MARGIN PROPAGATION AND FORWARD DECODING IN ANALOG VLSI

Shantanu Chakrabartty and Gert Cauwenberghs

ECE Department, Johns Hopkins University, Baltimore, MD 21218, USA
E-mail: {*shantanu,gert*}*@jhu.edu*

ABSTRACT

We propose margin propagation as an alternative to probability propagation in forward decoding. In contrast to sum-product probability propagation, margin propagation only incurs addition and subtraction in the computation and thus leads to reduced complexity of implementation. Simulations indicate that margin based forward decoding is more robust to input noise and parameter mismatch than sum-product probability decoding, and offers superior decoding performance. We also present an analog VLSI implementation of the margin propagation network independent of MOS device models, and provide experimental results from a prototype fabricated in a 0.5μ process.

1. INTRODUCTION

Analog probability propagation and decoding circuits have been shown to be superior to their digital counterparts in terms of speed and power consumption [1, 2, 3]. Such designs exploit inherent parallelism in architecuture, and trade off accuracy and simplicity of the interconnected modules. The circuit described in [1, 2] implements the sum-product algorithm by using the exponential characterstic of bipolar junction transistors or sub-threshold MOS transistors.

In this paper we present a novel decoding network based on a margin propagation algorithm. This work uses the normalization principle used in $Gini$-Support Vector Machines [4, 5] and extends it to decoding over arbitary graphs. The analog decoders that result from this approach inherit not only the benefits of the network in [1, 2] but has additional features:

1. The network operation does not depend on specific characteristics (square or exponential) of devices constituting the network. Thus, MOS transistors operating in the above threshold region can be used for applications of high-speed decoding. The device characteristic independent behavior also enables this principle to be directly employed in new emerging devices for instance in the field of nano-electronics.
2. The network can handle bi-directional current which results in a larger dynamic range of the input signal. In this paper however, the proposed MOS network is unidirectional for the sake of simplicity in illustration of proof of concept.
3. The normalization and hence the decoding principle is more robust to additive and scaling noise inherent in analog VLSI as compared to traditional probability normalization [6]. The primary reason for this is the robustness of propagation under margin operations as opposed to weighted linear summation through the graph. We will illustrate this robustness through examples in Section 3 and defer mathematical analysis to a later publication.

The paper is organized as follows. Section 2 describes the margin normalization procedure which is at the heart of the decoding procedure described in Section 3, and indicates its robustness through simulation. Section 4 presents analog VLSI circuits implementing margin propagation and Section 5 gives experimental results from a fabricated chip. Section 5 concludes with final remarks on the trade-offs of this procedure and its extensions.

2. MARGIN NORMALIZATION PROCEDURE

Given a vector of real values $g_i \in R$ representing confidence values over N classes indexed by $i \in \{1..N\}$ and a normalization factor γ, the margins P_i are obtained as

$$P_i = [g_i - Z]_+ \tag{1}$$

where $[x]_+ = max(x, 0)$ and the threshold Z is computed such that the following equation is satisfied:

$$\sum_i^N [g_i - Z]_+ = \gamma \tag{2}$$

It can be seen P_i/γ represents a valid probability measure over the classes i. Equation (2) can be solved using a reverse water-filling algorithm, popularly used in distortion rate theory [7] and involves sorting and search techniques [5]. In Section 4 we present an analog VLSI implementation that directly solves for Z by using (2) as its

This research is supported by a grant from The Catalyst Foundation, New York.

equilibrium criterion. When compared with conventional normalization defined on a vector of positive real values $g_i \in R^+$ and $P_i = g_i / \sum_i^N g_i$, the following properties are noteworthy:

- The distribution P_i obtained by (1) is offset insensitive whereas the distribution P_i obtained by conventional normalization is scale insensitive.
- For small values of the normalization constant γ the distribution P_i strongly favors the class with the higher confidence g_i and in extreme cases ($\gamma \to 0$) favors only the class with highest confidence while truncating the rest. The conventional normalization procedure on the other hand scales all the confidences equally. In analog VLSI systems truncation is natural and is governed by the noise floor.

3. MARGIN PROPAGATION ALGORITHM

The confidence measures P_i obtained through (1) can be used over graphs very similar to the probability propagation techniques described in [8]. In this section we apply it to a particular instance of a propagation algorithm which is popularly known in machine learning as the forward-recursion algorithm. We first describe the algorithm in its natural setting of its traditional normalization procedure. Given a fully connected graph consisting of S states, the probability of a state $i \in \{1, .., S\}$ at a time instant n is given by

$$P_i[n] = \sum_j P_j[n-1] P_{ji}[n] \tag{3}$$

where $P_{ji}[n]$ is the probability of transition from state j to i at time instant n. The method of obtaining these transition probabilities is application specific and can be generated by any probabilistic machine like neural networks [9], SVMs [10, 4], mixture of Gaussians, or table lookup.

In margin propagation algorithm the confidence estimates can be negative values $f_{ij}[n]$ which are combined through the following recursion:

$$\sum_i [f_{ij}[n] - Z_j[n]]_+ = P_j[n-1] \tag{4}$$

$$P_i[n] \to \sum_j [f_{ij}[n] - Z_j[n]]_+ \tag{5}$$

The first step (4) is margin normalization and the second step (5) is the accumulation step over the inter-connected graph. Like conventional recursion (3), the sum of measures P_i in (5) is conserved during propagation because $\sum_i P_i[n] = \sum_j P_j[n-1], \forall n$.

A simple experiment reveals the robustness of (5) over the (3). Figure 1 representd a finite-state machine used to recognize two sequences as shown in figure 2 where states

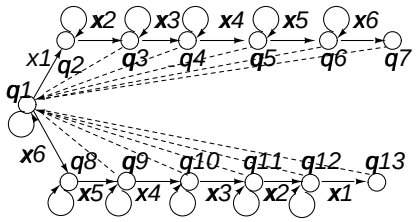

Fig. 1. *A finite state machine that recognizes sequences of pulses in a specific order.*

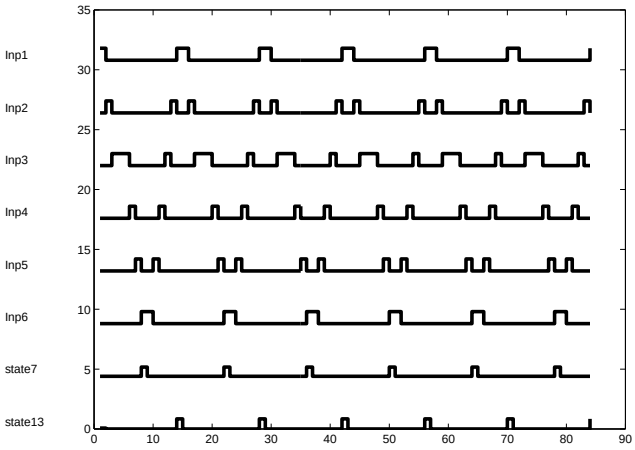

Fig. 2. *Measures $P_i[n]$ of states q_7 and q_{13} using both margin and probability propagation, with no noise present in the input signal and with infinite precision in the parameters.*

q_7 and q_{13} are triggered on the occurance of a top-down and a bottom-up sequence. Figure 2 also shows the response of both the decoding schemes in presence of clean inputs $\{x_i\}$ and precise model parameters. When relatively small amount of scaling and additive noise is added to the input with a small parameter mismatch the normalization schemes behave quite differently as shown in figures 3 and 4. The comparison reveals the robustness of margin decoding for analog VLSI.

4. ANALOG VLSI IMPLEMENTATION

A possible analog VLSI implementation of the margin propagating network is given in Figure 5. The network consists of an array of basic building blocks A_{ij} that compute $[I_{ij} - I_z]_+$, where I_{ij} are confidence values encoded as currents and I_z is the threshold value. F_j is the feedback circuit which determines the equilibrium condition in (2) with $\gamma = I_j$. A small signal analysis about the equilibrium reveals the stability and favorable speed of convergence to the equilibrium point.

Let g_{mk} be the transconductances of transistors M1-M4

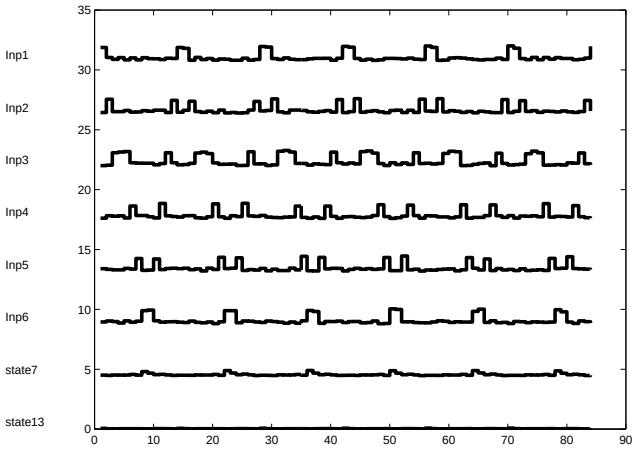

Fig. 3. *Measures $P_i[n]$ of states q_7 and q_{13} using probability propagation (3), in the presence of input noise and parameter inaccuracies.*

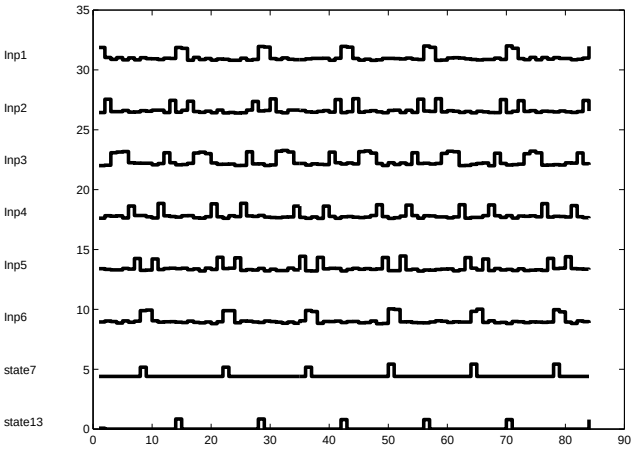

Fig. 4. *Measures $P_i[n]$ of states q_7 and q_{13} using margin propagation (5), in the presence of input noise and parameter inaccuracies.*

of $k = (ij)^{th}$ analog stage A_{ij} and the transconductances of transistors M7-M8 be g_{mp} and of M5 be g_{mn}. The sensitivity of the feedback signal V_z with respect to the an incremental change in input current I_{ij} is given by

$$\delta V_z/\delta I_{ij} \approx 1/(g_{mn}(1 + g_{mn}/g_{mp}) + \sum_{k\in S} g_{mk}) \qquad (6)$$

where S denotes the set of network elements whose $I_k - I_z > 0$. Note that V_z has to drive N input capacitances of M1 of the analog stages A_{ij}, where N is the total number of elements in a row of the normalization network. The following can be infered from equation (6):

- The sensitivity decreases as the number of active network elements $|S|$ decreases. This shows that if the network is pre-initialized to the condition $I_z = 0$ then it quickly converges to the neighborhood of the stable condition.

- The input capacitance driven by V_z remains fixed, and therefore the pole introduced by this capacitance shifts deeper in the right half plane as time progresses. This directly gives an indication of the stability of the network.

- As N becomes large then $\sum_{k\in S} g_{mk} \rightarrow |S|E[g_{mk}]$ where $|S|$ is the size of the set S and $E[g_{mk}]$ is the expected value of transconductance of the row elements. Relative random mismatch between transistors of the network element is thus compensated for large N

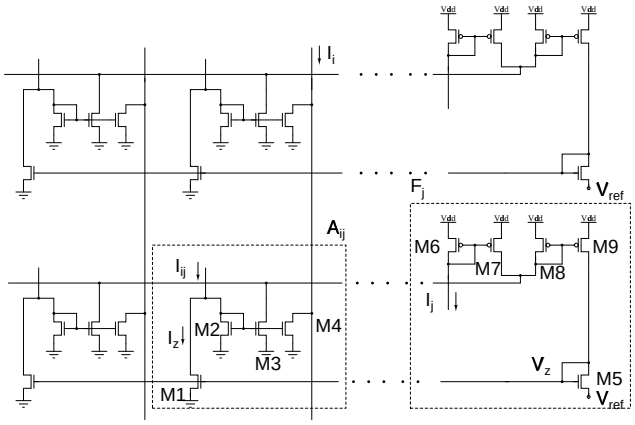

Fig. 5. *Analog network implementing margin normalization and propagation.*

5. RESULTS

A 24×24 margin normalization network was implemented as a part of a recognition system in a standard 0.5μ CMOS process. The transistors for this specific application were biased in weak inversion region to ensure low power dissipation, but weak inversion is not essential for correct operation. Figures 6 and 7 show normalized margins corresponding to the currents P1-P4 for four different row elements with respect to a reference current P1. The plots in figure 6 were obtained for a larger value of normalization factor γ as opposed to figure 7 which was obtained for a nominal value.

The following can be infered about the behavior of the normalization circuit :

- The higher the value of γ, the lower the sensitivity of

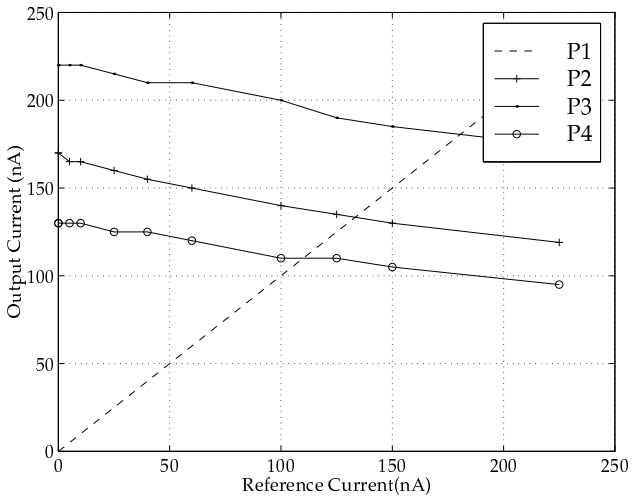

Fig. 6. *Output current values of different elements of the normalization network corresponding to a large value of the normalization constant γ.*

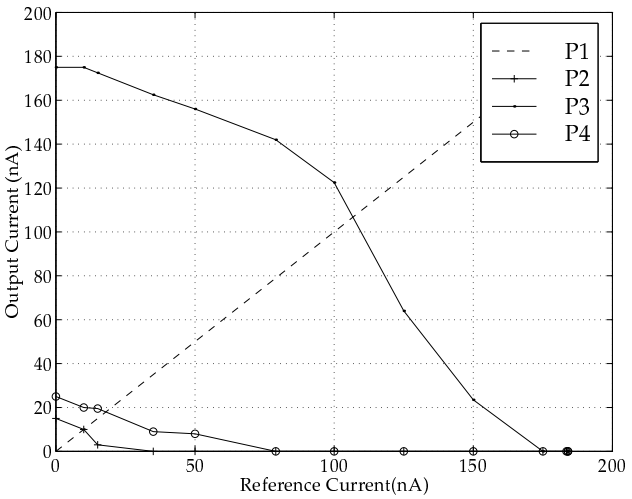

Fig. 7. *Output current values of different elements of the normalization network corresponding to a nominal value of the normalization constant γ.*

the output currents and vice-versa.

- Figure 7 shows that the output current sensivity increases when two probabilities/current values are close to each other and the behavior can be controlled by γ. Also the small currents $P2$ and $P4$ get truncated unlike in regular normalization where the model assumes that the current can be represented with infinite precision.

6. CONCLUSIONS

In this paper we presented a novel method of normalization, propagation and forward decoding based on relative margins between different confidence scores, as opposed to forward recursion of posterior propabilities. Using this scheme we proposed an analog network that can be applied to the design of decoders very similar to the approach presented in [1, 2]. Margin based decoders have an advantage over conventional probability propagation because it accounts for finite precision and noise in parameters and input signals, inherent in analog circuits. We also proposed an implementation of the network using MOS transistors and illustrated some of the characteristics of the network with experimental data from the fabricated circuits. The operation of the network is independent of device characteristics and only relies on accurate summation and substraction, which makes it more applicable to a broader class of devices that may include emerging technologies.

7. REFERENCES

[1] Loeliger, H-A. et. al, "Probability Propagation and Decoding in Analog VLSI *International Symposium of Information Theory*, Cambridge, MA 1998.

[2] Lustenberger, F., Helfenstein, M. et. al, "An Analog VLSI Decoding Technique For Digital Codes", *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 424-427, vol.2 1999

[3] Moerz, M., Gabara, T and Hagenauer, J, "An Analog 0.5μ BiCMOS tailbiting MAP decoder", *Solid-State Circuits Conference, 2000, ISSCC*, pp. 356-357, Feb 2000

[4] S. Chakrabartty and G. Cauwenberghs, "Forward Decoding Kernel Machines: A Hybrid HMM/SVM Approach to Sequence Recognition," *Proc. SVM'2002*, Lecture Notes in Computer Science, vol. **2388**, pp. 278-292, 2002.

[5] Chakrabartty, S. and Cauwenberghs, G. "Power Dissipation Limits and Large Margin in Wireless Sensors," *Proc. IEEE Int. Symp. Circuits and Systems*, Bangkok, Thailand 2003.

[6] Lustenberger, F., and Loeliger H-A, "On Mismatch Errors in Analog-VLSI Error Correcting Decoders", *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 198-201, vol.4 2001

[7] Cover T.A, Thomas J.A, *Elements of Information Theory*, John Wiley and Sons, 1991.

[8] Frey, B.J, Kschischang, F.R, Loeliger, H-A and Wiberg,N "Factor graphs and algorithms," *Proc. 35th Allerton Conf. on Communications, Control and Computing*, pp. 666-680, IL 1997.

[9] Bourlard, H and Morgan, N. *Connectionist Speech Recognition*, Kluwer Academic Publishers, 1994.

[10] Vapnik, V. *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.