iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65022

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Advanced Signal Processing Algorithms for IoT Devices: Real-Time Audio Classification with Resource-Constrained VLSI Architecture

Naveen Kumar V[1], Sathish S B[2], Venkatesh A[3]
[1]*Senior Scale Lecturer, Electronics & Communication Engineering, Government Polytechnic Arakere, Mandya, Karnataka, India*
[2]*Lecturer, Sound Recording & Engineering, Government Film & Television Institute, Hesaraghatta, Bengaluru, Karnataka, India*
[3] *Senior Scale Lecturer, Electronics & Communication Engineering, Government Polytechnic Nagamangala, Mandya, Karnataka, India*

Abstract: ***This study focuses on the investigation of enhancement of VLSI architectures for supporting real-time audio classification in IoT systems. Audio classification is considered an important element of IoT because there are a lot of cases when the detection of sound is absolutely crucial for the efficient functioning of the IoT system. Since IoT devices are restricted in terms of power supply, memory size, and computational power, this report preliminarily reviews dedicated VLSI architectures sufficient to meet these requirements.***
It also constrains the study to VLSI choices like ASICs, FPGAs and DSPs where each of the choices has its merits. For instance, ASICs provide power and specific application traditional solutions, FPGAs offer flexible prototyping and DSPs stand out as optimal for repetitive signal related operations. MFCC's and STFT are pointed out to be critical feature extraction roles for efficient, accurate audio representation and low complexity with the help of power optimization techniques as clock gating and dynamic voltage scaling.
Nevertheless, some existing issues have not been solved, such as making better tradeoff between the latency and accuracy and coping with the scaling requirements of the new applications. Future trends such as neuromorphic computing and TinyML have potential in improving efficiency and biologically inspired signal processing with advanced hardware software codesign are expected to lower power consumption and expand the scope of VLSI.
In summary, improvements on VLSI for IoT application such as audio classification make real time processing feasible meaning that IoT devices may infiltrate more fields. Innovative architectures, as well as emerging technologies are quickly enhancing the horizons of these capabilities, as this report shows.
Keywords: Real-time audio classification, IoT (Internet of Things) applications, VLSI (Very-Large-Scale Integration) architecture, ASIC (Application-Specific Integrated Circuit), FPGA (Field Programmable Gate Array), DSP (Digital Signal Processor), Feature extraction, MFCC (Mel-Frequency Cepstral Coefficients), STFT (Short-Time Fourier Transform).

I. INTRODUCTION

A. Context and Background

Recent years have brought newfound applications of audio classification technology in IoT devices beyond traditional use in smart surveillance to smart health and environmental monitoring systems. Audio classification is a type and subcategory of pattern recognition technology that enables systems to detect what an input audio includes and or whether the noise or sound is a spoken word, ambient sound or any specific noise and categorize it accordingly. This capability is invaluable in diverse sectors: for example, in urban centers sound categorization systems can identify and distinguish between different sounds for instance the shots fired or car accident. Respiratory sounds of patients can also be supervised in real-time for analyzing the condition of patients in the health care sector. Similarly, in environment monitoring, it can help in analyzing the pattern of wildlife, or track deforestation by change in acoustic environment.
This has however been occasioned by growing IoT devices across various applications, which has highlighted an important need for better hardware to enable real time processing of huge data in the form of audio.

©IJRASET: All Rights are Reserved

Previous techniques in audio classification strongly involve cloud computing resources because the many computational tasks such as features extraction and classification entail many computations. Nevertheless, reliance on cloud services causes growth of latency and privacy issues in such real-time critical applications. As a result, it has been perceived that carrying out audio classification routines on the IoT devices is both convenient, and more proficient than in comparison to the traditional systematic procedure, to address such issues, researchers have been concentrating on the creation of IoT devices equipped with onboard analysis tools.

A promising solution is found in Very Large-Scale Integration (VLSI) designs and even though millions of transistors are being placed on a single chip the hardware cost is still acceptable. VLSI makes it possible to design unique dedicated circuits for given operations such as audio categorization while considering the limitations of power, memory and computing that are characteristic of IoT gadgets. ASICs or FPGAs which are particular types of VLSI architecture have a special pathway for the efficient handling of audio signal processing functions. The approach lowers latency, helps to minimize data exposure to potentially unsafe Internet-connected hosts, and facilitates low-power operation—a requirement for many IoT devices powered by batteries in the field.

B. Difficulties in Real-Time Audio Classification

The extension of real-time audio classification to IoT devices has certain specialties, first of all, due to real-time requirements and strict hardware limitations of IoT systems. As with most other IoT gadgets, the various computational systems must strike a delicate balance between performance and energy efficiency, while the fixed power source dictates the need to achieve the longest possible battery life and achieve reliable and timely audio processing. The key challenges include:

1) *Computational Requirements:* Real-time audio classification process involves feature extraction and classification stage and these two stages require heavy computations. For example, spectrogram representation of the audio or extracting the Mel-Frequency Cepstral Coefficients (MFCCs) needs a lot of processing power. Thus, even current classification procedures which are becoming fairly popular to employ deep learning algorithms impose greater demands on computational capacities and, therefore, actual-time processing.
2) *Memory Constraints:* The constrained number of memory resources is another key characteristic of IoT devices. Important audio processing problems, feature extraction, and neural network classification involve using memory to store audio data, intermediate representation, and parameters. Pool of memory that can be addressed by the device type might not allow for real-time signal processing of high-quality audio data.
3) *Power Limitations:* IoT is used in devices that are generally run by batteries; real-time audio classification brings a significant load in power consumption. Real-time processing of audio data also consumes constant loads, consumes power and results in early battery discharge. High-energy algorithms and hardware implementation are therefore desirable features to maximize battery life while compromising on performance.
4) *Latency Requirements:* In many applications, including emergency response systems, the classification result of audio needs to be as fast as possible. The pros of high latency are that with high latency the issue can affect the efficiency of the IoT system especially if it is used in applications which require alerts or more frequent checks.

Considering these difficulties, development of proper signal processing algorithms which can be easily implemented in the limited VLSI platform is essential for real-time audio classification on IoT.

C. Objective and Scope

This paper highlights the objective and scope of the report.

This report presented the problem of performing real-time audio classification on IoT devices using signal processing techniques while considering the power-optimized VLSI design constraints. The goal is to find out how to extract and categorize audio features using a normal IoT device's constraints efficiently. Specifically, the report will:

Summarize current practice on analysis of signals and feature extraction which is useful in classifying audio data.

Study VLSI architecture design fundamentals and methods for enhancement of operation in real-time with applications that require low latency and power.

Suggest a strategy of how it will be implemented within the VLSI architecture providing for methods on how to improve computational complexity, memory organization and power consumption.

Show that the principles developed in this work can express IoT device models and perform simulations in an actual use case and evaluate the effectiveness of the proposed approach.

This paper's objective is to offer expertise and tangible strategies on IoT devices' resource-sustainable design for real-time audio classification in potential IoT and embedded system fields to enrich IoT knowledge.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

II. MACHINE LEARNING APPROACHES TO AUDIO CLASSIFICATION:

From a general understanding of signals and how sound can be classified, we come up with the following definition of Signal processing for Audio Classification.

Audio classification is generally a multiple-stage process because the task of identifying particular types of sound, sound events, or even spoken words in given sound samples deserves several forms of preparatory analysis. In general, the steps taken to classify an audio are acquisition of the signal, extraction of the Features and classification.

1) *Signal Acquisition:* It is a step that involves acquiring audio data by means of a variety of sensors, including microphones or as raw audio signals [1]. The audio signals are then sampled at a particular rate to transform the audio signals into digital signals for further analyses. Inability to select the right sampling rate compromises the quality and accuracy of the audio data results; thus, this step is very important.
2) *Feature Extraction:* After the audio signal is acquired, there is the process of feature extraction whereby relevant features concerning the sounds that are to be classified are extracted from the audio signal. The raw data signal is transformed into a lower-dimensional format that still preserves critical characteristics critical to classifying the signal, which can be easily understood by machine learning algorithms and deep learning neural networks.
3) *Classification:* In the feature extraction process, the features that are derived are then passed to the classifiers, which try to classify the audio data based on the patterns that it holds within features. It may occur with methods from the field of statistical and machine learning, including for example Support Vector Machines (SVM), or Hidden Markov Models (HMM) or in the field of deep learning with the use of Convolutional Neural Networks (CNN) or Long Short-Term Memory (LSTM) networks.

A. *Feature Extraction Techniques:*

In audio classification it is important to perform feature extraction because raw signal from an audio channel is not very easy to interpret by a classification model. Key techniques include:

1) *Short-Time Fourier Transform (STFT):*

a) Explanation: The STFT is a technique for describing the relationship between the frequency content of a signal and time. Employing Fourier transform to short overlapping sections of the signal, STFT introduces a two-dimensional time-frequency plot through which understanding the variation of frequency mixture of an audio signal is possible.
b) Computational Cost: Although useful for the analysis of non-stationary signals such as music, it can be computationally costly when a large number of overlapped windows is used. It is to notice that the proposed approach therewith defines the computational cost by the length of the windows and the overlap factor where there is always a trade-off between temporal and frequency resolution [2].
c) Applications: STFT finds application in speech processing and music genre classification, because it precisely analyzes the dynamic of these frequencies in captured signals such as speech or musical instruments.

2) *Mel-Frequency Cepstral Coefficients (MFCCs):*

a) *Explanation:* MFCCs are perhaps the most common features in audio classification, especially in speech recognition problems. The Mel scale, designed on the basis of the hearing range and sensation, pays more attention to the lower part of frepatial range. MFCCs are calculated by the Discrete Cosine Transform (DCT) on the logarithm of the STFT Mel-filtered power spectrum. This transformation results in attenuation of noise and quantization of data into coefficients that are assumptive to the human audial system.
b) *Relevance in Audio Classification:* MFCCs capture aspects of the timbre and pitch that are extremely useful for discriminating among different audio events. Due to the efficiency in modeling the spectral characteristics of the audio signals, they are particularly suitable for use in applications that demand low computational complexity such as speech recognition and classification of environmental sounds.

3) *Other Features:*

a) *Wavelet Transforms:* WT hold a large advantage over other transforms due to their multi-resolution ability which is useful for the time-frequency analysis of the audio signals. Wavelet transforms, on the other hand, employ variable size windows hence suitable for use in the analysis of transient signals with frequently changing frequencies. They are particularly useful in applications where both the high frequency and the low frequency signal is present as in music analysis.

©IJRASET: All Rights are Reserved

b) *Spectral Centroid:* This feature forms the 'center of gravity of the spectrum,' which is primarily seen as the sound-intensity level. Spectral band center position values above 0.4 suggest that the sounds are bright, high pitched while below 0.4 suggests that they are dark, low-pitched ones. Spectral centroid particularly is very widely applied in Audio and music genre classification.

c) *Zero-Crossing Rate (ZCR):* ZCR is designed to ascertain how often a signal switches positive/negative or conversely. It represents a single, effective measure of the level of randomness on a signal as well as the signal's frequency bandwidth. ZCR is especially important in the analysis of voiced and unvoiced speech sounds, and thus is valuable in the identification of voiceless consonants, and in the classification of music genres.

B. Classification Algorithm:

To achieve this, the extracted features are fed through classification algorithms to sort the audio data [3]. These algorithms include the more traditional machine learning and some of the innovative deep learning.

1) Traditional Methods:

a) *Support Vector Machines (SVM):* SVM is a type of supervised learning algorithm which decides on the hyperplane that provides best enhanced class spaces over classes in a given part of the feature space. In audio classification, SVMs are deployed with features such as MFCCs with the feature space often being high-dimensional, SVMs are best used where classes are linearly or nearly linearly separable. With regard to memory and computational cost, SVMs are favorable algorithms especially for devices with limited resources.

b) *Hidden Markov Models (HMM):* HMMs are statistical models that characterize temporal dependencies in sequences, and therefore it is used in speech recognition and phoneme classification. A specific HMM describes the audio signal as a set of states connected through transition probabilities thereby capturing temporal changes. They are especially suitable in the use cases where the data input is streams of real values of various lengths.

c) *Gaussian Mixture Models (GMM):* The GMMs resolve the probability distribution of features into a number of Gaussian mixtures. They are combined with HMMs to model the feature distribution in audio, especially in the speaker recognition system. They offer a statistical model by which the authors can model variability in audio features and still engage minimal computations.

2) Deep Learning Approaches:

a) *Convolutional Neural Networks (CNN):* CNNs have recently emerged as a preferred approach in audio classification because of their spatial learning in data. If applied to spectrograms or other representations in time-frequency domain, CNNs are able to discover most relevant patterns and features which are normally 'invisible' for other algorithms. For instance, deep CNNs have been employed in music genre recognition, environmental sound analysis, and real-time speech emotion recognition. It emerges from the fact that they can learn feature hierarchies in an automatic manner, but usually at large computational cost.

b) *Long Short-Term Memory Networks (LSTM):* LSTMs are a subcategory of RNNs optimized for processing sequential data with long term dependency or temporal data that includes audio data for speech or audio signal identification or feeling sensor data for identifying feeler data. That is why LSTMs avoid the drawbacks of traditional RNN and contain a memory cell, which can store essential information for an unlimited time, which is important when working with audio materials. However, LSTMs have a problem of being computationally expensive in most situations and may not be very effective when implemented in areas of limited resources.

c) *Transformer Networks:* In recent years, transformer models have been demonstrated to be effective for processing audio data especially in areas such as ASR and music application. Transducers make use of attention mechanisms to find out which parts of the sequence are rather essential thus helping the model not to have recurrent connections with the data of audio [4]. Transformers, in general, are computationally intensive to compute but are optimized, and some to meet IoT device constraints.

C. Balancing Complexity and Efficiency in Real-Time Applications:

The selection of the effectiveness of feature extraction and classification mechanisms in IoT devices is mainly dictated by the necessity to ensure high accuracy rates while consuming as few computing resources as possible. In low bandwidth environments, feature extraction is limited to certain methods like MFCCs or zero-crossing rate, or the classifier is bound to be a simple and efficient one like SVM classifiers. For more complicated audio signal processing, lightweight CNN or LSTM, these are called TinyML models, are being deployed to bring the power of Deep Learning with less memory and power requirements.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

III. REAL-TIME CONSTRAINTS IN INTERNET OF THINGS AND VERY LARGE-SCALE INTEGRATION

Real-Time Processing Requirements In this topic, it is possible to identify the general requirements for real time processing, and after that, explain what one or several of them can be relevant for the given topic.

Many IoT applications for real-time audio classification demand a setup that allows for quick analysis of the signals to support rapid compensation for events. Smart surveillance, voice activated control, and environmental monitoring are some of the general uses where low latency and high accuracy compute is necessary to support real-world applications. The primary processing requirements for real-time audio classification in IoT are:

1) *Low Latency:* In IoT applications, for instance in security and surveillance, applications must be able to recognize audio events within microseconds. There is delay between receiving an audio signal and the signal going through processing therefore the latency must be small. For instance, in surveillance applications, a delay in the identification of an alarm sound, makes the whole system culminate to nothing. Primary latency constraints for real-time systems are most likely in the milliseconds range, and these need fast, efficient processing throughout the signal processing and classification pathways.
2) *High Accuracy:* Real-time can work to deliver high classification accuracy while struggling to identify environmental noise, differences in sound intensity, or sounds that overlap with one another. This accuracy is crucial in screens such as health surveillance, where audio signals from respiratory or pulse sound needs to be categorized appropriately to warrant proper identification or response to an adverse event [5]. Small classification errors could potentially mean false alarms, or missed detections, so the focus was to achieve high accuracy classification while also keeping response time to a minimum.

A. Resource Constraints in IoT Devices:

A significant number of IoT devices, especially those which are built in distributed systems, implement these devices with limited resources. These constraints act to restrict the amount of memory, processing capability, and energy available and so present considerable problems when it comes to undertaking real-time audio classification systems.

1) *Memory Limitations:* Most of the IoT devices have a low processing power and integrated hardware memory that decides the level of computational algorithms that a smart device can handle. In fact, in audio classification, the buffers, feature extraction parameters, and model weights require storage. Frankly speaking, these are again more sophisticated models and they use greater memory so the process of choosing between the model with increased accuracy of classification has the moment of impossibility in IoT devices.
2) *Power Limitations:* Most of the IoT devices are battery operated or use energy harvesting, and therefore low power design is paramount. It is important to know about powers because they restrict the duration but also determine the type of algorithms and optimizations, at least some signal processing and machine learning models are power hungry. Real-time processing also increases the necessity for constant power, and thus low power using is required to extend the lifespan of the device.
3) *Limited Processing Capabilities:* IoT devices usually employ compact microcontrollers or processors known as microcontrollers for now, or mini-processors. There is still a problem with using such processing power for complex audio classification algorithms, especially for deep learning algorithms that require considerable computational power. Since most of the IoT devices lack hardware accelerators such as GPUs, there is a need for different methods that would allow for an efficient classification of incoming samples.

To overcome these constraints IoT devices have inherent advantages of efficient VLSI designs to accomplish needed resources against the performance specifications for real time audio classification.

B. VLSI implementation for Signal Processing:

VLSI (Very Large Scale Integrated) circuit has possibilities for the integration of the customized hardware for the needs of IoT real time [6]. Specific circuit architectures for specific audio processing tasks increase the speed of operation, decrease power consumption and better off/use memory resources in VLSI based designs. Key VLSI optimization techniques include:

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

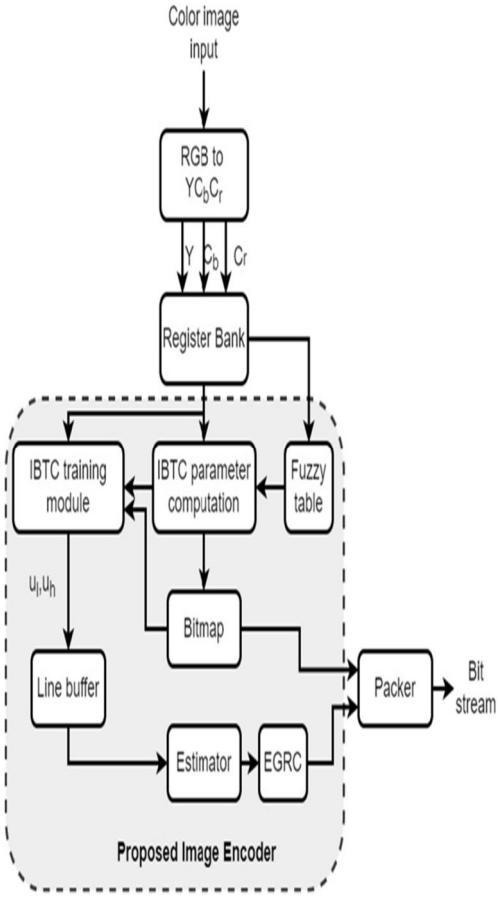


Fig 1: VLSI Implementation
(Source: Acquired from Google)

1) *Low-Power Design Techniques:* Energy is a critical issue in IoT devices, and with respect to energy efficiency, VLSI circuits for audio classification do not compromise power consumption.

- Clock Gating: This technique minimizes dynamic power as some areas in the circuit are switched off when they are unused, thereby "gating" or shutting out the clock to these regions.
- Power Gating: Power gating isolates power from entire sections of circuit that do not require computation and hence eliminates static power.
- Voltage Scaling: Decreasing signal voltages within circuits of VLSI consumes less power at the chip's cost at slightly lower levels of performance. This trade-off is switching voltage levels as needed according to processing loads required within the computer.

2) *Pipelining:* In pipelining, consecutive tasks in a processing stream are broken down into stages so as to allow them to occur at the same time. In audio classification, phases like signal acquisition, feature extraction and classification can all be done in different pipeline stages, so there is no jam at one stage waiting for it to process before allowing the next one to start.

- Efficiency: This way, pipelining reduces wait time and increases speed which makes it appropriate for real time applications.
- Latency Reduction: Since each stage is performed independently, pipelining is able to reduce overall latency to meet real-time targets of audio classification tasks [7].

3) *Parallelism:* As is the case with any form of computing, it is possible to design VLSI architectures to process in parallel so that several data points can be processed concurrently.

- Data-Level Parallelism: Data level parallelism allows many feature extraction or different classification tasks to be performed simultaneously which is useful for high numbers or frequent use of audio data.
- Instruction-Level Parallelism: Such approach allows to perform many narrow instructions at once, which is beneficial for complex tasks that require many arithmetic or similar operations.

©IJRASET: All Rights are Reserved

4) *Specific Proposals Relative to the Hardware:*

- Quantization: Quantization makes data representation easier by changing the size of numbers that are used during computations from 32 floating point bits or more to 8 bits or even binary. It is widely known that quantization impacts memory drastically and accelerates calculations, which is why it is used in resource-scarce environments for neural network deployment.
- Fixed-Point Arithmetic: While fixed-point operations are free from the exponential issues of floating-point arithmetic, they are accurate, inexpensive and power-efficient compared to floating-point arithmetic. Implementations in fixed point are also efficient and faster handling architectures in VLSI, which is a desirable attribute for real time applications. In terms of audio classification, with a view of reducing computational load, fixed point arithmetic is standardly sufficient to achieve a satisfactory precision.
- Application-Specific Integrated Circuits (ASICs): ASICs are very large-scale integrated circuits specially intended for a given operation [8]. In the classification of audio, ASICs prove advantageous in that they can be optimized for particular algorithms while being power-shavers but they are not ideal like FPGAs which are programmable processors.

5) *Use of Field-Programmable Gate Arrays (FPGAs):* FPGAs are a versatile type of VLSI, which allows for change in. They are perfect for real-time audio classification problems where certain characteristics of the system need to be changed for some time, such as in gaming applications by using dynamic configuration parameters in a fast-changing environment.

6) *Hybrid Architectures:* At other times there is added value in bringing on board both general purpose microcontrollers and specific dedicated hardware accelerators. The microcontroller deals with the general processing, while accelerators are used for segments of high computational requirements like feature extraction or some layers of the neural network.

IV. VLSI Architectures for Signal Processing:

Real-time audio classification in IoT devices is a high resource-demanding task, which necessitates adjustable and efficient VLSI architectures [9]. These architectures pay priorities to constraints of resources like power, memory and processing capability to deliver high precision, high speed work. This section describes the various sub-categories or VLSI architectures of IoT, management of memory and processor resources, approaches to energy-efficient design, as well as illustrating examples of successful VLSI implementations of audio signal processing for IoT applications.

A. *Architectural Designs for IoT:*

In large-scale interconnectivity, three main VLSI designs: ASICs, FPGAs and DSPs are worth highlighting for offering user-specific time-constrained signal processing for IoT class devices.

1) *Application-Specific Integrated Circuits (ASICs):*

a) Overview: ASICs are application specific integrated circuits that are closely designed to perform a particular task or set of tasks for example audio classification in IoT [10]. µP's, on the other hand, are general-purpose devices which, because of the need for flexibility in their tasks, lack the faster speed or low-power consumption of ASICs.

b) Advantages: Due to these reasons, ASICs consume lesser power and run faster compared to general-chips since they do not possess peripheral circuitry for redundant functions. The design of the model can be very simple, and still deliver quite accurate results for the audio classification tasks.

c) Disadvantages: ASICs have low adaptability; the devices are embedded systems that cannot be programmed for use in other applications once in the market, and as such, do not meet the dynamic IoT needs as efficiently as FPGAs.

2) *Field Programmable Gate Arrays (FPGAs):*

a) Overview: FPGAs are actual integrated circuits with a fixed array of logic blocks on a silicon substrate that can be reconfigured on the reconfiguration signal after manufacture to perform different digital logic functions than the original intended function. This reconfigurability makes FPGAs highly suitable for prototyping, designing systems in multiple iterations, designing systems that might need to be updated in real time or systems that require multiple processors.

b) Advantages: FPGAs can support different processing demands due to the reconfigurability property of the hardware solution. They enable concurrent processing, whereby multiple stages for signal processing can be run in parallel which is advantageous for IoT real-time audio streaming.

c) Disadvantages: FPGAs are power-hungrier than ASIC and take more board space and therefore not suitable for low power applications in stringent environments.

3) Digital Signal Processors (DSPs):

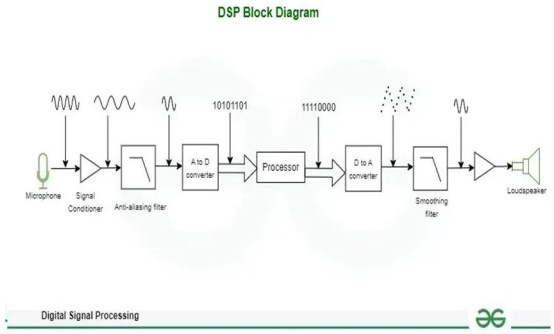


Fig 2: Digital Signal Processors (DSPs)
(Source: Acquired from Google)

a) Overview: DSPs are those microprocessors that are specifically designed for the signal processing activity such as filtering, FFT, and matrices [11]. Although not quite as adaptable as FPGAs or as fine-tuned as ASICs, DSPs are perfectly versatile and an ideal middle ground for a wide range of audio processing requirements in IoT.
b) Advantages: DSPs have quite high efficiency in implementing repeated operations on the signal and can work through complex algorithms with low power consumption and are suitable for audio classification.
c) Disadvantages: As stated before, DSPs are not as fast as ASICs in specific applications and typically dissipate more power than an ASIC created for similar work.

B. Memory and Processing Unit Improvement:

Specifically for real-time audio classification over the VLSI architecture, managing memory and processing units is more decisive. The following are strategies to enhance memory efficiency and streamline data processing in VLSI designs:

1) Memory Hierarchy Optimization:
a) Hierarchical Memory Structure: Thus, to improve the time needed to obtain data, a hierarchical memory design (for instance, incorporating cache memory) stores data most frequently used in multiple-access memory locations [12].
b) On-Chip vs. Off-Chip Memory: Avoiding utilization of off-chip memory for critically important information increases their availability in on-chip memory because gaining access to the on-chip memory requires less time and power consumption than off-chip memory.

2) Data Flow Optimization:
a) Data Locality: There are always methods applied in the conceptual and architectural design of VLSI so that data that is to be processed in the CPU is kept close to the CPU in order to minimize data transportation time.
b) Buffer Management: Effective scheduling methods are employed to eliminate constraints on input/output flows and provide continuous stream processing [13]. When it comes to real time audio processing circular buffers might be used in handling input streams of audio.

3) Processing Unit Enhancements:
a) Parallel Processing: The general organization of many VLSI architectures is based on multiple data streams or processing sections carrying out different aspects of a signal processing task in parallel [14]. For instance, feature extraction and classification stages in audio classification can benefit from parallel processing mechanisms.
b) Fixed-Point Arithmetic: The conversion from floating-point to fixed-point arithmetic helps the VLSI architectures offer more efficient performance by consuming less power, as fixed operations are easier to implement in architecture.

C. Energy-Efficient Designs:

Energy optimization is an essential factor for IoT devices, which are mostly deployed with either rechargeable batteries or energy-scavenging mechanisms only. Energy efficient VLSI designs across the various layers of design and abstraction provide ways to perform the signal processing required without consuming high amounts of power. Key techniques include:

1) Clock Gating:

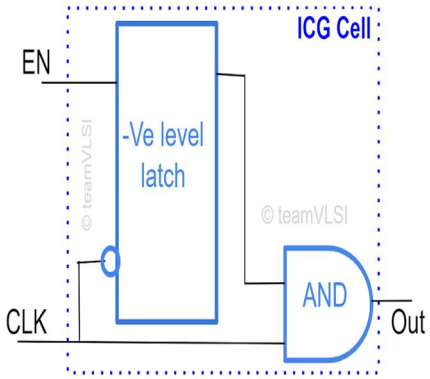


Fig 3: Integrated Clock Gating
(Source: Acquired from Google)

a) Principle: Clock gating lowers the switch activity by selectively disabling the clock signal to inactive areas in a power circuit. This technique assists in decreasing the amount of avoiding across the circuit, which is a primary reason for power consumption.

b) Application: When used in an audio classification system, clock gating is capable of eliminating the continuous processing of certain parts of the system where such processing is not required, that is, during periods of low activity or at night.

2) Power Gating:

a) Principle: Power gating is therefore one of the most efficient methods of lowering the extent of static power dissipation by removing the power from whole blocks of circuits that are not in use at that time [15]. This is especially beneficial in ultra-low-power use cases, a power-saving mechanism where energy-saving is crucial.

b) Application: For example, an audio classification VLSI circuit may incorporate power gating to off the feature extraction or classification part if there is no input signal for some time to save power.

3) Dynamic Voltage and Frequency Scaling (DVFS):

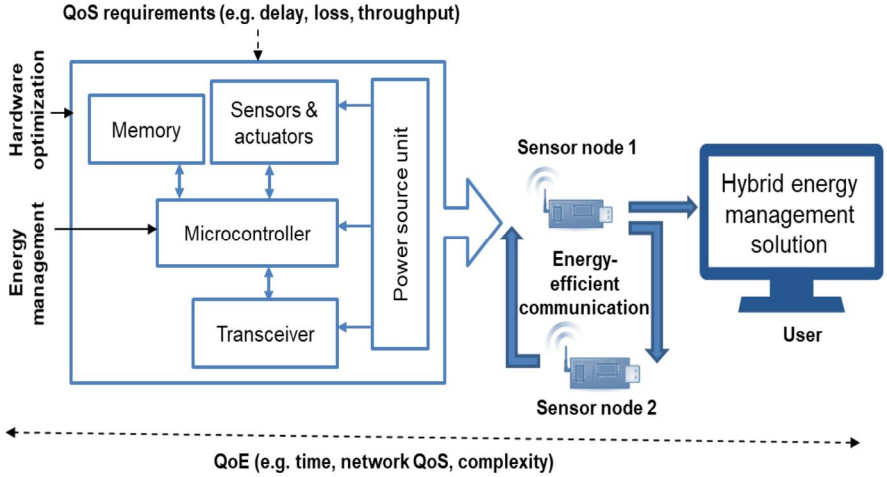


Fig 4: Dynamic Voltage and Frequency Scaling (DVFS)
(Source: Acquired from Google)

a) Principle: DVFS involves the manipulation of the voltage of a processing unit as well as the frequency of clock in relation to present processing requirements. The values of voltage and frequency can be decreased as far as possible in order to reduce energy dissipation especially when the computer is not heavily loaded.
b) Application: As for the real-time audio classification, DVFS can reduce power consumption during the time when the accuracy rate of classification is not very important, for example, during low noise or when the acoustical environment is stable.

D. Case Studies and Examples:

1) Edge-Optimized ASIC for Audio Processing:

a) Design: A special ASIC was developed to classify environmental sound into distinct categories relevant to a factory setting such as machinery noise or an alert sound. The ASIC implemented fixed point arithmetic throughout the signal processing stages and this resulted in high efficiency.
b) Result: This architecture was found to prove a reduction of energy consumption by as much as 70% lower compared with that of competitor microprocessor-based implementations all while preserving the classification accuracy to benefit a broad spectrum of battery-driven industrial IoT uses.

2) FPGA-Based Audio Classification for Surveillance Systems:

a) Design: An FPGA was programmed to operate a CNN for the purpose of accurately classifying sounds in real time in a surveillance system [16]. Concurrent processing of the audio features made it possible to quickly classify the audio events as gunshots, scream or the alarm, etc.
b) Result: Using the FPGA implementation, the algorithm was tested for classification speeds with a latency of less than 10ms rendering the solution ideal for real-time surveillance with low response time complexity. Although this design was more powerful than an ASIC as it consumed more power and was suitable for applications where reconfigurability and fast prototyping were inevitable.

3) DSP-Enhanced Wearable Health Monitoring Device:

a) Design: A DSP-based architecture was implemented in a wearable to classify respiratory sound for the detection signs of a respiratory disorder. They used MFCC for feature extraction and SVM for classification of the words [17].
b) Result: The DSP enabled real-time classification on a small battery and SVPs, going for more than 12 hours of continuous operation before recharging was necessary. The wearable device demonstrated that DSPs can offer high computational functionality with comparatively low energy demand in healthcare IoT.

V. CONCLUSION

Thus, the proposed real-time audio classification on VLSI architectures for IoT applications is possible and beneficial to improve functionality of devices that can perform important tasks in corresponding fields such as environmental control, healthcare and security. These applications are supported by VLSI designs including ASICs, FPGAs, and DSPs to enhance the power utilization efficiency, processing rates, and memory in IoT devices to overcome characteristic resource limitation. Methods like MFCC and STFT, applied to the feature extraction process reduce the amount of data in the audio stream and make it easier to process, with the improved efficiency owing to low power design methodologies such as clock gating and dynamic voltage scaling.

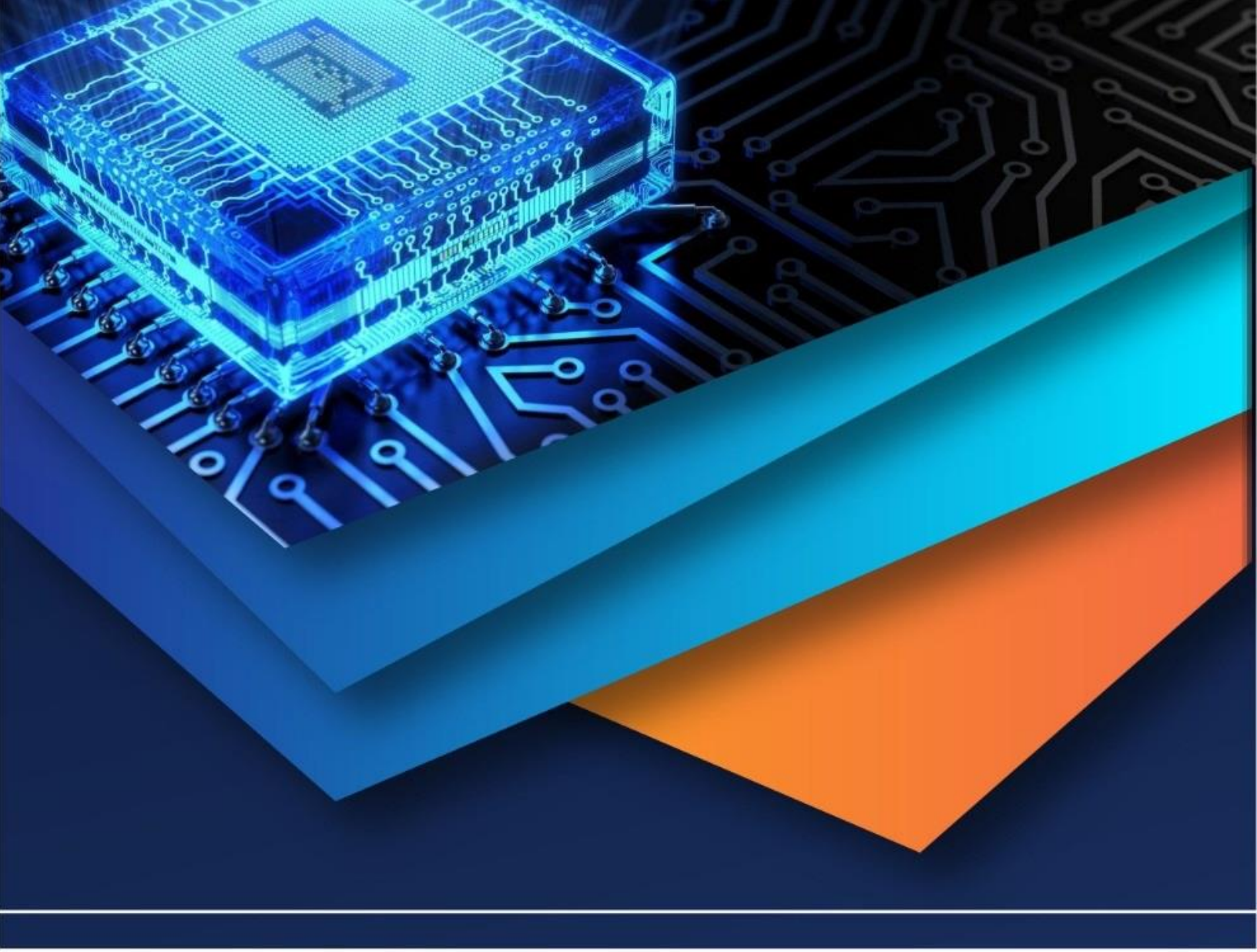
However, some issues are still observed and the first of them is the accuracy limitation with very low latency, the second is the scalability with fixed-function hardware implementations. Furthermore, adapting to change in needs of audio classification could be a bit challenging in some limited spaces. Though, innovative technologies offer some solutions to this problem. Neuromorphic computing for signals processing, which copies the approach of the brain, could be useful for audio recognition that requires low power yet high efficiency and TinyML enables to create lightweight machine learning directly on IoT devices thus minimizing the usage of external processors.

Additionally, there is a tendency for the development of bio-inspired signal processing and for further improvement of the ways to integrate hardware and software in VLSI systems. These innovations are designed to increase the capability of audio processing for IoT and to provide more growth for the IoT applications. Lastly, as VLSI technology advances, allocating real-time audio classification becomes even more reasonable and beneficial to expand IoT solutions in various industries.

REFERENCES

[1] Kornaros, G., 2022. Hardware-assisted machine learning in resource-constrained IoT environments for security: review and future prospective. IEEE Access, 10, pp.58603-58622. [1]
[2] Lamrini, M., Chkouri, M.Y. and Touhafi, A., 2023. Evaluating the performance of pre-trained convolutional neural network for audio classification on embedded systems for anomaly detection in smart cities. Sensors, 23(13), p.6227. [2]
[3] Góez, D., Beyazıt, E.A., Fletscher, L.A., Botero, J.F., Gaviria, N., Latré, S. and Camelo, M., 2024. Resource-Efficient Spectrum-Based Traffic Classification On Constrained Devices. IEEE Open Journal of the Communications Society. [3]
[4] Shuvo, M.M.H., Islam, S.K., Cheng, J. and Morshed, B.I., 2022. Efficient acceleration of deep learning inference on resource-constrained edge devices: A review. Proceedings of the IEEE, 111(1), pp.42-91. [4]
[5] Krishnamoorthy, R., Krishnan, K., Chokkalingam, B., Padmanaban, S., Leonowicz, Z., Holm-Nielsen, J.B. and Mitolo, M., 2021. Systematic approach for state-of-the-art architectures and system-on-chip selection for heterogeneous IoT applications. IEEE Access, 9, pp.25594-25622. [5]
[6] Hussain, H., Tamizharasan, P.S. and Rahul, C.S., 2022. Design possibilities and challenges of DNN models: a review on the perspective of end devices. Artificial Intelligence Review, pp.1-59. [6]
[7] Vora, N.R., Hajighasemi, A., Reynolds, C.T., Radmehr, A., Mohamed, M., Saurav, J.R., Aziz, A., Veerla, J.P., Nasr, M.S., Lotspeich, H. and Guttikonda, P.S., 2023. Real-Time Diagnostic Integrity Meets Efficiency: A Novel Platform-Agnostic Architecture for Physiological Signal Compression. arXiv preprint arXiv:2312.12587. [7]
[8] Kalwar, J.H. and Bhatti, S., 2024. Deep Learning Approaches for Network Traffic Classification in the Internet of Things (IoT): A Survey. arXiv preprint arXiv:2402.00920. [8]
[9] Yang, X., Liu, Z., Tang, K., Yin, X., Zhuo, C., Wei, Q. and Qiao, F., 2023. Breaking the energy-efficiency barriers for smart sensing applications with "Sensing with Computing" architectures. Science China Information Sciences, 66(10), p.200409. [9]
[10] Borrego-Carazo, J., Castells-Rufas, D., Biempica, E. and Carrabina, J., 2020. Resource-constrained machine learning for ADAS: A systematic review. IEEE Access, 8, pp.40573-40598. [10]
[11] Xu, T., 2021. Waveform-defined security: A low-cost framework for secure communications. IEEE Internet of Things Journal, 9(13), pp.10652-10667. [11]
[12] Islam, M.M., Nooruddin, S., Karray, F. and Muhammad, G., 2022. Internet of things: Device capabilities, architectures, protocols, and smart applications in healthcare domain. IEEE Internet of Things Journal, 10(4), pp.3611-3641. [12]
[13] Naveen, S., Kounte, M.R. and Ahmed, M.R., 2021. Low latency deep learning inference model for distributed intelligent IoT edge clusters. IEEE Access, 9, pp.160607-160621. [13]
[14] Ouda, H., Badr, A., Rashwan, A., Hassanein, H.S. and Elgazzar, K., 2022, May. Optimizing real-time ecg data transmission in constrained environments. In ICC 2022-IEEE International Conference on Communications (pp. 2114-2119). IEEE. [14]
[15] Chanal, P.M. and Kakkasageri, M.S., 2020. Security and privacy in IoT: a survey. Wireless Personal Communications, 115(2), pp.1667-1693. [15]
[16] Mazumder, A.N., Meng, J., Rashid, H.A., Kallakuri, U., Zhang, X., Seo, J.S. and Mohsenin, T., 2021. A survey on the optimization of neural network accelerators for micro-ai on-device inference. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 11(4), pp.532-547. [16]
[17] Vishwakarma, R., Monani, R., Rezaei, A., Sayadi, H., Aliasgari, M. and Hedayatipour, A., 2023, April. Attacks on continuous chaos communication and remedies for resource limited devices. In 2023 24th International Symposium on Quality Electronic Design (ISQED) (pp. 1-8). IEEE. [17]

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)