

19.6 A 40-to-80MHz Sub-4 μ W/MHz ULV Cortex-M0 MCU SoC in 28nm FDSOI with Dual-Loop Adaptive Back-Bias Generator for 20 μ s Wake-Up From Deep Fully Retentive Sleep Mode

David Bol¹, Maxime Schramme¹, Ludovic Moreau¹, Thomas Haine¹, Pengcheng Xu¹, Charlotte Frenkel¹, Rémi Dekimpe¹, François Stas², Denis Flandre¹

¹UCLouvain, Louvain-la-Neuve, Belgium, ²e-peas semiconductors, Louvain-la-Neuve, Belgium

Near-threshold circuits operating at ultra-low voltage (ULV) have matured with integration in commercial products such as ultra-low-power (ULP) MCUs for the IoT [1]. In this market, MCU design faces the key performance tradeoff between speed, active power, deep-sleep retention power and wakeup time, with the challenge of preserving it over PVT corners. We present a ULP MCU SoC in 28nm FDSOI codenamed SleepRunner, exploiting back biasing (BB) capability of FDSOI to push the performance tradeoff beyond the state of the art.

As shown in Fig. 1, the MCU logic at 0.4V (V_{DDL}) includes a Cortex-M0 CPU, an FFT accelerator, a wakeup interrupt controller (WIC) and various interfaces. A custom 32-kB ULP SRAM [2] is implemented as program memory (PMEM). It uses a custom single-ended 8-T bitcell based on negative-differential-resistance (NDR) MOS structures for ultra-low leakage and a 16-bank 512-bitcell/column divided-wordline architecture for ultra-low read access energy at 0.5V (V_{DDs}) of 1.6pJ per 32-bit word, at the cost of a limited density of 125kB/mm². Compared to the PMEM, the Cortex-M0 DMEM contributes less to the active power because of less frequent accesses. A high-density (HD) 32-kB foundry SRAM supplied at 0.8V (V_{DDH}) is thus selected as DMEM to limit die area. The main system clock (MCLK) with register-programmable frequency is generated on-chip from an external 12-MHz reference clock (REF_CLK). Embedded power management (ePM) includes switched-capacitor voltage regulators (SCVRs) to generate V_{DDL} , V_{DDs} and V_{DDH} internal supplies from the single 1.8V I/O voltage V_{DDIO} . They support $\pm 5\%$ output over/underdrive to expand the MCU frequency range from 48-72MHz (at nominal output voltage) to 40-80MHz.

To support 40-80MHz operation at ULV, LVT devices are used for logic and ULP SRAM. At ULV, upsizing their gate length (L_g) with poly biasing (PB) reduces the logic energy per cycle (E_{cycle}) at the minimum-energy point (MEP) thanks to improved subthreshold swing, lower DIBL and variability [3]. Fig. 2 shows that a 16-nm PB (PB16) yields an E_{cycle} close to 1pJ/cycle for frequencies below 10 MHz. Forward BB (FBB) applied to both the logic and ULP SRAM is used in active mode to shift the MEP with PB16 library to the target frequency range. SCVRs generating the ULV supplies from V_{DDIO} can only achieve high power conversion efficiency (PCE) in a limited output voltage range, which depends on their topology. We thus aim for 0.4V operation because it leads to efficient conversion with the divide-by-4 SCVR topology (V_{DDL}). To reach 72-MHz operation at 0.4V (Fig. 2), we use asymmetric FBB with a stronger PMOS BB (BBP) than NMOS BB (BBN), as PMOS are slower and have a lower back bias effect, with nominal (i.e. TT corner at 25°C) BBN/BBP of +1V/-2V. However, this high FBB level leads to 30- μ W logic leakage, which is fine for active mode (30% of the logic power) but prohibitive for deep-sleep mode. Therefore, BBN and BBP are driven by the on-chip FBB drivers to 0V in deep-sleep mode to kill leakage power, while preserving full state retention in both logic and ULP SRAM, which avoids SW initialization at wakeup.

To avoid conservative V_{DD} guardband for robust operation over PVT corners, we perform unified frequency/back bias regulation (UFBR) with an adaptive BB generator that dynamically tunes the BB to the current PVT conditions. Moreover, independent BBN and BBP regulation by two distinct loops enables preserving balanced rise/fall transitions in skewed process corners (e.g. slow NMOS/fast PMOS) [4]. This dual-loop UFBR is a key ingredient of SleepRunner MCU as its locking time sets the MCU wakeup time and its in-lock power contributes to the MCU active power. Fig. 3 shows the dual-loop UFBR architecture. BBN is tuned in a frequency-locked loop (FLL) to enable fractional ratios between MCLK and REF_CLK frequencies at low power. A counter senses the frequency of a critical-path replica tunable ring oscillator (TRO) used as both MCLK generator and f_{max} sensor [3,5,6], thus performing UFBR tolerant to V_{dd} droops. The actuation on BBN is based on a PWM-modulated charge pump (CP) with proportional control for fast startup. BBP is tuned in a delay-locked loop (DLL) with digital N/PMOS process imbalance sensors to detect skewed Faster NMOS/Slower PMOS (NFPS) conditions or the opposite (NSPF). A fast inaccurate version is used at startup and a slow accurate one is used when the lock is detected. The actuation is performed by a switched-cap negative CP (as in [5] but with 2 stages to support wide range asymmetric FBB) and a discharge pump (DP), with bang-bang control. Both BBN and BBP drivers are designed with I/O devices for wide-range BBN (0 to 1.8V) and BBP (0 to -3V) to fully compensate for PVT corners.

The challenge in the dual-loop design compared to AVS-based unified frequency/voltage regulation (UFVR) [3] is to ensure stable and fast lock despite their inter-dependency. Indeed, the large triple-well cap inside the logic and ULP SRAM tightly couples BBN and BBP nodes, which leads to instability. Fig. 3 shows that the use of an external decap on BBN ensures stable startup. A 2-nF decap (5 \times the parasitic coupling cap) avoids BBN overshoot, which speeds up the lock and consequently slightly reduces the wakeup energy. Fig. 3 also shows that this architecture can generate independent asymmetric BBN/BBP voltages to compensate for skewed corners.

For BBP regulation, previous N/PMOS process imbalance sensors were based on analog current comparison, which leads to a hard trade-off between DC power and response time [5,6]. In our BBP DLL, we designed a digital N/PMOS process imbalance sensor (Fig. 4) to overcome this challenge, while using only digital standard cells for good correlation to the logic to be calibrated. It is based on two delay lines with selective sensitivity to N/PMOS logic delay. Comparison of their delay provides binary information on the relative NFPS/NSPF conditions. The N/PMOS selective sensitivity is made by alternating strong ($\times 38$) and weak ($\times 2$) driving cells such that the low current of the weak cells has to charge the large input cap of the strong cells. Transistor stacking in NAND2/NOR2 with shorted inputs as weak cells further improves the selective sensitivity. However, the weak cells suffer from a high local V_t mismatch, which can lead to strong die-to-die (D2D) variations in the resulting closed-loop BBP voltage and thus significant leakage overhead (Fig. 4). We limit this effect in the accurate sensor (used in lock) through averaging by increasing the number of delay stages at the cost of slower sensing. The fast and accurate sensor versions can be sampled up to 6 and 3 MS/s for fast startup and loop response, respectively. Bang-bang control in the BBP DLL can lead to spurious CP activity in lock and thus MCU active power overhead. To avoid this, we added a small deadband in the process imbalance sensor with additional delays in the sensor comparison logic (Fig. 4). MCU timing closure was performed by taking these BBP deadband and variability into account.

SleepRunner was manufactured in 28nm FDSOI (Fig. 7). Measurement results in Fig. 5 confirm the FBB capability to preserve the MEP over a frequency range, while reaching 3 μ W/MHz (at least 2.9 \times lower than previous MCUs with similar speed and memory). Deep-sleep retention power is dominated by the HD SRAM, which can further be power gated by the WIC at the cost of a SW initialization at wakeup. Compared to [4,6], SleepRunner closed-loop adaptive FBB generation achieves at least 10 \times faster startup (<20 μ s) with full clock generation and deep-sleep mode integration. Finally, comparison in Fig. 6 shows that SleepRunner MCU achieves best-in-class power efficiency (280 DMIPS/mW), without compromising speed, deep-sleep full retention power nor wakeup time.

Acknowledgment

This work was supported by the Fonds européen de développement régional (FEDER) and the Wallonia within the « Wallonie-2020.EU » program and the Plan Marshall, as well as by the FRS-FNRS of Belgium.

References

- [1] AmbiqMicro, Apollo2 datasheet, on-line.
- [2] T. Haine et al., "An 8-T ULV SRAM macro in 28nm FDSOI with 7.4 pW/bit retention power and back-biased-scalable speed/energy trade-off", S3S conf., 2018.
- [3] D. Bol et al., "SleepWalker: 25-MHz 0.4-V Sub-mm² 7- μ W/MHz Microcontroller in 65-nm LP/GP CMOS", JSSC, vol. 48(1), pp. 20-32, 2013.
- [4] G. de Strel et al., "SleepTalker: A ULV 802.15.4a IR-UWB Transmitter SoC in 28-nm FDSOI", JSSC, vol. 52(4), pp. 1163-1177, 2017.
- [5] M. Blagojević et al., "A Fast, Flexible, Positive and Negative Adaptive Body-Bias Generator in 28nm FDSOI", Symp. VLSI Circuits, pp. 60-61, 2016.
- [6] A. Quellen et al., "A 2.5 μ W 0.0067mm² Automatic Back-Biasing Compensation Unit", ISSCC, pp. 304-305, 2018.

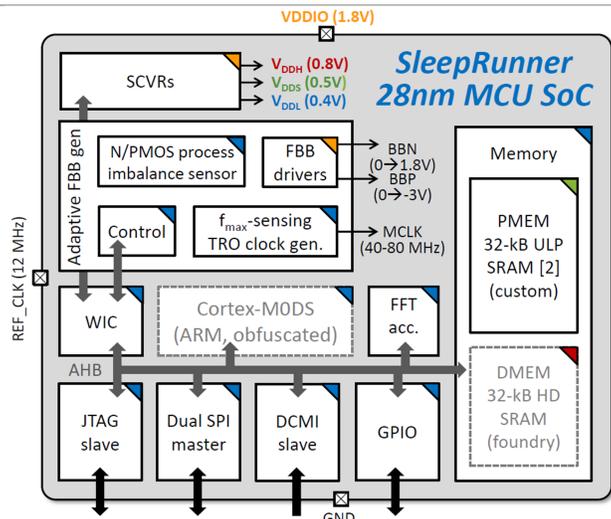
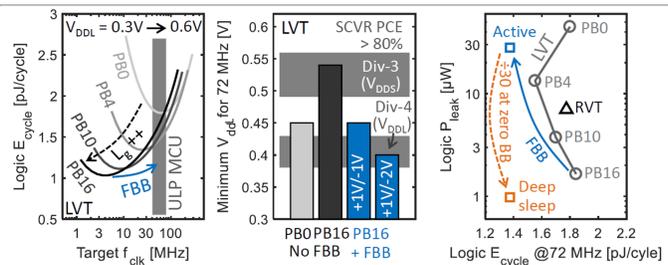


Figure 19.6.1: MCU SoC architecture showing the power domains.



Power mode	System clock MCLK	CPU clock	Logic & ULP SRAM	HD SRAM	SCVR range	Wake-up time
Active	40-80 MHz	40-80 MHz	Adapt. FBB	On	3x140μW	-
CPU sleep	40-80 MHz	Gated	Adapt. FBB	On	3x140μW	1 cycle
Deep sleep (retention: 64 kB + logic incl. CPU)	750 kHz (divided REF_CLK)	Gated	Zero BB	On	3x10μW	< 20 μs

Figure 19.6.2: Optimization of the trade-off for the logic between active energy per cycle (E_{cycle}) and deep-sleep leakage power (P_{leak}) under 72-MHz timing constraints (SPICE simulation results at TT corner, 25°C, with timing/power model calibrated from post-layout place/route results).

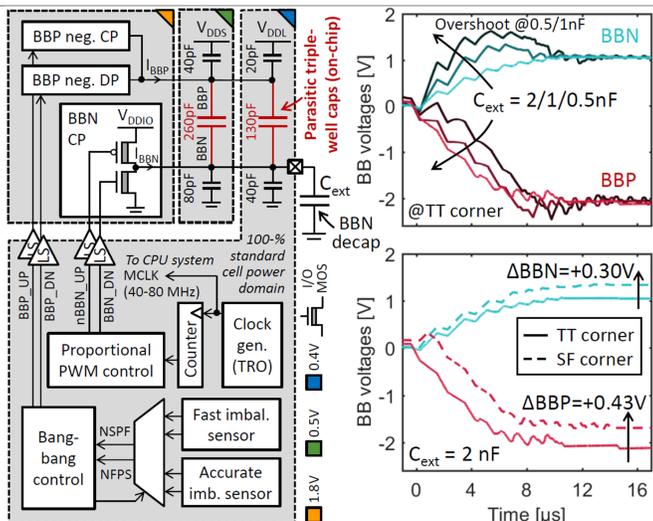


Figure 19.6.3: Dual-loop (FLL for BBN, DLL for BBP) UFBR architecture, startup showing the faster lock with a 2-nF external BBN decap and startup response to a skewed process corner (mixed-signal HDL/SPICE simulations at 25°C).

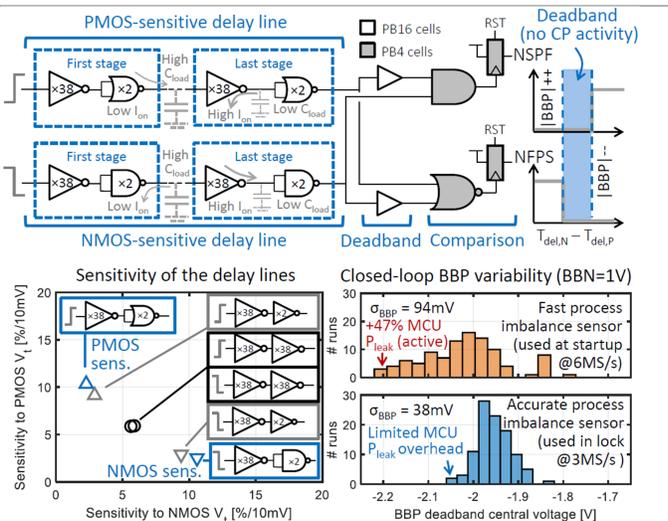


Figure 19.6.4: Schematic of the process imbalance sensor, sensitivity to NMOS/PMOS independent V_t variations and die-to-die variability of the closed-loop BBP voltage (100-run MC SPICE simulation results at TT corner, 25°C).

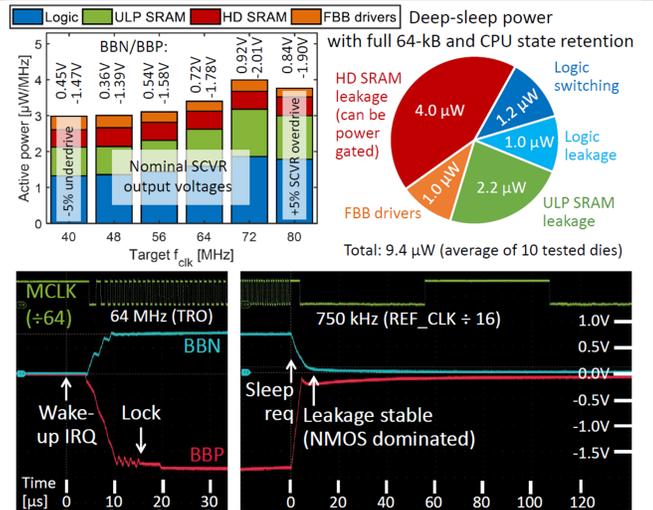


Figure 19.6.5: Power measurement results and power mode transition (adaptive FBB startup/stop). Logic power includes the adaptive FBB control, f_{max} -sensing TRO clock generator and process imbalance sensor.

	Bol, ISSCC, 2012	Myers, JSSC, 2015	Paul, JSSC, 2017	Lallement, APOLLO2, 2018	Ambiq, Apollo2, 2018	Uytterhoeven, ESSCIRC, 2018	This work
CMOS technology	65nm LP/GP	65nm LP	14nm FinFET	28nm FDSOI	40nm ULP eFlash	28nm FDSOI	28nm FDSOI
CPU	oMSP430	CM0+	x86 IA	CM0+	CM4F	Zscale	CM0DS
Memory	18kB SRAM	24kB SRAM	72kB SRAM, 16kB ROM	4kB SRAM, 4kB ROM	256kB SRAM, 1MB eFlash	64kB SRAM	64 kB SRAM
Closed-loop PVT compensation	✓	×	×	×	-	×	✓
Embedded power management (ePM)	✓	✓	×	×	✓	×	✓
CPU state retention in deep sleep mode	×	✓	✓ (S1), × (S0)	×	✓ (SS1), × (DS2)	×	✓
Max. frequency [MHz]	71	66	297	150	48	200	80
Deep-sleep retention power* [nW/kB]	95#	20	-	88#	914 (SS1) + 40 (DS2)##	-	147
Wake-up time	30 μs †	N/A	> 1 ms (S1), > 1 s (S0) ‡	-	1.9 μs (SS1), 16.5 μs (DS2)##	-	< 20 μs
Active power at MEP [μW/MHz]	6.1 @25 MHz	11.7 @0.7 MHz	27* @3.5 MHz	2.7 @16 MHz	44.3† @48 MHz	8.8 @22 MHz	3.0 @48 MHz
Peak efficiency [DMIPS/mW]	~66* @10 DMIPS	80 @0.65 DMIPS	~74* @7 DMIPS	× (not enough memory)	43† @92 DMIPS	114 @22 DMIPS	280 @40 DMIPS

* Computed with the CRO clock generator power. † Includes the ePM power as the numbers without ePM are not available. ‡ Without CPU state retention. ## These are estimations assuming 0.4 DMIPS/MHz for oMSP430 and 2 DMIPS/MHz for x86.

Figure 19.6.6: MCU characteristics and measured performance compared to state-of-the-art ULP MCUs. For fair comparison, the power is reported without ePM (thus here excluding the SCVR inefficiency).

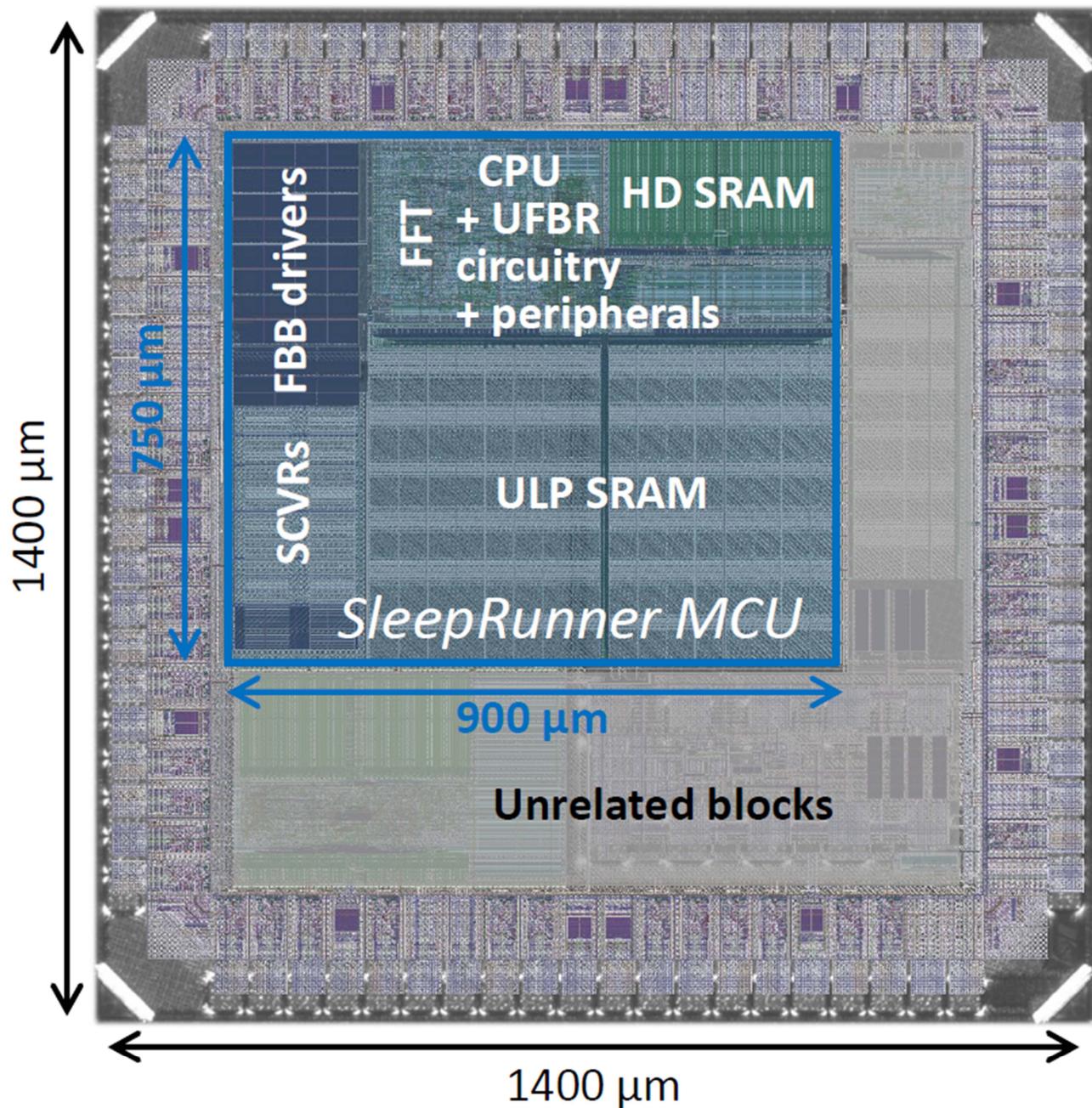


Figure 19.6.7: Microphotograph of the 2-mm² SleepRunner SoC die in 28nm FDSOI with superimposed layout view. Active MCU area is below 0.7 mm².