Infocommunications Journal

HTE 75 1949–2024

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

December 2024 Volume XVI Number 4 ISSN 2061-2079

Technically Co-Sponsored by

IEEE ComSoc IEEE Communications Society

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.
Infocommunications Journal is also included in the Thomson Reuters – Web of ScienceTM Core Collection, Emerging Sources Citation Index (ESCI)

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section

Supporters

FERENC VÁGUJHELYI – president, Scientific Association for Infocommunications (HTE)

The publication was produced with the support of the Hungarian Academy of Sciences and the NMHH

National Media and Infocommunications Authority • Hungary

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1051 Budapest, Bajcsy-Zsilinszky str. 12, Room: 502
Phone: +36 1 353 1027 • E-mail: info@hte.hu • Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Telecommunications and Media Informatics
Phone: +36 1 463 4189 • E-mail: pvarga@tmit.bme.hu

Subscription rates for foreign subscribers: 4 issues 10.000 HUF + postage

Publisher: PÉTER NAGY

HU ISSN 2061-2079 • Layout: PLAZMA DS • Printed by: FOM Media

Advances in Speech Recognition, Musical Anamnesis, Inter-ISP communication, MEC offloading and Software-Defined Radio

Pal Varga

Infocommunications is a vast domain, and our journal tries to capture the latest advances. 2024 was a turbulent year in many ways – full of technological advances and societal challenges. Regarding the Infocommunications Journal, 2024 was a very special year, as we handled a record number of 265 papers that has been submitted. This is the last issue, vith quite diverse topics – from speech recognition and transfer learning, through enhancing QoS of IoT devices by MEC offloading, Software-Defined Radios and inter-ISP communications. Let us briefly capture the topics in the December 2024 issue of the Infocommunications Journal.

Yan Meng and Peter Mihajlik investigate the use of adapter modules in pre-trained speech recognition models for cross-language transfer learning in low-resource automatic speech recognition tasks. Their study evaluates the impact of adapters on recognition accuracy, GPU memory consumption, and training duration across different models, including multilingual and target-language-trained foundational models. The results demonstrate that adapters enhance generalization while reducing overfitting, with significant gains in word error rate for models pre-trained in the target language. Moreover, adapters substantially reduce GPU memory usage during fine-tuning, highlighting their efficiency in optimizing resource-constrained training scenarios.

The paper by Marouane Myyara and his co-authors presents heuristic-based computation offloading algorithms to enhance Quality of Service (QoS) for IoT devices in Multi-access Edge Computing (MEC) environments. The proposed approach optimally offloads computational tasks from resource-constrained IoT devices to nearby edge locations, reducing latency and improving task execution performance. Evaluated using the EdgeCloudSim simulator, the algorithms demonstrated noticable reductions in service times and task failure rates compared to existing solutions. This work addresses key challenges in QoS management for MEC networks and contributes to advancing computation offloading strategies for resource-intensive IoT applications.

In their paper, Hari Krishnan S. and Syed Sadiqvali introduce a novel Finite Impulse Response (FIR) filter architecture based on Distributed Arithmetic (DA) and Look-Up Tables (LUTs) to enhance efficiency in Software-Defined Radio (SDR) systems. By addressing the limitations of traditional multiplier-based FIR filters, the proposed design reduces hardware complexity, power consumption, latency, and memory usage while improving throughput and bit error rate. The inclusion of a dynamic decimation factor and a highly adaptable Parallel Prefix Adder further optimizes the filter's performance, enabling flexible frequency response adjustments and faster partial product accumulation. Experimental results on an Artix-7 FPGA demonstrate significant improvements, including an operating speed of 260 MHz, power dissipation of 1 mW, delay of 190 ps, and throughput of 938.12 Mbps, making the DA-LUT-FIR filter a robust solution for real-time digital signal processing in future SDR applications.

Hamid Garmani et al. investigated interactions among ISPs and advertisers in an Information-Centric Networking (ICN) framework, focusing on collaborative caching of free content to enhance network performance and reduce distribution costs. Using game-theoretic models, they formulate ISP interactions as a non-cooperative game in their paper, proving the existence and uniqueness of a Nash equilibrium under certain conditions. They propose an iterative algorithm to guide ISPs toward equilibrium, balancing content caching strategies, pricing, and quality of service. Numerical simulations demonstrate that the approach benefits both ISPs and end-users, offering a scalable, economically incentivized solution for efficient content distribution in ICN environments.

In their paper, Fabian Simmank and his co-authors introduce Automated Musical Anamnesis (AMA), a scalable solution to streamline the manual and expertise-intensive process of gathering musical history for music-based interventions in dementia patients. By addressing cognitive and emotional challenges associated with dementia, AMA aims to enhance therapy relevance and reduce overstimulation, providing personalized and effective care. The study highlights the interdisciplinary nature of developing AMA, integrating methods from music therapy, technology, and cultural heritage to curate structured and meaningful digital repositories. This approach aims to expand access to music-based interventions, offering a cost-effective, sustainable, and inclusive therapeutic model to address the global rise in dementia cases.

Pal Varga is the Head of Department of Telecommunications and Artificial Intelligence at the Budapest University of Technology and Economics. His main research interests include communication systems, Cyber-Physical Systems and Industrial IoT, network traffic analysis, end-to-end QoS and SLA issues – for which he is keen to apply hardware acceleration and AI/ML techniques as well. He is advocating intent-based networking and the utilization of generative AI in network and service management. Besides being a member of HTE, he is a senior member of IEEE (both in ComSoc and IES). He is Editorial Board member in many journals, Associate Editor in IEEE TNSM, and the Editor-in-Chief of the Infocommunications Journal.

Assessing the Efficacy of Adapters in Cross-Language Transfer Learning For Low-Resource Automatic Speech Recognition

Yan Meng and Péter Mihajlik

Abstract—**In recent years, the application of adapter modules in large language models proved to be successful in reducing computing and memory costs during fine-tuning. In our paper, we apply adapters to the field of automatic speech recognition. Specifically, we add adapters to different pre-trained speech recognition models to evaluate their efficiency in cross-language transfer learning. In this study, the evaluations are extended to GPU memory consumption, training duration, and recognition accuracy. By comparing the effects of adapters added to different models, we further explore the impact of whether the foundational model was (pre-) trained in the target language.**

Index Terms—**Adapters, Whisper, Conformer, Fast Conformer, Cross-lingual transfer learning, speech recognition**

I. INTRODUCTION

With the implementation of neural networks in speech recognition, significant improvements have been achieved in neural speech-to-text (STT) models. When transformer [1] was initially proposed, it made significant progress mainly in machine translation tasks. Conformer [2] is an improvement and extension of the transformer model, used primarily in speech recognition tasks. Based on the LibriSpeech benchmark [3], Gulati et al. conducted experiments based on three different sizes of conformer models (small, medium, and large) [2]. The experimental results are improved by increasing the model parameters. The large-size Conformer model achieved excellent results with a word error rate (WER) of 2.1%/4.3% without using a language model and 1.9%/3.9% with an external language model on test/testother dataset. Fast Conformer [4] is a redesigned Conformer model with a novel downsampling schema, which is 2.8 times faster than the Conformer model. Fast Conformer is a newly proposed model, that has not been widely used in the training of low-resource data sets. Whisper [5] is a weakly-supervised model that can target multiple languages and tasks. The Whisper model has been trained on multiple language datasets, including Hungarian. Therefore, in automatic speech recognition tasks, high-quality results on specific distributions can be obtained without fine-tuning the Whisper model based on a Hungarian dataset.

Based on the English dataset, all the above models can achieve a lower word error rate (WER). For relatively low-resource languages such as Hungarian, direct training of ASR neural models from scratch is both time-consuming and may not result in optimal performances. Therefore, cross-lingual transfer learning methods are often applied for speech recognition tasks, especially for low-resource datasets ([6], [7], [8], [9], [10]). It means that when fine-tuning the model parameters, some pre-trained models are used to help models adapt to the distributional features of the language faster, e.g., fine-tuning the pre-trained English model based on the Hungarian dataset to obtain better WER results. However, as the number of model parameters increases, the cost required to fine-tune and train the model increases accordingly. Therefore, we need an approach to achieve better recognition performance of the model in a shorter time and with lower memory consumption.

To address these issues, especially for large-scale models, the Parameter Efficient Fine-Tuning (PEFT) method [11] has been proposed. Adapter is one of the core fine-tuning methods of the PEFT technique, which is a new module added between layers of a pre-trained network [11]. Adapter modules have two main features: a small number of parameters, and a near-identity initialization [12]. In every training step, the parameters of the original foundational model remain frozen, only all the parameters in the adaptors module are tuned. However their number is relatively small as compared to the foundational model parameters. This method effectively reduces the consumption of GPU memory in the training process. An increasing number of adapter types have been proposed for different types of basic models in ASR [13], [14]. In [15], Hou, et al. proposed a new adapter algorithm-based transformer structure for cross-lingual transfer learning, SimAdapter, and MetaAdapter, which was applied to parameter efficient cross-lingual speech adaptation. To explore the effect of adapters applied to the basic model, in [16], adapters were applied to both Transformer and Conformer architectures to comprehensively evaluate the adapter performance within the context of children's ASR. Huang, et al. [17] demonstrated that integrating adapters into End-to-End model can effectively mitigate catastrophic forgetting (CF), which is a common drawback of improving models through fine-tuning.

In this paper, we focus on studying the effect of adapters applied to the foundational model in cross-lingual transfer

Yan Meng is with the department of Telecommunications and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: yan.meng@edu.bme.hu).

Péter Mihajlik is with the department of Telecommunications and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary and the HUN-REN Hungarian Research Centre for Linguistics, Budapest, Hungary (e-mail: mihajlik@tmit.bme.hu).

DOI: 10.36244/ICJ.2024.4.1

TABLE I
CHARACTERISTICS OF DATABASE (BEA-BASE + COMMONVOICE).

	BEA-Base					CV12
	train-114	**dev-repet**	**dev-spont**	**eval-repet**	**eval-spont**	**test**
Length [hours]	71.2	0.65	4.02	0.95	4.91	6.80
Num of speakers	114	10	10	16	16	251
Num of segments	76 881	568	4 893	858	5 69	4 871
Num of characters	3.1M	28 467	154 994	43 448	197 738	250 709
Num of words	0.56M	4 110	27 939	6 229	35 178	35 485

learning based on the Hungarian dataset BEA-Base [18]. Meanwhile, by comparing different foundational models, we explore whether a foundational model pre-trained in the target language affects the effectiveness of adapter addition. Foundational models used in this paper are mainly Conformer-CTC, Fast Conformer [4], and Whisper [5]. All experiments are divided into two main aspects: first, the adapter module is added directly to the pre-trained model. Secondly, the adapter module is added to the model that has been fully fine-tuned. With these two types of experiments and using three different foundational models, we conclude that adapters significantly reduce GPU memory consumption. The application of adapter modules provided better results than full fine-tuning only in the Whisper model, suggesting that adding the adapter to models (pre-) trained already on the target language can perform better.

II. DATABASE AND BASELINE

A. Database

Throughout the experiment, the training and test dataset used is the Hungarian benchmark dataset BEA-Base [18]. The BEA-Base dataset contains both spontaneous speech and conversations with repeated elements, making it suitable for ASR research. The detailed dataset information of BEA-Base is shown in Table I, including training set train-114, verification set dev-spont, and test set eval-spont. In addition, the datasets include the repeated speech datasets dev-repet and eval-repet, as shown in Table I. To assess the generality of the model, an additional test dataset CommonVoice (CV) Hungarian v12.0 [19] was used in the study, which differs from BEA-Base in terms of recording conditions and speaker/speech diversity.

B. Fine-tuning baseline

The experimental baseline model comes from the paper [20], and the specific parameters can be referred to as the parameter setting in the paper. The baseline model is Conformer-CTC ([21], [22]) model from NeMo toolkit v1.62[1]. Conformer-CTC is a speech recognition model that combines Conformer architecture [2] and Connectionist Temporal Classification (CTC) [21] technology. The sizes of the Conformer-CTC model cited in this article are medium and large. The pre-trained model is in English, and the fine-tuning process uses the Hungarian dataset train-114 from BEA-Base for training. The results are shown in the first row of the two Conformer models in Table III.

The experiments were conducted on two distinct server configurations. The first configuration employed dual A6000 NVIDIA graphics cards, specifically designed for large-scale model experiments. Each card featured a substantial 48GB of memory and consumed 300 watts of power. In contrast, the second configuration utilized two NVIDIA Ge-Force GTX 1070 graphics cards tailored for lighter computational workloads. These cards were equipped with 8GB of memory, making them suitable for less memory-demanding tasks.

III. INCORPORATING ADAPTERS INTO ENGLISH PRE-TRAINED MODELS

In this section, Conformer-CTC model and adapter module are selected, provided from NVIDIA's NeMo toolkit v1.15.0[2]. There are two types of adapters used in experiments: Linear Adapter [11] and Multi-Head Attention Adapter [23]. In the subsequent exposition, the attention adapter is denoted as a tiny-attention adapter. The name of the adapter module is taken from the type of adapter function [12]. During fine-tuning experiments with adapters, only the parameters of the adapter module are updated, while the parameters of the original language model remain frozen. The linear adapter is a simple bottleneck structure feed-forward module [11]. Multi-Head attention adapter is an adapter model that combines multi-head self-attention [24] mechanism. The multi-head self-attention mechanism allows the model to focus on different parts of the input sequence separately under different contexts and positions.

Fast Conformer (FC) [4] was also selected for this experiment as the foundational model. But for the Fast Conformer experiment, all pre-trained models are provided by NVIDIA's NeMo toolkit v1.22.0[3].

A. Pre-trained Model with Adapters

For the Conformer experiments, we used three sizes of English pre-trained model provided by NVIDIA NeMo toolkit (STT En Conformer-CTC XLarge[4], STT En Conformer-CTC Large[5], STT En Conformer-CTC Medium[6]). During the ex-

[1] https://github.com/NVIDIA/NeMo/tree/v1.6.2

[2] https://github.com/NVIDIA/NeMo/tree/v1.15.0

[3] https://github.com/NVIDIA/NeMo/tree/v1.22.0

[4] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_conformer_ctc_xlarge

[5] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_conformer_ctc_large

[6] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_conformer_ctc_medium

TABLE II
CER(%) / WER(%) RESULTS BASED ON FINE-TUNING THE ENGLISH PRE-TRAINED CONFORMER AND FAST CONFORMER MODEL WITH LINEAR ADAPTER ADDED.

Foundational model	Total num params	Trainable params	BEA-Base dev-repet	BEA-Base dev-spont	BEA-Base eval-repet	BEA-Base eval-spont	CV12 test
FastConformer-Large	116M	0.52%	11.30 / 28.18	16.73 / 33.61	13.21 / 31.14	17.95 / 35.87	17.90 / 39.48
FastConformer-XLarge	610M	0.26%	8.84 / 23.50	14.40 / 31.11	10.56 / 26.70	15.14 / 32.25	16.13 / 37.28
Conformer-Medium	31.1M	0.98%	8.71 / 23.35	11.41 / 26.25	9.20 / 23.97	12.20 / 27.93	14.75 / 35.18
Conformer-Large	122M	0.50%	5.72 / 17.22	8.79 / 21.27	6.32 / 18.51	9.40 / 22.55	10.58 / 27.27
Conformer-XLarge	637M	0.25%	**4.19 / 12.21**	**8.29 / 20.03**	**4.77 / 13.28**	**8.73 / 21.08**	**9.53 / 25.22**

TABLE III
CER(%) / WER(%) RESULTS BASED ON TWO ROUNDS OF FINE-TUNING EXPERIMENTS ON ENGLISH PRE-TRAINED CONFORMER AND FAST CONFORMER MODELS OF DIFFERENT SIZES.

Foundational model	Total num params	Trainable params	Adapter type	BEA-Base dev-repet	BEA-Base dev-spont	BEA-Base eval-repet	BEA-Base eval-spont	CV12 test
FastConformer-Large	115M	100.00%	-	1.25 / 5.46	5.84 / 17.23	1.31 / 5.06	6.30 / 18.20	11.56 / 39.13
	116M	0.52%	Linear	1.25 / 5.43	5.83 / 17.18	1.30 / 5.02	6.32 / 18.19	11.55 / 39.02
	115M	0.06%	tiny-attention	1.23 / 5.36	5.82 / 17.20	1.30 / 5.05	6.28 / 18.19	11.54 / 39.00
FastConformer-XLarge	608M	100.00%	-	1.45 / 6.08	5.80 / 17.63	1.68 / 6.39	6.03 / 18.32	9.76 / 35.80
	610M	0.26%	Linear	1.46 / 6.11	5.79 / 17.60	1.69 / 6.44	6.07 / 18.38	9.81 / 35.87
	608M	0.03%	tiny-attention	1.47 / 6.11	5.79 / 17.59	1.70 / 6.45	6.06 / 18.37	9.80 / 35.88
Conformer-Medium	30.5M	100.00%	-	1.72 / 8.56	5.58 / 18.44	2.15 / 9.68	5.82 / 19.60	8.37 / 35.57
	30.8M	0.99%	Linear	1.72 / 8.54	5.58 / 18.44	2.15 / 9.65	5.83 / 19.61	8.36 / 35.57
	30.7M	0.12%	tiny-attention	1.72 / 8.56	5.58 / 18.44	2.15 / 9.68	5.82 / 19.60	8.37 / 35.57
Conformer-Large	121M	100.00%	-	**1.13 / 5.45**	**5.09 / 16.45**	**1.27 / 5.28**	**5.28 / 17.23**	**8.78 / 34.85**
	122M	0.50%	Linear	1.13 / 5.47	5.10 / 16.45	1.26 / 5.25	5.28 / 17.22	8.77 / 34.80
	121M	0.06%	tiny-attention	1.14 / 5.47	5.12 / 16.42	1.28 / 5.30	5.30 / 17.34	8.75 / 34.79

periments, we only use linear adapter modules combined with the English pre-trained model and then fine-tune the model for experiments based on the Hungarian BEA-Base train-114 dataset. For data augmentation, SpecAugment [25] and speed perturbation were applied, using the same configuration as [20]. Throughout the fine-tuning experimental process, for each experiment, we set the batch size to 16, and the learning rate to 0.001, and ran it on a GPU of the A6000 server for 100 epochs.

For the Fast Conformer experiments, we mainly used extra large and large-sized pre-trained English Fast Conformer models as foundational models (STT En Fast Conformer-CTC XLarge[7], STT En Fast Conformer-CTC Large[8]) and linear type adapter module. We use the same linear adapter module as the Conformer model experiment above. During the experiments, the batch size was set to 16, the learning rate was set to 0.001, and 100 training epochs were performed on one GPU of the A6000 server. Otherwise, the other settings were the same as the Conformer experiments described above. The experimental outcomes of the Conformer and Fast Conformer model are presented in Table II.

B. Fine-tuned Model with Adapters

This section focuses on adding adapters to the original model that has been fully fine-tuned based on the BEA-Base dataset to further explore the impact of adding adapters on the fully fine-tuned model. The whole experimental process consists of two rounds of fine-tuning experiments. First, the pre-trained English model was fully fine-tuned, and then the adapter module was added to the fully fine-tuned model to fine-tune it again. For the Conformer model, in the first round of the fine-tuning phase, we used the same parameters as [20], batch size of 32, 200 training epochs, etc. In the second fine-tuning phase, two types of adapter modules were used, linear and tiny-attention. During the fine-tuning process, we set the learning rate to 0.01, training epochs to 100, and other parameters consistent with the pre-trained model experiments in the previous subsection, batch size of 16. The experimental results are shown in Table III. For the Fast Conformer model, in the first fine-tuning phase, the batch size was 96, the learning rate was 0.01, and the experiments were trained in 150 epochs. In the second fine-tuning phase, two types of adapters were used, with a batch size of 96, a learning rate of 0.02, and 50 epochs training. The results of this experiment are displayed in Table III.

C. Result Analysis

Figure 1 shows the memory consumption of the two fine-tuning methods in the Conformer model within a single epoch. The blue line represents a large-sized Conformer model, and the orange line represents a medium-sized Conformer model. The dotted line represents the experimental results of full fine-tuning of the original pre-trained model, while the solid line represents the experimental results of fine-tuning the original

[7] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_fastconformer_ctc_xlarge

[8] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_fastconformer_ctc_large

pre-trained model with the addition of a linear adapter to it. Throughout the experiment, the batch size of training was set to 16, and GPU memory utilization was logged at intervals of 5 seconds. The results show that, regardless of model size, training with a linear adapter takes significantly less memory than direct full fine-tuning the Conformer model. In addition, when calculating the training duration of one epoch, it is clear that the training duration is relatively short for the linear adapter experiments.

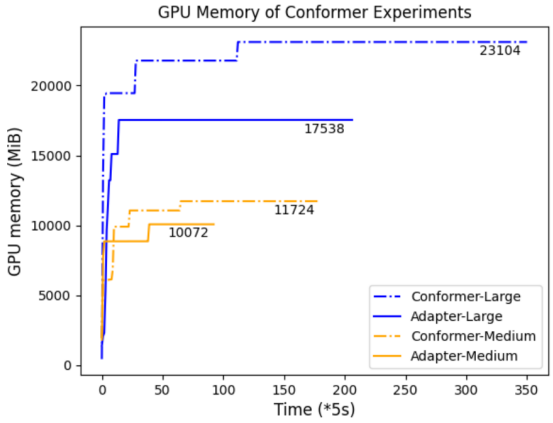


Fig. 1. GPU memory consumption from fine-tuning experiments on medium and large-sized English pre-trained Conformer models. Direct full fine-tuning the English pre-trained model is denoted by the dotted line, whereas incorporating adapters into the pre-trained model for fine-tuning is illustrated with the solid line.

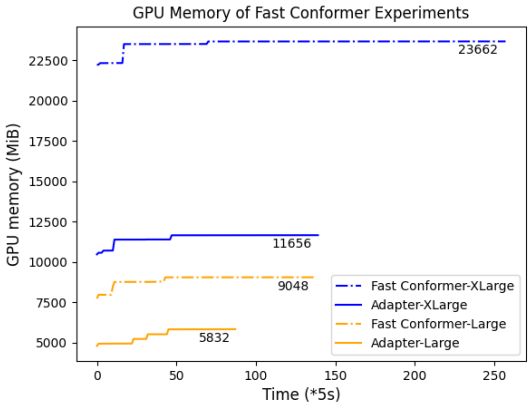


Fig. 2. GPU memory consumption from fine-tuning experiments on large-sized English pre-trained Fast Conformer models. Direct full fine-tuning the English pre-trained model is shown by the dotted line, whereas incorporating adapters into the pre-trained model for fine-tuning is illustrated by the solid line.

Figure 2 shows the GPU memory recorded every 5 seconds in two fine-tuning experiments using the English pre-trained extra large and large-sized Fast Conformer model. The dotted line indicates fully fine-tuning the pre-trained model directly, and the solid line indicates that the linear adapter is added to the pre-trained model for fine-tuning. The training batch size is also set as 16.

TABLE IV
RESULTS SHOW THE MAXIMUM GPU MEMORY CONSUMED IN THE FINE-TUNING EXPERIMENTS FOR CONFORMER AND FAST CONFORMER MODELS. CORRESPONDS TO THE VALUES IN FIGURE 1 AND 2.

Foundational model	Adapter	GPU memory (MiB)
Conformer-Large	-	23 104
	Linear	17 538
Conformer-Medium	-	11 724
	Linear	10 072
FastConformer-XLarge	-	23 662
	Linear	11 656
FastConformer-Large	-	9 048
	Linear	**5 832**

As can be observed from Figure 2, for fine-tuning linear adapter experiments, both the training duration and memory usage required were significantly reduced compared to the full fine-tuning experiment. Comparing the results of Figure 1 and Figure 2, it can be clearly seen that the training time of the Fast Conformer model is much shorter than that of the traditional Conformer model.

Observing Table II, When fine-tuning a cross-language pre-trained model with an adapter added, lower word error rates can also be achieved. However, for Fast Conformer model experiments, the results obtained are significantly worse than those of the Conformer model. We infer that the addition of the adapter to the Fast Conformer resulted in a poorer convergence of the overall model. By comparing the results of full fine-tuning experiments in Table III, the Fast Conformer results in a lower WER in a shorter training duration relative to the Conformer model, demonstrating the superiority of the Fast Conformer model. The overall analysis of Table III shows that for adding separately two types of adapter modules into the Fast Conformer and Conformer model, the experimental results have a relatively limited improvement on both CV and the BEA-Base datasets.

Comparing the results in the two tables, we found that adding adapters to the Conformer model and training for cross-lingual transfer learning did not show superior performance over the full fine-tuning results on the BEA-Base dataset. But it is worth noting that in the experiment of adding an adapter into the pre-trained Conformer model, the Xlarge-sized and large-sized models achieved lower character/word error rates (CER/WER) on the CommonVoice dataset, 9.53%/25.22% and 10.58%/27.27% respectively. By comparing the results in Table II (27.27%) and Table III (34.85%), the WER of the large-sized model on the CommonVoice dataset relatively decreased by about 21.7% when incorporating adapter modules. Comparing the results of the Fast Conformer model in the two tables, adding the adapter directly to the English pre-trained Fast Conformer model does not provide better results based on the BEA-Base dataset. Nevertheless, on the CommonVoice dataset, we still obtained similar WER results (39.48%) compared to the full fine-tuning results (39.13%). We infer that adding adapters to the foundational model for cross-lingual transfer learning did not significantly improve the word error rate on the original dataset. However, the model performs consistently on multiple datasets with wider

TABLE V
CER(%) / WER(%) ZERO-SHOT RESULTS OF DIFFERENT SIZES WHISPER MODEL ON HUNGARIAN DATASET.

Model	Total num params	BEA-Base				CV12
		dev-repet	dev-spont	eval-repet	eval-spont	test
whisper-small	242M	6.71 / 32.99	19.67 / 41.25	7.47 / 35.21	20.22 / 41.80	9.83 / 41.05
whisper-medium	764M	4.82 / 21.92	17.97 / 37.18	5.18 / 22.33	19.46 / 38.67	6.91 / 27.61
whisper-large-v2	1.54B	**3.74 / 17.54**	**17.06 / 33.17**	**3.99 / 18.04**	**17.06 / 32.76**	**5.27 / 20.41**

applicability and robustness. In contrast, fine-tuning tends to overfit the current dataset, resulting in poorer performance on external datasets. In addition, by comparing the results in two figures, we concluded that the adapter approach significantly reduces training time and GPU memory consumption.

IV. INCORPORATING ADAPTERS INTO MULTILINGUAL PRE-TRAINED MODELS

In the study of adding the adapter module into multilingual pre-trained models, we chose the typical weakly supervised model Whisper [5] as the foundational model. Different from the Conformer model experiments, the Whisper model has been trained on multiple languages, including Hungarian, and thus fine-tuning is a multilingual to monolingual transfer learning process.

This section carries out related experiments by using three Whisper models of different sizes. First, this study directly evaluate the speech recognition accuracy of Whisper models on Hungarian datasets. The CER and WER in the experimental results in Table V are normalized results. The models and codes used are from the Speech Brain [26], [27]. Among the zero-shot results, the experimental results of Whisper-large-V2 and Whisper-medium were quoted from paper [20]. Second, fine-tuning the Whisper model on the Hungarian dataset. Last, two types of Parameter-Efficient Fine-Tuning (PEFT) methods [11], namely Low-Rank Adaptation (LoRA) [28] and Adaptive Low-Rank Adaptation (AdaLoRA) [29] were used for the experiment. The implementation of the PEFT method aims to minimize the number of parameters while maintaining the performance of the model, thus improving the parameter efficiency. The code we used for train and evaluating the Whisper experiments in this section is available at https://github.com/MengYan0901/Whisper-Experiments.

This section focuses on the effect of integrating the Whisper model with two PEFT methods on the accuracy of speech recognition in Hungarian. Compared with the experimental group of the Conformer model, the effect of adding adapters to the foundational model trained by the target language in transfer learning is further discussed. Different sizes of Whisper models (Large-V2[9], Medium[10], Small[11]) were used and trained based on the BEA-Base train-114 dataset. In the testing phase, the CV test set and the BEA-Base test set were used to evaluate the performance of the model. Tests on multiple datasets aim to more comprehensively examine the experimental results of model training.

[9] https://huggingface.co/openai/whisper-large-v2
[10] https://huggingface.co/openai/whisper-medium
[11] https://huggingface.co/openai/whisper-small

A. Fine-tuning Whisper model

In this experiment, Whisper models are fine-tuned across three distinct scales. Specifically, for the Whisper-large-v2 model, the approach adopted in Paper [20] was used, where the encoder part was frozen and only the decoder part was fine-tuned. Conversely, for the Whisper-medium and Whisper-small models, fine-tune the parameters of the entire model. The outcomes of these experiments are shown in Table VI. The results of Whisper-V2 and Whisper-Medium are cited from the paper [20]. Regarding the experiments of the Whisper-Small model, the learning rate is set to 3e-4, the batch size to 16, and training is conducted for 5 epochs on one GPU of the A6000 server. The results distinctly demonstrate a notable decrease in word error rate (WER) as the number of model parameters increased, showing significant improvement, particularly within the CommonVoice dataset.

B. Incorporating Adapters into Whisper Model

Low-Rank Adaptation (LoRA) [28] is a popular Parameter-Efficient Fine-Tuning (PEFT) method [11]. When LoRA is used for downstream fine-tuning tasks, the parameters of the foundational model remain frozen throughout the training process, while the trainable rank decomposition matrices are integrated into each layer of the model transformer architecture [28]. This strategy significantly reduces the trainable parameters of the model. As indicated in the "Trainable params" column of Table VII, trainable parameters are only approximately 1% of the total model parameters, thereby reducing training duration and GPU memory usage. Adaptive Low-Rank Adaptation (AdaLoRA) [29] is a derivative of LoRA that manages the count of parameters introduced by LoRA. Throughout training, AdaLoRA will allocate parameters to different weight matrices based on the degree of adaptation to the task. The weight matrix that is more adaptable to the task will be assigned more parameters for training. As shown in the Trainable params column of Table VIII, trainable parameters only account for about 0.5% of all model's parameters.

Our training procedure follows the same configuration as the last fine-tuning section, but we change the learning rate to 1e-3 for both LoRA and AdaLoRA experiments, and the batch size remains unchanged at 16, still running on one GPU of the A6000 server for 5 epochs. By integrating two types of adapters (namely LoRA and AdaLoRA) into the Whisper model, the experimental results given in Tables VII and VIII show the impact of their application.

TABLE VI
CER(%) / WER(%). RESULTS BASED ON DIRECT FINE-TUNING THE WHISPER MODEL ON THE HUNGARIAN DATASET.

Model	Total num params	Trainable params	BEA-Base dev-repet	BEA-Base dev-spont	BEA-Base eval-repet	BEA-Base eval-spont	CV12 test
whisper-small	242M	100%	3.15 / 12.21	9.70 / 26.44	4.17 / 14.35	10.56 / 28.63	20.59 / 58.23
whisper-medium	764M	100%	1.31 / 5.38	7.96 / 18.83	1.50 / 4.90	9.33 / 20.60	7.83 / 27.93
whisper-large-v2	1.54B	58.75%	**1.01 / 4.45**	**7.10 / 16.96**	**1.23 / 4.37**	**8.46 / 18.69**	**6.19 / 23.69**

TABLE VII
CER(%) / WER(%). RESULTS BASED ON INCORPORATING LOW-RANK ADAPTATION (LORA) FOR FINE-TUNING ON THE HUNGARIAN DATASET, UTILIZING WHISPER AS THE FOUNDATIONAL MODEL.

Founditional model	Total num params	Trainable params	BEA-Base dev-repet	BEA-Base dev-spont	BEA-Base eval-repet	BEA-Base eval-spont	CV12 test
whisper-small	245M	1.44%	3.37 / 15.47	7.48 / 23.07	3.71 / 15.40	8.75 / 25.59	10.25 / 40.21
whisper-medium	773M	1.22%	2.12 / 10.02	5.80 / 17.91	2.61 / 10.71	6.36 / 19.37	7.66 / 31.29
whisper-large-v2	1.56B	1.01%	**1.65 / 7.13**	**4.74 / 14.89**	**1.68 / 6.66**	**5.08 / 15.56**	**5.91 / 24.22**

TABLE VIII
CER(%) / WER(%). RESULTS BASED ON INCORPORATING ADAPTIVE LOW-RANK ADAPTATION (ADALORA) FOR FINE-TUNING ON THE HUNGARIAN DATASET, UTILIZING WHISPER AS THE FOUNDATIONAL MODEL.

Founditional model	Total num params	Trainable params	BEA-Base dev-repet	BEA-Base dev-spont	BEA-Base eval-repet	BEA-Base eval-spont	CV12 test
whisper-small	243M	0.55%	4.32 / 19.49	9.21 / 26.72	4.84 / 20.23	9.35 / 27.46	10.71 / 41.68
whisper-medium	767M	0.46%	2.40 / 10.68	5.72 / 17.67	2.55 / 10.80	6.24 / 18.97	6.62 / 27.49
whisper-large-v2	1.55B	0.38%	**1.73 / 8.15**	**4.86 / 15.17**	**1.81 / 7.53**	**5.21 / 15.77**	**5.47 / 22.64**

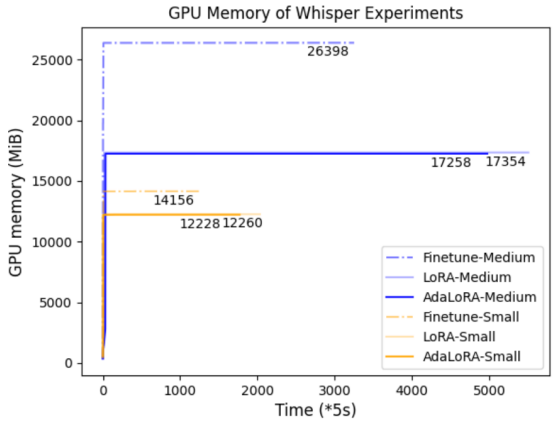


Fig. 3. GPU memory consumption from fine-tuning experiments on medium-sized and large-sized Whisper models in one epoch. Direct fine-tuning is indicated by the dotted line, incorporating LoRA into the model for fine-tuning is illustrated by the solid light blue line, and incorporating AdaLoRA into the model for fine-tuning is shown by the solid blue line.

TABLE IX
RESULTS SHOWS THE MAXIMUM GPU MEMORY CONSUMED IN THE FINE-TUNING EXPERIMENTS FOR WHISPER MODELS (CORRESPONDS TO THE VALUE IN THE FIGURE 3).

Foundational model	Adapter	GPU memory (MiB)
whisper-medium	-	26 398
	LoRA	17 354
	AdaLoRA	17 258
whisper-small	-	14 156
	LoRA	12 260
	AdaLoRA	**12 228**

C. Result Analysis

In Tables VI, VII and VIII, the data in bold font represent the best results obtained in the three experiments. It is clear that the performance of the model improves with increasing the number of model parameters whether LoRA or AdaLoRA is used. Comparing results in three tables, particularly in spontaneous speech tasks, both LoRA and AdaLoRA demonstrate superior performance, with the lowest WER 14.85%/15.16% on the dev-spont/eval-spont datasets of the BEA-Base dataset.

Figure 3 shows the memory occupation when fine-tuning the Whisper model in three different ways. Since the experiments with the whisper-large-V2 model were only trained on the decoder part of the model's parameters, it is not suitable for memory occupation comparisons. The results in Table IX correspond to the maximum memory occupied by each experiment in Figure 3. It can be observed from Figure 3 and Table IX that memory occupation can be significantly reduced when fine-tuning the Whisper model with LoRA or AdaLoRA. Especially for the medium-sized model, the memory occupation can be reduced by approximately 1/3, and the effect of reducing memory occupation is more significant as the model parameters increase. By comparing the CER/WER results of three tables, it can be concluded that for the Whisper model, the addition of an adapter module achieves a reduction in memory occupation and an improvement in speech recognition accuracy.

V. CONCLUSION

In this paper, we added adapter modules to different foundational models for automatic speech recognition tasks. When adding the adapter module to the foundational model that has not been trained in the target language, the test accuracy on the BEA-Base dataset was decreased while a significant improvement could be obtained on the CV dataset. We inferred that adapters can preserve the model's generalization ability without over-fitting the model to a specific training dataset, thus helping to maintain model excellence across multiple datasets. For a multilingual Whisper model (pre-) trained in Hungarian, adding the adapter module significantly reduced the word error rate on both datasets. The comparison of these two models shows that the addition of adapters can benefit from the foundational model (pre-) trained on the target language and achieve higher recognition accuracy. Furthermore, the addition of the adapter module showed a significant reduction in GPU memory consumption for all models during fine-tuning. Future research will further explore the adapter's efficacy for adding to other ASR models in cross-language transfer learning.

ACKNOWLEDGMENT

This research benefited greatly from the support provided by the Hungarian Linguistic Research Center in the development of the BEA-Base dataset. This work was supported partially by NKFIH-828- 2/2021(MILab), by the NVIDIA Academic Hardware Grant, and by the NKFIH K143075 and K135038 projects of the NRDI Fund. Thanks are also extended to the Budapest University of Technology and Economics and NVIDIA Academic Hardware Grant for their vital contribution including but not limited to hardware support.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017, DOI: 10.48550/arXiv.1706.03762.

[2] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu et al., "Conformer: Convolution-augmented transformer for speech recognition," *arXiv preprint arXiv:2005.08100*, 2020, DOI: 10.48550/arXiv.2005.08100.

[3] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2015, pp. 5206–5210, DOI: 10.1109/ICASSP.2015.7178964.

[4] D. Rekesh, N. R. Koluguri, S. Kriman, S. Majumdar, V. Noroozi, H. Huang, O. Hrinchuk, K. Puvvada, A. Kumar, J. Balam et al., "Fast conformer with linearly scalable attention for efficient speech recognition," in *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2023, pp. 1–8, DOI: 10.1109/ASRU57492.2023.1234567.

[5] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. Mcleavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 28 492–28 518, DOI: 10.48550/arXiv.2212.04356.

[6] S. Bansal, H. Kamper, K. Livescu, A. Lopez, and S. Goldwater, "Pre-training on high-resource speech recognition improves low-resource speech-to-text translation," *arXiv preprint arXiv:1809.01431*, 2018, DOI: 10.48550/arXiv.1809.01431.

[7] C. Yu, Y. Chen, Y. Li, M. Kang, S. Xu, and X. Liu, "Cross-language end-to-end speech recognition research based on transfer learning for the low-resource tujia language," *Symmetry*, vol. 11, no. 2, p. 179, 2019, DOI: 10.3390/sym11020179.

[8] J. Huang, O. Kuchaiev, P. O'Neill, V. Lavrukhin, J. Li, A. Flores, G. Kucsko, and B. Ginsburg, "Cross-language transfer learning, continuous learning, and domain adaptation for end-to-end automatic speech recognition," *arXiv preprint arXiv:2005.04290*, 2020, DOI: 10.48550/arXiv.2005.04290.

[9] J. Luo, J. Wang, N. Cheng, E. Xiao, J. Xiao, G. Kucsko, P. O'Neill, J. Balam, S. Deng, A. Flores et al., "Cross-language transfer learning and domain adaptation for end-to-end automatic speech recognition," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2021, pp. 1–6, DOI: 10.1109/ICME51207.2021.9428334.

[10] J. Zhao and W.-Q. Zhang, "Improving automatic speech recognition performance for low-resource languages with self-supervised models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1227–1241, 2022, DOI: 10.1109/JSTSP.2022.3184480.

[11] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," in *International Conference on Machine Learning*. PMLR, 2019, pp. 2790–2799, DOI: 10.48550/arXiv.1902.00751.

[12] J. He, C. Zhou, X. Ma, T. Berg-Kirkpatrick, and G. Neubig, "Towards a unified view of parameter-efficient transfer learning," *arXiv preprint arXiv:2110.04366*, 2021, DOI: 10.48550/arXiv.2110.04366.

[13] Y. Yu, C.-H. H. Yang, J. Kolehmainen, P. G. Shivakumar, Y. Gu, S. R. R. Ren, Q. Luo, A. Gourav, I.-F. Chen, Y.-C. Liu et al., "Low-rank adaptation of large language model rescoring for parameter-efficient speech recognition," in *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2023, pp. 1–8, DOI: 10.1109/ASRU57964.2023.10389632.

[14] N. Inoue, S. Otake, T. Hirose, M. Ohi, and R. Kawakami, "Elp-adapters: Parameter efficient adapter tuning for various speech processing tasks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024, DOI: 10.1109/TASLP.2024.3434445.

[15] W. Hou, H. Zhu, Y. Wang, J. Wang, T. Qin, R. Xu, and T. Shinozaki, "Exploiting adapters for cross-lingual low-resource speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 317–329, 2021, DOI: 10.1109/TASLP.2021.3138674.

[16] T. Rolland and A. Abad, "Exploring adapters with conformers for children's automatic speech recognition," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 12 747–12 751, DOI: 10.1109/ICASSP48485.2024.10447091.

[17] S. Vander Eeckt and H. Van Hamme, "Using adapters to overcome catastrophic forgetting in end-to-end automatic speech recognition," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5, DOI: 10.1109/ICASSP49357.2023.10095837.

[18] P. Mihajlik, A. Balog, T. E. Graczi, A. Kohari, B. Tarján, and K. Mady, "BEA-base: A benchmark for ASR of spontaneous Hungarian," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, Jun. 2022, pp. 1970–1977, DOI: 10.48550/arXiv.2202.00601.

[19] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common voice: A massively-multilingual speech corpus," *arXiv preprint arXiv:1912.06670*, 2019, DOI: 10.48550/arXiv.1912.06670.

[20] P. Mihajlik, M. S. Kádár, G. Dobsinszki, Y. Meng, M. Kedalai, J. Linke, T. Fegyó, and K. Mády, "What kind of multi-or cross-lingual pre-training is the most effective for a spontaneous, less-resourced asr task?" in *2nd Annual Meeting of the Special Interest Group on Under-resourced Languages: SIGUL 2023*, 2023, DOI: 10.21437/SIGUL.2023-13.

[21] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 369–376, DOI: 10.1145/1143844.1143891.

[22] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu et al., "Conformer: Convolution-augmented transformer for speech recognition," *arXiv preprint arXiv:2005.08100*, 2020, DOI: 10.48550/arXiv.2005.08100.

[23] H. Zhao, H. Tan, and H. Mei, "Tiny-attention adapter: Contexts are more important than the number of parameters," *arXiv preprint arXiv:2211.01979*, 2022, DOI: 10.48550/arXiv.2211.01979.

[24] E. Voita, D. Talbot, F. Moiseev, R. Sennrich, and I. Titov, "Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the 7 rest can be pruned," *arXiv preprint arXiv:1905.09418*, 2019, DOI: 10.48550/arXiv.1905.09418e.

[25] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "Specaugment: A simple data augmentation method for automatic speech recognition," *arXiv preprint arXiv:1904.08779*, 2019, DOI: 10.48550/arXiv.1904.08779.

[26] M. Ravanelli, T. Parcollet, P. Plantinga, A. Rouhe, S. Cornell, L. Lugosch, C. Subakan, N. Dawalatabad, A. Heba, J. Zhong, J.-C. Chou, S.-L. Yeh, S.-W. Fu, C.-F. Liao, E. Rastorgueva, F. Grondin, W. Aris, H. Na, Y. Gao, R. D. Mori, and Y. Bengio, "SpeechBrain: A general-purpose speech toolkit," 2021, DOI: 10.48550/arXiv.2106.04624.

[27] M. Ravanelli, T. Parcollet, A. Moumen, S. de Langen, C. Subakan, P. Plantinga, Y. Wang, P. Mousavi, L. D. Libera, A. Ploujnikov, F. Paissan, D. Borra, S. Zaiem, Z. Zhao, S. Zhang, G. Karakasidis, S.-L. Yeh, P. Champion, A. Rouhe, R. Braun, F. Mai, J. Zuluaga-Gomez, S. M. Mousavi, A. Nautsch, X. Liu, S. Sagar, J. Duret, S. Mdhaffar, G. Laperriere, M. Rouvier, R. D. Mori, and Y. Esteve, "Open-source conversational ai with SpeechBrain 1.0," 2024, DOI: 10.48550/arXiv.2407.00463.

[28] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021, DOI: 10.48550/arXiv.2106.09685.

[29] Q. Zhang, M. Chen, A. Bukharin, N. Karampatziakis, P. He, Y. Cheng, W. Chen, and T. Zhao, "Adalora: Adaptive budget allocation for parameter-efficient fine-tuning," *arXiv preprint arXiv:2303.10512*, 2023, DOI: 10.48550/arXiv.2303.10512.

Yan Meng is a Ph.D. student at Department of Telecommunications and Artificial Intelligence, Budapest University of Technology and Economics, Hungary. She received her management bachelor's degree in information management and information systems at Beijing Technology and Business University (2020) and master of science in computer science engineering at Budapest University of Technology and Economics (2022). Her current research interests include automatic speech recognition based on low-resource datasets, spontaneous speech recognition, Deep Learning, signal processing and acoustic models.

Péter Mihajlik PhD, is a senior research fellow at the Dept. of Telecommunications and Artificial Intelligence (TMIT), Faculty of Electrical Engineering and Informatics (VIK), Budapest University of Technology and Economics (BME). He also affiliated with the Hungarian Research Centre for Linguistics, Hungarian Research Network (HUN-REN) as part-time senior researcher. He got the MSc in Electrical Engineering in 1999 at BME and obtained the PhD in Automatic Speech Recognition in 2011. He is the head of Speech Recognition Group in SmartLabs/BME-TMIT.

Enhancing QoS for IoT Devices through Heuristics-based Computation Offloading in Multi-access Edge Computing

Marouane Myyara, Oussama Lagnfdi, Anouar Darif, and Abderrazak Farchane

***Abstract*—Multi-access Edge Computing (MEC) networks, particularly with the advent of 5G, aim to reduce latency and increase speed to meet the demands of resource-intensive applications in the Internet of Things (IoT), such as private wireless networks, online gaming, industry, and remote healthcare. These applications require guaranteed performance. However, while Quality of Service (QoS) management is well established in the Cloud, improving it remains a challenge in MEC en- vironments. This study addresses this challenge by proposing heuristic computation offloading algorithms for IoT-intensive devices in MEC networks. These algorithms aim to minimize service time while maximizing the QoS, taking into account tasks and resource characteristics to determine the optimal execution location for IoT device applications. We evaluated our approach using the EdgeCloudSim simulator, and the results demonstrate its superiority over existing solutions. Our approach significantly improves QoS by reducing the service time of IoT application tasks. This research fills a gap in efficient QoS improvement and contributes to advances in computation offloading strategies in MEC environments. It paves the way for enhanced performance of IoT applications in these networks.**

***Index Terms*—MEC, IoT, Computation Offloading, Quality of Service, Heuristic Algorithms, EdgeCloudSim.**

I. Introduction

The Internet of Things (IoT) is a rapidly expanding ecosystem of diverse physical objects connected through various networks, both wired and wireless [1]. It enhances internet utilization by linking mobile devices and sensors. However, IoT faces significant challenges, including high network latency, availability, and mobility, often mitigated through Cloud Computing [2]. Resource-constrained IoT devices struggle with limited processing power, memory, and battery life, complicating the execution of complex tasks. Additionally, the surge in IoT devices can overload networks accessing Cloud servers, hindering low-latency and high-capacity applications. To address these issues, edge computing paradigms, such as Mobile/Multi-access Edge Computing (MEC), have emerged.

MEC, introduced by the European Telecommunications Standards Institute (ETSI) [3], enhances edge intelligence and boosts processing and storage capabilities [4]. By bringing cloud functionalities closer to the Radio Access Network (RAN), it provides ultra-low latency and network context awareness. ETSI identifies IoT as a key use case for MEC [4], emphasizing the mutual benefits of their integration [2]. From the MEC perspective, IoT expands MEC services to various devices, while from the IoT viewpoint, MEC architecture offers computing resources closer to users. This integration significantly aids resource-constrained IoT devices by providing access to powerful computing at the network edge, enabling efficient task execution and improved service quality. According to [5], this integration provides three main benefits: reduced infrastructure traffic, lower application latency, and scalable network services. The key advantage is decreased latency through MEC, which shortens distances and transmission times between resources, facilitating efficient resource provision for processing IoT applications [6].

Despite growing interest in MEC for IoT, research on computation offloading remains limited. Recent studies focus on reducing latency in MEC networks [7], [8], [9], but offloading for resource-constrained IoT devices has received insufficient attention. Effective offloading strategies can lower latency, enhance service quality, and reduce reliance on centralized cloud systems. Optimizing MEC resources involves managing limited server capacity to minimize execution delays and improve user experience. Fair allocation mechanisms are essential due to the diverse interests of IoT users and edge servers. The dynamic nature of these systems highlights the need for ongoing research to develop robust resource allocation methods that maximize edge computing's potential and foster innovation.

This study aims to develop a heuristic-based offloading strategy that optimizes task execution time and improves QoS. This paper presents a novel heuristic-based offloading strategy that addresses the specific challenges faced by IoT devices. The paper's structure includes a review of related work (Section II), the system model and problem formulation (Section III), the proposed heuristic-based offloading strategy (Section IV), performance evaluation through simulation results (Section V), and a conclusion with future research perspectives (Section VI).

II. Related Work

The integration of IoT with MEC provides significant benefits, such as ultra-low latency, real-time data analytics, improved resource management, increased capacity, and enhanced scalability. By localizing computing capabilities, MEC reduces bandwidth needs and reliance on central Clouds, minimizing data transfers to remote data centers.

M. Myyara, O. Lagnfdi, A. Darif, and A. Farchane are affiliated with the Laboratory of Innovation in Mathematics, Applications, and Information Technology, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, 23000, Morocco. (E-mails: marouane.myyara@usms.ac.ma, lagnfdi.o@gmail.com, anouar.darif@gmail.com, a.farchane@gmail.com)

Manuscript received April 19, 2005; revised August 26, 2015.

DOI: 10.36244/ICJ.2024.4.2

Researchers have explored collaborative computation offloading and resource allocation schemes to enhance task processing efficiency in MEC systems. MEC facilitates efficient computation, load balancing, and latency reduction by migrating tasks to resource-rich infrastructures [10]. Recent studies include dynamic offloading algorithms that optimize user experience by minimizing service time and balancing workloads [11], and a deadline-aware scheduling algorithm that reduces execution time for critical tasks while considering task type and weight [12]. While many studies focus on optimizing task processing time, they often neglect the broader concept of service time, which encompasses both processing and transmission delays. This paper advances the field by introducing a novel algorithm that incorporates each task's deadline and latency tolerance to minimize service time while meeting QoS requirements.

Several research efforts have focused on optimizing computation offloading in MEC to enhance IoT device performance. One approach is the Lagrange duality resource optimization algorithm [13], which improves task offloading and resource allocation compared to traditional methods like random offloading and load balancing. This highlights the importance of efficient processing for real-time IoT applications, addressing service time and QoS requirements. A notable study [14] presents a collaborative computing framework that enables devices to partially process tasks across terminals, edge servers, and the Cloud using a pipeline-based offloading scheme. Additionally, various models have been developed to reduce latency and improve system efficiency, including an algorithm specifically designed to minimize execution latency [15].

Effective joint resource management between MEC and the central Cloud is crucial for meeting the service demands of IoT applications, particularly given the limited capabilities of edge devices compared to Cloud infrastructures [16]. Recent research has focused on MEC network workload orchestration and resource allocation strategies to improve IoT application performance [17]. For instance, [18] explores computation offloading and bandwidth distribution in IoT networks using graph-based models for resource optimization. Heuristic methods, such as the iterative heuristic mobile edge computing resource allocation algorithm [19], aim to enhance efficiency and minimize latency.

Despite advancements in MEC and IoT integration, challenges specific to resource-constrained IoT devices remain unaddressed. Our primary objective is to minimize task execution time while considering computing resource constraints and application requirements. Unlike prior studies, our research focuses on optimizing QoS for end IoT devices within an MEC framework. We propose a heuristic-based strategy for offloading and resource allocation that adheres to constraints while maximizing task execution efficiency, reducing latency, and effectively utilizing computing resources.

III. System Model and Problem Formulation

A. System Model

The MEC system model, illustrated in Figure 1, features a three-tier architecture: central cloud, MEC servers, and wirelessly connected IoT devices. IoT devices can offload resource-intensive computations to MEC servers or the cloud for efficient processing. Each MEC server is linked to a wireless access point or base station, covering a specific area and serving IoT users. These servers, equipped with sufficient hardware resources, connect to the Edge Orchestrator (EO) via a backhaul link, which manages infrastructure resources, server states, and capacities. Positioned close to users, the servers connect to the EO through a MAN. The central cloud consists of high-capacity servers provided by cloud service providers, accessible via a WAN network linking the EO to the upper layer. The offloading process begins with the device and is guided by the EO, considering workload distribution, computing resources, and network conditions.

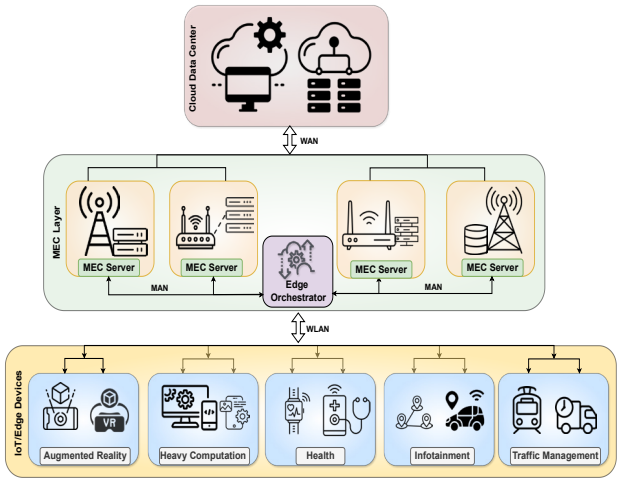


Fig. 1: Multi-layer Multi-access Edge Computing architecture.

B. Notation and Variables

Let $\mathcal{M}$ represent the set of MEC nodes, $\mathcal{C}$ denote the Central Cloud, and $\mathcal{D}$ the set of IoT devices, with IoT_i indicating device i. Each MEC node $m \in \mathcal{M}$ is associated with a set of Virtual Machines (VMs). The computing capacity of each VM, denoted by $\mathcal{F}$, is measured in MIPS (Millions of Instructions Per Second). Each IoT device i has one or more computation tasks, represented by $\mathcal{T}_i$, with each task $\tau_{i,j}$ characterized by a length $\mathcal{L}_{i,j}$, indicating the data generated for the task. The computational capacity required for task j from device i is $C_{i,j}$. These parameters are defined in an XML file in the EdgeCloudSim simulator [20], allowing for customization based on application characteristics.

C. Computation Offloading Model

In a multi-layer MEC environment, computation tasks can be executed locally on the IoT device, offloaded to MEC servers, or offloaded to central Cloud servers. The computation task offloading involves two sets of optimization variables:

- Binary Variable $\mathcal{A}$: This variable is defined as $\mathcal{A} = \{\alpha_{i,j} \mid i \in \mathcal{D}, j \in \mathcal{T}_i\}$. Here, $\alpha_{i,j}$ takes a value of 0 if task data $\tau_{i,j}$ is offloaded, and 1 if it is executed locally.

- Binary Variable $\mathcal{B}$: This variable is defined as $\mathcal{B} = \{\beta_{i,j} \mid i \in \mathcal{D}, j \in \mathcal{T}_i\}$. Here, $\beta_{i,j}$ equals 1 if the task is offloaded to an MEC server, and 0 if it is offloaded to the Cloud.

D. Task Computation Model

In this section, we focus on the task computation models in the MEC architecture, providing estimations for determining local processing times and remote processing times.

1) Local Computation: Local computation processes tasks directly on the user's IoT device, resulting in low latency since data does not need to be transferred to a remote server. However, if the task size is large or the device's computing capacity is limited, offloading may be necessary. Let $\mathcal{F}_{IoT_i}$ denote the computing power of the IoT device i in MIPS. The overall service time for local computation, denoted as $T^{IoT}_{\tau_{i,j}}$, can be calculated as:

$$T^{IoT}_{\tau_{i,j}} = \frac{\mathcal{L}_{i,j}}{\mathcal{F}_{IoT_i}} \tag{1}$$

The local computation delay depends on task size and computing power. Larger tasks or weaker computing power result in longer delays. Local computation has no transmission delays as data is processed on the user's IoT device.

2) Computation in MEC Servers: Computation in MEC servers leverages proximity to end-users to achieve low latency. Offloading tasks to nearby MEC servers typically reduces service time compared to local computation. However, task execution on MEC servers incurs delays due to wireless transmission from IoT devices. The total execution time on an MEC server includes both transmission and computation delays. Let $\mathcal{F}_{\text{MEC}_m}$ denote the computing power of the m-th MEC server. The execution time of task $\tau_{i,j}$ on MEC server m can be expressed as:

$$T^{m}_{\text{Com}(i,j)} = \frac{\mathcal{L}_{i,j}}{\mathcal{F}_{\text{MEC}_m}} \tag{2}$$

The total delay, considering both transmission and computation times, for transmitting the input data and receiving the computation results from the IoT layer to the MEC server through the wireless channel can be expressed as:

$$T^{\text{MEC}_m}_{\tau_{i,j}} = T^{m}_{\text{Com}(i,j)} + T^{m}_{\text{Up}(i,j)} + T^{m}_{\text{Dw}(i,j)} \tag{3}$$

where $T^{m}_{\text{Com}(i,j)}$ represents the computation time for task j on MEC server m, $T^{m}_{\text{Up}(i,j)}$ represents the upload data transfer time, and $T^{m}_{\text{Dw}(i,j)}$ represents the download data transfer time from the MEC server back to the IoT device.

3) Computation in Cloud Servers: Computation in Cloud servers offers high processing power and storage capacity but incurs higher latency due to the distance of data transfer. When MEC servers cannot process offloaded tasks promptly, they are sent to the Cloud server over the wireless network. The total delay in Cloud server computation consists of transmission and processing delays.

The computation time in the Cloud is given by:

$$T^{cloud}_{\text{Com}(i,j)} = \frac{\mathcal{L}_{i,j}}{\mathcal{F}_{\text{Cloud}}} \tag{4}$$

Where $\mathcal{F}_{\text{Cloud}}$ represents the computing power of the Cloud server in MIPS. Similar to the MEC server case, there is a transmission delay for uploading and downloading data, denoted as $T^{cloud}_{\text{Up}(i,j)}$ and $T^{cloud}_{\text{Dw}(i,j)}$ respectively. The total delay for offloading a task to the Cloud server is the sum of the computation time and the transmission delay:

$$T^{\text{Cloud}}_{\tau_{i,j}} = T^{cloud}_{\text{Com}(i,j)} + T^{cloud}_{\text{Up}(i,j)} + T^{cloud}_{\text{Dw}(i,j)} \tag{5}$$

Figure 2 illustrates task computation models for local computation, MEC servers, and Cloud servers, indicating task offloading and allocation in the MEC architecture. It's important to note that transmission delay in the Cloud server case is typically longer due to distance and potential network congestion.

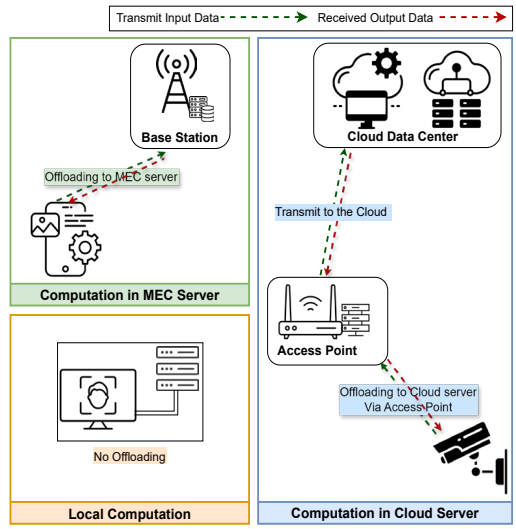


Fig. 2: An illustration of Computation Offloading type in MEC networks

E. Problem Formulation

The objective of this study is to minimize the execution time of computing tasks in an MEC system by optimizing task offloading and resource allocation. We consider application constraints, available computing capacities, and resources. Our problem is formulated as a minimization problem with an objective function that incorporates task constraints, resource capacities, and offloading decision variables:

$$\text{Minimize} \quad \mathcal{P} = \sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{T}_i} \sum_{m=1}^{\mathcal{M}} \alpha_{i,j} T^{IoT}_{\tau_{i,j}} + (1-\alpha_{i,j}) \times \left(\beta_{i,j} T^{\text{MEC}_m}_{\tau_{i,j}} + (1-\beta_{i,j}) T^{\text{Cloud}}_{\tau_{i,j}}\right) \tag{6}$$

Subject to:

$$\alpha_{i,j}, \beta_{i,j} \in \{0,1\} \quad , \forall i \in \mathcal{D}, \forall j \in \mathcal{T}_i \tag{6a}$$

$$\sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{T}_i} (1-\alpha_{i,j}) \times \beta_{i,j} C_{i,j} \leq \sum_{m=1}^{\mathcal{M}} \mathcal{F}_{\text{MEC}_m} \tag{6b}$$

$$\sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{T}_i} (1-\alpha_{i,j})(1-\beta_{i,j}) C_{i,j} \leq \mathcal{F}_{\text{Cloud}} \tag{6c}$$

$$\sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{T}_i} \alpha_{i,j} C_{i,j} \leq \sum_{i=1}^{\mathcal{D}} \mathcal{F}_{IoT_i} \tag{6d}$$

$$\sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{T}_i} \sum_{m=1}^{\mathcal{M}} (1-\alpha_{i,j}) \beta_{i,j} \mathcal{F}_{\text{MEC}_m} \leq \sum_{m=1}^{\mathcal{M}} Th_{\text{MEC}_m} \tag{6e}$$

The problem is subject to the following constraints: Constraint (6a) ensures that the decision variables $(\alpha_{i,j}, \beta_{i,j})$ are binary. Constraint (6b) guarantees that the server's computing capacity is sufficient if a task is offloaded to an MEC server. Constraint (6c) verifies that the capacity is adequate for tasks offloaded to the cloud server. Constraint (6d) ensures the device's computing capacity is sufficient for local execution. Finally, constraint (6e) restricts MEC server resource utilization to a specified threshold Th_{MEC_m}, typically set around 80%.

By addressing this optimization challenge, we can identify effective task offloading and resource allocation strategies that minimize total execution time while adhering to MEC system constraints. This problem can be formulated as an Integer Linear Programming (ILP) model, which provides optimal solutions [21] but faces high computational complexity, complicating real-time implementation for large-scale applications. To overcome this, heuristic strategies have been developed, balancing solution quality with computational efficiency. These methods enable real-time offloading decisions and resource optimization associated with ILP, ensuring system agility in managing dynamic task arrivals.

IV. HEURISTIC-BASED OFFLOADING STRATEGY FOR QOS IMPROVEMENT

We propose a **H**euristic-based **O**ffloading strategy to enhance **QoS** (HOQoS) in MEC environments. Our goal is to optimize offloading decisions for computing tasks while considering time constraints, resource limitations, and application requirements. Utilizing greedy heuristics, we explore the solution space to make optimal decisions for each task, factoring in predicted execution times across different architecture layers. By assessing the impact of offloading on overall service time, we aim for significant QoS improvement.

A key component of our strategy is the Offloading Decision Variable Identification (ODVI) algorithm (Algorithm 1), which determines the best offloading decision $(\alpha_{i,j}, \beta_{i,j})$ for each incoming task $\tau_{i,j}$. The algorithm initializes the minimum time as infinite and the decision variables as zero. For each task, it calculates the local execution time $T^{IoT}_{\tau_{i,j}}$ and estimates delays on MEC servers $T^{\text{MEC}_m}_{\tau_{i,j}}$ and in the Cloud $T^{\text{Cloud}}_{\tau_{i,j}}$. The task is assigned to the server with the shortest delay $T_{\min}$, while respecting the maximum acceptable delay constraint. The decision variables $(\alpha_{i,j}, \beta_{i,j})$ are then based on the chosen optimal location, focusing on individual tasks rather than overall execution time.

Algorithm 1 Offloading Decision Variable Identification (ODVI)

Require: $(\mathcal{M}, \mathcal{C}, \mathcal{D}, \mathcal{T}_i)$
Ensure: $(\alpha_{i,j}, \beta_{i,j})$ Offloading decision variables

$T_{\text{Min}(i,j)} \leftarrow \infty$
Decision variables $(\alpha_{i,j}, \beta_{i,j}) \leftarrow \emptyset$
for $i \in \mathcal{D}$ **do**
  for $j \in \mathcal{T}_i$ **do**
    for $m \in \mathcal{M}$ **do**
      $T_{\text{Min}(i,j)} \leftarrow \min(T_{\text{Min}(i,j)}, T^{IoT}_{\tau_{i,j}}, T^{\text{MEC}_m}_{\tau_{i,j}}, T^{\text{Cloud}}_{\tau_{i,j}})$
    end for
    if $T_{\text{Min}(i,j)} = T^{IoT}_{\tau_{i,j}}$ **then**
      $\alpha_{i,j} \leftarrow 1$
    else
      if $T_{\text{Min}(i,j)} = T^{\text{Cloud}}_{\tau_{i,j}}$ **then**
        $\alpha_{i,j} \leftarrow 0$
        $\beta_{i,j} \leftarrow 0$
      else
        $\alpha_{i,j} \leftarrow 0$
        $\beta_{i,j} \leftarrow 1$
      end if
    end if
  end for
end for
return $(\alpha_{i,j}, \beta_{i,j})$

The ODVI algorithm, while effective for individual task offloading, has limitations in global optimization and resource contention management. Its focus on single tasks can lead to suboptimal overall system performance and does not address resource contention among multiple tasks. To overcome these issues, we introduce the Virtual Machine Selection for Execution (VMSE) algorithm (Algorithm 2).

VMSE tackles the global optimization problem by considering overall system resource allocation when selecting a virtual machine for each incoming task $\tau_{i,j}$. It implicitly manages resource contention by prioritizing VMs with lower utilization. The algorithm evaluates the characteristics of incoming tasks against the current utilization of MEC servers and the Cloud, selecting the least utilized VM that meets the task's performance, resource, and time constraints. This integration significantly enhances QoS through efficient resource utilization and improved task execution efficiency.

To optimize task offloading and resource allocation, we developed the HOQoS algorithm (Algorithm 3), which integrates the features of the ODVI and VMSE algorithms. First, the ODVI algorithm evaluates available execution environments by considering MEC server performance and network latency, generating decision variable pairs based on task characteristics such as size, complexity, and time constraints. These pairs

Algorithm 2 Virtual Machine Selection for Execution (VMSE)

Require: Incoming task $\tau_{i,j}$, Node($R_{opt}[\tau_{i,j}]$)
Ensure: VM satisfying task's ($\tau_{i,j}$) resource requirements.

```
task ← τ_{i,j}
selectedVM ← null
bestUtili ← ∞
for host ∈ getInfrastructureHostList(Node(R_opt[τ_{i,j}])) do
    vmList ← getVmList(host)
    for vm ∈ vmList do
        if task.Utilization() ≤ vm.getCapacity() then
            currentUtili ← vm.getCurrentUtilization()
            if currentUtili < bestUtili then
                selectedVM ← vm
                bestUtili ← currentUtili
            else if currentUtili == bestUtili then
                selectedVM ← SelectVM(selectedVM,vm)
            end if
        end if
    end for
end for
return selectedVM
```

represent potential execution options for each task. Next, the VMSE algorithm selects the optimal execution location by assessing task characteristics and resource availability on MEC servers and in the Cloud. It evaluates each virtual machine's performance and chooses the one that best meets the task's requirements regarding performance, resource availability, and time constraints.

Algorithm 3 Heuristic-based offloading for improving quality of service (HOQoS)

Require: $((\alpha_{i,j}, \beta_{i,j}), \mathcal{M}, \mathcal{C}, \mathcal{D}, \mathcal{T}_{i,j})$
Ensure: Optimal Resource Allocation $R_{\text{opt}}[\tau_{i,j}]$ and selected VM $VM_{\text{selected}}[\tau_{i,j}]$

```
R_opt[τ_{i,j}] ← null
VM_selected[τ_{i,j}] ← null
for i ∈ 𝒟 do
    for j ∈ 𝒯_i do
        (α_{i,j}, β_{i,j}) ← Call Algorithm 1
        if α_{i,j} = 1 then                    ▷ Local allocation
            R_opt[τ_{i,j}] ← LocalNode
        else if β_{i,j} = 1 then               ▷ MEC allocation
            R_opt[τ_{i,j}] ← MECNode
        else                                   ▷ Cloud allocation
            R_opt[τ_{i,j}] ← CloudNode
        end if
        VM_selected[τ_{i,j}] ← Call Algorithm 2
    end for
end for
return R_opt and VM_selected[τ_{i,j}]
```

The HOQoS algorithm employs a sequential task processing strategy, handling tasks individually rather than in batches. This approach is crucial for achieving system agility and effectively managing dynamic task arrivals. By adapting resource allocation in real time, the algorithm ensures a rapid and optimized response, even with irregular or unpredictable task arrival rates. Thus, sequential processing is integral to the algorithm's ability to efficiently manage dynamic environments.

V. Performance Evaluation

In the performance evaluation section, we conducted simulations using the EdgeCloudSim simulator [20]. We analyze the performance of our MEC architecture using various measurement and evaluation methods.

A. Simulation Parameters

The simulation parameters used for our evaluations are summarized in Table I. The simulations are implemented in Java, and the generated plots are visualized using Matlab. The experiments are conducted on a computer with 4 Intel Core i7-9600U processors clocked at 2.59 GHz and 8 GB of RAM.

TABLE I
Simulation Parameters.

Parameter	IoT	MEC	Cloud
Number of devices	100 - 2,300	14	1
Number of hosts	1	1	1
Number of VMs per host	1	8	4
Number of Cores per VM	1	2	4
VM CPU Speed (in MIPS)	4,000	10,000	100,000

EdgeCloudSim utilizes four distinct application types to realistically simulate diverse scenarios, as detailed in [20]. Table II summarizes the characteristics of each application type: Heavy applications are characterized by high data transmission, high computational intensity, and low sensitivity to delays. Infotainment applications generally require moderate data transmission, high computational intensity, and low sensitivity to delays. AR/VR applications involve high data transmission volumes, moderate computational intensity, and high sensitivity to delays. Health applications typically feature moderate data transmission, moderate computational intensity, and moderate sensitivity to delays. Each application is assessed based on several criteria, including usage percentage, which indicates the share of each application in overall usage, and Cloud selection probability, which demonstrates the tendency to use the Cloud for their operation. The volumes of data uploaded and downloaded, as well as the task length, emphasize the requirements for bandwidth and processing.

TABLE II
Application Characteristics.

Characteristics	Heavy	Infotainment	AR/VR	Health
Task Length (GI)	45	15	9	3
Delay Sensitivity	0.1	0.3	0.9	0.7
Max. Delay Req. (s)	2	1.5	1	0.5
Data Upload (KB)	2500	25	1500	20
Data Download (KB)	200	1000	25	1250

In addition to the characteristics of the applications presented in Table II, it is important to specify the technical

parameters of the simulated environment. The bandwidth of the wireless local area network (WLAN) is set at 200 Mbps, while the bandwidth of the wide area network (WAN) is 15 Mbps. The propagation delay on the WAN is 0.1 seconds. These parameters also influence system performance and should be taken into account when evaluating the results of the simulation.

B. Simulation Results

This section presents the results of simulations conducted to evaluate the performance of different resource management approaches and computation offloading strategies in a high-density IoT environment. The studied approaches are:

- Only MEC: All tasks are executed on the MEC server. This approach is straightforward but may be limited in terms of resources and processing capacity.
- Only Cloud: All tasks are executed on the Cloud server. This approach offers high processing capacity but can lead to significant latency due to the distance between IoT devices and the cloud.
- Random: Tasks are randomly distributed between the MEC and Cloud servers. This implementation is simple but may not be optimal in terms of performance.
- DCOA-ST: A dynamic computation Offloading approach based on the service time [11]. This approach considers the current system state to decide where to execute tasks, potentially improving performance. It provides a systematic approach to task assignment but may lack flexibility in dynamic environments.
- LCDA*: The LCDA approach [12], which aims to minimize the execution time of delay-sensitive tasks while ensuring no deadline violations. Their algorithm selects servers and schedules tasks to optimize service time in a dynamic environment. While LCDA effectively adapts to changing workloads, it may not fully account for task sensitivity or latency constraints.

We can evaluate their relative advantages and limitations by comparing these methodologies with our HOQoS algorithm. HOQoS integrates factors such as execution time, task sensitivity to delays, latency requirements, resource utilization, and server processing capabilities. Through rigorous evaluations and performance comparisons, we illustrate the enhanced effectiveness of HOQoS in optimizing computation offloading and resource allocation within MEC environments, ultimately improving the quality of service for IoT applications.

1) Simulation Based on Average Service Time: We evaluate the average service time, a critical factor influencing service quality and user experience. Figure 3 illustrates the impact of the number of IoT devices on service times for different resource management approaches. For 100 devices, HOQoS and LCDA* show shallow service times of 0.8 seconds, while Only MEC and DCOA-ST display times of 0.9 seconds. At 900 devices, HOQoS maintains a service time of 1 second, while Only Cloud shows an alarming increase to 5.1 seconds, confirming its inadequacy in high-load scenarios. Other approaches, such as Random and DCOA-ST, exhibit times of 2.1 seconds and 1.15 seconds, respectively. For 1500 devices, HOQoS slightly increases to 1.2 seconds, while Only Cloud reaches 7.1 seconds. Finally, at 2300 devices, HOQoS reports a service time of 1.5 seconds, whereas Only Cloud remains high at 7.2 seconds.

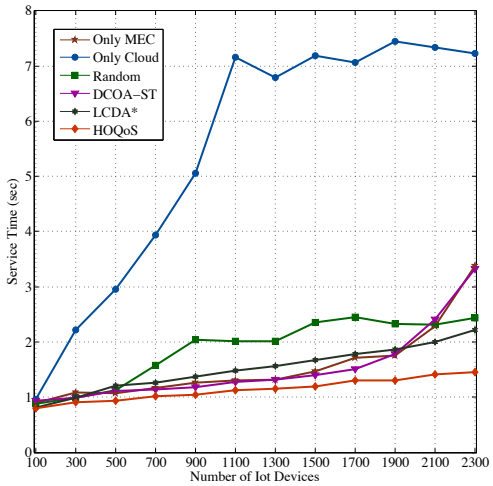


Fig. 3: Service time as a function of the number of IoT devices.

2) Simulation Based on Task Failure Rates: In assessing task execution success, task failure rates are key indicators influenced by factors such as excessive virtual machine usage and insufficient network bandwidth. Figure 4 illustrates how failure rates vary across algorithms with different numbers of IoT devices. The HOQoS approach consistently maintains low failure rates, even as device numbers increase, highlighting its robustness in high-density IoT environments. For 300 devices, HOQoS and Only MEC have low failure rates of 1%, while LCDA* has the highest at 10%. At 900 devices, Only Cloud shows a concerning 74% failure rate, whereas HOQoS maintains a reasonable 1.5%. With 1500 devices, Only Cloud's failure rate rises to 91%, confirming its inadequacy in high-load scenarios, while HOQoS remains at 2%. Finally, at 2300

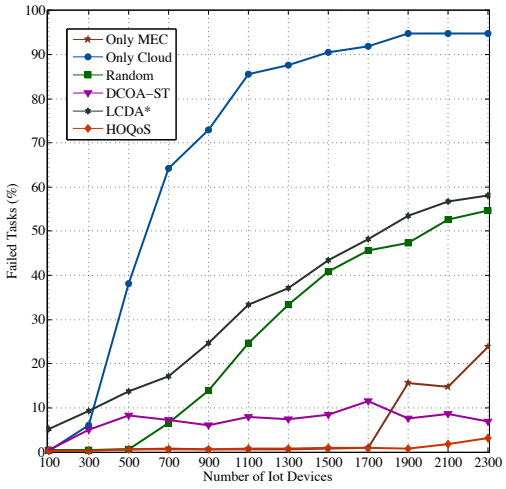


Fig. 4: Task failure rate as a function of the number of IoT devices.

devices, all failure rates increase, with Only Cloud reaching a critical 95% and HOQoS at 4%.

3) Simulation Based on VM Utilization Rates: Figure 5 highlights the impact of the number of IoT devices on the utilization of VMs at the MEC level. For 300 devices, HOQoS shows a utilization rate of 1%, while Only MEC reaches 3%. Other approaches, such as Random and DCOA-ST, exhibit utilization rates of 2% and 4.5%, respectively, while LCDA* reaches 5%. At 1100 devices, HOQoS maintains a rate of 2.5%, while Only MEC climbs to 12%. The Random and DCOA-ST approaches also show increases, reaching 5% and 13%, respectively. When the number of devices reaches 1700, HOQoS increases to 5%, while Only MEC rises to 23%, with LCDA* at 19% and DCOA-ST at 23%. Finally, at 2300 devices, HOQoS presents a utilization rate of 9.8%, contrasting sharply with Only MEC, which reaches 58%. Other approaches show increased utilization, with Random at 12% and DCOA-ST at 27%.

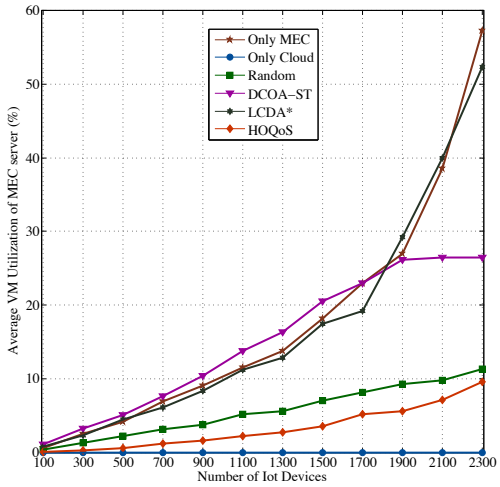


Fig. 5: MEC server utilization as a function of the number of IoT devices.

The utilization of VM at the Cloud level is presented in Figure 6, exhibiting slightly lower rates compared to MEC servers. For 500 devices, HOQoS shows a VM utilization rate of only 0.1%, while Only Cloud reaches 1.2%. Other approaches, such as Random and DCOA-ST, exhibit utilization rates of 0.8% and 1.2%, respectively, while LCDA* reaches 0.4%. At 1300 devices, HOQoS maintains a utilization rate of 0.1%, while Only Cloud climbs to 1.1%. The Random and DCOA-ST approaches also show increases, reaching 1% and 3.7%, respectively. When the number of devices reaches 1900, HOQoS increases to 0.2%, while Only Cloud rises to 1.1%, with LCDA* at 2.7% and DCOA-ST at 5.8%. Finally, at 2300 devices, HOQoS presents a utilization rate of 0.3%, contrasting sharply with Only Cloud, which reaches 1.4%. Other approaches show increasing utilization, with Random at 1% and DCOA-ST at 11.9%.

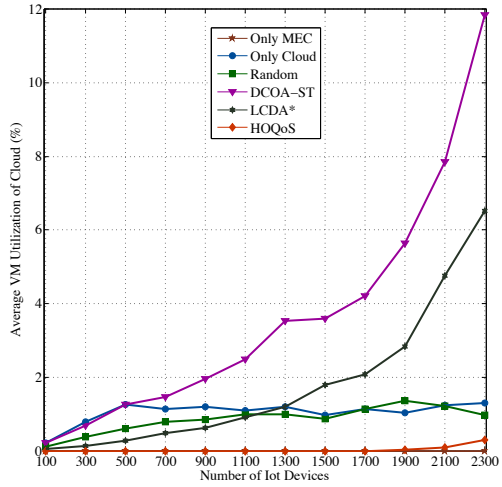


Fig. 6: Cloud server utilization as a function of the number of IoT devices.

The results demonstrate that HOQoS effectively manages both MEC and cloud server resources under increasing load. In contrast, other approaches, particularly Only MEC and Only Cloud, exhibit a significant rise in VM utilization, which could lead to challenges concerning performance and energy consumption. It is important to note that Only Cloud does not utilize MEC servers, while Only MEC does not utilize cloud servers; therefore, their VM usage is not considered in this analysis.

VI. Conclusion and Future Perspectives

This study aimed to improve service quality in Multi-access Edge Computing by offloading computational loads and allocating resources near IoT devices. The findings emphasize the importance of offloading tasks from IoT devices with limited computing capabilities to locations with adequate resources, thereby reducing latency. The proposed heuristic algorithms for task offloading and resource allocation consider task characteristics and resource availability, resulting in decreased service times and task failure rates. Future research will focus on developing an autonomous system that utilizes reinforcement learning and machine learning techniques to optimize task execution locations in complex MEC environments. Enhancing the algorithms' adaptive capabilities to respond dynamically to changes in network conditions and device capabilities will be a priority. Additionally, integrating advanced predictive analytics could enable proactive optimization of resource allocation and offloading strategies by forecasting future resource needs and user demands. These advancements aim to further enhance the performance and efficiency of MEC systems.

References

[1] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke," Middle-ware for Internet of Things: A Survey," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 70–95, Feb. 2016.
DOI: 10.1109/JIOT.2015.2498900.

[2] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on Multi-Access Edge Computing for Internet of Things Realization," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2961–2991, 2018. **DOI**: 10.1109/COMST.2018.2849509.

[3] European Telecommunications Standards Institute (ETSI), "Mobile-edge computing (MEC); Framework and reference architecture," ETSI GS MEC 003, 2019.

[4] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal, et al., "Mobile-edge computing introductory technical white paper," *White paper, mobile-edge computing (MEC) industry initiative*, vol. 29, pp. 854– 864, 2014.

[5] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017. **DOI**: 10.1109/COMST.2017.2705720.

[6] A. P. Jafari Pozveh and H. S. Shahhoseini, "IoT Integration with MEC," in *Mobile Edge Computing, Springer, Cham*, 2021, pp. 111–144. **DOI**: 10.1007/978-3-030-69893-5_6.

[7] M. Myyara, O. Lagnfdi, A. Darif, and A. Farchane, "A New Approach Based on Genetic Algorithm for Computation Offloading Optimization in Multi-Access Edge Computing Networks," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 4, pp. 4186–4194, 2024. **DOI**: 10.11591/ijai.v13.i4.pp4186-4194.

[8] C. Shu, Z. Zhao, Y. Han, G. Min, and H. Duan, "Multi-User Offloading for Edge Computing Networks: A Dependency-Aware and Latency-Optimal Approach," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1678–1689, Mar. 2020. **DOI**: 10.1109/JIOT.2019.2943373.

[9] C. Chen, B. Liu, S. Wan, P. Qiao, and Q. Pei, "An Edge Traffic Flow Detection Scheme Based on Deep Learning in an Intelligent Transportation System," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1840–1852, Mar. 2021. **DOI**: 10.1109/TITS.2020.3025687.

[10] A. Naouri, H. Wu, N. A. Nouri, S. Dhelim, and H. Ning, "A novel framework for mobile-edge computing by optimizing task offloading," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 13 065–13 076, 2021. **DOI**: 10.1109/JIOT.2021.3064225.

[11] M. Myyara, O. Lagnfdi, A. Darif, and A. Farchane, "Quality of Expe- rience Improvement and Service Time Optimization through Dynamic Computation Offloading Algorithms in Multi-access Edge Computing Networks," *International Journal of Computer Network and Information Security (IJCNIS)*, vol. 16, no. 4, pp. 1–16, 2024. **DOI**: 10.5815/ijcnis.2024.04.01.

[12] H. Choi, H. Yu, and E. Lee, "Latency-classification-based deadline-aware task offloading algorithm in mobile edge computing environments," *Applied Sciences*, vol. 9, no. 21, p. 4696, 2019. **DOI**: 10.3390/app9214696.

[13] C. Eang, S. Ros, S. Kang, I. Song, P. Tam, S. Math, and S. Kim, "Offloading Decision and Resource Allocation in Mobile Edge Computing for Cost and Latency Efficiencies in Real-Time IoT," *Electronics*, vol. 13, no. 7, p. 1218, 2024. **DOI**: 10.3390/electronics13071218.

[14] C. Kai, H. Zhou, Y. Yi, and W. Huang, "Collaborative cloud-edge-end task offloading in mobile-edge computing networks with limited communication capability," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 624–634, 2020. **DOI**: 10.1109/TCCN.2020.3018159.

[15] S. Wan, R. Gu, T. Umer, K. Salah, and X. Xu, "Toward Offloading Internet of Vehicles Applications in 5G Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4151–4159, Jul. 2021. **DOI**: 10.1109/TITS.2020.3017596.

[16] I. Sarrigiannis, K. Ramantas, E. Kartsakli, P.-V. Mekikis, A. Antonopoulos, and C. Verikoukis, "Online VNF Lifecycle Management in an MEC-Enabled 5G IoT Architecture," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4183–4194, May 2020. **DOI**: 10.1109/JIOT.2019.2944695.

[17] M. Myyara, O. Lagnfdi, A. Darif, and A. Farchane, "A Resource Allocation Strategy to Enhance User Experience for IoT Devices in Multi-access Edge Computing," in *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)*, 2024, pp. 1–7. **DOI**: 10.1109/ICDS62089.2024.10756444.

[18] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint Power and Time Allocation for NOMA–MEC Offloading," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6207–6211, Jun. 2019. **DOI**: 10.1109/TVT.2019.2907253.

[19] Z. Ning, P. Dong, X. Kong, and F. Xia, "A Cooperative Partial Computation Offloading Scheme for Mobile Edge Computing Enabled Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4804–4814, Jun. 2019. **DOI**: 10.1109/JIOT.2018.2868616.

[20] C. Sonmez, A. Ozgovde, and C. Ersoy, "Fuzzy workload orchestration for edge computing," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 769–782, 2019. **DOI**: 10.1109/TNSM.2019.2901346.

[21] P. W. Khan, K. Abbas, H. Shaiba, A. Muthanna, A. Abuarqoub, and M. Khayyat, "Energy efficient computation offloading mechanism in multi- server mobile edge computing—An integer linear optimization approach," *Electronics*, vol. 9, no. 6, p. 1010, 2020. **DOI**: 10.3390/electronics9061010.

Marouane Myyara received his B.Sc. in Electronic and Telecommunication Engineering in 2019 and M.Sc. in Telecommunication Systems and Computer Networks in 2021 from Sultan Moulay Slimane University, Beni Mellal, Morocco. He is currently a Ph.D. candidate at the Laboratory of Innovation in Mathematics, Applications, and Information Technology (LIMATI), Polydisciplinary Faculty, Sultan Moulay Slimane University, Morocco. His current research focuses on improving the performance of Multi-access Edge Computing networks (MEC), Cloud Computing, Computation Offloading, and the Internet of Things (IoT).

Oussama Lagnfdi received his B.Sc. in Physical Matter Science in 2020 and M.Sc. in Telecommunications Systems and Computer Networks in 2022 from Sultan Moulay Slimane University, Beni Mellal, Morocco. Currently, he is a Ph.D. candidate at the Laboratoire d'Innovation en Mathématiques et Applications et Technologies de l'Information (LIMATI), Polydisciplinary Faculty, Sultan Moulay Slimane University, Morocco. His ongoing research is focused on enhancing the performance of Internet of Things (IoT) and Mobile Edge Computing (MEC), Artificial Intelligence, Deep Learning, and Fuzzy Logic.

Anouar Darif received the bachelor in IEEA (Informatique Électrotechnique, Électronique and Automatique) from Dhar El Mahraz Faculty of Sciences at Mohamed Ben Abdellah University Fez, Morocco in 2005. He received the Diplôme d'Etudes Supérieures Approfondies in Computer Sciences and Telecommunications from the Faculty of Sciences Rabat in 2007. He received the Ph.D. degree in Computer Sciences and Telecommunications from the Faculty of Sciences of Rabat in 2015. He is currently a Research and Teaching Associate in the Multidisciplinary Faculty at the University of Sultan Moulay Slimane Beni Mellal, Morocco. His research interests include Wireless Sensor Networks (WSN), Mobile Edge Computing (MEC), Internet of Things (IoT), Cloud Computing, and Neural Networks.

Abderrazak Farchane received his B.Sc. in Computer Science and Engineering in June 2001 and M.Sc. in Computer Science and Telecommunication from the University of Mohammed V Agdal, Rabat, Morocco, in 2003. He obtained his Ph.D. in Computer Science and Engineering at ENSIAS, Rabat, Morocco. He is currently an Associate Professor of Computer Science in the Polydisciplinary Faculty, at Sultan Moulay Slimane University, Morocco. His areas of interest are Information Coding Theory, Cryptography, and Security.

Enhancing Signal Processing Efficiency in Software-Defined Radio Using Distributed Arithmetic and Look-Up Table-Based FIR Filters

Hari Krishnan S*, and Mr. S. Sadiqvali

***Abstract*—To meet the requirements of the wireless communication industry, digital communication systems require increasingly advanced coding and modulation technologies. Software-Defined Radio enables these advanced ideas to be easily adopted by such systems. The Finite Impulse Response filter is frequently used in wireless communication to pre-process detected signals to reduce noise by utilizing delay elements, multipliers, and adders. Traditional multiplier-based finite impulse response filter designs result in hardware-intensive multipliers that use a lot of space and energy and pose poor calculation speeds and low performance in throughput and latency. To overcome the existing issues, a novel Distributed Arithmetic with a Look Up Table-based FIR filter is proposed, which reduces the Bit Error Rate and latency and improves throughput by optimizing the channel equalizer as a crucial part of Software Defined Radio applications. Further, a key feature named the decimation factor is incorporated to dynamically alter the filter's output frequency response without altering the filter coefficients. Moreover, the worst-case critical route latency of partial product accumulation is reduced using a highly adaptable Parallel Prefix Adder. Additionally, the finite impulse response filters are integrated to decrease the number of Look-Up Tables, thereby saving time and memory. It also investigates the filter efficiency using faster multipliers and adders and validates it on an Artix-7 FPGA. As a result, the proposed model improved the filter's performance over the other existing designs by achieving an operating speed of 260 MHz, delay of 190 ps, power dissipation of 1 mW and throughput of 938.12 Mbps with the number of Look-Up Tables being 16504.**

***Index Terms*—Software-Defined Radio, Noise Removal, Finite Impulse Response, Digital Signal Processing, Field-Programmable Gate Array, and Wireless Communication.**

I. INTRODUCTION

Interference with symbols was one of the biggest problems with the digital structure. This suggests that in digital communication systems such as Software Defined Radio (SDR) applications, SDR technology offers the flexibility to implement adaptive filtering and real-time signal processing, allowing for more effective interference management. This capability enhances overall communication reliability, making SDR a valuable tool in modern digital systems.

*Hari Krishnan S / Sanskrithi School of Engineering, Behind Sri Sathya Sai Super Speciality Hospital, Puttaparthi, Sathya Sai District, Andhra Pradesh 515134 (e-mail: hariprofsse@gmail.com)

Mr. S. Sadiqvali / Sanskrithi School of Engineering, Behind Sri Sathya Sai Super Speciality Hospital, Puttaparthi, Sathya Sai District, Andhra Pradesh 515134 (e-mail: syedb6933@gmail.com)

DOI: 10.36244/ICJ.2024.4.3

However, interference with symbols can lead to distorted channels, which may result in errors in data transmission. Interference can occur due to various factors such as noise, signal degradation, or electromagnetic interference. Distorted channels can result in errors in data transmission, making it essential to employ techniques to mitigate interference and restore the integrity of the transmitted symbols [1-3]. As the speed of data transfer systems goes up, Digital Signal Processing (DSP) needs high-speed communication systems. In DSP computers, the speed couldn't go beyond 1 GHz and it plays a crucial role in enhancing the quality of communication by processing and analyzing signals in real-time. The limitation mentioned(1 GHz) likely pertains to a specific context or technology, as modern DSP systems can operate at much higher frequencies, depending on the application [4]. To support multiple bit rates in communication systems, it's necessary to develop new designs. Different bit rates require different modulation schemes, coding techniques, and signal processing methods. Adapting to multiple-bit rates allows for more flexible and efficient data transmission in modern communication systems [5]. Pipelining and parallel processing are techniques used to improve the speed and efficiency of computers in optical transmission systems. Pipelining involves breaking down tasks into stages, allowing for parallel execution of different stages simultaneously. Parallel processing utilizes multiple processing units to perform tasks in parallel, increasing overall system throughput. These techniques are essential for handling the high data rates in optical communication systems [6].

Equalizers are devices or algorithms used on the receiving end of a communication system to reduce distortions like Inter-Symbol Interference (ISI). ISI occurs when symbols interfere with each other in a digital communication signal due to channel characteristics. Equalizers help in recovering the original symbols by compensating for the distortion caused by ISI, thus improving the overall data reception quality [7]. Equalizing Decision Feedback (EDF) is a nonlinear equalization technique used to mitigate ISI in communication systems. It works by making decisions about received symbols and then using these decisions to feedback information to the equalizer to compensate for post-cursor ISI. By doing so, EDF helps in reducing errors caused by

ISI and, consequently, improves the SNR. This technique is particularly effective in scenarios where ISI is a significant challenge.

Maximum Likelihood Sequence Detection (MLSD) is another technique used for ISI mitigation [8]. It considers all possible symbol sequences and selects the one with the highest likelihood, thereby reducing errors caused by ISI. MLSD is especially effective in situations with severe ISI or complex modulation schemes. In a digital communication signal, post-cursor ISI that is, ISI that appears after the primary symbol is eliminated in part by both MLSD and EDF. Through their efficient handling of post-cursor ISI, these methods help to enhance the quality of signals that are received. Combining the various access strategies is probably going to produce the greatest results in terms of enhancing security, increasing data transfer speed, and reducing ISI [9]. These techniques can also address problems related to the way noise spreads, especially in the presence of spectral nulls. Spectral nulls are frequencies where the signal has little or no energy, making them susceptible to noise interference. EDF and MLSD can help mitigate the impact of noise, including noise related to spectral nulls, thus improving overall signal quality [10]. In addition to mitigating ISI and improving SNR, Speed- testing techniques are used. It could refer to the computational complexity and processing speed required for implementing EDF and MLSD algorithms. These algorithms can be computationally intensive, so optimizing their speed is crucial for real-time applications [11-12]. The improvement of Speed and Design Technology is necessary to keep up with the demands of modern high-speed communication systems, where fast and efficient equalization is crucial for reliable data transmission [13].

Distributed Arithmetic (DA) architecture is becoming more popular in DSP. This architecture is chosen due to its simplicity in design and its utilization of Look-Up Tables (LUTs) and transfer build blocks for obtaining partial products. DA architecture offers advantages in terms of efficient hardware implementation, making it a suitable choice for various DSP applications [14]. DA architecture uses two-way binary code complements or offsets for representing the filter coefficients and input values. Binary code complements involve representing numbers as positive and negative complements, which simplifies arithmetic operations. Using offsets can also simplify operations by shifting the input values within a certain range. These techniques contribute to the efficiency and simplicity of DA-based DSP algorithms [15]. To reduce the amount of memory required by a DA Finite Impulse Response (FIR) filter, various strategies are suggested. Memory Divisions refer to dividing the memory resources into smaller blocks or segments to reduce the overall memory footprint, which can help optimize memory usage while still performing the filtering operations efficiently. In addition, Different Memory Bank Approaches are utilized with varying access speeds or capacities to efficiently store and access data. This approach can be particularly useful when dealing with large datasets and complex FIR filter structures [16-17].

The Look-Up Table (LUT) decomposition scheme is used to simplify the LUT structures in FIR filters based on the DA architecture. However, it may come at the cost of using a few extra filters. The trade-off between complexity and resource usage is common in DSP design. Over the past few decades, there have been efforts to improve the performance and efficiency of filters, including those based on DA architecture. Advances in filter design, algorithm optimization, and hardware capabilities have led to more efficient DSP solutions. These improvements help narrow the field of DA architecture as a powerful and viable design choice for various filter applications [18-19]. The DA architecture is particularly well-suited for FIR filters that are based on Decision Feedback Equalizers (DFEs). DFEs are used to mitigate ISI (Inter-Symbol Interference) in communication systems. The DA architecture can offer an efficient and effective way to implement FIR filters within DFEs, contributing to the overall performance and reliability of communication systems [20]. Hence there is a need to design a novel FIR filter for improving the quality of wireless communication applications.

The Major contributions in this paper are given as follows:

- To design an FIR filter based on DA-LUT multiplier for quicker multiplier and faster adder to improve the speed of filter operation and to create a channel equalizer as part of an SDR application and apply it to FIR for validation.
- To use a decimation factor that dynamically modifies the output frequency response of the filter, a highly adaptive parallel prefix adder (PPA) to lower the worst-case critical route latency and verify the filter efficiency on Artix-7 FPGA.

The content of the paper is organized as section 2 describes the literature survey, section 3 describes the proposed design and its working process, and section 4 discusses the proposed design simulation, performance, and comparative analysis. Finally, section 5 concludes the paper.

II. LITERATURE SURVEY

Kumar et al [21] developed a new architecture for a 2-D block FIR filter by using the DA algorithm, which was renowned for its effective design of the multiply and accumulate block. The DA-LUT has a hardware-based architecture that enables the 2-D FIR filter's architecture to be changed. Additionally, sharing occurs among DA-LUTs at different levels as a result of block processing. In order to simplify the hardware complexity of DA-LUT, a common DA-LUT was created for block inputs. Additionally, the systolic architectures in the suggested design were decreased over the designs that already exist thanks to memory overlapping. By separating the internal block into parallel and small blocks for higher-order 2-D FIR filters, the complexity of DA-LUT was decreased. However, building hardware for DSP applications was more challenging, and it requires specialized knowledge and resources.

Amrita Rai [22] proposed a 4-bit FIR filter used in Digital Signal Processing (DSP) employing completely adiabatic technology (PAL) to decrease all parametric performance and power consumption. The designs of a completely adiabatic, low-power, high-speed FIR filter to that of CMOS filters were compared. Reversible logic was used in the design of the PAL FIR filter, and CADENCE digital lab was used to simulate and synthesise it for a variety of parameters, including changes in supply voltage, load capacitance, and transition frequency. This architecture used a logarithmic multiplier to lower the hardware needs and adiabatic technology to provide low power dissipation. However, achieving low area efficiency was more challenging.

Prashanth et al [23] discussed the design of the DA- FIR filter system construct, which was built on an architecture with tightly coupled co-processor-based data processing units. The designed DA-based FIR filter was implemented on field programmable gate array (FPGA) using a series of LUT accesses to simulate multiply and accumulate processes. The proposed filter was implemented using the very high-speed integrated circuit hardware description language (VHDL), and the design is confirmed via simulation. In this study, two optimization strategies were discussed, and the improvements produced were applied to the LUT layer and architecture extractions. The suggested approach provides an optimized design in the form of average LUT minimizations, populated slice reductions, and gate minimization for a discrete impulse response filter. However, combining digital and analog components was challenging, since they have different design constraints, voltage levels, and noise considerations. Hence ensuring seamless integration between the two is crucial.

Maamoun et al [24] proposed an effective high-order FIR filter structure with simultaneous DSP and LUT decreased utilization for FPGA based applications. Also considered was the real-time update of the filter coefficients. Both the speed and the structure of the FPGA were effectively utilized to accomplish these goals. In order to achieve more computation sequences, the difference between the needed input sampling frequency and the FPGA's permitted maximum frequency was handled. Furthermore, the pipelining and selection of the input samples make full use of the unique FPGA Look-up-table Shift-Register (LUT-SR) architecture and internal connections. Reconfigurable filter coefficients were handled by FPGA Block RAMs (BRAMs), and FPGA DSP slices were used to compute the output data of the BRAMs and multiplexers. A single unit was employed for simultaneous control to synchronize the LUT multiplexer selection with the BRAM unit addressing. However, meeting real-time processing requirements is more challenging.

Shrivastava et al [25] proposed an efficient architecture for the DA algorithm-based two-dimensional (2-D) adaptive FIR filter. Practically all DA-based filter topologies demand LUT. The structure that creates the LUT value corresponding to the input, based on adders and logic gates, replaces the RAM- or ROM-based LUT in the proposed filter architecture. As a result, in DA-based realization, the MAC unit needs fewer logic gates and adders. Additionally, the architecture's memory-sharing idea lessens the latency components. Furthermore, the parallel division of the internal MAC block for the DA decomposition, which provides a greater level of flexibility and parallelism in the proposed design, reduces the complexity of the LUT hardware of higher-order filters. The filter coefficient weights were updated using the 2-D delayed Least Mean Square (LMS) algorithm. However, processing two- dimensional signals introduces challenges related to data handling, such as memory organization and data flow management.

Lakshmaiah et al [26] proposed a modified version of the delayed μ-law proportionate normalised least mean square (DMPNLMS) method. This method is an adapted form of the μ-law proportionate normalised least mean square (MPNLMS). To minimise the silicon area, the technique was implemented through the use of a parallel prefix logarithmic adder of the Ladner-Fischer type. VLSI architecture was implemented and simulated using MATLAB, the Vivado suite, and Cadence RTL and Genus Compiler for complementary metal-oxide-semiconductor (CMOS) 90 nm technology nodes. The DMPNLMS approach showed increased stability, a faster rate of convergence, and a decrease in mean square error. However, the proposed LMS algorithm was sensitive to input noise and outliers and hence ensuring the filter remains robust in noisy environments is more complex.

Khan et al [27] designed an LMS algorithm based on the steepest descent technique presented with a potential expansion to its power-normalized LMS version and examined its convergence features. The design and development of non-pipelined ADF systems was accomplished by transforming the coefficient update equation of the LMS algorithm via TC DA and OBC DA. The LUT pre-decomposition approach was utilised by the suggested architectures to improve throughput performance. It allowed the deconstructed LUTs to be updated concurrently using the same mapping approach. Additionally included was an effective fixed-point quantization model for assessing suggested structures from a practical standpoint. However, minimizing power consumption while maximizing throughput is a constant challenge.

Murthy et al [28] presented multiple methods for designing reconfigurable finite impulse response (RFIR) filters. Software-defined radio (SDR) applications were appropriate for programmable FIR filter designs based on DA. Reusing registers, multipliers, and adders as well as optimizing other factors including area, power consumption, speed, throughput, latency, and flip-flop and slice hardware utilizations were the key contributions of reconfiguration. In light of the aforementioned factors, the efficient design of the building blocks was optimized for the RFIR filter. However, achieving high filter performance, such as sharp roll-off and minimal distortion, is challenging, especially when trying to optimize for reconfigurability.

Rammohan et al [29] presented the decimation filter's hardware implementation and architecture for use in hearing aids. Using Matlab Simulink, the CIC, half band filter, and corrector FIR filters were created and tested for real-time implementation. When compared to a basic decimation filter, the suggested decimation filter architecture uses a compressor adder-based DA FIR filter, which reduces the amount of hardware needed by 69% and reduces power consumption by 83%. FIR filters were used in decimation filters for audio applications because they make it simple to establish a linear phase. However, in the future, a linear phase across the entire band is needed.

Uma et al [30] focused on applying DA based FIR filters to remove baseline drift and muscle artifact noise. An area-efficient modified DA-based FIR filter was used to filter out noise and has no LUTs. The modified DA-based FIR filter's performance was contrasted with that of the traditional DA FIR filter. Baseline Wander noise, Muscle Artefact noises, and an arbitrary real-time ECG record are all extracted from the MIT-BIH noise stress test database. Signal to Noise Ratio (SNR) and Mean Square Error (MSE) output metrics were used to assess both filters' performance. The redesigned DA-based FIR filter yields good output SNR and low MSE for baseline wander noise reduction. However, a filter designed for a stationary noise model was not as effective in removing non-stationary noise components.

Nirmala et al [31] proposed a shared LUT updating system for a reconfigurable offset-binary code (OBC) DA-based FIR filter. With each additional filter, the LUTs in DA grow exponentially larger. A way to lessen this significant memory consumption for higher-order filters was a shared LUT-based DA structure. The shared LUT updating method that was being suggested makes use of LUT partitioning, which divides coefficients into small length vectors and significantly reduces the size of LUTs. CMOS 90 nm technology was used to synthesize the suggested DA filter with Synapsis ASIC Design Compiler. When compared to prior designs, the suggested design delivers high speed at a smaller ADP. However, high-speed, low-area OBC-based decimation filters were quite complex to design and implement, especially when dealing with high-order delta-sigma modulators.

Șorecău et al [32] introduced the SDR measurement system for real-time spectrum monitoring. It enabled channel power and complementary cumulative distribution function measurements. It was validated against a high-performance spectrum analyser (SA) in a laboratory setting and successfully captured signals from modern communication standards. The results demonstrated the SDR system's capability to perform real-time measurements and provided valuable insights into signal behaviour, highlighting its potential for advanced spectrum analysis. However, achieving optimal performance across diverse conditions remains a challenge.

Radu et al [33] proposed a system for identifying the modulation of complex radio signals using an artificial intelligence model integrated with a cloud-based platform. The implementation controls a software-defined radio platform to generate and receive real modulated signals, demonstrating the viability of cloud computing for signal processing tasks. The results indicate a high degree of success in identifying certain modulation types, allowing users to access the system from anywhere with an internet connection. However, a significant limitation is the challenge of improving model accuracy under varying signal-to-noise ratios.

From the analysis, [21] building hardware for DSP applications was more challenging, [22] does not attain low power and area efficiency, [23] challenges in combining digital and analog components, [24] does not meet real-time processing requirements, [25] data handling problem obtained, [26] need to ensure the filter in noisy environments, [27] minimizing power consumption is a constant challenge, [28] does not achieve high filter performance, [29] linear phase over the entire band is required, [30] not effective in removing non-stationary noise components, and [31] quite complex in high-speed, low-area OBC-based decimation filters. For [32] it is difficult to perform under various circumstances and [33] indicates that increasing model accuracy at different signal-to-noise ratios is a challenge. Hence, to overcome the aforementioned issues and to enhance the performance of DA-FIR filter, a new novel approach has to be proposed.

III. FIR FILTERS IN SOFTWARE DEFINED RADIO WIRELESS COMMUNICATION SYSTEMS

FIR filters are commonly used in wireless communication systems for various purposes, including signal processing and noise reduction. These filters are used to shape or modify the frequency response of signals to improve communication quality. In SDR systems, the Finite Impulse Response filters play a crucial role in the channelization process, which involves extracting narrowband channels from a wideband signal. These FIR filters must be designed to operate at high sampling rates and handle large-order filters to meet stringent adjacent channel attenuation specifications. The design of FIR filters for SDR applications often focuses on achieving a balance between reconfigurability, high-speed operation, and low power consumption, which are essential for next-generation wireless communications. Advanced FIR filter designs utilize techniques like Distributed Arithmetic and Residue Number Systems (RNS) to improve performance. For instance, DA-based FIR filters can offer significant area delay and energy efficiency improvements, making them suitable for high-throughput implementations. Similarly, RNS-based FIR channel filters can be reconfigured to adapt to various channel filtering specifications, providing speed improvements and complexity reduction compared to traditional methods. These innovations in FIR filter design contribute to the versatility and efficiency of SDR systems, enabling them to support multi-standard wireless communication protocols.

A. Problem Statement and Motivation for the Research

In a traditional FIR filter design, the filter coefficients are multiplied with delayed versions of the input signal, and

the results are summed to produce the filtered output. This operation requires multipliers, delay elements, and adders. Multiplier-based FIR filters typically require dedicated hardware multipliers, which is expensive in terms of both space and energy consumption. This is a significant drawback, particularly in applications where hardware resources are constrained. The use of dedicated multipliers leads to slower calculation speeds, especially for high-speed signal processing. Multipliers tend to be relatively slow compared to other operations, which limits the filter's performance in applications that require real-time processing. Due to their hardware-intensive nature and slow speed potential, traditional multiplier-based FIR filters suffer from low throughput and high latency. This is problematic in applications where the timely processing of signals is crucial. Hence, a novel FIR filter design that mitigates these issues is imperative for enhancing the efficiency and responsiveness of SDR systems.

B. Proposed Design of Novel FIR Filter for an Effective SDR System

To address the limitations of traditional multiplier-based FIR filters, a new innovative approach called **Distributed Arithmetic (DA) with Look Up Table-based FIR filter (DA-LUT-FIR)** is proposed, in which enhanced efficiency of FIR filters is typically achieved by optimizing the computational components of the filter, such as the multipliers and adders thereby significantly speeding up the filter's operation. This study utilizes an Artix-7 FPGA to implement and test an optimized FIR filter design. The Artix-7 FPGA provides high throughput and low latency, crucial for SDR systems, along with robust DSP resources that efficiently handle complex DA operations without relying on traditional multipliers, thus eliminating common bottlenecks. Its optimized power efficiency reduces energy consumption, making it ideal for energy-sensitive applications. Additionally, the Artix-7's high memory bandwidth and extensive LUT resources enhance processing speed and conserve resources, supporting the filter's high- performance demands. Notably, the filter's output frequency response can be dynamically adjusted through a decimation factor, all while keeping the filter coefficients unchanged. A highly adaptive parallel prefix adder is used to lower the worst-case critical route latency of partial product accumulation. FIR filters reduce the amount of LUTs, thereby conserving memory and processing time. To boost performance while cutting down on processing time, this study also suggests limiting the number of coefficients read in parallel for FIR filter operations.

In this proposed design, DA optimizes multiply-and-accumulate (MAC) operations in the FIR filters where instead of directly multiplying filter coefficients with input samples, it precomputes partial products and stores them in memory (LUTs or RAMs). During filter operation, it efficiently combines these precomputed values to compute the final output thereby significantly reducing the need for multipliers, which are resource-intensive in FPGA implementations. Traditional FIR filters rely on multipliers and adders to compute the convolution of input samples with filter coefficients but the LUT-based FIR filters used in this proposed approach replace multipliers with precomputed LUT entries, which store the results of coefficient multiplication thereby avoiding expensive multiplication operations and achieving area and power savings. The decimation factor used in this design dynamically adjusts the output frequency response of the filter and hence, if the original filter operates at a higher sample rate, decimation reduces it to match the desired output rate. The parallel prefix adder efficiently computes the sum of partial products and limits the number of coefficients read in parallel during filter operations. By distributing the addition process across multiple stages, this PPA carefully manages the parallelism and reduces critical path delays thereby striking a balance between throughput and resource utilization and improving performance.

The combination of these technical innovations, including the use of Distributed Arithmetic, dynamic decimation factor, parallel prefix adder, reduced LUT utilization, and coefficient parallelization optimization, collectively improve the efficiency and performance of FIR filters in the proposed DA-LUT-FIR approach. These enhancements enable more efficient and high-performance FIR filtering solutions, particularly for applications where resource constraints and real-time processing requirements are critical, such as in SDR systems.

a. Block Level Diagram of the proposed FIR Filter

Figure 1 depicts the overall block-level diagram for the proposed approach. The architecture of a proposed DA-LUT-based FIR filter typically involves several key components and stages. The filter receives input data, which is the signal to be filtered and it is typically in the form of discrete samples.

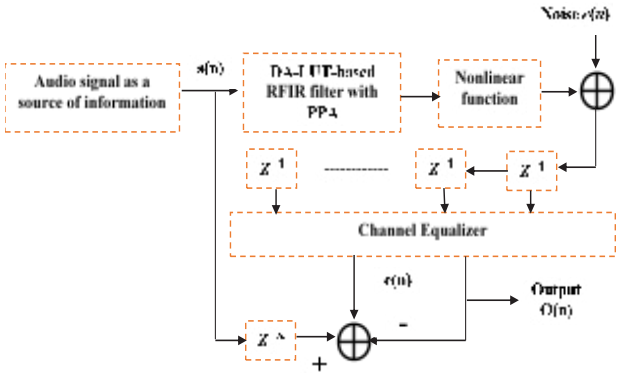


Figure 1: Proposed Block Level Diagram

The FIR filter uses a set of filter coefficients (taps) that determine the filter's behavior. These coefficients are usually constants and define the filter's impulse response. The core of the DA-LUT-based FIR architecture is the use of Look-Up Tables (LUTs), which store precomputed values i.e., the result of multiplying each possible input value by each filter coefficient. The number of LUTs is typically minimized for efficiency. A multiplexer is used to select the appropriate LUT entry based on the current input data value and it effectively "looks up" the precomputed result for the current data value and coefficient. Instead of using traditional multipliers, the DA-LUT-based FIR filter uses multiplier-less multiplication.

The selected LUT entry is treated as a partial product and then the accumulator sums up the partial products obtained from the multiplier less multiplication. This accumulation process continues for multiple data samples, producing the filtered output. The final output of the filter is the result of the accumulation process and it represents the filtered version of the input signal. Depending on the proposed design and application, a decimation stage is added to reduce the output data rate, and is often used in cases where the filter output does not need to retain all the input data points. To enhance performance and throughput, the architecture incorporates parallel processing, which involves the processing of multiple data points and coefficients simultaneously, further improving filter speed. The architecture is highly customizable, allowing for adjustments such as filter length, word length, and the number of LUT entries to be tailored to specific application requirements. The number of LUTs required for a DA-LUT- FIR filter scales with the filter length, particularly if all coefficient multiplications are independently handled. Overall, the DA-LUT-based FIR architecture is designed to efficiently perform filtering operations by utilizing precomputed values stored in LUTs and minimizing the need for traditional multiplication hardware. This results in an efficient and hardware-friendly FIR filter suitable for SDR applications.

C. DA-LUT-FIR filter Formulation

Typically, DA is a well-known FIR filter method, which focuses especially on the computation of the sum of products, often known as the vector dot product that includes several crucial DSP filtering and frequency-shifting operations prompted by the possibilities of the Artix-7 FPGA look-up table architecture. To determine the total number of products needed for FIR filters, DA effectively uses LUTs, shifters, and adders. The DA-LUT-FIR filter Formulation for analysis is discussed in further sections below.

a. Distributed Arithmetic (DA) Computation

When the coefficients of the FIR filter are known, the DA resolves the computation of the internal product, and the output of an FIR filter is given by the convolution sum in equation (1):

$$Y[n] = \sum_{k=0}^{K-1} h[k]x[n-k] \tag{1}$$

Where,

- Y[n] is the output signal at time n
- h[k] are the filter coefficients
- x[n−k] is the input signal at time n−k
- K is the number of filter coefficients (filter length)

DA shifts the computation from the traditional method of directly calculating the product of the input signal and filter coefficients to a method that relies on bit-level manipulations. This is especially efficient in FPGAs where LUTs can store precomputed values. The step-by-step DA process is given in the following equations.

Decompose the input data x[n−k] into its binary representation. For simplicity, each input sample is represented by B bits as given in equation (2):

$$x[n-k] = \sum_{b=0}^{B-1} x_b[n-k] \tag{2}$$

Here, $x_b[n-k]$ represents the b^{th} bit of the input sample x[n−k].

Then precompute all possible values of the partial products $h[k] \cdot x_b[n-k]$ for each bit position and store them in LUTs. This reduces the real-time computation to simple LUT lookups and bit-shifting operations. The precomputed partial products for each bit position are accumulated across all filter taps which is given in equation (3):

$$Y = \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} h[k].x_b[n-k].2^b \tag{3}$$

This step involves shifting the precomputed values according to their bit positions and summing them up using adders. Finally, the DA-based FIR filter output is represented by equation (4) as given below:

$$Y = \sum_{k=0}^{K-1} h[k].x'[k] \tag{4}$$

Here, x'[k] is a function of the bit-level decomposition of the input data x[n–k].

Moreover, in the context of FIR filters, particularly with the proposed DA-LUT-FIR approach, representing input data using a two's complement B-bit binary format is crucial. This representation accommodates both positive and negative values and provides a precision level determined by the number of bits B. The formulation of this representation is provided in the below section.

b. Distributed Arithmetic (DA) for FIR Filters with Two's Complement Representation

In the proposed DA-LUT-FIR filter, the input data x[k] is represented using a two's complement BBB-bit binary representation, allowing for an accurate representation of both positive and negative values. The two's complement representation of x[k] is crucial for handling negative numbers in FIR filter calculations which is given in equation (5);

$$x'[k] = -2^B X_B[k] + \sum_{b=0}^{B-1} x_b[k] \cdot 2^b \tag{5}$$

Where, $x_b[k]$ is b^{th} bit of $x[k]$, $X_B[k] \in \{0,1\}$.

Substituting equation (5) in equation (4),

$$Y = \sum_{k=0}^{K-1} h[k]\left(-2^B X_B[k] + \sum_{b=0}^{B-1} x_b[k] \cdot 2^b\right)$$

$$= -2^B \sum_{k=0}^{K-1} h[k]X_B[k] + \sum_{b=0}^{B-1} 2^b \sum_{k=0}^{K-1} h[k]x_b[k]$$

$$Y = -2^B f(h[k].X_B[k]) + \sum_{b=0}^{B-1} 2^b f(h[k]x_b[k])$$

The final output of the FIR filter with two's complement representation is given in equation (6); Thus, from equation (6), the simplified input data of B-Bit binary data is given below in equation (7):

$$f(h[k]x_b[k]) = \sum_{k=0}^{K-1} h[k]x_b[k] \qquad (7)$$

Therefore, the filter coefficient is further stored at LUT and it is addressed by $X_b[k]$. This reduces entry and summation with LUT of the MAC blocks of FIR filters. The digital filters are made with the use of registers, memory resources, and a scale accumulator to perform this arithmetic. One of the key features of the proposed implementation is the use of a small number of LUTs which is discussed in the following section.

c. Minimizing the LUT size

This proposed approach suggests that the design optimizes the LUTs' usage, potentially by reusing or sharing LUT resources for multiple coefficients to minimize the LUT size while maintaining accuracy. Reducing the number of LUTs leads to savings of hardware resources and power consumption. To execute FIR filter operations efficiently, the proposed approach allows parallel access to multiple coefficients in LUTs. This means that multiple coefficients are accessed simultaneously to perform filter calculations. Parallelism in coefficient access leads to a significant reduction in the processing delay and also enhances the filter's throughput, making it suitable for real-time applications. By minimizing the number of LUTs and enabling parallel access to coefficients, the filter processes the data with lower latency, which is essential for real-time processing.

d. Decimation and Parallel Prefix Adder

Additionally, the frequency response of the filter output is dynamically altered using the decimation factor. Decimation is a process in DSP where the sampling rate of a signal is reduced. By changing the decimation factor, the effective bandwidth and characteristics of the filter output are adjusted without modifying the filter coefficients. It's emphasized that the filter coefficients remain unchanged while altering the frequency response through decimation. This suggests that the filter is designed to be flexible in its application, allowing for real-time adjustments without the need for recalculating or modifying the filter coefficients. To reduce the worst-case critical path time during partial product accumulation, a highly customizable parallel prefix adder is implemented. It is a type of digital adder that efficiently adds multiple numbers of LUT in parallel. By customizing this adder, the design optimizes its performance for the specific requirements of the FIR filter design.

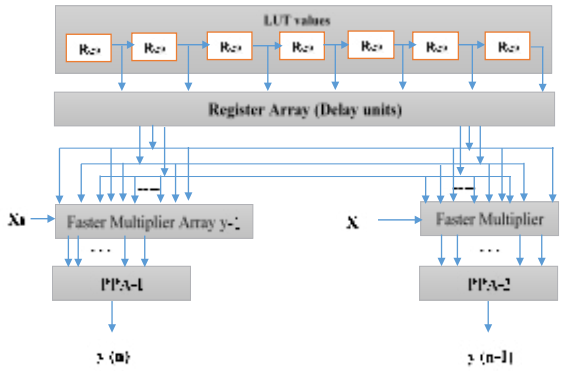


Figure 2: DA-LUT-based RFIR filter with PPA

Figure 2 shows the DA-LUT-based Reconfigurable Finite Impulse Response (RFIR) filter with PPA. The DA with LUT-based RFIR filter, when combined with Power, Performance, and Area (PPA) considerations, offers a versatile and efficient approach to DSP. In order to achieve optimal power efficiency, high performance and minimal hardware footprint, the DA-LUT-based RFIR filter is provided, in which the RFIR filter is a powerful tool for processing digital signals. It allows for adaptability, making it ideal for a wide range of applications in wireless communication systems. This not only reduces power consumption but also accelerates the processing speed of the DA-RFIR filter. The LUT stores precomputed products of filter coefficients and input data, thereby effectively transforming more multipliers into simple LUT. The optimization problem formulation for channel equalizer in terms of the objective function and system constraints is discussed below.

e. Optimization Problem Formulation

The primary objective of this research is to minimize the Bit Error Rate (BER) and latency while maximizing the throughput of the FIR filter system, which is integral to the performance of SDR applications. This section clearly outlines the optimization objectives and constraints associated with the channel equalizer in SDR applications.

Objective Function:

The goal is to minimize latency and BER while maximizing throughput. This can be mathematically represented as:

Objective: min BER (H, X, C), min Latency (H, X, C), and max Throughput (H, X, C)

Where,

- H represents the filter coefficients.
- X denotes the input data.
- C symbolizes the system configurations, including the decimation factor and hardware resources.

Constraints:

Hardware Resource Constraint: The total number of Look-Up Tables (LUTs) and slices used should not exceed the available resources on the Artix-7 FPGA.

$$LUTs(H) \leq 17{,}000$$
$$Slices(H) \leq 10{,}000$$

Power Consumption Constraint: The power dissipation should be within acceptable limits for SDR applications.

$$Power\ (H, X, C) \leq 100\ mW$$

Latency Constraint: The latency must be minimized while ensuring it supports real-time data processing.

$$Latency\ (H, X, C) \leq 20\ ns$$

Throughput Constraint: The filter must maintain a high throughput to handle real-time data processing.

$$Throughput\ (H, X, C) \geq 900\ Mbps$$

Optimization Approach:

Filter Coefficient Optimization: Use DA to precompute possible outcomes for each filter coefficient, reducing the need for real-time multipliers and thus decreasing latency and power consumption.

Parallel Processing: Implement parallel prefix adders to handle partial product accumulations efficiently, enhancing throughput.

Dynamic Decimation: Adjust the decimation factor dynamically to balance the trade-off between processing speed and frequency response.

Adaptive Channel Equalization: Optimize the channel equalizer settings to minimize BER by dynamically adjusting the filter coefficients in response to changing channel conditions.

This approach is not only space-efficient but also reduces the need for resource-intensive multiplication hardware. By carefully analyzing and optimizing the DA-RFIR filter's architecture, the best trade-offs between power efficiency, high performance, and a minimal hardware footprint are achieved. The DA-LUT-based RFIR filter is fine-tuned with PPA, this ensures that signal processing applications operate at peak efficiency, and deliver results that meet the most demanding requirements. With the DA-LUT-based RFIR filter, the power of PPA is unlocked, thereby making it possible to process digital signals with unparalleled efficiency. To perform a performance analysis and optimization of a LUT layer, the proposed model follows a systematic process involving the LUT, identifying bottlenecks, and implementing optimizations. Thereby, the existing technique's drawbacks are overcome by this proposed method. In the next section, the performance and comparison of the proposed method are discussed.

IV. RESULT AND DISCUSSION

In this section, the results for the proposed DA-LUT-FIR filter are presented and engaged in a thorough discussion on its performance and efficiency. The filter was designed and implemented to address the challenges associated with finite impulse response filtering while harnessing the power of distributed arithmetic and lookup tables for optimized multiplication.

A. Experimental Setup

The simulation results are discussed below. This work has been implemented in the MATLAB working platform using the following system specifications.

Software	: MATLAB
OS	: Windows 10 (64-bit)
Processor	: Intel i5
RAM	: 8GB RAM

B. Simulated output of the proposed method

The proposed structure has been added to Xilinx System Generator and Matlab Simulink. For the execution of the proposed design, a string of channel impulses that have been BPSK message modulated for implementation is considered. The signal was transmitted to adaptive DFE for ISI error correction and noise removal. The algorithm is programmed directly into the FPGA integrated within the SDR. This allows for efficient processing and real-time performance, utilizing the FPGA's parallel processing capabilities while minimizing latency and maximizing throughput.

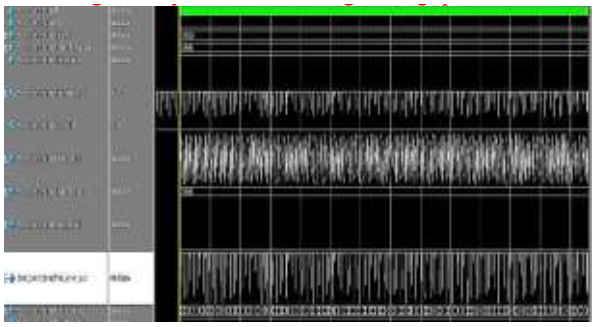


Figure 3: Verilog output of FIR

Figure 3 depicts the Verilog output of the FIR filter. This FIR input module is responsible for receiving the incoming digital data stream and buffering it for processing. It feeds the data into the filter's main processing engine. Multiplier and Accumulator components perform the core filtering operation. The multiplier module multiplies each data sample by the corresponding coefficient, and the accumulator sums up these products to produce the filter's output. The coefficients used by the filter are stored in a memory module. This memory is accessed based on the current position of the sliding window. To slide the window over the input data, there is a control module that manages the window's position and ensures the correct samples are selected for multiplication. Finally, the filtered output data is sent to the output module, which makes it available for further processing.

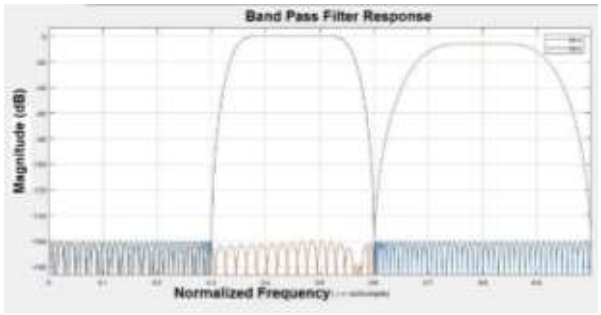


Figure 4: FIR Output Response

Figure 4 depicts the output response for the FIR filter which is characterized by its ability to effectively filter and modify the input signals in a precise and controlled manner. As the input signal progresses through the filter, it encounters each tap and undergoes a series of multipliers and adders. At each tap, the input is multiplied by the corresponding coefficient, and the results are summed together. It represents the frequency response of two bandpass filters where the blue trace (M=1) illustrates a filter that allows a lower range of frequencies to pass through, effectively filtering out frequencies outside this range. Conversely, the orange trace (M=2) demonstrates a filter with a passband at higher frequencies. The graph clearly outlines the effective frequency ranges for each filter, with the passbands being the regions where the magnitude does not exhibit significant attenuation.

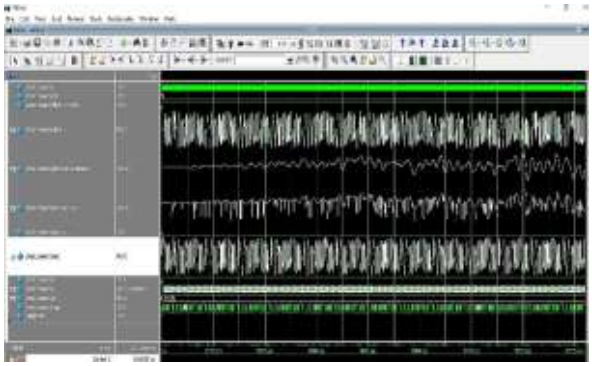


Figure 5: LMS filter output

Figure 5 shows the Least Mean Square (LMS) filter output for the proposed approach. In this figure, the top waveform, labelled "test_bench/data_in" represents the gradient of the input error signal for the first test bench. This is crucial as it indicates how the LMS filter's predictions deviate from the desired outcome. The waveform below, labelled "test_bench/desired_response" is the target or reference signal that the LMS filter aims to replicate or predict accurately. Then the next waveform, labelled "test_bench/FILTER_OUT," shows the output of the filter applied to the first test bench's data. This output is what the LMS filter has produced as its prediction or filtered signal. As the input signal flows through the LMS filter, a remarkable transformation takes place and also this filter armed with its adaptive capabilities meticulously analyses the incoming data in real time. It constantly refines its internal coefficients to minimize the error between the desired signal and the filtered output.

C. Performance metrics of the proposed methodology

The performance metrics collectively provide a comprehensive evaluation of the proposed DA-LUT-FIR filter, helping to assess its efficiency, effectiveness, and suitability for specific applications. The effectiveness of the proposed design is discussed in this section by analysing the performance parameters such as delay, power consumption, number of slice registers, AND/OR gates, LUTs, frequency and the number of adders used.

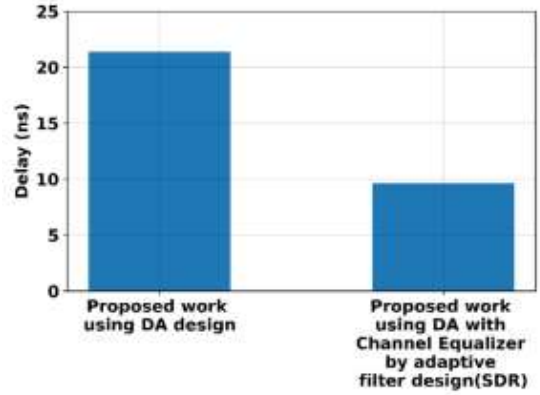


Figure 6: Delay of the proposed design

The graph in figure 6 compares the delay between two FIR filter designs for SDR systems. The first design, using traditional Distributed Arithmetic (DA), has a delay of about 21.41 ns. The second design, which incorporates DA with an adaptive Channel Equalizer, shows a significantly reduced delay of approximately 9.627 ns. This improvement is due to more efficient MAC operations, dynamic adjustment of filter coefficients, and optimized coefficient management. Consequently, the adaptive design ensures faster data processing and lower latency, making it ideal for real-time applications.

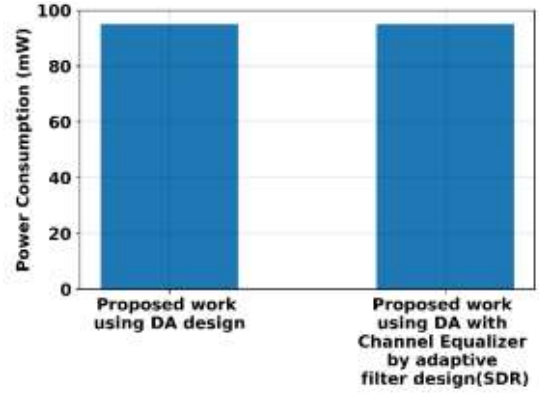


Figure 7: Power consumed by the proposed design

The graph in figure 7 compares the power consumption of the proposed FIR filter design. Both designs show similar power consumption, close to 95 mW. However, the proposed work using DA with an adaptive Channel Equalizer achieves this power efficiency while also significantly reducing delay, as seen in the previous graph. This is due to the optimized use of resources, efficient coefficient management, and reduced need for multipliers, which all contribute to maintaining power

consumption at a low level without sacrificing performance. This balance ensures that the adaptive design is suitable for real-time applications, offering both speed and energy efficiency.

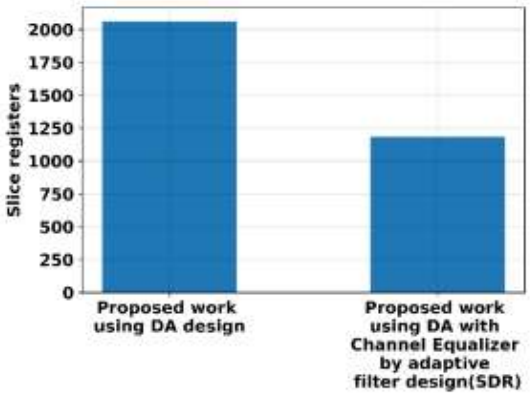


Figure 8: Number of slice registers used in the proposed design

The figure 8 illustrates the number of slice registers used in the proposed work in which using DA design requires 2062 slice registers, whereas the proposed work using DA with an adaptive Channel Equalizer significantly reduces this number t0 1182. The reduction in slice registers is by the efficient design of the adaptive filter, which optimizes resource usage by minimizing redundant or unnecessary computations. This leads to a more streamlined architecture, reducing the hardware complexity and enhancing overall efficiency without compromising performance.

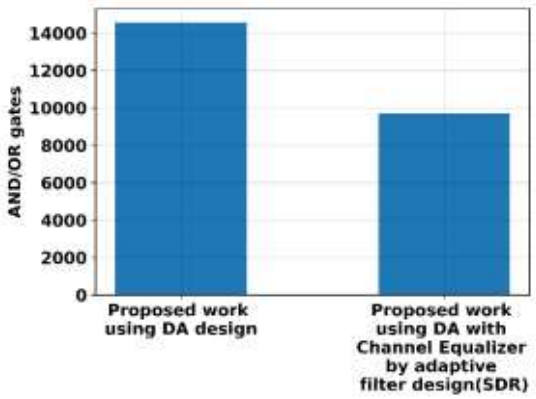


Figure 9: Number of gates used in the proposed approach

The figure 9 shows the number of AND/OR gates used in two different FIR filter designs. The proposed work using DA design requires 14,568 gates, while the proposed work using DA with an adaptive Channel Equalizer uses significantly fewer gates, totalling 9,715. This reduction in the number of gates is because of the adaptive Channel Equalizer's ability to streamline the logic design, minimizing the need for excess logical operations. As a result, the overall gate count is reduced, leading to a more efficient and compact design that maintains functionality while lowering the hardware complexity.

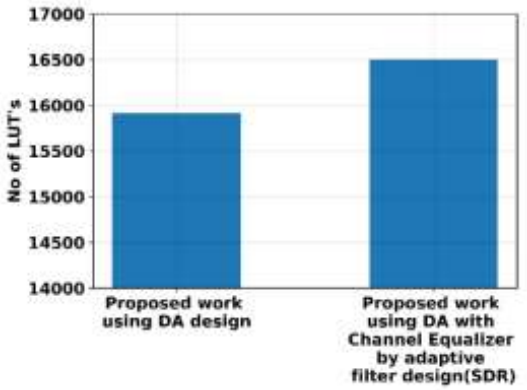


Figure 10: Number of LUTs used in the proposed design

The graph in figure 10 illustrates the number of LUTs used in two different FIR filter designs where the proposed work using DA design requires 15,914 LUTs, while the proposed work using DA with an adaptive Channel Equalizer uses slightly more, totalling 16,504 LUTs. This slight increase in the number of LUTs is by the additional complexity introduced by the adaptive Channel Equalizer, which enhances the system's ability to dynamically adjust to varying channel conditions. Although there is a small increase in the number of LUTs, this trade-off results in improved performance and adaptability, making the design more robust and efficient in handling diverse signal environments.

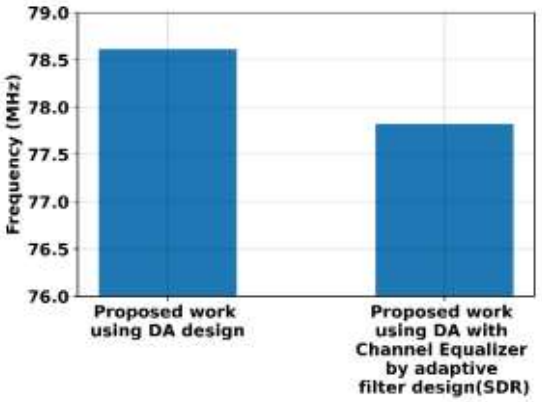


Figure 11: Frequency of the proposed design

The figure 11 compares the operating frequencies of filters in two different designs. The proposed work using DA design attains a frequency of 78.617 MHz, while the proposed work using DA with an adaptive Channel Equalizer operates at a slightly reduced frequency of 77.825 MHz. This minor decrease in frequency is due to the added complexity and functionality of the adaptive Channel Equalizer, which allows the system to better adapt to varying channel conditions. Despite the small reduction in frequency, the enhanced adaptability and performance benefits of the adaptive equalizer outweigh this trade-off, resulting in a more robust and versatile system.

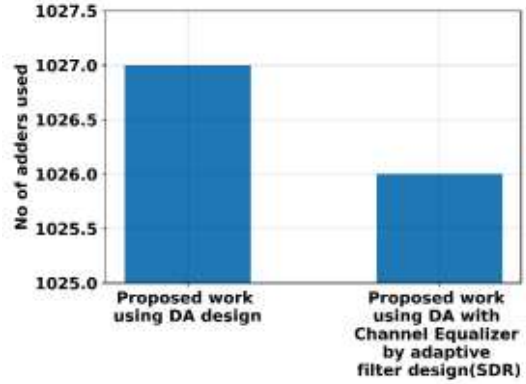


Figure 12: Number of adders used in the proposed filter design

Figure 12 shows the comparison between two proposed designs for digital adders, one employing a conventional DA design and the other integrating an adaptive Channel Equalizer (CE). Interestingly, the latter design, featuring the DA with an adaptive CE, demonstrates a slight reduction in the number of adders required compared to the former, with counts of 1026 and 1027, respectively. This marginal decrease is attributed to the enhanced efficiency achieved through the adaptive CE, which dynamically adjusts to channel variations, optimizing the performance and reducing the demand for additional adders. Hence, while both designs offer competitive functionality, the incorporation of adaptive CE showcases a subtle but notable improvement in resource utilization.

D. Comparison of the proposed methodology

This section highlights the proposed method's performance by comparing it to the outcomes of existing approaches and showing their results based on various metrics. The performance of the existing approaches such as conventional DA-based filter, LUT-Less 2, Separated LUT-DA, and DA-LUT using buffer [34], GBoost Classifier, Light GBM and Gradient Boosting [33] GFSK, GMSK and BPSK OFDM [35], are compared to that of the proposed DA-based LUT-FIR filter.

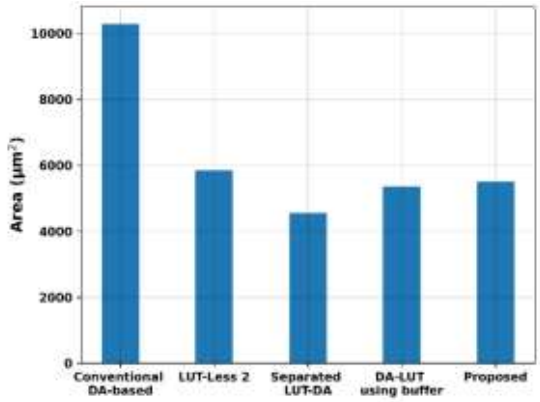


Figure 13: Comparison of area of various filter designs

Figure 13 compares the area efficiency of proposed DA-LUT-FIR filter model with existing approaches, including the conventional DA-based filter, LUT-Less 2, Separated LUT-DA, and DA-LUT using buffer. The conventional DA-based filter occupies 10279 μm², while LUT-Less 2, Separated LUT-DA, and DA-LUT using buffer require 5854 μm², 4554 μm², and 5356 μm², respectively. In contrast, the proposed model achieves a compact area of 5500 μm² by optimally combining precomputed products of filter coefficients with input data, thereby simplifying complex multiplication operations. With streamlined footprint, the DA-LUT-FIR filter offers a significant advancement for efficient signal filtering.

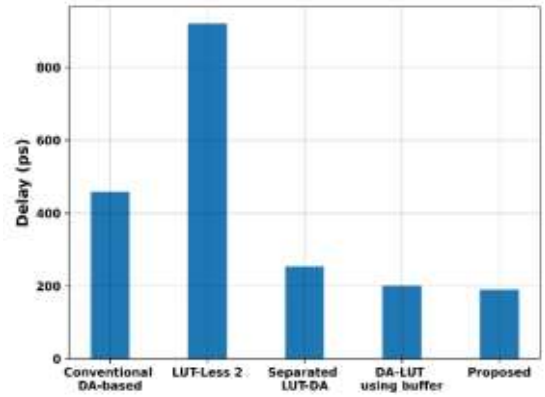


Figure 14: Comparison of Delay

Figure 14 compares the time delay of the proposed DA-LUT-FIR filter model with existing techniques, including the conventional DA-based filter, LUT-Less 2, Separated LUT-DA, and DA-LUT using buffer. The conventional DA-based filter experiences a delay of 459 ps, while LUT-Less 2, Separated LUT-DA, and DA-LUT using buffer have delays of 920 ps, 254 ps, and 201 ps, respectively. In contrast, the proposed approach achieves an impressive delay of just 190 ps. By using the DA-LUT architecture, this innovative filter minimizes the delays typically associated with more resource-intensive FIR filter implementations, showcasing its potential for enhancing performance in time-sensitive scenarios.

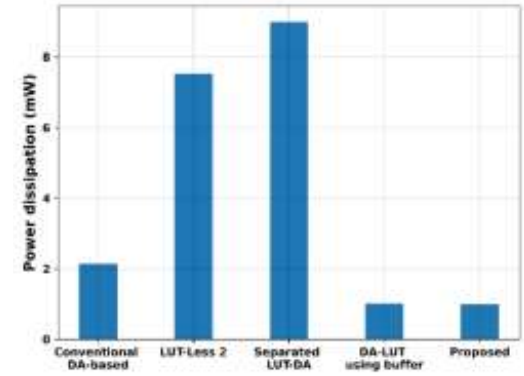


Figure 15: Comparison of Power dissipatio

Figure 15 illustrates the power dissipation of the proposed DA-LUT-FIR filter, which is particularly lower than that of traditional FIR filters, making it an appealing option for SDR applications. Existing approaches, including the conventional DA-based filter, LUT-Less 2, Separated LUT- DA, and DA-LUT using buffer, show power dissipation values of 2.14 mW, 7.52 mW, 8.99 mW, and 1.02 mW, respectively. In contrast, the proposed model achieves a power dissipation of just

1 mW. By integrating DA and LUT technologie, the proposed filter minimizes power consumption, making it ideal for power-sensitive environments in SDR applications.

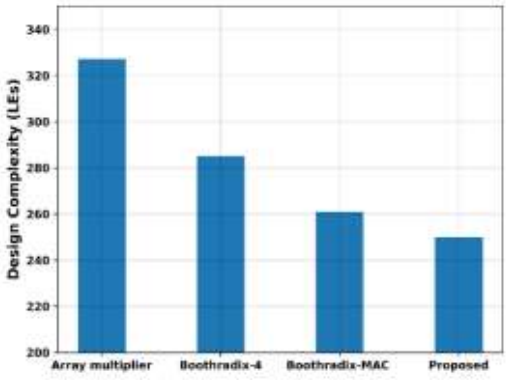


Figure 16: Comparison of Design Complexity

Figure 16 compares the design complexity of the proposed DA-LUT-FIR filter model with existing approaches, including the array multiplier, booth radix-4, and booth radix-MAC unit, which exhibit design complexities of 327 LE, 285 LE, and 261 LE, respectively. Compare to this, the proposed model achieves a design complexity of just 250 LE. This reduction demonstrates the innovative nature of the DA-LUT-FIR filter, significantly decreasing the inherent complexity typically associated with conventional FIR filters and highlighting its efficiency in filter design.

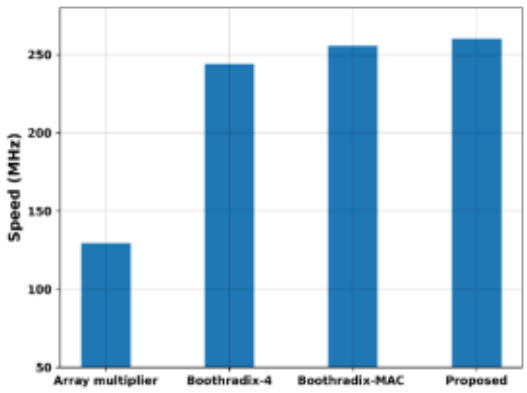


Figure 17: Comparison of processing speed

Figure 17 compares speed of the proposed DA-LUT-FIR filter model with existing models, including the array multiplier, booth radix-4, and booth radix-MAC unit, which achieve speeds of 129.57 MHz, 244.02 MHz, and 255.43 MHz, respectively. The proposed model reaches a speed of 260 MHz. By integrating DA with the LUT approach, it accelerates multiplication operations through precomputed values stored in its LUT, eliminating the need for resource-intensive multipliers.

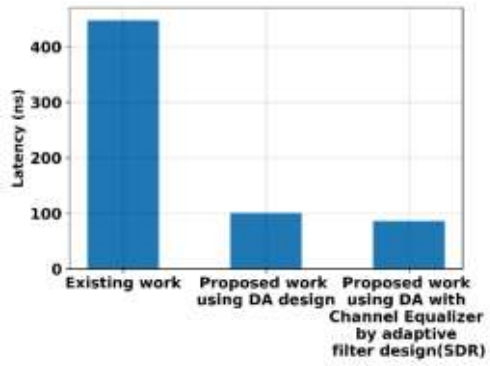


Figure 18: Comparison of latency

Figure 18 provides a comparative analysis of latency in nanoseconds across different FIR filter designs. The baseline latency for 'Existing work' is approximately 448 ns. In contrast, the 'Proposed work using DA with Channel Equalizer by adaptive filter design (SDR)' achieves a remarkable reduction in latency to around 86.126 ns. The 'Proposed work using DA design' shows a slight increase in latency, yet it remains significantly lower than the existing work, with a latency of approximately 101 ns. The proposed designs have significantly reduced latency, making them crucial for applications requiring quick response times.

TABLE I
OVERALL TABLE FOR PERFORMANCE ANALYSIS AND COMPARISON

Parameter	Existing work	Proposed work using DA design	Proposed work using DA with Channel Equalizer by adaptive filter design (SDR)
Block Size	8	8	8
Filter Length	64	64	64
FF	1656	752	952
Delay	56 ns	21.41 ns	9.627 ns
Area (Slices)	839936	6503	8421
Power (Vdd = 1.8V)	251.2 mW	95 mW	95mW
Slice Registers	6144	2062	1182
No of LUT	16,129	15,914	16504
AND/OR gates	190464	14568	9715
Throughput	142.4 Mbps	633.062 Mbps	938.12 Mbps
Frequency	65.7 MHz	78.617 MHz	77.825 MHz
Latency	56ns*8=448 ns	Product of delay and size of data= 12.637ns*8=101.096 ns	Product of delay and size of data=08.527ns*8=86.126 ns
No of adders used	2077	1027	1026
Area-delay product	839936*56ns = 47036416 ns	4145*21.42 ns= 88785.9 ns	8421*9.627 ns = 81068.967 ns
Power-delay product	251.20 mW *56 ns= 14067.2 mW/ns	0.095*21.42 mW = 2.0349 W/ns	0.095*9.627 ns= 0.914565 ns

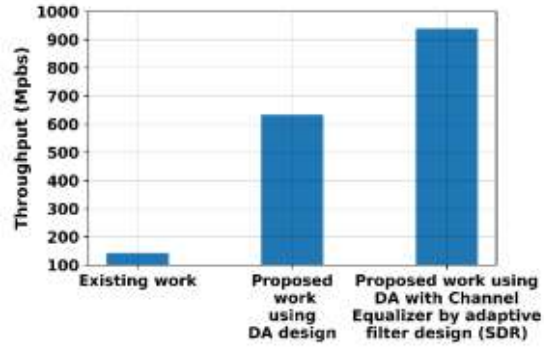


Figure 19: Comparison of Throughput

Figure 19 presents a comparative analysis of throughput performance across three different designs. The Existing work demonstrates modest throughput, significantly below 142.4 Mbps. The 'Proposed work using DA design' shows a substantial improvement, achieving a throughput of 633.062 Mbps. Furthermore, the 'Proposed work using DA with Channel Equalizer by adaptive filter design (SDR)' showcases an impressive throughput close to 938.12 Mbps. The proposed designs significantly improve throughput by integrating an adaptive filter and channel equalization in software-defined radio, doubling the previous models' throughput.

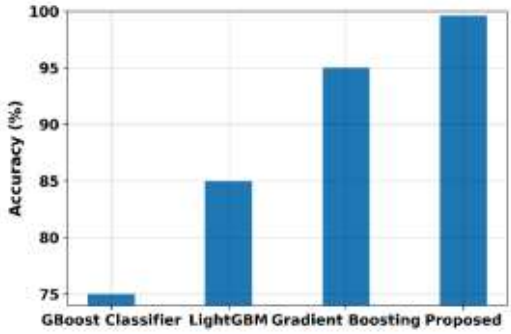


Figure 20: Comparison of Accuracy

Figure 20 compares the accuracy of the proposed DA-LUT-FIR filter model with existing models, including GBoost Classifier, Light GBM and Gradient Boosting, which exhibit accuracy rates of 75%, 85%, and 95%, respectively. The proposed model achieves a significantly higher accuracy of 98%. This improvement underscores the innovative design of the DA-LUT-FIR filter, which not only enhances performance but also minimizes the errors typically associated with traditional filtering methods, highlighting its effectiveness in digital signal processing applications.

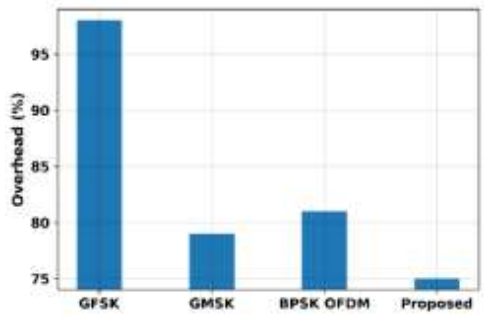


Figure 21: Comparison of Overhead

Figure 21 compares the design complexity of the proposed DA-LUT-FIR filter model with existing models, including GFSK, GMSK and BPSK OFDM, which exhibit design complexities of 98%, 77% and 82% respectively. In contrast, the proposed model achieves a significantly lower overhead of 74%. This reduction demonstrates the innovative nature of the Enhanced Intellectual PMU Controller, significantly decreasing the overhead typically associated with traditional methods and highlighting its efficiency in electric drive applications.

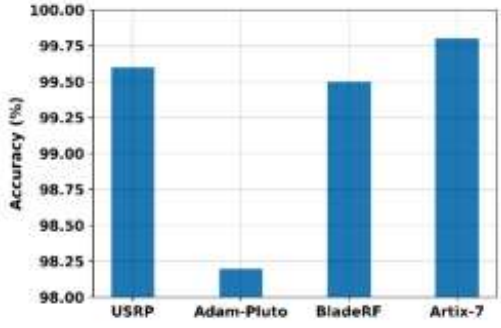


Figure 22: Comparison of accuracy using different FPGA Models

Figure 22 presents an accuracy comparison of various SDR platforms utilizing different FPGA models, including USRP [36], Adalm Pluto [37], and BladeRF [38], which achieve accuracy rates of 99.6%, 98.2%, and 99.5%, respectively. In contrast, the proposed model using the Artix-7 FPGA demonstrates a significantly higher accuracy of 99.8%. This substantial improvement highlights the advanced capabilities of the Artix-7 FPGA in enhancing the performance of SDR applications, effectively minimizing errors associated with traditional models and showcasing its potential in delivering superior digital communication outcomes.

Overall, the proposed models' performance is analyzed and compared with the existing approaches such as conventional DA-based filter, LUT-Less 2, Separated LUT-DA, DA-LUT using buffer array multiplier, boothradix-4, and boothradix-MAC, GBoost Classifier, Light GBM, Gradient Boosting, GFSK, GMSK and BPSK OFD. While comparing the proposed approach with existing models, the proposed approach achieves the best result of delay 190ps, the power dissipation of 1mW, the design complexity attains the value of 250 LE, the processing speed of 260MHz, reduced latency

and overhead of 86 ns and 74%, increased throughput and accuracy of 938.12 Mbps and 99%. Hence the proposed method effectively reduces the noise from SDR applications and enhances the performance of throughput and latency in DA-LUT-based FIR filters.

E. Overall Performance Analysis

The overall performance analysis of the proposed work and its comparison with other existing works is summarized in table 1.

V. CONCLUSION

This study presents a comprehensive evaluation of a novel FIR filter architecture based on Distributed Arithmetic and Look-Up Tables, implemented on an Artix-7 FPGA. This DA-LUT-FIR filter design addresses several key limitations of traditional multiplier-based FIR filters, which often suffer from high hardware complexity, significant power consumption, and slower processing speeds. This proposed filter was implemented with quicker multipliers and adders thereby decreasing bit error rate and latency which in turn helps to boost the throughput of data given in bits. Additionally, the decimation factor frequently changes the FIR filter coefficients, allowing filters to vary their frequency response. According to the experimental findings, fewer LUT for FIR filter coefficients result in less memory usage and latency. The employment of a highly adaptable parallel prefix adder during partial product accumulation was another factor that contributed to the decreased latency. The use of DA and LUTs in the architecture proves to be a powerful combination, delivering remarkable performance improvements and making the filter highly suitable for real-time digital signal processing tasks. The numerical findings from this study—such as the operating speed of 260 MHz, power dissipation of 1 mW, delay of 190 ps, and throughput of 938.12 Mbps—demonstrate substantial improvements over existing methods. These results make the DA-LUT-FIR filter a highly suitable choice for real-time digital signal processing tasks, contributing significantly to the advancement of FIR filter design for future SDR systems.

REFERENCES

[1] Abraham, J., Venusamy, K., Judice, A., Shaik, H. and Suriyan, K., Research, challenges and opportunities in software define radio technologies. *Int J Reconfigurable & Embedded Syst ISSN, 2089*(4864), p. 4864. **DOI**: 10.11591/ijres.v12.i2.pp260-268

[2] Mori, S., Mizutani, K. and Harada, H., 2023. Software-defined radio-based 5G physical layer experimental platform for highly mobile environments. *IEEE Open Journal of Vehicular Technology*, *4*, pp. 230–240. **DOI**: 10.1109/OJVT.2023.3237390

[3] P. P. Sundar, D. Ranjith, T. Karthikeyan, V. V. Kumar, & B. Jeyakumar, "Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture," *International Journal of Speech Technology*, vol. 23 no. 2, pp. 287–296, 2020. **DOI**: 10.1007/s10772-020-09686-y

[4] N. J. Grande, & S. Sridevi, "ASIC implementation of shared LUT-based distributed arithmetic in FIR Filter," In *2017 International Conference on microelectronic devices, Circuits and Systems (ICMDCS). IEEE*, pp. 1–4, 2017, August. **DOI**: 10.1109/ICMDCS.2017.8211705

[5] G. N. Jyothi, K. Sanapala, & A. Vijayalakshmi, "ASIC implementation of distributed arithmetic-based FIR filter using RNS for high-speed DSP systems," *International Journal of Speech Technology*, pp. 1–6, 2020. **DOI**: 10.1007/s10772-020-09683-1

[6] K. Vijetha, & B. R. Naik, "High-performance area-efficient DA-based FIR filter for concurrent decision feedback equalizer," *International Journal of Speech Technology*, vol. 23, no. 2, pp. 297–303, 2020. **DOI**: 10.1016/j.eswa.2024.123488

[7] P. Kumar, P. C. Shrivastava, M. Tiwari, & G. R. Mishra, "High-throughput, the area-efficient architecture of 2-D block FIR filter using the distributed arithmetic algorithm," *Circuits, systems, and signal processing*, vol. 38, no. 3, pp. 1099–1113, 2019. **DOI**: 10.1007/s00034-018-0897-2

[8] M. Sumalatha, P. V. Naganjaneyulu, & K. S. Prasad, "Low power and low area VLSI implementation of Vedic design FIR filter for ECG signal de-noising," *Microprocessors and Microsystems*, vol. 71, p. 102 883, 2019. **DOI**: 10.1016/j.micpro.2019.102883

[9] K. Bagadi, C. V. Ravikumar, K. Sathish, M. Alibakhshikenari, B. S. Virdee, L. Kouhalvandi, K. N. Olan-Nuñez, G. Pau, C. H. See, I. Dayoub, and P. Livreri, "Detection of signals in MC–CDMA using a novel iterative block decision feedback equalizer," *IEEE Access*, vol. 10, pp. 105 674–105 684, 2022. **DOI**: 10.1109/ACCESS.2022.3211392

[10] S. R. Rammohan, N. Jayashri, M. A. Bivi, C. K. Nayak, & V. R. Niveditha, "High-performance hardware design of compressor adder in DA-based FIR filters for hearing aids," *International Journal of Speech Technology*, vol. 23, no. 4, pp. 807–814, 2020. **DOI**: 10.1007/s10772-020-09759-y

[11] S. F. Ghamkhari, & M. B. Ghaznavi-Ghoushchi, "A New Low Power Schema for Stream Processors Front-End with Power-Aware DA-Based FIR Filters by Investigation of Image Transitions Sparsity," *Circuits, Systems, and Signal Processing*, pp. 1–23, 2021. **DOI**: 10.1007/s00034-020-01632-2

[12] M. E. Meybodi, H. Gomez, Y. C. Lu, H. Shakiba, and A. Sheikholeslami, "Design and implementation of an on-demand maximum-likelihood sequence estimation (MLSE)," *IEEE Open Journal of Circuits and Systems*, vol. 3, pp. 97–108, 2022. **DOI**: 10.1109/OJCAS.2022.3173686

[13] T. V. Padmavathy, S. Saravanan, & M. N. Vimalkumar, "Partial product addition in Vedic design-ripple carry adder design for filter architecture for electrocardiogram (ECG) signal de-noising application," *Microprocessors and Microsystems*, vol. 76, p. 103 113, 2020. **DOI**: 10.1016/j.micpro.2020.103113

[14] B. Pandey, N. Pandey, A. Kaur, D. A. Hussain, B. Das, & G. S. Tomar, "Scaling of output load in energy efficient FIR filter for green communication on ultra-scale FPGA," *Wireless Personal Communications*, vol. 106, no. 4, pp. 1813–1826, 2019. **DOI**: 10.1007/s11277-018-5717-2

[15] S. Yergaliyev, and M.T. Akhtar, "A Systematic Review on Distributed Arithmetic-Based Hardware Implementation of Adaptive Digital Filters," *IEEE Access*, 2023. **DOI**: 10.1109/ACCESS.2023.3304234

[16] W. M. Salama, M. H. Aly, and E.S. Amer, "Underwater optical wireless communication system: Deep learning cnn with noma-based performance analysis," *Optical and Quantum Electronics*, vol. 55, no. 5, p. 436, 2023. **DOI**: 10.1007/s11082-023-04638-7

[17] P. Chowdari Ch, and J. B. Seventline, "Implementation of distributed arithmetic-based symmetrical 2-D block finite impulse response filter architectures," *F1000Research*, vol. 12, p. 1182, 2023. **DOI**: 10.12688/f1000research.126067.1

[18] B. U. V. Prashanth, M. R. Ahmed, & M. R. Kounte, "Design and implementation of DA FIR filter for bio-inspired computing architecture," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, p. 1709, 2021. **DOI**: 10.11591/ijece.v11i2.pp1709-1718

[19] S. Sridevi, & R. Dhuli, "ASIC Implementation of Linear Periodically Time Varying Filter by Thread Decomposition," *In International Conference on Advances in Electrical and Computer Technologies. Singapore: Springer Nature Singapore*, pp. 775–788, 2020, October. **DOI**: 10.1007/978-981-15-9019-1_67

[20] A. Gorantla, & T. Kudithi, "ASIC Implementation of Linear Equalizer Using Adaptive FIR Filter," *International Journal of e-Collaboration*, vol. 16, no. 4, 2020. **DOI**: 10.1007/978-981-15-9019-1_67

[21] P. Kumar, P. C. Shrivastava, M. Tiwari, & G. R. Mishra, "High-throughput, the area-efficient architecture of 2-D block FIR filter using the distributed arithmetic algorithm," *Circuits, systems, and signal processing*, vol. 38, no. 3, pp. 1099–1113, 2019. DOI: 10.1007/s00034-018-0897-2

[22] A. Rai, "An optimization of low power 4-bit PAL FIR filter using adiabatic techniques," *Sādhanā*, vol. 48, no. 2, p. 84, 2023. DOI: 10.1007/s12046-023-02132-0S

[23] B. U. V. Prashanth, M. R. Ahmed, & M. R. Kounte, "Design and implementation of DA FIR filter for bio-inspired computing architecture," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, p. 1709, 2021. DOI: 10.11591/ijece.v11i2.pp1709-1718

[24] M. Maamoun, A. Hassani, S. Dahmani, H. Ait Saadi, G. Zerari, N. Chabini, & R. Beguenane, "Efficient FPGA-based architecture for high-order FIR filtering using simultaneous DSP and LUT reduced utilization," *IET Circuits, Devices & Systems*, 2021. DOI: 10.1049/cds2.12043

[25] P. C. Shrivastava, P. Kumar, M. Tiwari, & A. Dhawan, "Efficient Architecture for the Realization of 2-D Adaptive FIR Filter Using Distributed Arithmetic," *Circuits, Systems, and Signal Processing*, vol. 40, no. 3, pp. 1458–1478, 2021. DOI: 10.1007/s00034-020-01539-y

[26] G. S. Lakshmaiah, C. K. Narayanappa, L. Shrinivasan, and D. M. Narasimhaiah, "Efficient very large-scale integration architecture design of proportionate-type least mean square adaptive filters," *Int J Reconfigurable & Embedded Syst*, vol. 13, no. 1, pp. 69–75, 2024. DOI: 10.11591/ijres.v13.i1.pp69-75

[27] M. T. Khan, M. A. Alhartomi, S. Alzahrani, R. A. Shaik, and R. Alsulami, "Two distributed arithmetic based high throughput architectures of non-pipelined LMS adaptive filters," *IEEE Access*, vol. 10, pp. 76 693–76 706, 2022. DOI: 10.1109/ACCESS.2022.3192619

[28] C. S. Murthy, & K. Sridevi, "Optimized DA-reconfigurable FIR filters for software defined radio channelizer applications," *Circuit World*, vol. 47, no. 3, pp. 252–261, 2021. DOI: 10.1108/CW-11-2020-0332

[29] S. R. Rammohan, N. Jayashri, M. A. Bivi, C. K. Nayak, & V. R. Niveditha, "High performance hardware design of compressor adder in DA based FIR filters for hearing aids," *International Journal of Speech Technology*, vol. 23, pp. 807–814, 2020. DOI: 10.1007/s10772-020-09759-y

[30] A. Uma, & P. Kalpana, "ECG Noise Removal Using Modified Distributed Arithmetic Based Finite Impulse Response Filter," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 5, pp. 1444–1452, 2021. DOI: 10.1166/jmihi.2021.3770

[31] Y. Nirmala, B. Bhaygya, & B. Saimani, "High speed low area OBC DA based decimation filter for hearing aids application". DOI: 10.1007/s10772-019-09660-3

[32] Şorecău, M., Şorecău, E., Sarbu, A. and Bechet, P., 2023. Real-time statistical measurement of wideband signals based on software defined radio technology. *Electronics*, *12*(13), p. 2920. DOI: 10.3390/electronics12132920

[33] Radu, F., Cotfas, P. A., Alexandru, M., Bălan, T. C., Popescu, V. and Cotfas, D. T., 2023. Signals Intelligence System with Software-Defined Radio. *Applied Sciences*, 13(8), p. 5199. DOI: 10.3390/app13085199

[34] E. Chitra, & T. Vigneswaran, "An efficient low power and high speed distributed arithmetic design for FIR filter," *Indian Journal of Science and Technology*, vol. 9, no. 4, pp. 1–5, 2016. DOI: 10.17485/ijst/2016/v9i4/79055

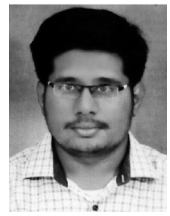
[35] Molla, D. M., Badis, H., George, L. and Berbineau, M., 2022. Software defined radio platforms for wireless technologies. *IEEE Access*, *10*, pp. 26 203–26 229. DOI: 10.1109/ACCESS.2022.3154364

[36] Alashqar, A., Mesleh, R. and Alshawaqfeh, M., 2023, June. Digital Communication Software-Defined Radio-Transceiver Implementation Using MATLAB and USRP. In *2023 International Wireless Communications and Mobile Computing (IWCMC)* (pp. 929–934). IEEE. DOI: 10.1109/IWCMC58020.2023.10182939

[37] Üngüder, Ö. Ö., 2023. *Real-Time Chat Application with ADALM-PLUTO Software Defined Radio* (Doctoral dissertation, Hochschule Rhein-Waal). https://opus4.kobv.de/opus4-rhein-waal/frontdoor/index/index/docId/1852

[38] Terris-Gallego, R., Fernandez-Hernandez, I., López-Salcedo, J. A. and Seco-Granados, G., 2023. E1-E6 SDR Platform Based on BladeRF for Testing Galileo-Assisted Commercial Authentication Service. *Engineering Proceedings*, *54*(1), p. 29. DOI: 10.3390/ENC2023-15428

S. Hari Krishnan obtained his B.E Degree in Electronics & Communication Engineering from Mepco Schlenk Engineering College, Sivakasi, Tamilnadu and M.E Degree in VLSI DESIGN from Karpagam University, Coimbatore, Tamilnadu. His Field of interest includes Digital Signal Processing and VLSI Design. He has organized and attended more number of Conferences, seminars, Workshops, Symposiums, Faculty development programmes, Project expos.. He has published several articles in international and national conferences and also he guided various UG projects. Currently, he is an Associate Professor in the Department of ECE at Sanskrithi School Of Engineering, Puttaparthi, Andra Pradesh, India.



Syed Sadiq Vali has received Bachelor of technology degree (B.Tech) in Electronics and Communication Engineering(ECE) from Sri Krishna Devaraya University, Aanantapur, Andhra Pradesh, in 2017, M.Tech in VLSI system design from JNTU Anantapur in 2021 and pursuing Ph.D degree from mohan Babu University, Tirupati. Currently, he is an Assistant Professor in the department of Electronics and Communication Engineering at Sanskrithi School of Engineering, Puttaparthy, Andra Pradesh, India. Research Interests includes: System-on-Chip Architecture design and Low-Power VLSI systems.

Analysis of Interactions Among ISPs in Information Centric Network with Advertiser Involvement

Hamid Garmani, Mohamed El Amrani, Driss Ait Omar, Mohamed Baslam, and Hicham Zougagh

***Abstract*—In response to the escalating volume of Internet traffic, scalability challenges have emerged in content delivery. Information-Centric Networking (ICN) has emerged as a solution to accommodate this surge in traffic by leveraging caching. Collaborative caching within ICN is pivotal for enhancing network performance and reducing content distribution costs. However, current pricing strategies on the Internet do not align with ICN interconnection incentives. This paper delves into the economic incentive caching of free content among various types of ICN providers, including advertisers and Internet service providers (ISPs). Specifically, we employ game-theoretic models to analyze the interaction between providers within an ICN framework, where providers are incentivized to cache and share content. Content popularity is modeled using a generalized Zipf distribution. We formulate the interactions among ISPs as a non-cooperative game and, through mathematical analysis, establish the existence and uniqueness of the Nash equilibrium under certain conditions. Additionally, we propose an iterative and distributed algorithm based on best response dynamics to converge towards the equilibrium point. Numerical simulations demonstrate that our proposed game models yield a win-win solution, showcasing the effectiveness of our approach in incentivizing collaborative caching of free content within ICN.**

***Index Terms*—Pricing, ICN, ISP, Caching, advertisers, Nash equilibrium, Game Theory.**

I. INTRODUCTION

Since its inception in the 1960s, the Internet has increasingly become integral to people's lives. Currently, the predominant mode of content delivery involves distributing content from a single source to multiple users (similar to multicast), facilitating the dissemination of multimedia files from creators (such as Netflix, IPTV, and various websites) and the sharing of user-generated content on platforms like Facebook, Weibo, YouTube, and Youku. The exponential growth in users and service demands has led to significant improvements in the backbone network infrastructure, including the deployment of numerous routers, high-speed transmission technologies, and private network systems. These enhancements aim to achieve more efficient data delivery across the Internet [1].

Traditionally, content delivery has relied on the client-server model. However, a key solution that offers superior performance in terms of lower access latency, higher data transfer rates, and reduced costs compared to the client-server model involves moving content from the original server to the edge of the Internet, known as a local replica server. A prime example of this approach is the Content Delivery Network (CDN) using IP multicast, which addresses the core challenge of Internet infrastructure [2]. CDN works by distributing content from the original server to end-users via replica servers strategically placed across the network, alleviating backbone network congestion and enhancing service quality. Content stored and served through replica servers is meticulously selected to achieve near 100% hit rates in some instances. Consequently, CDN implementation can significantly reduce access delays, boost content distribution rates, and minimize network bandwidth utilization [3][4][5].

As network traffic continues to escalate and quality of service (QoS) demands become more complex, maintaining optimal CDN performance has become costly due to the need for a large number of edge servers and ongoing operational expenses [6]. Additionally, the agility of server deployment is limited, resulting in prolonged setup times and challenges in selecting appropriate locations with sufficient capacities. A significant drawback of CDN is the lack of real-time network condition information, leading to suboptimal user assignment decisions and exacerbating network bottlenecks [7][8]. The rapid surge in network traffic has strained existing CDN systems, highlighting limitations in IT infrastructure and storage space availability.

ICN is gaining momentum as a prospective Internet framework, representing a departure from host-centric communication towards user-driven data retrieval [9]. Its inception can be traced back to TRIAC, as proposed by Cheriton and Gritter in 2000. Subsequent advancements include the development of Data-Oriented Network Architecture (DONA) by [10]. Related technologies such as Content-Centric Networking (CCN) and Named Data Networking (NDN) complement ICN's objectives. In ICN, users prioritize the content itself over its origin [11]. This stems from a recognition that contemporary Internet usage increasingly emphasizes data dissemination rather than host-to-host communication. ICN thus aims to address this trend by offering a new, more efficient model. Interest-oriented networking within ICN utilizes dynamic content caching at the network layer, facilitating reliable and scalable content delivery. This approach accelerates the timely provision of information to end-users. However, there are diverse approaches to ICN distribution strategies, with ongoing research exploring various means to overcome additional challenges in Internet architecture, such as mobility management and security enforcement [12]. A fundamental aspect of ICN is its

Hamid Garmani, Mohamed El Amrani, Driss Ait Omar, Mohamed Baslam, and Hicham Zougagh are with the Sultan Moulay Slimane University, TIAD Laboratory, Beni Mellal, Morocco.

Manuscript received December 1, 2012; revised August 26, 2015.

Corresponding author: H. Garmani (email: garmani.hamid@gmail.com).

DOI: 10.36244/ICJ.2024.4.4

utilization of in-network caching to enhance the efficiency of content dissemination and information retrieval across network entities. From a caching perspective, ICN cache exhibits distinct characteristics such as transparency to applications, ubiquity, and fine-grained content caching, setting it apart from conventional Web and CDN caching methods [13][14][15].

In-network caching is recognized for its potential to yield advantages for users, ISPs, and content providers (CPs). By caching content on nearby routers, users can experience reduced delays in accessing content, consequently enhancing their overall quality of experience. Viewed from the ISP's standpoint, in-network caching offers the benefit of reducing outgoing traffic directed towards neighboring ISPs or content providers, thereby decreasing overall bandwidth consumption. Consequently, studies [16][17][18] advocate for ISPs to implement infrastructures equipped with built-in caching capabilities. Furthermore, for content providers, fulfilling requests from the on-path caching router can alleviate the burden on the content provider, conserve bandwidth resources, and mitigate network congestion. Particularly noteworthy is that in situations where the content provider experiences temporary downtime, the network's built-in content store can continue delivering content to users, thereby bolstering the network's resilience to some extent.

Given the pivotal role of in-network caching in ICN, scholars have devoted considerable attention to content caching and distribution. A key consensus emphasizes the necessity for various ICNs to collaborate to enhance overall network performance. However, due to the self-interested nature of ICN entities (such as ISPs and CPs), these entities may refrain from caching or forwarding data, and ISPs may hesitate to deploy ICN if profitability is not assured. Therefore, to promote content distribution and sharing, it becomes imperative to identify a suitable win-win economic model that benefits all entities involved.

In this study, we delve into caching free content, alongside pricing and Quality of Service (QoS) strategies adopted by ISPs, leveraging insights from content popularity. We explore Nash strategies within a non-cooperative game framework among the aforementioned entities, utilizing a probabilistic model that assumes users' requests conform to the generalized Zipf distribution.

Our analysis rigorously establishes the existence and uniqueness of Nash equilibrium in this non-cooperative game among ISPs. This equilibrium signifies a stable state wherein ISPs lack incentives to alter their strategies. Thus, our model not only offers economic incentives for caching content but also ensures the stability of the economy and fosters economic growth. We substantiate our analysis with numerical results demonstrating the mutual benefits accrued by both ISPs and users through caching investment. These findings underscore the viability of our approach in fostering symbiotic relationships within the caching ecosystem.

The remainder of this paper is structured as follows. Section 2 provides an overview of related research. In Section 3, we delineate the system model employed in our study. Section 4 is dedicated to formulating and analyzing the non-cooperative game under consideration. Subsequently, Section 5 presents the numerical results derived from our analysis. Finally, Section 6 offers conclusions drawn from our findings.

II. Related Work

The existing literature has investigated various issues pertaining to caching within small cell networks, cloud radio access networks, CPs, and ISPs. The authors in [19] introduced two strategies for uplink transmission in distributed Beyond Fifth Generation small cell networks. The initial scheme prioritizes content matching to eradicate duplicate contents within distributed caches. Meanwhile, the second scheme reallocates non-duplicated cached contents among distributed caches, considering their available space and content size. These methods target boosting energy and spectral efficiency by minimizing redundant uploads and optimizing distributed content caching. Ultimately, the goal is to enhance content delivery efficiency. In [20] the authors presented a cooperative caching strategy leveraging mobile prediction, and social awareness to enable collaborative content decision-making among edge devices. They utilized the long short-term memory network to forecast vehicle trajectories, calculate social relationships through content similarity and contact rates among vehicle users to identify potential caching nodes, and employ deep reinforcement learning to determine the ultimate caching decision. In [21] the authors introduced a content-centric network-based content transfer strategy tailored for the vehicle-to-grid network. Specifically, they leveraged the mobility patterns of Electric Vehicles to capture inter-contact times between node pairs and devised a caching placement scheme aimed at maximizing the content offloading ratio.

The authors in [22] tackled the vehicular service caching issue by leveraging vehicle surroundings' function-features and transportation correlations. Initially, they utilized these features to estimate service preferences, ensuring high hit ratios for cached services. Subsequently, they framed the vehicular service caching problem as a constrained optimization challenge and devised a collaboration mechanism among Roadside Units, utilizing transportation correlations from vehicle trajectories. Finally, they introduced two vehicular service caching techniques based on Gibbs sampling to optimize network delays for vehicular services and mitigate the adverse effects of vehicle mobility. The authors in [23] introduced a labeled graph partitioning scheme for distributed edge caching, which hinges on frequent query patterns. This scheme first generates frequent query patterns from users' historical query subgraphs and then conducts labeled graph partitioning based on these patterns. This ensures that the resulting labeled graph, divided among edge servers, optimally serves frequent user queries while maintaining a minimum edge cut. In [24] the authors introduced a collaborative caching strategy employing federated learning. Initially, federated learning is employed to predict user preferences across distributed nodes, enabling the development of an efficient content caching policy. Subsequently, the allocation of caching resources to optimize video provider costs is framed as a Markov decision process, and a reinforcement learning approach is utilized to optimize caching decisions. The authors in [25] proposed a

video caching and transcoding strategy for delay-constrained content delivery in multi-tier wireless networks. Particularly for multimedia services whose content can be encoded into multiple quality versions and consists of multiple chunks, they presented an approach of caching chunks of an identical file separately in different network layers with different qualities. The authors decomposed the joint optimization problem into caching and transcoding subproblems.

The authors in [26] developed a model considering the preferences of both service providers and requesters to optimize cache content sharing via content-centric network communications. They formulated a decentralized matching problem incorporating the joint transmit power of both content providers and requesters. Lastly, the authors introduce a distributed blind matching algorithm, implemented through a smart contract deployed on the Ethereum network, thereby eliminating the need for an intermediary authority. In [27] the authors explored the dynamic allocation of cache resources and price determination problem within a caching system comprising multiple content producers as buyers and multiple competing ICN cache providers as sellers of caching resources. They introduced a novel reverse auction-based caching and pricing scheme designed to maximize the caching benefits for content producers. The authors in [28] suggest an economic pricing strategy to tackle caching resource management challenges in 5G wireless networks, aiming to address limitations in throughput, latency, and reliability. Furthermore, they explore this approach through an oligopolistic multi-market framework derived from Cournot, Stackelberg, and Bertrand models. In [29] the authors explored the utilization of an in-network caching model involving multiple CPs. They investigated the collective objective of maximizing CPs' profits and increasing their market share to capture a substantial customer base. This problem is modeled as a non-cooperative game among the CPs. The authors in [30] delve into the dynamics between users, ISPs, and CPs. Then they scrutinize the caching and pricing strategies employed by each entity within the context of Nash equilibrium. By comparing and analyzing these strategies, they aimed to determine which charging model best aligns with the requirements of network entities in the ICN environment. In [31] the authors delve into a non-cooperative game involving an ISP and a CP. Within this framework, the ISP adjusts its caching strategy while the CP can influence its pricing strategy. They notably highlight that the direction of the side-payment (from the ISP to the CP) in an ICN setup differs from that in the existing Internet model, which follows a host-centric communication model. The authors in [32] proposed an examination of the dynamic interactions within a duopoly model featuring two ISPs. The competition between these ISPs revolves around their pricing strategies and the quantity of cached items, with both entities characterized as bounded rational. In [33] the authors investigated and analyzed the non-cooperative game dynamics between ISPs and CPs within the context of ICN. They proposed establishing peering links between access ISPs to facilitate collaborative caching. Their findings indicate that cache allocation can be reasonably managed based on end-user demand, allowing each entity to determine an optimal pricing strategy according to its equilibrium point. The authors in [16] characterized content popularity using the generalized Zipf distribution and employed a non-cooperative game framework to determine caching and pricing strategies.

From a technical standpoint, ICN presents considerable potential as a future network architecture. The successful deployment of ICN hinges on the development of effective pricing and caching mechanisms. While most existing research has primarily focused on caching paid content, where ISPs generate revenue from both network and content access, our study fills a gap by exploring pricing and caching strategies for free content. Using game theory and Nash equilibrium, we analyze these interactions. Furthermore, our research introduces an advertiser model to investigate the economic dynamics between advertisers and ISPs, where ISPs earn revenue from advertisers and network access. This approach offers a novel perspective by concentrating on free content and the role of advertisers, providing insights that extend beyond previous studies.

III. PROBLEM MODELING

We examine a streamlined networking market featuring a single CP, N ISPs, and a considerable number of users. In this setup, all users can exclusively access the CP content through the network infrastructure offered by their respective ISP, with the CP serving as the content source for the users. Figure 1 illustrates the financial transactions among diverse entities at different price points. The network economy relies on three key factors: pricing, caching, and QoS. Assuming that each ISP has access to all content, it can choose to cache either the entire requested content or a portion of it. Consider L as the quantity of items sold by the CP. The caching strategy implemented by each ISP is represented by k_{il}, where it assumes a value of 1 if ISP_i opts to cache item l, and 0 if ISP_i chooses not to cache item l. Each ISP_i establishes two distinct prices: (1) the network price per unit of data p_{s_j}, for transporting the content to users; and (2) price per attention $p_{a_{il}}$ for advertising within the cached content. Each ISP_i assigns a bandwidth Ψ_i and promotes a QoS q_{s_i} to users. In representing user behavior, we have assumed that the demand for content at each ISP is a linear function of the strategies employed by all ISPs. A detailed summary of notations is presented in Table I.

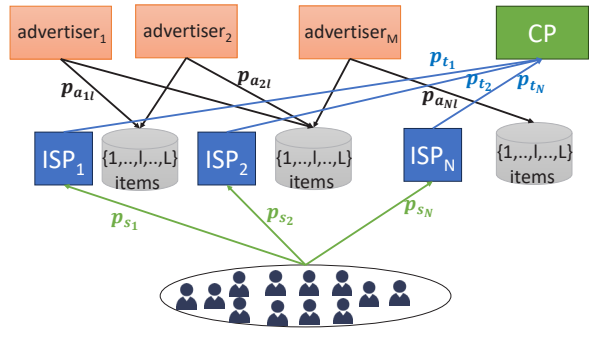


Fig. 1. Model architecture.

TABLE I
SUMMARY OF NOTATION.

Notation	Description
N	Number of ISPs.
L	The number of items the CP would sell.
L_i	Number of items that the CP sells to ISP_i.
M_{il}	Number of advertisers for item l in ISP_i cache.
B_{il}	Budget over a specified time frame (such as daily, weekly, or monthly).
v_{il}	Valuation assigned by advertisers for item l, representing their willingness to pay for ISP_i advertisement slots.
$\bar{v}_{il}$	Maximum value of v_{il}.
p_{s_i}	Network access price of ISP_i.
$p_{a_{il}}$	Price charged per attention by the ISP_i for item l.
q_{s_i}	Quality of service (QoS) of ISP_i.
$D_{a_{il}}$	Demand for attention from advertisers to the ISP_i for item l.
C_{il}	Cost paid by ISP_i for caching item l.
γ_i	Backhaul bandwidth cost of ISP_i.
p_{t_i}	Fee that the ISP_i pays to the CP when requesting content from it.
α_i^j	Sensitivity of ISP_i to price p_{s_j} of ISP_j.
β_i^j	Sensitivity of ISP_i to QoS q_{s_j} of ISP_j.
d_i	Total potential demand of users of ISP_i.
D_i	Demand of ISP_i.
Ψ_i	Backhaul bandwidth of ISP_i.
l^η	Rank of item l.
η	Skewness of the popularity distribution.
χ_{il}	Non-negative constant.
k_{il}	The caching strategy implemented by each ISP, where $k_{il} = 1$ if ISP_i opts to cache item l, and $k_{il} = 0$ if ISP_i chooses not to cache item l.
ISP	Internet Service Provider.
CP	Content provider.

A. Content Popularity

Let L denote the total number of free items provided by CPs, with each item characterized by a measure of popularity represented by the probability of requests for it. We adopt a model where the popularity of content remains uniform across all users. As observed in prior studies (e.g., [34] [35]), we model the probability of requests using the generalized Zipf distribution function, defined as follows:

$$\phi_l = A^{-1} l^{-\eta} \tag{1}$$

where $A = \sum_{l=1}^{L} l^{-\eta}$, $l^{-\eta}$ is the rank of item $l - th$, and η is the skewness of the popularity distribution. Each item is ranked according to their popularity, with item l representing the $l - th$ most popular item, where $l = 1$ denotes the most popular item, and $l = L$ represents the least popular item. The reason we have used the generalized Zipf distribution function is twofold: (1) experimental results show a reasonable fit to the Zipf model, and (2) the generalized Zipf distribution function offers analytical tractability [16].

B. Demand model

We define D_i as the average demand from all users for service of ISP_i. The demand D_i is a linear function of the strategies employed by all ISPs. The linearity means that the relationship between the demand and strategies is proportional, i.e., a change in strategy will result in a directly proportional change in demand. This simplifies the mathematical representation of the interactions between ISPs, making it easier to model their competitive behavior. However, this simplification is common in economic models to focus on strategic interactions without the complexities of nonlinear demand. In addtion, most paper in the literature model the demande as a liniear function ([36], [37], [16]). The demand D_i is influenced by various factors including price p_{s_i}, and QoS q_{s_i}. Additionally, the demand function is affected by the competitive landscape, taking into account the prices p_{s_j} and QoS q_{s_j}, $j \neq i$ offered by ISP_j rivals of ISP_i . Ultimately, D_i diminishes as the price p_{s_i} rise, and grows with an increase in the prices p_{s_j} for $j \neq i$. Conversely, it escalates with an enhancement in QoS q_{s_i}, and declines in response to an increase in QoS q_{s_j} for $j \neq i$.

Subsequently, the demand functions D_i are formulated as follows:

$$D_i = d_i - \alpha_i^i p_{s_i} + \beta_i^i q_{s_i} + \sum_{j=1, j\neq i}^{N} \left(\alpha_i^j p_{s_j} - \beta_i^j q_{s_j} \right) \tag{2}$$

where d_i represents the potential demand from users. Here, α_i^j and β_i^j are positive constants that denote, respectively, ISP_j sensitivity to the price and the QoS offered by ISP_j.

Assumption 1 *The sensitivity α verifies:*

$$\alpha_i^i \geq \sum_{j=1, j\neq i}^{N} \alpha_i^j \tag{3}$$

The sensitivity β verifies:

$$\beta_i^i \geq \sum_{j=1, j\neq i}^{N} \beta_i^j \tag{4}$$

Assumption 1 is essential to guarantee the uniqueness of the Nash equilibrium.

This paragraph delineates the economic interactions between advertisers and the ISP_i concerning the content stored in their cache. It discusses a scenario where there are M_{il} advertisers for item l, each operating with a predetermined budget B_{il} over a specified time frame (such as daily, weekly, or monthly). Additionally, advertisers assign a valuation v_{il} of item l, representing their willingness to pay for ISP_i advertisement slots. The valuation v_{il} of advertisers for item l is typically considered private information, making it difficult to ascertain. However, we make the assumption that we possess knowledge of the probability density function of advertisers' valuations, denoted as $x(v_{il})$, defined over the domain $[0, \bar{v}_{il}]$. Correspondingly, the cumulative distribution function is denoted as $X(v_{il})$. Additionally, we assume that the valuations of all advertisers are independent and identically distributed. Let $p_{a_{il}}$ represent the price charged per attention by the ISP_i for item l. We define $D_{a_{il}}$ as the demand for attention from advertisers to the ISP_i for item l. Therefore, $D_{a_{il}}$ can be represented as follows [38]:

$$D_{a_{il}} = \frac{M_{il} B_{il}}{p_{a_{il}}} prob(v_{il} \geq p_{a_{il}}) \tag{5}$$

with $prob(v_{il} \geq p_{a_{il}}) = 1 - X(p_{a_{il}})$

$$D_{a_{il}} = \frac{M_{il} B_{il}}{p_{a_{il}}} (1 - X(p_{a_{il}})) \tag{6}$$

We assume that v_{il} adheres to a uniform distribution within the $[0, \bar{v}_{il}]$ range. This means that every value within the range $[0, \bar{v}_{il}]$ is equally likely. In this case, v_{il} can take any value between 0 and $\bar{v}_{il}$, with no preference for one value over another. By assuming a uniform distribution, the model becomes easier to work with mathematically. Uniform distributions often lead to more straightforward calculations and analyses compared to more complex or skewed distributions.

This assumption leads us to derive the expression for the cumulative distribution function $X(p_{a_{il}})$ as

$$X(p_{a_{il}}) = \frac{p_{a_{il}}}{\bar{v}_{il}} \tag{7}$$

Subsequently, the demand $D_{a_{il}}$ takes the following form:

$$D_{a_{il}} = \frac{M_{il}B_{il}}{p_{a_{il}}}(1 - \frac{p_{a_{il}}}{\bar{v}_{il}}) \tag{8}$$

The optimal price $p_{a_{il}}$, is achieved when $D_{a_{il}} = (1 + \chi_{il}k_{il})\phi_l D_i$:

$$\sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})\phi_l D_i = \frac{M_{il}B_{il}}{p_{a_{il}}}\left(1 - \frac{p_{a_{il}}}{\bar{v}_{il}}\right) \tag{9}$$

Then

$$p_{a_{il}} = \frac{M_{il}B_{il}\bar{v}_{il}}{M_{il}B_{il} + \sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})\phi_l D_i \bar{v}_{il}} \tag{10}$$

C. Utility

The net profit for ISP_i is essentially the contrast between the total revenue and the fee paid:

$$\begin{aligned} U_i =& p_{s_i}\sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})D_i + \sum_{l=1}^{L_i}\phi_l[k_{il}p_{a_{il}}D_{a_{il}} \\ &- p_{t_i}D_i(1 - k_{il}) - C_{il}k_{il}(1 + \chi_{il}k_{il})D_i] \\ &- \gamma_i(L_i - \sum_{l=1}^{L_i}k_{il})\Psi_i \end{aligned} \tag{11}$$

where $p_{s_i}\sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})D_i$ is the income generated from network access. Recall that caching may serve as an incentive to increase content consumption. The new content demand, denoted as $(1 + \chi_{il}k_{il})D_i$, is influenced by the proportion of cached content. The term $\chi_{il}k_{il}$ represents the variation in demand for the content from the ISP cache. C_{il} represents the cost associated with caching item l. p_{t_i} is the fee that the ISP remits to the CP when seeking content, as the direction of the side-payment (from the ISP to the CP) in an ICN differs fundamentally from the current Internet model. $\sum_{l=1}^{L_i}\phi_l k_{il}p_{a_{il}}D_{a_{il}}$ is the revenue of ISP_i from advertisers. $\sum_{l=1}^{L_i}\phi_l p_{t_i}(1 + \chi_{il}k_{il})D_i(1 - k_{il})$ is the transmission fee. $C_{il}k_{il}(1 + \chi_{il}k_{il})D_i$ is the caching cost. γ_i represents the cost of unit backhaul bandwidth paid by ISP_i. Ψ_i denotes the backhaul bandwidth necessary to fulfill demand D_i. $(L_i - \sum_{l=1}^{L_i}k_{il})\Psi_i$ is the backhaul bandwidth indispensable for meeting the demand $(L_i - \sum_{l=1}^{L_i}k_{il})D_i$. We present the QoS as the expected delay, which is computed using the Kleinrock function corresponding to the delay in an M/M/1 queue with FIFO discipline or an M/G/1 queue under processor sharing, as described in [36]. Similar to [36], instead of using the actual delay, we consider its reciprocal of its square root :

$$q_{s_i} = \frac{1}{\sqrt{Delay_i}} = \sqrt{\Psi_i - \sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})D_i} \tag{12}$$

signifying that

$$\Psi_i = q_{s_i}^2 + \sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})D_i \tag{13}$$

By substituting equations (10) and (13) into equation (11), the utility of ISP_i becomes:

$$\begin{aligned} U_i =& p_{s_i}\sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})D_i + \sum_{l=1}^{L_i}\phi_l \\ &[\frac{B_{il}M_{il}\bar{v}_{il}k_{il}(1 + \chi_{il}k_{il})\phi_l D_i}{B_{il}M_{il} + \sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})\phi_l D_i \bar{v}_{il}} \\ &- p_{t_i}(1 + \chi_{il}k_{il})D_i(1 - k_{il}) - C_{il}k_{il}(1 + \chi_{il}k_{il})D_i] \\ &- \gamma_i(L_i - \sum_{l=1}^{L_i}k_{il})\left(q_{s_i}^2 + \sum_{l=1}^{L_i}(1 + \chi_{il}k_{il})D_i\right) \end{aligned} \tag{14}$$

IV. GAMES FORMULATION

The non-cooperative price QoS game is $G = [\mathcal{N}, \{\mathscr{P}_{s_i}, \mathscr{Q}_{s_i}\}, \{U_i(.)\}]$, where $\mathcal{N} = \{1, ..., N\}$ is the set of ISPs, $\mathscr{P}_{s_i}$ is the price strategy set of ISP_i and $\mathscr{Q}_{s_i}$ is the QoS strategy set of ISP_i. The strategy spaces $\mathscr{P}_{s_i}$ and $\mathscr{Q}_{s_i}$ are compact and convex sets. Thus, $\mathscr{P}_{s_i} = \left[\underline{p}_{s_i}, \overline{p}_{s_i}\right]$ and $\mathscr{Q}_{s_i} = \left[\underline{q}_{s_i}, \overline{q}_{s_i}\right]$. Let the price vector $\mathbf{p}_s = (p_{s_1}, ..., p_{s_N})^T \in \mathscr{P}_s^N = \mathscr{P}_{s_1} \times \mathscr{P}_{s_2} \times ... \times \mathscr{P}_{s_N}$, QoS vector $\mathbf{q}_s = (q_{s_1}, ..., q_{s_N})^T \in Q_s^N = \mathscr{Q}_{s_1} \times \mathscr{Q}_{s_2} \times ... \times \mathscr{Q}_{s_N}$.

Considering rationality of service providers, the Nash equilibrium concept is the natural concept solution of the non-cooperative price QoS game. We first will investigate the Nash equilibrium solution for the induced game as defined in the previous section. We will show that a Nash equilibrium solution exists and is unique by using the theory of concave games [39]. We recall that a non-cooperative game G is called concave if all players' utility functions are strictly concave with respect to their corresponding strategies [39].

According to [39] a Nash equilibrium exists in a concave game if the joint strategy space is compact and convex, and the utility function that any given player seeks to maximize is concave in its own strategy and continuous at every point in the product strategy space. Formally, if the weighted sum of the utility functions with non-negative weights:

$$\Xi = \sum_{i=1}^{N} x_i U_i \tag{15}$$

is diagonally strictly concave, this implies that the Nash equilibrium point is unique. The notion of diagonal strict concavity means that an individual user has more control over its utility function than the other users have on it, and

is proven using the pseudo-gradient of the weighted sum of utility functions, [39].

Acoriding to [40] if a concave game fulfills the dominance solvability condition:

$$-\frac{\partial^2 U_i}{\partial p_{s_i}^2} - \sum_{j=1,j\neq i}^{N} \left| \frac{\partial^2 U_i}{\partial p_{s_i} \partial p_{s_j}} \right| \geq 0 \tag{16}$$

the non-cooperative game G admits a unique Nash equilibrium.

A. Price game

The game G in price is defined for fixed $\mathbf{q}_s \in \mathscr{Q}_s$ as $G(\mathbf{q}_s) = [\mathscr{N}, \{\mathscr{P}_{s_i}\}, \{U_i(., \mathbf{q}_s)\}]$.

Definition 1 *A price vector $\mathbf{p}_s^* = (p_{s_1}^*, ..., p_{s_N}^*)$ is a Nash equilibrium of the game $G(\mathbf{q}_s)$ if:*

$$\forall (i, p_{s_i}) \in (\mathscr{N}, \mathscr{P}_{s_i}), U_i(p_{s_i}^*, \mathbf{p}_{s_{-i}}^*, \mathbf{q}_s) \geq U_i(p_{s_i}, \mathbf{p}_{s_{-i}}^*, \mathbf{q}_s)$$

Theorem 1 *For each $\mathbf{q}_s \in \mathscr{Q}_s$, the game $[\mathscr{N}, \{\mathscr{P}_{s_i}\}, \{U_i(., \mathbf{q}_s)\}]$ admit a unique Nash Equilibrium.*

The second order derivative of the utility with respect to the prices is :

$$\frac{\partial^2 U_i}{\partial p_{s_i}^2} = -2\alpha_i^i \sum_{l=1}^{L_i} (1 + \chi_{il} k_{il}) - \sum_{l=1}^{L_i} \phi_l \left[\frac{2(\alpha_i^i)^2 B_{il}^2 M_{il}^2 \overline{v}_{il}^2 k_{il}^2 (1 + \chi_{il} k_{il})^2 \phi_l^2}{\left(B_{il} M_{il} + \sum_{l=1}^{L_i} (1 + \chi_{il} k_{il}) \phi_l D_i \overline{v}_{il} \right)^3} \right] \leq 0 \tag{17}$$

The second derivative of U_i is consistently negative, it signifies the concavity of U_i and, consequently, assures the existence of a Nash equilibrium point within the game $G(\mathbf{q}_s)$.

We rely on the following proposition, applicable to concave games [40]: In the case where a concave game fulfills the dominance solvability condition:

$$-\frac{\partial^2 U_i}{\partial p_{s_i}^2} - \sum_{j=1,j\neq i}^{N} \left| \frac{\partial^2 U_i}{\partial p_{s_i} \partial p_{s_j}} \right| \geq 0 \tag{18}$$

the game $G(\mathbf{q}_s)$ admits a unique Nash equilibrium.

The mixed partial is written as:

$$\frac{\partial^2 U_i}{\partial p_{s_i} \partial p_{s_j}} = \alpha_i^j \sum_{l=1}^{L_i} (1 + \chi_{il} k_{il}) + \sum_{l=1}^{L_i} \phi_l \left[\frac{2\alpha_i^i \alpha_i^j B_{il}^2 M_{il}^2 \overline{v}_{il}^2 k_{il}^2 (1 + \chi_{il} k_{il})^2 \phi_l^2}{\left(B_{il} M_{il} + \sum_{l=1}^{L_i} (1 + \chi_{il} k_{il}) \phi_l D_i \overline{v}_{il} \right)^3} \right] \geq 0 \tag{19}$$

Then,

$$-\frac{\partial^2 U_i}{\partial p_{s_i}^2} - \sum_{j=1,j\neq i}^{N} \left| \frac{\partial^2 U_i}{\partial p_{s_i} \partial p_{s_j}} \right| = (2\alpha_i^i - \sum_{j=1,j\neq i}^{N} \alpha_i^j) \times \sum_{l=1}^{L_i} (1 + \chi_{il} k_{il}) + (\alpha_i^i - \sum_{j=1,j\neq i}^{N} \alpha_i^j) \times \sum_{l=1}^{L_i} \phi_l \left[\frac{2(\alpha_i^i) B_{il}^2 M_{il}^2 \overline{v}_{il}^2 k_{il}^2 (1 + \chi_{il} k_{il})^2 \phi_l^2}{\left(B_{il} M_{il} + \sum_{l=1}^{L_i} (1 + \chi_{il} k_{il}) \phi_l D_i \overline{v}_{il} \right)^3} \right] \geq 0 \tag{20}$$

Thus, the game $G(\mathbf{q}_s)$ admits a unique Nash equilibrium. □

B. QoS game

The game G in QoS is defined for fixed $\mathbf{p}_s \in \mathscr{P}_s$ as $G(\mathbf{p}_s) = [\mathscr{N}, \{\mathscr{Q}_{s_i}\}, \{U_i(\mathbf{p}_s, .)\}]$.

Definition 2 *A QoS vector $\mathbf{q}_s^* = (q_{s_1}^*, ..., q_{s_N}^*)$ is a Nash equilibrium of the game $G(\mathbf{p}_s)$ if:*

$$\forall (i, q_{s_i}) \in (\mathscr{N}, \mathscr{Q}_{s_i}), U_i(\mathbf{p}_s, q_{s_i}^*, \mathbf{q}_{s_{-i}}^*) \geq U_i(\mathbf{p}_s, q_{s_i}, \mathbf{q}_{s_{-i}}^*)$$

Theorem 2 *For each $\mathbf{p}_s \in \mathscr{P}_s$, the game $[\mathscr{N}, \{\mathscr{Q}_{s_i}\}, \{U_i(\mathbf{p}_s, .)\}]$ admit a unique Nash Equilibrium.*

The second order derivative of the utility with respect to the prices is :

$$\frac{\partial^2 U_i}{\partial q_{s_i}^2} = -2\gamma_i (L_i - \sum_{l=1}^{L_i} k_{il}) \leq 0 \tag{21}$$

The second derivative is negative, it signifies the concavity of U_i and, consequently, assures the existence of a Nash equilibrium point within the game $G(\mathbf{p})$.

To demonstrate uniqueness, we adopt the approach outlined in [39] and establish the weighted sum of utility:

$$\vartheta(q_{s_k}, x) = \left[x_1 \nabla U_1 \left(q_{s_1}, q_{s_{-1}} \right), ..., x_N \nabla U_N \left(q_{s_N}, q_{s_{-N}} \right) \right]^T \tag{22}$$

The pseudo-gradient of (22) is

$$\Xi(q_s, x) = \sum_{i=1}^{N} x_i U_i \left(q_{s_i}, q_{s_{-i}} \right) \tag{23}$$

The Jacobian matrix J of the pseudo-gradient (23) is

$$\mathbb{J} = \begin{bmatrix} x_1 \frac{\partial^2 U_1}{\partial q_{s_1}^2} & x_1 \frac{\partial^2 U_1}{\partial q_{s_1} \partial q_{s_2}} & \cdots & x_1 \frac{\partial^2 U_1}{\partial q_{s_1} \partial q_{s_N}} \\ x_2 \frac{\partial^2 U_2}{\partial q_{s_2} \partial q_{s_1}} & x_2 \frac{\partial^2 U_2}{\partial q_{s_2}^2} & \cdots & x_2 \frac{\partial^2 U_2}{\partial q_{s_2} \partial q_{s_N}} \\ \vdots & \vdots & \ddots & \vdots \\ x_N \frac{\partial^2 U_N}{\partial q_{s_1} \partial q_{s_2}} & x_N \frac{\partial^2 U_N}{\partial q_{s_N} \partial q_{s_2}} & \cdots & x_N \frac{\partial^2 U_N}{\partial q_{s_N}^2} \end{bmatrix} \tag{24}$$

Then

$$\mathbb{J} = \begin{bmatrix} -2r_1 & 0 & \cdots & 0 \\ 0 & -2r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -2r_N \end{bmatrix} \tag{25}$$

where $r_i = \gamma_i(L_i - \sum_{l=1}^{L_i} k_{il})$.

We have J is negative definite and according to [39], the functions $\psi(q_s, x)$ is diagonally strictly concave. Consequently, the uniqueness of the Nash equilibrium point is established. □

C. Learning Nash equilibrium

The best response algorithm, as outlined in [41], operates through a series of rounds. In each round, starting from the second onwards, each ISPs observes the prices (resp QoS) selected by its adversaries in prior rounds. Subsequently, the ISP incorporates these observed prices into its decision-making process to adjust its own price. Algorithm 1 provides a concise summary of the steps involved in the best response learning, guiding each ISPs towards finding the Nash equilibrium.

Algorithm 1 Best response Algorithm

1: Initialize vectors $\rho(0) = [\rho_1(0), ..., \rho_N(0)]$ randomly;
2: **For each** ISP_i at time instant t computes:
- $\rho_i(t+1) = \underset{\rho_i \in \mathscr{W}_i}{\text{argmax}} \, (U_i(\rho_i(t)))$.

3: **If** $\forall i \in \mathscr{N}$, $|\rho_i(t+1) - \rho_i(t)| < \epsilon$, then STOP.
4: **Else**, $t \leftarrow t+1$ and go to step (2).

Such as:

- ρ denotes the vector price $\mathbf{p}_s = (p_{s_1}, ..., p_{s_N})$ or vector QoS $\mathbf{q}_s = (q_{s_1}, ..., q_{s_N})$.
- $\mathscr{W}_i$ denotes the policy profile price $\mathscr{P}_{s_i}$ or policy profile QoS $\mathscr{Q}_{s_i}$.

V. NUMERICAL ILLUSTRATION

In this section, we shift our focus towards leveraging our analytical insights. Our strategy involves conducting a numerical analysis of the gaming market, incorporating the previously identified utility functions of the ISPs. To demonstrate, we consider a network scenario that includes three ISPs aiming to maximize their profits. Tables II and III present the system parameter values and environment parameter settings considered in this numerical study.

TABLE II
PARAMETERS SETTING USED FOR NUMERICAL EXAMPLES.

$\alpha_1^1 = \alpha_2^2 = \alpha_3^3$	$\alpha_i^j,\ i \neq j$	$\beta_1^1 = \beta_2^2 = \beta_3^3$	β_i^j
0.7	0.3	0.7	0.3
$\overline{p}_{s_1} = \overline{p}_{s_2} = \overline{p}_{s_3}$	$\underline{p}_{s_1} = \underline{p}_{s_2} = \underline{p}_{s_3}$	$L_1 = L_2 = L_3$	d_1
1000	1	35	200
$\overline{q}_{s_1} = \overline{q}_{s_2} = \overline{q}_{s_3}$	$\underline{q}_{s_1} = \underline{q}_{s_2} = \underline{q}_{s_3}$	$M_1 = M_2 = M_3$	d_2
1000	1	18	250
$B_1 = B_2 = B_3$	$\overline{v}_1 = \overline{v}_2 = \overline{v}_3$	$\gamma_1 = \gamma_2 = \gamma_3$	d_3
1000	300	5	300
$C_1 = C_2 = C_3$	$pt_1 = pt_2 = pt_3$	$\chi_1 = \chi_2 = \chi_3$	η
8	10	4	0.8
$ps_1 = ps_2 = ps_3$	$qs_1 = qs_2 = qs_3$	N	
300	250	3	

TABLE III
ENVIRONMENT PARAMETER SETTINGS.

Configuration	Parameters
Operating System	Windows 10 Pro 64-bit
Processor	Intel(R) Core(TM) i5-6300U CPU @ 2.40GHz, 4 cores
Memory	12 GB
Software	MATLAB R2020b.

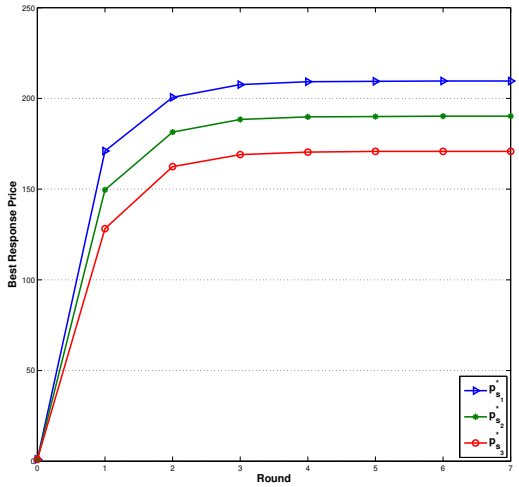


Fig. 2. Achieving price at the Nash equilibrium point.

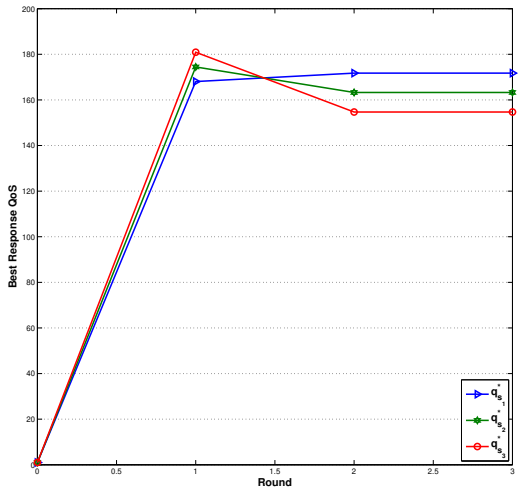


Fig. 3. Achieving QoS at the Nash equilibrium point.

Figures 2 and 3 illustrate the swift convergence towards Nash equilibrium in terms of price and QoS. The rapid

convergence indicated by the low number of iterations required highlights the efficiency of the system in stabilizing. This swift adjustment suggests that ISPs quickly adapt their strategies to reach a stable point where no single ISP can benefit by unilaterally changing its strategy. Moreover, the presence and uniqueness of the Nash equilibrium point confirm that the system settles at a stable state where all ISPs' strategies are mutually optimal. This stability is crucial for predicting long-term outcomes and for strategic decision-making in competitive scenarios.

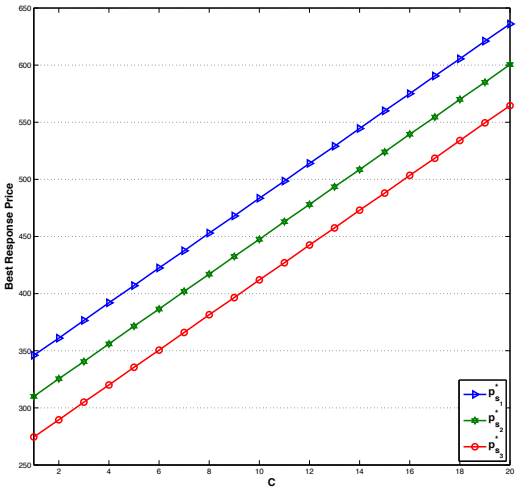


Fig. 4. Equilibrium price as a function of the caching cost.

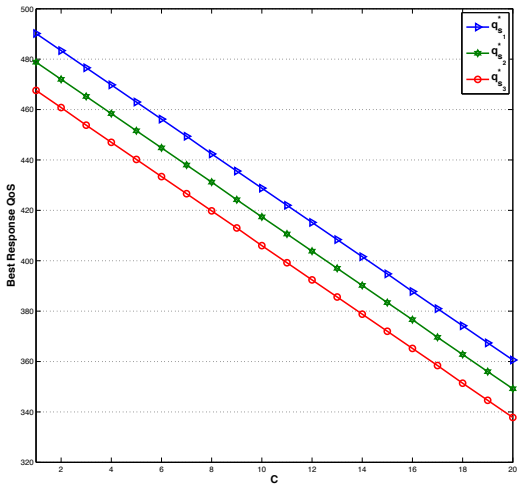


Fig. 5. Equilibrium QoS as a function of the caching cost.

Figures 4 and 5 provide insights into how Nash equilibrium price and QoS are influenced by rising caching costs for two ISPs. As the cost of caching content increases, ISPs tend to raise their prices while simultaneously reducing QoS. This trend arises because the higher expenses associated with caching make it less financially viable for ISPs to maintain extensive caching. Consequently, ISPs are compelled to forward more content requests to CP, which leads to higher transmission fees and greater backhaul bandwidth costs. To offset these increased costs, ISPs increase their prices. Moreover, to manage the additional burden of backhaul bandwidth expenses, ISPs lower the QoS they offer. These results underscore the financial strain that increased caching costs impose on ISPs, driving them to adjust their pricing strategies and QoS in response. The figures highlight the delicate balance ISPs must maintain between caching costs, pricing, and QoS to manage operational expenses and remain competitive.

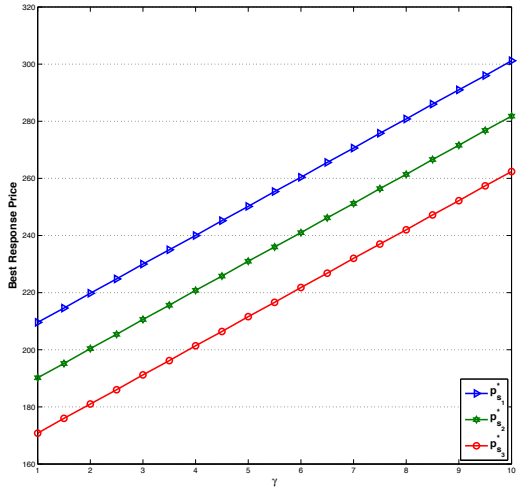


Fig. 6. Equilibrium price as a function of the bandwidth cost.

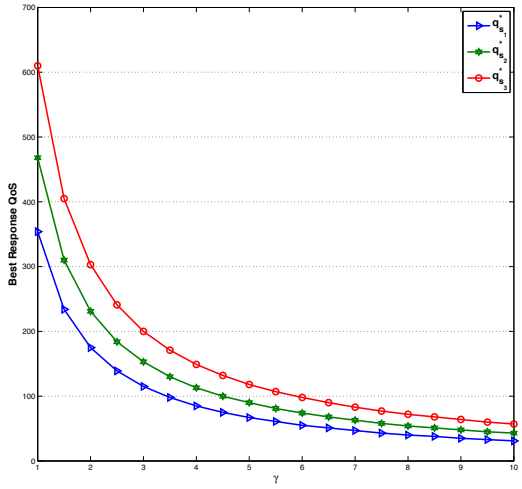


Fig. 7. Equilibrium QoS as a function of the bandwidth cost.

In Figures 6 and 7, we present the impact of bandwidth cost on price and QoS at Nash equilibrium. It is observed that the equilibrium price for both ISPs increases in response to higher bandwidth costs. Conversely, the equilibrium QoS for all ISPs decreases as the bandwidth cost rises. When the network owner sets a lower cost for bandwidth, ISPs tend to invest in more bandwidth to offer improved QoS and competitive pricing. However, as bandwidth costs escalate, ISPs opt to elevate their prices and reduce QoS to offset the increased cost of bandwidth.

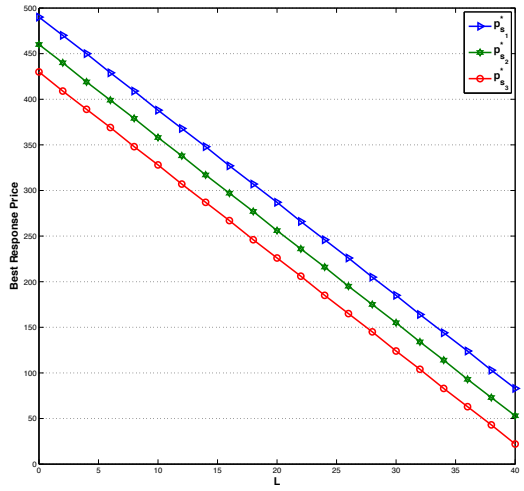


Fig. 8. Equilibrium price as a function of the number of cached items.

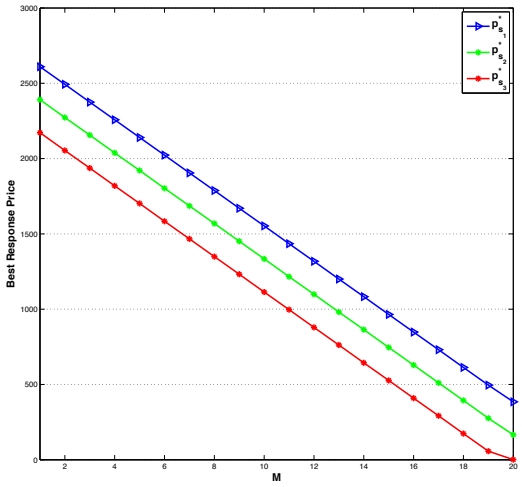


Fig. 10. Equilibrium price as a function of the quantity of advertisers.

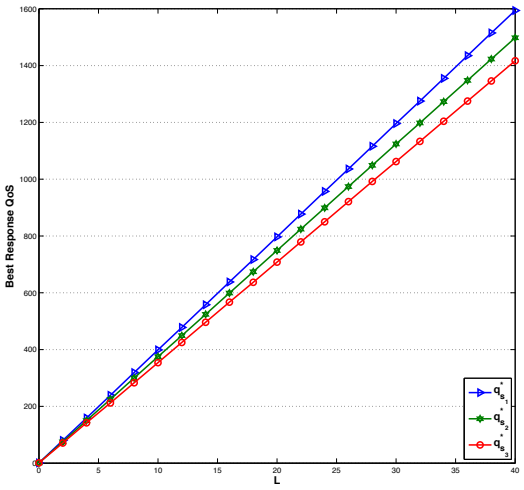


Fig. 9. Equilibrium QoS as a function of the number of cached items.

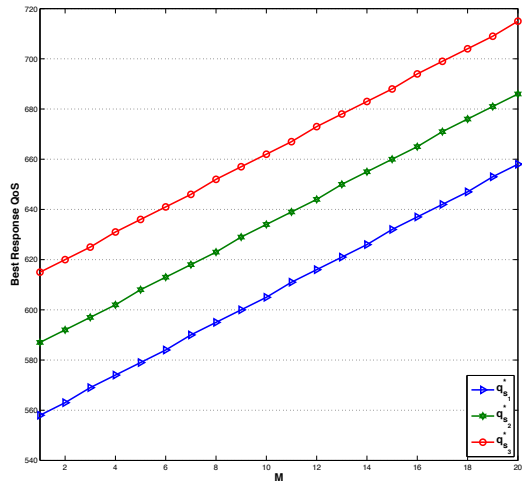


Fig. 11. Equilibrium QoS as a function of the quantity of advertisers.

Figures 8 and 9 reveal a significant relationship between the number of cached items, network access price, and QoS. As the number of items cached by ISPs increases, both the network access price and QoS improve. This is because a larger cache allows a greater proportion of content requests to be fulfilled directly from the ISP's cache, leading to lower transmission fees and reduced bandwidth costs. With these cost reductions, ISPs see an increase in revenue, which enables them to lower their access prices. The lower prices, in turn, attract more users, further boosting demand. Additionally, the enhanced QoS resulting from a larger cache improves user satisfaction and engagement. These findings underscore how increasing the cache size can create a positive feedback loop: it reduces costs, allows for lower pricing, improves QoS, and ultimately drives greater user demand and revenue for ISPs.

Figures 10 and 11 illustrate the impact of the number of advertisers on QoS and pricing for ISPs. As the number of advertisers increases, both QoS improves and prices decrease. This trend occurs because a higher number of advertisers generates more advertising revenue for the ISPs. With increased income from ads, ISPs are able to invest more in enhancing their QoS, leading to better QoS for users. Additionally, the increased revenue enables ISPs to lower their prices, making their services more attractive to users. This relationship aligns with real-world scenarios where greater advertising revenue provides ISPs with the financial flexibility to improve QoS and adjust pricing strategies to boost user demand and competitiveness.

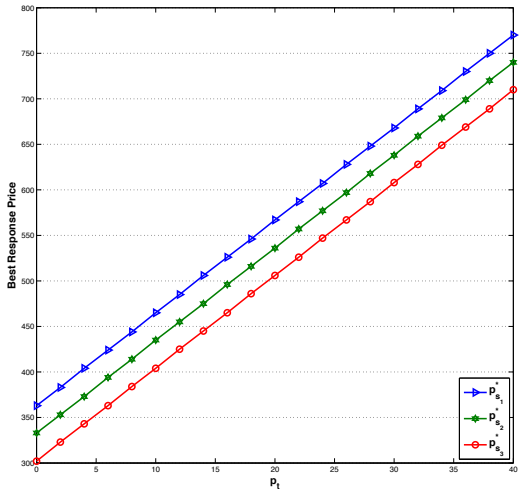


Fig. 12. Equilibrium price as a function of the transmission price.

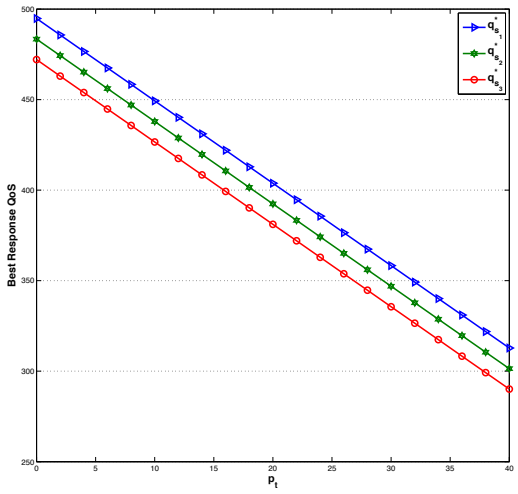


Fig. 13. Equilibrium QoS as a function of the transmission price.

Figures 12 and 13 illustrate the impact of transmission price on both the pricing and QoS provided by three ISPs. The data show that as the transmission price rises, ISPs reduce their prices and improve QoS. This counterintuitive outcome occurs because higher transmission prices lead to increased revenue from transmission fees, which boosts the ISPs' overall income. With greater financial resources, ISPs can afford to lower their access prices, making their services more attractive to consumers. Simultaneously, they invest more in bandwidth and infrastructure to enhance QoS. This dual strategy aims to maximize user demand by providing better QoS at lower prices, demonstrating how revenue from transmission fees can be leveraged to benefit users and improve competitive positioning.

VI. Conclusion

In this paper, we established an analytical framework to study the distribution of accessible content within an ICN environment encompassing multiple ISPs and advertisers. The interactions among ISPs were examined through a non-cooperative game, where each ISP has control over the amount of free content cached in the network and can adjust its strategies, including network access price and QoS. We assumed that the popularity of content adheres to a generalized Zipf distribution. We rigorously proved the existence and uniqueness of the Nash equilibrium in a competitive ICN market. This finding is significant as it suggests that a stable solution, accompanied by suitable economic incentives for collaborative caching, is attainable within the ICN paradigm. Furthermore, we outlined a learning mechanism enabling each ISP to swiftly and accurately discover its equilibrium policies. Our simulation results underscored the benefits of caching investment for both ISPs and end-users, demonstrating the efficacy of our proposed approach in fostering mutually advantageous outcomes within the ICN ecosystem. Results from this work can be further extended to more general network scenarios, particularly when considering non-linear demand and when valuation follows a normal distribution.

References

[1] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse?," in *Passive and Active Network Measurement: 9th International Conference, PAM 2008, Cleveland, OH, USA, April 29-30, 2008. Proceedings 9*, pp. 1–10, Springer, 2008, DOI: 10.1007/978-3-540-79232-1_1.

[2] M. Mosko, "Labeled content information draft-mosko-icnrg-ccnxlabeledcontent-0 1," *IETF draft*, 2015.

[3] A. J. Abu, B. Bensaou, and J. M. Wang, "Interest packets retransmission in lossy ccn networks and its impact on network performance," in *Proceedings of the 1st ACM Conference on Information-Centric Networking*, pp. 167–176, 2014, DOI: 10.1145/2660129.266014.

[4] R. Halloush, H. Liu, L. Dong, M. Wu, and H. Radha, "Hop-by-hop content distribution with network coding in multihop wireless networks," D*igital Communications and Networks*, vol. 3, no. 1, pp. 47–54, 2017, DOI: 10.1016/j.comcom.2016.12.002.

[5] M. Urueña, I. Soto, I. Martinez-Yelmo, and M. Calderon, "Effect of content popularity, number of contents and a cellular backup network on the performance of content distribution protocols in urban vanet scenarios," *Computer Communications*, vol. 99, pp. 13–23, 2017, DOI: 10.1016/j.comcom.2016.12.002.

[6] M. Elkotob and K. Andersson, "Challenges and opportunities in content distribution networks: A case study," in *2012 IEEE Globecom Workshops*, pp. 1021–1026, IEEE, 2012, DOI: 10.1109/GLOCOMW.2012.6477717.

[7] B.-N. Cheng, G. Kuperman, P. Deutsch, L. Mercer, and A. Narula-Tam, "Group-centric networking: addressing information sharing requirements at the tactical edge," *IEEE Communications Magazine*, vol. 54, no. 10, pp. 145–151, 2016, DOI: 10.1109/MCOM.2016.7588284.

[8] M. Mangili, F. Martignon, and A. Capone, "Performance analysis of content-centric and content-delivery networks with evolving object popularity," *Computer Networks*, vol. 94, pp. 80–98, 2016, DOI: 10.1016/j.comnet.2015.11.019.

[9] E. Borgia, R. Bruno, and A. Passarella, "Reliable data delivery in icn-iot environments," *Future Generation Computer Systems*, vol. 134, pp. 271– 286, 2022, DOI: 10.1016/j.future.2022.04.004.

[10] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 181–192, 2007, DOI: 10.1145/1282380.1282402.

[11] G. Zhang, Y. Li, and T. Lin, "Caching in information centric networking: A survey," *Computer networks*, vol. 57, no. 16, pp. 3128–3141, 2013, DOI: 10.1016/j.comnet.2013.07.007.

[12] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos, "A survey of information-centric networking research," *IEEE communications surveys & tutorials*, vol. 16, no. 2, pp. 1024–1049, 2013, DOI: 10.1109/SURV.2013.070813.00063.

[13] J. M. Caldeira, J. J. Rodrigues, P. Lorenz, and S. Ullah, "Impact of sensor nodes scaling and velocity on handover mechanisms for healthcare wireless sensor networks with mobility support," *Computers in Industry*, vol. 69, pp. 92–104, 2015, DOI: 10.1016/j.compind.2014.09.002.

[14] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache "less for more" in information-centric networks (extended version)," *Computer Communications*, vol. 36, no. 7, pp. 758–770, 2013, DOI: 10.1016/j.comcom.2013.01.007.

[15] H. Lee and A. Nakao, "User-assisted in-network caching in information-centric networking," *Computer Networks*, vol. 57, no. 16, pp. 3142–3153, 2013, DOI: 10.1016/j.comnet.2013.07.008.

[16] M. Hajimirsadeghi, N. B. Mandayam, and A. Reznik, "Joint caching and pricing strategies for popular content in information centric networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 3, pp. 654–667, 2017, DOI: 10.1109/JSAC.2017.2672161.

[17] V. G. Douros, S. E. Elayoubi, E. Altman, and Y. Hayel, "Caching games between content providers and internet service providers," *Performance Evaluation*, vol. 113, pp. 13–25, 2017, DOI: 10.1016/j.peva.2017.04.006.

[18] H. Garmani, D. Ait Omar, M. El Amrani, M. Baslam, and M. Jourhmane, "The effect of caching on cp and isp policies in information-centric networks," *Mobile Information Systems*, vol. 2020, pp. 1–18, 2020, DOI: 10.1155/2020/8895271.

[19] W. Ur Rehman, M. M. A.-E. Sufyan, T. Salam, A. Al-Salehi, Q. E. Ali, and A. H. Malik, "Cooperative distributed uplink cache over b5g small cell networks," *Plos one*, vol. 19, no. 4, p. e0299690, 2024, DOI: 10.1371/journal.pone.0299690.

[20] C. Kan, H. Wu, L. Xing, and H. Ma, "Cooperative caching strategy based mobile vehicle social-aware in internet of vehicles," *Transactions on Emerging Telecommunications Technologies*, vol. 34, no. 7, p. e4792, 2023, DOI: 10.1002/ett.4792.

[21] A. Miglani and N. Kumar, "Maximized content offloading for content centric v2g networks," in *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, pp. 318–323, IEEE, 2024, DOI: 10.1109/CCNC51664.2024.10454759.

[22] C. Ling, W. Zhang, Q. Fan, Z. Feng, J. Wang, R. Yadav, and D. Wang, "Cooperative service caching in vehicular edge computing networks based on transportation correlation analysis," *IEEE Internet of Things Journal*, 2024, DOI: 10.1109/JIOT.2024.3382723.

[23] P. Wang, S. Li, G. Sun, C. Zhou, C. Gao, S. Qiu, T. Tao, and Q. Zhang, "Labeled graph partitioning scheme for distributed edge caching," *Future Generation Computer Systems*, vol. 153, pp. 492–504, 2024, DOI: 10.1016/j.future.2023.12.008.

[24] Y. Wang, J. Chen, et al., "Collaborative caching in edge computing via federated learning and deep reinforcement learning," *Wireless Communications and Mobile Computing*, vol. 2022, 2022, DOI: 10.1155/2022/7212984.

[25] D. Kim, D.-w. Seo, and M. Choi, "Edge caching and computing of video chunks in multi-tier wireless networks," *Journal of Network and Computer Applications*, p. 103889, 2024n DOI: 10.1016/j.jnca.2024.103889.

[26] A. Miglani and N. Kumar, "A blockchain based matching game for content sharing in content-centric vehicle-to-grid network scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2023, DOI: 10.1109/TITS.2023.3322826.

[27] F. Khandaker, W. Li, S. Oteafy, and H. Hassanein, "Reverse auction-based dynamic caching and pricing scheme in producer-driven icn," in *ICC 2023-IEEE International Conference on Communications*, pp. 119–124, IEEE, 2023, DOI: 10.1109/ICC45041.2023.10278905.

[28] B. Assila, A. Kobbane, and J. Ben-Othman, "Improving caching resource management: A pricing economic approach using cournot, bertrand, and stackelberg game models," *International Journal of Communication Systems*, vol. 36, no. 1, p. e4358, 2023, DOI: 10.1002/dac.4358.

[29] A. B. Rahman, P. Charatsaris, E. E. Tsiropoulou, and S. Papavassiliou, "Information-centric networking cache memory allocation: A network economics approach," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*, pp. 1259–1264, IEEE, 2023, DOI: 10.1109/GLOBECOM54140.2023.10437315.

[30] Q. Zheng, J. Lin, W. Yan, Z. Xu, Q. Shi, and X. Tan, "Research on icn caching and pricing strategies under the package billing model," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pp. 4667–4672, IEEE, 2022, DOI: 10.1109/GLOBECOM48099.2022.10000995.

[31] T.-M. Pham, "Analysis of isp caching in information-centric networks," in *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future (RIVF)*, pp. 151–156, IEEE, 2015, DOI: 10.1109/RIVF.2015.7049891.

[32] M. El Amrani, H. Garmani, D. Ait Omar, M. Baslam, B. Minaoui, et al., "Analysis of the chaotic dynamics duopoly game of isps bounded rational," *Discrete Dynamics in Nature and Society*, vol. 2022, 2022, DOI: 10.1155/2022/6632993.

[33] J. Nie, Q. Zheng, F. Yang, Z. Xu, and Q. Shi, "Cache pricing mechanism for icn on isp peering," in *2022 5th International Conference on Hot Information-Centric Networking (HotICN)*, pp. 13–18, IEEE, 2022, DOI: 10.1109/HotICN57539.2022.10036171.

[34] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pp. 1–14, 2007, DOI: 10.1145/1298306.1298309.

[35] C. Fricker, P. Robert, J. Roberts, and N. Sbihi, "Impact of traffic mix on caching performance in a content-centric network," in *2012 Proceedings IEEE INFOCOM Workshops*, pp. 310–315, IEEE, 2012, DOI: 10.1109/INFCOMW.2012.6193511.

[36] R. El Azouzi, E. Altman, and L. Wynter, "Telecommunications network equilibrium with price and quality-of-service characteristics," in T*eletraffic science and engineering*, vol. 5, pp. 369–378, Elsevier, 2003, DOI: 10.1016/S1388-3437(03)80182-8.

[37] E. Altman, A. Legout, and Y. Xu, "Network non-neutrality debate: An economic analysis," in *NETWORKING 2011: 10th International IFIP TC 6 Networking Conference, Valencia, Spain, May 9-13, 2011, Proceedings, Part II 10*, pp. 68–81, Springer, 2011, DOI: 10.1007/978-3-642-20798-3_6.

[38] J. J. Liu and D. M. Chiu, "Mathematical modeling of competition in sponsored search market," in *Proceedings of the 2010 Workshop on Economics of Networks, Systems, and Computation*, pp. 1–6, 2010, DOI: 10.1145/1879082.18790.

[39] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica: Journal of the Econometric Society*, pp. 520–534, 1965, DOI: 10.2307/1911749.

[40] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 41–52, 2009, DOI: 10.1109/MSP.2009.933496.

[41] D. Ait Omar, M. Outanoute, M. Baslam, M. Fakir, and B. Bouikhalne, "Joint Price and QoS Competition with Bounded Rational Customers," in *Networked Systems: 5th International Conference, NETYS 2017, Marrakech, Morocco, May 17-19, 2017, Proceedings* (A. El Abbadi and B. Garbinato, eds.), pp. 457–471, 2017, DOI: 10.1007/978-3-319-59647-1_33.

Hamid Garmani is a Professor of computer science in the Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Morocco. His research interests include network economics, network security, applications of game theory in wireless networks, and radio resource management.

Mohamed El Amrani is a Professor of computer science in the Polydisciplinary Faculty, Sultan Moulay Slimane University, Morocco. His research interests include network economics, network security, applications of game theory in wireless networks, and radio resource management.

Driss Ait Omar is a Professor of computer science in the Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Morocco. His research interests include network economics, network security, applications of game theory in wireless networks, and radio resource management.

Mohamed Baslam is a Professor of computer science in the Faculty of Sciences and Technology, Sultan Moulay Slimane University, Morocco. His current research interests include performance evaluation and optimization of networks based on game-theoretic and queuing models, applications in communication/transportation and social networks, such as wireless flexible networks, bio-inspired and self-organizing networks, and economic models of the Internet and yield management.

Hicham Zougagh is a Professor of computer science in the Faculty of Sciences and Technology, Sultan Moulay Slimane University, Morocco. His current research interests include the security, routing protocols, QOS in ad hoc networks, sensor networks, Vanets, IoT and IA, and network economics.

Towards Automated Musical Anamnesis for Music-based Intervention in Dementia Patients

Fabian Simmank, Katarzyna Grebosz-Haring, Thomas Ballhausen, Christian Thomay, Martin Biallas, and Markus Tauber

***Abstract*—Dementia is a neurodegenerative disease affecting millions worldwide, leading to cognitive decline and difficulties in daily activities. Music-based interventions offer a promising, cost-effective, non-pharmacological approach to improving quality of life for people with dementia. However, understanding both preferred and familiar music, as well as individual music affinity, is crucial to avoid overstimulation and ensure meaningful engagement. Developing a protocol for musical anamnesis, which gathers a patient's musical history and hearing health, demands significant manual effort and expertise, limiting its scalability. An automated approach could enhance the sustainability of music-based therapy by reducing therapist time while maintaining relevance and preference evaluation. Here, we introduce Automated Musical Anamnesis (AMA), a personalized, scalable intervention combining interdisciplinary methods to support people with dementia.**

***Index Terms*—Dementia, P4 Medicine, Music Therapy, Digital Therapeutics, Music information retrieval**

I. Introduction

Dementia is a neurodegenerative condition that affects millions worldwide, leading to memory loss, cognitive decline, and emotional dysregulation, which significantly reduce well-being and quality of life while increasing family caregiver distress [1]. Due to the limited efficacy and potential side effects of medications, non-pharmacological, effective, scalable, and feasible interventions are now a priority. Music-based interventions have emerged as a promising, cost-effective, non-pharmacological approach for dementia patients by tapping into preserved musical memories [2], [3]. These interventions encompass various approaches, including music therapy, where licensed music therapists design programs involving active participation (e.g., instrumental play, singing) or receptive engagement (e.g., listening). Another prevalent practice involves the independent use of pre-recorded music (music listening intervention) due to its accessibility and cost-effectiveness. Music listening interventions are not only highly accepted but also easily implemented. However, studies suggest that such interventions must meet at least two prerequisites for optimal application. First, they must engage individual autobiographical musical memories, which requires knowledge of both preferred and familiar music that can elicit such memories. Second, and perhaps more importantly, an individual's affinity for music listening must be assessed to ensure that under- or overexposure to music stimulation is avoided, thereby guaranteeing high-quality listening experiences. Additionally, with age-related hearing decline and the prevalence of conditions such as tinnitus, hyperacusis, or hearing loss—along with the potential use of hearing aids—extra care must be taken to tailor music programs to the personal needs of each patient.

Developing a successful protocol for musical anamnesis, which involves gathering a patient's musical history and hearing health, demands substantial manual effort and expertise, making it challenging to ensure scalability. Traditional anamnesis methods rely on direct interviews or questionnaires [4], [5]. Consequently, there is considerable potential for improvement in this area, leveraging both therapeutic expertise and advancements in digital technology. Automating the process of musical anamnesis can facilitate the collection and analysis of vast amounts of data, incorporating personal preferences, past musical experiences, and associated memories. This can lead to a more precise and faster understanding of an individual's musical preferences, enabling tailored therapeutic interventions. Furthermore, automating the anamnesis process can enhance personalization, adapting music-based therapy to the specific needs and preferences of each dementia patient while improving accuracy by minimizing human error and bias.

Automated systems can analyze extensive musical databases and generate personalized playlists or musical interventions aligned with an individual's background and emotional connections to music. This level of personalization has the potential to strengthen therapeutic effects and increase patient engagement while improving accessibility to music-based interventions for individuals with dementia, their family caregivers, and healthcare professionals.

This paper outlines the potential of Automated Musical Anamnesis (AMA) and demonstrates how AMA can contribute to the effectiveness and sustainability of music-based adjuvant therapies for dementia. Beyond establishing this position, the paper aims to advance the field by presenting the

Fabian Simmank is with Research Studios Austria FG, Vienna, Austria, (e-mail: fabian.simmank@researchstudio.at)

Katarzyna Grebosz-Haring ist with Mozarteum University Salzburg und Paris Lodron University Salzburg, Interuniversity Organisation Science & Art, Austria (e-mail: katarzyna.grebosz-haring@plus.ac.at)

Thomas Ballhausen is with Mozarteum University Salzburg, Institute for Open Arts, Austria (e-mail: thomas.ballhausen@moz.ac.at)

Christian Thomay is with Research Studios Austria FG, Vienna, Austria, (e-mail: christian.thomay@researchstudio.at)

Martin Biallas is with Lucerne University of Applied Sciences and Arts, School of Engineering and Architecture, Switzerland (e-mail: martin.biallas@hslu.ch)

Markus Tauber is with Research Studios Austria FG, Salzburg, Austria, (e-mail: markus.tauber@researchstudio.at)

Manuscript received 31.01.2024
revised Manuscript received 25.06.2024

DOI: 10.36244/ICJ.2024.4.5

state of the art in music-based interventions while proposing a novel combination of methods and tool support, detailed in subsequent sections. Within this scope, the paper contributes to the paradigm of P4 medicine (predict, prevent, personalize, and participate) by emphasizing a personalized approach in flexible environments and a data-driven foundation, potentially enabling early detection and mitigation of symptoms in the long run. The emergence of P4 medicine, which relies on advanced sensors and low-cost digital instruments to build a new healthcare ecosystem, presents a significant opportunity with vast market potential for the European electronic components and systems industry. This includes over 25,000 SME MedTech companies across Europe and forms part of the strategic research agenda of the European ECS Industry Associations [6].

Taken together, this convergence of various fields and technological approaches—coined as a new digital reality—highlights a compelling case for infocommunications [7]. The paper is structured as follows: Section II and Section III review the state of the art in musical interventions for dementia and musical anamnesis, respectively. Section IV discusses relevant psychophysiological parameters and the sensors used to measure them. Section V presents the proposed computational modeling approach. Section VI explores the tool support from a holistic user perspective. Finally, Section VII highlights the societal relevance of the solution.

II. Current Status of Musical Intervention for Dementia

Dementia encompasses various degenerative and chronically progressive brain disorders that result in memory impairments as well as behavioral and psychological disturbances, often accompanied by high comorbidity with conditions such as depression and agitation.

Previous research suggests that music-based interventions are associated with positive outcomes at both psychological and physical levels. For instance, these interventions can elicit positive emotions [8], [9], uplift mood [10], and induce both arousing and relaxing experiences [11]–[13]. Furthermore, they have been shown to alleviate stress and anxiety and are linked to reduced cortisol levels [14]–[16]. These findings have inspired the development of practical treatment programs for adjuvant therapy targeting a broader spectrum of psychosomatic conditions and neurodegenerative illnesses [17].

The field of neurocognition provides valuable insights into musical processing and its emotional effects, which often involve activation changes in the brain's core emotion-processing structures. These findings form a crucial basis for understanding cerebral music processing and its potential clinical applications, particularly for patients with neurodegenerative disorders [17]–[19]. Although no research has specifically addressed dementia patients' assessment of digital media performance quality, relevant insights emerge from a community project involving adolescents undergoing psychiatric treatment: In this project, Mozart's works were arranged innovatively with multimedia support from a collective of artists. The intervention demonstrated individual benefits of the arts for these adolescents, including reduced psychopathological symptoms, improved self-esteem, and better emotional and behavioral regulation, including media consumption habits [20].*

Regarding dementia, music-based interventions show beneficial outcomes across various domains (see Fig. 1 for a summarizing overview). Listening to individualized, personally relevant music is particularly promising for dementia patients, as it can elicit emotional responses and tap into autobiographical memories tied to life experiences [2], [21], [22]. Studies indicate that familiar receptive music, when combined with cognitive training or physical exercise, improves overall cognitive performance compared to standard treatments. Notable improvements have been observed in attention, executive functions, orientation, verbal memory, and episodic memory. Additionally, these interventions have shown positive effects on mood, reducing symptoms of agitation, anxiety, and depression [2], [23].

However, the type of dementia may influence therapeutic responses. For example, patients with frontotemporal degenerative dementia often struggle more with emotional associations in music than those with Alzheimer's disease. This suggests that the effectiveness of interventions may vary based on the specific type of dementia. It is hypothesized that cognitive and emotional gains from music-based interventions may stem from either music-induced dopamine release and activation of the brain's reward system or stimulation of the parasympathetic nervous system [2]. Nevertheless, the causal relationship between music-induced mood improvements and underlying neurological changes remains unclear.

Imaging studies provide insights into the neurophysiological mechanisms behind musical interventions in dementia [17]. For example, research has shown that familiar music activates the medial prefrontal cortex in healthy individuals, and this brain region degenerates more slowly in Alzheimer's disease. This slower degeneration may explain why patients with Alzheimer's can recognize familiar songs and retrieve personally significant episodic memories even in the later stages of the disease [21], [24]–[26]. Additionally, King et al. [27] using functional magnetic resonance imaging (fMRI), demonstrated that listening to preferred music excerpts activates the supplementary motor area, a region associated with memory for familiar music and less affected in the early stages of Alzheimer's. They also noted increased activity in the cerebral cortex and cerebellum, which are associated with sensory and attention-related functions.

In summary, existing evidence supports receptive music as a positive factor in the treatment of dementia. However, the development of a standardized protocol for musical anamnesis

*"How to Find Myself Through Mozart": A project by Katarzyna Grebosz-Haring and Belinda Plattner in cooperation with the inter-university institution "Wissenschaft & Kunst" of the Paris Lodron University Salzburg and the University Mozarteum Salzburg, as well as the University Clinic for Child and Adolescent Psychiatry of the Paracelsus Medical Private University and the art collective "gold extra"

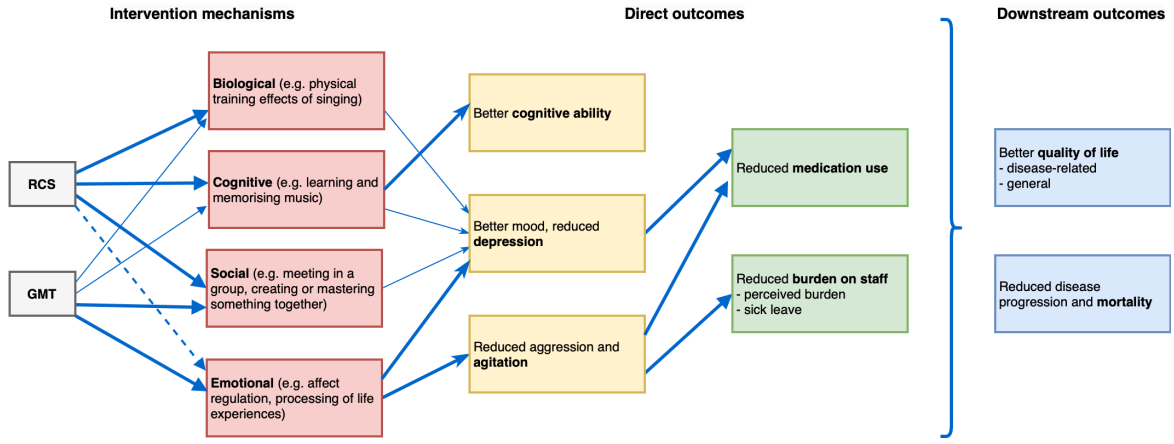


Fig. 1: Mechanisms and processes of music interventions in dementia patients. RCS: Recreational choir singing, GMT: Group music therapy

remains essential to ensure the effectiveness and scalability of these interventions.

III. Anamnesis as a process: State of the art

Anamnesis in music-based interventions aims to gather information about a patient's individual music preferences, performing history, and habits. By understanding these preferences, an appropriate selection of music and environmental factors can be prepared for subsequent music listening sessions. Typically, anamnesis is conducted through interviews and questionnaires, either via telephone or face-to-face, to identify personally relevant music for each individual. This information may be collected from family members, nursing staff, or directly from participants, provided they can articulate their preferences. Additionally, short clips of musical selections—representing genres, artists, or songs suggested in the questionnaire—may be played to the individuals. Patients are then observed for behavioral indicators such as straightening of posture, increased alertness, engagement, smiling, or moving to the rhythm of the music [28]–[30]. Relevant information includes favorite songs and artists, preferred musical genres and styles, and favored musical epochs. For patients with dementia, the primary focus is on music activation programs rather than the conventional quality of musical performance. This approach often requires extensive effort and may necessitate multiple sessions. When patients are unable to express their preferences or emotional associations verbally, trained therapists must rely on non-verbal behavior and body signals to draw conclusions. It is evident that achieving success in this area demands substantial individualized counseling time. Depending on the communicative abilities of the patient, information from proxies such as relatives or caregivers should also be considered in constructing appropriate playlists. In cases where such information is unavailable, research on autobiographical musical memory can help identify music that the patient might recognize or respond to. The advent of music streaming services has further facilitated this process, enabling immediate access to and selection of music without logistical challenges.

IV. Making use of methods: Psycho-Physiology and Sensorics

As outlined above, there is extensive evidence supporting the use of music-based interventions as adjuvant therapy. The integration of digital information technology introduces the potential for scaling these interventions, enabling broader accessibility to larger target groups in less time. Consequently, the upcoming sections present a tool support framework for AMA, detailing methods and technologies from various domains. This approach leverages cognitive abilities and digital tools to augment or substitute lost capabilities. It integrates technological and psychological expertise, along with specific requirements, as a prerequisite for clinical trials. This entails a psycho-physiological approach to measure cognitive and emotional states, translating these into parameters used to describe retrievable features of music. The methods it builds on and components required will be described in the following. The automated interpretation of cognitive capabilities requires the sensor-based assessment of associated somatic and behavioral expressions, comprising psycho-physiological and behavioral indicators: Psycho-physiological indicators, such as heart rate variability (HRV), skin conductance, and blood pressure, etc., correlate with attentional mechanisms and emotions, providing insights into arousal and valence [31], [32]. These measures exploit the body's physical reactions during or in response to cognitive activities, offering continuous data for noninvasive and noninterruptive analysis of user-stimulus interactions. However, certain psycho-physiological measures face limitations, such as high obtrusiveness (e.g., muscular tension [33], blood sugar [34]), insufficient temporal resolution (e.g., galvanic skin response [35], [36]), or poor transferability between individuals (e.g., ECG, HRV [37]). Among these, pupil dilation emerges as the most promising indicator of cognitive states [38]–[40]. Behavioral expressions

of cognitive capabilities involve observable activities related to (i) information perception, such as visual attention, gaze behavior (e.g., saccadic eye movements, fixations [41], [42], head movements [42], [43]), and (ii) descriptive qualities of task and movement execution, such as steadiness and coordination of hand movements [44], [45]. The envisioned digitization of AMA focuses on analyzing and interpreting observable changes in overt behavior during music listening. Overt behavior includes body movements or posture changes observable by others, which may indicate relaxation versus tension, positive versus negative emotion, or engagement with the music (e.g., rhythmic tapping of hands, feet, or fingers). Metrics like gaze behavior and cognitive load can further assess attention and cognitive activation during music perception [28], [46]–[48].

To achieve this, relevant behaviors can be measured using various sensors, including optical sensors (e.g., cameras with skeleton tracking), body-worn sensors (e.g., accelerometers for activity recognition), and remote or head-worn eye trackers. The trade-off between sensor proximity and invasiveness influences patient acceptance and compliance. Body-worn sensors offer high data quality and immersive somatic experiences but may face low acceptance due to their invasive application and complexity, limiting scalability. Conversely, remote sensing approaches, such as camera-based systems and remote eye trackers, offer better scalability and less obtrusiveness but at the cost of reduced data quality.

A low-level remote sensing setup, such as remote eye tracking or movement detection, can serve as a starting point. Mixed or virtual reality (XR) technologies offer a promising avenue by unifying applications and digital environments [49], [50]. A harmonized XR application—such as a headset with integrated eye tracking, visual input, and interaction devices—provides an immersive environment that enhances user experience while minimizing tedious setups. Built-in sensors within XR devices are pre-calibrated, reducing the need for aligning disparate systems. Virtual reality (VR) in particular allows to aim for a broader range of positive effects when implemented effectively [51]–[54]. Successful use cases have proven the applicability of VR in medical settings [55] as well as in education [56], [57], and professional settings [58]. Thus, a target-oriented development approach, prioritizing acceptance and compliance, is essential for obtaining valid and reliable results. The use of XR headsets offers significant advantages for initial data collection due to their advanced sensor capabilities, such as integrated eye tracking and immersive engagement. However, their size and discomfort during prolonged use present challenges for patient acceptance. To address this, we propose utilizing XR headsets exclusively in the initial stage to establish a reliable baseline of cognitive and physiological responses to musical stimuli. Following the baseline assessment (and based on the patient's preferences), the process may transition to less intrusive technologies, such as remote sensors or wearable devices, for subsequent sessions.

Based on current research and available measurements, an evaluation design for AMA development should consider the following requirements: a) Intervention type: receptive interventions b) Focus of the studies: identification of behavioral and somatic reactions to presented music c) Therapy Approach: individual listening d) Intervention setting: at nursing institutions / doctor setups e) Evaluation scale: quantification of behavioral and physiological reaction to music f) Experiment design: 3 stage process for iterative refinement of music preference model.

V. Novel Combination: Computational Modeling of Musical Preference

Understanding the factors that influence individual music preferences has been widely studied over the past decade, identifying variables such as age [59], [60], gender [61], [62], cognitive style [12], and personality [60], [63]. Recent approaches to modeling user music preferences have evolved from correlating personality traits with genres or styles to leveraging finer-grained content-level features derived from the audio itself [64]. AMA builds on these advancements by employing convolutional neural networks (CNNs) to extract high-level features from intermediate network layers [65]. AI-driven models will be utilized to interpret and quantify behavioral changes on a numerical scale from 0 to 100. This process involves the multi-dimensional mapping of behavioral descriptions onto a one-dimensional score. To achieve this, AMA integrates methods from behavioral and physiological interpretation of human attention, creating supervised machine learning models for classification tasks based on multi-modal input vectors. The interpretation of reactions to music samples will be combined with music similarity models from the literature to iteratively refine sample selection for study execution. This iterative process leverages network analysis of graph representations to identify which music samples provide the most significant insights into an individual's music preference model, thereby continuously improving the computational representation of musical taste. Beyond model architecture, the approach offers flexibility for integrating elements of interaction and more complex intervention designs. For instance, the acceptance of music supporting a particular mood may vary based on an individual's current state. In a calm state, an individual may prefer calm music to maintain that state, or they may seek more arousing and activating music. The proposed approach relies on robust cognitive measures, such as attention and overt reactions, while acknowledging that more complex interrelationships and interactions with music should be explored in future iterations. These advancements extend beyond the scope of traditional anamnesis and will require further refinement.

Fig. 2 shows a design of a tool support system built by the authors to prepare such an interacting environment. It contains an XR device enabling the users to interact in a game-like setting in order to trigger emotion- and/or cognition-based reactions that can be used to gain insights into the user's emotional and cognitive states while performing a task and listening to music. Psycho-physiological measures such as built-in eyetracking as well as heart rate measurement are

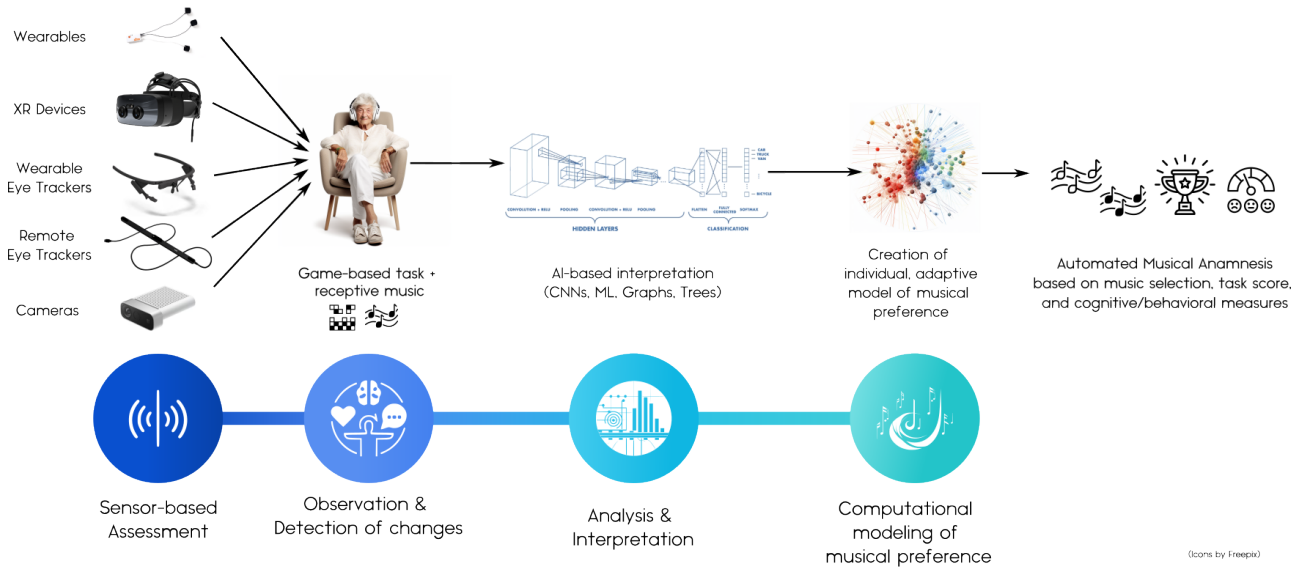


Fig. 2: Conceptual design of a tool support solution developed to conduct pilot studies for model development. A prototype was created around a reaction game with increasing difficulty built with Unity engine, a Varjo XR3 headset with eyetracking capabilites, and a self-developed protocol for structured data collection and preprocessing.

taken into account. The tool support is able to play different kinds of music to the user and analyze acquired data separately. Results will be processed into the music recommender being under development. As such, the demonstrator will be used to conduct experimental pilot studies preceding clinical studies and thereby enable the researchers to create the model described above. Further details of the setup with regard to the user-centered approach applied are described in the following section.

VI. Holistic Tool Support: Integrating User-Centered Design, Methodologies, and Ethical Considerations

In a broader context, the overarching goal of an AMA system is to become a widely adopted tool in care settings. Achieving this objective necessitates the application of various methodologies and best practices. A user-centered design (UCD) approach is particularly suitable, as it involves a structured process that incorporates input from target group individuals throughout all phases of system development, including defining the context of use, establishing requirements, designing solutions, and conducting evaluations. Best practices emphasize the inclusion of all relevant stakeholders [66], [67]. Depending on the deployment environment, different stakeholder groups must be considered. While patients with dementia and music therapists are the primary users, informal caregivers (e.g., family members) and professional caregivers must also play integral roles. In nursing home settings, additional representation from staff, such as care managers and facility management (covering IT infrastructure, care documentation, and billing), should be included in the UCD process. Complementing the UCD approach with methods specifically designed for individuals with dementia is crucial to ensuring user acceptance, addressing ethical considerations, and creating solutions for patients across all stages of the condition. Tailored methods are particularly necessary for individuals with advanced dementia. The importance of the care dyad—comprising the patient with dementia and their caregiver—is emphasized in individualized care settings [68], [69]. Principles of "compassionate design" for cognitively impaired individuals, as proposed by [70], can also be applied to an AMA system. These principles advocate for designs that stimulate the senses, are highly personalized, and foster connections between people.

Methods for analyzing the acceptance of technological solutions are well-established and applicable to the development of an AMA system. The Technology Acceptance Model (TAM), which focuses on perceived usefulness and ease of use [71], and the Unified Theory of Acceptance and Use of Technology (UTAUT), which considers performance expectancy, effort expectancy, social influence, and facilitating conditions [72], are particularly relevant. While alternative technology acceptance models exist, TAM and UTAUT are widely used in healthcare [73] and are well-suited for optimizing system development. While newly developed solutions for people with dementia have the potential to enhance quality of life, it is imperative to ensure they do not compromise privacy, freedom, or human rights [74]. Consequently, the development of an AMA system will incorporate the ethical adoption model proposed by [75], which is based on five pillars: inclusive participatory design,

emotional alignment, adoption modeling, ethical standards assessment, and education and training. The publication offers 18 recommendations derived from this model.

P4 medicine (predict, prevent, personalize, and participate) provides an overarching framework that encompasses all relevant objectives for developing holistic solutions in digital health and therapeutics. By emphasizing prevention and home care, P4 medicine aligns with the needs of patients suffering from chronic conditions while supporting aging well and value-based healthcare. These principles, as outlined in the Strategic Research and Innovation Agenda by the experts of the European Electronic Components and System (ECS) community [6], form a robust foundation for all user-centered and ethical considerations in the development of the AMA system.

VII. Conclusion and Final Remarks

Dementia presents a fundamental challenge to society as a whole and to each affected individual. Musical treatment approaches are among the few interventions capable of successfully activating cognitive, behavioral, and emotional resources in dementia patients, even in late stages, thereby enhancing well-being and quality of life. Significant progress has been made in understanding both the neurodegenerative effects of dementia and the potential of mitigating therapies.

However, music-based interventions face scalability challenges due to the extensive manual effort required for musical anamnesis. From a societal perspective, the urgency to develop scalable therapeutic approaches is heightened by the expected global increase in dementia cases. The development of automated processes to facilitate and objectify musical anamnesis holds the potential to deliver substantial societal impact by providing effective treatments to a broader patient population.

This paper has presented the state-of-the-art research and processes in music therapy, with a particular focus on musical anamnesis. It has explored how dementia affects cognitive and emotional capabilities, the existing interventions, and the potential benefits of these approaches. By introducing a range of methods and technological solutions, the paper outlines all necessary components for building an automated musical anamnesis tool support system. Developing an AMA solution addresses a highly interdisciplinary challenge, requiring the integration and combination of diverse methodologies. In this context, a content-driven approach necessitates a deep understanding of music and its associated contexts, while also addressing issues of inclusion, care, and sustainability. The availability of cultural content in a free and structured manner is a crucial factor, emphasizing the importance of carefully curated digitized artifacts enriched with metadata, narrativity, and interconnectedness. These elements enable the meaningful use of repositories and foster the use of both digitized and digital-born cultural heritage. For this endeavor to succeed, culture must be positioned at the core of all development steps and processes, ensuring that technological advancements go beyond functioning merely as tools for distribution or commercialization. We therefore aim at a renewed understanding of culture tech to ensure an intertwinedness of culture and technology that transcends the concept of a mere toolbox to help distribute or market content. That also includes a broader and more informed understanding of cultural heritage that is not only material for the worst-case stress test of technological developments, but rather at the core of all described endeavours [76].

Music intervention is an efficient, cost-effective, and easily applicable rehabilitation strategy for treating dementia. The proposed approach seeks to automate the labor-intensive process of musical anamnesis and content selection, enabling scalable applications and amplifying the impact of musical interventions in dementia care. It will build upon cognitive, behavioral and physiological indicators to identify interaction with presented music to automatically build models of musical preference of subjects. This innovation addresses the primary limitation of musical intervention in dementia—human effort—by providing a scalable solution that broadens accessibility and effectiveness.

References

[1] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. Cohen-Mansfield et al., "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, no. 10113, pp. 2673–2734, 2017. DOI: 10.1016/S0140-6736(17)31363-6

[2] A. J. Sihvonen, T. Särkämö, V. Leo, M. Tervaniemi, E. Altenmüller, and S. Soinila, "Music-based interventions in neurological rehabilitation," *The Lancet Neurology*, vol. 16, no. 8, pp. 648–660, 2017. DOI: 10.1016/S1474-4422(17)30168-0

[3] J. T. Van der Steen, H. J. Smaling, J. C. Van der Wouden, M. S. Bruinsma, R. J. Scholten, and A. C. Vink, "Music-based therapeutic interventions for people with dementia," *Cochrane Database of Systematic Reviews*, no. 7, 2018. DOI: 10.1002/14651858.CD003477.pub4

[4] L. A. Gerdner, "Individualized music for dementia: Evolution and application of evidence-based protocol," *World journal of Psychiatry*, vol. 2, no. 2, p. 26, 2012. DOI: 10.5498/wjp.v2.i2.26

[5] D. P. Schoenfelder and L. Gerdner, "Evidence-based guideline. individualized music for elders with dementia," *Journal of Gerontological Nursing*, vol. 36, no. 6, pp. 7–15, 2010.

[6] AENEAS, EPoSS and INSIDE, "*Electronic Components and Systems (ECS) Strategic Research and Innovation Agenda (ECS-SRIA) 2024*," https://www.smart-systems-integration.org/system/files/document/ECS-SRIA-2024.pdf

[7] P. Baranyi, "Special issue on internet of digital and cognitive reality: Applications and key challenges," *Infocommunications Journal*, 2023.

[8] K. Kallinen and N. Ravaja, "Emotion perceived and emotion felt: Same and different," *Musicae Scientiae*, vol. 10, no. 2, pp. 191–213, 2006. DOI: 10.1177/102986490601000203

[9] V. N. Salimpoor, M. Benovoy, G. Longo, J. R. Cooperstock, and R. J. Zatorre, "The rewarding aspects of music listening are related to degree of emotional arousal," *PloS one,* vol. 4, no. 10, p. e7487, 2009. DOI: 10.1371/journal.pone.0007487

[10] J. Völker, "Personalising music for more effective mood induction: Exploring activation, underlying mechanisms, emotional intelligence, and motives in mood regulation," *Musicae Scientiae*, vol. 25, no. 4, pp. 380–398, 2021. DOI: 10.1177/1029864919876315

[11] L. Bernardi, C. Porta, and P. Sleight, "Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence," *Heart*, vol. 92, no. 4, pp. 445–452, 2006. DOI: 10.1136/hrt.2005.064600

[12] D. M. Greenberg, S. Baron-Cohen, D. J. Stillwell, M. Kosinski, and P. J. Rentfrow, "Musical preferences are linked to cognitive styles," *PloS one*, vol. 10, no. 7, p. e0131151, 2015. DOI: 10.1371/journal.pone.0131151

[13] J. Panksepp and G. Bernatzky, "Emotional sounds and the brain: the neuro-affective foundations of musical appreciation," *Behavioural Processes*, vol. 60, no. 2, pp. 133–155, 2002. DOI: 10.1016/S0376-6357(02)00080-3

[14] A. Dell'Anna, M. Leman, and A. Berti, "Musical interaction reveals music as embodied language," *Frontiers in Neuroscience*, vol. 15, p. 667838, 2021. DOI: 10.3389/fnins.2021.667838

[15] H. Fukui and M. Yamashita, "The effects of music and visual stress on testosterone and cortisol in men and women." *Neuro Endocrinology Letters*, vol. 24, no. 3-4, pp. 173–180, 2003.

[16] G. Kreutz, C. Q. Murcia, and S. Bongard, "Psychoneuroendocrine research on music and health: an overview," *Music, Health, and Wellbeing*, pp. 457–476, 2012. DOI: 10.1093/acprof:oso/9780199586974.003.0030

[17] K. Grebosz-Haring, G. Bernatzky, F. Wendtner, and G. Kreutz, "Musikhören bei Depression und Demenz: Von der Hirnforschung zur klinischen Anwendung," In *Musik und Gesundheit: Psychologische, neurologische und klinische Perspektiven*, G. Bernatzky and G. Kreutz, Eds. Berlin, Germany: Springer Berlin Heidelberg, 2024, pp. 139–152. DOI: 10.1007/978-3-662-67506-9_9.

[18] S. Garrido, C. J. Stevens, E. Chang, L. Dunne, and J. Perz, "Music and dementia: individual differences in response to personalized playlists," *Journal of Alzheimer's Disease*, vol. 64, no. 3, pp. 933–941, 2018. DOI: 10.3233/JAD-180084

[19] M. C. Hillebrand, L. Weise, and G. Wilz, "Immediate effects of indi- vidualized music listening on behavioral and psychological symptoms of dementia: A randomized controlled trial," *International Journal of Geriatric Psychiatry,* vol. 38, no. 3, p. e5893, 2023. DOI: 10.1002/gps.5893

[20] F. Wallner, M. Karner, and T. Ballhausen, Eds., *Spot On MozART. Visualizing Music in the Digital Age*. Wien: Verlag für moderne Kunst, 2023.

[21] M. El Haj, L. Fasotti, and P. Allain, "The involuntary nature of music-evoked autobiographical memories in alzheimer's disease," *Consciousness and cognition*, vol. 21, no. 1, pp. 238–246, 2012. DOI: 10.1016/j.concog.2011.12.005

[22] E. M. Long, "An innovative approach to managing behavioral and psychological dementia," *The Journal for Nurse Practitioners,* vol. 13, no. 7, pp. 475–481, 2017. DOI: 10.1016/j.nurpra.2017.05.003

[23] M. A. Gaviola, K. J. Inder, S. Dilworth, E. G. Holliday, and I. Higgins, "Impact of individualised music listening intervention on persons with dementia: A systematic review of randomised controlled trials," Australasian journal on ageing, vol. 39, no. 1, pp. 10–20, 2020. DOI: 10.1111/ajag.12642

[24] S. Eschrich, T. F. Münte, and E. O. Altenmüller, "Unforgettable film music: the role of emotion in episodic long-term memory for music," BMC neuroscience, vol. 9, no. 1, pp. 1–7, 2008. DOI: 10.1186/1471-2202-9-48

[25] P. Janata, "The neural architecture of music-evoked autobiographical memories," Cerebral Cortex, vol. 19, no. 11, pp. 2579–2594, 2009. DOI: 10.1093/cercor/bhp008

[26] C. Moreno-Morales, R. Calero, P. Moreno-Morales, and C. Pintado, "Music therapy in the treatment of dementia: A systematic review and meta-analysis," *Frontiers in medicine*, vol. 7, p. 160, 2020. DOI: 10.3389/fmed.2020.00160

[27] J. King, K. Jones, E. Goldberg, M. Rollins, K. MacNamee, C. Moffit, S. Naidu, M. Ferguson, E. Garcia-Leavitt, J. Amaro et al., "Increased functional connectivity after listening to favored music in adults with alzheimer dementia," *The journal of prevention of Alzheimer's disease,* vol. 6, pp. 56–62, 2019. DOI: 10.14283/jpad.2018.19

[28] T. T. Lineweaver, T. R. Bergeson, K. Ladd, H. Johnson, D. Braid, M. Ott, D. P. Hay, J. Plewes, M. Hinds, M. L. LaPradd et al., "The effects of individualized music listening on affective, behavioral, cognitive, and sundowning symptoms of dementia in long-term care residents," *Journal of aging and health*, vol. 34, no. 1, pp. 130–143, 2022. DOI: 10.1177/08982643211033407

[29] L. Gerdner, "An individualized music intervention for agitation," *Journal of the American Psychiatric Nurses Association*, vol. 3, no. 6, pp. 177–184, 1997. DOI: 10.1177/107839039700300603

[30] L. A. Gerdner, "Effects of individualized versus classical "relaxation" music on the frequency of agitation in elderly persons with alzheimer's disease and related disorders," *International Psychogeriatrics*, vol. 12, no. 1, pp. 49–65, 2000. DOI: 10.1017/S1041610200006190

[31] S. Gabay, Y. Pertzov, and A. Henik, "Orienting of attention, pupil size, and the norepinephrine system," *Attention, Perception, & Psychophysics*, vol. 73, pp. 123–129, 2011. DOI: 10.3758/s13414-010-0015-4

[32] B. Gollan, B. Wally, and A. Ferscha, "Automatic human attention estimation in an interactive system based on behavior analysis," in *Proceedings of the 15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*, 2011, pp. 40–54.

[33] B. H. Cohen, R. J. Davidson, J. A. Senulis, C. D. Saron, and D. R. Weisman, "Muscle tension patterns during auditory attention," *Biological psychology*, vol. 33, no. 2-3, pp. 133–156, 1992. DOI: 10.1016/0301-0511(92)90028-S

[34] S. H. Fairclough and K. Houston, "A metabolic measure of mental effort," *Biological psychology*, vol. 66, no. 2, pp. 177–190, 2004. DOI: 10.1016/j.biopsycho.2003.10.001

[35] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä, and N. Ravaja, "A review of the use of psychophysiological methods in game research," *Journal of Gaming & Virtual Worlds*, vol. 3, no. 3, pp. 181–199, 2011. DOI: 10.1386/jgvw.3.3.181_1

[36] C. L. Bethel, K. Salomon, R. R. Murphy, and J. L. Burke, "Survey of psychophysiology measurements applied to human-robot interaction," in *RO-MAN 2007 – The 16th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2007, pp. 732–737. DOI: 10.1109/ROMAN.2007.4415182

[37] P. Nickel and F. Nachreiner, "Psychometric properties of the 0.1 hz component of hrv as an indicator of mental strain," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 12. SAGE Publications Sage CA: Los Angeles, CA, 2000, pp. 2–747. DOI: 10.1177/154193120004401284

[38] B. Gollan and A. Ferscha, "Modeling pupil dilation as online input for estimation of cognitive load in non-laboratory attention-aware systems," *COGNITIVE 2016: The Eighth International Conference on Advanced Cognitive Technologies and Applications*, 2016.

[39] O. E. Kang, K. E. Huffer, and T. P. Wheatley, "Pupil dilation dynamics track attention to high-level information," *PloS one*, vol. 9, no. 8, p. e102463, 2014. DOI: 10.1371/journal.pone.0102463

[40] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources." *Psychological bulletin*, vol. 91, no. 2, p. 276, 1982. DOI: 10.1037/0033-2909.91.2.276

[41] S. P. Liversedge and J. M. Findlay, "Saccadic eye movements and cognition," *Trends in cognitive sciences*, vol. 4, no. 1, pp. 6–14, 2000. DOI: 10.1016/S1364-6613(99)01418-7

[42] J. E. Hoffman and B. Subramaniam, "The role of visual attention in saccadic eye movements," *Perception & psychophysics*, vol. 57, no. 6, pp. 787–795, 1995. DOI: 10.3758/BF03206794

[43] E. I. Knudsen, "Fundamental components of attention," *Annu. Rev. Neurosci*., vol. 30, pp. 57–78, 2007. DOI: 10.1146/annurev.neuro.30.051606.094256

[44] W. Zhang, L.-F. Low, M. Schwenk, N. Mills, J. D. Gwynn, and L. Clemson, "Review of gait, cognition, and fall risks with implications for fall prevention in older adults with dementia," *Dementia and geriatric cognitive disorders*, vol. 48, no. 1-2, pp. 17–29, 2019. DOI: 10.1159/000504340

[45] E. Scherder, W. Dekker, and L. Eggermont, "Higher-level hand motor function in aging and (preclinical) dementia: its relationship with (instrumental) activities of daily life–a mini-review," *Gerontology*, vol. 54, no. 6, pp. 333–341, 2008. DOI: 10.1159/000168203

[46] M. Stolte, B. Gollan, and U. Ansorge, "Tracking visual search demands and memory load through pupil dilation," *Journal of Vision*, vol. 20, no. 6, pp. 21–21, 2020. DOI: 10.1167/jov.20.6.21

[47] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988. DOI: 10.1207/s15516709cog1202_4

[48] M. Pomplun and S. Sunkara, "Pupil dilation as an indicator of cognitive workload in human-computer interaction," in *Human-Centered Computing*. CRC Press, 2019, pp. 542–546.

[49] A. Hamad and B. Jia, "How virtual reality technology has changed our lives: An overview of the current and potential applications and limitations," *International Journal of Environmental Research and Public Health*, vol. 19, no. 18, p. 11278, 2022. DOI: 10.3390/ijerph191811278

[50] M. Slater and M. V. Sanchez-Vives, "Enhancing our lives with immersive virtual reality," *Frontiers in Robotics and AI*, vol. 3, p. 74, 2016. DOI: 10.3389/frobt.2016.00074

[51] M. C. Howard and M. B. Gutworth, "A meta-analysis of virtual reality training programs for social skill development," *Computers & Education*, p. 103707, 2020. DOI: 10.1016/j.compedu.2019.103707

[52] T. Tursø-Finnich, R. O. Jensen, L. X. Jensen, L. Konge, and E. Thinggaard, "Virtual reality head-mounted displays in medical education—a systematic review," *Simulation in Healthcare*, 2022. DOI: 10.1097/SIH.0000000000000063

[53] G. S. Ruthenbeck and K. J. Reynolds, "Virtual reality for medical training: the state-of-the-art," *Journal of Simulation*, vol. 9, pp. 16–26, 2015. DOI: 10.1057/jos.2014.14

[54] I. Horváth, Á. B.Csapó, B. Berki, A. Sudár, and P. Baranyi, "Definition, background and research perspectives behind 'cognitive aspects of virtual reality'(cVR)," *Infocommunications Journal*, no. SP, pp. 9–14, 2023. DOI: 10.36244/ICJ.2023.SI-IODCR.2

[55] L. Li, F. Yu, D. Shi, J. Shi, Z. Tian, J. Yang, X. Wang, and Q. Jiang, "Application of virtual reality technology in clinical medicine," *American Journal of Translational Research*, vol. 9, no. 9, p. 3867, 2017.

[56] Y.-L. Chen, "The effects of virtual reality learning environment on student cognitive and linguistic development," *The Asia-Pacific Education Researcher*, vol. 25, pp. 637–646, 2016. DOI: 10.1007/s40299-016-0293-2

[57] L. Jensen and F. Konradsen, "A review of the use of virtual reality head-mounted displays in education and training," *Education and Information Technologies*, vol. 23, pp. 1515–1529, 2018. DOI: 10.1007/s10639-017-9676-0

[58] C. H. Wijkmark, I. Heldal, and M. M. Metallinou, "Virtual reality support of cognitive processes in firefighter skills training," in *2022 1st IEEE International Conference on Cognitive Aspects of Virtual Reality (CVR)*. IEEE, 2022, pp. 000 027–000 034. DOI: 10.1109/CVR55417.2022.9967649

[59] A. Bonneville-Roussy, P. J. Rentfrow, M. K. Xu, and J. Potter, "Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood." *Journal of Personality and Social Psychology*, vol. 105, no. 4, p. 703, 2013. DOI: 10.1037/a0033770

[60] B. Ferwerda, M. Tkalcic, and M. Schedl, "Personality traits and music genre preferences: how music taste varies over age groups," in *1st Workshop on Temporal Reasoning in Recommender Systems (RecTemp) at the 11th ACM Conference on Recommender Systems, Como, August 31, 2017*., vol. 1922. CEUR-WS, 2017, pp. 16–20.

[61] D. M. Greenberg, M. Kosinski, D. J. Stillwell, B. L. Monteiro, D. J. Levitin, and P. J. Rentfrow, "The song is you: Preferences for musical attribute dimensions reflect personality," *Social Psychological and Personality Science*, vol. 7, no. 6, pp. 597–605, 2016. DOI: 10.1177/1948550616641473

[62] W. McCown, R. Keiser, S. Mulhearn, and D. Williamson, "The role of personality and gender in preference for exaggerated bass in music," *Personality and individual differences*, vol. 23, no. 4, pp. 543–547, 1997. DOI: 10.1016/S0191-8869(97)00085-8

[63] T. Schäfer and C. Mehlhorn, "Can personality traits predict musical style preferences? A meta-analysis," *Personality and Individual Differences*, vol. 116, pp. 265–273, 2017. DOI: 10.1016/j.paid.2017.04.061

[64] A. B. Melchiorre and M. Schedl, "Personality correlates of music audio preferences for modelling music listeners," in *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*, 2020, pp. 313–317. DOI: 10.1145/3340631.3394874

[65] M. Sheikh Fathollahi and F. Razzazi, "Music similarity measurement and recommendation system using convolutional neural networks," *International Journal of Multimedia Information Retrieval*, vol. 10, pp. 43–53, 2021. DOI: 10.1007/s13735-021-00206-5

[66] E. Duque, G. Fonseca, H. Vieira, G. Gontijo, and L. Ishitani, "A systematic literature review on user centered design and participatory design with older people," in *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, ser. IHC '19. New York, NY, USA: Association for Computing Machinery, 2019. DOI: 10.1145/3357155.3358471

[67] S. Kujala and M. Kauppinen, "Identifying and selecting users for user-centered design," in *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, ser. NordiCHI '04. New York, NY, USA: Association for Computing Machinery, 2004, pp. 297–303. DOI: 10.1145/1028014.1028060

[68] N. R. Turner and C. Berridge, "How I want technology used in my care: Learning from documented choices of people living with dementia using a dyadic decision making tool," *Informatics for Health and Social Care*, vol. 48, no. 4, pp. 387–401, 2023, pMID: 37675938. DOI: 10.1080/17538157.2023.2252066

[69] C. Berridge, N. R. Turner, L. Liu, K. I. Fredriksen-Goldsen, K. S. Lyons, G. Demiris, J. Kaye, and W. B. Lober, "Preliminary Efficacy of Let's Talk Tech: Technology Use Planning for Dementia Care Dyads," *Innovation in Aging*, vol. 7, no. 3, p. igad018, 03 2023. DOI: 10.1093/geroni/igad018

[70] A. T. Cathy Treadaway and J. Fennell, "Compassionate design for dementia care," *International Journal of Design Creativity and Innovation*, vol. 7, no. 3, pp. 144–157, 2019. DOI: 10.1080/21650349.2018.1501280

[71] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989. DOI: 10.2307/249008

[72] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003. DOI: 10.2307/30036540

[73] A. A. AlQudah, M. Al-Emran, and K. Shaalan, "Technology acceptance in healthcare: A systematic review," *Applied Sciences*, vol. 11, no. 22, 2021. DOI: 10.3390/app112210537

[74] B. Bennett, F. McDonald, E. Beattie, T. Carney, I. Freckelton, B. P. White, and L. Willmott, "Assistive technologies for people with dementia: ethical considerations," *Bulletin of The World Health Organization*, vol. 95, no. 11, pp. 749–755, 5 2017. DOI: 10.2471/blt.16.187484

[75] J. M. Robillard, I. Cleland, J. Hoey, and C. Nugent, "Ethical adoption: A new imperative in the development of technology for dementia," *Alzheimer's & Dementia*, vol. 14, no. 9, pp. 1104–1113, 2018. DOI: 10.1016/j.jalz.2018.04.012

[76] M. Tauber, B. Gollan, C. Schmittner, and T. Ballhausen, "Social inclusion, health and content," *ERCIM News*, pp. 9–10, 2022.

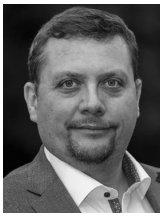
Fabian Simmank works as a Key Researcher at Research Studios Austria Forschungsgesellschaft. He received his PhDs in human biology and economics investigating neural correlates of human behavior in judgment and decision tasks, resp. in moral economic paradigms concerning social value orientation. He has worked on a range of research and consulting projects in the scope of marketing, health, and human factors bridging neuroscientific and psychological approaches to derive relevant insights for industry and consumers.

Katarzyna Grebosz-Haring is currently a Senior Scientist in the Inter-University Organization Arts and Knowledges, Paris Lodron University of Salzburg / Mozarteum University Salzburg and at the Department of Art History, Musicology and Dance Studies at the Paris Lodron University of Salzburg. She studied music pedagogy, music therapy, violin and music and movement pedagogy in Poland and Austria, received her doctorate in musicology at the Mozarteum University Salzburg and habilitated in systematic musicology at the Paris Lodron University of Salzburg. She is director of the Salzburg Institute for Arts in Medicine (SIAM) and co-chair of the International Network for the Critical Appraisal of Arts and Health Research. Psychophysiological, socio-psychological and aesthetic issues of music and arts with special reference to vulnerable groups, such as children and young people with mental illnesses or older people with neurodegenerative diseases are in the centre of her research interests. More: https://scholar.google.com/citations?user=w0sOp48AAAAJ

Thomas Ballhausen is a writer, philosopher and university lecturer at the Institute of Open Arts at Mozarteum University Salzburg. His current research interests include Comparative Media Studies, (Post-) Digitality and Language-Based Artistic Research. In addition to his work in European research contexts, Ballhausen has repeatedly and successfully engaged in interdisciplinary collaborations in fields as diverse as film, music, and new media art. Among his most recent publications are the monographs Nachtaktiv. Versuch über das Cahier“ (2023) and Fiction Fiction. Language Arts and the Practice of Spatial Storytelling“ (together with E. Peytchinska; 2023).

Christian Thomay received his diploma degree in physics from the University of Vienna in 2011 and his PhD in experimental particle physics from the University of Bristol in 2015. He is currently leading the studio Pervasive Computing Applications at the Research Studios Austria FG as Operative Studio Director. In this role, he has worked on a range of research projects in the scope of pervasive computing and attention-aware systems, including projects enhancing flight pilot training towards support using evidence-based performance and attention metrics, industrial assistance systems, and data science projects analzying large amounts of sensor data in medical use-cases, such as robotic-assisted physical therapy and medical training systems.

Martin Biallas is a lecturer at the iHomeLab, part of the Lucerne University of Applied Sciences and Arts. He specializes in assistive technologies, human-system interaction, and physiological monitoring, focusing on innovative solutions that emphasize interdisciplinary collaboration and ethical, user-centered design. He earned his Dipl.-Ing. (TH) degree in electrical engineering and communication technologies, with a specialization in biomedical engineering, from the University of Karlsruhe (TH). Later, in 2011, he completed his Dr. sc. ETH degree in biomedical engineering at the Swiss Federal Institute of Technology (ETH Zurich).



Markus Tauber works as Chief Scientific Officer at Research Studios Austria Forschungsgesellschaft. Between 2015 until 2021, he worked as FH-Professor for the University of Applied Sciences Burgenland, where he held the position: director of the MSc program “Cloud Computing Engineering” and led the research center “Cloud and Cyber-Physical Systems Security”. From 2012 until 2015, he worked at the Austrian Institute of Technology (AIT). Amongst other activities, he was the coordinator of the FP7 Project “Secure Cloud computing for CRitical infrastructure IT” - (www.seccrit.eu). From 2004 to 2012, he was working at the University of St Andrews (UK), where he worked as a researcher on various topics in the area of network and distributed systems and was awarded a PhD in Computer Science for which he was working on “Autonomic Management in Distributed Storage Systems”.

2025 IEEE International Mediterranean Conference on Communications and Networking

7-10 July 2025 // Nice, France

CALL FOR PAPERS

IEEE MeditCom is the conference of the IEEE Communications Society serving the Mediterranean area and surrounding countries. It gathers visionary researchers in academia and industry from all over the world to the shores of the Mediterranean Sea. IEEE MeditCom 2025 will feature a comprehensive and timely technical program that addresses many of the outstanding challenges in the areas of communications and networking. Submission of original technical papers is solicited on a wide range of research topics encompassing theoretical and applied research in, but not limited to, the following areas:

- 5G/6G Systems and Networks
- Antennas, Propagation, and Channel Modeling
- Big Data and Machine Learning for Communications
- Cloud Communications and Data-Center Networks
- Coding/Decoding Theory and Techniques
- Cognitive Radio and Dynamic Spectrum Access
- Communication and Information Theory
- Edge Computing, Edge Intelligence, and Fog Networks
- Energy Efficient Communications and Computing
- Image, Speech, and Signal Processing for Communications
- Integrated Sensing and Communications
- Internet of Things, Smart Grids, and Vehicular Networks
- Cellular and Cell-Free Massive MIMO
- Millimeter-Wave, Sub-Terahertz, and Terahertz Communications
- 3D Networks: Unmanned Aerial Vehicles and High Altitude Platforms
- Molecular and Nanoscale Communications
- Network Applications, Services, and Management
- Network Architecture, SDN, NFV
- Next-Generation Multiple Access
- Next-Generation Physical, Link, and Network Layers Techniques
- Optical Communications and Networks
- Performance Evaluation, Simulation, Testbeds and Prototypes
- QoE/QoS Support and Cross-Layer Design
- Quantum Communications and Computing
- Reconfigurable Intelligent Surfaces, Holographic MIMO, Wave-Based Processing for Smart Environments
- Satellite and Space Communications
- Security, Privacy, Trust and Blockchain
- Semantic and Goal-Oriented Communications
- Smart Grids and Energy Networks
- Underground and Underwater Communications

IMPORTANT DATES

Submissions Deadline:	3 March 2025
Acceptance Notification:	12 May 2025
Camera-Ready Submission:	26 May 2025

IEEE ComSoc IEEE Communications Society

2025 IEEE MeditCom Organizing Committee

General Co-Chairs
Hikmet Sari, NJUPT, China
Dirk Slock, EURECOM, France

TPC Co-Chairs
Angela Zhang, Chinese Univ. of Hong Kong, HK
Davide Dardari, University of Bologna, Italy

Workshop Co-Chairs
Guan Gui, NJUPT, China
Marwa Chafii, NYU Abu Dhabi, UAE

Tutorial Co-Chairs
Periklis Chatzimisios, International Hellenistic University, Greece
Mutlu Koca, Bogazici University, Turkey

Panel Co-Chairs
Yan Chen, Huawei, Canada
Emilio Calvanese Strinati, CEA-LETI, France

Keynotes Chair
Stefano Bregni, Politecnico di Milano, Italy

Operations and Finance Chair
Adlen Ksentini, EURECOM, France

Publication Chair
Bouziane Brik, University of Sharjah, UAE

Web & Social Media Chair
Ejder Bastug, Nokia Bell Labs, France

Publicity Co-Chairs
Ranga Rao Venkatesha Prasad, Delft University of Technology, The Netherlands
Eirini Eleni Tsiropoulou, University of New Mexico, USA
Tomoaki Ohtsuki, Keio University, Japan
Nury Gabriela Ramirez Cely, HCL Technologies, Mexico

Senior Conference Planner
Tina Gaerlan, IEEE ComSoc, USA

www.ieee-meditcom.org

2025 IFIP/IEEE Network Operations and Management Symposium (NOMS 2025)
12-16 May, 2025, Honolulu, HI, USA
"Managing the Wave to Global Connectivity"

Second IEEE Workshop on Generative AI for Network Management (GAIN)

• • • CALL FOR WORKSHOP PAPERS • • •

The 2025 edition of GAIN will again bring together academic researchers from various disciplines (communication networks, data science, operational research) and practitioners from industry. The workshop welcomes scientific papers as well as industrial use case papers. The considered topics in GenAI for network management are initially structured along the well-accepted FCAPS model for network and network service management. However, the workshop is not limited to these topics, and novel contributions on using Generative AI in the operation and management of networks and services are welcome.

Fault Management
- Predictive Maintenance: Using generative AI for proactive network management
- Network Troubleshooting using Generative AI, incl. root cause analysis and resolution
- Monitoring using Generative AI: Using generative AI for efficient monitoring of network resources

Configuration Management
- Network Configuration Automation with Generative AI
- Automated Network Design and Deployment using Generative AI
- Generative AI for Traffic Management: Optimizing network traffic engineering through AI

Accounting
- Ethical Considerations: Addressing Privacy and Security Concerns in AI-Based Network Management
- Ensuring fairness between network users using generative AI for optimal resource allocation

Performance
- Network Optimization with AI: Leveraging generative algorithms for network efficiency
- Dynamic Resource Allocation: Leveraging generative AI for efficient network resource management
- Efficient Network Data Analysis using Generative AI

Security
- AI-Based Security Protocols: Developing next-generation network security strategies
- Anomaly Detection and Response: Utilizing generative AI for enhanced network security

Use Cases
- Generative AI for management of IoT, Wireless/RAN, or Core, and Cloud-to-Edge networking

General
- Prompt Engineering for Network Management Using LLMs
- Robustness and Reliability of Generative AI for net. management (incl. benchmarks and datasets)
- Scalability Orchestration, Testing and Validation of Generative AI for Network Management

Submission and Important Dates:

Submission site:
GAIN 2025 full papers: https://jems3.sbc.org.br/events/224/355/submit
GAIN 2025 short papers: https://jems3.sbc.org.br/events/224/356/submit
Paper Submission Deadline: **Jan. 17, 2025**
Notification of Acceptance: **Feb. 28, 2025**
Final Camera Ready: **Mar. 14, 2025**

Workshop organisers:

- Alberto Leon-Garcia, Univ. of Toronto, CA *(alberto.leongarcia@utoronto.ca)*
- Pal Varga, Budapest Univ. of Technology and Economics, HU *(pvarga@tmit.bme.hu)*
- Kurt Tutschku, Blekinge Inst. of Technology, SE *(ktt@bth.se)*

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on

https://journals.ieeeauthorcenter.ieee.org/
Then click: "IEEE Author Tools for Journals"
- "Article Templates"
- "Templates for Transactions".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *Index Terms (Keywords)*. For the final version of accepted papers, please send the short cvs and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information. When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue
g) Document Object Identifier (DOI)

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984. DOI: 10.1109/TEI.1984.298778

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to submit their papers electronically via the following portal address:

https://www.ojs.hte.hu/infocommunications_journal/about/submissions

If you have any question about the journal or the submission process, please do not hesitate to contact us via e-mail:

Editor-in-Chief: Pál Varga – pvarga@tmit.bme.hu

Associate Editor-in-Chief:

József Bíró – biro@tmit.bme.hu

László Bacsárdi – bacsardi@hit.bme.hu

J3C 2025

IFAC Joint Conference on Computers, Cognition and Communication

15th-18th September 2025 @ Padova, Italy

CALL FOR PAPERS

The 1st IFAC Joint Conference on Computers, Cognition, and Communication will be held in Padova, Italy, from 15-18 September 2025. J3C serves as a platform for scientists, researchers, and practitioners to discuss their cutting-edge research results and findings, shape future directions and developments, and exchange knowledge and perspectives in the fields of intelligent control, automation & robotics, cyber-physical systems, Artificial Intelligence & control, control via communication networks, engineering, and its integration into industry and society. It constitutes the primary forum for cross-industry, multidisciplinary research and provides an opportunity for a unique and rich cultural experience complemented by excellent technical and social programs. J3C 2025 is the flagship event for the IFAC CC on Computers, Cognition, and Communication (CC3) and will consist of 3 sub-events:

1st IFAC Workshop on Engineering and Architectures of Automation Systems (EAAS 2025): Automation systems are subject to substantial change. Architectures are becoming more flexible and virtualized. Communication channels no longer just transfer individual data points but comprehensive data models. Digital twins form the backbone of modern applications in the cloud or at the edge. The role of humans is also changing from observer to an integral part of a human-centered production environment. EAAS core topics: digital twins, system architectures of automation systems, information models in engineering and operation, application of artificial intelligence, autonomous and cyber-physical systems.

7th IFAC Conference on Intelligent Control and Automation Sciences (ICONS 2025): The ICONS conference delves into computational intelligence methods for modeling, system identification, and control. Key topics include neural networks, evolutionary computing, fuzzy techniques, swarm intelligence, reinforcement learning, and training algorithms. Focus areas cover machine learning, deep learning, neurodynamic optimization, brain-computer interfaces, and human-machine interaction. Hybrid systems, smart sensors, actuators, data fusion, intelligent agents, swarm robotics, and autonomous systems are also featured. Applications span transport, medical, biomedical, aerospace, automation, manufacturing, process control, industry, smart cities.

7th IFAC Symposium on Telematics Applications (TA 2025): The TA symposium focuses on the expanding capabilities in web technology, telecommunication, and machine-to-machine communication. It explores how Cyber-Physical Systems, the Industrial Internet, and the Internet of Things enhance traditional remote control and create new applications. Emphasizing the need for sound theory, methods, proof of concepts, and industrial standardization, the symposium offers a platform for academic and industrial communities to address new challenges, share solutions, and discuss future research. Topics include telematic methods, networked systems, control through networks, cyber-physical systems, Industry 4.0/5.0, interoperability.

J3C 2025 will feature keynote addresses, tutorials, and technical presentations and special sessions, all of which will be included in the registration. Special session proposals and technical papers are solicited on all cited aspects, from theory to applications. All accepted papers will appear in the proceedings of the meeting, will be hosted on-line on the IFACPapersOnLine.net website and will be indexed in SCOPUS.

IFAC TC Main-Sponsors: TC3.1, TC3.2, TC3.3
IFAC TC Co-Sponsors: TC1.2, TC1.5, TC2.1, TC4.1, TC4.3, TC5.1, TC5.3, TC6.1, TC6.4, TC7.3, TC7.5, TC8.2, TC9.4

Schedule:
Special Session Proposals:
15 March 2025
Submission of full papers:
15 April 2025
Notification of acceptance:
25 June 2025
Camera-ready paper & author registration:
10 July 2025
Early registration deadline:
10 July 2025

National Organizing Committees (NOC):
General Chair – Gian Antonio Susto (ITA)
EAAS Co-Chair – Federico Tramarin (ITA)
EAAS Vice Chair Industry – Gilberto Pin (ITA)
ICONS Co-Chair – Alessandro Beghi (ITA)
ICONS Vice C. Industry – Chiara Masiero (ITA)
TA Co-Chair – Angelo Cenedese (ITA)
TA VC Industry – Dave Cavalcanti (USA)

International Program Committees (IPC):
IPC Chair – Shuzhi Sam Ge (SGP)
EAAS Co-Chair – Mike Barth (DEU)
EAAS VC Industry – Frank Maurer (DEU)
ICONS Co-Chair – Kevin Guelton (FRA)
ICONS VC Industry – Diego Romeres (USA)
TA Co-Chair – Lei Ma (CHN)
TA Vice Chair Industry – Ming Jiang (CHN)

Editors & Other Main Organizers:
CC3 Coordinator – Thierry Marie Guerra (FRA)
CC3 Conference Vice-chair – Ulrich Jumar (DEU)
Vice Program Chair – Séan McLoone (GBR)
EAAS Editor – Jus Kocijan (SVN)
ICONS Editor – Roberto Oboe (ITA)
TA Editor – Alexander Fay (DEU)

j3c.org

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its 1000 individual members, the Scientific Association for Infocommunications (in Hungarian: Híradástechnikai és Informatikai Tudományos Egyesület, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society.

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **FERENC VÁGUJHELYI** • *elnok@hte.hu*

Secretary-General: **GÁBOR KOLLÁTH** • *kollath.gabor@hte.hu*

Operations Director: **PÉTER NAGY** • *nagy.peter@hte.hu*

Address: H-1051 Budapest, Bajcsy-Zsilinszky str. 12, HUNGARY, Room: 502

Phone: +36 1 353 1027

E-mail: *info@hte.hu*, Web: *www.hte.hu*