Word-line and Charge-pump modeling of NAND Flash using Standard CMOS Logic Process

Hyun Sik Jeong[1] and Seong Hwan Cho[a]
Department of Electrical Engineering, KAIST
E-mail : [1]hyunsik.jeong@kaist.ac.kr

Abstract - **The industry has increased the word line (WL) layer to improve the density of 3D NAND flash. As the WL layer increases, the energy consumed to generate the boosted WL voltage has also increased. In 3D NAND flash, understanding the block configuration and operation of a word line driver (WL driver) with high energy consumption should be preceded to implement a 3D NAND flash with low energy consumption.**

This study presents a conventional WL driver for triple-level cell 3D NAND flash to understand the operation and classify energy consumption. The conventional WL driver for a 56-WL layer is fabricated in 180nm UHV process, and it consumes 141.15nJ from a 2.2V during 1 unit of program pulse and verify period.

***Keywords*—3D NAND flash, Charge pump, High voltage regulator, Voltage step-up system, Word line driver, WL driver**

I. INTRODUCTION

In the era of big data, 3D NAND flash has increased its density as a solution suitable for storage applications with large-capacity storage characteristic. PUC is a technique for placing peripheral circuits under a cell array [1]. Increasing the number of strings or stacking more WL layers increases the density [2]. A common method in the industry to increase the density is to stack more WL layers. Fig. 1 shows the trend of the number of WL stacks published as papers by companies over the past 5 years.

The increase in WL capacitive load due to more stacked WL layers increases energy consumption proportionally during 3D NAND flash operation. Low energy consumption is one of the important factors to consider in applications (e.g. mobile device, data center, PC, and other electronic devices) using 3D NAND flash. So it is essential to reduce energy while increasing density.

In the following, the structure and core operation (i.e. read and program) of the triple-level cell 3D NAND flash are presented, and the operation of the conventional WL driver is presented.

A 3D NAND flash cell array consisting of 176 main WLs and 8Kbyte bit lines (BLs), including WLs (SSL and DSL) for string selection, is shown in Fig. 2 [2-3]. The core operation of the read and program is to supply a boosted voltage to a one-WL for the selected cells and access the data of the cell connected to the 8Kbyte BLs. 175-WLs for unselected cells act as switches by supplying boost voltages not to affect the operation of the selected cells.

Fig. 3 shows the timing diagram of read and program operation [4-5]. Since the cell's threshold voltage (cell Vth) of the triple-level cell 3D NAND flash does not exceed 5V, the main WL voltage supplied to the unselected cell's gate is 5V or higher. Therefore, charge pumps and high voltage regulators (HV regulators) are used to supply the boosted voltage to the WL, and energy consumed by the charge pump is large due to low energy efficiency.

Fig. 4 shows the energy consumed in the WL driver of the 3D NAND flash. The 3D NAND flash consumes different energy depending on the read and program operation and how the 3D NAND flash with multiple strings supplies voltage to the WL and BL. According to the control methods used in 3D NAND flash products in the industry, WL drivers consume up to 50%. Therefore, it is essential to reduce the energy of WL drivers to reduce the energy of 3D NAND flash.

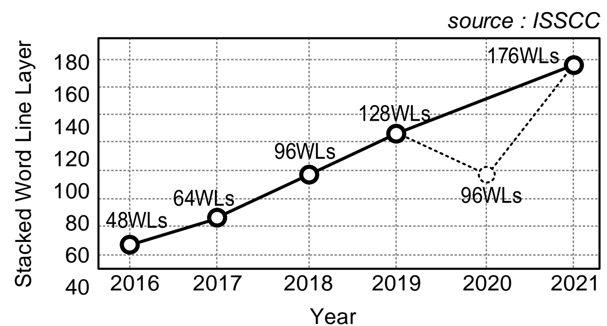


Fig. 1. Trend of stacked WL layer of 3D NAND flash

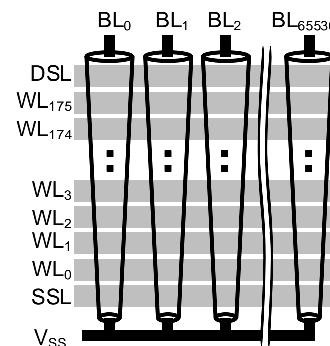


Fig. 2. A string cell array in 3D NAND flash

a. Corresponding author; chosta@kaist.ac.kr

Manuscript Received Jul. 14, 2021, Revised Sep. 26, 2021, Accepted Sep. 29, 2021

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/bync/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

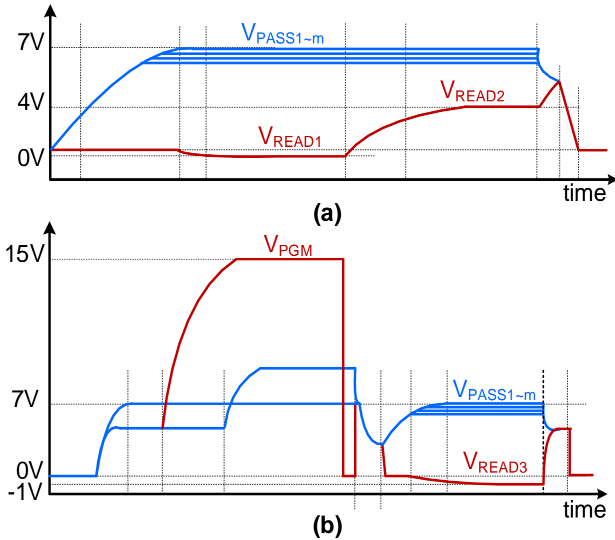


Fig. 3. 3D NAND flash operation (a) read and (b) program

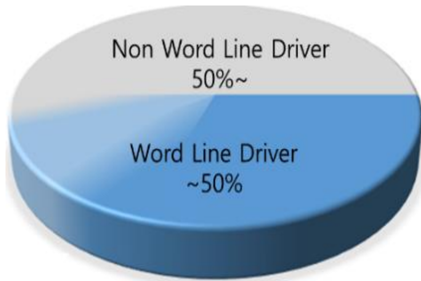


Fig. 4. Energy consumption during read and program operation of 3D NAND flash

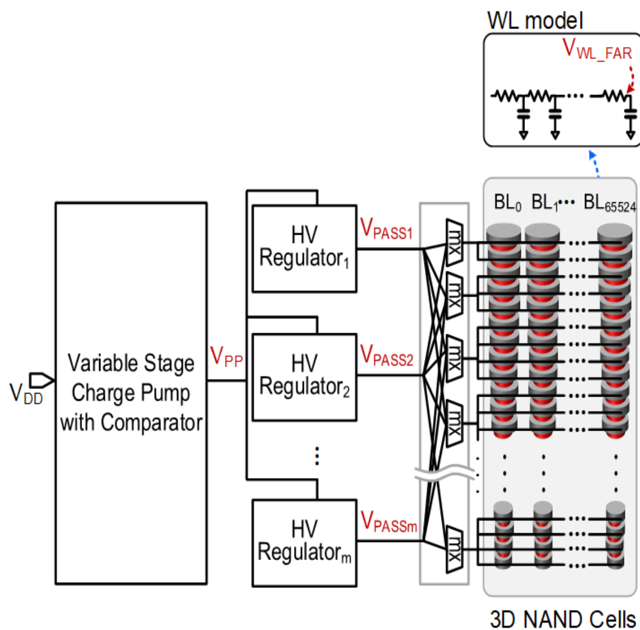


Fig. 5. Conventional word line driver and word line

In this study, to compare the energy improvement amount of the proposed WL driver in the future, the operation of the conventional WL driver is confirmed with the prototype charge pump and the modeled RC delay circuits in the standard CMOS logic process instead of the 3D NAND flash process.

II. DESIGN METHODOLOGY

A. *Conventional word line driver*

A block diagram of a conventional WL driver for supplying boosted voltage to unselected WLs is shown in Fig. 5 [6-7]. To generate V_{PP}, a comparator controls the pumping driver's clock of the multi-stage charge pump. V_{PP} is used as the supply voltage of HV regulators which supply the regulated voltage (V_{PASS}) to WLs.

3D NAND flash requires many HV regulators. This is because the channel thickness and gate area of the upper and lower layers of the 3D cell are different. When the same WL voltage is applied, the current drivability of the cell is different. Therefore, different WL voltages are supplied to the gate of the cell for the same current drivability.

The drop-out voltages of high voltage regulators is shown in Fig. 6. V_{PP} is determined by considering the dropout voltage of the HV regulator based on the highest value among the output voltages (V_{PASS}s) of the HV regulators. Therefore, except for the HV regulator that outputs the highest V_{PASS}, HV regulators consume unnecessary energy equal to the difference $\triangle V_{PASS}$ from the highest V_{PASS}.

V_{PP} is set to a voltage between $5V_{DD}$ (Maximum output voltage of 4-stage charge pump) and $6V_{DD}$ (Maximum output voltage of 5-stage charge pump) instead of $6V_{DD}$ to ensure current drivability of the charge pump and to reduce the energy dissipated in C_{VPP} during transients.

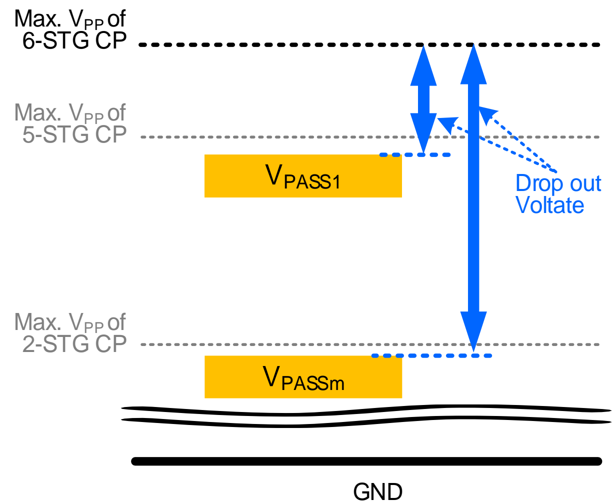


Fig. 6. Drop-out voltage of high voltage regulator

B. *Conventional variable stage charge pump*

The variable stage charge pump with comparator is shown in Fig.7. The logic controller of the 3D NAND flash controls the suitable number of the charge pump unit to be operated according to the predefined V_{PP} level. Note that the higher the V_{PP}, the more charge pump units operate. For example, if the signal E_4 supplied by the logic controller is 'high' and E_5~E_n and E_3 are 'low,' only CP_1~CP_4 among the charge pump units of the variable stage charge pump are operated. Note that operating the optimal pump stage is to reduce the energy consumed in the V_{PP}.

C. Conventional high voltage regulator

The WL with 8Kbyte BLs is modeled as a distributed RC circuits (e.g. total resistance of a WL=more than 100kΩ and total capacitance of a WL=more than 10 pF). Depending on the 3D NAND flash operation, several to tens of WLs are supplied by one HV regulator, and the load varies with the 3D NAND flash operation.

Fig.8 is a schematic and bode plot of a HV regulator [7]. To minimize current consumption, the supply power of the error amplifier is V_{DD} instead of V_{PP}. V_{PASS} is regulated using NMOS as the pass transistor. So, the V_{PASS} node has the third pole located at high frequency, and the first pole and the second pole are formed at node A and node B, respectively. The HV regulator uses NMOS as the pass transistor to form the second pole is as follows. In the initial stage of 3D NAND flash operation, the HV regulator is a transient state in which voltage is boosted to the WL. After that, the voltage of the WL is close to the desired V_{PASS}, and the far-end WL voltage becomes almost the same as the desired V_{PASS}. In the steady-state, the current supply of the HV regulator becomes unnecessary. Therefore, when PMOS is used as the pass transistor, the r_O of the output terminal becomes large, and a low-frequency pole is generated along with a large capacitor of WLs. Eventually, there are two poles in the low frequency, resulting in unstable frequency stability. However, NMOS as the pass transistor of the HV regulator reduces energy efficiency by requiring more charge pump stages due to the large drop-out voltage.

To use a PMOS as a pass transistor, add a source follower to the gate of the output PMOS to reduce the overall loop gain. However, it is not a desirable solution as the current used by V_{PP} increases and the offset voltage due to gain errors increases. It can also be stabilized by reducing the gain of the error amplifier, but this is also not a good solution. After all, HV regulators have no choice but to use NMOS as the pass transistor for frequency stability.

The output NMOS of the HV regulator is supplied with a high voltage V_{PP}, so use an NMOS that operates even at high voltage. High voltage depletion type NMOS and high voltage enhancement type NMOS are available.

When a high voltage depletion type NMOS is used as a pass transistor, it usually operates even if the voltage difference between V_{PP} and V_{PASS} is small. However, when V_{PASS} outputs a high voltage around 10V, V_{th} approaches 0V

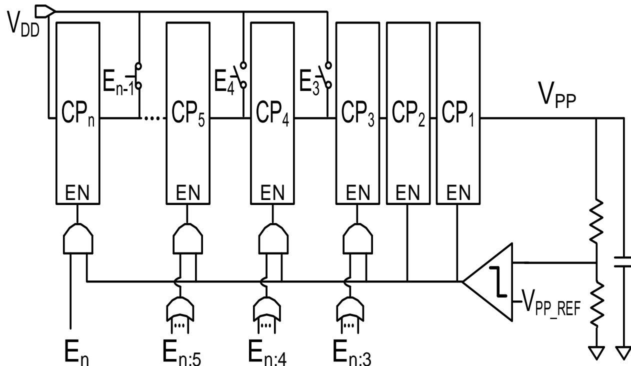


Fig. 7. Variable stage charge pump with comparator

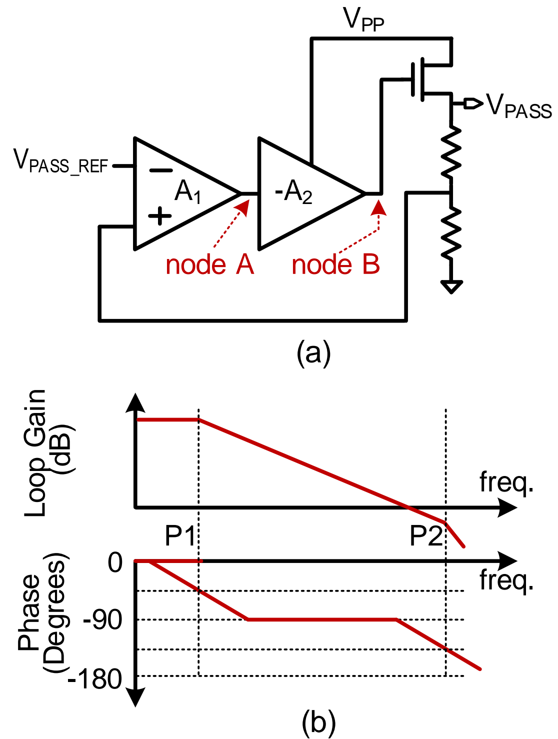


Fig. 8. High voltage regulator (a) schematic (b) bode plot

due to the bulk effect. The current drivability is small due to the characteristics of the high voltage depletion type NMOS, so the gate voltage must be increased to compensate for the current drivability. For this reason, the charge pump stage increases because V_{PP} must use a sufficiently high voltage compared to V_{PASS}. Moreover, it is difficult to output a low V_{PASS} voltage around 4V due to a negative V_{th}. Therefore, for low V_{PASS} voltage output, high voltage enhancement type NMOS should be used.

When a high voltage enhancement type NMOS is used as a pass transistor, V_{PP} and V_{PASS} must have $V_{th}+\Delta V$ of the NMOS for the HV regulator to operate normally. As a result, the HV regulator used for low V_{PASS} uses high voltage enhancement type NMOS, which more boosted V_{PP} than using high voltage depletion type NMOS. Eventually, it causes more energy consumption because an additional charge pump stage must be used.

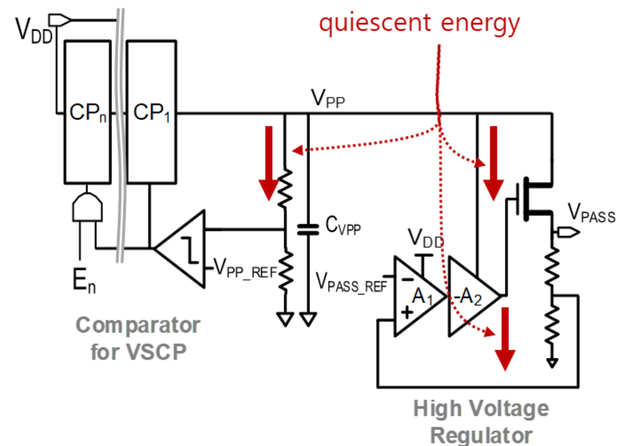


Fig. 9. Quiescent energy of comparator for charge pump and high voltage regulator

D. Quiescent energy of comparator for charge pump and high voltage regulators

Fig.9 shows the quiescent current consumed by the HV regulator and comparator to regulate the charge pump. Both circuits consume V_{PP} current, which draws V_{DD} current as a multiple of 'charge pump stage+1'. As mentioned in the previous section, it is essential to reduce the quiescent current of HV regulators because many HV regulators operate. In particular, when the 3D NAND flash performs a plane interleaving operation, it is essential to reduce the standby current of the HV regulator because the HV regulator operates in multiples of the number of using planes.

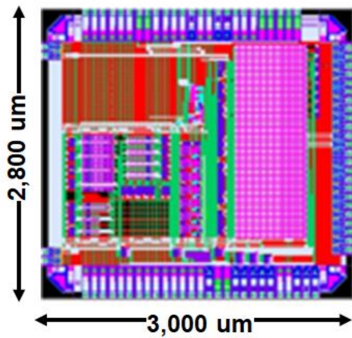


Fig. 10. Layout of WL driver and modeled WL for 3D NAND flash

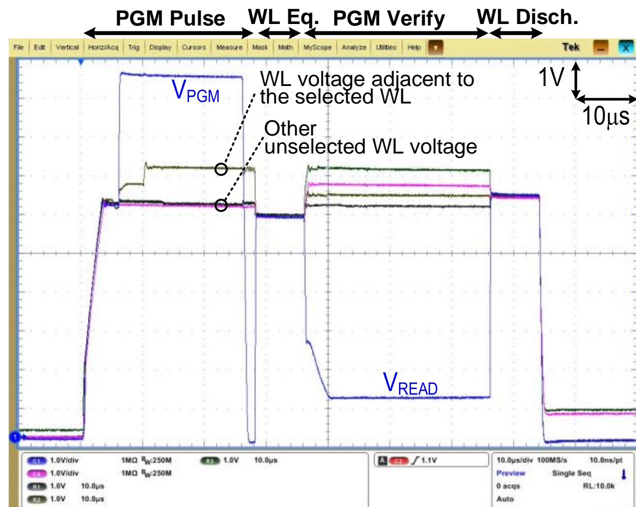


Fig. 11. Measured output voltages of high voltage regulators

TABLE I. Performance Summary

	Value
Process	180nm
V_{DD}	2.2 V
V_{PASS}	4~7 V
Maximum # of CP stage	5
C_{FLYING}	8 pF.
Frequency	40 MHz
$C_{WL\text{-}to\text{-}GND}$ / $C_{WL\text{-}to\text{-}WL}$	10.1 pF / 3.0 pF
R_{WL}	131 kΩ
Implemented WLs	56

TABLE II. Energy Breakdown of Conventional WL Driver

	Value (nJ)
Charge Pump for WLs	52.45
High voltage regulators	85.01
Reference Generators	3.7
Total	141.15

III. RESULTS AND DISCUSSIONS

The conventional WL driver has been fabricated in a 180nm UHV process. 56-WLs can be modeled as distributed RC circuits with a purely capacitive load as the gate leakage of the 3D NAND flash cell is very small. The capacitance values are C_{VPP}=14pF, C_{VPASS}=5pF for each V_{PASS}, and the flying capacitor of the charge pump is 8pF.

To generate 8.2V, the highest voltage among V_{PASS}, 5 stages of the variable stage charge pump operate to enerate a boosted V_{PP}. The output voltage of the seven V_{PASS}S is 4V to 7V.

To verify the functionality of the conventional WL driver for 3D NAND flash and obtain its energy consumption, program (PGM) command that complies with the industry standard is applied to the chip. During 28us, when a program pulse (i.e. write operation) is applied, the selected WL (9.4V), 2-WLs adjacent to the selected WL (7V), and the WL except for the edge WL (4V) are driven at 6V.

And during PGM verify whose duration is 30μs, all WLs except the selected WL (1V) and edge WLs (4V) are grouped equally into 4 and driven with 6V, 6.3V, 6.6V, and 7V. The WL equalization where the PGM pulse ends and the discharge period where the verify ends are each 8us. The voltage waveform during PGM pulse and PGM Verify is shown in Fig.11, and V_{PASS} is linearly regulated. The simulated energy consumption during this entire operation for the whole system is 141.15nJ. The performance summary of the conventional WL driver is shown in Table I. The energy breakdown of conventional WL driver is shown in Table II.

IV. CONCLUSION

In this study, a conventional WL driver for triple-level cell 3D NAND flash is implemented. During 1 unit of program pulse and verify period, seven HV regulators supplying boosted voltage to unselected WLs consume 85.01nJ. The energy consumed while supplying the boosted voltage to the modeled 56-WLs with 8Kbyte BLs of 1 string is 52.45nJ. As the number of BLs, strings, and stacked WLs increases, the energy consumed by the WL driver increases. To reduce the energy consumed in the WL driver, a method is needed to reduce the charge pump stage and lower the quiescent current of the HV regulators.

ACKNOWLEDGMENT

The chip fabrication and EDA tool were supported by the IC Design Education Center (IDEC) in Korea.

REFERENCES

[1] Hwang Huh et al., “A 1Tb 4b/Cell 96-Stacked-WL 3D NAND Flash Memory with 30MB/s Program Throughput Using Peripheral Circuit Under Memory Cell Array Technique,” 2020 IEEE International Solid - State Circuits Conference - (ISSCC), San Francisco, CA, 2020, pp. 220-221.

[2] Jae-Woo Park et al., “A 176-Stacked 512Gb 3b/Cell 3D-NAND Flash with 10.8Gb/mm^2 Density with a Peripheral Circuit Under Cell Array Architecture,” 2021 IEEE International Solid - State Circuits Conference - (ISSCC), San Francisco, CA, 2021, pp. 422-423.

[3] Ki-Tae Park et al., “Three-Dimensional 128 Gb MLC Vertical NAND Flash Memory With 24-WL Stacked Layers and 50 MB/s High-Speed Programming,” IEEE Journal of Solid-State Circuits - (JSSC), VOL. 50, NO. 1, January 2015, pp. 204-213.

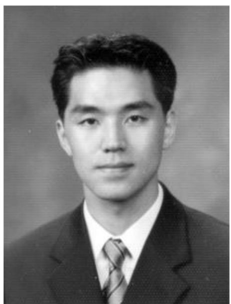
[4] Sam-Kyu Won et al., “High-voltage word line generator for low-power program operation in NAND flash memories,” IEEE Asian Solid-State Circuits Conference 2011 - (ASSCC), Jeju, South Korea, 2011, pp. 169-172.

[5] Jong Hak Yuh, "Techniques for the fast settling of word lines in NAND flash memory", US Patent App. 20120075931, SANDISK TECHNOLOGIES, Inc., 2012.

[6] Raymond Zeng et al., “A 172mm^2 32Gb MLC NAND flash memory in 34nm CMOS,” 2009 IEEE International Solid-State Circuits Conference - (ISSCC), San Francisco, CA, 2009, pp. 236-237

[7] Kim et al., "Soft Pre-charge H/V Switch for 3D Solid State Drive with NAND Flash Memories," *International SoC Design Conference,* pp. 339-342, June 2013.

Hyunsik Jeong received the B.S. and M.S. degree in electrical engineering from Hanyang University, South Korea, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree with the Korea Advanced Institute of Science and Technology, South Korea. His main interests are low-power and small area circuits for NAND Flash.



SeongHwan Cho (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1997 and 2002, respectively. In 2002, he joined Engim, Inc., Acton, MA, USA, where he was involved in data converters and phased-locked loop design for the IEEE 802.11abg WLANs. From 2011 to 2012, he was a Research Scientist with Marvell Semiconductor, Inc., Santa Clara, CA, USA. From 2016 to 2017, he was a Research Scientist with Google, London, U.K. Since 2004, he has been with the Department of Electrical Engineering, KAIST, where he is currently a Professor. His current research interests include analog and mixed-signal circuits for low-power communication systems, health-care devices, and CMOS sensors. Dr. Cho has served on the Technical Program Committee for several IEEE conferences, including the International Solid-State Circuits Conference (ISSCC), the Symposium on very-large-scale integrated (VLSI), and the Asian Solid-State Circuits Conference. He was a co-recipient of the 2009 IEEE Transactions on Circuits and System Society Guillemin-Cauer Best Paper Award and the 2012 ISSCC Takuo Sugano Award for Outstanding Far-East Paper. He has twice received the Outstanding Lecturer Award from KAIST. He was an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS— I. He was a Distinguished Lecturer of the IEEE Solid-State Circuits Society.