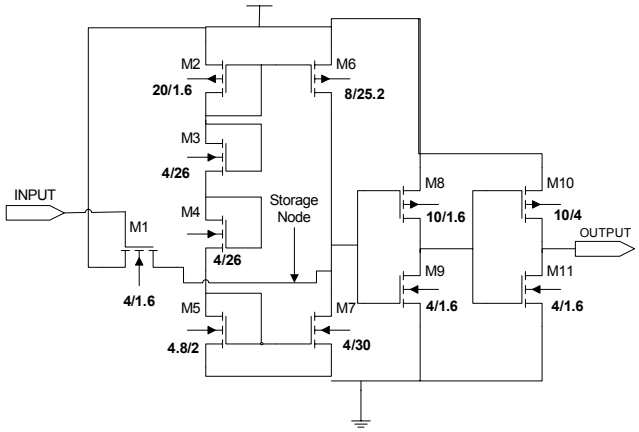

Figure 1: Basic Non-inverting Forgetful Latch Circuit. By eliminating M10 and M11 an inverting forgetful latch is obtained. Typical W/L ratios are shown in the figure.

[1] This work is supported by the NSF-Idaho EPSCoR Program and by the National Science Foundation under award number EPS-0132626.

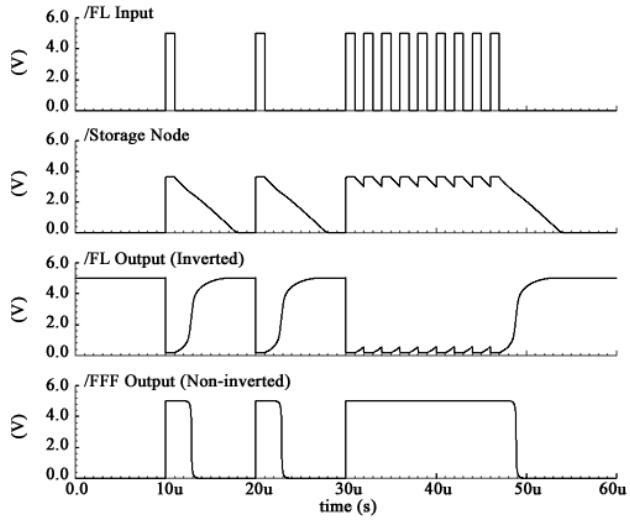

Figure 2: Pulse-rate dependence of the response of the basic forgetful latch.

signaling activity at the FL input and is a characteristic used in constructing the various other signal processing functions implemented using forgetful logic. The FL output pulse width for a single isolated input pulse is given by

$$\tau = \frac{C\left(V_{DD} - V_{SP} - V_t\right)}{I} + \tau_{in} \qquad (1)$$

where τ is the output pulse width, C is the total gate capacitance at the storage node, V_{DD} is the power supply voltage, V_{SP} is the switching threshold of M8-M9, V_t is the threshold of the n-channel device, I is the drain current of M7, and τ_{in} is the width of the input pulse. At input pulse rates above a threshold given by $1/(\tau + \tau_{in})$ the logic ceases to be "forgetful."

III. THE FORGETFUL FLIPFLOP

The cascade of two inverting forgetful latches, typically with different design values for τ, comprises a forgetful flipflop (FFF). The circuit is shown in Figure 3. M2-M5 comprise the bias stick. M1 and M6-M9 comprise the first FL, while M10-M14 comprise the second FL. Under quiescent conditions the output is LOW and the storage node at the drain of M12 is charged. τ at M12 is set to be larger than that of M7 such that the second FL cannot respond to single input pulses at the gate of M1. Rather, an input tetanus is required before the FFF output will respond.

The number of input pulses in the tetanus and the minimum input pulse rate required to evoke an output response from the FFF depends on the relative values of τ for the two stages. It is possible to achieve a wide range in the length of the tetanus required and in the delay-to-output assert and pulse width of the FFF output pulse. As a matter of terminology, we refer to FFF designs that respond relatively quickly and have output pulses that reset shortly

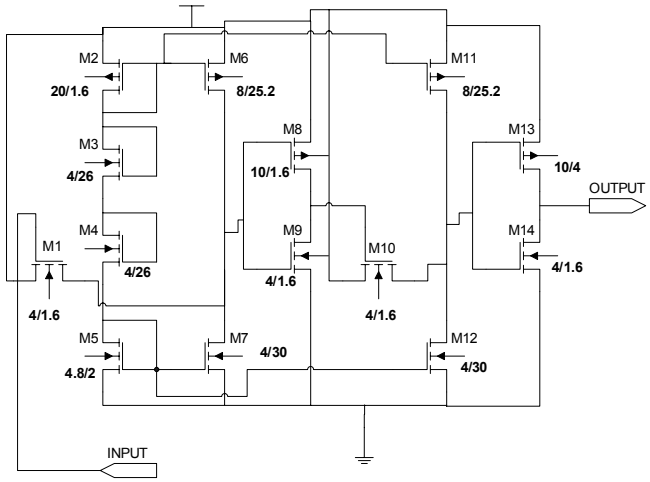

Figure 3: Basic forgetful flipflop (FFF) circuit with typical W/L values.

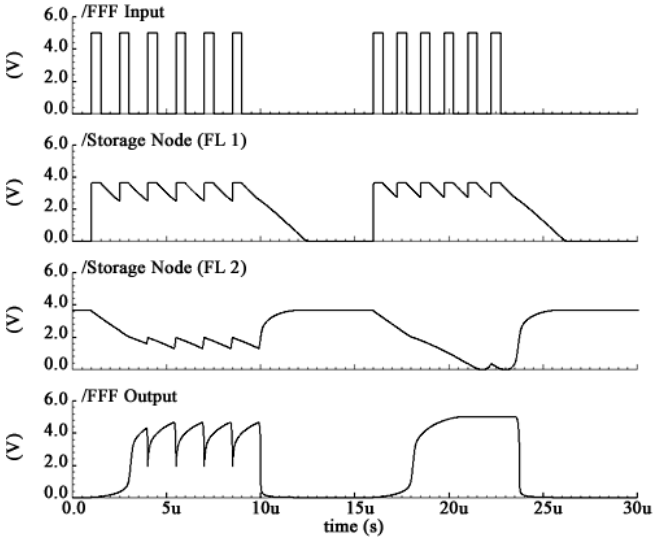

Figure 4: Input/output responses for 333 kpps and 400 kpps input pulse rates for a forgetful flipflop designed to produce a facilitation response. The leftmost output response is called a "twitch" response; the rightmost response is called a "complete" response. The FFF is designed to not respond to input pulse rates below 200 kpps.

after the end of the tetanus as a "facilitation" response; designs that require a longer tetanus or which hold the output pulse HIGH for a longer period of time after the end of the tetanus are called "augmentation" responses.

The basic action of the FFF is illustrated in Figure 4 for a design that implements a facilitation response. The FFF circuit which produces this response ignores input pulse trains that arrive at a pulse rate of less than 200 kpps and has a peak output response of only 1 volt for input pulse rates of 250 kpps when the input pulses are 1 μsec wide. The input pulse rates shown in this figure are 333 kpps and 400 kpps, respectively. Figure 5 graphs the time the FFF output remains above 1 volt as a function of input pulse rate for input pulses of 1 μsec width. (1 volt is the minimum synapse threshold for the artificial neurons used as application examples in sec. IV).

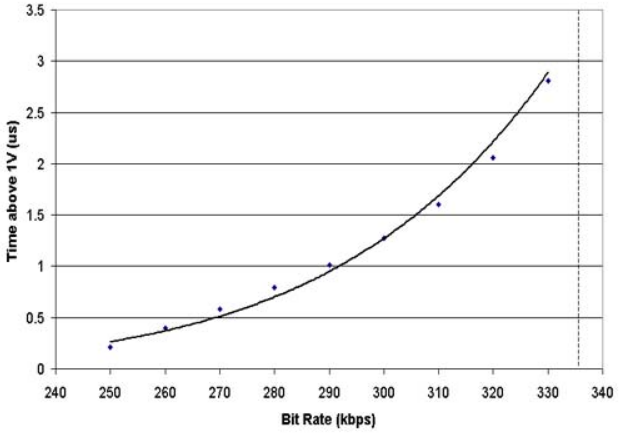

Figure 5: FFF output pulse width (> 1 volt) vs. input pulse rate for a continuous input tetanus of 1 μsec-wide input pulses for the circuit illustrated by figure 4. For input pulse rates above 360 kpps the FFF exhibits a complete (that is, dc-level) output response.

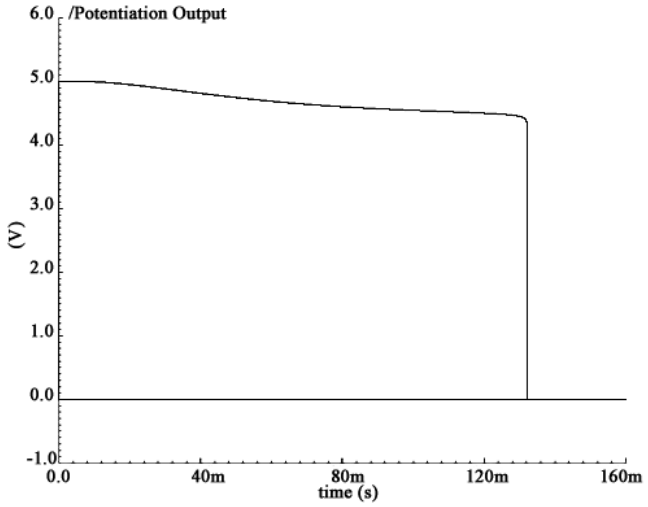

Figure 7: Response of a potentiation FFF.

A simple addition to the basic FFF produces the ability to maintain an active HIGH-level output signal for a sizable fraction of a second. The circuit is shown in Figure 6 below. M1 – M14 comprise a standard FFF. M15 – M18 implement a long-term memory FFF (LT-FFF). Under quiescent conditions, a LOW-level output turns on "keeper" transistor M16 and keeps the storage node at M17-M18 charged to V_{DD}. When a HIGH-level input is applied to M15 the storage node is discharged and the output goes HIGH. After the gate of M15 returns to a LOW value, leakage current through M16 slowly recharges the storage node. The storage time for the LT-FFF is determined by the switching threshold V_{SP} for M18. We call the response of this circuit a "potentiation" response.

Figure 7 illustrates a typical potentiation response. An input tetanus of 1 μsec pulses at 500 kpps was applied to the circuit of figure 6 for 18 μsec. The tetanus was then terminated. The LT-FFF output went high at approximately 10 μsec and maintained this high-level output state for 132 msec. In our work, we typically use LT-FFF designs for potentiation response in the range from about 20 msec up to the response illustrated in figure 7.

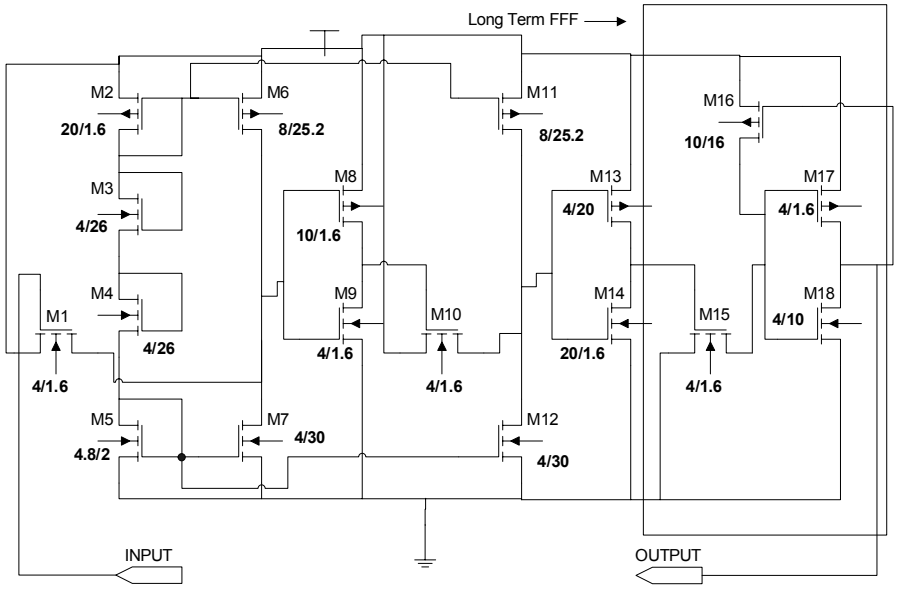

Figure 6: Potentiating forgetful flipflop. M1-M14 comprise two cascaded inverting forgetful latches. M15-M18 comprise a forgetful element with long-term memory. When the output signal is LOW, M16 maintains the gates of inverter M17-M18 at a HIGH level. When the gate of M15 receives a HIGH signal, M15 discharges the gates of M17 and M18 and the output goes HIGH. In this state, leakage current through M16 slowly recharges the gate voltages of M17-M18 after the input to M15 goes LOW. HIGH-level outputs from this FFF can be maintained for more than 100 msec.

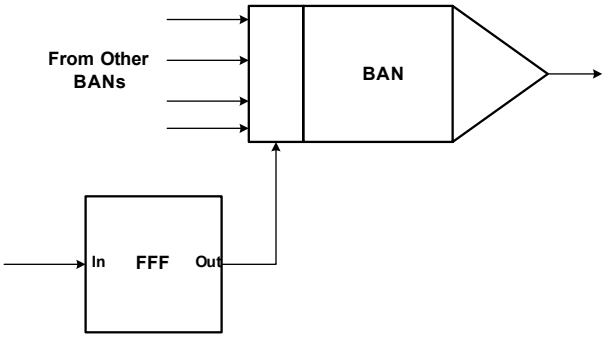

Figure 8: FFF used to increase sensitivity of a BAN neuron. A HIGH output from the FFF adds a DC bias to the input of the leaky integrator (LI) in the BAN. This additional bias decreases the number of synchronous synaptic inputs required to evoke an AP from the BAN.

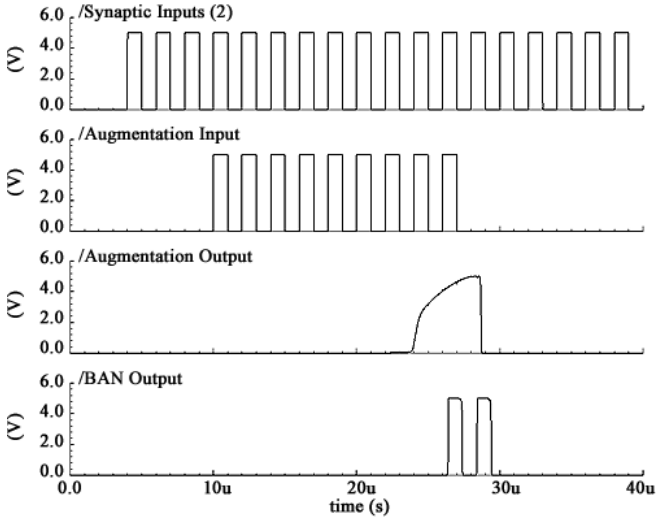

Figure 9: Waveforms for augmentation of firing sensitivity of a BAN for the circuit of figure 8. The top trace shows the two synchronous synaptic inputs to the BAN. The second trace shows the input to the FFF. The third trace is the FFF output. The bottom trace is the BAN output. By replacing the FFF with a LT-FFF, augmentation of firing sensitivity can be maintained for a longer time period after the FFF input ceases.

IV. APPLICATIONS OF FORGETFUL LOGIC

In this section we illustrate some of the applications of forgetful logic in pulse-mode neural networks. The neuron element used is a previously reported design known as a biomimic artificial neuron (BAN) [13], [16]. The first application is the use of a FFF to increase the sensitivity of a neuron to excitatory synaptic inputs. The circuit is illustrated in Figure 8. The BAN was designed such that a minimum of four synchronous synaptic inputs is required to fire an action potential (AP). A FFF output is applied to a synaptic input with the synaptic weight set such that: 1) the FFF cannot by itself stimulate an AP from the BAN, and 2) when the FFF input is HIGH two other synchronous synaptic inputs suffice to produce an AP. Figure 9 shows two synchronous BAN inputs, the input pulse train to the FFF, the FFF output, and the BAN output. In this illustration, the FFF was designed to respond after a 7-pulse tetanus at 500 kpps before

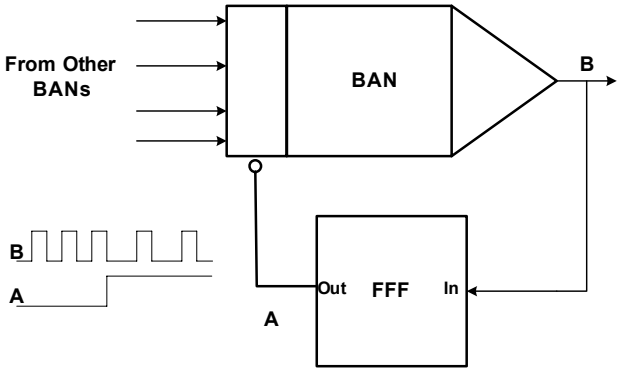

Figure 10: FFF used as feedback to an inhibitory synapse to produce accommodation in the BAN output firing rate. Conceptual waveforms are shown in the figure. A high-rate output at B eventually induces a high output from the FFF, which is fed back to an inhibitory synapse. This feedback lowers the firing rate at B. If firing rate B is slowed sufficiently, the FFF will eventually go inactive, thereby re-enabling the higher firing rate.

augmenting the sensitivity of the BAN. The augmentation input would remain applied so long as the FFF continued to receive the input tetanus. By replacing the FFF with a LT-FFF, augmentation of the BAN inputs can be maintained for a much longer period of time after the FFF input ceases. This technique can be used to enable specific cell groups of BAN neurons to implement re-configurable neurocomputing functional units. Similarly, by applying the FFF output to an inhibitory BAN input [16], the sensitivity of the BAN to synaptic inputs is reduced and, if the inhibitory weight of the BAN is large enough, can even be suppressed entirely (disabling of BAN cell assemblies). It should also be noted that because the FFF acts as a filter to low firing-rates, the augmentation action can be made frequency-selective. This has potential application for rate-dependent binding code specifications in pulse-mode neural networks.

A trivial variation on this scheme can be used to produce an accommodation response from a BAN neuron. This is illustrated in Figure 10. Assume that a firing response is induced in the BAN such that the firing rate at B is high enough to invoke a response in the FFF. When the FFF output goes HIGH, its signal is applied to an inhibitory synaptic input at the BAN, thereby reducing the BAN firing rate [16]. This mode of pulse coding is called an accommodation response by biologists and is frequently observed in numerous biological neurons. If the rate at B is reduced sufficiently (by selection of the inhibitory synaptic weight), the FFF, which acts as a high-pass rate filter, will eventually de-assert its output, thereby re-enabling the higher firing rate.

By combining positive feedback from a FL with negative feedback from a FFF, a BAN can be made to exhibit burst firing patterns. This is illustrated in Figure 11. Here the synaptic weight at A is set high enough that the FL signal invokes an AP from the BAN. Because the FL output pulse is wider than that of the BAN, the BAN re-triggers after its refractory period and re-fires [16].

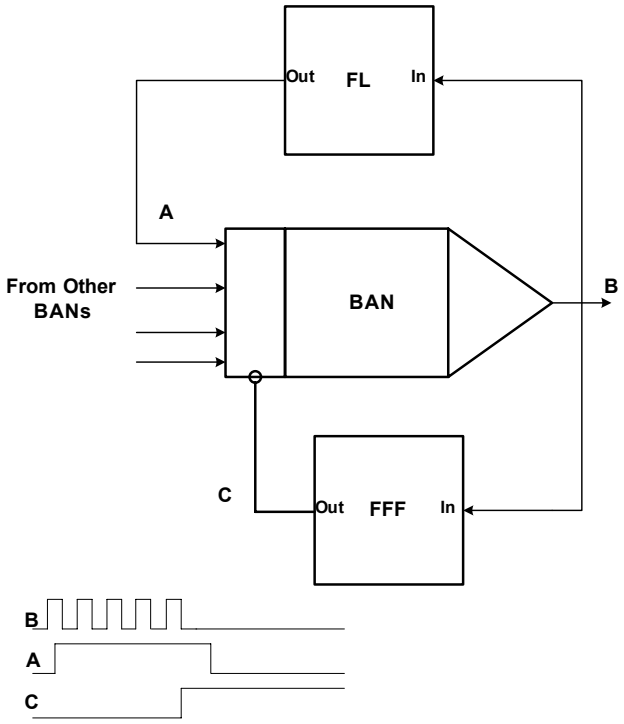

Figure 11: Forgetful logic used to turn a BAN integrate-and-fire cell into a bursting cell. Forgetful latch FL is applied to an excitatory synapse having a synaptic weight high enough to ensure re-firing of the BAN. After a burst length determined by the design of the FFF, FFF signal C is asserted at an inhibitory synapse. The weight of this synapse is set sufficiently high to ensure that C inhibits further firing. Firing at B can resume after the FFF output discharges and returns to the LOW state. Waveform schematics are shown in the figure.

After a number of pulses at B determined by the design of the FFF, the output at C is asserted at an inhibitory synapse. The synaptic weight of this synapse is set high enough to ensure that C completely inhibits further firing. After the FFF discharges, C is de-asserted and the BAN can again respond to its other synaptic inputs.

The BAN design responds to inhibitory synaptic inputs differently than excitatory synapses [16]. In particular, the response time for inhibitory BAN inputs is faster than that of the excitatory synapses because of the method used to discharge the BAN's leaky integrator (LI). This difference can be exploited to obtain the linking field behavior of an Eckhorn neural network [11] using integrate-and-fire BAN devices. The scheme is illustrated in Figure 12. An inverting FL is used as the feedback device from the second layer of the Eckhorn network. Its output is therefore normally HIGH and is applied to inhibitory synapses in the first (and elsewhere in the second, see [11]) layer. The synaptic weight of this input is set so that it is not high enough to prevent the BANs from firing in response to sufficient excitation of their synaptic inputs. When the second-layer BAN fires, the output of the inverting FL is de-asserted, which effectively raises the sensitivity of the BANs to their excitatory inputs. This mimics the linking field effect of a conventional Eckhorn neuron.

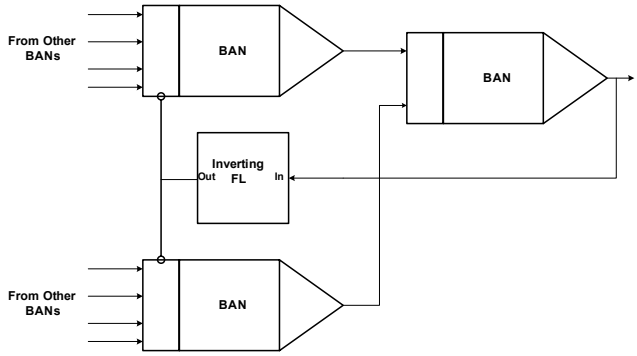

Figure 12: Mimicking the linking field effect of an Eckhorn neural network using an inverting FL and integrate-and-fire BAN neurons. It is to be noted that in most reported Eckhorn network designs, the linking field time constant is short compared to the feeding field time constant. This requirement is satisfied by the relatively short pulse duration of the forgetful latch, as shown in figure 2.

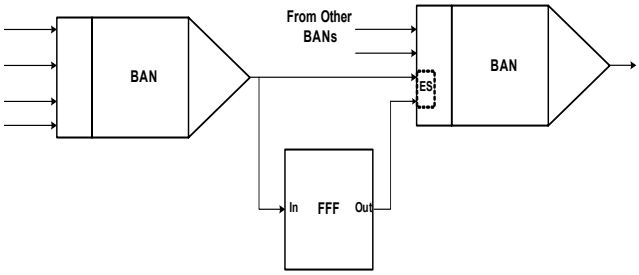

Figure 13: Short-term synaptic weight modulation using a FFF. The standard synaptic input of a BAN is modified by adding an additional control input to which the FFF is connected. When the FFF output goes HIGH, this input switches additional current to the synaptic input, thereby increasing the synaptic weight. The actual application of synaptic current to the BAN's LI is controlled by the direct connection to the source BAN. The FFF output goes high only in response to a tetanus at its input of sufficiently high frequency to invoke an output response from the FFF.

As a final application example, a FFF can be used to obtain short-term modulation of synaptic weights. The scheme is illustrated in Figure 13. To implement weight modulation, a trivial modification must be made to the standard BAN synaptic input discussed in [13], [16]. In the standard BAN design, a HIGH level input at a synapse switches current to an internal summing resistor at which the voltage input to the BAN's LI is obtained. To make an elastic synapse (ES), all that is required is that a second switch, which routes additional current through the main synaptic switch, be added. When the FFF output goes HIGH, this switch is activated, thereby adding to the synaptic current produced by the direct connection between BANs. The synaptic weight of a BAN is determined by the total current switched to the summing resistor. With aperiodic or low-rate input pulses, the FFF output remains LOW. However, the FFF will respond to a high-frequency tetanus by asserting its output as shown in the earlier figures.

V DISCUSSION

Test devices were fabricated using the 5V, 1.5 um MOSIS® process. SPICE simulations were carried out

using BSIM3v3 (level 8) modeling. Agreement between simulation predictions and measured device performance was excellent. Actual W/L ratios of the devices are application dependent, but typical values are shown in the circuit figures provided above. Total power dissipation excluding the common bias stick (5 uW) in typical applications averages less than 0.5 uW per FFF at 500 kHz operation. In the worst-case applications we have looked at (not illustrated in this paper) [17] total power dissipation can be as much as 5 uW for some FFFs in the network.

The circuit designs are centered at the process design center. We simulated process variation effects over the process-specified "4-corners" variation range. All devices operated over this range with worst case process-induced pulse width variations in FLs of +43% (slow-slow corner)/ -22 % (fast-slow corner).

In summary, this paper introduced forgetful logic and illustrated its application to pulse-mode neural networks. The well-known integrate-and-fire neuron has for many years been the most popular hardware implementation for artificial neurons owing to its simplicity. However, it has also been long recognized that the I&F neuron is somewhat limited in the types and methods of information encoding it is capable of achieving. Forgetful logic was developed in order to provide a richer repertoire of signal encoding capabilities and to provide a simple means of short-term synaptic weight modulation to support work in dynamic link architectures.

Because forgetful logic is still quite new, it is presently not clear what the full range of its potential as a signal processing element in pulse-mode neural network circuitry will eventually prove to be. We have carried out additional and as yet unpublished or not widely distributed work exploring what can be done using forgetful logic circuits, and our results so far have been encouraging [17]. From our preliminary work, it appears to be possible to implement artificial neurons entirely from forgetful logic. We have early, and so far successful, results with implementing all four major classes of Wilson neuron models [18] using only forgetful logic. The application examples provided in this paper illustrate the ease with which forgetful logic processing elements can augment the basic capabilities of the integrate-and-fire neuron. The hardware implementation of forgetful logic elements is barely more complex than standard logic elements, and so forgetful logic at this time appears to hold significant promise as a cost-effective means for implementing pulse-mode neural networks in VLSI.

VI REFERENCES

[1] Hubel, D.H. and Wiesel, T.N., "Functional architecture of macaque monkey visual cortex," *Proc. Roy. Soc.* Ser. B., vol. 198, pp. 1-59, 1977.

[2] White, E.L., *Cortical Circuits*, Boston, MA: Birkhauser, 1989.

[3] Edelman, G.M., "Group selection and phasic reentrant signaling: A theory of higher brain function," in *The Mindful Brain* (G.M. Edelman & V.B. Mountcastle, eds.) Cambridge, MA: MIT Press, 1978, pp. 51-100.

[4] Pearson, J.C., Finkel, L.H. and Edelman, G.M., "Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection," *J. Neurosci.*, vol. 7, pp. 4209-4223, 1987.

[5] Linsker, R., "From basic network principles to neural architecture: Emergence of orientation columns," *Proc. Nat. Acad. Sci. USA*, vol. 83, pp. 8779-8783, 1986.

[6] Anderson, C.H. and Van Essen, D.C., "Shifter circuits: A computational strategy for dynamic aspects of visual processing," *Proc. Nat. Acad. Sci. USA*, vol. 84, pp. 6297-6301, 1987.

[7] C.v.d. Malsburg, "Dynamic link architecture," in *Handbook of Brain Theory and Neural Networks* (M. Arbib, ed.) 2nd ed., Cambridge, MA: MIT Press, 2003, pp. 365-368.

[8] C.v.d. Malsburg, "The correlation theory of brain function," in *Models of Neural Networks II* (E. Domany, J.L. van Hemmen, K. Schulten, eds.), NY: Springer-Verlag, 1994, pp. 95-120.

[9] G.A. Carpenter and S. Grossberg, "ART 3: hierarchical search using chemical transmitters in self-organizing pattern recognition architectures," *Neural Networks*, vol. 3, pp. 129-152, 1990.

[10] R.C. Malenka and S.A. Siegelbaum, "Synaptic plasticity," in *Synapses* (W.M. Cowan, T.C. Südof, C.F. Stevens, eds.), Baltimore, MR, John Hopkins University Press, 2001, pp. 393-454.

[11] R. Eckhorn, H.J. Reitboeck, M. Arndt, & P. Dicke, "Feature linking via synchronization among distributed assemblies: Simulations from cat visual cortex," *Neural Computat.*, vol. 2, pp. 293-307, 1990.

[12] M. Abeles, *Corticonics*, Cambridge, UK: Cambridge University Press, 1991.

[13] B.C. Barnes, R.B. Wells and J.F. Frenzel, "PWM characteristics of a capacitor-free integrate-and-fire neuron," *IEE Electron. Letters*, vol. 39, no. 16, Aug. 2003, pp. 1191-1193.

[14] B. Sharon and R.B. Wells, "VLSI implementation of neuromime pulse generator for Eckhorn neurons," *IEE Electron. Letters*, vol. 40, no. 18, Sept., 2004, pp. 1143-1144.

[15] B. Sharon and R.B. Wells, "An Eckhorn neuron dendrite circuit VLSI implementation," to appear.

[16] B.C. Barnes and R.B. Wells, "A versatile pulse-mode biomimic artificial neuron using a capacitor-free integrate-and-fire technique," *Proc. 29th An. Conf. IEEE Indus. Electron. Soc.* (IECON'03), Nov. 2-6, 2003, pp. 2968-2972.

[17] A. Bhattacharya, *Forgetful Logic and Its Applications in the VLSI Implementation of Pulse-Coded Neural Networks*, M.S. Thesis, the University of Idaho, Moscow, ID, April, 2004.

[18] H.R. Wilson, "Simplified dynamics of human and mammalian neocortical neurons," *J. Theor. Biol.*, vol. 200, pp. 375-388, 1999.