Extended Abstracts of the 2002 International Conference on Solid State Devices and Materials, Nagoya, 2002, pp. 254-255

D-5-2

Fully Parallel Nearest Manhattan-Distance-Search Memory With Large Reference-Pattern Number

Yuji Yano, Tetsushi Koide, and Hans Jürgen Mattausch

Research Center for Nanodevices and Systems, Hiroshima University, 1-4-2 Kagamiyama, Higashi-Hiroshima 739-8527, Japan
Phone: +81-824-24-6265 Fax: +81-824-22-7185 e-mail: koide@sxsys.hiroshima-u.ac.jp

1. Introduction

Finding the nearest-match between an input-data word of W bit length and a number R of reference-data words is a basic operation for pattern recognition [1] as well as data compression [2]. The nearest-match or winner is defined by the minimum with respect to a distance measure. Practically important are the Hamming (data strings, voice patterns, black/white pictures) and the Manhattan (gray-scale or color pictures) distance. Previously, we demonstrated an efficient architecture for fully-parallel minimum Hamming-distance search [3]. We also proposed an encoding approach to exploit Hamming-distance-search hardware for winner-search according to the Manhattan-distance [4], which has, however, advantages only up to 4bit binaries. Conventional partially parallel Manhattan-distance-search hardware based on multiple SRAMs and external distance calculation plus winner-take-all circuity (WTA) [5] has drawbacks with respect to integration density and short nearest-match times. To overcome these drawbacks, we propose a dedicated mixed analog-digital fully-parallel associative-memory architecture for nearest Manhattan-distance search with > 4bit binaries.

2. Associative Memory Architecture for Nearest Manhattan-Distance Search

Figure 1 shows a block diagram of the compact associative memory with fast fully-parallel match capability according to the Manhattan distance. The concept for the memory-field is illustrated in Fig. 2. Digital k-bit subtraction and absolute-value calculation units (UC) compare the W binaries, each with k-bit, in all rows of the memory field in parallel with the reference data. The k-bit subtraction circuit can be realized on the basis of a ripple carry adder circuit. In the test-chip design, we use a newly devised compact circuit to minimize its design area. The structure diagram of the winner-line-up amplifier (WLA) is shown in Fig. 3. We improved the WLA circuit proposed previously in [3] so as to obtain a large regulation range for feedback stabilization and relatively low power dissipation. The new WLA achieves this larger regulation range for feedback stabilization and eliminates the inefficient possibilities of under- or over-regulation by a maximum-gain region which self-adapts to the winner input C_{win}. The signal follower provides the necessary high driving current for scaling to a large number of reference patterns R. Low power dissipation is achieved by an individual power regulation from the signal-regulation units for each input-signal source.

Fig. 4 shows a circuit, which implements the new WLA according to Fig. 3 in CMOS technology. The transistor-count is only 6 transistors per row. A modified version of the fast minimum circuit proposed by Opris et al [6] is applied for combined feedback generation and distance amplification. The minimum function is used in the feedback loop and an intermediate node in each row circuitry is used for the distance-amplified WLA-output LA_i.

Distance amplification and self-adaptation of the maximum-gain region work as follows: Since the winner-row's WC-output C_{win} is lowest, transistor p_{3win} has the largest current-source capability, which must be balanced by the current-sink capability of transistor n_{2win}. Thus the gate voltage F_a of n_{2win}, common to all rows, has to rise appropriately and is controlled by the winner row. This in turn is only possible, if the gate voltage of the source follower n_{3win}, being also the output voltage LA_{win}, rises highest. The mechanism works independent of the absolute value of C_{win} and provides the self-adaptability of the maximum-gain region. A gain of about 20-50 over a wide range of absolute C_{win} input voltages is thus achieved. The voltage follower in Fig. 4 is equivalent to the signal followers in Fig. 3 and provides a sufficient driving current for scalability to large row numbers, which are necessary especially for codebook-based data-compression systems.

The WTA-circuit implemented in the test chip is depicted in Fig. 5. It is of O(R) complexity and needs just 17 transistors per row. In order to reduce the negative effects form fabrication induced miss-match of corresponding transistors in different rows and to improve the reliability, we adopt 5 stages of the common-source WTA-configuration proposed by Lazzaro et al. [7], which amplify winner-loser distances by voltage-current-voltage transformations. In the 1st stage, the current for the winner row is highest, because it has the largest WLA-output voltage. This highest current is then transformed into the lowest output-voltage of the 1st stage, while the output voltages of all other rows are substantially suppressed. The intermediate stages performs a similar voltage-current-voltage transformation and a further amplification of the winner-loser differences. The winner voltage is again lowest after the 5th stage. The final decision circuit consists of inverters with an adjusted switching threshold. It generates a "1" for the winner row and a "0" for each loser row.

3. Simulated Performance and Test-Chip Design

The test chip was designed in a 0.35µm CMOS technology with 3 metal layers and contains 128 reference words with 16 binaries each 5-bit long. The pattern length is estimated sufficient up to the rather high-end applications of full-color video-signal compression with 4x4 pixel blocks, after rounding of each color to a 5 bit representation. Due to a chip-area limitation, it was only possible to integrate 128 rows, which is, however, a factor 4 larger than previously reported [3,4]. Fig. 6 shows layout images of the designed associative-memory. The data of the test-chip for the associative-memory is summarized in Table 1.

Figure 7a shows the simulated nearest-match times as a function of the distance between winner and input-data word. The data for winner to nearest-loser distances of 1, 2, 5 and 10 bit is plotted. The individual current-regulation (see Fig. 4) in the VDD power supply connection of each word comparator (WC) limits the maximum WC-currents. In addition it helps to improve the static signal differences $|C_{win}-C_{lo}|$ between winner and loser rows on the WC output lines. In Fig. 7b, the simulation results of these signal differences are illustrated.

For the application to e. g. full-color video-signal compression a codebook size of about 2048 reference patterns and an operation frequency of 1-2 MHz is necessary [5]. Considering preprocessing of rotation, flipping, and inverting of the input data, the codebook size can be reduced to 256 reference patterns. Taking into account that the area for the input-pattern circuit remains the same, we extrapolate an area of about 17.2mm^2 and a power-dissipation of about 130mW (at 1 MHz) for nearest Manhattan-distance-search memory with 256 reference patterns in 0.35µm CMOS technology. If we furthermore extrapolate the test-chip data to a state-of-the-art 0.13µm CMOS technology with 1.2V power-supply, we expect an integration area of about 6.4 mm^2 and a power dissipation of about 47.3mW (at 1 MHz) for the associative-memory core of a full-color real-time motion-picture compression system.

4. Conclusion

Associative-memory architecture for fully-parallel minimum-distance search according to the Manhattan-distance is proposed and successfully verified by a chip design in 0.35µm CMOS. The 8.60mm² test-circuit with 128 reference patterns and 496 equivalent bit per pattern, has high performance, equivalent to a 32bit computer with 20 GOPS/mm², at low power dissipation of just 30.2mW/mm².

Acknowledgment

The test-chip in this study has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo with the collaboration by Rohm Corporation and Toppan Printing Corporation.

References

[1] D. R. Tveter, The Pattern Recognition Basis of Artificial Intelligence, Los Alamitos, CA: IEEE Computer Society, (1998)

[2] A. Gersho and R. M. Gray, Vector Quantization and Signal Compression, Boston, MA: Kluwer Academic, (1992)

[3] H.J. Mattausch et al., IEEE J. Solid-State Circuits, vol. 37, (2002) pp.218-227

[4] H.J. Mattausch et al., Symposium on VLSI Circuits, (2002) in press

[5] T. Nozawa et al., IEEE J. Solid-State Circuits, vol. 35, (2000) pp.1744-1751

[6] I. E. Opris, IEEE Trans. Circuits and Systems II, vol. 45, (1998) pp. 137-141

[7] J. Lazzaro et al., in Advances in Neural Information Processing Systems, I.D.S. Touretzky Ed., San Mateo, CA: Morgan Kaufmann, (1989)

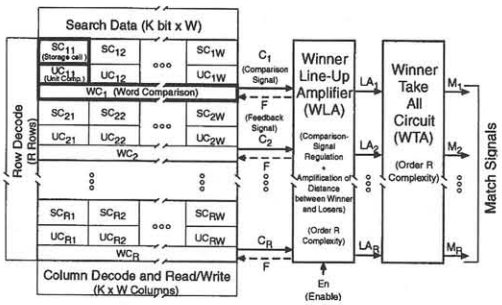

Fig. 1: Block diagram of the compact associative memory with fast fully-parallel match capability according to the Manhattan distance.

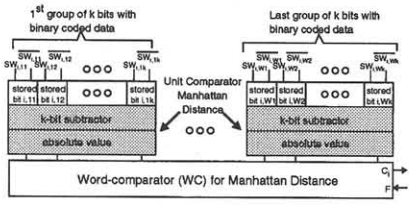

Fig. 2: Architecture of the memory-field of a mixed digital/analog associative-memory for Manhattan-distance search.

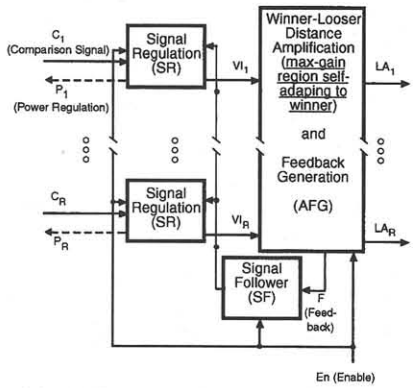

Fig. 3: Structure diagram of the winner line-up amplifier (WLA).

Table 1: Characteristics of the designed associative memory.

Distance Measure	Manhattan-Distance
Reference-Patterns	128 Pattern (16 binary each 5-bit long)
Design Area	8.60 mm² (2.32mm*3.71mm)
Nearest-Match-Unit Area	0.99 mm² =11.5% of Design Area (4.45% for WLA, 7.05% for WTA)
Nearest-Match Range	0 – 496 ((2⁵-1)*16) Equivalent Bit Input-Winner Distance
Nearest-Match Times	< 240ns
Power Dissipation	< 260mW at 4MHz (< 2.03mW per Ref. Pattern)
Technology	0.35µm, 2-Poly, 3-Metal CMOS
Supply Voltage	3.3V

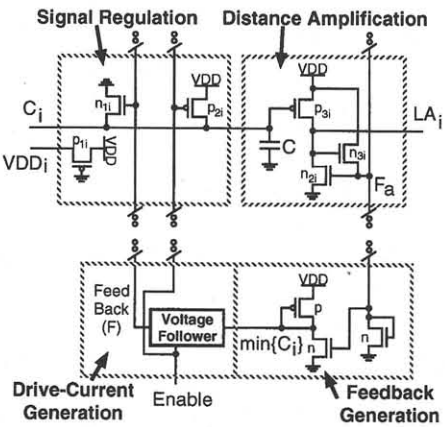

Fig. 4: Circuitry of the winner-line-up amplifier (WLA) with self-adapting maximum-gain region, following automatically the winner-row output C_{win} and thus eliminating the inefficient possibilities of under- or over-regulation.

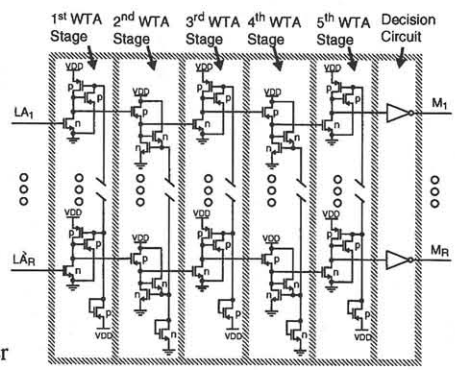

Fig. 5: Winner-take-all (WTA) circuit with 17 transistors per row of the associative memory as used for the test chip.

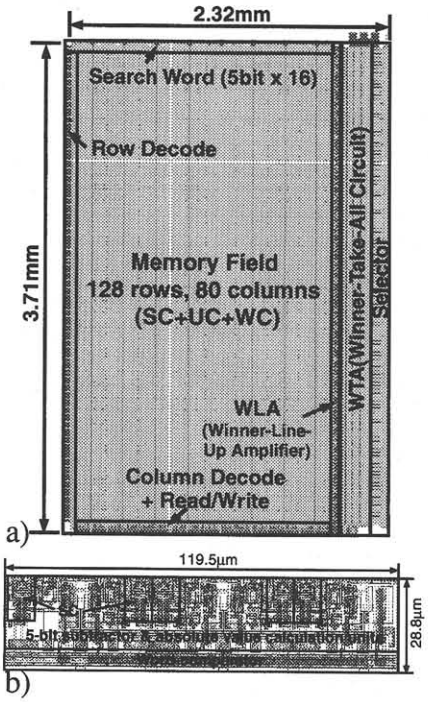

Fig. 6: Layout of the associative-memory for the Manhattan-distance, designed in 0.35µm CMOS technology. (a)whole circuit, (b) memory-field block for one 5bit binary.

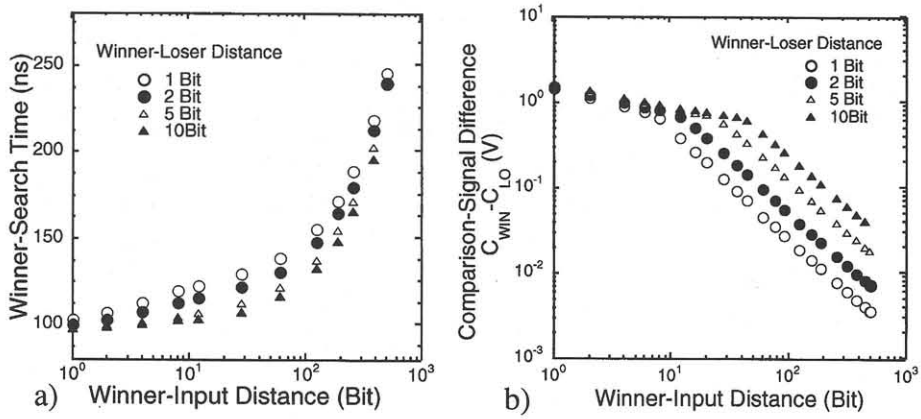

Fig. 7: Simulated performance of the designed test chip. a) Winner-search times as a function of winner distance to input and to nearest loser. b) Difference of the regulated comparison signals of winner row and nearest loser row. This difference is large for "good" matches and decreases as function of match quality.