Leakage and Process Variation Effects in Current Testing on Future CMOS Circuits

Ali Keshavarzi, James W. Tschanz, Siva Narendra, and Vivek De
Intel Labs

Kaushik Roy
Purdue University

Charles F. Hawkins
University of New Mexico

W. Robert Daasch
Portland State University

Manoj Sachdev
University of Waterloo

Barriers to technology scaling, such as leakage and parameter variations, challenge the effectiveness of current-based test techniques. This correlative multiparameter test approach improves current testing sensitivity, exploiting dependencies of transistor and circuit leakage on operating frequency, temperature, and body bias to discriminate fast but intrinsically leaky ICs from defective ones.

FOR FUTURE CMOS technology generations, supply and threshold voltages will have to scale together to sustain high performance, limit energy consumption, control power dissipation, and maintain reliability. These continual scaling requirements pose several technology, circuit design, and testing challenges. Controlling process variation and leakage has become critical in designing and testing ICs. Die-to-die and intradie parameter variations—which are worsening with technology scaling—affect IC clock frequency and leakage power distributions. These effects are more pronounced at low supply voltages (V_{CC}). Technology scaling also affects various aspects of VLSI testing.[1] Specifically, elevated transistor leakage and excessive parameter variations in scaled process technologies threaten the feasibility of leakage-based (I_{DDQ}) tests.

Researchers have proposed several methods for adapting I_{DDQ} testing to scaled technologies—reverse body bias (RBB), current signatures, ΔI_{DDQ} testing, and transient current testing (see the sidebar, "Other I_{DDQ} test methods"). Our data suggests adopting a correlative multiparameter test solution. For high-performance IC applications, we propose looking at leakage in the context of the circuit's maximum operating frequency. This approach doesn't rely on the absolute value of the current. Added parameters of temperature and body bias further improve our test technique's defect detection sensitivity. Leakage averaging and variance

0740-7475/02/$17.00 © 2002 IEEE

reduction techniques, such as nearest-neighbor residual, should mitigate parameter variation limitations. Finally, we advocate the adaptive body bias technique to enhance manufacturing yield by compensating for die-to-die and intradie parameter variations and their effect on a circuit's leakage and frequency.

Leakage in the context of frequency

Our approach to testing intrinsically leaky, scaled ICs differs from other state-of-the-art solutions, yet complements them. We correlate intrinsic transistor device and circuit parameters to develop multiparameter test solutions. An example is correlating leakage (I_{DDQ}) to frequency (F_{MAX}) in a collection of ICs. Here, our test solution depends critically on characterizing and quantifying change in I_{DDQ} as a function of change in F_{MAX}. Elsewhere, we correlated the leakage current (I_{DDQ}) of a 32-bit microprocessor to its maximum clock frequency (F_{MAX}) in 0.35-micron technology.[1] We reexamined this relationship for 150-nm technology on a test chip circuit.[2] This test IC gave us more flexibility and degrees of freedom, such as the availability of body terminals, to explore the fundamental concepts before developing actual product-specific tests. The roots of the correlation between leakage and frequency exist in device physics; thus, for a circuit designed on a given process technology, that relationship can be established and characterized.[1]

Figure 1 (next page) shows a plot of log I_{DDQ} versus F_{MAX}, without RBB applied ($RBB = 0$ V), in our 150-nm technology. We collected this data from more than 100 ICs on two wafers. This semilog curve plots the relationship between test-chip ring oscillator (RO) normalized I_{DDQ} leakage and its normalized maximum operational frequency. (We normalized this figure with respect to the lowest chip leakage and frequency.)

The data indicates a linear relationship between log I_{DDQ} and RO frequency. We observed this linear dependency across the range of natural variation in transistor and circuit parameters in our ICs. We did not intentionally skew any parameters, such as V_T or L (transistor threshold voltage and transistor channel length), so our experiment shows the natural range of parameter variation. For this collection of ICs, a 35% increase in the frequency of the circuit under test increased the I_{DDQ} by 4.1 times at room temperature. In other words, the circuit that was faster by 35% also had leakage 4.1 times higher. Slower circuits have lower leakage; faster circuits have higher leakage. This fundamental relationship between an IC's

Other I_{DDQ} test methods

Motivated primarily by increasing adverse device leakage trends, researchers have recently reported several methods for sustaining the effectiveness of I_{DDQ} testing for scaled technologies—among them are reverse body bias (RBB), current signatures, ΔI_{DDQ} testing, and transient current testing.

Applying a reverse body bias creates a low-leakage I_{DDQ} test mode.[1,2] Gattiker and Maly suggested sorting I_{DDQ} test vectors in ascending order; an abrupt discontinuity in the current level indicates a defect.[3] Maxwell and colleagues demonstrated the effectiveness of current signatures with silicon data.[4] Thibeault and Miller each proposed a ΔI_{DDQ} test technique for uncovering defects.[5,6] Conceptually, the ΔI_{DDQ} technique is similar to the current signatures technique, where a sudden elevation in current level indicates a defect. Other proposals involve transient current test techniques.[7,8]

References

1. M. Sachdev, "Deep Submicron I_{DDQ} Testing: Issues and Solutions," *Proc. European Design & Test Conf.* (ED&TC 97), IEEE CS Press, Los Alamitos, Calif., 1997, pp. 271-278.
2. A. Keshavarzi, K. Roy, and C.F. Hawkins, "Intrinsic Leakage in Low-Power Deep Submicron CMOS ICs," *Proc. Int'l Test Conf.* (ITC 97), IEEE Press, Piscataway, N.J., 1997, pp. 146-155.
3. A. Gattiker and W. Maly, "Current Signatures: Applications," *Proc. Int'l Test Conf.* (ITC 97), IEEE Press, Piscataway, N.J., 1997, pp. 156-165.
4. P. Maxwell et al., "Current Ratios: A Self-Scaling Technique for Production I_{DDQ} Testing," *Proc. Int'l Test Conf.* (ITC 99), IEEE Press, Piscataway, N.J., 1999, pp. 738-746.
5. C. Thibeault, "An Histogram-Based Procedure for Current Testing of Active Defects," *Proc. Int'l Test Conf.* (ITC 99), IEEE Press, Piscataway, N.J., 1999, pp. 714-723.
6. A.C. Miller, "I_{DDQ} Testing in Deep Submicron Integrated Circuits," *Proc. Int'l Test Conf.* (ITC 99), IEEE Press, Piscataway, N.J., 1999, pp. 724-729.
7. M. Sachdev, V. Zieren, and P. Janssen, "Defect Detection with Transient Current Testing and Its Potential for Deep Submicron ICs," *Proc. Int'l Test Conf.* (ITC 98), IEEE Press, Piscataway, N.J., 1998, pp. 204-213.
8. B. Kruseman, P. Janssen, and V. Zieren, "Transient Current Testing of 0.25-µm CMOS Devices," *Proc. Int'l Test Conf.* (ITC 99), IEEE Press, Piscataway, N.J., 1999, pp. 47-56.

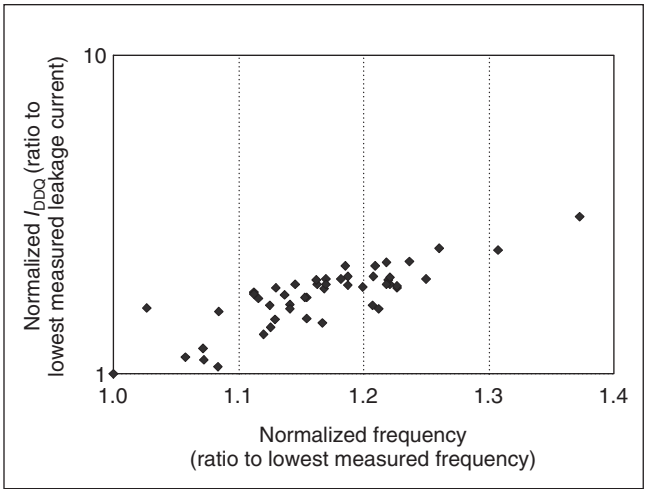

Figure 1. Linear relationship between frequency and the logarithm of leakage: normalized ring oscillator (RO) circuit frequency versus I_{DDQ} leakage, at room temperature without applied body bias.

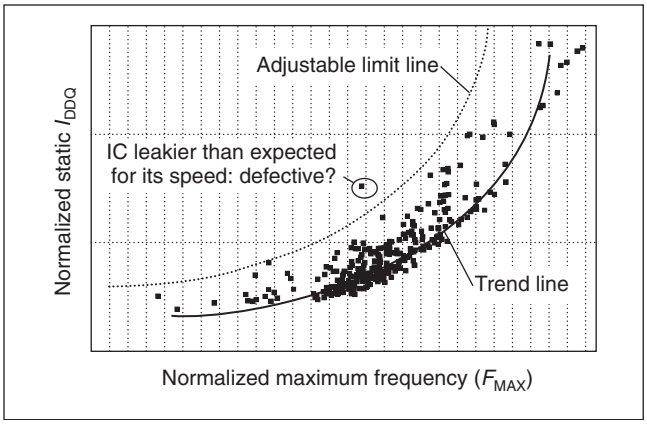

Figure 2. Two-parameter test: microprocessor I_{DDQ} versus F_{MAX}, with trend line and test limit.[1] ICs that fall substantially beyond the limit line are defective.

maximum operating frequency and its aggregate leakage is essential to the concept of two-parameter testing.

Multiple-parameter testing

Now let's look at a multiparameter test solution based on the intrinsic leakage-to-frequency relationship with the natural occurrence of parameter variations. Multiparameter testing correlates a parameter such as circuit leakage to another parameter, such as the circuit's frequency or temperature, and uses a third variable if necessary. For example, incorporating body bias or temperature can enhance the test's sensitivity. Multiple-parameter testing is a low-cost alternative to the methods described in the sidebar for discriminating fast, intrinsically leaky ICs from defective ones. This method has a good signal-to-noise ratio for defect detection, defining the leakage of defective ICs as the signal, and the background leakage as noise.

During characterization and test development, we measure and plot leakage (I_{DDQ}) and maximum operating frequency (F_{MAX}) against each other for many ICs. Figure 2 shows a trend line superimposed on the measured data, along with an adjustable limit line. We determine the leakage limit, which is frequency dependent, by statistical analysis; the dependency trend line defines the limit line's shape (see Figure 2). Then, we determine whether a measured IC is defective, depending on where it lies on our Figure 2 graph with respect to the leakage limit line. If, for a given frequency, an IC has substantially higher leakage than forecast by the intrinsic dependency, we classify that IC as defective. The IC circled in Figure 2 is only questionable, because its leakage is not *substantially* higher than expected. This IC is a candidate for further examination.

Thus, the two-parameter test limit distinguishes fast and slow dies from defective ones. This test improves the signal-to-noise ratio for defect detection for high-performance ICs with high background (intrinsic) leakage levels. If the determinations are doubtful, we can seek other variables to enhance the test's defect discrimination sensitivity. We must first establish the frequency-dependent leakage limit during the test development. Then, in the actual test, we measure each IC's frequency and leakage as a single-point measurement. Next, we compare this single-IC data against the relationship (the frequency-dependent limit established beforehand). No parameter sweep is necessary at test. A multiparameter test method can use any two parameters—or any number of parameters.

We originally proposed a two-parameter test[1] that measured I_{DDQ} and F_{MAX} parameters and compared them to a precharacterized, already-established I_{DDQ}-versus-F_{MAX} curve (similar to Figures 1 and 2). The channel length of

the ICs measured in Figure 2 was intentionally skewed smaller during fabrication to increase the leakage, and the data comes from multiple wafers and lots. Consequently, the data in Figure 2 incorporates a much broader range of variation than the data in Figure 1, which consists only of the natural range of variation in the unmodulated baseline parameters. The curve in Figure 2 would have been linear (in a semilog plot) if we had plotted it for a more limited range in frequency, if we had not skewed the channel length, or if we had only considered natural parameter variation.

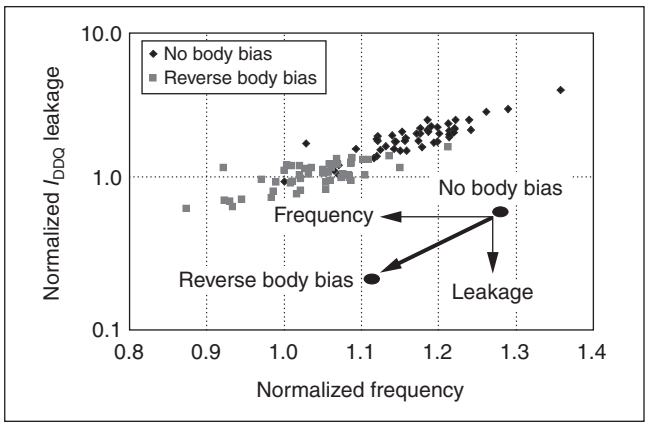

Figure 3. IC leakage versus frequency with and without reverse body bias of 0.5 V at room temperature (27.7 °C). RBB can increase the sensitivity of multiparameter testing.

Parameters improving test sensitivity: Body bias and temperature

Because the leakage-to-frequency correlation is intrinsic, varying transistor, circuit, and environment parameters such as temperature and body bias causes predictable changes in this dependency. We have used this concept to improve the sensitivity of the two-parameter test. RBB lowers an IC's leakage and reduces its speed.[1] Lowering the temperature increases the transistor switching speed and reduces its leakage current. Applying temperature, RBB, or both, as additional parameters improves the signal-to-noise ratio of our I_{DDQ}-versus-F_{MAX} test.

RBB alters the fundamental I_{DDQ}-versus-F_{MAX} relationship and statistics. Figure 3 shows the shift in RO leakage and frequency resulting from the application of 0.5 V of RBB to the ICs of Figure 1. The arrows in Figure 3 represent the direction of the shifts in speed and leakage resulting from applying RBB. On average, for all the ICs we tested, applying 0.5 V of RBB reduced leakage by 1.8 times while reducing speed by 10%. Because defective ICs respond differently to RBB, this parameter can function as a sensitivity knob, as we have discussed in other publications.[2]

However, RBB does not scale with technology and hence is not very effective. RBB provides minimal leakage reduction, and hence has limited application for our 150-nm technology and beyond.[3]

We next studied temperature as a possible additional parameter to improve test sensitivity. Figure 4 plots leakage versus frequency of the same ICs as a function of two temperatures. The arrow in Figure 4 shows the direction of leakage and speed change as the ICs cool.

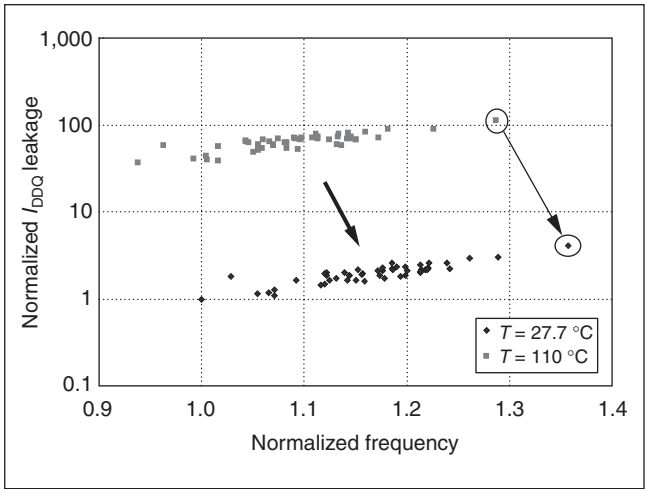

Figure 4. IC leakage versus frequency at room temperature (27.7 °C) and at a hot temperature (110 °C), without reverse body bias. Temperature is another possible parameter for multiparameter testing. The circled IC is defective.

Now we'll use a defective IC to demonstrate temperature's defect detection capability, showing how this sensitivity parameter improves our test's signal-to-noise ratio. We emulated a defective IC by adding a bridge 1-MΩ resistor between V_{DD} and V_{SS}. Figure 5 (next page) shows a good separation between the intrinsic populations of ICs at two tempera-

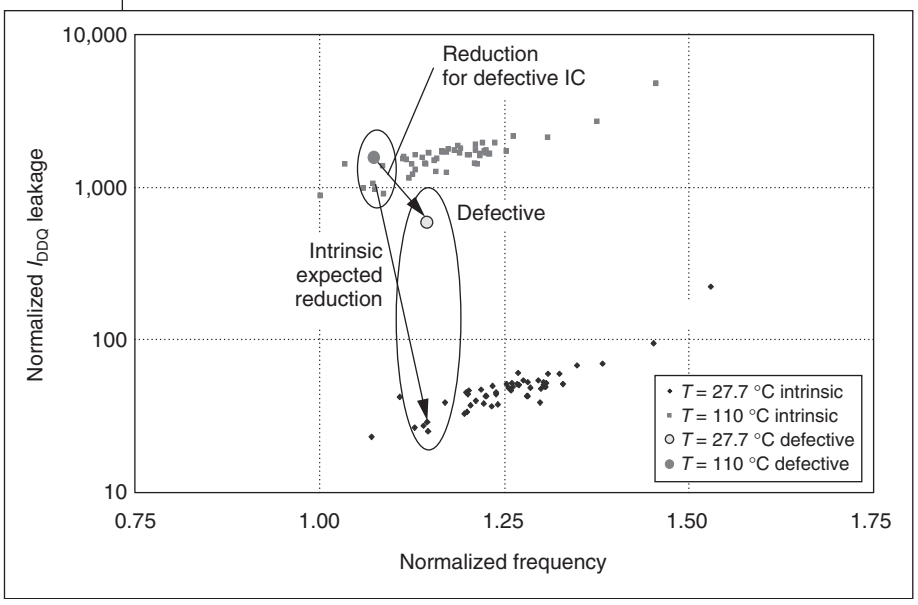

Figure 5. Defect sensitivity improvement by temperature. Defective and defect-free IC leakage versus ICs' measured frequency at 27.7 °C and 110 °C, without applied body bias.

tures studied for defect sensitivity improvement. No intrinsic data overlap (coming from natural variations in frequency versus leakage) occurred for the two temperatures—room temperature (27.7 °C) and hot (110 °C).

The leakage of the original IC circled in Figure 5 for our defect sensitivity study reduced intrinsically by 36 times when we dropped the temperature from hot to room temperature. When we added the defect to this IC, the leakage of this emulated defective IC, shown in Figure 5, increased by 1.6 times because of the extra leakage path between V_{CC} and V_{SS}. The defective IC's data point still belonged to the hot leakage-versus-frequency population plot, making it a challenge to detect purely by the adjustable limit concept. In other words, two-parameter testing with the proposed adjustable frequency-dependent leakage limit lacked the necessary sensitivity to detect and isolate this defective IC. However, when we lowered the temperature, the defective IC's leakage decreased by a factor of only 2.6, so that at room temperature, the defective IC remained outside the main population of frequency-versus-leakage behavior.

We quantified the gain in the test sensitivity by taking the ratio of intrinsic leakage reduction (36 times) to the amount of leakage reduction for the defective IC (2.6 times). The signal-to-noise ratio improved by more than an order of magnitude (36×/2.6× = 13.8).

Variance reduction

Nearest-neighbor current averaging attempts to address the problem of excessive parameter variance and elevated leakage.[4] Fundamentally, this technique uses location and geometric pattern information to establish estimators and residuals as a new test parameter. The nearest-neighbor residual technique (NNR) complements the multiple-parameter test concept. Here, we investigate NNR and apply it to the leakage-versus-frequency technique. NNR subtracts multiple leakage measurements on the circuit under test (the average of several I_{DDQ} vectors, given in Figure 6a) from the median of the measured leakage of adjacent neighboring dies (see Figure 6b); this local region, an average of eight neighboring dies, accounts for background leakage.

We use the resulting computed residual leakage (given in Figure 6c), which is ideally frequency independent, in our test decision criterion. Test limit selection is rather simple, using a single threshold leakage value with no need for an adjustable test limit. Figure 6c, which plots NNR as a function of frequency, exposes more defective ICs as compared to Figure 6a. Thus, NNR improves sensitivity by taking into account locality and the intrinsic leakage of neighboring dies, and so highlights current increases due to the presence of defects. Hence, NNR improves the limitations posed by excessive parameter variation.

Adaptive body bias

So far, our main focus in this article has been on the defect detection capability of several proposed test solutions for future scaled technologies. In these solutions, we have primarily used leakage versus frequency with a frequency-adjusted leakage test limit. We developed this adjustable test limit to let us correctly deter-

mine the functionality of fast (and hence leaky) dies, rather than discarding them on the basis of leakage only. Thus, these techniques indirectly improve a process' yield and bin split (the number of high-frequency parts).

We have shown the application of a frequency-adjusted test limit when no body bias is used, as well as the improvement made possible by using a constant reverse body bias. However, as we mentioned, RBB becomes less effective as technology scales to smaller feature sizes.

Recently, we studied the application of forward body bias (FBB) as an enabler for scaling CMOS technologies.[5] FBB improves transistor short-channel effects and reduces parameter variations; it also modulates IC leakage more effectively than RBB. Therefore, a constant FBB, similar to our application of RBB, can improve the leakage-versus-frequency test sensitivity. This technique could become an attractive replacement for temperature as a means of improving test sensitivity.

Figure 7 (next page) shows measurements for 62 dies at high temperature in a 150-nm CMOS technology. The figure shows three distributions: no body bias, 500-mV RBB, and 400-mV FBB. Both RBB and FBB affect the leakage and frequency of the die distribution, but the effect of FBB is much more significant. When we apply 400 mV of FBB to each die on the wafer, we completely separate the leakage-versus-frequency distributions, improving our signal-to-noise ratio for defect detection.

An advantage of using body bias (either RBB or FBB) rather than temperature for defect isolation is that the external body bias can change a part's frequency and the leakage, compensating for the effects of process parameter variations. Each fabricated die must meet two constraints: one on frequency and one on total power. The frequency constraint represents the part's minimum operating frequency. Any die with a frequency below this minimum value cannot be sold. At the same time, each die must meet a maximum power constraint, dictated by the system's thermal design. Any die that violates this maximum power constraint must be discarded as well. Rather than rejecting dies that do not meet one of these two constraints, manufacturers can use body bias to modulate their frequency and leakage power. This is the motivation behind the adaptive body bias (ABB) technique.[6]

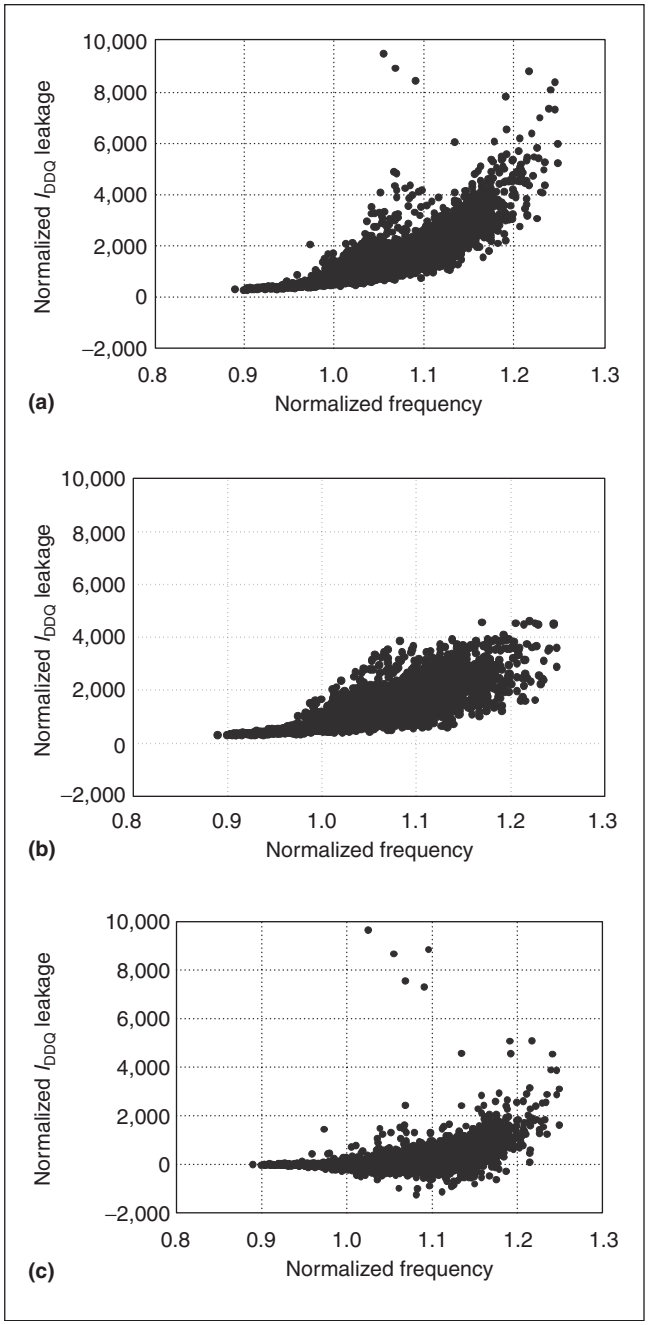

Figure 6. The nearest-neighbor residual (NNR) technique improves the sensitivity of multiparameter testing. Average leakage versus frequency (speed) for several I_{DDQ} vectors (a) is subtracted from median local neighborhood leakage versus frequency (speed) for eight local neighbors (b), to yield NNR leakage versus frequency (speed) (c).

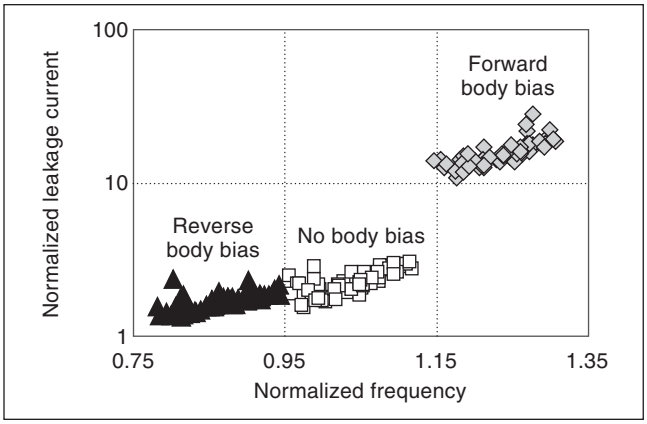

Figure 7. Leakage versus frequency for no body bias, reverse body bias, and forward body bias. FBB more effectively separates the die populations.

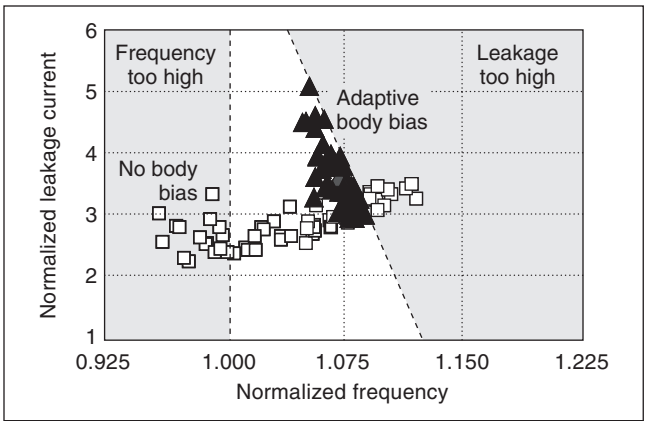

Figure 8. Leakage versus frequency, for no body bias and adaptive body bias: An application of ABB makes more of the dies acceptable.

Figure 8 shows the same 62 dies as in Figure 7. In the population with no body bias (NBB), dies on the left of the distribution violate the minimum frequency constraint and are not acceptable. Dies on the right, above the leakage limit line, violate the total power constraint and must be rejected as well. This example uses a total power density limit of 20 W/cm^2, typical of a mobile microprocessor design. The leakage limit line is sloped because, for a given fixed total power limit, dies with higher frequency must have lower leakage to meet the constraint. Therefore, only 50% of these dies are initially acceptable.

Applying the ABB scheme involves finding for each die on the wafer the body bias value (for both PMOS and NMOS transistors) that will maximize the die frequency while meeting the two constraints. We can apply this body bias using an on-chip generator or an off-chip source. Figure 8 shows the result of applying this optimum body bias to each die, in the group labeled ABB. The slow dies receive FBB to increase their speed, whereas the leaky dies receive RBB to reduce their leakage. As a result, 100% of the dies become acceptable, and the mean die frequency improves. In addition, this technique reduces the frequency variation (σ/μ, where σ is the standard deviation, and μ is the mean representing the center of the population) from 4.1% for NBB to 0.7% for ABB.

Thus, if designs include body bias capability, manufacturers can improve defect detection by using a constant body bias during testing, and they can improve bin split by using an adaptive body bias to maximize frequency. Of course, defect isolation must occur before application of adaptive body bias—otherwise each die would receive a unique body bias, making defects difficult to detect.

ELEVATED LEAKAGE and parameter variations are inherent elements of aggressively scaled devices and technologies. Our test methods exploit the intrinsic dependencies of transistor and circuit leakage on circuit operating frequency, temperature, and body bias to differentiate between normal fast ICs and defective ones. We have used these fundamental dependencies to develop sensitive correlative multiparameter test techniques to scale along with challenges imposed by technology and transistor advancements. Our adaptive body bias technique gives us a means of improving bin split in the environment of increased leakage and parameter variations resulting from continued process scaling. ■

Acknowledgments

We thank Shekhar Borkar, Brad Bloechel, and Jaume Segura for valuable discussions. Kaushik Roy's research was partially supported by the Marco Gigascale Silicon Research Center.

References

1. A. Keshavarzi, K. Roy, and C.F. Hawkins, "Intrinsic Leakage in Low-Power Deep Submicron CMOS ICs," *Proc. Int'l Test Conf.* (ITC 97), IEEE Press, Piscataway, N.J., 1997, pp. 146-155.
2. A. Keshavarzi et al., "Multiple-Parameter CMOS IC Testing with Increased Sensitivity for I_{DDQ}," *Proc. Int'l Test Conf.* (ITC 00), IEEE Press, Piscataway, N.J., 2000, pp. 1051-1059.
3. S. Yang et al., "A High-Performance 180-nm Generation Logic Technology," *IEDM Tech. Dig.,* 1998, p. 197.
4. W.R. Daasch et al., "Variance Reduction Using Wafer Patterns in I_{DDQ} Data," *Proc. Int'l Test Conf.* (ITC 00), IEEE Press, Piscataway, N.J., 2000, pp. 189-198.
5. A. Keshavarzi et al., "Forward Body Bias for Microprocessors in 130-nm Technology Generation and Beyond," *Proc. VLSI Circuits Symp.,* IEEE Press, Piscataway, N.J., 2002, pp. 312-315.
6. J. Tschanz et al., "Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage," *Proc. Int'l Solid-State Circuits Conf.* (ISSCC 02), IEEE Press, Piscataway, N.J., 2002, pp. 422-423.

Ali Keshavarzi is a staff engineer and senior researcher at Intel's Microprocessor Research Laboratories (MRL), Portland, Oregon. His research interests include low-power and high-performance circuit techniques and device structures for future generations of microprocessors. Keshavarzi has a PhD in electrical engineering from Purdue University.

James W. Tschanz is a circuits researcher at Intel Laboratories in Hillsboro, Oregon. He is also an adjunct faculty member at the Oregon Graduate Institute in Beaverton. His research interests include low-power digital circuits, design techniques, and methods for tolerating parameter variations. Tschanz has an MS in electrical engineering from the University of Illinois at Urbana-Champaign.

Siva Narendra is a staff engineer at Intel Laboratories. He is also an adjunct faculty member in the Department of Electrical and Computer Engineering, Oregon State University, Corvallis. His research areas include low-voltage MOS circuits and the effect of MOS parameter variation on circuit design. Narendra has a PhD in electrical engineering and computer science from Massachusetts Institute of Technology.

Vivek De is a principal engineer and manager of low-power circuit technology research at Intel's MRL. His research interests are in the area of low-power design. De has a PhD in electrical engineering from Rensselaer Polytechnic Institute.

Kaushik Roy is a professor of electrical and computer engineering at Purdue University. His research interests include VLSI design and CAD, VLSI testing and verification, and reconfigurable computing. Roy has a PhD in electrical and computer engineering from the University of Illinois at Urbana-Champaign.

Charles F. Hawkins is a professor in the Electrical and Computer Engineering Department at the University of New Mexico. His research interests include IC electronics, VLSI design and testability, IC reliability, and IC failure analysis. Hawkins has a PhD in electrical engineering from the University of Michigan.

W. Robert Daasch is an associate professor of electrical and computer engineering and codirector of the Integrated Circuits Design and Test Laboratory at Portland State University. His research interests include digital and analog IC design and test. Daasch has a PhD in physical chemistry from the University of Washington.

Manoj Sachdev is a professor in the Electrical and Computer Engineering Department at the University of Waterloo, Canada. His research interests include low-power and high-performance digital circuit design, mixed-signal circuit design, and test and manufacturing issues for ICs. Sachdev has a PhD in electrical engineering from Brunel University, UK.

■ Direct questions and comments about this article to Ali Keshavarzi, Mail Stop EY2-07, 5350 NE Elam Young Parkway, Hillsboro, OR 97124-6461; ali.keshavarzi@intel.com.