An Ultra-Low-Voltage Sensor Node Processor with Diverse Hardware Acceleration and Cognitive Sampling for Intelligent Sensing

Xin Liu, *Member, IEEE*, Jun Zhou, *Senior Member, IEEE*, Chao Wang, *Member, IEEE*, Kah-Hyong Chang, Jianwen Luo, Jingjing Lan, Lei Liao, Yat-Hei Lam, Yongkui Yang, Bo Wang, Xin Zhang, *Member, IEEE*, Wang Ling Goh, *Senior Member, IEEE*, Tony Tae-Hyoung Kim, *Senior Member, IEEE, and* Minkyu Je, *Senior Member, IEEE*

***Abstract*—An energy-efficient sensor node processor is presented for intelligent sensing in internet of things (IoT) applications. To achieve ultra-low energy consumption and satisfying performance, the proposed processor incorporates an ARM Cortex-M0 RISC core and diverse hardware accelerators, including discrete wavelet packet trans-form engine, FIR filtering engine, FFT engine, and CORDIC engine, to accelerate signal processing tasks. At the architecture level, dual-bus architecture with automatic bus sensing and reconfigurable memory access scheme are proposed. At the circuit level, digitally-assisted cognitive sampling and ultra-low-voltage operation with in-situ timing-error monitoring techniques are employed. When applied to neural spike classification and vehicle speed detection, the proposed SNP consumes only 39 and 29 pJ/cycle, respectively.**

***Index Terms*—Sensor node processor, cognitive, low power**

I. Introduction

With the emerging intelligent sensing in internet of things (IoT) applications, widespread use of intelligent and miniaturized sensors is expected in our near-future daily lives. They benefit applications such as personal healthcare, public transportation, and environment monitoring. As a key component of the intelligent sensor system, the on-chip processor realizes system control and digital signal processing to add intelligence into the sensor system. Correspondingly, there is an emerging category of processor family, i.e., sensor node processor (SNP) [1] – [4], where the energy efficiency is the design focus because the power budget becomes extremely critical for battery-supplied or battery-less sensors. This type of processor requires dedicated architecture and circuit design for energy-efficient control and processing as well as higher level of integration for system miniaturization.

Manuscript received in February 2015.

X. Liu, J. Zhou, C. Wang, K.-H. Chang, J. Luo, J. Lan, L, Liao, Y.-H. Lam, and X. Zhang are with the Institute of Microelectronics, A*STAR (Agency for Science, Technology and Research), 117685 Singapore (e-mail: zhouj@ime.a-star.edu.sg; liux@ime.a-star.edu.sg).

Y. Yang and B. Wang are with the Institute of Microelectronics, A*STAR, 117685 Singapore, and also with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore.

W. L. Goh and T. T.-H. Kim are with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore.

M. Je is with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 711-873, Korea.

This work was supported by A*STAR (Agency for Science, Technology and Research) SERC (Science and Engineering Research Council), Singapore under the grant no. 102 171 0163.

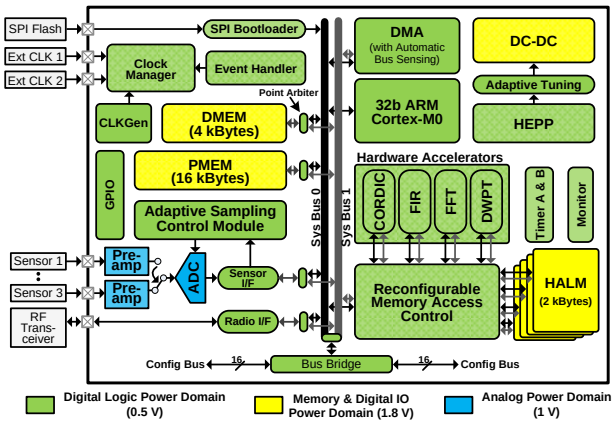


Figure 1. System architecture of the proposed sensor node processor.

Most of existing SNPs achieve low energy consumption by applying general low-power techniques such as power gating and ultra-low-voltage (ULV) operation [1] – [4]. However, the energy efficiency and performance is limited as the general low power techniques do not fully utilize the features of intelligent sensing. To further improve the energy efficiency and performance, in this paper, an SNP with parallel bus architecture, diverse hardware accelerators, cognitive sampling, and variation-resilient ULV operation is proposed for power-constrained intelligent sensing applications, such as biomedical signal processing, vehicle-status monitoring, and environmental sensing. This paper is extended from the short version of our work in [5]. In the following sections, we first present the proposed SNP design in Section II. The chip implementation and experiment results are provided in Section III. Finally, conclusions are drawn in Section IV.

II. Proposed SNP Design

A. System Architecture

The system architecture of the proposed SNP is illustrated in Fig. 1. It incorporates a 32-bit ARM Cortex-M0 RISC core for programming flexibility. It can switch between active and sleep modes through dynamic clock gating. There are 4

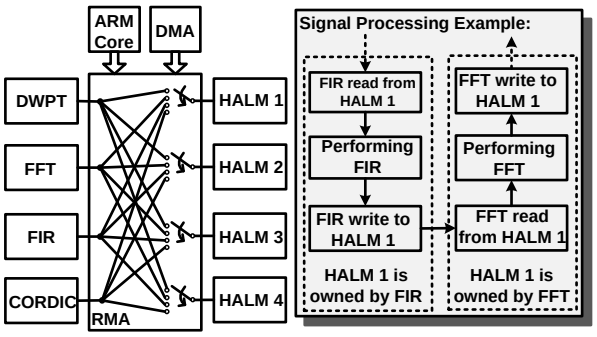


Figure 2. Proposed reconfigurable local memory access for hardware accelerators.

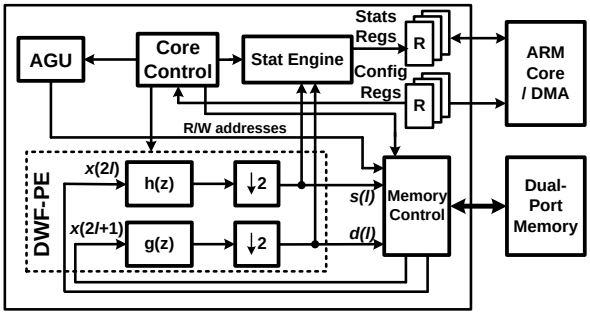


Figure 3. Reconfigurable DWPT engine.

kBytes data memory (DMEM) and 16 kBytes program memory (PMEM). Dual-bus architecture is developed to facilitate parallel processing. Diverse hardware accelerators are developed to speed up the commonly used signal processing tasks in intelligent sensing applications. The hardware accelerators include discrete wavelet packet transform (DWPT) engine, finite-impulse-response (FIR) filtering engine, fast Fourier transform (FFT) engine, and coordinate rotation digital computer (CORDIC) engine. A radio interface module is developed to communicate with the external RF transceiver. To achieve high level of integration for system miniaturization, this SNP integrates multi-channel analog front-ends, ADC, sensor interface, and DC-DC. The 9-bits SAR ADC is used to achieve satisfying performance and low power consumption. The DC-DC is a Buck-converter based design. In-situ timing error monitoring technique with adaptive voltage and clock scaling is applied to reduce the design margin and enhance variation resilience for ULV operation. Key design efforts are described in details below.

B. Dual-Bus Architecture with Automatic Bus Sensing

To realize comprehensive signal processing, the RISC core and diverse hardware accelerators are required to operate together. For improving the throughput of the system, we propose the dual-bus architecture (i.e. System Bus 0 and 1) to facilitate parallel processing of the RISC core and hardware accelerators as shown in Fig. 1. In our SNP design, event-driven processing is adopted. Unlike the existing design in [4], the event-driven processing is applied not only to the RISC core, but also to the diverse hardware accelerators. Most of the processing tasks can be performed quickly in the hardware accelerators while the RISC core stays in sleep mode to save power. From time to time the RISC core is woken up to process some complex tasks when hardware accelerators are still operating. In this case both hardware accelerators and RISC require the control of the system bus for intermediate data transfer. The bus control of hardware accelerators is through a direct memory access (DMA) controller. Before data transfer, DMA automatically senses the available buses from system bus 0 and 1. When the RISC core is in sleep mode, DMA can choose either system bus for data transfer and event-driven processing. When System Bus 0 is occupied by the RISC core, DMA can switch to System Bus 1 automatically so that system throughput can be improved. To prolong the sleep time of the RISC core for improving energy efficiency, the task transition among different hardware accelerators can be coordinated through the linked descriptor (LD) features of DMA. The information for a series of DMA operations such as the source-/destination- addresses, the total data size and mode of data transfer etc., are stored in a series of LDs which are preloaded into the SRAM. When the RISC core goes to sleep mode, based on a resultant interrupt received from the different event handler, the DMA controller can control the signal processing flow according to the preloaded LDs. By doing so, the wake up frequency of the RISC core can be minimized, resulting in significant reduction of system energy consumption.

C. Reconfigurable Local Memory Access

In our proposed SNP, each hardware accelerator is assigned with one memory (2 kBytes), i.e., hardware accelerator local memory (HALM). Therefore, the input/output data for each hardware accelerator can be stored in its HALM, instead of using global DMEM. By doing so, the data switching frequency on system buses can be reduced, resulting in energy consumption reduction of memory access. In existing designs (e.g. [1]), the ownerships of the local memories for different hardware accelerators are fixed. However, during comprehensive signal processing, frequent data exchange is expected between different hardware accelerators. It is not energy efficient if one HALM is solely owned by one hardware accelerator while other hardware accelerators cannot access it, as the processed data by one hardware accelerator has to be transferred to another HALM for processing, causing extra energy consumption. To address this limitation and further reduce the energy consumption, a reconfigurable memory access (RMA) scheme is developed, as shown in Fig. 2. RMA can reconfigure the ownerships of HALMs among different hardware accelerators. Therefore, one hardware accelerator can save the processed data into its own HALM. Subsequently, the ownership of this HALM can be re-allocated to another hardware accelerator so that it can access this HALM and fetch the data for further signal processing. In Fig. 2, for example, HALM 1 is firstly assigned to FIR module for data writing and reading. Following that, the ownership of HALM 1 is changed to FFT module. In such case, the data after FIR filtering can be directly read out from HALM 1 to perform FFT, and thus the energy for data transfer between FIR and FFT HALMs can be avoided. The setting of RMA can also be programmed by software for operation flexibility. In addition, the HALMs can also be accessed by the RISC

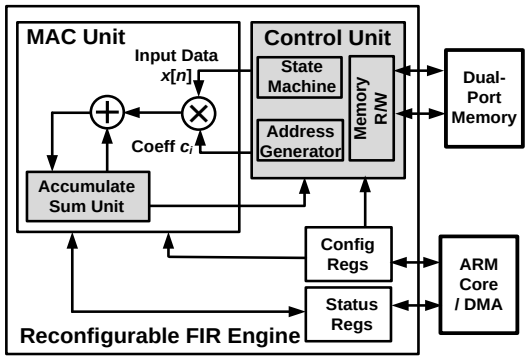


Figure 4. Reconfigurable FIR filtering engine.

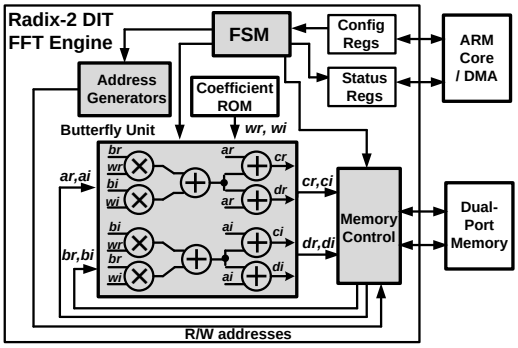


Figure 5. Memory-based FFT engine.

core through RMA control module, as shown in Fig. 1.

D. Diverse Hardware Accelerators

Intelligent sensing requires comprehensive signal processing consisting of some common processing tasks such as time/frequency analysis and de-noising. To reduce the workload of the RISC core and increase the processing speed, diverse hardware accelerators are developed and integrated in the proposed SNP. A reconfigurable 256-point DWPT engine is adopted to accelerate the time-frequency analysis required by many sensing applications, as shown in Fig. 3 [6]. It uses a discrete wavelet filter processing element (DWF-PE) to perform frame-based multi-scale decomposition with reconfigurable time-frequency resolution and decomposition tree structure. A reconfigurable FIR filtering engine with memory-based processing architecture and reusable multiply-accumulator (MAC) unit is developed to achieve variable-length filtering, as shown in Fig. 4. The filter coefficients and the number of filter taps of FIR filter are fully reconfigurable through configuration registers. In Fig. 5, a memory-based radix-2 FFT engine is developed for accelerating tasks related to frequency domain analysis. In Fig. 6, a fast-convergence CORDIC engine is developed, to speed up operations including sin(x), cos(x), exp(x), ln(x) and sqrt(x). An optimal-angle-selection logic is used to determine the optimal rotation angle for the next iteration by comparing the residual angle of each iteration with optional angle values stored in a look-up table. As a result, the number of iterations is significantly reduced, compared to the conventional CORDIC designs [7].

E. Digitally-Assisted Cognitive Sampling

For a large number of intelligent sensing applications such

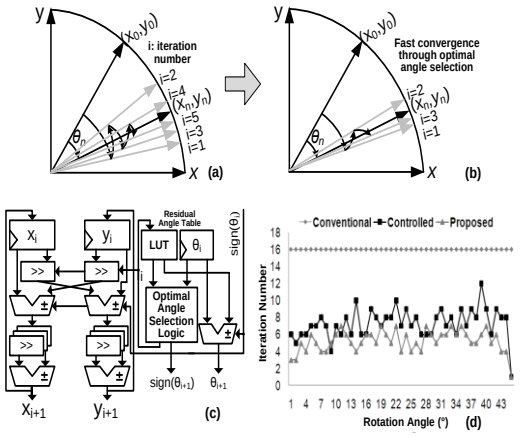
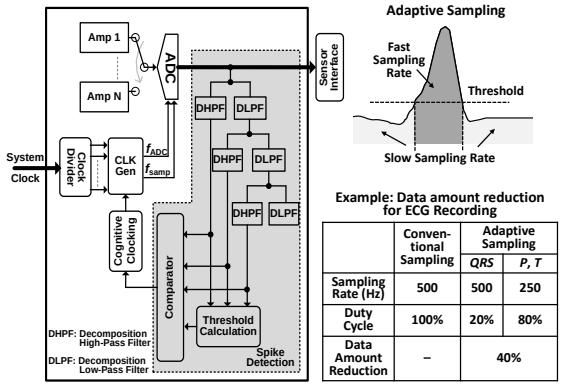


Figure 6. CORDIC engine: (a) operation of the conventional CORDIC, (b) operation of the proposed CORDIC, (c) architecture of the proposed CORDIC, (d) iteration comparison.



	Conventional Sampling	Adaptive Sampling	
		QRS	*P, T*
Sampling Rate (Hz)	500	500	250
Duty Cycle	100%	20%	80%
Data Amount Reduction	–	40%	

Figure 7. Digitally-assisted cognitive sampling technique.

as biomedical and environmental monitoring, the input signals are spike-like [8], [9]. In this type of signals, the low-frequency components occupy the major portion of the signal period, while the high-frequency components (i.e. spike) occupy a small portion of the signal period. To save power consumption in sensing, processing and transmitting this type of signals, we adopt a digitally-assisted cognitive sampling technique as shown in Fig. 7. A spike-detection module based on wavelet transform is used to remove the noise and detect spike portions. The detection result is then used to control the ADC sampling rate by cognitive clocking control [8]. The spike portions are sampled with high sampling rate, while low sampling rate is used for the non-spike portions. For example, when applied to ECG recording application (Fig. 7), a sampling rate of 500 Hz is used for the *QRS* complex and 250 Hz is used for the *P* and *T* waves. As a result, a data amount reduction of 40% is achieved. This does not only reduce the ADC power consumption, but also reduces the amount of the data to be processed and transmitted, resulting in additional energy saved in the signal processing and transmission, because the energy consumption on these parts are nearly proportional to the data amount.

F. Ultra-Low Voltage Operation Techniques

In our SNP design, multiple power domains are used in order to improve the energy efficiency, as shown in Fig. 1.

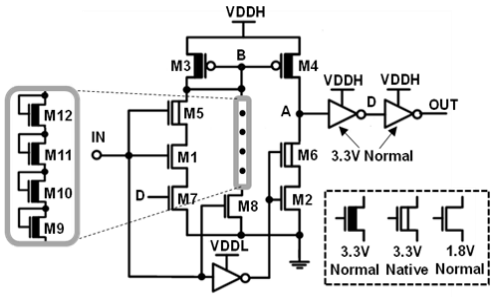


Figure 8. Level shifter from ultra-low voltage to I/O voltage.

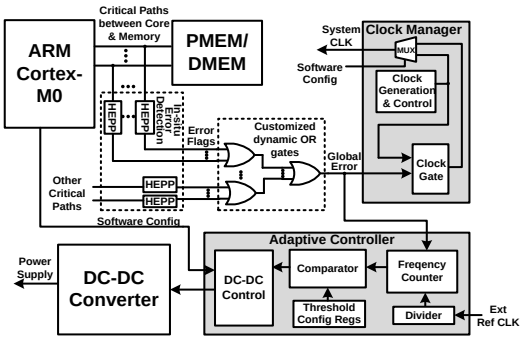


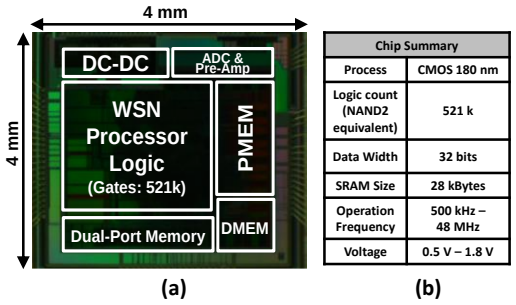
Figure 9. In-situ timing-error monitoring and adaptive voltage scaling.

The overall digital logic operates at ULV (i.e. 0.5 V) to achieve ultra-low power consumption. Commercial SRAM modules are used which operate at nominal voltage (1.8 V). Digital IOs are also supplied by nominal voltage. The analog modules are supplied by separate analog VDD (1 V). A ULV standard cell library is developed by re-characterizing a commercial library at ULV. The library is also pruned by excluding the cells with large fan-in as these cells tend to have large variability or function issue at ultra-low voltage. In addition, an ULV level shifter based on modified Wilson current mirror is adopted for voltage conversion between the different power domains [10]. As shown in Fig. 8, this level shifter can significantly improve the delay and power consumption while achieving a wide voltage conversion range. It also employs mixed-V_t device and device sizing aware of inverse narrow width effect to further improve the delay and power consumption. A half-path-based error prediction and prevention (HEPP) technique with adaptive voltage tuning is employed in our SNP to deal with the PVT variation under ULV [11], as shown in Fig. 9. It detects timing errors (late transitions) at the half-path-delay point and performs global clock gating to prevent timing failures. The error frequency is measured for adaptive voltage tuning via the on-chip DC-DC converter. Compared to [12], our technique does not introduce large MIN delay constraint while being able to deal with sudden errors. Furthermore, the error prevention is done through clock gating, making it applicable to general digital designs without design conversions.

III. Chip Implementation and Experiment Results

The proposed SNP is fabricated in 0.18 µm CMOS process to benefit from low leakage power, process variation, and costs. Fig. 10 shows the chip micrograph, chip summary and demonstration PCB prototype of the developed SNP.

Fig. 11 shows the measurement results of the fabricated



Chip Summary	
Process	CMOS 180 nm
Logic count (NAND2 equivalent)	521 k
Data Width	32 bits
SRAM Size	28 kBytes
Operation Frequency	500 kHz – 48 MHz
Voltage	0.5 V – 1.8 V

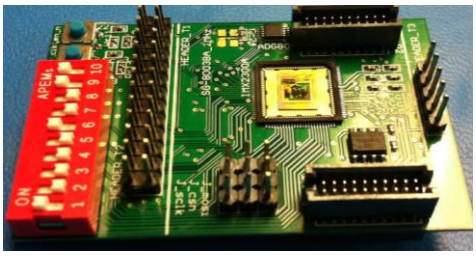


Figure 10. Developed SNP: (a). Chip micrograph; (b). Chip summary; (c). Demonstration PCB prototype.

SNP for two typical intelligent sensing applications, i.e., bio-medical and automotive applications. In Fig. 11(a), the SNP is used to detect and classify the neural spikes from the input neural signals. Denoising is performed on the input signal by the FIR filtering engine. Neural spike is detected by firmware running in the RISC core, and then the important morphology features are extracted by the DWPT engine. Finally, spike classification is performed by the RISC core and CORDIC engine. From Fig. 11(a), it is observed that spikes can be accurately identified into 3 different types according to their morphology features. The average dynamic power for this application is 19.5 µW with 500 kHz operating frequency and 0.5 V operating voltage. In Fig. 11(b), the SNP is used to detect the vehicle speed from the signal acquired by an accelerometer in an advanced tire monitoring system (ATMS). After denoising by the FIR filtering engine, the FFT engine is used for spectrum analysis. After that, the fundamental frequency extraction is done by the RISC core, and the vehicle speed is calculated by the RISC core and CORDIC engine. The average dynamic power for this application is 14.5 µW with 500 kHz operating frequency and 0.5 V operating voltage. Fig. 11 also shows the modules that are activated in different applications, and thus the measured energy consumptions are mainly contributed by these modules.

Fig. 12 illustrates the operation flow of the proposed SNP by the measured waveforms from logic analyzer. At the beginning, program is loaded from external flash memory, followed by the DSP configuration. After that, the sensor interface continuously acquires digitalized data from ADC module. The acquired data are temporarily saved in the buffer in the sensor interface module. When the amount of buffered data reaches a predefined threshold (configurable by software), the data is transferred into DMEM, and RISC core will wake up for signal processing together with hardware accelerators. After processing, extracted data is sent to radio interface for wireless data transmission, and RISC core goes

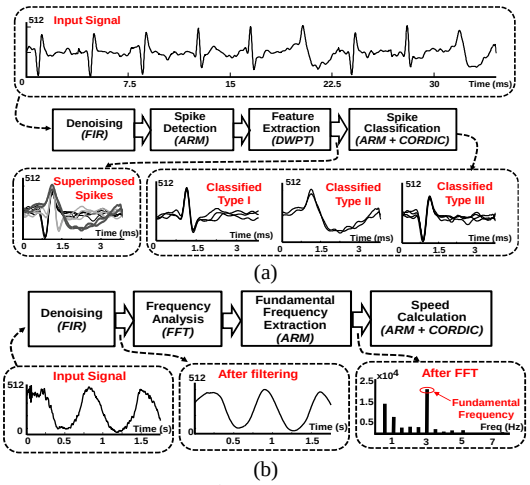


Figure 11. Measurement results from two applications: (a) neural spike classification, (b) vehicle speed detection.

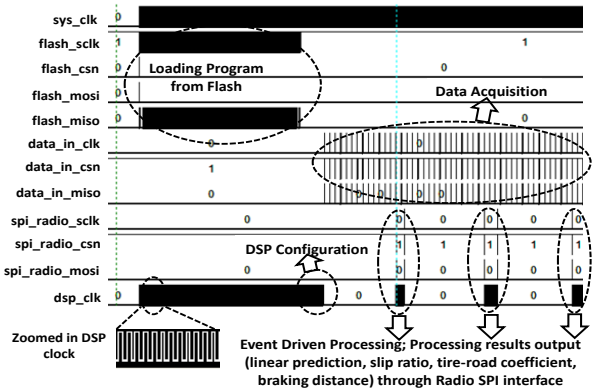


Figure 12. Measurement results of the developed SNP for tire monitoring.

back to sleep mode. It can be observed from Fig. 12 that the DSP clock is periodically woken up. The event driven signal processing is used to achieve significant power saving.

The performance comparison between the proposed SNP and other state-of-the-art designs is presented in Table I. The proposed SNP achieves comparatively low energy consumption when performing comprehensive signal processing through the proposed design techniques. For example, the LD features in DMA minimize the wake up frequency of the RISC core for internal data transfer and task transition. Diverse hardware accelerators share the work load with the RISC core and make the signal processing more energy efficient. Reconfigurable local memory access reduces the energy consumption for data transfer between hardware accelerators. Dual-bus architecture with automatic bus sensing enables parallel processing with higher energy efficiency. Digitally-assisted cognitive sampling reduces the sampling power and the amount of data to be processed and transmitted. Ultra-low voltage design is adopted to further reduce the energy consumption on the circuit level. As power gating cells are not available in the used 0.18 µm process technology, the leakage power is 12 µW. This can be reduced by using advanced process technology. Meanwhile, satisfying performance is achieved for intelligent sensing applications. Finally, system miniaturization is achieved by higher level system integration.

TABLE I. COMPARISION WITH THE STATE-OF-THE-ART SNPS

SNP	[1]	[2]	This Work
Technology (nm)	130	40	180
Voltage (V)	0.5 – 1	0.4 – 1.1	0.5 – 1.8
Clock Frequency (MHz)	0.1 – 10	1 – 150	0.5 – 48
DSP Core	16-bit RISC	VLIW	32-bit RISC
Hardware Accelerators	Median, FFT, FIR, CORDIC	No	FIR, DWPT, FFT, CORDIC with RMA and dual-bus
Cognitive Sampling	No	No	Yes
On-chip Analog Components	No	DC-DC	Pre-amp, ADC, DC-DC
In-Situ Error Monitoring and Adaptive Voltage Tuning	No	Yes, no error correction	Yes
Minimum Energy/Cycle (pJ)	35 @0.5V, 100kHz (EEG)	30 @0.4V, 1MHz (ECG)	29 @0.5 V, 500kHz (Car) 39 @0.5 V, 500kHz (Neural)

IV. CONCLUSIONS

A sensor node processor for intelligent sensing in IoT applications has been designed. To achieve high energy-efficiency and satisfying performance, several architecture- and circuit-level design techniques are developed. It consumes only 39 and 29 pJ/cycle for neural spike classification and vehicle speed detection applications, respectively.

REFERENCES

[1] J. Kwong, *et. al*., "An energy-efficient biomedical signal processing platform," *IEEE JSSC*, vol. 46, no.7, pp. 1742–1753, July 2011.

[2] M. Konijnenburg, *et. al*., "Reliable and energy-efficient 1MHz 0.4V dynamically reconfigurable SoC for ExG applications in 40nm LP CMOS," *IEEE ISSCC*, Feb. 2013, pp. 430–431.

[3] X. Liu, *et. al*., "Ultra-low-energy near-threshold biomedical signal processor for versatile wireless health monitoring," *IEEE A-SSCC*, Nov. 2012, pp. 381–384.

[4] M. Ashouei, *et. al*., "A voltage-scalable biomedical signal processor running ECG using 13pJ/cycle at 1MHz and 0.4V," *IEEE ISSCC*, Feb. 2011, pp. 332–333.

[5] J. Zhou, *et. al*., "A 0.5V 29pJ/cycle sensor node processor for intelligent sensing applications," *IEEE ISOCC*, Nov. 2014, pp. 70 – 71.

[6] C. Wang, *et. al*., "Near-threshold energy- and area-efficient reconfigurable DWPT/DWT processor for healthcare-monitoring Applications," *IEEE TCAS-II*, vol. 62, pp. 70–74, Jan. 2015.

[7] S. Wang, *et. al*., "Critically damped CORDIC algorithm," *IEEE MWSCAS*, Aug. 1994, pp. 236–239.

[8] X. Liu, *et. al*., "A 457 nW near-threshold cognitive multi-functional ECG processor for long-term cardiac monitoring" *IEEE JSSC*, vol. 49, pp. 2422 – 2434, Nov. 2014.

[9] M. Trakimas, *et. al*., "An adaptive resolution asynchronous ADC architecture for data compression in energy constrained sensing applications," *IEEE TCAS-I*, vol. 58, pp. 921–934, May 2011.

[10] J. Zhou, *et. al*., "An ultra-low voltage level shifter using revised wilson current mirror for fast and energy-efficient wide-range voltage conversion from sub-threshold to I/O voltage," *IEEE TCAS-I*, vol. 62, pp. 697–706, Mar. 2015.

[11] J. Zhou, *et. al*., "HEPP: a new in-situ timing-error prediction and prevention technique for variation-tolerant ultra-low-voltage designs," *IEEE A-SSCC*, Nov. 2013.

[12] H. Fuketa, *et. al*., "Adaptive performance compensation with in-situ timing error predictive sensors for subthreshold circuits," *IEEE TVLSI*, vol. 20, no. 2, pp. 333–343, Feb. 2012.