

An Extra-Regular, Compact, Low-Power Multiplier Design Using Triple-Expansion Schemes and Borrow Parallel Counter Circuits

Rong Lin, Department of Computer Science, SUNY-Geneseo, Geneseo, NY 14454

Ronald B. Alonzo, Department of ECE, University of Rochester, Rochester, NY 14627

Abstract This paper presents an extra regular, complexity-reduced, high-performance pipelined multiplier architecture, using newly proposed multiplier triple-expansion schemes. It is based on a parallel counter circuitry, called borrow parallel counter, which utilizes 4-b 1-hot encoded signals and borrow bits, i.e. bits weighted 2, as building blocks for a compact, low-power, large parallel multiplier implementation. Exemplified by a 54 x 54-bit multiplier, the scheme allows large multipliers to be generated from smaller multipliers, tripling the size in each expansion (6 x 6-bit to 18 x 18-bit to 54 x 54-bit). This significantly reduces the complexity of state of the art designs and achieves full-self-testability without sacrificing high-performance. The preliminary circuit layouts and simulations have demonstrated and confirmed the potential of the new architecture.

1. Introduction

The high-speed large, say $n \times n$ -bit for $n > 32$, pipelined high-performance multiplier designs, often have the following major disadvantage: they are (Booth [11] and non-Booth) constructed based on the schemes of generation and reduction of a single large partial product bit matrix, usually with Wallace tree structure [10] processing in parallel. The schemes are intrinsically irregular and not exhaustively self-testable (requiring built-in test circuits). This is due to that the initial partial product bit matrix has a triangle (or trapezoid) shape, and the multiplier circuits have low controllability and observability for test, particularly for the most commonly used Booth multipliers. The area cost, power cost, layout cost and the test cost in dealing with the irregularity are significant.

The functions of conventional multipliers are divided into three stages. The first stage is the generation stage of the partial products, followed by the adding stage of the partial products. The last stage is the final addition. Since the last stage usually employs a standard fast adder, it is often excluded in the discussion.

Two recently proposed designs, as the typical examples of the improved conventional architectures, are rectangular-styled wallace tree multiplier (RSWM for short) [2] and limited switch dynamic logic multiplier (LSDL) [1].

The RSWM design proposes a rectangular Wallace-tree construction method. In this method, the partial products are divided into two groups and added in the opposite direction. The partial products in the first group are added downward, and the partial products in the second group are added upward. This method eliminates the dead area that occurs in a general Wallace tree design. It also optimized the carry propagation between the two groups to realize high speed and a simple layout. Applying the method to a 54 x 54-bit multiplier, a 980 $\mu\text{m} \times 1000 \mu\text{m}$ (0.98 mm^2) area size and a 600-MHz clock speed have been achieved using 0.18 μm CMOS technology.

The LSDL design proposes a method merging precharged dynamic logic into the input of every latch, which differs circuits merging logic and latches described in [3]. In [3], clocks are used to tri-state the output of a static logic gate, while in LSDL clocks are used to control precharge and evaluation phases of dynamic logic and latch the outputs. It allows most of speed advantages of dynamic logic to be preserved while eliminating most of the traditional dynamic logic power penalty. The LSDL design achieves a 2.2 GHz 53 x 54 pipelined multiplier, fabricated in 0.13 μm CMOS technology with an area of 315 $\mu\text{m} \times 495 \mu\text{m}$ (0.155 mm^2) which reduces the area required by RSWM design by 50% (scaled for technology) also ceases the operation frequency.

Both RSWM and LSDL are Booth encoded Wallace tree design and have yielded multipliers with great performance and cost reduction in terms of area or area-power. However, the design complexity in both cases are increased accordingly.

RSWM design uses high-speed redundant binary (RB) architecture [3], and complex optimization process as well as extra area for carry-signal propagation to add upward partial products in the lower-bit group. While LSDL design requires well controlled dynamic circuit and clock design with proper pulses, long enough for evaluation of dynamic logic

— The work was supported, in part, by National Science Foundation under grant CCR 0073469; and by NYSTAR, (MDC) under grant 02332105.

and short enough to prevent significant leakage on the dynamic node. Furthermore, the design requires relative expensive custom process in laying out of the most of its circuits. Finally, a building test circuitry is required for both design.

The new multiplier triple-expansion scheme [5] adopted in this design is developed based on the previous work [4, 6], with improved performance through using a novel partial product bit matrix decomposition method and extra compact, low-power large parallel counter circuits. Comparing with conventional large Booth multipliers, it is highly regular, compact in layout, exhaustively testable without an extra built-in test circuit.

The decomposition and re-arrangement of the bit matrices provided by this new scheme significantly reduces the number of recursive levels required for the construction of large multipliers. Actually no more than two levels of recursive circuits are needed for efficient construction of a multiplier including, 36x36-b, 54x54-b, 64x64-b and 72x72-b. The building block of the proposed multiplier is a novel CMOS parallel counter circuitry, utilizing 4-b 1-hot encoded signals, and borrow bits, i.e. bits weighted 2. The borrow parallel counter circuits greatly simplify the structures of small multipliers (as a single array of almost identical counters), improve the compactness and effective circuit layout. The circuits style contributes significantly to the efficient implementation of the triple expanded multipliers.

Based on the preliminary layouts and simulations, the proposed 54x54-b pipelined multiplier could be implemented in an area of $434.8 \times 769.5 = 334578.6 \mu\text{m}^2$ with a 0.18 nm technology, achieving a 1GHz at 1.8V supply and a good low-power performance. The area is 37.9% of the area of RSWM design, or 75.8% of the LSDL area (scaled for technology).

Our experimental work (with Cadence Analog Affirma tools) has preliminarily demonstrated all the unique features of the design. The triple-expansion schemes and the new counter circuits based pipelined multiplier architecture have shown the potential and interest to the new generation of complexity-effective microprocessor designs.

2. The 18 x 18 multiplier

Figures 1 and 2 illustrate how an 18 x 18-b virtual multiplier, meaning two instead of one output numbers being produced, is constructed by nine 6 x 6-b small multipliers and five adders, using a trisect decomposition approach. Two 18-b input numbers are first trisected, partitioned, and distributed to nine 6 x 6-b multipliers, where the 6 x 6 partial product matrices are generated and the nine 12-b products are produced. The weighted bits of the nine products are then added by the

adders. The weight range of each bit group received by the adders are indicated by a number, 1 to 5, at the top of each adder (or receiver) block (see Figure 1). Adder-3a adds three 6-bit numbers to result in the final sum's bits 6 to 11 (and carries to adder-5a). Adder-5a then adds five 6-bit numbers (and the carry-ins) to result in the final sum's bits 12 to 17 (and carries to adder-5b). Similarly, adder-5b (adding 5 6-b numbers) and adder-3b (adding 3 6-b numbers) result in final sum's bits 18 to 23 and bits 24 to 29 respectively. The carry-out bits from adder-5b will be added by the last adder, adder-c, to result in the 6 MSBs. Usually no addition is needed for the output bits 0 to 5. All 36 bits of the product have been correctly produced.

Since small multipliers are independent of receiving inputs (trisected segments of the input numbers) and doing multiplications, they can be re-arranged to minimize the interconnection between the small multipliers and the CSAs (with 2 levels of 3:2 and 4:2 counters plus a latch for each output bit). The block re-positioned such a triple-expanded 18 x 18 multiplier schematic, along with its inputs distribution, is illustrated in Figure 2. The inputs, two 18-b numbers J and K, are trisected into segments: a, b, c and x, y, z, each of 6 bits. They are distributed to the 18 blocks. The final addition is not needed when it serves as a virtual multiplier for constructing a 54 x 54 larger multiplier.

3. The 54 x 54 multiplier

When the circuit scheme is applied recursively for one more level, it results in the 54 x 54-b multiplier (see Figure 3). It comprises nine 18 x 18-b triple-expanded virtual multipliers and a level of CSAs adders (called level-2 CSAs), which is a row of 2 levels of binary (4, 2) and (6, 2) counters plus latches, residing at the bottom of the 54 x 54-b multipliers. The outputs (two-number pairs) of the CSA adders are sent to the fast final adder which is not included. It requires three stages of pipelined operations: (1) base (i.e. 6x6-b) virtual multiplication, (2) level-1 (i.e. 18x18-b) bit reduction, and (3) level-2 bit reduction. Since all these three operations require comparable delays, the scheme fits well for a 3-stage (or 3.5-stage) pipelining and multiply-accumulate implementations. Two output numbers of each 18x18 multiplier are routed to the CSAs in parallel, passing through zero or three or six rows of 6x6 multipliers. Since the height of each 6x6 is made as short as possible (see Figure 10), the interconnection distance is minimized (Figures 2 3).

Any efficient small multipliers may be considered as bases for the triple expansion to yield large multipliers. We have adopted two types of 6x6 multipliers as shown in Figures 10 and 11 respectively in the preliminary simulation study. The first is tiny (3,2)-(4,2) counter (ref. to [6]) based Wallace-tree style multiplier. The

second is a borrow parallel small multiplier which is basically a single array of borrow parallel counters, The counter circuits will be described in detail in the next section. Both multipliers receive two 6-bit input numbers, J and K, and generate a small partial product bit matrix and then reduce it into two numbers P (p10-p0) and Q (q10-q5), so that $J*K=P+ Q*2^5$. The (4,2)-(3,2) based 6x6 multiplier uses slightly fewer transistors, while the borrow parallel 6x6 multiplier has a more compact layout and mainly performs logic with 4b-1-hot signals that feature lower switching activity and use fewer hot lines.

4. The 4-b 1-hot borrow parallel counters

The parallel counter circuits utilizing 4-b (bit) 1-hot, or non-binary encoded signals, i.e. three 0 bits plus one 1 bit representation of integers ranging from 0 to 3 (ref. to Table 1), are interesting and superior in several aspects (including speed and power), compared with traditional binary counters for multiplier designs [4-6]. However, in order to reduce 7 bits into 3 or 2 bits, the previously proposed circuits require 8 to 10 additional transistors for signal type conversion, from non-binary to binary. The new family of circuits, called borrow parallel counters including 5_1, 5_1_1, 6_1, 6_0, does not need type conversion, and requires a minimum number of transistors with a large ratio of nMOS/pMOS, and yet shows superior layout and performance. As shown in Figures 4 and 5, the counter not only utilizes both 4-b 1-hot signal encoding and borrow bits, i.e. input bits weighted 2 instead of 1, but also provides an embedded full adder adding non-binary (4-b, 1-hot) and binary signals without type conversion, This leads to a significant reduction in circuit complexity. The circuit has a potential to become a new type of building block, replacing traditional (2, 2), (3, 2), i.e. half-adder, full-adder, and (4, 2) parallel counters for some arithmetic processor designs.

The parallel counter shown in Figure 4 is designated 5_1 borrow parallel counter, with 5 input bits A1-A5, and bit A5 weighted 2 (note that its input arrow in the symbol is offset). The new circuit and its variants possess the following three features: (1) Each counter reduces 5 or 6 input bits (one or two of them are borrowed bits) into 2 output bits (with a few in-stage carry in and out bits) in high speed. (2) The majority of the transistors are gated by, or used to pass, 4-b 1-hot signals (ref. to Figure 6), which leads to the reduction of both switching activities and the flow of hot signals by about a half of the normal [4-6]. (3) The ratio of nMOS/pMOS is 2.4 (instead of 1 for traditional CMOS) and a compact layout can be achieved easily.

Table 1 shows the 4-b 1-hot encoding scheme. The unique bit positions determine the values of a 4-b 1-hot signal. The change of R value from one to another causes the change of bit-values in no more than 2 lines,

which reduces switching activity of the circuit. Also at any logic stage there is only one hot bit on four signal lines, which reduces static leakage power. Figure 5 shows a full adder circuit which adds three bits s0, s1 and q, represented by two 4-b 1-hot signals and a binary signal without type conversion. The components and the typical application of the 5_1 counters are shown in Figures 7 to 9.

The 5_1 parallel counter comprises seven components (refer to Figures 4, 6 and 7): (1) The 4-b 1-hot signal encoder which encodes $(A1+A2+A3+A4) \bmod 4$ into $R=s0'+2s1'$ (intermediate results s0' and s1' not shown); (2) Adding-A5 that adds Xi, s1' and A5. Note that $s0+A5 \bmod 2=s0$; no change for s0 is one of advantages of using borrow bits; (3) Q-generator that generates $q= (A1+A2+A3+A4+ 2A5)/4$; (4) R-restoration (or R res) that restores non-full swing 4-b 1-hot signal R into a full swing one; (5), (6), and (7) Three stages (components) of the embedded full adder circuit as detailed in Figures 5 and 7 to 9. Each 5_1 counter co-works with its upper and lower neighbor 5_1 counters, as shown in Figure 8, to produce two output bits S and C. Because s0, s1, and q within each counter are weighted 1, 2, and 4 respectively. the actual s0, s1, and q being added by the full adder are from three adjacent columns with s0 in the highest column, thus they have the same weight. There is no explicit data type conversion and the output is in binary form.

The circuit simulations have shown the superiority of the new counters, comparing with the conventional ones in all aspects including delay, area, and power dissipation. This will be clearer when the circuits are applied in small multiplier designs. The 5_1 counter uses 78 transistors, about two third are nMOS cells, and 56 out of 78 (73%) transistors are either gated by, or used to pass, 4-b 1-hot signals, leading to a significant reduction in power-consuming activities. The counter implements arithmetic equation E1 and five logic equations below:

$$A1 + A2 + A3 + A4 + 2A5 = s0 + 2s1 + 4Q \quad (E1)$$

$$Xo=s0; \quad Yo=Xi \text{ xor } s1; \quad Zo=Xi; \quad S = Yi \text{ xor } Q;$$

$$C= Zi \text{ and } Y\bar{i} \text{ or } Q \text{ and } Yi.$$

Here s0, s1, Q are temporary parameters, and Xo, Yo, Zo and Xi, Yi, Zi are in-stage carry (out/in) bits. The close variants of the 5_1 counter are denoted by 5_1_1, 6_1 and 6_0, which are similar to 5_1, except the number of borrow bits and the component for encoding those bits are slightly different. There is little change in complexity between 5_1 and 5_1_1 as well as between 6_1 and 6_0. The main application of the proposed borrow counters is, in a unique way, to reduce in parallel the height of a weighted bit matrix, with significant

new features which is interesting to efficient VLSI implementations of arithmetic circuit designs.

Borrow parallel counts may be used for efficient partial product bit reduction for large (say 32b or larger) multiplier designs. For example, a 96 transistor 6-1 borrow parallel counter (two output buffers may not be needed) can replace 4 full adders or two (4, 2) counters, possessing all advantages as described above without an increase in circuit transistor count. The simulation results for 5-1 and 5-1-1 borrow parallel counters are given in Table 2.

5. The base 6 x 6 borrow parallel multipliers

As a building block, the 6x6-b borrow parallel (virtual) multiplier shown in Figure 11 produces 17 output bits, or two numbers instead of one. Such an output form has two advantages: First, it is fast. When the 7 LSBs outputs are produced (through a ripple carry style process) the second 10 MSBs outputs are about ready (through carry save process). Second, it is good for regular inter-connection and CSA bit reduction; refer to Figures 2 and 3, the two output groups of each base 6x6 block are accurately separated with lower weighted group as a 6-b number, while the higher weighted group as two 5-b numbers.

The multiplier is an array with five borrow parallel counters. The small borrow parallel multiplier possesses the following new features compared with conventional binary full-adder based counterparts: (1) It is basically a single array of identical counters with a simple layout, since the "borrow-effect" naturally re-arranges the bits being processed so that the actual bits to each column are balanced. (2) It requires minimal line connections, since basically only a single counter is used in each column. (3) It gives the nearly same delay for almost all output bits, except a few faster outputs at two ends, therefore little cost is needed in transistor sizing and delay equalization. The delay of the circuit of Figure 11 is about 0.7ns or 2 times a (4, 2) delay. Table 2 shows the summary and comparisons of the small multiplier circuits .

6. The experimental work and concluding remarks

The layouts of the 5-1 (and 5-1-1) counters and the 6 x 6 multiplier in 180 μm CMOS technology (3 metal layers) with areas of $12.87 \times 16.0 \mu\text{m}^2$ and $26.5 \times 85.5 \mu\text{m}^2$ respectively have been experimentally made (Figures 12 and 13). The current studies are focusing on design of two (level-1 and level-2) CSA blocks as shown in Figures 2 and 3. Our preliminary results have shown that both blocks are basically regular structured and can be layout with a straightforward simplicity. The size (including output latches) of level-1 block is estimated as $34.2 \times 85.5 \times 3 \mu\text{m}^2$. The size of level-2 block is

about $48.7 \times 85.5 \times 9 \mu\text{m}^2$. The overall pipelined 54 x 54 multiplier can be layout (4-metal-layer) in a rectangular area with a height of $((26.5 + 5) \times 3 + 34.2) \times 3 + 48.7 = 434.8 \mu\text{m}$, and a width of $85.5 \times 9 = 769.5 \mu\text{m}$, or area of $434.8 \times 769.5 = 334578.6 \mu\text{m}^2$. The area is about 37.9% of the area $882000 \mu\text{m}^2$ of RSWM multiplier [2] (excluding the final adder about 10% of the total area of $980000 \mu\text{m}^2$), or 75.8% of the area of LSDL multiplier [1] (scaled for technology).

The complexity reduction of the design can be seen from the high regularity of the multiplier logic scheme. Eighty one identical 6 x 6 small multipliers, serving as building blocks, are simply organized in a 9 x 9 matrix form. The nine identical level-1 CSA adder blocks plus a single level-2 CSA block require minimal custom design work load for optimal layouts. It is quite easy to organize the inputs routine network and three levels pipeline interconnection nets in Manhattan cells structure.

The advantages of the design in terms of complexity-effectiveness, compared with the designs of RSWM [2] and LSDL [1] may include (1) simpler CMOS technology and layout; (2) significantly less amount of custom design work load; (3) significant area reduction without sacrificing high-performance: an expected pipeline frequency of 1GHz can be achieved; (4) low-power achieved through using the compact 4-b 1-hot counter circuitry; (5) modular and repeated components; (6) self-testable: It is directly provided by the triple expansion logic scheme. The regular decomposition of partial product bit matrix enables the circuit possessing high controllability and observability for test, without using a built-in circuit. Exhaustive test can be done by testing 81 6 x 6 small multipliers separately, along with 9 level-1 CSA adder blocks and the level-2 adder block. The test vector length is practically feasible (refer to algorithm of [7]).

REFERENCES

- [1] Robert Montoye, Wendy Belluomini, Hung Ngo, Chandler McDowell, Jun Sawada, Tuyet Nguyen, Brian Veraa, James Wagoner, Mike Lee, A double precision floating point multiplier. Proc. of 2003 *IEEE ISSCC*, February, 2003.
- [2] N. Itoh, Y. Naemura, H. Makino, Y. Nakase, T. Yushihara, Y. Horiba, A 600MHz, 54 x 54-bit multiplier with Rectangular-Styleled Wallace Tree, *IEEE JSSCs*, Vol 35, No2, February 2001.
- [3] Daniel W. Dobberpuhl, Richard T. Witek, Randy Allmon, Robert Anglin, David Bertucci,...Sharon Britton, Linda Chao, Robert A. Conrad, Daniel E.

Dever, Bruce Gieseke, Soha M. N. Hassoun, Gregory W. Hoepfner, Kathryn Kuchler, Maureen Ladd, Burton M. Leary, Liam Madden, Edward J. McLellan, Derrick R. Meyer, James Montanaro, Donald A. Priore, Vidya Rajagopalan, Sridhar Samudrala, and Sribalan Santhanam, A 200-MHz 64-b Dual-Issue CMOS Microprocessor, *IEEE JSSCs*, Vol 27, No11, November 1992.

- [4] R. Lin, Reconfigurable Parallel Inner Product Processor Architectures, *IEEE T LSI*, Vol. 9, No 2. April, 2001. pp. 261-272.
- [5] R. Lin, Borrow parallel counters and borrow parallel small multipliers and Triple-expanded multipliers. New Tech. Disclosures of SUNY, August, 2002 (US Patent pendings).
- [6] R. Lin, Low-Power High-Performance Non-Binary CMOS Arithmetic Circuits, in *Proc. of 2000 IEEE Workshop on SIGNAL PROCESSING SYSTEMS (SiPS)*, Lafayette, Louisiana, October, 2000. pp. 477-486.

- [7] R. Lin and M. Margala, Novel design and verification of a 16 x 16-b self-repairable reconfigurable inner product processor, in *Proc. of 12th Great Lakes Symposium on VLSI*, NYC, April, 2002.
- [8] C. F. Law, S. S. Rofail, and K. S. Yeo, "A Low-Power 16x16-b Parallel Multiplier Utilizing Pass-Transistor Logic", *IEEE J. of Solid-State Circuits*, vol. 34, no. 10, pp. 1395-1399, Oct 1999.
- [9] G. Goto, A. Inoue, R. Ohe, S. Kashwakura, S. Mitarai, T. Tsuru, and T. Izawa, A 4.1-ns compact 54x54-b multiplier utilizing sign-select Booth encoders, *IEEE Journal of Solid-State Circuits*, Vol. 32; No 11, Nov 1997.
- [10] C. S. Wallace, A suggestion for a fast multiplier, *IEEE Trans. Electronic Computers*, Vol. Ec-13, 1964, pp. 14-17.
- [11] A. D. Booth, A signed binary multiplication technique, *Quart. J. Mech. Appl. Math.*, vol 4, 1951.

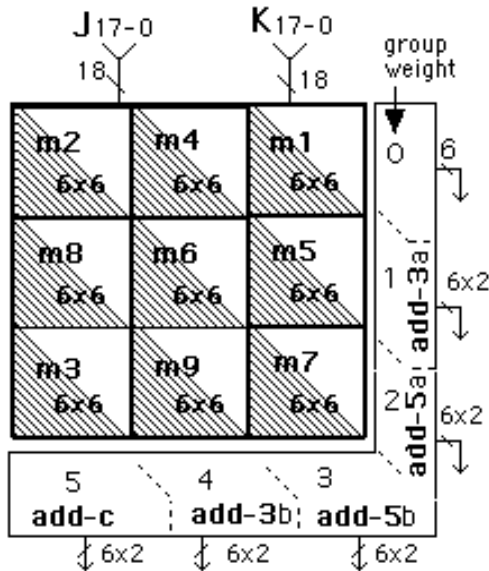


Figure 1. The trisect-decomposing an 18x18 product partial matrix.

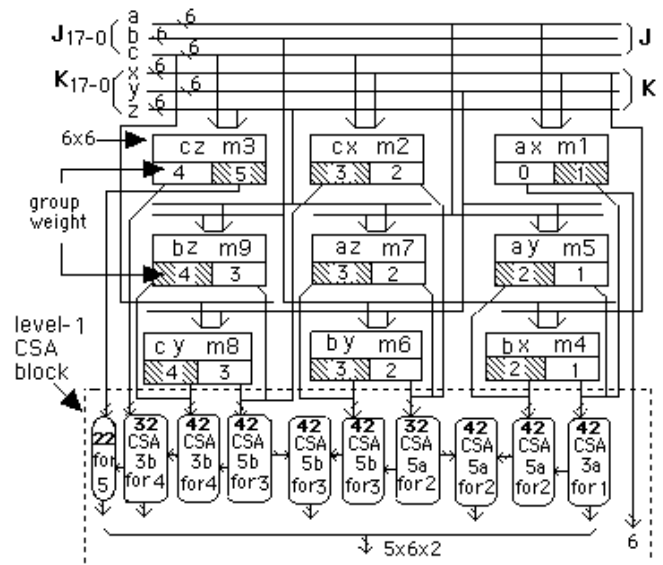


Figure 2. The triple-expanded 18x18-b multiplier (the implementation of Figure 1 with CSA outputs).

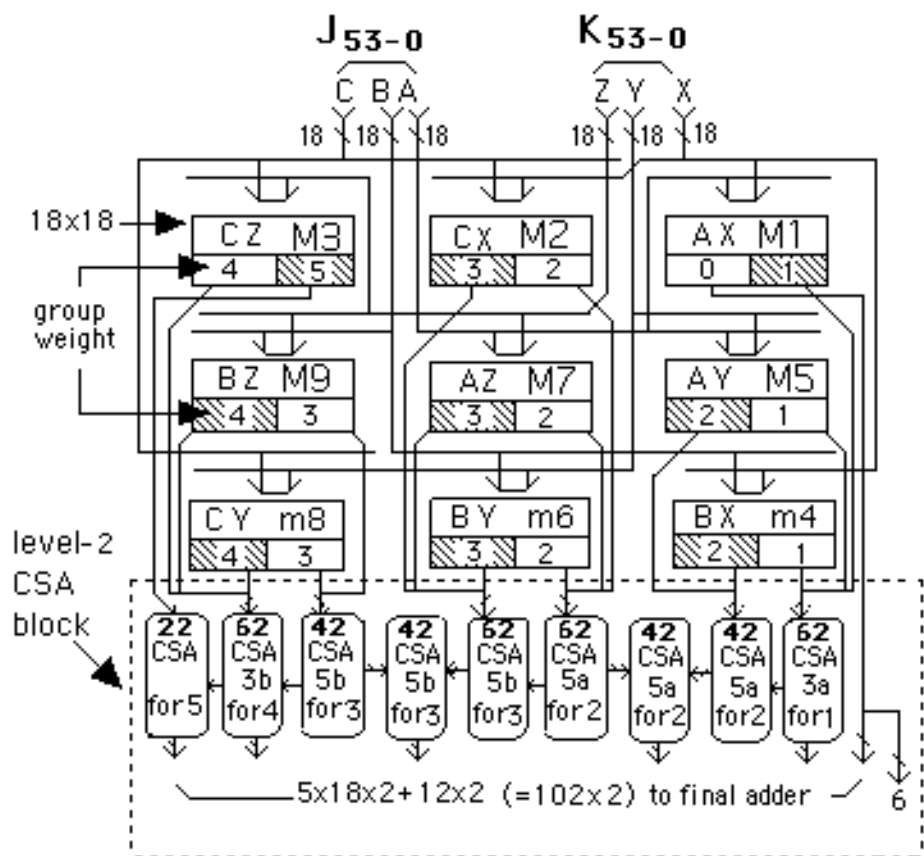


Figure 3. The triple-expanded 54x54 multiplier block diagram.

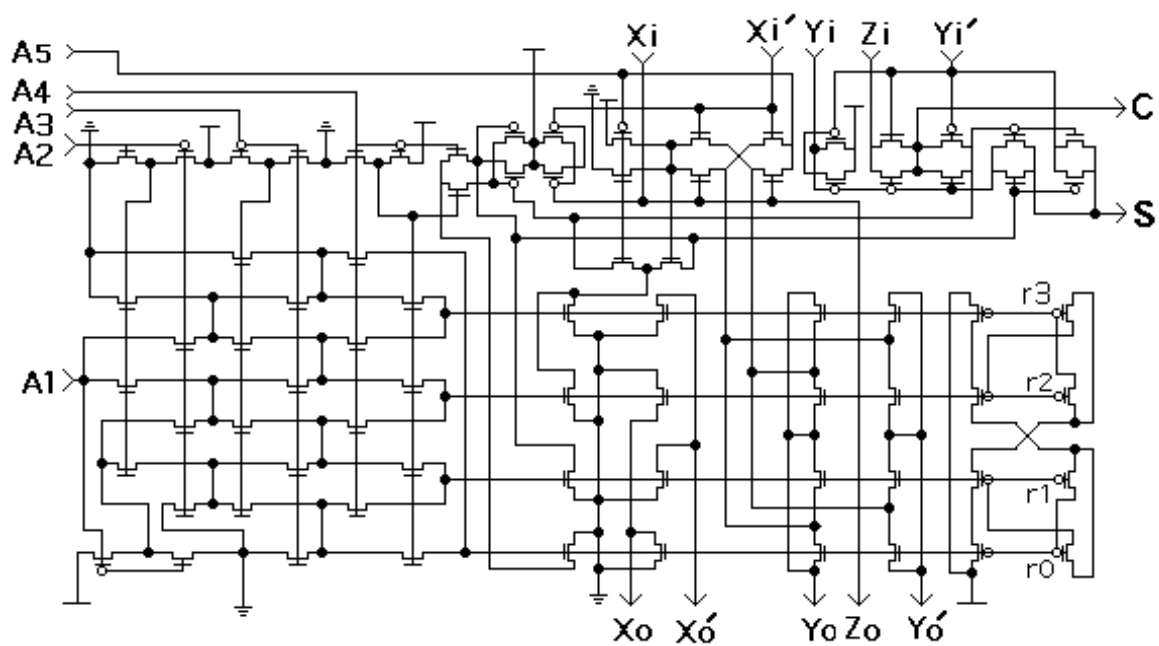


Figure 4. The 5_1 borrow parallel counter. $R=(r_0 r_1 r_2 r_3)$ is a 4-b 1-hot encoded signal. A_5 is a borrow bit, X_i/X_o , Y_i/Y_o , Z_i/Z_o are in-stage in/out bits.

Table 1. The 4-b 1-hot signal encoding

r_3 $R =$ r_2 r_1 r_0	0 0 0 1	0 0 1 0	0 1 0 0	1 0 0 0
decimal value of R	0	1	2	3
binary value of R = s1s0	00	01	10	11
binary value of s0 (encoded by R)	0	1	0	1
binary value of s1 (encoded by R)	0	0	1	1

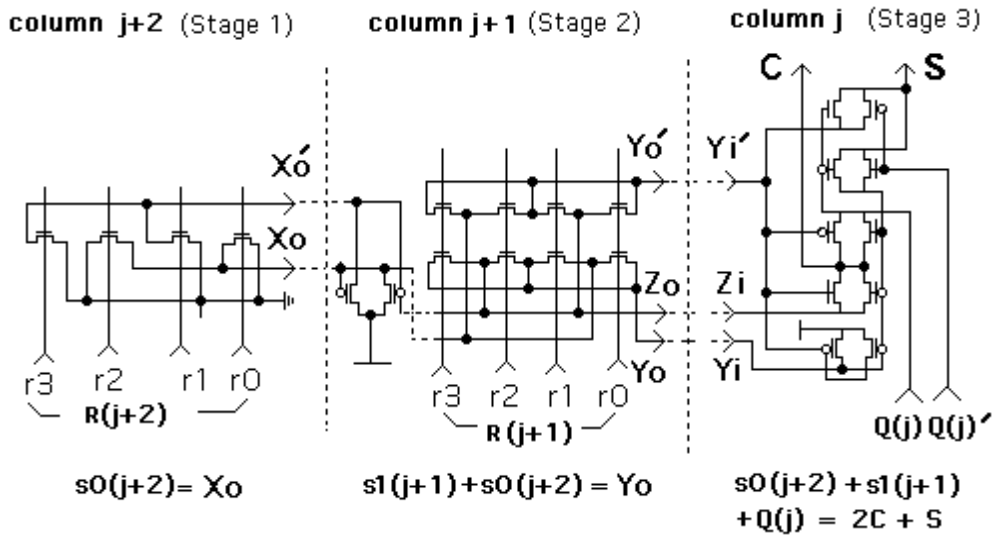


Figure 5. The full adder adding three bits, one binary and two 4-b 1-hot encoded bits, without a type conversion.

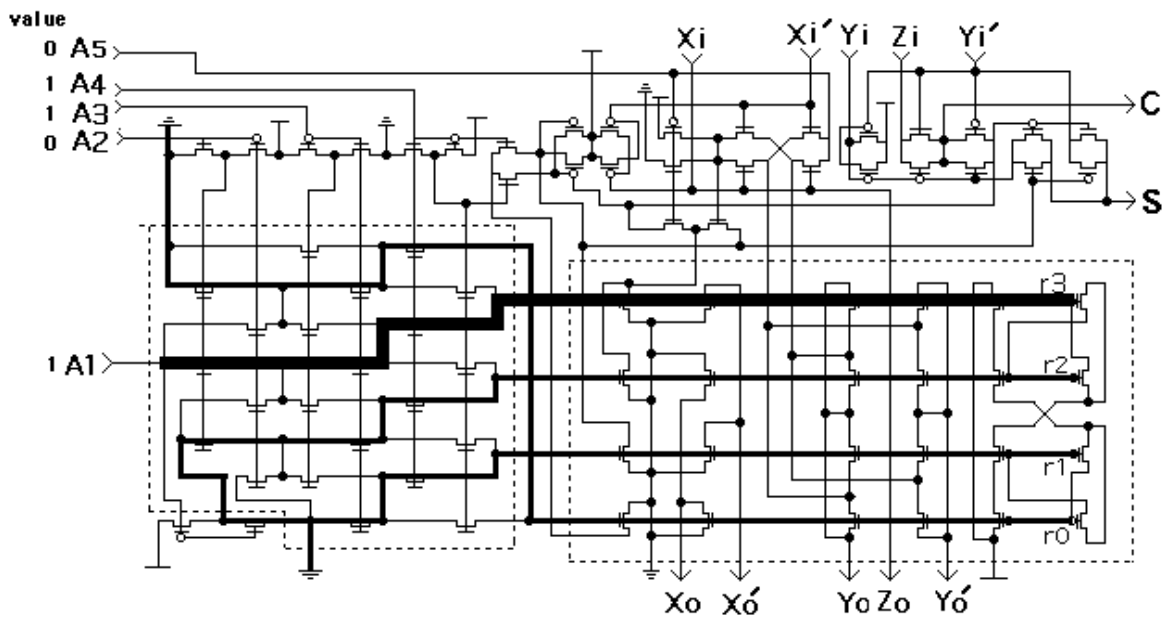


Figure 6. The low-power features of the 5-1 borrow parallel counter. The bold lines show the 4-b 1-hot signal. The double bold line is for the 1-hot bit. The transistors in right (left) dotted box are gated by, (used to pass) the 4-b 1-hot signal, which reduces switching activities and leakage.

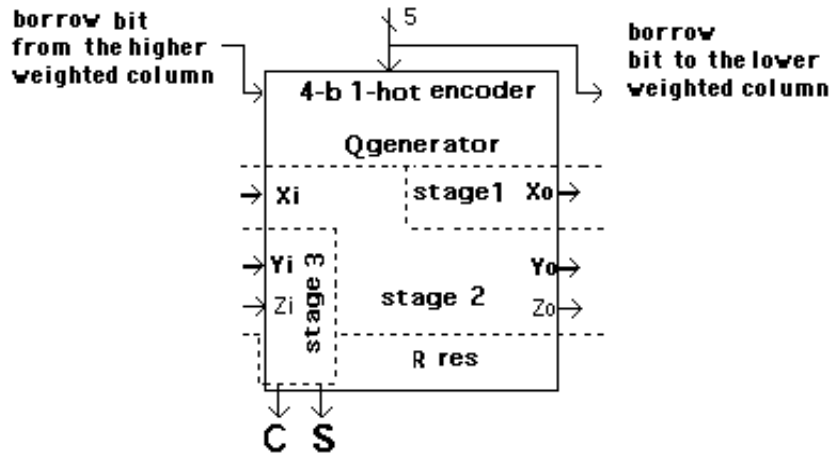


Figure 7. The functional structure of the 5₁ parallel counter.

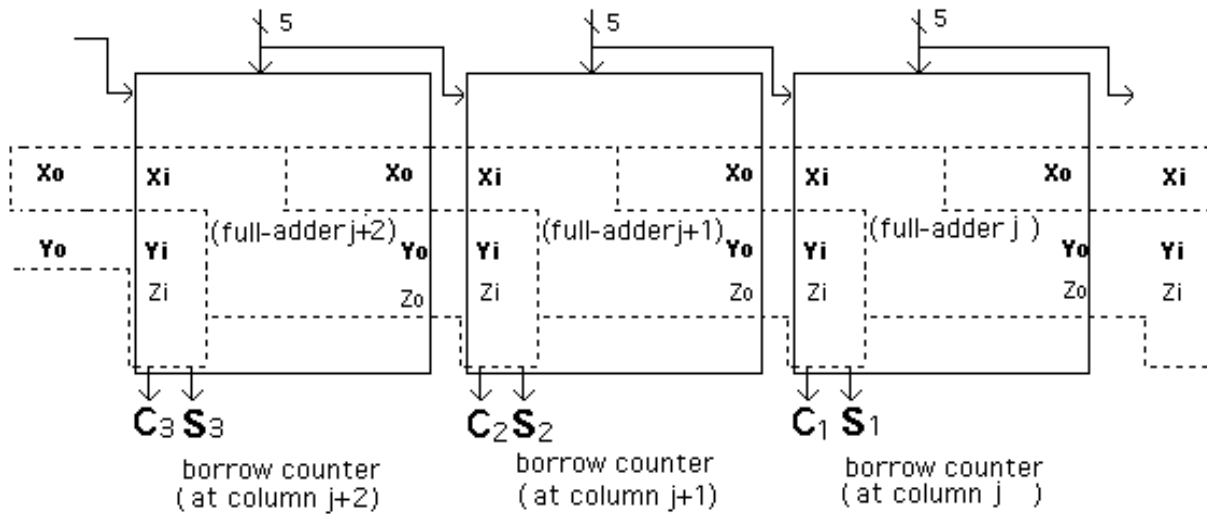


Figure 8. A typical application of 5₁ counter array.

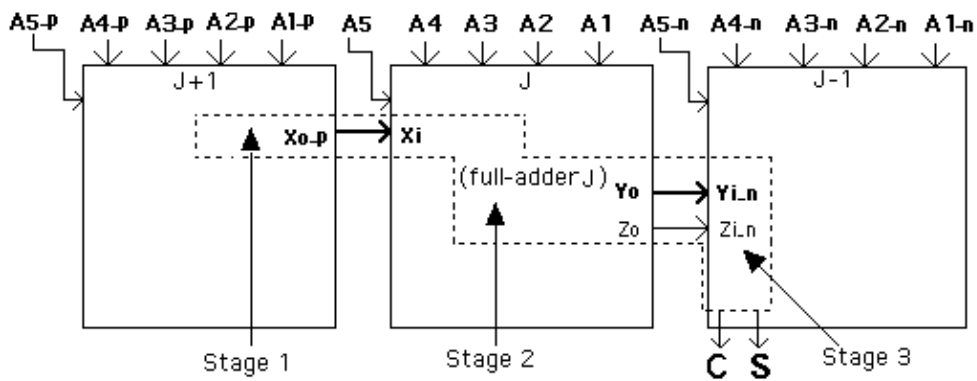


Figure 9. A full-adder embedded in three contiguous borrow parallel counters.

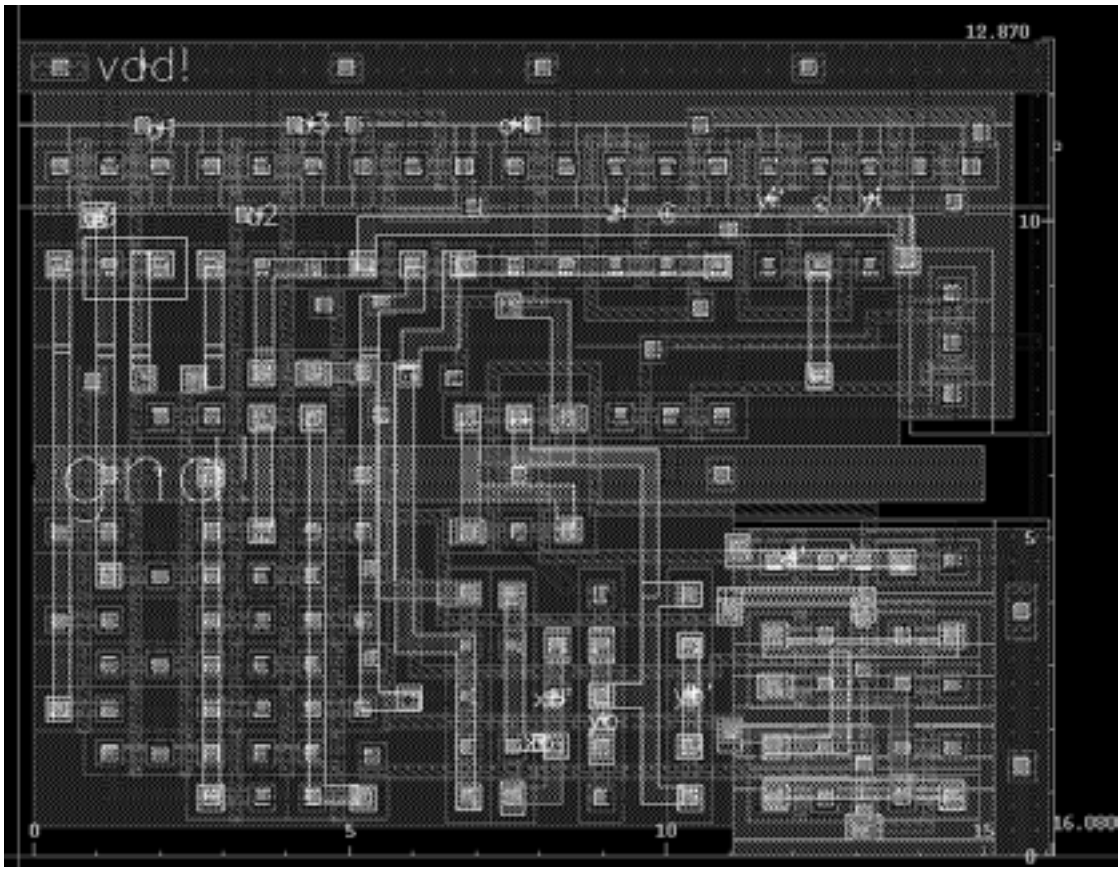


Figure 12. The layout of the 5-1 borrow parallel counter (refer to Figure 4, with two output buffers added).

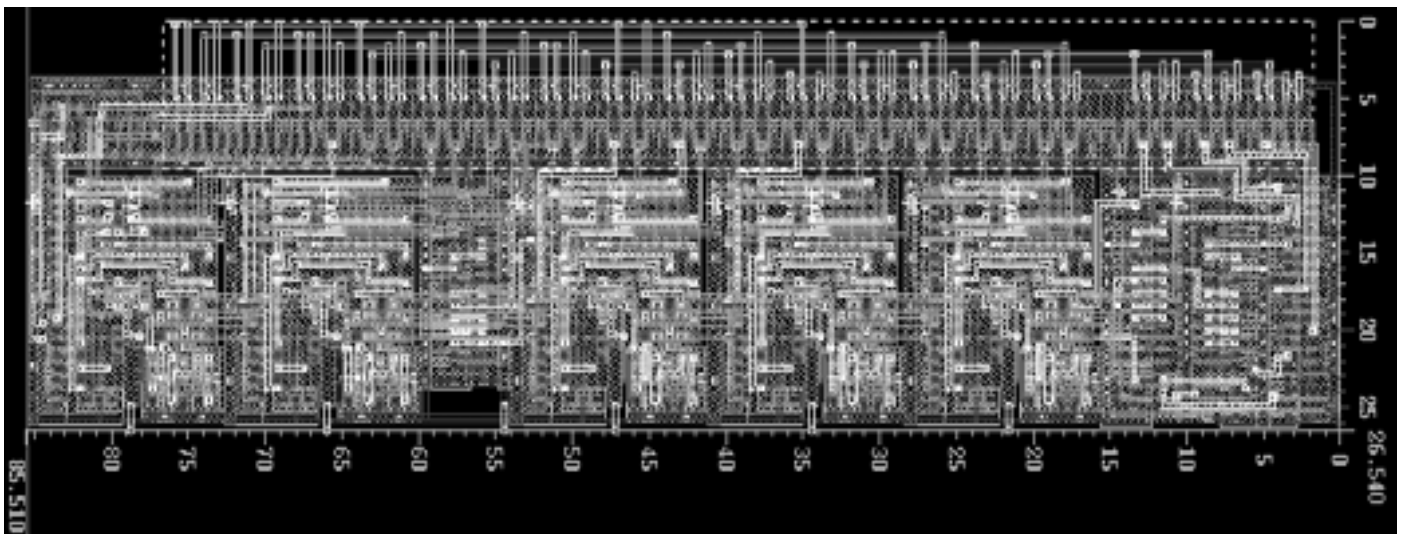


Figure 13. The layout of the 6 x 6 multiplier (refer to Figure 11).