

一种适用于 MIMO 系统 MMSE 检测的 LDLT 分解硬件实现

梁子豪¹, 刘 辉², 贺光辉¹

(1 上海交通大学 电子信息与电气工程学院, 上海 200240; 2 上海卫星通信工程研究所, 上海 201109)

摘 要: 提出了一种适用于 MIMO 系统 MMSE 检测的 LDLT 分解算法的硬件实现. 通过三角阵形式的脉动阵列完成对待求逆矩阵的 LDLT 分解操作, 合理利用待求逆矩阵的共轭对称性, 输入与输出的矩阵均仅需为下三角矩阵. 同时, 针对 LDLT 分解算法中的除法, 采用求倒结构来避免使用除法器. 同时合理安排时序, 使脉动阵列结构中求倒结构能够分时复用, 并采用查找表压缩技术, 明显减小了求倒结构的硬件开销. 仿真结果和实现报告表明该硬件实现方法在保证吞吐率和误码率性能的前提下, 以较低的硬件资源消耗完成了 MMSE 检测算法中的矩阵分解.

关键词: MIMO; MMSE 检测; LDLT 分解; SBTM 算法

中图分类号: TN929. 5

文献标识码: A

文章编号: 1000-7180(2018)08-0082-05

A Hardware Implementation of LDLT Decomposition for MMSE Detection in MIMO System

LIANG Zi-hao¹, LIU Hui², HE Guang-hui¹

(1 School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 2 Shanghai Institute of Satellite Engineering, Shanghai 200240, China)

Abstract: A Hardware implementation of LDLT decomposition is proposed for MMSE detection in MIMO system. A systolic array of triangle shape is adopted for matrix inversion. By making use of the property of the Hermitian matrix to be inversed, the input and output can just be lower triangular matrix. Moreover, aiming at the divisions in the LDLT decomposition algorithm, a reciprocal architecture is proposed to avoid dividers. By arranging the timing, time-division multiplexing of the reciprocal architecture can be achieved. And by using look-up table compression technique, the hardware utilization of the reciprocal architecture can be reduced obviously. The simulation result and the implementation result show that the proposed hardware implementation can compute LDLT decomposition of MMSE detection in low hardware utilization, under the premise of high throughput and good bit error rate performance.

Key words: MIMO; MMSE detection; LDLT decomposition; SBTM algorithm

1 引言

多输入多输出 (multiple-input multiple-output, MIMO) 技术是现代通信技术的核心技术之一. 在发射端和接收端均采用多根天线, 可以在不提高高频带宽度的情况下达到更高的数据传输速率^[1]. 而 MIMO 系统中的检测算法则是整个系统中非常重要的一环. 由于发射端和接收端均有多根天线, 且

信道中存在着噪声, 所以接收端的接收信号是发射向量混叠再加上噪声的结果. 检测算法则是将接收向量还原回发射信号的算法. 在众多检测算法中, 最小均方差估计方法 (minimum mean square error, MMSE) 是非常常见的一种线性检测算法. 根据信道矩阵和噪声能量, MMSE 检测可以保证对发射向量的估计与发射向量之间的均方差最小. 但是 MMSE 检测中包含了矩阵求逆, 对硬件实现具有一

定的挑战性.在保证性能的情况下,如何设计高吞吐量、低资源消耗的检测器成为了 MIMO 技术研究的一个难点.

发射端和接收端的天线数量均较少的一般 MIMO 系统的信道不具有信道硬化的性质,因而在矩阵求逆上通常采用直接法,即利用各种矩阵分解将待求逆矩阵分解为较容易求逆的形式再进行求逆.由于待求逆矩阵的共轭对称性,LDLT 分解可以节省较多的硬件资源,但是其分解算法中包含了除法,不利于硬件实现.

为此,一种适用于 MIMO 系统 MMSE 检测的 LDLT 分解的硬件实现方法被提出.为了保证系统的吞吐量,采用脉动阵列结构实现 LDLT 分解算法.针对分解算法中存在的除法,将除法改写为求倒和乘法,避免直接使用除法器.同时,为了减少求倒结构的硬件开销,需要合理安排时序,将求倒结构完全地分时复用;同时采用查找表压缩算法,将求倒结构的硬件资源消耗进一步降低.最终 FPGA 实现的结果表明,提出的硬件实现方法可以在保证吞吐率的情况下,以较低的硬件资源消耗完成 MMSE 检测中的 LDLT 分解.

2 背景

2.1 MIMO 模型

现考虑 MIMO 系统的发射天线数量为 M_t ,接收天线数量为 M_r .

发射端的原始数据经过编码、交织以及调制之后,会得到相应的复数符号.经过串并转换之后,发射端的天线会同时发射一组规模为 $M_t \times 1$ 的复数发射向量,记为 x .发射过程中,以信道矩阵 $\mathbf{H}$ 描述信道增益, $\mathbf{H}$ 是规模为 $M_r \times M_t$ 的复数矩阵.同时,会有加性高斯白噪声与接收向量叠加,以 n_0 表示. n_0 是规模为 $M_r \times 1$ 的复数向量,每个元素均独立地服从均值为 0,方差为 σ^2 的高斯分布.最终,接收端收到规模为 $M_r \times 1$ 的复数接收向量,以 y 表示,整个 MIMO 系统可用以下公式描述:

$$y = \mathbf{H}x + n_0 \quad (1)$$

2.2 MMSE 检测算法

文献[2]指出,MMSE 检测算法的主要思想是同时考虑消除各发射天线的发射信号间的干扰以及抑制噪声.其目标是求出一个矩阵 $\mathbf{G}_{\text{MMSE}}$,使得检测模块输出的对发射向量的估计值 $\hat{x} = \mathbf{G}_{\text{MMSE}}y$ 与其真实值 x 之间的均方差最小,用如下公式描述:

$$\mathbf{G}_{\text{MMSE}} = \arg \min_{\mathbf{G}} \|\mathbf{G} \cdot y - x\| \quad (2)$$

公式(2)的解由以下公式给出:

$$\mathbf{G}_{\text{MMSE}} = (\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I})^{-1} \mathbf{H}^H \quad (3)$$

式中, $\mathbf{I}$ 是规模为 $M_t \times M_t$ 的单位矩阵; σ^2 表示加性高斯白噪声 n_0 的能量.可以看出,MMSE 检测算法需要对矩阵 $\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I}$ 进行求逆,该矩阵是一个规模为 $M_t \times M_t$ 的复数共轭对称矩阵.

2.2 LDLT 分解算法

LDLT 分解算法是 Cholesky 分解算法的改进型,其分解对象是共轭对称矩阵.LDLT 分解算法相比于 Cholesky 分解算法的主要优势是不需要进行开方操作.

假设 $\mathbf{A}$ 为待分解矩阵,LDLT 分解的算法流程如下所示:

for $k = 1:n$ (4)

$\mathbf{A}(k,k) = 1/\mathbf{A}(k,k)$ (5)

$\text{temp}(k+1:n,k) = \mathbf{A}(k+1:n,k)$ (6)

$\mathbf{A}(k+1:n,k) = \mathbf{A}(k+1:n,k) \cdot \mathbf{A}(k,k)$ (7)

for $j = k+1:n$ (8)

$\mathbf{A}(j:n,j) = \mathbf{A}(j:n,j) -$ (9)

$\text{temp}(j:n,k) \cdot \mathbf{A}^H(j,k)$

end (10)

end (11)

该算法中的除法已经分别被改写成(5)中的求倒和(6)中的乘法.

LDLT 分解的形式为 $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^H$,其中 $\mathbf{L}$ 是单位下三角矩阵, $\mathbf{L}^H$ 是其共轭转置矩阵, $\mathbf{D}$ 为对角矩阵.由于 $\mathbf{A}$ 本身具有共轭对称性,所以其下三角矩阵已经包含了整个矩阵的信息.因此,上述算法的输入为 $\mathbf{A}$ 的下三角矩阵;输出同样为下三角矩阵,输出矩阵的对角线上的元素构成 $\mathbf{D}$,其余元素构成 $\mathbf{L}$ 的非对角元素,由于 $\mathbf{L}$ 是单位下三角矩阵,所以其对角元素无需存储.

3 LDLT 分解的硬件实现

3.1 基于脉动阵列结构的 LDLT 分解架构设计

为了保证高吞吐量,采用三角形的脉动阵列结构完成 LDLT 分解操作.如图 1 所示,以 4×4 规模的 MIMO 为例,该脉动阵列结构包含两种处理单元,一种是三角形斜边上的处理单元,称为边界处理单元(boundary processing element, BPE),用圆形表示;其余的为内部处理单元(internal processing element, IPE),用矩形表示. BPE 负责完成算法中的(5)和(7),因此 BPE 包含了求倒结构和复数乘法器. BPE 将上方的第一个输入求倒,将后续数据均

人在两个查找表中查得两个值,最后将两个值相加得到最终结果.两个查找表存储的值分别用以下公式求得:

$$a_0(x_0,x_1)=1/(1+x_0+x_1+\delta_2+\delta_3) \tag{13}$$

$$a_1(x_0,x_2)=\frac{(\delta_2-x_2)}{(1+x_0+\delta_1+\delta_2+\delta_3)^2} \tag{14}$$

式中,

$$\delta_1=2^{-n_0-1}-2^{-n_0-n_1-1} \tag{15}$$

$$\delta_2=2^{-n_0-n_1-1}-2^{-n_0-n_1-n_2-1} \tag{16}$$

$$\delta_3=2^{-n_0-n_1-n_2-1}-2^{-n_0-n_1-n_2-n_3-1} \tag{17}$$

因此,只需要将 a_0 和 a_1 存入两个查找表.此时,查找表的深度由 $2^{n_0+n_1+n_2}$ 降低为 $2^{n_0+n_1}+2^{n_0+n_2}$.

文献[5]同时还指出,由于存放 a_1 的查找表中的数值具有一定的对称性,该查找表的大小还能再压缩一半.其理论依据来自于以下公式:

$$a_1(x_0,x_2)=-a_1(x_0,2\delta_2-x_2) \tag{18}$$

该公式很容易证明,不加赘述.由于 a_1 的对称性,当 x_2 为 $2\delta_2-x_2$ 的补码时, $a_1(x_0,x_2)$ 同样是 $a_1(x_0,2\delta_2-x_2)$ 的补码.因此,在进行存储时,输入位宽可以舍去 x_2 的最高位,查找表大小缩小为一半.在进行查找时,先判断 x_2 的最高位是0还是1,若为0,则直接查找;若为1,则先将 x_2 其余位取补码再查找,将查到的值再取补码作为结果.

整个求倒结构及利用SBTM算法优化后的查找表细节示意图如图2所示:

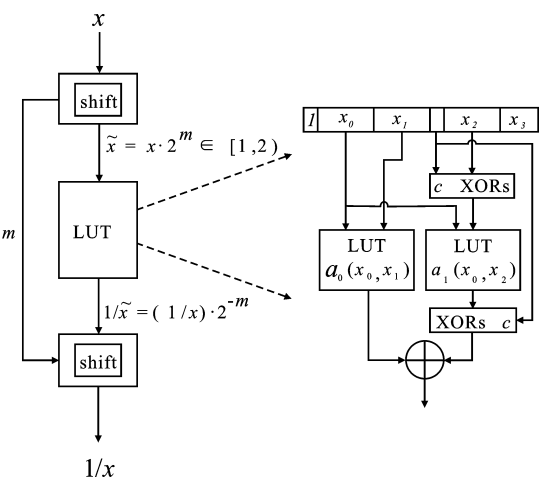


图 2 求倒结构及查找表细节

利用SBTM算法压缩查找表的效果和对 x 的具体分段相关.在数据位宽相同的情况下,压缩比如下所示:

$$ratio=\frac{2^{n_0+n_1+n_2}}{2^{n_0+n_1}+2^{n_0+n_2}-1} \tag{19}$$

举例说明:假设查找表存储的数据位宽均相同,

一个普通的查找表地址位宽为12,即深度为4 096,若将地址位宽划分为5,4,3,那么最终两个查找表深度分别为512和128,共640.该情况下总压缩比为6.4.

4 仿真及实现结果

4.1 BER 性能

在第三章提出的脉动阵列结构的LDLT分解架构中,在求倒结构中采用SBTM算法优化查找表大小之后,会对误比特率(bit error rate, BER)性能造成一定的影响.因此,在仿真中将使用了SMTM优化的LDLT分解结构与不使用优化的结构进行BER性能对比.信道模型为理想的随机信道,发射天线数量与接收天线数量均为8,噪声为加性高斯白噪声,发射端采用码率为5/6、多项式为[131 171]的卷积码编码与64QAM调制,两者在接收端均采用MMSE检测、硬判决,使用BCJR(Baul, Cocke, Jelinek and Raviv)算法译码.同时,不加优化的查找表深度为8 192,宽度为18,采用SBTM优化的查找表取 $n_0=5, n_1=4, n_2=4, n_3=0$.仿真结果如所示:

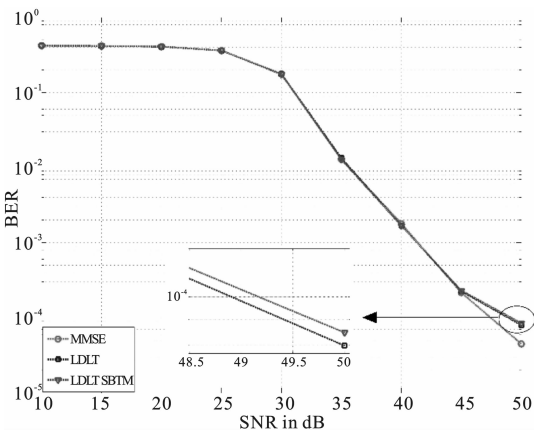


图 3 BER 性能对比图

从图3可以看出,在BER为 10^{-4} 的时候,采用SBTM算法压缩的查找表相比不加压缩的查找表性能损失不到0.5 dB.根据以上参数可得,不加优化时,查找表大小为144 kb,优化之后大小为13.5 kb,压缩比达到了10.67.可以证明采用SBTM算法可以在性能损失很小的情况下明显减少查找表大小.

4.2 FPGA 实现结果

利用Xilinx Vivado 2014.4.1工具对本文提出的设计在Virtex 7-XC7VX690T FPGA芯片上进行实现,目标MIMO系统规模是 8×8 .设计内的大部

分变量都量化为 16 bits,部分变量被量化为 18 bits. 脉动阵列结构的求倒结构完全分时复用,求倒结构的输入位宽为 13 bits,输出位宽为 18 bits,利用 SBTM 算法进行优化,取 $n_0=5, n_1=4, n_2=4, n_3=0$. 实现结果如下表所示:

表 1 FPGA 实现的资源消耗

资源种类	LUT	Slice	FF	DSP
消耗数量	6540	2106	6246	128

查找表大小为 13.5 kb. 由于该 FPGA 片上的 Block Memory 最小为 18 kb,而使用 SBTM 算法需要 2 个查找表,所以在 FPGA 上实现的时候需要消耗 2 个 18 kb 的 Block Memory.

本文提出的 LDLT 分解结构在系统时钟为 200 MHz 的情况下有 0.379 ns 的裕量,即系统时钟最高能够支持 216 MHz. 同时,该结构支持每个时钟周期输入一个矩阵列向量的吞吐率.

5 结束语

本文提出了一种适用于 MIMO 系统 MMSE 检测的 LDLT 分解算法的硬件实现. 该硬件实现充分利用了待分解矩阵的共轭对称性,仅需对其下三角矩阵进行操作,节省了硬件资源;同时针对不利于硬件实现的除法,提出了基于 SBTM 算法优化的求倒结构,并通过调整处理单元的 latency 使求倒结构能够完全分时复用,进一步减小了硬件资源的开销. 由于该求倒结构充分利用了待求倒数据的位宽,所以能在仅使用较小的查找表的情况下达到良好的数据精度,解决了近期文献中为了取得较好数据精度而需要在求倒结构中使用的查找表过大的问题.

参考文献:

[1] Burg A, Haene S, Perels D, et al. Algorithm and VL-SI architecture for linear MMSE detection in MIMO-OFDM systems[C]. IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings. IEEE, 2006:4.

[2] Chockalingam A, Rajan B S. Large MIMO Systems [M]. IEEE, 2013.

[3] Studer C, Fatch S, Seethaler D. ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation[J]. IEEE Journal of Solid-State Circuits, 2011, 46(7): 1754-1765.

[4] Gestner B, Anderson D V. Single Newton-Raphson iteration for integer-rounded divider for lattice reduction algorithms[C] // Circuits and Systems, 2008. MWS-CAS 2008. 51st Midwest Symposium on. IEEE, 2008: 966-969.

[5] Schulte M J, Stine J E, Wires K E. High-speed reciprocal approximations[C]// Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on. IEEE, 1997, 2: 1183-1187.

作者简介:

梁子豪 男,(1993-),硕士研究生. 研究方向为 MIMO 系统的检测与硬件实现. E-mail:lzhhdzw@163.com.

刘 辉 男,(1984-),硕士,工程师. 研究方向为卫星通信系统.

贺光辉 男,(1980-),副研究员. 研究方向为通信系统 MIMO 技术、大规模集成电路等.

(上接第 81 页)

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition[C]. In: International Conference on Learning Representations (ICLR), 2015.

[8] G. Cheng, J. Han, P. Zhou, and L. Guo. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014,98(12):119-132.

作者简介:

吴佳祥 男,(1995-),硕士研究生. 研究方向为计算机视觉. E-mail:wjxzju@foxmail.com.

刘 辉 男,(1984-),硕士,工程师. 研究方向为卫星通信系统.

贺光辉 男,(1980-),博士,副教授. 研究方向为无线通信和多媒体系统的信号处理算法优和 VLSI 架构设计.