

# Energy Efficient Embedded Video Processing Systems

Muhammad Usman Karim Khan  
Muhammad Shafique • Jörg Henkel

# Energy Efficient Embedded Video Processing Systems

A Hardware-Software Collaborative  
Approach

 Springer

Muhammad Usman Karim Khan  
IBM Deutschland Research &  
Development GmbH  
Böblingen, Germany

Muhammad Shafique  
Institute of Computer Engineering  
Vienna University of Technology  
Vienna, Austria

Jörg Henkel  
Department of Computer Science  
Karlsruhe Institute of Technology  
Karlsruhe, Germany

ISBN 978-3-319-61454-0      ISBN 978-3-319-61455-7 (eBook)  
DOI 10.1007/978-3-319-61455-7

Library of Congress Control Number: 2017944437

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Multimedia Systems	2
1.1.1	Multimedia Processing Architectures	4
1.2	Fundamentals of Video Processing	5
1.2.1	Video Compression	6
1.3	Design Complexity of a Video System	8
1.3.1	The Dark Silicon Problem	9
1.3.2	SRAM Aging	10
1.4	Video System Design Challenges	11
1.4.1	Software Layer Challenges	14
1.4.2	Hardware Layer Challenges	14
1.5	Limitations of State-of-the-Art	14
1.6	Design and Optimization Methods Discussed in This Book	15
1.6.1	Software Layer Design	17
1.6.2	Hardware Layer Design	18
1.6.3	Open-Source Tools	20
1.7	Book Outline	21
	References	21
<b>2</b>	<b>Background and Related Work</b>	25
2.1	Overview of Video Processing	25
2.1.1	Intra- and Inter-frame Processing	28
2.2	Overview of Video Coding	29
2.2.1	H.264/AVC and HEVC	32
2.2.2	Parallelization	36
2.2.3	DVC and HDVC	38
2.3	Technological Challenges	40
2.3.1	Dark Silicon or Power Wall	41
2.3.2	NBTI-Induced SRAM Aging	42
2.3.3	Other Challenges	44

2.4	Related Work . . . . .	45
2.4.1	Video System Software . . . . .	45
2.4.2	Video Systems Hardware . . . . .	50
2.5	Summary of Related Work . . . . .	57
	References . . . . .	57
<b>3</b>	<b>Power-Efficient Video System Design . . . . .</b>	<b>67</b>
3.1	System Overview . . . . .	67
3.1.1	Design Time Feature Support . . . . .	69
3.1.2	Runtime Features and System Dynamics . . . . .	71
3.2	Application and Motivational Analysis . . . . .	72
3.2.1	Video Application Parallelization . . . . .	73
3.2.2	Workload Variations . . . . .	75
3.2.3	HEVC Complexity Analysis . . . . .	76
3.3	Hardware Platform Analysis . . . . .	79
3.3.1	Heterogeneity Among Computing Nodes . . . . .	79
3.3.2	Memory Subsystem . . . . .	80
3.3.3	Analysis of Different Aging Balancing Circuits . . . . .	83
3.4	Summary . . . . .	86
	References . . . . .	87
<b>4</b>	<b>Energy-Efficient Software Design for Video Systems . . . . .</b>	<b>89</b>
4.1	Power-Efficient Application Parallelization . . . . .	89
4.1.1	Power-Efficient Workload Balancing . . . . .	91
4.2	Compute Configuration . . . . .	94
4.2.1	Uniform Tiling . . . . .	94
4.2.2	Non-uniform Tiling . . . . .	96
4.2.3	Frequency Estimation ( $f_{k,m}$ ) . . . . .	99
4.2.4	Maximum Workload Estimation ( $\alpha_{k,m}$ ) . . . . .	100
4.2.5	Self-Regulated Frequency Model . . . . .	100
4.2.6	Retiling . . . . .	104
4.3	Application Configuration . . . . .	104
4.3.1	HEVC Application Configurations . . . . .	104
4.3.2	HEVC Configuration Tuning . . . . .	106
4.3.3	HEVC Parameter Mapping . . . . .	107
4.4	Workload Balancing on Heterogeneous Systems . . . . .	112
4.4.1	System Model . . . . .	112
4.4.2	Load Balancing Algorithm . . . . .	113
4.5	Resource Budgeting for Mixed Multithreaded Workloads . . . . .	116
4.5.1	Hierarchical Resource Budgeting . . . . .	118
4.5.2	Intra-Cluster Power Budgeting $p_{i,j}$ . . . . .	120
4.5.3	Inter-Cluster Power Budgeting $p_i$ . . . . .	121
4.5.4	Selection of Cluster Size . . . . .	123
	References . . . . .	124

- 5 Energy-Efficient Hardware Design for Video Systems . . . . .** 127
  - 5.1 Custom Video Processing Architectures . . . . . 127
    - 5.1.1 Memory Analysis and Video Input . . . . . 128
    - 5.1.2 Video Preprocessing . . . . . 129
    - 5.1.3 DDR Video Write Master . . . . . 129
    - 5.1.4 Heterogeneous Computing Platform . . . . . 131
  - 5.2 Accelerator Allocation and Scheduling . . . . . 132
    - 5.2.1 Accelerator Sharing on Multi-/Many-Core . . . . . 132
    - 5.2.2 Multicast Video Processing Hardware . . . . . 139
  - 5.3 Efficient Hardware Accelerator Architectures . . . . . 142
    - 5.3.1 Low Latency H.264/AVC Encoding Loop . . . . . 142
    - 5.3.2 Distributed Hardware Accelerator Architecture . . . . . 153
  - 5.4 Hybrid Video Memory Architectures . . . . . 157
    - 5.4.1 AMBER Memory Hierarchy . . . . . 158
    - 5.4.2 NVM Reference Memory Architecture . . . . . 159
    - 5.4.3 Energy Management of NVM Reference Memories . . . . . 161
    - 5.4.4 System Computation Flow . . . . . 163
  - 5.5 Energy-Efficient Anti-aging Circuits for SRAM . . . . . 164
    - 5.5.1 Memory Write Transducer (MWT) . . . . . 166
    - 5.5.2 Aging-Aware Address Generation Unit (AGU) . . . . . 167
    - 5.5.3 Aging Controller . . . . . 168
    - 5.5.4 Generalization and Applicability . . . . . 170
    - 5.5.5 Sensitivity Analysis of SRAM Anti-aging Circuits . . . . . 171
  - References . . . . . 173
- 6 Experimental Evaluations and Discussion . . . . .** 175
  - 6.1 Parallelization and Workload Balancing . . . . . 175
    - 6.1.1 Software Architecture and Simulation Setup . . . . . 176
    - 6.1.2 Compute and Application Configuration  
for Uniform Tiling . . . . . 177
    - 6.1.3 Compute Configuration with Non-uniform Tiling . . . . . 182
    - 6.1.4 Workload Balancing on Heterogeneous Platforms . . . . . 184
  - 6.2 Resource Budgeting . . . . . 185
    - 6.2.1 Experimental Setup . . . . . 186
    - 6.2.2 Results and Discussion . . . . . 187
  - 6.3 Memory Subsystem . . . . . 188
    - 6.3.1 AMBER: Hybrid Memories . . . . . 189
    - 6.3.2 SRAM Anti-aging Circuits . . . . . 191
  - References . . . . . 199
- 7 Conclusion and Future Outlook . . . . .** 201
  - 7.1 Software-Level Techniques . . . . . 201
  - 7.2 Hardware-Level Techniques . . . . . 203

- 7.3 Further Improvements . . . . . 205
  - 7.3.1 Approximate Computing . . . . . 205
  - 7.3.2 GPU-Based Acceleration . . . . . 206
  - 7.3.3 Reliability and Workload Management . . . . . 206
  - 7.3.4 Generalization . . . . . 207
- References . . . . . 207
- Appendices . . . . . 211**
- Index . . . . . 233**