JOURNAL OF SEMICONDUCTOR TECHNOLOGY AND SCIENCE, VOL.16, NO.4, AUGUST, 2016
http://dx.doi.org/10.5573/JSTS.2016.16.4.405

ISSN(Print) 1598-1657
ISSN(Online) 2233-4866

K-Nearest Neighbor Associative Memory with Reconfigurable Word-Parallel Architecture

Fengwei An, Keisuke Mihara, Shogo Yamasaki, Lei Chen, and Hans Jürgen Mattausch

Abstract—**IC-implementations provide high performance for solving the high computational cost of pattern matching but have relative low flexibility for satisfying different applications. In this paper, we report an associative memory architecture for k nearest neighbor (KNN) search, which is one of the most basic algorithms in pattern matching. The designed architecture features reconfigurable vector-component parallelism enabled by programmable switching circuits between vector components, and a dedicated majority vote circuit. In addition, the main time-consuming part of KNN is solved by a clock mapping concept based weighted frequency dividers that drastically reduce the in principle exponential increase of the worst-case search-clock number with the bit width of vector components to only a linear increase. A test chip in 180 nm CMOS technology, which has 32 rows, 8 parallel 8-bit vector-components in each row, consumes altogether in peak 61.4 mW and only 11.9 mW for nearest squared Euclidean distance search (at 45.58 MHz and 1.8 V).**

Index Terms—**Pattern matching, k nearest neighbor, reconfigurable vector-component parallelism, programmable switching circuits, dedicated majority vote circuit, clock mapping concept**

I. INTRODUCTION

Pattern recognition in mobile or wearable devices has attracted much attention for a wide range of applications. Accordingly, various special-purpose hardware implementtations for pattern recognition, e.g. artificial neural networks (ANNs) [1-6], support vector machines (SVMs) [7-9], and so on, have been proposed and significantly outperform software implementations with respect to recognition speed. Additionally, it has been proven that a hardware implementation achieves higher energy efficiency than a comparable sequential software implementation. Nearest neighbor search (NNS) is one of the most basic algorithms in pattern recognition to classify unknown samples [10].

The high computational costs of the minimal distance searching with O(dn) where a d-dimensional feature vector is classified among n reference vectors by brute-force search, is the main limitation of the usage of the NNS classifier. Usually, the distances between a test sample and references are often defined as the Euclidean distance. A K nearest neighbor (KNN with K>1) classifier finds the k most similar (smallest distance) reference samples that are assumed to be closely related to an unknown input sample. Then, the input-sample class is assigned to the most frequent class among these k reference samples through a majority vote. Recently, hardware implementations for KNN developed in [11, 12] achieved high performance in classification speed since KNN has intrinsic massive vector-parallelism. In [11], a FPGA-based linear-array architecture was designed for KNN featuring Manhattan-distance metric and higher parallelization than achievable with a CPU or a GPU. The mixed digital/analog solution for k nearest-neighbor search (without a majority-vote circuit) in [12] has good area efficiency. However, reliability may become insufficient for scaled-down technologies due to

Manuscript received Feb. 28, 2016; accepted Jun. 7, 2016
Hiroshima University, Japan
E-mail : anfengwei, chen, hjm@hiroshima-u.ac.jp

analog current-mode calculation of the Euclidean distance.

Apart from the speed performance, a KNN associative memory with high flexibility in number and dimensionality of reference feature vectors can also cover multiple applications. The reported KNN associative memory with reconfigurable word-parallel architecture features the following properties. (a) Mapping of the minimal squared Euclidean distance into a clock-based time domain for high search speed. (b) Control switches for match-signal connections from clocked search circuits to achieve reconfiguration of reference-vector storage and search circuitry. (c) Majority-voting circuit associated with the switches and high class-number flexibility.

The contents of this paper are organized as follows: Section II describes the reconfigurable associative memory architecture for kNN classification. Section III presents the experimental results of the proposed kNN hardware. Finally, we conclude in Sect. IV.

II. Reconfigurable Word-Parallel Architecture for KNN

1. K-Nearest Neighbor (KNN) Algorithm

KNN was introduced by Fix and Hodges [13] and is a well-known classifier with Bayes error rate [14]. It has been widely used in pattern recognition and image processing applications, such as text categorization [15], gene classification [16], content-based image retrieval [17], image compression [18], and so on.

$$c = \arg\min \left\| \mathrm{REF}_i - IN \right\| \tag{1}$$

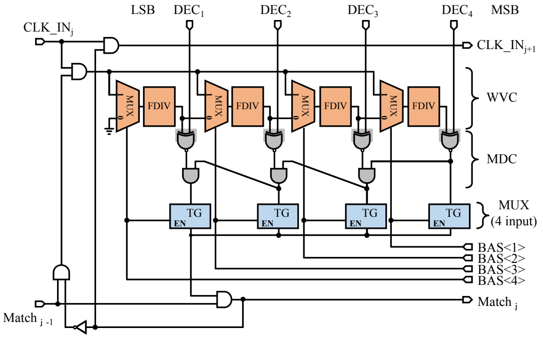
$$ED = \sqrt{\sum_{j \le d}^{0<i\le n} \left(REF_{ij} - IN_j \right)^2} \tag{2}$$

Given are a training set ***REF*** of n reference vectors ***REF***= {ref$_1$, ref$_2$, …, ref$_n$} and an unknown input vector ***IN*** defined in a d-dimensional space. Then, the nearest neighbor (NN) classifier, which represents a non-parametric statistical method, assigns ***IN*** to the class of its closest neighbor from the non-preprocessed ***REF*** in terms of a distance metric according to (1). On the other hand, ***IN*** is assigned by the KNN classifier to the class that has a majority among the k closest neighbor vectors. Here, the Euclidean distance (ED) in (2) was proved to have high efficiency in practical applications.

The main drawback of KNN is the high computational cost of *O*(dn) in case of a brute force search (BFS). In general, BFS is a basic search method which computes all distances between ***IN*** and the n reference vectors of the training set ***REF*** with d dimensionality. Finally, the class of ***IN*** is determined by a sorting algorithm among the resulting k nearest neighbors. Hardware implementtation is a good solution for this computational problem due to its intrinsic parallelizability. However, traditional method with conventional fully-digital circuits use adders and comparators and thus consume a large number of resources which causes a bad cost-efficiency relation. In this paper, a clock mapping concept [19-23] implements both the summation of component differences and the distance comparison among the references without using the conventional circuits, while still retaining the advantages of digital processing.

2. Clock Mapping Concept with Weighted-value Counters for Minimal Distance Searching

Since the root operation in (1) has no effect on the result for the k smallest distances but has high cost in hardware, the squared Euclidean distance (SED) for each vector component is used and calculated by shift operation and partial-product addition [19-23] with good area-efficiency by reusing the full adder for absolute-difference calculations of *IN* and *REF$_i$*. The distance units compute the squared absolute differences (SAD=(REF_i-IN)^2) of each vector component in parallel. Then, the SADs are partly summarized by the dimension-extension circuits (DECs) [19] to achieve additional dimension flexibility. In this paper, the clock mapping concept transforms the DEC outputs of partly summed SADs directly into the clock-number domain without completely accumulating SED (∑SAD). The outputs from each DEC are for this purpose connected to a corresponding distance evaluation unit (DEU). Rather than adders and comparators, a weighted value counter (WVC) is applied in DEU for minimal-distance searching (see Fig. 1) with lower power dissipation and smaller chip area. Each bit of the DEU consists of a 1-bit frequency divider (FDIV) (with 21 transistors [24]), a



FDIV: Frequency Divider, MDC: Match Detection Circuit, WVC: Weighted Value Counter, TG: Transmission Gate

Fig. 1. Example of a 4-bit DEU for minimal distance searching with a WVC. For each bit of the DEU a 1-bit frequency divider, a multiplexer, a XOR gate, a AND gate, and a transmission gate.

multiplexer, a XOR gate, an AND gate, and a transmission gate. The straightforward clock counting method suffers from the computational worst-case cost when the distance of each component is 2^{2N} (each feature-vector component has N bits) as described in [20].

The WVC concept reduces the required clock number by first comparing the most significant bits (MSB) of the DEC output avoiding the influence of the lower-value bits. The clock counting controlled by bit-activator [21, 22] signal (BAS) starts from MSB and switches down to a lower-value bit, whenever a match signal of any row is asserted for the higher-value bits. Correctness of the search result is achieved by considering all non-matching higher-value bits after each inclusion of lower-value bits in the search. Accordingly, the clock mapping method with the WVC can reduce the search-clock number to a linear increase $2N \times (d+1) - 1$ with N for the squared Euclidean distance search in the worst case [21, 22].

For the example of 4-bit vector components, shown in Fig. 1, the FDICs in the WVC are initialized to 0 before each search. First, the match-detection-circuit (MDC) comparison between the DEC outputs and the clock counting status of the FDIVs is restricted to the MSBs of all vector components. For the example of R reference vectors with M vector components, the search starts from the first component of all R reference vectors in parallel. It takes at most 1 clock cycle until the MDC of the first component detects a MSB match so that the search can continue with the second component. The MSB-based search continues thus at most M clock cycles until at least one reference vector issues a match signal from the MDC of its last component to the bit-activator (BA) [21, 22]. Then, the BA uses 1 clock cycle for BAS generation to expand clock counting to the next significant bit of all reference vectors in parallel, and clock counting starts again from the first component. The match processing, which takes again at most M clock cycles, is now considering the two most significant bits. In other words, the BA expands previous clock counting up to the kth significant bit by additional inclusion of the (k-1)th most significant bit, when a match signal up to the kth bit is received for at least one reference vector. In the case of M vector components, a maximum of M+1 clock cycles is therefore needed before clock counting can continue with the next lower significant bit. The minimal distance search is completed after the winner detection is expanded to the least significant bit (LSB). The completion match signal is issued from the DEU circuit of the last component of the winning reference vector to the BA circuit. Finally, the BA then outputs a global match signal for the corresponding reference vector.

3. Programmable Switches for Reconfiguring Reference Storage and Vector Dimensionality

VLSI-implementations often have high performance for solving the computational cost of KNN but have usually low flexibility for satisfying different target applications. A reconfigurable associative memory (RASM) concept is developed to reconfigure the reference storage and vector dimensionality with vector-component (word) parallelism by programmable switches (PS). RASM is a complementary solution to the DEC method [14-19]. Assume that the RASM has elements arranged in R rows and M columns, which contain SRAM cells for *p* vector components, *p* vector-distance computing units (DCUs) and one DEU. As shown in Fig. 2, this example can be configured into 6 combinations (1-row, 12-d; 2-row, 6-d; 3-row, 4-d; 4-row, 3-d; 6-row, 2-d; 12-row, 1-d) for reference-vector number and dimension by placing switches between the elements. In case of *d*-dimensional feature vectors, (*R*×*M*×*p*)/*d* vectors can be processed to find the minimal distance by appropriately reconfiguring the switches between the elements. SRAM cells are associated with the functional logic circuits for high efficiency. The configuring signal (CS$_i$) of the multiplexing switch, illustrated in Fig. 3, is initialized by pre-stored information

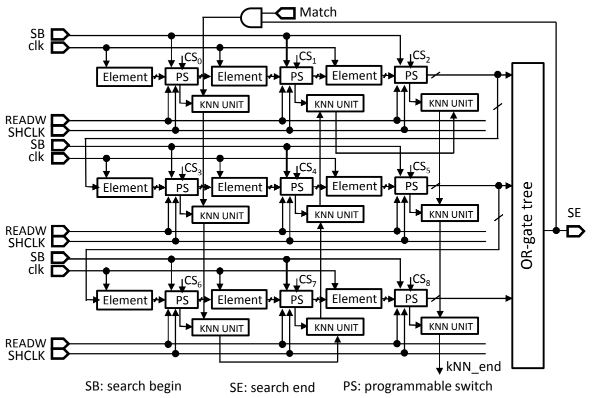


Fig. 2. Overview diagram of the reconfigurable associative memory architecture for KNN with a 3-row, 3-column example for the arrangement of basic elements. Each element contains SRAM cells for *p* vector components, *p* DCUs and one DEU.

in the memory. The switches provide a flexibility in the number of reference vectors and their dimensionality.

Each PS mainly controls the connection of the match signals ($Match_i$) for every element, OR tree, and final winner-signal reading as shown in Fig. 3. The shift-winner signal (SHW) is an enable signal to transfer the winner reading sequentially to the next element. Once the search end (SE) signal is asserted, read winner (READW) becomes the clock signal of the D-FF for storing the match signal. The match signals ($Match_i$ and MOR_i) of the neighboring elements of the current switch (PS_i) are connected when CS_i=1; in other words, these elements are assigned to the same reference vector (or row). Before the rising edge of the SEL signal, which is a delayed SE signal, the match signal ($Match_{i-1}$) of the prior element is selected by the multiplexer and then is latched in the D-FF.

In case of distinguishing two different feature vector, the prior element of the switch is the end of one reference vector (row). The next element then becomes head of the next reference vector and the match-signal input of its first element's DEU becomes the search begin (SB) signal, when CS_i=0. Additionally, the OR gate in PS_B is used as first level of the OR gate tree (OGT) for asserting the SE signal while the OGT is a part of the bit activator circuit [16, 17]. The D-FF plays as a part of a shift register once SEL is asserted. Namely, the winner stored in the D-FF of the first switch is read by the shift clock (SHCLK) after the falling edge of READW.

In this work, the RASM is arranged in *R* rows and *M* columns while each column contains elements with *p*

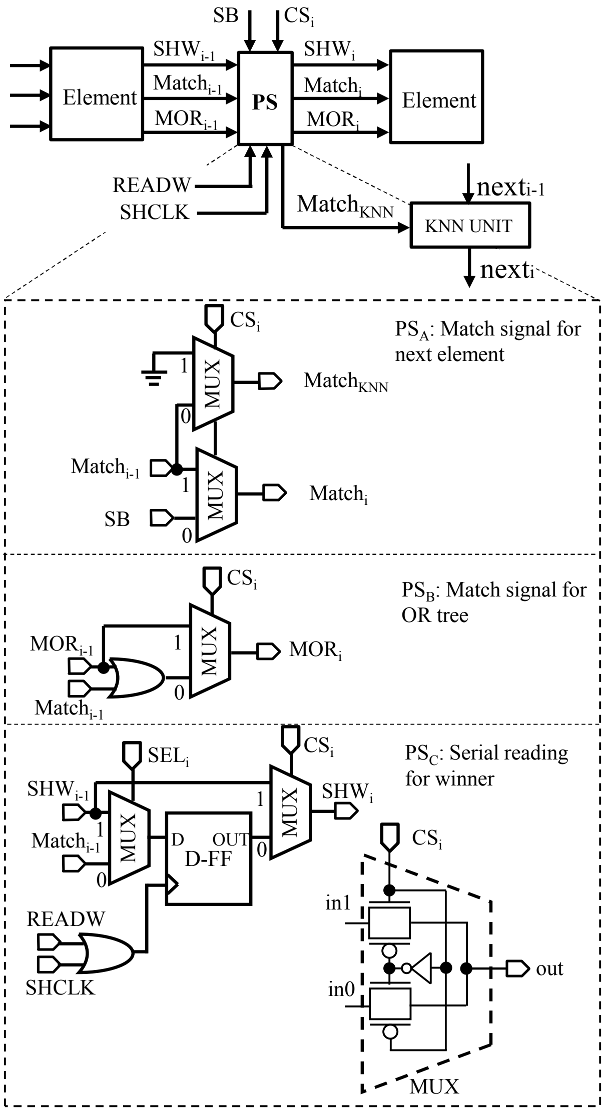


Fig. 3. Multiplexing-switch architecture for reconfiguration of reference vector number and dimension.

vector components. As a result, $[(R\times M\times p)/d)]$ (integer obtained by rounding off $(R\times M\times p)/d)$ reference vectors with *d* dimensions can be configured. However, for the developed architecture implementation, *d* should be chosen as a multiple of *p*. If less than the configurable reference-vector number is required, the DCUs of the unused elements are set to the maximum possible distance (i.e. are filled with 1's). Invalidation of the unused elements is another possibility, but has not been implemented yet.

Due to the PSs, RASM provides high flexibility in the reference number and the dimensionality of feature vectors while each PS consumes only about 60

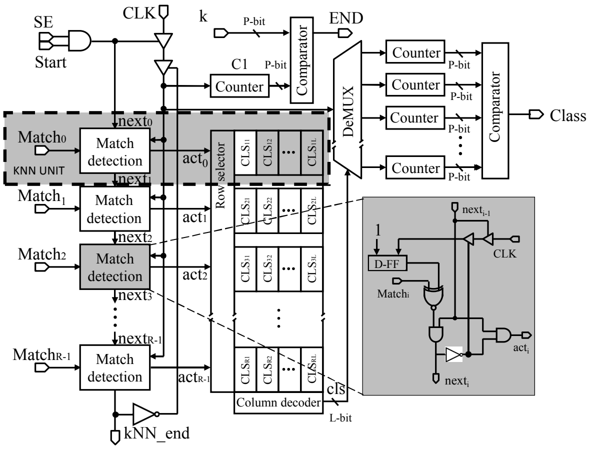


Fig. 4. Diagrams of majority vote circuit and match detection unit for the associative memory with distributed KNN unit.

transistors. Furthermore, the minimal SED can be found with high energy efficiency due to the clock mapping concept.

4. Majority Vote Circuit for KNN

The KNN classifier assigns ~~to~~ the class of an unknown input by a majority vote with the most voted label ~~in~~ among its k closest reference vectors. A dedicated majority-vote circuit of Fig. 4 is developed to find the label with largest vote value. When a match signal is detected through the OR tree, the clock counting of the DEU is terminated. Then, the match detection circuits (MDC) in Fig. 4 identify the k winner rows. The D-FF in each MDC is initialized with 0 and compared to the match signal. The next signal is conducted to the next row if the match signal (Match_i) of the current row is 0. On the other hand, when Match_i is 1, the row selection signal (act_i), which is in fact a read signal of the class information storage, is assert, the class information is read out from the class label storage (CLS) cells and the vote counter specified by the class information through the de-multiplexer (DeMUX) is counting up. Simultaneously, the D-FF is set to 1 since its clock is enabled. In the next clock cycle, a next match signal is detected by the above described process. The voting procedure is completed when the counter C1 is activated k times. In other words, k clock cycles are required for the majority vote.

As described in the previous section II.3, CS_i inputs of the PSs configure head and internal nodes of reference vectors by binary 0 and 1, respectively (Fig. 2). The distributed form of the majority vote circuit (see Fig. 5) is embedded into RASM associated with the PSs for high flexibility. In case of CS_i=1, the neighboring elements of the current switch (PS_i) are nodes within a reference pattern and the match signal for KNN ($\mathrm{Match}_{\mathrm{KNN}}$ in Fig. 3) is connected to ground (i.e. logic 0) for indicating a reference-vector-internal KNN UNIT, which is not used for the present configuration. Otherwise, in case of CS_i=0, the left element is configured as the tail of a reference vector and the right element becomes the head of the next reference vector. At the same time, the local KNN UNIT is activated for processing match signals from the tail element of the reference vector.

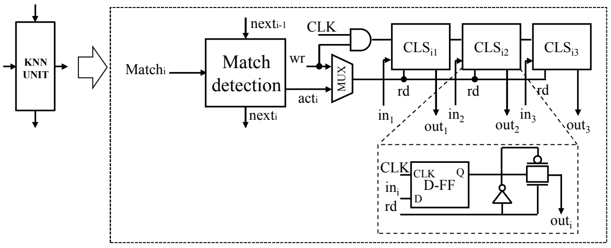


Fig. 5. Distributed form of the majority vote circuit for its embedding into the RASM.

Each local KNN UNIT has about 70 transistors and is composed of a comparator, a D-FF, five logic gates, and the storage cells (each is a L-bit register) for the class information. The remaining parts of the majority-vote circuit, except for the distributed local KNN units, namely, column decoder, DeMUX circuit, comparator for kth nearest neighbor detection, and all counters, are implemented globally. The output of each KNN unit is the class information to drive the DeMUX circuit and the corresponding vote counters. The test-chip design of the complete architecture was carried out as a full-custom design.

III. Chip Realization and Experimental Results

1. Architecture Implementation and Chip Fabrication

The test chip of the proposed RASM architecture with 32 rows and 8 vector components of 8-bit in each row ~~is~~ was designed and fabricated in 180 nm CMOS technology. Fig. 6 shows the chip photomicrograph and Table 1 lists the chip specifications. In particular, a DEC

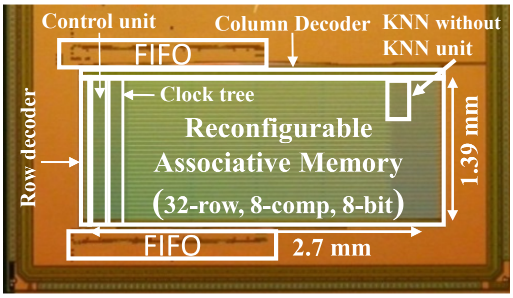


Fig. 6. Photomicrograph of the fabricated highly-flexible kNN classifier based on a reconfigurable word-parallel associative memory.

Table 1. Specification of the fabricated chip in 180nm for KNN associative memory

Performance items	Specific
Distance metric	Euclidean distance
Technology	CMOS 180 nm
Supply voltage	1.8 V
Parallelism	256-word
Total power dissipation (mW)	61.4 (45.58 MHz)
Power for search (mW)	11.9
Search time (average)	4.38 μs
Core area (mm^2)	3.75

[19] with E=24-bit, which can extend the search capability to 2048-dimensional feature vectors, is also implemented in this work.

On the other hand, the match-signal propagation mainly determines the delay of the critical path in the case where the currently evaluated distance bit is different in the 1st element and equal in all other elements. To be specific, the critical-path delay contains the propagation of the AND at the DEU input, the WVC of the 1st element, the AND gates in the match-signal path of all reference-vector elements, the multiplexers in the switches between these elements and the final OR-tree for clock-counting termination. In this work, the path delay through the AND gates in the match-signal path of the DEUs, the switches plus KNN UNITs between DEUs, and the OR-gate tree is measured at 7.79 ns. And, the measured delay within the DEU until the output of the WVC for the currently evaluated distance bit is about 0.59 ns. The match signal in the DEU has 0.2 ns increased delay per bit due to the PSs in comparison to the previous work in [19], while the PSs provide much higher flexibility and applicability for multiple applications. As a result, the maximum working frequency is about 120 MHz according to the measured critical-path delay.

As explained in section II.2, the worst-case clock number for searching the minimal distance with the clock mapping method is $2N\times(d+1)-1$ where each d-dimensional reference vector has N-bit components. The prototype chip realizes N=8-bit component precision. For indicating the worst case, in the test reference pattern with nearest distance, only the squared absolute difference of the first vector component has a large value (128^2) and for the remaining seven vector components the squared absolute differences are zero. Furthermore, in the pattern with second nearest distance, the squared absolute differences for the first, second, third, fourth, fifth and sixth vector components are 127^2, 15^2, 5^2, 2^2, 1 and 1, respectively, while the remaining two components are zero. As a result, the match signal in one DEU has to pass through the entire path of all MDCs and the MUXs of the internal PSs which leads to the critical-path delay. On the other hand, the best search case is that the winner reference has no difference to the unknown input vector for all components (distance is 0), which results in a delay of only one AND gate and one transmission gate.

The KNN unit has a class label with 3-bit (L=3-bit, 8 class categories can be expressed), a k of 4-bit (P=4-bit, so that $k\leq15$ is possible) and consumes an area of about 418.7 μm^2 in the prototype chip. Each reference vector is associated with a class label which is initialized in the storage cells of the KNN unit. Usually, several reference vectors are categorized in the same class so that the reference number is normally larger than the class number (32>8). As a result, the DCU, containing arithmetic circuits associated with memory cells, consumes the largest part of 81.3% of the total area. On the other hand, DEU with clock counting concept spends only 16% area, while the KNN UNIT and the switch circuit take only 1.4% and 1.3% of the total area, respectively. In particular, the majority vote circuit with the distributed KNN UNIT requires just about 2% chip area. In other words, the clock-mapping concept has clearly higher area and energy efficiency than a conventional solution with adders and comparators which are used in [11] for the simple Manhattan distance, as illustrated in Table 2. In general, a 1-bit static-ripple-carry full adder (SRCFA) has 28 transistors and a 1-bit logic comparator contains 22 transistors. For example, in order to summarize eight 24-bit DEC outputs and find

Table 2. Performance comparison list

	This work	FPGA solution [11]	Mixed A/D with current expression [12]	Digital solution [19]
Distance metric	Euclidean	Manhattan	Euclidean	Euclidean
Technology	0.18 µm 1P5M	0.13 µm	0.35 µm 2P3M	0.18 µm 1P5M
Parallelism	256	128	1024	256
KNN	Yes	Yes	No (k-th)	No
Power (mW)	11.9 (1.8 V@45.58MHz)	4700 @100MHz	195 (3.3 V)	5.02 (1.8 V@42.9MHz)
Search time	8.76 µs	0.03 µs	>>0.96 µs	>>30 µs
Normalized power (@100MHz)	13.5 mW	4700 mW	-	12.4 mW
Search time (@100MHz)	3.99 µs	0.03 µs	>>0.96 µs	>>12.9 µs
Power-delay product (nJ)	53.9	141	>>187.2	>>160
Area (mm^2)	3.75	22 K logic cells	5.12	3.51

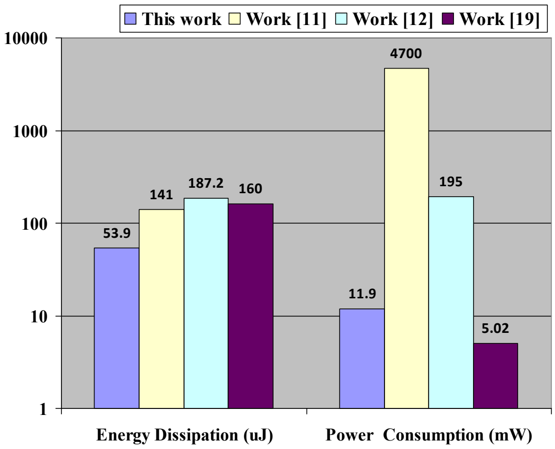


Fig. 7. Energy and power efficiency comparison with previous works.

(comparison) the nearest match as in the case of our test-chip implementation, four 24-bit, two 25-bit, and one 26-bit SRCFAs for each row, plus one 26-bit comparator for every two rows with about 0.06 mm^2 are required to implement a full-digital solution. Meanwhile, in the clock mapping concept, 8 cascaded 24-bit DEUs with 0.01 mm^2 are used to implement the complete distance searching. It is obvious that the full-digital method uses 6 times more area than the clock mapping concept, resulting in addition in an equivalently higher power dissipation.

While the DCU with arithmetic circuitry consumes most of the area, the distance computation by DCUs, which requires only at most 8 clock cycles per DCU in the case of 8-bit words, uses more than 80% of the total power dissipation. The power dissipation is measured to be 61.4 mW (at 45.58 MHz, 1.8 V supply voltage), using the on-chip ring oscillator and average search configurations for the DCU and DEU. On the other hand, PSs, KNN-UNITs and MVC consume only 11.9 mW during KNN classification.

2. Experimental Results and Comparison

Regarding to applications with feature vectors of more than 8 dimensions, additional circuitries for partially loading new vector-components for processing and storing intermediate distance results are required in [19]. Furthermore, this work as a complimentary solution for extending the applicability and flexibility of [19] and outperforms it in both speed and power-delay product for target applications with more than 8-dimensional feature-vectors. The implementation of PSs can overcome the deficiency of the storage of reference vectors. On the other hand, a higher parallelism requires more hardware resources, in particular, the implementation for adders and comparators. As shown in Fig. 7, due to the implementation of PSs and distributed KNN UNITs, the power consumption of this work is 2.4x higher than that of [19] while the energy dissipation is more than 3x lower than that of [19]. If the dimensionality of the target application becomes larger, the energy efficiency of this work continues to become better than that of [19].

As described in section III.1, arithmetic circuitry, i.e. DCU, with large chip-area consumption certainly uses also a much larger part of the power, where the area of the DCU is 5 times larger than that of the DEU applying the clock counting concept. Indeed, higher parallelism leads to faster processing speed but consumes also much more resources, as verified by e.g. the full-digital FPGA implementation in Table 2. However, this work can achieve higher search speed by massive parallelism without huge area and power problems due to clock-based distance-mapping concept. Even though the analog/digital solution [12] has better area efficiency than the reported work, Si-area and transistor length in [12] cannot be scaled down easily since distances are expressed by voltage differences. Further improvement via an advanced low voltage CMOS technology is therefore not possible for the work of [12].

V. Conclusions

In this paper, we proposed a reconfigurable associative memory (RASM) concept with vector-component parallel architecture for k nearest neighbor (KNN) classification. A clock-counting solution with high power/area efficiency for nearest Euclidean distance search and a distributed circuitry for class-determination by majority voting are applied. Furthermore, programmable switches (PSs) enable configurability of reference-vector storage with respect to vector number and dimensionality, resulting in the implementation possibility of many different applications on the same integrated hardware. A prototype chip is designed and fabricated in 180 nm CMOS technology to demonstrate this flexibility and the power/area efficiency of RASM through experimental verification.

Acknowledgments

This research was supported by grant 25420332 from the ministry of Science and Education, Japan. The VLSI-chip was fabricated through the chip fabrication program of VDEC, the University of Tokyo in collaboration with, Rohm, Synopsys, and Cadence.

References

[1] J. Liu, M. A. Brooke, and K. Hirotsu, “A CMOS feedforward neural-network chip with on-chip parallel learning for oscillation cancellation,” *Neural Networks, IEEE Transactions on,* 13 (5), pp. 1178–1186 , 2002.

[2] J. Liu, M. A. Brooke, and K. Hirotsu, “A CMOS feedforward neural-network chip with on-chip parallel learning for oscillation cancellation,” *Neural Networks, IEEE Transactions on,* 13 (5), pp. 1178–1186 , 2002.

[3] P. Arena et al., “A CNN-based chip for robot locomotion control,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 52 (9), pp. 1862–11, 2005.

[4] H. Li, D. Zhang, and S. Foo, “A Stochastic Digital Implementation of a Neural Network Controller for Small Wind Turbine Systems,” *Power Electronics, IEEE Transactions on,* 21 (5), pp. 1502–1507, 2006.

[5] T. Koickal et al., “Analog VLSI Circuit Implementation of an Adaptive Neuromorphic Olfaction Chip,” *Circuits and Systems I: Regular Papers, IEEE Transactions on,* 54 (1), pp. 60–73, 2007.

[6] F. An, et al, “VLSI realization of learning vector quantization with hardware/software co-design for different applications,” *Japanese Journal of Applied Physics*, 54, 04DE05, 2015.

[7] D. Anguita, A. Bon, and S. A Ridella, “A digital architecture for support vector machines: theory, algorithm, and FPGA implementation,” *Neural Networks, IEEE Transactions on,* 14 (5), pp. 993-1009, 2003.

[8] D. Anguita and A. Boni, “Improved neural network for SVM learning,” *Neural Networks, IEEE Transactions on,* vol. 13, pp. 1243–1244, 2002.

[9] S. S. Keerthi and E. G. Gilbert, “Convergence of a Generalized SMO Algorithm for SVM Classifier Design,” *Machine Learning*, Vol. 46, pp. 351–360, 2002.

[10] C. Kyrkou, and T. Theocharides, “A Parallel Hardware Architecture for Real-Time Object Detection with Support Vector Machines,” *Computers, IEEE Transactions on,* 61 (6), pp.831-842 (2012).

[11] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *Information Theory, IEEE Transactions on,* 13, pp. 21-27, (1967).

[12] E. S. Manolakos, and I. Stamoulias, “IP-cores design for the kNN classifier”, *in Proc. ISCAS*, pp. 4133-4136, 2010.

[13] Md. A. Abedin, et al, “Mixed digital-analog associative memory enabling fully parallel nearest Euclidean distance search”, *Japanese Journal of Applied Physics*, 46, pp.2231-2237, 2007.

[14] E. Fix and J. L. Hodges. Discriminatory analysis, nonparametric discrimination: Consistency properties.

Technical Report 4, USAF School of Aviation Medicine, Randolph fiels, TX, 1951

[15] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory*, IT-13(1): 21–27, 1967.

[16] S. Jiang, G. Pang, M. Wu, L. Kuang, An improved K-nearest-neighbor algorithm for text categorization, *Expert Systems with Applications*, Vol. 39 (1), pp. 1503-1509, 2012.

[17] F. Pan, B. Wang, X. Hu, and W. Perrizo, “Comprehensive vertical sample-based knn/lsvm classification for gene expression analysis,” *J. Biomed. Inform.*, vol. 37, pp. 240–248, 2004.

[18] H. Zhang, A. C. Berg, M. Maire, and J. Malik, “SVM-KNN: Discriminative nearest neighbor classification for visual category recognition,” *in International Conference on Computer Vision and Pattern Recognition,* pp. 2126-2136, 2006.

[19] K. W. Hung and W. C. Siu, "Novel DCT-Based Image Up-Sampling Using Learning-Based Adaptive KNN MMSE Estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 12, pp. 2018-2033, 2014.

[20] F. An, et al, “A Coprocessor for Clock-Mapping-based Nearest Euclidean distance Search with Feature Vector Dimension Adaptability”, *in Proc. CICC,* pp. 1-6 , 2014

[21] S. Sasaki, Masahiro Yasuda, and H. J. Mattausch., “Digital associative memory for word-parallel Manhattan-distance-based vector quantization”, *in Proc. ESSCIRC,* pp. 185-188, 2012.

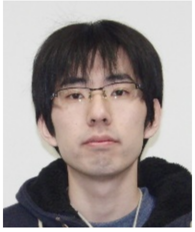
[22] T. Akazawa, S. Sasaki, and H. J. Mattausch, Word-Parallel Coprocessor Architecture for Digital Nearest Euclidean Distance Search, *in Proc. ESSCIRC,* 2013, pp. 267-270.

[23] T. Akazawa, S. Sasaki, and H. J. Mattausch, “Associative memory architecture for word-parallel smallest Euclidean distance search using distance mapping into clock-number domain,” *Japanese Journal of Applied Physics*, 53, 04EE16, 2014.

[24] F. An, K. Mihara, S. Yamasaki, L. Chen, and H. J. Mattausch, “Word-parallel Associative Memory for k-Nearest-Neighbor with Configurable Storage Space of Reference Vectors,” IEEE Asian Solid-State Circuits Conference, pp. 14-4, 2015.

[25] E. A. Vittoz, et al, “Silicon-gate CMOS frequency divider for electronic wrist watch”, *Solide-State Circuits, IEEE Journal of,* 7 (2), pp. 100-104, 1972.

Fengwei An received his M.S. and PhD from Hiroshima University, Japan, in March 2010 and March 2013 respectively. He has been an Assistant Professor at Hiroshima University in graduate school of engineering since December 2013. His research interests include energy-efficiency image recognition algorithms, and low-power circuit for embedded systems.



Keisuke Mihara received the B.S., from Hiroshima University in 2014. Now he is master student and his main research is on the associative memory design for nearest neighbor search.

Shogo Yamasaki received B.S. and M.S. from Hiroshima University in 2013 and 2015, respectively. His main research is on the associative memory design for nearest neighbor search.

Lei Chen received PhD from Hiroshima University, Higashi-Hiroshima, Japan, in 2012. She is currently pursuing her post-doctoral research at the HiSIM Research Center, Hiroshima University. Her research interests include circuit design based on organic thin-film transistors and image processing VLSI.

Hans Jürgen. Mattausch received the doctor degree from the University of Stuttgart, Stuttgart, Germany, in 1981. He is a Professor at the Research Institute for Nanodevice and Bio Systems and the Graduate School for Advanced Sciences of Matter, Hiroshima University, Higashi-hiroshima, Japan. From 1982 to 1996 he was with Siemens AG in Munich, Germany, where he was involved in the development of CMOS technology, memory and telecommunication circuits, power semiconductor devices, chip-card ICs and compact models. Since 1996 he is a Professor with Hiroshima University, researching in the fields of VLSI design, nano-electronics and compact modeling. Dr. Mattausch is a senior member of IEEE and a member of IEICE.