

A PROGRAMMABLE ANALOG RADIAL-BASIS-FUNCTION BASED CLASSIFIER

Sheng-Yu Peng, Yu Tsao, Paul E. Hasler, and David V. Anderson

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA 30332

ABSTRACT

A 16×16 programmable analog radial-basis-function (RBF) based classifier is demonstrated. The distribution of each feature is modeled by a Gaussian function, which is realized by a proposed floating-gate bump circuit having bell-shaped transfer characteristics. The maximum likelihood, mean, and variance of the distribution are stored in floating-gate transistors and are independently programmable. By cascading these floating-gate bump circuits, the overall transfer characteristics approximate a multivariate Gaussian distribution with a diagonal covariance matrix. An array of these circuits constitutes a compact RBF-based classifier. When followed by a winner-take-all circuit, the analog classifier can implement vector quantization. Automatic gender identification is implemented on a 16×16 analog vector quantizer chip as one possible audio application of this work. The performance of the analog classifier is comparable to that of digital counterparts. The proposed approach can be at least two orders of magnitude more power efficient than the digital microprocessors at the same task.

Index Terms— Analog classifier, Gaussian distribution, Radial basis function, Vector quantizer, Bump circuit

The aggressive scaling of silicon technologies has led to transistors and many sensors becoming faster and smaller. The trend toward integrating sensors, interface circuits, and microprocessors into a single package or into a single chip is more and more prevalent. Fig. 1(a) illustrates the block diagram of a typical microsystem, which receives analog inputs via sensors and performs classification, decision-making, or, in a more general term, information-refinement tasks in the digital domain. Although fabrication and packaging technologies enable an unprecedented number of components to be packed into a small volume, the accompanying power density can be higher than ever, which has become one of the bottle-neck factors in the microsystem development. If the information-refinement tasks can be performed in the analog domain with less power consumption, the specifications for the analog-to-digital-converters, which are usually power-hungry, can be relaxed. In some cases, analog-to-digital conversion can be avoided altogether. The system can hence achieve higher power efficiency.

In this paper, we demonstrate a highly compact and

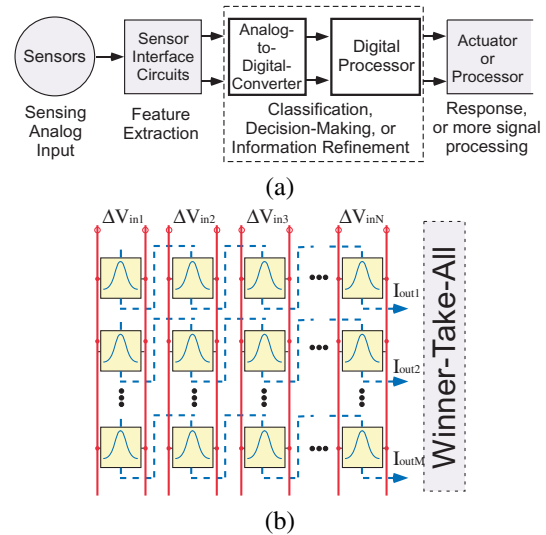


Fig. 1. (a) The block diagram of a typical microsystem. (b) An array of the proposed floating-gate bump cells, which is an RBF-based classifier, followed by a winner-take-all circuit constitute a highly compact and power-efficient analog vector quantizer.

power-efficient, programmable analog radial-basis-function (RBF) based classifier. It can serve the functions indicated inside the dashed box in Fig. 1(a) and is at least two orders of magnitude more power efficient than the digital counterparts. As illustrated in Fig. 1(b), the analog RBF-based classifier is composed of an array of proposed floating-gate bump cells having bell-shaped transfer characteristics that can realize the Gaussian distribution functions. The height, the width, and the center of a bump circuit transfer curve, which represent the maximum likelihood, the variance, and the mean of a template distribution respectively, can be independently programmed. The ability to program these three parameters empowers the classifiers to fit into different scenarios with the full use of statistical information up to the second moment.

When the RBF-based classifier is followed by a winner-take-all (WTA) stage, it results in an analog vector quantizer, which classifies the input data to the most representative template. In this paper, we conduct the automatic gender identification experiment on a resulting analog vector quantizer chip as a demonstration of one possible application of our analog RBF-based classifier. Other possible applications include en-

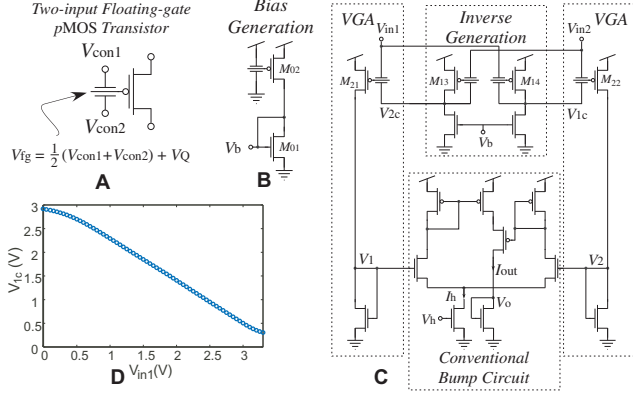


Fig. 2. **A:** The symbol for a two-input floating-gate transistor. **B:** The schematic of the bias generation block. **C:** The schematics of the proposed floating-gate bump circuit. **D:** The transfer characteristic of the inverse generation block.

environment classification in hearing aids, image pattern recognition, and chemical sensing, etc.

1. THE PROGRAMMABLE BUMP CIRCUIT

Floating-gate transistors in the circuits have two equal-size input capacitors and the symbol is shown in Fig. 2A. The schematics of the bias generation and the proposed floating-gate bump circuit are shown in Fig. 2B and C. The new bump circuit is composed of an inverse generation block, two variable gain amplifiers (VGA), and a conventional bump circuit [1]. The inverse generation block provides the complementary input voltages to the VGA, as shown in Fig. 2A. If the floating-gate charges on M_{02} , M_{13} and M_{14} are matched, then the output current will be independent of the input common-mode level. The height of the bell-shaped transfer curve is set by the tail current, I_{h1} , of the conventional bump circuit. The width can be adjusted by varying the gain of the VGA. A detailed description of the circuit has been given in [2].

The magnitude of the VGA gain decreases exponentially with the common-mode charge on M_{21} and M_{22} and hence the width of the bell-shaped transfer curve increases exponentially. We can program the differential charge on M_{21} and M_{22} to vary the center of the bell-shaped transfer curve, and program the common-mode charge to tune the width. The technique to precisely program the charges in a floating-gate transistor array was described in [3]. Because the template information is stored in a pair of floating-gate transistors as in [4, 5], this circuit has the potential to implement adaptive learning algorithms with not only an adaptive mean but also an adaptive variance.

All of the following results are from a 16×16 (16 templates in a 16-dimensional feature space) analog vector quantizer chip, which was fabricated in a $0.5 \mu\text{m}$ CMOS process. The common-mode charge of a floating-gate bump circuit is

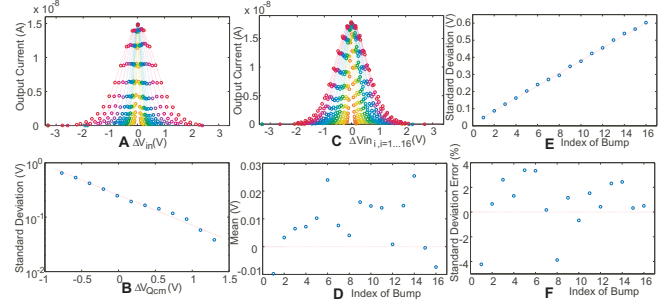


Fig. 3. **A:** Comparison between the measured 1D bumps (circles) and the corresponding Gaussian fits (dashed lines) of a single floating-gate bump circuit. **B:** The exponential relation between the extracted standard deviation and the floating-gate common-mode charge level. **C:** Comparison between the measured 1-D bumps (circles) and the corresponding Gaussian fits (dashed lines) from 16 different floating-gate bump circuits in the same template. **D:** The offsets are within 26mV. **E:** Comparison between the target (dashed line) and measured (circles) standard deviations. **F:** The programing errors of the standard deviations are within 5%.

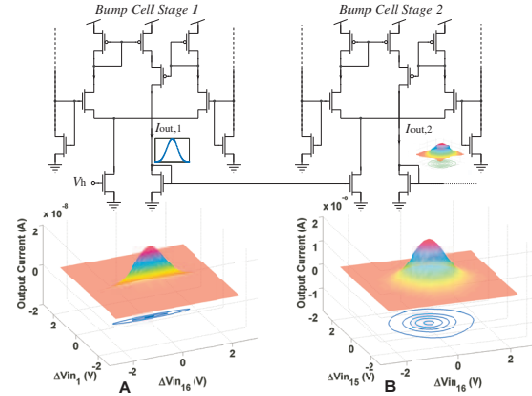


Fig. 4. A multivariate Gaussian function with a diagonal covariance matrix can be implemented by cascading the bump circuits. **A:** The first and the 16th bumps programmed in Fig. 3C are swept in a 2-D space. **B:** The 15th and the 16th bumps programmed in Fig. 3C are swept in a 2-D space.

programmed to several levels and the measured results are compared with the correspondent Gaussian fits in Fig. 3A. The extracted standard deviation is exponentially related to the common-mode charge, as shown in Fig. 3B. The minimum achievable standard deviation is 40mV, which depends on the maximum gain of the VGA. After the characterization process, 16 different floating-gate bump circuits in the same template can be precisely programmed as shown in Fig. 3C. The offsets of these 16 bump circuits are within 26mV, as shown in Fig. 3D. The measured standard deviations are compared with the targets in Fig. 3E. The programmed standard deviation errors are less than 5%, as shown in Fig. 3F.

The output current of the previous bump circuit is duplicated and fed into the next stage as its tail current to imple-

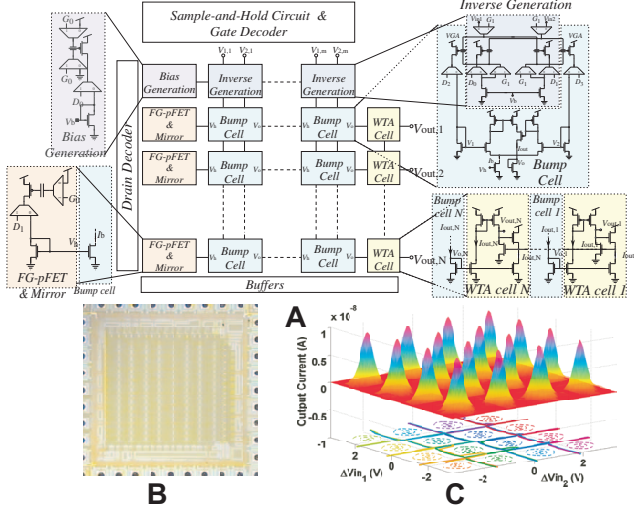


Fig. 5. A: The architecture of the resulting programmable analog vector quantizer. **B:** The micrograph of a 16×16 analog vector quantizer. **C:** 16 templates are programmed to have same variances and heights, and are evenly spaced in a 2-D space. The distributions are superposed in a 3-D plot. The thick lines at the bottom plane are the boundaries determined by the WTA outputs.

ment a multivariate Gaussian function with a diagonal covariance matrix, as shown in Fig. 4. Any two of the 16 bump circuits in the same template can be swept in a 2-D space while others remain constant to visualize the resulting bivariate distribution in a 3-D plot. Two examples of the 3-D plots measured from the floating-gate bump circuits, which are programmed as in Fig. 3C, are also shown in Fig. 4.

2. A 16×16 ANALOG VECTOR QUANTIZER

To implement an analog vector quantizer, a current mode winner-take-all circuit is placed after the floating-gate bump cell array. To control the maximum likelihood of each template in the RBF-based classifier, an “FG-pFET & Mirror” block is inserted in front of the first floating-gate bump circuit. The complete architecture and schematics of the analog vector quantizer are shown in Fig. 5A. Most of the multiplexers and the overhead circuitries for floating-gate programming are at the peripheries of the bump cell array. Consequently, the system can be easily scaled up and is highly compact.

The micrograph of the 16×16 analog vector quantizer – occupying area less than $1.5 \times 1.5 \text{ mm}^2$ – is shown in Fig. 2B. In Fig. 2C, 16 templates are programmed to have the same variances and heights and are evenly programmed in a 2-D space. The thick lines at the bottom plane indicate the boundaries determined by the winner-take-all circuit.

Receiver operating characteristic (ROC) curves, which indicate the whole range of the operating characteristics and provide a richer measure of classification performance than

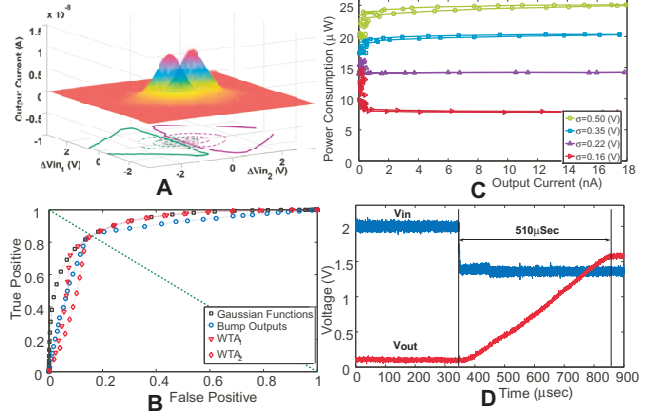


Fig. 6. A: To characterize the classifier performance, two templates are programmed to have variances of 0.5V with separation of 1V . The corresponding Gaussian functions are used as the actual pdf’s of two classes to calculate ROC curves. **B:** The ROC curves of the Gaussian functions (squares), bump output currents (circles) and WTA output voltages (triangles and diamonds). **C:** The relation between the power consumption of a single floating-gate bump cell and the output current. **D:** The transient response of the analog vector quantizer when the maximum output current is set to 10 nA .

scalar measures, are adopted to characterize our classifier performance. In the evaluation experiment, two templates are programmed to have standard deviations of 0.5V with a separation of 1V in a 2-D plane as shown in Fig. 6A. The corresponding Gaussian distributions are used as the actual probability density functions (pdf) of these two classes. Comparing these two Gaussian pdf’s using different thresholds renders an ROC curve, which is used as the evaluation reference. With the knowledge of the class distributions, comparing two output currents of the analog RBF-based classifier using different thresholds generates an ROC curve for the 2-D bumps. Comparing each of the two WTA output voltages with different thresholds generates two ROC curves that characterize the performance of the analog vector quantizer. The ROC areas under these four curves in Fig. 6B are 0.921 , 0.869 , 0.898 , and 0.876 , respectively. The equal error rates (EER), which are the usual operating points, of these four curves are 0.160 , 0.160 , 0.159 , and 0.159 . At the EER point, the performance of our RBF-based classifier is indistinguishable from that of an ideal Gaussian-function-based classifier.

3. POWER EFFICIENCY COMPARISON

The power consumption of the floating-gate bump cell is proportional to the output current and the width of the transfer curve. As shown in Fig. 6C, the larger the extracted standard deviation, the more the power consumption. The power consumption can be reduced by choosing larger transistor dimensions, which also alleviates the mismatch problem. The speed

of a single bump stage can be estimated indirectly by applying input step from different bump cell stages. In Fig. 6D, the WTA output step response is measured when a voltage step is applied from the first bump cell stage. When the maximum output current is 100 nA, the response time of a single bump cell is estimated as 0.65 μ sec.

We use the metric of millions of multiply accumulates per second per milli-watt (MMAC/s/mW) to compare the efficiency of our analog system with that of the digital hardware. Since the efficiency of the bump cells dominates the performance when the system is scaled up, we consider the single bump cell only. Each Gaussian function is estimated as 10 MACs and can be evaluated by a bump cell in 0.65 μ s with the power consumption of approximately 30 μ W. This is equivalent to 513 MMAC/s/mW. The performance of commercial low-power DSP microprocessors ranges from 1 MMAC/s/mW to 10 MMAC/s/mW. If the comparison is expanded to include the WTA function and if the WTA circuit is also optimized, the efficiency of this analog approach can be at least two to three orders of magnitude better than digital microprocessors at the same task. Moreover, this power analysis has not included the power reduction from the analog-to-digital converters, which is a major factor.

4. AUDIO CLASSIFICATION DEMONSTRATIONS

To demonstrate one possible application of this work, we use the analog vector quantizer to implement an automatic gender identification (AGI) classifier, which can be used in automatic speech or speaker recognition systems to enhance the performance. With the available number of templates and feature dimensions, eight 14-variate Gaussian components are used to characterize one specific gender. A winner-take-all voting scheme makes the final decision. The experiment is conducted on the Aurora-2 database [6], which is a standardized database for speech recognition research. Four hundred utterances from the training set are used to train the models by means of the maximum likelihood criterion. The speech data is windowed to 100msec frames and parameterized into 14 order MFCCs, consisting of 13 cepstral coefficients along with a logarithmic energy value. Although these features are prepared from a computer in our demonstration, they can be provided from an analog Cepstrum generator, as proposed in [7]. Therefore, a highly power-efficient analog audio recognizer front-end is feasible. One thousand utterances from the testing set are used to evaluate the performance. The confusion matrix is presented in Table 1. The accuracy of the ideal model on the testing set is 73.7% and the accuracy obtained from the analog vector quantizer is 69.8%.

5. CONCLUSION

In this paper, we demonstrate a new programmable floating-gate bump circuit, of which the height, the center and the

Table 1. AGI Results

	Gaussian Classifier		Analog Vector Quantizer	
	Counts as Male	Counts as Female	Counts as Male	Counts as Female
Male	389	111	374	126
Female	152	348	176	324

width of the bell-shaped transfer characteristics can be programmed individually. Based on the new floating-gate bump circuit, a compact 16 \times 16 analog vector quantizer is fabricated and tested. The performance of the classifiers are evaluated and the results are comparable to the digital system. The efficiency of this analog approach is two orders of magnitude better than digital microprocessors. An automatic gender identification system is demonstrated by using this analog vector quantizer with an accuracy of just under 70%.

6. REFERENCES

- [1] T. Delbruck, "Bump circuits for computing similarity and dissimilarity of analog voltage," in *IEEE Proceedings of the International Neural Network Society*, Oct. 1991, pp. 475–479.
- [2] S.-Y. Peng, P. E. Hasler, and D. V. Anderson, "An analog programmable multi-dimensional radial basis function based classifier," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 54, no. 10, pp. 2148–2158, 2007.
- [3] A. Bandyopadhyay, G.J. Serrano, and P. Hasler, "Adaptive algorithm using hot-electron injection for programming analog computational memory elements within 0.2% of accuracy over 3.5 decades," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 9, pp. 2107–2114, 2006.
- [4] P. Hasler, "Continuous-time feedback in floating-gate mos circuits," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48, no. 1, pp. 56–64, 2001.
- [5] D. Hsu, M. Figueroa, and C. Diorio, "A silicon primitive for competitive learning," *Advances in Neural Information Processing Systems*, vol. 13, pp. 713–719, 2001.
- [6] H. G. Hirsh and D. Pearce, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. ISCA ITRW ASR*, 2000.
- [7] P. Hasler, P. D. Smith, D. Graham, R. Ellis, and D. V. Anderson, "Analog floating-gate, on-chip auditory sensing system interfaces," *IEEE Journal of Sensors*, vol. 5, no. 5, pp. 1027–1034, 2005.