

文章编号:1007-5321(2022)01-0063-06

DOI:10.13190/j.jbupt.2021-092

基于异构计算加速的开源 5G 架构

朱子坤, 章晨宇, 王鲁晗, 路兆铭, 温向明

(1. 北京邮电大学 信息与通信工程学院, 北京 100876; 2. 北京邮电大学 网络体系构建与融合北京市重点实验室, 北京 100876;
3. 北京邮电大学 先进信息网络北京实验室, 北京 100876)

摘要: 针对通用处理器下开源第 5 代移动通信系统(5G)处理性能差的问题,研究了开源 5G 的处理特性并提出了基于异构计算加速的架构。该架构将空口协议栈上层与物理层进行分离,并模块化开源 5G 物理层功能。在保留通用处理器对于上层协议栈流程处理优势的基础上,引入异构硬件对计算资源需求大的物理层功能模块进行加速。提出异构物理层的概念并设计了计算资源适应性分配的算法。其次,首次实现了基于 OpenAirInterface 平台的异构加速开源 5G 软件化平台。对信道编解码和调制模块进行实际测试的结果显示,相比于仅在通用处理器下测试结果,异构计算加速架构能够显著地提升物理层模块的处理性能。

关键词: 移动通信; 开源第 5 代移动通信系统; 异构计算加速; 异构物理层

中图分类号: TN929.5

文献标志码: A

Open Source 5G Architecture Based on Heterogeneous Computing Acceleration

ZHU Zishen, ZHANG Chenyu, WANG Luhan, LU Zhaoming, WEN Xiangming

(1. School of Information and Communications Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China;
3. Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Aiming at the problem of poor processing performance of open source of the fifth generation of mobile communications system(5G) under general processor, the characteristics of open source 5G are studied and the architecture of heterogeneous computing acceleration is proposed. This proposed architecture separates the upper layer and the physical layer of the air interface protocol stack, and modularizes physical layer functions of the open source 5G. Utilizing the advantages of general processor on processing upper protocol stack, heterogeneous hardware is introduced to accelerate physical layer function modules, which is demanding for computing resources. The concept of heterogeneous physical layer and the algorithm of adaptive allocation of computing resources are introduced into the architecture. Then, the heterogeneous accelerated open source 5G software platform is implemented for the first time based on OpenAirInterface platform. The actual tests of channel codec and modulation modules show that the proposed heterogeneous acceleration architecture can significantly improve the processing performance of physical layer compared with general purpose processor architecture.

Key words: mobile communication; open source the fifth generation of mobile communications system architecture; heterogeneous computing acceleration; heterogeneous physical layer

收稿日期: 2021-05-17

基金项目: 国家重点研发计划项目(2019YFB1803303); 中央高校基本科研业务费项目(2020RC04)

作者简介: 朱子坤(1998—), 男, 硕士生。

通信作者: 王鲁晗(1989—), 男, 讲师, 邮箱: wluhan@bupt.edu.cn。

随着第5代移动通信系统(5G, the fifth generation of mobile communication system)的不断发展、推进以及初步部署,移动通信系统需要适应业务场景多样化、网络功能快速更新迭代的新需求。开源5G的发展推动了无线网络的软硬件解耦,为学者和业界提供了无线网络更新和部署的新思路,成为无线网络发展的主要趋势。

现阶段的开源5G软件定义无线电(SDR, software define radio)平台多数是在通用硬件平台上实现的。在OpenAirInterface(OAI)平台中,5G无线网络侧的协议栈流程处理以及信号处理均是在x86架构的平台上实现。在开源5G SDR平台中,物理层(PHY, physical layer)的处理性能关乎整个平台的处理能力。但是随着网络数据的指数增长、网络功能的不断扩展,仅仅在x86架构的平台下实现的软件物理层的处理性能难以满足5G无线网络的需求,与基于专用集成电路的商用设备仍有一定的差距^[1]。

随着硬件技术的高速发展,现场可编程逻辑门阵列(FPGA, field programmable gate array)、图形处理单元(GPU, graphics processing unit)、数字信号处理(DSP, digital signal process)等异构硬件能够更好地适应不同功能与不同场景下的处理需求。异构硬件具有强大的计算能力并能够对某些具有特殊计算特征的计算任务进一步优化。高性能的异构硬件能够胜任SDR平台无线网络系统协议栈的处理工作,满足高速无线网络对高处理性能的需求^[2]。特异性的异构硬件能够更好地适应开源5G物理层模块多样化的计算特征以加速物理层乃至整个开源5G SDR平台。以中央处理单元(CPU, central processing unit)作为主处理中心,处理无线网络上层协议栈,异构硬件作为物理层加速处理器的开源5G架构成为提升开源5G SDR平台处理性能的可行方案。

1 相关工作

部分文献在OAI平台的基础上研究开源5G的实际部署与应用。Gavrilovska等^[3]提出了协议软件化、资源虚拟化的高灵活性5G云接入网架构来实现高效的虚拟化无线网络部署。有学者从业务应用的角度出发,研究OAI平台下雾无线网络与增强现实业务的灵活结合,5G网络的云原生部署测试、人工智能应用在5G虚拟化网络上的测试以及车联网应用测试^[4-7]。大量针对开源5G部署应用的研究

都面临一个不可避免的问题,即如何高效地利用底层硬件资源来提升开源5G SDR平台的性能。

Fişne等^[8]提出了一种基于信号侦听与感知的高性能SDR架构设计。该架构中加入了实时处理的GPU和DSP异构硬件用于优化处理信号的感知速度。Hussain等^[9]和Fang等^[10]从高性能SDR架构设计以及高性能异构部署环境2个方向进行考虑,提出了一种异构多核下可编程的架构,并对数据流的处理进行优化。该方案以FPGA和DSP等硬件构建多核的部署环境,并引入数据流可编程机制,提升SDR平台数据处理能力。Marojevic等^[11-12]从异构硬件的计算资源分配角度出发,通过将底层多种类型异构硬件的计算资源虚拟化,以最小化无线网络功能模块的处理时间作为优化目标,进行建模并设计资源映射算法。

在众多使用异构硬件加速的SDR平台研究的基础上,研究了开源5G SDR平台的特性,引入了具有异构物理层的开源5G架构。该架构使用CPU处理无线网络协议栈上层,并通过多种计算特性不同的异构硬件加速PHY的信号处理。在此基础上,实现了基于OAI平台的CPU-GPU异构计算加速的开源5G SDR平台。通过对信道编解码以及调制模块的实际测试,证明了该架构能显著地提升开源5G SDR平台的性能。

2 架构设计

2.1 总体架构

现今的开源5G架构按照无线接入网侧协议栈的划分而设计。各个协议子层之间相对独立、协同工作。用户设备通过射频单元(RU, radio unit)与基站的RU进行无线信号收发,然后依次递交各个协议子层进行相应处理。在开源5G SDR平台的实际开发与部署测试中可知,包括媒体接入控制层子层在内的无线接入网络上层中的信号流程处理大多不需要过多的计算资源,CPU完全可以胜任,且相比于GPU等硬件更适合无线接入网络上层的信号流程处理任务。但对于大量信号发生与处理的PHY层则需要更多的计算资源。

针对PHY层对计算能力的需求,设计了基于异构计算加速的开源5G架构。在该架构中,尽可能避免无线接入网络上层的改动以保留CPU对于其流程处理的优势,同时引入异构物理层(H-PHY, heterogeneous physical layer)以加速PHY层的处理。

2.2 H-PHY

如图 1 所示, H-PHY 可以分为 3 个部分, 即功能模块、适配层和异构计算平台。

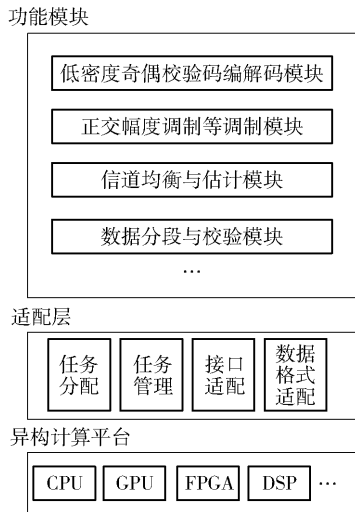


图 1 H-PHY 架构图

2.2.1 功能模块

H-PHY 的各个信号处理模块, 包括低密度奇偶校验码 (LDPC, low density parity check code) 编解码、正交振幅调制 (QAM, quadrature amplitude modulation)、信道均衡与估计以及数据分段与校验等信号处理模块。各个模块相对独立且在 H-PHY 中被视为不同的待处理任务。在 H-PHY 中, 对异构计算平台中不同的异构硬件进行编号, 用不同的整数序号表示, 并根据不同异构硬件的性能, 如 GPU 的核数, 抽象其算力的大小, 并用一个数值表示。记任务集为 $J, J = \{j_1, j_2, \dots, j_N\}$ 。则第 n 个任务为

$$j_n = (q_n, c_n^{\min}, c_n^{\max}, f_n)$$

其中: q_n 是一个整数序号, 表示与任务 j_n 计算特性相适应的异构计算平台; c_n^{\min} 和 c_n^{\max} 分别为该任务所需的最大算力和最小算力值; f_n 为该任务的运算优先级, 是一个整数, f_n 越大运算优先级越高。设 c_n 为该任务最终分配得到的算力值。

2.2.2 异构计算平台

异构计算平台是由 CPU, GPU, FPGA 或 DSP 等异构硬件所组成的硬件计算处理平台。异构计算平台内的各个异构硬件通过高速串行计算机扩展总线 (PCIe, peripheral component interconnect express) 进行高速连接。不同的异构硬件具有不同的计算处理优势, 因此, 需要对其进行适应性的任务分配。记异构计算平台中异构硬件的集合为 $P, P = \{p_1, p_2, \dots,$

$p_M\}$ 。则第 m 个异构硬件为

$$p_m = (q_m, c_m^{q_m})$$

其中: q_m 为异构硬件 p_m 的编号, $c_m^{q_m}$ 为 p_m 的总算力。

因此, 异构计算平台的总算力为 $C = \sum_{m=1}^M c_m^{q_m}$ 。

2.2.3 适配层

适配层作为中间层需要对功能模块以及异构计算平台进行连接, 其功能如下:

1) 任务分配。H-PHY 中不同的功能模块具有不同的计算特性与计算能力的需求。因此, 适配层需要根据不同功能模块的特性将其分配给相适应的异构硬件。

任务分配算法

- 1 根据每个任务的 f_n , 对 J 进行从大到小排列, 优先分配优先级高的任务
- 2 对任务 j_n , 有 $c_n^{\min} \leq c_n \leq c_n^{\max}$
- 3 对于所有指向 p_m 的任务都有 $\sum_{q_n=q_m} c_n \leq c_m^{q_m}$
- 4 在满足 2, 3 的前提下, 将高优先级的任务 j_n 分配到 q_n 对应的 p_m 上
- 5 若 p_m 分配的任务过多导致不满足 3, 则任务 j_n 的 q_n 随机更改为另一序号
- 6 回到 2, 直到不需要进行 q_n 更改
- 7 根据上述约束条件, 优化 $\max_{J, P} \sum_{n=1}^N c_n$ 得到最大总算力以及分配方案

2) 任务管理。H-PHY 中的各个模块大致呈现出串行的处理关系, 因此适配层必须记录与管理不同模块和异构硬件之间的映射以协调串行处理流程的高效运作。对于每条任务, 使用任务编号、任务时序以及任务窗口对任务进行管理。任务编号用于唯一标识不同任务便于进行识别。每个任务都被赋予一个任务时序用于标识当前任务与其他任务的时序关系。当相邻时序任务串行执行时, 各任务所在的异构硬件无需再将数据回传到 CPU, 再传输到其他异构硬件中, 避免了数据回传 CPU 造成的时延。任务时序会进行动态的调整以适应计算需求的变化。任务窗口则是用来记录任务的未响应时间, 当窗口超时, 适配层会休眠该任务并释放资源, 达到资源的高效利用。

3) 接口适配。当任务 j_n 在不同的异构硬件中运行时, 由于不同异构硬件计算特性的差别, j_n 的软件化实现不同, 从而产生软件接口不统一。所提架

构加入接口适配的功能,通过调整不同异构硬件间的数据流传输格式来解决接口不适配的问题。

4)数据格式适配。不同的异构硬件需要采用不同的数据处理格式以达到最佳性能,例如 DSP 更适合进行 Q1.15 一类的定点数运算;GPU 需要使用 HALF_2 的浮点数格式。因此,适配层需要在不同的异构硬件之间进行数据格式的适配。

3 CPU-GPU 的开源 5G 平台

GPU 的并行处理以及流处理优势使得其能适应物理层大带宽模块的计算需求。基于 OAI 平台,搭建了 CPU-GPU 的异构计算加速开源 5G SDR 平台。使用上文所述的异构计算架构,该平台将无线接入网络上层以及部分物理层的处理交由 CPU 完成,将物理层中 LDPC 编解码模块以及 QAM 调制模块在 GPU 中运行。

3.1 LDPC 编解码模块

LDPC 编解码是上下行共享信道进行数据收发的必须模块。LDPC 是线性分组码,其编码过程需要进行大量的矩阵乘加运算。LDPC 的解码则采用置信传播的软译码算法,即最小和译码算法。该算法将 LDPC 码分为比特位与校验位两部分,并在两部分之间进行迭代运算,需要进行大量的并行浮点数运算。

LDPC 编解码在 CPU-GPU 异构加速开源 5G SDR 平台中的实现如下:

CPU 将待编码或待解码数据以及编解码参数传输给 GPU。GPU 首先对码块数据进行复制并存放在显存之中。显存中的待计算数据被视为张量数据。根据编解码的相应参数,GPU 创建张量描述以及工作空间。其中,张量描述用于表示张量数据的格式与大小,工作空间则是 GPU 用来进行 LDPC 编解码计算的空间。然后,GPU 调用显存中编解码核函数实现 LDPC 编解码功能。同时,GPU 采用流处理的方式,并结合并行多核运算以加速整个处理过程。最后,GPU 将处理好的数据返回给 CPU。

3.2 QAM 调制模块

传统的软件 QAM 调制模块基于数据的移位运算以及查找进行设计,当数据流很大时,CPU 难以并行处理多数据查找。

与 CPU-GPU 下 LDPC 编解码模块的实现相似,GPU 接收到 CPU 传来的待调制数据后,将输入放入显存之中,并根据调制参数设置相应的张量描述与

工作空间。然后 GPU 调用 QAM 调制核函数进行 QAM 调制。与 LDPC 编解码核函数不同的是,QAM 调制核函数需要处理大量移位、比较与查找运算。

4 仿真测试

4.1 测试环境

使用型号为 Core i9 @ 3.1 GHz 16C32T 的 CPU 和型号为 V100 的 GPU 组建 CPU-GPU 异构计算加速的开源 5G SDR 平台,简称为 CPU-GPU 平台,并对其进行测试。同时,将仅使用型号为 Core i5 @ 2.6 GHz 4C8T 的 CPU 和型号为 Core i9 @ 3.1 GHz 16C32T 的 CPU 运行的开源 5G SDR 平台作为对比平台,以下简称 2.6 GHz CPU 平台和 3.1 GHz CPU 平台。测试并对比不同平台下 LDPC 信道编解码模块与 QAM 调制模块的运行时间和吞吐量。

4.2 LDPC 编解码测试

根据第三代合作伙伴计划(3GPP,3rd generation partnership project)技术规范(TS, technical specification)38.212 标准,选用基本图样一(BG1, base graph one)以及偏移值为 384 作为 LDPC 编码所使用的校验矩阵,其码块长度为 8 448 bits。对 CPU-GPU 平台及不同主频的 CPU 对比平台进行测试,结果如图 2 所示。其中,仅 GPU 内运行时间为在 GPU 中 LDPC 编码所需要的时间。通过对该时间的统计,可间接得出数据在 CPU-GPU 平台中用于数据传输损耗的时间。

由图 2 可知,CPU 平台中的编码性能会随着主频的升高而有所提升,但即使在 3.1 GHz CPU 平台中,编码一个码块仍需 230 μ s。用每秒处理的比特数表示模块吞吐量,则 3.1 GHz CPU 平台中 LDPC 编码的吞吐量为 36.73 Mbit/s。而在 CPU-GPU 平台中,从数据自 CPU 输出到 GPU,到数据再返回 CPU 只需 63 μ s,即吞吐量为 134.1 Mbit/s。相比之下,使用异构计算加速架构的 CPU-GPU 平台的吞吐量有显著提升。

由图 2 还可知,GPU 中 LDPC 编码所需的时间约为 42 μ s,则 CPU-GPU 平台中数据传输损耗时间,即 CPU 与 GPU 间 PCIe 数据传输时间以及 CPU 与 GPU 中数据存放到内存与显存消耗的时间之和,约为 21 μ s。

同样对偏移值为 384 的 BG1 的 LDPC 码块进行解码,当信噪比为 30 dB 时,不同运行平台的 LDPC 解码时间对比如图 3 所示。由图可知,CPU 平台中

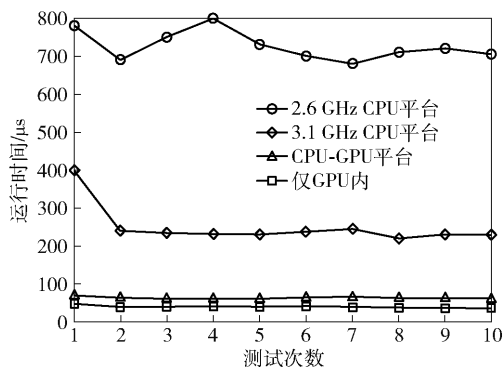


图2 不同运行平台中 LDPC 编码性能对比

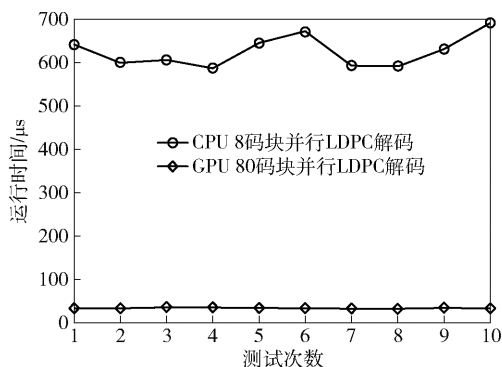


图4 3.1 GHz CPU 平台与 GPU 中 LDPC 并行解码性能对比

LDPC 的解码性能同样随着 CPU 主频的增加而提升。LDPC 解码模块在 2.6 GHz CPU 平台与 3.1 GHz CPU 平台中消耗的时间分别为 938.6 μs 与 389.3 μs , 吞吐量分别为 9 Mbit/s 与 21.7 Mbit/s。而在 GPU 中, LDPC 解码只需要 24 μs , 可达到 352 Mbit/s 的高吞吐量。在 GPU 中进行 LDPC 解码的吞吐量较 3.1 GHz CPU 平台提升了 16.22 倍。

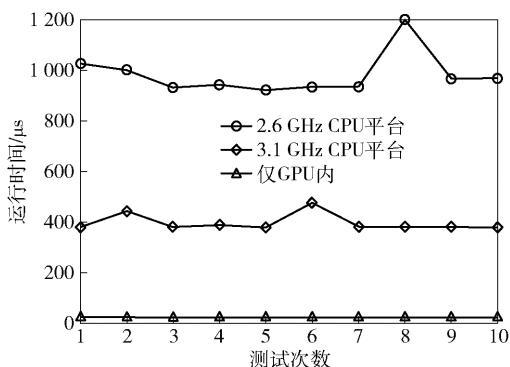


图3 不同运行平台中 LDPC 解码性能对比

为了充分发挥 GPU 强大的并行处理能力,在其他解码条件与上述相同的情况下,更改 3.1 GHz CPU 平台中的信道解码模块为 8 码块并行解码,与 GPU 中 80 码块并行解码作对比。结果如图 4 所示。由图可知,3.1 GHz CPU 平台中 LDPC 8 码块并行解码的时间约 643.6 μs , 吞吐量约为 105 Mbit/s, 而 GPU 中 80 码块并行解码的时间为 33 μs , 吞吐量可达到 20.5 Gbit/s。

CPU-GPU 平台中解码数据传输损耗时间如图 5 所示。并行解码 80 个码块时,数据传输损耗时间约为 70 μs ,则 CPU-GPU 平台中 CPU 与 GPU 的数据传输接口的业务数据交换带宽为 9.65 Gbit/s,能够覆盖 3GPP TS 38.214 中所有调制编码方式等级。综上,在 CPU-GPU 平台中 80 码块并行解码的总运行

时间为 103 μs ,吞吐量为 6.56 Gbit/s。虽然 CPU 与 GPU 之间的数据传输损耗时间占总运行时间的 68%,但 CPU-GPU 平台中 LDPC 并行解码的吞吐量仍然远高于 3.1 GHz CPU 平台的并行解码吞吐量。CPU-GPU 平台对 LDPC 解码性能的提升十分显著。

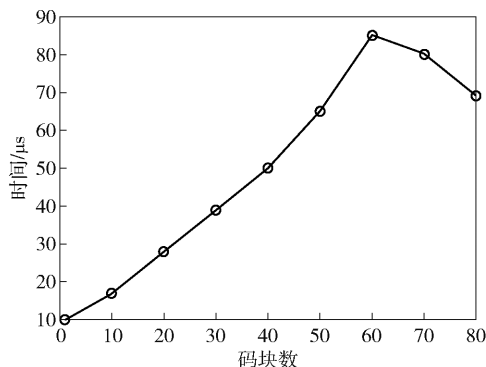


图5 CPU-GPU 平台中数据传输损耗时间

4.3 QAM 调制测试

在对 QAM 调制的测试中,以算力需求最大的 256QAM 调制为例,分别统计 CPU-GPU 平台中输出大约 36 000 个 256QAM 符号以及 3.1 GHz CPU 平台中输出远小于该数量的 256QAM 符号的处理时间,结果如图 6 所示。

由图可知,GPU 中进行 QAM 调制的运行时间约为 14 μs ,每个 256QAM 符号由 8 bits 数据组成,即 GPU 中 QAM 调制可达 20.6 Gbit/s 的数据吞吐量。相比于 CPU 平台约 1 000 μs 的运行时间,GPU 中 QAM 调制的性能得到了显著提升。根据图 6 中数据交换时间的曲线,可得 CPU-GPU 平台中,数据从 CPU 到 GPU 的单次交换时间,即传输损耗时间约为 80 μs ,则总交换时间为 160 μs ,换算业务数据交换带宽为 1.8 Gbit/s,占总运行时间的 92%,较

LDPC 解码具有更高的相对传输损耗。

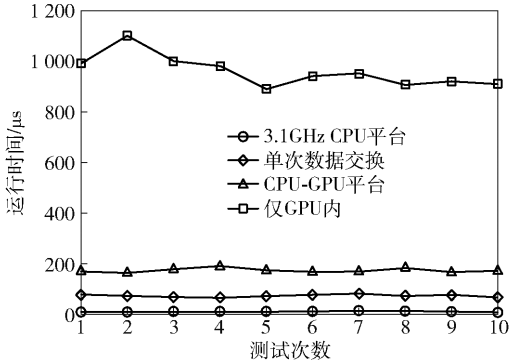


图6 不同运行平台中256QAM调制性能对比

5 结束语

提出了一种基于异构物理层的高性能异构计算加速的开源5G架构,以解决开源5G物理层在传统x86架构平台中处理性能不足的问题。基于OAI平台,首次研发了CPU-GPU异构计算加速的开源5G SDR平台,并对物理层LDPC编解码与QAM调制模块进行测速。测试结果显示,LDPC的解码可达到6.56 Gbit/s, GPU内QAM数据吞吐量可达到20.6 Gbit/s,显著提升了开源5G SDR平台的性能。但异构硬件之间的数据传输损耗时间占比较大的问题仍然需要更好的解决方式。在未来的工作中,该架构会更好的定义不同模块之间的接口,避免较大的传输时延,并考虑将多个串行物理层模块移植到GPU中去,避免数据在不同硬件平台间的来回交换。同时,随着异构计算加速的开源5G架构对不同异构硬件的支持以及开源5G SDR平台性能的进一步提升,人工智能相关的算法与应用可以更好地与无线网络相结合。

参考文献:

- [1] HAN W T, KNOPP R. Open AirInterface: a pipeline structure for 5G[C]//2018 IEEE 23rd International Conference on Digital Signal Processing. Shanghai: IEEE Press, 2018: 1-4.
- [2] ALAWIEH M, KASPAREK M, FRANKE N, et al. A high performance FPGA-GPU-CPU platform for a real-time locating system[C]//2015 23rd European Signal Processing Conference. Nice: IEEE Press, 2015: 1576-1580.
- [3] GAVRILOVSKA L, RAKOVIC V, ICHKOV A, et al. Flexible C-RAN: radio technology for 5G[C]//2017 13th International Conference on Advanced Technologies, Systems and Services in Telecommunications. Nis: IEEE Press, 2017: 255-264.
- [4] MAI S H, LIU Y Q. Implementation of web AR applications with fog radio access networks based on openair-interface platform[C]//2019 5th International Conference on Control, Automation and Robotics. Beijing: IEEE Press, 2019: 639-643.
- [5] REKOPUTRA N M, HARWAHYU R, SARI R F. Performance study of OpenAirInterface 5G system on the cloud platform managed by juju orchestration[C]//2019 International Seminar on Research of Information Technology and Intelligent Systems. Yogyakarta: IEEE Press, 2019: 211-216.
- [6] NAHUM C V, DE NÓVOA MARTINS PINTO L, TAVARES V B, et al. Testbed for 5G connected artificial intelligence on virtualized networks[J]. IEEE Access, 2020, 8: 223202-223213.
- [7] MANCO J, BAÑOS G G, HÄRRI J, et al. Prototyping V2X applications in large-scale scenarios using OpenAir-Interface[C]//2020 IEEE Vehicular Networking Conference. New York: IEEE Press, 2020: 1-4.
- [8] FIŞNE A, ÖZSOY A. Grafik processor accelerated real time software defined radio applications[C]//2017 25th Signal Processing and Communications Applications Conference. Antalya: IEEE Press, 2017: 1-4.
- [9] HUSSAIN T, KHAN M, REHMAN M U, et al. A high performance software defined radio system architecture and development environment for a wide range of applications[C]//2018 International Conference on Computing, Mathematics and Engineering Technologies. Sukkur: IEEE Press, 2018: 1-5.
- [10] FANG X, CHEN S M. The design and algorithm mapping of a heterogeneous multi-core processor for SDR[C]//APCCAS 2008 - 2008 IEEE Asia Pacific Conference on Circuits and Systems. Macao: IEEE Press, 2008: 1086-1089.
- [11] MAROJEVIC V, REVES X, GELONCH A. Computing resource management for SDR platforms[C]//IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications. Berlin: IEEE Press, 2005: 685-689.
- [12] MAROJEVIC V, BALLESTE X R, GELONCH A. A computing resource management framework for software-defined radios[J]. IEEE Transactions on Computers, 2008, 57(10): 1399-1412.