Flexible, Self-Adaptive Sense-and-Compress SoC for Sub-microWatt Always-On Sensory Recording

Flexible, Self-Adaptive Sense-and-Compress SoC for Sub-microWatt Always-On Sensory Recording

Jaro De Roose, *Member, IEEE*, Haoming Xin, *Member, IEEE*, Ahmed Hallawa, Gerd Ascheid, *Life Senior Member, IEEE*, Pieter J. A. Harpe, and Marian Verhelst, *Senior Member, IEEE*

***Abstract*—We present a 5-sensor, fully integrated sensing system with interchangeable sensors and programmable configuration to create a sub-microWatt multisensor node that can tackle a wide range of sensing applications. Furthermore, the sensor node is capable of autonomously adapting its configuration to the application requirements hence minimizing system power. Such self-reconfiguration is enabled at low overhead by developing an automated offline optimization strategy, in combination with an autonomous embedded configuration controller, using the concept of behavioral trees (BTs). The resulting fully integrated platform consumes a maximum of 321 nW when sampling at 500 Hz and 3025 nW at 8 kHz. Furthermore, we demonstrate the end-to-end autonomous optimization flow for two different applications exploiting different sensors: 1) human activity recognition using accelerometers and 2) machine listening using a microphone. Both use cases demonstrate that the introduced system and methodology reduces the power by more than a factor 2 without losing significant application detection accuracy.**

***Index Terms*—Automated optimization, behavioral tree (BT), flexible sensor node, low overhead flexible hardware, ultralow power.**

I. Introduction and SotA

Miniature multisensor IoT systems with small batteries require continuous gathering of multisensor information at minimal power consumption in a multitude of applications, such as human activity tracking [1], machine listening [2], voice activity detection [3], human health monitoring [4], etc. Besides the energy efficiency, also the monetary cost of sensor nodes is a critical element, mainly consisting of development and production cost. This development cost could be amortized when creating a single, versatile sensor node, capable of reconfiguring to the requirements of a wide range of applications. Some works have created multisensor platform [5], with configurable sensor interfaces [6], or even a complete reconfigurable biomedical SoC with interchangable sensors [7] to cover many biomedical applications. Yet, these designs are not optimized for energy efficiency. Other recent work, on the other hand, has introduced self-reconfiguration techniques, enabling the sensor node to tune down its sensing effort, and hence power consumption, at runtime whenever circumstances tolerate this. A few examples are the adaptive decimation of ECG signals [8], usage of compressed sensing [9], or the online tuning of health monitoring sensor efforts based on the early warning score, a measure of potential health danger a patient is in [4]. These works effectively reduce power through online adaptivity, yet the designs are custom-made to their application, putting the burden back on the development cost.

Manuscript received May 18, 2020; revised July 14, 2020; accepted August 9, 2020. Date of publication August 20, 2020; date of current version September 29, 2020. This article was approved by Associate Editor Stefan Rusu. This work was supported in part by EU H2020 under Grant 66534; and in part by EU-ERC under Grant ERC-2016-STG-71503.
(Corresponding author: Jaro De Roose.)

Jaro De Roose and Marian Verhelst are with the Department of Electrical Engineering, KU Leuven, 3000 Leuven, Belgium (e-mail: jaro.deroose@esat.kuleuven.be).

Haoming Xin and Pieter J. A. Harpe are with the Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands.

Ahmed Hallawa and Gerd Ascheid are with the Department of Electrical Engineering, RWTH Aachen, 52074 Aachen, Germany.
Digital Object Identifier 10.1109/LSSC.2020.3018382

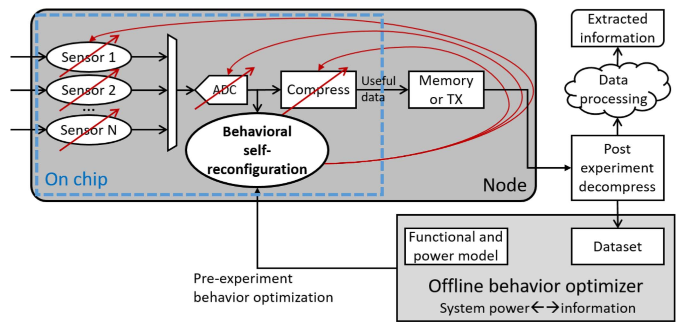

Fig. 1. Information-goal-oriented automated optimization scheme.

To overcome this conflict, this letter aims to achieve energy efficiency over a wide range of applications, by automatically tailoring sensor node configurations to the application. This is both pursued at design time, to limit design effort and cost, as well as at run-time, enabling online self-reconfiguration to adapt to online changes. This is achieved by letting the node learn its optimal "behavior," capturing how it should adapts its setting to external stimuli.

II. Self-Configuration Through Offline Optimized Behavioral Trees

Fig. 1 shows the system concept of the proposed automated online reconfigurable sensor nodes. The first important building block is a configurable sensor node, whose power scales down when required performance decreases (Section III). Second, we require a control center that executes the learned behavior to reconfigure the sensor node in function of online sensor stimuli. This is implemented in the form of a behavioral tree (BT), which conditionally changes sensor configurations based on online calculated features from the incoming sensory data (Section IV). This allows the node to achieve its most energy-efficient operation for a given performance target. The optimal BT is determined automatically at low cost, using a dataset of the targeted application as training material. Using a model of the sensor node and the training dataset, an optimizer simulates the impact of a deployed BT on the system power consumption. Furthermore, it calculates a figure-of-merit that indicates the quality of the application-targeted information (e.g., % of steps recognized for a step counter application). This allows the optimizer to optimize the sensor node behavior toward the desired tradeoff between power consumption and information quality (Section V).

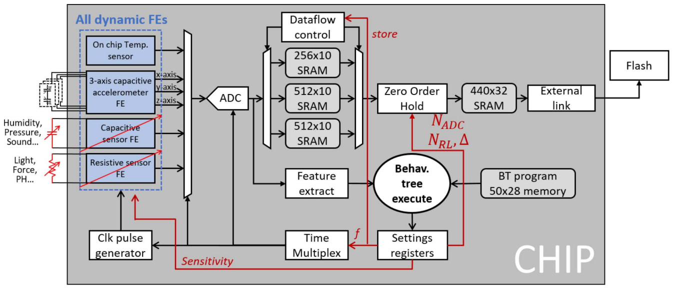

Fig. 2. Hardware of the flexible sensor system overview.

III. Flexible and Scalable Hardware

The developed sensor node targets configurable multisensory recording and lossy data compression, before storing the compressed data in a memory (e.g., for later wireless transmission or readout). Specifically, the implemented multisensor chip (Fig. 2) can connect to up to 5 passive resistive or capacitive sensors, whose analog signals are time-multiplexed into one ADC and stored in a 5 × 256-sample buffer bank. Once the buffer is full, the samples are compressed using zero-order hold (ZOH): each sample from a sensor is compared to the previously stored sample of the same sensor, if the difference is larger than a configurable parameter Δ, the new sample is stored, combined with the run length (=numbers of samples passed since the last stored sample) [10]. The compressed data is stored in an output buffer which is used for the data transfer to an external ULP flash memory [11].

The sensor platform supports power scalability through six dynamic tuning knobs that can be changed for each of the five sensors individually each time the buffer is full: sensor sensitivity [12], sample frequency, bits stored per sample, compression harshness (Δ), bits used for RL encoding (more bits allow longer possible run length and thus higher compression ratio), and sensor data deletion.

As leakage is dominant in the targeted sensing applications with <10 kSps, all parallel hardware that leaks is eliminated. As such, all five sensors share the same ADC, feature extraction, and compression hardware. Furthermore, the sample buffer allows to power gate the compression logic and memory while sampling. Since the sample buffer is a source of leakage, it is banked to be able to powergate segments corresponding to unused sensors. All flexibility in the design comes with minimal power and area overhead.

IV. Embedded Behavioral Tree Implementation

The node's behavior, its self-configuration, is controlled using a BT [10] (Fig. 3, top left), which is built using three types of nodes.

1) Action nodes (AN) are the nodes that reconfigure the sensor node, i.e., determine all six settings of a sensor.
2) Condition nodes (CN) are nodes that evaluate online sensor stimuli. They have access to five hardware-friendly features per sensory stream (Fig. 3): a) mean; b) mean-average-deviation; c) zerocrossing; d) slope; and e) peak-to-peak information. The CN compares a feature to a learned threshold and outputs success/failure.
3) Composite nodes (CPN) connect the CNs to ANs to create sequences of ANs and CNs. The CPN executes its children nodes from left to right until either the first success or the first failure of a CN (see Fig. 3, top left).

The complete BT can be translated to if-else code which has to be programmed into the SoC for each application. To execute this code, without consuming too much overhead power, a small specialized

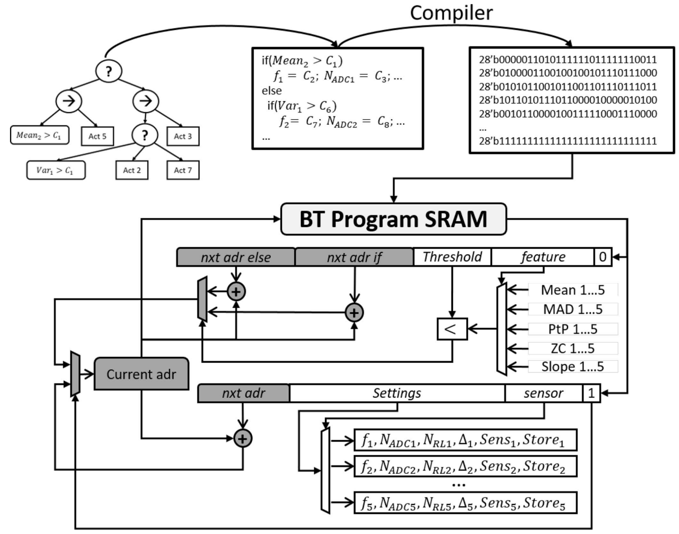

Fig. 3. Low power BT implementation enabling online adaptivity.

processor is implemented on chip, shown in Fig. 3. A custom compiler compiles the BT with one binary word per AN or CN, containing all relevant information to be used by the processor. AN commands, starting with a "1," contain the targeted sensor, the requested settings for this sensor and to what BT node the processor should jump next. CN lines, starting with a "0," contain the feature that needs to be compared, the comparison threshold, and to which BT node the processor jumps if the CN is successful, resp. failing. The BT program is executed at powerup and each time the buffer is full, hence changing the SoC configuration whenever a new set of features becomes available.

V. Offline Behavioral Tree Training

Using 1) training data, consisting of sensor data of the targeted application, 2) a model of the sensor node, and 3) figure-of-merit that indicates the quality of the application-targeted information, a BT is optimized to achieve the best tradeoff between the information metric and power consumption. To converge efficiently to an optimal BT from virtually infinite possible BTs, training is performed in three steps.

1) The algorithm creates a pool of suitable ANs (Fig. 4 left) using an evolutionary algorithm (EA) that identifies sensor configurations that are Pareto-optimal (power versus figure-of-merit) when applied statically throughout the whole simulation. Ten of the configurations across this Pareto front are selected and added to the action pool, to keep the action pool compact.
2) The algorithm creates a pool of potential CNs by statistically analyzing the sensory features calculated on the training set and choosing a select number of thresholds, which cover the expected range of their respective feature (Fig. 4 upper right).
3) A BT synthesizer uses CPNs to tie ANs and CNs from the pools together to form BTs and a second EA automatically evolves the BTs to an application-optimal BT (Fig. 4 lower right).

This application-optimal BT is then translated to if-else code, compiled and loaded into the sensor node. Hence the sensor node has optimal online reconfiguration without any manual design effort.

VI. Measurement Results

The complete sensing system depicted in Fig. 2 has been integrated in a 3-mm^2 65-nm LP CMOS chip (Fig. 5). We first show the system's measured power consumption in various configurations

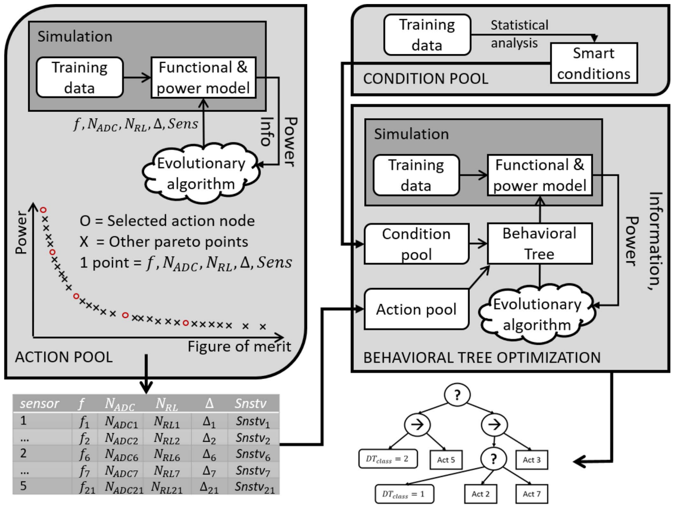

Fig. 4. Automated behavior training using previous data of the targeted application in three steps: action pool creation, condition pool creation, and BT training.

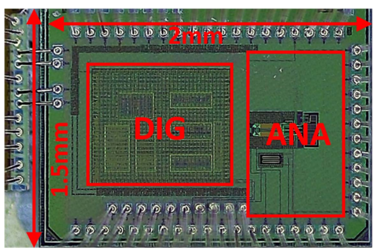

Fig. 5. Chip photograph

$$Power = SPS * EPS + OBL * OBO + \#CD * EPB + IBL(\#SS) + APS * EPA + Leakage$$

Parameter	Value
Energy Per Sample	Acc= 20pJ, Light=33pJ, Mic =30pJ
Output Buffer Leakage	28nW
Input Buffer Leakage	0, 10, 13, 24, 27, 37 nW
Energy Per Byte	32pJ (on chip) +54pJ (mem)
Energy Per Activation	135nJ

SPS=Samples Per Second
OBO=Output Buffer On
#CD= amount of Compressed Data
#SS = Sensors stored
EA = Activations Per Second

Fig. 6. Power model.

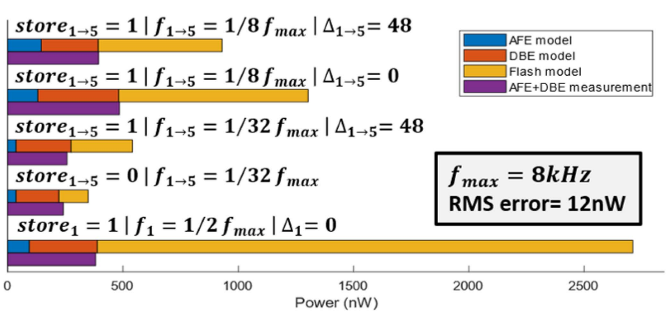

Fig. 7. Correlation between power measurements and power model.

and its correlation with the derived power model, followed by the automated deployment of the system in two specific applications.

1) Power Consumption and Model: The power consumption of the chip (without flash) is measured for various configurations at 8-kHz sampling frequency at 23 °C, and shown in Fig. 6. The system has a minimum power of 110 nW@500 Hz and a maximum power of >2500 nW@8 kHz, showing >20× power scalability. Furthermore, the base power only measures 110 nW@500 Hz, consisting mainly of leakage.

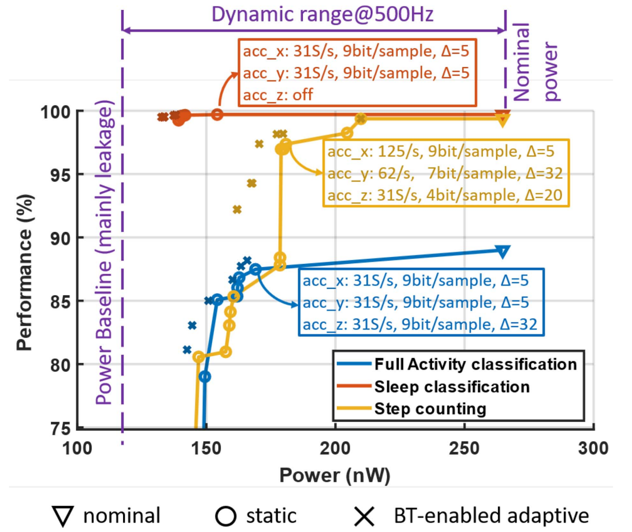

Fig. 8. Multiple HAR applications and their performance-power tradeoff in simulation.

To support the automated optimal behavior learning, a power model has been constructed for the complete sensor node, including an external signal-level-simulated flash memory [11]. The model and its parameters are shown in Fig. 6. The power model is correlated with the chip measurements, shown in Fig. 7, demonstrating a correspondence with a root-mean-square error of 12 nW over five configurations.

2) Automated Application-Adaptivity: The low-cost adaptivity to specific applications is showcased through deployment in two different types of applications: 1) human activity recognition (HAR) and 2) machine fault detection through machine listening.

Using a public dataset [1], we consider three specific HAR applications: 1) full activity recognition; 2) sleep recognition; and 3) step counting, to showcase how our automated flow can customize the sensor node to application specifics. The application-specific information figure-of-merit for activity and sleep recognition is obtained by the smoothened output of a neural-network-based classifier operating on the dataset's inherent features. Step detection, on the other hand, is triggered when the combined magnitude of the three accelerometer axis crosses the average signal value. The baseline reference corresponds to operating at nominal settings: a static Nyquist sampling sensor node with lossless compressing only using the 3-axis accelerometer data. Fig. 8 shows the results of the automated configuration flow. The triangles show nominal settings, circles show the Pareto front found by the EA optimizing static settings or "flexible" solutions (solutions found for the action pool), and crosses show solutions achieved through BT-enabled online reconfiguration, or "adaptive" solutions. Using the automated application-adaptivity flow, up to 80% of the dynamic power can be saved, depending on the difficulty of the task and the targeted performance.

Next, we consider a machine listening application with microphone sensors to show the sensor node's and the framework's versatility. We select a pump anomaly classification case from the MIMII dataset resampled at 8 kHz [2] (machine id 6, 0-dB SNR, post-processing using mel features and a neural network). The base accuracy is expressed as area under the curve (AUC). Fig. 9 shows the power performance tradeoff we obtain using our framework, showing more than a factor 3 power reduction by flexibility alone.

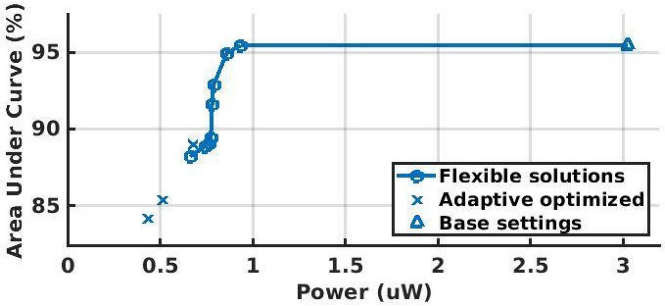

Fig. 9. Machine listening application.

	JSSC 2012 [9]	JSSC 2014 [7]	JSSC 2015 [5]	JSSC 2019 [3]	This work	
System	Digital compressed sensing block	All type sensor FE +digital filtering + TX	3 Channel ECG + motion artifact reduction	Low power voice activity detection	Programmable R and C type sensor FE + compression + ULP Flash memory	
Application adaptable	No	Yes	No	No	**Yes**	
# Sensors	1	4	6	1	5	
Technology	90nm	0.35µm	0.18µm	0.18µm	65nm	
Area	2mm²	11.25mm²	49mm²	17.55mm²	3mm²	
ADC resolution	8 bit	9.4 ENOB	13.5 ENOB	8 bit	10 bit	
Sampling Speed	20kS/s	200S/s/sensor	250/s/sensor	1kS/s	500S/s (MUX)	8kS/s (MUX)
VDD	0.6V	1.8V	1.2V	1.4V, 0.6V	0.8V	
Power AFE	-	178.55µW	233µW	69nW	**13.8nW**	**220nW**
Power DBE	1.9µW	2.6µW	112µW	73nW	**170nW**	**1145nW**
Power Data Transfer	-	762µW (TX)	-	-	**138nW**	**2200nW**

Fig. 10. State-of-the-art comparison. This work can be used for a multitude of applications at very low power levels.

3) SotA Comparison: Fig. 10 compares this letter to the state-of-the-art. This hardware is unique as a reconfigurable multisensor platform, which works down to microWatt level, while dealing with a multitude of applications.

VII. Conclusion

The work presented in this letter is a complete multisensor system with changeable sensor setup and flexible behavior reconfiguration, to maximize automated, low-cost reusability of the same platform over multiple applications. The system consumes 321 nW for HAR applications without optimized settings at 500 Hz and 3025 nW for machine listening at 8 kHz. From this baseline, specific application experiments demonstrated the ease to adapt the settings and the dynamic behavior of the complete sensing system in an automated way to any new application. Moreover, the EA- and BT-enabled adaptivity pushes dynamic power by a factor 2 for all demonstrated applications without significant accuracy loss.

Acknowledgement

Special thanks to Imec's IC-link for back-end support.

References

[1] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Proc. Int. Workshop Ambient Assist. Living (IWAAL)*, Dec. 2012, pp. 216–223.

[2] H. Purohit *et al.*, "MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," 2019. [Online]. Available: arXiv:1909.09347.

[3] S. Oh *et al.*, "An acoustic signal processing chip with 142-nW voice activity detection using mixer-based sequential frequency scanning and neural network classification," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 3005–3016, Nov. 2019.

[4] A. Anzanpour *et al.*, "Self-awareness in remote health monitoring systems using wearable electronics," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, Lausanne, Switzerland, Mar. 2017, pp. 1056–1061.

[5] N. Van Helleputte *et al.*, "A 345 μW multi-sensor biomedical soc with bio-impedance, 3-channel ECG, motion artifact reduction, and integrated DSP," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 230–244, Jan. 2015.

[6] J. Xi, C. Yang, A. Mason, and P. Zhong, "Adaptive multi-sensor interface system-on-chip," in *Proc. IEEE SENSORS*, Daegu, South Korea, 2006, pp. 50–53.

[7] Y. Huang *et al.*, "A self-powered CMOS reconfigurable multi-sensor SoC for biomedical applications," *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 851–866, Apr. 2014.

[8] C.-I. Ieong, M. Li, M.-K. Law, P.-I. Mak, M. I. Vai, and R. P. Martins, "A 0.45 V 147–375 nW ECG compression processor with wavelet shrinkage and adaptive temporal decimation architectures," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 4, pp. 1307–1319, Apr. 2017.

[9] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012.

[10] A. Hallawa, J. De Roose, M. Andraud, M. Verhelst, and G. Ascheid, "Instinct-driven dynamic hardware reconfiguration: Evolutionary algorithm optimized compression for autonomous sensory agents," in *Proc. Genet. Evol. Comput. Conf.*, Jul. 2017, pp. 1727–1734.

[11] Q. Dong *et al.*, "11.2 A 1Mb embedded NOR flash memory with 39μW program power for mm-scale high-temperature sensor nodes," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, Feb. 2017, pp. 198–199.

[12] H. Xin, M. Andraud, P. Baltus, E. Cantatore, and P. Harpe, "A 0.34-571nw all-dynamic versatile sensor interface for temperature, capacitance, and resistance sensing," in *Proc. IEEE 45th Eur. Solid-State Circuits Conf. (ESSCIRC)*, Cracow, Poland, 2019, pp. 161–164.