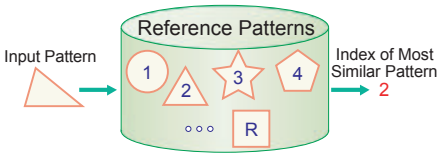
Associative Memory for High-Speed Nearest Hamming/Manhattan Distance Search with Large Reference Pattern Number

Hiroshima University

Yuji Yano, Tetsushi Koide, and Hans Jürgen Mattausch

Research Center for Nanodevices and Systems, Phone: +81-824-24-6265, Fax: +81-824-22-7185

Associative Memory Functionality



- Finding the nearest match pattern among R reference patterns.
- Important for pattern recognition (Hamming) and codebook based data compression (Manhattan).

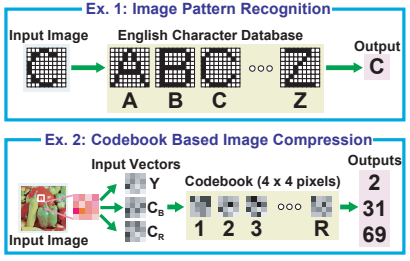
Hamming Distance

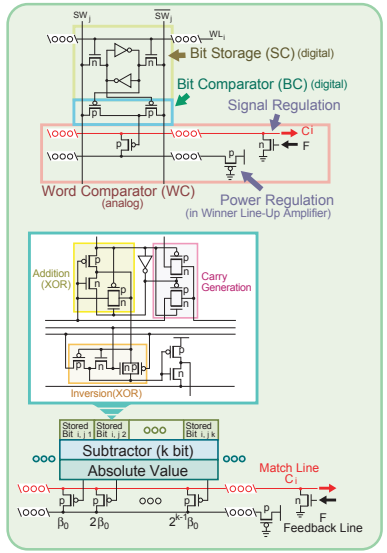
$$D_{Hamm,\,i} = \sum_{j=1}^{W} (IN_j \oplus REF_{i,j})$$

Manhattan Distance

$$D_{Manh,\,i} = \sum_{j=1}^{W} |IN_j - REF_{i,j}|$$

- Minimum Hamming distance search for pattern recognition with binary (black/white) images
- Manhattan distance for many applications with color/gray-scale images

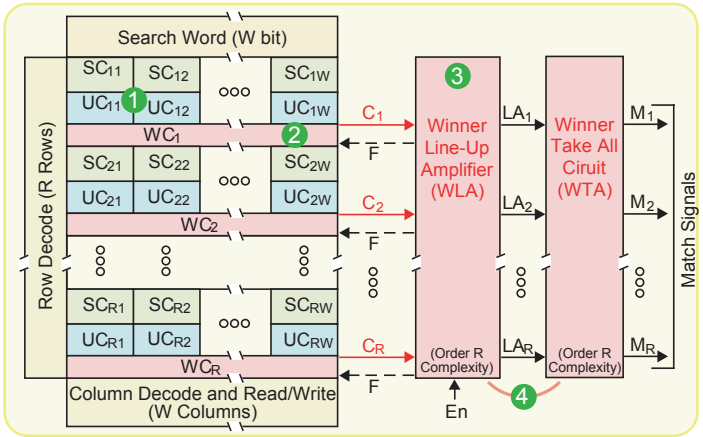


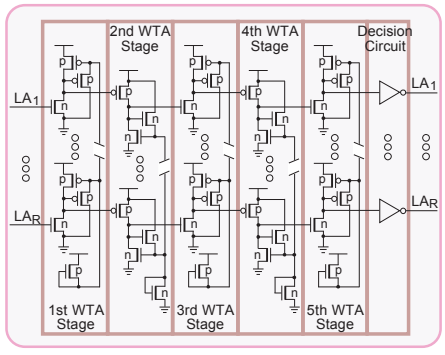


- K-bit digital subtractor and absolute-value calculator are needed to realize the Manhattan-distance-search.
- All unit comparators and all word comparators calculate the distance between search word and stored word in parallel.

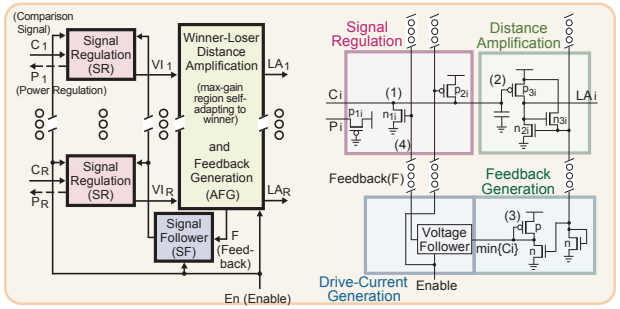
Associative Memory Architecture

1. K-bit subtractors and absolute value calculators (UC, WC) within the memory field.
2. A fast and static analog-current-encoding of the word-comparison results.
3. Improved self-adapting regulation circuit to the point of the largest winner-loser distance amplification for all search cases.
4. Winner search circuit (WLA, WTA) with only O(R) complexity.



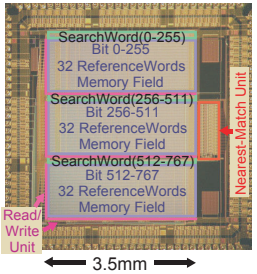


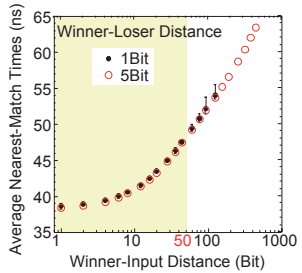
- Each stage amplifies the differences by a voltage-current-voltage transformation.
- Enough amplification magnitude (by a factor 20-50)

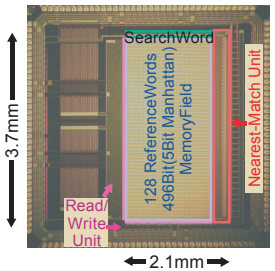


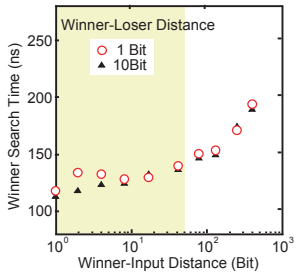
- The current on C_{win} is the smallest and the voltage of C_{win} is the lowest.
- The current-source capability of p_{3win} is the largest and the voltage of LA_{win} is the highest.
- The feedback voltage of F is approximately equal to that of C_{win}.
- For larger winner-input distance the voltage F becomes higher and n_{1i} have larger current-sink capability.

Associative Memory Chip





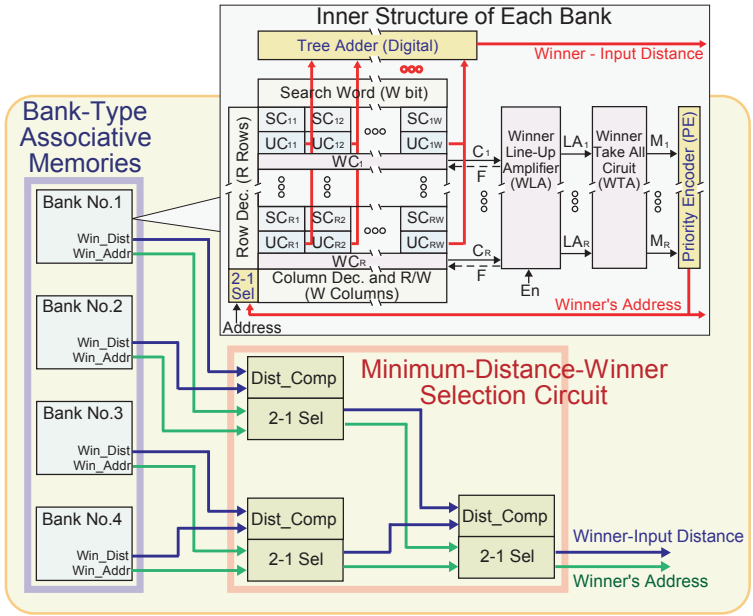


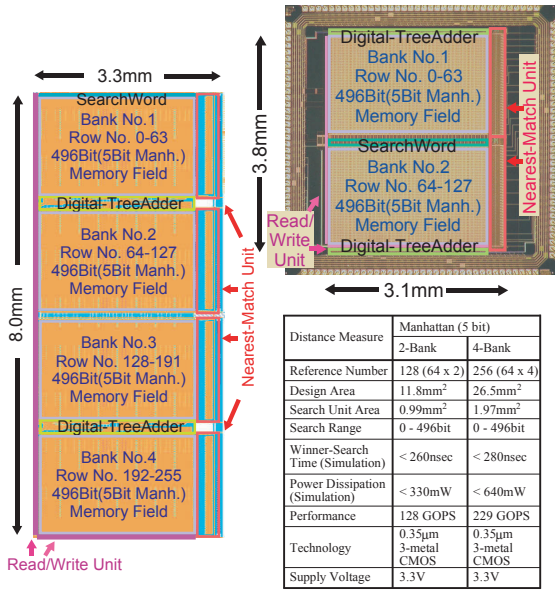


Distance Measure	Hamming	Manhattan (5 bit)
Memory Field	32 x 768	128 x 80
Technology	0.6μm CMOS	0.35μm CMOS
Area	9.11 mm^2	8.6 mm^2
Search Range	0 - 400 bit	0 - 480 bit
Winner-Search Time (Measured)	< 70 nsec	< 190 nsec
Performance	1.34 TOPS	160 GOPS
Power Dissipation	43 mW	91 mW
Supply Voltage	3.3V	3.3V

- 9.11mm^2 test chip designed in 0.6μm CMOS technology (Hamming) and 7.49mm^2 test chip designed in 0.35μm CMOS technology (Manhattan).
- High-speed minimum distance search at < 70ns (Hamming) and at < 190ns (Manhattan).
- Low power dissipation of < 43mW (Hamming) and of < 91mW (Manhattan).

Bank-Type Associative Memory





Distance Measure	Manhattan (5 bit)	
	2-Bank	4-Bank
Reference Number	128 (64 x 2)	256 (64 x 4)
Design Area	11.8mm^2	26.5mm^2
Search Unit Area	0.99mm^2	1.97mm^2
Search Range	0 - 496bit	0 - 496bit
Winner-Search Time (Simulation)	< 260nsec	< 280nsec
Power Dissipation (Simulation)	< 330mW	< 640mW
Performance	128 GOPS	229 GOPS
Technology	0.35μm 3-metal CMOS	0.35μm 3-metal CMOS
Supply Voltage	3.3V	3.3V

- Each of the 4 banks searches its local winner independently.
- The minimum distance winner selection circuit determines the global winner among 4 local winners.
- Each bank has the circuitry (PE, tree adder, digital distance output port) for digital calculation of the local winner.

- 11.8mm^2 and 26.5mm^2 test chips for bank-type associative memories are designed in 0.35μm CMOS technology with 3 metal layers.
- 2/4 bank associative memories have the function of minimum Manhattan distance search among 128/256 reference patterns.
- Each chip has high performance at low power dissipation.

Conclusions

- **Associative memories without and with (for large pattern number) bank-type architectures are proposed for fully-parallel minimum distance search.**
- **Test chips are designed in 0.6μm (Hamming) and in 0.35μm CMOS technologies.**
- **Measured data indicates sufficient performance for application in mobile real-time systems.**