Energy Efficiency is Not Enough: Towards a Batteryless Internet of Sounds

Vincent Lostanlen, Antoine Bernabeu, Jean-Luc Béchennec, Mikaël Briday, Sébastien Faucou, Mathieu Lagrange

HAL Id: hal-03324622
https://hal.science/hal-03324622

Submitted on 23 Aug 2021

Energy Efficiency is Not Enough: Towards a Batteryless Internet of Sounds

Vincent Lostanlen, Antoine Bernabeau, Jean-Luc Béchennec,
Mikaël Briday, Sébastien Faucou, Mathieu Lagrange
vincent.lostanlen@ls2n.fr
Université de Nantes, Centrale Nantes, CNRS, LS2N, F-44000 Nantes, France

ABSTRACT

This position paper advocates for digital sobriety in the design and usage of wireless acoustic sensors. As of today, these devices all rely on batteries, which are either recharged by a human operator or via solar panels. Yet, batteries contain chemical pollutants and have a shorter lifespan than electronic components: as such, they hinder the autonomy and sustainability of the Internet of Sounds at large. Against this problem, our radical answer is to avoid the use of batteries altogether; and instead, to harvest ambient energy in real time and store it in a supercapacitor allowing a few minutes of operation. We show the inherent limitations of battery-dependent technologies for acoustic sensing. Then, we describe how a low-cost Micro-Controller Unit (MCU) could serve for audio acquisition and feature extraction on the edge. In particular, we stress the advantage of storing intermediate computations in ferroelectric random-access memory (FeRAM), which is nonvolatile, fast, endurant and consumes little. As a proof of concept, we present a simple-minded detector of sine tones in background noise, which relies on a fixed-point implementation of the fast Fourier transform (FFT). We outline future directions towards bioacoustic event detection and urban acoustic monitoring without batteries nor wires.

CCS CONCEPTS

• **Hardware** → **Sound-based input / output**; *Digital signal processing*; **Power estimation and optimization**; **Impact on the environment**; • **Computer systems organization** → *Real-time systems*; • **Networks** → *Sensor networks*.

KEYWORDS

acoustic sensor networks, batteryless computing, digital sobriety, intermittent computing, nonvolatile random access memory

1 INTRODUCTION

The term "Internet of Sounds" (IoS) refers to the integration of audio engineering technologies into the application layer of the Internet [42]. This topic is quickly gaining momentum in the digital industry, chiefly because of two parallel trends: cheaper recording and communication on one hand [34], cheaper computing and storage on the other hand [29]. In particular, the first trend lowers the deployment cost of acoustic sensor networks, *i.e.* arrays of autonomous recording units which stream acoustic information in real time [8]. Meanwhile, the second trend boosts the scalability of audio content analysis on large-scale databases [20]. Together, these improvements push the IoS beyond its historical scope of transmitting speech and music to encompass new domains: conservation science [28], Industry 4.0 [35], meteorology [46], structural health monitoring [27], and urban planning [9], to name a few.

The mass production of inexpensive "audio things" heralds a promising future for scientific research on the IoS. However, it also poses a fundamental threat to ecological sustainability: like any other system of information technology (IT), the IoS contains non-biodegradable materials and consumes electrical energy. Specifically, loudspeaker and microphone parts consist of rare earth elements such as neodymium and dysprosium while audio plugs are plated with precious metals such as platinum, palladium, and gold [11]. Moreover, always-on devices such as "smart speakers" and urban acoustic sensors continuously perform intensive computations, *e.g.* Fast Fourier Transforms (FFTs) [17] and neural network prediction [25]. These examples illustrate the urgent need for a more responsible design and usage of IoS technologies.

Mitigating the ecological damage caused by the IoS is a two-fold problem: managing electronic waste (e-waste) [6] and improving energy efficiency [1]. Of these two avenues of research, the latter is beginning to receive some attention: for instance, "small-footprint keyword spotting" is emerging as a central audio processing task for the smart speaker industry [13]. On a related note, a recent publication has estimated the electrical consumption of training state-of-the-art systems for end-to-end speech recognition [33], thus initiating a long-needed ethical debate on the footprint of machine listening. Even so, the former question of reducing the *material* footprint of IoS terminals remains largely undiscussed [37]. Indeed, although multiple publications in the humanities have alarmed about the environmental cost of the IoS (particularly digital music [3, 15]), we electrical engineers and computer scientists have yet to collectively take action for "Green(er) IT" in the specific domains of sound and music computing and semantic audio.

Evaluating the environmental impact of any given commercial product requires a complete life-cycle assessment (LCA) which should encompass the extraction of raw materials *(cradle)* as well as

recycling or disposal *(grave)*. This is by no means simple for a single piece of audio hardware, let alone for the IoS at large. Indeed, LCA demands expertise in a wide range of scientific domains, including mineralogy, econometrics, sociology, and law [31]. Furthermore, one methodological difficulty of LCA resides in the partial attribution of global environmental impacts to a specific human activity [16]. Smartphones, for example, are certainly IoS terminals, but they also serve other purposes: text messaging, GPS navigation, video streaming, and so forth. Likewise, IoS bitstreams rely on a physical infrastructure that is most often shared with other network nodes; this infrastructure includes copper cables, optical fibers, modems, and so forth [40]. Hence, the task of coming up with a number, be it in tons or in joules, that summarizes the environmental impact of the IoS appears as nothing but Sisyphean.

With this caveat in mind, it remains possible to take action at the local level of product engineering. This is because the pressure of demand at the cradle and the outflow of e-waste at the grave are both consequences of the volume of goods shipped at the "factory gate". In the case of the IoS, this volume is already high and ever-increasing dangerously. The implications for our field are clear: we should embrace digital sobriety by manufacturing fewer objects while making them more durable.

In this context, our article proposes a new research orientation for the Internet of Sounds, specifically regarding the design of wireless acoustic sensors. As of today, these sensors require batteries which are recharged asynchronously, either by a human operator or via a solar panel. Yet, the limited lifespan of batteries hinder the sustainability of acoustic sensing off the grid. Against such a limitation, our radical answer is to avoid the use of batteries altogether; instead, we harvest ambient (solar) energy in real time. The immediate benefit of this approach is that, unlike batteries, solar panels can operate autonomously for decades. Thus, the prospect of a building IoS devices without wires nor batteries could, in the future, reduce the material flow at the factory gate, thereby also reducing our dependency on raw materials as well as the amount of e-waste.

At the same time, batteryless computing involves challenges of its own. First, low-cost solutions for energy harvesting, such as miniature solar panels, provide a few milliwatts of electrical power. Thus, wireless communication must be kept to a minimum: instead of transmitting raw audio, the sensor should perform low-bitrate feature extraction "on the edge". Secondly, for lack of any batteries, the energy supply to the sensor is intermittent. To function properly, the sensor should anticipate power losses by monitoring its own consumption and scheduling tasks accordingly. Our article discusses these two challenges and paves the way towards solving them in a real-world deployment setting.

Section 2 reviews prior work on energy-efficient acoustic sensor networks and points out the drawbacks of battery dependency in this context. Section 3 presents some of the key technologies which make batteryless IoT devices possible. Section 4 reports the results of our proof of concept: an FFT-based detector of sine waves. Lastly, Section 5 outlines two potential applications of our research program: bioacoustics and urban acoustics.

2 LIMITATIONS OF PRIOR WORK

Recent advances in digital signal processing and solid-state circuits have considerably alleviated the electrical consumption of audio content analysis on embedded systems. These advances have undeniable merits in terms of energy efficiency and constitute a necessary part of the effort towards "Green IT". That being said, we note that the prospect of making durable wireless acoustic sensors involves more than energy efficiency: ideally, the sensor should not only perform energy harvesting and edge computing but also implement communication protocols, manage power losses gracefully, and accommodate software updates. Yet, the state of the art on low-power machine listening tends to neglect these considerations, which are crucial for usability testing and life-cycle analysis.

2.1 Battery-powered acoustic loggers: Song Meter and AudioMoth

With over 20k recorders sold to date, the "Song Meter" product line by Wildlife Acoustics is arguably the industry standard in remote sensing for ecology and conservation[1]. The SM4 ($849) is the model with the longest autonomy. The power consumption of the SM4 is of the order of 100 mW during acquisition and 1 mW at idle [21]. Its weatherproof case encloses four D cells, hence about 800–100 Wh assuming that the batteries are alkaline. According to the manufacturer, this energy supply translates to an autonomy of 510 hours, either continuously or divided in short acquisition segments over several months. In comparison, the Song Meter Mini is priced at $499 and runs for 210 hours with four AA batteries. Lastly, the Song Meter Micro is priced at $249 and runs for 150 hours with three AA batteries (see Figure 3 in Appendix).

At a lower price tag ($60), the AudioMoth[2] combines an ARM Cortex-M4F MCU and a 256-kilobyte SRAM chip, thus allowing to process audio in real time at sample rates up to 384 kHz [21] (see Figure 4 in Appendix). The AudioMoth is powered by three AA batteries. Interestingly, the AudioMoth spends most of its energy on writing audio data onto the external microSD card, rather than on audio acquisition itself. The real-time power consumption of the AudioMoth lies within the range 17–70 mW, depending on the required sample rate and the type of microSD card. Like the SM4, the AudioMoth microcontroller can stay at idle between acquisition segments: this reduces its power consumption to 80 μW. The manufacturers of the AudioMoth state that, in most practical use cases, the limiting factor to the autonomy of the AudioMoth is not the depletion of batteries but the storage capacity of the microSD card (32 gigabytes).

From the observations above, it stands that the main purpose of battery-powered devices such as Song Meter and AudioMoth is to conduct short-term campaigns of audio acquisition in remote areas. These devices are particularly convenient for experimental research in bioacoustics or eco-acoustics because they can be deployed virtually anywhere and operate according to a predefined schedule. For example, they can record nocturnal flight calls from migratory birds between sunset and sunrise.

However, a major drawback of Song Meter and AudioMoth is that they lack wireless connectivity. As a result, retrieving the audio

---

[1] Official website: https://www.wildlifeacoustics.com
[2] Official website: https://www.openacousticdevices.info

data requires the intervention of a human operator every few weeks or so. Therefore, Song Meter and AudioMoth are unsuitable for real-time acoustic monitoring. Furthermore, as advocated in the introduction, the presence of non-rechargeable batteries in these sensors hinders their ecological sustainability. To overcome these drawbacks, one solution is to recharge batteries *in situ* by means of a renewable energy source such as a solar panel.

2.2 Repurposed smartphones: RFCx Guardian

A successful example of a real-time acoustic sensor with built-in energy harvesting is the RFCx Guardian (see Figure 5 in Appendix). This device is made by Rainforest Connection, a 501(c) nonprofit organization[3] which aims to detect the sounds of illegal logging in the forests of Ecuador, Indonesia, and Philippines. The RFCx Guardian comprises a repurposed Huawei smartphone which is encased in a weatherproof box and equipped with an omnidirectional microphone. The smartphone does not perform any feature extraction nor sound event detection on the edge: instead, it transmits compressed audio to a central server via quad-band radio, typically 2G or 3G.

The RFCx Guardian is typically perched on a tree and surmounted by eight solar panels. However, we note that the solar panels do not power the smartphone directly: instead, they recharge a battery which in turn powers the smartphone. This battery has a lithium–iron–phosphate ($LiFePO_4$) cathode[4]; which, in comparison with lithium–cobalt oxyde ($LiCoO_2$) or lithium–nickel oxyde ($LiNiO_2$), has a lower electrical voltage (3.2 V) and energy density (325 Wh/L). On the flip side, LiFePO4 is safer, more durable, and incurs a lesser ecological damage: indeed, iron (Fe) is a more abundant metal than nickel (Ni) and cobalt (Co) while being less toxic [19]. Given that the design of wireless acoustic sensors is primarily driven by requirements of autonomy rather than weight, LiFePO4 appears as a judicious choice of rechargeable battery.

The RFCx Guardian demonstrates that it is possible to build off-the-grid nodes for the Internet of Sounds which may remain active for more than a year under remote control. Furthermore, the choice of repurposing 2010-generation smartphone instead of building an acoustic sensor from a brand new MCU and modem is certainly laudable. Even so, this choice comes at the detriment of ecological sustainability for other parts of the sensor: batteries and solar panels in particular. Indeed, smartphones have a relatively high electrical consumption, especially when transmitting data over 2G: of the order of 1.5 W according to the manufacturer. To keep the smartphone powered on a 24/7 basis, the RFCx Guardian must be able to recharge quickly and retain 35–40 watt–hours of energy, assuming 3–4 hours of sunlight per day; hence the need for a large $LiFePO_4$ battery and eight solar panels instead of one.

Another prototype, named SAFE [39], also proposes to combine a large solar panel, a deep-cycle battery, and 3G connectivity. The main difference between RFCx Guardian and SAFE is that the former operates on smartphone hardware whereas the latter operates on a Raspberry Pi computer. We also note that SAFE uses a 64GB microSD card as a high-capacity buffer (200 hours of audio) in case of connectivity losses. With these differences in mind, the

RFCx Guardian and SAFE belong to the same category in terms of autonomy and ecological footprint.

2.3 Compact recurrent neural networks

A recent publication by Cerutti et al. [12] offers an inspiring example of real-time audio classification on the edge. The authors have trained a deep learning model on top of a pre-trained audio embedding (VGGish [20]) to classify urban sounds. Then, they have integrated the resulting classifier onto a low-power and low-complexity MCU: an ARM Cortex M4, *i.e.* the same type of hardware as AudioMoth. To this end, the authors have applied a sequence of computational techniques: student–teacher training with multi-stage knowledge distillation, 8-bit quantization, and firmware implementation with the CMSIS-NN library. On the UrbanSound8k dataset [38], their edge computing node achieves competitive classification accuracy (68%) while having a power consumption of 5.5 mW. In comparison, the original VGGish-based model reaches 75% classification accuracy but requires 800 times as many operations and is therefore unfit for low-power devices.

A shortcoming of the approach proposed by [12] is that the resilience to power losses remains undiscussed. Specifically, the authors perform knowledge distillation between a gated recurrent unit (GRU) and a recurrent neural network (RNN). By definition, RNNs have an internal state which is estimated on the fly at prediction time and updated recursively from one spectrogram frame to the next. Yet, the authors propose to store the value of this internal state as a buffer in the random-access memory (RAM) of the Cortex-M4 MCU. Unfortunately, this type of memory is volatile: any loss of power corrupts the values in the buffer irreversibly and non-deterministically. In other words, the data corruptions which arise during power losses will affect the response of the recurrent neural network when the power supply resumes. This issue is all the more serious when designing off-the-grid sensors: indeed, the energy that is provided by solar panels is intermittent and typically fluctuates depending on time of day, weather, and season.

2.4 Analog signal processing and spiking neural networks

While the knowledge distillation of deep learning models in MCUs aims at a power consumption of a few milliwatts (see subsection above), the methodology of spiking neural networks (SNN) operates in an even more energy-efficient regime: that is, of the order of one microwatt. The key idea behind SNNs is to encode the flow of neural information across artificial synapses as a spike train, much like biological neurons.

In this way, the power consumption of SNNs is proportional to the average firing rate of all neurons at any point in time. Because neural networks tend to learn sparse representations, nonzero firing rates are found in few neurons at once; hence an adaptive routing of electrical supply onto the synapses which are barely necessary for machine prediction [45]. In contrast, conventional implementations of artificial neural networks encode synaptic information by means of volatile RAM buffers, whose power consumption remains high even if many values are zero.

The resort to SNNs is particularly beneficial in applications where the sensor is constantly active and the acoustic events of

---

[3]Official website: https://rfcx.org
[4]Source: https://news.ycombinator.com/item?id=16664175

interest are rare. This is the case, for example, in smart homes, where appliances respond to human speech after having detected a specific voice command, known as "wake word". Indeed, wake words last for less than a second and typically appear a few times a day at most: conversely, whenever the home environment is silent, power consumption reduces accordingly. State-of-the-art systems for always-on keyword spotting combine SNNs with techniques in analog feature extraction: for example, it is possible to approximate mel-frequency spectrograms with band-pass filters, clipping amplifiers, and half-wave rectifiers. In this context, a recent publication [44] has improved the robustness of keyword spotting to the presence of background noise by prototyping a nonlinear circuit which approximates per-channel energy normalization (PCEN) via an integrate-and-fire (IAF) scheme.

The new generation of mixed-signal (analog–digital) and neuromorphic architectures for pattern recognition pushes the energy efficiency of acoustic sensors to an extreme level. For example, a single AA battery (3.9 W h) would contain enough energy to power the SNN-based printed circuit board of [44] for centuries. At first glance, this calculation may seem to render the debate on energy-harvesting sensors altogether moot. Yet, besides the power consumption of edge computing, wireless connectivity in the Internet of Things remains costly in terms of energy.

We also note that manufacturing solid-state circuits demands a high investment, which can only be financially amortized by relying on economies of scale. While keyword spotting chips are found in billions of devices worldwide, other use cases for the Internet of Sounds operate at a more "sober" regime of a few thousand devices. For example, urban noise pollution monitoring and bioacoustic conservation require customizable and reprogrammable solutions: yet, it is unclear to what extent the aforementioned ultra-low-power devices can be maintained and upgraded by end users.

As a consequence, we believe that in parallel with the very large-scale integration (VLSI) of specialized microwatt devices, the development of reprogrammable milliwatt devices will continue to play an important role in the Internet of Sounds, and that they will require dedicated methods for energy management.

3 INTERMITTENT COMPUTING

In an effort to achieve digital sobriety, we choose to avoid the use of batteries. We propose to directly use energy extracted from the environment, simply integrating a supercapacitor as a buffer between the harvesting device and the MCU to avoid abrupt power loss. By nature, the energy supplied by the environment is fluctuating. In addition, a full charge of the supercapacitor can power the MCU for a much shorter period of time than a battery would. In such a system, power losses must therefore be included in the model of computation as normal events and not as exceptional failures. In the literature, this model of computation is identified as intermittent computing [7, 30].

3.1 Nonvolatile random-access memory

The central goal of intermittent computing is to recover from power outages gracefully; that is, to preserve the execution context throughout periods of standby so that it can be resumed later. As explained in Section 2.3, this is not possible with MCUs such as

Cortex-M4, whose random-access memory is entirely volatile. To circumvent this problem, it is necessary to equip the MCUs with a form of non-volatile random-access memory (NVRAM). The role of the NVRAM is to perform so-called "checkpoints" of the execution context; *i.e.* backup copies of the contents of volatile memory as well as the processor's registers.

The design of NVRAM hardware seeks a tradeoff between speed and energy efficiency. On one hand, writing to NVRAM must be fast enough to allow frequent execution of checkpoints without affecting the latency of the program flow. On the other hand, the electrical consumption of checkpointing should remain negligible in comparison with edge computing and network connectivity. In this way, a dedicated runtime may correctly restore the context and continue the execution once the power is back.

3.2 Flash memory vs. ferroelectric RAM

Nowadays, the most common form of nonvolatile memory in portable devices is Flash: indeed, Flash memory is cheap, scalable, and resilient to mechanical shocks. Unfortunately, Flash is too slow to serve for intermittent checkpointing: a write access takes between 200 and 500 µs. Furthermore, Flash suffers from a limited lifespan: around 10^5 cycles for the most durable hardware type (*i.e.* single-level cell) [10].

To overcome the shortcomings of Flash memory, new forms of NVRAM are currently under development [32], such as: spin–transfer torque magnetic RAM (STT-MRAM) [2], phase–change RAM (PCRAM) [36], and ferroelectric RAM (FRAM or FeRAM) [23]. Unlike STT-MRAM and PCRAM, FeRAM has reached a sufficient level of maturity to be readily available in MCUs, where it replaces Flash memory; or as separate circuits with a serial or parallel interface. Another advantage of FeRAM is that it is resistant to gamma radiation as well as magnetic field exposure, unlike storage under the form of electrical charges.

FeRAM cell capacitors rely on a ferroelectric material, most often a lead–zirconium–titanate (PZT) ceramic compound. Indeed, by applying an external electric field on PZT, one may reverse its spontaneous electric polarization, and thus write digital information in a nonvolatile way. This operation also allows to read the current state of the PZT compound: in other words, every read access to FeRAM is destructive and requires an overwrite, which is automatically handled by the memory controller.

In comparison with Flash memory, FeRAM has a much faster access time: 125 ns for a read–write access. As a result, checkpointing 64 kilobytes to FeRAM takes about 4 ms. We note however that FeRAM remains slower than volatile RAM, either static (0.2–2 ns per write access) or dynamic (10 ns). Besides, FeRAM has a much higher endurance than Flash memory: it allows $10^{15} \sim 10^{16}$ read–write cycles, hence a lifespan that is expressed in decades.

3.3 FeRAM-enabled MCU

The choice of execution platform for batteryless Internet of Sound nodes is constrained by the availability of NVRAM in the MCU. This constraint discards many of the most popular hardware options for machine learning on the edge: Raspberry Pi, Arduino,

ARM Cortex-M, and BeagleBone series. However, one manufacturer (Texas Instruments) has developed FeRAM-enabled MCUs in its MSP430 series, wherein it replaces Flash memory[5].

In particular, the MSP430FR5994 is a 16-bit RISC with a clock frequency of 16 MHz which is sold at approximately $4 per unit (see Figure 6 in Appendix). The maximum amount of available FeRAM is equal to 256 kB. According to the manufacturer [22], its electrical consumption is around 118 µA per MHz under a 3 V power supply during operation and 500 nA during standby.

One major challenge of working with this MCU resides in its small amount of static (volatile) RAM: eight kilobytes. This means that even the most compact version of recurrent neural network by [12] would still not fit the constraints of MSP430FR5994. That being said, we note that there is a growing research interest on the training and compilation of kilobyte-sized machine learning models [18] with successful applications to keyword spotting [26]. These recent publications suggest that the SRAM limitation of the MSP430FRx series does not preclude the development of batteryless machine learning systems in the near future.

Furthermore, it is possible to use part of the FeRAM (256 kB) as working memory if the amount of SRAM (8 kB) happens to be insufficient. This comes at the cost of a higher energy consumption and lower performance due to the slower read–write access.

Note that allocating all of the working data in FeRAM would protect it against power outages. Yet, a checkpointing mechanism would still be necessary for the volatile parts of the system state: *i.e.* the processor's registers as well as those of the peripherals. Lastly, ensuring memory consistency when restoring a checkpoint and properly handling replay of operations involving peripherals remains a challenging problem, requiring the use of advanced software technologies at runtime [41].

3.4 Low-energy accelerator

Besides the availability of nonvolatile RAM for energy-efficient checkpointing, a major appeal behind working with MSP430FR5994 resides in the availability of numerical routines for signal processing, including: matrix multiplication, fast Fourier transforms (FFT), and filtering with finite and infinite impulse responses (FIR and IIR). These operations are supported by a dedicated hardware engine known as low-energy accelerator (LEA). The LEA is a subsystem of the MCU which can run simultaneously with the CPU while sharing 4 kB of static volatile RAM. It performs fixed-point arithmetic with either 16 or 32 bits of precision.

According to the manufacturer [43], a complex FFT on 256 samples is about 35 times faster on the LEA than on the CPU: 715 µs vs. 24.5 ms. This implies that the LEA can extract the short-term Fourier transform of an audio signal in real time, even at sample rates above 10 kHz. Furthermore, the same benchmark reports that the LEA consumes almost 30x less energy than the CPU for the same operation: 2 µJ vs. 69.5 µJ. For these two reasons, we believe that LEA-enabled MSP430FRx MCUs have the potential to withstand a research agenda on the topic of audio content analysis under intermittent energy supply.

---
[5]Official website: https://ti.com/

3.5 Wireless connectivity

Once the relevant information is extracted from the measurements, it must be transmitted to the Internet. Given the power constraints, it seems natural to turn to low-power wide-area-network (LPWAN) solutions such as NB-IoT, DAHS7, or LoRAWAN [5].

However, it should be noted that even with these technologies, transmitting a bit requires between a few tens to a few hundreds of microjoules, depending on the conditions, which is several orders of magnitude higher than executing an instruction on an ultra-low power MCU (a few nanojoules). Thus, given the limited capacities of the MCUs used, it is always relevant to maximize the number of computation tasks assigned to an intermittent system. On the one hand, this allows to make the best use of the energy extracted from the environment which is wasted when the supercapacitor is full and no activity is to be executed. On the other hand, after a power loss, it is always possible to resume a computation, whereas a transmission must be completely restarted.

4 PROOF OF CONCEPT

We aim to initiate a research program on the topic of intermittent computing for batteryless devices in the Internet of Sounds. In doing so, we bring together perspectives from audio signal processing and real-time systems. From this new idea, the development of a full-fledged acoustic sensor network without wires nor batteries is likely to take several years. Nevertheless, we can already present a proof of concept which indicates the feasibility of audio signal processing in FeRAM-equipped devices such as the MSP430FR5994.

Our proof of concept consists in detecting a sine wave of unknown frequency when mixed with uniform white noise. While we acknowledge that this task does not reflect the difficulty of real-world acoustic event detection, we stress that its role is not to serve as an end application but as a simple test bed for the fast Fourier transform (FFT) on embedded hardware.

4.1 Fast Fourier Transform

We consider an input signal of the form:

$$\boldsymbol{x}(t) = a_s \sin(2\pi\omega_s t) + a_n \boldsymbol{n}(t), \tag{1}$$

where the frequency ω_s is the variable of interest and $\boldsymbol{n}$ is a realization of white noise. The factors a_s and a_n control the amplitudes of signal and noise respectively. Within a discrete-time setting, we define $\boldsymbol{x}$ as a vector of length $T = 256$ and $\boldsymbol{n}$ as a sequence of independent uniform random variables in the range $[-1; 1]$.

An approximate value of ω_s may be simply estimated from its discrete Fourier transform $\mathbf{X}$ by seeking the frequency bin of maximum magnitude. While a naive implementation of the DFT has a time complexity of $\Theta(T^2)$, Fast Fourier Transform (FFT) algorithms achieve the same result with a complexity of $\Theta(T \log T)$.

4.2 Hardware implementation

We implement the method described above on an MSP430FR5994 MCU, whose development kit includes non-volatile ferroelectric random access memory (FeRAM), a 0.22F supercapacitor for energy supply, and a low-energy accelerator (LEA) for signal processing.

The digital signal processing (DSP) library which is provided by Texas Instruments for the MSP430FR5994 supports two numeric
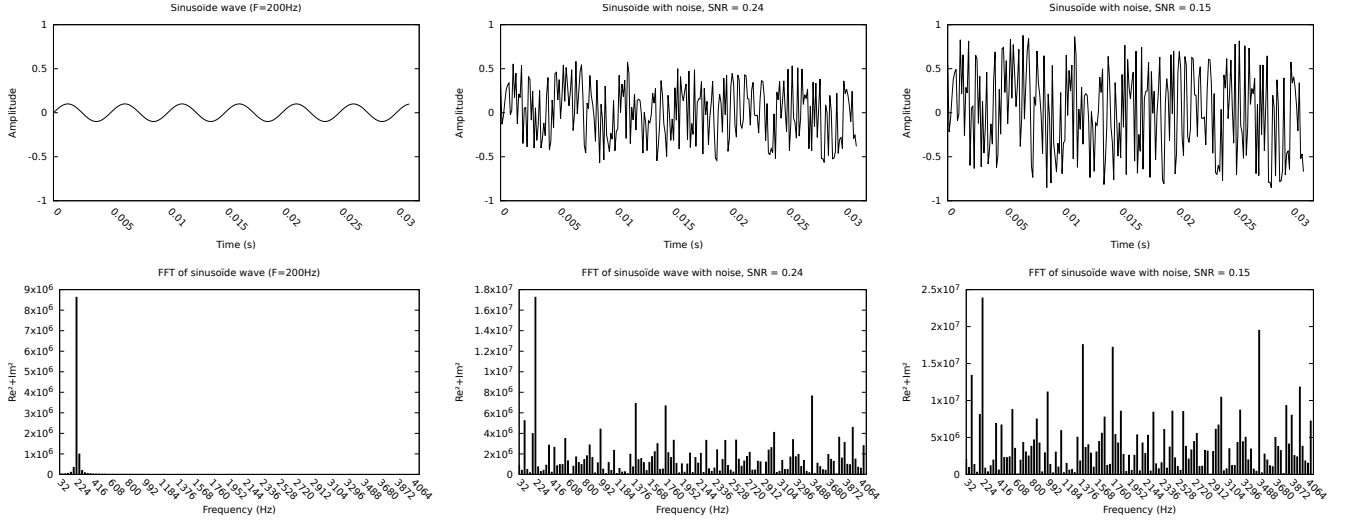
Figure 1: Top row: three sine waves of frequency $200\,\text{Hz}$**, amplitude** a_s = 0.1**, and duration** $31.25\,\text{ms}$**; with uniform random noise of amplitude** a_n **equal to 0 (left), 0.5 (center), and 0.8 (right) respectively. Bottom row: corresponding discrete Fourier transform (DFT) magnitudes, as computed by the low-energy accelerator (LEA) of a batteryless MSP430FR5994 MCU.**

types: 16-bit signed integer (Q15) and 32-bit signed integer (IQ31). Note that both of these types operate in fixed-point, rather than floating-point, arithmetic. After mapping to the range $[-1, 1)$, the rounding errors of Q15 and IQ31 are of the order of 3×10^{-5} and 5×10^{-10} respectively.

A disadvantage of fixed-point arithmetic resides in the relatively limited range of admissible values, causing a risk of overflow in the DFT. To circumvent this problem, the DSP library proposes "auto-scaling" versions of FFT, which monitor the output **X** for overflow and rescale it by a factor of two if necessary.

4.3 Experimental benchmark

At a sampling rate of f_s =8192 Hz, the window length T would correspond to a duration of 31.25 ms. As an example value, we set the fundamental frequency ω_s equal to 200 Hz, *i.e.* a period of 40.96 samples. We set the amplitude of the signal to a_s = 0.1 throughout our experiment.

Figure 1 illustrates the response of the DFT magnitude operator $|\mathbf{X}|^2$ for different values of noise amplitude a_n: 0 (left), 0.5 (center), and 0.8 (right). We represent x in Q15 format and read out the response **X** via the microUSB port of the development kit. In all three cases, we verify that the bin of greatest magnitude is $\widetilde{\omega}$ = 6; *i.e.* the closest integer to $256 \times 200/8192 = 6.25$. Therefore, we confirm that it is possible to program a digital circuit to carry out a non-trivial form of audio content analysis while operating without wires nor batteries.

Furthermore, as noted in Section 3.4, the LEA is about 30 times more energy-efficient and 35 times faster than the CPU when executing FFTs. By using the LEA, we manage to compute around 185k FFT sequences with a single charge of the capacitor.
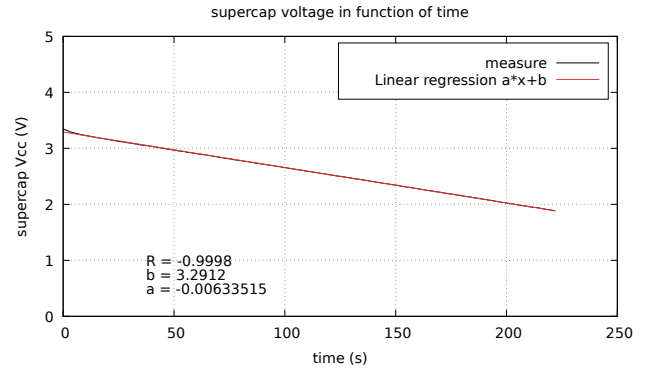
Figure 2: Drop in supply voltage as the FFT is run on the LEA

4.4 Measurement of electrical consumption

As the system should manage energy and must be able to maintain computational progress through power loss, measuring and modeling energy is needed. Today, most MCUs include an analog-digital converter (ADC) that can be used to measure the voltage of the supercapacitor that supplies the platform.

The energy E stored in the supercapacitor is equal to

$$\frac{1}{2} \times C \times V_{\text{cc}}^2 \tag{2}$$

where C and V_{cc} are capacitance and voltage respectively.

Moreover, the voltage drops linearly during MCU operation because the power consumed by the platform is also related to the voltage by the following equation:

$$P = f_{\text{clk}} \times C_L \times V_{cc}^2, \tag{3}$$

where P is the power drawn by the MCU, f_{clk} is the frequency of the MCU, C_L is the sum of the capacitors charged and discharged during the operation of the MCU and V_{cc} is the supply voltage of the MCU. Here, we neglect the static power consumption which remains negligible for this kind of MCU and which can however be evaluated by measuring the power consumption when the MCU is in low-power mode with the clock halted but where the CPU remains powered.

This can be extended to peripheral including the LEA. The linear voltage drop of the power supply is confirmed by a measurement campaign we conducted on several MCU subsystems. For example, Figure 2 shows the measurement of the supply voltage, and therefore the voltage across the supercapacitor, in the MSP430FR5994 launchpad board while the LEA is performing FFTs. We can also see that, by performing only FFTs, the platform can be supplied more than 200 seconds with a single charge of the supercapacitor. During this time, the platform is able to compute more than 185k FFT sequences. Assuming non-overlapping short-term Fourier transform (STFT) frames, this number converts to 95 minutes of audio at a sample rate of 8192 Hz.

To extend Section 3.5, we measured the electrical consumption of a low power long range transceiver module that feature LoRAWAN specification. While powering the LoRa module, we could maintain the execution up to 80 sec (the LoRa module is powered but in sleep mode) and only 8 sec when transmitting messages at full output power. This correspond to a voltage drop of 15 mV/ sec and 121 mV/ sec respectively. We can compare those metrics to the electrical consumption of a heavy computing application where we could maintain the execution up to 240 sec, corresponding to a 6 mV/ sec voltage drop. This highlight the fact that transmitting data cost more energy than executing.

The interval of measurement from 3.3 V to 1.8 V is related to the specification from the manufacturer of the MCU we used. A supply range from 3.6 V to 1.8 V is recommended but we chose to start from 3.3 V as it is the nominal voltage reference for the MCU, below the 1.8 V brown-out voltage the MCU shuts down.

5 FUTURE PERSPECTIVES

While the section above has demonstrated the feasibility of audio content analysis on a batteryless MCU, the question of turning them into fully autonomous nodes for the Internet of Sounds remains open. We now outline future research towards this goal.

5.1 Power loss management

In case of power loss, an intermittent computing system must retain its execution context: in this way, it can resume its computations at the point where they were previously suspended once sufficient energy is available again (see Section 3). However, not all activities can be suspended. For example, signal sampling or radio communication should not be suspended, as such a suspension would lead to a computation or transmission error. On the contrary, any computation that is independent of the external environment, such as feature extraction on a signal that has already been acquired, may certainly be suspended.

Besides, we should also take into account the cost of checkpointing to nonvolatile memory. As a result, it may be optimal not to suspend activities whose memory overhead is high but power consumption is low in comparison with that of checkpointing.

In an intermittent computing system, non-suspendable activities can only be launched if the platform has enough energy to complete the activity. Thus, it will be necessary to estimate the remaining operating time before the power supply drops out. Such an estimation requires 3 things:

(1) A consumption model for each subsystem: CPU, LEA, ADC, timers, radio, and so forth;
(2) An usage model for each activity which captures the time during which the CPU and the peripherals are used by the activity;
(3) An usage model for checkpointing itself, in which the volatile memory that is needed for power loss recovery is clearly expressed.

As we saw in Section 4.4, we expect consumption models would be mostly linear forecasts of voltage drop. The consumption model of an activity would be obtained by summing the voltage drops associated with each subsystem the activity uses. In this way, it would then be possible to estimate the time at which the voltage reaches the minimum operating voltage of the MCU and therefore to start the activity safely if it does not support a suspension of operation.

5.2 Application to conservation biology

Bioacoustic sensors, also known as autonomous recording units (ARUs), have a vital role to play against the rapid decline of biodiversity worldwide. Indeed, these sensors provide a minimally invasive sampling of natural habitats, thus providing key information about the relative abundancy and migratory patterns of vocalizing species. However, the current generation of bioacoustic sensors is not yet fully autonomous (see Section 2).

In this context, we envision batteryless computing as a key technology for the next generation of bioacoustic sensors. Specifically, our objective is to build a prototype which performs time–frequency analysis, per-channel energy normalization, and simple-minded event detection; *e.g.* via template matching. We leave the integration of small-footprint deep learning systems, *e.g.* species classifiers [14, 24], as a long-term goal.

5.3 Application to urban acoustics

Unlike natural habitats, urban areas do not impose drastic constraints on the deployment of acoustic sensor networks. Indeed, most cities in the world have a reliable electrical grid as well as infrastructures for high-bandwidth communication. That being said, we believe that batteryless acoustic sensors will not only benefit data-driven research in remote locations but also in cities. Indeed, these sensors could be easily displaced across neighborhoods every few months depending on policy, or even reused in a different city after a few years. This reactive approach would be more sustainable than existing platforms such as CENSE [4] or SONYC [8], whose dependency on wires implies a fixed topology.

6 CONCLUSION

Batteries hinder the durability and ecological sustainability of wireless nodes in the Internet of Sounds. Against this problem, we

propose a research agenda towards solar-powered acoustic sensors without wires nor batteries. The obvious drawback of our proposition is that these sensors could not be guaranteed to operate 24/7: therefore, they should not be used in safety-critical systems. Yet, some applications in conservation bioacoustics or urban acoustics are well suited to intermittent computing on the edge. Future research is needed to evaluate the usability of batteryless acoustic sensors in conjunction with energy harvesting and wireless communication in a real-world application context.

REFERENCES

[1] Anders SG Andrae and Tomas Edler. 2015. On global electricity usage of communication technology: trends to 2030. *Challenges* 6, 1 (2015), 117–157.
[2] Dmytro Apalkov, Alexey Khvalkovskiy, Steven Watts, Vladimir Nikitin, Xueti Tang, Daniel Lottis, Kiseok Moon, Xiao Luo, Eugene Chen, Adrian Ong, Alexander Driskill-Smith, and Mohamad Krounbi. 2013. Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM). *J. Emerg. Technol. Comput. Syst.* 9, 2, Article 13 (May 2013), 35 pages. https://doi.org/10.1145/2463585.2463589
[3] David Arditi. 2018. Digital subscriptions: The unending consumption of music in the digital era. *Pop. Mus. Soc.* 41, 3 (2018), 302–318.
[4] Jérémy Ardouin, Ludovic Charpentier, Mathieu Lagrange, Félix Gontier, Nicolas Fortin, David Ecotière, Judicael Picaut, and Christophe Mietlicky. 2018. An innovative low cost sensor for urban sound monitoring. In *Proc. INTER-NOISSE*, Vol. 258. Institute of Noise Control Engineering, 2226–2237.
[5] Wael Ayoub, Abed Ellatif Samhat, Fabienne Nouvel, Mohamad Mroue, and Jean-Christophe Prévotet. 2019. Internet of Mobile Things: Overview of LoRaWAN, DASH7, and NB-IoT in LPWANs Standards and Supported Mobility. *IEEE Communications Surveys Tutorials* 21, 2 (2019). https://doi.org/10.1109/COMST.2018.2877382
[6] Cornelis P Baldé, Vanessa Forti, Vanessa Gray, Ruediger Kuehr, and Paul Stegmann. 2017. *The global e-waste monitor 2017: Quantities, flows and resources.* United Nations University, International Telecommunication Union, and ….
[7] D. Balsamo, A. S. Weddell, G. V. Merrett, B. M. Al-Hashimi, D. Brunelli, and L. Benini. 2015. Hibernus: Sustaining Computation During Intermittent Supply for Energy-Harvesting Systems. *IEEE Embed. Syst. Letters* 7, 1 (2015). https://doi.org/10.1109/LES.2014.2371494
[8] Juan P Bello, Claudio Silva, Oded Nov, R Luke Dubois, Anish Arora, Justin Salamon, Charles Mydlarz, and Harish Doraiswamy. 2019. SONYC: A system for monitoring, analyzing, and mitigating urban noise pollution. *Comm. of the ACM* 62, 2 (2019), 68–77.
[9] Patrizia Bellucci, Laura Peruzzi, and Giovanni Zambon. 2017. LIFE DYNAMAP project: The case study of Rome. *Applied Acoustics* 117 (2017), 193–206.
[10] Jalil Boukhobza, Stéphane Rubini, Renhai Chen, and Zili Shao. 2017. Emerging NVM: A Survey on Architectural Integration and Research Challenges. *ACM Trans. Des. Autom. Electron. Syst.* 23, 2, Article 14 (Nov. 2017), 32 pages. https://doi.org/10.1145/3131848
[11] Dylan T Buechler, Nadezhda N Zyaykina, Cole A Spencer, Emily Lawson, Natasha M Ploss, and Inez Hua. 2020. Comprehensive elemental analysis of consumer electronic devices: Rare earth, precious, and critical elements. *Waste Manag.* 103 (2020), 67–75.
[12] Gianmarco Cerutti, Rahul Prasad, Alessio Brutti, and Elisabetta Farella. 2020. Compact recurrent neural networks for acoustic event detection on low-energy low-complexity platforms. *IEEE J. Sel. Top. Sig. Process.* 14, 4 (2020), 654–664.
[13] Guoguo Chen, Carolina Parada, and Georg Heigold. 2014. Small-footprint keyword spotting using deep neural networks. In *Proc. IEEE ICASSP*. IEEE, 4087–4091.
[14] Jason Cramer, Vincent Lostanlen, Andrew Farnsworth, Justin Salamon, and Juan Pablo Bello. 2020. Chirping up the right tree: Incorporating biological taxonomies into deep bioacoustic classifiers. In *Proc. IEEE ICASSP*. IEEE, 901–905.
[15] Kyle Devine. 2019. *Decomposed: The political ecology of music.* MIT Press.
[16] Mine Ercan, Jens Malmodin, Pernilla Bergmark, Emma Kimfalk, and Ellinor Nilsson. 2016. Life-Cycle Assessment of a Smartphone. In *Proc. ICT for Sustainability.* Atlantis Press, 124–133.
[17] Félix Gontier, Mathieu Lagrange, Pierre Aumond, Arnaud Can, and Catherine Lavandier. 2017. An Efficient Audio Coding Scheme for Quantitative and Qualitative Large Scale Acoustic Monitoring Using the Sensor Grid Approach. *Sensors* 17, 12 (2017), 2758.
[18] Sridhar Gopinath, Nikhil Ghanathe, Vivek Seshadri, and Rahul Sharma. 2019. Compiling kB-sized machine learning models to tiny IoT devices. In *Proc. ACM SIGPLAN PLDI*. 79–95.
[19] Robert A Goyer and Thomas W Clarkson. 1996. Toxic effects of metals. *Toxicology* 23 (1996), 813–858.
[20] Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al. 2017. CNN architectures for large-scale audio classification. In *Proc. IEEE ICASSP*. IEEE, IEEE, New York, 131–135.
[21] Andrew P Hill, Peter Prince, Evelyn Piña Covarrubias, C Patrick Doncaster, Jake L Snaddon, and Alex Rogers. 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Meth. Ecol. Evol.* 9, 5 (2018), 1199–1211.
[22] Texas Instruments. 2016. *MSP430FR599x, MSP430FR596x Mixed-Signal Microcontrollers.* Technical Report. Texas Instruments.
[23] Hiroshi Ishiwara. 2012. Ferroelectric Random Access Memories. *J. Nanosc. Nanotech.* 12, 10 (October 2012), 7619–7627.
[24] Stefan Kahl, Connor M Wood, Maximilian Eibl, and Holger Klinck. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inf.* 61 (2021), 101236.
[25] Sangeeta Kumari, Dhrubojyoti Roy, Mark Cartwright, Juan Pablo Bello, and Anish Arora. 2019. Edge-L3: Compressing L3-Net for Mote Scale Urban Noise Monitoring. In *Proc. IEEE IPDPSW*. IEEE, 877–884.
[26] Aditya Kusupati, Manish Singh, Kush Bhatia, Ashish Kumar, Prateek Jain, and Manik Varma. 2018. FastGRNN: A fast, accurate, stable and tiny kilobyte-sized gated recurrent neural network.
[27] Ákos Lédeczi, Thomas Hay, Péter Volgyesi, D Robert Hay, András Nádas, and Subash Jayaraman. 2009. Wireless acoustic emission sensor network for structural monitoring. *IEEE Sensors J.* 9, 11 (2009), 1370–1377.
[28] Vincent Lostanlen, Justin Salamon, Andrew Farnsworth, Steve Kelling, and Juan Pablo Bello. 2019. Robust sound event detection in bioacoustic sensor networks. *PLOS ONE* 14, 10 (2019), e0214168.
[29] Chris Mack. 2015. The multiple lives of Moore's law. *IEEE Spectrum* 52, 4 (2015), 31–31.
[30] K. Maeng, A. Colin, and B. Lucia. 2017. Alpaca: Intermittent Execution Without Checkpoints. *Proc. ACM Program. Lang.* 1 (2017). https://doi.org/10.1145/3133920
[31] Kevin Marquet, Jacques Combaz, and Francoise Berthoud. 2019. Introduction aux impacts environnementaux du numérique. *Bull. Soc. Inf. Fr.* (2019), 85–97.
[32] Jagan Singh Meena, Simon Min Sze, Umesh Chand, and Tseung-Yuen Tseng. 2014. Overview of emerging nonvolatile memory technologies. *Nanoscale Research Letters* 9, 1 (2014), 526. https://doi.org/10.1186/1556-276X-9-526
[33] Titouan Parcollet and Mirco Ravanelli. 2021. The Energy and Carbon Footprint of Training End-to-End Speech Recognizers. (April 2021).
[34] Judicaël Picaut, Arnaud Can, Nicolas Fortin, Jeremy Ardouin, and Mathieu Lagrange. 2020. Low-cost sensors for urban noise monitoring networks—A literature review. *Sensors* 20, 8 (2020), 2256.
[35] Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. 2019. MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection. In *Proc. DCASE*. DCASE, New York.
[36] S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y.-C. Chen, R. M. Shelby, M. Salinga, D. Krebs, S.-H. Chen, H.-L. Lung, and C. H. Lam. 2008. Phase-change random access memory: A scalable technology. *IBM J. Res. Dev.* 52, 4.5 (2008), 465–479. https://doi.org/10.1147/rd.524.0465
[37] Serenella Sala and Malgorzata Goralczyk. 2013. Chemical footprint: A methodological framework for bridging life cycle assessment and planetary boundaries for chemical pollution. *Int. Env. Assess. Manag.* 9, 4 (2013), 623–632.
[38] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. 2014. A dataset and taxonomy for urban sound research. In *Proc. ACM MM*. 1041–1044.
[39] Sarab S Sethi, Robert M Ewers, Nick S Jones, Christopher David L Orme, and Lorenzo Picinali. 2018. Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. *Methods in Ecology and Evolution* 9, 12 (2018), 2383–2387.
[40] Nicole Starosielski. 2015. *The undersea network.* Duke University Press.
[41] Milijana Surbatovich, Brandon Lucia, and Limin Jia. 2020. Towards a formal foundation of intermittent computing. *Proc. ACM Program. Lang.* 4, OOPSLA (2020), 163:1–163:31. https://doi.org/10.1145/3428231
[42] Luca Turchet, György Fazekas, Mathieu Lagrange, Hossein S Ghadikolaei, and Carlo Fischione. 2020. The Internet of Audio Things: State of the art, vision, and challenges. *IEEE Internet of Things J.* 7, 10 (2020), 10233–10249.
[43] Evan Wakefield. 2016. *Benchmarking the Signal Processing Capabilities of the Low-Energy Accelerator on MSP430 MCUs.* Technical Report. Texas Instruments. https://www.ti.com/lit/an/slaa698b/slaa698b.pdf
[44] Dewei Wang, Sung Justin Kim, Minhao Yang, Aurel A Lazar, and Mingoo Seok. 2021. A Background-Noise and Process-Variation-Tolerant 109 nW Acoustic Feature Extractor Based on Spike-Domain Divisive-Energy Normalization for an Always-On Keyword Spotting Device. In *Proc. IEEE ISSCC*, Vol. 64. IEEE, 160–162.
[45] Jibin Wu, Yansong Chua, Malu Zhang, Haizhou Li, and Kay Chen Tan. 2018. A spiking neural network framework for robust sound classification. *Front. Neur.* 12 (2018), 836.
[46] Zhongxiang Zhao, Eric A D'Asaro, and Jeffrey A Nystuen. 2014. The sound of tropical cyclones. *J. of Phys. Ocean.* 44, 10 (2014), 2763–2778.

7 APPENDIX

Figure 3: Top: Song Meter SM4. Bottom-left: Song Meter Mini. Bottom-right: Song Meter Micro. Photographs reproduced from the official website of Wildlife Acoustics.
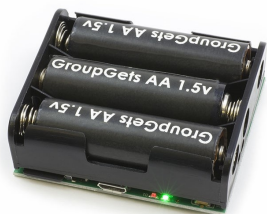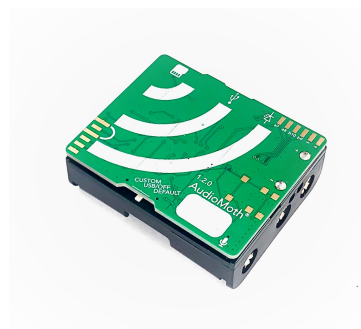

Figure 4: Left: Printed circuit board (PCB) of the AudioMoth sensor. Right: battery holder. Photographs reproduced from the official website of Open Acoustic Devices.

Figure 5: The RFCx Guardian acoustic sensor, equipped with eight solar panels and deployed in a tree. Photograph credit: Aya Naboulsi, Rainforest Connection.
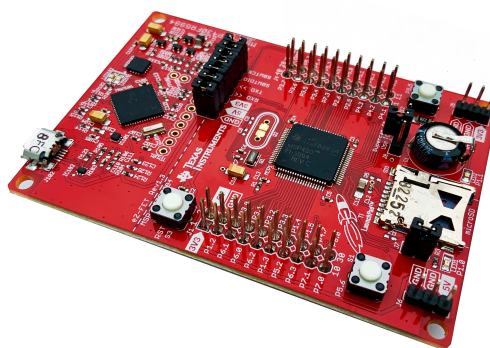

Figure 6: Development kit for the MSP430FR5994 MCU, including embedded FeRAM, a super capacitor, an analog-to-digital converter (ADC), a microSD card slot, and various communication ports.