25.7 Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage

James Tschanz, James Kao[1], Siva Narendra, Raj Nair, Dimitri Antoniadis[1], Anantha Chandrakasan[1], Vivek De

Microprocessor Research Labs, Intel Corporation, Hillsboro, OR
[1]Massachusetts Institute of Technology, Cambridge, MA

Die-to-die and within-die (WID) device parameter variations, which are becoming worse with technology scaling, impact clock frequency and leakage power distributions of microprocessors in volume manufacture [1]. In particular, they limit the percentage of processors that satisfy both minimum frequency requirement and maximum active power (switching and leakage) constraint. Their impacts are more evident at the low supply voltages (V_{CC}) used in mobile processors where the power budget is limited by constraints imposed by heat removal, power delivery and battery life.

A test chip is implemented in a 150nm CMOS technology to evaluate effectiveness of the bidirectional (forward and reverse) adaptive body bias (ABB) technique for minimizing impacts of parameter variations on processor frequency and active leakage power (Figure 25.7.1) [2]. The test chip contains 21 "subsites" distributed over 4.5x6.7mm^2 in two orthogonal orientations. Each subsite has: (i) a circuit block (CUT) containing key circuit elements of a microprocessor critical path; (ii) a replica of the critical path whose delay is compared against an externally applied target clock frequency (ϕ) by a phase detector; (iii) a counter which updates a 5b digital code based on the phase detector output; and (iv) a "resistor-ladder D/A converter + op-amp driver" which, based on the digital code, provides one of 32 different body bias values to pMOS transistors in both the CUT and the critical path delay element (Figure 25.7.2). For an externally applied nMOS body bias, this on-chip circuitry generates the pMOS body bias that minimizes leakage power of the CUT while meeting a target clock frequency, as demonstrated by measurements (Figure 25.7.3). Different ranges of unidirectional – forward (FBB) or reverse (RBB) – or bidirectional body bias values can be selected by using appropriate values of V_{REF} and V_{CCA}, and by setting a counter control bit.

Clock frequency, switching power and active leakage power of the 21 CUTs per die are measured independently at 0.9V V_{CC} and 110°C, for 62 dies on a wafer. Die clock frequency is the minimum of the CUT frequencies, and active leakage power is sum of the CUT leakages. When no body bias (NBB) is used, 50% of the dies meet both the minimum frequency requirement and the maximum active leakage constraint set by a total power density limit of 20W/cm^2 (Figure 25.7.4). Using 0.2V forward body bias (FBB) allows all of the dies to meet the minimum frequency requirement, but most of them fail to satisfy the leakage constraint. As a result, only 20% of the dies are acceptable even though variations are reduced slightly by FBB due to improved short-channel effects [2].

Bidirectional ABB is used for both nMOS and pMOS devices to increase the percentage of dies that meet both frequency requirement and leakage constraint. For each die, a single combination of nMOS and pMOS body bias values is used to maximize clock frequency without violating the leakage power limit. As a result, die-to-die frequency variations (σ/μ) are reduced by an order of magnitude, and 100% of the dies become acceptable (Figure 25.7.4). In addition, 30% of the dies are now in the highest frequency bin allowed by the power density limit.

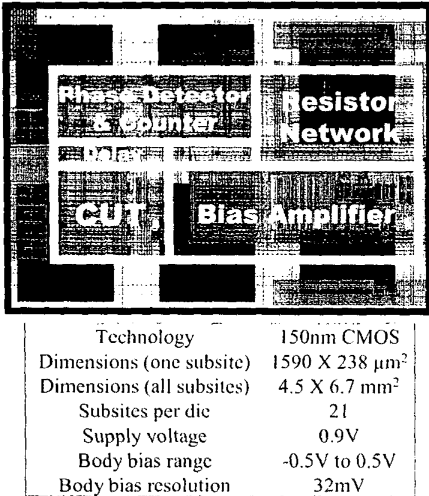
In a simpler ABB scheme, within-die variations can be neglected and the required body bias for a die can be determined from measurements on a single CUT [2]. However, test chip measurements show that as the number of critical paths (N_{CP}) on a die increases, WID delay variations cause both μ and σ of the die frequency distribution to become smaller (Figure 25.7.5). This is consistent with statistical simulation results indicating that the impact of WID parameter variations on die frequency distribution is significant [1]. As N_{CP} exceeds 14, there is no change in the frequency distribution with N_{CP}. Therefore, using measurements of 21 critical paths on the testchip is sufficiently accurate for obtaining frequency distributions of microprocessors which contain 100s of critical paths. Previous measurements on 49-stage ring oscillators show that σ of the WID frequency distribution is 4x smaller than s of the device saturation current (I_{ON}) distribution [2]. However, measurements on the test chip containing 16-stage critical paths show that σs of WID critical path delay distributions and nMOS/pMOS I_{ON} distributions are comparable (Figure 25.7.5). Since typical microprocessor critical paths contain 10-15 stages, and this number is reducing by 25% per generation, impact of within-die variations on frequency is becoming more pronounced [3]. This is further evidenced by the fact that the number of acceptable dies reduces from 100% to 50% in the simpler ABB scheme, which neglects within-die variations, although die count in the highest frequency bin increases from 0% to 11% compared with NBB.

The ABB scheme, which compensates primarily for die-to-die parameter variations by using a single nMOS/pMOS bias combination per die, can be further improved to compensate for WID variations as well. In this WID-ABB scheme, different body bias combinations are used for different circuit blocks on the die. A triple-well process is needed for nMOS implementation. For each CUT, the nMOS body bias is varied over a wide range using an off-chip bias generator. For each nMOS bias, the on-chip circuitry determines the pMOS bias that minimizes leakage power of the CUT while meeting a particular target frequency. The optimal nMOS/pMOS bias for the CUT at a specific clock frequency is then selected from these different bias combinations as the one that minimizes CUT leakage. This produces a distribution of optimal body bias combinations for the CUTs on a die at a specific clock frequency. If the die leakage power exceeds the limit at that frequency, the target frequency is reduced and the process is repeated until the maximum frequency is found where the leakage constraint is also met.

WID-ABB reduces σ of the die frequency distribution by 50%, compared to ABB (Figure 25.7.6). In addition, virtually 100% of the dies are accepted in the highest possible frequency bin, compared to 30% for ABB. Distribution of optimal nMOS/pMOS body bias combinations for a sample die in the WID-ABB scheme reveals that while RBB is needed for both pMOS and nMOS devices, FBB is used mainly for the pMOS devices (Figure 25.7.7). In addition, body bias values in the range of 0.5V RBB to 0.5V FBB are adequate. Finally, measurements (Figure 25.7.7) show that ABB and WID-ABB schemes need at least 300mV and 100mV body bias resolutions, respectively, to be effective. The 32mV bias resolution provided by the on-chip circuitry in the test chip is, therefore, sufficient for both ABB and WID-ABB.

Acknowledgements:
Authors thank S. Menon, B. Bloechel, M. Borkar, G. Dermer for measurements and lab support; J. Rattner, S. Borkar, R. Hofsheier, F. Pollack for encouragement and support.

References:
[1] K. Bowman et al., "Impact of Die-to-Die and Within-Die Parameter Fluctuations on the Maximum Clock Frequency Distribution", ISSCC Digest of Technical Papers, pp. 278-279, Feb. 2001.
[2] M. Miyazaki et al., "A 1000-MIPS/W Microprocessor using Speed-Adaptive Threshold-Voltage CMOS with Forward Bias", ISSCC Digest of Technical Papers , pp. 420-421, 2000.
[3] V. De, S. Borkar, "Technology and design challenges for low power and high performance", 1999 ISLPED Dig. Tech. Papers, pp. 163-168, 1999.



Technology	150nm CMOS
Dimensions (one subsite)	1590 X 238 μm²
Dimensions (all subsites)	4.5 X 6.7 mm²
Subsites per die	21
Supply voltage	0.9V
Body bias range	-0.5V to 0.5V
Body bias resolution	32mV

Figure 25.7.1: Chip micrograph (subsite).

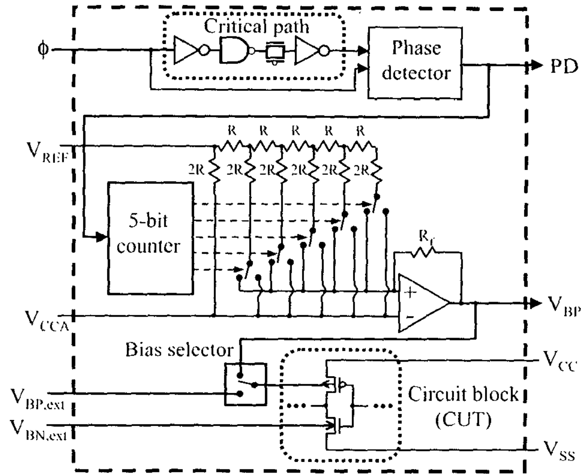
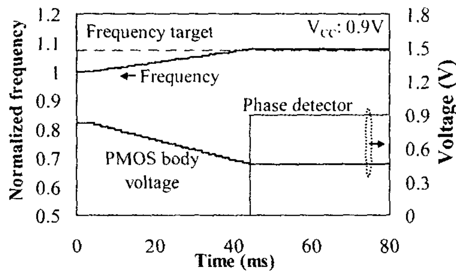


Figure 25.7.2: Block diagram of adaptive body bias circuit.



Bias Mode	Condition	Range
NBB → FBB	$V_{CCA} = V_{CC}$ $V_{REF} > V_{CCA}$	FBB: 0 → V_{REF}-V_{CCA}
NBB → RBB	$V_{CCA} = V_{CC}$ $V_{REF} < V_{CCA}$	RBB: 0 → V_{CCA}-V_{REF}
FBB → RBB	$V_{CCA} < V_{CC}$ $V_{REF} < V_{CCA}$	FBB: V_{CC}-V_{CCA} → RBB: $2V_{CCA}$-V_{REF}-V_{CC}

Figure 25.7.3: Adaptive body bias operation and bias modes.

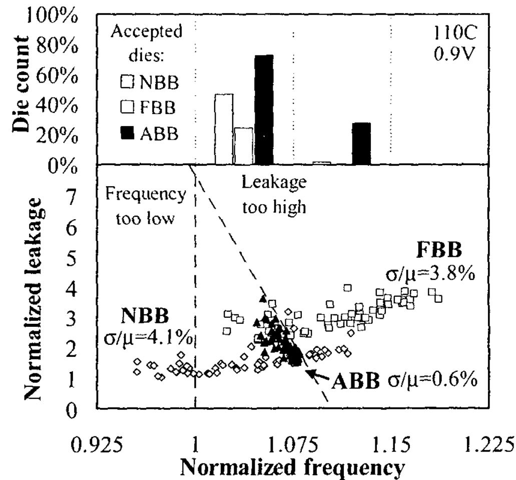


Figure 25.7.4: Variation in leakage and frequency for no body bias (NBB), 0.2V forward body bias (FBB) and die-to-die adaptive body bias (ABB).

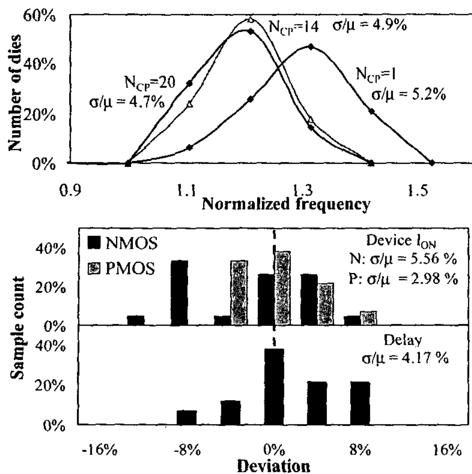


Figure 25.7.5: (a) Die frequency vs. number of critical paths. (b) Within-die I_{ON} and frequency variaation.

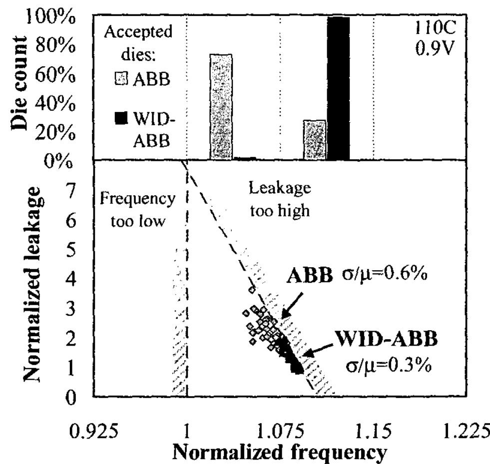
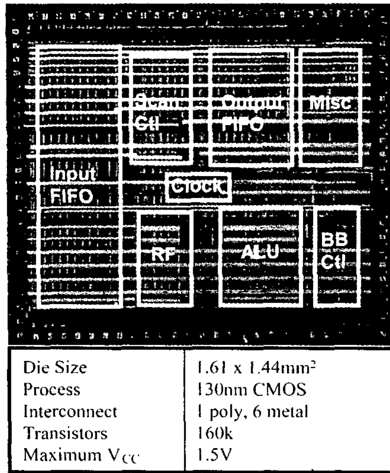


Figure 25.7.6: Variation in leakage and frequency for adaptive body bias (ABB) and within-die adaptive body bias (WID-ABB).

Continued on Page 478



Die Size	1.61 x 1.44mm²
Process	130nm CMOS
Interconnect	1 poly, 6 metal
Transistors	160k
Maximum V_{CC}	1.5V

Figure 25.2.7: Die characteristics.

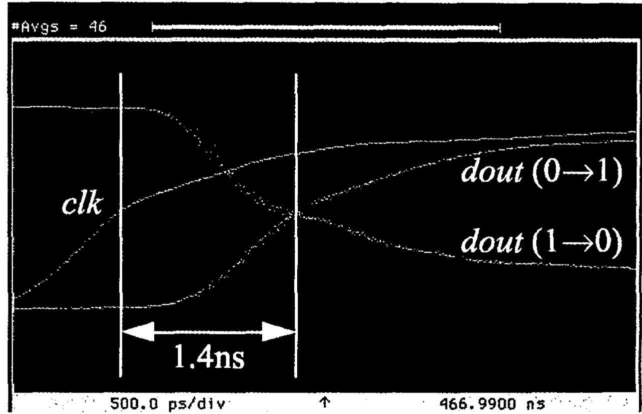


Figure 25.4.7: Oscilloscope waveform showing access time.

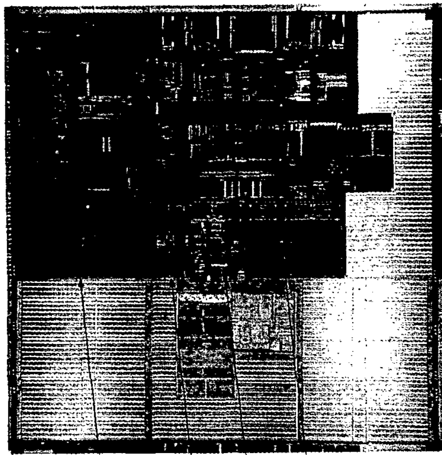


Figure 25.5.6: Die micrograph.

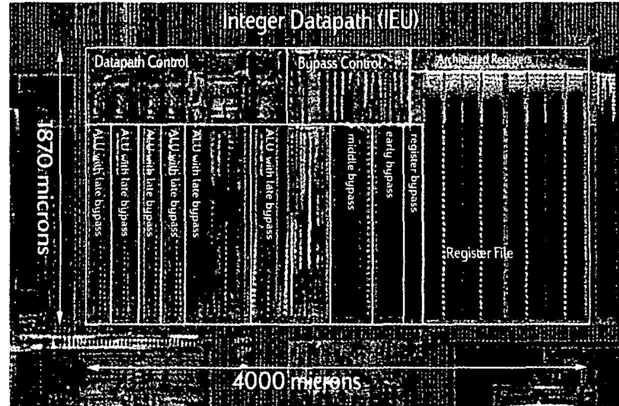
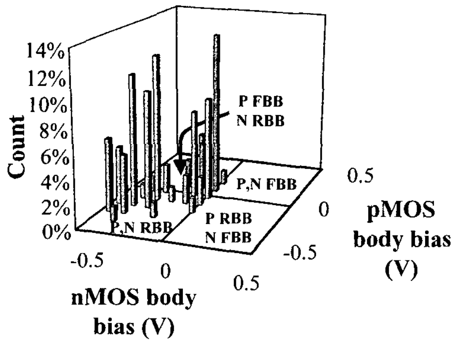


Figure 25.6.7: Die micrograph.



Bias resolution	Die-to-die ABB dies, F > 1	Die-to-die ABB σ/μ	Within-die ABB dies, F > 1.075	Within-die ABB σ/μ
0.5	79 %	2.87 %	2 %	1.89 %
0.3	100 %	1.47 %	66 %	0.50 %
0.1	100 %	0.58 %	97 %	0.25 %

Figure 25.7.7: (a) Histogram of applied bias voltage. (b) Impact of bias resolution on ABB effectiveness.

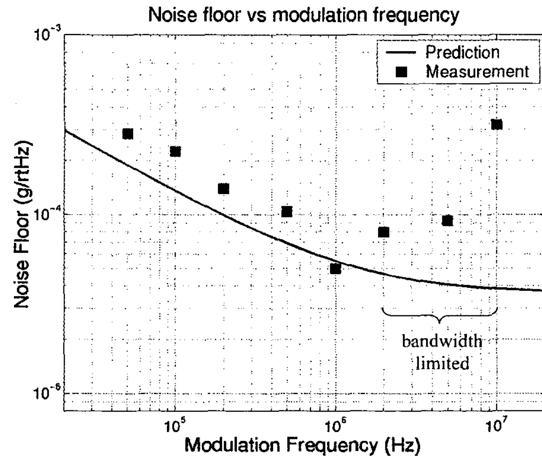


Figure 26.2.7: Measured and predicted noise floor versus modulation frequency.