

OSIC检测中高效排序QR分解FPGA实现

王海麟¹, 冯献礼¹, 辜方林¹, 高明柯², 赵海涛¹

(1. 国防科技大学电子科学学院, 长沙 410073; 2. 中国电子科技集团公司第32研究所, 上海 201808)

摘要: 排序连续干扰消除 (Ordered successive interference cancellation, OSIC) 是多输入多输出 (Multiple input multiple output, MIMO) 系统中一种常用的信号检测算法, 但该算法的吞吐量、时延等指标受制于信道矩阵逆运算。因此, 计算复杂度低且能高速实现矩阵求逆分解预处理是算法硬件实现的关键。本文采用对信道矩阵进行排序正交三角 (Orthogonal triangle, QR) 分解的矩阵预处理硬件加速方案, 其中排序过程引入对复值 1 范数的快速估计方法消除复数模计算, QR 分解过程利用深度流水线化坐标旋转数字计算 (Coordinate rotation digital computer, CORDIC) 迭代方法消除 Givens 旋转过程中的元素矢量化, 计算置零旋转角度, 实现了面向 QR 分解的可复用 Givens 旋转结构的流水线电路结构设计, 使矩阵分解过程中无需乘法器。仿真结果表明, 本文所提 OSIC 改进算法误比特率性能与基于信噪比的 OSIC 检测算法性能基本一致, 所提的基于 CORDIC 迭代的 Givens 旋转结构能够高度分时复用, 显著提升系统并行度并极大减少资源占用, 系统设计时钟最高能达到 250 MHz, 矩阵分解吞吐量能达到 1.88 M Matrices/s, 能够满足 4 天线及以上 MIMO 系统接收端吞吐量和时延需求。

关键词: 多输入多输出信号检测; 排序连续干扰消除; 排序 QR 分解; Givens 旋转; 现场可编程门阵列
中图分类号: TN911.7 **文献标志码:** A

Efficient FPGA Implementation of Sorted QR Decomposition in OSIC Detection

WANG Hailin¹, FENG Xianli¹, GU Fanglin¹, GAO Mingke², ZHAO Haitao¹

(1. College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China; 2. The 32nd Research Institute, China Electronic Technology Group Corporation, Shanghai 201808, China)

Abstract: Ordered successive interference cancellation (OSIC) is a commonly utilized signal detection algorithm in multiple input multiple output (MIMO) systems. However, the algorithm's performance in terms of throughput and latency is constrained by the computational complexity of the channel matrix inverse operation. Therefore, matrix inverse decomposition pre-processing with low computational complexity and high speed is the key to hardware implementation of the algorithm. In this paper, we adopt a hardware-accelerated matrix pre-processing scheme for sorted orthogonal triangle (QR) decomposition of the channel matrix, in which the sorting process introducing a fast estimation method for complex-valued 1-norm to eliminate complex modulus computation. The QR decomposition process uses a deeply pipelined coordinate rotation digital computer (CORDIC) iterative method to eliminate the element vectorization and nulling rotation angle computation in the Givens rotation process, thus a pipeline circuit structure with a reusable Givens rotation structure for QR decomposition is designed, obviating the necessity for multipliers

in the matrix decomposition process. Simulation results demonstrate that the OSIC enhancement algorithm proposed achieves the bit error rate (BER) performance comparable to that of the signal-to-noise ratio-based OSIC detection algorithm. The CORDIC iterative Givens rotation structure proposed in this paper can achieve highly time-sharing multiplex. It significantly improves the system parallelism and reduces the resource consumption, and the system design clock attains up to 250 MHz, and the matrix decomposition throughput reaches 1.88 M Matrices/s, meeting the processing throughput and latency requirements of 4 or more antennas MIMO systems at the receiver.

Key words: multiple input multiple output(MIMO) signal detection; ordered successive interference cancellation(OSIC); sorted QR decomposition; Givens rotation; field programmable gate array (FPGA)

引言

随着无人机及无人机集群等新型平台的快速涌现,以及其应用的爆炸式增长,无线通信的带宽需求越来越大,通信信号处理技术不断推陈出新。多输入多输出(Multiple input multiple output, MIMO)技术是提升系统容量的一种有效手段,并取得了广泛应用^[1]。MIMO系统性能受多方面影响,其中信号检测技术的优劣至关重要。一方面,为了保证MIMO系统的通信性能,要求信号检测能够克服噪声、信道等因素影响,具有优异的检测性能;另一方面,实际工程应用中需要考虑计算复杂度和硬件资源消耗的影响,特别是随着基站天线和用户数量的大幅增加,系统接收端信号检测算法复杂度会急剧增大^[2]。因此,设计低复杂度信号检测算法具有重要的现实需求和应用前景。

对于MIMO信号的检测算法,按检测效果可以分为最优检测算法和次优检测算法。最大似然(Maximum likelihood, ML)检测是性能最优检测算法^[3],但该算法求解时需要遍历所有的信号空间,当天线数量增多时,算法的复杂度呈指数形式增长。次优检测算法可以分为线性检测算法和非线性检测算法。线性检测算法有迫零(Zero forcing, ZF)检测算法、最小均方误差(Minimum mean square error, MMSE)检测算法^[3],它们在利用信道加权矩阵抑制其他天线发射信号时会将噪声放大,影响检测性能。非线性检测算法往往能平衡计算复杂度和检测性能需求,已经成为目前研究的重点。排序连续干扰消除(Ordered successive interference cancellation, OSIC)算法是一种重要的非线性检测方法^[4],其基本思想是每次只检测一层信号并进行排序,从而逐层消除上层信号干扰,但其检测过程中存在复杂矩阵伪逆运算,不利于硬件实现,正交三角分解(Orthogonal triangle decomposition, QRD)算法^[5-6]为矩阵逆运算提供了一种低复杂度实现方式,在OSIC检测算法中得到了广泛应用。

Wubben等^[7]首先提出了基于Gram-Schmidt正交化的排序QR分解检测(Sorted QR decomposition, SQRD)算法,随后基于Householder变换和基于Givens旋转的排序正交三角(Orthogonal triangle, QR)分解检测算法也被提出^[8],这3种检测算法都不需要进行矩阵逆运算。尽管由于排序方式的问题,它们的检测顺序不一定最优,但它们均较大程度降低了复杂度^[9],且性能损失较小。图1比较了3种算法在复数

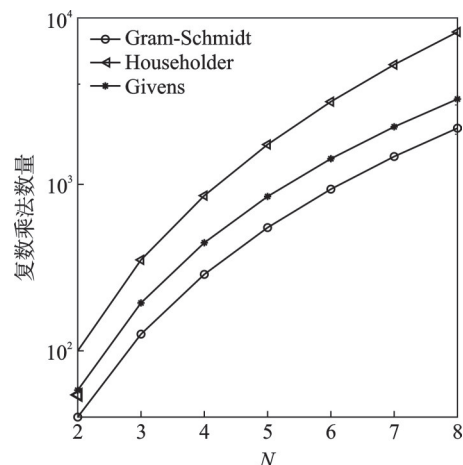


图1 分解算法复杂度比较

Fig.1 Comparison of complexity of decomposition algorithms

乘法方面随 $N \times N$ 矩阵维数的增长复杂度变化。可以看出 Householder 复杂度最高, Gram-Schmidt 变换最低。而考虑到 Givens 旋转数值计算结果稳定, 且计算流程可替代性强, 因此本文对 Givens 旋转实现排序 QR 分解进行研究。

排序 QR 分解的硬件实现也是一个极为重要的问题。基于 Givens 旋转的 QR 分解将原信道矩阵 H 乘上一系列的 Givens 旋转矩阵后将下三角元素都变为 0。现有研究大多采用三角收缩阵列作为硬件实现结构^[8-17]。由于旋转矩阵一般用三角函数(或用倒数、平方根)表示, 硬件实现不便, 因此优化方案研究较多。一般可分为两类: (1) 采用旋转数字计算方法 (Coordinate rotation digital computer, CORDIC)^[18], 该类方案主要考虑了 CORDIC 模块只进行加法和移位操作就能实现三角函数等计算, 复杂度较低, 可以实现较高的数据吞吐量。但是由于 CORDIC 使用多级迭代单元串联进行计算, 所以 QR 分解的处理时延受迭代次数的限制。相关文献进行了缩短处理时延的研究, 文献[9, 17]对常用三角收缩阵列结构进行了单元化计算, 基本计算单元仍然包括复数乘法, 而且数据经常需要缓存转发, 占用硬件资源较多。文献[19]设计了基于 CORDIC 的主从对结构, 提高了数据处理流水化程度, 并且将每次 Givens 旋转变得相互独立, 提高了矩阵分解的可扩展性。例如文献[20]提出基于锦标赛的复 Givens 旋转方法, 同时进行两个以上元素的处理, 在较大尺寸的矩阵中进一步减小了时延, 吞吐量更高。(2) 不采用 CORDIC 方案的研究主要为了降低计算复杂度。例如文献[21]对三角收缩阵列单元 Givens 旋转部分中的倒数与平方根计算采用分段多项式近似计算。文献[22]计算三角函数采用二元数值函数近似法, 直接计算出 Givens 矩阵, 并使用最小二乘回归方法进一步提高近似精度。此外, 对于复数矩阵 QR 分解问题, 文献[18, 21]均考虑将复数矩阵进行实数等效后再进行分解, 这降低了矩阵下三角化复杂度, 而文献[16, 19]等将复数矢量化成实数后再分解的方案, 减少了数据的交叉使用。

本文主要针对排序 QR 分解的低硬件资源消耗现场可编程门阵列 (Field programmable gate array, FPGA) 实现开展研究, 具体工作如下:

(1) 分析了基于排序 QR 分解的 OSIC 检测算法基本原理, 并着重分析基于 Givens 旋转实现排序 QR 分解的处理流程。

(2) 提出了一种基于深度流水化 CORDIC 迭代的低硬件资源消耗的 Givens 旋转实现方法, 显著降低了实现复杂度。

(3) 完成了 OSIC 检测方法的 FPGA 设计与实现, 通过软硬件仿真验证了本文所提硬件实现结构的高效性。

1 MIMO 信号检测

1.1 MIMO 系统模型

对于常见 MIMO 系统, 设 n_R 和 n_T 分别为接收天线数量和发送天线数量 ($n_R \geq n_T$), 第 (j, i) 个元素 h_{ji} ($i = 1, 2, \dots, n_T; j = 1, 2, \dots, n_R$) 表示发送天线 i 到接收天线 j 间的信道增益系数, 且为复数, 那么复信道矩阵 H 表示为

$$H = \begin{bmatrix} h_{11} & \cdots & h_{1n_T} \\ \vdots & \ddots & \vdots \\ h_{n_R 1} & \cdots & h_{n_R n_T} \end{bmatrix} \quad (1)$$

发送信号和接收信号分别表示为矢量形式: $\mathbf{s} = [s_1, s_2, \dots, s_{n_T}]^T$, $\mathbf{r} = [r_1, r_2, \dots, r_{n_R}]^T$, $n_R \times n_T$ 的 MIMO 系统表示为

$$\mathbf{r} = H\mathbf{s} + \mathbf{n} \quad (2)$$

式中: $\mathbf{n} = [n_1, n_2, \dots, n_{n_r}]^T$ 为加性高斯白噪声, 协方差 $E\{\mathbf{nn}^H\} = \sigma_n^2 \mathbf{I}_{n_r}$, σ_n^2 为噪声功率。

1.2 OSIC 检测

OSIC 检测算法将接收天线信号分成多路数据流后, 再进行零化和干扰消除操作^[10]。零化即通过 ZF 或 MMSE 进行检测并排序, 干扰消除操作即每一层信号检测前先减去上一层信号对该层的干扰。

假定先正确检测出第 1 层发射符号 s_1 , 估计值为 \hat{s}_1 , 已知信道矩阵 $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{n_t}]$, 第 1 层干扰为 $\mathbf{h}_1 \hat{s}_1$, 则现在第 2 层剩余信号为

$$\hat{\mathbf{r}}_1 = \mathbf{h}_1(s_1 - \hat{s}_1) + \mathbf{h}_2 s_2 + \dots + \mathbf{h}_{n_t} s_{n_t} + \mathbf{n} \quad (3)$$

以此类推, 直至检测出所有的发送信号, 以 $n_t = 4$ 为例, 其检测顺序如图 2 所示。

1.3 排序 QR 分解

OSIC 检测过程需要反复地进行矩阵求伪逆、排序、消元过程, 计算复杂度和硬件资源消耗极大, 给工程实现带来巨大挑战。排序 QR 分解是避免上述过程、降低计算复杂度的重要方法^[11]。

对式(1)中矩阵 \mathbf{H} 进行 QR 分解, 计算公式为

$$\mathbf{H} = \mathbf{Q}\mathbf{R} \quad (4)$$

式中: 酉矩阵 $\mathbf{Q}_{n_r \times n_t}$ 满足 $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}^{[12]}$, $\mathbf{R}_{n_t \times n_t}$ 为上三角矩阵, 有

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n_t} \\ 0 & R_{22} & \dots & R_{2n_t} \\ 0 & 0 & \dots & R_{3n_t} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & R_{n_t n_t} \end{bmatrix} \quad (5)$$

将式(2)左乘 \mathbf{Q}^H 得

$$\mathbf{y} = \mathbf{Q}^H \mathbf{r} = \mathbf{Q}^H \mathbf{H} \mathbf{s} + \mathbf{Q}^H \mathbf{n} = \mathbf{R} \mathbf{s} + \mathbf{v} \quad (6)$$

式中: $\mathbf{R} = \mathbf{Q}^H \mathbf{H}$, $\mathbf{v} = \mathbf{Q}^H \mathbf{n}$ 。对比式(3)可以得到, 对 \mathbf{H} 预处理后的信号依然可以按 OSIC 算法处理流程来检测。

式(6)可以展开矩阵形式为

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n_t} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n_t} \\ 0 & R_{22} & \dots & R_{2n_t} \\ 0 & 0 & \dots & R_{3n_t} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & R_{n_t n_t} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_{n_t} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{n_t} \end{bmatrix} \quad (7)$$

接下来考虑排序过程, 根据 OSIC 算法思想, 在符号判决前先对矩阵 \mathbf{R} 排序^[13], 以保证每一步的信噪比(Signal-to-noise ratio, SNR)都是按从大到小排列, 即有

$$\text{SNR}_i = \frac{E\{|s_i|^2\} |R_{ii}|^2}{E\{|v_i|^2\}} \sim |R_{ii}|^2 \quad (8)$$

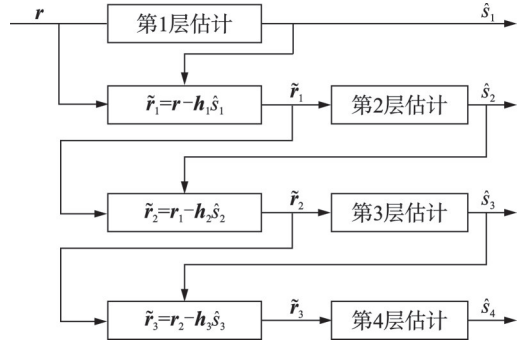


图 2 4×4 天线 OSIC 检测示意图
Fig.2 Schematic of 4×4 antenna OSIC detection

从判决过程来看译码顺序从最后一层开始从下往上进行,那么矩阵 R 应该满足 $|R_{ii}| (i = n_T, n_T - 1, \dots, 1)$ 从大到小排列。

2 OSIC 检测中排序 QR 分解硬件结构设计

2.1 基于 Givens 旋转的排序 QR 分解

首先考虑 QR 分解前排序过程,1.3 节表明检测顺序与排序顺序相反,在对 R 对角元素的计算过程中需要以计算矩阵列向量 1 范数 $\|x\|_1 = \sum_i |x_i|$ 为依据实现 $|R_{ii}|$ 最小化,而这需要计算每个复元素的模值。文献[14]给出了一种精度要求不高时估算复数模的方法,计算公式为

$$|x_i| \approx \alpha \cdot \max(|\operatorname{Re}(x_i)|, |\operatorname{Im}(x_i)|) + \beta \cdot \min(|\operatorname{Re}(x_i)|, |\operatorname{Im}(x_i)|) \quad (9)$$

式中: $\operatorname{Re}(\cdot)$ 表示复数实部, $\operatorname{Im}(\cdot)$ 表示复数虚部。一般取 $\alpha = 1, \beta = 1/4$ 便于硬件通过移位操作实现。

接下来具体阐述 Givens 旋转过程^[15],将 $[a, b]^T$ 右乘 Givens 旋转矩阵,即有

$$G(\theta)[a, b]^T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad (10)$$

$$\cos \theta = \frac{a}{\sqrt{|a|^2 + |b|^2}}, \quad \sin \theta = \frac{b}{\sqrt{|a|^2 + |b|^2}} \quad (11)$$

经过旋转后的 a 变为 $\sqrt{|a|^2 + |b|^2}$, b 变为 0。对实向量 $x = [x_1, x_2, \dots, x_p]^T$ 将第 $j-1, j$ 个元素乘上该矩阵就能使 x_j 变为 0。

更进一步,考虑复向量 x ,注意到复数实部和虚部同样可以乘上该矩阵使虚部变为 0,此为分解流程设计带来极大便利,此时

$$G(\theta) = \begin{bmatrix} (\cos \theta)^* & (\sin \theta)^* \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (12)$$

可见,原矩阵 H 乘上 Givens 旋转矩阵后只影响第 $j-1, j$ 行并只使一个元素置零,这样经过多次旋转后其对角线以下的值可全部被置零,最终得到上三角矩阵 R 。

仿真伪代码逻辑如算法 1 所示:

算法 1 基于 Givens 旋转变换的 SQRD

- (1) $Q = 0, R = H, P = I_{n_R \times n_T}$
- (2) For $i = 1, 2, \dots, n_T$
- (3) $k = \operatorname{col}(\operatorname{argmin}_{l=i, \dots, n_T} \|R(*, l)\|)$ (计算矩阵列向量范数)
- (4) 变换 R, P 中的第 k 列
- (5) END
- (6) For $i = 1, 2, \dots, n_T$
- (7) For $j = n_R: -2:i + 1$
- (8) 由下至上计算第 j 和第 $j-1$ 元素实虚部得到旋转矩阵 $G_{ji}^1, G_{(j-1)i}^1$
- (9) 第 j 和第 $j-1$ 个元素变为实数,同一行元素作角度相同的变换。
- (10) 计算第 j 和第 $j-1$ 行元素旋转矩阵 G_{ji}^2 。
- (11) 将第 i 列向量的第 j 行元素消去。

(12) END

(13) 变换后的矩阵 R 中, 向量 $R(i:n_R, i)$ 从 $i + 1$ 开始的元素都变为 0。

(14) END

因为 Q 在所提检测算法中没有被使用, 所以不必计算。

2.2 FPGA 硬件实现

图 3 展示了本文 FPGA 设计顶层结构, 主要包括了一个初始化(Initial)模块、一个排序(Sort)模块、一个控制(Control)模块和一个 Givens 模块的 4 个时刻分时复用。为了方便分析, 以 4×4 的 MIMO 系统 OSIC 检测为例进行说明。信道矩阵 H 为 4×4 复矩阵, 接收信号向量 r 为 4×1 矢量, 矩阵 H 为流水化逐列输入, 向量 r 在 H 输入完成时刻再输入。矩阵 H 经过初始化模块后记录下其矩阵列输入的顺序, 再进入排序模块以计算其 1 范数为依据将 H 重排列, 然后经过 4 级 Givens 旋转模块, 每级 Givens 通过 Control 模块控制 CORDIC 基本单元的矢量 (Vec) 和旋转 (Rot) 模式以及数据消元顺序, 将矩阵对应列下三角元素进行置零, 最终输出 R 上三角矩阵。随后 r 重复矩阵 H 的消元过程, 得到输出向量映射 $y = Q^H r$, 同时输出排序后的列置换顺序 P , 以便进行下一步译码。

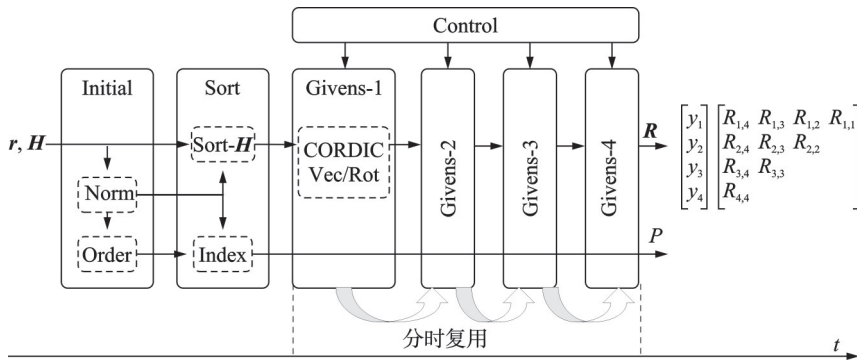


图 3 顶层模块

Fig.3 Top module

2.2.1 初始化模块

初始化模块将 H 输入列向量进行范数计算并标记序号, 输入数据第 1 列 h_1 后计数值 Order 就置零, 同时 h_1 进入 Norm 单元分别先取实部和虚部的绝对值 (即符号位都置零), 再延迟一个时钟周期后就可以计算列范数 $Norm_i$ 并存储到寄存器中, 再经过一个时钟周期计算出 $Norm_{i+1}$ 就可以将 $Norm_i$ 输入到 Sort 模块开始进行比较排序, 同时对应的 Order 计数也输入到 Sort 模块中。考虑到输入向量映射 y 和上三角矩阵 R 同步, 在 Initial 模块中还需要对接收信号向量 r 进行缓存, 延迟输入时钟为排序所花费的时钟。

2.2.2 排序模块

排序模块如图 4 所示, 该电路由 1 个计数器、1 个寄存器、1 个比较器和 1 个多路选择器组成。由于电路结构简单, 这里可以同时使用两个该模块进行最大值和最小值的寻找。

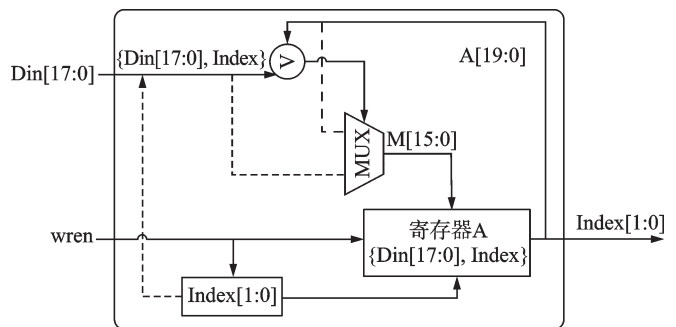


图 4 排序模块

Fig.4 Sort module

当使能信号 $wren$ 有效时数据输入, 前一个数据输入到该模块后与其序号 $Index$ 组合 ($Index$ 范围为 $0\sim 3$), 存储在寄存器 A 中, 下一个输入的数据与前一个数据进行比较, 若小于 (或大于) 前一个值则寄存器 A 的值更新, 同时计数器 $Index$ 的值也会更新, 否则保持原来的值不变。这样矩阵 H 对应的列范数 $Norm$, 输入完毕 (共 4 个时钟周期) 就能同时找到并输出最小和最大列范数的所在列序号。更进一步, 同样利用该排序模块结构, 在寻找最大值的过程中同时寻找次大值, 只是这时寄存器 A 中的值每次应该为当前所找到的最大值, 这样在第 5 个时钟周期就能找出其中最大、次大、最小值的顺序, 那么剩下的值就是次小值。

值得注意的是, 列范数计算了 4 个输入数据和, 所以位宽扩大 2, 而排序结果输出仅为序号, 便于直接寻址, 同时每次比较都是将对应序号 $Index$ 包含进去, 此时即使计算出的列范数相等, 其比较结果也会保持序号小的在序号大的前面, 并不会对排序结果产生影响。

2.2.3 深度流水化CORDIC迭代结构设计

对于Givens旋转中的角度变换过程, 硬件中常采用CORDIC算法^[18]。图5展示了CORDIC基本单元, 其包含 Vec 和 Rot 模式, 通过 Control 模块输出 Mode 信号控制其模式改变, Sign 中存储角度旋转符号 (用 ± 1 表示)。本文对 CORDIC 迭代单元结构进行了寄存器优化, 如图 6 所示, 利用该结构进行角度旋转只需进行简单的加减、移位操作, 极大地提高了流水化程度。

将一路数据连续输入, 每个时钟进行 1 次迭代过程以处理 1 个元素。当第 1 个元素输入时, CORDIC 工作在矢量模式, 将输入的该元素的实部和虚部分别进行一次加减和移位操作来进行一级迭代, 并将最高位符号位进行异或后延迟一个时钟后送入 Sign 存储, 计算后数据直接传递到下一级迭代, 迭代总次数由具体需求确定, 综合考虑 MIMO 检测性能后可定为 12 次^[19]。在第 1 个元素在第 i 级进行矢量操作的同时, 第 2 个元素已经进入到第 $i-1$ 级, 这时第 $i-1$ 级控制信号 Mode 就改变模式为旋转模式, 根据第 $i-1$ 级的存储的 Sign 符号值判断旋转方向, 同步进行第 2 个元素加减和移位操作, 这样后一个元素始终只比前一个元素延迟输入 1 个时钟周期, 实现了数据的深度流水化计算。对于一行输入 4 个元素, 第 1 个元素迭代完成需要 12 个时钟周期, 后 3 个元素依次延迟 1 个时钟周期输出, 最后将输出数据存储, 总共需要 $12+3+1=16$ 个时钟周期。

2.2.4 基于CORDIC迭代的Givens旋转模块

Givens 旋转模块是对输入矩阵进行 QR 分解的核心部分, 该模块通过 2.1 节介绍的旋转消元方法, 将 Sort 模块输出的排序后矩阵 $sort-H$ 的下三角元素消除, 也可以完成 r 到 y 的映射。如图 7 所示, 图中 C 代表复数, R 代表实数, 下标仅代表元素顺序, 不作为实际数值变化。消元操作分为两步, 第 1 步将输入的复数进行矢量化变为实数, 第 2 步对输入的两个实数进行置零旋转, 从而消除 1 个元素。

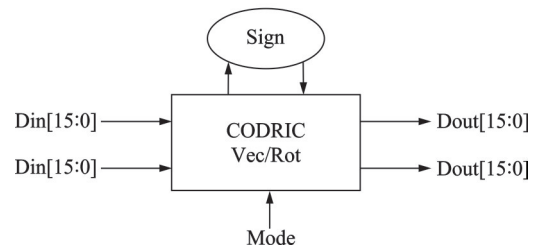


图5 CORDIC基本单元结构

Fig.5 CORDIC basic unit structure

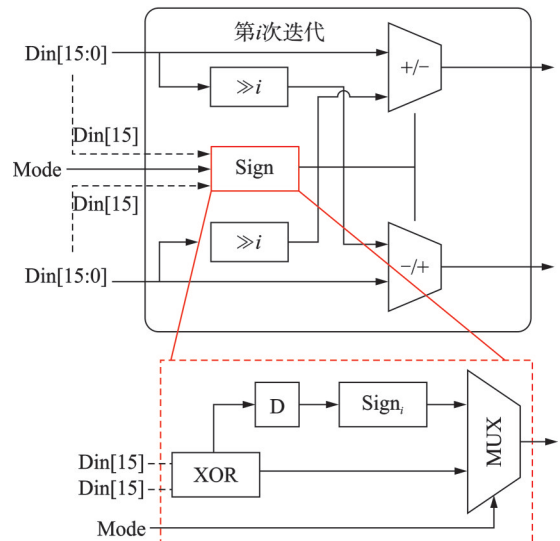


图6 CORDIC迭代单元结构

Fig.6 CORDIC iterative unit structure

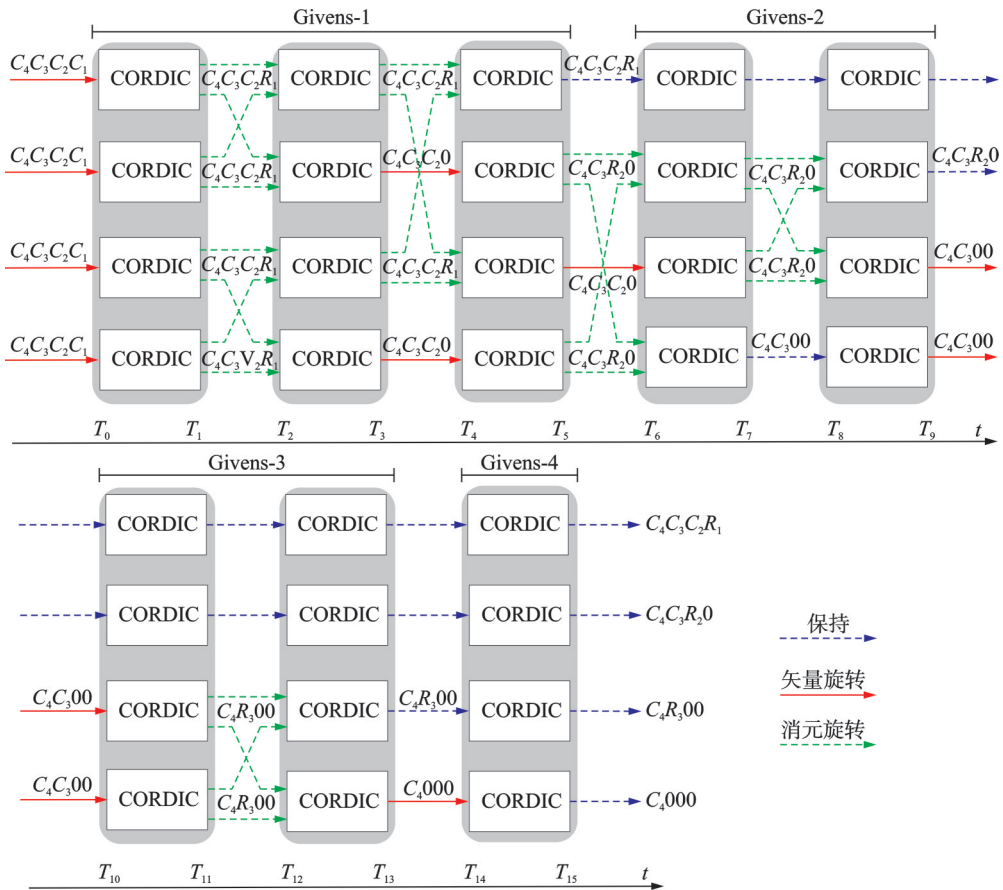


图7 基于CORDIC迭代的Givens旋转结构

Fig.7 Givens rotation structure based on CORDIC iterative

针对4路输入数据将CORDIC单元并行使用4个,组成1个Givens模块。对于Givens-1模块,输入第1列向量时首先控制信号为矢量模式,CORDIC第1级迭代开始就将第1个列向量元素进行矢量化操作(图7中红色箭头表示),紧随其后输入的列向量则改变控制信号为旋转模式,与第1级进行一样的旋转变换(图7中绿色箭头表示),这样就完成了同一路数据流的流水线式变化。对于后续不同路数据的消元操作,需要在Control模块的控制下选择数据映射顺序,将某两路的数据流重新映射到对应位置上的CORDIC模块输入信号,比如Givens-1第2步流程就通过数据重映射排列来消去第2行和第4行的第1列元素,第3步就消去了第3行的第1列元素,这时第1行就不再参与后续计算,Control模块给出让其保持输出的信号(图7中蓝色箭头表示),以此类推,Givens-2~Givens-4分时逐列完成矢量化和消元操作,每一步流程对应的具体操作都已在图7通过不同颜色的线型箭头进行标识。

最终输出上三角化的矩阵R。值得注意的是,对于后续输入接收信号r映射y重复进行上述过程必须调整CORDIC模块为旋转模式,其余操作均相同。

对于4×4信道矩阵H的三角化过程,根据图7中的分解流程,仅需8×16=128个时钟周期。再加入排序过程,总共需要128+5=133个时钟周期。由此可见,通过并行化和深度流水线处理,能够显著降低系统的处理时延,进而提升吞吐量。

3 实现结果分析

3.1 Matlab 仿真结果

对ZF、MMSE、基于信噪比的OSIC检测与本文算法进行了Matlab仿真性能比较。参数设置为 4×4 MIMO、16阶正交幅度调制(16-ary Quadrature Amplitude Modulation, 16QAM)、瑞利平坦衰落信道加性高斯白噪声、蒙特卡洛方法次数为10 000。如图8所示,本文实现的算法与基于信噪比的OSIC信号检测算法误比特率相近,证明了本文所提算法的可行性。

3.2 FPGA 实现结果

通过Modelsim进行寄存器传输级(Register transfer level, RTL)仿真验证本文设计结构的正确性,并与Matlab仿真结果进行比较分析。

首先随机生成 4×4 复信道瑞利衰落矩阵 H ,将产生的 $(-1 \sim 1)$ 范围内数据乘上8 192后进行取整量化,量化后的数据保存为txt文件,再导入测试激励文件然后进行仿真。在运算过程中,数据量化位宽也会对系统性能产生影响,考虑到浮点数在硬件实现的不便,本文对计算数据都做了定点化处理,接下来通过具体分析仿真流程中的量化误差来确定数据位宽。根据第2.2节的流程介绍,首先输入矩阵排序过程,可知输入数据分为I、Q两路将实部和虚部分流输入,经过排序后将输入数据列交换后输出,这一部分不存在误差。

接下来进行QR分解过程,这一过程Matlab和Modelsim仿真结果往往存在个位数级别的误差,这是因为现场可编程门阵列(Field-programmable gate array, FPGA)计算有符号数的过程中,对于负数的取整往往会存在1 bit的量化误差,在CORDIC迭代过程中这种误差被累积至最终结果中,但此误差在可接受范围内。

在完成了QR分解后,CORDIC迭代过程中保存的角度变化符号就能够再次用于对接收信号 r 进行重映射,即对 r 重复进行QR分解中的旋转过程,即 r 的映射过程再次利用了 H 进行QR分解过程中已经存储的角度旋转符号Sign。采用均方误差(Mean Squared Error, MSE)来评价量化误差,如图9所示。从图9可以得到数据位宽为16 bit时MSE达到最小,因此数据采用16 bit量化(1符号位+3位整数+12位小数)。

RTL级仿真已经验证了本文设计结构的可实现性,接下来进行FPGA综合产生实际电路后,对比资源占用和电路时序等情况分析该结构的可行性,使用Vivado 2015.4在平台Zynq-7000(xc7z045ffg900-1)上对各模块资源进行评估。

表1中对主要计算单元的资源利用情况进行对比,可以看出本文排序QR分解方案将乘法器使用资源减少到0,优化效果突出。表2中本文方法与相关文献比较了输入 4×4 复矩阵时系统整体资源。与文献[17]比较, Slice资源减少84.5%,查找表(Look up table, LUT)资源减少78%,触发器(Flip flop, FF)资源减少91%,而且没有使用DSP48,但处理时钟几乎一

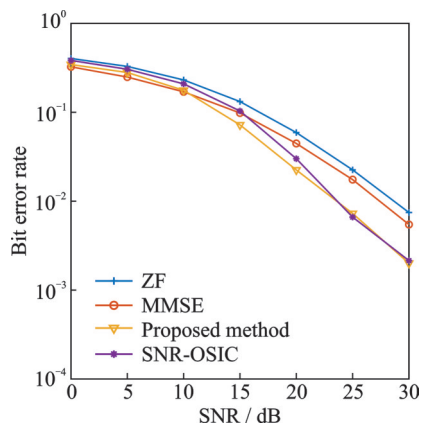


图8 检测算法性能比较

Fig.8 Comparison of performance of detection algorithms

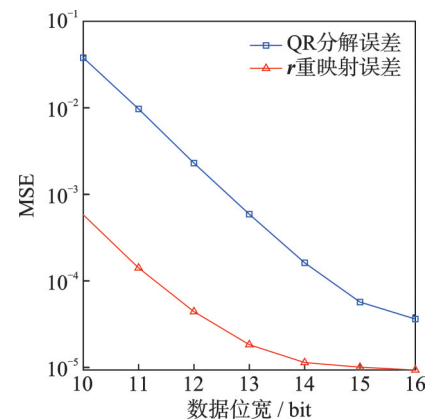


图9 数据量化位宽对计算误差影响

Fig.9 Effect of data quantization bit width on MSE

表1 计算单元资源利用

Table 1 The resource utilization of computing units

比较指标	文献[17]		本文	
	CORDIC	乘法	CORDIC	乘法
SliceLUTs	606	1 274	585	0
SliceFFs	626	1 412	603	0

表2 整体资源比较

Table 2 Resource comparison of the entire design

类别	文献[9]	文献[[17]	文献[19]	本文
平台	Virtex7	Virtex7	Virtex6	Zynq-7 000
矩阵维度	4×4复	4×4复	4×4复	4×4复
方案	Givens旋转 +CORDIC	Givens旋转 +CORDIC	Givens旋转 +CORDIC	Givens旋转 +CORDIC
操作	排序	排序	不排序	排序
综合频率 MHz	333	330	128	250
Slice	20 315	9 078	1 851	1 409
Slice LUTs	71 737	31 468	—	4 694
Slice FFs	78 984	35 983	—	3 230
DSP48	448	48	—	0
处理时钟 cycles	—	132	137	133
吞吐量	—	2.5	0.934	1.88
单位面积吞吐量	—	275	505	1 334

注:吞吐量:M Matrices/s 单位面积吞吐量=吞吐量/Slice

致,吞吐量仅减少24.8%,而单位面积吞吐量提升约3.85倍。同文献[19]相比,Slice资源减少23.9%,而吞吐量提升约1倍,单位面积吞吐量提升约1.64倍。

表3为模块时序分析。根据表3可知,当整个OSIC模块需要主频时钟达到250 MHz(4 ns)时,利用本文所提硬件结构,此时最差负时序裕量(Worst negative slack, WNS)和最差保持时序裕量(Worst hold slack, WHS)分别还有0.509 ns、0.041 ns的设计剩余,能够满足高吞吐量的设计需求。

表3 时序分析

Table 3 Chronological analysis

设计频率	WNS/ns	WHS/ns
250 MHz(4 ns)	0.509	0.041

4 结束语

针对MIMO系统OSIC信号检测算法中矩阵逆运算存在硬件实现方面复杂度高的难题,本文采用基于Givens旋转进行排序QR分解的矩阵分解方案,提出了基于1范数估算的排序优化方法和利用深度流水化CORDIC迭代实现矩阵元素矢量化和旋转操作的方案,显著降低了计算复杂度。仿真结果表明,本文所提算法误比特性能与常用基于信噪比的OSIC检测算法性能基本一致。在FPGA具体实现过程中,设计了硬件电路结构并仿真验证了RTL级功能,分析资源、时序占用。结果对比表明,CORDIC模块几乎不增加额外资源消耗,而实现过程中不再需要复数乘法计算资源开销。以4×4复矩

阵分解为例,由于所提结构能够分时复用支持高度并行,设计的系统最高时钟能达到 250 MHz,吞吐量达到了 1.88 M Matrices/s。与其他文献方案比较,关键资源如 Slice 至多能减少 84.5%,其他系统资源占用也极大减少,而计算延迟没有明显增长,单位面积吞吐量至少能提升约 1.64 倍。本文所提结构能够满足 4 天线及以上系统接收端处理吞吐量和时延需求,在高速 MIMO 系统中的应用具有很大潜力。

参考文献:

- [1] 尤肖虎, 潘志文, 高西奇, 等. 5G 移动通信发展趋势与若干关键技术[J]. 中国科学:信息科学, 2014, 44(5): 551-563.
YOU Xiaohu, PAN Zhiwen, GAO Xiqi, et al. 5G mobile communication development trend and some key technologies[J]. Science China: Information Science, 2014, 44(5): 551-563.
- [2] LIU L B, PENG G Q, WEI S J. Massive MIMO detection algorithm and VLSI architecture[M]. Beijing: Science Press, 2019: 23-24.
- [3] CHATAUT R, AKL R. Massive MIMO systems for 5G and beyond networks—Overview, recent trends, challenges, and future research direction[J]. Sensors, 2020, 20(10): 2753.
- [4] 张娟. Massive MIMO 系统信号检测算法研究[D]. 重庆:重庆邮电大学, 2021.
ZHANG Juan. Research on signal detection algorithm for massive MIMO system[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2021.
- [5] 庄勇. MIMO 系统中基于 QR 分解算法的信号检测[D]. 南京:南京信息工程大学, 2017.
ZHUANG Yong. Signal detection based on QR decomposition algorithm in MIMO system[D]. Nanjing: Nanjing University of Information Engineering, 2017.
- [6] SHAHABUDDIN S, ISLAM M, SHAHABUDDIN M, et al. Matrix decomposition for massive MIMO detection[C]// Proceedings of 2020 IEEE Nordic Circuits and Systems Conference (NorCAS). Norway: IEEE, 2020:1-6.
- [7] WUBBEN D, BOHNKE R, VOLKER K, et al. MMSE extension of V-BLAST based on sorted QR decomposition[C]// Proceedings of 2003 IEEE the 58th Vehicular Technology Conference. Orlando, USA: IEEE, 2003: 508-512.
- [8] 马晓双. 矩阵 QR 分解的 FPGA 设计研究[D]. 成都: 电子科技大学, 2017.
MA Xiaoshuang. Research on FPGA design of matrix QR decomposition[D]. Chengdu: University of Electronic Science and Technology of China, 2017.
- [9] 陈健, 庄耀宇, 杨丹, 等. 基于 FPGA 的改进的排序 QR 分解实现[J]. 湖南大学学报(自然科学版), 2022, 49(10): 8-16.
CHEN Jian, ZHUANG Yaoyu, YANG Dan, et al. Improved FPGA-based implementation of sorted QR decomposition[J]. Journal of Hunan University (Natural Science Edition), 2022, 49(10): 8-16.
- [10] 冯皎, 胡冬顺, 李鹏. MIMO 系统中基于信号可靠性判定的 OSIC 检测[J]. 电讯技术, 2020, 60(9): 1020-1027.
FENG Jiao, HU Dongshun, LI Peng. OSIC detection based on signal reliability determination in MIMO systems[J]. Telecommunications Technology, 2020, 60(9): 1020-1027.
- [11] 陈礼锐. 面向未来无线通信的 MIMO 检测预处理算法及硬件结构研究[D]. 长沙:国防科技大学, 2020.
CHEN Lirui. Research on MIMO detection preprocessing algorithm and hardware structure for future wireless communication [D]. Changsha: National University of Defence Technology, 2020.
- [12] 魏木生, 李莹, 赵建立. 广义最小二乘问题的理论与计算[M]. 2 版. 北京: 科学出版社, 2020: 270-271.
WEI Musheng, LI Ying, ZHAO Jianli. Theory and computation of generalised least squares problems[M]. 2nd ed. Beijing: Science Press, 2020: 270-271.
- [13] 付向聪. 室内高速 MIMO-OFDM 系统接收机信号检测技术研究[D]. 西安: 西安电子科技大学, 2020.
FU Xiangcong. Research on receiver signal detection technology for indoor high-speed MIMO-OFDM system[D]. Xi'an: Xidian University, 2020.
- [14] AUGER F, LOU Z, FEUVRIE B. Multiplier-free divide, square root, and log algorithms [DSP tips and tricks][J]. IEEE Signal Processing Magazine, 2011, 58(4): 122-126.
- [15] LU H Y, YEN M H, CHANG C W, et al. Efficient hardware implementation of CORDIC-based symbol detector for GSM MIMO systems: Algorithm and hardware architecture[J]. IEEE Access, 2022, 10: 114232-114241.

- [16] MALTSEV A, PESTRETSOV V, MASLENNIKOV R. Triangular systolic array with reduced latency for QR-decomposition of complex matrices[C]//Proceedings of 2006 IEEE International Symposium on Circuits and Systems (ISCAS). Kos: IEEE, 2006: 4.
- [17] 庄耀宇. 基于FPGA的排序算法研究及其在MIMO检测中的应用[D]. 上海: 上海大学, 2022.
ZHUANG Yaoyu. Research on FPGA-based sorting algorithm and its application in MIMO detection[D]. Shanghai: Shanghai University, 2022.
- [18] LEE H, KIM H, CHOJ M. Low-latency implementation of CORDIC-based sorted QR decomposition for high-speed MIMO-OFDM system[C]//Proceedings of 2018 the 28th International Conference on Radio Electronic. Prague, The Czech Republic: [s.n.], 2018: 1-4.
- [19] ZHAO W, LIN J, CHAN S C. Throughput/area efficient FPGA implementation of QR decomposition for MIMO systems [C]//Proceedings of 2016 IEEE International Conference on Digital Signal Processing (DSP). Beijing: IEEE, 2016: 522-526.
- [20] LEE M W, YOON J H, PARK J. High-speed tournament givens rotation-based QR decomposition architecture for MIMO receiver[C]//Proceedings of 2012 IEEE International Symposium on Circuits and Systems (ISCAS). Korea (South): IEEE, 2012: 21-24.
- [21] SANTOS L C, ATOCHE A C, CASTILLO J V, et al. An improved hardware design for matrix inverse based on systolic array QR decomposition and piecewise polynomial approximation[C]//Proceedings of 2015 International Conference on Reconfigurable Computing and FPGAs. Riviera Maya, Mexico: IEEE, 2015: 1-6.
- [22] RUST J, SEIDEL P, KNOOP B, et al. Hardware-efficient QR-decomposition using bivariate numeric function approximation [J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2017, 64(12): 3150-3159.

作者简介:



王海麟(2001-),男,硕士研究生,研究方向:无线通信信号处理,E-mail:wang-hailin@nudt.edu.cn。



冯献礼(2001-),男,硕士研究生,研究方向:无线通信信号处理。



辜方林(1986-),通信作者,男,博士,副研究员,研究方向:可靠无线通信,E-mail:gufanglin@nudt.edu.cn。



高明柯(1980-),男,博士研究生,高级工程师,研究方向:智能计算,无人集群智能。



赵海涛(1981-),男,教授,博士生导师,研究方向:多信道认知组网、无人集群通信。

(编辑:陈珺)