文章编号：1003-0530(2022)11-2221-13

一体化信号处理与先进处理架构展望

梁兴东[1,2] 李焱磊[1,2] 刘云龙[1] 郭宇豪[1] 解玉凤[3] 徐兴元[4] 刘 柳[1,2] 刘文成[1,2]

（1. 中国科学院空天信息创新研究院，微波成像技术国家级重点实验室，北京 100190；
2. 中国科学院大学电子电气与通信工程学院，北京 100049；3. 复旦大学微电子学院，上海 200433；
4. 北京邮电大学电子工程学院，北京 100876）

摘 要： 多功能一体化系统是电子信息技术领域的重要发展方向之一，一体化信号处理是其中的关键技术，对实现各种功能之间资源共享与高效协同具有重大意义，同时也从算法到处理架构提出了新的要求。本文首先对一体化信号处理的研究现状进行了归纳总结，给出一种基于一体化信号的多功能系统模型，指出进一步开发空域资源是当前一体化信号设计的主要研究方向之一，并且给出相关信号处理的典型算法；然后，针对上述算法进行了核心算子提炼，指出该类算法以大规模矩阵运算为主要特征，对处理架构提出高算力、高能效的需求；最后，针对上述算法的核心算子设计了基于存内计算和光子计算的处理架构，由于避免了数据搬移并采用高性能模拟计算模式，因此可以大幅提升算力和能效，为一体化信号处理先进架构设计提供了新的技术途径。

关键词： 一体化信号处理；时空频多维联合波形设计；存内计算；光子计算

中图分类号： TN951 **文献标识码：** A **DOI：** 10. 16798/j. issn. 1003-0530. 2022. 11. 001

引用格式： 梁兴东，李焱磊，刘云龙，等. 一体化信号处理与先进处理架构展望[J]. 信号处理，2022，38(11)：2221-2233. DOI：10. 16798/j. issn. 1003-0530. 2022. 11. 001.

Reference format：LIANG Xingdong，LI Yanlei，LIU Yunlong，et al. Prospect of integrated signal processing and advanced processing architecture [J]. Journal of Signal Processing，2022，38 (11)：2221-2233. DOI: 10. 16798/j. issn. 1003-0530. 2022. 11. 001.

Prospect of Integrated Signal Processing and Advanced Processing Architecture

LIANG Xingdong[1,2] LI Yanlei[1,2] LIU Yunlong[1] GUO Yuhao[1] XIE Yufeng[3] XU Xingyuan[4]
LIU Liu[1,2] LIU Wencheng[1,2]

(1. National Key Laboratory of Microwave Imaging Technology，Aerospace Information Research Institute，Chinese Academy of Science，Beijing 100190，China；2. School of Electronic，Electrical and Communication Engineering，University of Chinese Academy of Sciences，Beijing 100049，China；3. School of Microelectronics，Fudan University，Shanghai 200433，China；4. School of Electronic Engineering，Beijing University of Posts and Telecommunications，Beijing 100876，China)

Abstract: The multifunctional integrated system is one of the development directions in the electronic information technology field. Integrated waveform，the crucial technology in the multifunctional integrated system，has great significance for resource sharing and efficient cooperation between various functions. Meanwhile，new requirements are raised from algorithm to processing architecture. In this study，the research status of the integrated waveform processing is summarized on the comb，and an integrated signal processing system model is suggested. The further development of spatial resources is pointed out as a substantial direction in the research of integrated waveform processing. And the typical algorithms of correlation signal processing are given. Then，the large-scale matrix operation is proposed as the key characteristic by abstracting the core

收稿日期：2022-08-16；修回日期：2022-11-30
基金项目：国家自然科学基金(62271471)

operator of this kind of integrated signal processing method. The corresponding system processing architecture is required to have high computing power and high energy efficiency. Finally, the In-Memory computing and photonic computing advanced architectures are designed to solve the above requirement. By reducing the data movement and employing high-performance analog computing mode, these two architectures significantly improve the capability of high computing power and high energy efficiency, providing a novel technical approach for integrated waveform advanced processing architecture research.

Key words: integrated signal processing; temporal-spatial-spectral multi-dimensional joint waveform; in-memory computing; photon computing

1 引言

随着电子信息技术的飞速发展,为了满足不断涌现的各种应用需求,多功能一体化电子信息系统成为大势所趋[1-8]。为了提升系统的感知能力和反应速度,需要在同一平台上搭载雷达、通信终端等多种电子信息系统。这些系统在提高系统综合性能的同时,也会导致体积、重量和功耗大为增加;并且由于各系统间缺乏统一的规划设计,系统冗余和频谱冲突等问题非常突出,因此,多功能一体化系统成为解决上述难题的不二选择[3-4]。此外,在5G/6G移动通信[5-8]中,多功能一体化系统同样具有广阔的应用前景,智慧家居[7]、自动驾驶[8]等应用要实现传感器之间高速率通信,同时要具备环境感知能力。为了解决各种功能之间由于频谱冲突造成的电磁空间资源紧张问题,必须有效解决多功能一体化信号处理问题[9-11]。

一体化信号处理主要包括发射端的一体化信号设计与实时生成[12-18]和接收端的信号分离与协同处理[19-21]。发射端一体化信号设计与生成具体是指:通过对信号幅度、频率和相位等参数的配置,形成同时承载雷达探测和无线通信等功能的一体化信号,并在实际应用中利用高性能计算实时生成一体化信号波形。为了同时满足多种功能的需求,需要联合时间、频率、空间等维度资源以提供更多的自由度。接收端的信号分离与处理具体是指:根据一体化信号中各功能的承载方式,在接收端完成不同功能信号的分离,利用雷达探测、通信解调等相关处理方法实现相应的功能。随着一体化信号维度数量的增加,一体化信号处理对系统的算力需求呈几何级数增长。同时,搭载于轻小型平台的一体化信号处理系统具有广阔的应用前景[22],在此类应用中系统的尺寸、重量和功耗(size, weight and power, SWaP)严格受限。因此,一体化信号处理要求系统架构具有高算力、低功耗(即高能效)的特征。

在采用冯·诺依曼架构的系统中,由于总线的传输带宽受限,因此系统难以满足一体化信号处理的算力需求。此外,系统中一次单精度的基本运算只需要几皮焦耳的能量,而从存储器中进行数据检索和搬移则需要消耗上千皮焦耳的能量,远超于计算所需的能量。因此,冯·诺依曼架构的系统能效极低,无法满足一体化系统高能效的需求。针对上述需求目前可供选择的方案包括:增加处理单元(GPU)[23]、对指令进行流水化设计(DSP)[24]、采用数据流驱动(FPGA)[25]和采用面向特定领域的处理架构(DSA)[26]等。GPU增加处理单元提高了并行处理的规模,虽然能够大幅增加算力,但是其每个处理单元仍采用串行处理方式,导致系统功耗过大。DSP采用哈佛架构,在冯·诺依曼架构的基础上,通过增加总线数量的方式提高了系统传输带宽,故而更适合计算密集型的应用场合,但其串行处理的特点导致提升算力只能依靠主频的提高和核心数量的增加,从而限制了算力和能效的进一步提高。FPGA采用岛式架构,具备硬件可编程的能力,但这种架构限制了其工作频率的提升,同时冗余的布线资源造成了额外的功耗,导致无法大幅提升处理能效。近年来DSA技术发展迅速,面向卷积神经网络这一特定域的处理需求,谷歌研发出张量处理器(TPU);类似地,针对一体化信号处理的高算力、高能效处理需求,应研发相应的特定域处理架构。

2 一体化信号处理及其算力需求分析

如前所述,一体化系统信号处理主要包括发射端的一体化信号设计与实时生成和接收端的一体化信号分离与协同处理(如图1所示)。基于电磁波承载物理信息的本质,雷达探测、无线通信等多功能的同时实现离不开发射端的一体化信号设计与生成,通过对信号的幅度、频率、相位、空间导向矢量等可调参数进行编码设计,使得一体化信号具备高效的频谱资源利用率和更加灵活的功能配置能

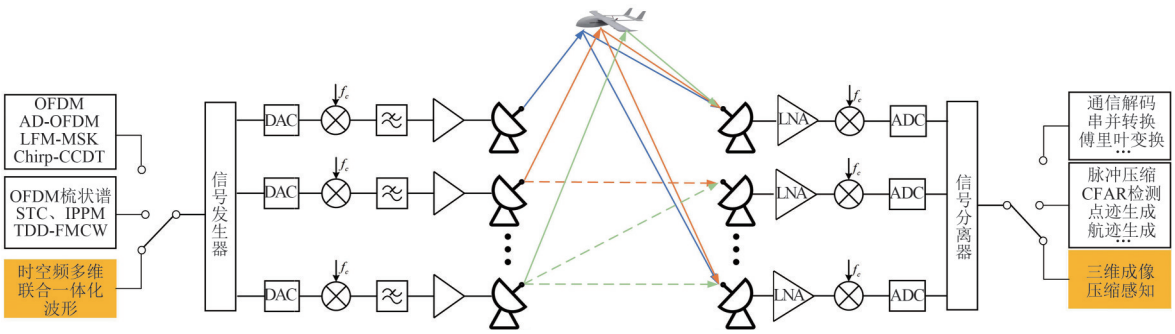

图 1　基于一体化信号的多功能系统模型

Fig. 1　Multifunctional system model based on integrated signal

力。根据信号资源的利用方式，一体化波形包括共用波形和复用波形两类，其中共用波形分为基于雷达波形的共用波形[18, 27]和基于通信波形的共用波形[17, 28-30]，复用波形分为时频复用波形[12-14, 31-32]和时空频多维联合波形[15-16, 33-34]等。

基于雷达波形的共用波形[18, 27]通过对常用雷达波形（如线性调频信号）的相位、幅度或脉冲重复间隔进行编码以携带通信信息，在接收端通过脉冲压缩等雷达信号处理方法实现雷达探测等功能，并根据编码方式对接收信号进行解码获取通信信息，算力需求与单功能处理方法相当。基于通信波形的共用波形[17, 28-30]可直接利用通信波形（如正交频分复用（Orthogonal Frequency Division Multiplexing，OFDM）信号）来完成通信和探测功能，其中探测功能主要通过基于匹配滤波或失配滤波[30]的脉冲压缩来实现，整个过程主要涉及线性卷积（向量乘法）、向量加法和傅里叶变换等运算，算力需求与单功能处理方法相当。

时频复用波形[12-14, 31-32]将时间、频率等维度资源分割成相互正交的子集，分别加载传统雷达波形和通信波形。以 OFDM 梳状谱一体化波形实现探测、通信功能[32]为例，在发射信号生成时，可直接利用逆傅里叶变换完成一体化信号快速生成，在接收端进行接收信号处理时，可直接利用傅里叶变换提取所有子载波的信息，子载波分离难度低，算力需求与单功能处理方法相当。时空频多维联合波形[15-16, 33-34]是联合时间、频率、空间等维度资源的一体化信号设计方法，具有在任意空间、任意时间、任意频段生成任意信号的潜力。该方案不再限制一体化波形所属类别，充分开发波形设计可利用的自由度，在空间相参合成各功能的指定波形；接收端联合多个节点进行协同处理，保留数据空间结构特性，获得相参处理增益，整个过程中涉及大量矩阵乘法、矩阵分解、矩阵求逆（求伪逆）等算子，算力需求高达 TFLOPS 甚至 PFLOPS 量级。

基于雷达波形的共用波形、基于通信波形的共用波形和时频复用波形在信号生成与处理中，面临的计算压力与单功能处理压力相当，以对长度为 N 的通信共用波形进行傅里叶变换运算为例，其计算复杂度为 $O\left(N\cdot\log_2 N\right)$，利用现有处理架构即可快速完成计算；而发射端的时空频多维联合波形设计和接收端的多维信号处理因信号维度的增加，给一体化系统带来了巨大的计算压力。因此，本文将对时空频多维联合波形的信号处理方法、算力需求进行深入分析，在此基础上提出先进架构实现方案。

2.1　发射波形设计及其算力需求分析

在时频复用的基础上，为了充分挖掘空域资源，P. M. McCormick 等人提出了一种基于数字阵列天线的远场辐射发射设计（Far-Field Radiated Emission Design，FFRED）信号模型[33]，通过多通道联合波形设计，将雷达信号与通信信号同时辐射至指定方向，其实现方式如图 2 所示。

它以相参合成雷达波形和通信波形为目标对一体化波形进行约束，综合考虑功率放大器的工作效率，建立一体化信号优化模型，

$$\begin{aligned}&\min_{\boldsymbol{S}}\left\|\boldsymbol{X}\right\|_{\mathrm{F}}^{2}\\&\text{s.t.}\quad \boldsymbol{C}^{\mathrm{H}}\boldsymbol{X}=\boldsymbol{G}\\&\left|x_m(n)\right|=\left|x_p(q)\right|\quad \forall n,q=0,\cdots,N-1\\&\qquad m,p=0,\cdots,M-1\end{aligned}\tag{1}$$

其中，$\boldsymbol{X}\in\mathbb{C}^{M\times N}$ 为一体化信号矩阵，$\boldsymbol{C}\in\mathbb{C}^{M\times K}$ 为阵列流行矩阵，$\boldsymbol{G}\in\mathbb{C}^{K\times N}$ 为期望功能波形矩阵，M、N、K 分别为阵元个数、采样点数与多功能目标个数。

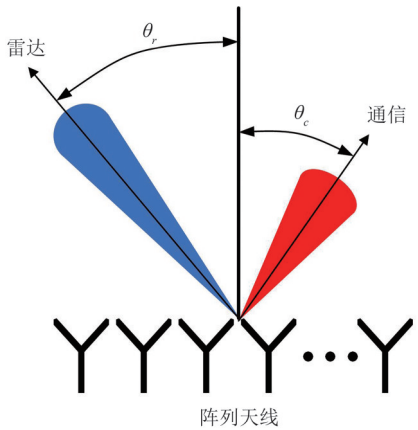

图 2 FFRED模型场景示意图[33]

Fig. 2 Schematic diagram of FFRED model scenario[33]

在对一体化信号优化模型求解时，优化模型为非凸模型，故将其拆分为两个可计算解析解的子凸优化模型迭代优化，直至满足收敛条件。FFRED模型处理流程如表1所示。

表 1 FFRED模型处理流程

Tab. 1 The processing chart of FFRED model

FFRED模型
输入：阵列导向矩阵 $\boldsymbol{C}$，期望雷达信号 $\boldsymbol{g}_r$，期望通信信号 $\boldsymbol{g}_c$，功率效率 ρ，迭代次数 ζ；
输出：发射波形矩阵 $\boldsymbol{X}$。
1. 初始化发射波形矩阵 $\boldsymbol{X}_0$；
2. 计算最小范数解 $\boldsymbol{X}_* = \boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C}\right)^{-1}\boldsymbol{G}$ 及其功率 $P_* = \left\|\boldsymbol{X}_*\right\|_{\mathrm{F}}^2 / MN$；
3. 计算信号幅度 $\gamma = \sqrt{P_* / 1-\rho}$；
4. 令 $i = i + 1$；
5. 等式约束下求解最小二乘问题，获得解 $\tilde{\boldsymbol{X}}_i = \boldsymbol{X}_* + \left(\boldsymbol{I}_M - \boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^{\mathrm{H}}\right)\boldsymbol{X}_{i-1}$；
6. 恒模约束下求解最小二乘问题，获得解 $\boldsymbol{X}_i = \gamma \exp\left(j\angle\left(\tilde{\boldsymbol{X}}_i\right)\right)$；
7. 总迭代次数大于 ζ，迭代停止，否则返回步骤4。

在步骤2中，主要涉及浮点级精度复数的矩阵乘法和矩阵求逆两种运算，矩阵乘法的操作数为 $16MK^2 + 8MNK$，矩阵求逆的操作数为 $16K^3$；在步骤5中，主要涉及矩阵乘法、矩阵求逆和矩阵加法三种运算，其中矩阵乘法的操作数为 $8M^2N + 8M^2K + 16MK^2$，矩阵求逆的操作数为 $16K^3$，矩阵加法的操作数为 $2M^2 + 2MN$；在步骤6中，主要涉及恒模运算和标量乘法两种，它们的操作数为 $16MN$；同时，步骤4至步骤7共需要迭代执行 ζ 次，对应的计算复杂度也将增大 ζ 倍。根据实际应用需求，取各参数的典型值如下：$M = 256$、$N = 1 \times 10^6$、$K = 2$、$\zeta = 20$，在2秒的相干处理时间内，整个优化过程的算力需求约为5.33 TFLOPS，其中矩阵加法和矩阵求逆等运算的计算压力较低，利用现有架构即可满足在线实时生成约束，而矩阵乘法的算力需求巨大，高达5.28 TFLOPS，约占据整个算力需求的99%。

2.2 接收信号处理及其算力需求分析

在接收端通过对分布式多节点接收信号或多通道接收信号进行相参处理，充分挖掘空域维度资源，实现雷达探测和无线通信等能力的提升。以多通道雷达三维成像为例，其主要任务为对距离-方位-俯仰三维信号进行反问题求解处理。面对三维观测数据，若采用传统方法，需将三维数据向量化处理，即使利用压缩感知算法降低采样率，计算过程中矩阵运算和向量运算仍需要耗费大量的计算和存储资源。根据回波数据的高维结构特性，邱伟将其定义为三阶张量，直接将压缩感知理论应用于张量数据，充分利用其内在的结构特征进行处理，有利于降低字典矩阵的内存消耗，进一步提高高维数据处理效率[35-36]。下面将对该算法的流程进行简要介绍。

在压缩感知框架下，接收数据与目标三维图像可以表示为

$$\mathcal{Z} = \mathcal{X} \times_1 \boldsymbol{\Theta}_r \times_2 \boldsymbol{\Theta}_c \times_3 \boldsymbol{\Theta}_v \tag{2}$$

其中，$\mathcal{Z}$ 为稀疏采样数据张量，大小为 $M_r \times M_c \times M_v$，$\mathcal{X}$ 为待求解的目标三维图像数据张量，大小为 $N_r \times N_c \times N_v$，$\boldsymbol{\Theta}_r \in \mathbb{C}^{M_r \times N_r}$、$\boldsymbol{\Theta}_c \in \mathbb{C}^{M_c \times N_c}$、$\boldsymbol{\Theta}_v \in \mathbb{C}^{M_v \times N_v}$ 分别为距离、方位、俯仰对应感知矩阵。

根据压缩感知理论，$\mathcal{X}$ 的重构模型为

$$\begin{aligned} &\min_{\mathcal{X}} \left\|\mathcal{X}\right\|_0 \\ &\text{s.t.} \quad \mathcal{Z} = \mathcal{X} \times_1 \boldsymbol{\Theta}_r \times_2 \boldsymbol{\Theta}_c \times_3 \boldsymbol{\Theta}_v \end{aligned} \tag{3}$$

利用SL0算法对该模型进行重构，算法流程如表2所示。

在步骤1中，主要涉及张量模式积、矩阵乘法和矩阵求逆三种运算，其中张量模式积的操作数为 $8 \times \left(N_r M_c M_v M_r + N_r N_c M_v M_c + N_r N_c N_v M_v\right)$，矩阵乘法的操作数为 $16 \times \left(N_r M_r^2 + N_c M_c^2 + N_v M_v^2\right)$，矩阵求逆的操作数为 $16 \times \left(N_r^3 + N_v^3 + N_v^3\right)$；在步骤2~步骤4中，主要涉及标量乘法、张量加法运算，操作数为 $2N_r N_c N_v$；在步骤5中，主要涉及张量模式积和张量加法两种运算，其中张量模式积的操作数为 $8 \times \left(N_r M_r M_c M_v + N_r N_c M_c M_v + N_r N_v N_v M_v + M_r N_r N_c N_v +\right.$

表 2　张量-SL0法处理流程

Tab. 2　The processing chart of tensor-SL0 method

张量-SL0法
输入：稀疏采样数据张量 $\mathcal{Z}$，感知矩阵 $\boldsymbol{\Theta}_r$、$\boldsymbol{\Theta}_c$、$\boldsymbol{\Theta}_v$，递减因子 η，迭代步长 μ，循环次数 K，迭代阈值 ρ_Δ，迭代次数 I；
输出：目标三维数据张量 $\mathcal{X}$。
1. 初始化目标数据张量 $\mathcal{X}_0 = \mathcal{Z} \times_1 \boldsymbol{\Theta}_r^\dagger \times_2 \boldsymbol{\Theta}_c^\dagger \times_3 \boldsymbol{\Theta}_v^\dagger$；
2. 初始化递减序列 $\rho_0 = 2 \cdot \max(\mathcal{X})$；
3. 计算高斯平滑函数 $\Delta\mathcal{X} = \left[\delta_{n_r n_c n_v}\right]$，其中 $\delta_{n_r n_c n_v} \triangleq x_{n_r n_c n_v} \cdot \exp\left(-\left\|x_{n_r n_c n_v}\right\|^2 / 2\rho_i\right)$；
4. 更新目标张量 $\mathcal{X} = \mathcal{X} - \mu\Delta\mathcal{X}$；
5. 向可行集投影 $\mathcal{X} = \mathcal{X} - \left(\mathcal{X} \times_1 \boldsymbol{\Theta}_r \times_2 \boldsymbol{\Theta}_c \times_3 \boldsymbol{\Theta}_v - \mathcal{Z}\right) \times_1 \boldsymbol{\Theta}_r^\dagger \times_2 \boldsymbol{\Theta}_c^\dagger \times_3 \boldsymbol{\Theta}_v^\dagger$
6. 将步骤3~步骤5重复执行 K 次；
7. 令 $i = i + 1$，$\rho_i = \eta\rho_{i-1}$；
8. 判断若 $\rho_i \leqslant \rho_\Delta$ 或 $i = I$，输出目标张量，否则重复执行步骤3~步骤7。

$M_r M_c N_c N_v + M_r M_c M_v N_v\big)$，张量加法的操作数为 $2 \times \big(M_r M_c M_v + N_r N_c N_v\big)$。假设 $N_r = N_c = N_v = 800$，$M_r = M_c = M_v = 500$，在整个运算中，张量模式积运算几乎占据了全部的算力开销，在2秒的相干处理时间内算力需求为6.2 TFLOPS，而张量模式积的本质仍为矩阵乘法。因此，多维一体化信号处理导致一体化系统面临较大的计算负担，需要设计适用于矩阵乘法的处理架构。

3　面向一体化信号处理的架构分析

一体化信号处理中多维信号涉及大量的矩阵乘法运算，对处理架构提出高算力需求，同时端平台自身存在SWaP约束，因此一体化信号处理系统的架构需要具备高算力、高能效的能力。现有主流处理器主要包括以下三个方面：1）通用处理器，如采用冯·诺依曼架构的CPU和GPU、采用哈佛架构的DSP等；2）采用数据流驱动的硬件可编程处理器，如FPGA；3）面向特定领域的专用加速器，如采用脉动阵列架构的TPU。上述三类主流处理器虽然能够满足一体化信号处理提出的TFLOPS量级高算力需求，但随之造成功耗急剧增加，无法满足端平台的SWaP约束，使得基于这几类处理器的一体化信号系统面临能效低的问题。而面向未来的先进处理架构，如以模拟信号为信息载体进行计算的存内计算、光子计算，具备兼顾高算力和高能效的潜力。因此，我们分别设计了适用于一体化信号处理的存内计算和光子计算先进架构，并与现有架构实现矩阵乘法运算的能效进行了对比，展示了其在一体化信号处理中的价值。

3.1　通用处理器

3.1.1　CPU

CPU采用的是冯·诺依曼架构，如图3所示，冯·诺依曼架构由运算器、控制器、存储器、输入设备以及输出设备组成。在程序的执行过程中，计算机先从内存中取出第1条指令，通过控制器的译码器接收指令的要求，再从存储器中取出数据，将数据给到运算器中，然后进行指定的运算和逻辑操作等，随后按照指令中的地址把结果送到内存中，接下来取出第2条指令执行，直到遇到停止指令。因此，在冯·诺依曼架构中程序被编码为数据存储在存储器中，需要运行时只需从存储器中依次取出、执行即可，这极大地降低了编程的难度，使得冯·诺依曼架构具有较高地灵活性。然而这种从存储器中读取指令和数据执行的设计也使得冯·诺依曼架构天然地受到信息传输带宽的影响。以IBM公司的Power9为例，当其进行各种DeepSpeech基准测试的通用矩阵运算时，可以在130 W功耗下实现486 GFLOPS的最高算力，对应的性能功耗比为1.62 GFLOPS/W[37]。当利用CPU进行多维信号处理时，大量的数据搬移将极大地增加冯·诺依曼架构系统的延迟和能量消耗，限制系统的算力和能效。

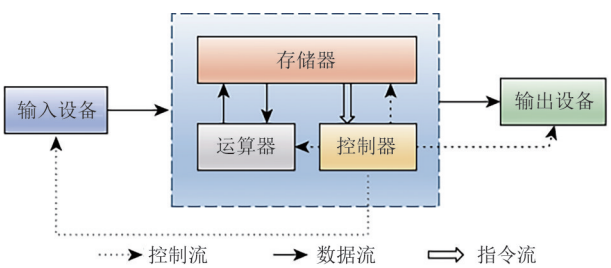

图3　冯·诺依曼计算架构[38]

Fig. 3　Von Neumann computing architecture[38]

3.1.2　GPU

GPU是当前主流加速器之一，从最初用作图形处理器到后来用于通用计算加速，在数据中心加速等应用的推动下，GPU的性能有了显著的提高[23]，架构如图4所示。与CPU相比，GPU去掉了复杂的控制电路和大量的片上高速缓存，能够集成大量的计算核心，这种通过众核方式增加并行度的计算架构，使得GPU更适合大规模同质化数据的并行处理。

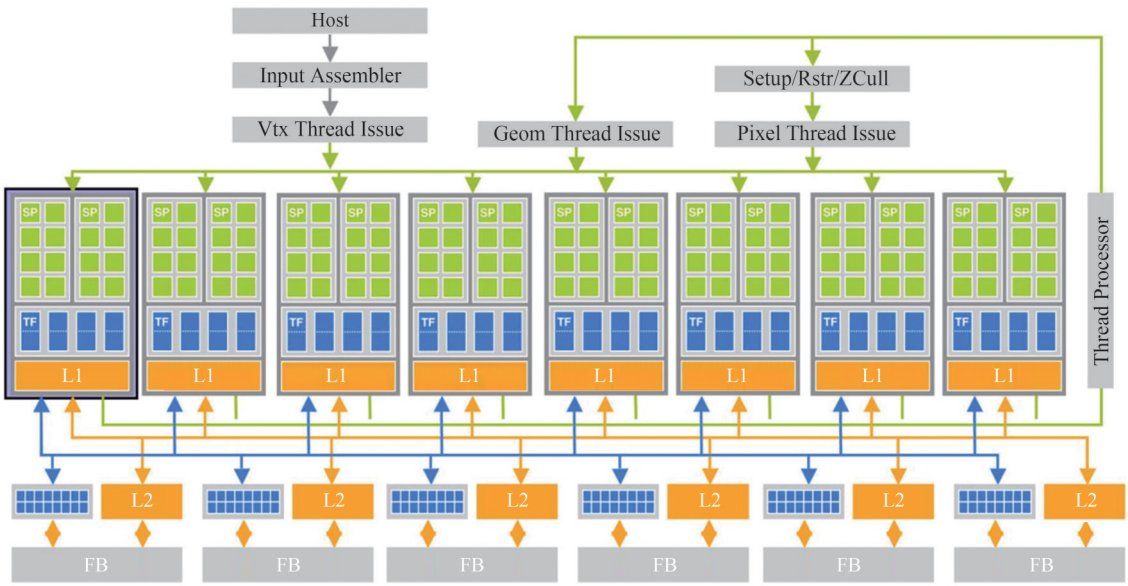

图4 GPU架构示意图[23]

Fig. 4 GPU architecture diagram[23]

以Nvidia公司的V100为例，在进行各种DeepSpeech基准测试的通用矩阵运算时，可以在300 W功耗下实现7.8 TFLOPS的最高算力，对应的性能功耗比为26 GFLOPS/W。虽然GPU可以通过集成更多的核心和更大的内存带宽提高了算力，但由于每个计算核心仍采用串行处理方式，计算核心数量的增加会导致功耗增大，其能效优势并不明显，不适合一体化信号处理这种需要高能效的应用场景。

3.1.3 DSP

DSP是数字信号处理常用的处理器之一[24]，采用如图5所示的哈佛架构。与CPU指令和数据共用同一存储器不同，该架构将指令和数据分开存储，并对指令进行了流水线优化设计，同时集成了数字信号处理常用的乘法器硬件电路，使得DSP完成计算的指令周期大大缩短，提高了对数字信号处理的算力，适用于计算密集型的应用场景。以TI公司的C66XX系列DSP组成板卡为例，由6片DSP组成的模块进行矩阵运算时，能够在267.1 W功耗下实现938.21 GFLOPS的算力，对应的性能功耗比为3.51 GFLOPS/W[39]。虽然DSP能够为数字信号处理提供高计算精度，但是其串行处理的特点导致算力的提升只能依靠主频的提高和核心数量的增加，限制了DSP算力和能效的进一步提高，不能满足一体化信号处理需求。

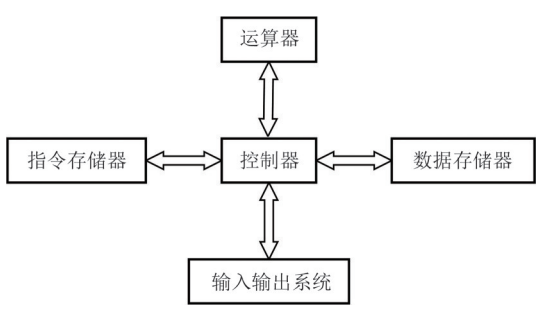

图5 哈佛架构示意图[40]

Fig. 5 Harvard architecture diagram[40]

3.2 FPGA

与冯·诺依曼架构的控制流驱动不同，目前主流的FPGA芯片大多采用岛式架构来实现数据流驱动的方式，如图6所示[25]。逻辑块(Logic Block，LB)中成孤岛式分布，各个LB之间通过可编程布线资

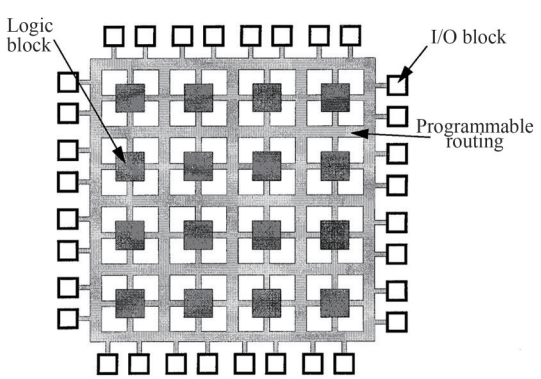

图6 通用FPGA架构示意图[25]

Fig. 6 A generic architecture of FPGA[25]

源连接，芯片与外界通过输入/输出块（Input/Output Block，I/O Block）进行数据的传输。在FPGA中，待处理的数据在时钟信号的驱动下可以直接流入LB中的运算单元进行计算，不再需要通过控制器的指令去进行数据的读写，运算单元的利用率相较于传统的冯·诺依曼得到了提高，而且众多的LB可以通过编程实现不同的运算功能也使得FPGA可以同时进行不同的运算，具有较高的并行度。以Xilinx公司的Ultrascale+系列的VU3P为例，在进行各种DeepSpeech基准测试的通用矩阵运算时，可以在23 W功耗下实现194 GFLOPS的最高算力，对应的性能功耗比为8.43 GFLOPS/W。然而FPGA的岛式结构让其具备硬件编程能力同时，这种灵活编程的能力使得FPGA的布线资源存在复杂、冗余等问题，增加了额外的功耗与延迟，从而限制了FPGA的工作频率与能效的提升。受限于此，FPGA并不能满足一体化信号处理的需求。

3.3　DSA

DSA是面向不同特定域需求研发的专用处理架构，例如TPU是谷歌研发的一款面向数据中心卷积神经网络（CNN）计算特定域的专用加速器[26]，其架构框图如图7所示。TPU架构的核心是采用脉动架构实现的矩阵乘法单元，高速缓存为矩阵乘法单元提供高带宽的数据流，使得TPU可以持续不断地进行矩阵乘法运算，脉动架构提高了矩阵乘法运算的访存效率，数据复用降低了功耗，使得TPU具备高算力和低功耗的能力。以TPU-V2为例，可以在280 W的功耗下实现将近20 TFLOPS的算力，性能功耗比可以达到71.43 GFLOPS/W[41]。TPU满足了CNN计算中较低精度（通常是Int8）下大量矩阵乘法等矩阵运算的加速需求，算力和能效相比GPU大幅提升，但由于TPU是面向CNN加速应用场景的，其计算精度无法满足一体化信号处理需求。

3.4　新型先进处理架构

存内计算、光子计算等以模拟信号作为信息载体进行计算的架构有计算速度快、能耗低等优势，具有很高的应用潜力。然而以模拟信号进行计算的架构受限与硬件技术，存在计算精度低（目前的精度大多是8比特整型）的问题，还无法满足一体化信号处理32位浮点的需求。但是相信，未来随着硬件技术的提升，模拟计算的精度会逐渐提升，从而满足一体化信号处理的需求。

3.4.1　存内计算架构

早在20世纪90年代，就已经有了存内计算（Compute in Memory，CIM）的架构概念，受到技术等因素的限制，当时存内计算架构并没有得到广泛的应用。后来，随着CMOS和存储技术的发展，以及人工智能的兴起，存内计算架构再次受到了人们的关注，其中比较知名的存内计算架构有FlexRAM[42]、DIVA[43]、Sandwich-RAM[44]、memristor-based CNN[45]

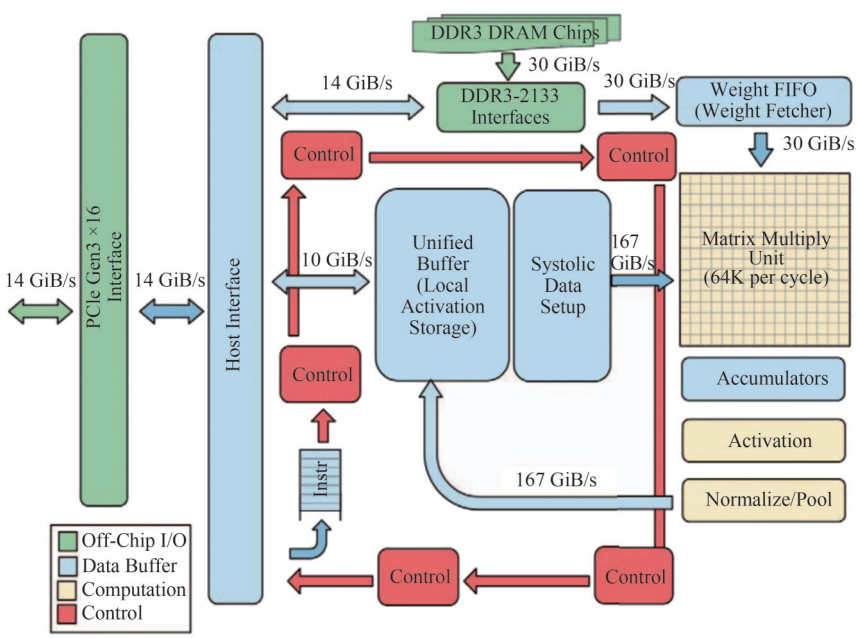

图7　TPU架构示意图[26]

Fig. 7　TPU architecture diagram[26]

等。存内计算架构的原理如图 8 所示，它将计算单元放入存储单元中，直接使用内存单元（如 SRAM、忆阻器等）的电阻、电流与电压关系进行计算。相较于传统的冯·诺依曼架构，由于存内计算架构中的计算单元与存储单元的结合更为密切，因此存内计算架构可以很好地减少数据搬移，从而降低能耗，提升系统性能。

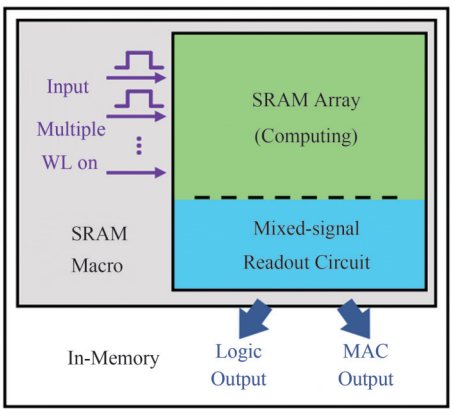

图 8 存内计算架构[46]

Fig. 8 In-Memory computing architecture[46]

目前，存内计算架构主要还是针对人工智能领域中的算法加速提出的，它们的精度大多是 INT8 等[45, 47-49]，如普林斯顿大学研究团队[47]提出了基于存内计算架构的深度神经网络加速器，解决了深度神经网络中大规模矩阵向量乘法的数据搬移问题，该架构在 1 bit 精度下的能效达到了 866 TOPS/W；清华大学研究团队[45]基于忆阻器实现了卷积神经网络并用来进行图像识别，计算能效达到了 11.014 TOPS/W；德克萨斯大学研究团队[48]提出的 8 bit 卷积存内计算架构，每个时钟周期最多可以实现 175 次乘累加运算，能效达到 14.4 TOPS/W。虽然现有的存内计算架构还无法满足一体化信号处理的需求，但其表现出的高能效特点使其在一体化信号处理研究中具有巨大的潜力。

3.4.2 光子计算架构

光子计算架构与存内计算架构类似，即数据在硬件系统中的实时位置与进行运算的位置相同，因而规避了冯·诺依曼瓶颈。此外，宽达数十太赫兹的光谱也为高速运算提供了充足的带宽，通过密集波分复用、空分复用、时分复用等光电信息技术手段，光子计算架构的并行度也可大幅提升，进而可实现万亿次运算每秒（TOPS）量级的超高单核运算速度。此外，模拟无源的光子架构也具有实现更高能效比的潜力，能量效率可达到 1 pJ/运算。因而，光子计算架构在模拟信号智能处理方面有广阔的应用空间。

目前国内外研究机构已对光子计算架构展开了深入研究，加州大学研究团队基于空间透镜光学实现了深度衍射神经网络[50]，牛津大学研究团队基于相变材料实现了并行矩阵运算[51]，麻省理工学院研究团队基于集成无源光学干涉器阵列实现了矩阵运算[52]，法国 FEMTO-ST 研究团队利用时分复用构建了光子水库运算结构[53]，澳大利亚斯威本科技大学团队提出并实现了基于时间、波长交织的光子卷积加速器[54]。其中澳大利亚斯威本科技大学团队提出的光子卷积加速器算力可以达到 11.3 TOPS，相较于高速的光学神经网络（Optical Neural Network，ONN），算力提升了 500 倍，原理如图 9 所示。输入向量 $\boldsymbol{X}$ 被编码在电信号的强度上，卷积核由一个长度为 $\boldsymbol{R}$ 的权向量 $\boldsymbol{W}$ 表示，该向量被编码在光梳的功率上。将带有向量 $\boldsymbol{X}$ 的电信号通过电光调制器（EOM）调制到光频梳上，然后通过色散延迟传播，相邻波长间延迟一个元素的持续时间，最后通过光电二极管对信号进行求和，即可得到 $\boldsymbol{X}$ 和 $\boldsymbol{W}$ 之间卷积的结果 $\boldsymbol{Y}$。

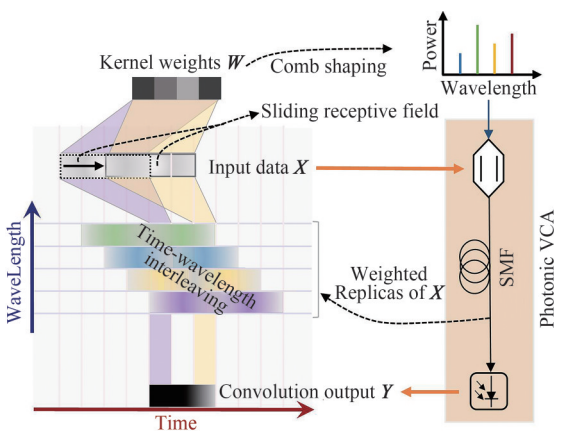

图 9 卷积的工作原理[54]

Fig. 9 The working principle of convolution[54]

3.5 面向一体化信号处理架构的算力和能效比较

在一体化信号处理中经常需要单精度浮点级的运算，且常常涉及到复数运算，而现有的先进架构无法满足一体化信号处理技术的需求，因此我们设计了支持浮点级复数矩阵乘法运算的存内计算架构和支持矩阵乘法运算的光子计算架构，并与表 3 所示的现有主流处理器的典型器件进行能效对比。

表3　主流处理器的典型器件

Tab. 3　Typical components of mainstream processors

架构类型	典型器件
CPU	Power9
GPU	V100
DSP	C66XX
FPGA	VU3P
TPU	TPU-v2

基于存内计算实现复数矩阵乘法 $\boldsymbol{R}=\boldsymbol{X}\times\boldsymbol{Y}$ 的架构如图10所示，使用一个脉动阵列来完成复数矩阵的乘法运算时，脉动阵列的每一计算单元需要完成复数的乘加操作，因此可以将复数的乘加操作分解为2个实数的乘加操作，分两个周期完成，其中实数的乘加主要为浮点数的乘加。浮点数的乘加可分解为指数部分和尾数部分，尾数部分是乘法计算，由存内计算乘加单元完成，指数部分由CMOS电路完成，最后两部分运算数据经过整合后为浮点乘加运算结果。

基于上述架构，我们初步设计了8×8复矩阵乘法运算，并分析BFP16精度和FP32精度下的存内计算性能，其结果如表4所示，相较于TPU和FPGA分别BFP16精度下提升了6.85与7.59倍。存内计算架构的算力可随着矩阵乘法规模的扩大进一步增加，例如对于64×64复矩阵乘法运算的存内计算加速器，其算力相较于8×8的存内计算加速器在算力上提升了64倍，可以在BFP16精度下达到745 GFLOPS，通过9片加速器并行处理即可满足一体化信号处理中TFOPLS量级的高算力需求，同时芯片规模的增加对存内计算能效的影响很小，所以存内计算在高算力的同时兼顾了高能效的需求。因此我们认为存内计算架构在一体化信号处理中具有巨大的应用潜力，未来随着计算精度的进一步提升，存内计算架构会得到广泛地应用。

表4　存内计算性能分析

Tab. 4　In-Memory computing performance analysis

精度	功耗（W）	算力（GFLOPS）	能效（GFLOPS/W）
BFP16	0.02077	11.64	560.22
FP32	0.02677	1.94	72.45

光子计算架构实现矩阵乘法的工作原理如图11所示，其中列向量 $\boldsymbol{A}$ 被编码在光梳的功率上，将矩阵 $\boldsymbol{B}$ 中的元素进行排列加载至电信号上。将该电信号通过EOM调制到光频梳上，然后通过色散延迟传播，相邻波长间延迟一个元素的持续时间，最后通过光电二极管进行求和。对光电二极管求和的结果按照相应的间隔进行提取，再进行排列，就可以得到矩阵 $\boldsymbol{B}$ 与列向量 $\boldsymbol{A}$ 的计算结果列向量 $\boldsymbol{C}$。因此，通过重复将不同的向量编码至光梳的功率上，然后重复上述操作，就可以得到两个矩阵相乘的结果，从而实现矩阵乘法的功能。

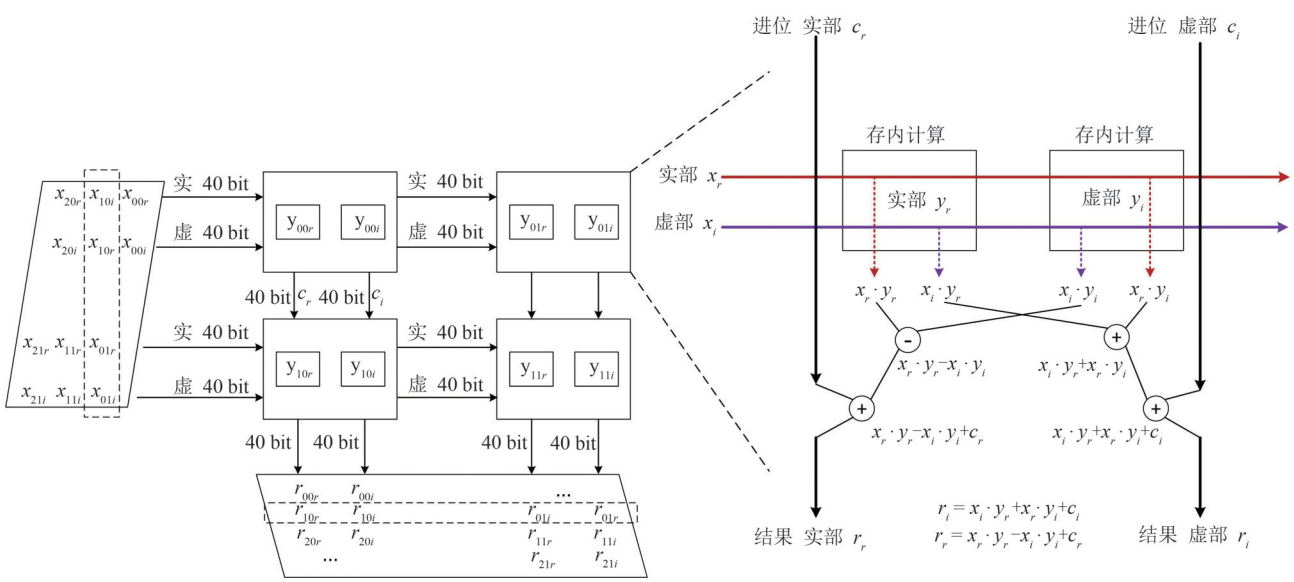

图10　基于存内计算架构的矩阵乘法

Fig. 10　Matrix multiplication based on In-Memory computing architecture

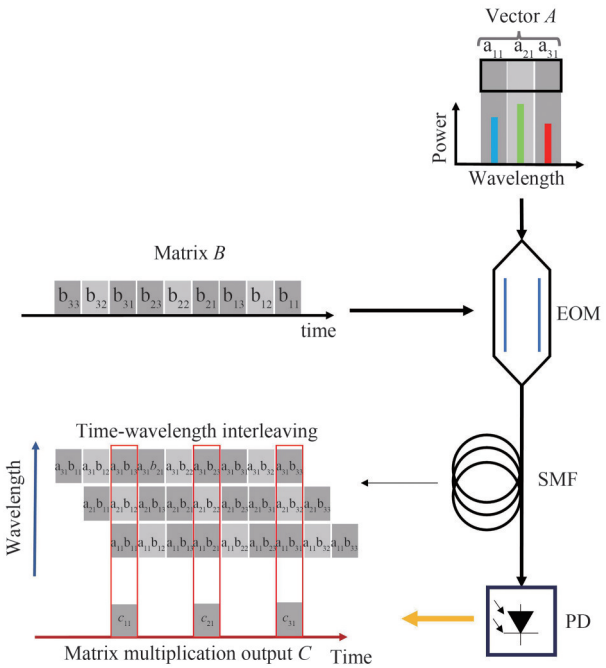

图 11 基于光子计算架构的矩阵乘法

Fig. 11 Matrix multiplication based on photonic computing architecture

我们采用吞吐量对光子计算架构的算力进行评估，即计算输出数据速率与每个输出码元所需运算次数的乘积。光子卷积加速器的输出数据速率为62.9 GBaud/s，每个卷积核可同时支持9根光梳进行运算，所以每个输出码元由9次加法与9次乘法运算得到，该加速器共有十个并行卷积核，因此最终算力为62.9G×(9+9)×10=11.322 TOPS。如果用该加速器进行矩阵乘法操作，则有效的输出码元为原来的1/9，最终算力仍有1.258 TOPS。未来通过进一步扩展频域、空间等维度的并行度，可以大幅度提升光子加速器的算力。例如，通过使用S、L、C三个光通信波段，可利用的频谱宽度可以达到20 THz，从而支持405个50 GHz间隔的并行波长通道。结合偏振复用与10路空分复用，整体算力可达62.9G×405×2×2×10=1.019 POPS。由于光计算架构为存算一体的模拟架构，无需数据往复读取，因而其功耗主要来源于光源。采用自泵浦克尔光频梳产生技术，光频梳所需能耗低至100 mW，总能耗预计小于1 W，因而未来总体能效预计可达1 W/1POPS=1 fJ/OPS。由表5可知。光子计算架构在算力和能效上均远高于其他架构，因此在高算力一体化信号处理的应用中具有很高的应用潜力。然而受限于硬件技术，目前光子计算架构的精度只有INT8，还无法满足一体化信号处理单精度浮点的需求。但是我们相信，未来随着硬件技术以及算法的改进，光子计算架构终会广泛地应用于一体化信号处理中。

表 5 架构性能功耗比分析

Tab. 5 Analysis of architecture performance power consumption ratio

架构类型	功耗（W）	算力（GOPS）	精度	能效（GOPS/W）
CPU	130	486	双精度	1.62
GPU	300	7800	双精度	26
DSP	267.1	938.21	双精度	3.51
FPGA	23	194	单精度	8.43
TPU	280	20000	半精度	71.43
CIM	0.02077	11.64	半精度	560.42
CIM	0.02677	1.94	单精度	72.45
optical CA	1	1019000	整型	1019000

4 结论

多功能一体化系统利用一体化信号，在同一框架下通过硬件复用和波形共享的方式，同时满足雷达探测和通信信息传输等功能需求，可有效缓解频谱冲突，提高系统的集约性。本文通过分析一体化信号处理的研究现状和发展规律，指出时空频联合多维波形设计是一体化信号研究的发展方向之一。从发射端的一体化信号设计与生成、接收端的信号分离与处理两个方面，对时空频联合多维波形一体化信号处理的计算法复杂度进行了分析，指出其具有高维、高计算复杂度的特征，现有处理架构无法满足一体化信号处理需求。基于存内计算和光子计算等技术设计的新型先进专用处理架构，具备高算力、高能效的特征，为未来一体化信号处理及其先进处理架构研究提供了技术途径。

参考文献

[1] ZHANG J A, RAHMAN M L, WU K, et al. Enabling joint communication and radar sensing in mobile networks—a survey[J]. IEEE Communications Surveys & Tutorials, 2022, 24(1): 306-345.

[2] 梁兴东, 李强, 王杰, 等. 雷达通信一体化技术研究综述[J]. 信号处理, 2020, 36(10): 1615-1627.
LIANG Xingdong, LI Qiang, WANG Jie, et al. Joint wireless communication and radar sensing: Review and future prospects[J]. Journal of Signal Processing, 2020, 36(10): 1615-1627. (in Chinese)

[3] JACYNA G M, FELL B, MCLEMORE D. A high-level overview of fundamental limits studies for the DARPA SSPARC program [C]//2016 IEEE Radar Conference (RadarConf). Philadelphia, PA, USA. IEEE, 2016:1-6.

[4] CHIRIYATH A R, PAUL B, BLISS D W. Radar-communications convergence: Coexistence, cooperation, and co-design[J]. IEEE Transactions on Cognitive Communications and Networking, 2017, 3(1): 1-12.

[5] LIYANAARACHCHI S D, RIIHONEN T, BARNETO C B, et al. Optimized waveforms for 5G-6G communication with sensing: theory, simulations and experiments[J]. IEEE Transactions on Wireless Communications, 2021, 20(12): 8301-8315.

[6] WILD T, BRAUN V, VISWANATHAN H. Joint design of communication and sensing for beyond 5G and 6G systems[J]. IEEE Access, 2021, 9(30): 845-857.

[7] ZHOU Zimu, WU Chenshu, YANG Zheng, et al. Sensorless sensing with WiFi[J]. Tsinghua Science and Technology, 2015, 20(1): 1-6.

[8] ZHANG Q, SUN H, WEI Z, et al. Sensing and communication integrated system for autonomous driving vehicles [C]// IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2020: 1278-1279.

[9] WANG Jie, LIANG Xingdong, CHEN Longyong, et al. First demonstration of airborne MIMO SAR system for multimodal operation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.

[10] 刘永军, 廖桂生, 李海川, 等. 电磁空间分布式一体化波形设计与信息获取[J]. 中国科学基金, 2021, 35(5): 701-707.
LIU Yongjun, LIAO Guisheng, LI Haichuan, et al. Distributed integrated waveform design and information acquisition in electromagnetic space[J]. Bulletin of National Natural Science Foundation of China, 2021, 35(5): 701-707. (in Chinese)

[11] 刘凡, 袁伟杰, 原进宏, 等. 雷达通信频谱共享及一体化:综述与展望[J]. 雷达学报, 2021, 10(3): 467-484.
LIU Fan, YUAN Weijie, YUAN Jinhong, et al. Radar-communication spectrum sharing and integration: Overview and prospect[J]. Journal of Radars, 2021, 10(3): 467-484. (in Chinese)

[12] WANG Jie, LIANG Xingdong, CHEN Longyong, et al. Joint wireless communication and high resolution SAR imaging using airborne MIMO radar system[C]//IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019: 2511-2514.

[13] WANG Jie, LIANG Xingdong, CHEN Longyong, et al. First demonstration of joint wireless communication and high-resolution SAR imaging using airborne MIMO radar system[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 6619-6632.

[14] WANG Jie, XIN Yue, LIANG Xingdong, et al. Interpulse phase modulation waveform scheme for spaceborne MIMO SAR systems[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(6): 4051-4066.

[15] LIU Fan, ZHOU Longfei, MASOUROS C, et al. Toward dual-functional radar-communication systems: Optimal waveform design[J]. IEEE Transactions on Signal Processing, 2018, 66(16): 4264-4279.

[16] HASSANIEN A, AMIN M G, ZHANG Y D, et al. Dual-function radar-communications: information embedding using sidelobe control and waveform diversity[J]. IEEE Transactions on Signal Processing, 2016, 64(8): 2168-2181.

[17] LIU Yongjun, LIAO Guisheng, XU Jingwei, et al. Adaptive OFDM integrated radar and communications waveform design based on information theory[J]. IEEE Communications Letters, 2017, 21(10): 2174-2177.

[18] BERGGREN F, POPOVIĆ B M. Joint radar and communications with multicarrier chirp-based waveform[J]. IEEE Open Journal of the Communications Society, 2022, 3: 1702-1718.

[19] HAVARY-NASSAB V, SHAHBAZPANAHI S, GRAMI A. Joint receive-transmit beamforming for multi-antenna relaying schemes[J]. IEEE Transactions on Signal Processing, 2010, 58(9): 4966-4972.

[20] DING Zihang, XIE Junwei. Joint transmit and receive beamforming for cognitive FDA-MIMO radar with moving target [J]. IEEE Sensors Journal, 2021, 21(18): 20878-20885.

[21] ZHANG J A, LIU Fan, MASOUROS C, et al. An overview of signal processing techniques for joint communication and radar sensing[J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(6): 1295-1315.

[22] KNOWLES J. DARPA to develpo multifunction RF system for group 3 UASs[J]. The Journal of Electronic Defense, 2016, 39(6): 15-15.

[23] OWENS J D, HOUSTON M, LUEBKE D, et al. GPU computing[J]. Proceedings of the IEEE, 2008, 96(5): 879-899.

[24] 阮进, 曾浩, 高远. 基于DSP平台并行计算方式浅析[J]. 科技创新导报, 2018, 15(31): 106, 108.
RUAN Jin, ZENG Hao, GAO Yuan. Analysis of parallel computing approach based on DSP platform[J]. Science and Technology Innovation Herald, 2018, 15(31): 106, 108. (in Chinese)

[25] BETZ V, ROSE J, MARQUARDT A. Architecture and CAD for Deep-Submicron FPGAS [M]. Boston, MA:

Springer US, 1999.

[26] JOUPPI N P, YOUNG C, PATIL N, et al. In-datacenter performance analysis of a tensor processing unit [C]// 2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture. Toronto, ON, Canada. IEEE, 2017: 1-12.

[27] CHEN Xingbo, WANG Xiaomo, XU Shanfeng, et al. A novel radar waveform compatible with communication[C]// 2011 International Conference on Computational Problem-Solving (ICCP). Chengdu, China. IEEE, 2011: 177-181.

[28] 刘永军, 廖桂生, 杨志伟. 基于OFDM的雷达通信一体化波形模糊函数分析[J]. 系统工程与电子技术, 2016, 38(9): 2008-2018.
LIU Yongjun, LIAO Guisheng, YANG Zhiwei. Ambiguity function analysis of integrated radar and communication waveform based on OFDM[J]. Systems Engineering and Electronics, 2016, 38(9): 2008-2018. (in Chinese)

[29] ZHU Shengkun, LI Xiaobai, YANG Ruijuan, et al. Adaptive waveform optimization method for OFDM radar communication jamming[C]//2021 IEEE International Conference on Consumer Electronics and Computer Engineering. Guangzhou, China. IEEE, 2021: 600-605.

[30] 张霄霄, 梁兴东, 王杰, 等. 融合失配处理和LMS滤波的雷达通信一体化OFDM信号距离旁瓣抑制技术[J]. 信号处理, 2021, 37(9): 1727-1738.
ZHANG Xiaoxiao, LIANG Xingdong, WANG Jie, et al. Range sidelobe suppression using mismatching and LMS adaptive filter for radar communication integrated OFDM signal[J]. Journal of Signal Processing, 2021, 37(9): 1727-1738. (in Chinese)

[31] MOGHADDASI J, WU Ke. Multifunctional transceiver for future radar sensing and radio communicating data-fusion platform[J]. IEEE Access, 2016, 4:818-838.

[32] WANG Jie, LI Yanlei, LIANG Xingdong, et al. Multidimensional waveforms for joint wireless communication and high resolution SAR systems[C]//2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), 2019: 1-4.

[33] MCCORMICK P M, BLUNT S D, METCALF J G. Simultaneous radar and communications emissions from a common aperture, Part I: Theory[C]//2017 IEEE Radar Conference (RadarConf). Seattle, WA, USA. IEEE, 2017: 1685-1690.

[34] JIANG Mengchao, LIAO Guisheng, YANG Zhiwei, et al. Integrated radar and communication waveform design based on a shared array[J]. Signal Processing, 2021, 182: 107956.

[35] 邱伟. 基于压缩感知的多维度雷达成像方法研究[D]. 长沙: 国防科学技术大学, 2014.
QIU Wei. Research on multidimensional radar imaging methodology by means of compressive sensing[D]. Changsha: National University of Defense Technology, 2014. (in Chinese)

[36] QIU Wei, ZHOU Jianxiong, ZHAO Hongzhong, et al. Three-dimensional sparse turntable microwave imaging based on compressive sensing[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(4): 826-830.

[37] DIAMANTOPOULOS D, HAGLEITNER C. HelmGemm: managing GPUs and FPGAs for transprecision GEMM workloads in containerized environments[C]//2019 IEEE 30th International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2019: 71-74.

[38] 蔡晓军, 栾峻峰, 申兆岩, 等. 面向冯·诺依曼计算机的指令执行虚拟仿真设计与探讨[J]. 实验技术与管理, 2022, 39(5): 89-93.
CAI Xiaojun, LUAN Junfeng, SHEN Zhaoyan, et al. Design and discussion of virtual simulation of instruction execution for von Neumann computer [J]. Experimental Technology and Management, 2022, 39(5): 89-93. (in Chinese)

[39] MITRA G, BOHMANN J, LINTAULT I, et al. Development and application of a hybrid programming environment on an ARM/DSP system for high performance computing[C]//2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2018: 286-295.

[40] Wikimedia Commons contributors. File:Harvard architecture.svg[OL]. Wikimedia Commons: 2020.09.06, 03:22 UTC [cited 2022.10.28]. Available from: https://commons. wikimedia. org/w/index. php?title=File: Harvard_architecture.svg&oldid=449340684.

[41] ZHOU Yangjie, YANG Mengtian, GUO Cong, et al. Characterizing and demystifying the implicit convolution algorithm on commercial matrix-multiplication accelerators [C]//2021 IEEE International Symposium on Workload Characterization (IISWC), 2021: 214-225.

[42] TORRELLAS J. FlexRAM: Toward an advanced Intelligent Memory system: A retrospective paper [C]//2012 IEEE 30th International Conference on Computer Design. Montreal, QC, Canada. IEEE, 2012: 3-4.

[43] DRAPER J, KANG C W, KIM I, et al. The architecture of the DIVA processing-in-memory chip[C]//Proceedings of the 16th international conference on Supercomputing - ICS '02. New York, New York, USA. New York: ACM Press, 2002.

[44] YANG Jun, KONG Yuyao, WANG Zhen, et al. 24.4 sandwich-RAM: An energy-efficient in-memory BWN architecture with pulse-width modulation [C]//2019 IEEE International Solid- State Circuits Conference -. San Francisco, CA, USA. IEEE, 2019: 394-396.

[45] YAO Peng, WU Huaqiang, GAO Bin, et al. Fully hardware-implemented memristor convolutional neural network[J]. Nature, 2020, 577(7792): 641-646.
[46] JHANG C J, XUE Chengxin, HUNG J M, et al. Challenges and trends of SRAM-based computing-in-memory for AI edge devices[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2021, 68(5): 1773-1786.
[47] VALAVI H, RAMADGE P J, NESTLER E, et al. A 64-tile 2.4-Mb In-Memory-computing CNN accelerator employing charge-domain compute[J]. IEEE Journal of Solid-State Circuits, 2019, 54(6): 1789-1799.
[48] ROHAN J N, KULKARNI J P. Systolic-RAM: scalable direct convolution using In-Memory data movement[J]. IEEE Solid-State Circuits Letters, 2022, 5: 142-145.
[49] KHADDAM-ALJAMEH R, MARTEMUCCI M, KERSTING B, et al. A multi-memristive unit-cell array with diagonal interconnects for In-Memory computing[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021, 68(12): 3522-3526.
[50] LIN Xing, RIVENSON Y, YARDIMCI N T, et al. All-optical machine learning using diffractive deep neural networks[J]. Science, 2018, 361(6406): 1004-1008.
[51] FELDMANN J, YOUNGBLOOD N, KARPOV M, et al. Parallel convolutional processing using an integrated photonic tensor core[J]. Nature, 2021, 589(7840): 52-58.
[52] SHEN Yichen, HARRIS N C, SKIRLO S, et al. Deep learning with coherent nanophotonic circuits[J]. 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM), 2017: 189-190.
[53] LARGER L, BAYLÓN-FUENTES A, MARTINENGHI R, et al. High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification[J]. Physical Review X, 2017, 7: 011015.
[54] XU Xingyuan, TAN Mengxi, CORCORAN B, et al. 11 TOPS photonic convolutional accelerator for optical neural networks[J]. Nature, 2021, 589(7840): 44-51.

作者简介

梁兴东　男，1973年生，吉林桦甸人。中国科学院空天信息创新研究院，微波成像技术国家级重点实验室副主任，研究员，博士生导师。主要研究方向为新概念新体制雷达通信一体化系统、高分辨率合成孔径雷达系统、干涉合成孔径雷达系统、成像处理及应用、实时信号处理。
E-mail：xdliang@mail.ie.ac.cn

李焱磊　男，1983年生，河北定兴人。中国科学院空天信息创新研究院，微波成像技术国家级重点实验室，研究员，硕士生导师。主要研究方向为新体制雷达信号处理、一体化信号波形设计、可重构异构处理架构、穿墙感知雷达技术。
E-mail：lily002954@aircas.ac.cn

刘云龙　男，1988年生，河北武强人。中国科学院空天信息创新研究院，微波成像技术国家级重点实验室，助理研究员。主要研究方向为机载SAR精细化定标处理、实时成像处理、一体化数字处理系统。
E-mail：liuyl003299@aircas.ac.cn

郭宇豪　男，1994年生，河北崇礼人。中国科学院空天信息创新研究院，微波成像技术国家级重点实验室，助理研究员。主要研究方向为新概念新体制雷达通信一体化系统、高分辨率合成孔径雷达系统。
E-mail：guoyh@aircas.ac.cn

解玉凤　女，1980年生，山东潍坊人。复旦大学微电子学院副教授，博士生导师。主要研究方向为新型存储器电路、存内计算电路设计。
E-mail：xieyf@fudan.edu.cn

徐兴元　男，1992年生，内蒙古人。北京邮电大学电子工程学院教授，博士、硕士生导师。主要研究方向为智能光计算、克尔光频梳、硅基光子学。
E-mail：xingyuanxu@bupt.edu.cn

刘　柳　女，1996年生，河北石家庄人。中国科学院大学电子电气与通信工程学院，中国科学院空天信息创新研究院，博士研究生，主要研究方向为一体化信号设计。
E-mail：liuliu18@mails.ucas.ac.cn

刘文成　男，1997年生，陕西西安人。中国科学院大学电子电气与通信工程学院，中国科学院空天信息创新研究院，博士研究生。主要研究方向为地面运动目标检测、运动目标成像以及FPGA算法开发。
E-mail：liuwencheng19@mail.ucas.edu.cn