A 3.8-ns CMOS $16 \times 16$-b Multiplier Using Complementary Pass-Transistor Logic

KAZUO YANO, MEMBER, IEEE, TOSHIKII YAMANAKA, MEMBER, IEEE, TAKASHI NISHIDA, MASAYOSHI SAITO, MEMBER, IEEE, KATSUHIRO SHIMOHI GASHI, MEMBER, IEEE, AND AKIHIRO SHIMIZU

Abstract — A 3.8-ns 257-mW CMOS $16 \times 16$-b multiplier with a supply voltage of 4 V is described. A complementary pass-transistor logic (CPL) is proposed and applied to almost the entire critical path. The CPL consists of complementary inputs/outputs, an nMOS pass-transistor logic network, and CMOS output inverters. The CPL is twice as fast as conventional CMOS due to lower input capacitance and higher logic functionality. Its multiplication time is the fastest ever reported, including bipolar and GaAs IC's, and it can be enhanced further to 2.6 ns with 60 mW at 77 K.

I. INTRODUCTION

The speed of CMOS devices, which were used mainly in low-power high-density LSI's, has increased drastically with the rapid progress in miniaturization. CMOS speed is getting close to that of Si bipolar technology; for example, with submicrometer CMOS technology, a 9-ns 1-Mb SRAM [1] and a 7.4-ns $16 \times 16$-b multiplier [2] have been reported. However, the recent progress in fast engineering workstations and real-time digital-signal processing requires faster CMOS speed.

A multiplier is an essential element in any digital-signal processing circuit and constitutes the critical path in DSP and FPU LSI's. Recently, versatile microprocessor chips also have begun to contain multipliers, which is made possible by the rapid progress in integration technology. Therefore, the demand for improved multiplier performance is increasing. In addition, multipliers are designed and fabricated as benchmarks for demonstrating various high-speed technologies, e.g., Si bipolar [3], GaAs [4], and Josephson junction devices.

This paper describes a fast 3.8-ns 0.5-um CMOS $16 \times 16$-b multiplier that is implemented as a test vehicle for investigating a new circuit technique for high-speed CMOS-based logic circuits. A new family of advanced differential CMOS logic, called complementary pass-transistor logic (CPL), is proposed and fully utilized on almost the entire critical path to achieve very high speeds. The circuit techniques and the technology used in this multiplier have the potential for 100-MHz operation of 32- to 64-b floating-point multiplication, thus enabling very fast DSP's, FPU's and ASIC's. First, the concept of CPL is introduced in Section II, and then CPL implementation of the multiplier is described in Section III. Device fabrication is described in Section IV, and finally the performance of the fabricated multiplier is shown in Section V.

II. CPL: CONCEPT AND EXAMPLES

Several differential CMOS logic families such as cascode voltage switch logic (CVSL) [5] (Fig. 1) and differential split-level logic (DSL) [6] have been proposed for CMOS circuit speed improvement. These have the common features of complementary data inputs/outputs, an nMOS logic tree, and a pMOS cross-coupled load, which together can reduce input capacitance, increase logic functionality, and sometimes eliminate inverter circuits. Therefore, these logic families can increase speed. However, the actual advantage of CVSL circuits is less than that anticipated in the original paper, as clarified in [7]. This is because the pMOS cross-coupled latch cannot easily be inverted due to the regenerative property of the latch. High-speed inversion of the pMOS latch is possible only when the gate width of the pMOS is sufficiently small. However, a small
gate width severely degrades the pull-up transit time. DSL is faster than conventional CMOS, however at the expense of static power consumption [7].

The main concept behind CPL is the use of an nMOS pass-transistor network for logic organization, and elimination of the pMOS latch, as shown in Fig. 2. CPL consists of complementary inputs/outputs, an nMOS pass transistor logic network, and CMOS output inverters. The pass transistors function as pull-down and pull-up devices. Thus the pMOS latch can be eliminated, allowing the advantage of the differential circuits to be fully utilized. Because the high level of the pass-transistor outputs (nodes \( Q \) and \( \overline{Q} \)) is lower than the supply voltage level by the threshold voltage of the pass transistors, the signals have to be amplified by the output inverters. At the same time, the CMOS output inverters shift the logic threshold voltage and drive the capacitive load. The logic threshold shift is necessary because the logic threshold voltage of the output inverter is lower than half the supply voltage, due to the lowering of the high signal level.

A pMOS latch can also be added to CPL, as shown in Fig. 2, to decrease static power consumption, as opposed to the conventional pull-up function. In this case, the pMOS gate width can be designed to be minimum, as long
as the pull-up function is completed in the given cycle time.

Arbitrary Boolean functions can be constructed from
the pass-transistor network by combining four basic circuit
modules: an AND/NAND module, an OR/NOR module, an
XOR/XNOR module, and a wired-AND/NAND module. These
are shown in Fig. 3(a) in which the XOR/XNOR module is
used once [8]. One attractive feature of CPL is that the
complementary outputs are produced by the simple four-
transistor circuits. Because inverters are unnecessary in
CPL circuits, the number of critical-path gate stages can
be reduced. Note that these various functions are produced
by an identical circuit configuration with only a change
of input configuration. This property of CPL is apparently
suitable for masterslice design. Logic functions for three
and greater inputs can also be easily constructed similarly
to two-way logic. Examples of three-way input logics are
shown in Fig. 3(b).

As an example of more complex logic circuits, a CPL
full adder is shown in Fig. 4. A conventional CMOS full
adder [9] is also shown in Fig. 4 for comparison. In the
CPL full adder, both the sum logic and carry logic are
structured on the CPL concept by combining the basic
modules in Fig. 3. The sum logic comprises two XOR/XNOR
modules, whereas the carry logic comprises three wired-
AND/NAND modules. The output inverters are “overhead,”
in the sense that they are needed whether the circuit has
one, two, or many inputs. Therefore designing with com-
plex logic functions in a gate is adopted to minimize the
overall device count and delay time.

Since pMOS can be eliminated in logic construction in
the CPL, the input capacitance is about half of the
conventional CMOS configuration, thus achieving higher
speed and lower power dissipation. Moreover, the power-
ful logic functionality of CPL due to the multilevel pass-
transistor network realizes complex Boolean functions
efficiently with a small number of MOS transistors, thus
further reducing area and delay time. In fact, the transistor
count in the CPL full adder is 28, whereas in the CMOS it
is 40. The areas required for these full adders based on the
half-micrometer design rule are 4218 μm² (37 × 114 μm²)
for CPL and 4730 μm² (55 × 86 μm²) for conventional
CMOS. The actual area reduction rate is smaller than the
transistor-count reduction rate, because the interconnec-
tion area is not proportional to the transistor count. Fur-
ther, the complementary input/output function eliminates
the internal inverter to provide XOR input in the full adder,
thus reducing the number of critical-path gate stages.

To compare the full-adder performance between CPL
and CMOS, circuit simulations are performed using the
half-micrometer device parameters at a supply voltage of
4 V. The 4-V supply was chosen because it is the maximum
voltage at which half-micrometer CMOS devices are
immune to hot-carrier degradation. The simulated worst-case
delay time (which refers to situations where the input
signals are such that circuit operation is slowest) of the
CPL full adder is as short as 0.26 ns, which is 2.5 times
faster than the conventional CMOS.

The simulated power dissipation as a function of supply
voltage is shown in Fig. 5. The CPL consumes 30% less
power than CMOS with a 4-V supply mainly due to
smaller input capacitance. The effect of power reduction in
CPL is more significant at lower supply voltages. This is
because the logic swing of the pass-transistor outputs
(nodes $Q$ and $\bar{Q}$ in Fig. 2) is smaller than the supply voltage level. The dynamic power resulting from charging and discharging the capacitances in CPL circuits is given as

$$P = C_{OUT} \cdot V_{CC}^2 \cdot f + C_{INT} \cdot V_{CC} (V_{CC} - V_{TN}) \cdot f$$  \(1\)

where $P$ is the dynamic power in the CPL circuit, $C_{OUT}$ is the output capacitance driven by the output inverters, $C_{INT}$ is the internal capacitance (at nodes $Q$ and $\bar{Q}$ in Fig. 2), $V_{CC}$ is the supply voltage, $V_{TN}$ is the threshold voltage of the nMOS pass transistors, and $f$ is the operating frequency. On the other hand, as is well known, the power dependence on supply voltage in CMOS circuits is described by squared dependence:

$$P' = C_{OUT} \cdot V_{CC}^2 \cdot f$$  \(2\)

where $P'$ is the dynamic power in the CMOS circuit, and $C_{OUT}$ is the total capacitance in the CMOS circuit. In the full-adder circuits in Fig. 4, the sum of $C_{OUT}$ and $C_{INT}$ is 20% smaller than $C_{OUT}$. In addition, at low supply voltages the threshold voltage effect of the second term in (1) further reduces the dynamic power dissipation.

The following are key points in CPL circuit design:

1) controlling the threshold voltage of the nMOS pass transistors to a lower value than the threshold voltage of pMOS; and

2) designing the logic threshold voltage of the output CMOS inverter to a lower value than $V_{CC}/2$.

These points improve the speed, static power dissipation, and noise margins of CPL circuits by offsetting the lowering of the high signal level at the pass-transistor output nodes. The full-adder delay time as a function of gate width ratio between pMOS and nMOS in output inverters $W_p/W_n$ has a minimum as shown in Fig. 6. The optimum gate width ratio exists at the ratio of about 0.75, which is much less than that of the ordinary CMOS inverters (1.5-2).

The CPL circuit can also be applied to carry-lookahead logic, which is indispensable for fast ALU’s and multipliers. The CPL 4-b carry-propagate circuit and its single-ended counterpart are shown in Fig. 7. The logic functions
of these carry-propagate circuits are expressed by

\[ C_a = G_a + P_4 \cdot \left[ G_5 + P_5 \cdot (G_2 + P_2 \cdot G_1) \right] + P_4 \cdot P_5 \cdot P_2 \cdot P_1 \cdot C_0 \]  

\[ C_b = A + B \cdot C_0 \]

where \( C_a \) is the carry input from the lower carry-lookahead unit, \( C_b \) is the carry output to the higher carry-lookahead unit, \( P_j \)'s \((j = 1 \ldots 4)\) are the carry-propagate signals, and \( G_j \)'s are the carry-generate signals. The OR logic in (4) can be expressed by the simple pass-transistor circuits in Fig. 7 by considering the following relationship:

\[ A \cdot B = 0. \]

Within a single inverter delay time, the CPL carry-lookahead unit can quickly transfer the carry output to the upper unit \((C_a \text{ and } C_b)\) after receiving the carry input from the lower unit \((C_a \text{ and } C_b)\). Complementary carry data in the CPL unit can be inverted simply by twisting the carry lines. The simulated delay time is only 0.15 ns/4 b. By contrast, the single-ended counterpart requires two-stage inverter delay. Because in actual adders and multipliers these units are connected in series (for a 64-b adder, 16 units have to be connected) and constitute the critical path, the delay reduction of a factor of 2 significantly reduces the total adder delay time.

III. MULTIPLIER ARCHITECTURE

The 16×16-b multiplier was designed using a parallel multiplication architecture, as shown in Fig. 8(a). A Wallace-tree adder array and a CLA adder were used to minimize the critical-path gate stages. There are a total of 8500 transistors in an active area of 1.3×3.1 mm², whereas the area including bonding pads is 1.6×4.5 mm². The transistor count is less than that of a full CMOS counterpart [2], mainly because the transistor count in the full adder is less than that of the CMOS full adder. The critical path consists of a partial-product generator, six full adders, lookahead carry logic, three carry-propagate circuits, and a final full adder as shown in Fig. 8(b).

IV. DEVICE FABRICATION

The multiplier and the ten-stage full-adder chains are fabricated with double-level-metal 0.5-μm CMOS technology. The minimum feature size is 0.5 μm, and the gate oxide thickness is 12.5 nm. In this fabricated 0.5-μm CMOS device optimized for CPL, the nMOS pass transistors are designed to have a threshold voltage of 0 V, whereas the other nMOS and pMOS have a threshold voltage of 0.4 and −0.4 V, respectively. This threshold control reduces the static power dissipation and delay time. The drain saturation current per 10-μm width is 4.6 mA for nMOS and −2.6 mA for pMOS. The interconnection metal consists of first-level W and second level Al. The W was adopted for its high immunity to electromigration.

V. PERFORMANCE RESULTS

A microphotograph of the multiplier and the full-adder chains is shown in Fig. 9. Ten-stage full-adder chains using CPL full adders and CMOS full adders were designed and fabricated to compare the actual performance. The measured worst delay time as a function of supply voltage is shown in Fig. 10.

The delay dependences are similar for CMOS and CPL. Thus CPL can be used at supply voltages at least as low as those for CMOS. The results agree well with the simula-
The completed multiplier chips were probe tested at wafer level. The fully functional chips were mounted on the 68-pin pin-grid-array ceramic package, followed by waveform observation of the clock inputs and the product outputs through source-follower circuits on the chip. The test was performed up to the maximum frequency limit (250 MHz) of the pulse generator. Speed performance was measured using the worst-case pattern, $\text{FFFF} \times 8001 - \text{7FFF} \times 8001$. Circuit simulation was also performed to confirm that this pattern consists of a series of worst-case operations of the full adder and the CLA. The multiplication time was measured at both room and liquid-nitrogen temperatures. The latter was considered for high-end applications, such as supercomputers [10], [11]. The maximum multiplication times were 3.8 and 2.6 ns at room and liquid-nitrogen temperatures, respectively. The measured waveforms are shown in Fig. 11. The observed signal amplitude is larger at lower temperatures because the on-chip source-follower gain is larger. The multiplication time versus product output is shown in Fig. 12. $P_{30}$ gives the longest delay time as expected from our simulations. There is very little variation on the product output from $P_{22}$ to $P_{32}$, which means the carry propagates very quickly in the CPL carry-lookahead circuit. Power dissipation was 257 and 60 mW at 300 and 77 K, respectively, for 10-MHz operation with a pattern of $\text{FFFF} \times \text{FFFF} - \text{0000} \times \text{FFFF}$.

The multiplication times at both room and liquid-nitrogen temperatures are compared with the published 16×16-b multipliers in Fig. 13. The minimum multiplication time before this work was 41 ns with a power dissipation of 6.2 W and was realized by GaAs high electron mobility
transistors (HEMT’s). Previously, CMOS multipliers dissipated power one order of magnitude lower than the others. However, the multiplication time was 2–3 times larger than those of fast multipliers. By contrast, the present CPL’s multiplication times at both room and liquid-nitrogen temperatures are faster than those of any other devices. In addition, the power dissipation is much lower than those of the other devices. The features of this multiplier are summarized in Table I.

The estimated performance of a multiplier using CPL and the half-micrometer CMOS devices is shown in Fig. 14, considering that the bit length is enlarged and floating-point multiplication functions are included. The architecture is assumed to be a combination of Booth’s algorithm and Wallace-tree adder array. The bit length and the depth of the carry-lookahead unit are assumed to be optimized for the mantissa bit length. The variations of process, supply voltage, and temperature are considered simply by multiplying the empirical factor by the typical delay time. The multiplication time increases with increasing mantissa bit length. However, the 64-b floating-point multiplication time is estimated to be 10 ns at room temperature and 6 ns at liquid-nitrogen temperature. Thus, operation at over 100 MHz is possible even at room temperature if the multiplier architecture and the carry-lookahead circuit configurations are carefully optimized for high-speed operation.

### VI. Conclusions

This paper described a fast 16 × 16-b multiplier using a new differential CMOS logic family, CPL. In CPL, differential logic is constructed without pMOS latching load, enabling a speed more than twice as fast as conventional CMOS. The power dissipation is also smaller due to smaller input capacitance. The multiplier is the fastest ever reported at both 300 K and 77 K, proving that the half-micrometer CMOS technology fully utilizing CPL has a speed which is at least competitive with those of other fast devices with a much smaller power dissipation at room temperature, and is faster at liquid-nitrogen temperature. These results also demonstrate that half-micrometer CMOS devices fully utilizing CPL have a performance potential of a 100-MHz repetition rate for floating-point multiplication by carefully optimizing the multiplier architecture for high-speed operation. Therefore, very high-speed MPU’s, DSP’s, FPU’s, and ASIC’s are possible.

### Acknowledgment

The authors wish to thank Y. Sakai and O. Minato from Hitachi Semiconductor Development Center and T. Masuhara, T. Nakagawa, T. Baji, K. Kaneko, T. Sawase, and K. Ishibashi from Hitachi Central Research Laboratory for their useful suggestions and discussions. The authors are also greatly indebted to K. Yagi and the device processing staff members from Hitachi Central Research Laboratory for their support throughout the sample fabrication, and to A. Kawamata from Hitachi VLSI Engineering Corporation and K. Ueda from Hitachi Central Research Laboratory for their layout design.

### References


Kazuo Yano (M'87) was born in Sakata, Japan, on October 18, 1959. He received the B.S. and M.S. degrees in physics from Waseda University, Tokyo, Japan, in 1982 and 1984, respectively.

He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, in 1984. Since 1984 he has been engaged in research on the physics, limitations, reliability, and circuits of low-temperature CMOS and bipolar devices. His current interests include fast circuits for CMOS/BiCMOS VLSI processors and limitations of ULSI systems.

Mr. Yano is a member of the Japan Society of Applied Physics and the Institute of Electronics, Information and Communication Engineers of Japan.

Masayoshi Saito (S'83-M'83) was born in Saitama, Japan, on October 28, 1954. He received the B.S., M.S., and Dr. Eng. degrees in electronics from Tohoku University, Sendai, Japan, in 1978, 1980, and 1984, respectively.

In 1984 he joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, where he is engaged in research on MOS gate electrodes and refractory metal interconnections for VLSI. Dr. Saito is a member of the IEEE Electron Devices Society, the Electrochemical Society, and the Japan Society of Applied Physics.

Toshiaki Yamanaka (M'87) was born in Tokyo, Japan, in 1957. He received the B.S. and M.S. degrees in electrical engineering from the University of Electro-Communications, Tokyo, Japan, in 1980 and 1982, respectively.

In 1982 he joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan. He has been engaged in the research and development of CMOS and high-density static RAM processes. Mr. Yamanaka is a member of the IEEE Electron Devices Society and the Solid-State Circuits Council.

Katsuhiro Shimohigashi (M'77) was born in Fukuoka, Japan, on September 21, 1946. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Kyushu University, Fukuoka, Japan, in 1969, 1971, and 1988, respectively.

He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, in 1971. Since 1973 he has been working on the circuit and device design of dynamic MOS memories. He was a Visiting Research Scholar at Stanford University, Stanford, CA, from September 1977 to August 1978, where he worked on submicrometer device physics and the modeling of VLSI devices. Currently, he is supervising the development of CMOS/BiCMOS devices and circuits for advanced logic/memory applications, nonvolatile memories, and low-temperature devices for future ULSIs. Dr. Shimohigashi is a member of the IEEE Electron Devices Society and the Institute of Electronics, Information and Communication Engineers of Japan.

Takashi Nishida received the B.S. and M.S. degrees in metallurgy from the University of Tokyo in 1972 and 1974, respectively.

He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, in 1974, and worked on the research and development of field emission cathode, CO$_2$ laser-assisted dielectric film deposition and InSb thin-film transistors. He is currently engaged in multilevel interconnection technology for ULSSI's. He is also responsible for CMOS SRAM and BiCMOS process integration.

Akihiro Shinizu was born in Tochigi, Japan, on October 19, 1958. He received the B.S. degree in material physics from the University of Hiroshima, Hiroshima, Japan, in 1980.

He was with Hitachi Microcomputer Engineering Ltd. from 1981 to 1984, where he researched in MOS device technology. Since 1985 he has been working for the Hitachi VLSI Engineering Corporation, Tokyo, Japan. His current research interests are in submicrometer MOS device technology for SRAM and reliability physics, such as hot-carrier effects. Mr. Shinizu is a member of the Japan Society of Applied Physics.