REFERENCES

[1] N. Benowitz *et al.*, "An advanced fault isolation system for digital logic," *IEEE Trans. Comput.*, vol. C-24, pp. 489–497, May 1975.
[2] T. Sridhar, "A new parallel test approach for large memories," in *Proc. Int. Test Conf.*, 1985, pp. 462–470.
[3] T. Mano *et al.*, "Circuit techniques for a VLSI memory," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 5, pp. 463–469, Oct. 1983.
[4] J. Yamada, "Selector-line merged built-in ECC technique for DRAM's," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 5, pp. 868–873, Oct. 1987.
[5] R. M. Tanner, "Fault-tolerant 256K memory designs," *IEEE Trans. Comput.*, vol. C-33, pp. 314–322, Apr. 1984.
[6] F. I. Osman, "Error-correction techniques for random-access memories," *IEEE J. Solid-State Circuits*, vol. SC-17, no. 5, pp. 877–881, Oct. 1982.
[7] W. W. Peterson and E. J. Weldon, in *Error Correcting Codes*. Cambridge, MA: MIT Press, 1972.

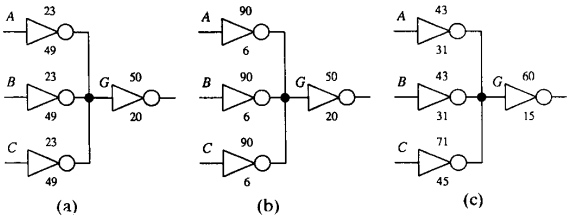

Fig. 1. Three GCMOS circuits with identical topology, but different functions: (a) OR3, (b) AND3, and (c) $A \cdot B + C$. Number above inverter is p-transistor width; number below is n-transistor width. All channel lengths = 3 μm.

Ganged CMOS: Trading Standby Power for Speed

KENNETH J. SCHULTZ, ROBERT J. FRANCIS, AND KENNETH C. SMITH, FELLOW, IEEE

***Abstract*—This correspondence presents ganged-CMOS logic (GCMOS), a technique employing CMOS inverters with their outputs shorted together, driving one or more encoding inverters. These encoding inverters, serving to quantize the nonbinary signal at the "ganged" node, effectively buffer it from external circuitry, thus allowing locally smaller noise margins. As demonstrated by two novel adders, GCMOS achieves higher speeds and lower input capacitances than static CMOS, at the expense of higher static power dissipation.**

I. INTRODUCTION

Fully complementary CMOS static circuits dissipate negligible dc power, can operate asynchronously, and do not require the routing of clock signals [1]. However, static circuits are generally slower than dynamic circuits. One way of overcoming this deficiency is to trade off standby power consumption for speed. Johnson [2] recently presented a novel CMOS NOR gate using inverters with their outputs shorted together. The design of such NOR's involves transistor ratioing to set appropriate high and low output levels.

The static power–speed trade-off is acceptable in localized applications, where there is a demonstrated need for a special function to operate particularly quickly. The static power dissipation is thus kept physically isolated to a few locations on an IC, and may in fact be insignificant in an environment where conventional static circuits operate near their maximum frequency, and thereupon dissipate considerable dynamic power. It may be noted, moreover, that the local nature of the concept does not preclude the use of other speed-enhancing techniques, and thus its use can provide the incremental delay improvement which may make a design feasible.

This correspondence extends the concepts in [2] to include buffering of the shorted or "ganged" node, thereby allowing the realization of more complex gates, and thus, the idea called "ganged-CMOS logic" (GCMOS). A number of sample circuits are presented. In particular, two novel adders are described and compared with an accepted conventional implementation. The ganged-CMOS adders provide lower input capacitance and faster carry propagation, for equally sized layouts.

Manuscript received December 13, 1989.
The authors are with the Department of Electrical Engineering, University of Toronto, Toronto, Ont., M5S 1A4, Canada.
IEEE Log Number 9035670.

II. GANGED CMOS

By buffering the ganged node with a simple CMOS inverter, a number of advantages are obtained. First of all, the ganged node is effectively isolated from external circuitry—its value is neither transmitted on long interconnect wires and corrupted by noise, nor does it drive complex gates, where any voltage exceeding a transistor threshold can cause a logic error. Essentially, one can tolerate much lower noise margins on a local node than on a global node. This benefit is enhanced by the inverter's inherent encoding action—its high gain results in a sharp distinction between low and high inputs. Furthermore, the inverter's switching point, while dependent on the square root of the p-n ratio, can be varied adequately by adjusting transistor geometries.

Fig. 1 shows the same circuit topology repeated three times; transistor widths are changed to realize three different functions. The first, in Fig. 1(a), implements an OR gate; only one of the three inputs need be high for the ganged node G to be forced well below the switching point of the buffer inverter. Similarly, the dimensions shown in Fig. 1(b) implement an AND gate. The circuit in Fig. 1(c) implements the logic function $A \cdot B + C$; the inverter driven by C is essentially "twice" as strong as those driven by A and B.

GCMOS results in a lower transistor count for more complex functions, as demonstrated in the examples that follow. Although it is true that the exclusive use of inverters limits the area-saving parallel and series layout of transistors, the lower transistor count overcomes this area deficiency. For further area savings, it is possible to group n-channel and p-channel transistors in common tubs, such that the inverter's two transistors are not constrained to be physically adjacent. It is also possible to alter p- and n-area requirements by enhancing the encoding-inverter threshold (for example, using narrower p-transistors for both input and encoding inverters, if the encoding-inverter switching point is lowered).

III. GLAD: GANGED-LOGIC ADDER WITH DOUBLE GANGED NODES

Two ganged nodes, with their input and output inverters weighted to implement different functions, are present in the adder circuit shown in Fig. 2. The full adder is realized using

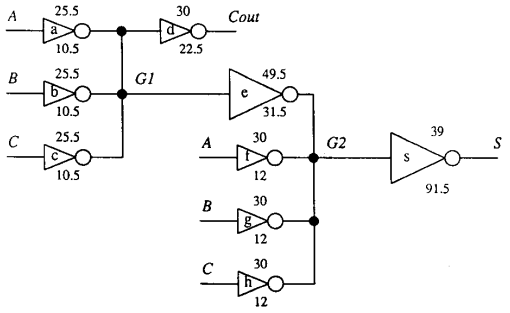

Fig. 2. GLAD adder. Transistor widths shown; all lengths = 3 μm.

TABLE I
8-b ADDER COMPARISON

property	(units)	Hampel	GLAD	GLAQ
no. transistors		224	144	192
area	(mm^2)	0.144	0.138	0.141
A/B input cap	(fF)	900	248	105
C input cap	(fF)	620	248	105
τ_C	(ns)	5.4	3.8	3.8
τ_S	(ns)	14	14	15
8-bit time	(ns)	52	41	42
degraded time	(ns)	57	50	52
static power	(mW)	0	120	35-65
20 MHz power	(mW)	12	130	75

only 18 transistors; this is far fewer than the number required for a conventional fully complementary CMOS static adder (26 or 28) [1], and is four fewer than the 22 needed for a DCVS adder [3].

Inverters INV_a, INV_b, and INV_c, and the $G1$ node's buffer inverter, INV_d, are scaled for minimum carry-propagation delay; a two-of-three majority function is realized by these inverters. The other five inverters implement the function $\overline{C_{out}}\cdot(A+B+C)+A\cdot B\cdot C$. A first pass at transistor-width selection is performed by selecting output levels and noise margins, and doing a quick hand analysis using basic transistor equations. SPICE simulations are then used to arrive at optimum dimensions. A rather low switching voltage was selected for INV_s in order to minimize the layout area by keeping p-channel transistor widths low.

In order to evaluate the performance of the GLAD, the circuit was laid out in a 3-μm CMOS process and extracted, and simulations were run on an 8-b configuration using HSPICE. Results are compared with those obtained from a conventional CMOS adder by Hampel *et al.* [4] with approximately the same layout area; results are shown in Table I. The 8-b addition time was determined as $\tau_S + 7\times\tau_C$, where τ_S is the worst-case 1-b sum time (to 90% of the output transition, driving a 1-pF load). Input capacitance is much less for the GLAD, and the carry-propagation delay τ_C is considerably less as well. Note that for longer word lengths, the speed advantage of the GLAD would become even more pronounced. The obvious drawback of the GLAD is its extremely high static power dissipation.

Because noise margins are quite narrow compared to conventional implementations, further simulations were carried out using the Monte-Carlo technique in order to ensure correct operation over a range of processing variations and supply voltages. Thirty passes were used; if the circuit functions correctly for all 30 trials, there is a 99% probability that it will also operate correctly over 80% of all component values in the given distribution [5]. While V_{DD} was varied from 4.75 to 5.25 V ($\pm$5%), simultaneous Gaussian distributions were used for transistor threshold voltages (σ_{VT} = 3%) and gate oxide thicknesses (σ_{OX} = 5%). Voltage distributions are shown for node $G1$ in Fig. 3(a), and for node $G2$ in Fig. 3(b). Upper and lower distributions are for node voltages; the middle distributions are for the encoding inverter switching points. Over the entire parameter range, noise margins (referenced to the encoding inverter's switching points) are maintained at a minimum of 280 mV. In the worst case, 8-b addition time degrades to 50 ns. Note that even for the Hampel adder, a degradation to 57 ns occurs.

The values of σ_{VT} and σ_{OX} above certainly take into account intra-chip variations, and should also be adequate for characterizing variations from batch to batch in a high-quality process, of the kind which is a normal prerequisite for high-speed high-volume circuit production. As derived in [8], simulation with variations in only the two process variables above implicitly takes into account most possible process variations, because of statistical independence.

There may be a concern regarding the low noise margins required for correct operation. However, it must be noted that it is the nature of high-speed circuits (especially dynamic ones) that glitches exceeding V_{tn} or V_{tp} can result in logic errors. This usually necessitates specification of setup and hold times on inputs relative to a clock; correct operation is thus inherently dependent on process variations, which may lead to threshold voltage changes and/or potential race conditions.

IV. GLAQ: Ganged-Logic Adder with a Quaternary Ganged Node

If a given GCMOS node can sustain more than two-stable voltage values, they can be distinguished by multiple buffer inverters, each with a different switching threshold. In this case, the GCMOS node is actually a multiple-valued logic (MVL) node [6]. The major emphasis in MVL circuit and system design traditionally concerns the use of multiple-valued signals (rather than simple binary ones) as a means by which to increase information-transfer density by reduction of the number of signal-interconnection lines. Rather less attention has been paid to the fact that the succinct encoding that MVL implies is potentially also of use on a more *local* scale. This local application of MVL is readily achieved using GCMOS [7]. The net result is the use of an MVL node to enable GCMOS to simultaneously realize multiple functions.

Fig. 4 shows a schematic of the GLAQ adder [7]. All inputs and outputs are binary, but the node Q is quaternary (four-valued). Ideally, if the inputs are all low, node Q will be high (V_{DD}); if one of the inputs is high, the voltage at node Q is 2/3 V_{DD}; if two are high, it is 1/3 V_{DD}; if all inputs are at V_{DD}, node Q is at zero. The three output inverters have different switching thresholds, producing the functions 1-of-3 (INV_d), 2-of-3 (INV_e), and 3-of-3 (INV_f). Unfortunately, the "ideal levels" given above are impossible to obtain in practice; with one input at V_{DD} and two at GND, the voltage at Q will be so high that the switching threshold of INV_d will have to be unreasonably high (requiring $W_{pd} \gg W_{nd}$) in order to differentiate this signal from V_{DD}. A similar situation occurs at the low end.

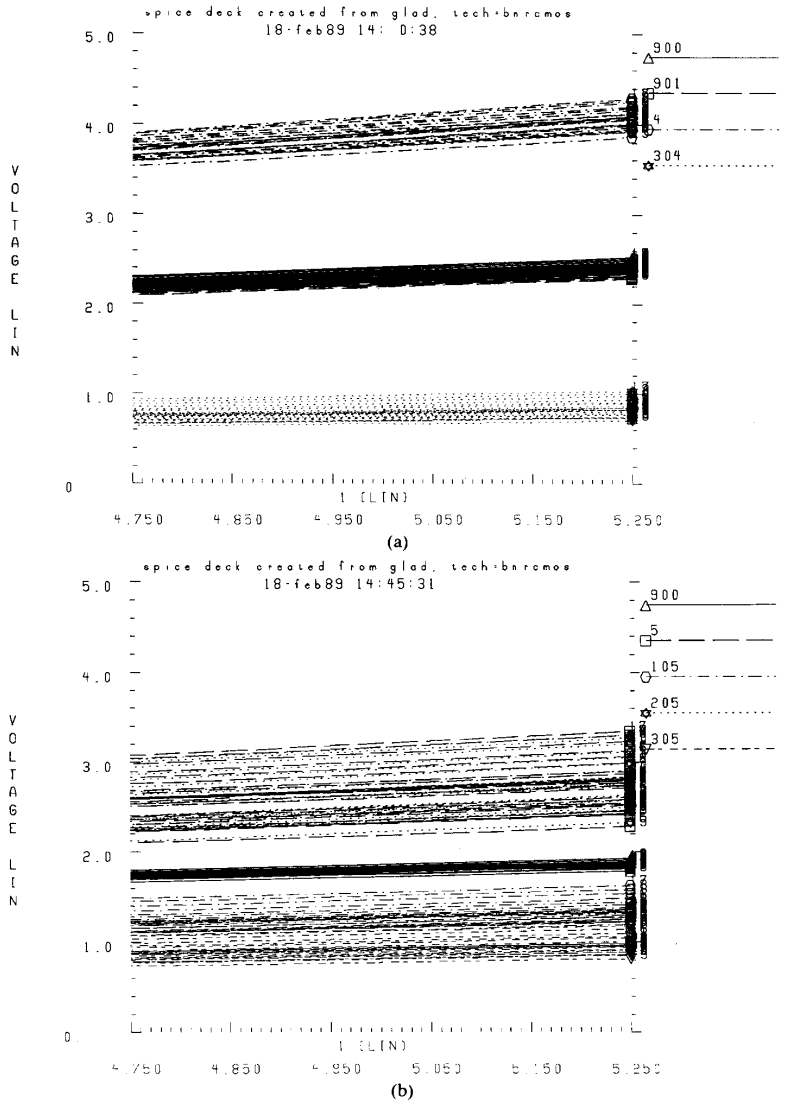

Fig. 3. Monte Carlo distributions of node voltages and encoding-inverter switching points: (a) GLAD node $G1$, and (b) GLAD node $G2$.

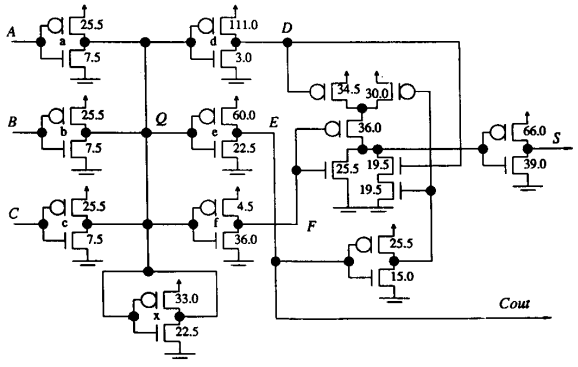

Fig. 4. GLAQ adder. Transistor widths shown; all lengths = 3 μm.

This problem is overcome by connecting INV_x as shown in Fig. 4. This inverter provides negative feedback at Q, which tends to pull all stable voltages closer to a ratio-dependent central value, thus enabling differentiation between 0- and 1-of-3, as well as 2- and 3-of-3. The four possible voltages on Q can be adjusted by selecting values of W_{na}, W_{pa}, W_{nx}, and W_{px}, where $W_{na} = W_{nb} = W_{nc}$ and $W_{pa} = W_{pb} = W_{pc}$. The widths can be determined roughly by hand analysis; they must be optimized using SPICE. The same observation applies to transistor ratios of the output inverters $\text{INV}_{d,e,f}$. Table II presents ratios of the encoding inverters for suitable switching thresholds. Larger noise margins are maintained for the 2-of-3 inverter (INV_e), since it must operate as quickly as possible to minimize τ_C. This is an example of the obvious threshold-voltage–speed trade-off that must be considered when designing a GCMOS gate.

TABLE II
GLAQ ENCODING-INVERTER INFORMATION

inverter	V_{IL}	threshold	V_{IH}	W_n	W_p
INV_d	3.16	3.60	3.98	3.0	111.0
INV_e	1.82	2.56	3.16	22.5	60.0
INV_f	0.68	1.43	1.82	36.0	4.5

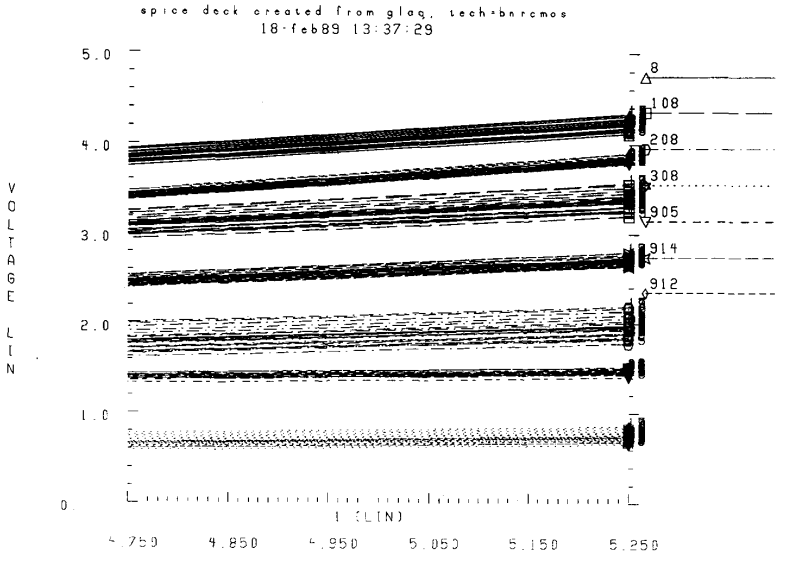

Fig. 5. GLAQ node-Q Monte Carlo distributions. Stable node voltages alternate with switching points.

A simple fully complementary CMOS gate is used to realize the sum, according to the expression $S = F + D \cdot \bar{E}$, where $E = C_{out}$.

In order to determine the merit of the circuit, an 8-b GLAQ was laid out, extracted, and simulated for comparison with both the Hampel adder and the GLAD. The reader is referred to Table I for a performance comparison. Because each of its inputs is applied to only one inverter, the GLAQ has the lowest input capacitance. The GLAQ has a slightly slower 8-b addition time than the GLAD, but consumes significantly less power. Standby power dissipation is highly dependent on the state of the adder inputs.

Monte Carlo simulations were again carried out, using the same processing and power-supply variations that were used for the GLAD. As shown in Fig. 5, noise margins of at least 220 mV are maintained at the ganged quaternary node. In the worst case, 8-b addition time degrades to 52 ns.

V. CONCLUSIONS

This correspondence has presented ganged-CMOS logic (GCMOS), a technique employing CMOS inverters with their outputs ganged together, driving one or more encoding inverters. These encoding inverters effectively buffer the "ganged" node from external circuitry, thus allowing smaller noise margins. As demonstrated by two novel adder structures, GCMOS achieves higher speed and lower input capacitances at the expense of high static power dissipation. Monte-Carlo simulations have shown that extremely tight process control is not needed to ensure correct operation; however, it is required to obtain optimum circuit performance. The degree of control is not inconsistent with that required by other techniques used (separately or in conjunction) in speed-critical parts of high-performance systems.

REFERENCES

[1] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*. Reading, MA: Addison-Wesley, 1985.
[2] M. G. Johnson, "A symmetric CMOS NOR gate for high-speed applications," *IEEE J. Solid-State Circuits*, vol. 23, pp. 1233–1236, 1988.
[3] K. M. Chu and D. L. Pulfrey, "A comparison of CMOS circuit techniques: Differential cascode voltage switch logic versus conventional logic," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 528–532, 1987.
[4] D. Hampel, K. E. McGuire, and K. J. Prost, "CMOS/SOS serial–parallel multiplier," *IEEE J. Solid-State Circuits*, vol. SC-10, pp. 307–313, 1975.
[5] *HSPICE Simulation Report 10*, Meta-Software, Campbell, CA, Mar. 1988.
[6] K. C. Smith, "Multiple-valued logic: A tutorial and appreciation," *IEEE Computer*, vol. 21, no. 4, pp. 17–27, Apr. 1988.
[7] K. J. Schultz and K. C. Smith, "A CMOS binary adder using a quaternary ganged-logic internal node," in *Proc. 19th Int. Symp. Multiple-Valued Logic* (Guangzhou, P. R. C.), May 29–31, 1989, pp. 356–359.
[8] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterization and modelling of mismatch in MOS transistors for precision-analog design," *IEEE J. Solid-State Circuits*, vol. SC-21, pp. 1057–1066, 1986.