

DOI:10.19651/j.cnki.emt.2211998

基于分块转置的二维快速傅里叶变换的 FPGA 设计<sup>\*</sup>姚佳辰<sup>1</sup> 马睿<sup>2</sup> 杨晓华<sup>1</sup> 黄艳艳<sup>1</sup> 耿乐<sup>2</sup>

(1.南通大学理学院 南通 226019; 2.北京大学长三角光电科学研究院 南通 226000)

**摘要:** 离散二维快速傅里叶变换被广泛应用于数字图像处理,对工程领域具有重要意义。通常 2D-FFT 使用行列分解计算,即先沿着行计算一维快速傅里叶变换,再沿列计算。由于现场可编程门阵列的数据传输带宽以及相关存储硬件的物理结构特性的限制,该方案不满足高分辨图像实时处理的需求。采用行 FFT-转置-行 FFT 的方案,虽减少计算过程中直接内存访问控制器的等待时间且能提高 2D-FFT 的计算效率,但目前矩阵转置实现有很大的局限性。传统的设计使用加载和存储指令来完成矩阵的换位。提出一种基于快速分块转置的 2D-FFT 方案,通过搭建转置模块与四路并行 1D-FFT 模块,充分利用 FPGA 片上资源以降低延时。实验基于 Xilinx Kintex UltraScale FPGA,在相同的时钟频率以及并行条件下,对比不同的 2D-FFT 计算方案。在实验误差范围内,本文提出的解决方案使计算效率提升约 15 倍。

**关键词:** 离散二维傅里叶变换;可编程阵列逻辑;矩阵转置;并行计算

**中图分类号:** TN911.73 **文献标识码:** A **国家标准学科分类代码:** 510.4030

FPGA-based design of two-dimensional fast Fourier transformation  
via block transposingYao Jiachen<sup>1</sup> Ma Rui<sup>2</sup> Yang Xiaohua<sup>1</sup> Huang Yanyan<sup>1</sup> Geng Le<sup>2</sup>

(1. College of Science, Nantong University, Nantong 226019, China;

2. Yangtze Delta Institute of Optoelectronics, Peking University, Nantong 226000, China)

**Abstract:** Two-dimensional discrete fast Fourier transformation is widely used in digital image processing, which is of great significance in engineering field. Usually, 2D-FFT is computed using column decomposition, that is, a row-wise 1D-FFT followed by another column-wise one. Due to the limitation of data transmission bandwidth of field programmable gate array and the physical structure characteristics of related storage hardware, this method cannot meet the requirement of real-time processing of high-resolution images. The scheme of row FFT-transposed-row FFT can reduce the waiting time of direct memory access controller in the computation process and improve the computational efficiency of 2D-FFT, but the existing implementation of matrix transposition has significant limitations. Traditional design uses load and store instructions to complete the transposition of a matrix. This paper proposes a 2D-FFT scheme based on fast block transposition. By building a transposition module and a four-way parallel 1D-FFT module, the FPGA on-chip resources are fully utilized, thus the delay is reduced. The experiment is based on Xilinx Kintex UltraScale FPGA, and under the same clock frequency and parallel conditions, different 2D-FFT calculation schemes are compared. Within the experimental error range, the solution proposed in this paper improves the computational efficiency by about 15 times.

**Keywords:** two-dimensional discrete Fourier transform; field programmable gate array; transposed matrix; parallel computing

## 0 引言

离散傅里叶变换(discrete Fourier transform, DFT)被

广泛运用于科学计算中,快速离散傅里叶变换(fast Fourier transform, FFT)是 DFT 的快速算法。其中,一维离散傅里叶变换(one-dimensional fast Fourier transform, 1D-DFT)

收稿日期:2022-11-08

<sup>\*</sup> 基金项目:国家自然科学基金(12004199)、南通基础科学研究计划项目(JC2021014)资助

常用于信号处理(digital signal processing, DSP)领域<sup>[1]</sup>,如多普勒雷达测速系统<sup>[2]</sup>,二维离散傅里叶变换(two-dimensional discrete Fourier transform, 2D-DFT)用于各种需要频域分析的成像应用,如光学显微<sup>[3]</sup>、核磁共振(magnetic resonance imaging, MRI)<sup>[4]</sup>、合成孔径雷达(synthetic aperture radar, SAR)处理<sup>[5]</sup>。随着 sCMOS 相机成像工艺的提升,图像的像素也随之增加。在数字全息成像与医学成像中,常见的图像已经达到  $1\ 024\ \text{pixels} \times 1\ 024\ \text{pixels}$ 。因此需要新的快速算法和架构来支持高分辨率图像的 2D-DFT。光学衍射层析成像<sup>[6]</sup>(optical diffraction tomography, ODT)解算过程中,需要对大量高分辨率的图像计算 2D-DFT。为加速工程化中的图像处理速度,提高 2D-FFT 的运算速度尤为关键。

目前计算图像 2D-FFT 方式可以划分为软件方案和硬件方案。其中软件方案主要是通过图形处理器(graphics processing unit, GPU)对高分辨率图像数据进行 2D-FFT 处理<sup>[7]</sup>,硬件解决方案包括使用专用 FFT 处理芯片<sup>[8]</sup>和现场可编程门阵列(field programmable gate array, FPGA)<sup>[9]</sup>。GPU 是通过数千个小内核实现加速处理的软件处理器,在进行内存访问过程中需要占用大量时钟周期,导致数据时序等待<sup>[10]</sup>。专用 FFT 处理芯片,首先,其制造成本相当高,不利降本增利。其次,一旦芯片被制造,功能和性能就无法再改变,增加了后续的研发成本。FPGA 的出现使 2D-FFT 的实现有了新的替代方案。FPGA 通过并发和流水两种技术实现更大的并行度减少时序等待,还具有可定制性与可重构性,减少开发成本,更适合工程化嵌入式应用。目前 FPGA 广泛应用于各种嵌入式信号处<sup>[11]</sup>和图像处理系统如:智能相机、医疗图像重建<sup>[12]</sup>,其中 FFT 是关键组成部分。

基于既往研究结论,如 Allaf 等<sup>[13]</sup>提出了一种针对低成本 FPGA 技术的  $256 \times 256$  小规模数据量 Radix-4 的 2D FFT 系统的开发方案。利用 DSP 算法和行列技术在 FPGA 上具有并行处理、结构灵活、集成度高、速度快等优点。该体系结构使用可重构计算来验证两种正交方法,以权衡硬件成本和性能。Chen 等<sup>[14]</sup>提出 3D 存储器集成 FPGA 上的 2D-FFT 系统架构,并以动态数据布局来优化 3D 架构来实现 2D-FFT,以吞吐量和延迟为指标评估其性能。所提出的架构通过最大化平衡 FPGA 上的外部存储器和 FFT 内核之间的带宽来实现高吞吐量。通过采用片上排列网络实现动态数据布局,极大地降低了行激活开销引起的延迟,从而显著提高了性能。除此之外, Li 等<sup>[15]</sup>提出了一种设计根据 MRI 数据流的特点进行了定制的二维 FFT 算法的 FPGA 处理器,实现了对低分辨率的多层图像数据集进行 2D-FFT 处理的加速。其方法是针于多层图像切片处理,不具有普适性且纯 Verilog 硬件逻辑开发耗时、开发难度大,不适合工程化。

本文中提出了一种基于快速分块转置 2D-FFT 的 FPGA 架构。解决了现存 FPGA 模块化设计中搭建完模型后,通过中央控制器(Microblaze)控制大数据量进行 2D-FFT 运算时耗的问题。本文进行 3 组实验,对实验数据进行对比分析后发现,基于快速分块转置 2D-FFT 的 FPGA 架构能够优化耗时问题,加速高分辨率图像的 2D-FFT。

## 1 基本原理

### 1.1 离散傅里叶变换的定义

设  $N$  个点的离散时间序列为,其中的,其中 1D-DFT 的定义为:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot E_N^{nk}, k = 0, 1, \dots, N-1 \quad (1)$$

其中,  $E_N = e^{-j2\pi/N}$ 。  $N$  点 DFT 的计算复杂度为  $O(N^2)$ ,  $N$  点 FFT 的计算复杂度为  $O(N \log_2 N)$ 。

设  $N \times N$  点的二维离散时间序列为  $x(i_1, i_2)$ , 其中  $i_1, i_2 = 0, 1, \dots, N-1$ 。其 2D-DFT 的定义为:

$$X(k_1, k_2) = \sum_{i_1=0}^{N-1} \sum_{i_2=0}^{N-1} x(i_1, i_2) \cdot E_N^{k_1 i_1 + k_2 i_2} \quad (2)$$

$k_1, k_2 = 0, 1, \dots, N-1$ 。

根据式(1)、(2), 2D-DFT 可以分解为行列计算, 可视为多次 1D-DFT, 如式(3)、(4):

$$Y(k_1, i_2) = \sum_{i_1=0}^{N-1} x(i_1, i_2) \cdot E_N^{k_1 i_1} \quad (3)$$

其中,  $k_1 = 0, 1, \dots, N-1$ 。

$$X(k_1, k_2) = \sum_{i_2=0}^{N-1} Y(k_1, i_2) \cdot E_N^{k_2 i_2} \quad (4)$$

其中,  $k_2 = 0, 1, \dots, N-1$ 。

通过行列分解, 2D-DFT 可以通过先执行  $N$  行 1D-DFT, 然后再执行  $N$  列 1D-DFT 来计算。如果使用 2D-FFT 实现 2D-DFT, 则复杂度为  $O(N^2 \log_2 N)$ 。

### 1.2 离散傅里叶变换的矩阵乘法形式

长度为  $M$  的 1D-DFT 可以通过变换和置换进行分解和计算。将 1D-DFT 用向量的形式表示如下:

$$[\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2 \cdots \mathbf{X}_{M-1}]^T = \mathbf{F}_M \cdot [\mathbf{x}_0 \mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_{M-1}]^T \quad (5)$$

其中,  $\mathbf{F}_M$  是旋转因子矩阵。

由上述表达式可以看出 1D-DFT 本质就是稀疏矩阵  $\mathbf{F}_M$  与一维数据矩阵的乘积。 $\mathbf{F}_M$  定义如下:

$$\mathbf{F}_M = \mathbf{P}_{M,p} (\mathbf{I}_p \otimes \mathbf{F}_n) \mathbf{D}_M (\mathbf{F}_p \otimes \mathbf{I}_n) \quad (6)$$

$$\mathbf{D}_M(j, j) = \mathbf{W}_M^{(j \bmod n) \cdot \lfloor j/n \rfloor} j \in [0, M-1] \quad (7)$$

$$\mathbf{A}_m \otimes \mathbf{B}_n = [\mathbf{a}_{k,l} \mathbf{B}_n]_{0 \leq k, l < m} \mathbf{A}_m = [\mathbf{a}_{k,l}]_{0 \leq k, l < m} \quad (8)$$

其中,  $\mathbf{N} = p \cdot m$  (其中  $p, m$  为整数)。 $\mathbf{I}_n$  是  $n \times n$  的单位矩阵,  $\mathbf{D}_M$  是旋转因子的对角矩阵,  $\otimes$  表示为张量积,  $\mathbf{P}_{N,p}$  表示步长为  $p$  的置换。

根据式(3)、(4), 2D-DFT 用矩阵描述如下:

$$\mathbf{Y} = \mathbf{F}_M^T \cdot \mathbf{X} \cdot \mathbf{F}_N = \mathbf{F}_M \cdot \mathbf{X} \cdot \mathbf{F}_N \quad (9)$$

其中, 输入  $\mathbf{X}$  和输出  $\mathbf{Y}$  的大小为  $M \times N$ ,  $\mathbf{F}_M$  与  $\mathbf{F}_N$  旋

转因子矩阵是对角矩阵。

使用矩阵转置分解：

$$Y = F_M \cdot X \cdot F_N = (X^T \cdot F_M)^T \cdot F_N = Z^T \cdot F_N \quad (10)$$

$$Z = X^T \cdot F_M \quad (11)$$

首先对输入的大小为  $M \times N$  的矩阵  $X$ ，如式(11)做原矩阵每一行的行 FFT，结果为  $Z$ ，再对  $Z$  矩阵做转置，对转置后的结果再做每行 FFT，得出结果  $Y$  如式(10)所示。即把行列分解的 2D-FFT 通过转置变化变为两次行 FFT，实现 2D-FFT。因此，基于以上分析，基于行列分解的 2D-FFT 可以进一步简化为行 FFT-转置-行 FFT 的过程。

## 2 系统设计

在本节中，首先论述在 2.1 节中的 FPGA 框架不适用于大数据量的 2D-FFT 处理。故分别在 2.2、2.3 节中搭建基于行列分解的 2D-FFT 框架、基于外部原地转置的 2D-FFT 框架，由于大数据量存储于外部存储器中存在访问数据耗时间问题，在 2.4 节提出了基于快速分块转置的 2D-FFT 架构，并对框架中的数据传输流程以及模块进行描述。

### 2.1 基于 FPGA 的小数据量 2D-FFT 框架

在 FPGA 中部署行列分解的 2D-FFT，输入的数据存储在 FPGA 的片外同步随机存取器 (static random-access memory, SRAM) 中，然后依次对二维数据的行与列执行 1D-FFT。图 1 给出基于 SRAM 的外部存储器的基本架构。

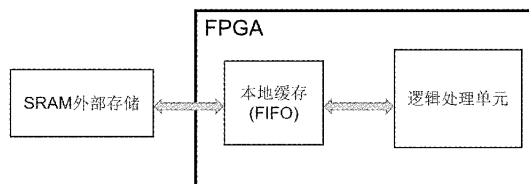


图 1 基于同步随机存取存储器的外部存储器的基本架构

输入数据中固定数量的行与列被加载到 FPGA 本地缓存中依次进行行列 1D-FFT，如图 2 所示，图像大小为  $N \times N$ ， $P$  为并行比例系数，当  $P$  取不同的值，影响 1D-FFT 行列并行的个数。执行多次后，直至遍历所有行列数据。由于 FPGA 访问 SRAM 中存储数据的行列时间相同，因此这种实现方式对于较小的图像数据有优异的表现。然而，面对大量的数据，SRAM 解决方案由于其存储空间和高成本以及占用空间面积大而变得不切实际。

同步动态随机存取内存 (synchronous dynamic random-access memory, SDRAM) 外部存储器是大数据存储的一个方案，由于其高吞吐量和高容量，SDRAM 成为大多数 FPGA 板上的标准组件。

SDRAM 由于其物理结构特性。其具有非统一的访问时间，访问同一行中的连续元素具有低延迟特性，而访问不同行中的数据，即访问列数据具有高延迟特性<sup>[16]</sup>。因此，如果使用行列分解算法来实现 2D-FFT，执行行方向 1D-

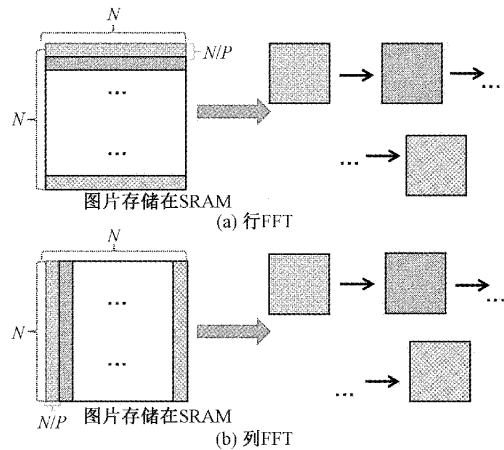


图 2 行列一维快速傅里叶变换

FFT 比列方向 1D-FFT 快得多，因为 SDRAM 内存物理结构有利于沿行进行数据访问。因此，对 SDRAM 中列数据访问是 2D-FFT 实现的主要瓶颈。尽管 DDR4 SDRAM 可以实现外部存储器的读取和写入具有更高数据速率，但是现有的存储器接口设计未能利用 SDRAM 设备的带宽用于列 FFT。即便使用自定义内存接口设计，对于高分辨率图像的传统行列分解 2D-FFT 实现也会导致极高的延时。

### 2.2 基于行列分解的 2D-FFT 框架

如图 3 所示的框架是基于式(3)、(4)。在 2.1 节中已说明 SRAM 在大数据量应用场景下不适用，所以将多帧图像数据存储在外部的 DDR4 SDRAM 中。首先以图 3 步骤①中将 SDRAM 的图像数据按行读取，在中央控制单元的调度下行数据通过高速吞吐量外设存储接口传入 FPGA 中，行数据以流传输的形式快速写入 Xilinx 提供的 1D-FFT IP 核，1D-FFT IP 核的参数按照实际输入图片的每行像素点来确定。完成一帧图像数据的 1D-FFT 后，数据返回 SDRAM 的源地址。然后以步骤②通过中央处理器的调度将 SDRAM 中数据以列读取方式重复上述行 FFT 环节，直至完成一帧图像的 2D-FFT。

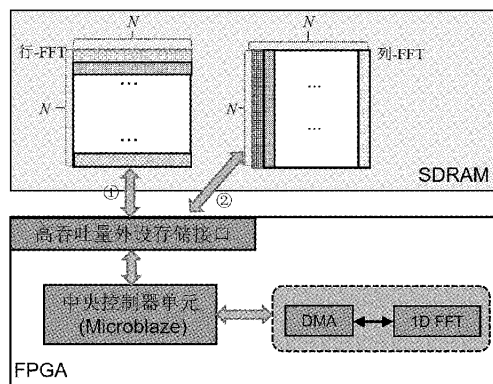


图 3 基于行列分解的 2D-FFT 框架

### 2.3 基于外部原地转置的 2D-FFT 框架

由于直接内存访问器 (direct memory access, DMA) 访

问连续地址数据的效率更高,在行方向 FFT 和列方向 FFT 之间插入转置步骤,可以原来的列方向 FFT 转变为行方向 FFT。最常见的转置方式是直接原地按元素转置。如图 4 所示的框架是基于式(10)、(11),与 2.2 节中行列分解的 2D-FFT 框架相比,图 3、4 中的①都是将 SDRAM 中数据按行送入 FPGA 中进行行 1D-FFT。图 4 中的步骤②表示直接将数据在 SDRAM 中进行原地转置操作,再将转置完的数据如步骤③再次按行传入 FPGA(此处步骤③与步骤①一致)。

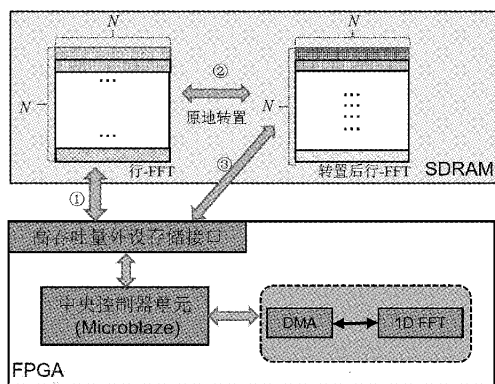


图 4 基于原地转置的 2D-FFT 框架

#### 2.4 基于快速分块转置的 2D-FFT 架构

在外部 SDRAM 进行原地转置没有从本质上解决问题。为了充分利用 FPGA 的电路特性,可以基于电路的形式实现转置。然而,由于片上资源(BlockRam、LUTs、FF)有限,对于  $1024 \text{ pixels} \times 1024 \text{ pixels}$  的图像没有足够缓存空间进行转置,需要将图像数据传输到 FPGA 片上,利用片上逻辑以及存储空间进行大矩阵的分块转置。通过带有快速转置的 2D-FFT 算法可以有效解决二维大数据量行列 FFT 算法数据访问延时问题,并且可以有效的映射到基于 FPGA 的框架中。

用于实现基于快速分块转置 2D-FFT 的 FPGA 架构如图 5 所示。由多个模块构成,这些模块可以大致分为特定模块和基础模块组件。例如数据处理器模块是专门为 2D-FFT 应用而设计的,故为特定模块组件。图中高吞吐量内存接口单元、中央处理器单元、主机通信单元和串口调试单元为基础模块组件,为特定模块组件提供高速控制。

##### 1) RTL 级转置模块设计

FPGA 模块化设计思想的基础是 AXI 总线,转置模块是以 AXI 总线中的 AXI4-Stream 接口协议设计的 RTL 级模块。AXI4-Stream 接口协议与 AXI4、AXI4-Lite 接口协议相比没有地址项,允许无限制的数据突发传输规模,减少延迟,实现高速流式处理数据。图 6 所示为 AXI4-Stream 的时序图,其中 S 表示从端口,M 表示主端口。

基于 AXI4-Stream 协议设计的转置模块具有主从端口。当转置模块的从端口接收到上层模块主端口发出的 S\_TVALID 信号,此时转置模块进入准备收数状态。S\_

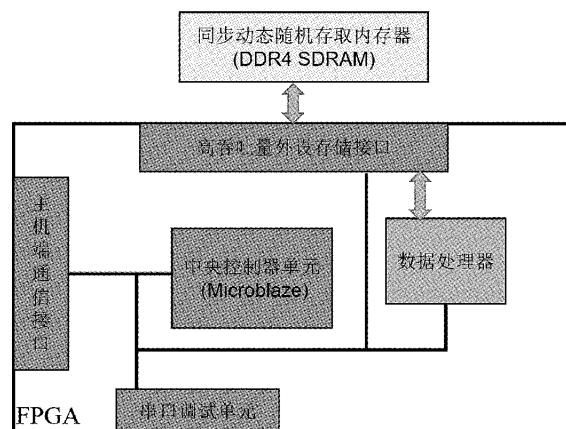


图 5 基于快速二维转置 FFT 的 FPGA 架构

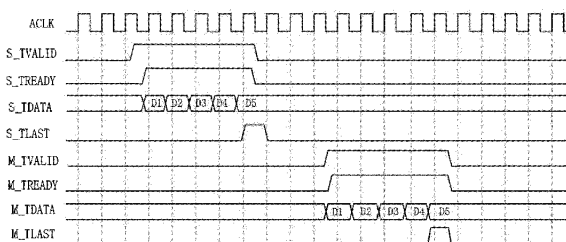


图 6 AXI4-Stream 时序图

TREADY 信号由 S\_TVALID 信号以及预设写入数据个数的阈值决定。只有在时钟上升沿到来且 S\_TREADY 与 S\_TVALID 同时为高电平时,模块进入接受数据状态。当写入的数据达到设定阈值时,S\_TVALID 与 S\_TREADY 信号变为低电平,S\_LAST 信号由上层模块发出保持一个时钟周期的高电平,表示转置模块成功接收完上层模块所需传输的数据量。等到下一个时钟上升沿,模块进入转置状态,数据被送入数据选择器中进行转置,无论多大数据的转置仅需一个时钟周期。转置完成后进入数据发送状态,当 M\_TVALID 与 M\_TREADY 同时为高时,数据开始按时钟依次发出,当数据发送完毕,M\_TLAST 被拉高一个时钟周期,表示输出端口完成转置数据传输。

如图 7 所示的分块转置操作:

- (1) 将存储在 SDRAM 中的  $N \times N$  大小的二维图像数据分割成  $M \times M$  个数据块,每块大小为  $N_1 \times N_1$ 。
- (2) 通过中央直接内存访问器(central direct memory access,CDMA)取  $N_1 \times N_1$  个数据。
- (3) 将数据通过直接内存访问器送入转置模块进行转置。
- (4) 将转置后的数据通过直接内存访问器送回源地址覆盖原数据。
- (5) 重复上述操作直至整张图像转置完成。

##### 2) 并行转置 FFT 模块

如图 8 所示,并行 FFT 处理流程框图中的虚线部分是 1D-FFT 并行模块,并行模块由直接存储器访问器与一维

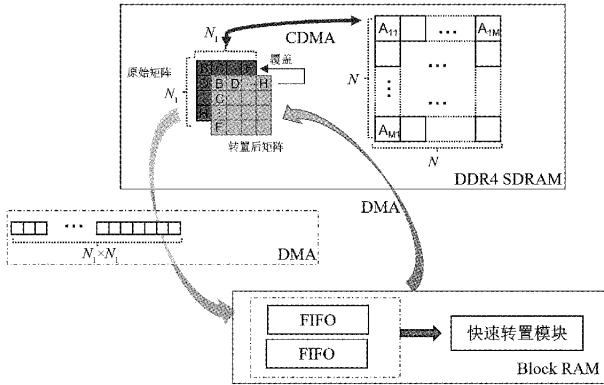


图 7 分块转置操作

FFT IP 核构成,实现 DDR4 SDRAM 存储的原始数据进入 FPGA 进行并行 FFT 处理。在如图 5 的设计中,采用 Xilinx 提供的基于 AXI-Stream 协议的 FFT IP 核。利用 FPGA 可并行的特点并结合实际使用 FPGA 开发板提供的可编程资源以及实际执行效率,最终,对一维 FFT 执行四路并行处理。

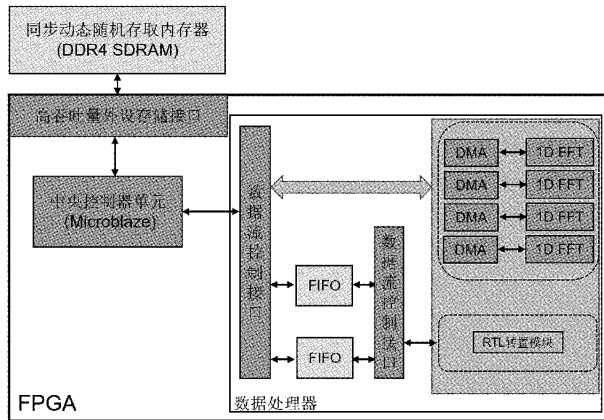


图 8 并行 FFT 处理流程框图

输入数据存储在外部存储器(DDR4 SDRAM)中,通过 FPGA 端高吞吐量外设存储接口接受传递的数据,将数据通过按行快速访问的方式并行的将 4 行数据送入 1D-FFT 并行模块,再将做完 FFT 的图像数据传回外部存储器,此时二维图像的 1D-FFT 完成。通过第 2 节的式(10)、(11)可知,此时需要对已经做完 1D-FFT 的图像数据做转置操作。由于 FPGA 片上资源的限制,对图像数据采取分块转置的方式,部分数据被加载到 FPGA 本地存储器中进行处理。用两个相同的本地存储器作为乒乓处理操作的缓冲器。这些本地存储器在 FPGA 上用 FIFO 实现。此处乒乓处理操作可以有效的减少数据等待时间以及实现数据的无缝缓冲和处理。

将分块后的数据等量的加载到 FPGA 片上缓冲器中,经过乒乓处理操作,将数据送入转置模块。这种模式对于访问同步动态随机存储器(如 DDR3、DDR4 SDRAM)有利,因为动态存储器只支持行快速突发访问。重复这些操作直到遍历整个数据。

### 3) 中央控制器单元

Xilinx 公司的 FPGA 开发板对中央控制器提供两种方案,软核处理器(system on programmable chip, SOPC)与硬核处理器(system on chip, SOC)。中央控制器用于将输入数据加载到外部存储器,以及控制整个系统运行。

SOPC 技术起源于 Xilinx 的竞品公司 Altera 公司是基于 FPGA 片上系统级设计方式。此技术是通过 FPGA 的通用逻辑搭建的一个软核 CPU 系统。Xilinx 公司的 Microblaze, Altera 公司的 Nios-II 皆是软核处理器。SOC 技术在设计时,在 FPGA 板上额外增加了处理器。硬核技术不会消耗 FPGA 本身的资源且彼此在逻辑上是相互独立。

本文使用 Xilinx 公司的 RISC 处理器软核 Microblaze。被 Xilinx 优化后的 Microblaze 具有灵活的可配置性,与 FPGA 硬件逻辑有良好的交互性,能够满足本文的实现,相较于硬核处理器软核软核处理器具有低成本性、外设可扩展性,有利于工程化降本增利。此外,串口模块用于提供一个基本终端,以显示状态和调试。主机端通信接口,目前使用 PCIE3.0 与主机进行数据高速传输。高吞吐量外设接口使用 DDR4 MIG IP 核,加速工程化设计。

### 3 实验结果及分析

实验使用的集成设计环境是 Vivado2019.1,实验使用的是基于 Xilinx 的 Kintex-UltraScale 架构的 AXKU041 开发板,具有丰富的外设资源如:USB-UART 接口、DDR4、QSPI-FLASH、JTAG 调试口、PCIE3.0 × 8,其中 DDR4 SDRAM 的型号为 MT40A512M16LY。

RTL 级转置模块的时序仿真图如图 9 所示,仿真仅为验证基于 AXI Stream 协议的矩阵转置模块的正确性,故仿真输入的时钟频率为 100 MHz、转置矩阵大小为 4 × 4,多维矩阵逻辑不变。

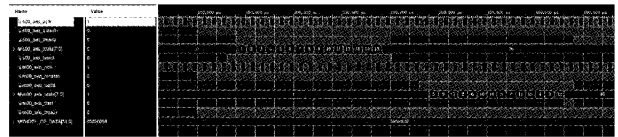


图 9 转置模块仿真波形图

在 BlockDesign 中搭建基于 Microblaze 软核处理器的快速分块转置的 2D-FFT。外部存储器输入一帧相机采集到的 ODT 图像数据,数据大小为 1 024 pixels × 1 024 pixels 的通道 64 bit,在 FPGA 中设计的 RTL 级转置模块会占用 FPGA 大量的逻辑资源以及片上存储资源,在考虑资源下,实验将分块转置矩阵的大小分别设为 64 × 64、128 × 128、256 × 256 来对比图像进行 2D-FFT 中分块矩阵大小对 2D-FFT 速率的影响,3 次实验均使用四路并行 FFT,通过时间函数打印时间,结果表明分块矩阵设置为 256 × 256 时 2D-FFT 耗时仅为 22.32 ms 时间如表 1 所示。

表 1 分块转置 2D FFT 时间

芯片型号	分块矩阵大小	位宽/bit	时钟/MHz	像素大小/pixels	时间/ms
Kintex-UltraScale	64×64				65.79
	128×128	64	300	1 024×1 024	36.82
	256×256				22.32

表 2 给出无并行行列分解 2D-FFT、原地转置 2D-FFT 和四路并行行列分解 2D-FFT、原地转置 2D-FFT 的时间以及本文方法在分块矩阵大小为 256×256 下四路并行快速分块转置 2D-FFT 的运算时间。前 4 组实验两两对比说

明加入并行 FFT 对执行时间影响不大,表明 1D-FFT 本身耗时极小,耗时主要源于外部存储器数据访问与传输。运用本文快速分块转置方法与表 2 中最少耗时 352.78 ms 相比,使得运算速度提升了 15.8 倍,效率提高了 93.68%。

表 2 运算时间

芯片型号	实验	位宽/bit	时钟/MHz	图像大小/pixels	时间/ms
Kintex-UltraScale	行列分解 2D-FFT				484.23
	四路并行行列分解 2D-FFT				362.44
	原地转置 2D-FFT	64	300	1 024×1 024	471.31
	四路并行原地转置 2D-FFT				352.78
	本文方法				22.32

为进一步验证 FPGA 输出结果与 Matlab 中 fft2 函数结果的一致性,本文将频谱转化为幅度谱用均方误差 (mean square error, MSE) 评估两种结果的差异。此外,ODT 解算过程中,需要提取相应 K 值(二维幅度谱中最大值坐标)用于后续解算,准确的 K 值对高质量成像至关重要。因此,在评估本文算法时,关注了幅度谱中最大值所在位置与标准结果之间的差异。

MSE 如式(12)所示,其中  $y_i$  为标准值,  $\hat{y}_i$  表示 FPGA 测量值,  $n$  为图像的像素个数。

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (12)$$

欧氏距离如式(13)所示。

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (13)$$

具体而言,实验中将多幅大小为 1 024 pixels×1 024 pixels 的单通道 64 bit 图像样本分别输入到 MATLAB 和 FPGA。表 3 分别给出各个图像样本的由 FPGA 输出的幅度谱和幅度谱峰值坐标与 MATLAB 中结果的差异。

由表 3 结果看出,幅度谱的 MSE 误差仅为 0.094 1。

表 3 误差分析表

图像样本	频谱峰值坐标(Matlab)	频谱峰值坐标(FPGA)	坐标间欧氏距离	幅度谱均方误差
1	(326,1 001)	(326,1 001)	0	0.095 1
2	(256,660)	(256,660)	0	0.094 2
3	(283,705)	(283,705)	0	0.094 5
4	(223,870)	(223,870)	0	0.094 8
5	(204,685)	(204,685)	0	0.092 9
6	(288,674)	(288,674)	0	0.092 8
7	(65,657)	(64,657)	1.00	0.093 4
8	(473,1 003)	(473,1 003)	0	0.093 8
9	(301,908)	(301,908)	0	0.094 7
10	(238,741)	(238,741)	0	0.094 5
平均值	—	—	0.10	0.094 1

而且在 1 024 pixels×1 024 pixels 的图像中,幅度谱峰值位置平均误差约为 0.10 pixels,不影响后续图像解算,以上误差在实际工程的允许范围内。

## 4 结 论

本文针对高分辨率 ODT 图像在 FPGA 中的行列分解

2D-FFT 框架以及外部原地转置的 2D-FFT 框架运算耗时问题,提出了一种基于快速分块转置 2D-FFT 的 FPGA 架构,充分利用 FPGA 并行特点,通过乒乓操作在 Xilinx Kintex-UltraScale 系列开发板上实现了对高分辨率 ODT 图像进行 2D-FFT 的加速,并且将基于快速分块转置 2D-FFT 的 FPGA 架构运算的结果与 Matlab 进行对比,在误差允许范围内,验证了结果的正确性,运算速度提升了 15.8 倍,效率提高了 93.68%。有利于图像处理的工程化开发,具有良好的应用前景。

### 参考文献

- [1] 宋奋韬,王梦莹,付志远. FPGA 发展概论[J]. 科技信息, 2012(23): 145-145.
- [2] 张建宏,武锦辉,刘吉,等. 基于 FPGA 的多普勒雷达测速系统[J]. 国外电子测量技术, 2019, 38(12): 72-75.
- [3] TIAN C, LI H, GAO X. Photo-realistic 2D expression transfer based on FFT and modified poisson image editing[J]. Neurocomputing, 2018, 309(2): 1-10.
- [4] WEST B L, FESSLER J A, WENISCH T F. Jigsaw: A slice-and-dice approach to non-uniform FFT acceleration for MRI image reconstruction[C]. IEEE, 2021, DOI:10.1109/IPDPS49936.2021.00081.
- [5] GUO Y, SUO Z, JIANG P, et al. A fast back-projection SAR imaging algorithm based on wavenumber spectrum fusion for high maneuvering platforms [J]. Remote Sensing, 2021, 13(9): 1649-1667.
- [6] 穆书奇,董大山,施可彬. 无标记光学成像技术[J]. 激光与光电子学进展, 2022, 59(12): 11-31.
- [7] PRATX G, LEI X. GPU computing in medical physics: A review[J]. Medical Physics, 2011, 38(5): 2685-2697.
- [8] ELANGO, K, MUNIANDI K. Hardware implementation of FFT/IFFT algorithms incorporating efficient computational elements [J]. Journal of Electrical Engineering and Technology, 2019, 14(4): 1717-1721.
- [9] NGUYEN N H, KHAN S A, KIM C H, et al. A high-performance, resource-efficient, reconfigurable parallel-pipelined FFT processor for FPGA platforms [J]. Microprocessors and Microsystems, 2018, 60: 96-106.
- [10] ARONOV A, KAZAKEVICH L, MACK J, et al. 5G NR LDPC decoding performance comparison between GPU & FPGA platforms [C]. IEEE Long Island Systems, IEEE, 2019, DOI:10.1109/LISAT.2019.8816821.
- [11] 赵精博,张军,史晓锋. 基于 FPGA 的飞行校验多模信号采集与处理设计[J]. 电子测量技术, 2014, 37(3): 102-106.
- [12] RYMARCZYK T, KOSIOR A, P TCHÓRZEWSKI, et al. Image reconstruction in electrical impedance tomography using a reconfigurable FPGA system[J]. Journal of Physics: Conference Series, 2021, 1782(1): 12033-12040.
- [13] ALLAF A F, DAWWD S A. FPGA based reconfigurable 2D FFT system [J]. Al-Rafadain Engineering Journal, 2011, 19(3): 76-88.
- [14] CHEN R, SINGAPURA S G, PRASANNA V K. Optimal dynamic data layouts for 2D FFT on 3D memory integrated FPGA [J]. The Journal of Supercomputing, 2017, 73(2): 652-663.
- [15] LI L, WYRWICZ A M. PARALLEL 2D FFT implementation on FPGA suitable for real-time MR image processing [J]. Review of Scientific Instruments, 2018, 89(9): 93706-93709.
- [16] MCKEE S A, KLENKE R H, WRIGHT K L, et al. Smarter Memory: Improving Bandwidth for Streamed References[J]. Computer, 2012, 31(7): 54-63.

### 作者简介

姚佳辰, 硕士, 主要从事 FPGA 系统开发、光学工程方面的研究。

E-mail: JCYao@stmail.ntu.edu.cn

马睿, 副研究员, 主要从事全内反射显微镜, 超分辨单分子定位显微镜研究。

E-mail: rma@ydioe.pku.edu.cn

杨晓华, 博士研究生, 教授, 博士生导师, 主要从事原子分子物理、光学、光电子学方面的研究。

E-mail: xhyang@ntu.edu.cn

黄艳艳, 博士, 副教授, 主要从事手征特异材料的电磁传输特性及光调制效应的研究。

E-mail: yyhuang@ntu.edu.cn

耿乐, 助理研究员, 主要从事 FPGA 开发、基于深度学习的图像、信号处理算法研究。

E-mail: gengle@ydioe.pku.edu.cn