New Wireless Technologies for Next-Generation Internet-of-Things

A Dissertation Presented

by

Nan Cen

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Electrical Engineering

Northeastern University
Boston, Massachusetts

September 2019
To my family.
Contents

List of Figures v
List of Tables vii
List of Acronyms viii
Acknowledgments xii
Abstract of the Dissertation xiii

1 Introduction 1
  1.1 Dissertation Outline ................................................................. 2

2 Inter-view Motion Compensated Joint Decoding for Compressively-Sampled Multi-
  View Video Streams 4
  2.1 Related Work ................................................................. 6
  2.2 Preliminaries ................................................................. 8
    2.2.1 CS Acquisition ................................................................. 8
    2.2.2 CS Recovery ................................................................. 8
  2.3 System Architecture ............................................................ 9
  2.4 Joint Multi-view Decoding ....................................................... 11
    2.4.1 K-view Decoding ................................................................. 11
    2.4.2 Inter-view Motion Compensated Side Frame .................................. 12
    2.4.3 Fusion Decoding Algorithm .................................................... 14
    2.4.4 Blind Video Quality Estimation .............................................. 14
  2.5 Performance Evaluation ......................................................... 18
  2.6 Summary ........................................................................... 21

3 Low-Power Multimedia Internet of Things through Compressed Sensing based Multi-
  view Video Streaming 23
  3.1 Related Works .................................................................. 26
  3.2 Preliminaries .................................................................. 27
    3.2.1 Compressed Sensing Basics .................................................. 27
    3.2.2 Rate-Distortion Model for Compressive Imaging ........................ 29
3.3 CS-based multi-view Coding
   Architecture Design .................................................. 29
   3.3.1 Cooperative Block-level Rate-adaptive Encoder ................. 31
   3.3.2 Independent Decoder ............................................. 33
   3.3.3 Centralized Controller .......................................... 33
3.4 End-to-End Rate-Distortion Model .................................. 35
3.5 Network Modeling Framework ........................................ 37
3.6 Performance evaluation .............................................. 44
   3.6.1 Evaluation of CS-based Multi-view Encoding/Decoding Architecture .... 44
   3.6.2 Evaluation of Power-efficient Compressive Video Streaming .......... 50
3.7 Summary ............................................................. 51

4 LiBeam: Throughput-Optimal Cooperative Beamforming for Indoor Visible Light Networks 53
   4.1 Related Work ....................................................... 56
   4.2 System Model and Problem Formulation ............................ 57
   4.3 Globally Optimal Solution Algorithm .............................. 62
      4.3.1 Overview of The Algorithm .................................. 62
      4.3.2 Convex Relaxation ............................................. 63
      4.3.3 Variable Partition ............................................. 65
   4.4 Testbed Development ................................................ 65
   4.5 Performance Evaluation ............................................ 68
      4.5.1 Simulation Results ............................................. 68
      4.5.2 Experimental Evaluation .................................... 69
   4.6 Summary ............................................................ 72

5 LANET: Visible-Light Ad Hoc Networks 75
   5.1 LANET: Visible-Light Ad Hoc Networks .......................... 77
      5.1.1 Main Design Challenges ...................................... 78
      5.1.2 LANETs vs Traditional MANETs ............................... 80
   5.2 Envisioned Applications ............................................ 82
   5.3 LANET Node Architecture .......................................... 84
      5.3.1 Node Architecture ............................................. 84
      5.3.2 Front-end Hardware ........................................... 86
      5.3.3 Existing VLC Testbeds ........................................ 87
   5.4 Physical Layer ...................................................... 89
      5.4.1 Existing Modulation Schemes .................................. 90
      5.4.2 Open Research Issues ......................................... 93
   5.5 Medium Access Control Layer (MAC) .............................. 94
      5.5.1 Existing Visible Light MACs .................................. 94
      5.5.2 MAC for LANETs ............................................... 97
      5.5.3 Standardization: MAC of IEEE 802.15.7 ....................... 99
      5.5.4 Open Research Issues ......................................... 100
   5.6 Network Layer ...................................................... 101
      5.6.1 Open Research Problems ..................................... 102
List of Figures

2.1 Muti-view encoding/decoding architecture. ........................................ 10
2.2 Block diagram of side frame generation. .......................................... 11
2.3 (a) original, (b) independently reconstructed, (c) generated side frame, and (d) fusion decoded 5th frame of Exit; Measurement rate is set to 0.2. ............... 15
2.4 (a) original, (b) independently reconstructed, (c) generated side frame, and (d) fusion decoded 25th frame of Vassar; Measurement rate is set to 0.15. .... 16
2.5 PSNR comparison for CS-views (a) view 1, (b) view 3, and (c) view 4, and SSIM comparison for CS-views (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.3 of Vassar. ......................................................... 17
2.6 PSNR comparison for CS-views (a) view 1, (b) view 3, and (c) view 4, and SSIM comparison for CS-views (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.1 of Exit. ................................................................. 18
2.7 PSNR comparison for CS-views (a) view 1, (b) view 3, and (c) view 4, and SSIM comparison for CS-views (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.2 of Ballroom. ......................................................... 19
2.8 Video quality estimation results for different video sequences: (top) Ballroom, (bottom) Exit. ........................................................................................................ 21
3.1 Encoding/decoding architecture for multi-hop CS-based multi-view video streaming. .......................................................................................... 30
3.2 Block Sparsity: (a) Original image, (b) Block-based DCT coefficients of (a). 30
3.3 Comparison of (a) PSNR, (b) the number of transmitted bits, and (c) the compression rate between approaches with and without mean subtraction. .......... 31
3.4 Rate-Distortion curve fitting for Vassar view 2 sequence. ................. 36
3.5 PSNR against frame index for (a) view 1, (b) view 2 (R-view), (c) view 3, and (d) view 4 of sequence Vassar. ............................................................... 42
3.6 PSNR against frame index for (a) view 1, (b) view 2 (R-view), (c) view 3, and (d) view 4 of sequence Exit. ................................................................. 43
3.7 Rate-distortion comparison for frame 75 of Vassar sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4. ..................................................... 45
3.8 Rate-distortion comparison for frame 9 of Exit sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4. ..................................................... 46
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Average Pearson correlation coefficient for <em>Vassar</em> five views</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>Improved average PSNR (dB) when selecting different <em>Vassar</em> views as R-view</td>
<td>35</td>
</tr>
<tr>
<td>3.3</td>
<td>PSNR and SSIM comparison for <em>Vassar</em> eight views</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Network Scenario 1</td>
<td>69</td>
</tr>
<tr>
<td>4.2</td>
<td>Network Scenario 2</td>
<td>70</td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison between LANETs and MANETs</td>
<td>79</td>
</tr>
<tr>
<td>5.2</td>
<td>Representative existing VLC testbeds</td>
<td>88</td>
</tr>
<tr>
<td>5.3</td>
<td>Visible Light Modulation Schemes</td>
<td>89</td>
</tr>
<tr>
<td>5.4</td>
<td>Summary of MAC protocols for VLC</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Acronyms

ACK  Acknowledgment
ACN  Availability Confirmation
AP   Access Point
ARQ  Automatic Reply Request
ART  Availability Request
BD   Block Diagonalization
BE   Backoff Exponent
BER  Bit Error Rate
BI   Beacon Interval
CAP  Contention Access Period
CC   Convolutional Coding
CCA  Clear Channel Assessment
CDMA Code Division Multiple Access
CFP  Contention Free Period
CSMA/CA Carrier Sense Multiple Access/Collision Avoidance
CSMA/CD Carrier Sense Multiple Access/Collision Detection
CSK  Color-Shift Keying
CTS  Clear-to-send
DMT  Discrete Multi-Tones
DCO-OFDM Direct-Current Optical Orthogonal Frequency Division Multiplexing (OFDM)
D2D  Device to Device
DoA  Direction of Arrival
DoD  Department Of Defence
ECC  Error-Correction Code
EMI  Electromagnetic Interference
FEC  Forward Error Correction
FSO  Free Space Optics
FOV  Field Of View
GPS  Global Positioning System
GTS  Guaranteed Time Slots
IM/DD  Intensity-Modulation Direct-Detection
IoT  Internet of Things
IR  Infra-Red
ISM  Industrial, Scientific and Medical
ISR  Intelligence, Surveillance, and Reconnaissance
LANET  Visible-Light Tactical Ad-Hoc Networking
LED  Light Emitting Diode
LOS  Line of Sight
LPI/LPD  Lower Probability of Intercept/Lower Probability of Detection
MAC  Medium Access Control
MA-DMT  Multiple Access Discrete Multi-Tones
MANET  Mobile Ad Hoc Network
MUI  Multi-User Interference
MU-MIMO  Multi-User Multiple-Input Multiple-Output
MU-MISO  Multi-User Multiple-Input Single-Output
NAV  Network Allocation Vector
NB  Number of Backoffs
MC-CDMA  Multi-carrier Code Division Multiple Access (CDMA)
NLOS  non-Line Of Sight
NRL  Naval Research Labs
OC  Optical Carrier
OCDMA  Optical Code-Division Multiple Access
OFDM  Orthogonal Frequency Division Multiplexing
OFDMA  Orthogonal Frequency Division Multiple Access
OOC  Optical Orthogonal Codes
O-OFDMA  Optical Orthogonal Frequency Division Multiple Access
O-OFDM-IDMA  Optical Orthogonal Frequency Division Multiplexing Interleave Division Multiple Access
OOK  On-Off Keying
OWC  Optical Wireless Communication
OWMAC  Optical wireless MAC
PD  Photon Detector
PHR  PHY Header
PHY  Physical
PRO-OFDM  Polarity Reversed Optical OFDM
PSDU  PHY Service Data Unit
QAM  Quadrature Amplitude Modulation
QoS  Quality of Service
RA  Random Access
RES  Reserve Sectors
RF  Radio Frequency
RLL  Run Length Limited
ROC  Random Optical Codes
RS  Reed-Solomon
SACW  Self-Adaptive minimum Contention Window
SD  Superframe Duration
SNR  Signal-to-Noise Ratio
SWaP  (Size, Weight, and Power)
TDD  Time Division Duplex
TDMA  Time Division Multiple Access
THP  Tomlinson-Harashima Precoding
VLC  Visible Light Communication
VPPM  Variable Pulse Position Modulation
VPAN  Visible-light communication Personal Area Network
USRP  Universal Software Radio Peripheral
UV  Ultraviolet
UVC  Ultraviolet Communication
V2I  Vehicle to Infrastructure
V2V  Vehicle to Vehicle
WiFi  Wireless Fidelity
WSN  Wireless Sensor Network
ZF  Zero Forcing
4B6B  4-bit to 6-bit encoded symbols
8B10B  8-bit to 10-bit encoded symbols
First and foremost, I would like to extend my most sincere gratitude to my advisor, Professor Tommaso Melodia, for his support, guidance, patience, and encouragement through the years of my Ph.D. studies. He enlightened me what it means to be a true researcher, and taught me many important lessons. He supported me in every aspect in my Ph.D. years. I learned a lot from his ways of thinking and philosophy of life. He has been a true mentor to me. The experience with him has profoundly influenced, and will continue to guide me in the years to come.

I would like to thank my committee members, Professor Kaushik Roy Chowdhury, Professor Stefano Basagni and Professor Yunsi Fei. Thanks for their valuable time, interest and help for my research and job-hunting. They always provided me insightful questions and comments to my dissertation.

I would like to thank all my colleagues in Wireless Networks and Embedded Systems (WiNES) Lab. Special thanks to my collaborators: Professor Zhangyu Guan, Emrecan Demirors, Neil Dave. The WiNESers are all my special friends during these years.

Last but not least, I would like to thank my family for all their continuous support and encouragement during my Ph.D. study. This dissertation would not have been possible without their love!
Abstract of the Dissertation

New Wireless Technologies for Next-Generation Internet-of-Things

by

Nan Cen

Doctor of Philosophy in Electrical and Computer Engineering
Northeastern University, September 2019
Dr. Tommaso Melodia, Advisor

The explosion of the Internet of Things (IoTs) will result in billions of heterogeneous, low-power and low-complexity devices, and will enable diverse sets of applications, ranging from pervasive surveillance systems, health-care, smart cities, precision agriculture, industrial automation as well as military, and expanding over air, space, water, underground as well as in the human body. Along with the pervasive expansion and innovation of the IoT, researchers are faced with a plethora of technical challenges, including: (i) Low-power low-complexity algorithms are required for capability- and resource-limited IoT devices, where processing large amounts of sensed data is impossible, especially for multimedia data. (ii) Scaling out zillions of mobile devices, machines and objects in IoT in a few available bands in legacy radio spectrum will inevitably lead to the dreaded spectrum crunch problem.

Towards addressing these challenges, we first propose a new paradigm for multi-view encoding and decoding based on Compressed Sensing (CS), which reduces the computational complexity for resource-limited IoT devices, where processing large amounts of sensed data is impossible, especially for multimedia data. (ii) Scaling out zillions of mobile devices, machines and objects in IoT in a few available bands in legacy radio spectrum will inevitably lead to the dreaded spectrum crunch problem.

Towards addressing these challenges, we first propose a new paradigm for multi-view encoding and decoding based on Compressed Sensing (CS), which reduces the computational complexity for resource-limited IoT devices, where processing large amounts of sensed data is impossible, especially for multimedia data. Based on the proposed CS encoding/decoding architecture, a power-minimizing delivery algorithm in multi-path multi-hop networks is further proposed to reduce the power consumption, thus prolonging the lifetime of “things” in IoT.

We then investigate on a clean-slate wireless communication technology, visible-light networking, to alleviate the spectrum crunch crisis problem. We first propose LiBeam, throughput-optimal cooperative beamforming for indoor infrastructure visible light networks, with the objective to provide throughput-optimal WiFi-like downlink access to users in indoor visible light networks through a set of centrally-controlled and partially interfering light emitting diodes (LEDs). We then propose a new visible-light ad hoc networking (LANET) paradigm, based on which a software-defined LANET testbed is developed with resilience and reconfigurability, with the potential to enable cutting-edge applications (e.g., military, intelligent transportation systems.)
Chapter 1

Introduction

The Internet of Things (IoTs) envision a world-wide, interconnected network of smart physical entities, which will greatly impact and benefit our lives. In the next few years, cars, kitchen appliances, televisions, smartphones, utility meters, intra-body sensors, thermostats, and almost anything we can imagine will be accessible from anywhere on the planet [1]. The Revolution brought by the IoT will be similar to the building of roads and railroads which powered the Industrial Revolution of the 18th to 19th centuries [2] - and is expected to radically transform the education, health-care, smart home, manufacturing, mining, commerce, transportation, and surveillance fields, just to mention a few [3].

As IoT penetrates in every aspect of our lives, the demand for wireless resources will accordingly increase in an unprecedented way. Sensors are everywhere and the trend will only continue. As the number of connected devices swells beyond an expected 30 billion by 2020, which will generate a global network of "things" of dimensions never seen before. As a result, a huge amount of sensed data are pouring into limited bandwidth internet, which will certainly bring a plethora of challenges in front of researchers.

- **Low-power, Low-complexity.** IoT devices are usually capability- and resource-limited in terms of CPU, memory and power, which makes it impossible to process large amounts of sensing data, especially for multimedia data.

- **Spectrum Crunch Crisis.** As only a few bands in the legacy radio spectrum are available to the wireless carriers, scaling out zillions of mobile devices, machines and objects in IoTs will inevitably lead to the dreaded spectrum crunch problem.
CHAPTER 1. INTRODUCTION

To address these challenges, algorithms and communication schemes must be redesigned to dynamically accommodate for the fast-paced requirements of next-generation IoT devices. The objective of my research is to design low-power low-complexity algorithms for IoT devices and to investigate new spectrum technologies (e.g., based on visible light communications (VLC)) to alleviate the spectrum crunch crisis. So far, my research has focused on modeling, optimization and control of sensor and ad hoc networks, with applications to wireless multimedia networks, visible light ad hoc networks, and drone ad hoc networks. Currently, I am working on designing and developing software-defined infrastructure-less visible-light ad hoc networks.

1.1 Dissertation Outline

In Chapter 2, we design a novel multi-view video encoding/decoding architecture for wirelessly multi-view video streaming applications, e.g., 360 degrees video, Internet of Thing (IoT) multimedia sensing, among others, based on distributed video coding (DVC) and compressed sensing (CS) principles. Specifically, we focus on joint decoding of independently encoded compressively-sampled multi-view video streams. Based on the proposed joint reconstruction method, we also derive a blind video quality estimation technique that can be used to adapt online the video encoding rate at the sensors to guarantee desired quality levels in multi-view video streaming.

In Chapter 3, to address low-power and low-complexity challenges in Internet of Multi-media Things (IoMTs), we propose a new encoding and decoding architecture for multi-view video systems based on Compressed Sensing (CS) principles, composed of cooperative sparsity-aware block-level rate-adaptive encoders, feedback channels and independent decoders. Based on the proposed encoding/decoding architecture, we further develop a CS-based end-to-end rate distortion model by considering the effect of packet losses on the perceived video quality. We then introduce a modeling framework to design network optimization problems in a multi-hop wireless sensor network.

In Chapter 4, we study how to provide throughput-optimal WiFi-like downlink access to users in indoor visible light networks through a set of centrally-controlled and partially interfering light emitting diodes (LEDs). This chapter first proposes a mathematical model of the cooperative visible-light beamforming (LiBeam) problem, presented as maximizing the sum throughput of all VLC users. Then, we solve the resulting mixed integer nonlinear nonconvex programming (MINCoP) problem by designing a globally optimal solution algorithm based on a combination of branch and bound framework as well as convex relaxation techniques. We then design for the first time a large
programmable visible light networking testbed based on USRP X310 software-defined radios, and experimentally demonstrate the effectiveness of the proposed joint beamforming and association algorithm through extensive experiments.

In Chapter 5, we propose visible-light ad hoc networks - referred to as LANETs to alleviate the spectrum crunch problem in overcrowded RF spectrum bands. This chapter discusses typical architectures and application scenarios for LANETs and highlights the major differences between LANETs and traditional mobile ad hoc networks (MANETs). Enabling technologies and design principles of LANETs are analyzed and existing work is surveyed following a layered approach. Open research issues in LANET design are also discussed, including long-range visible light communication, full-duplex LANET MAC, blockage-resistant routing, VLC-friendly TCP and software-defined prototyping, among others.

Finally, Chapter 6 concludes this dissertation.
Chapter 2

Inter-view Motion Compensated Joint Decoding for Compressively-Sampled Multi-View Video Streams

Traditional multi-view video coding techniques, e.g., MVC H.264/AVC, can achieve high compression ratio by adopting intra-view and inter-view prediction, thus resulting in extremely complex encoders and relatively simple decoders. Recently, a multi-view extension of HEVC (MV-HEVC) was proposed to achieve higher coding efficiency by adopting improved flexible coding tree units (CTUs). [4], [5], [6], [7] propose an efficient parallel framework based on many-core processors for coding unit partitioning tree decision, motion estimation, deblocking filter, and intra-prediction, respectively, thus achieving many fold speedups compared with current existing parallel methods. However, typical wirelessly multi-view video streaming applications emerging in recent years such as 360 degrees video, and those encountered in Internet of Thing (IoT) multimedia sensing scenarios [8], [9], [10], [11], [12] are usually composed of low-power and low-complexity mobile devices, smart sensors or wearable sensing devices. 360 degrees video enables immersive "real life", "being there" experience for users by capturing the 360 degree view of the scene of interest, thus requiring higher bitrate than conventional video because it supports a significantly wider field of view. IoT multimedia sensing also needs to simultaneously capture the same scene of interest from different viewpoints and then transmit it to a remote data warehouse, database or cloud for further processing or rendering. Therefore, they need to be based on architectures with relatively simple encoders, while there are less constraints at the decoder side. To address these challenges, so-called
Distributed Video Coding (DVC) architectures have been proposed in the last two decades, where the computational complexity is shifted to the decoder side by leveraging architectures with simple encoders and complex decoder to help offload resource-constrained sensors.

Compressed Sensing (CS) is another recent advancement in signal and data processing that shows promise in shifting the computational complexity at the decoder side. CS has been proposed as a technique to enable sub-Nyquist sampling of sparse signals, and it has been successfully applied to imaging systems [13] [14] since natural imaging data can be represented as approximately sparse in a transformed domain, e.g., through discrete cosine transform (DCT) or discrete wavelet transform (DWT). As a consequence, CS-based imaging systems allow the faithful recovery of sparse signals from a relatively small number of linear combinations of the image pixels referred to as measurements. Recent CS-based video coding techniques [15] [16] [17] [18] [19] have been proposed to improve the reconstruction quality in lossy channels. Therefore, CS has been proposed as a clean-slate alternative to traditional image or video coding paradigms since it enables imaging systems that sample and compress data in a single operation, thus resulting in low-complexity encoders and more complex decoders, which can help offload the sensors and further prolong the lifetime of the mobile devices or sensors.

In this context, our objective is to develop a novel low-complexity multi-view coding/encoding architecture for wirelessly video streaming applications, e.g., 360 degrees immersive video, IoT multimedia sensing, among others, where devices or sensors are usually equipped with power-limited battery. However, current existing algorithms are mostly based on the MVC h.264/AVC or MV-HEVC architecture, which involves complex encoders (motion estimation, motion compensation, disparity estimation, among others) and simple decoder, and is thus not suitable to low-power multi-view video streaming applications. To address this challenge, we propose a novel mult-view encoding/decoding architecture based on compressed sensing theory, where video acquisition and compressing are implemented in one step through low-complexity and low-power compressive sampling (i.e., simple linear operations) while complex computations are shifted to the decoder side. Thus this proposed architecture is more suitable to the aforementioned multi-view scenarios compared with the conventional coding algorithm. To be specific, at the encoder end, one view is selected as a key view (K-view) and encoded at a higher measurement rate; while the other views (CS-views) are encoded at relatively lower rates. At the decoder end, the K-view is reconstructed using a traditional CS recovery algorithm, while the CS-views are jointly decoded by a novel fusion decoding algorithm based on side information generated by a new proposed inter-view motion compensation scheme. Based on the proposed architecture, we develop a blind quality
estimation algorithm and apply it to perform feedback-based rate control to regulate the received video quality.

We claim the following contributions:

- **Side information generated by inter-view motion compensation.** We design a motion compensation algorithm for inter-view prediction, based on which we propose a novel side information generation method that uses the initially reconstructed CS-view and the reconstructed K-view.

- **CS-view fusion reconstruction.** State-of-the-art joint reconstruction methods either use side information [20] as sparsifying basis or use it as the initial point of the developed joint recovery algorithm [21]. Differently, we operate on the measurement domain and propose a novel fusion reconstruction method by padding measurements resampled from side information to the original received CS-view measurements. Then, traditional sparse signal recovery methods can be used to perform the final reconstruction of CS-view by using the resulting measurements.

- **Blind quality estimation for compressively-sampled video.** To guarantee the CS-based multi-view streaming quality is not trivial since original pixels are not only unavailable at the encoder end but also not available at the decoder side. Therefore, how to estimate the reconstruction quality as accurate as possible plays fundamental roles on the quality-assured rate controlling. Based on the proposed reconstruction approach, we develop a blind quality estimation approach, which further can be used to effectively guide the rate adaptation at the encoder end.

The reminder of the chapter is organized as follows. In Section 3.1 related works are discussed. In Section 3.2 we briefly review the basic concepts used in compressed imaging system. In Section 2.3 we introduce the overall encoding/decoding compressive multi-view video streaming framework, and in Section 2.4 we describe the inter-view motion compensation based multi-view fusion decoder. The performance evaluations are presented in Section 2.5 and in Section 5.9 we draw the main conclusions.

### 2.1 Related Work

**CS-based Mono-view Video.** In recent years, several mono-view video coding schemes based on compressed sensing principles have been proposed in the literature [16] [17] [18] [20] [22] [23] [24].
These works mainly focus on single view CS reconstruction by leveraging the correlation among successive frames. For example, [21] proposes a distributed compressive video sensing (DCVS) framework, where video sequences are composed of several GOPs (group of pictures), each consisting of a key frame followed by one or more non-key frames. Key frames are encoded at a higher rate than non-key frames. At the decoder end, the key frame is recovered through the GPSR (gradient projection for sparse reconstruction) algorithm [25], while the non-key frames are reconstructed by a modified GRSR where side information is used as the initial point. Based on [21], the authors further propose dynamic measurement rate allocation for block-based DCVS. In [20], the authors focus on improving the video quality by constructing better sparse representations of each video frame block, where Karhunen-Loeve bases are adaptively estimated with the assistance of implicit motion estimation. [23] and [22] consider the rate allocation and energy consumption under the above-mentioned state-of-the-art mono-view compressive video sensing frameworks. [16] and [17] improve the rate-distortion performance of CS-based codecs by jointly optimizing the sampling rate and bit-depth, and by exploiting the intra-scale and inter-scale correlation of multiscale DWT, respectively.

**CS-based Multi-view Video.** More recently, several proposals have appeared for CS-based multi-view video coding [26] [27] [28] [29]. In [26], a distributed multi-view video coding scheme based on CS is proposed, which assumes the same measurement rates for different views, and can only be applied together with specific structured dictionaries as sparse representation matrix. A linear operator [27] is proposed to describe the correlations between images of different views in the compressed domain. The authors then use it to develop a novel joint image reconstruction scheme. The authors of [28] propose a CS-based joint reconstruction method for multi-view images, which uses two images from the two nearest views with higher measurement rate of the current image (the right and left neighbors) to calculate a prediction frame. The authors then further improve the performance by way of a multi-stage refinement procedure [29] via residual recovery. The readers are referred to [28] [29] and references therein for details. Differently, in this work, we propose a novel CS-based joint decoder based on a newly-designed algorithm to construct an inter-view motion compensated side frame. With respect to existing proposals, the proposed framework considers multi-view sequences encoded at different rates and with more general sparsifying matrixes. Moreover, only one reference view (not necessarily the closest one) is selected to obtain the side frame for joint decoding.

**Blind Quality Estimation.** Ubiquitous multi-view video streaming of visual information and the emerging applications that rely on it, e.g., multi-view video surveillance, 360 degrees video, and IoT
multimedia sensing, require an effective means to assess the video quality because the compression methods and the error-prone wireless links can introduce distortion. Peak Signal-to-Noise Ratio (PSNR) and SSIM (Structural Similarity) [30] are examples of successful image quality assessment metrics; which however require full reference image at the decoder end. In many applications such as surveillance scenarios, however, the reference signal is not available to perform the comparison. Especially, when compressed sensing is used, the reference signal may not even be available at the encoder end. Readers are referred to [31] [32] and references therein for good overviews of image quality assessment (FR-IQA) and non-reference (blind) image quality assessment (NR-IQA) for state-of-the-art video coding methods, e.g., H.264/AVC, respectively. Yet, to the best of our knowledge, we propose for the first time a NR-IQA scheme for compressive imaging systems.

2.2 Preliminaries

In this section, we briefly introduce the basic concepts of compressed sensing for signal acquisition and recovery as applied to compressive video streaming systems.

2.2.1 CS Acquisition

We consider the image frame signal vectorized and represented as \( x \in \mathbb{R}^N \), with \( N = H \times W \) denoting the number of pixels in one frame, with \( H \) and \( W \) representing the dimensions of the captured scene. The element \( x_i \) of \( x \) represents the \( i \)-th pixel in the vectorized signal representation. As mentioned above, CS-based sampling and compression are implemented in a single step. We denote the sampling matrix as \( \Phi \in \mathbb{R}^M \times N \), with \( M \ll N \). Then, the acquisition process can be expressed as

\[
y = \Phi x,
\]

where \( y \in \mathbb{R}^M \) represents the measurements and the vectorized compressed image signal.

2.2.2 CS Recovery

Most natural images can be represented as a sparse signal in some transformed domain \( \Psi \), e.g., DWT or DCT, expressed as

\[
x = \Psi s,
\]

where \( s \in \mathbb{R}^N \) denotes the sparse representation of the image signal. Then, we can rewrite (3.1) as

\[
y = \Phi x = \Phi \Psi s.
\]
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

If $s$ has $K$ non-zero elements, we refer to $x$ as a $K$-sparse signal with respect to $\Psi$.

In [13], the authors proved that if $A \triangleq \Phi \Psi$ satisfies the so-called Restricted Isometry Property (RIP) of order $K$,

$$
(1 - \delta_k)||s||_{l_2}^2 \leq ||As||_{l_2}^2 \leq (1 + \delta_k)||s||_{l_2}^2,
$$

with $0 < \delta_k < 1$ being a small “isometry” constant, then we can recover the optimal sparse representation $s^*$ of $x$ by solving the following convex optimization problem

$$
P_1: \begin{array}{c}
\text{Minimize} \\
\|s\|_0 \end{array} \\
\text{Subject to:} \quad y = \Phi \Psi s \quad (2.5)
$$

by taking only

$$
M = c \cdot K \log(N/K) \quad (2.6)
$$

measurements according to the uniform uncertainty principle (UUP), where $c$ is some predefined constant. Then, $x$ can be obtained as

$$
\hat{x} = \Psi s^*. \quad (2.7)
$$

However, Problem $P_1$ is NP-hard in general, and in most practical cases, measurements $y$ may be corrupted by noise, e.g., channel noise or quantization noise. Then, most state-of-the-art works rely on $l_1$ minimization with a relaxed constraint in the form of

$$
P_2: \begin{array}{c}
\text{Minimize} \\
\|s\|_1 \\
\text{Subject to:} \\
||y - \Phi \Psi s||_2 \leq \epsilon \end{array} \quad (2.8)
$$

to recover $s$. Note that $P_2$ is also a convex optimization problem [33]. The complexity of reconstruction is $O(M^2 N^{3/2})$ if solved by interior point methods [34]. Moreover, researchers interested in sparse signal reconstruction have developed more efficient solvers [25] [35] [36]. For measurement matrix $\Phi$, there are two types, Gaussian random and deterministic. Readers are referred to [20] [37] and references therein for details about Gaussian random and deterministic measurement matrix constructions.

2.3 System Architecture

We consider a multi-view video streaming system equipped with $N$ cameras, with each camera capturing the same scene of interest from different perspectives. At the source nodes, each
captured view is encoded and transmitted independently and jointly decoded at the receiver end. The proposed CS-based N-view encoding/decoding architecture is depicted in Figure 2.1 with $N > 2$.

At the encoder side, we first select one of the considered views as a reference (referred to as K-view) for other views (referred to as CS-views). The frames of the K-view and of the CS-view are encoded at a measurement rate of $R_k$ and $R_{cs}$, respectively. According to the asymmetric distributed video coding principle, the reference view (i.e., K-view) is coded at a higher rate than the non-reference views (i.e., CS-views). In the following, we assume that $R_{cs} \leq R_k$. The size of the scene of interest is denoted as $H \times W$ (in pixels), with the number of total pixels being $N = H \times W$. The K-view frame (denoted as $x_k \in \mathbb{R}^N$) is compressively sampled into a measurement vector $y_k \in \mathbb{R}^{M_k}$ with measurement rate $\frac{M_k}{N} = R_k$, and the CS-view frame $x_{cs} \in \mathbb{R}^N$ is sampled into $y_{cs} \in \mathbb{R}^{M_{cs}}$ with $\frac{M_{cs}}{N} = R_{cs}$. Readers are referred to [38] and references therein for details of the encoding procedure.

At the decoder side, the reconstruction of K-view frames is only based on the received K-view measurements. To reconstruct a CS-view frame, we propose a novel inter-view motion compensated joint decoding method. We first generate a side frame based on the received K-view and CS-view measurements. Then, we fuse the initially received measurements of the CS-view frame with the newly sampled measurements from generated side frame through the proposed novel fusion...
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

Figure 2.2: Block diagram of side frame generation.

algorithm. In the following section, we describe the joint multi-view decoder in detail.

2.4 Joint Multi-view Decoding

In this section, we discuss the proposed joint multi-view decoding method. The frames of the K-view are first reconstructed to serve as a reference for the CS-view reconstruction procedure.

2.4.1 K-view Decoding

Denote the received measurement vector of any frame of the K-view video sequence as \( \hat{y}_k \in \mathbb{R}^{M_k} \) (i.e., a distorted version of \( y_k \) considering the joint effects of quantization, transmission errors, and packet drops due to playout deadline violation). Based on CS theory as discussed in Section 3.2 the K-view frame can be simply reconstructed by solving the following convex optimization problem (sparse signal recovery)

\[
P_3: \min_{s \in \mathbb{R}^N} \|s\|_1 \\
\text{Subject to:} \quad \|\hat{y}_k - \Phi_k \Psi s\|_2^2 \leq \epsilon
\]

(2.9)

and then by mapping \( \hat{x}_k = \Psi^* s^* \), with \( \Phi_k \) and \( \Psi \) representing the K-view sampling matrix and the sparsifying matrix, respectively. Here, \( \epsilon \) denotes the predefined error tolerance, and \( s^* \) represents the reconstructed coefficients (i.e., the minimizer of (2.9)).
2.4.2 Inter-view Motion Compensated Side Frame

Motivated by the traditional mono-view video coding schemes, where motion estimation and compensation techniques are used to generate the prediction frame, we propose an inter-view motion estimation and compensation method for a multi-view video coding scenario. The core idea behind the proposed technique for generating the side frame is to compensate the reconstructed high-quality K-view frame \( \hat{x}_k \) through an estimated inter-view motion vector. To obtain a more accurate inter-view motion estimation vector, we first down-sample the received K-view measurements \( \hat{y}_k \) to obtain the same number of measurements as the number of received CS-view measurements. Then, we use these down-sampled K-view measurements to reconstruct a lower-quality K-view that has the equivalent level of quality as the initially reconstructed CS-view frame. Next, we compare the preliminary reconstructed CS-view with the reconstructed lower-quality K-view to obtain the side frame. Below, we elaborate on the main components of the side frame generation method as illustrated in Fig. 2.2.

**CS-view initial reconstruction.** We denote \( \hat{y}_{cs} \) and \( \Phi_{cs} \) as the received distorted version of CS-view frame measurements and the corresponding sampling matrix, respectively. By substituting \( M_{cs} \) received measurements \( \hat{y}_{cs} \), \( \Phi_{cs} \) and \( \hat{x}_{cs} \) into (2.9), a preliminary reconstructed CS-view frame (denoted as \( \hat{x}^{p}_{cs} \)) can be obtained by solving the corresponding optimization problem.

**K-view down-sampling and reconstruction.** As mentioned above, the reconstructed K-view frame has higher quality than the preliminary reconstructed CS-view. To achieve higher accuracy in the estimation of the inter-view motion vector, we propose to first down-sample the received K-view measurement vector \( \hat{y}_k \) to obtain a new K-view frame with the same (or comparable) reconstructed quality with respect to \( \hat{x}^{p}_{cs} \). Experiments were conducted to validate this approach, which results in more accurate motion vector estimation than the originally reconstructed K-view frame \( \hat{x}_k \).

Since \( R_{cs} \leq R_k \) as stated in Section 2.3, without loss of generality, we consider the CS-view sampling matrix \( \Phi_{cs} \) to be a sub-matrix of \( \Phi_k \). Then, down-sampling can be achieved by selecting from \( \hat{y}_k \) only measurements corresponding to \( \Phi_{cs} \), which is equivalent, apart from transmission errors and quantization errors, to sampling the original K frame with the matrix used for sampling the CS frame. The down-sampled K-view measurement vector and the corresponding reconstructed k-view frame with lower quality are denoted as \( \hat{y}^d_k \) and \( \hat{x}^d_k \), respectively.

**Inter-view motion vector estimation.** With the preliminary reconstructed CS-view frame \( \hat{x}^{p}_{cs} \) and the reconstructed down-sampled quality-degraded K-view frame \( \hat{x}^d_k \), we can then estimate the inter-view motion vector by comparing \( \hat{x}^{p}_{cs} \) and \( \hat{x}^d_k \). The detailed inter-view vector estimation procedure
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

is as follows. First, we divide \( \hat{x}_{cs} \) into a set \( B_{cs}^0 \) of blocks with block size \( B_{cs}^0 \times B_{cs}^0 \) (in pixel). For each current block \( i_{cs} \in B_{cs}^0 \), within a predefined search range \( p \) in the lower-quality K-frame \( \hat{x}_k^d \), a set \( B_k^d(i_{cs}, p) \) of reference blocks, each with the same block size \( B_{cs}^0 \times B_{cs}^0 \), can be identified based on existing strategies [39], e.g., exhaustive search (ES), three step search (TSS), or diamond search (DS). Then, we calculate the mean of absolute difference (MAD) between block \( i_{cs} \in B_{cs}^0 \) and any block \( i_k \in B_k^d(i_{cs}, p) \), which is defined as

\[
MAD_{i_{cs}i_k} = \frac{\sum_{m=1}^{B_{cs}^0} \sum_{n=1}^{B_{cs}^0} \| v_{cs}(i_{cs}, m, n) - v_k^d(i_k, m, n) \|}{B_{cs}^0 \times B_{cs}^0},
\]

(2.10)

with \( v_{cs}(i_{cs}, m, n) \) and \( v_k^d(i_k, m, n) \) denoting the value of the pixels at \( (m, n) \) in block \( i_{cs} \in B_{cs}^0 \) and \( i_k \in B_k^d(i_{cs}, p) \), respectively. Next, the best matching block denoted by \( i_k^* \) has the minimum MAD, which can be obtained by solving

\[
i_k^* = \arg \min_{i_k \in B_k^d(i_{cs}, p)} MAD_{i_{cs}i_k},
\]

(2.11)

with \( MAD_{i_{cs}i_k^*} \) being the corresponding minimum MAD value.

In the single view scenario [40], it is sufficient to search for the block corresponding to the minimum MAD (i.e., block \( i_k^* \)) to estimate the motion vector. However, in the multi-view case, the best matching block \( i_k^* \) is not necessarily a proper estimation of block \( i_{cs} \) due to the possible “hole” problem (i.e., an object that appears in a view is occluded in other views), which can be rather severe.

To address this challenge, we adopt a threshold-based policy. Let \( MAD_{th} \) represent the predefined MAD threshold, which can be estimated online by periodically transmitting a frame at a higher measurement rate. Denote \( \Delta m(i_{cs}) \) and \( \Delta n(i_{cs}) \) as the horizontal and vertical offset (aka motion vector, in pixel) of the block \( i_k^* \) relative to the current block \( i_{cs} \). Then, if a block \( i_k^* \) can be found satisfying \( MAD_{i_{cs}i_k^*} \leq MAD_{th} \), then the current block \( i_{cs} \) is marked as referenced with motion vector \( (\Delta m(i_{cs}), \Delta n(i_{cs})) \); Otherwise, the block is marked as non-referenced.

**Inter-view motion compensation.** After estimating the inter-view motion vector, the side frame \( x_{si} \in \mathbb{R}^N \) can then be generated by compensating the initially reconstructed CS-view frame \( \hat{x}_{cs}^p \), with above-estimated motion vector \( (\Delta m(i_{cs}), \Delta n(i_{cs})) \) for each block in \( B_{cs}^0 \), and the reconstructed high-quality K-view frame \( \hat{x}_k^1 \). The detailed procedure of compensation is as follows. First, we initialize the side frame \( x_{si} \) to \( x_{si} = \hat{x}_{cs}^p \). Then, we replace each referenced block \( i_{cs} \) by using the

\[1\] Note that we estimate the motion vector based on the quality-degraded K-view frame, but compensate the initially reconstructed CS-view frame using the K-view frame at the original reconstructed quality.
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

corresponding block from the initially reconstructed high-quality K-view frame \( \hat{x}_k \) with the estimated motion vector \((\Delta m(i_{cs}), \Delta n(i_{cs}))\).

### 2.4.3 Fusion Decoding Algorithm

The side frame, aka side information, plays a very significant role in state-of-the-art CS-based joint decoding approaches, acting as the initial point \([21]\) of the joint recovery algorithm or sparsifying basis \([20]\). Differently, we explore a novel joint decoding method by directly adopting the side information in the measurement domain. Specifically, we propose to fuse the received CS-view measurements \( \hat{y}_{cs} \) and the measurements resampled from the above generated side-frame \( x_{si} \) to obtain a new measurement vector for further reconstruction of the CS-view. The key idea is to involve more measurements with the assistance of the side frame to further improve the reconstructed quality. This is achieved by generating CS measurements by sampling \( x_{si} \), appending the generated measurements to \( \hat{y}_{cs} \), and then reconstructing a new CS-view frame based on the combined measurements.

To sample the side frame, we use a sampling matrix \( \Phi \), with \( \Phi_{cs} \) and \( \Phi_k \) both being a sub-matrix of \( \Phi \). We then select a number \( R_{si} \times H \times W \) of the resulting measurements, with \( R_{si} \) representing the predefined measurement rate for the side frame. The value of \( R_{si} \) depends on the amount of CS-view measurements \( \hat{y}_{cs} \) that have already been received. Experiments have been conducted to verify the intuitive conclusion that larger \( R_{cs} \) implies to smaller \( R_{si} \). The experiments show that if a sufficient number of CS-view measurements is received at the decoder to result in acceptable reconstruction quality, adding more measurements and combining them from the side frame will result in the introduction of more noise, ultimately reducing the video quality of the recovered frame. Based on experimental evidence, we set \( R_{si} \) as

\[
\left\{ \begin{array}{ll}
R_{si} = 1 - R_{cs}, & \text{if } R_{cs} \leq 0.5 \\
R_{si} = 0.6 - R_{cs}, & \text{if } 0.5 < R_{cs} \leq 0.6 \\
R_{si} = 0, & \text{if } R_{cs} > 0.6
\end{array} \right. \tag{2.12}
\]

With the newly generated \( R_{cs} + R_{si} \) measurements \( \hat{y}_{cs} \), following optimization problem \((2.9)\), the final jointly reconstructed CS-view frame (denoted by \( \hat{x}_{cs} \)) can be obtained.

### 2.4.4 Blind Video Quality Estimation

A natural question for the newly designed multi-view codec is: how good is the reconstructed video quality? As stated in Section \([3.1]\) how to assess the reconstruction quality at the
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

Figure 2.3: (a) original, (b) independently reconstructed, (c) generated side frame, and (d) fusion decoded 5th frame of Exit; Measurement rate is set to 0.2.

decoder end without original reference frames is substantially an open problem, especially for CS-based video coding systems where the original pixels are not available either at the transmitter or at the receiver side. To address this challenge, we propose a blind video quality estimation method within the proposed compressively-sampled multi-view coding/decoding framework described above.

Most state-of-the-art quality assessment metrics, e.g., PSNR or SSIM, are based on the comparison between a-priori-known reference frames and the reconstructed frames in the pixel domain. In this context, we propose to blindly evaluate the quality in the measurement domain by adopting an approach similar to that used to calculate PSNR. The detailed procedure is as follows.

First, the reconstructed CS-view frame $\hat{x}_{cs}$ is resampled at the CS-view measurement rate $R_{cs}$, with the same sampling matrix $\Phi_{cs}$, thus obtaining $M_{cs}$ new measurements denoted by $\hat{y}_{cs}$. Then, the measurement-domain PSNR of $\hat{x}_{cs}$ with respect to the original frame $x_{cs}$ (which is not available even at the encoder side) can be estimated by comparing the measurement vector $\hat{y}_{cs}$ and $y_{cs}$, as
Figure 2.4: (a) original, (b) independently reconstructed, (c) generated side frame, and (d) fusion decoded 25th frame of Vassar; Measurement rate is set to 0.15.
Figure 2.5: PSNR comparison for CS-views (a) view 1, (b) view 3, and (c) view 4, and SSIM comparison for CS-views (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.3 of Vassar.

\[
\text{PSNR} = 10 \log_{10} \left( \frac{(2^n - 1)^2}{\text{MSE}} + \Delta \text{PSNR} \right), \quad (2.13)
\]

where \( n \) is the number of bits per measurement, and

\[
\text{MSE} = \frac{\| \hat{Y}_{cs} - Y_{cs} \|^2}{M^2_{cs}}. \quad (2.14)
\]

In (2.13), \( \Delta \text{PSNR} \) is a compensation coefficient that has been found to stay constant or vary only slowly for each view in the conducted experiments. Hence, it can be estimated online by periodically transmitting a CS-frame at a higher measurement rate.

The proposed blind estimation technique can then be used to control the encoder to dynamically adapt the encoding rate by adaptively increasing or decreasing the rate to guarantee the perceived video quality at the receiver side.
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

Figure 2.6: PSNR comparison for CS-views (a) view 1, (b) view 3, and (c) view 4, and SSIM comparison for CS-views (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.1 of Exit.

2.5 Performance Evaluation

In this section, we experimentally study the performance of the proposed compressive multi-view video decoder by evaluating the perceptual quality, PSNR and SSIM. Three multi-view test sequences are used, i.e., Vassar, Exit and Ballroom representing scenarios with slow, moderate and fast movement characteristics, respectively. The spatial dimension for each frame is 320 × 240 (in pixel). All experiments are conducted only on the luminance component.

At the encoder side, the sampling matrixes $\Phi_k$, $\Phi_{cs}$ and $\Phi$ are implemented with Hadamard matrixes. At the decoder end, TSS [41] is used for motion vector estimation, with block size and search range set to $B = 16$ and $p = 32$, respectively. In the blind video quality estimation algorithm the value of $\Delta$ PSNR is set to 6 and 2.9 for Ballroom and Exit, respectively. GPSR [25] is used to solve $P_3$ in (2.9).

The inter-view motion-compensated side frame generation approach and the fusion decoding method for CS-view frames are two of the main contributions of the chapter. To evaluate
CHAPTER 2. COMPRESSED-SENSING BASED JOINT DECODING

the effectiveness, we compare the following four approaches: i) the proposed inter-view motion compensated side frame based fusion decoding method for CS-view frame (referred to as MC fusion), ii) the GPSR joint decoder proposed in [21] by adopting the side frame generated by the proposed inter-view motion compensation method (referred to as MC joint GPSR), iii) the GPSR joint reconstruction by adopting initially reconstructed CS-view frame as side frame (referred to as joint GPSR) and iv) independent decoding method (referred to as Independent) used as a baseline.

Figure 2.7: PSNR comparison for CS-views (a) view 1, (b) view 3, and (c) view 4, and SSIM comparison for CS-views (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.2 of Ballroom.

First, we evaluate the improvement of CS-view perceptual quality of the proposed MC fusion decoding method compared with Independent reconstruction approach by considering a specific frame as an example, i.e., the 5th frame of Exit and the 25th frame of Vassar. 2-view scenario is considered, where view 1 is set as K-view with measurement rate 0.6 and view 2 is CS-view. Results are illustrated in Fig. 2.3 and Fig. 2.4. We observe that the blurring effect in the independently reconstructed frame is mitigated through joint decoding. Taking the regions of the

Joint GPSR is the base line for MC joint GPSR which is used to validate the effectiveness of the proposed inter-view motion compensation based side frame.
person, bookshelf and photo frame in Fig. 2.3(b) and (d), and almost the whole regions in Fig. 2.4(b) and (d) as examples, we can see that the video quality improvement is noticeable, which corresponds to an improvement in PSNR from 28.17 dB to 29.58 dB and 25.81 dB to 27.87 dB, respectively, and in an improvement in SSIM of 0.09 (from 0.75 to 0.84) and 0.14 (from 0.60 to 0.74), respectively. The block effect introduced by the block-based side frame generation method (shown in Fig. 2.3(c) and Fig. 2.4(c)) is not observed in the reconstructed frame in Fig. 2.3(d) and Fig. 2.4(d) since the proposed fusion decoding algorithm operates in the measurement domain.

Then, we consider the 4-view scenario, views 1, 2, 3 and 4. Without loss of the generality, view 2 is selected as K-view and the other three as CS-views. We then compare the achieved SSIM and PSNR for the first 50 frames of Vassar, Exit, Ballroom. We set three different CS-view measurement rates 0.3, 0.1 and 0.2 for Vassar, Exit, Ballroom, respectively. The results are illustrated in Figs. 2.5, 2.6 and 2.7 with respect to PSNR and SSIM. We observe that the proposed MC fusion decoding method and MC joint GPSR outperform significantly joint GPSR and Independent decoding approaches by up to 1.5 dB and 0.16 in terms of PSNR and SSIM, respectively. MC fusion (blue curve) and MC joint GPSR (pink curve) have similar performance for the tested three multi-view sequences. This observation demonstrates the effectiveness of the proposed fusion decoding method for CS-view; it also showcases the effectiveness of the side frame generated by the proposed inter-view motion compensated side frame. For the Vassar test sequence with CS-view encoding rate 0.3, MC joint GPSR is slightly better than MC fusion by no more than 0.3 dB and 0.03 in terms of PSNR and SSIM. Instead, for Exit with 0.1 encoding rate and Ballroom with 0.2 measurement rate sequences, MC joint GPSR and MC fusion achieve almost the same performance. We can also see that joint GPSR (black curve) proposed for single view video odd and even frames joint decoding just slightly outperforms Independent (red curve), which shows that joint GPSR is not suitable for the multi-view scenario and the importance of the side frame that acts as the initial point for the joint GRSR recovery algorithm.

Finally, to evaluate the proposed blind quality estimation method, we transmit the CS-view sequence over simulated time-varying channels with a randomly generated error pattern. The K-view is assumed to be correctly received and reconstructed. A setting similar to [23] is considered for CS-view transmission, i.e., the encoded CS-view measurements are first quantized and packetized. Then, parity bits are added to each packet. A packet is dropped at the receiver if detected to contain errors after a parity check. Here, we consider the Ballroom and Exit sequences as an example. The simulation result is depicted in Fig. 2.8 where the top figure refers to Ballroom, while the bottom refers to Exit. Different from the results in Figs. 2.6 and 2.7 where the measurement rate is set to 0.1
and 0.2, respectively, in Fig. 2.8, the actual received measurement rate is varying between 0.1 and 0.6 because of the randomly generated error pattern, which further results in varying PSNR. Through comparing the estimated PSNR (blue line) with real PSNR (red dot) for 100 successive frames, we can conclude that the proposed blind estimation within our joint decoding of independently encoding framework is rather precise, with an estimation error of 4.32% for Ballroom and of 6.50% for Exit, respectively. With the proposed quality estimation approach, the receiver can provide precise feedback to the transmitter to guide dynamic rate adaptation.

2.6 Summary

In this chapter, we proposed an inter-view motion compensated side frame generation method for compressive multi-view video coding systems, and based on it, a novel fusion decoding approach for CS-view frame was developed. At the decoder end, a side frame is first generated and then resampled to obtain measurements and then appended after the received CS-view measurements. With the newly combined measurements, the state-of-the-art sparse signal recovery algorithm GPSR
is used to obtain a final reconstructed CS-view frame. Extensive simulation results show that the proposed \textit{MC fusion} decoder outperforms the independent CS-decoder in the case of fast-, moderate- and low-motion scenarios. The efficacy of the proposed side frame is also validated by adopting the existing \textit{joint GPSR} with the proposed inter-view motion compensated side frame as the initial reconstruction point. Based on the proposed multi-view joint decoder, we also developed a video quality assessment metric (operating in the measurement domain) without reference frames for CS video systems. Experimental results with wireless video streaming scenario validated the accuracy of the proposed blind video quality estimation approach.
Chapter 3

Low-Power Multimedia Internet of Things through Compressed Sensing based Multi-view Video Streaming

Low power multimedia wireless sensing systems have enabled a plethora of new services and applications such as virtual reality (VR) based 360 degree video as well as other Internet-of-Things sensing scenarios with multimedia streaming. These applications are usually based off of low-power and low-complexity mobile devices, smart multimedia sensors or wearable sensing devices. 360 degree video enables immersive “real life”, “being there” experience for users by capturing 360 degree view of the scene of interest, thus requiring higher bandwidth than conventional video because it supports a significantly wider field of view (FoV). IoT multimedia sensing also needs to simultaneously capture the same scene of interest from different viewpoints and then transmit it to a remote data warehouse, database, or cloud for further processing or rendering. Therefore, natural system architectures for these applications need to be based on relatively simple encoders, while there are less constraints at the decoder side.

While there has been intense research and considerable progress in wireless video sensing systems, how to enable real-time quality-aware power-efficient multi-view video streaming in large-scale, possibly multi-hop, wireless networks of battery-powered embedded devices is still a substantially open problem. State-of-the-art Multi-view Video Coding (MVC) technologies such as MVC H.264/AVC [42,43] are mainly based on predictive encoding techniques, i.e., selecting one

\footnote{360 degree video, also known as immersive video or spherical video, senses the real world scene in an omnidirectional way.}
frame (referred to as reference frame) in one view (referred to as reference view), based on which they perform motion compensation and disparity compensation to predict other intra-view and inter-view frames, respectively. As a consequence, they are characterized by the following fundamental limitations when applied to multi-view streaming in multi-hop wireless sensor networks:

**Large storage space, high power consumption and encoder complexity on embedded devices.** State-of-the-art MVC technologies incorporating inter-view and intra-view prediction require extra storage space for reference views and frames. They also induce intensive computational complexity at the encoder, which further results in high processing load or additional cost for specialized processors (to perform operations such as motion estimation and compensation) and high power consumption.

**Prediction-based encoding techniques are vulnerable to channel errors.** In predictive encoding approaches, errors in independently encoded frames can lead to error propagation on the predictively encoded frames, which is especially detrimental in wireless networks with lossy links, where best-effort delivery scheme with simple error detection schemes such as UDP are usually adopted [9]. Therefore, to guarantee multi-view video streaming quality, a desirable MVC framework should allow graceful degradation of video quality as the channel quality decreases.

Recently, so-called compressed sensing (CS) techniques have been proposed that are able to reconstruct image or video signals from a relatively “small” number of (random or deterministic) linear combinations of original image pixels, referred to as measurements, *without collecting the entire frame* [13, 14], thereby offering a promising alternative to traditional video encoders by acquiring and compressing video or images simultaneously at very low computational complexity for encoders [38]. This attractive feature motivated a number of works that have applied CS to video streaming in low-power wireless surveillance scenarios. For example, [20, 23, 24] mainly concentrate on single-view CS-based video compression, by exploiting temporal correlation among successive video frames [20, 24] or considering energy-efficient rate allocation in WMSNs with traditional CS reconstruction methods [23]. In [22], we showed that CS-based wireless video streaming can deliver surveillance-grade video for a fraction of the energy consumption of traditional systems based on predictive video encoding such as H.264. In addition, [23] illustrated and evaluated the error-resilience property of CS-based video streaming, which results in graceful quality degradation in wireless lossy links. A few recent contributions [26, 44–46] have proposed CS-based multi-view video streaming techniques, primarily focusing on an independent-encoder and joint-decoder paradigm, which exploits the implicit correlation among multiple views at the decoder side to improve the resulting video quality using complex joint reconstruction algorithms.

From a systems perspective, how to allocate power-efficient rates to different views for a
required level of video quality is another important open problem in wirelessly networked multi-view video streaming systems. Very few algorithms have been reported in the literature to address this issue. For example, \cite{47} and \cite{48} have looked at this problem by considering traditional encoding paradigms, e.g., H.264 or MPEG4; these contributions focus on video transmission in single-hop wireless networks and provide a framework to improve power efficiency by adjusting encoding parameters such as quantization step (QS) size to adapt the resulting rate.

To bridge the aforementioned gaps, in this chapter we first propose a novel CS-based multi-view coding and decoding architecture composed of cooperative encoders and independent decoders. Unlike existing works \cite{26,44,45}, the proposed system is based on independent encoding and independent decoding procedures with limited channel feedback information and negligible content sharing among camera sensors. Furthermore, we propose a power-efficient quality-guaranteed rate allocation algorithm based on a compressive Rate-Distortion (R-D) model for multi-view video streaming in multi-path multi-hop wireless sensor networks with lossy links. Our work makes the following contributions:

**CS-based multi-view video coding architecture with independent encoders and independent decoders.** Different from state-of-the-art multi-view coding architectures, that are either based on joint encoding or on joint decoding, we propose a new CS-based sparsity-aware independent encoding and decoding multi-view structure, that relies on lightweight feedback and inter-camera cooperation.

- **Sparsity estimation.** We develop a novel adaptive approach to estimate block sparsity based on the reconstructed frame at the decoder. The estimated sparsity is then used to calculate the block-level measurement rate to be allocated with respect to a given frame-level rate. Next, the resulting block-level rates are transmitted back to the encoder through the feedback channel. The encoder that is selected to receive the feedback information, referred to as reference view (R-view), shares the content with other non-reference views (NR-views) nearby.

- **Block-level rate adaptive multi-view encoders.** R-view and NR-views perform the block-level CS encoding independently based on the shared block-level measurement rate information. The objective is to not only implicitly leverage the considerable correlation among views, but also to adaptively balance the number of measurements among blocks with different sparsity levels. Our experimental results show that the proposed method outperforms state-of-the-art CS-based encoders with equal block-level measurement rate by up to 5 dB.

**Modeling framework for CS-based multi-view video streaming in multi-path multi-hop wireless sensor networks.** We consider a rate-distortion model of the proposed streaming system that
captures packet losses caused by unreliable links and playout deadline violations. Based on this model, we propose a two-fold (frame-level and path-level) rate control algorithm designed to minimize the network power consumption under constraints on the minimum required video quality for multi-path multi-hop multi-view video streaming scenarios.

The rest of the chapter is organized as follows. In Section 3.1, we discuss related works. In Section 3.2, we review a few preliminary notions. In Section 3.3, we introduce the proposed CS-based multi-view video encoding/decoding architecture. In Section 3.4, we discuss the modified R-D model, and in Section 3.5, we present a modeling framework to design optimization problems of multi-view streaming in multi-hop sensor networks based on the end-to-end R-D model and propose a solution algorithm. Finally, simulation results are presented in Section 3.6, while in Section 5.9 we draw the main conclusions and discuss future work.

3.1 Related Works

**CS-based Single-view Video.** In the past few years, several single-view video coding schemes based on compressed sensing principles have been proposed in the literature [24] [20] [23] [22] [16] [17] [18]. These works mainly focus on single view CS reconstruction by leveraging the correlation among successive frames. For example, [21] proposes a distributed compressive video sensing (DCVS) framework, where video sequences are composed of several GOPs (group of pictures), each consisting of a key frame followed by one or more non-key frames. Key frames are encoded at a higher rate than non-key frames. At the decoder end, the key frame is recovered through the GPSR (gradient projection for sparse reconstruction) algorithm [25], while the non-key frames are reconstructed by a modified GRSR where side information is used as the initial point. Based on [21], the authors further propose dynamic measurement rate allocation for block-based DCVS. In [20], the authors focus on improving the video quality by constructing better sparse representations of each video frame block, where Karhunen-Loeve bases are adaptively estimated with the assistance of implicit motion estimation. [16] and [17] improve the rate-distortion performance of CS-based codecs by jointly optimizing the sampling rate and bit-depth, and by exploiting the intra-scale and inter-scale correlation of multiscale DWT, respectively.

**CS-based Multi-view Video.** More recently, several proposals have appeared for CS-based multi-view video coding [26] [46] [27] [28] [29] [49] [50]. In [26], a distributed multi-view video coding scheme based on CS is proposed, which assumes the same measurement rates for different views, and can only be applied together with specific structured dictionaries as sparse representation matrix.
A linear operator [27] is proposed to describe the correlations between images of different views in the compressed domain. The authors then use it to develop a novel joint image reconstruction scheme. The authors of [28] propose a CS-based joint reconstruction method for multi-view images, which uses two images from the two nearest views with higher measurement rate of the current image (the right and left neighbors) to calculate a prediction frame. The authors then further improve the performance by way of a multi-stage refinement procedure [29] via residual recovery. The readers are referred to [28] [29] and references therein for details. Disparity-based joint reconstruction for multi-view video is also proposed in [49] and [50], where different reconstruction methods, i.e., residual-based and total variation based approaches are adopted, respectively. In our previous work [46], we proposed a motion-aware joint multi-view video reconstruction method based on a newly designed interview motion compensated side information generation approach. Differently, in this article, we propose a novel CS-based independent encoding and independent decoding architecture for multi-view video systems based on newly-designed cooperative sparsity-aware-block-level rate adaptive encoders.

Energy-efficient CS-enabled Video streaming. Several articles have investigated energy-constrained compressively-sampled video streaming. In [22], an analytical/empirical rate-energy-distortion model is developed to predict the received video quality when the overall energy available for both encoding and transmission of each frame is fixed and limited and the transmissions are affected by channel errors. The model determines the optimal allocation of encoded video rate and channel coding rate for a given available energy budget. [51] proposes a cooperative relay-assisted compressed video sensing systems that takes advantage of the error resilience of compressively-sampled video to maintain good video quality at the receiver side while significantly reducing the required SNR, thus reducing the required transmission power. Different from the previous works, which mainly aims at single-view single path CS-based video streaming, in this article, we consider CS-based multi-view video streaming in multi-path multi-hop wireless sensor networks.

3.2 Preliminaries

3.2.1 Compressed Sensing Basics

We first briefly review basic concepts of CS for signal acquisition and recovery, especially as applied to CS-based video streaming. We consider an image signal vectorized and then represented as \( x \in \mathbb{R}^N \), where \( N = H \times W \) is the number of pixels in the image, and \( H \) and \( W \) represent the
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

dimensions of the captured scene. Each element $x_i$ denotes the $i^{th}$ pixel in the vectorized image signal representation. Most natural images are known to be very nearly sparse when represented using some transformation basis $\Psi \in \mathbb{R}^{N \times N}$, e.g., Discrete Wavelet Transform (DWT) or Discrete Cosine Transform (DCT), denoted as $x = \Psi s$, where $s \in \mathbb{R}^N$ is sparse representation of $x$. If $s$ has at most $K$ nonzero components, we call $x$ a $K$-sparse signal with respect to $\Psi$.

In CS-based imaging system, sampling and compression are executed simultaneously through a linear measurement matrix $\Phi \in \mathbb{R}^{M \times N}$, with $M \ll N$, as

$$y = \Phi x = \Phi \Psi s,$$  \hfill (3.1)

with $y \in \mathbb{R}^M$ representing the resulting sampled and compressed vector.

It was proven in [13] that if $A \triangleq \Phi \Psi$ satisfies the following Restricted Isometry Property (RIP) of order $K$,

$$(1 - \delta_k)||s||_2^2 \leq ||As||_2^2 \leq (1 + \delta_k)||s||_2^2,$$  \hfill (3.2)

with $0 < \delta_k < 1$ being a small “isometry” constant, then we can recover the optimal sparse representation $s^*$ of $x$ by solving the following optimization problem

$$P_1: \text{Minimize } ||s||_0 \quad \text{Subject to: } y = \Phi \Psi s$$  \hfill (3.3)

by taking only

$$M = c \cdot K \log(N/K)$$  \hfill (3.4)

measurements, where $c$ is some predefined constant. Afterwards, $x$ can be obtained by

$$\hat{x} = \Psi s^*.$$  \hfill (3.5)

However, problem $P_1$ is NP-hard in general, and in most practical cases, measurements $y$ may be corrupted by noise, e.g., channel noise or quantization noise. Then, most state-of-the-art work relies on $l_1$ minimization with relaxed constraints in the form

$$P_2: \text{Minimize } ||s||_1 \quad \text{Subject to: } ||y - \Phi \Psi s||_2 \leq \epsilon$$  \hfill (3.6)

to recover $s$. Note that $P_2$ is a convex optimization problem. Researchers in sparse signal reconstruction have developed various solvers [25, 35, 36]. For example, the Least Absolute Shrinkage and Selection Operator (LASSO) solver [35] can solve problem $P_2$ with computational complexity $O(M^2N)$. We consider a Gaussian random measurement matrix $\Phi$ in this chapter.
3.2.2 Rate-Distortion Model for Compressive Imaging

Throughout this chapter, end-to-end video distortion is measured as mean squared error (MSE). Since Peak Signal-to-Noise Ratio (PSNR) is a more common metric in the video coding community, we use $\text{PSNR} = 10\log_{10}(255^2 / \text{MSE})$ to illustrate simulation results. The distortion at the decoder $D_{\text{dec}}$ in general includes two terms, i.e., $D_{\text{enc}}$, distortion introduced by the encoder (e.g., not enough measurements and quantization); and $D_{\text{loss}}$, distortion caused by packet losses due to unreliable wireless links and violating playout deadlines because of bandwidth fluctuations. Therefore,

$$D_{\text{dec}} = D_{\text{enc}} + D_{\text{loss}}.$$  (3.7)

To the best of our knowledge, there are only a few works [23] that have investigated rate-distortion models for compressive video streaming, but without considering losses. For example, [23] expands the distortion model in [52] to CS video transmission as

$$D(R) = D_0 + \frac{\theta}{R - R_0},$$  (3.8)

where $D_0$, $\theta$ and $R_0$ are image- or video-dependent constants that can be determined by linear least squares fitting techniques; $R = \frac{M}{N}$ is the user-controlled measurement rate of each video frame.

3.3 CS-based multi-view Coding

Architecture Design

In this section, we introduce a novel encoding/decoding architecture design for CS multi-view video streaming. The proposed framework is based on three main components: (i) cooperative sparsity-aware block-level rate adaptive encoder, (ii) independent decoder, and (iii) a centralized controller located at the decoder. As illustrated in Fig. 5.2, considering a two-view example, camera sensors acquire a scene of interest with adaptive block-level rates and transmit sampled measurements to the base station/controller through a multi-path multi-hop wireless sensor network. Then, the centralized controller calculates the relevant information and feeds it back to the selected R-view. The R-view then shares the limited feedback information with the other one - NR-view. The architecture can be easily extended to $V \geq 2$ views.
Different from existing compressive encoders with equal block measurement rate [20,23], the objective of the proposed framework is to improve the reconstruction quality by leveraging each block’s sparsity as a guideline to adapt the block-level measurement rate. We next describe how to implement the proposed paradigm by discussing each component in detail.

Figure 3.1: Encoding/decoding architecture for multi-hop CS-based multi-view video streaming.

Figure 3.2: Block Sparsity: (a) Original image, (b) Block-based DCT coefficients of (a).
Figure 3.3: Comparison of (a) PSNR, (b) the number of transmitted bits, and (c) the compression rate between approaches with and without mean subtraction.

### 3.3.1 Cooperative Block-level Rate-adaptive Encoder

To reduce the computational burden at encoders embedded in power-constrained devices, most state-of-the-art multi-view proposals focus on developing complex joint reconstruction algorithms to improve the reconstruction quality. Differently, in our architecture we obtain improved quality only through sparsity-aware encoders.

To illustrate the idea, Figure 3.2(b) depicts the sparse representation of Fig. 3.2(a) with respect to block-based DCT transformation. We can observe that sparsity differs among blocks, e.g., the blocks within the coat area are more sparse than others. According to basic compressed sensing theory in Section 3.2.1, (3.4) indicates that the number of required measurements is inversely proportional to the sparsity $K$. Therefore, we propose to adapt the measurement rate at the block level according to sparsity information, i.e., more measurements will be allocated to less-sparse blocks, and vice versa.

In our work, the number of required measurements $M_{vf}^i$ for block $i$ in frame $f$ of view $v$, $1 \leq i \leq B$, is calculated based on the sparsity estimated at the centralized controller and sent back via a feedback channel. Here, $B = \frac{N}{N_b}$ denotes the total number of blocks in one frame with $N$ and $N_b$ being the total number of pixels in one frame and block, respectively. Assume that we have received $\{M_{vf}^i\}_{i=1}^B$. Then, the encoding process is similar to (3.1), described as

$$y_{vf}^i = \Phi_{vf}^i x_{vf}^i,$$

where $y_{vf}^i \in \mathbb{R}^{M_{vf}^i}$ and $\Phi_{vf}^i \in \mathbb{R}^{M_{vf}^i \times N_b}$ are the measurement vector and measurement matrix for block $i$ in frame $f$ of view $v$, respectively; $x_{vf}^i \in \mathbb{R}^{N_b}$ represents the original pixel vector of
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

block \( i \). From (3.9), we can see that \( M^i_{vf} \) varies among blocks from 1 to \( N_b \), thereby implementing block-level rate adaptation. In Section 3.6, the simulation results will show that this approach can improve the quality by up to 5 dB compared with independent encoder and independent decoder method.

**Mean value subtraction.** The CS-based imaging system acquires and compresses each frame simultaneously through simple linear operations as in (3.1). Therefore, it can help reduce the energy consumption compared with traditional signal acquisition and encoding approaches (e.g., H.264/AVC) that are based on complicated motion estimation and motion compensation operations. However, the compression rate of CS is not as high as traditional encoding schemes \[22\]. There is clearly an energy-consumption trade-off between the compression rate and the bit transmission rate. \[22\] analyzes the rate-energy-distortion for compressive video sensing encoder. To improve the compression rate, we perform **mean value subtraction**, which can further help reduce the number of transmitted bits. How to obtain the mean value \( \bar{m} \) will be discussed in Section 3.3.3. Since the original pixels are not available at the compressive encoder, we perform the **mean value subtraction in the measurement domain**. First, we establish a mean value vector \( m \in \mathbb{R}^{N_b} \) with dimensions the same as \( x^i_{vf} \), and where each element is equal to \( \bar{m} \). Then, we use the same block-level measurement matrix \( \Phi^i_{vf} \) to sample \( m \) and then subtract the result from \( y^i_{vf} \) as

\[
\ddot{y}^i_{vf} = y^i_{vf} - \Phi^i_{vf} m = \Phi^i_{vf} (x^i_{vf} - m).
\]  

(3.10)

After sampling, \( \ddot{y}^i_{vf} \) is transmitted to the decoder. From (3.10), we can see that the proposed mean value subtraction in the measurement domain is equivalent to subtraction in the pixel domain.

Next, to validate the effectiveness of mean value subtraction, we take the Vassar sequence as an example. We select a uniform quantization method. The forward quantization stage and the reconstruction stage can be expressed as \( q = \text{sgn}(x) \cdot \left| \frac{|x|}{\Delta} + \frac{1}{2} \right| \) and \( \hat{q} = \Delta \cdot q \), respectively. Here, \( x, q, \hat{q} \) and \( \Delta \) represent original signal, quantized signal, de-quantized signal and quantization step size, respectively. Figure 3.3 shows a comparison of PSNR, the number of transmitted bits and the compression rate with and without mean subtraction, where a measurement rate 0.2 is used, and the total bits in the original frame are \( 320 \times 240 \times 8 = 614400 \) bits. Quantization step sizes from the set \{1, 2, 3, 4, 8, 16, 32, 64, 128, 256\} are selected. From Fig. 3.3(a), we can observe that mean subtraction has a negligible effect on the reconstruction quality and there is no significant quality degradation when the quantization step size is less than 32. This is because the value of measurement is up to thousand and tens of thousand compared to original pixel value with maximum 255. Figures
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

3.3(b) and (c) illustrate that with mean subtraction the total number of bits transmitted for one frame is significantly reduced by up to 30 kbits compared to not using mean subtraction, which corresponds to an improvement in compression rate from 0.2391 to 0.1902.

**Cooperation via sparsity pattern sharing.** Multi-view video streaming is based on reducing the redundancy among views captured by arrays of camera sensors that are assumed to be close enough to each other. Most state-of-the-art literature adopts the concept of distributed system coding architecture [53, 54], where a reference view transmits more measurements than other non-reference views and then the receiver jointly decodes by exploiting the implicit correlation among views. Instead, we allow the encoders to explicitly cooperate to a certain extent. For example, the R-view selected by the centralized controller will periodically receive feedback information, i.e., \( \{M_i\}_{i=1}^B \) and \( \bar{m}_i \), and then share it with the NR-views in the same group. Since camera sensors in the same group are assumed to be close enough to each other, the block sparsity among views will be correlated. By using the same sparsity information, we can directly exploit multi-view correlation at the encoders, thus resulting in a clean-slate compressive multi-view coding framework with simple encoders and simple decoders but with improved reconstruction quality.

### 3.3.2 Independent Decoder

As mentioned above, the proposed framework results in relatively simple decoders. At each decoder, the received \( \hat{y}_{vf}^i \), distorted version of \( \tilde{y}_{vf}^i \) because of the joint effects of quantization, transmission errors, and packet drops, will be independently decoded. The optimal solution \( s_{vf}^{i,\star} \) can be obtained by solving

\[
P_3 : \quad \text{Minimize} \quad ||s_{vf}^i||_1 \\
\text{Subject to:} \quad ||\hat{y}_{vf}^i - \Phi_{vf}^i \Psi_b s_{vf}^i ||_2 \leq \epsilon, \quad (3.11)
\]

where \( \Psi_b \in \mathbb{R}^{N_b \times N_b} \) represents the sparsifying matrix (2-D DCT in this work). We then use (3.5) to obtain the reconstructed block-level image \( \hat{x}_{vf}^i \), by solving \( \hat{x}_{vf}^i = \Psi_b s_{vf}^{i,\star} \). Afterward, \( \{\hat{x}_{vf}^i\}_{i=1}^B \) can be simply reorganized to obtain the reconstructed frame \( \hat{x}_{vf} \).

### 3.3.3 Centralized Controller

The centralized controller is the key component at the receiver, which is mainly in charge of selecting the R-view and estimating sparsity and mean value required to be sent back to the transmitter. Additionally, the controller is also responsible for implementing the power-efficient
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

multi-path rate allocation algorithm discussed in Section 3.5. Next, we introduce the three key functions executed at the controller in sequence, i.e., $R$-view selection, sparsity estimation, and mean value estimation.

**R-view selection.** The controller selects a view to be used as reference view ($R$-view) among views in the same group and then sends feedback information to the selected $R$-view. For this purpose, the controller first calculates the Pearson correlation coefficients among the measurement vectors of any two views as

$$
\rho_{mn} = \text{corr}(\hat{y}_{mf}, \hat{y}_{nf}), \quad \forall m \neq n, \ m, n = 1, \ldots, V,
$$

(3.12)

where $\hat{y}_{mf}$ is the simple cascaded version of all $\hat{y}_{mf}$ and $\text{corr}(\hat{y}_{mf}, \hat{y}_{nf}) \triangleq \frac{\text{cov}(\hat{y}_{mf}, \hat{y}_{nf})}{\sigma_{mf} \sigma_{nf}}$. Then, view $m^*$, referred to as $R$-view, is selected by solving

$$
m^* = \arg \max_{m=1,\ldots,V} \tilde{\rho}_m,
$$

(3.13)

where $\tilde{\rho}_m \triangleq \frac{1}{V-1} \sum_{n \neq m} \rho_{mn}$. The reconstructed frame $\hat{x}_{v,f}$ of the $R$-view is then used to estimate the block sparsity $K_i$ and the frame mean value $\bar{m}$ for block $i$.

Next, we take the Vassar 5-view scenarios as an example, Table 3.1 shows the calculated $\tilde{\rho}_m$. We can see that the average Pearson correlation coefficient of view 3 is the largest. Therefore, view 3 is selected as $R$-view. Moreover, to elaborate how much quality gain we can obtain if the other views except view 3 are selected as $R$-view, we also set them as $R$-view and calculate the average improved PSNR, respectively, as shown in Table 3.2. We can observe that the improved average PSNR is proportional to $\tilde{\rho}_m$, where selecting view 3 as $R$-view results in the highest improved average PSNR gain, i.e., 1.6674 dB. For this case, because the Vassar multi-view sequences used here is captured by parallel-deployed cameras with equal spacing, we obtain the same result, i.e., view 3 as $R$-view, as if we were to choose simply the most central sensor. However, for scenarios with cameras that are not parallel-deployed with unequal spacing, selecting the most central sensor is not necessarily a good choice.

**Table 3.1:** Average Pearson correlation coefficient for Vassar five views.

<table>
<thead>
<tr>
<th></th>
<th>View 1</th>
<th>View 2</th>
<th>View 3</th>
<th>View 4</th>
<th>View 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{\rho}_m$</td>
<td>0.8184</td>
<td>0.8988</td>
<td>0.9243</td>
<td>0.8973</td>
<td>0.8435</td>
</tr>
</tbody>
</table>

**Sparsity estimation.** Since the original frame in the pixel domain is not available, we propose to estimate sparsity based on the reconstructed frame $\hat{x}_{v,f}$ as follows. By solving the optimization
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

Table 3.2: Improved average PSNR (dB) when selecting different Vassar views as R-view.

<table>
<thead>
<tr>
<th>R-view</th>
<th>View 1</th>
<th>View 2</th>
<th>View 3</th>
<th>View 4</th>
<th>View 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>1.2312</td>
<td>1.6241</td>
<td>1.6674</td>
<td>1.6167</td>
<td>1.3833</td>
</tr>
</tbody>
</table>

problem P3 in (3.11), we can obtain the block sparse representation $s_{i,f}^{\star}$ and then reorganize $\{s_{i,f}^{\star}\}_{i=1}^{B}$ to get the frame sparse representation $s_{v,f}^{\star}$ periodically. The sparsity coefficient $K^i$ is defined as the number of non-zero entries of $s_{v,f}^{\star}$. However, natural pictures in general are not exactly sparse in the transform domain. Hence, we introduce a predefined percentile $p_s$, and assume that the frame can be perfectly recovered with $N \cdot p_s$ measurements. Based on this, one can adaptively find a threshold $T$ above which transform-domain coefficients are considered as non-zero entries. The threshold can be found by solving

$$
\frac{||\max(|s_{v,f}^{\star}| - T, 0)||_0}{N} = p_s. \quad (3.14)
$$

Then, we apply $T$ to each block $i$ to estimate the block sparsity $K^i$ as

$$
K^i = ||\max(|s_{v,f}^{\star}| - T, 0)||_0. \quad (3.15)
$$

According to (3.4) and given the frame measurement rate $R$, $M_{i,f}$ can then be obtained as

$$
M_{i,f}^i = \frac{K^i \log_{10}(\frac{N_b}{K^i})}{\sum_{i=1}^{B} K^i \log_{10}(\frac{N_b}{K^i})} N R. \quad (3.16)
$$

**Mean value estimation.** Finally, the mean value $\bar{m}$ can be estimated from $\bar{x}_{v,f}$ as

$$
\bar{m} = \frac{1}{N} \sum_{i=1}^{N} \bar{x}_{v,f}(i). \quad (3.17)
$$

With limited feedback and lightweight information sharing, implementing block-level rate adaptation at the encoder without adding computational complexity can improve the reconstruction performance of our proposed encoding/decoding paradigm. This claim will be validated in Section 3.6 in terms of Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity (SSIM) [30].

3.4 End-to-End Rate-Distortion Model

To handle CS-based multi-view video streaming with guaranteed quality, a rate-distortion model to measure the end-to-end distortion that jointly captures the effects of encoder distortion
and transmission distortion as stated in (3.7) is needed. To this end, we modify the R-D model (3.8) proposed in [23] by adding a packet loss term to jointly account for compression loss and packet loss in compressive video wireless streaming systems. In traditional predictive-encoding based imaging systems, the importance of packets is not equal (i.e., I-frame packets have higher impact than P-frame and B-frame packets on the reconstructed quality). Instead, each packet in CS-based imaging systems has the same importance, i.e., it contributes equally to the reconstruction quality. Therefore, the packet loss probability $p_{\text{loss}}$ can be converted into a measurement rate reduction through a conversion parameter $\kappa$ and considered into the rate-distortion performance, described as

$$D_{\text{dec}} = D_{\text{enc}} + D_{\text{loss}} = D_{0} - \frac{\theta}{R - \kappa p_{\text{loss}} - R_{0}}.$$  

(3.18)

However, how to derive captured-scene-dependent constants $D_{0}$, $\theta$, and $R_{0}$ in (3.18) is not trivial. The reasons are listed as follows:

1) Packet loss rate plays a fundamental role in the modified R-D model. In multi-view video streaming in multi-path multi-hop wireless network, how to model the packet loss rate as accurately as possible is still an open problem. In Section 3.5 we describe our proposed packet loss probability model in detail.

2) The original pixel values are not available at the receiver end and even not available at the
transmitter side in compressive multi-view streaming systems. To address this challenge, we develop a simple but very effective online estimation approach to obtain these three fitting parameters. We let the R-view periodically transmit a frame at a higher measurement rate, e.g., 60% measurement \[^2\] and after reconstruction at the decoder side, the reconstructed frame is considered as the original image in the pixel domain. We then resample it at different measurement rates and perform the reconstruction procedure again. Finally, approximate distortion in terms of MSE can be calculated between the reconstructed frame at lower measurement rates and the reconstructed frame with 60% measurements.

We take the Vassar view 2 sequence as example. According to the above-mentioned online rate-distortion estimation approach, a measurement rate of 0.6 is selected. Figure 3.4 illustrates the simulation results, where the black solid line is the rate-distortion curve fitted through a linear least-square approach. To evaluate this approach, we calculate the distortion value for frames 1, 4 and 80 at different measurement rates and then compare them with the estimated rate-distortion curve, where ground-truth distortion values are depicted as red pentagrams, blue squares and green pluses compared to the black line (estimated rate-distortion curve), respectively. We can observe that model (3.18) matches well the ground-truth distortion values.

Next, in Section 3.5 we further validate the effectiveness of the R-D model by applying it to the design of a modeling framework for compressive multi-path wireless video streaming, where a power-efficient problem is presented as an example.

### 3.5 Network Modeling Framework

We consider compressive wireless video streaming over multi-path multi-hop wireless multimedia sensor networks (WMSNs). Based on the R-D model developed in Section 3.4, we first formulate a video-quality-assured power minimization problem, and then solve the resulting nonlinear nonconvex optimization problem by proposing an online solution algorithm with low computational complexity.

**Network model.** In the considered WMSN there are a set \( V \) of camera sensors at the transmitter side, with each camera capturing a video sequence of the same scene of interest, and then sending the sequence to the server side through a set \( Z \) of pre-established multi-hop paths. Denote \( L^z \) as the set of hops belonging to path \( z \in Z \), with \( d^{z,l} \) being the hop distance of the \( l^{th} \) hop in \( L^z \). Let \( V = |V| \), \( Z = |Z| \), and \( L^z = |L^z| \) represent cardinality of sets \( V \), \( Z \) and \( L^z \), respectively. The following three

\[^2\]Based on CS theory, image reconstructed by using 60% measurement can result in basically the original image, i.e., the differences between the reconstructed image and the original image cannot be perceived by human eyes.
assumptions are considered:
- **Pre-established routing**, i.e., the set of multi-hop paths $Z$ is established in advance through a given routing protocol (e.g., AODV [55]) and does not change during the video streaming session.
- **Orthogonal channel access**, i.e., there exists a pre-established orthogonal channel access, e.g., based on TDMA, FDMA, or CDMA, and hence concurrent transmissions do not interfere with each other [56].
- **Time division duplexing**, i.e., each node cannot transmit and receive simultaneously, implying that only half of the total air-time is used for transmission or reception.

At the receiver side, the video server concurrently and independently decodes each view of the received video sequences, and based on the reconstructed video sequences it then computes the rate control information and sends the information back to camera sensors for actual rate control. For this purpose, we define two types of video frames, Reference Frame (referred to as $R$-frame) and Non-Reference Frame (referred to as $NR$-frame). An $R$-frame is periodically transmitted by the $R$-view; all other frames sent out by the $R$-view and all frames transmitted by the $NR$-views are categorized as $NR$-frames. Compared to an $NR$-frame, an $R$-frame is encoded with equal or higher sampling rate and then sent to the receiver side with much lower transmission delay. Hence, an $R$-frame can be reconstructed with equal or higher video quality and used to estimate sparsity pattern information, which is then fed back to video cameras for rate control in encoding the following $NR$-frames. For the $R$-view, we consider a periodic frame pattern, meaning that the $R$-view camera encodes its captured video frames as $R$-frames periodically, e.g., one every 30 consecutive frames.

In the above setting, our objective is to minimize the average power consumption of all cameras and communication sensors in the network with guaranteed reconstructed video quality for each view, by jointly controlling video encoding rate and allocating the rate among candidate paths. To formalize this minimization problem, next we first derive the packet loss probability $p_{loss}$ in (3.18).

**Packet loss probability.** According to the proposed modified R-D model (3.18), packet losses affect the video reconstruction quality because they introduce an effective measurement rate reduction. Therefore, effective estimation of packet loss probability at the receiver side has significant impact on frame-level measurement rate control.

In real-time wireless video streaming systems, a video packet can be lost primarily for two reasons: i) the packet fails to pass a parity check due to transmission errors introduced by unreliable wireless links, and ii) it takes too long for the packet to arrive at the receiver side, hence violating the maximum playout delay constraint. Denoting the corresponding packet loss probability as $p_{per}$ and
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

$p_{\text{dly}}$, respectively, the total packet loss rate $p_{\text{loss}}$ can then be written as

$$p_{\text{loss}} = p_{\text{per}} + p_{\text{dly}}.$$  

In the case of multi-path routing as considered above, $p_{\text{per}}$ and $p_{\text{dly}}$ in (3.19) can be further expressed as

$$p_{\text{per}} = \sum_{z \in \mathcal{Z}} \frac{b_z}{b} p_{\text{per}}^z, \quad (3.20)$$

$$p_{\text{dly}} = \sum_{z \in \mathcal{Z}} \frac{b_z}{b} p_{\text{dly}}^z, \quad (3.21)$$

where $p_{\text{per}}^z$ and $p_{\text{dly}}^z$ represent the packet loss rate for path $z \in \mathcal{Z}$ due to transmission error and delay constraint violation, respectively; $b$ and $b_z$ represent total video rate and the rate allocated to path $z \in \mathcal{Z}$, respectively.

Since each path $z \in \mathcal{Z}$ may have one or multiple hops, to derive the expressions for $p_{\text{per}}^z$ and $p_{\text{dly}}^z$ in (3.20) and (3.21), we need to derive the resulting packet error rate and delay violation probability at each hop of path $z \in \mathcal{Z}$, denoted as $p_{\text{per}}^{z,l}$ and $p_{\text{dly}}^{z,l}$, respectively. For this purpose, we first express the feasible transmission rate achievable at each hop. For each hop $l \in \mathcal{L}^z$ along path $z \in \mathcal{Z}$, let $G^{z,l}$ and $N^{z,l}$ represent the channel gain that accounts for both path loss and fading, and the additive white Gaussian noise (AWGN) power currently measured by hop $l$, respectively. Denoting $P^{z,l}$ as the transmission power of the sender of hop $l$, then the attainable transmission rate for the hop, denoted by $C^{z,l}(P^{z,l})$, can be expressed as [57]

$$C^{z,l}(P^{z,l}) = \frac{W}{2} \log_2 \left( 1 + K \frac{P^{z,l}G^{z,l}}{N^{z,l}} \right), \quad (3.22)$$

where $W$ is channel bandwidth in Hz, calibration factor $K$ is defined as

$$K = \frac{-\phi_1}{\log(\phi_2 p_{\text{ber}})}, \quad (3.23)$$

with $\phi_1, \phi_2$ being constants depending on available set of channel coding and modulation schemes, and $p_{\text{ber}}$ is the predefined maximum residual bit error rate (BER). Then, if path $z \in \mathcal{Z}$ is allocated video rate $b_z$, for each hop $l \in \mathcal{L}^z$, the average attainable transmission rate should be equal to or higher than $b_z$, i.e.,

$$\mathbb{E}[C^{z,l}(P^{z,l})] \geq b_z, \quad (3.24)$$
with \( \mathbb{E}[C_{z,l}(P_{z,l})] \) defined by averaging \( C_{z,l}(P_{z,l}) \) over all possible channel gains \( G_{z,l} \) in (3.22).

Based on the above setting, we can now express the single hop packet error rate \( p_{\text{per}}^{z,l} \) for each hop \( l \in \mathcal{L}_z \) of path \( z \in \mathcal{Z} \) as,

\[
p_{\text{per}}^{z,l} = 1 - (1 - p_{\text{ber}})^L,
\]

where \( L \) is the predefined packet length in bits. Further, we characterize the queueing behavior at each wireless hop as in [58] using a M/M/1 model to capture the effects of channel-state-dependent transmission rate \( (3.22) \) single-hop queueing delay. Denoting \( T_{z,l}^{z,l} \) as the delay budget tolerable at each hop \( l \in \mathcal{L}_z \) of path \( z \in \mathcal{Z} \), the resulting packet drop rate due to delay constraint violation can then be given as [59]

\[
p_{\text{dly}}^{z,l} = \mathbb{E}[C_{z,l}(P_{z,l})] - b \frac{\sum_{l \in \mathcal{L}_z} T_{z,l}^{z,l}}{L},
\]

with \( \mathbb{E}[C_{z,l}(P_{z,l})] \) defined in (3.24). For each path \( z \in \mathcal{Z} \), the maximum tolerable end-to-end delay \( T_{\text{max}}^{\text{z}} \) can be assigned to each hop in different ways, e.g., equal assignment or distance-proportional assignment [60]. We adopt the same delay budget assignment scheme as in [60].

Finally, given \( p_{\text{per}}^{z,l} \) and \( p_{\text{dly}}^{z,l} \) in (3.25) and (3.26), we can express the end-to-end packet error rate \( p_{\text{per}}^{z} \) and delay violation probability \( p_{\text{dly}}^{z} \) in (3.20) and (3.21) as, for each path \( z \in \mathcal{Z} \),

\[
p_{\text{per}}^{z} = \sum_{l \in \mathcal{L}_z} p_{\text{per}}^{z,l}, \forall z \in \mathcal{Z},
\]

\[
p_{\text{dly}}^{z} = \sum_{l \in \mathcal{L}_z} p_{\text{dly}}^{z,l}, \forall z \in \mathcal{Z},
\]

by neglecting the second and higher order product of \( p_{\text{per}}^{z,l} \) and of \( p_{\text{dly}}^{z,l} \). The resulting \( p_{\text{per}}^{z} \) and \( p_{\text{dly}}^{z} \) provide an upper bound on the real end-to-end packet error rate and delay constraint violation probability. The approximation error is negligible if packet loss rate at each wireless hop is low or moderate. Note that it is also possible to derive a lower bound on the end-to-end packet loss rate, e.g., by applying the Chernoff Bound [61].

**Packet loss to measurement rate.** After having modeled \( p_{\text{loss}} \), we now concentrate on determining \( \kappa \) to convert \( p_{\text{loss}} \) to measurement rate reduction (referred to as \( R_d = \kappa \cdot p_{\text{loss}} \)). First, parameter \( \tau = \frac{1}{QN} \) is defined to convert the amount of transmitted bits of each frame to its measurement rate \( R \) used in the (3.18), with \( Q \) being the bit-depth for each measurement. We assume that \( b \) is equally distributed among \( F \) frames within 1 second for all \( V \) views, i.e., the transmitted bits for
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

each frame is \( b/F/V \). Thus, measurement rate \( R \) for each frame of each view is equal and defined as \( R = \tau b/F/V \). Then, we can define \( \kappa \) as

\[
\kappa = \tau L \left\lfloor \frac{b/F/V}{L} \right\rfloor ,
\]

and rewrite (3.18) as

\[
D_{\text{dec}} = D_0 - \frac{\theta}{\tau b/F/V - \kappa p_{\text{loss}} - R_0}.
\]

Problem formulation. Based on (3.30), we formulate, as an example of applicability of the proposed framework, the problem of power consumption minimization for quality-assured compressive multi-view video streaming over multi-hop wireless sensor networks, by jointly determining the optimal frame-level encoding rate and allocating transmission rate among multiple paths, i.e.,

\[
P_4 : \quad \text{Minimize} \quad \sum_{z \in Z} \sum_{l \in L^z} P^{x,l}
\]

\[
\text{Subject to:} \quad b = \sum_{z \in Z} b^z
\]

\[
D_{\text{dec}} \leq D_t
\]

\[
0 < \tau b/F/V - \kappa p_{\text{loss}} \leq 1
\]

\[
0 \leq P^{x,l} \leq P_{\max}, \quad \forall l \in L^z, \quad z \in Z,
\]

where \( D_t \) and \( P_{\max} \) represent the constraints upon distortion and power consumption, respectively. Here, (3.33) and (3.34) are the constraints for required video quality level and total measurement rate not lower than 0 and higher than 1, respectively. In fact, the optimization problem \( P_4 \) is non-convex because the distortion constraint is non-convex. Solving it directly will be computationally expensive due to the large space of \( b \). Therefore, in the following, we design a solution algorithm to find the solution to the problem in real time.

Solution Algorithm. The core idea of the solution algorithm is to iteratively control video encoding and transmission strategies at two levels, i.e., adjusting video encoding rate for each frame (frame level) and allocating the resulting video data rate among different paths (path level). In each iteration, the algorithm first determines at the frame level the minimum video encoding rate required to achieve predefined reconstructed video quality, i.e., \( b \) in (3.33); and then determines at the path level the optimal routing strategy with minimal power consumption, i.e., \( b^z \) for each path \( z \in Z \).
At the frame level, given the current total video encoding rate $b$ and assigned rate $b^z$ for each path $z \in Z$, the algorithm estimates the video construction distortion $D_{\text{dec}}$ based on (3.19)-(3.30). Then, if the video quality constraint in optimization problem $P_4$ can be strictly satisfied, i.e., the inequality holds in (3.33), it means that power consumption can be further reduced by reducing the total video encoding rate $b$, e.g., by a predefined step $\Delta b$, while keeping the distortion constraint (3.33) still satisfied. Otherwise, if constraint (3.33) is violated, we need to reduce reconstructed video $D_{\text{dec}}$ by increasing the video encoding rate $b$ hence transmission power. Whenever there are changes with the total encoding rate $b$, it triggers at the path level rate allocation among different paths. For example, if $b$ is increased by $\Delta b$, the increased amount of video data rate is allocated to the path that results in minimum increase of power consumption, and vice versa.
As the above procedure goes on, the resulting video distortion $D_{\text{dec}}$ is maintained fluctuating around, ideally equal to, the predefined maximum tolerable distortion $D_{\text{max}}$. Hence, we approximately solve the optimization problem $P4$ formulated in (3.31)-(3.35), and the resulting power consumption provides an upper bound on the real minimum required total power. The algorithm is summarized in Algorithm 1. Next, in Section 3.6 we validate the effectiveness of the proposed solution algorithm through extensive simulation results.

Figure 3.6: PSNR against frame index for (a) view 1, (b) view 2 (R-view), (c) view 3, and (d) view 4 of sequence Exit.
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

3.6 Performance evaluation

The topology includes a certain number $V$ camera sensors and pre-established paths with random number of hops between camera sensors and the receiver. The frame rate is $F = 30$ fps, and the R-view periodically sends the R-frame every second. At the sparsity-aware CS independent encoder side, each frame is partitioned into $16 \times 16$ non-overlapped blocks implying $N_d = 256$. A measurement matrix $\Phi_{vf}^i$ with elements drawn from independent and identically distributed (i.i.d) Gaussian random variables is considered, where the random seed is fixed for all experiments to make sure that $\Phi_{vf}^i$ is drawn from the same matrix. The elements of the measurement vector $\tilde{y}_{vf}^i$ are quantized individually by an 8-bit uniform scalar quantizer and then transmitted to the decoder. At the independent decoder end, we use $\Psi_b$ composed of DCT transform basis as sparsifying matrix and choose the LASSO algorithm for reconstruction motivated by its low-complexity and excellent recovery performance characteristics. We consider two test multi-view sequences, Exit and Vassar, which are made publicly available \[62\]. In the sequences considered, the optical axis of each camera is parallel to the ground, and each camera is 19.5 cm away from its left and right neighbors. A spatial resolution of $(H = 240) \times (W = 320)$ is considered. Exit and Vassar are indoor surveillance and outdoor surveillance videos, respectively. The texture change of Exit is faster than that of Vassar, i.e., the block sparsity of Exit changes more quickly.

3.6.1 Evaluation of CS-based Multi-view Encoding/Decoding Architecture

We first experimentally study the performance of the proposed CS-based multi-view encoding/decoding architecture by evaluating the PSNR (as well as SSIM) of the reconstructed video sequences. Experiments are carried out only on the luminance component. Next, we illustrate the performance comparisons among (i) traditional Equal-Block-Measurement-Rate Independently Encoding and Independently Decoding approach (referred to as EBMR-IEID), (ii) the proposed sparsity-aware Adaptive-Block-Measurement-Rate Independently Encoding and Independently Decoding approach (referred to as ABMR-IEID) and (iii) Independently Encoding and Jointly Decoding (referred to as IEJD) proposed in \[45\] which selects one view as reference view reconstructed by traditional CS recovery method, while other views are jointly reconstructed by using reference frame.

Figures \[3.5\] and \[3.6\] show the PSNR comparisons of 50 frames for views 1, 2, 3 and 4 of Vassar and Exit multi-view sequences, where a 0.3 measurement rate for each view of ABMR-IEID and EBMR-IEID is selected. To assure fair comparison, the measurement rate of each view in IEJD is also set to 0.3. Besides, according to the R-view selection algorithm, view 2 is chosen as the R-view
Figure 3.7: Rate-distortion comparison for frame 75 of Vassar sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4.

Table 3.3: PSNR and SSIM comparison for Vassar eight views.

<table>
<thead>
<tr>
<th>View #</th>
<th>ABMR-IEID</th>
<th>EBMR-IEID</th>
<th>IEJD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR (dB)</td>
<td>SSIM</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td>1</td>
<td>33.6675</td>
<td>0.8648</td>
<td>30.0883</td>
</tr>
<tr>
<td>2</td>
<td>33.7768</td>
<td>0.8686</td>
<td>30.3459</td>
</tr>
<tr>
<td>3</td>
<td>34.1934</td>
<td>0.8771</td>
<td>30.6265</td>
</tr>
<tr>
<td>4</td>
<td>33.5766</td>
<td>0.8696</td>
<td>30.4168</td>
</tr>
<tr>
<td>5</td>
<td>33.3030</td>
<td>0.8624</td>
<td>30.1011</td>
</tr>
<tr>
<td>6</td>
<td>34.2191</td>
<td>0.8846</td>
<td>30.6803</td>
</tr>
<tr>
<td>7</td>
<td>32.9924</td>
<td>0.8575</td>
<td>29.8250</td>
</tr>
<tr>
<td>8</td>
<td>32.3376</td>
<td>0.8472</td>
<td>29.3713</td>
</tr>
</tbody>
</table>
for this scenario. Since the R-view transmits the R-frame periodically, i.e., per second, and for the first frame of each period, the encoder will not encode them based on sparsity pattern, therefore we can observe drops occurred periodically in Fig. 3.5(b) and Fig. 3.6(b). For the Vassar sequences, as illustrated in Fig. 3.5 we can see that the proposed method ABMR-IEID outperforms the traditional approach EBMRIEID and IEJD by up to 3.5 dB and 2.5 dB in terms of PSNR, respectively. For the Exit sequences, Figure 3.6 shows improvement in the reconstruction quality of ABMR-IEID compared with EBMRIEID and IEJD fluctuates more than that of Vassar video, with increased PSNR varying from 5 dB to 2 dB and from 4 dB to 1 dB, respectively. This phenomenon occurs because of the video-based features, i.e., the texture of Exit changes faster than in Vassar. In other words, the proposed scheme is more robust in surveillance scenarios where the changes of texture are

Figure 3.8: Rate-distortion comparison for frame 9 of Exit sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4.
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

Figure 3.9: SSIM comparison for frame 75 of Vassar sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4.

less severe. However, we can eliminate this phenomenon by transmitting R-frames more frequently. Figures 3.5 and 3.6 also depict performance improvement on NR-views (views 1, 3 and 4 here), i.e., by sharing the sparsity information between R-view and NR-views, correlation among views is implicitly exploited to improve the reconstruction quality.

We then illustrate the rate-distortion characteristics of ABMR-IEID, EBMR-IEID and IEJD. Figures 3.7 and 3.8 show the comparisons of Vassar and Exit 4-view scenario, where the 75th frame of Vassar and 9th frame of Exit are taken as example, respectively. Evidently, ABMR-IEID outperforms significantly EBMR-IEID and IEJD, especially as the number of measurements increases. Since view 2 is selected as reference view, aka K-view for IEJD, we set a fixed measurement rate 0.6 for the K-view [45], therefore, a platform is observed in view 2 for IEJD method. We can observe
that at measurement rate 0.4, ABMR-IEID can improve PSNR by up to 4.4 dB and 2.4 dB, not only on R-view but also on NR-views for both video sequences. In the experiments, as the number of views increases, ABMR-IEID can still obtain significant PSNR gain compared to EBMR-IEID; while the performance of IEJD degrades faster as the distance to R-view increases, which can be apparently observed from Figs. 3.5, 3.6 and 3.7 where view 4 has distance 2 to R-view but with relatively lower PSNR gain compared to views 1 and 3. SSIM comparisons are also conducted for Vassar and Exit, as shown in Figures 3.9 and 3.10. IEJD outperforms ABMR-IEID and EBMR-IEID when measurement rate is below 0.2, while above 0.2, the SSIM performance of IEJD decreases fast, even worse than EBMR-IEID. The proposed ABMR-IEID method outperforms the other two
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

Figure 3.11: Reconstructed frame 25 of view 3 by (a) ABMR-IEID, (b) EBMR-IEID, (c) IEJD, and reconstructed frame 25 of view 7 by (d) ABMR-IEID, (e) EBMR-IEID, and (f) IEJD.

methods till measurement rate $0.6^3$ with improvement up to 0.05 for both test sequences.

Next, we extend the scenario to 8 views on Vassar, where view 4 is selected as R-view, and the measurement rate is set to 0.35 for all views. Figure 3.11 shows the specific reconstructed image comparison, where the left column illustrates the reconstructed frame 25 of view 3 and view 7 by ABMR-IEID, respectively. The middle column shows the reconstructed images by EBMR-IEID, and the left columns shows the results obtained by using IEJD. We can observe that the quality of images located in the left column is much better than that in the right two columns (e.g., the curtain in the 2nd floor and person in the scene, and etc.). Furthermore, Table 3.3 shows the detailed PSNR and SSIM value comparison between ABMR-IEID and EBMR-IEID and IEJD for frame 25 of 8 views. From Fig. 3.11 and Table 3.3, we can see that ABMR-IEID also works well on 8 views compared to ABMR-IEID and EBMR-IEID, with PSNR and SSIM improvement up to 3.5 dB and 0.05, respectively. However, the IEJD method proposed in [45] does not perform well on 8 views, where the gain is almost negligible.

$^3$Based on CS theory, image reconstructed by using 60% measurement can result in basically the original image, therefore, the SSIM of ABMR-IEID almost converges to that of EBMR-IEID.
CHAPTER 3. COMPRESSED-SENSING BASED LOW-POWER IOMT

3.6.2 Evaluation of Power-efficient Compressive Video Streaming

The following network topologies are considered: i) 2-path scenario with 2-hop path 1 and 1-hop path 2; ii) 3-path scenario with 2-hop path 1, 1-hop path 2 and 2-hop path 3. We assume bandwidth $W = 1$ MHz for each channel. The maximum transmission power at each node is set to 1 W and the target distortion in MSE is 50. We also assume the maximum end-to-end delay is $T_{\text{max}} = 0.5$ s assigned to each hop proportional to the hop distance. To evaluate PE-CVS (referred to as the proposed power-efficient compressive video streaming algorithm proposed in Algorithm 1), we compare it with an algorithm (referred to as ER-CVS) that equally splits the frame-level rate calculated by PE-CVS onto different paths.

![Figure 3.12: 2-path Scenario: (a) Total power consumption comparison, (b) Saved power consumption by PE-CVS compared to ER-CVS.](image)

Figures 3.12 and 3.13 illustrate the total power consumption comparison between PE-CVS and ER-CVS and the saved power by PE-CVS compared to ER-CVS for 2-path and 3-path topologies, respectively. From Figs. 3.12(a) and 3.13(a), we see that PE-CVS (depicted in red line) results in less power consumption than ER-CVS (black dash line) for both cases. At some points, the total power consumption of PE-CVS and ER-CVS is almost the same. This occurs because the path-level bit rates calculated by PE-CVS are equal to each other. Since ER-CVS uses frame-level rate obtained from PE-CVS and equally allocates it to each path, thereby resulting in the same power consumption. As shown in Figs. 3.12(b) and 3.13(b), the histograms apparently show that PE-CVS saves more power than ER-CVS, up to 170 mW.
3.7 Summary

We have proposed a novel compressive multi-view video coding/decoding architecture - cooperative sparsity-aware independent encoder and independent decoder. We also introduced a central controller to do the sparsity pattern estimation, R-view selection, mean value estimation and implement network optimization algorithms. By introducing limited channel feedback and enabling lightweight sparsity information sharing between R-view and NR-views, the encoders independently encode the video sequences with sparsity awareness and exploit multi-view correlation to improve the reconstruction quality of NR-views. Based on the proposed encoding/decoding architecture, we developed an R-D model that considers the packet loss effect in CS video streaming in WSNs. Then, we studied a modeling framework to design network optimization algorithms, where packet loss rate for a multi-hop multi-path sensor network and the conversion from packet loss rate to the measurement rate reduction are derived. Finally, we presented a power-efficient algorithm. Extensive simulation results showed that the designed compressive multi-view framework can considerably improve the video reconstruction quality with minimal power consumption.
Algorithm 1 Solution Algorithm

Data: Predefine $T_{\text{max}}$, $\{p_{\text{delay}}^z\}$, $\{N_z^z\}$, target distortion $D_t$ and distortion error tolerance $D_e$, total bits $b$, incremental bits $\Delta b$. Set $\{P_z^z\} = 0$, $\{b^z\} = 0$, $D_{\text{dec}} = 0$

Result: Obtain $\{P_z^z\}$ and $\{b^z\}$ when $|D_{\text{dec}} - D_t| \leq D_e$

while true do
  Initialize $P_z^z(0) = 0$, $\{b^z(0)\} = 0$;
  for $t = 1 : b/\Delta b$ do
    Allocate $\{b^z(t)\} = \{b^z(t - 1)\} + \Delta b$ to each path $z$ to calculate $\{P_z^z(t)\}$ for each hop $l \in L_z$;
    Calculate total power consumption for path $z$: $P_z^z(t) = \sum_{l \in L_z} P_z^z(t)$;
    Finally allocate $\Delta b$ to path $m$ satisfying $m = \arg\min_{m \in Z} (P_m^m(t) - P_m^m(t - 1)), \forall m \in Z$;
    Set $b^z(t) = b^z(t - 1)$, $z \neq m, z \in Z$;
    Set $P_z^z(t) = P_z^z(t - 1)$, $z \neq m, z \in Z$;
  end
  Calculate $D_{\text{dec}}$ using (3.30);
  if $|D_{\text{dec}} - D_t| \leq D_e$ then
    Output $\{P_z^z\}$ and $\{b^z\}$;
    break;
  else
    if $(D_{\text{dec}} - D_t) > D_e$ then
      $b = b + \Delta b$;
    end
    if $(D_{\text{dec}} - D_t) < -D_e$ then
      $b = b - \Delta b$;
    end
  end
end
Chapter 4

LiBeam: Throughput-Optimal Cooperative Beamforming for Indoor Visible Light Networks

Indoor visible light communications (VLC) are a promising technology to alleviate the problem of an increasingly overcrowded RF spectrum, especially in unlicensed spectrum bands [63–67]. Unlike RF communications, VLC relies on a substantial portion of unregulated spectrum ranging from 375 THz to 750 THz, providing bandwidth orders of magnitude ($10^4$) wider than the available radio spectrum. In recent years, while there have been significant advances in understanding and designing efficient physical layer techniques (e.g., modulation schemes) [68] [69], the problem of designing optimized strategies to provide high-throughput WiFi-like access through VLC comms in indoor environments is still largely unexplored. To bridge this gap, in this article we focus on downlink indoor scenarios and study techniques to provide VLC-based wireless access to multiple concurrent users with optimized throughput using a set of centrally-controlled partially interfering LEDs.

There are multiple challenges to be addressed to provide high-throughput indoor visible light networking. First, VLC link quality is significantly affected by the imperfect, possibly time-varying, alignment between the communicating devices [70]. Hence, it is difficult to maintain reliable high-quality VLC links. Second, the link quality is degraded by the presence of mutual interference among adjacent partially interfering LEDs. Third, VLC links can easily get blocked because of the inherent low penetration of light. For these reasons, most existing work has focused either on
To address these challenges, in this chapter we propose LiBeam, a new cooperative beamforming scheme for indoor visible light networking. In a nutshell, LiBeam uses multiple LEDs collaboratively to serve the same set of users thus reducing the interference among users and hence enhancing the quality of the visible light links.

**Cooperative Visible Light Beamforming.** VLC systems commonly exploit intensity modulation and direct detection (IM/DD), where an electrical signal is transformed into a real nonnegative waveform that carries no phase information to drive LEDs [63]. As a result, the conventional phase-shift-based RF beamforming techniques cannot be directly applied to VLC systems.

A few recent efforts have been made focused on VLC beamforming [75–77]. For example, Kim et al. propose in [76] time-division multiple access (TDMA) optical beamforming by using a specially-designed optical component, referred to as the spatial light modulator (SLM). In [77], the authors present a multiple-input-single-output (MISO) transmit beamforming system using a uniform circular array (UCA) as transmitter. Ling et al. propose a biased beamforming for multicarrier multi-LED VLC systems in [75]. However, these existing VLC beamforming techniques cannot be directly applied to indoor visible light downlink access networks, because (i) the existing lighting infrastructure is not easily modified by adding some special optical components or custom designed LEDs; (ii) existing beamforming schemes haven’t considered the interference among users, and hence are not suitable for indoor visible light networking with densely-deployed partially interfering LEDs.

In contrast to prior work, in this chapter we propose a new beamforming technique to reduce the effects of interference among users in visible light networks using off-the-shelf LEDs. Specifically, our objective is to control the visible light signals so that they add constructively at the desired receiver if carrying the same information, and add destructively otherwise. Since it is difficult (if not impossible) to directly control the phase of the carrier signal (which is visible light here) as in traditional RF domain, we propose to control the beamforming weights (i.e., the amplitude and initial phase) of the baseband electrical modulating signal, and then use the resulting beamed electrical signal to modulate the visible light signal. Using aforementioned beamforming technique, we then propose LiBeam, a cooperative beamforming scheme for indoor visible-light downlink access network, as shown in Fig. 4.1, based on which the LEDs form multiple clusters, with each
CHAPTER 4. LIBEAM

Figure 4.1: Indoor visible light networking with cooperative beamforming.

cluster serving a subset of the users by jointly determining the LED-user association strategies and the beamforming vectors of each LED cluster.

We claim the following main contributions:

- **Cooperative beamforming.** We formulate mathematically the cooperative beamforming problem with the control objective of maximizing the sum throughput of users in indoor visible-light downlink access networks, by jointly controlling the LED-user association and the beamforming vectors of the LEDs.

- **Globally-optimal solution algorithm.** To solve the resulting mixed integer nonlinear nonconvex programming (MINCoP) problem, we design a globally optimal solution algorithm based on a combination of the branch and bound framework and convex relaxation techniques.

- **Programmable visible light networking testbed.** We design for the first time a programmable indoor visible light networking testbed based on USRP X310 software-defined radios with a custom-designed optical front-end. The testbed consists of three main components: network control host, SDR control host, and VLC hardware and front-ends.
CHAPTER 4. LIBEAM

- Experimental performance evaluation. We experimentally demonstrate the effectiveness of the proposed cooperative beamforming scheme through extensive experiments.

The remainder of the chapter is organized as follows. We review the related work in Section 4.1 and then present the mathematical model of the cooperative beamforming scheme in Section 4.2. The globally optimal solution algorithm is then described in Section 4.3. In Section 4.4 we discuss the design of the programmable visible-light networking testbed. Then, simulation and experimental performance evaluation results are presented in Section 4.5 and finally we draw main conclusions in Section 5.9.

4.1 Related Work

There is a growing body of literature on visible light communications, mainly focusing on designing efficient physical layer techniques (e.g., modulation schemes) [71] [78] [79]. Recently, several results on visible light beamforming [73] [75]-[77] [80] and visible-light communication testbeds [81]-[84] have been presented. For example, [76] proposes a TDMA optical beamforming system based on a special optical component (SLM) to mechanically steer the light beams to the desired user. In [77], the authors propose a new indoor positioning system by adopting a uniform circular array (UCA) LEDs as transmitter to increase positioning accuracy. Ling et al. propose in [75] a beamforming scheme by jointly determining the DC bias of each LED and the beamforming vectors to maximize the sum throughput for OFDM multicarrier VLC system. In [80], a beamforming scheme is proposed to improve the secrecy performance under the assumption that there are multiple LED transmitters and one legitimate user. Most of these approaches are designed for specific application scenarios, without considering a network scenario with mutual interference introduced by multiple densely-deployed LEDs.

On the experimental front, a few platforms have been proposed in recent years for rapid prototyping of VLC communications. In [84], a software-defined single-link VLC platform utilizing WARP is presented. Gavrinca et al. prototype in [83] a USRP-platform-based visible light communication system based on the IEEE 802.15.7 standard. The authors of [81] and [82] present OpenVLC and the improved version OpenVLC1.0 based on Beagle-Bone Black (BBB) board, with the objective of being a starter kit for low-cost and low-data-rate VLC research. Most of these existing testbeds are focused on single-link demonstrations, where a networking perspective is not the core focus. To the best of our knowledge, no large-scale programmable indoor visible-light networking prototypes
have been proposed so far.

4.2 System Model and Problem Formulation

We consider an indoor visible light downlink access network scenario as illustrated in Fig. 4.1, where a set of LED transmitters form multiple clusters and in each cluster LEDs cooperatively transmit signal to the associated user. The set of LED transmitters is denoted as $\mathcal{N}$, with $|\mathcal{N}| = N$ being the number of LED transmitters, and the set of visible-light users is denoted as $\mathcal{U}$, with $|\mathcal{U}| = u$ representing the number of total users in the room. We assume that the LED transmitters are installed on the ceiling at pre-defined locations, straightly facing downwards. We also assume that the information of location, azimuth angle and elevation angle of the users can be obtained by the devices themselves [85]. As shown in Fig. 4.1, the azimuth angle (denoted as $\alpha$) of a vector is the angle between the $x$-axis and the orthogonal projection of the vector onto the $xy$-plane. The elevation angle (denoted as $\varepsilon$) is the angle between the vector and its orthogonal projection onto the $xy$-plane.

**IM/DD Channel.** We consider an intensity modulation and direct detection (IM/DD) model, as illustrated in Fig. 4.2, which is often modeled as a baseband linear system [86] as

$$ Y(t) = RX(t) \otimes h(t) + N(t), \quad (4.1) $$

where $X(t)$ and $Y(t)$ denote the instantaneous input power and the output current, respectively; $R$ represents the detector responsivity; $N(t)$ is channel noise$^2$ and the symbol $\otimes$ denotes the convolution operation. Unlike RF wireless channels, the frequency selectivity of the channel in VLC networks is mostly a consequence of hardware impairments of the transmit/receive devices (e.g., LEDs and PDs) rather than caused by the multipath nature of RF wireless channels. Moreover, the frequency selective characteristics of optical devices is substantially static and independent of the users’ positions or orientations. However, the average received power is much more dynamic and is significantly dependent on the position and orientation of the user devices. Therefore, in this article, we assume that the visible-light channel is frequency non-selective, i.e.,

$$ h(t) = H_0 \delta(t), \quad (4.2) $$

$^2$ $N(t)$ usually follows signal-independent additive Gaussian distribution [87].
CHAPTER 4. LIBEAM

Figure 4.2: (a) Transmission and reception in a visible light link with IM/DD, (b) Geometry LOS propagation model.

where $\delta(\cdot)$ is the dirac delta function and $H_0$ denotes the static gain of the impulse response of the visible-light gain and follows the Lambertian radiation pattern [88], given as

$$H_0 = \begin{cases} \frac{A(m+1)}{2\pi r^2} \cos^m(\theta) T_s(\psi) g(\psi) \cos(\psi) & 0 \leq \psi \leq \Psi, \\ 0 & \text{otherwise}, \end{cases}$$

(4.3)

where $A$ is the physical area of the PD, and $m$ is the Lambertian emission index and is given by the semi-angle $\psi_{1/2}$ at half illuminance power of an LED as $m = \frac{\ln 2}{\ln \cos(\psi_{1/2})}$. As illustrated in Fig. 4.2(b), $r$ is the distance between a transmitter and a receiver, $\theta$ is the irradiance angle, $\psi$ is the incidence angle, and $\Psi$ denotes the field of view of PD. $T_s(\psi)$ and $g(\psi)$ represent the gain of an optical filter and the gain of an optical concentrator [88], respectively. Then, the channel model in (4.1) can be rewritten as

$$Y(t) = R H_0 X(t) + N(t).$$

(4.4)

Orientation- and Location-based Link Status. In visible-light networks, the field of views are limited for both LEDs and visible-light user receivers (i.e., photodetector (PD)). Therefore, LEDs and users may be out-of-FOV from each other, i.e., the transmit-receive link may not exist for some LED-user pairs. Therefore, determining the link status among LED-user pairs is the fundamental step in visible light networking. We denote the location and orientation information for the $n$-th LED transmitter as $P^n = [x^n, y^n, z^n, \alpha^n, \epsilon^n]$, with $1 \leq n \leq N$. Accordingly, the location and orientation information for the $j$-th LED user is denoted as $P^u = [x^u, y^u, z^u, \alpha^u, \epsilon^u]$. 
CHAPTER 4. LIBEAM

with $1 \leq u \leq U$. Since the LEDs are installed on the ceiling and straightly face downwards, the irradiance angle (denoted as $\theta^u_n$) from $n$-th LED to $u$-th user can be calculated as

$$\theta^u_n = \arctan \left( \frac{\| V_{-z} \times V^u_n \|_2}{V^T_{-z} V^u_n} \right),$$

with $V_{-z} = [0, 0, -1]^T$ being the unit norm vector of the $n$-th LED, $V^u_n = [x^n, y^n, z^n]^T - [x^n, y^n, z^n]^T$ representing the vector that points to the $u$-th user from the $n$-th LED transmitter, and $\arctan \left( \cdot \right)$ is the function used to calculate the four-quadrant inverse tangent in degree [89].

Accordingly, the incidence angle $\psi^u_n$ from $n$-th LED to the $u$-th user is calculated as

$$\psi^u_n = 90 - \arctan \left( \frac{\| V_u \times V^u_n \|_2}{V^T_u V^u_n} \right),$$

where $V_u$ is the unit vector of user, calculated based on the obtained orientation information of $u$-th user as $V_u = [\cos(\alpha^u) \cos(\varepsilon^u), \sin(\alpha^u) \cos(\varepsilon^u), \cos(\varepsilon^u)]^T$, and $V^u_n = [x^n, y^n, z^n]^T - [x^n, y^n, z^n]^T$ is the vector pointing to the $n$-th LED from the $u$-th user.

With $\theta^u_n$ and $\psi^u_n$, we then can determine if there exists a transmit-receive link between the $n$-th LED and the $u$-th user, as follows:

$$l_{n,u} = \begin{cases} 1, & \theta^u_n \leq \Theta, \psi^u_n \leq \Psi, \\ 0, & \text{Otherwise}, \end{cases}$$

with $l_{n,u}$ representing the link status between LED $n$ and user $u$, and $\Theta$ and $\Psi$ represent the FOV of LEDs and users, respectively. We denote $l = \{ l_{n,u} \mid 1 \leq n \leq N, \ 1 \leq u \leq U \}$ as the set of the link status between LEDs and users.

**LED-User Association.** In this article, we consider single-guest service for LED transmitters, i.e., each LED can serve at most one user in each cooperative transmission. Denote the LED-user association vector as $\mu = \{ \mu_{n,u} \mid n \in \mathcal{N}, u \in \mathcal{U} \}$, where $\mu_{n,u} = 1$ if LED $n$ is selected to serve user $u$ and a link exists between them, i.e., $l_{n,u} = 1$, and $\mu_{n,u} = 0$ otherwise. Then, we have

$$\mu_{n,u} = \{ 0, 1 \}, \forall n \in \mathcal{N}, \forall u \in \mathcal{U},$$

$$\sum_{n \in \mathcal{N}} \mu_{n,u} = 1, \forall u \in \mathcal{U},$$

$$\mathcal{N}_u \triangleq \{ n \mid n \in \mathcal{N}, \mu_{n,u} = 1 \}, \forall u \in \mathcal{U},$$

$$\mathcal{N}_u^l \triangleq \{ n \mid n \in \mathcal{N}, l_{n,u} = 1 \}, \forall u \in \mathcal{U}. $$
Cooperative Transmission With Beamforming. Denote $d_{n,u}$ as the symbol to be transmitted to the $u$-th user from $n$-th LED. We assume $d_{n,u}$ is zero mean normalized to the range $[-1, 1]$. At the $n$-th LED transmitter, to enable cooperative beamforming, $d_{n,u}$ is multiplied by beamforming weight $w_{n,u}$. Furthermore, to make the resulting input electrical signal positive, a bias $B$ needs to be added to $d_{n,u}w_{n,u}$. Then, we obtain the input electrical signal from LED $n$ to user $u$ as

$$y_{n,u} = d_{n,u}w_{n,u} + B. \quad (4.12)$$

To ensure the nonnegativity of $y_{n,u}$, we need

$$|d_{n,u}w_{n,u}| \leq B, \forall n \in N, \forall u \in U. \quad (4.13)$$

In IM/DD visible-light system, the emitted light intensity is proportional to the input signal. Therefore, without loss of generality, we assume that the emitted light intensity equals the input signal and represented the same as in (4.12).

Light carrying signal propagates from the LED to the user where we only consider the line-of-sight (LOS) propagation path. The channel gain from the $n$-th LED to the $u$-th user is given by

$$h_{n,u} = \begin{cases} \frac{A_u(m+1)}{2\pi r_{n,u}^2} \cos^m(\theta_u^n)T_s(\psi_u^n)g(\psi_u^n) \cos(\psi_u^n) & 0 \leq \psi_u^n \leq \Psi, \\ 0 & \text{otherwise}, \end{cases} \quad (4.14)$$

where $\theta_u^n$ and $\psi_u^n$ denote the incidence and irradiance angles between the $n$-th LED transmitter and user $k$, respectively, and $r_{n,u}$ represents the distance between the $n$-th transmitter and the $u$-th user.

Let $w_u = [w_{1,u}, w_{2,u}, \ldots, w_{N,u}]$ denote the beamforming vector for the $u$-th user, and $w = [w_1, w_2, \ldots, w_U]^{T}$ represent the beamforming weights matrix. Let $h_u = [h_{1,u}, h_{2,u}, \ldots, h_{N,u}]$ denote the channel gain vector for the $u$-th user, and $h = [h_1, h_2, \ldots, h_U]^{T}$ represent the channel matrix. After removing the DC component at the PDs of the users, the received signal at the $u$-th user can be written as

$$r_u = \sum_{n \in N_u} d_{n,u}w_{n,u}h_{n,u} + \sum_{n \in (N_u' \setminus N_u)} d_{n,k}w_{n,k}h_{n,k} + z_u^2,$$

$$= (h_u^H) w_u^H d_u^H + (h_u^H) w_u^H d_u^H + z_u,$$  \quad (4.15)

where the first term $(h_u^H) w_u^H d_u^H$ is the desired signal, the second term $(h_u^H) w_u^H d_u^H$ is the interference from other users, and $z_u$ denotes the power of noise at user $u$. In VLC, $z_u$ is considered to be Gaussian distributed with zero-mean and variance $\sigma_u^2$ \[63\]. The other symbols in (4.15) are defined as
CHAPTER 4. LIBEAM

\[
\begin{align*}
    h_u^\mu &= \mu_u \circ h_u, \quad \forall u \in U, \quad (4.16) \\
    w_u^\mu &= \mu_u \circ w_u, \quad \forall u \in U, \quad (4.17) \\
    d_u^\mu &= \mu_u \circ d_u, \quad \forall u \in U, \quad (4.18) \\
    h_u^l &= (l_u - \mu_u) \circ h_u, \quad \forall u \in U, \quad (4.19) \\
    w_u^l &= (l_u - \mu_u) \circ \sum_{u \in U} w_u^\mu, \quad \forall u \in U, \quad (4.20) \\
    d_u^l &= (l_u - \mu_u) \circ \sum_{u \in U} d_u^\mu, \quad \forall u \in U, \quad (4.21)
\end{align*}
\]

where \( \circ \) represents Hadamard product and \( d_u = [d_{1,u}, d_{2,u}, \ldots, d_{N,u}] \) denotes the transmitted signal vector for the \( u \)-th user.

**Signal-to-Interference-plus-Noise Ratio (SINR).** In indoor visible-light networks, multiple transmissions usually occur concurrently, thus introducing mutual interference at the receiver side. Therefore, the notion of SINR is adopted in this work to measure the signal quality at the user end. Denote \( \gamma_u \) as the SINR for user \( u \), then it can be given as

\[
\gamma_u = \frac{B^2 (h_u^\mu)^T w_u^\mu (w_u^\mu)^T h_u^\mu}{z_u + B^2 (h_u^l)^T w_u^l (w_u^l)^T h_u^l}. \quad (4.22)
\]

**Problem Statement.** The network control objective can be stated as maximizing the sum utility of indoor visible-light downlink access network by jointly considering the position and orientation, FOVs of the LED transmitters and users, the LED-user association vectors, as well as the beamforming vectors for cooperative transmission and interference cancellation, subject to the following constraints:

- **Signal amplitude constraints:** To ensure the nonnegativity of the electrical signal input to the LEDs and to maintain linear current-to-light conversion, the amplitude of the transmitted signal is constrained as (4.13).

- **Beamforming weight coefficients:** To avoid violating the constraints of the modulated signal amplitude, when introducing beamforming weights, the following constraints should be satisfied:

\[
\begin{align*}
    |w_u^\mu| &\leq B, \quad (4.23) \\
    |w_u^l| &\leq B. \quad (4.24)
\end{align*}
\]
Define \( l = \{ l_{n,u} | n \in \mathcal{N}, u \in \mathcal{U} \} \) as the link status with respect to position, orientation and FOV of LEDs and users. Denote \( \mu = \{ \mu_{n,u} | n \in \mathcal{N}, u \in \mathcal{U} \} \) and \( w = \{ w_{n,u} | n \in \mathcal{N}, u \in \mathcal{U} \} \) as LED-user association and the beamforming vectors, respectively. Further define \( P_N = [P^1, P^2, \ldots, P^n] \) and \( P_U = [P^1, P^2, \ldots, P^U] \) as the location and orientation information of LEDs and users. The network control problem can then be formulated as

**Problem 1:** Given: \( \Gamma, P_N, P_U, \Theta, \Psi, l \)

Maximize \[ f(\mu, w) = \sum_{u \in \mathcal{U}} R_u(\mu, w) \] \hspace{1cm} (4.25)

Subject to: \( (4.8), (4.9), (4.13), (4.16) \sim (4.21), (4.23), (4.24) \).

with \( R_u = \log_2(1 + \gamma_u) \) representing the achievable throughput of user \( u \).

### 4.3 Globally Optimal Solution Algorithm

As stated in Sec. 4.2, the social objective of the indoor multi-user visible-light network control problem is to maximize the sum throughput of the users by jointly controlling LED-user association strategies and the cooperative beamforming vectors, as presented in Problem 1. In (4.25), the individual SINR \( \gamma_u \) is a nonconvex function with respect to LED-user association vector \( \mu \) and the beamforming vectors \( w \). Moreover, the LED-user association variable \( \mu \) can only take binary values. Therefore, the resulting network control problem is a mixed nonlinear nonconvex programming (MINCoP) problem, for which there is in general no existing solution algorithm that can be used to obtain the global optimum in polynomial computational complexity. In this chapter, we design a globally optimal solution algorithm based on a combination of the branch and bound method and of convex relaxation techniques [90] [91].

#### 4.3.1 Overview of The Algorithm

The objective of the proposed algorithm is to solve the MINCoP formulated in Problem 1 by exploiting branch-and-bound framework [92]. With this approach, we aim to search for an \( \epsilon \)-optimal solution, with \( \epsilon \in (0, 1] \) being the predefined optimality precision that can be set as close to 1 as we wish. Denote \( Q_0 = \{ \mu, w | \text{constraints in (4.25)} \} \) as the feasible set of the initial problem (4.25), and \( U^*(Q_0) \) as the global optimum of problem (4.25) over \( Q_0 \), then our objective is to search iteratively for \( U \) so that \( U(Q_0) \geq \epsilon U^*(Q_0) \).
To this end, the algorithm maintains a set $Q = \{Q_i, i = 0, 1, 2, \ldots\}$ of subproblems by iteratively partitioning feasible set $Q_0$ into a series of smaller subsets $Q_i$. During the iterations, the algorithm also maintains a global upper bound $\bar{U}_{\text{glb}}(Q_0)$ and a global lower bound $\underline{U}_{\text{glb}}(Q_0)$ on $U^*(Q_0)$ so that

$$\underline{U}_{\text{glb}}(Q_0) \leq U^*(Q_0) \leq \bar{U}_{\text{glb}}(Q_0).$$

(4.26)

The global upper and lower bounds are updated as follows:

$$\bar{U}_{\text{glb}}(Q_0) = \max\{\bar{U}_{\text{glb}}(Q_i), i = 1, 2, \ldots\},$$

(4.27)

$$\underline{U}_{\text{glb}}(Q_0) = \max\{\underline{U}_{\text{glb}}(Q_i), i = 1, 2, \ldots\}.$$  

(4.28)

Then, if $\underline{U}_{\text{glb}}(Q_0) \geq \epsilon \bar{U}_{\text{glb}}(Q_0)$, it indicates that the predefined optimality precision $\epsilon$ is achieved, and then the algorithm terminates and sets the optimal sum rate to $U^*(Q_0) = \underline{U}_{\text{glb}}(Q_0)$. Otherwise, the algorithm chooses a sub-domain from $Q$ and partition it into two sub-domains. In our algorithm, we select sub-domain $Q_i$ with the highest local upper bound, i.e., $i = \arg\max_i \bar{U}_{\text{glb}}(Q_i)$. Based on the global bounds update criterion in (4.27) and (4.28), the gap between the two global bounds converges to 0 as the partition progresses. Furthermore, from (4.26), $\underline{U}_{\text{glb}}(Q_0)$ and $\bar{U}_{\text{glb}}(Q_0)$ converge to the global optimum $U^*(Q_0)$.

### 4.3.2 Convex Relaxation

Because the problem formulated in Sec. 4.2 is nonconvex, a key step in the algorithm described above is to obtain a relaxed but convex version of the original problem (4.25) and the subproblems resulting from the partition, so that a tight local upper bound $\bar{U}_{\text{glb}}(Q_i)$ can be easily computed for each of them. To this end, we first relax the LED-user association variables $\mu_{n,u}$, $n \in \mathcal{N}$, $u \in \mathcal{U}$ in (4.25), which take binary values only, by allowing each LED to serve multiple user nodes. Then the constraint in (4.25) can be rewritten as

$$0 \leq \mu_{n,u} \leq 1 \quad \forall n \in \mathcal{N}, \forall u \in \mathcal{U},$$

(4.29)

and the individual throughput $R_u$ in problem (4.25) can be further expressed as
According to composition rule (i.e., composition operations preserve convexity) in convex optimization \cite{93}, the first and second parts (including the minus sign) in (4.30) are convex and concave, respectively. Therefore, a convex relaxation of (4.30) can be obtained by approximating the logarithm operation in the concave part of (4.30) using a set of linear functions. To this end, we first replace \( z_u + B^2(h_u^T w_u)(w_u^T h_u^T) \) in the second part of (4.30) with \( t \), then \( \log_2(z_u + B^2(h_u^T w_u)(w_u^T h_u^T)) \) in (4.30) can be represented as \( \log_2(t) \) subject to \( t \geq (z_u + B^2(h_u^T w_u)(w_u^T h_u^T)) \). Then \( \log_2(t) \) can be further relaxed using a segment and three tangent lines \cite{93}. 

\[
R_u = \log_2(1 + \gamma_u) \\
= \log_2\left(1 + \frac{B^2(h_u^T w_u)(w_u^T h_u^T)}{z_u + B^2(h_u^T w_u)(w_u^T h_u^T)}\right) \\
= \log_2\left(\frac{z_u + B^2(h_u^T w_u)(w_u^T h_u^T)}{z_u + B^2(h_u^T w_u)(w_u^T h_u^T)}\right) \\
= \log_2(z_u + B^2(h_u^T w_u)(w_u^T h_u^T)) \\
\text{subject to } t \geq (z_u + B^2(h_u^T w_u)(w_u^T h_u^T)) \\
\]
CHAPTER 4. LIBEAM

Then the original MINCoP problem in (4.25) can be reformulated as a convex problem as

Problem 2: Given: \( \Pi, \mathbf{P}_N, \mathbf{P}_{I}, \Theta, \Psi, I \)

Maximize \( f = \sum_{u \in I} R_{ua}(\mu, w) \), \hspace{1cm} (4.35)

Subject to: \( (4.9), (4.13), (4.16) \sim (4.21), (4.23), (4.24), (4.29) \)

with \( R_{ua} \) representing the relaxed convex version of \( R_u \) in (4.25). As variable partition progresses, the association variable \( \mu_{n,u} \) becomes fixed to either 0 or 1 in all subproblems, for which the optimal beamforming weights \( w \) can be obtained by solving a convex programming problem (4.35).

4.3.3 Variable Partition

Variable partition can be conducted by partitioning association variable \( \mu \) and the beamforming variables \( w \). For example, given a subproblem \( Q_i \), by fixing association variable \( \mu_{n,u} \) subproblem \( Q_i \) can be partitioned into two subproblems with feasible set \( Q_{i,1} = \{(\mu, w) \in Q_i | \mu_{n,u} = 0\} \) and \( Q_{i,2} = \{(\mu, w) \in Q_i | \mu_{n,u} = 1\} \), respectively. For the beamforming vectors, say \( w_{n,u} \in [w_{n,u}^{\text{min}}, w_{n,u}^{\text{max}}] \) for LED \( n \) to user \( u \), the partition can be conducted by splitting \( w_{n,u} \) from the half, resulting in two subproblems with feasible sets

\[
Q_{i,1} = \{(u, w) \in Q_i | w_{n,u} \in [w_{n,u}^{\text{min}}, w_{n,u}^{\text{mid}}]\}, \hspace{1cm} (4.36)
\]

\[
Q_{i,2} = \{(u, w) \in Q_i | w_{n,u} \in [w_{n,u}^{\text{mid}}, w_{n,u}^{\text{max}}]\}, \hspace{1cm} (4.37)
\]

where \( w_{n,u}^{\text{mid}} \triangleq \frac{w_{n,u}^{\text{min}} + w_{n,u}^{\text{max}}}{2} \).

4.4 Testbed Development

As discussed in Sec. 4.1 most of existing visible-light testbeds are focused on single-link implementation. To the best of our knowledge, we design for the first-time a large programmable indoor visible-light networking prototype, which can support arbitrary \( N \) nodes.

Overall Diagram. The prototyping diagram is illustrated in Fig. 4.3 following a hierarchical architecture with three tiers, i.e., network control host, SDR control host and VLC hardware and front-ends. At the top tier of the hierarchical architecture is the network control host, where the designed optimization solution algorithms are executed. The output of this tier is a set of optimal variables, which will then be sent to each of the SDR control hosts. At the second tier, the
programmable protocol stack (PPS) is installed on each of the SDR control hosts. With the optimal variables received from the network control host, the PPS will be compiled to generate operational code to control at network run time the VLC hardware and front-ends of the third tier. Finally, each of the VLC hardware and front-ends (i.e., USRP) receives the baseband samples from its control host via Gigabit Ethernet (GigE) interface and then sends them over the air with transmission parameters specified in the control commands from the SDR control hosts.

**Network Control Host.** The network control host is a Dell OPTIPLEX 9020 desktop running Windows 10 pro. On the host the networking optimization algorithms designed in Sec. 4.3 are executed to solve the cooperative beamforming problem formulated in (4.25). The output of the algorithms is the optimized LED-user association vector and beamforming vectors.

**SDR Control Host.** As shown in Fig. 4.3, the programmable protocol stack (PPS) is installed on each of the SDR control hosts, which are Dell XPS running Ubuntu 16.04. The PPS has been developed in Python on top of GNU Radio to provide seamless controls of USRPs. The developed PPS covers PHY and link layers currently, and can be easily extended to upper layers in future. As illustrated in Fig. 4.4, the architecture of the LiBeam node has been developed based on PPS to verify the effectiveness of the designed visible-light networking prototype. At the physical layer, a wide set of modulation schemes can be supported, including On-Off Keying (OOK), Gaussian minimum-shift keying (GMSK), binary phase-shift keying (BPSK), among others. The programmable parameters at this layer include modulation schemes, transmission power, and beamforming weights, among others. At the link layer, besides fragmentation/defragmentation, network-to-physical address translation, reliable point-to-point frame delivery, cooperative transmitter access control and LED cluster formation are particularly designed for LiBeam.
VLC Hardware and Front-ends. The hardware components of each LiBeam node and the snapshot of the LiBeam testbed are illustrated in Fig. 4.5. The LiBeam testbed is designed based on USRP X310 software-defined radios. The motherboard of each USRP X310 has four wideband daughterboard slots that support bandwidth of up to 120 MHz within DC - 6 GHz frequency. We currently use two slots of the motherboard to accommodate LFTX and LFRX daughterboards for visible light signal transmission and reception, while the remaining two slots are reserved for future extension, for example, RF/VLC coexistence prototype, MIMO VLC implementation.

At the transmitter side, we use a Bivar L2-MLW1-F LED with 125° field of view (FOV). We build an transconductance amplifier based LED driver from scratch to drive the LED, which mainly consists of a bias-T and a RF NPN transistor. The bias-T is used to combined the modulated AC waveform from USRP X310 and the DC bias that meets the minimum voltage requirement to light up the LED.

At the receiver side, we use Thorlabs PDA36A with FOV 90°, which can detect light with
CHAPTER 4. LIBEAM

wavelength ranging from 350 to 1100 nm. PDA36A features a built-in low-noise transimpedance amplifier (TIA) with switchable gain and it can support bandwidth from DC to 12 MHz. The PDA36A consequently converts the received photons into real-valued digital samples and then sends them to the SDR control host for post-processing.

4.5 Performance Evaluation

In this section, we first evaluate the proposed solution algorithm through simulations, and then we further validate experimentally the effectiveness of LiBeam over the designed prototype through testbed experiments.

4.5.1 Simulation Results

We first evaluate the performance of the solution algorithm proposed in Sec. 4.3 by considering an indoor area of $5 \times 5 \times 5$ m$^3$, where $N = \{3, 4, \ldots, 9\}$ LEDs serve $U = \{2, 3, 4, 5\}$ visible-light users. The altitude of the LEDs are set to 5 meters, emulating scenarios where all LEDs are mounted on the ceiling, straightly facing downwards. The FOVs of LED and user PD are both set to $2/3\pi$. The PD’s physical area and responsivity are $10^{-5}$ m$^2$ and 0.5 A/W, respectively. The average noise power is set to $6.4640e^{-17}$ W. Results are obtained by randomly generating network topologies with a given number of LEDs and users, i.e., positions of LEDs, positions and orientations of users.

Figure 4.6 shows the convergence of the proposed solution algorithm with 3-LED 2-user and 5-LED 2-user scenarios. It can be seen that the proposed algorithm can converge very fast to the global optimum of the MINCoP problem formulated in (4.25), in around 70 and 90 iterations in Figs. 4.6a and (b), respectively.

In Fig. 4.7 we then compare the performance with respect to the network spectral efficiency of the proposed solution algorithm (aka, Joint Optimization) with other two strategies, i.e., w/o Association and Greedy. In w/o Association, the LED-user association is randomly generated. And in Greedy, the LED-user association is determined according to the best channel gain rule and the selected LED transmitting with maximum power. It can be seen that the joint network control achieves the highest spectral efficiency in almost all of the tested network topologies. When the randomly generated LED-user association of w/o Association strategy is occasionally the same as the Joint Optimization scheme, they will achieve the same network spectral efficiency. Results also
CHAPTER 4. LIBEAM

Figure 4.6: Global upper and lower bounds of the globally optimal solution algorithm for network topology with (a) 3 LEDs and 2 users and (b) 5 LEDs and 4 users.

Table 4.1: Network Scenario 1

<table>
<thead>
<tr>
<th>Number Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LED position (m)</td>
<td>(5, 0, 0)</td>
<td>(5, 1, 0)</td>
<td>(5, 1.5, 0)</td>
<td>(5, 3, 0)</td>
</tr>
<tr>
<td>User position 1 (m)</td>
<td>(3, 1, 0)</td>
<td>(3, 3.5, 0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User position 2 (m)</td>
<td>(3, 1, 0)</td>
<td>(3, 2, 0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

show that when the LED-user association generated by Greedy is better than that of w/o Association, Greedy can slightly outperform w/o association, for example in network topology instance 13. To make the result clearer, Fig. 4.8 shows the increase of the network spectrum efficiency achievable by Joint Optimization compared to w/o Association and Greedy. We can clearly see that the proposed Joint Optimization algorithm outperforms the other two strategies, particularly the Greedy strategy.

4.5.2 Experimental Evaluation

As shown in Fig. 4.5, we set up the experimental testbed by using the software-defined programmable visible light networking node introduced in Sec. 4.4 to validate the proposed cooperative beamforming solution algorithm in indoor visible light networks. We designed two different
networking scenarios (i.e., 4 LEDs 2 users and 4 LEDs 3 users) as shown in Tables 4.1 and 4.2, respectively. In each network scenario, two different user position sets are used, where users in the first set are more densely deployed than in the second set. Without loss of generality, users’ PDs straightly face towards the plain where LEDs located, with the azimuth and elevation angles being $\varepsilon = 90^\circ$ and $\alpha = 90^\circ$, respectively. Due to the limited bandwidth of the LED, 40 kHz bandwidth is set for each USRP. After modulation, the data is sampled at sampling rate of 800 kHz. The communication range in the experiments is set to 5 m. According to the specifications of the hardware components used in the experiments, the FOVs of LED and PD are $125^\circ$ and $90^\circ$, respectively. The PD’s active physical area is $1.3 \times 10^{-5}$ m$^2$.

Before conducting the experiments, we first test the visible-light instantaneous channel response by using GoldSequence preamble. The results are shown in Fig. 4.9 obtained by sending
10000 preambles. We can see that the visible-light channel is almost stable once the position of the
LED and user as well as the corresponding optical parameters (e.g., PD active area, orientations of
LEDs and PDs) are fixed, which is also satisfied the channel model presented in Sec. 4.2.

We then test the effectiveness of the proposed Joint Optimization algorithm in terms of
sum utility, by comparing it to the other two suboptimal network control strategies: w/o Association
and Greedy algorithms. Figures 4.10 and 4.11 report the average end-to-end throughput (in terms
of packets/s) achievable in the two tested network scenarios. The packet length in the experiments
is set to 1500 bits. We observe that the proposed joint optimization method outperforms the other
two methods in most of the tested instances, and up to 95.9% sum utility gain can be achieved in
network scenario 2. In Fig. 4.10, for the second user position set, Joint Optimization achieves the
same performance as w/o Association. This is because the w/o Association method may randomly
select the same LED-user association as Joint Optimization. Figures 4.10 and 4.11 also show that
more-densely-deployed users would suffer from severer mutual interference, resulting in lower
average sum utility compared to the cases where users are deployed farther away from each other,
especially with the Greedy method. This is because, with the Greedy algorithm, the transmitter with
the best channel gain will be selected with the maximum power to transmit data to the desired user,
thus resulting in higher interference to other users, especially when users are closer to each other. As a result, no packet can be successfully delivered with the Greedy method in the second test instance of the two network scenarios.

Figure 4.12 provides a closer look at the contrasting behaviors in terms of the corresponding instantaneous throughput resulting from Joint Optimization, w/o Association and Greedy for the first user position set in network scenarios 1 and 2, respectively. It can be seen from Figs. 4.12(a) and (b) that, the instantaneous throughput obtained from these three methods are stable at some level, without or with little fluctuations only. These results are consistent with the observations in Fig. 4.9 where the instantaneous channel response is almost stable. We can also see that the proposed Joint Optimization method always outperforms the other two methods in terms of instantaneous throughput in real-time running experiments.

4.6 Summary

We have proposed LiBeam, a new cooperative beamforming approach for indoor visible light networks with the objective of maximizing the sum throughput of the VLC users by jointly
determining the user-LED association strategies and the beamforming vectors of the LEDs. We mathematically formulated the cooperative beamforming problem and a globally optimal solution algorithm has been designed to solve the problem. A programmable visible light networking testbed has also been developed, on which the effectiveness of the proposed LiBeam was validated through extensive simulation as well as experimental performance evaluation.
CHAPTER 4. LIBEAM

Figure 4.11: Average sum utility of network scenario 2.

Figure 4.12: Instantaneous throughput comparison for the first user position set of (a) network scenario 1 and (b) network scenario 2.
Chapter 5

LANET: Visible-Light Ad Hoc Networks

The proliferation of advanced multimedia devices and services is causing significant growth in demand for bandwidth and spectrum resources. While new portions of the radio frequency (RF) electromagnetic spectrum are being made available and are increasingly leveraged to meet this demand, RF communications inevitably suffer from problems including spectrum crunch, co-channel interference, vulnerability to eavesdroppers, among others [94][95]. Moreover, RF-based communications are not always permitted because of the potential dangerous effect of Electromagnetic Interference (EMI), which occurs when an external device generates radiations that affect electrical circuits through electromagnetic induction, electrostatic coupling, or conduction. For example, cellular and WiFi emissions are prohibited in airplanes during takeoff and landing because electromagnetic radiations can interfere with onboard radios and radars; electronic equipment can emit unintentional signals that allow eavesdroppers to reconstruct processed data at a distance by means of directional antennas and wideband receivers.

Optical communications have attracted significant attention as a valid alternative over legacy RF-based wireless communications. Optical communications are classified in two main categories, fiber-based and optical wireless communications (OWCs). Fiber-based systems are frequently employed in the backbone network cabling because of their robustness, reliability and high-rate in delivering large amounts of data. OWCs are rapidly growing in popularity as an emerging and promising wireless technology capable of high speed data transfer over short distances [96][97]. An optical wireless-based system relies on optical radiation to deliver information in free space, with wavelengths included in the Infra-red (IR), visible-light, and ultraviolet (UV) bands. In the last decades, OWCs have been deployed in medium to long communication distance environments, e.g., OWC has been applied for inter-chip connection as short-range transmission...
while visible light communication (VLC) found applications in medium-range indoor wireless access. Moreover, inter-building connections can be established using IR communications whereas ultraviolet communications (UVCs) have been recently adopted in outdoor non-line-of-sight scenarios and specifically for ad-hoc and wireless sensor networks (WSNs). Recently, satellite communications and deep-space applications based on OWC have been demonstrated, especially for military applications [98]. In particular, the recent rapid increase in the use of LEDs for lightning has paved the way for the development of new communication systems based on leveraging visible light as a communication medium. That is, LEDs can act as illumination devices as well as information transmitters at the same time, thus delivering data by digitally modulating the emitted light beam intensity at a very fast rate [99]. In this article, we discuss challenges, basic principles, state-of-the-art, open research directions and possible solutions in the design of Visible-Light Ad Hoc Networks (LANETs), i.e., infrastructure-less (e.g., sensor, ad hoc) wireless networks based on visible light links.

A few survey papers on optical and visible light communications (VLCs) [100–107] have appeared in the past few years, mainly focused on physical and link layers or specific VLC applications. For example, in [100] Karunatilaka et al. discuss physical layer techniques to enhance the performance of LED-based indoor VLC systems, including modulation schemes and circuit design, among others. In [101], the authors survey existing VLC channel models and provide insights on the theoretical basis for VLC system design. In [102], Yoo et al. discuss existing VLC-based positioning systems, while in [103] the authors focus on VLC receiver design for automotive applications. In [108], Tsonev et al. survey the development of Li-Fi systems in cellular networks utilizing OFDM as well as link-layer schemes. Similarly, in [104] the authors review transmitter and receiver design for visible light communication systems, physical layer techniques, medium access techniques and visible light indoor applications (e.g., indoor localization, gesture recognition, among others). This article differs from the above-mentioned papers in the following ways: (i) we mainly focus on visible light ad hoc networking, which is substantially unexplored; (ii) we provide a comprehensive review of protocol design at all layers of the networking protocol stack; (iii) we discuss challenges and applications for visible light ad hoc networks; (iv) we discuss a potential software-defined visible light ad hoc network (LANET) architecture and discuss possible solutions to implement each component.

The rest of this paper is organized as follows. In Section 5.1, we provide a high-level comparison between LANETs and traditional MANETs, and highlight major factors that need to be re-considered in LANET design, and then discuss enabled applications in Section 5.2. In Section 5.3, we discuss available hardware devices and technologies that can be used to build LANETs, and
then present the overall architecture of LANET and discuss possible design challenges. Through Sections 5.4-5.8, we discuss the state of the art in VLC-based networking and highlight possible open research issues in LANET design following a layered approach, from physical layer up to transport layer. We finally draw conclusions in Section 5.9.

5.1 LANET: Visible-Light Ad Hoc Networks

Visible light ad hoc networks (LANETs) refer to *infrastructure-less mobile ad hoc networks* where LANET nodes are wirelessly connected using single-/multi-hop visible light links, configure their protocol stacks in a cross-layer, online and software-defined manner, and adapt to various networking environments (e.g., air/ground/underwater) by switching among different frontend transceiver devices. Two examples of LANETs are illustrated in Fig. 5.1 for civilian (e.g., Internet of Things,
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

environmental sensing, vehicular communications, smart homes, disaster rescue operations, among others) and military applications [66], respectively. In this section we discuss major challenges in the design of LANETs, as well as the main characteristics of LANETs by comparing it with traditional RF-based wireless networks.

5.1.1 Main Design Challenges

Optical wireless communications, particularly visible light spectrum, has found many applications in short-, medium-, as well as long-range communications in the last decade. These include inter-chip connections, indoor wireless access, as well as satellite and deep-space applications, among others [98, 105]. However, while there has been significant advancement in understanding efficient physical layer design for visible-light point-to-point links, the core problem of developing efficient networking technology specialized for visible-light networks is substantially unaddressed. One of the main challenges is that VLC relies on optical radiations to deliver information in free space through a substantial portion of unregulated spectrum between 400 and 800 THz, with corresponding wavelengths in the Infra-Red (IR), visible light, and Ultraviolet (UV) bands [105]. This makes VLC substantially different from RF-based communications in terms of communication range, transmission alignment and shadowing effect, ambient light interference and receiver noise, and VLC ad hoc networking, among others.

Short Communication Range. Because of the limited propagation range of short-wavelength signals, the transmission range of VLC is relatively short (typically a few meters), compared to RF propagation distances ranging from tens of meters (WiFi) to kilometers (LoRa) [100, 104]. When increasing the link distance, for a given desired level of reliability the achievable data rate decays sharply, thus limiting the number of applications where VLC high data rate transmissions can be employed.

Transmission Alignment and Shadowing Effect. Because of the low penetration of light, while visible light signals in adjacent rooms do not interfere with each other, this also presents several limitations. First, the transmitter and the receiver must be aligned to each other, especially for line of sight (LOS) short distance communications with small field of views (FOVs), and this is challenging especially if LANET nodes are moving [70]. Second, VLC link quality can be significantly degraded because of shadowing effects caused by obstructing objects, e.g., mobile human bodies [109].

Ambient Light Interference And Receiver Noise. Noise and interference in VLC are mainly caused by exposure of the receiver to direct sunlight and by the presence of other sources of
illuminations (i.e., other LED sources, fluorescent and bulb lamps) that cause shot noise and consequently decrease the Signal-to-Noise Ratio (SNR). In turn, the receiver can be affected by thermal noise caused by the pre-amplification chain.

**Lack of Well-established Channel Models.** Factors that affect the performance of visible light links include *free space loss, absorption, scattering, scintillation noise induced by atmospheric turbulence*[^1] and alignment between transmitters and receivers, among others. Different from RF, channel modeling for visible light links is still largely based on preliminary empirical measurements, especially for outdoor non-line-of-sight (NLOS) environments. The applicability of existing theoretical channel models in the design of LANETs still needs to be verified and tested in different transmission media.

**VLC Ad Hoc Networking.** Existing work on VLC mostly focuses on increasing the data rate for a single VLC link using advanced modulation schemes. However, VLC ad hoc networking with a large number of densely co-located VLC links (i.e., LANETs) is still substantially unexplored because of the unique characteristics of VLC, including intense modulation/direct detection (IM/DD) channel model, FOV based directionality, low-penetration, among others. To the best of our knowledge, there are no existing architectures and protocols designed specifically for LANETs.

[^1]: Scintillation noise induced by atmospheric turbulence will affect the performance of outdoor VLC-based applications, such as free-space tactical field applications, ad hoc vehicular communications, disaster rescue applications, among others.

<table>
<thead>
<tr>
<th>Property</th>
<th>MANET</th>
<th>LANET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Consumption</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>Regulated, Limited</td>
<td>Unlimited (400nm ~ 700nm)</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Access Point</td>
<td>Illumination/Signaling LED</td>
</tr>
<tr>
<td>EMI</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Security</td>
<td>Reduced</td>
<td>Higher</td>
</tr>
<tr>
<td>Mobility</td>
<td>High</td>
<td>Reduced</td>
</tr>
<tr>
<td>Line of Sight</td>
<td>Not required</td>
<td>Strictly required</td>
</tr>
<tr>
<td>Technology</td>
<td>Mature</td>
<td>Early stage</td>
</tr>
<tr>
<td>Coverage - Range</td>
<td>Medium - Long</td>
<td>Narrow - Short</td>
</tr>
</tbody>
</table>

Table 5.1: Comparison between LANETs and MANETs.
5.1.2 LANETs vs Traditional MANETs

Similar to traditional RF-based MANETs, LANETs also have the ability to self-organize, self-heal, and self-configure. Because of the unique characteristics of visible light compared to RF signals, in LANETs visible light point-to-point links require mutual alignment of transmitters and receivers given the directivity of light signal propagation, which is not easy to obtain with mobile nodes; communication links in LANETs can be easily interrupted by intermittent blockage since light does not propagate through opaque materials. In Table 5.1, we summarize the differences between LANETs and MANETs, in terms of critical aspects including transmitter and receiver, spectrum regulation, network capacity, spatial reuse, security and costs.

- **Transmitter and Receiver.** In MANETs, the front-end components of each node are typically antenna-based, operating at high frequency. In contrast, simple LED luminaires and photodetectors (PDs) or imaging sensors are typically adopted as transmitters and receivers in LANET. They are relatively simple and inexpensive devices that operate in the baseband and do not require frequency or sophisticated algorithms for the correction of radio frequency impairments, e.g., phase noise and IQ imbalance \[105\]. As a consequence, SWaP (size, weight, and power)\[footnote:2\] and cost of front-end components involved in LANET systems are often lower than equivalent MANET systems.

- **Spectrum Regulation.** The visible light spectrum is mostly unused for delivering information, which implies potential high throughput and an opportunity to alleviate spectrum congestion, particularly evident in the Industrial, Scientific and Medical (ISM) band. The bandwidth available in the visible light portion of the electromagnetic spectrum is considerably larger than the radio frequency bandwidth, which ranges from 3 kHz to 300 GHz. The availability of this mostly unused portion of spectrum provides the opportunity to achieve high data rates through low-cost multi-user broadband communication systems. VLC solutions could be complementary to traditional RF systems and alleviate the spectrum congestion that especially impacts the ISM band.

- **Network Capacity.** In MANETs, all the nodes usually operate in a shared wireless channel with a single radio at each node, where the number of channels, the operating frequency,
and maximum transmit power are stringently regulated\cite{118}, and consequently the network capacity is unavoidably limited and affected by co-located networks. LANETs, instead, can rely on a substantial portion of unlicensed and currently unregulated spectrum as described above, which have the potential to make significant capacity available for networked operations.

• **Spatial Reuse.** Visible light cannot pass through opaque objects, thus resulting in low penetration. Moreover, in contrast to omnidirectional RF communications, because of predefined limited field of view (FOV) of LEDs, visible light links are typically directional. This provides a higher degree of spatial reuse with respect to omnidirectional transmissions typically used in RF. For example, since light cannot propagate outside of a closed room, there is no interference from VLC signals in adjacent rooms. Because of this unique characteristic of VLC, most existing MAC and network layer MANET protocols cannot be directly applied to LANETs and hence need to be redesigned, including neighbor discovery and route selection, among others.

• **Security.** Since they operate in dynamic distributed infrastructure-less configurations without centralized control, MANETs are vulnerable to various kinds of attacks, ranging from passive attacks such as eavesdropping to active attack such as jamming\cite{119}. Differently, in LANETs, the inherent security property that stems from the spatial confinement (low penetration and restricted FOVs) of light beams, will enable secure communications since jammers or eavesdroppers can be easily spotted than in legacy RF communication.

• **Costs.** As discussed above, LANETs are more cost-efficient than MANETs because of much simpler front-end devices (e.g., LEDs, PDs) compared to RF solutions for transmitting, sampling and data processing. Moreover, nodes in MANETs are usually battery-powered to enable communications in the absence of a fixed infrastructure. The sensing unit, the digital processing unit and the radio transceiver unit are the main consumers of the battery energy, and therefore more sophisticated energy-efficient algorithms, e.g., energy-efficient MAC or routing schemes\cite{120} \cite{121}, are needed, which are however challenging in such resource-limited and infrastructure-less MANETs. Differently, LEDs used as transmitters in LANETs highlight themselves by high energy efficiency, longevity, and environment-friendly factor enabled by recent tremendous advances in LED technologies\cite{105}. Moreover, VLC manifests its low-power baseband processing property, which further results in low-cost LED devices compared to high-frequency passband RF front-end antennas.
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

5.2 Envisioned Applications

LANETs have a great potential for enabling a rich set of new civilian and military applications, as illustrated in Fig. 5.1, ranging from low-latency high-bandwidth indoor communications and outdoor intelligent transportation networking, to highly secure Lower Probability of Intercept/Lower Probability of Detection (LPI/LPD) operations under high network density and jamming conditions, among others. Just name a few examples in the following.

- **Intelligent Transport Systems.** One of the most promising outdoor applications of LANETs is for ad hoc vehicular communications [122][123], including Vehicle to Infrastructure (V2I), Infrastructure to Vehicle (I2V) and Vehicle to Vehicle (V2V) communications. LANETs can be employed to design intelligent transport systems with better road safety. For V2V, a communication link can be established using head and tail lights or photo-diodes and image sensors at the receiver side, while for V2I the urban infrastructures (e.g., traffic lights, street lights) can be utilized for transmitting useful information related to current circulation of traffic including vehicle safety, traffic information broadcast and accident signalling. Additionally, in vehicular ad hoc networks (VANET), the network topology is highly dynamic and often large-scale. This makes realizing visible-light VANETs more challenging because of the limited FoV, and the relatively short transmission ranges [124]. Moreover, different from legacy RF-VANETs, the quality of visible links can be significantly degraded by weather conditions, including fog and rain, among others.

- **Internet of Things.** The vision of Internet of Things (IoTs) anticipates that large amounts of mobile embedded devices and/or low-cost resource-constrained sensors will communicate with each other via the Internet. To allow networking among a massive number of devices, the communication system must be ubiquitous, low-cost, and bandwidth and energy efficient. Infrastructureless LANETs are a promising choice for communication in the Internet of Things because of its inherent advantages as discussed in Section 5.1.2 e.g., orders of magnitude available bandwidth, reusing ubiquitously existing lighting infrastructure, low-cost front-end devices, among others. Therefore, LANETs can easily enable a wide range of IoT services, such as localization, smart home, smart city, air/land/sea defense, among others.

- **D2D Communications.** Device to Device (D2D) communications are rapidly emerge in recent years [125]. Beyond the crowded RF spectrum, LANETs are a promising candidate to support D2D communications. VLC-D2D applications [126] can use LEDs and PDs or LCD screens
and camera sensors. The ubiquitous presence of LCD screens and surveillance cameras in urban environments creates numerous opportunities for practical D2D applications since information can for example be encoded in display screens while camera sensors can record and decode data using image processing techniques [127].

- Indoor Positioning. Recently proposed VLC-based indoor localization schemes have shown improved performance, in terms of accuracy, given the higher density of LEDs as compared to Wi-Fi access points [128] [129] [130]. To set up a light-weight indoor positioning network, LANET-enabled sensors can be organized to form an ad hoc network with a tree-like structure (i.e., having a sensor connected to a LAN as the root node) and a simplified protocol stack only providing basic data transfer and routing functionalities that can be run on devices with limited resources. The authors in [130] design a VLC-based indoor positioning system, aiming at avoiding interference among a large amount of ad-hoc deployed light sources without any explicit coordination. This scheme could be tested in LANET in future.

- RF-Suppressed Applications. LANETs can provide a reliable and accurate solution for data transmission in scenarios where RF communications are suppressed or prohibited, like hospital and climbing/landing airplanes. For example, wireless technology is applied in hospitals for updating information related to patient records, collecting data in a real-time way from handheld patient devices, detecting changes in a patient’s condition, and also for observing medical images via medical equipment (e.g., ultrasound). There, security and safety are essential to maintain confidentiality of patient records, and to ensure that only authorized personnel have access to the data being transferred wirelessly while limiting the interference to those interference sensitive medical devices like EMI.

- Military Applications. In last decades, the most common optical/visible light communication for military applications employ short-range transmissions [131]. In recent years, the emerging of VLC has shown promising advancements making possible the extensive deployment of VLC for military communication strategies [66]. The use of VLC is turned out to be beneficial in the tactical field with enhanced network capacity and better resistance against adversary jamming, and the research is focused in this direction by military organizations and defense companies. Novel and advanced visible light-based military applications include personal area networks, warfighter-to-warfighter communication, vehicular networks, underwater networks, and space applications including inter-satellite and deep-space links. For example, in underwa-
5.3 LANET Node Architecture

In this section, we discuss the two major components of LANET nodes, i.e., hardware and protocol stack as shown in Fig. 5.2, by describing a general reference architecture for LANET nodes. We first review existing frontend hardware components with a particular emphasis on transmitters and receivers that can be used to develop versatile LANET platforms in different environments, e.g., air/space, ground and underwater.

5.3.1 Node Architecture

To date, as we will discuss in Section 5.3.3, there is no existing testbed fully considering VLC-based networking with cross-layer optimized protocol stack (from physical up to transport layer). To bridge this gap, we discuss a potential solution for VLC ad hoc networking, i.e., a software-defined LANET architecture that supports fully flexible and reconfigurable networking.\footnote{We are currently working on the proposed software-defined LANET architecture and more details and results will be discussed in our future work.}
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

based on visible light communications. As shown in Fig. 5.2, each LANET node consists of two main modules: (i) **LANET protocol stack**, which includes cross-layer network optimizer and a software-defined programmable visible light networking protocol stack, from physical up to transport layer, and (ii) **LANET hardware**, which consists of fixed firmware and user-customized control logic, signal processing chain circuit and LANET front-end (e.g., LED and PD).

- **LANET Protocol Stack**: In LANETs, each node is installed a programmable protocol stack, which implements networking functionalities across multiple layers in a software-defined fashion to enable fast and intelligent adaptability. The protocol stack has a modular structure, where different functional blocks, such as timing functionalities, medium accessing functionalities, routing functionalities, among others, can be designed and upgraded independently and conveniently.

  Cross-layer design is an effective way to optimally leverage dependencies between protocol layers to obtain performance gains. In LANETs, the programmable protocol stack is driven using a cross-layer optimizer, which adaptively controls and reconfigures on-the-fly the network parameters based on the results of cross-layer optimization to maximize network utility (i.e., throughput, energy consumption, re-routing, among others), e.g., channel-aware adaption of link layer transmission schemes and multi-user channel access strategies [132–134].

- **LANET Hardware**: While different software-defined radio devices have been adopted in existing VLC testbeds, including USRP, WARP and BBB boards (see Table 5.2), these devices failed to achieve a good tradeoff between fast and flexible prototyping, high-performance signal processing capability and low cost of the device [82–84]. To resolve this issue, some new family of software-defined devices can be used in LANET development, e.g., Nutaq MicroZed, which integrates FPGA and ARM processors into a single board to enable real-time signal processing without requiring large-size FPGA (hence with reduced cost) and without turning to external host (hence with reduced signal processing delay).

As shown in Fig. 5.2 in LANETs LED and PD are used as transmitter and receivers, respectively. **Medium absorption property** of the networking environment is one of the most important factors in selecting proper transceiver devices. For example, in the atmosphere environment, the absorption is inversely promotional to the wavelength [135], i.e., the absorption of violet/blue light is stronger than red light in air. While Blue LED has been proven to be the best choice for the receiving transceiver because deep ocean water typical exhibits a minimum
absorption at this wavelength \cite{136}. The selection of PD will be based on the types of LED selected, the sensitivity of the application requirement, among others.

### 5.3.2 Front-end Hardware

Because of advancements in LED technologies, LEDs outperform conventional light sources or fluorescent bulbs in terms of energy-efficiency, longevity, switching speed and environment-friendliness. All of these advantages motivate the research on visible light communication and enable low-cost VLC systems. To implement the communication function of LEDs, the driver circuit should be modified to modulate data through the use of emitted light, which may help improve the performance \cite{137}. Existing LEDs can be classified into three categories as follows:

- **Phosphor Converted LEDs (pc-LED)** employ a yellow phosphor coating covered upon a blue LED to produce white light. By modifying the thickness of the phosphor layer, different white colors, such as warm-white, neutral-white or cool-white can be produced. Pc-LEDs are cheaper and less complex compared to other LEDs (e.g., RGB LED, Micro Led, etc.). However, their bandwidth is limited to a few MHz because of the low phosphor conversion efficiency \cite{100}.

- **RGB LEDs** utilize three LED chips emitting Red, Green and Blue (RGB) to produce white light. By controlling the intensities of different LED chips, color control can be achieved. Compared to low-cost and low-complexity pc-LED, the cost of RGB LED is higher but with wider achievable bandwidth of 10-20 MHz \cite{138}.

- **Micro LEDs (\(\mu\) LED)** have been used to develop high data rate VLC testbed with much higher bandwidth compared with pc-LED and RGB LED (usually above 300 MHz) and with the resulting achievable data rate up to 3 Gbit/s \cite{71}.

For receiving devices, three types of light receivers have been used: PD, imaging sensors and LEDs.

- **Photodetectors (PDs)** are a semiconductor devices that convert the received light signal into electrical current. Currently, basic PIN and more complex, expensive (about four times the cost of the PIN) Avalanche PD (APD) have attracted more interest for the development of visible light testbeds. APDs has been shown to be more suitable for long range communication as a high speed receiver in high bandwidth applications and bit rates since their internal gain
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

can result in higher SNR \[139\]. However, the high-cost is inevitable compared with PIN PDs. As demonstrated in \[115\], by using APD the data rate has been almost doubled compared with \[114\] where basic PIN PD is adopted.

- **Imaging Sensor**, aka camera sensor, can also be used to receive light signals. However, to enable high-resolution photography, the number of PDs must be very large, which greatly increases the cost of the resulting testbed. Besides, due to low sampling rate, image sensors can only provide limited data rate (a few kbit/s) \[100\]. Therefore, image sensors are not suitable to develop cost-efficient LANETs.

- **LEDs** have been used not only as transmitters but also receivers \[140, 141\]. The most compelling advantage of using LEDs as receivers is to further reduce the cost of the systems but with possibly complemented data rate of up to 12 kbit/s and highly limited FoV \[81\]. For developing visible light networks like LANETs, LEDs as receiver is not recommended.

5.3.3 Existing VLC Testbeds

Visible light ad hoc networking technologies are still in their infancy, with the core problem of developing flexible networking protocol stacks and resource control algorithms specialized for visible-light networks still substantially unaddressed. To see this, next we briefly review several software-defined VLC-based testbed available in existing literature \[70, 81–84\].

**Software-defined single link VLC testbeds.** A software-defined single-link VLC platform utilizing WARP is presented in \[70\]. At transmitter side, the AC waveform is generated by OOK modulation scheme on the software-defined modulation on WARP, then fed to a baseband filter and then converted to analog signal by adding a DAC board (EMC150) on WARP. Besides, a Bias-Tee module is used to build the driver circuit to combine the AC signals and DC power to drive the LED. At the receiver side, PD and ADC are used to receive light signal and convert it to real-valued signal for post processing in WARP. The supported bit rate of such single link platform is from 500 Kbps to 4 Mbps. Similarly, \[84\] also implements ACO-OFDM and DCO-ODFM single-link VLC testbed.

**IEEE 802.15.7 standard based VLC testbeds.** In \[83\], the authors prototype a visible light communication system based on the IEEE 802.15.7 standard. The transmitter of the low-cost software-defined system consists of USPR platform, an amplification stage, the LED driver circuit and a commercial pc-LED. The transmitted data is modulated in the PC and then delivered to USRP over Ethernet link to do DAC. At the receiver side, PD (e.g., ThorLabs PDA36A) delivers the received
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

<table>
<thead>
<tr>
<th>Testbed</th>
<th>Hardware</th>
<th>Topology</th>
<th>Layer Involved</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al [70]</td>
<td>WARP</td>
<td>single link</td>
<td>PHY</td>
<td>data rate 500Kbps to 4Mbps</td>
</tr>
<tr>
<td>Qiao et al [84]</td>
<td>WARP</td>
<td>single link</td>
<td>PHY</td>
<td>ACO-OFDM</td>
</tr>
<tr>
<td>Gavrincea et al [83]</td>
<td>USRP</td>
<td>single link</td>
<td>PHY</td>
<td>IEEE 802.15.7 standard based</td>
</tr>
<tr>
<td>Wang et al [81, 82]</td>
<td>BeagleBone Black board</td>
<td>single link</td>
<td>MAC and PHY</td>
<td>low cost, low data rate</td>
</tr>
</tbody>
</table>

Table 5.2: Representative existing VLC testbeds

signal to the USPR receiving platform, where the signal is sampled and then passed to the PC for demodulation. Similar to above discussed [70] [84], only single visible link has been implemented without considering networking development including techniques in the MAC layer, network layer and transport layer.

Low-cost low-data-rate OpenVLC testbeds. [82] presents OpenVLC1.0, an improved version of OpenVLC [81]. OpenVLC1.0 is an open source, flexible, software-defined, and low-cost platform for research in VLC networks. OpenVLC1.0 mainly consists of three parts: i) BeagleBone Black (BBB) board, ii) OpenVLC1.0 cape and iii) OpenVLC1.0 driver. BBB is a low-cost development platform running Linux for implementing quick communication prototyping. The cape is front-end transceiver that can be plugged into the BBB, including high power LED (HL), low power LED (LL) and PD to be switched to transmit or receive light signals according to application requirements. The driver is used to implement the software solutions for VLC networking, where currently key primitives at MAC and PHY layers are implemented such as signal sampling, symbol detection, coding/decoding, channel contention and carrier sensing. The data rate around 12 kb/s over 4-5 meters is validated using the proposed OpenVLC1.0. OpenVLC1.0 can be adopted as a starter kit for low-cost and low-data-rate VLC research.

We summarize the above-discussed representative testbeds in Table 5.2, from which we can see that most existing VLC testbeds have been focusing on understanding and designing efficient physical layer technology for visible light point-to-point links [70,81,84], or designing simple MAC schemes based on the IEEE 802.15.7 VLC standard [81,82]. As discussed in Section 5.1.1 and Section 5.3.3, unlike protocol design for RF communications, visible light networking technologies are substantially unexplored because of unique VLC wireless links. Next, we discuss those enabling technologies and highlight possible open research issues at each layer of LANET protocol stack.
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

Table 5.3: Visible Light Modulation Schemes

<table>
<thead>
<tr>
<th>Modulation</th>
<th>References</th>
<th>Computation</th>
<th>Power Efficiency</th>
<th>Bandwidth Efficiency</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Carrier Modulation (SCM)</td>
<td>OOK</td>
<td>low</td>
<td>medium</td>
<td>medium</td>
<td>low to moderate data rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>medium</td>
<td>low</td>
<td>high</td>
<td>medium data rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>complex</td>
<td>high</td>
<td>low</td>
<td>medium data rate</td>
</tr>
<tr>
<td>Multiple Carrier Modulation (MCM)</td>
<td>OFDM</td>
<td>complex</td>
<td>low</td>
<td>high</td>
<td>multiuser high data rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color Domain Modulation</td>
<td>CSK</td>
<td>complex</td>
<td>medium</td>
<td>high</td>
<td>multiuser high data rate</td>
</tr>
</tbody>
</table>

5.4 Physical Layer

Unlike RF systems where signal can be modulated in terms of amplitude, frequency and phase, in VLC it is the intensity (aka instantaneous power) of the visible light that is modulated [108], i.e., intensity modulation (IM). Correspondingly, demodulation is typically based on direct detection (DD), where a photodetector produces an electrical current proportional to the received instantaneous light power, i.e., proportional to the square of the received electric field [142]. This combination of modulation techniques is referred to as IM/DD (Intensity Modulation / Direct Detection). As discussed in the previous sections, LEDs may have dual functions, illumination and communication. Different from indoor communication using visible light spectrum, where illumination is the primary function [107], in LANETs illumination may not be as important as in indoor applications. This means that flicker mitigation and dimming support for comfortable indoor living environment are not core considerations in the modulation process of LANETs.

Flicker mitigation aims to eliminate the phenomenon that human eyes can observe the flickering of the light, which can be avoided by using waveforms whose lowest frequency components are far greater than the flicker fusion threshold of the human eyes (which is typically less than 3 kHZ).
5.4.1 Existing Modulation Schemes

In this section, we discuss the state-of-the-art IM/DD modulation schemes adopted at the PHY layer for visible light communication system. As summarized in Table 5.3, existing VLC modulation schemes can be classified into single carrier, multi-carrier and color domain modulation schemes. We will compare the main VLC modulation schemes from the perspective of power efficiency, bandwidth efficiency, and implementation complexity.

5.4.1.1 Single Carrier Modulation

Single carrier modulation techniques were first proposed for IM/DD wireless infrared communication [86]. For example, on-off keying (OOK), pulse amplitude modulation (PAM), and pulse position modulation (PPM) are easily implemented for LANET systems. In general, single carrier modulation schemes are suitable for LANETs where low-to-moderate data rate are required [152].

On-Off Keying (OOK). OOK is the most common and simplest modulation technique for IM/DD in VLC, where higher or lower intensity of light represents a 1 or 0 bit [88]. Both OOK non-return-to-zero (NRZ) and OOK return-to-zero (RZ) can be applied. Since OOK-RZ has twice the bandwidth requirement of OOK-NRZ and does not support sample clock recovery at the receiver [143], OOK-NRZ has been more widely used in VLC systems [113] [114] [115] [137] [78]. In [113] the authors present a 10 Mbit/s visible light information broadcasting system with maximum communication distance 3.6 m based on message signboard with four LED arrays. [114] and [115] demonstrate a visible light link operating at 125 Mbit/s over a 5 m communication distance by adopting blue-filtering with analogue equalization at the receiver and an improved 230 Mbit/s visible link with OOK-NRZ by using an APD instead of the PIN photodiode, respectively. More recently, in [137] a 300 Mbit/s line-of-sight visible light link using OOK-NRZ over 11 m is demonstrated with 600 nm LED and off-the-shelf PIN PD by proposed 2-cascaded Schottky diodes-capacitance current-shaping drive circuit. In [78], an OOK-NRZ based visible link with maximum transmission speed 477 of Mbit/s over 0.5 m by using a commercially available red LED and a proposed LED driver with a simple pre-emphasis circuit and a low-cost PIN PD is demonstrated.

Pulse Amplitude Modulation (PAM). PAM is a more generalized OOK (the simplest 2-PAM is namely OOK modulation) [144]. In PAM, multiple intensity levels are defined to represent various amplitudes of the signal pulse. However, multiple intensity levels may undergo nonlinearity
in terms of LEDs luminous efficacy, depending on the color of LED emission on input current and temperature [100].

**Pulse Position Modulation (PPM).** PPM divides a symbol duration into $L$ equal time slots and a single pulse is transmitted in each of the $L$ slots, where the position of the pulse represents different transmitted symbols. PPM can improve the power efficiency compared with OOK but at the expense of an increased bandwidth requirement and greater complexity [100]. Therefore, to overcome the lower spectral efficiency and data rate limitations, some variants of PPM, e.g., Multi-pulse PPM (MPPM) [145] and Overlapping PPM (OPPM) [146], are proposed. MPPM and OPPM can not only achieve higher spectral efficiency but also provide dimming control. Besides, Variable PPM (VPPM) [107] is another important variant of PPM, adopted in standard IEEE 802.15.7 (which will be discussed later in this section), where the duty cycle (pulse width) of the transmitted symbol can be adjusted according to the dimming level requirements. Recently, other variations based on MPPM, such as OMPPM [147] and EPPM [148] are also proposed to further either improve the spectral efficiency or provide arbitrary dimming control levels. Because of the low data rate of PPM and the low relevance of dimming control in LANETs, we will not discuss PPM-based modulation schemes in detail, interested readers are referred to [100] and references therein for more information.

### 5.4.1.2 Multi-carrier Modulation

Compared to single carrier modulation, multi-carrier modulation can achieve high aggregate bit rates and improved bandwidth efficiency at the cost of reduced power efficiency because increasing the number of subcarriers also increases the DC offset to avoid clipping [88]. Orthogonal Frequency Division Multiplexing (OFDM) and its variants, as the typical multi-carrier modulation techniques, are widely adopted in the existing VLC systems.

OFDM is first demonstrated in [154] for visible light communications. OFDM can help combat inter-symbol interference (ISI) and multi-path fading while significantly boosting the achievable data rate over wireless links. To date, the highest data rates achieved in visible light communications by utilizing OFDM is up to 3 Gbit/s over 0.05 m [71] where a single LED is adopted.

Different from original OFDM in RF systems, where complex-valued bipolar signals are generated, in IM/DD based visible light communications only real-valued signals are acceptable. Therefore, conventional OFDM techniques for RF need to be modified for VLC systems. To convert
bipolar signals to unipolar, there are two major techniques: i) DC-biased Optical OFDM (DCO-
OFDM) [142] and ii) Asymmetrically-Clipped Optical OFDM (ACO-OFDM) [149]. In ACO-OFDM,
only odd subcarriers are used to modulate data, while in DCO-OFDM all the subcarriers are adopted
by adding a DC bias to make the signal positive. It is shown in [79] that ACO-OFDM is more
efficient than DCO-OFDM in average optical power for constellations from 4 QAM to 256 QAM
because the DC bias used in DCO-OFDM is less power efficient; but DCO-OFDM outperforms
ACO-OFDM in spectrum efficiency since ACO-OFDM uses only half of the subcarriers to carry data.
Recently, Unipolar OFDM (U-OFDM) [150] and asymmetrically clipped DC biased optical OFDM
(ADO-OFDM) [151] are proposed to overcome the limitations of DCO-OFDM and ACO-OFDM.

5.4.1.3 Color Shift Keying (CSK)

CSK was defined in the latest IEEE 802.15.7 standard [153] by using multi-color LEDs,
which is similar to frequency shift keying in that bit patterns are encoded to color (wavelength)
combinations. Specifically, the transmitted bit corresponds to a specific color in the CIE 1931 [155]
coordinate. The IEEE 802.15.7 standard divides the spectrum into 7 color bands from which
the RGB sources can be picked from, and the picked wavelength bands determine the vertices of
a triangle inside which the constellation points of the CSK symbols lie. The color point for each
symbol is generated by modulating the intensity of RGB chips. However CSK cannot be used in
a VLC system where the source is a pc-LED [100] (which is one of the most common sources of
light in an illumination system). Moreover, implementation of CSK requires a more complex circuit
structure [100].

5.4.1.4 Standardization of Physical Layer: IEEE 802.15.7

IEEE 802.15.7 standard [153] has specified at the PHY layer three types of VLC techniques,
including in total 30 modulation and coding schemes for different applications with different desired
data rates, as discussed as follows.

- **Physical (PHY)** I is designed for outdoor applications with low data rates. This mode uses
  OOK and VPPM along with Reed-Solomon (RS) and Convolutional Coding (CC) for Forward
  Error Correction (FEC). The operating data rates vary from 11.67 kbit/s to 266.6 kbit/s with
  support for 11.67 kbit/s at 200 kHz being mandatory.

---

[6]The CIE 1931 color space chromaticity diagram represents all the colors visible to the human eyes with their
chromaticity values $x$ and $y$. 

---

92
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

- **PHY II** has been designed for outdoor applications with moderate data rates. **PHY II** uses the same modulations and Run Length Limited (RLL) code as **PHY I** but supports only RS coding for FEC. **PHY II** supports data rate ranging from 1.25 Mbit/s to 96 Mbit/s. All **PHY II** VPPM modes shall use 4-bit to 6-bit encoded symbols (4B6B) encoding, while all OOK **PHY II** modes use 8-bit to 10-bit encoded symbols (8B10B) with DC balance.

- **PHY III** uses CSK for applications equipped with multiple light sources and color filtered photo detectors. The data rates vary from 12 Mbit/s to 96 Mbit/s. **PHY III** supports RS coding for FEC.

5.4.2 Open Research Issues

In the physical layer of LANETs, the following two research directions can be identified to further enhance capacity and power efficiency of visible light communications.

- **High Power Efficiency.** Besides free space loss, other factors, including absorption and atmospheric conditions, can considerably reduce the intensity of visible light for outdoor applications. Moreover, in ad hoc networking, low energy consumption is often a critical factor since network devices are usually battery powered. Examples include mesh networks of unmanned aerial vehicles (UAVs), sensors or communication devices in disaster recovery scenarios, tactical field devices, among others. Therefore, intuitively, new physical layer techniques enabling higher power efficiency are needed. Although [156] and [157] have pioneered research on low-power consumption, this line of work for visible-light wireless communications is still in its infancy.

- **Long Communication Range.** Visible light has the potential to provide high data rate communications. For example, [69] and [71] demonstrated a 4.5 Gbit/s RGB-LED based WDM indoor visible light communication system and a 3 Gbit/s single gallium nitride μLED OFDM-based wireless VLC link, respectively. However, the communication ranges are only 1.5 m and 0.05 m. For LANETs, mainly operating in outdoor environments, significantly longer ranges are a key requirement. [72] proposes to use a polarized-light intensity modulation scheme to increase the transmission range, up to 40 meters, with very limited data rate, i.e., 76 bytes per second. [158] and [159] can achieve data rate 210 Mbps and 400 Mbps respectively at bit error rates of $10^{-3}$ over distances in the order of 100 meters, at the cost of increased system complexity. In [158], a collimating lens for optical antennas is designed and optimized by
using Taguchi method. In [159], advanced OFDM modulation schemes, pre-equalization, reflection cup, convex lenses, and receiver diversity are adopted to boost the data rate over 100 meter distance. There is clearly a trade-off among the data rate, transmission range and system complexity scintillation noise induced.

5.5 Medium Access Control Layer (MAC)

There has been limited work specifically on Medium Access Control (MAC) for visible light communications. The few existing MAC schemes for Visible Light Communication (VLC), as summarized in Table 5.4, are mainly based on approaches blindly drawn from RF communications, such as Carrier Sense Multiple Access/Collision Detection (CSMA/CD) (also adopted in IEEE 802.15.7 [153]) or Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA), cooperative MAC and OFDMA, unfortunately without considering specific VLC channel characteristics and challenges. Additionally, most of the existing MAC schemes have been designed to enable point-to-point VLC and hence are not easily extendable to LANET. Some of these MAC schemes are discussed below.

5.5.1 Existing Visible Light MACs

**CSMA-based Channel Access** [160–162]. In [160], the authors propose a full-duplex Medium Access Control (MAC) protocol with Self-Adaptive minimum Contention Window (SACW) that delivers higher throughput from the central node to the terminal nodes in a star topology. The proposed algorithm still uses the basic slotted CSMA/CA mechanism as in [153] with adaptive contention window. The objective of SACW MAC is to allow the central node to monitor the data traffic to increase the probability of full-duplex operation. The authors of [161] also propose a high speed full-duplex MAC protocol based on CSMA/CD by considering a start topology with Access Point (AP) at the center and multiple terminal nodes trying to communicate with the AP. Another example of VLC using CSMA/CA is in [162], which uses LED to transmit and receive to reduce hardware cost and size. This work uses Light Emitting Diode (LED) charged in reverse bias to receive the incoming light.

**Cooperative MAC** [163]. A cooperative MAC protocol is proposed in [163] to reduce latency and for on-demand error correction. The sender and receiver will initiate a cooperative mechanism to find relay nodes when the direct link does not provide the required bandwidth to meet
## Chapter 5. LANET: Visible-Light Ad Hoc Networks

### Table 5.4: Summary of MAC protocols for VLC

<table>
<thead>
<tr>
<th>MAC Protocol</th>
<th>Medium Access Method</th>
<th>Topology/Operation Modes</th>
<th>Other Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE 802.15.7</td>
<td>CSMA/CA</td>
<td>peer-to-peer, star, broadcast</td>
<td>Standardization for VLC</td>
</tr>
<tr>
<td>SACW MAC [160]</td>
<td>CSMA/CA</td>
<td>star</td>
<td>Full-duplex</td>
</tr>
<tr>
<td>Lin et al [161]</td>
<td>CSMA/CD</td>
<td>star</td>
<td>Full-duplex</td>
</tr>
<tr>
<td>Schmid et al [162]</td>
<td>CSMA/CA</td>
<td>peer-to-peer</td>
<td>LED-to-LED</td>
</tr>
<tr>
<td>Cooperative MAC [163]</td>
<td>CSMA/CA</td>
<td>peer-to-peer</td>
<td>Cooperative relay</td>
</tr>
<tr>
<td>Broadcasting MAC [164]</td>
<td>TDMA</td>
<td>broadcast</td>
<td>Frame synchronization and supports QoS</td>
</tr>
<tr>
<td>OWMAC [165]</td>
<td>TDMA</td>
<td>star, with unicast, broadcast, &amp; multicast</td>
<td>84 Mb/s data rates</td>
</tr>
<tr>
<td>Dang et al [166]</td>
<td>OFDMA</td>
<td>star</td>
<td>Comparison of O-OFDMA &amp; O-OFDM-IDMA</td>
</tr>
<tr>
<td>Ghimire et al [167]</td>
<td>OFDMA-TDD</td>
<td>star</td>
<td>Self-organising interference management</td>
</tr>
<tr>
<td>Chen et al [168]</td>
<td>DCO-OFDM</td>
<td>Indoor downlink transmission</td>
<td>Spectral efficiency of 5.9 bits/s/Hz</td>
</tr>
<tr>
<td>Bykhovsky et al [169]</td>
<td>DMT</td>
<td>star</td>
<td>Interference-constrained subcarrier reuse</td>
</tr>
<tr>
<td>Shoreh et al [169]</td>
<td>MC-CDMA with PRO-OFDM</td>
<td>star</td>
<td>Handles dimming using PRO-OFDM</td>
</tr>
<tr>
<td>He et al [170]</td>
<td>OCDMA with OOC</td>
<td>peer-to-peer, star</td>
<td>Bipolar-to-Unipolar encoding and decoding</td>
</tr>
<tr>
<td>Gonzalez et al [171]</td>
<td>OCDMA with ROC</td>
<td>peer-to-peer, star</td>
<td>Specific design of OOC, higher complexity</td>
</tr>
<tr>
<td>Chen et al [172]</td>
<td>OCDMA with CSK</td>
<td>peer-to-peer, star</td>
<td>Mobile phone camera used as receiver</td>
</tr>
<tr>
<td>Yu et al [173]</td>
<td>MU-MISO</td>
<td>Cooperative broadcasting</td>
<td>ZF algorithm using generalized inverse</td>
</tr>
<tr>
<td>Pham et al [174]</td>
<td>MU-MISO</td>
<td>Cooperative broadcasting</td>
<td>ZF algorithm using optimal precoding</td>
</tr>
<tr>
<td>MU-MIMO (BD) [175]</td>
<td>MU-MIMO</td>
<td>star</td>
<td>Precoding using BD algorithm</td>
</tr>
<tr>
<td>MU-MIMO (THP) [176]</td>
<td>MU-MIMO</td>
<td>star</td>
<td>Precoding using THP algorithm</td>
</tr>
</tbody>
</table>

the Quality of Service (QoS) requirement. Once cooperative mode is initiated, the sender broadcasts a RelayRequest. Nodes within range save the sender’s identification number. Next, the destination broadcasts a RelayRequest. Nodes that receive both RelayRequests will broadcast its information to sender and destination if the node decides to be a relay. The relay overhears the sender’s packets
and saves them till an Acknowledgment (ACK) is received from the destination. If the ACK is not received, the relay transmits the saved packets to the destination.

**Orthogonal Frequency-division Multiple Access (OFDMA) [68, 166–168].** Recently, the OFDM used in the PHY layer of VLC has been extended to enable multi-user access through Orthogonal Frequency Division Multiple Access (OFDMA). In [166], authors compare the Bit Error Rate (BER) performance, receiver complexity and power efficiency of two multicarrier-based multiple access schemes namely, Optical Orthogonal Frequency Division Multiplexing Interleave Division Multiple Access (O-OFDM-IDMA) and Optical Orthogonal Frequency Division Multiple Access (O-OFDMA). The authors of [167] evaluate a self-organizing interference management protocol implemented inside an aircraft cabin. The goal of the work is to allocate time-frequency slots (referred to as chunks) for transmitting data in an Intensity-Modulation Direct-Detection (IM/DD)-based OFDMA-Time Division Duplex (TDD) systems. Another OFDMA technique for indoor VLC cellular networks is analyzed in [68] using Direct-Current Optical OFDM (DCO-OFDM) as multi-user access scheme. In [168], the authors propose a heuristic subcarrier reuse and power redistribution algorithm to improve the BER performance of conventional Multiple Access Discrete Multi-Tones (MA-DMT) used for VLC.

**Code Division Multiple Access (CDMA) [169–172, 177, 178].** There have been several contributions aimed at employing CDMA in VLC. A system using Multi-carrier CDMA (MC-CDMA) along with OFDM platform is proposed in [169]. The proposed design uses Polarity Reversed Optical OFDM (PRO-OFDM) to overcome the inherent light-dimming problem associated with using CDMA with visible light. In this design a unipolar signal is either added or subtracted to the minimum or maximum current respectively in the LED’s linear current range to provide various levels of dimming. In [170], the authors discuss how Gold sequences and Wash-Hadamard sequences can be adapted for VLC Optical Orthogonal Codes (OOC) [177] comprising of sequences of 0s and 1s have also been explored as a prime candidate to establish Optical Code-Division Multiple Access (OCDMA) for visible light communication. Since as the number of users increases in the system, it becomes challenging to generate OOC for each user, Random Optical Codes (ROC) have been proposed as an alternative, even though they do not provide optimal performance [171, 178]. There have also been efforts to combine Color-Shift Keying (CSK) modulation and OCDMA to enable simultaneous transmission to multiple users [172].

**QoS-Based MAC.** In [164], the authors propose a QoS-based slot allocation to enhance the broadcasting MAC of IEEE 802.15.7 standard. They use a super frame structure similar to the standard. When a new channel wants to join the AP, it sends a traffic request to the access point along
with its QoS parameters (data rate, maximum burst traffic, delay requirements and buffer capacity). Optical wireless MAC (OWMAC) \[165\] is a Time Division Multiple Access (TDMA) based approach aimed at avoiding collision, retransmission and overhead due to control packets. In OWMAC, each node reserves time slot and advertises the reservation using a beacon packet. OWMAC also employs Error-Correction Code (ECC) in their ACK to ensure that retransmission are reduced to corrupted ACK packets. This protocol is designed to handle start like topologies.

MU-MIMO \[173–176, 179–181\]. An alternative method uses multiple LED arrays as transmitters to serve multiple users simultaneously \[173, 174\]. In contrast to the RF counterpart, the VLC signal is inherently non-negative leading to the necessity of modifying the design of the Zero Forcing (ZF) precoding matrix. In \[173\], a ZF precoder is chosen in the form of specific generalized inverse of the channel matrix known as the pseudo-inverse. The authors of \[174\] recognize that the pseudo-inverse may not be the optimal precoder. Accordingly, they design an optimal ZF precoding matrix for both the max-min fairness and the sum-rate maximization problems. Block Diagonalization (BD) algorithm \[179\] has also been used to design the precoding for Multi-User Multiple-Input Multiple-Output (MU-MIMO) VLC system \[175\] to eliminate Multi-User Interference (MUI) and its performance has been evaluated in \[180\]. Finally, Tomlinson-Harashima Precoding (THP) \[181\] has been utilized in \[176\] to achieve better BER performance compared to the block diagonalization algorithm in VLC systems.

MAC protocols \[68, 160, 161, 166–168\] that are designed for centralized operation in a star topology are not easily extensible to LANETs. Cooperative operations like in \[163\] can be employed in LANETs but cannot be the primary MAC protocol used to negotiate reliable medium access. Techniques based on CDMA or MU-MIMO are suitable for centralized networks as it may be complex to negotiate different codes for each link in a distributed network. Similarly, QoS-based techniques can be used to improve a stable MAC protocol that has been primarily designed to overcome inherent problems of LANETs such as deafness, blockage and hidden node problem. These problems are described in detail in Section \[5.5.4\].

5.5.2 MAC for LANETs

A MAC protocol for LANETs (VL-MAC) is proposed in \[182\] to alleviate problems caused by hidden nodes, deafness and blockage while maximizing the use of full-duplex links. VL-MAC introduces the concept of opportunistic link establishment in contrast to traditional methods where a forwarding node is chosen before the negotiation for channel access begins. A utility based
opportunistic three-way handshake is employed to efficiently negotiate medium access. First, a node chooses the optimal transmission sector, i.e., the “direction” that maximizes the probability of establishing a link even when some of the neighbors are affected by blockage or deafness. Since full-duplex communication is inherent to VLC, the utility function is also used favors the establishment of full-duplex communication links. The full-duplex transmission or busy tone along with power control employed by the proposed MAC protocol is aimed at mitigating the hidden node problem. All these factors contribute towards maximizing the throughput of Visible-Light Tactical Ad-Hoc Networking (LANET). The timing diagram and an example of three-way handshake procedure is depicted in Fig. 5.3 and Fig. 5.4 respectively. The node that initiates communication is called the initiator and the node that accepts communication link is called the acceptor.

Consider four nodes $A$, $B$, $C$ and $D$ as shown in Fig. 5.4, among which $B$ and $C$ are the initiators with packets to be transmitted and $A$ and $D$ are prospective acceptors. Once a node has packets to transmit, it has to choose a sector to transmit such that it maximizes the initiator’s utility function ($U_{ini}$). This is a joint function of backlog and the achievable forward progress through the chosen sector. Accordingly, $B$ and $C$ choose the sector corresponding to their maximum $U_{ini}$. In this example, assume that both choose the same sector. Nodes $B$ and $C$ choose a random backoff depending on their $U_{ini}$ and broadcast an Availability Request (ART) packet if the channel is idle. As shown in Fig. 5.4, both $A$ and $D$ listen to control packet during the corresponding sector duration. On reception of ART, $A$ and $D$ will calculate their respective acceptor’s utility function, $U_{acp}$. Next, $A$ and $D$ choose the initiator ($B$ or $C$), initiator’s session and acceptor’s session for potential full-duplex communication such that it maximizes their respective $U_{acp}$. As shown in Fig.
5.3 and Fig. 5.4, A transmits a Availability Confirmation (ACN) to the chosen initiator (A chooses B in this case) after a random backoff which is dependent on $U_{acp}$. The initiators B and C listen for ACN from A and D. In this example, the ACN from A is received by intended node B and overheard by C. Accordingly, B transmits Reserve Sectors (RES) packet to reserve time required to complete the transmission. Node C learns that it was not chosen for transmission by overhearing the ACN and hence defers access. Similarly, D overhears the RES packet and returns idle. Performance evaluation studies show up to 61% increase in throughput and significant improvement in the number of full-duplex links established with respect to CSMA/CA.

5.5.3 Standardization: MAC of IEEE 802.15.7

The IEEE 802.15.7 MAC protocol [153] is designed to support three different topologies, namely peer-to-peer, star and broadcast considered by IEEE 802.15.7, as shown in Fig. 5.5. In a peer-to-peer topology, each node is capable of communicating with any other node within its coverage area. One node among the peers need to act as a coordinator. This could be determined in multiple ways for example, by being the first to initiate communication on the channel. As shown in Fig. 5.5 a star topology consists of a single coordinator communicating with several child nodes. Each star network operates independently of other networks by choosing a unique Visible-light communication Personal Area Network (VPAN) identifier within its coverage area. Any new child node uses the VPAN identifier to join the star network. Finally, in the broadcast mode the communication is uni-directional and does not need address or formation of a network. Visibility support is also provided across all topologies to mitigate flickering and maintain the illumination function in the absence of communication or in the idle or receive modes of operation [153].

Active and passive scan are performed by nodes across a specified list of channels to listen
for beacon packets and form VPANs. While every node should be capable of passive scan, the coordinator should be able to perform active scan. An active scan is used by a prospective coordinator to locate any active coordinator within the coverage area and select a unique identifier before starting a new VPAN. To perform an active scan over a specified set of logical channels, the node switches to the required channel and sends out a beacon request. Next, it enables the receiver such that only beacon packets are processed. The passive scan is similar to active scan but nodes do not send out the beacon request. The passive scan is envisioned to be used in star or broadcast topologies while the active scan is for peer-to-peer topologies. Beacon packets are also used to synchronize with the coordinator. In VPANs that do not support the use of beacons, polling is used to synchronize with the coordinator.

5.5.4 Open Research Issues

From the above discussion we can see that existing VLC MAC protocols consider primarily point-to-point link or simple multicast or broadcast access where a master node serves as coordinator. In LANETs, VLC-enabled nodes are networked together via possibly multi-hop visible light links in an ad hoc fashion to support various applications spanning terrestrial, underwater, air as well as space domains, for which the MAC design is more challenging. Several open research issues are identified below.

- **Deafness Avoidance.** When the VLC receiver is oriented towards a segment of the space, it is unable to receive from all the remaining segments. This situation is referred to as deafness. Thus, a node may try to initiate communication with its neighbor who is experiencing deafness with respect to the node, leading to additional delays during the contention phase. Additionally, the list of instantaneous neighboring nodes may change if the system has a Field Of View (FOV) that changes direction. Hence, appropriate synchronization procedures need to be included in the MAC protocol to coordinate between the prospective neighbors.

- **Hidden Node Detection.** Classic challenges like hidden node problem amplified in LANETs because of directionality. Control packets like Clear-to-send (CTS) transmitted by a receiver may not be received by nodes because of limited FOV. When a node that does not receive the CTS tries to initiate communication with the receiver, it causes interference to the ongoing communication leading to collisions. Furthermore, traditional virtual carrier sensing using Network Allocation Vector (NAV) has to be modified to take advantage of spatial reuse.
Because of the above challenges, it is necessary to design channel dependent MAC protocols specifically to leverage the characteristics of VLC.

- **Channel-aware VLC MAC.** Directionality is a key distinguishing feature of VLC. Larger FOV result in more diffused links (i.e., with light reflected by objects between transmitter and receiver), which in turn leads to higher attenuation. Therefore, VLC systems with high-rate transmission cannot have large FOV. Moreover, sudden communication discontinuity (blockage) may happen during the contention phase and communication stage. This will result in frequent re-connect problem, which will further cause increase in the contention payload and degradation of the effective throughput. VLC devices need to operate at a wide range of power levels to satisfy lighting or other requirements. This implies that a channel-aware MAC protocol is required to negotiate and operate at appropriate configuration (i.e. wavelength, data rates or modulation) to maintain the link under different scenarios.

- **Full-duplex capability.** Unlike typical Radio Frequency (RF) transceiver systems equipped with a single antenna to transmit or receive, VLC devices are usually equipped with a LED for transmission and a Photon Detector (PD) for reception making these devices inherently capable of full-duplex communication. Therefore, MAC protocols designed for LANETs should be able to take advantage and utilize the full-duplex links to improve the network throughput.

### 5.6 Network Layer

Routing at the network layer will play a significant role on the performance of LANETs and have a major influence on the overall network throughput. However, most of the existing work in visible light communication is confined to point-to-point communication or a cooperative relay based communication [162, 163]. To the best of our knowledge, multi-hop routing for visible light ad-hoc networking is substantially unexplored. There are two major challenges:

- **Blocking of Service.** In LANETs, one of the most important characteristics of visible light communications is that signal penetration through any non-transparent objects is physically impossible. We refer to this problem as blocking of service. For example, in traditional routing schemes in RF-based MANET, links with the best quality are generally selected [183, 184]. However, best-quality links may not be inside the previous hop’s FOV or some objects may appear as obstacles over one link after the routing decision. In these cases, the best routes determined by traditional routing schemes may not be desirable.
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

- **Limited Route Lifetime.** Route maintenance is important in any ad-hoc network due to possible route failures caused by impaired channel, node failures, among other reasons. This problem is magnified in LANETs because of blockage caused by obstacles or deafness caused by directionality as described in Section 5.5. The nodes in a LANET must rapidly adapt to route failures and dynamically find alternate path to the destination.

To address these challenges, we identify three possible research directions in the design of LANET network layer.

5.6.1 **Open Research Problems**

- **Proactive LANET Routing.** In proactive or table-driven routing protocols, each node maintains routing information for the entire network. Usually, in an omnidirectional network, the nodes may use broadcast messages regularly to learn changes in topology and routes. In a directional network, this becomes challenging and time intensive due to deafness and the need to exchange messages in every sector. In LANETs, the problem is further aggravated due to the limited route lifetime discussed earlier. Therefore, there is a constant need to update routes but at the same time, it is extremely challenging to learn changes in the network in an efficient manner.

All these factors render it extremely difficult to maintain updated routing tables for the entire network.

- **Reactive LANET Routing.** In reactive routing protocols, the routes are discovered when a source requires to transmit a packet to a destination and eliminates the need to maintain routing tables at every node. Although reactive protocols reduce communication overhead and power consumption, they lead to higher delays. It is difficult to discover all possible routes due to the narrow FOV and without an adequate neighbor discovery scheme that overcomes blocking. After route discovery, it becomes important to select the optimal route to maximize the overall throughput of the network. Depending on the device, a dynamic routing protocol should consider the interaction between routing and channel selection with help of a cross-layer controller.

- **MAC-aware Routing.** Due to the frequent reconnect problem, routing in LANETs relies heavily on MAC layer to maintain the links for uninterrupted transmission. Thus, repeated interaction between the network layer and the MAC layer becomes crucial, inducing the need for a cross-layer controller. While directionality enables spatial reusability, it also poses serious
challenges during neighbor discovery and route selection. For example, during the neighbor
discovery phase, some nodes may be overlooked due to deafness. This will reduce the number
of potential opportunistic routes available to the node in a LANET as compared to a traditional
MANETs. Thus, an efficient neighbor discovery technique and a dynamic routing algorithm
has to be uniquely designed for LANETs.

5.7 Transport Layer

The main objective of transport layer protocols is to provide end-to-end communication
services with, among other functionalities, reliability support and congestion avoidance. To achieve
reliable transmission, a transport layer protocol, say TCP [185], detects packet loss either caused by
transmission errors or network congestion and then sends an ACK to the sender to acknowledge the
successful reception of the packet or NACK message to request retransmissions; and regulates the
maximum data rate a sender is allowed to inject into the network to avoid congestions.

In past years, transport layer protocols has been extensively discussed focusing on wireless
multimedia sensor networks [186], cognitive radio networks [187], delay and disruption tolerant
networks [188], and wireless video streaming networks [134], among others. These protocols in
existing literature however are not suitable to (at least are not the optimal for) LANETs because
of the special characteristics of visible light communications, including directionality, intermittent
availability and predictability. Next, we discuss the applicability of existing transport layer protocols
and the necessary modifications to address the unique challenges in LANETs.

5.7.1 Existing Transport Layer Protocols

Existing transport-layer protocols [189–195] can be categorized into three classes, UDP,
TCP and TCP-friendly protocols, and application-/network-specific protocols, as illustrated in
Fig. 5.6

- UDP is a simple connectionless but unreliable transport layer transmission scheme, which
  provides a minimum set of transport layer functionalities without any guarantee of delivery,
  order of packets, or congestion control. Because of its timeliness, UDP protocol has been

Unlike radio-frequency-based communications, where the wireless channels can be considerably faded by multi-path
transmissions, in LANETs VLC links are largely dominated by LOS transmissions and the resulting wireless channel
quality can be much more stable than its RF counterparts and hence is easier to predict. By predicting the channel quality
of the links belonging to a route, transport layer protocols can response in a proactive manner to the route outages, e.g., by
allocating higher data rate to routes with higher predicated throughput if multiple routes are available.
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

Transport Layer

UDP

TCP/TCP-friendly Protocols

Loss-based
- Veno
- HSTCP
- Scalable TCP

Delay-based
- DCA TCP
- TCP Vegas
- Fast TCP

Loss-delay-based
- TCP-Illinois

Figure 5.6: Existing transport layer protocols.

typically used in applications that are delay sensitive but packet loss tolerable, e.g., real-time video streaming, online gaming, and VOIP in wired and radio networks. However, the protocol does not suit well to LANETs due to its indiscriminate packet dropping. Particularly, in mobile LANETs each VLC link can be only intermittently available with link outage at a level of seconds, and the resulting burst packets dropping may cause considerable QoS degradation that can be even fatal the dropped packets are key packets (e.g., packets of intra-coded video frames). Multi-path routing can be used to account for link outages, however UDP protocol does not provide any guarantee of receive order of packets.

- TCP/TCP-Friendly Protocols. Different from UDP, TCP protocols provide connection-oriented, reliable and ordered packet delivery [185], and hence it is more favorable to account for the link outages and multi-path routing in LANETs. We discuss three classes of TCP protocols, loss-based, delay-based and their combinations, and discuss their applicability in LANETs. The congestion control in loss-based TCP protocols, including Reno TCP [196] and its enhancements [197, 198], has the form of additive-increase/multiplicative-decrease (AIMD), e.g., the well known slow start and exponential backoff mechanisms. While AIMD-based congestion control has been remarkably successful since Reno first developed in 1988, as pointed out in [190], it may eventually become the performance bottleneck in newly evolved wireless networks with high bandwidth-delay product (BDP), such as LANETs. Roughly speaking, if BDPs are high it can be too slow for the transport layer protocols based on AIMD to converge to the optimal transmission size. To date, up to 3 Gbits/s over 5 cm VLC link [71] and 300Mbits/s over VLC links of tens meters [137] have been be achieved. By jointly taking the advantages of directionality and predictability of VLC links, LANETs are envisioned to have the potential to unlock the capacity of wireless ad hoc networks, typically resulting in

104
large BDPs.

Therefore, delay-based TCPs are more suitable to LANETs since they have been proven to outperform loss-based TCPs in networks with large BDPs [190]. These protocols adjust the transmission window size based on the measured end-to-end delay: increase the window size if the delay increases and decrease the window size otherwise. Because the network congestion can be indicated more accurately, network resources can be almost fully used with increased network throughput. Main problems of delay-based TCPs are that, they are incompatible with the standard TCPs, and may lead to unfair network resource allocation if it coexists with loss-based TCPs. A possible solution, as in [199], is to design transport layer protocols by jointly considering packet loss and delay.

**Transport Layer of LANETs.** To date, there are only few research work focusing on transport layer protocol design and performance evaluation in VLC networks [200–203]. In [200], Mai et al. study the effects of link layer protocols on the performance of TCP over VLC networks. Automatic-repeat request, selective repeat (ARQ-SR) protocol is considered at the link layer, and they find that TCP throughput can be considerably affected by the ISI and reflection of visible light signals, and ARQ-SR could significantly improve the achievable TCP throughput if the number of re-transmissions is properly selected. In [201], Kushal et al. present a visible-light-based protocol to provide reliable machine-to-machine communications. A flow control algorithm similar to TCP has been integrated into the proposed protocol to deal with dynamic ambient brightness. Different from standard TCP, the flow control algorithm there adjusts the packet size based on if previous packets can be successfully delivered. Through experiment results, with given communication distance and angular variation of transmitter, a sharp drop off in packet delivery ratio can be observed if the packet size exceeds certain threshold, which calls for a joint optimization of packet size at transport layer and communication link distance at physical layer. In [202,203], Sevincer, Bilgi, et al. discuss the effects of intermittent alignment-misalignment behaviors of VLC links at physical layer on the TCP stability at transport layer. They argue that a special buffer should be introduced to make the physical layer more tolerable to the intermittency, and hence mitigate the link-layer packet loss and further make the transport layer protocols less sensitive to the intermittency. Since larger buffer may increase queueing delay, a trad-off needs to be achieved at transport layer between route connectivity and end-to-end delay.
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

5.7.2 Open Research Issues

The performance of transport layer protocols can be considerably affected by the unique characteristics of LANETs at lower layers, including intermittent link connectivity, transceiver angular variation, and the channel-dependent layer-2 strategies, among others. Next, we identify the following open research issues at transport layer of LANETs.

- **Blockage-Aware LANET Transport Protocol Design.** In traditional ad hoc wireless networks, dynamic network topology changes are usually caused by the unrestricted mobility of the nodes in the network, which will further lead to frequent changes in the connectivity of wireless links and hence rerouting at the network layer. If the frequent route reestablishment time is greater than the retransmission timeout (RTO) period of the TCP sender, then the TCP sender assumes congestion in the network, and retransmits the lost packets, and initiates the congestion control algorithm. This phenomenon may be even severer in LANETs because visible light links are easily blocked. Frequent blockage will further introduce dynamic changes of the topology. Therefore, how to design blockage-aware LANET transport protocols is challenging and substantially unexplored.

- **Application-Specific Transport Protocols.** LANETs have a great potential to support a diverse set of multimedia applications, and the transport layer protocols can be designed by considering the requirements of specific applications in terms of reliability, throughput, delay, mobility, energy efficiency, among others. For example, to ensure reliable delivery of key frames for video streaming, multiple-path transport protocol can be used and then transmit the packets of key frames through different paths; consequently, the probability of a whole key frame is dropped due to VLC link outage along multiple paths can be considerably reduced.

5.8 Cross-layer Design

In previous sections, we have discussed existing research work and remaining open issues at different layers of the network protocol stack of LANETs. The lessons learned from the discussions are that, the unique visible light communications impose both challenges and opportunities in the design of LANETs, and it calls for cross-layer design to address these challenges and to exploit the new opportunities. Next, we first classify existing research activities in cross-layer design in LANETs, and then point out future research directions.
5.8.1 Existing Cross-Layer Research Activities

- **Joint Link and Physical Layers.** The objectives of jointly considering link and physical layer in VLC networks design are to (i) improve the achievable throughput by designing channel-aware link layer transmission schemes [70] and multi-user channel access strategies [204–208]; (ii) mitigate the negative effects of visible light channels on link stability and availability, e.g., use intra-frame bidirectional transmission in favor of easier transmitter-receiver alignment [209], reduce the SNR fluctuations of VLC channels through LED lamp arrangement [210]; and (iii) enable seamless handover in VLC networks by accurately sensing mobile users [211].

- **Joint Network, Link and Physical Layers.** Network layer can be designed together with lower layer protocols to mitigate the limitations of VLC in transmission distance and directionality, and hence to extend the coverage and enhance the reliability VLC networks. In [212], WU et al. design a multi-hop multi-access VLC network, where the source node searches for a multi-hop path if the direct link is blocked; in [213], Liu et al. show that improved end-to-end delivery ratio can be achieved by using multi-path routing to account for the intermittent blockage problem of VLC links in vehicular visible light communication (V2LC) networks. It is shown that the capacity of VLC networks can be considerably enhanced by establishing multiple concurrent full-duplex paths to take the advantage of directional transmissions [214]. In [215], Ashok et al. propose a visual MIMO physical layer transmission scheme that has a great potential to extend the communication distance in mobile visual light networks; challenges imposed by visual MIMO on the design of MAC and Network protocol layers have also been discussed.

- **Joint Transport and Link Layers.** As discussed in Section 5.7, transport layer has been overlooked in existing literature with only few performance evaluation results reported [200] [203], and we believe it is an important research direction to incorporate transport layer into the cross-layer design of VLC networks.

It can be noticed that cross-layer optimization of VLC networks is still in its infancy, with most existing research focusing on simulation-/experiments-based performance analysis of protocols at different network layers [200] [203] [206] [208], or treating the cross-layer optimization problems heuristically without theoretically guaranteed optimality and convergence of the resulting cross-layer algorithms and protocols [210] [212] [215]. To date, there is still no mature systematic methodologies that can be used to design cross layer network protocols for infrastructure-less visible light communi-
CHAPTER 5. LANET: VISIBLE-LIGHT AD HOC NETWORKS

cation networks, which we believe is a key research direction towards LANETs. Next, we discuss the challenges with cross-layer design for LANETs based on software-defined networking (SDN), a newly emerging network design architecture.

5.8.2 Open Research Issues: Software-Defined LANETs

The notion of software defined networking (SDN) has been recently introduced to simplify network control and to make it easier to introduce and deploy new applications and services as compared to classical hardware-dependent approaches [216]. The main ideas are (i) to separate the data plane from the control plane; and (ii) to introduce novel network control functionalities that are defined based on an abstract and centralized representation of the network. Software defined networking has been envisioned as a way to programmatically control networks based on well-defined abstractions.

So far, however, most work on SDNs has concentrated on commercial infrastructure-based wired networks, with some recent work addressing wireless networks. However, applications of software-defined networking concepts to infrastructureless wireless networks such as LANETs are substantially unexplored. The reasons are multi-fold:

- Essentially, the distributed control problems in LANETs are much more complex and hard to separate into basic, isolated functionalities (i.e., layers in traditional networking architectures). Similar to traditional wireless ad hoc networks [132][133][217][218], as discussed above in this section, control problems in LANETs involve making resource allocation decisions at multiple layers of the network protocol stack that are inherently and tightly coupled because of the shared wireless radio transmission medium; conversely, in software-defined commercial wired networks one can concentrate on routing at the network layer in isolation.

- Moreover, in the current instantiations of this idea, SDN is realized by (i) removing control decisions from the hardware, e.g., switches, (ii) by enabling hardware (e.g., switches, routers) to be remotely programmed through an open and standardized interface, e.g., Openflow [219], and (iii) by using a centralized network controller to define the behavior and operation of the network forwarding infrastructure. This unavoidably requires a high-speed fronthaul infrastructure to connect the edge nodes with the centralized network controller, which is typically not available in LANETs where network nodes need to make distributed, optimal, cross-layer control decisions at all layers to maximize the network performance while keeping the network scalable, reliable, and easy to deploy.
Clearly, these problems cannot be solved with existing approaches, and calls for new approaches following which one can design protocols for LANETs in a software-defined, distributed, and cross-layer fashion.

5.9 Summary

In this paper, we studied the basic principles and challenges in designing and prototyping visible-light ad hoc networks (LANETs). We first examined emerging visible light communication (VLC) techniques, discussed how VLC can be used to enable a diverse set of new applications, and analyzed the main differences between LANETs and traditional MANETs. We then examined currently available VLC devices, testbed and existing physical and MAC layer protocols and the related standardization activities at these two layers. In network layer, we discussed the challenges in route establishment caused by the directionality of visible light link and its narrow FOV, and in transport layer we compared existing congestion control protocols and pointed out that none of them can suit well in LANETs. Finally, we pointed out that it is essential to develop a systematic cross-layer design methodology towards unlocking the capacity of wireless ad hoc networks via LANETs, and the challenges to accomplish software-defined LANETs were also discussed.
Chapter 6

Conclusion

This dissertation studied new wireless technologies for next-generation IoT. We focused on two tasks: (1) low-power low-complexity algorithms design for resource-constrained IoT devices, and (2) new wireless technology investment, i.e., VLC to alleviate spectrum crowded problem from the perspective of Internet.

In Chapter 2, we proposed a novel joint decoding algorithm for independently encoded compressively-sampled multi-view video streams. We also derived a blind video quality estimation technique that can be used to adapt online the video encoding rate at the sensors to guarantee desired quality levels in multi-view video streaming. Extensive simulation results of real multi-view video traces show the effectiveness of the proposed fusion reconstruction method with the assistance of SI generated by an inter-view motion compensation method. Moreover, they also illustrate the blind quality estimation algorithm can accurately estimate the reconstruction quality.

In Chapter 3, a new independent encoding independent decoding architecture for compressive multi-view video systems, composed of cooperative sparsity-aware block-level rate-adaptive encoders, limited feedback channels and independent decoders. A network modeling framework is also proposed to minimize the power consumption. Extensive performance evaluation results show that the proposed coding framework and power-minimizing delivery scheme are able to transmit multi-view streams with assured video quality at lower power consumption.

In Chapter 4, mathematical model of the cooperative visible-light beamforming (LiBeam) problem for indoor visible light networks is proposed, presented as maximizing the sum throughput of all VLC users. A networking testbed based on USRP X310 software-defined radios is developed. Simulation and experimental performance evaluation results indicate that 95% utility gain can be achieved compared to suboptimal network control strategies.
CHAPTER 6. CONCLUSION

In Chapter 5 we proposed a typical architecture for visible-light ad hoc networks (LAN-ETs). Application scenarios, enabling technologies and protocol-based design principles, and open research issues are discussed.

In my future research, I will continue studying new technologies for next-generation IoT from the perspective of low-complexity, low-power and new available spectrums.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY

International Workshop on Mobile Computing Systems and Applications (HotMobile), New Mexico, USA, February 2015.


BIBLIOGRAPHY


[174] T. V. Pham and A. T. Pham, “Max-Min Fairness and Sum-Rate Maximization of MU-VLC Local Networks,” in Proc. of IEEE Globecom Workshops (GC Wkshps), San Diego, CA, USA, Dec 2015.

BIBLIOGRAPHY


129
BIBLIOGRAPHY


BIBLIOGRAPHY


132