Massive MIMO Detection Algorithm and VLSI Architecture

Leibo Liu · Guiqiang Peng ·
Shaojun Wei

Massive MIMO Detection Algorithm and VLSI Architecture

Science Press
Beijing

Springer

Leibo Liu
Institute of Microelectronics
Tsinghua University
Beijing, China

Guiqiang Peng
Institute of Microelectronics
Tsinghua University
Beijing, China

Shaojun Wei
Institute of Microelectronics
Tsinghua University
Beijing, China

ISBN 978-981-13-6361-0 ISBN 978-981-13-6362-7 (eBook)
https://doi.org/10.1007/978-981-13-6362-7

Jointly published with Science Press, Beijing, China
The print edition is not for sale in China Mainland. Customers from China Mainland please order the print book from: Science Press.
ISBN of the China Mainland edition: 978-7-03-060210-7

Library of Congress Control Number: 2019930265

Translation from the Chinese language edition: 大规模MIMO检测算法VLSI架构— —专用电路及动态重构实现 by Leibo Liu, Guiqiang Peng, Shaojun Wei, © Science Press 2019. Published by Science Press. All Rights Reserved.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

As one of the core technologies for future mobile communications, the massive MIMO technology can effectively improve the network capacity, enhance the network robustness, and reduce the communication latency. However, the complexity of baseband processing increases sharply as the number of antennas increases. Therefore, the design of high-performance massive MIMO baseband processing chips, especially the design of massive MIMO detection chips featuring low complexity and high parallelism, has become a technical bottleneck that restricts the broad application of the massive MIMO technology in communications systems.

This book first introduces the process of the team's research on efficient massive MIMO detection algorithms and circuit architectures. On the basis of the analysis on the existing massive MIMO detection algorithms, the team has optimized the algorithms from different aspects such as computation complexity and parallelism, conducted mathematical theoretical analyses and proven that the massive MIMO detection optimization algorithms proposed by the team have the advantages of low complexity and high parallelism and can fully satisfy the requirements for detection accuracy. Finally, by using the ASIC as a carrier, the team has verified that the chips based on the proposed massive MIMO detection algorithms feature high energy efficiency, high area efficiency, and low detection error.

In the process of designing the massive MIMO detection chip, we learned that the massive MIMO detection chips based on the ASIC are suitable only for application scenarios with very high requirements for the processing speed; however, some application scenarios require massive MIMO detection chips to have certain flexibility and scalability so that the massive MIMO detection chips can support different standards, algorithms, and antenna sizes and adapt to the evolution of standards and algorithms. After we conducted certain analyses, we believe that the reconfigurable computing architecture is a very promising solution. On the basis of the analyses on and common feature extraction from a large number of existing massive MIMO detection algorithms, the team has designed the data channels and configuration channels that are applicable to massive MIMO detection algorithms,

involving PEs, interconnections, storage mechanisms, context formats, and configuration methods. Thus, the team has completed the design of a massive MIMO detection reconfigurable processor.

The massive MIMO detection reconfigurable processor may also be applicable to future wireless communications systems such as Beyond 5G. There are three main reasons: First, wireless communication algorithms are now developed in the repeated iteration and optimization processes. In the process of solving the limitation problem of commercial algorithms, the update of an algorithm, no matter whether it is an optimized algorithm or a newly designed algorithm, has a strong logical continuation relationship, which provides an internal logical basis for design of the reconfigurable processor architecture. Second, the design for PEs and PEAs of the massive MIMO detection reconfigurable processors fully considers the requirements for flexibility and scalability so that the PEs and PEAs can meet the hardware requirements and foreseeable future needs of various algorithms at present. Third, the design methodology is applicable to all the massive MIMO detection reconfigurable processors. Therefore, the hardware implementation requirements for future algorithms can be met. Hence, after corresponding algorithm analyses are conducted, the optimization and design of the reconfigurable processor architecture based on the design methodology will become a universal process.

This book consists of seven chapters. Chapter 1 introduces the development trend of wireless communication technologies including the development and research status of the massive MIMO technology and the MIMO detection technology, analyzes the advantages and disadvantages of the MIMO detection chip based on the ASIC and instruction-level architecture processor in aspects such as performance, power consumption and flexibility, proposes the dynamic reconfigurable chip technology for MIMO detection, and analyzes the feasibility for implementing the proposed technology. Chapters 2 and 3 introduce the linear massive MIMO detection algorithm and the corresponding circuit architecture, respectively, and analyze the advantages of the linear detection optimization algorithm proposed by the team from different aspects such as algorithm convergence, computation complexity, and detection performance. The experimental results have shown that the circuit designed on the basis of the algorithm proposed by the team has higher energy efficiency and area efficiency, and thus verified that the optimization algorithm proposed by the team is more suitable for hardware implementation. Chapters 4 and 5 introduce the nonlinear massive MIMO detection algorithm with high detection accuracy and the corresponding circuit architecture, respectively, and compare the nonlinear massive MIMO detection algorithm proposed by the team with other algorithms from different aspects such as algorithm convergence, computation complexity, detection performance, and experimental results. The results have shown that the complexity of the algorithm proposed by the team is within the acceptable range while the algorithm implements high detection accuracy. Chapter 6 provides detailed information on the dynamic reconfigurable chip for massive MIMO detection. First, the chapter uses the reconfigurable computing architecture as the target hardware platform to analyze mainstream massive MIMO detection algorithms at present, including common logic extraction from algorithms, feature extraction of data types and parallelism analysis

on algorithms. Then, the chapter provides a detailed analysis on the hardware architecture design of the dynamic reconfigurable chip for massive MIMO detection from different aspects of data channels and configuration channels, and introduces the design method for the hardware architecture specific to the massive MIMO detection algorithm. Chapter 7 provides an outlook on application of the VLSI architecture for massive MIMO detection on the server, mobile terminal and edge computing sides.

This book embodies the nearly 6-year collective wisdom of the wireless communication baseband processor team from the Institute of Microelectronics of Tsinghua University. Thanks to the classmates and colleagues of the team members including Peng Guiqiang, Wang Junjun, Zhang Peng, Wei Qiushi, Tan Yingran, Yang Haichang, Wang Pan, Wu Yibo, Zhu Yihong, Xue Yang, Li Zhaoshi, Yang Xiao, Ding Ziyu, and Wang Hanning for their participation. Thanks to our engineers Wang Yao, Ying Yijie, Kong Jia, Chen Yingjie, Wang Guangbin, Wang Lei, Li Zhengdong, Luo Senpin, Jin Yu, et al. for their participation. Thanks to Prof. Wei Shaojun for his support for and guidance to the preparation of this book. Thanks to Editor Zhao Yanchun from Science Press for her suggestions on the publication of this book. Finally, I give thanks to my wife and children for their understanding and tolerance of my work. Without their support, it is hard to imagine how I could finish this work. They are also an important impetus for my future efforts and progress.

Beijing, China Leibo Liu
August 2018

Contents

1 Introduction 1
1.1 Application Requirements 2
1.1.1 Typical Applications in Future 2
1.1.2 Communication System Requirements 6
1.2 Mobile Communication and MIMO Detection 9
1.2.1 Development of Communication Technologies 9
1.2.2 Key 5G Technologies 11
1.2.3 MIMI Baseband Processing 17
1.2.4 Difficulties in Massive MIMO Detection 23
1.3 Status Quo of MIMO Detection Chip Research 24
1.3.1 ISAP-Based MIMO Detection Chip 24
1.3.2 ASIC-Based MIMO Detection Chip 31
1.3.3 Limitation of Traditional MIMO Detection Chips 45
1.4 Dynamic Reconfigurable Chip Technologies of MIMO Detection 47
1.4.1 Overview of Reconfigurable Computing 47
1.4.2 Status Quo of Dynamic Reconfiguration Chip of MIMO Detection 57
References 66

2 Linear Massive MIMO Detection Algorithm 71
2.1 Analysis of Linear Detection Algorithm 71
2.2 Neumann Series Approximation Algorithm 74
2.2.1 Algorithm Design 75
2.2.2 Error Analysis 76
2.2.3 Complexity and Block Error Rate 80
2.3 Chebyshev Iteration Algorithm 83
2.3.1 Algorithm Design 83
2.3.2 Convergence 89
2.3.3 Complexity and Parallelism 93

2.3.4 Bit Error Rate . . . 95
2.3.5 Analysis on Channel Model Impact . . . 96
2.4 Jacobi Iteration Algorithm . . . 100
2.4.1 Weighted Jacobi Iteration and Convergence . . . 100
2.4.2 Complexity and Frame Error Rate . . . 104
2.4.3 Analyses on Channel Model Effects . . . 107
2.5 Conjugate Gradient Algorithm . . . 110
2.5.1 Algorithm Design . . . 110
2.5.2 Convergence . . . 113
2.5.3 Initial Iteration Value and Search . . . 114
2.5.4 Complexity and Parallelism . . . 118
2.5.5 Symbol Error Rate . . . 120
References . . . 122

3 Architecture of Linear Massive MIMO Detection . . . 125
3.1 NSA-Based Hardware Architecture . . . 125
3.1.1 VLSI Top-Level Structure . . . 125
3.1.2 Approximate Inversion and Matched Filtering Module . . . 126
3.1.3 Equalization and SINR Module . . . 128
3.1.4 IFFT and LLR Module . . . 128
3.1.5 Inverse Module Based on Cholesky Decomposition . . . 129
3.2 Chebyshev Iteration Hardware Architecture . . . 130
3.2.1 VLSI Top-Level Structure . . . 131
3.2.2 Initial Module . . . 131
3.2.3 Iterative Module . . . 133
3.2.4 LLR Module . . . 134
3.2.5 Experimental Results and Comparison . . . 135
3.3 Hardware Architecture Based on Weighted Jacobi Iteration . . . 138
3.3.1 VLSI Top-Level Architecture . . . 138
3.3.2 Diagonal Systolic Array . . . 140
3.3.3 WeJi Module . . . 142
3.3.4 LLR Module . . . 145
3.3.5 Experimental Result and Comparison . . . 145
3.4 Hardware Architecture Based on Conjugate Gradient Method . . . 153
3.4.1 VLSI Top-Level Structure . . . 153
3.4.2 Input/Output Module . . . 154
3.4.3 Multiplication Module . . . 155
3.4.4 Iterative Module . . . 157
3.4.5 Experimental Results and Comparison . . . 158
References . . . 161

4 Nonlinear Massive MIMO Signal Detection Algorithm 165
4.1 Conventional Nonlinear MIMO Signal Detection Algorithm 166
4.1.1 ML Signal Detection Algorithm 166
4.1.2 SD Signal Detection Algorithm and *K*-Best Signal Detection Algorithm 168
4.2 CHOSLAR Algorithm 172
4.2.1 System Model 172
4.2.2 QR Decomposition 173
4.2.3 Lattice Reduction 175
4.2.4 Cholesky Preprocessing 176
4.2.5 Improved *K*-Best Detector and Its Performance Simulation 184
4.2.6 Summary and Analysis 190
4.3 TASER Algorithm 192
4.3.1 System Model 192
4.3.2 Semi-definite Relaxation 194
4.3.3 Algorithm Analysis 195
4.3.4 Performance Analysis 198
4.3.5 Computational Complexity 199
References 201

5 Architecture for Nonlinear Massive MIMO Detection 205
5.1 CHOSLAR Hardware Architecture 206
5.1.1 VLSI Architecture 206
5.1.2 Implementation Results and Comparison 213
5.2 TASER-Based Hardware Architecture 217
5.2.1 Architecture Overview 217
5.2.2 PE 219
5.2.3 Implementation Details 220
5.2.4 FPGA Implementation Result 221
5.2.5 ASIC Implementation Results 224
References 227

6 Dynamic Reconfigurable Chips for Massive MIMO Detection 229
6.1 Algorithm Analysis 229
6.1.1 Algorithm Analysis Method 230
6.1.2 Common Features of Algorithms 231
6.1.3 Computing Model 232
6.2 Data Path 233
6.2.1 Structure of Reconfigurable PEA 235
6.2.2 PE Structure 238
6.2.3 Shared Memory 245
6.2.4 Interconnection 247

6.3 Configuration Path 286
6.3.1 Control Design 287
6.3.2 Master Control Interface 288
6.3.3 Configuration Controller 290
6.3.4 Design of Configuration Package 292
6.3.5 Mapping Method 297
References 304

7 Prospect of the VLSI Architecture for Massive MIMO Detection 307
7.1 Prospect of Server-Side Applications 307
7.1.1 Outline of 5G Communications Characteristics 307
7.1.2 Outline of the Server-Side Characteristics 309
7.1.3 Server-Side Application 310
7.2 Prospect of Mobile-Side Application 317
7.2.1 Application of ASIC-Based Detection Chips 319
7.2.2 Application of Reconfigurable Detection Chips 323
7.3 Prospect of Applications of Edge Computing 326
7.3.1 Concept of Edge Computing 327
7.3.2 Application of Detection Chips in the Edge Computing 330
References 334

Abbreviations

ACC	Accumulator
ADPLL	All digital phase-locked loop
AHB	Advanced high-performance bus
ALU	Arithmetic logical unit
AMBA	Advanced microcontroller bus architecture
AR	Augmented reality
ARM	Advanced RISC machine
ASIC	Application-specific integrated circuit
ASIP	Application-specific instruction set processor
AU	Arithmetic unit
BB	Branch and bound
BLER	Block error rate
BPSK	Binary phase-shift keying
BTS	Base transceiver station
CBU	Column-broadcast unit
CC	Convolutional coding
CDMA	Code-division multiple access
CG algorithm	Conjugate gradient algorithm
CGLS	Conjugate gradient least square
CGRA	Coarse-grained reconfigurable array
CHEST	Channel estimation
CHOSLAR	Cholesky sorted QR decomposition and partial iterative lattice reduction
CM	Complex multiplication
CORDIC	Coordinate rotation digital computer
CoREP	Common Reports
CP	Cyclic prefix
CPA	Control program assist
CPLD	Complex programmable logic device
CPU	Central processing unit

CSG	Closed subscriber group
CSI	Channel state information
CSIR	Receiver channel state information
DDR	Double data rate
DMA	Direct memory access
DSP	Digital signal processor
DVFS	Dynamic voltage and frequency scaling
ELP	Energy latency product
EMI	Electromagnetic interference
EPD	Expectation propagation detection
FBMC	Filter bank based multicarrier modulation
FBS	Forward–backward splitting
FDD	Frequency-division duplexing
FEC	Forward error correction
FER	Frame error rate
FFT	Fast Fourier transform
FIR	Finite impulse response
FPGA	Field-programmable gate array
FSM	Finite-state machine
GI	Guard interval
GPP	General purpose processor
GPU	Graphics processing unit
GR	Givens rotation
GSM	Global System for Mobile communication
HART	Highway addressable remote transducer
HDL	Hardware description language
HEVC	High Efficiency Video Codec
HMD	Head-mounted display
HT	Householder transformation
i.i.d.	Independent identically distributed
I/O	Input/output
IaaS	Infrastructure as a Service
IASP	Instruction set architecture processor
IFFT	Inverse fast Fourier transform
IIC	Intra-iterative interference cancellation
IoT	Internet of things
ISI	Intersymbol interference
ISP	Internet Service Provider
JED	Joint channel estimation and data detection
JTAG	Joint Test Action Group
LBC	Lower bound of cost
LDPC	Low-density parity-check code
LLC	Last level cache
LLR	Log likelihood ratio
LPF	Low-pass filter

LR	Lattice reduction
LTE	Long-term evolution
LUD	LU decomposition
LUT	Lookup table
M2M	Machine to machine
MAC	Multiply and accumulate
MDA	Multimode detection architecture
MEC	Mobile edging computing
MF	Matched filtering
MIMO	Massive multiple-input multiple-output
ML algorithm	Machine language algorithm
MMSE	Minimum mean square error
MMSE-SIC	Minimum mean square error-successive interference cancelation
MPD	Message passing detector
MWD	Multi-window display
NI	Network interface
NoC	Network on chip
NP problem	Nondeterministic polynomial problem
NSA algorithm	Neumann series approximation algorithm
NTL	Network topology link
OCD	Optimized coordinate descent
OFDM	Orthogonal frequency-division multiplexing
OFDMA	Orthogonal frequency-division multiple access
opcode	Operation code
OSG	Open subscriber group
PaaS	Platform as a Service
PAR	Peak-to-average ratio
PARSEC	Princeton Application Repository for Shared-Memory Computers
PCBB	Priority and compensation factor oriented branch and bound
PDA	Probabilistic data association
PE	Processing element
PEA	Processing element array
PILR	Partial iterative lattice reduction
PIP	Picture in picture
PLL	Phase-locked loop
PSD	Positive semidefinite
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase-shift keying
RADD	Real-valued addition
RAM	Random-access memory
RBU	Row-broadcast unit
RC	Reliability benefits
RCM	Reliability cost model
REM	Reliability efficiency model

RISC	Reduced instruction set computer
RMUL	Real-valued multiplication
RSN	Resource node
RTL	Resistor transistor logic
SA	Simulated annealing
SaaS	Software as a Service
SC-FDMA	Single-carrier frequency-division multiple access
SD	Sphere decoding
SD algorithm	Standard deviation algorithm
SDP	Semidefinite program
SDR	Software-defined radio
SER	Symbol error rate
SIMD	Single instruction, multiple data
SINR	Signal-to-interference-plus-noise-ratio
SNR	Signal-to-noise ratio
SoC	System on chip
SRAM	Static random-access memory
TASER	Triangular approximate semidefinite relaxation
TDD	Time-division duplexing
TDMA	Time-division multiple access
TGFF	Task graph for free
TSMC	Taiwan Semiconductor Manufacturing Company
UBC	Upper bound of cost
UMTS	Universal Mobile Telecommunications System
VLIW	Very long instruction word
VLSI	Very-large-scale integration
VOPD	Video object plane decoder
VR	Virtual reality
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless local area network
ZF	Zero frequency
ZF-DF	Zero-forcing decision feedback