Design and evaluation of 6T SRAM layout designs at modern nanoscale CMOS processes

Dimitrios Balobas
Department of Informatics
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
dmpalomp@csd.auth.gr

Nikos Konofaos
Department of Informatics
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
nkonofao@csd.auth.gr

***Abstract*—Six layout variations of the 6T SRAM cell are examined and compared. The comparison includes four conventional cells, plus the thin cell commonly used in industry and a recently proposed ultra-thin cell. The layouts of the cells are presented and corresponding memory arrays are implemented at 65, 45 and 32 nm using 3-metal CMOS n-well process. The obtained designs are compared in terms of area, power dissipation and read/write delay, using proper BSIM4 level simulations. The thin cell presents the best results regarding area efficiency and delay. In terms of power dissipation, it performs poorly at 65 and 45 nm but appears to be the best at 32 nm, presenting great improvement with downscaling. The ultra-thin cell provides a more lithographically friendly alternative to the thin cell, with lower power dissipation at 65 and 45 nm and higher at 32 nm. Overall, it performs worse in area and power relative to most conventional designs and gets worse with downscaling.**

Keywords—SRAM; layout; 6T cell; memory array; delay; power;

I. Introduction

SRAM design is becoming increasingly challenging with each new technology node. The most pressing issues arising from scaling are increased static power, cell stability concerns, reduced operating margins, robustness and reliability, and testing [1]. Despite the growing challenges of lithography and variability, though, the 6T SRAM cell size has scaled well over five process generations [2]. In this work, various layout implementations of the 6T cell, as well as 16 bit memory arrays of each corresponding cell type, are designed at 65, 45, and 32 nm and evaluated in terms of area, power dissipation and read/write delay, using suitable simulation. The results are compared in order to derive a potential optimum performance and observe the effects of scaling in each design.

II. Cell Categorization

According to the categorization made by Ishida et al [3], the 6T SRAM cells are divided into four variations that result from the different placement of the two inverters constituting the core of the 6T cell. The first type consists of two sub-types, making a total of five basic cells: type 1a [4, 5], type 1b [6], type 2 [7], type 3 [8] and type 4 [9]. Amongst the conventional 1-3 types, type 2 is the most popular cell design which has been widely used until the 90 nm generation. Due to lithography limitations with deeper nanoscaling, it was replaced by the lithographically friendly type 4 cell, also known as the thin cell [10], which has been the industry standard since 65 nm [2, 11]. This cell is long and skinny, reducing the critical bitline capacitance at the expense of longer wordlines. Ishida's categorization has been recently expanded to include a type 5 category, introducing the type 5 ultra-thin cell [11], which, compared to the thin cell, is said to offer lower bit line capacitance, reduced metal complexity and notchless design for improved resistance to alignment induced device mismatch. The cell categories and corresponding types are shown in Fig. 1. From now on, the cells will be referred to as T1a, T1b, T2, T3, T4 and T5.

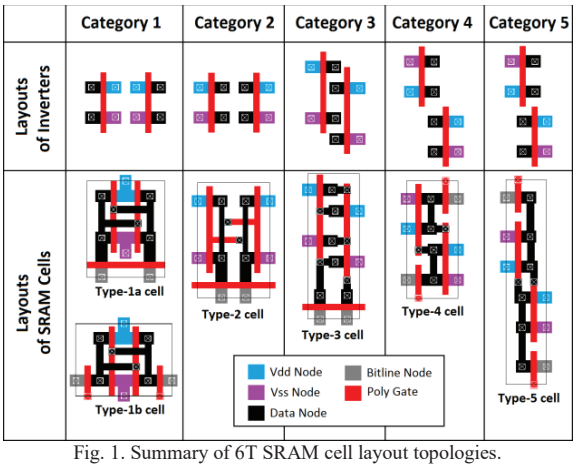

Fig. 1. Summary of 6T SRAM cell layout topologies.

III. SRAM Layout Design

A. Cell Design and Sizing

For the layout design of all cells we use a standard 3-metal CMOS n-well process, with each cell implemented following the same design rules at 65, 45, and 32 nm. While the T4 and T5 cells originally use up to two levels of metal and one level of local interconnect (trench contacts), in this work they are implemented with three levels of metal instead. The layouts of all cells are shown in Fig. 2.

To ensure both read stability and writability, the transistors must satisfy certain ratio constraints. The nMOS driver transistors in the cross-coupled inverters must be strongest, the nMOS access transistors must be of intermediate strength, and

UMC 1449
UMC v. AICP

the pMOS pullup transistors must be weak [12]. To achieve a good layout density, all of the transistors must be relatively small. In this work, we use the same sizing for all types of the examined SRAM cells. The length of all transistors is the minimum, 2λ. The width is 3λ, 4λ and 6λ for the pullup, access and pulldown transistors, respectively.

B. Array Design and Area Comparison

The cells presented above are to be used for the construction and evaluation of memory arrays, thus we used each cell type to design 4x4 (16-bit) SRAM arrays. Every array type is implemented with the maximum area efficiency that the corresponding cell can provide, given the design rules followed. Hence, some cells are properly flipped horizontally or vertically in order to partially merge and overlap with neighboring cells. This results in different cells sharing the same polysilicon, diffusion or n-well areas, as well as metal wires and contacts. Furthermore, n-well taps and substrate contacts may be shared among multiple cells for additional area efficiency. The connections inside the cells are implemented with metal-1 wires and polysilicon gates, while the I/O routing (wordlines and bitlines) is implemented with metal-2 and metal-3 wires. The layouts of the 16-bit arrays are shown in Fig. 3.

After comparing the layouts of the various 16-bit SRAM architectures, it can be safely assumed that the T4 thin cell presents the highest area efficiency, as shown in Fig. 4. The T4 cells overlap on all four sides, thus saving significant area by shared diffusions and contacts. The T5 cells also overlap on all sides, but they leave a lot of area unoccupied between them, resulting in an area-inefficient structure. Indicatively, at the 32 nm, the T4 array covers an area of 3.186 μm^2, which is 7.97%, 13.14%, 14.9%, 32.6% and 36.0% less than the T1a, T3, T2, T5 and T1b designs, respectively. The T1a, T3 and T2 cells are close at 3.462, 3.668, and 3.744 μm^2. The T5 ultra-thin cell performs worse than most basic cells, at 4.730 μm^2. The T1b cell results in the largest layout at 4.984 μm^2. A similar analogy can be derived for the 65 and 45 nm circuits. The area and bit density of the SRAM arrays is shown in Table I.

TABLE I. AREA AND BIT DENSITY OF SRAM ARRAYS

	65 nm		45 nm		32 nm	
SRAM Type	***Area (μm^2)***	***Bit Density (μm^2/bit)***	***Area (μm^2)***	***Bit Density (μm^2/bit)***	***Area (μm^2)***	***Bit Density (μm^2/bit)***
T1a	18.849	1.178	6.154	0.385	3.462	0.216
T1b	27.136	1.696	8.861	0.554	4.984	0.312
T2	20.386	1.274	6.657	0.416	3.744	0.234
T3	19.970	1.248	6.521	0.408	3.668	0.229
T4	17.346	1.084	5.664	0.354	3.186	0.199
T5	25.754	1.610	8.410	0.526	4.730	0.296

IV. SIMULATIONS AND RESULTS

The SRAM cells, as well as the 16-bit SRAM memory arrays, are simulated under varying conditions, to calculate and compare their performance in terms of propagation delay and power dissipation. For all the designs and simulations, a BSIM4 level model for low-leakage nMOS and pMOS transistors is used at the 65, 45, and 32 nm. Furthermore, all simulations are performed under room temperature (27° C), at an operating frequency of 1GHZ, meaning that the word line is inserted every 1 ns to begin a new read/write cycle. The supply and input voltage is set to 1.0 V for the 65 and 45 nm simulations and 0.8 V for the 32 nm simulations.

A. Read/Write Delay of Cells

To calculate the delay of the write operation, two cases must be considered: writing '0' when the cell contains '1' and writing '1' when the cell contains '0'. In each case, the delay is calculated between the insertion of the word line and the switching of the data node to the new input. The pullup transistors are smaller than the driver transistors, hence the 'write 1' delay is higher than the 'write 0' delay. The average value of these two cases is calculated for each cell. When writing the same value to the cell, there is no delay to be measured.

To calculate the delay of the read operation, an external circuit has to be used for signal sensing. In this simulation, we use a large signal sensing method, specifically a pair of HI-skew inverters connected to the bit lines. The transistor sizes for the inverters are: Wp = 9λ, Wn = 4λ, Lp = Ln = 2λ. The delay is calculated between the insertion of the word line and the switching of the bitline inverter's output node to 1 when reading 0, or the switching of ~bitline inverter's output node to 1 when reading 1. The average value of 'read 0' and 'read 1' is calculated for each cell. The simulation results regarding the write and read delay of the cells are summarized in table 2.

TABLE II. READ AND WRITE DELAY OF SRAM CELLS

	65 nm – 1.0 V		45 nm – 1.0 V		32 nm – 0.8 V	
SRAM Type	***Read delay (ps)***	***Write delay (ps)***	***Read delay (ps)***	***Write delay (ps)***	***Read delay (ps)***	***Write delay (ps)***
T1a	8	7.5	6	7.5	5	6.5
T1b	8	7	6	7	6	6.5
T2	8	7	6	6.5	5	6
T3	8	6.5	6	7	6	6
T4	8	6	6	6	5	5.5
T5	8	7	6	7	6	6

B. Power Dissipation of Cells and Arrays

When a memory cell is active, six possible operations can occur: write 0 when data = 0, write 0 when data = 1, write 1 when data = 0, write 1 when data = 1, read 0, read 1. In each case, a different amount of power is dissipated. To calculate the average power dissipation of the cell, proper bit sequences are inserted to the bitlines to cover all the possible transactions. More specifically, the repeating sequence of transactions that the cell performs is: write 0 (writing 0 when data = 1), write 0 (writing 0 when data = 0), read (reading 0), write 1 (writing 1 when data = 0), write 1 (writing 1 when data = 1), read (reading 1). The results are shown in Table 3. All memory arrays are simulated under the same scenario, comprising a sequence of 4 write cycles, 4 read cycles and another 4r write and read circles, for a total of 16 ns. The lines are written and then read consecutively. Certain 4-bit words are used so that the input sequence is identical in every array's simulation. Additionally, the input sequences are properly set so that no external circuitry is needed for addressing, precharging e.g. The results are shown in Table 4.

TABLE III. POWER DISSIPATION OF SRAM CELLS AND ARRAYS

	65 nm – 1.0 V		45 nm – 1.0 V		32 nm – 0.8 V	
SRAM Type	***Cell power (μW)***	***Array power (μW)***	***Cell power (μW)***	***Array power(μW)***	***Cell power (μW)***	***Array power (μW)***
T1a	0.263	2.029	0.113	0.911	0.063	0.557
T1b	0.326	2.489	0.149	1.232	0.066	0.560
T2	0.304	1.774	0.126	0.779	0.069	0.522
T3	0.301	2.047	0.123	0.870	0.068	0.569
T4	0.283	2.167	0.092	0.985	0.056	0.492
T5	0.328	2.103	0.141	0.947	0.076	0.599

UMC 1449
UMC v. AICP

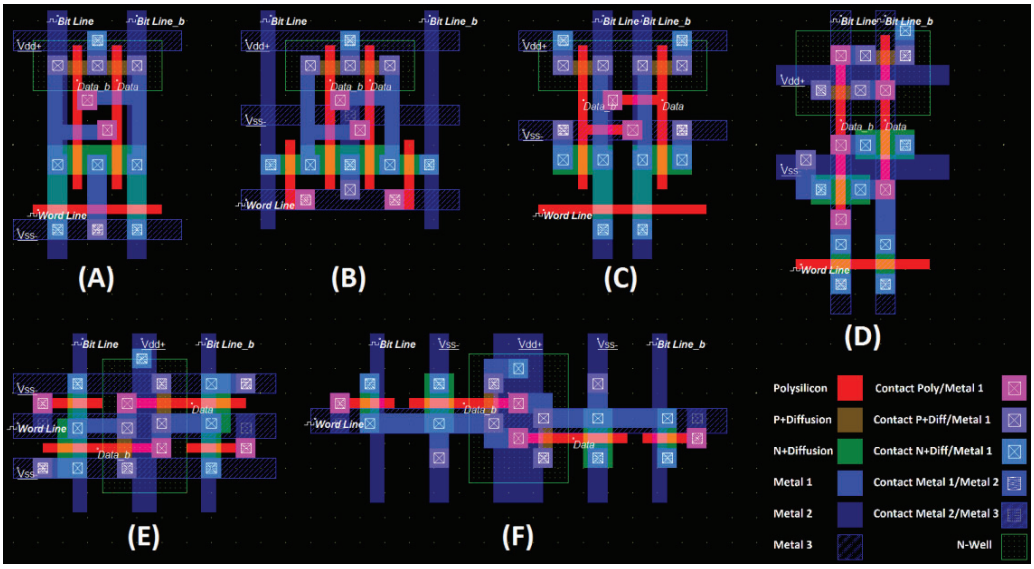

Fig. 2. Layout of Type 1a (A), Type 1b (B), Type 2 (C),Type 3 (D), Type 4 (E) and Type 5 (F) SRAM cells.

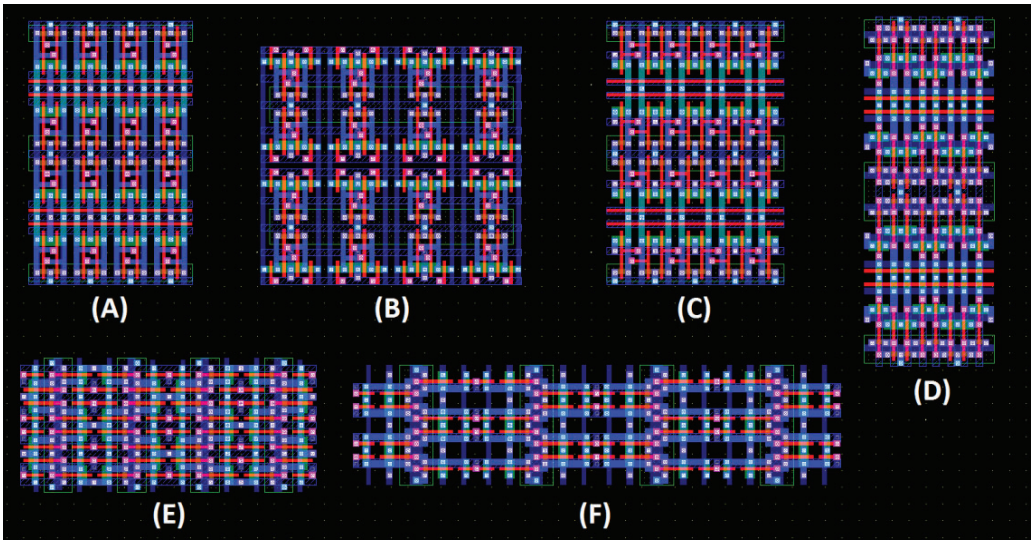

Fig. 3. Layout of Type 1a (A), Type 1b (B), Type 2 (C),Type 3 (D), Type 4 (E) and Type 5 (F) 16-bit SRAM memory array.

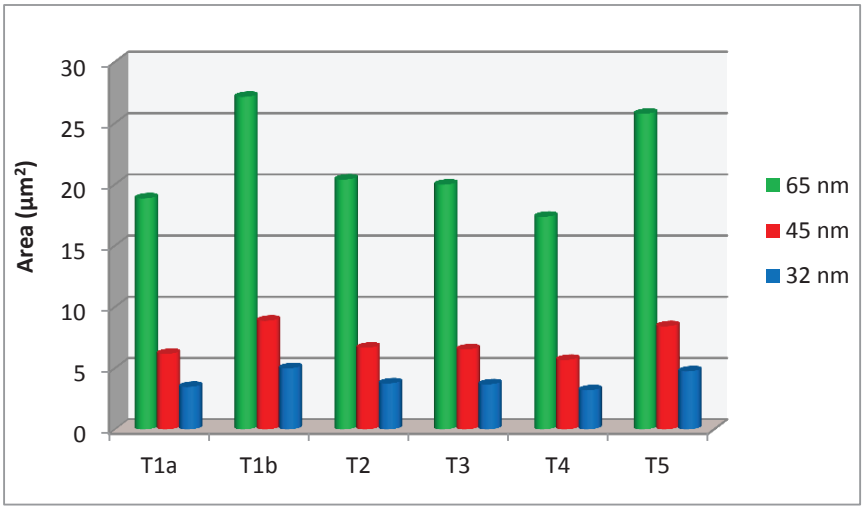

Fig. 4. Area of 16 bit SRAM arrays

C. Results

Regarding the single cell simulations, the results we obtained present little deviation among different designs, since the 6T SRAM cell is a small circuit and all cells are identical at the transistor level. In addition, read delay strongly depends on the sensing method that is used, which was the same in all cases. Nonetheless, it can be assumed that the T4 cell performs best in terms of power dissipation (except for 65 nm where it ranks second) and write delay. This can be attributed to its compact design with small wire and diffusion capacitances. An important thing to note is that read/write delay is hardly affected with downscaling while the power dissipation drops significantly from 65 to 45 and to 32 nm. The cell simulation results for read delay, write delay and power dissipation are shown in Fig. 5, 6 and 7, respectively.

A more reliable comparison can be derived from the 16-bit array simulations, where the results seem to vary a lot among SRAM types and relative to scaling. Hence, the ranking from best to worst in terms of power dissipation is: T2, T1a, T3, T5, T4, T1b for 65 nm, T2, T3, T1a, T5, T4, T1b for 45 nm and T4, T2, T1a, T1b, T3, T5 for 32 nm. The T2 array is the best at 65 and 45 nm and second best at 32 nm, proving to be a power-efficient layout design in all cases. The T4 array is the best at 32 nm but performs poorly at 65 and 45 nm, being fifth in rank. The T5 array performs better than T4 at 65 and 45 nm, but overall worse than most conventional designs, thus being ranked fourth at 65 and 45 nm and last at 32 nm. The array simulation results are shown in Fig. 8.

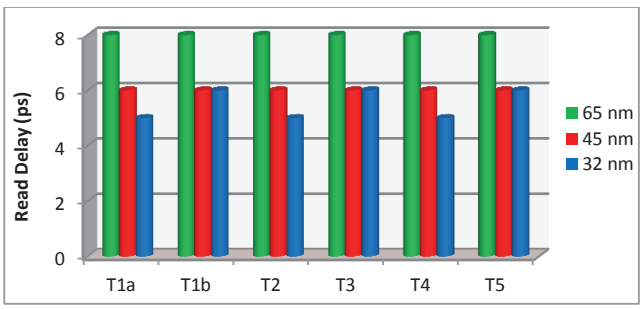

Fig. 5. Read delay of SRAM cells

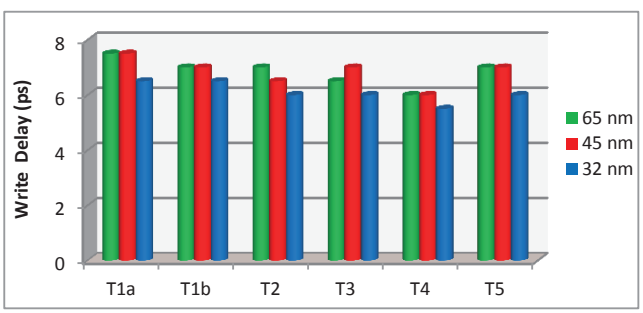

Fig. 6. Write delay of SRAM cells

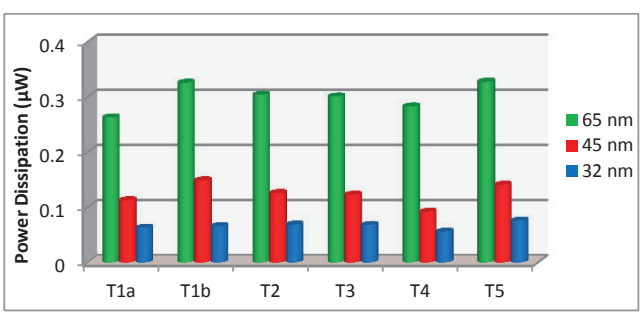

Fig. 7. Power dissipation of SRAM cells

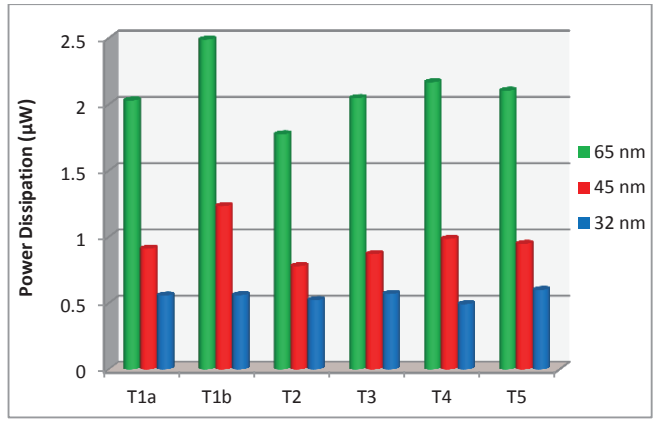

Fig. 8. Power dissipation of 16 bit SRAM arrays

V. Conclusions

Various types of 6T SRAM cell layout architectures and corresponding 4X4 16-bit arrays have been implemented and compared at the 65, 45 and 32 nm, in terms of area efficiency and simulation performance. The T4 cell seems to be the most viable layout topology for further development, since it seems to get comparatively better with downscaling. It presented the best overall performance in terms of read/write delay, the lowest power dissipation at 32 nm and the highest area/bit density efficiency. The recently proposed T5 cell, even though it provides a more lithographically friendly alternative to the T4, introduces a significant penalty in area and performance relative to most conventional designs, and seems to perform worse with downscaling.

References

[1] B.H.Calhoun, Yu Cao, Xin Li, Ken Mai, L.T. Pileggi, R.A.Rutenbar, K.L.Shepard, "Digital circuit design challenges and opportunities in the era of nanoscale CMOS," Proceedings of the IEEE, vol. 96, issue 2, February 2008, pp. 343–365.

[2] Neil HE Weste, David Money Harris, CMOS VLSI design: a circuits and systems perspective, Addison-Wesley, fourth edition, 2011.

[3] M.Ishida, T.Kawakami, A.Tsuji, N.Kawamoto, M.Motoyoshi, N.Ouchi, "A novel 6T-SRAM cell technology designed with rectangular patterns scalable beyond 0.18 um generation and desirable for ultra high speed operation," IEEE Int. Electron Devices Meet. (1998) 201-204.

[4] M.Woo, et al, "A High Performance 3.97pm2 CMOS SRAM Technology Using Self-Aligned Local Interconnect and Copper Interconnect Metallization," Symp. on VLSI Tech., p.12 (1998).

[5] Y.Takao, et al, "A 4-μm^2 Full-CMOS SRAM Cell Technology for 0.2-μm High Performance Logic LSIs," Symp. on VLSI Tech., p.1 I (1997).

[6] M. Helm, et al, "A Low Cost, Microprocessor Compatible, 18.4 μm^2, 6-T Bulk Cell Technology for High Speed SRAMs," Symp. on VLSI Tech., p.65 (1993).

[7] Y.Sambonsugi, T.Maruyama, K. Yano, H.Sakaue, H.Yamamoto, E. Kawamura, S.Ohkubo, Y.Tamura, T.Sugii, "A Perfect Process Compatible 2.491 μm^2 Embedded SRAM Cell Technology for 0.13 μm Generation CMOS Logic LSIs," Symp. on VLSI Tech., p.62 (1998).

[8] K.Noda, et al, "A 2.9μm^2 Embedded SRAM Cell with Co-Salicide 847 Direct-Strap Technology for 0.18μm High Performance CMOS Logic," IEDM Tech. Dig., p.847 (1997).

[9] K. Osada et al., "Universal-VDD 0.65-2.0-V 32-kB cache using a voltage-adapted timing-generation scheme and a lithographically symmetrical cell," JSSC, vol. 36, no. 11, Nov. 2001, pp. 1738–1744.

[10] M.Khare et al., "A high performance 90nm SOI technology with 0.992 mm^2 6T-SRAMcell," Proc. Intl. Electron Devices Meeting, 2002, pp. 407–410.

[11] R.W.Mann and B.H.Calhoun, "New category of ultra-thin notchless 6T SRAM cell layout topologies for sub-22 nm," Proceedings of the International Symposium on Quality Electronic Design, pp. 1–6, 2011.

[12] E.Grossar, M.Stucchi, K.Maex, W.Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," IEEE Journal of Solid-State Circuits 41 (11) (2006) 2577–2581.

UMC 1449
UMC v. AICP

View publication stats