University of Southampton Research Repository

UNIVERSITY OF SOUTHAMPTON

Signal Processing for Ultra Fast Next Generation Metallic Access Network

by

Yangyishi Zhang

Supervisor: Prof. Lajos Hanzo

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Engineering and Physical Science
School of Electronics and Computer Science

January 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCE
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Yangyishi Zhang

With the escalating demand for high speed, high reliability, low latency, low cost and ubiquitous connectivity, the telecommunications industry is entering a new era where the ultimate optimality of the current wireline-wireless access network has to be achieved. Regarding the current wireline network paradigm, dominated by the copper-based digital subscriber lines (DSL) technology, multi-Gigabit data rate is the ambitious design objective at the customer end for the forthcoming ITU-T G.mgfast standard. In order to prepare for the new challenges in the era of total network convergence, both the wireline and the wireless community must be able to think beyond their respective conventions and learn from each other if necessary. Overall, the current DSL-based wireline network architecture is prone to the mutual interference resulting in far-end crosstalk (FEXT). The recently expanded 424/848 MHz spectrum of the ambitious G.mgfast project introduces far higher FEXT than that over the current 212/30 MHz G.fast/VDSL2 band. Additionally, the coexistence of multiple standards will also cause 'alien' FEXT. In the face of these new challenges, most state-of-the-art techniques, such as linear transmit precoding or Reed-Solomon coding, fail to satisfy the escalating performance requirements. In this thesis, we conceive a general architecture for next-generation DSL networks based on recent technological advances in key areas of access networks. More specifically, we conceive a dynamic spectrum management (DSM) aided signal processing framework relying on vectored transmission, spectrum balancing and error control.

Specifically, the excessive FEXT cannot be combated by low-complexity linear transmit precoding for downstream vectoring. However, as shown in the field of wireless communications, using lattice reduction as a signal space remapping technique significantly improves the performance of traditional multi-user detectors (MUD) and of the respective multi-user precoders (MUP). These promising techniques have largely remained unexploited in commercial wireless communications, due to their complexity in the face of the rapidly fluctuating wireless channels. However, they are eminently suitable for quasi-static copper channels. Hence we design powerful lattice-reduction-aided MUP

(LRMUP) as well as an optimal sphere encoder and its low-complexity variants. Furthermore, we propose the concept of vectoring mapping regions (VMR) as the transmitter-side dual counterpart of the classic *decision regions* conceived for receiver-side signal detection. Based on the VMR, we propose a near-optimal multi-level DSM algorithm relying on the family of LRMUPs, which strikes a more favourable sum rate vs. fairness trade-off than the state-of-the-art non-linear Tomlinson-Harashima precoder (THP).

On the other hand, impulsive noise (IN) contamination constitutes another challenge in next-generation DSL networks. In particular, the copper loops used for transmitting DSL data may become contaminated by other high-power undesired wireless signals such as IN and radio-frequency interference (RFI). As a counter-measure, we amalgamate forward error correction (FEC) with sophisticated automatic repeat request (ARQ) techniques for conceiving hybrid-ARQ (HARQ) schemes. Given that low density parity check (LDPC) codes are likely to be used in G.mgfast, we propose a novel low-complexity HARQ protocol based on the class of New Radio (NR) LDPC codes conceived for 5G cellular wireless systems, while exploiting the available *a posteriori* knowledge concerning the IN events at the receiver, in order to minimize the SNR degradation due to IN. Specifically, based on our proposed Tanner graph characterization as well as on the corresponding extrinsic information transfer (EXIT) chart, we propose an EXIT-chart-aided deferred iteration (DI) decoder switching technique, as well as a pair of MI-thresholding-aided curtailed iteration (CI) early stopping strategies, in order to minimize the potentially futile or redundant decoding operations.

Declaration of Authorship

I, Yangyishi Zhang , declare that the thesis entitled and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: [1], [2] and [3]

Signed:..

Date:...

Acknowledgements

I would like to sincerely thank my primary supervisor Prof. Lajos Hanzo and former supervisor Dr. Rong Zhang for their invaluable guidances, suggestions and critique throughout my entire research journey as a PhD student. They have significantly influenced my perspective on scientific research as a career and helped me tremendously to adapt to the rigorous and critical mindset of a true scientist. I could not have been on the right way to being a real scientist without their selfless support. Secondly, I would like to thank my industrial collaborator Dr. Anas F. Al Rawi at British Telecom, for his irreplaceable contributions from an industrial perspective, as well as for kindly providing me the internship opportunity at BT in order to help me gain a lot of industrial insight. Additionally, I would also like to thank Prof. Rob Maunder and Mr. Shao Shuai for the insightful discussions and for the invaluable collaborative effort in the study of channel coding in 5G systems. Furthermore, I would like to thank Prof. Lie-Liang Yang, Dr. Mohammed El-Hajjar, Dr. Jiankang Zhang and Dr. Tong Bai for their kind feedback with respect to the developing of this thesis' main contributions. Finally, I would like to express my deepest gratitude to my family who have been giving me the most moral and financial support throughout my entire academic life as a PhD student.

To my parents,
Prof. B. Zhang and Prof. X. Yang,
and my fiancée, L. Lin

Contents

List of Acronyms

ACK	Acknowledgement
ACM	Adaptive Coding and Modulation
AI	Artificial Intelligence
APP	A Posteriori Probability
APPLE	Approximate Perturbation Aided Lattice Encoder
ARQ	Automatic Repeat reQuest
ATP	Aggregate Transmit Power
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BGIN	Bernoulli Gaussian Impulsive Noise
BLER	Block Error Rate
CAPEX	Capital Expenditure
CC	Chase Combining
CI	Curtailed Iteration
CIR	Carrier-to-Interference Ratio
CN	Check Node
CO	Central Office
CP	Cyclic Prefix
CPE	Custom Premise Equipment
CR	Cyclic Redundancy
CRB	Cramer-Rao Bound
CSI	Channel State Information
CT	Communication Technology
CVP	Closest Vector Problem
CVPP	Closest Vector Problem with Preprocessing
DFE	Decision Feedback Equalization
DFSE	Depth-First Sphere Encoder
DI	Deferred Iteration
DMT	Discrete Multitone
DP	Distribution Point
DPC	Dirty Paper Coding
DSB	Dynamic Spectrum Balancing

DSL	Digital Subscriber Lines
DSM	Dynamic Spectrum Management
DSP	Digital Signal Processing
EC	Echo Cancellation
EMBB	Enhanced Mobile Broadband
EXIT	Extrinsic Information Transfer
EZF	Extended Zanatta-Filho
FD	Frequency Domain
FDE	Frequency Domain Equalizer
FDX	Full Duplexing
FEC	Forward Error Correction
FEXT	Far-End Crosstalk
FMCA	Fixed-Mobile Convergence Alliance
FSD	Fixed-complexity Sphere Decoder
FTTC	Fibre to the Cabinet
FTTdp	Fibre to the Distribution Point
FTTH	Fibre to the Home
FTTX	Fibre to the X
GFDM	Generalized Frequency Domain Multiplexing
HARQ	Hybrid Automatic Repeat Request
HKZ	Hermite-Korkine-Zolotareff
IAT	Inter-Arrival Time
IDFT	Inverse Discrete Fourier Transform
ISDN	Integrated Service Digital Network
ICI	Inter-Carrier Interference
IDEALE	Irregular Differential Evolution Aided Lattice Encoder
IFP	Integer Forcing Precoding
IN	Impulsive Noise
IR	Incremental Redundancy
ISI	Inter-Symbol Interference
ISP	Internet Service Provider
IT	Information Technology
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
IWF	Iterative Water-Filling
KBSE	$\mathcal{K}$-Best Sphere Encoder
LDGM	Low Density Generator Matrix
LDPC	Low Density Parity Check
LLL	Lenstra-Lenstra-Lovász
LLR	Log Likelihood Ratio
LLU	Local Loop Unbundling
LMS	Least Mean Square

LR-THP	Lattice-Reduction-aided Tomlinson-Harashima Precoding
LR-ZFP	Lattice-Reduction-aided Zero-Forcing Precoding
LRMUP	Lattice-Reduction-aided Multi-User Precoding
MAP	Maximum A Posteriori
MI	Mutual Information
MIMO	Multiple-Input-Multiple-Output
MLD	Maximum Likelihood Detection
MMSE	Minimum Mean Square Error
MUD	Multi-User Detection
MUP	Multi-User Precoding
NACK	Negative Acknowledgement
NEXT	Near-End Crosstalk
NFV	Network Function Virtualization
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplexing
ONU	Optical Network Unit
OSB	Optimal Spectrum Balancing
PAM	Pulse Amplitude Modulation
PAPR	Peak-to-Average Power Ratio
PCM	Parity Check Matrix
POTS	Plain Old Telephone Service
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
RCoF	Reverse Compute-and-Forward
REIN	Repetitive Electrical Impulse Noise
RFI	Radio Frequency Interference
RS	Reed-Solomon
SDN	Software Defined Networking
SER	Symbol Error Rate
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-and-Noise Ratio
SIVP	Shortest Independent Vector Problem
SLNR	Signal-to-Leakage-and-Noise Ratio
SNR	Signal-to-Noise Ratio
SP	Sum-Product
SQRD	Sorted QR Decomposition
SRA	Seamless Rate Adaptation
SSB	Static Spectrum Balancing
SVP	Shortest Vector Problem

TD	Time-Domain
TDD	Time-Division Duplexing
TDSL	Terabit DSL
THP	Tomlinson-Harashima Precoding
TxPSD	Transmit Power Spectral Density
URLLC	Ultra Reliable Low Latency Communications
V-BLAST	Vertical Bell-Lab Space Time
VCE	Vectoring Control Entity
VF	Vectoring Feedback
VMR	Vectoring Mapping Region
VN	Variable Node
VP	Vector Perturbation
WZC	Wyner-Ziv Coding
ZFP	Zero-Forcing Precoding

Chapter 1

The Next Generation Metallic Access Network

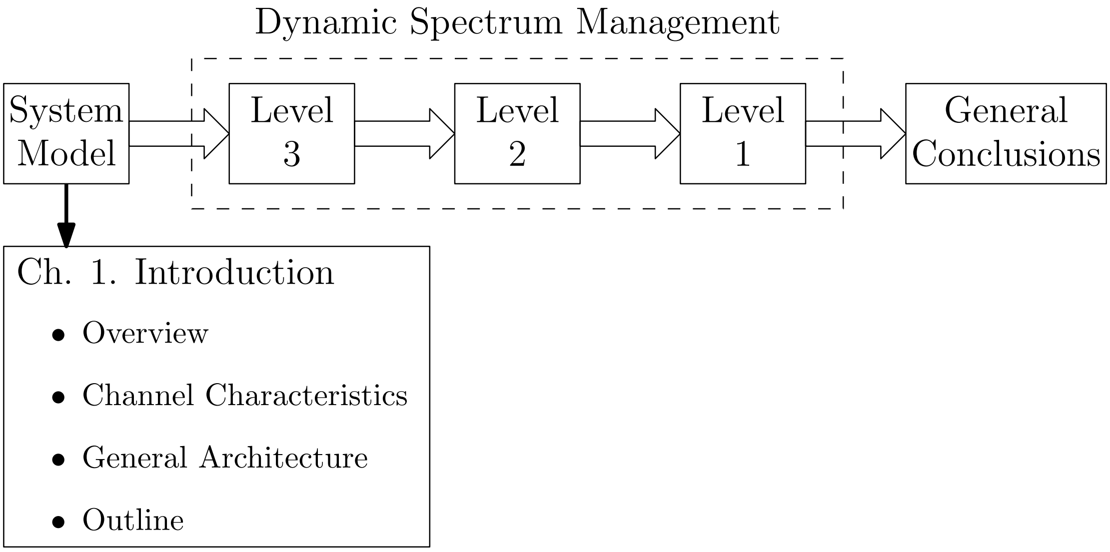


FIGURE 1.1: Outline of Ch. 1.

1.1 The Fixed-Mobile Convergence

With the arrival of the next generation cellular wireless standard and the emerging of the Internet of Things, the evolution of the communications network has reached a critical point, where the requirement for high speed, high reliability, low latency, low cost and ubiquitous connectivity requires seamless universal convergence of the currently fragmented network infrastructure. However, unifying the telecommunications network is not a new concept of the 5G era. First established in 2004, the Fixed-Mobile Convergence Alliance (FMCA), which consisted of more than 20 influential telecommunications network operators around the globe, made a dedicated effort to make this convergence a reality. Unfortunately, due to a paucity of customer demand back then, the alliance was eventually disbanded in the year 2010 not long after the launch of the 4G mobile networks.

However, unifying the wireline and wireless access networks goes way beyond reviving the legacy of the FMCA. The universal convergence entails a *multi-dimensional* [4] overhaul, spanning the entire existing communications ecosystem that aims for bridging the fragmented sections of conventional networks, such as copper and fibre in fixed wireline broadband [5], as well as the cloud and the edge in wireless cellular broadband [6]. Furthermore, the convergence is expected to ultimately lead to a seamless end-to-end system from wireline to wireless [7, 8, 9, 10]. On the other hand, converging the network infrastructure demands a corresponding convergence in the solution domain as well. Based on information theory, a seamless integration of information technology (IT) and communication technology (CT) is anticipated to be a future-proof solution for the telecommunications community [11]. The convergence of IT-CT may be facilitated by machine learning techniques [12, 13, 14] and software defined networking (SDN) [15, 16, 17, 18] for an *intelligent* and *flexible* next-generation network architecture.

The universal convergence of communication networks means that the performance bottleneck between each network-segment has to be eliminated for approaching the holistically optimal network performance. These requirements impose challenges on the future generations of both the wireline and the wireless broadband networks. In particular, the wireline network is typically tasked with high-rate ultra-reliable communication that covers significantly longer range than a wireless cellular network. The investigations conducted in [19][5] have explicitly demonstrated that the next generation wireline access network, providing distributed gateways and backhaul for mobile devices will be responsible for the majority of future access network traffic. Therefore, optimizing the performance of the next generation wireline network is extremely important both for wired broadband access itself, and also for other indoor wireless networks such as Wi-Fi and visible light communication systems.

1.2 Overview of Broadband Wireline Communications

The family of wireline access network implementations based on the digital subscriber lines (DSL) technology has been dominating the global fixed broadband market [24] since the dêbut of the original integrated service digital network (ISDN) in the 1980s. As a flexible technology, DSL is constantly evolving to meet the escalating demand for high-speed ultra-reliable communications. Initially deployed over the established plain old telephone service (POTS) network to exploit vacant baseband spectrum for Internet services, the wireline broadband network has developed through multiple generations distinguished by their performance metrics and architecture. According to the Interna-

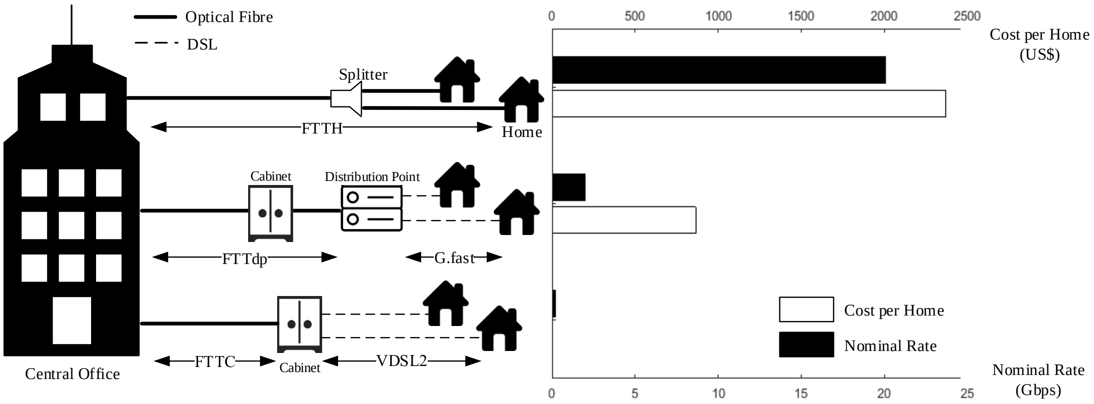


Figure 1.2: Nominal aggregate 'upstream+downstream' data rate and deployment cost for each FTTx-type hybrid wireline broadband architecture, showing the coexistence of FTTH (top), 'FTTdp + G.fast' (middle), 'FTTC + VDSL2' (bottom). An universally established VDSL2 environment, known as 'brownfield' [20] is assumed. The nominal rates are from their respective ITU-T standard [21, 22, 23].

tional Telecommunication Union (ITU), there are four main DSL generations (Fig. 1.3) which we can classify into two eras:

1.2.1 Legacy DSL Era

As the widely recognized ancestor of the DSL family, the specifications of the integrated services digital network (ISDN) were mostly established in 1976. ISDN was originally intended for unifying the global telephony and data communication networks. However, the implementation of ISDN in reality mainly targeted telephony services and low-speed data transmission, at a rate of approximately 160 kbps over the local telephone loop. The massive ISDN deployment around the world was eventually shown to hinder the evolution of broadband wireline communications, because ISDN was not future-proof in terms of supporting high-speed wireline access networks.

Prior to the introduction of the optical network unit (ONU), the first two universally defined generations of the DSL standards, HDSL (ITU-T G.991.x series) and ADSL (ITU-T G.992.x series), also entirely relied on copper-based local telephony loops from the central office (CO) all the way to the end users' customer premise equipment (CPE). Similar to ISDN, HDSL supported both telephony and basic data transmission at a rate of 1544 or 2048 kbps to the same customer, but not simultaneously, owing to the associated mutual interference. Compatibility with the POTS was not achieved until the standardization of discrete multi-tone (DMT) modulation in the ADSL series. DMT initialized the trend of spectrum expansion for all future DSL generations in order to attain ever-increasing data rates. However, due to the high propagation loss of the local loops, the legacy DSL architecture can only support a total data rate of 25 Mbps, covering a maximum radius of 1.5 km from the CO [25], which is considerably lower than the average requirement of 50-70 Mbps for typical ultra-high-definition TV streaming (using ITU-T H.265 encoding [26]). As a result, the legacy DSL technologies experienced a considerable decline in their global market share, except for Africa [24].

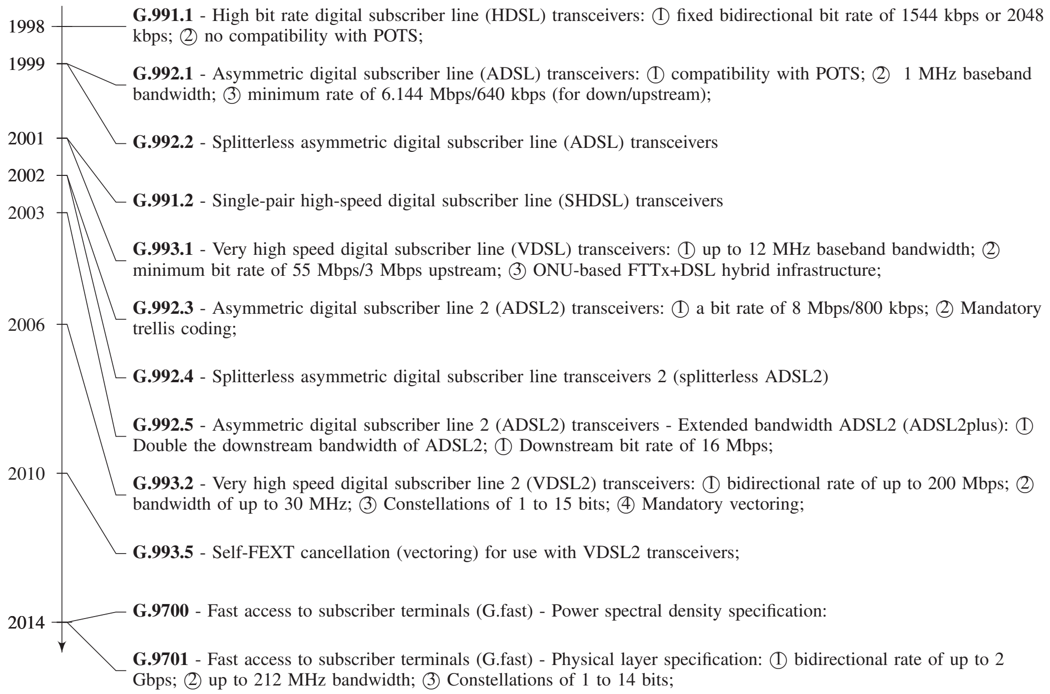


FIGURE 1.3: Timeline of the ITU-T DSL standards by date of initial publication. Based on the performance and architecture difference, there are four generations consisting of the HDSL (G.991 series), ADSL (G.992 series), VDSL (G.993 series) and G.fast (G.970x series).

1.2.2 Fibre to the X (FTTx) Era

The development of optical fibre technology has revolutionized the conventional wireline access networks. Utilization of the intermediate data relay sites known as the ONUs, which connect to the CO via fibre, significantly boosts both the data rate and reach of wireline broadband beyond the limit of the legacy DSL. The current wireline broadband ecosystem can generally be depicted as a coexistence of multiple fibre-copper hybrid deployments known as fibre-to-the-X (FTTx) (Fig. 1.2). In general, due to the superior reliability of optical fibre over conventional copper loops for long-distance communications, the ONU is gradually moving closer to the CPEs, resulting in a shift from the fibre-to-the-cabinet (FTTC) based VDSL (ITU-T G.993.x series) to the fibre-to-the-distribution-point (FTTdp) based G.fast (mainly ITU-T G.9701 [22]). In VDSL2, a total data rate of up to 200 Mbps with 2.5 km coverage has been achieved, whereas G.fast supports a total data rate of up to 2 Gbps for 100m loops and 200 Mbps for a maximum of 500m coverage.

In preparation for the multi-Gigabit requirement of the next generation access network, the general performance of wireline access networks should be fibre-like, which has led to massive fibre-to-the-home (FTTH) roll-out in developing countries such as China [24]. However, as investigated in [20], FTTH incurs considerably more capital expenditure (CAPEX) than the 'FTTdp + G.fast' hybrid model, even though the latter also requires an extensive deployment of distribution points (DP) due to the reduced coverage of G.fast. The rate versus cost trade-off depicted in Fig. 1.2 characterizes the average case over different geotypes [20], where in rural and suburban areas a more significant cost reduction can be achieved by using FTTdp. Therefore, FTTH deployment is highly situational and it is in general not an economical solution. Furthermore, the advent of powerful communication processors [27] and network function virtualization (NFV) [28] creates the possibility of reaching fibre-like performances using the existing FTTx architecture. The proposition of Cioffi in [29] suggested that new transmission modes in the millimetre-wave band accompanied by enhanced digital signal processing (DSP) techniques can potentially achieve Terabit performance target (i.e. TDSL) using the DSL framework.

1.3 Channel Characteristics

Despite the fact that copper-based communication systems have had a long history within the telecommunications industry thanks to the popularization of POTS and DSL, the understanding of the twisted-wire channel is still limited. Due to the renewed interest in boosting the performance of DSL, alternative signal propagation modes over twisted copper are being investigated by both the research community [29, 30, 31, 32] and the standardization body [33], such as the phantom mode and the waveguide mode. Although the quasi-static nature of a DSL channel is intuitive, the modelling of its transfer characteristics is far from trivial, especially with the discovery of new operational modes. This section is dedicated to a comparative overview of the transfer characteristics of the copper channel for cutting-edge DSL.

1.3.1 Direct Channel

The premise of conventional twisted-copper communication methods is the two-port network model relying on the classic transmission line theory. In the simplest case, a pair of twisted copper wires, i.e. a copper pair, carrying differential voltages constitutes a *direct channel*[1]. Given the length of the copper pair d, as well as the propagation constant $\gamma(f)$ of the two-port network as a function of the differential voltage's frequency f, then

---

[1] In this treatise, we make the realistic assumption that all connections other than those directly associated with CPEs consist of optical fibre as seen in Fig. 1.2, therefore the multi-segment DSL case discussed in [34] can be skipped.

the transfer function $H(f;\, d)$ of this *single-pair* channel can be formulated as:

$$H(f;\, d) = e^{-d\gamma(f)}, \tag{1.1}$$

if the pair is perfectly terminated. In general, $H(f;\, d)$ is a gradually decaying function of both f and d[2]. Many existing DSL channel models proposed for the VDSL/VDSL2 band are inapplicable both to G.fast and to the next generation G.mgfast band [35], mainly because they do not consider the substantial change of the twisted pair's self-coupling characteristics at high frequencies, where the signal wavelength and the copper pair's twist length become comparable. In practice, these (average-case) channel models are not suitable candidates for the network operator. Because they do not exactly match the actual channel transfer function of particular copper pairs, they cannot be used to assess the critical worst-case performance and they cannot be mitigated by time-diversity[3]. For this reason, channel measurements are also extensively used by the DSL research community in order to acquire realistic performance estimates.

1.3.2 Crosstalk Channel

In areas that are close to the DP, the copper pairs connecting CPEs from different houses are bundled together as a large *multi-pair* DSL binder. However, because the POTS network was originally designed for carrying 3.4 kHz voice signals, the high-frequency signals of DSL may cause considerable electromagnetic leakage and unintended coupling between different pairs within a DSL binder. Due to the binder sheath which reduces the alien interferences from outside sources, the unintended coupling *inside* a binder becomes the dominant source of *multi-user interference* and constitutes the *crosstalk channel*. In particular, the forward interference travelling in the same direction as the direct channel signal is termed the *far-end crosstalk* (FEXT), whereas the returning interference travelling in the opposite direction of the direct channel signal is termed the *near-end crosstalk* (NEXT) (Fig. 1.4). We note that there are in fact many other *external* sources of interference in multi-pair DSL, most of which are however neither measurable nor static. Therefore, they do not constitute interfering *channels* and are more widely classified as *noise*, with the rare exception of alien crosstalk both from other DSL binders [36, 37, 38] and from the co-existing power line communication systems [39][40].

The mathematical modelling of crosstalk channels is considerably different from that of direct channels. In general, the coupling characteristics between adjacent copper pairs is not universally deterministic, subject to the exact geometry of the DSL binder's interior, as well as to the dielectric behaviour of the binder's sheath. Hence, the frequency response

---

[2] The length of a copper wire is constant after being placed, therefore d is usually considered a 'parameter' of the transfer function.

[3] Time-diversity is a common phenomenon in wireless communications where the severity of the worst-case performance is naturally mitigated due to the time dependence of wireless channels.

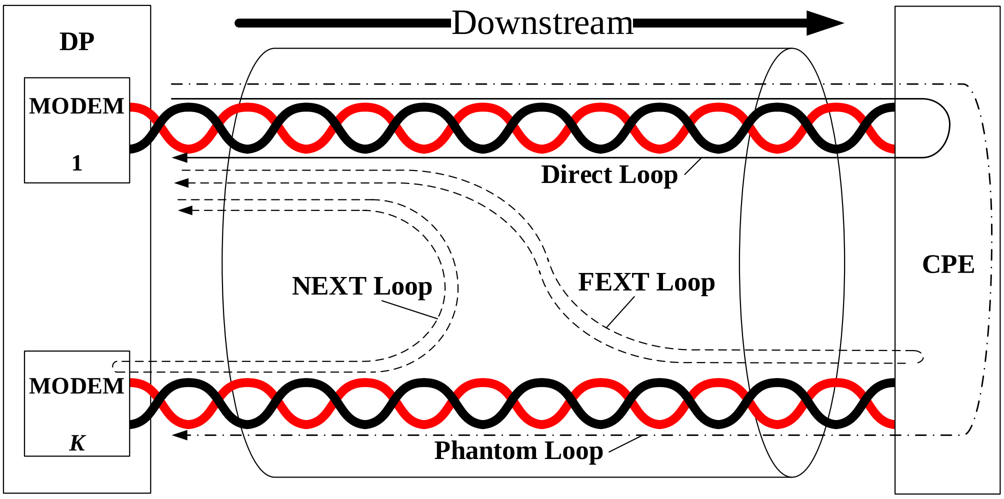


FIGURE 1.4: Virtual local loop illustration of DSL channels operating in transmission line mode. Both the forward path and the return path are shown for each virtual loop. Signals transmitted by modem 1 are received at the end of each virtual loop's forward path. The CPE must be a single entity or a set of coordinated entities in order to utilize the phantom loop.

of crosstalk channels is commonly modelled by stochastic functions [41][42], which can then be used to generate a particular crosstalk channel realization. Regardless, general practical bounds were established using the '1% worst-case' power spectral density (PSD) model for FEXT and NEXT[4], formulated respectively as [43]:

$$|H_F(f;\,d)|^2 = |H(f;\,d)|^2(\frac{K-1}{49})^{0.6}(9\times 10^{-20})d\cdot f^2\,, \tag{1.2}$$

$$|H_N(f;\,d)|^2 = (\frac{K-1}{49})^{0.6}\cdot 10^{-13}f^{1.5}\,, \tag{1.3}$$

where it is shown that the PSD of crosstalk channels increases with the number of encapsulated copper pairs K. For FEXT, the interfering signal must travel the full length of the DSL binder, whereas NEXT is a form of returning interference, as seen in Fig. 1.4. In general, the gain of the NEXT channel diminishes with the distance from the transmitter. Eq. (1.3) represents the integral of this diminishing returning interference over the full length d of the DSL binder, and thus the effect of d is eliminated in the model. On the other hand, the channel gain of FEXT over the direct channel $|H_F(f;\,d)|^2/|H(f;\,d)|^2$, known as the equal level FEXT [44], is shown to increase with the frequency in Eq. (1.2), implying the deterioration of multi-user interference at high frequencies (Fig. 1.5). We refer the readers to further references for more in-depth modelling of crosstalk channels [44, 45, 46]. The excessive level of FEXT at high frequencies has led to some of the most pronounced research challenges in the wireline communications community.

---

[4] The crosstalk channel PSD predicted by this model is guaranteed to be lower than 99% of the realizations.

1.3.3 Phantom Channel

The phantom and common mode signalling [31, 47, 48], which have not been assessed by the standardization bodies until recently [33], relies on the hardware infrastructure of the well-established multi-pair bonding technology [49, 50, 51]. Akin to the multi-antenna transceivers of wireless communications, corporate buildings may have access to a whole binder of copper pairs for a set of coordinated CPEs or a single large CPE such as a server. It is plausible that multi-pair bonding can be used to glean multiplexing gain from the additional spatial dimensions. However, the main benefit of multi-pair bonding is that of obtaining additional virtual channels beyond the physical direct channels and the crosstalk channels.

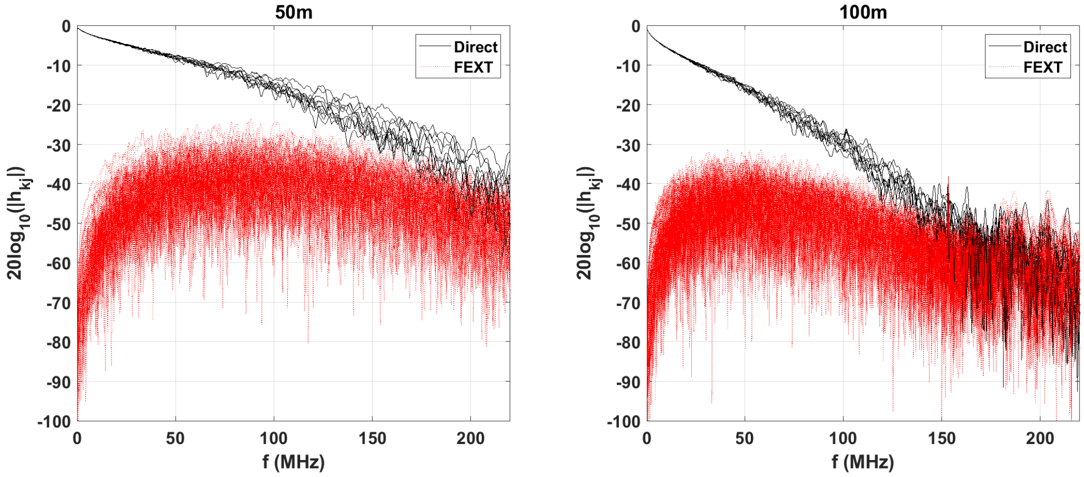


FIGURE 1.5: Frequency response of the direct link channels and FEXT link channels for the forthcoming 212 MHz G.fast profile of a 50-meter (left) and a 100-meter (right) cable, each containing 10 twisted pairs.

For a binder of K copper pairs, there are in theory K *common-mode* direct channels in addition to the K differential-mode direct channels, if each of the $2K$ wires is *virtually* paired with a different wire from another adjacent copper pair [30]. Moreover, the common mode signals, constituted by the sum of two voltages of the same sign, should inherently have improved the energy efficiency because they carry higher power than the differential mode signals. However, the proposed common mode operation [30] requires heavy modifications to the DSL binders, which makes it impractical. Instead, a better approach to exploiting the ‘hidden’ circuits of a multi-pair binder is the proverbial ‘phantom-mode’ transmission.

For a two-pair binder, there exists a third, ‘phantom’ channel which consists of the two existing copper pairs as its ‘wires’. The phantom-mode signal equals to the voltage difference between the means of the two differential voltages carried by the two physical copper pairs [31] (Fig. 1.6). For a K-pair binder, we can utilize up to $K - 1$ phantom channels by pairing the physical channels. Therefore, phantom channels can be modelled as conventional transmission lines similar to the physical direct channels, and the multi-pair bonding technology may also benefit the hybrid group of channels for gleaning more multiplexing gain [52]. On the other hand, phantom channels are orthogonal to the physical channels and therefore the phantom mode does not interfere with the usual differential mode. The phantom mode channel is depicted in Fig. 1.4.

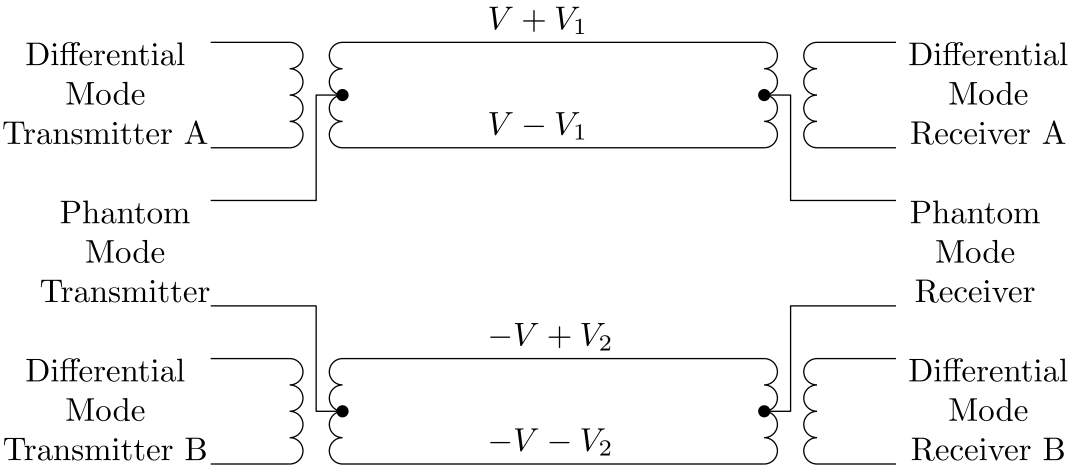


Figure 1.6: The phantom circuit in a two pair system. Here, the physical differential channel A carries a voltage of V_1 whilst channel B carries V_2. By taking the difference of each channel's common mode voltage, the virtual circuit becomes a non-interfering third channel carrying a differential voltage of V

1.3.4 Plasmon Polariton Channel

The state-of-the-art DSL technologies relying on channel models of the previous types have limited performance even under the G.mgfast specification, compared to the achievable rates of FTTH systems, as seen in Fig. 1.2. In essence, the superiority of optical fibre comes from its role as an unconventional waveguide rather than a straightforward two-port transmission line in terms of the propagation of electromagnetic waves. Fortunately, it was recently shown that metal wires may be used for signalling in *surface plasmon polariton* mode (Fig. 1.7) at the Terahertz (THz) frequency band [53][54] in order to imitate the behaviour of visible light in optical fibre. The pioneer study of Cioffi *et al.* in [29] has attempted to utilize this particular waveguide signalling mode within the multitude of existing copper wires in DSL binders in order to realize the ultra fast TDSL and therefore align the copper access network performance with that of the fibre access networks.

Due to the complete change of signal propagation mode, the transfer characteristics associated with this waveguide channel are dramatically different from that associated with the transmission line channel, even though their transfer functions may depend on the same set of variables (1.1) [29]. In the THz band, the wavelength of electromagnetic waves falls in the millimetre range, which is comparable to the radii of the existing copper

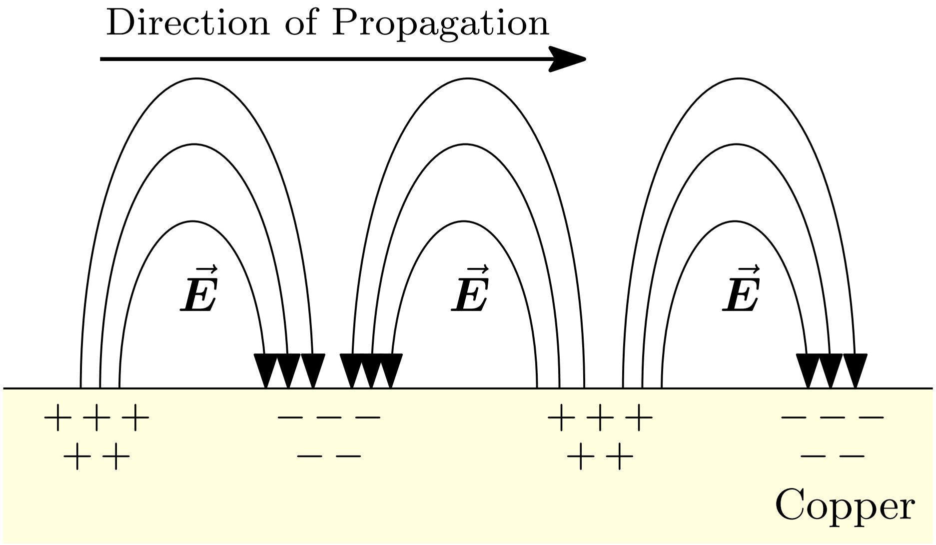


FIGURE 1.7: Surface plasmon polariton guided by copper (adapted from [55]).

pairs. Therefore, THz signals can travel along the surfaces of copper wires (plasmon) as well as in the surrounding free space (polariton). For free space electromagnetic waves guided by the copper wires in DSL binders, the other closely spaced wires of the binder, as well as the metallic sheath if there is one, may be used as reflectors. These natural reflectors can prevent surface waves from leaking when the designated waveguide bends, and hence the total internal reflection effect of optical fibres may theoretically be recreated in TDSL. The physical modelling of these effects are more thoroughly investigated in [56, 57, 58].

In general, the new operational modes of DSL are not sufficiently investigated at the moment even under laboratory conditions. The significant potential predicted for the somewhat distant future is far from its theoretical completeness and industrial fruition. We will therefore not consider these new types of channels in further details. However, we should bear in mind that the performance of metallic access networks is ultimately limited by the inherently low capacity of conventional transmission line channels. Therefore, further research on revolutionary signalling modes would have immense practical value for next-generation metallic access networks.

1.4 General Architecture

The overall hostile environment of DSL access networks is the combined result of severe frequency-selectivity, intensive self-crosstalk and coloured noise. Additionally, because of the limited computation power of signal processors in the past, the suboptimal performances of legacy and current DSL standards are also largely due to inefficient exploitation of the available spectrum. The channel gain is an exponentially decaying function of the binder length d (Eq. (1.1)), which hence results in the critical *rate vs. reach* trade-off,

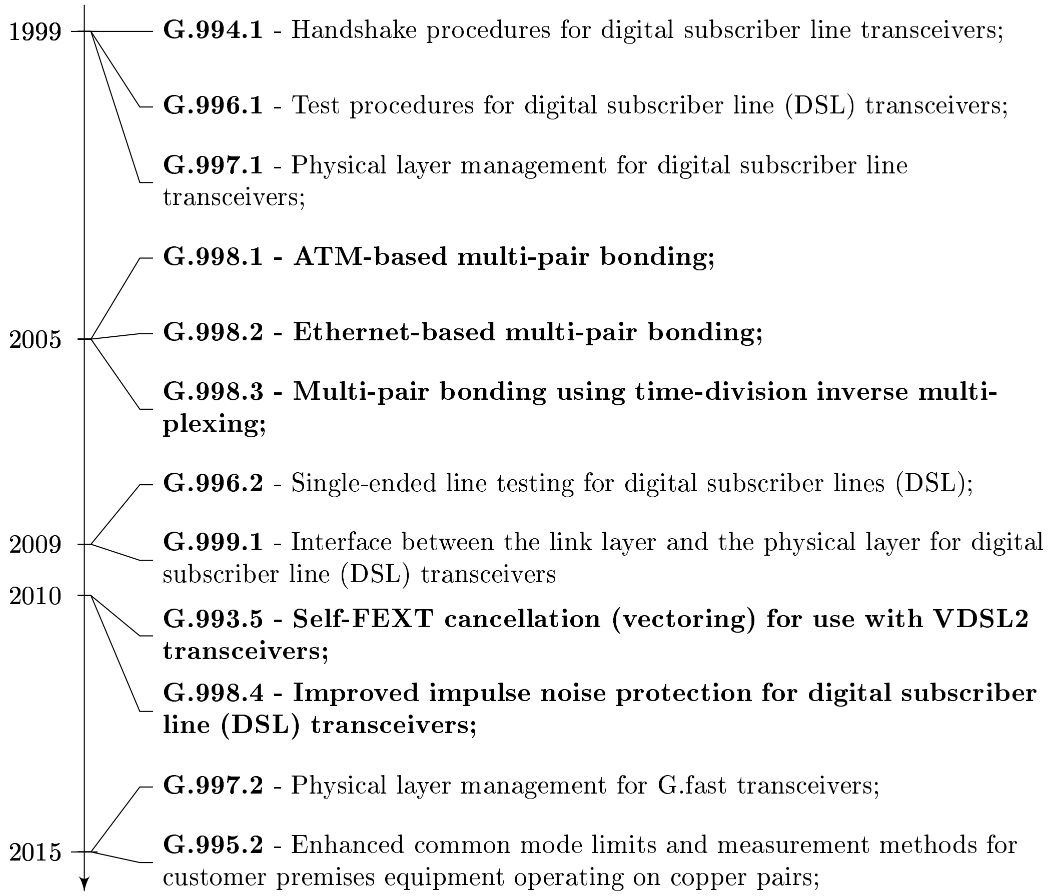


FIGURE 1.8: Timeline of the ITU-T DSL standards for supplementary technologies by date of initial publication. Bold entries represent technologies which are used to boost bandwidth efficiency.

as well as in poor fairness for CPEs far away from their parent DP.

Overcoming the performance bottleneck of the telephony-based DSL networks requires a cohesive set of signal processing components for each stage of the networks. The ITU-T G.99x series established a family of supplementary technologies that are designed for improving the bandwidth efficiency and quality of service (QoS) of DSL systems, which is highlighted in Fig. 1.8. In this section, for the completeness of our system design, we will critically appraise their respective state-of-the-art as well as review the technological advances, which are under consideration for both next-generation wireless and future wireline access networks (Fig. 1.9). From an industrial and a commercial perspective, the expected technological advances in the forthcoming generation of wireline access networks are investigated in further details in [59][60].

We should note that some of the signal processing techniques of wireline communication systems are conceptually similar to those of their wireless counterpart, albeit for historical reasons they may use different terminologies. In particular, the commercial metallic access network is principally similar to a multi-user spatial multiple-input-multiple-output (MU-MIMO) system, where each user also employs multi-carrier modulation, with the

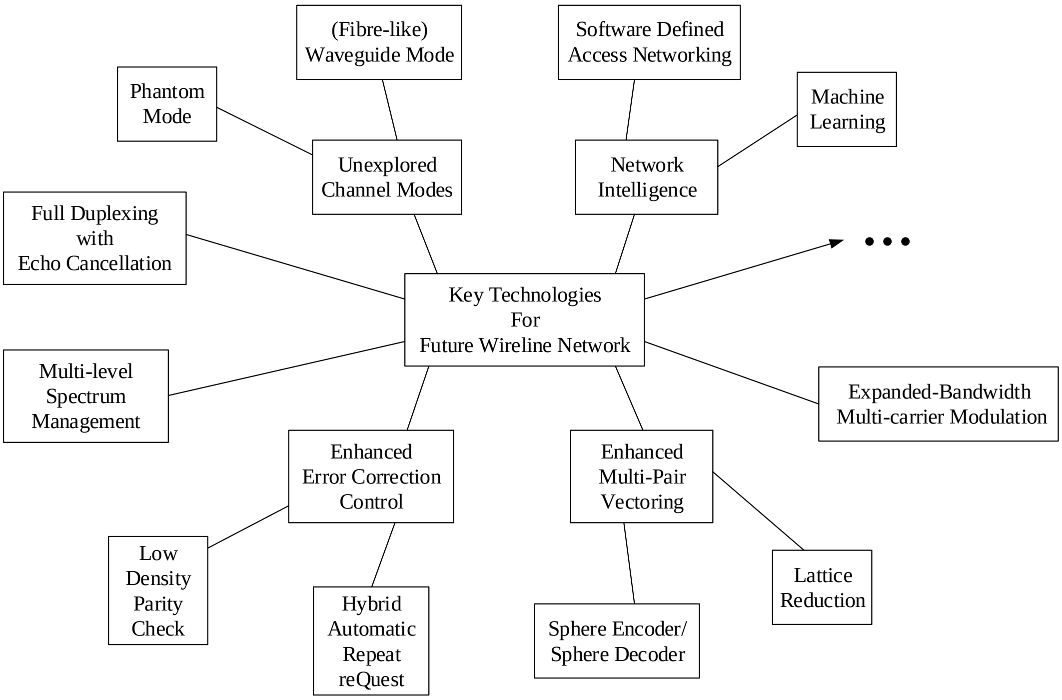


FIGURE 1.9: Key technologies in consideration for G.mgfast and/or future generation wireline access network.

downstream being a broadcast system and the upstream being a multiple access system. More specifically, the bit stream associated with each user is firstly encoded by error correction codes. The encoded bit stream is then mapped to quadrature amplitude modulated symbols, which are subsequently modulated onto the orthogonal frequency division multiplexing (OFDM) subcarriers. Since the metallic access network operates in the baseband spectrum dispensing with an RF carrier, in contrast to its wireless counterpart, the subcarriers described here are typically termed as *tones* and this particular multi-carrier scheme is referred to as the discrete multi-tone modulation in metallic wireline communications. Moreover, the family of dynamic spectrum management techniques investigated in the later chapters of this treatise constitute the dual counterparts of multi-user interference cancellation (cf. Ch. 2) as well as of adaptive modulation (Ch. 3) and coding (Ch. 4) schemes.

1.4.1 Modulation

From the ADSL era onwards, DSL broadband access became compatible with the POTS. The enabling technology accounting for this feature is the DMT modulation scheme [43], whose passband variant is more commonly known as OFDM in wireless and as orthogonal wavelength-division multiplexing (OWDM) in optical communications community. Additionally, the severe frequency-selectivity of DSL ((1.1)(1.2)(1.3) and Fig. 1.5), which would otherwise cause strong inter-symbol interference (ISI), can be conveniently removed by DMT, resulting in a multitude of frequency bins having small bandwidth and negligible frequency-selectivity, which are known as the *tones*. Each tone carries a trellis-coded subsymbol from a designated constellation. In G.fast, the constellation is associated with a given order of quadrature amplitude modulation (QAM). Since DMT operates in the baseband, the absence of carrier-related problems such as frequency and phase offset allows a choice of up to 2^{15}-QAM [61].

In the physical layer architecture characterized by Fig. 1.10, the downstream transmission of a typical K-pair T-tone G.fast system is shown, assuming that each CPE is connected to the DP via a single twisted pair. The end-to-end system operates in frequency domain, whilst the transmission segment between the front ends of the DP and the CPEs is in time domain. The role of the MUP and the frequency-domain equalizers (FDE) will be discussed in the following chapters. In order to guarantee tone orthogonality and therefore avoiding inter-carrier interference (ICI), each time-domain subsymbol, obtained via the inverse discrete Fourier transform (IDFT) of the frequency-domain subsymbols, must be transmitted with a sufficiently long cyclic prefix (CP) attached [62]. Upon reception of the time-domain DMT symbol, the CP is removed before the remainder of the DMT symbol is transformed back into frequency domain by DFT. In principle, the duration of CP should be at least identical to the DSL channel's delay spread, which is true in G.fast. Therefore, it is reasonable to assume that the top system of Fig. 1.10 consists of T independent layers of frequency-domain systems having K users each.

Due to the uncertainty of the channel characteristics at very high frequencies far above the current 212 MHz G.fast ceiling, e.g. for 424 or 848 MHz G.mgfast profiles, as well as to the strict power constraints over these bands[5], the enormous peak-to-average-power ratio (PAPR) of conventional DMT with long CP overhead might not be tolerable for future performance requirements. For the sake of good power efficiency, the generalized frequency-division multiplexing (GFDM) technique [64] may be used as an alternative, due to its reduced CP usage and lower implementation complexity compared to OFDM/DMT yet using a similar structure (as showcased in [65]). Furthermore, zero ICI is still achievable in GFDM if Dirichlet filters are used [66]. Nonetheless, the assumption that each frequency-domain channel is orthogonal holds valid. However, with the degradation of channel quality over both the binder length and the tone frequency, the average performance achieved by DMT or GFDM will eventually reach a critical point, where further spectrum expansion stops contributing to the increase of sum rate. The authors of [67] demonstrated that the total QoS will suffer even with the aid of capacity-approaching turbo codes, when the spectrum is expanded beyond a critical frequency.

---

[5] The power constraints of DSL are mostly enforced for electromagnetic compatibility with radio broadcast services as seen in [63], therefore we may anticipate similar conditions for the currently undefined bands.

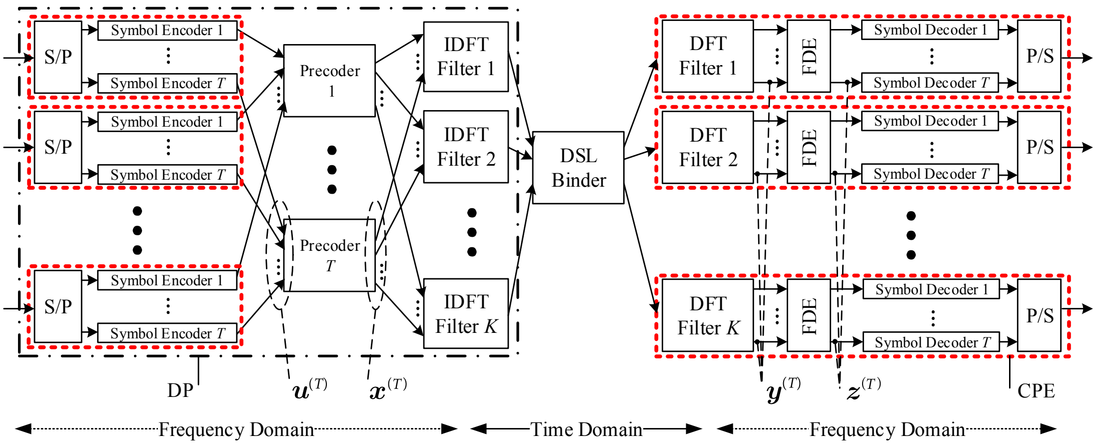


FIGURE 1.10: Physical layer architecture of multi-user G.fast (Copyright [22])

1.4.2 Vectoring

Besides frequency selectivity, which is mitigated by DMT modulation, another implication of Fig. 1.5 is the fast decrease of the carrier-to-interference ratio (CIR), which has already caused interference problems in the wideband DSL deployment of VDSL2 and particularly in G.fast. Since the modems at the ONU (cabinet or DP) are co-located, and usually coordinated by the same Internet service provider (ISP), they may invoke FEXT removal mechanism referred to as *vectoring* (or *vectored transmission*) [68], subject to the availability of channel state information (CSI). Due to the orthogonality of DMT modulated subsymbols, the system diagram of Fig. 1.10 may be decoupled into T instances of K-user subsystems, which we formulate as:

$$\boldsymbol{y}^t = \boldsymbol{H}^t \boldsymbol{x}^t + \boldsymbol{n}^t \quad (t = 1, 2, ..., T), \tag{1.4}$$

where the $K \times 1$ vectors $\boldsymbol{y}^t$, $\boldsymbol{x}^t$ and $\boldsymbol{n}^t$ represent the received symbol vector, transmitted symbol vector and the noise vector of tone t. If $\boldsymbol{n}^t$ is 'white', it has the same PSD across all frequency tones. However, in practice there might be other sources of noise which are not white. The $K \times K$ MIMO transfer matrix $\boldsymbol{H}^t$ is the frequency domain multi-user DSL channel of the tth tone. In DSL, the diagonal entry $h^t_{k,k}$ in $\boldsymbol{H}^t$ is the direct channel gain associated with pair k, whilst the off-diagonal elements $h^t_{k,l}$ for $k \neq l$ represent the FEXT channels contaminating pair k. With the emerging of G.mgfast, the DSL network becomes more vulnerable to alien FEXT due to the coexistence of multiple standards, as well as to the typical local loop unbundling (LLU) problem investigated in [69]. The general definition of a vectored transmission group does not deal with alien FEXT. However, the power of SDN/NFV has the promise of flexible traffic management and signal coordination between multiple ISPs [17][70] in the future, so that vectoring may be expanded to remove 'alien' FEXT (which will then become domestic).

In the broader context, vectoring is defined as the highest level (level-3) of the holistic dynamic spectrum management (DSM, Tab. 1.1) paradigm for DSL networks. Overall, level-1 error control relying on both forward error correction (FEC) codes (with interleaving) and hybrid automatic repeat request (HARQ) provides essential standalone protection against coloured background noise. On the other hand, spectral load balancing (level-2 strategy) is often used collaboratively with vectoring (after its introduction in ITU-T G.993.5) for both sum rate optimization and for combating the crosstalk of DSL systems. In the literature, their joint implementation is sometimes simply referred to as vectoring, which is reminiscent of the jointly optimized power and rate allocation under multi-user beamforming in wireless systems.

TABLE 1.1: Dynamic Spectrum Management Level Definitions [71]

Level	Functionality
0	N/A
1	Single-pair-based Error Control Configuration (cf. Ch. 4 and also in [72])
2	Multi-pair-based Spectral Load Balancing and Optimization (cf. Ch. 3)
3	Multi-pair-based Vectored Signal Transmission and Reception (cf. Ch. 2)

By definition, the transmission link spanning from the CPE to the DP/CO is called the upstream, whilst the link from the DP/CO to the CPE is termed as the downstream. During downstream transmission, FEXT is pre-cancelled by the MUP, while in its upstream counterpart the FEXT channel is equalized by the MUD. In DSL deployments, downstream vectoring is typically valued more than its upstream counterpart, because the *co-located* downstream transmitters interfere with each other immediately at the output port without being attenuated by the copper. This results in far stronger mutual coupling than the CPEs do in the upstream scenario. We will mainly consider downstream vectoring throughout the remainder of this thesis, unless explicitly stated otherwise.

We simplify the discussion in the remainder of this section by neglecting the tone index t for one of the T orthogonal tones. The generic precoder as portrayed in Fig. 1.10 usually consists of a *FEXT canceller* and a *power controller* in order to maximize the throughput of a DSL binder, even though the precoder may refer only to the canceller part when spectral load optimization is not invoked. Additionally, modulo type vectoring schemes also utilize back end *signal coolant*. When relying on an optimized choice of the coolant-controller-canceller trio, each CPE's received symbol y_k becomes mutually independent. Therefore the equalized symbol vector $\boldsymbol{z}$ can be readily obtained from $\boldsymbol{y}$ without collaborative channel equalization. The general structure of a vectored DTU

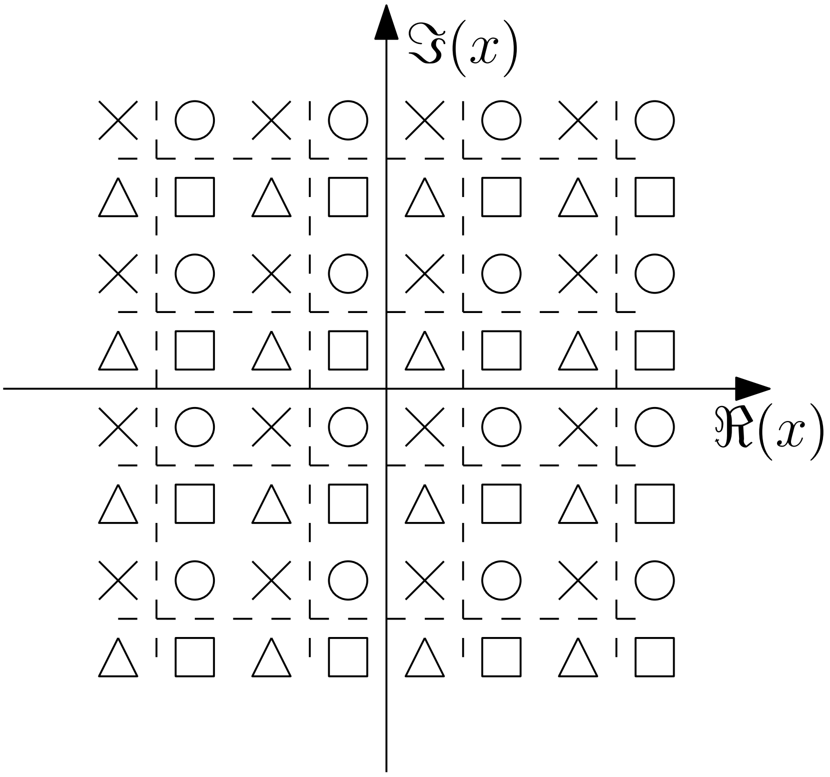


FIGURE 1.11: The effect of the signal coolant in non-linear vectoring. The four possible states of the 4QAM are depicted using one of the four symbols. Given that the original 4QAM constellation constitutes a square centred at the origin, introducing the perturbation vector results in an infinitely expanded QAM constellation. After perturbation, each message symbol from the original 4QAM is remapped onto one of the locations denoted by the same symbol on the infinite QAM constellation. Each vector of QAM symbols is repositioned according to the minimum power criterion (cf. Sec. 2.4.2) From the perspective of the modulo receiver, all locations denoted by the same symbol on the infinite QAM constellation are indistinguishable.

transmitter is given by a cascade of the three modules formulated as:

$$\boldsymbol{x} = \boldsymbol{G}\boldsymbol{A}\tilde{\boldsymbol{u}}. \tag{1.5}$$

1. **Signal Coolant**. As it will be discussed in Ch. 2, the transmit signal cooling process typically relying on Galois field operations (e.g. modulo reduction) serves the purpose of reducing the precoded signal power at the transmitter's output, akin to how coolant prevents excessive temperature increase in motors and reactors. Applying the back-end signal coolant $\boldsymbol{l}$ to the users' message symbol vector $\boldsymbol{u}$ will reduce the power boost incurred by the FEXT canceller, normally as a consequence of optimized lattice mapping. The coolant $\boldsymbol{l}$ results in a remapping of $\boldsymbol{u}$ to the expanded signal space (Fig. 1.11). The distributed CPE receivers can remove the effect of $\boldsymbol{l}$ and detect each received symbol y_k independently without the full knowledge of $\boldsymbol{l}$, which is analogous to the 'vaporization' of liquid coolant.

2. **Power Controller**. Represented by the non-negative real-valued diagonal matrix $\boldsymbol{A}$, it is used for coordinating the transmit power allocated to each message symbol

u_k in order to achieve a similar quality of service for all users, while simultaneously maximizing the throughput under the limit of the TxPSD mask and the ATP budget. The power allocation policy corresponding to the controller $\boldsymbol{A}$ is given by $\boldsymbol{P} = \boldsymbol{A}^T\boldsymbol{A}$. We should note that the *power allocation* carried out here is different from the *power control* of wireless communication systems, where the overall transmit power of an entire system is scaled with respect to other wireless systems operating in close proximity, for compliance with electromagnetic compatibility requirements.

3. **FEXT Canceller**. The front-end FEXT canceller $\boldsymbol{G}$ is a linear filter that, together with the signal coolant, maps the message vector $\boldsymbol{u}^i$ to a multi-dimensional signal space related to the inverse $\boldsymbol{H}^{-1}$ of the invertible CSI matrix $\boldsymbol{H}$. Hence the equivalent channel between the message symbol vector $\boldsymbol{u}$ and the received symbol vector $\boldsymbol{y}$ is interference-free. In certain vectoring schemes, the CPE receivers may require an additional equalizer to compensate for the uncorrelated direct channels. Based on the design of signal cooling and/or the choice of the receiver optimization criterion (e.g. zero forcing), the set of interference-free subchannels obtained from FEXT cancelling may have drastically different average quality.

1.4.3 Duplexing

In order to boost the bandwidth efficiency of DSL links beyond the Shannon limit of *a single channel use*, recent proposals [5][73] have suggested simultaneous upstream and downstream transmissions over the same bandwidth, i.e. full duplexing (FDX), for the forthcoming G.mgfast. Furthermore, FDX is also a strong candidate which has been widely studied in the wireless context [74, 75, 76]. By definition, FDX allows doubled channel use within a single DMT symbol duration. However, the capacity gain due to FDX is typically less than 100% [77][78] as a consequence of the resultant strong self-interference, which consists of signal reflection (due to imperfect receiver side impedance matching, sometimes known as the echo), as well as of NEXT in DSL systems. We characterize the channel utility paradigm of FDX systems in Fig. 1.12. In particular, FDX achieves 100% efficiency in both frequency and time domain at the expense of losing orthogonality in the interference domain.

Echo cancellation (EC) [79] and NEXT cancellation [80] are critical techniques for guaranteeing the performance of multi-pair FDX DSL systems. Akin to FEXT cancellation, NEXT cancellation is only possible at the DP side, or for either a subgroup of coordinated CPEs or a single CPE connected to multiple twisted pairs (the upstream dual of Fig. 1.4). On the other hand, since EC is only associated with a single line, it does not require coordinated transmission with other transceivers. However, when the CPEs are not co-located, the CPE side NEXT is shown to be consistently lower than the DP

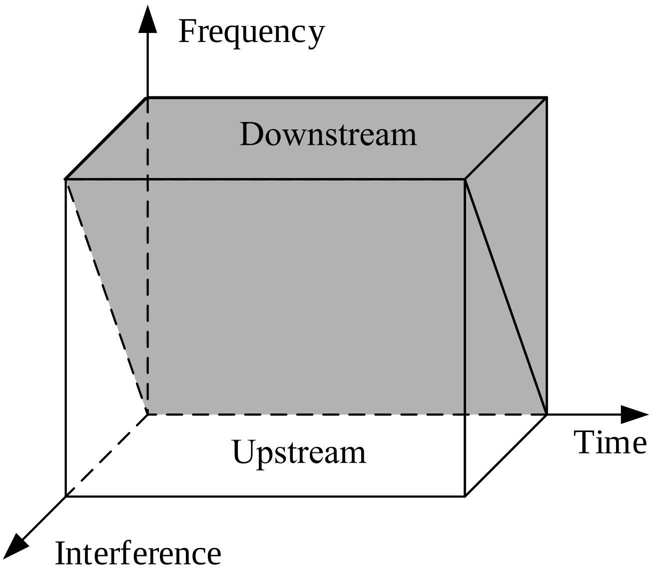


FIGURE 1.12: Illustration of the upstream downstream channel separation in FDX systems. Similar to a two-user non-orthogonal multiple access (NOMA) system, the time and bandwidth efficiency both achieve 100%, at the expense of losing orthogonality in the interference domain. Therefore, using FDX cannot achieve doubled information rate compared to using TDD.

side NEXT [5]. In general, EC/NEXT cancellation can be done in both the time and the frequency domain. The FDX operation of a single G.mgfast transceiver unit (CPE or one modem in DP) is depicted in Fig. 1.13. Overall, cancelling the self-interference in frequency domain is simpler to implement, but the signal reception and transmission must be properly aligned in time. On the other hand, self-interference cancellation in the time domain does not require synchronized signalling, but the complexity of the time domain approach is higher. Additionally, self-interference cancellation may also be implemented in time domain at the analogue front end, if the analogue-to-digital converter becomes saturated in the face of strong self-interference. The reader is redirected to [5] for more details on the multi-pair FDX design as well as for the performance study of G.mgfast, where a throughput increase of nearly 100% is observed for FDX compared to TDD using the same bandwidth. In the case of coexisting G.fast/mgfast deployment, it has also been shown in [81] that a throughput increase of more than 50% is achievable.

1.4.4 Error Control

Besides alien FEXT, *Impulsive noise* (IN) constitutes another category of impairments, whose non-stationary nature cannot be accurately captured by the relatively static CSI. Furthermore, the measurable additive white Gaussian noise (AWGN) floor is typically at -150 dBm/Hz in DSL systems, which is considerably higher than the -174 dBm/Hz AWGN floor characterized for common cellular wireless systems. In general, the noise sources cannot be mitigated by vectoring techniques. Hence, impulsive noise protection

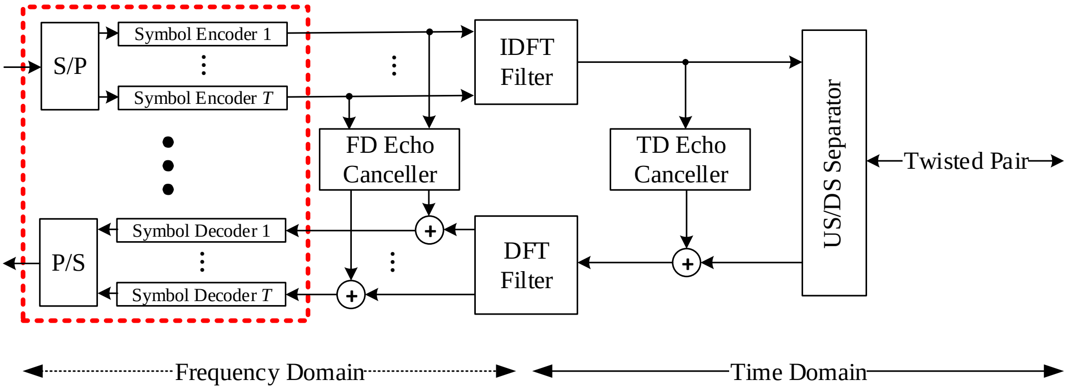


FIGURE 1.13: Single-pair T-tone full duplexing DSL transceiver model. EC may be done in frequency domain or time domain. The upstream and downstream transmission is separated by a special hybrid circuit. In the multi-pair case[5], the NEXT canceller is included in the EC module.

is implemented by means of forward error correction coding and retransmission, besides adding noise margin in the transceiver design [82]. The bursty nature of IN and its influence on the performance of DSL systems is characterized in [83]. On the other hand, alien impairment in DSL systems such as the *radio frequency interference* (RFI) caused by other wireless systems in close proximity must be handled by error correction coding as well [84]. Aerial sections of a DSL network, such as drop wires, are susceptible to RF interference, acting just like antennas. Unlike vectoring, error correction coding generally follows the same rules for both upstream and downstream transmissions. The error correction paradigm for the next generation G.mgfast systems consists of the following strategies:

- **Channel Coding**: The performance of the standardized channel coding approach, relying on Reed-Solomon (RS) coding aided trellis coded modulation, is unable to satisfy the increasing demand for high QoS. Therefore, capacity-approaching coding scheme such as low density parity check (LDPC) codes are employed as the successor. Additionally, for IN protection, interleaving will remain an effective approach due to its inherent capability of dispensing bursty errors. Even though interleaving is typically applied in the frequency domain, time domain techniques may be specifically designed to combat the non-uniform spreading of time domain IN in the frequency domain upon demodulation [85].

- **Automatic Repeat reQuest**: When relying on transmitting cyclic redundancy (CR) along with the payload, the channel decoder becomes capable of identifying the incorrectly received DMT symbols and triggering retransmission of the same payload until either the DMT symbol in question is correctly received or the maximum number of retransmissions is reached. ARQ has demonstrated great potential in boosting the throughput of DSL systems [86]. In the future, hybrid ARQ schemes exploiting the availability of past retransmissions are also under consideration. An

LDPC coded solution is specifically investigated in [87].

As mentioned before in Sec. 1.4.2, error control constitutes one of the three most widely recognized signal processing challenges for DSL systems, together with dynamic load balancing and vectored transmission [71] (Tab. 1.1). The reader is redirected to [72] and the references therein for a survey of (impulsive) noise mitigation and the error control paradigm in the next generation wireline networks. Further investigation of optimized error control for next generation DSL networks is specified in Ch. 4.

1.4.5 Network Intelligence

In recent years we have seen a rapid increase in the research of artificial intelligence (AI) and machine learning, which are revolutionizing many areas in communications, such as resource management [88] and routing optimization [89][90] in SDN-aided system architectures, where the network characteristics can be efficiently learned. Therefore, the machine learning assisted SDN constitutes an intelligent solution for the converged access network architecture [12].

However, harnessing these powerful tools in the existing multi-standard multi-ISP based DSL network paradigm requires further research. As mentioned in Section 1.4.2, DSL access networks are particularly susceptible to alien crosstalk both from LLU and from other binders in close proximity, in addition to the severe in-binder FEXT and NEXT. In [17], a software-defined, open access network infrastructure was proposed for mitigating the contaminating effects of LLU. Specifically, a management interface, maintained by a third party, is employed for coordinating the traffic associated with different ISPs sharing a single DSL binder. Subsequently, the third party coordinator may invoke vectoring for conveying the transmitted data across the ISPs using the same binder, subject to authentication for preventing violation of privacy.

We should note that, apart from the aforementioned macroscopic benefit of AI for the DSL-aided future access network infrastructure, AI can also be used for tackling the challenges associated with other parts of the system architecture we have discussed. In particular, evolutionary algorithm aided MUDs [91] and deep neural network assisted MUDs [92] have already been conceived for MIMO systems, in addition to the 'auto-encoder' type of end-to-end system design of [93]. However, we should note that the near-optimality of these particular applications has only been characterized for low-dimensional systems. Later in Sec. 2.4, we will develop a novel differential evolution aided lattice mapping technique, which retains near-optimality even in sufficiently large-scale systems.

1.5 Outline of Thesis

1.5.1 Structure of Thesis

In this thesis, we proposed several novel solutions for the main signal processing challenges associated with next-generation ultra-high-speed metallic access networks. The main focus of the thesis is the three levels of DSM as a result of the massive channel quality degradation in the recently expanded high frequency spectrum of both G.mgfast and the second revision of G.fast.

In the spirit of encouraging further research into combating the hostile crosstalk-infested channel environment and therefore providing high performance Internet services in the converged access network era, we firstly review the multi-user precoding (MUP) vs. the multi-user detection (MUD) duality followed by an overview of the MUP state-of-the-art in Ch. 2, in the hope that the fifty years' worth of literature of multi-user detection [94] will accelerate the development of powerful and efficient vectoring techniques. Secondly, following the unique stability characteristics of the DSL channels, we present an extensive study of the lattice theory basics and of lattice basis reduction, based on which we derived a pair of novel MUPs capable of achieving near-optimal performance in large-scale systems even in hostile scenarios. Additionally, based on the author's novel graphical modelling, an original performance study of joint vectoring and dynamic spectrum balancing (DSB) is presented in Ch. 3 for DSL systems under realistic power constraints. Furthermore, Ch. 4 proposes a novel reduced-latency LDPC-coded HARQ scheme based on the powerful tools of extrinsic information transfer (EXIT) charts and on early stopping of the FEC decoder. Finally, Ch 5 concludes this thesis as well as providing suggestions for future research directions.

1.5.2 Novel Contributions

It is important to point out that even though a lot of the studies in this thesis were conducted in the context of wireline systems, the majority of them are also applicable in the broader context of channels. For example, the study of generic lattice problems is of high practical relevance in capacity-achieving channel coding and even beyond the classic telecommunications context, such as in quantum-attack-resistant cryptography. The list of original contributions of this thesis is as follows:

- An original semi-analytical tool referred to as the *vectoring mapping graph* which characterizes the behaviour of multi-user preprocessing/precoding (termed as MUP in this thesis) is developed. The model directly reflects the fundamental duality between MUP and MUD, especially for the non-linear category which relies on solving hard lattice problems.

- A novel algorithm is conceived for solving one of the fundamental lattice problems, namely the closest vector problem (CVP). By conceiving the hybrid of the CVP algorithmic families of enumeration and sieving [95], the new algorithm combines their advantages and hence significantly outperforms the sphere decoder based benchmarker under hostile channel conditions. Even though the algorithm mainly targets large-scale metallic access networks, the general framework is suitable for optimal maximum likelihood decoding in a wide range of other applications;
- The first known *spectrum balancing/adaptive rate and power allocation* algorithm for wireline systems relying on lattice reduction aided MUP *under the peak power constraint* is presented. The algorithm characterizes the sum rate performance more realistically than Shannon's noisy channel coding theorem;
- A novel hybrid ARQ scheme is proposed for LDPC-coded systems defined under the 5G channel coding specifications. Based on both EXIT chart predictions and early decoder stopping, the proposed scheme is capable of both avoiding potentially futile iterative decoding attempts and mitigating erroneous convergence during iterative decoding.

1.5.3 List of Publications

1. Y. Zhang, R. Zhang, A. F. A. Rawi, and L. Hanzo, "Approximate perturbation aided lattice encoding (APPLE) for G.fast and beyond," *IEEE Access*, vol. 6, pp. 53 438–53 451, 2018
2. Y. Zhang, R. Zhang, J. Zhang, T. Bai, A. F. Al Rawi, M. Moonen, and L. Hanzo, "Far-end crosstalk mitigation for future wireline networks beyond G.mgfast: A survey and an outlook," *IEEE Access*, vol. 8, pp. 9998–10 039, 2020
3. S. Shao, Y. Zhang, R. G. Maunder, and L. Hanzo, "3D EXIT chart analysis and scheduling of 3GPP New Radio LDPC decoder," *IEEE Trans. Veh. Technol. (to be submitted)*, 2020

Chapter 2

Vectored Transmission for Multi-Pair Networks

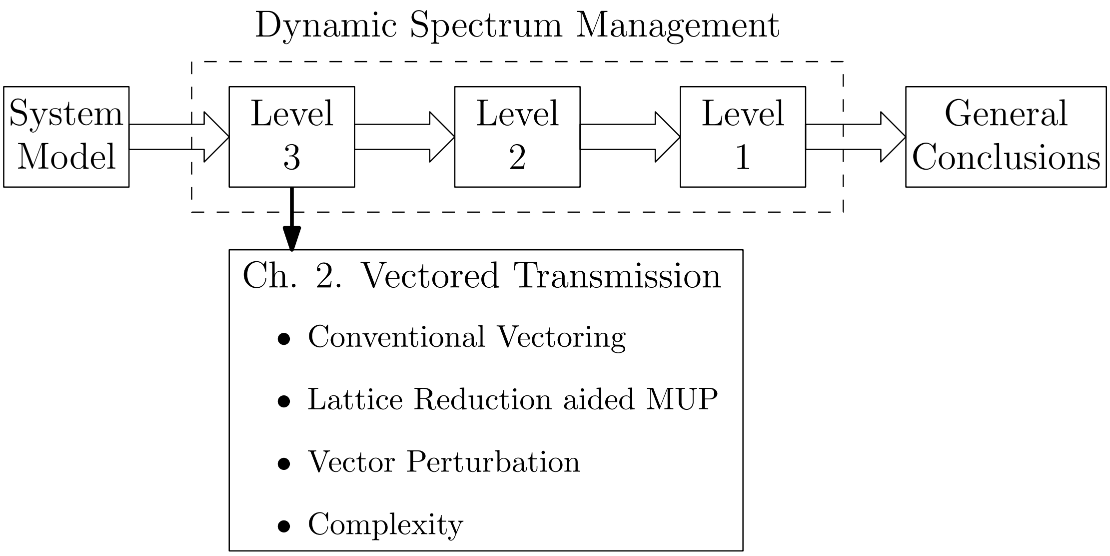


FIGURE 2.1: Outline of Ch. 2.

2.1 Vectoring Feedback

Our ability to accurately (pre-)compensate for FEXT using vectoring relies on the CSI knowledge. In G.fast, the vectoring control entity (VCE in Fig. 2.2) is responsible for obtaining the prerequisite CSI knowledge mainly for the MUP, whereas MUD configuration is vendor discretionary. Without loss of generality, we will consider a frequency-domain subsystem whose tone index is neglected (Fig. 2.2).

Under the specifications of [22], the operations of downstream vectoring are divided into the initialization stage for CSI acquisition and the main operational stage. The initialization stage invokes a training-aided channel estimation technique using the vectoring

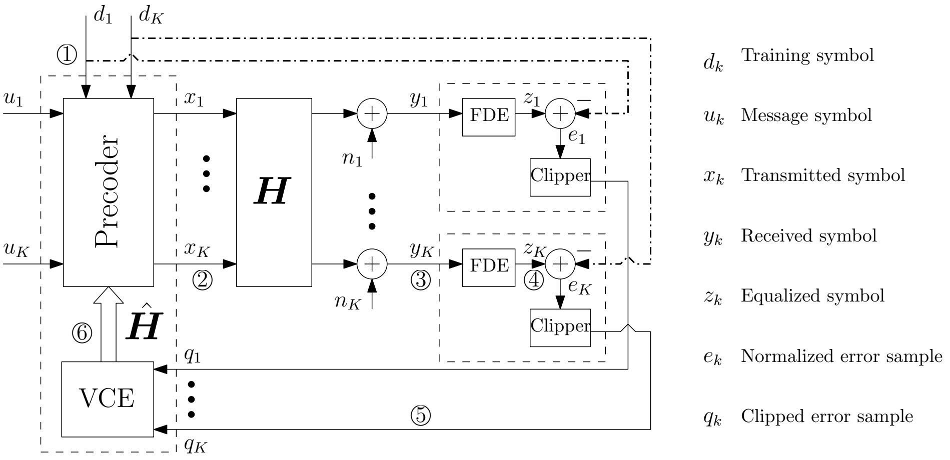


FIGURE 2.2: Downstream vectoring feedback and control loop Copyright [22].

feedback loop of Fig. 1.10. The full loop consists of the following blocks for the single-tone example:

① The training symbol vector $\boldsymbol{d} = [d_1, d_2, \cdots, d_K]^T$ is fed into the uninitialized precoder. The identity information of $\boldsymbol{d}$ is transmitted as part of the training sequence (termed as the probe sequence in [22]), therefore the training symbol vector is virtually known to the CPEs.

② The transmitted symbol vector $\boldsymbol{x}$ is normally a function of $\boldsymbol{d}$ and the estimated downstream CSI $\hat{\boldsymbol{H}}$. During initialization, the uninitialized precoder forwards $\boldsymbol{d}$, which is directly fed into the channel as the transmitted symbol vector $\boldsymbol{x}$. Hence the transmitted symbol of each customer is uncorrelated.

③ The symbol vector $\boldsymbol{y}$ received at the decentralized CPEs (right side of Fig. 1.10) is contaminated by the frequency-domain channel $\boldsymbol{H}$ and thus $\boldsymbol{y}$ contains information about the CSI. From each customer's perspective, each received element y_k in (1.4) can be rewritten as:

$$y_k = h_{k,k} d_k + \sum_{j=1, j \neq k}^{K} h_{k,j} d_j + n_k. \tag{2.1}$$

Each received symbol contains information both about the channel of the direct link (the first term of (2.1)), as well as about all FEXT channels coupled to the said link (the second term of (2.1)).

④ In order to identify the influence of the channel on the training symbol vector, the error between the equalized symbol vector $\boldsymbol{z}$ and the training symbol vector $\boldsymbol{d}$ is recorded as the normalized error sample vector $\boldsymbol{e} = \boldsymbol{z} - \boldsymbol{d}$. For QAM based systems, $\Re(\boldsymbol{e})$ and $\Im(\boldsymbol{e})$ are stored as separate quantities. Since the uninitialized single user equalizer w_k is not invoked, the equalized symbol vector $\boldsymbol{z}$ is identical to $\boldsymbol{y}$.

⑤ Due to the limited bandwidth of the feedback channel, $\boldsymbol{e}$ must be quantized using the quantization format defined in [22]:

$$\boldsymbol{q} = \max\left[-2^{B_{\max}}, \min\left(\lfloor \boldsymbol{e} \cdot 2^{N_{\max}-1} \rfloor, 2^{B_{\max}} - 1\right)\right], \tag{2.2}$$

where $N_{\max}$ and $B_{\max}$ represent the number of bits that control the quantisation step size and the maximum quantisation range of the clippers, respectively. The clipped error sample vector $\boldsymbol{q}$ is reported back to the VCE on the ONU side (left side of Fig. 1.10) via the feedback channel.

⑥ The VCE attempts to deduce the original received symbol vector $\boldsymbol{y}$ by reconstructing $\hat{\boldsymbol{y}} = \boldsymbol{d} + \boldsymbol{q}/2^{N_{\max}-1}$. When the VCE receives a sufficient number of clipped error sample vectors, it will be able to produce an estimate $\hat{\boldsymbol{H}}$ of the CSI $\boldsymbol{H}$. For the $K \times K$ channel matrix $\boldsymbol{H}$, at least K clipped error sample vectors (thus containing K^2 clipped error samples) are required, since no unique solution of the K^2 channel coefficients exists, if there are less than K^2 linearly independent equations (in the format of (2.1)).

The above vectoring feedback scheme is overall sub-optimal in terms of achieving the Cramer-Rao bound (CRB) [99][100], as a result of the loss from quantization (2.2) and the feedback channel ⑤. Firstly, the training process may also be triggered at the request of the VCE afterwards in order to update the CSI estimate $\hat{\boldsymbol{H}}$ and subsequently the precoder. Therefore, the time-invariant nature of the channel matrix $\boldsymbol{H}$ may be tackled by regularly updated channel estimation [101][102]. Secondly, since G.fast currently uses time-division duplexing (TDD) to separate upstream and downstream transmissions, the VCE may exploit the channel's reciprocity [103] to acquire the downstream CSI based on the upstream CSI estimate, the latter of which trivially approaches the CRB [67]. We should note that the in-band full duplexing operation to be introduced in G.mgfast also benefits from this reciprocity.

It is also worth noting that multi-pair channel characterization is a critical issue for correctly initializing the VCE. Even though the DSL channel is quasi-static over time, due to the DSL loop's sensitivity to mutual coupling, all copper connections within the binder that shall be active for data communication must be trained simultaneously during a single initialization session. Additionally, simply truncating the VCE's estimated transmit CSI matrix in case the binder is only partly activated will typically lead to imperfect FEXT cancellation due to CSI mismatch, even if the perfect full transmit CSI matrix is available to the VCE.

In the remainder of this chapter, we will discuss the transmit precoding algorithms that have been verified in DSL environments. A multi-user transmit precoding problem in essence constitutes the dual counterpart of the multi-user equalization problem of wireless systems, and conversely the majority of the precoding algorithms have a dual pair for equalization. This kind of relationship has been shown for linear algorithms in [104]. The main difference between precoding and equalization is considered to be their information transfer characteristics, since the latter assumes a Gaussian type conditional probability distribution, while the former operates on the premise of the conditional probability distribution related to the CSI. An extensive review of multi-user equalization algorithms can be found in [94], which assists us in selecting a subset, whose duals are applicable to downstream vectoring.

2.2 The MUP-MUD Duality

It is widely acknowledged that there exists a duality between the uplink and the downlink of cellular wireless networks. In particular, if the channel reciprocity holds between the uplink and the downlink, while the sum of uplink tranmit power constraints equals the downlink transmit power constraint, then the *optimal transmission and reception strategies are equivalent* in both directions, despite the difference in the capacity regions [105]. In the DSL standards (e.g. [63]), downstream transmission obeys per-pair power constraints similar to its upstream counterpart, thus the duality becomes strict-sense. Nonetheless, the upstream-downstream duality specifically focuses on the duality between the optimal signalling strategies for each direction, which is essentially the duality between MUP and MUD, a different duality of independent research interest.

The MUP-MUD duality originated from a pair of achievable bounds for the *coding with side information* problem, achieved by Dirty Paper coding (DPC) [106] and Wyner-Ziv coding (WZC)[1] [109], respectively. As portrayed in Fig. 2.3, DPC achieves optimal encoding/precoding using 'blind' (as in having no knowledge of the interference $\boldsymbol{s}$ in Fig. 2.3) decoding/detection, while WZC achieves optimal decoding/detection using 'blind' encoding/precoding. Namely, given a Gaussian distributed source $\boldsymbol{u} \sim \mathcal{N}(0, \sigma_u^2)$ and a Gaussian distributed noise source $\boldsymbol{n} \sim \mathcal{N}(0, \sigma_n^2)$, then the Shannon limit of the coded system is $\mathcal{C} = \log_2(1 + \sigma_u^2/\sigma_n^2)$ for both WZC and DPC, regardless of the variance of the interference $\boldsymbol{s}$. The WZC-DPC duality characterizes the holy grail of a multi-user system, i.e. "*(non-causally) known interference does not matter even if one side is blind*". However, the MUP-MUD duality can be further expanded to sub-optimal schemes in an algorithmic and structural sense. For instance, the widely studied linear schemes, such as the zero forcing and the minimum mean square error estimator, are shown to constitute MUP-MUD pairs [104][110].

In this section and the next, we will discuss the existing MUP algorithms in the current DSL communications paradigm, as well as reviewing a family of powerful MUP algorithms based on lattice reduction. Based on the MUP-MUD duality, we hope that

---

[1] We note that WZC was originally a solution of the distributed source coding problem. However, as mentioned in [107], its lossless counterpart Slepian-Wolf coding [108] has a strong duality with channel coding.

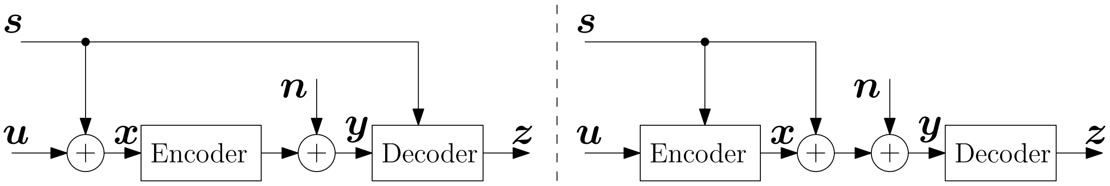


FIGURE 2.3: A special case of Wyner-Ziv coding (left) versus Dirty Paper coding (right).

our study of the more advanced MUP techniques will inspire further exploration into the subject of downstream vectoring techniques in an algorithmic sense, where the rich literature of MUD [94] will assist us tremendously in selecting algorithms whose duals are applicable to downstream vectoring. Overall, the main difference between MUP and MUD is considered to be their information transfer characteristics, since the latter assumes a Gaussian type conditional probability distribution due to white noise, while the former usually does not rely on nor has access to the noise statistics. Therefore, MUD algorithms that strongly exploit the knowledge of noise statistics such as [111] are generally inapplicable to MUP. Another problematic aspect is that MUP does not benefit from using soft information, which is a consequence of the limited bandwidth of each tone. For this reason, iterative MUD algorithms which rely on log-likelihood ratio (LLR) feedback such as the turbo MUD of [67] is also impractical for MUP design.

A classification of the candidate MUP algorithms that we will cover in this survey is portrayed in Fig. 2.4. Despite the theoretical optimality of DPC, the processing delay incurred by the precoder also has an impact on the final achievable throughput. Therefore, it is sometimes preferable to use a low-complexity suboptimal precoder to a high-complexity near-optimal one. This design choice is usually justified when the FEXT power is low, as in VDSL2, where the performance gap between the optimal and suboptimal precoding schemes is negligible. Even in the 106 MHz version of G.fast, low-complexity linear precoding is eminently suitable for downstream vectoring. In this section, we will cover the basics of the conventional linear and non-linear precoding algorithms, based on extensions of both the classic zero forcing precoding (ZFP) and the Tomlinson-Harashima precoding (THP) [112][113], respectively.

We should note that the widely recognized sum rate upper bound, obtained by using either the maximum ratio combining or its downstream counterpart, the maximum ratio transmission[2], is only capable of characterizing the *best-case single-user performance*, where both the direct channel and every FEXT channel associated with the designated user are gleaned for obtaining the specific user's maximum multiplexing gain. In maximum ratio transmission, the FEXT channels coupled into the direct are exploited, whereas in maximum ratio combining, the FEXT channels leaking from the direct are gleaned. In realistic multi-user wireline systems, it is impossible for all users to achieve their respective best-case performance simultaneously using the maximum ratio strate-

---

[2] These bounds are also known as the matched filter bounds, or single-user bounds [114].

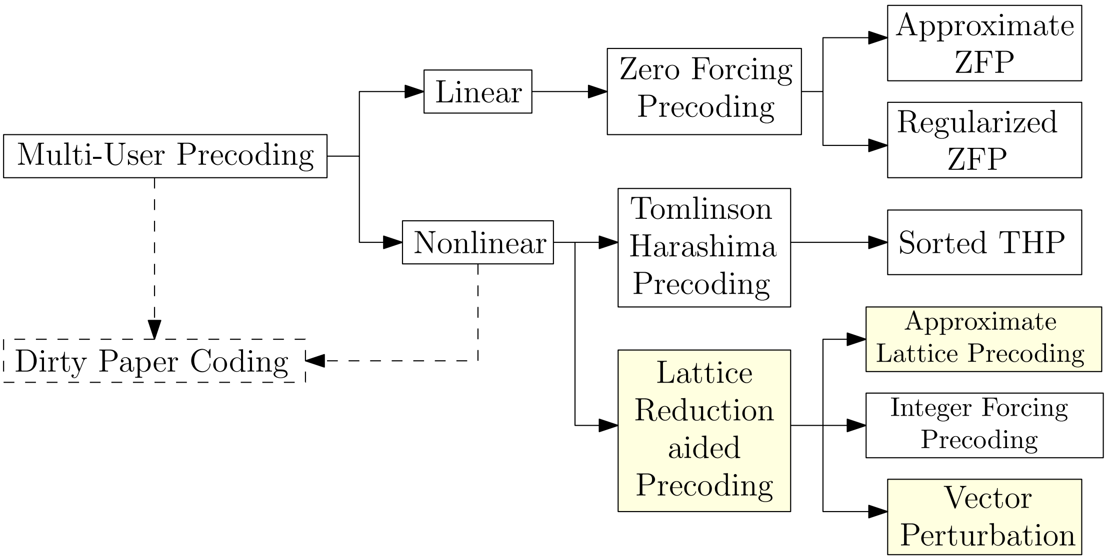


FIGURE 2.4: Family tree of the MUP techniques. Dirty paper coding characterizes an idealized non-linear scheme which achieves capacity.

gies. Therefore, we will not classify them as 'multi-user' precoding techniques. Instead, we may consider those as beamformers suitable for the single-CPE multi-pair scenario of Fig. 1.4.

2.3 Conventional MUP aided Vectoring

2.3.1 Linear Precoding

2.3.1.1 Zero Forcing

In order to *force the FEXT to zero* at the CPEs of Fig. 1.10, an obvious choice of the precoder $\boldsymbol{G}$ would be *channel inversion*. Given an invertible $K \times K$ DSL channel $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_K]$, we have:

$$\boldsymbol{x} = \boldsymbol{G}\boldsymbol{u}, \tag{2.3}$$

where $\boldsymbol{G} = \boldsymbol{H}^H(\boldsymbol{H}\boldsymbol{H}^H)^{-1} = \boldsymbol{H}^{-1}$. As implied by Fig. 1.5, $\|\boldsymbol{h}_k\|^2$ $\forall k$ is small in the high-frequency band, which subsequently suggest that its inverse $\boldsymbol{G}$ usually has a high power. Since G.fast systems operate under a strict transmit PSD (TxPSD) mask, the operations in (2.3) generally result in violation of the TxPSD mask and therefore we have to normalize the transmitted symbols. The TxPSD mask imposes a power limit on a *per tone per pair* basis [63], and hence we generally have to normalize the transmitted symbols with respect to the peak TxPSD per tone over all pairs $k = 1, 2, \cdots, K$. But for analytical tractability, we can gain meaningful insights into the performance of precoding by limiting the sum of TxPSD over all pairs with the aid of a scalar power assignment policy, similar to the total power constraint of an antenna group in wireless communications. This is also known in wireline communications as *static* spectrum balancing (SSB, as opposed to DSB) since there is no optimization involved. In fact, it is shown in [96] that the two TxPSD characterization methods have a simple geometric relationship in the signal space. Meanwhile, scalar power assignment can also significantly reduce the complexity of spectrum balancing, in which case practical seamless rate adaptation (SRA) techniques [115] may be efficiently invoked. Assuming $\gamma = \|\boldsymbol{H}^{-1}\boldsymbol{u}\|^2$, the combined *equivalent channel* seen by the message symbol vector $\boldsymbol{u}$ is a normalized identity matrix $\boldsymbol{I}/\sqrt{\gamma}$:

$$\begin{aligned} \boldsymbol{y} &= \frac{\boldsymbol{H}}{\sqrt{\gamma}}(\boldsymbol{H}^{-1}\boldsymbol{u}) + \boldsymbol{n} \\ &= \frac{\boldsymbol{u}}{\sqrt{\gamma}} + \boldsymbol{n}, \end{aligned} \tag{2.4}$$

which results in amplified noise for the constellation demappers, hence decreases the detector's signal-to-noise ratio (SNR). More explicitly, the equalized symbol vector becomes:

$$\boldsymbol{z} = \boldsymbol{u} + \boldsymbol{n}\sqrt{\gamma}. \tag{2.5}$$

An Example

For the forthcoming G.fast 212 MHz standard, an example of the basic features of downstream vectoring with perfect DP-side transmit CSI is depicted in Fig. 2.5 and Fig. 2.6. In the two-customer test system, each customer is assigned one of the two pairs within the binder and the AWGN has a PSD of -150 dBm/Hz. The message signal transmitted via the first pair (P1) $u_1(f)$ has a flat PSD of -50 dBm/Hz and a constant phase of π over the full 212 MHz bandwidth, whereas the second pair (P2) transmits the blank message $u_2(f) = 0$. In the first scenario, $u_1(f)$ is injected directly into P1 without precoding. In the second scenario, ZFP-based vectoring is applied both to $u_1(f)$ and to $u_2(f)$ for all frequencies, and the resultant signals $x_1(f)$ and $x_2(f)$ are injected into P1 and P2, respectively. A TxPSD mask of -50 dBm/Hz per pair is introduced for all frequencies.

When P2 has no signal injection, the signal received at the output of P2 consists solely of FEXT from P1. Injecting the vectored signal $x_1(f)$ and $x_2(f)$ eliminates said FEXT, since the PSD of the FEXT signal is seen to be reduced to the noise level in Fig. 2.5, while the phase of the message $u_1(f)$ is recovered in Fig. 2.6. However, in order to comply with the TxPSD mask, the power of the signal received at the output of P1 is penalized. The SNR penalty shown in Fig. 2.5 is the direct consequence of the channel's degradation upon increasing the frequency. More specifically, as formulated in (2.5), the equivalent noise seen at the receiver end of P2 is amplified by the normalization constant. As a consequence, for a channel matrix in the high frequency spectrum whose inverse will significantly boost the transmit power, the flat TxPSD mask will result in a high normalization constant, which causes high noise amplification and an SNR penalty.

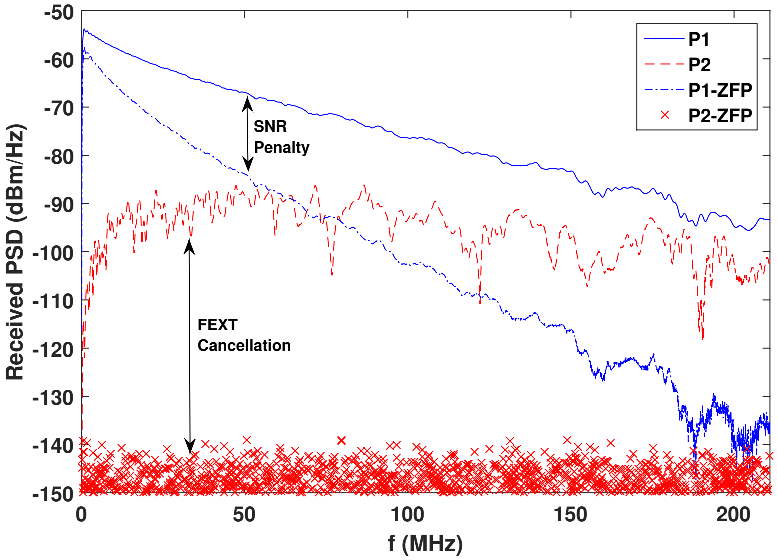


FIGURE 2.5: Received PSD at the output of a two-pair binder. The intended message $u_1(f)$ carried by pair 1 (P1) is an impulse-like signal with flat PSD of -50 dBm/Hz and a constant phase of π spanning the 212 MHz band while pair 2 (P2) carries blank message $u_2(f) = 0$. The transmitted signal is normalized to the flat TxPSD mask of -50 dBm/Hz per pair and the DP has perfect non-causal downstream channel knowledge. Comparing with the case of plain transmission without vectoring, employing ZFP reduces the FEXT signal picked up by P2 to the noise level, while the received PSD of P1 is also penalized as a result of channel inversion.

Eq. (2.3) is attractive in terms of its low complexity. Additionally, the near-optimality of ZFP was also widely recognized during the legacy DSL era [116][117] due to the diagonally dominant structure of $\boldsymbol{H}$ for all tones over the entire bandwidth of up to 30 MHz. However, applying ZFP to the tones beyond 106 MHz will result in excessively amplified noise due to the TxPSD mask. Therefore, it is plausible that the equivalent parallel AWGN channels characterized by (2.5) is not capacity-approaching. In order to reduce the capacity loss of ZFP, a number of *linear* ZFP improvements can be used. An efficient relative of ZFP is constituted by the *diagonalizing precoder* of [118], which generates a diagonal (but not identity) channel matrix when combined with $\boldsymbol{H}$. Namely, the following decomposition of $\boldsymbol{H}$ is invoked:

$$\boldsymbol{H} = \boldsymbol{D}\boldsymbol{L}, \tag{2.6}$$

where $\boldsymbol{D}$ is a diagonal matrix and the diagonal of $\boldsymbol{L}$ is a unit vector. Diagonalizing precoding avoids full channel inversion at the transmitter, which results in a reduced equivalent noise of $\|\boldsymbol{L}^{-1}\boldsymbol{u}\|\boldsymbol{D}^{-1}\boldsymbol{n}$ rather than the noise contribution of $\|\boldsymbol{H}^{-1}\boldsymbol{u}\|\boldsymbol{n}$ of ZFP in (2.5), given the assumption that CSI knowledge is perfect at both the DP and the CPEs.

Since the processing delay incurred by MUP strongly affects the achievable sum rate, as well as the realization of ultra-reliable low-latency communications (URLLC) in next

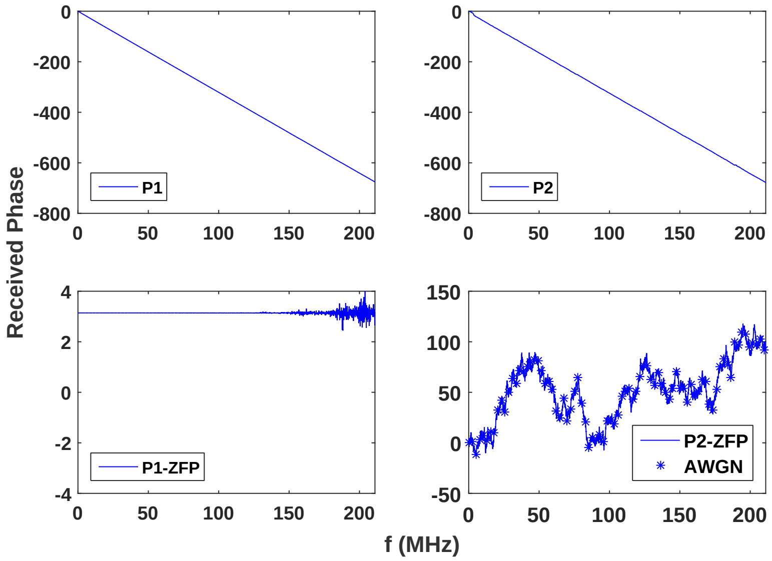


FIGURE 2.6: Received phase at the output of a two-pair binder. The system configuration of Fig. 2.5 is used. Without vectoring, the phase response shows the linear phase of the direct channel of P1 as well as the FEXT channel coupling into P2. Using ZFP, the constant phase of $u_1(f)$ is recovered at the output except for high frequency channels where the the signal is heavily distorted by AWGN. Meanwhile, the phase response at the output of P2 is the phase of the observed AWGN sample.

generation access networks, it is crucial to maintain minimum MUP complexity. Historical approaches to low complexity MUP in the DSL literature include relaxing the ZF criterion, which led to partial/approximate cancellation schemes such as those of [119][120]. However, their performances heavily depend on the diagonal dominance of the channel matrices. With the evolution of signal processing hardware and parallel computing, the complexity saving of partial/approximate ZFP becomes negligible, while their performance loss compared to full cancellation is expected to become significantly more pronounced in G.fast/mgfast. Therefore, we will not discuss these schemes further, because their performance is well bounded by the standard ZFP.

2.3.1.2 Regularized Zero Forcing

In the ideal noiseless scenario, the zero forcing criterion is optimal for the QAM constellation demappers at the receiver side. However, in the presence of both noise and FEXT, ZFP is unable to strike an attractive trade-off between noise enhancement and residual FEXT. A common linear improvement of the ZFP usually chooses to optimize the received signal-to-interference-and-noise ratio (SINR) of $\boldsymbol{z}$ for achieving an improved trade-off, by regularizing the precoding matrix $\boldsymbol{H}^{-1}$. Formally, the regularized version of ZFP, which maximizes the received SINR [121] is formulated as:

$$\boldsymbol{G}_{\text{mSINR}} = \boldsymbol{H}^H(\boldsymbol{H}\boldsymbol{H}^H + \alpha \boldsymbol{I}_K)^{-1}, \tag{2.7}$$

where the optimal choice of the positive constant α is $\alpha = K\sigma_n^2$ [121]. Coincidentally, (2.7) is also the closed form solution of the minimum mean square error (MMSE) signal reception criterion in the limit of large binder sizes K, corresponding to minimizing the following cost function:

$$\boldsymbol{G}_{\text{MMSE}} = \arg\min_{\boldsymbol{G}} E\{\|(\boldsymbol{HG} - \boldsymbol{I}_K)\boldsymbol{u} + \boldsymbol{n}\|^2\}. \tag{2.8}$$

Extending the result of (2.8), the authors of [122, 123, 124] recently proposed more flexible precoder designs by using the weighted MMSE criterion, which may be customized to achieve different optimization criteria such as the sum rate, QoS or fairness. The flexible configuration of weighted MMSE makes it a preferable candidate for joint vectoring-spectrum-balancing, i.e. multi-level DSM. Another precoding scheme closely related to the max-SINR precoding scheme optimizes the received signal-to-leakage-and-noise ratio (SLNR) [125][126]. In contrast to max-SINR precoding, which minimizes the total received interference $|\sum_{j=1,j\neq k}^{K} h_{k,j}x_j|^2$ of (2.1) *from* other pairs, max-SLNR precoding minimizes the total leakage $|\sum_{k=1,k\neq j}^{K} h_{k,j}x_j|^2$ coupled *into* other pairs. As a result of the TDD/FDX channel reciprocity in G.fast/mgfast, max-SLNR precoding for the downstream actually constitutes the dual counterpart of max-SINR detection for the upstream [127] and vice versa.

On a historical note, it is worth mentioning that instead of computing the closed form Wiener filtering matrix, the solution of (2.8) may also be found using classic iterative algorithms such as the least mean square (LMS) method, or alternatively using the simpler sign error feedback scheme proposed in [101][128]. However, the success and fast convergence of these iterative algorithms heavily rely on the conditioning of the channel matrices. Given the evaded near-orthogonality of the frequency domain channel matrices in G.fast/mgfast, the performance of these iterative algorithms in next-generation wireline access networks has to be further investigated.

2.3.1.3 Remarks

Despite the paucity of literature on use cases and on the specific analysis of regularized ZFP in the area of wireline communications, its appealing simplicity and general performance advantage over ZFP in large-scale multi-pair DSL networks has been qualitatively intimated by the asymptotic analysis of [129]. Furthermore, it was shown in [130] that for wireless MIMO channel matrices, which do not exhibit diagonal dominance, the gain of regularized ZFP over plain ZFP (and its simplified extensions) does not vanish for high SNRs σ_u^2/σ_n^2. Unfortunately, these low complexity, linear precoding schemes are known to be suboptimal in wireless systems [131], where the channel matrices are far from orthogonal. Since the DSL channel matrices also explicitly evaded orthogonality in the high frequency band, the suitability of linear MUP schemes for high performance wireline communications in the 100+ MHz band may become much more dependent upon the optimization of sophisticated joint vectoring-and-spectrum-balancing strategies, as reported by the investigations of [132][133], and also by the spatially-targeted hybrid MUP-beamforming technique of [134]. These techniques may become even more plausible for low utility rate wireline networks, where the number of active pairs at any given time is sparse compared to the binder size. In particular, it is plausible that the optimal MUP design converges to the linear maximum ratio transmitter, when the network becomes more sparse.

2.3.2 Tomlinson-Harashima Precoding

2.3.2.1 ZF-THP

As a dual counterpart of both successive interference cancellation (SIC) and of decision feedback equalisation (DFE), the THP relies on triangular factorization of $\boldsymbol{H}$ at the transmitter. This version obeying the ZF reception criterion is sometimes referred to as the ZF-THP. Following the design of [135][68] seen in Fig. 2.7, for the $(K \times K)$-element square matrix $\boldsymbol{H}$, we invoke the QR decomposition of its conjugate transpose given by:

$$\boldsymbol{H}^H = \boldsymbol{Q}\boldsymbol{R}, \tag{2.9}$$

where $\boldsymbol{R}$ is a $(K \times K)$-element upper triangular matrix, and $\boldsymbol{Q}$ is a unitary matrix, i.e. $\boldsymbol{Q}^H\boldsymbol{Q} = \boldsymbol{I}$. Instead of using $\boldsymbol{x} = \boldsymbol{H}^{-1}\boldsymbol{u}$ as we did in ZFP, we define $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{x}'$ as the final transmitted symbol vector. The channel output $\boldsymbol{y}$ in this case is:

$$\boldsymbol{y} = \boldsymbol{R}^H\boldsymbol{x}' + \boldsymbol{n}. \tag{2.10}$$

By defining the diagonal matrix $\boldsymbol{\Sigma} = \text{diag}[\text{diag}(\boldsymbol{R}^H)] = \text{diag}\{r_{1,1}, r_{2,2}, \cdots, r_{K,K}\}$, we have a *naive* FEXT pre-canceller in the form of:

$$\boldsymbol{\Sigma}\boldsymbol{u} = \boldsymbol{R}^H\boldsymbol{x}'. \tag{2.11}$$

Eq. (2.11) is favourable in that it exploits the triangular structure of $\boldsymbol{R}^H$. Since $\boldsymbol{R}^H$ is now a *lower* triangular matrix, the first message symbol u_1 does not experience FEXT, hence its leakage into all subsequent pairs from the same cable may be determined and subtracted, which now makes the second pair FEXT-free. In contrast to the THP scheme relying on optimized sorting, as it will be discussed in Sec. 2.3.2.2, the indexing of the twisted pairs here follows a randomly seeded sequence that is fixed for this specific vectored group of wired connections. However, using the naive successive FEXT pre-cancellation will inevitably amplify the unnormalized TxPSD. Therefore, we may recursively apply FEXT subtraction and *modulo reduction* to construct the pre-equalized symbol vector $\boldsymbol{x}'$ as follows:

$$x'_k = \begin{cases} u_1 & k = 1 \\ \Gamma_\phi \left[u_k - \sum\limits_{m=1}^{k-1} \dfrac{r_{k,m}}{r_{k,k}} x'_m \right] & k = 2, 3, \cdots, K. \end{cases} \tag{2.12}$$

The corresponding matrix notation is:

$$\boldsymbol{x}' = \Gamma_\phi \left[\boldsymbol{u} + (\boldsymbol{I} - \boldsymbol{\Sigma}^{-1}\boldsymbol{R}^H)\boldsymbol{x}' \right], \tag{2.13}$$

where $\Gamma_\phi[\cdot]$ represents the complex version of a real-valued modulo-ϕ reduction for a complex vector $\boldsymbol{a}$:

$$\Gamma_\phi[\boldsymbol{a}] = \Re(\boldsymbol{a}) - \phi \lfloor \frac{\Re(\boldsymbol{a})}{\phi} + \frac{1}{2} \rfloor + j(\Im(\boldsymbol{a}) - \phi \lfloor \frac{\Im(\boldsymbol{a})}{\phi} + \frac{1}{2} \rfloor). \tag{2.14}$$

The implication of the modulo operation is that the signal space is now partitioned into non-intersecting congruent regions relying on tessellation, which are essentially the Galois field representation of a QAM constellation (Fig. 1.11, also cf. Sec. 3.1). The in-phase part and quadrature-phase part of a QAM symbol drawn from rectangular constellations usually share a common modulo base. If we denote the *minimum phaser spacing* and *maximum amplitude* of a rectangular QAM constellation by ξ and c respectively, then we have $\phi = 2c + \xi$. For square M-QAM, the modulo base is simplified to $\phi = \xi\sqrt{M}$. The substitution of (2.14) in (2.12) shall guarantee that all elements of $\boldsymbol{x}'$ are located inside the square bounded by $(-\phi/2, \phi/2] + j(-\phi/2, \phi/2]$, therefore strictly confines the average power of $\boldsymbol{x}'$. In fact, the output of the modulo-ϕ operation constitutes a uniformly distributed signal set over the square region. Meanwhile, the phase rotator $\boldsymbol{Q}$ does not cause any power enhancement. We may therefore expect that the final unnormalized TxPSD of $\boldsymbol{x}$ is nearly identical to the average power density of $\boldsymbol{u}$.

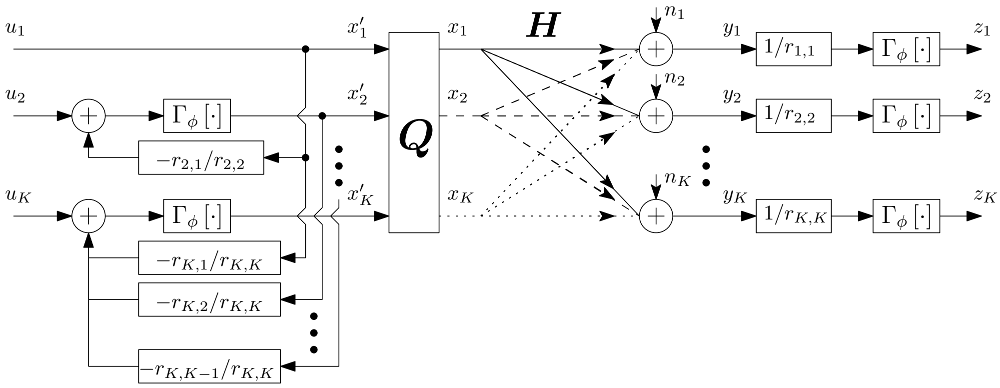


FIGURE 2.7: The THP transceiver scheme. The modulo operators Γ_ϕ at both ends of the transceiver, as well as the distribution of square QAM symbols, result in the spectral efficiency loss of THP (Sec. 2.3.2.3).

By applying element-wise modulo reduction and single user FDE, we can recover $\boldsymbol{z}$ from

$\boldsymbol{y}$ in (2.10) as:

$$\boldsymbol{z} = \boldsymbol{u} + \Gamma_{\phi}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{n}\right]. \tag{2.15}$$

Since the equalization matrix $\boldsymbol{\Sigma}^{-1}$ represents the inversion of the direct links only, it will not amplify the received noise $\boldsymbol{n}$ nearly as much as ZFP did in (2.5). On the other hand, $\Gamma_{\phi}[\cdot]$ transforms the received Gaussian noise into modulo-Gaussian detection noise and thus increases the detection noise. However, the latter usually has little or no impact on the performance of the THP, especially when the detection SNR is high. It should also be noted that the differential entropy of $\boldsymbol{z}$ is identical to that of $\boldsymbol{u}$, which generally results in information loss at low SNRs. This is known as the modulo loss (cf. Section 2.3.2.3).

2.3.2.2 Sorted THP

As shown in Eq. (2.15), the detection SNR required by the QAM demappers is related to the gain of the (rotated) direct channel $r_{k,k}$. Because of the DFE nature of the THP, the magnitude of $r_{k,k}$ and thus the detection SNR usually degrades with the user's index k during transmit precoding. The standard ZF-THP is therefore subject to the *worst-case dominance* effect, because the system's overall performance will typically be dominated by the last user having the worst performance. On the other hand, the DFE structure of THP is also subject to *error propagation* effects, when the realistic imperfect CSI feedback protocol of Fig. 1.10 is invoked. In particular, any erroneously encoded QAM symbol will result in consecutive errors for all subsequent symbols to be transmitted over the same tone. For these reasons, it is critical to have an optimal ordering of the twisted pairs for ensuring that the performance loss associated with THP's DFE structure is minimized. Given a specific channel matrix, we can optimize the system by initializing the sorted THP to the specific pair associated with the worst direct channel for ensuring that its corresponding detection SNR is maximized. Several propositions have already been suggested using the sorted THP for next generation wireline access networks [136][137].

Mathematically, a binary *permutation matrix* $\boldsymbol{E}$ is used for indicating the extra sorting during channel triangularization, so that the *minimum detection SNR per vector is maximized* [138]. Given the optimal choice $\boldsymbol{E}^{*}$, we have:

$$\boldsymbol{H}^{H}\boldsymbol{E}^{*} = \boldsymbol{Q}\boldsymbol{R}, \text{ where } \boldsymbol{E}^{*} = \arg\max_{\boldsymbol{E}}\left[\min\{r_{1,1}^{2}, \cdots, r_{K,K}^{2}\}\right], \tag{2.16}$$

while for a fixed channel $\boldsymbol{H}$ and any given $\boldsymbol{E}$, the product of all direct channel gain is constant:

$$\det(\boldsymbol{H}^{H}\boldsymbol{H}) = \det(\boldsymbol{R}\boldsymbol{E}^{H}\boldsymbol{E}\boldsymbol{R}^{H}) = \det(\boldsymbol{R})\det(\boldsymbol{R}^{H}) = \prod_{k=1}^{K}|r_{k,k}|^{2}. \tag{2.17}$$

Hence, max-min optimization compensates the worst pair's performance at the expense of good pairs. As a consequence, the largest performance gap between the pairs diminishes. Another sorting scheme conceived in [139] for multi-antenna wireless systems prioritizes the good channels, while turning off those having hostile channel conditions, so that the remaining active links achieve a higher sum-throughput than the entire group did before. However, this technique is only beneficial for point-to-point MIMO systems where *fairness* is irrelevant.

Under the constraint of (2.17), (2.16) represents the process of iteratively minimising $|r_{k,k}|$ of line k for $k = 1, 2, \cdots, K$. A very similar problem was also repeated for the sequential detection of space time codes [140]. The V-BLAST solution was found to be inefficient due to multiple inversions of the channel in the algorithm. Consequently, the authors of [141] proposed the *sorted QR decomposition* (SQRD) to solve the same problem, which also showed that SQRD solves the same user ordering problem by investing at most 60% of the complexity required by V-BLAST, while only imposing negligible loss on the detection performance.

2.3.2.3 Remarks

As a very popular non-linear precoding technique representing the original vectored transmission proposal [68], the performance of the THP is an extensively studied subject in multi-user communication systems. As for the ZF-THP scheme of Section 2.3.2.1, it is widely recognized that there is a gap between the maximum achievable bandwidth efficiency of the THP using square constellations and the Shannon limit. Without loss of generality, the three dominant sources of information loss of the THP for *the ideal AWGN channel* were studied in [142]. We should however note that the following types of loss are inherent to THP[3] and they are therefore not due to the specific types of communication channels. In fact, the wireline communications industry has already started investigating the fundamental limitations of the benchmark ZF-THP as shown in [143][144]. For AWGN channels, these limitations are portrayed in Fig. 2.8 and described as follows:

- **Modulo Loss**: Due to the modulo operation (2.14) at the receiver, each QAM symbol to be demapped to (coded) bits is distributed within the square region bounded by the modulo base ϕ. Therefore, $\boldsymbol{z}$ and $\boldsymbol{u}$ have identical differential entropy. As a consequence, the maximum bandwidth efficiency of the THP will be upper bounded by $\log_2(\sigma_u^2/\sigma_n^2)$ rather than by the Shannon limit of $\log_2(1+\sigma_u^2/\sigma_n^2)$. A study of this particular phenomenon in a G.fast environment was presented in [144]. Fortunately, the lost term '1' inside the logarithm operator may be regained if we regularize the channel decomposition of (2.9) in a way similar to (2.7) to

---

[3] In the following sections, we may notice that they are in fact related to the underlying modulo arithmetic and to the uniform distribution of QAM symbols, and therefore the analysis is applicable to non-linear MUPs in general.

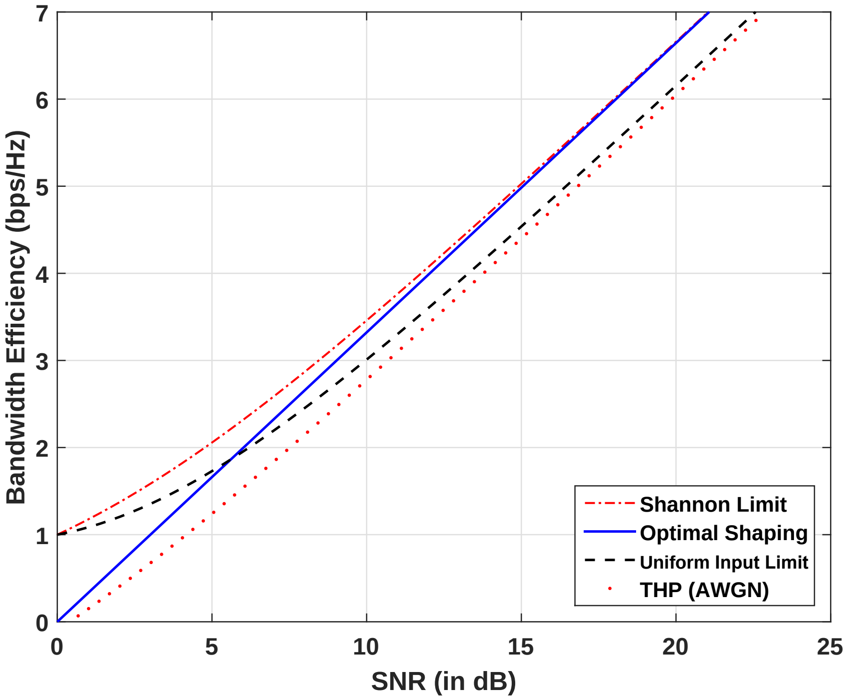


FIGURE 2.8: The bandwidth efficiency limit of the ZF-THP in interference-free AWGN channel.

obtain precoding matrices based on the MMSE-THP scheme instead [145]. As characterized by the 'optimal shaping' in Fig. 2.8, the modulo loss is the most pronounced one in the low to medium SNR regime.

- **Precoding Loss**: Due to the modulo operation at the transmitter and the particular choice of the modulo base, the convex hull of the transmitted symbols occupies a slightly larger volume in the signal space than the specific constellation that they come from. As shown previously in Section 2.3.2.1, this precoding loss is negligible (precisely $(M-1)/M$ for large M) and it converges to one, when the constellation order M tends to infinity. Since the precoding loss of the THP mainly affects low-order constellations, it becomes most prominent in the low-SNR regime.

- **Shaping Loss**: Unlike the modulo loss and the precoding loss, the shaping loss is known to be associated with the square *shaped* QAM constellations, rather than with the non-linearity of the THP transmitter and the modulo receivers. Consequently, the shaping loss occurs even in linear precoding schemes such as ZFP. As characterized by the 'Uniform input limit' seen in Fig. 2.8, the shaping loss grows to a constant of $1/2 \cdot \log_2(2\pi e/12) \approx 0.255$ bps/Hz or equivalently 1.53 dB in the high-SNR regime. More generally, the shaping loss refers to the ratio of differential entropies [146] between a given constellation distribution and the infinite-dimensional hypersphere of the same volume, because the infinite-dimensional hypersphere distribution is capacity-achieving for AWGN channels ([142]). Even though the case

> portrayed in Fig. 2.8 represents a single AWGN channel, we should note that the same shaping loss also applies to DSL and general wireless channels. However, it has been discovered that for channels with memory (such as fibre optical channels), the shaping loss associated with square constellations can be as high as 1.88 dB [147]. Interestingly, we will see in Section 2.4 that the aforementioned SNR penalty associated with MUP and DSL channels strongly resembles the shaping loss in a geometric sense.

The near-optimal information rate of the THP is readily recognized and highly appreciated by both the wireline and wireless communications community [68, 148, 149]. However, to what degree the optimality of the THP in practical multi-pair DSL systems is approached heavily relies on sophisticated DSB [150] strategies, such as the classic adaptive coding and modulation (ACM) protocols of wireless systems. Therefore, we also have to carefully assess the performance vs. complexity trade-off attained by DSB in order to quantify the practical performance of the THP and those of the other MUPs.

2.4 Lattices in Telecommunications

The concept of lattices is among the most fundamental and influential analytical tools in information theory. Lattice based methodologies are usually among the optimal candidates in a wide range of IT and CT related areas such as *quantum-attack-resistant* cryptography [151], *capacity-achieving* channel coding [152] and more relevantly, *optimal* MUP/MUD design [142]. We should note that even though lattice coding and the family of lattice reduction aided MUPs (LRMUP) to be investigated in this section both exploit the geometric goodness of lattices, their approaches are rather distinct. In the case of lattice coding, we have to construct a lattice codebook, which gives us the desired properties of good channel codes. By contrast, LRMUPs usually exploit the existing lattice structure spawned by the multi-user channel. Popularized by the celebrated Lenstra-Lenstra-Lovász [153] algorithm, the recent developments [154][155] in lattice reduction algorithms and LRMUPs demonstrate that they have significant practical value and potential in the telecommunication industry.

2.4.1 Multi-user System as a Lattice

In the complex-valued K-dimensional Euclidean space $\mathbb{C}^K$, the $(K \times K)$-element *generator matrix* $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_K]$ spawns a *lattice* $\mathcal{L}(\boldsymbol{G})$ whose column vectors $\boldsymbol{g}_k$ represent the *basis*. By definition, lattices are periodic arrangements of discrete points. As a consequence, there is an infinite number of legitimate basis for any given lattice $\mathcal{L}(\boldsymbol{G})$, where

$K > 1$. Therefore, the points of the lattice are formulated as:

$$\mathcal{L}(\boldsymbol{G}) = \{\boldsymbol{G}\boldsymbol{l} : \boldsymbol{l} \in \mathbb{G}^K\}, \tag{2.18}$$

where $\mathbb{G}$ denotes the set of all complex-valued integers and $\boldsymbol{G}\boldsymbol{l}$ is the standard matrix-vector multiplication. Using the above definition (2.18), it may be readily seen that the K-pair T-tone system of (1.4) is closely related to the union of linear spans of K-dimensional lattices $\mathcal{L}(\boldsymbol{H}^t)$ from T independent signal spaces, i.e. $\bigcup_{t=1}^{T} \mathrm{span}(\mathcal{L}(\boldsymbol{H}^t))$ or simply $\bigcup_{t=1}^{T} \mathrm{span}(\boldsymbol{H}^t)$. Meanwhile, since the majority of number theory problems originate from the real-valued domain, the complex-valued lattices are usually decoupled into real-valued ones. In particular, we may decouple one of the T signal spaces in $\mathbb{C}^K$ into $\mathbb{R}^{2K}$ using the following transformation:

$$\begin{bmatrix} \Re(\boldsymbol{y}) \\ \Im(\boldsymbol{y}) \end{bmatrix} = \begin{bmatrix} \Re(\boldsymbol{H}) & -\Im(\boldsymbol{H}) \\ \Im(\boldsymbol{H}) & \Re(\boldsymbol{H}) \end{bmatrix} \begin{bmatrix} \Re(\boldsymbol{x}) \\ \Im(\boldsymbol{x}) \end{bmatrix} + \begin{bmatrix} \Re(\boldsymbol{n}) \\ \Im(\boldsymbol{n}) \end{bmatrix}. \tag{2.19}$$

As mentioned in the definition above, a lattice of at least two dimensions has an infinite amount of basis. However, the fundamental parallelotopes constructed by two legitimate basis $\boldsymbol{G}_1 \neq \boldsymbol{G}_2$ of the same lattice have the same volume given by:

$$vol[\mathcal{L}(\boldsymbol{G})] = \sqrt{\det(\boldsymbol{G}_1^H \boldsymbol{G}_1)} = \sqrt{\det(\boldsymbol{G}_2^H \boldsymbol{G}_2)}, \tag{2.20}$$

where it may be observed that the pair of generator matrices are related by the unimodular transformation matrix $\boldsymbol{Z}$:

$$\boldsymbol{G}_1 \boldsymbol{Z} = \boldsymbol{G}_2, \text{ where } \det(\boldsymbol{Z}) = \pm 1. \tag{2.21}$$

Since the desirable diagonal-dominance and quasi-orthogonality of low-frequency DMT channels is no longer achieved over the majority of the wide G.fast/mgfast spectrum, increasing the grade of orthogonality for these badly shaped channels emerges as a valuable performance-boosting strategy. A lattice basis is called *reducible* if performing lattice reduction on it does not produce the identity matrix $\boldsymbol{Z} = \boldsymbol{I}$. A widely accepted measure of basis orthogonality for a lattice $\mathcal{L}(\boldsymbol{G})$ having the generator matrix $\boldsymbol{G}$ is termed as the *orthogonality defect* formulated as:

$$\delta(\boldsymbol{G}) = \frac{\prod_{k=1}^{K} \|\boldsymbol{g}_k\|}{vol[\mathcal{L}(\boldsymbol{G})]}, \tag{2.22}$$

where we can explicitly see that the most orthogonal basis is simultaneously the basis with the *shortest* vectors for any given lattice. A set of perfectly orthogonal basis vectors $\boldsymbol{G}^*$ satisfies $\delta(\boldsymbol{G}^*) = 1$. The primary task of LR is therefore to find the shortest basis of a given lattice (Minkowski's criterion [156]), or one of the shorter ones (LLL criterion [153]), using a known but long basis. Motivated readers might like to consult the survey and tutorial in [155]. A more specific survey on the LLL reduction algorithm is also available in [154].

An Example

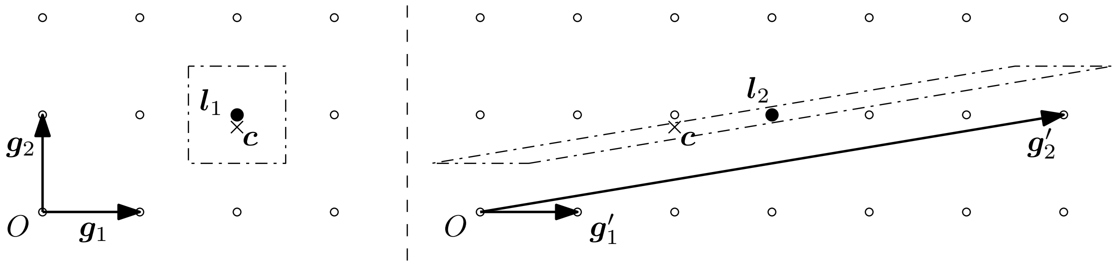


Figure 2.9: The effect of lattice reduction on quantizing a complex number. For the optimal basis on the left, the quantization region associated with any given complex-valued integer is the set of complex numbers that have the shortest Euclidean distances to said integer then to any other complex-valued integer. The minimum distance criterion is not satisfied by the basis on the right, therefore the quantization result is erroneous. This criterion can only be satisfied by orthogonal basis, and by the Voronoi cell (cf. Section 2.4.2) of lattices that do not have an orthogonal basis.

Let us consider the simple problem of quantizing a complex number $\boldsymbol{c} = (2, 0.9\jmath)$ to the nearest complex-valued integer. The correct answer $\boldsymbol{l}_1 = (2, \jmath)$ is obtained using the conventional basis of $\mathbb{C}$, i.e. $\boldsymbol{g}_1 = (1, 0\jmath), \boldsymbol{g}_2 = (0, \jmath)$. This is because $\boldsymbol{c}$ falls inside the quantization region of $\boldsymbol{l}_1$ (left of Fig. 2.9). However, as seen on the right side of Fig. 2.9, if a bad/long basis such as $\boldsymbol{g}'_1 = (1, 0\jmath), \boldsymbol{g}'_2 = (6, \jmath)$ is used, then $\boldsymbol{c}$ falls in the quantization region of $\boldsymbol{l}_2 = (3, \jmath)$. Therefore, the quantizer will erroneously consider $\boldsymbol{l}_2$ as the nearest integer neighbour of $\boldsymbol{c}$. As shown in [142], the quantization problem we have considered here represents the foundation of a wide range of non-linear MUP algorithms.

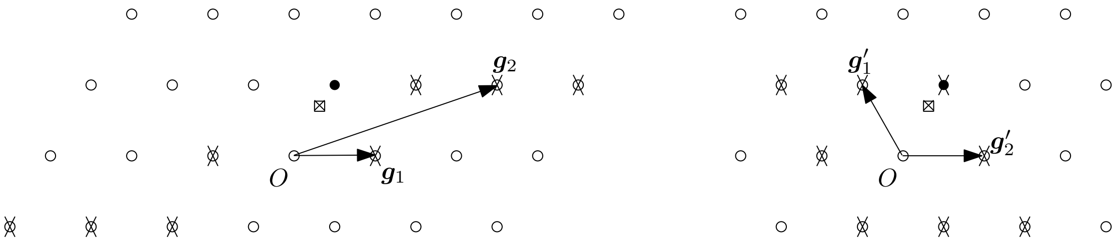


Figure 2.10: The effect of lattice reduction on the complexity of enumerating the closest lattice point problem $\arg\min_{\boldsymbol{l}} \|\boldsymbol{G}(\boldsymbol{u}+\boldsymbol{l})\|^2$. The crossed square represents the given point $-\boldsymbol{G}\boldsymbol{u}$ and the solid dot represents the closest lattice point to $-\boldsymbol{G}\boldsymbol{u}$. The lattice point O is given as the first heuristic solution during the sphere encoder's initialization, while the eight crossed circles on each side are the minimum integer combinations of the respective basis vectors. Therefore these are the first batch of lattice points to be enumerated. Compared with the case of the reduced basis on the right side where the final solution can be found within the first eight enumerations, finding the closest lattice point over the long basis on the left cannot be completed within the first eight enumerations. Therefore lattice reduction decreases the complexity of enumerating the closest lattice point.

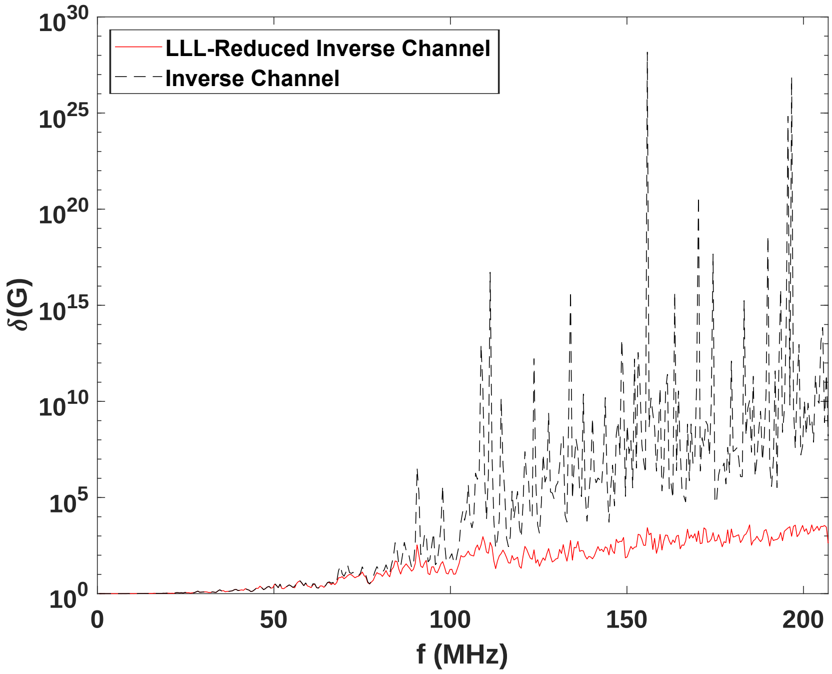


FIGURE 2.11: Orthogonality defect $\delta(\boldsymbol{G})$ of inverse DMT channels $\boldsymbol{G}$ for a 10 pair binder, and the effect of lattice reduction.

As an illustration in Fig. 2.11, we depict the quality of the basis associated with the inverse channel matrices, as well as that of the LLL-reduced basis, for a 10 pair DSL binder over the 212 MHz bandwidth. Observe from Fig. 2.11 that the orthogonality improvement offered by LLL reduction is quite significant, starting from around 90 MHz. As proven in the pioneering contribution of Babai [157] and empirically shown later by our simulations in Ch. 3, the LLL-reduced basis determines the lower bound of the worst-case performance of heuristic algorithms (e.g. rounding-off [157]) associated with MUPs (e.g. ZFP). By contrast, for an arbitrary basis the worst-case performance remains unbounded. Moreover, since lattice reduction is invoked at the initialization stage as preprocessing, its complexity overhead does not detrimentally affect the performance of DSL systems.

2.4.2 The Essence of Vectoring and the Duality

Previously, we have shown that both the optimization criterion as well as the design of MUP and MUD may be viewed as each other's dual pairs based on the WZC-DPC duality. In the following, we will see that the algorithmic core of MUP and MUD may also be interpreted as each other's dual pairs, which is a by-product of using lattice reduction strategies for multi-user systems.

In terms of the MUD design for upstream DSL, the widely recognized optimal decision criterion is the maximum likelihood (ML) criterion, or the maximum a posteriori (MAP) criterion if the data symbol at the source $\boldsymbol{u}$ *does not* follow the uniform distribution. Formally, the ML detector (MLD) optimizes the following non-convex cost function:

$$\boldsymbol{z}_{opt} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{z}\|^2,\ \boldsymbol{z} \in \mathcal{M}, \tag{2.23}$$

where $\mathcal{M}$ represents a square QAM constellation. Since $\mathcal{M}$ represents a scaled version of (a subset of) the integer lattice $\mathbb{Z}^{2K}$, it is readily seen that the main task of (2.23) is to synthesize a point in $\mathcal{L}(\boldsymbol{H})$ that is the closest to another given point $\boldsymbol{y}$ in $\mathbb{R}^{2K}$. This is one of the most influential and fundamental problems in the development of lattice theory which is termed as the *closest vector problem* (CVP) [158][159] that captures the essence of MUD. By extension, the other MUD algorithms may be conceived as heuristic approaches to an approximate solution of (2.23). Nonetheless, each MUD algorithm uniquely defines a tessellation pattern, and more importantly, a *decision region* surrounding each point of $\mathcal{L}(\boldsymbol{H})$. As shown in [160], the shape of the decision region directly determines the optimality of the associated MUD. It is well-known that the decision region associated with MLD is the basis-invariant Voronoi cell of $\mathcal{L}(\boldsymbol{H})$ [158]. The Voronoi cell $\mathcal{V}(\boldsymbol{H},\hat{\boldsymbol{y}}_i)$ centred at $\hat{\boldsymbol{y}}_i \in \mathcal{L}(\boldsymbol{H})$ is defined as:

$$\mathcal{V}(\boldsymbol{H},\hat{\boldsymbol{y}}_i) = \{\boldsymbol{y} \in \mathbb{R}^{2K} :\ \|\boldsymbol{y} - \hat{\boldsymbol{y}}_i\| \le \|\boldsymbol{y} - \hat{\boldsymbol{y}}_j\| \quad \forall i \neq j\}. \tag{2.24}$$

By comparison, any other form of the decision region is suboptimal, because they violate the minimum Euclidean distance criterion (cf. Fig. 2.9). From a similar perspective, we will now characterize the optimal criterion for MUP.

With respect to the MUP design for downstream DSL, it is known from (2.5)(2.15) and Fig. 1.5 that the SNR penalty is the dominant problem over the wide bandwidth of G.fast (also in [161][96]) compared to the losses discussed in Section 2.3.2.3. Therefore, it is reasonable to consider the transmit signal power as the cost function for MUP optimization. For example, we may use the simple formulation of [162] by introducing a perturbation vector defined as:

$$\boldsymbol{l}_{opt} = \arg\min_{\boldsymbol{l}} \|\boldsymbol{G}(\boldsymbol{u} + \boldsymbol{l})\|^2, \tag{2.25}$$

where $\boldsymbol{G}$ is the inverse channel matrix. Under the principle of DPC, the choice of the perturbation vector $\boldsymbol{l}$ should be limited to the set of scaled integers $\phi\mathbb{Z}^{2K}$ so that the modulo receivers $\Gamma_\phi[\cdot]$ can reconstruct the transmitted signal without knowing the exact value of $\boldsymbol{l}$. The same principle applies to the THP. For analytical simplicity, uniformly distributed *continuous* constellations are widely considered in non-linear MUP analysis [145, 163, 96]. Let $\mathcal{U}$ be a continuous square set formulated as:

$$\mathcal{U} = \{u :\ -\phi/2 \le \Re(u) < \phi/2, -\phi/2 \le \Im(u) < \phi/2\ \}. \tag{2.26}$$

We now immediately notice the strong similarity between (2.23) and (2.25). More explicitly, in (2.25), we seek to synthesize a lattice point $\boldsymbol{Gl}$ in the lattice $\mathcal{L}(\boldsymbol{G})$ that is the closest one to the given point $\boldsymbol{Gu}$. It was also shown in [163] that the output of the MUP $\boldsymbol{x} = \boldsymbol{G}(\boldsymbol{u} + \boldsymbol{l})$ is uniformly distributed over $\mathcal{V}(\boldsymbol{G}, \boldsymbol{0})$. This MUP design is known as the vector perturbation (VP) [162] precoding.

Based on the concept of the decision region of MUDs, the authors of [96] proposed the dual concept of the vectoring mapping region (VMR). Because $\mathcal{L}(\boldsymbol{G})$ and $\mathcal{L}(\boldsymbol{H})$ constitute a pair of dual lattices [155], the VMR and the MUD decision regions are closely related. Without loss of generality, the VMRs of all the MUPs investigated in this thesis are portrayed in Fig. 2.12, where the lattice is spawned by the inverse of a two-user system employing pulse amplitude modulation (PAM) in $\mathbb{R}^2$. We note that this limitation is a result of the fact that we cannot graphically demonstrate the VMRs of a two-user QAM system, because it is in $\mathbb{R}^4$.

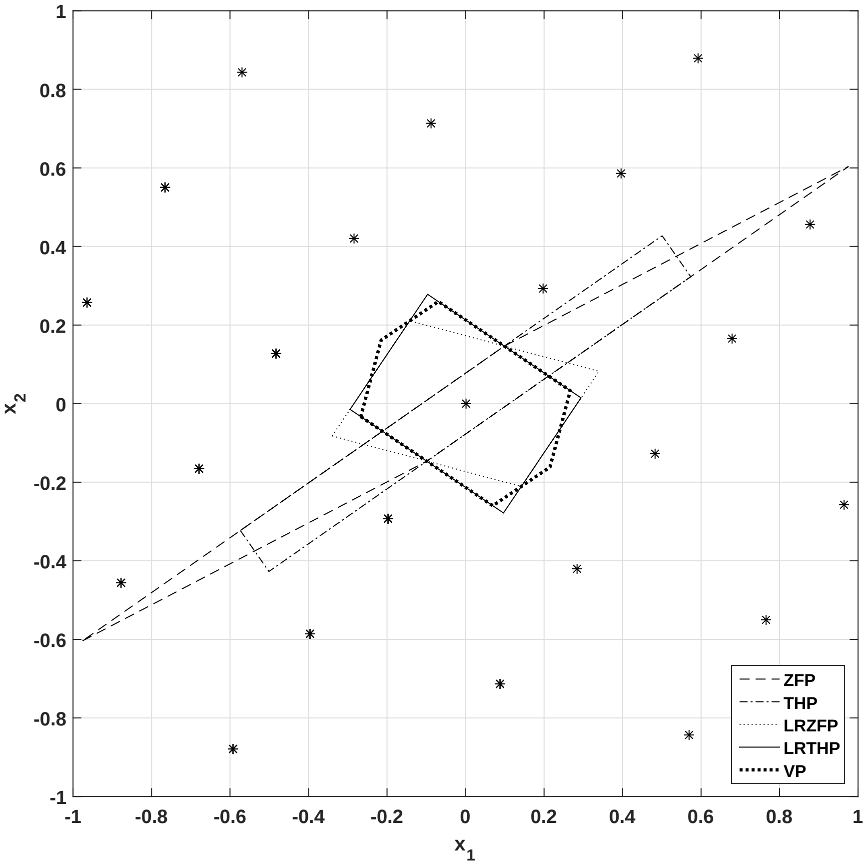


FIGURE 2.12: Geometric boundaries indicating the VMRs associated with the MUPs for a two-pair PAM based system. The basis of the lattice in $\mathbb{R}^2$ is spawned by the (*LLL-reducible*) channel inverse matrix, and its lattice points are shown. The joint input to the MUPs is the uniformly distributed input $\mathcal{U}$ of (2.26), which approximates the union of two independent PAM constellations. The *second moment* of each VMR indicates the average encoded signal power, while the *second moment per dimension* indicates the average encoded signal power per user. These properties are graphically characterized by the degree of 'isotropy' of the VMRs.

2.4.3 Lattice Reduction

As portrayed in the examples of Sec. 2.4, the family of lattice reduction algorithms has been specifically developed with one goal in mind: to find the basis of a lattice having the shortest possible vectors with respect to the infinite sets of legitimate bases under moderate complexity (time) and buffer size (space). The beneficial geometric and numerical properties of a short basis has strong positive impact on the performance of lattice-based algorithms. However, finding the shortest basis effectively corresponds to solving a cascade of consecutive SVPs, which results in a set of lattice vectors whose lengths represent the successive minima. Additionally, each vector must also be at least linearly independent[4] in order to constitute a legitimate basis. These requirements have been the evaluation criteria for all variants of the optimal lattice reduction algorithms, such as Helfrich's algorithm [164] for finding a Minkowski-reduced basis and Kannan's algorithm [165] for finding a Hermite-Korkine-Zolotareff (HKZ) reduced basis. However, unfortunately, the SVP exhibits an exponentially growing complexity as a function of the lattice dimension $2K$. For MUD and MUP applications, the complexity overhead caused by the optimal lattice reduction preprocessing is prohibitive. As a result, it has not been lavishly discussed in the MIMO literature.

Suboptimal lattice reduction algorithms such as the LLL algorithm reduces the average-case complexity order to a polynomial function of the lattice dimension, at the expense of finding slightly longer bases. Nonetheless, the lengths of the vectors from an LLL-reduced basis remain upper bounded, which makes the LLL algorithm immensely popular in the communications community. However, wireline systems can afford a higher pre-processing overhead in favour of lower processing delay at runtime, while only imposing a negligible performance loss. In this section, we will comparatively investigate the quality of the Minkowski-reduced and the LLL-reduced basis in the interest of finding how much preprocessing is practically required.

Minkowski Reduction

Before discussing the algorithm conceived for finding a Minkowski-reduced lattice basis, we first have to discuss the associated evaluation criterion and *why* the Minkowski-reduced basis is the optimal one. In particular, Minkowski's successive minima of a general lattice is defined as follows:

Definition 2.1. For a lattice $\mathcal{L}(\boldsymbol{G}) \in \mathbb{R}^{2K}$ and $1 \leq k \leq 2K$, Minkowski's kth successive minimum $\lambda_k(\boldsymbol{G})$ is the radius of the smallest $2K$-dimensional origin-centred hypersphere, which contains at lest k linearly independent vectors from $\mathcal{L}(\boldsymbol{G})$.

It is a trivial result that $\lambda_1(\boldsymbol{G})$ is simply the shortest non-zero lattice vector. Even

---

[4] In dimensions $2K > 4$, successive minima of a lattice do not necessarily constitute a basis.

though Minkowski's successive minima do not necessarily constitute a legitimate basis in dimensions higher than 4, the Minkowski-reduced basis always contains the shortest possible set of lattice vectors. More specifically, we have:

Definition 2.2. A lattice basis $\boldsymbol{G}^*$ is termed as being Minkowski-reduced, if each member vector $\boldsymbol{g}_k^*$ $\forall 1 \leq k \leq 2K$ has the minimum Euclidean length among all lattice vectors $\boldsymbol{g}_k$, where $\{\boldsymbol{g}_1, \cdots, \boldsymbol{g}_k\}$ are expandable into a legitimate basis of $\mathcal{L}(\boldsymbol{G})$. The expandable criterion is formulated recursively in [164] as follows:

$$\|\boldsymbol{g}_k\|_2 = \min\{\|\boldsymbol{G}^{(k)}\boldsymbol{l}\|_2 : \boldsymbol{l} \in \mathbb{Z}^{2K}, \gcd(l_k, \cdots, l_{2K}) = 1\}. \tag{2.27}$$

In (2.27), $\boldsymbol{G}^{(k)}$ represents a basis of $\mathcal{L}(\boldsymbol{G})$ whose first $k-1$ columns may be expanded to a Minkowski-reduced basis. The operator $\gcd(\cdots)$ takes the greatest common divisor of the operands. Based on Minkowski's successive minima, we may now both upper-bound and lower-bound the quality of the Minkowski-reduced basis by [166]:

$$\lambda_k^2(\boldsymbol{G}) \leq \|\boldsymbol{g}_k\|_2^2 \leq \max\left[1, (5/4)^{2K-4}\right] \lambda_k^2(\boldsymbol{G}), \quad 1 \leq k \leq 2K. \tag{2.28}$$

In practice, it is found that the upper bound of (2.28) is too loose for the average-case, especially in higher dimensions. This particular limitation has made it rather difficult to theoretically assess the achievable quality of the Minkowski-reduced basis without first carrying out the reduction in advance.

Lenstra-Lenstra-Lovász Reduction

The celebrated LLL reduction is the first algorithm conceived for finding a lattice basis whose member vectors have bounded lengths, at a polynomial complexity order on average. If we consider the unitary-triangular decomposition $\boldsymbol{G} = \boldsymbol{QR}$ similar to that of (2.9), then the LLL reduction criterion may be formulated as follows based on their original contribution in [153]:

Definition 2.3. A lattice basis $\boldsymbol{G}$ is said to be LLL-reduced, if its upper triangular component $\boldsymbol{R}$ from the QR decomposition satisfies the following conditions:

$$|r_{i,j}| \leq |r_{i,i}|/2, \qquad 1 \leq i < j \leq 2K, \text{ and} \tag{2.29}$$

$$\omega\, r_{i-1,i-1}^2 \leq r_{i,i}^2 + r_{i-1,i}^2\,, \qquad 1 < i \leq 2K, \tag{2.30}$$

where the reduction factor is $\omega \in (1/4, 1)$. Eq. (2.29) is commonly referred to as the size condition, while (2.30) is simply called the LLL condition. It should be noted that (2.30) becomes closer to the HKZ condition, when opting for $\omega \to 1$. It was shown in

[167] that the polynomial complexity order of finding an LLL-reduced basis is retained even when choosing the optimal boundary value of $\omega = 1$. It is now apparent that the lower complexity of finding an LLL-reduced lattice basis is a benefit of the locality of the condition (2.30), since each element is solely compared against its adjacent entries in the triangular component $\boldsymbol{R}$ of the generator matrix. Therefore the LLL algorithm may be capable of avoiding the solution of multiple global SVPs. Additionally, it was proven in [168] that the size-condition (2.29) is irrelevant in the lattice reduction aided preprocessing if followed by an enumeration-based decoding process. In the seminal contribution [153], it was also shown for an LLL-reduced basis that $\|\boldsymbol{g}_k\|$ may also be upper and lower bounded by Minkowski's respective successive minimum $\lambda_k(\boldsymbol{G})$ as follows:

$$(\omega - \frac{1}{4})^{k-1}\lambda_k^2(\boldsymbol{G}) \leq \|\boldsymbol{g}_k\|_2^2 \leq (\omega - \frac{1}{4})^{1-2K}\lambda_k^2(\boldsymbol{G}), \quad 1 \leq k \leq 2K. \tag{2.31}$$

When compared to (2.28), the upper bound in (2.31) is less tight. However, the empirical studies conducted in [166] have characterized the superiority of Minkowski reduction over LLL in practical systems, such as in the MUD for Rayleigh faded wireless MIMO systems.

Performance Evaluation

In order to theoretically characterize the general performance bound of lattice reduction, we have to rely on a critical lattice constant. In lattice theory, the classic Hermite constant dictates the best-case construction of lattices of any given dimension. In a $2K$-dimensional Euclidean space, the Hermite constant $\mathcal{H}$[5] is formulated as [154]:

$$\mathcal{H}_{2K} = \sup\{\frac{\lambda_1^2(\boldsymbol{G})}{vol[\mathcal{L}(\boldsymbol{G})]^{1/K}}\} \qquad \forall\, \boldsymbol{G} \in \mathbb{R}^{2K\times 2K}, \tag{2.32}$$

where $\sup(\cdot)$ is the supremum operator. Intuitively, $\mathcal{H}$ defines the maximum normalized length of the shortest non-zero vector in any lattices of a given dimension. From a max-min optimization perspective, $\mathcal{H}$ is also a measure of the maximum isotropy of all tessellating shapes (e.g. the VMRs of Fig. 2.12) of a particular dimension. As we will see in Ch. 3, isotropy is a critical factor for determining the achievable performance of (particularly non-linear) MUP algorithms. As an example, in wireless communications, it is known that a hexagonal arrangement of base stations has the optimal area efficiency, mainly because the Hermite constant of $\mathcal{H}_2 = 4/3$ is achieved in hexagonal lattices. However, the exact value of $\mathcal{H}$ is only known from $\mathbb{R}^1$ to $\mathbb{R}^8$ as well as in $\mathbb{R}^{24}$.

For empirical comparisons, we complement the investigations conducted for Fig. 2.11 by comparing the orthogonality defect of both Minkowski-reduced and LLL-reduced inverse channel matrices over the 424 MHz G.mgfast spectrum. When the frequency exceeds the 212 MHz upper bound stipulated in the current G.fast standard, using lattice reduction

---

[5] We use the notation $\mathcal{H}$ in order to avoid confusion with the normalized transmit power γ in this chapter

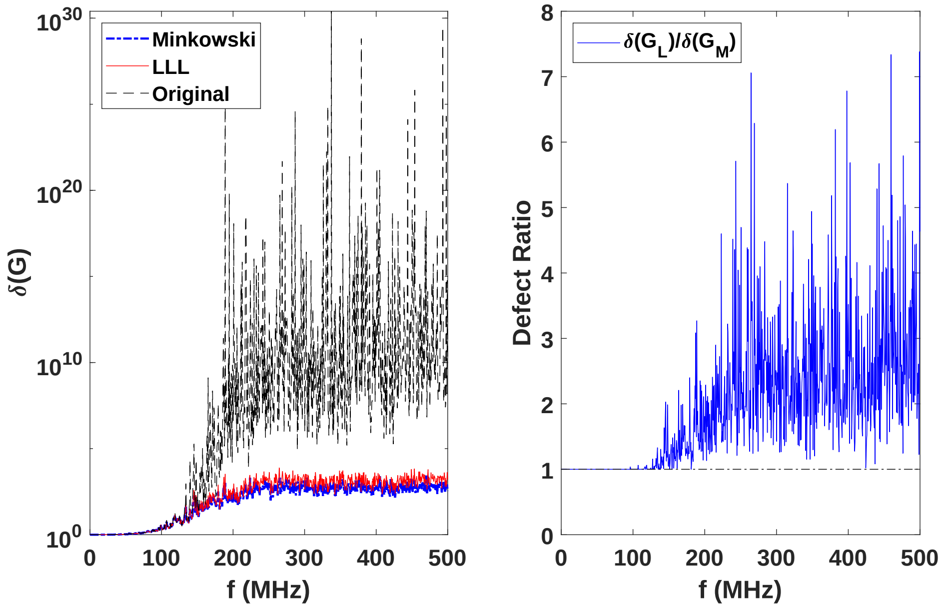


Figure 2.13: Performance comparison between Minkowski- and LLL-reduction. It is shown that the orthogonality defect of LLL-reduced basis is higher than that of Minkowski-reduced basis. For a 10-pair, 50m binder occupying up to 500 MHz bandwidth, the average defect ratio of LLL-reduced basis to Minkowski-reduced basis is approximately 3 when choosing a near-optimal LLL reduction factor of $\omega = 1 - 1e^{-5}$.

for initializing the downstream MUP becomes a promising approach, especially due to the polynomial complexity order of LLL reduction.

2.5 Approximate Lattice Precoding

Lattice Reduction aided ZFP

In order to mitigate the noise enhancement of the original ZFP, we may construct an LLL-reduced inverse channel $\boldsymbol{F} = \boldsymbol{H}^{-1}\boldsymbol{Z}$ as the precoder, in which case we have to shift the K-dimensional point $\boldsymbol{u}$ to a different position due to the unimodular transformation. Denoting the unimodular transformed message symbol vector as $\tilde{\boldsymbol{u}}$, we have:

$$\boldsymbol{x} = \frac{1}{\sqrt{\gamma}}\boldsymbol{F}\tilde{\boldsymbol{u}} \quad \text{where} \quad \tilde{\boldsymbol{u}} = \Gamma_\phi\left[\boldsymbol{Z}^{-1}\boldsymbol{u}\right]. \tag{2.33}$$

The corresponding equalized symbol vector is:

$$\boldsymbol{z} = \boldsymbol{u} + \Gamma_\phi\left[\boldsymbol{n}\sqrt{\gamma}\right], \tag{2.34}$$

where $\gamma = \|\boldsymbol{F}\tilde{\boldsymbol{u}}\|^2$. In this case, the expected performance improvement mainly accrues from the assumption that the inverse of any high-frequency DMT channel never constitutes a shorter basis than an LLL-reduced one. Otherwise, (2.34) would exhibit an even higher detection noise than (2.5) due to the modulo operation. The transceiver structure of the LR-ZFP scheme is given in Fig. 2.14. In comparison to the THP based transceiver of Fig. 2.7, the unimodular matrix filter $\boldsymbol{Z}^{-1}$ of the LR-ZFP encoder on the left side of Fig. 2.14 substitutes the decision feedback loop of the THP encoder on the left side of Fig. 2.7. The remaining components are structured similarly in both cases, except for the extra direct channel equalizers $(1/r_{k,k})$ on the right side of Fig. 2.7. Since matrix filters can be applied with the advent of parallel computing, whereas the decision feedback loop cannot, the LR-ZFP encoder would incur a lower processing delay than the THP encoder. In this case, the LR-ZFP encoder assumes the coolant to be 'multiplicative' rather than additive, the latter of which will be presented later in the form of a perturbation process (see Section 2.7).

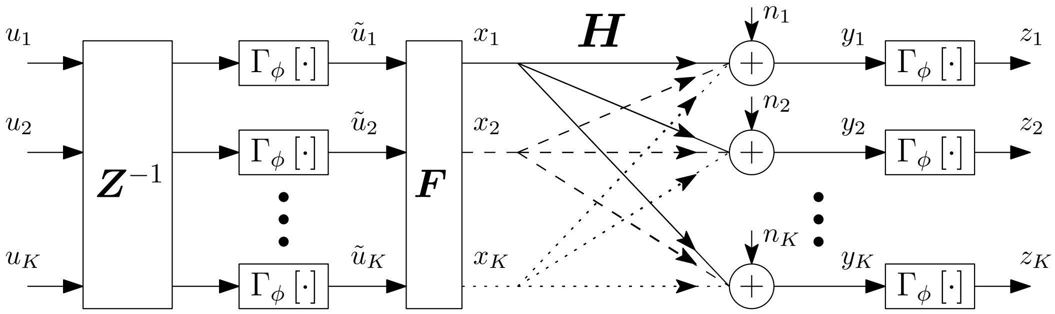


FIGURE 2.14: LR-ZFP transceiver scheme shown in $\mathbb{C}^K$.

Lattice Reduction aided THP

Since the sorted THP may be considered to constitute a low-level LR-THP, we may also employ a triangular decomposition while replacing the sorting procedure by the more powerful LLL reduction, as formulated in:

$$\boldsymbol{H}^H = \boldsymbol{S}\boldsymbol{B}\boldsymbol{Z}^{-1}, \tag{2.35}$$

where $\boldsymbol{S}$ has orthogonal columns and $\boldsymbol{B}$ is of an upper triangular structure, with $\mathrm{diag}(\boldsymbol{B}) = \mathrm{diag}(\boldsymbol{I})$. As seen in [169], it is possible to find the maximum according to (2.16) on top of (2.35). However, Chang. *et al* [170] proved that a combination of LLL reduction and SQRD would be similar to using LLL reduction alone, since the LLL algorithm naturally includes a sorting operation, whose standalone performance is similar to that of SQRD. Therefore, in our case we will dispense with any additional sorting.

Denoting the unimodular transformed message symbol vector as $\tilde{\boldsymbol{u}} = \Gamma_\phi\left[\boldsymbol{Z}^H\boldsymbol{u}\right]$, we have the precoded symbol vector $\boldsymbol{x}$ given by:

$$\boldsymbol{x} = \frac{1}{\sqrt{\gamma}}\boldsymbol{S}^{-H}\boldsymbol{x}', \text{ where } \boldsymbol{x}' = \Gamma_\phi\left[\tilde{\boldsymbol{u}} + (\boldsymbol{I} - \boldsymbol{B}^H)\boldsymbol{x}'\right], \tag{2.36}$$

which amounts to the exact form of the equalized symbol vector $\boldsymbol{z}$ as we obtained in

(2.34), except that in the case of the LR-THP, $\boldsymbol{S}^{-H}$ results in an even lower power amplification than the precoding matrix $\boldsymbol{F}$ does in LR-ZFP. The transceiver structure of the LR-THP scheme is given in Fig. 2.15. The LR-THP encoder on the left side of Fig. 2.15 may be considered as a hybrid of the THP encoder on the left side of Fig. 2.7 and the LR-ZFP encoder on the left side of Fig. 2.14. In fact, the main difference between LR-THP encoding and LR-ZFP encoding, the decision feedback encoding loop, is akin to the difference between the conventional THP encoding and the conventional ZFP encoding. As the LR-ZFP transceiver of Fig. 2.14 may be considered as a conventional ZFP transceiver (from $\tilde{u}_k$ to y_k in Fig. 2.14) implemented inside a unimodular mapping process, the LR-THP transceiver may also be conceptually considered as a conventional THP transceiver[6] (from $\tilde{u}_k$ to z_k in Fig. 2.15) implemented inside a unimodular mapping process.

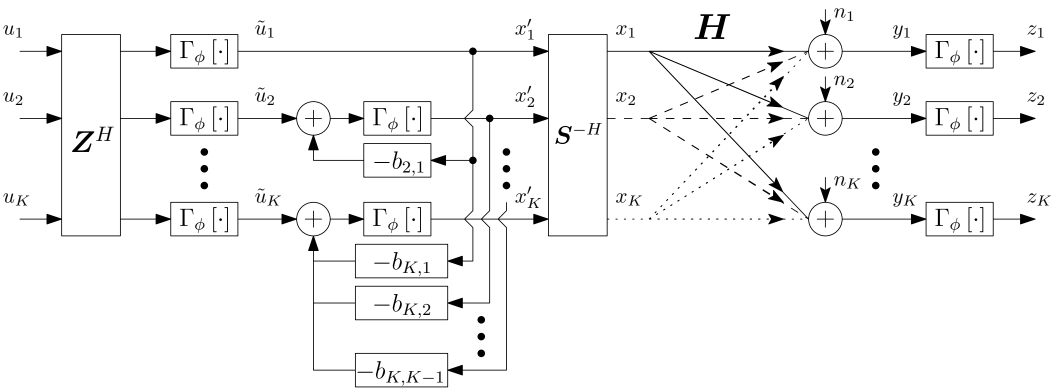


FIGURE 2.15: LR-THP transceiver scheme shown in $\mathbb{C}^K$, $b_{i,j}$ is the element on the ith row and jth column of $\boldsymbol{B}^H$.

2.6 Integer Forcing Precoding

As an emerging lattice reduction aided multi-user technique, the integer forcing precoding (IFP) [171] solution provides a novel yet familiar perspective of approximate lattice precoding. Conceptually, IFP is inspired by the *reverse compute-and-forward* (RCoF) protocol [172] used in wireless systems, which is also based on the parallel modulo Gaussian channel principle of all non-linear MUPs. The IFP scheme proposed in [171] may in fact be considered as a generalization of the aforementioned approximate lattice precoding schemes. By comparing the schematic of the LR-ZFP in Fig. 2.14 to that of the LR-THP in Fig. 2.15, it is plausible that both of the approximate lattice precoding schemes have three common building blocks at the transmitter besides the power controller. Additionally, the coolant may be viewed as two consecutive blocks in both approximate lattice precoding and IFP. In specific, these are generalized as follows:

---

[6] This is somewhat different from the structure of Fig. 2.7 because the independent equalizers $1/r_{k,k}$ on the right side of Fig. 2.7 can be 'transferred' to the ONU side on the left if the decomposition of (2.9) produces a triangular matrix having unit diagonal.

- **Integer-valued filter**. In conventional approximate lattice precoding, this is exactly the unimodular transformation matrix, which is the solution of the SVP associated with the act of lattice reduction itself. However, instead of using a strictly invertible integer-valued matrix, IFP relaxes the criterion and only requires the integer-valued matrix to be invertible over Galois fields. The underlying lattice problem is the shortest independent vectors problem (SIVP) [173], whose solution does not necessarily constitutes a basis, but is obtainable via lattice reduction strategies [174].
- **Lattice coding**. The lattice mapping of IFP is not necessarily carried out over either the multi-user channel lattice or its dual lattice. By contrast, the codebook is constructed using a pair of nested lattices [175][176] with the standard 'construction A' approach of [177]. The fundamental properties of the lattice codebooks are ultimately bounded by the Hermite constant as presented in (2.32). Recent research in lattice coding shows that we may achieve the same goal using only a single 'good' lattice [152]. Nonetheless, the coding procedure itself is ultimately a CVP solver.
- **Linear beamformer**. In the special case, this component is identical to the FEXT canceller. For approximate lattice precoding, the front-end linear beamformer is obtained via factorizing the reduced lattice basis. IFP allows for more relaxed choices of this front-end, subject to the specific power constraint of the system. In particular, if the identity matrix is used, then the IFP scheme becomes equivalent to the standard RCoF [171].

IFP bridges the lattice coding approach and the lattice reduction approach, the two main applications of lattice theory in telecommunications. In particular, IFP has the potential of borrowing good practices from lattice coding for improving the MUP performance beyond that of the conventional approximate lattice precoding. As a consequence of the more flexible lattice coding design, the VMR of IFP is not generally deterministic and it may also overlap with the VMR associated with another LRMUP. More interestingly, the distributed IFP-aided receivers can 'virtually' cooperate, which is contrary to the common belief, namely that the distributed downstream receivers cannot cooperate. Specifically, cooperative signal processing is typically associated with (multi-user) matrix filtering. In IFP, the receivers have to acquire the columns of the integer-valued matrix during initialization, which virtually enables matrix filtering. This particular aspect may lead to other cooperative distributed receiver designs relying on quantized information (such as CSI) gleaned from the downstream transmitter.

2.7 Vector Perturbation

Even though Babai had shown in [157] that the performance gap between LLL based approximate lattice decoding and MLD is reasonably small, attaining the optimal performance is still relevant, if we want to reach the maximum bandwidth efficiency in the forthcoming G.mgfast and beyond. Reflecting on the optimal MUP criterion of (2.25), the MUP algorithm for finding the optimal transmit power coolant (cf. 1.4.2) is known as the vector perturbation [162] scheme. The most widely recognized solver of the underlying exact CVP problem is more generally known as the sphere encoder[7]. We note that the cost function of (2.25) may be optimized using other near-optimal heuristic methods, as in the MUPs of [96][178]. However, the (near-)exact solution is valuable in that it allows full exploitation of the multi-dimensional signal space spawned by the multi-pair channels, because the performance gap between the suboptimal MUPs and the optimal MUP can be enormous for DSL binders enclosing a large number of copper pairs[8]. In particular, with the popularization of SDN and NFV, future wireline networks may be able to consolidate different signal spaces resulting from LLU for all vectored groups simultaneously, in order to regain the loss imposed by alien FEXT. In the following sections, we will investigate the members of the sphere encoder family proposed for finding a near-optimal additive cooling signal according to the basic formulation of (1.5).

In Fig. 2.16, we present the equivalent schematic of (downstream) VP and that of the (upstream) MLD in the original $\mathbb{C}^K$. To demonstrate explicitly the underlying duality principle, we partition the MLD's sphere decoder into two consecutive stages. In this equivalent MLD, the difference vector $\boldsymbol{m}$ from the received vector $\boldsymbol{y}$ to its nearest lattice vector $\boldsymbol{H}\boldsymbol{z}_{\mathrm{ML}}$ is calculated first, which is then subtracted from $\boldsymbol{y}$. In this case, applying the linear ZF MUD becomes optimal. By comparison, the VP preprocessing scheme constitutes an extension of the ZFP, where the cooling process (from u_k to x'_k) is invoked in order to optimally reduce the transmit power at the output of the ZFP's encoder $\boldsymbol{G}$.

The exact solution of (2.25) is found by the enumeration-based sphere encoder. The original *depth-first* type sphere decoder was conceived in [179], which was then further improved both in [180] and [181]. The breadth-first type [182, 183, 184] and the more recent best-first type [185, 186, 187] enumeration algorithms were conceived as more hardware-friendly versions, since they can be readily pipelined to have a fixed throughput. All versions of the enumeration algorithms were claimed to approach the MLD performance, relying on a simple fact of the CVP problem, namely that if a lattice point is the closest one to another point in $\mathbb{R}^{2K}$ within a hypersphere region centred at the given point, then said lattice point is also the closest to the given point within the entire signal space. However, choosing the optimal radius of such a hypersphere is a hard lattice problem [188]. This problem is synonymous with finding the *covering radius* of the given lattice, whose complexity was investigated in [189]. If the radius is too large, then there

---

[7] For MLD this is typically called the sphere decoder. As seen in the difference between (2.23) and (2.25), the exact CVP solver operates in a finite lattice for MLD while the VP scheme can potentially work in infinite lattice. To avoid ambiguity, the latter is denoted as sphere encoder here.

[8] In fact, as a known result in lattice theory, both the exact and the near-optimal solution (whose Euclidean distance to the true optimal is upper bounded by a constant) of the CVP are obtained in exponential complexity order. Conversely, solutions found by low-complexity heuristic algorithms usually have exponentially increasing Euclidean distance to the true optimal as a function of the lattice dimension.

are too many lattice points to be enumerated in the hypersphere, whilst if the radius is too small, then the CVP search will fail and we are forced to extend the radius to restart the enumeration. Additionally, in the breadth- and best-first versions (BFSE), tuning the 'breadth' $\mathcal{K}$ optimally is also a difficult problem, if both the MLD-like performance and the minimum buffer size are desired.

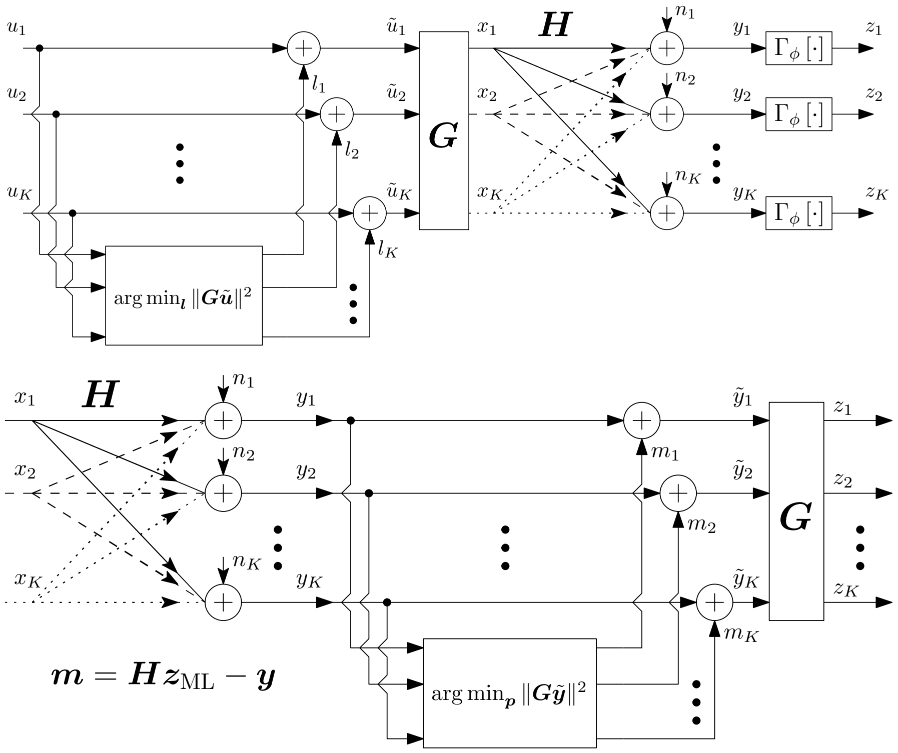


FIGURE 2.16: Duality view of the optimal MUP and MUD, based on vector perturbation (top) and maximum likelihood detection (bottom).

In this section, we will characterize the classic depth-first [180] and best-first [184] sphere encoders as a pair of well-established examples. Additionally, based on the claimed superiority of another class of sieving-based CVP algorithms conceived for large-scale lattice based cryptography, we propose a new approach to CVP which strikes an attractive trade-off between the buffer-size requirement and complexity for large-scale multi-pair DSL networks. For a more in-depth analysis of the sphere encoder, the reader is redirected to the contribution of [190] and the references therein.

2.7.1 Depth-First Sphere Encoder

The sphere encoder consists of the lattice reduction preprocessing of Section 2.5 followed by the search-tree-based enumeration. In the example of Fig. 2.10, the lattice is spawned in the 2D plane $\mathbb{R}^2$ having the generator matrix $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2]$. In order to determine the closest lattice point (the black dot) to the given point $\bar{\boldsymbol{u}} = -\boldsymbol{G}\boldsymbol{u}$, the sphere encoder commences by transforming the long basis of $\boldsymbol{G}$ into a short one $\boldsymbol{G}' = [\boldsymbol{g}'_1, \boldsymbol{g}'_2]$ using lattice reduction. The main purpose of such a transformation is to accelerate the subsequent enumeration process as proven in [170].

```
Input: Lattice Generator $\boldsymbol{G}$, a given point $\bar{\boldsymbol{u}} \in \mathbb{R}^{2K}$, the sphere radius $\sqrt{\beta} = \infty$;
Output: Integer vector $\boldsymbol{l} \in \mathbb{Z}^{2K}$, s.t. $\bar{\boldsymbol{u}} \in \mathcal{V}(\boldsymbol{G}, \boldsymbol{G}\boldsymbol{l})$;
$[\boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{Z}] = \text{Reduction}(\boldsymbol{G})$;
$\boldsymbol{v} \leftarrow \boldsymbol{Q}^H \bar{\boldsymbol{u}}/\phi$;
$c_{2K} \leftarrow v_{2K}/r_{2K,2K}$;
$l'_{2K} \leftarrow \lceil c_{2K} \rfloor$;
$k \leftarrow k - 1$;
while true do
    if $\sum_{j=k+1}^{2K} r_{j,j}^2 (l'_j - c_j)^2 < \beta$ then
        if $k > 1$ then
            $c_k \leftarrow (v_k - \sum_{j=k+1}^{2K} r_{k,j} l'_j)/r_{k,k}$;
            $l'_k \leftarrow \lceil c_k \rfloor$;
            $k \leftarrow k - 1$;
        else A valid intermediate $\boldsymbol{l}'$ is found
            $\beta \leftarrow \|\boldsymbol{v} - \boldsymbol{R}\boldsymbol{l}'\|^2$;
            $k \leftarrow 2$;
            Choose the next value for $l'_2$ in order;
        end
    else
        $k \leftarrow k + 1$;
        if $k > 2K$ then
            Terminate;
        end
        Choose the next value for $l'_k$ in order;
    end
end
$\boldsymbol{l} \leftarrow \phi \boldsymbol{Z} \boldsymbol{l}'$;
```

Algorithm 1: Depth-First Sphere Encoder

By performing QR decomposition following lattice reduction $\boldsymbol{G} = \boldsymbol{QRZ}^{-1}$ on the inverse channel matrix $\boldsymbol{G} = \boldsymbol{H}^{-1}$, we can map the sphere encoding problem of (2.25) from a general matrix $\boldsymbol{G}$ onto a triangular one $\boldsymbol{R}$ by rotation as follows:

$$\boldsymbol{l}'_{opt} = \arg\min_{\boldsymbol{l}'} \|\boldsymbol{v} - \boldsymbol{R}\boldsymbol{l}'\|^2, \text{ where } \boldsymbol{v} = \boldsymbol{Q}^H \bar{\boldsymbol{u}}/\phi, \ \boldsymbol{l}' = \boldsymbol{Z}^{-1}\boldsymbol{l}/\phi. \tag{2.37}$$

The constant ϕ is removed so that $\boldsymbol{l}'$ becomes an integer-valued vector in $\mathbb{Z}^{2K}$. For the initialization of the sphere encoder, an approximate solution to (2.37) can be easily obtained by solving the respective least squares problem, where we solve $\boldsymbol{l}'$ for a real-valued vector in $\mathbb{R}^{2K}$ first, and then we round off the entries of the real-valued vector to their respective nearest integer. Denoting the Euclidean distance from $\boldsymbol{v}$ to this approximate solution as $\sqrt{\beta}$, we may enumerate the lattice points within the $2K$-dimensional hypersphere of radius $\sqrt{\beta}$ centred at $\boldsymbol{v}$. Let $r_{k,j}$ be the entry at the kth row and jth column of the upper-triangular matrix $\boldsymbol{R}$ of (2.37). Then we may define the following

DFE structure as in [168]:

$$c_k = \begin{cases} \dfrac{v_k}{r_{k,k}} & k = 2K, \\ \dfrac{v_k}{r_{k,k}} - \sum\limits_{j=k+1}^{2K} \dfrac{r_{k,j}}{r_{k,k}} l'_j & k = 2K-1, \cdots, 1. \end{cases} \tag{2.38}$$

Similar to (2.12), Eq. (2.38) has a bottom-up decision feedback structure. Therefore, the sphere constraint may be invoked in each iteration of (2.38), corresponding to the subspaces having lower than $2K$ dimensions. A necessary condition for a lattice point to fall inside the hypersphere centred at $\boldsymbol{v}$, i.e. $\|\boldsymbol{v} - \boldsymbol{R}\boldsymbol{l}'\|^2 < \beta$, is formulated as [168]:

$$r_{k,k}^2 (l'_k - c_k)^2 < \beta - \sum_{j=k+1}^{2K} r_{j,j}^2 (l'_j - c_j)^2, \quad k = 2K, 2K-1, \cdots, 1. \tag{2.39}$$

Eq. (2.39) represents a tree structure from the root node of $k = 2K$ to the leaf node of $k = 1$, in which we may compute the partial Euclidean distance of $\sum_{j=k+1}^{2K} r_{j,j}^2 (l'_j - c_j)^2$ from the enumerated lattice point to $\boldsymbol{v}$ in the $(2K - k + 1)$-dimensional subspace. The sphere encoder progresses towards the leaf node by one step whenever (2.39) is satisfied, hence the algorithm is depth-first. Otherwise, it backs up towards the root node by one level to evaluate a different lattice point. The evaluation at each node is one-dimensional along the integer axis $\mathbb{Z}$ in the Schnorr-Euchner sequence [180] $l'_k, l'_k - 1, l'_k + 1, l'_k - 2, l'_k + 2, \ldots$, starting from the integer $l'_k = \lceil c_k \rfloor$, where $\lceil \cdot \rfloor$ is the rounding-off operator. Whenever the leaf node of $k = 1$ is reached, the radius $\sqrt{\beta}$ of the hypersphere boundary is updated and the corresponding state represents a new candidate lattice point. The final solution of (2.37) is the last lattice point obtained before the sphere encoder terminates its operation.

2.7.2 Parallel Sphere Encoder Designs

The complexity order of the depth-first sphere encoder grows exponentially with the dimension of the lattice. Moreover, the depth-first approach involves the sequential enumeration of the lattice points and the real-time update of the hypersphere constraint during enumeration. These characteristics make it difficult to convert the search tree in the DFSE into a pipelined structure in order to accelerate the search process using parallel computing. Without modifying the basic SIC-like framework implied in (2.38), in this section we will compare a pair of parallel sphere encoder designs.

2.7.2.1 Sub-Tree Enumeration

In [191], an asynchronous parallel enumeration strategy based on runtime search tree partitioning was proposed, which may be viewed as interactively executing multiple instances of Alg. 1 for exchanging the hypersphere boundary condition during runtime.

The basic concept of the algorithm is summarized as follows. During the full search, each valid node in the tree is expanded into a sub-tree, which is then immediately assigned to one of the inactive parallel threads referred to as a sub-enumeration thread (SET) until all SETs from the parallel pool become active. As a result, each SET operates asynchronously. Whenever a SET successfully reaches a leaf node, it updates the radius of the global hypersphere boundary in real time. The updated boundary condition is then passed to all active SETs in the parallel pool. Due to the shorter constraint length, a fraction of the active SETs may terminate earlier. The terminated SETs become inactive until a new sub-tree is spawned by another active SET. The whole algorithm terminates, when no more SETs are in operation.

It is plausible that this particular strategy achieves a speed-up factor proportional to the size of the parallel pool in the simulations reported by the authors of [191]. However, the massive real-time message-passing overhead due to both sub-tree assignment and global boundary update makes the algorithm less appealing for hardware implementations. These complications may also result in premature termination and suboptimal performance depending on the parallel pool's hardware operations.

2.7.2.2 Best-First Enumeration

In contrast to both the DFSE and the SET-based strategy, the best-first sphere encoder only progresses through the search tree (from the root $k = 2K$ to the leaf $k = 1$) once, dispensing with any backwards travel. In the search tree's kth level, the BFSE, notably the $\mathcal{K}$-best variant, enumerates $\mathcal{K}_k$ elements in the Schnorr-Euchner sequence for each of the $\mathcal{K}_{k+1}$ surviving paths from level $k + 1$. The $\mathcal{K}_k\mathcal{K}_{k+1}$ extended paths are then sorted in ascending order based on the accumulated partial Euclidean distances (the second term on the right hand side of (2.39)), in which case the first $\mathcal{K}_k$ paths are admitted to level $k - 1$ in the search tree. The process is repeated until the leaf node is reached, where the $\mathcal{K}_1$ best lattice vectors are selected as the final encoder output.

Generally, a BFSE does not necessarily have an exponentially growing complexity with the dimension of the lattice. Instead, the complexity of the BFSE increases linearly with the lattice dimension if we choose a small breadth $\mathcal{K} << 2K$ per enumeration level. However, this approach is likely to prevent the BFSE from finding the global optimum, unless the breadth $\mathcal{K}$ is dynamically configured to increase exponentially, when progressing through the search tree, so that the probability of prematurely discarding the globally optimal choice approaches zero. Moreover, the BFSE follows the SIC architecture just like the DFSE, which results in a hard complexity lower bound.

2.7.3 Irregular Differential Evolution Aided Lattice Encoder

In the era of intelligence-aided and software-defined networking, AI is regarded as a powerful data-driven technique of tackling hard practical problems that do not have an analytically tractable mathematical model. In this section, we propose a novel MUP scheme relying on both lattice reduction and a variant of the intelligent *differential evolution* algorithm. In our proposed *irregular differential evolution aided lattice encoder* (IDEALE), the powerful lattice reduction preprocessing provides us with a reference coolant vector. Furthermore, if the reference coolant vector fails to achieve the system's target QoS, IDEALE will proceed to call the differential evolution aided progressive sieving routine in order to find the optimal coolant vector during retransmissions, as requested by ARQ. The success of this hybrid strategy also reflects the recent progresses made in lattice-based cryptography, where the deterministic enumeration-based strategy and the randomized sieving-based strategy are combined for solving CVP more efficiently than using either strategy individually. The two stages of IDEALE are described in details as follows.

2.7.3.1 Babai's Bounds

As we have discussed, the main problem of using ZFP over DSL channels having both poor direct channel gain and bad conditioning is the SNR penalty incurred by channel inversion. The same problem applies to THP without the aid of joint spectrum balancing. As characterized by the VMRs associated with ZFP and THP in Fig. 2.12 and by the lattice quality performance in Fig. 2.13, it may be readily seen from a lattice theoretical perspective that the problem of ZFP is the unbounded length of the vectors constituting the lattice basis. Therefore the vectors located within the ZFP's VMR may have a worst-case length of infinity. Consequently, the approximate solution of the CVP of (2.25) found by ZFP may have an arbitrarily large distance from the exact CVP solution.

Fortunately, it was theoretically shown in Babai's pioneering contribution [157] that lattice reduction can upper bound the worst-case performances of suboptimal heuristic MUP algorithms. More specifically, in the case of finding the coolant vector for the VP-based MUP encoder in (2.25), a pair of suboptimal heuristic algorithms may be used for finding approximate solutions relying on (LLL-)reduced lattice basis, which are termed as the *rounding-off* and the *nearest-plane*[157] solutions, respectively[9]. We note that the pair of operations strongly resembles the encoding steps carried out in the approximate lattice precoding schemes of Sec. 2.5. Nonetheless, based on the pair of LLL reduction criteria in (2.29) and (2.30), the following bounds may be defined for the performance of the rounding-off and the nearest-plane algorithms with respect to (2.25), respectively.

---

[9] For historical reasons, Babai's heuristics are commonly associated with the lattice reduction prerequisite, even though the original algorithm proposed in [157] did not have such a requirement.

In particular, for an LLL-reduced $2K$-dimensional lattice basis $\boldsymbol{G}$ found with a reduction factor of $\omega = 3/4$, we have [157]:

Theorem 2.4. *The* rounding-off *algorithm finds a lattice vector* $\boldsymbol{G}\boldsymbol{L}_{\mathrm{RO}}$ *that is near the target vector* $\boldsymbol{G}\boldsymbol{u}$ *within the constraint of* $\|\boldsymbol{G}(\boldsymbol{u} - \boldsymbol{l}_{\mathrm{RO}})\| \leq [1 + 4K(9/2)^{K}]\|\boldsymbol{G}(\boldsymbol{u} - \boldsymbol{l}_{\mathrm{SE}})\|$*, where* $\boldsymbol{G}\boldsymbol{L}_{\mathrm{SE}}$ *is the optimal solution given by the sphere encoder.*

Theorem 2.5. *The* nearest-plane *algorithm finds a lattice vector* $\boldsymbol{G}\boldsymbol{L}_{\mathrm{NP}}$ *that is near the target vector* $\boldsymbol{G}\boldsymbol{u}$ *within the constraint of* $\|\boldsymbol{G}(\boldsymbol{u} - \boldsymbol{l}_{\mathrm{NP}})\| \leq 2^{K}\|\boldsymbol{G}(\boldsymbol{u} - \boldsymbol{l}_{\mathrm{SE}})\|$*, where* $\boldsymbol{G}\boldsymbol{L}_{\mathrm{SE}}$ *is the optimal solution given by the sphere encoder.*

The coefficients on the right hand side of the inequalities in both Th. 2.4 and 2.5 may be interpreted as a ‘proximity factor’ [160] with respect to the shortest possible Euclidean distance to the target vector. The pair of upper bounds naturally becomes tighter when $\omega \to 1$. Recent results in [160][166] have provided tighter bounds for both cases, however these bounds remain in the exponential order in terms of their lattice dimension. Therefore, it is currently more practical to use semi-analytical models such as the VMRs (cf. Ch. 3) to characterize their average-case performances.

2.7.3.2 Differential Evolution Aided Progressive Sieving

In lattice-based cryptography, *sieving* is a class of widely-recognized CVP solving strategies. Specifically, sieving-based CVP algorithms commence from a large list of *randomly-selected* lattice vectors and subtract them from each other in order to progressively obtain better solutions until the global optimum is found. In contrast to the *optimal* enumeration strategies such as DFSE (Alg. 1), sieving requires large buffer space in order to find the global optimum. Despite the exploitation of spatial resources, sieving still has an exponentially increasing complexity order in the lattice dimension. However, if lattice preprocessing is allowed, then the complexity of solving the corresponding *closest vector problem with preprocessing* (CVPP) via sieving may become similar to or even lower than that of the classic enumeration strategies. This is because in enumeration algorithms, the lattice vectors are visited following a deterministic path from the initial solution to the global optimum, whilst in sieving, there is a non-zero probability of skipping at least some of the intermediate solutions due to the randomness of the algorithm. Sieving becomes more efficient when its enumeration based counterpart has to traverse many intermediate lattice vectors.

Due to hardware limitations in the DP of G.mgfast, it is impractical to have a buffer whose size is an exponential function of the number of copper pairs in a massive DSL binder, while also being a linear function of the number of discrete tones. Therefore, in order to realistically exploit the randomness of sieving, an irregular version of the differential evolution algorithm is proposed for progressively generating and sieving lattice vectors under ‘weakly guided randomness’. The concept of the conventional (continuous) differential evolution algorithm [192] is redefined here in the lattice problem domain. Specifically, we define the mutation operator as a random jump between a pair of adjacent lattice vectors, and the crossover operator as a recombination of the component vectors (drawn from the lattice basis) between two lattice vectors. Furthermore, we exchange their order of appearance in the algorithm in order to ‘guide’ the algorithm’s random progression, which will in return reduce the buffer size requirement and increase the probability of finding the optimal solution. For optimizing the cost function (2.25), the basic operations of the algorithm are described as follows (Fig. 2.17):

1. **Initialization**. Babai’s rounding-off algorithm is performed, which finds an initial reference coolant vector $\boldsymbol{l}_{\mathrm{b}}$. If the QoS target cannot be satisfied with the aid of $\boldsymbol{l}_{\mathrm{b}}$, then the members of the initial population $\{\boldsymbol{l}_1\}$ are initialized according to:

   $$\boldsymbol{l}_{p,1} = \begin{cases} \boldsymbol{l}_{\mathrm{b}} & p = 1, \\ \boldsymbol{l}_{\mathrm{b}} + \boldsymbol{\delta}_{p-1} & p = 2, 3, \cdots, N_p, \end{cases} \tag{2.40}$$

   where N_p is the size of the population, while $\boldsymbol{\delta}_p$ represents randomly-generated sparse vectors whose non-zero entries are drawn from ± 1. This initialization approach ensures the near-optimality of the initial population. If the buffer supports a large N_p, then $\boldsymbol{\delta}_p$ is allowed to draw uniformly distributed random integers from a wider symmetric interval such as $[-2, 0) \cup (0, 2]$. Due to the lack of ergodicity of lattice vectors, allowing a relatively dense fluctuation vector $\boldsymbol{\delta}_p$ over small N_p may increase the risk of significantly reducing the quality of the initial population $\{\boldsymbol{l}_1\}$. This drawback outweighs the small rise of diverging from a locally optimal lattice vector.

2. **Crossover** Also known as *discrete recombination*, this specific operation uses a crossover mask M_C to select and recombine the component vectors, i.e. subspace projections, from a randomly-selected pair of members of the current generation g. Specifically, the ‘child set’ in the gth generation $\{\tilde{\boldsymbol{l}}_g\}$ is produced as follows:

   $$\tilde{\boldsymbol{l}}_{p,g} \supset \tilde{l}_{k,p,g} = \begin{cases} l_{k,p,g}, & \text{if } \operatorname{rand}(0,1) < M_C, \\ l^{\Pi}_{k,p,g}, & \text{otherwise}, \end{cases} \tag{2.41}$$

   where $\{\boldsymbol{l}_g\}_{\Pi} \supset \boldsymbol{l}^{\Pi}_{p,g} \supset l^{\Pi}_{k,p,g}$ represents a random permutation Π of the ‘parent set’ $\{\boldsymbol{l}_g\}$, and $\operatorname{rand}(0,1)$ represents a uniformly distributed random variable within the range $(0, 1)$. If all of the member vectors in the ‘parent set’ are sufficiently fit and non-overlapping, they should be close to each other in the lattice. As a result, the probability of producing significantly worse member vectors for the ‘child set’ becomes low, if we simply recombine the component vectors.

3. **Mutation** In contrast to the regular definition of mutation in the differential evo-

lution algorithm, we define the specific operation as a random jump over the lattice grid, which is characterized by the LLL-reduced basis obtained during initialization. This modification addresses a potential problem associated with using the regular mutation strategy in our context. More specifically, the difference of two close lattice vectors may be very far away from either of them, which does not constitute a good 'trial set' $\{\hat{\boldsymbol{l}}_{p,g}\}$. In particular, the mutation operation is defined for a random movement vector $\tilde{\boldsymbol{\delta}}_p$ as follows:

$$\hat{\boldsymbol{l}}_{p,g} = \begin{cases} \tilde{\boldsymbol{l}}_{p,g} + \tilde{\boldsymbol{\delta}}_p, & \text{if } \gamma(\tilde{\boldsymbol{l}}_{p,g} + \tilde{\boldsymbol{\delta}}_p) - \gamma(\tilde{\boldsymbol{l}}_{p,g}) < 0, \\ \tilde{\boldsymbol{l}}_{p,g}, & \text{otherwise}, \end{cases} \tag{2.42}$$

Unlike in (2.40), the random jump $\tilde{\boldsymbol{\delta}}_p$ in (2.42) does not have to be based on a sparse vector, but its entries must be selected strictly from $\{-1, 0, 1\}$ in order to guarantee that the change is only by a single grid position. Moreover, the mutation strategy in (2.42) does not rely on calculating the difference of the member vectors. Instead, it uses the difference of the cost function values, i.e. the Euclidean distance from the target vector, in order to produce a 'trial set' that is at least as good as the 'child set'. This is the principal stage that characterizes the 'progressive sieving' concept. However, this operation incurs an additional N_p cost function evaluations compared to the conventional mutation.

4. **Selection** In order to determine the member vectors to be retained for the $(g+1)$st generation, the cost of the 'trial set' is compared against that of the 'parent set'. Because the costs of both the 'trial set' and of the 'parent set' are known, there are no additional cost function evaluations at this stage. Similar to (2.42), each member vector of the $(g+1)$st generation may be formulated as:

$$\boldsymbol{l}_{p,g+1} = \begin{cases} \hat{\boldsymbol{l}}_{p,g}, & \text{if } \gamma(\hat{\boldsymbol{l}}_{p,g}) - \gamma(\boldsymbol{l}_{p,g}) < 0, \\ \boldsymbol{l}_{p,g}, & \text{otherwise}. \end{cases} \tag{2.43}$$

When the new 'parent set' of the $(g+1)$st generation is filled, the cost of the member vectors is compared to that of the current best member. The best member vector is updated if the best member in the new 'parent set' has a lower cost than the existing one, where the 'converged generation counter' Δg is reset to zero. If no new best member can be found, then Δg is increased by one. The algorithm terminates when Δg reaches its designated limit $\Delta g_{\max}$.

2.7.3.3 Performance

It is now trivial that the performance of MUP strongly depends on the power efficiency of the transmitter. More specifically, the optimality of the solution we find for (2.25)

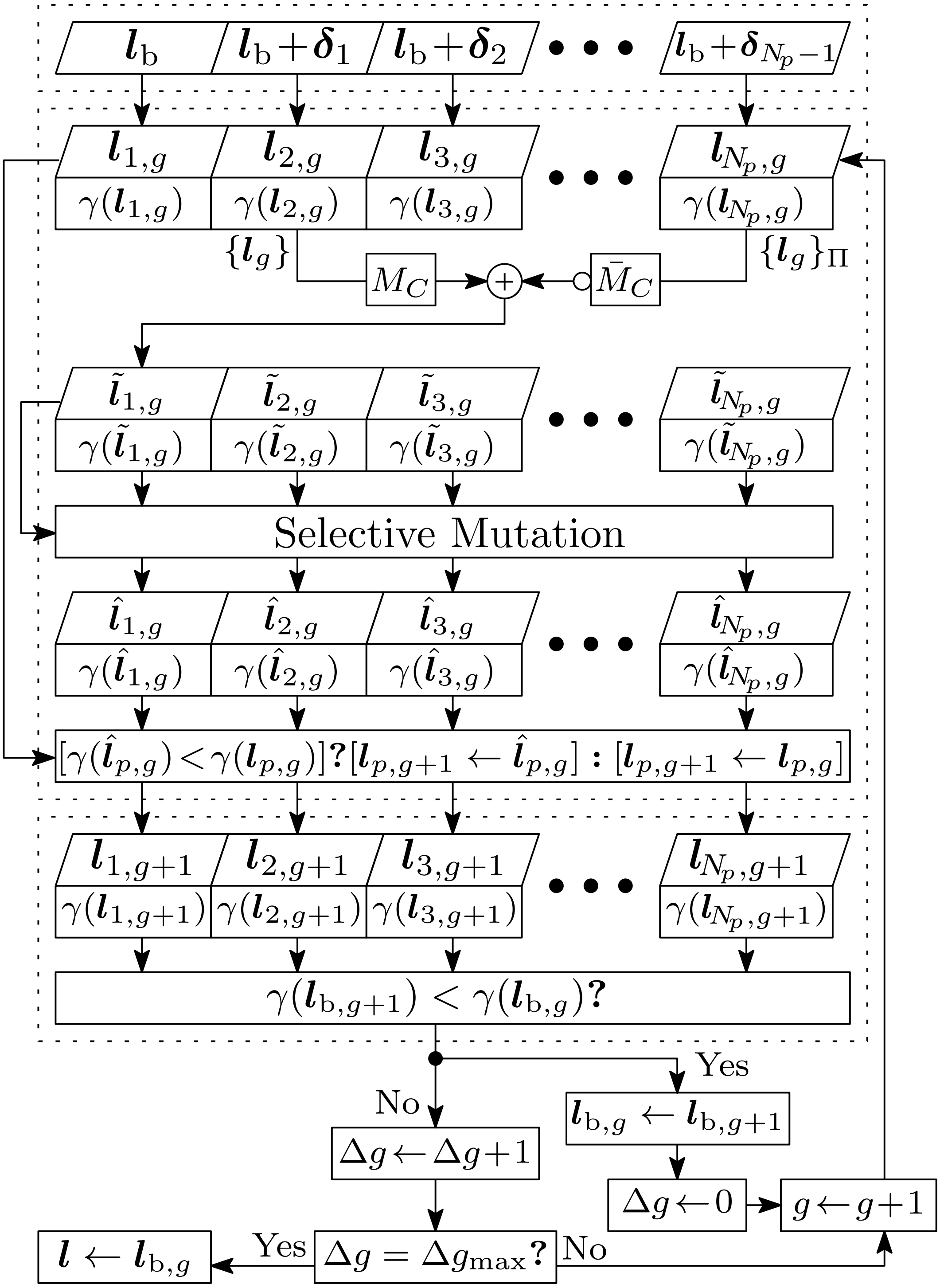


FIGURE 2.17: System diagram of irregular differential evolution aided lattice encoder (IDEALE). The initial population is found using Babai's rounding off algorithm [157] plus a random sparse binary vector $\boldsymbol{\delta}_p$ for $p = 1, 2, \cdots, N_p - 1$.

decides how much power budget is still available for signalling after the MUP is applied. From a lattice-theoretical perspective, the equivalent performance measure becomes the minimum Euclidean distance in the multi-user signal space. In this respect, a pair of closely related metrics is widely considered in the lattice literature, namely the *proximity factor* we mentioned in Th. 2.4 and the *success probability* investigated in [170]. Using the formulation in Th. 2.4, the success probability of a CVP algorithm $\mathcal{A}$ is defined as $P(\boldsymbol{l}_{\mathcal{A}} = \boldsymbol{l}_{\mathrm{SE}})$. Here, we relax the definition by including the influence of the proximity factor, which is formulated as:

Definition 2.6. For a given proximity factor $\mathcal{F}$, the success probability of an algorithm $\mathcal{A}$ solving (2.25) is defined as $P(\|\boldsymbol{G}(\boldsymbol{u}+\boldsymbol{l}_{\mathcal{A}})\| \leq \mathcal{F}\|\boldsymbol{G}(\boldsymbol{u}+\boldsymbol{l}_{\mathrm{SE}})\|)$, where $\boldsymbol{l}_{\mathrm{SE}}$ is the solution found by the optimal DFSE.

For IDEALE, we plot its performance in Fig. 2.18, which is averaged over the lattices spawned by a 20-pair DSL binder over the 212-424 MHz band for the interest of the quasi-worst-case performance in the forthcoming G.mgfast. The entries of $\boldsymbol{u}$ are drawn from the (square) 16-QAM constellation. In spite of the low success probability of finding the global optimum, IDEALE is capable of finding solutions of $\mathcal{F} < 1.15$ with a higher than 90% probability. Moreover, IDEALE is almost guaranteed to find a solution of $\mathcal{F} < 1.3$ according to Fig. 2.18. By contrast, we may observe that the $\mathcal{K}$-best sphere encoder with a choice of $\mathcal{K} = 15$ yields a moderately higher success probability than IDEALE under a given $\mathcal{F}$. As we will see later, such an improvement comes at the expense of increased complexity.

2.8 Other Lattice-Aided MUP Variants

Albeit VP is capable of achieving the optimal MLD performance and it functions as a simple attachment to the widely deployed ZFP, it exhibits a high PAPR and a potentially prohibitive sphere encoding complexity. In terms of PAPR, the spliced constellation of Fig. 1.11 suggests that the dynamic range required by the front-end filter $\boldsymbol{G}$ is significantly increased, especially since the constellation expansion depends on the worst-case channel quality of all tones. Therefore we in general need more expensive circuitry for the DP of G.mgfast using the VP-based approach. As a design alternative of sphere encoding, Zhang *et al.* [161] proposed the expanded constellation mapping scheme that solves the closest point problem of a *finite lattice*, which is more closely related to the *sphere decoding* based implementation of MLDs. The sphere decoding algorithm naturally works on a finite lattice, where the candidate lattice points are confined in a pre-determined region. However, sphere decoding usually has a higher complexity than sphere encoding because of the extra overhead of the former required for remapping the invalid solutions, which are found outside the boundaries of the finite lattice [193]. Meanwhile, the family of powerful lattice reduction techniques such as LLL are generally not favoured in sphere decoding, since it would be difficult to keep track of the boundary of the finite lattice. Instead, less sophisticated transformations such as the SQRD should be used, which however increases the complexity of the enumeration process. Regarding other characteristics of the existing G.fast system with respect to its potential practical deployment in VP, its backwards compatibility with linear precoding based receivers was addressed in [194], while a particularly robust vectoring feedback error mitigation arrangement was proposed in [195].

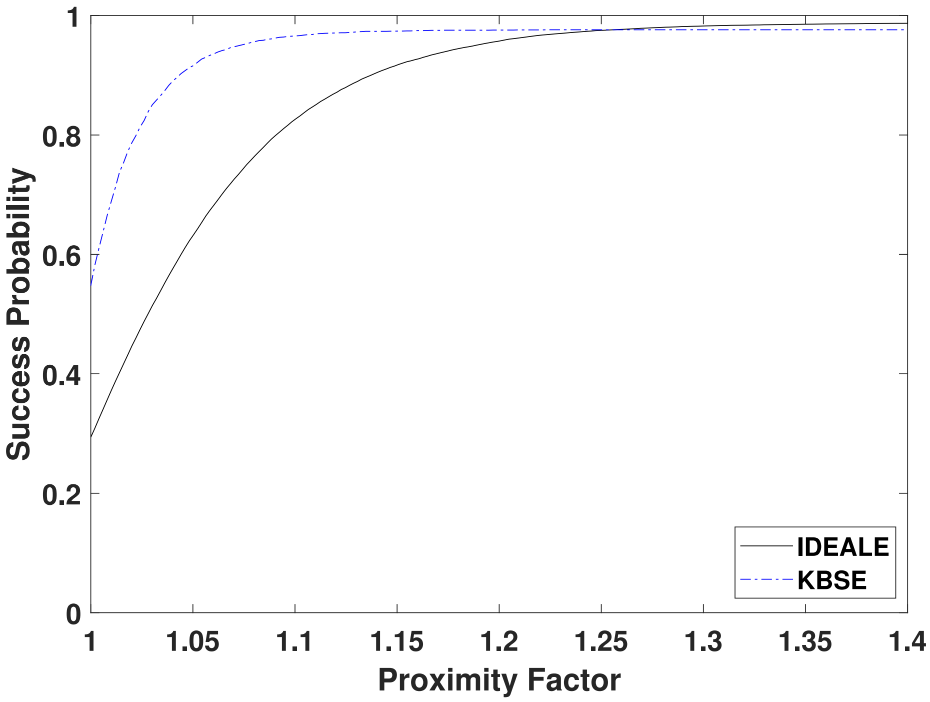


FIGURE 2.18: The success probability of IDEALE and the $\mathcal{K}$-best sphere encoder as a function of the nominal proximity factor for solving (2.25) over the 212-424 MHz bandwidth in a 20-pair 50 m DSL binder.

As a well-investigated subject, the DFSE has both an average-case and a worst-case complexity that grows exponentially with the system's dimensionality [196], which constitutes a substantial disadvantage compared to the deterministic polymonial (quadratic) complexity order of the conventional linear and nonlinear vectoring schemes, as well as compared to the IFP (depending on the choice of the lattice coding scheme) and to the family of approximate lattice precoding schemes, when ignoring the complexity overhead of initialization. This had led to renewed efforts invested in reducing the complexity of the conventional sphere decoder/encoder [197, 198, 199], as well as to the optimization of the lattice reduction preprocessing [200]. Theoretically, the currently provable best-case algorithm for finding the exact solution of CVP has a complexity order of $\mathcal{O}(2^{\mathcal{O}(2K)})$ relying on a buffer size of $\mathcal{O}(2^{\mathcal{O}(2K)})$ [201][202], given that the solver has *a priori* knowledge of the Voronoi cell. However, the particular solver invoked in this case is very impractical for high throughput transmission compared to the sphere encoder, and the performance difference is negligible for medium-sized systems.

Since the VP scheme constitutes a nonlinear expansion of the linear ZFP, the concept of the reinforced versions of ZFP may also be used for enhancing the performance of VP. The MMSE criterion based VP scheme has been proposed relying on either a continuous-and-discrete hybrid perturbation [203][204] strategy, or a regularized channel inversion strategy [162][205]. As seen in Section 2.3.2.3, this is to overcome the same modulo loss caused by the non-linear receivers. Additionally, as studied in [206], VP is a flexible scheme that, when expanded over multiple symbol durations, can fully regain the 'shaping loss' via the construction of nested lattice constellations. The expansion of VP over time allows for a more efficient exploitation of the non-causal CSI knowledge (towards the original concept of DPC), whose gain was also empirically characterized in [207][208].

2.9 Complexity

In this section, we consider the complexity of the aforementioned MUP algorithms. Since the system requires initialization, we will briefly investigate the complexity of the initialization (excluding that of the DSB-related power controller configuration) and that of the actual encoding process independently.

The initialization of LR-THP and LR-ZFP mainly involves the LLL-reduction, which has an average-case complexity order of $\mathcal{O}[(2K)^2\log(2K)]$. LR-THP does not require additional QR decomposition which THP-like schemes do, because it is naturally handled by the LLL-reduction. The initialization of ZFP and THP requires matrix inversion and triangular decomposition (both imposing a complexity order of $\mathcal{O}(K^3)$ for complex-valued matrices), respectively. On average, the LLL algorithm has a slightly higher complexity order. However, since DSL channels are quasi-static, the complexity of initialization can be ignored for practical purposes.

During the symbol encoding process, the FEXT canceller used for all vectoring schemes operating in $\mathbb{C}^K$ is a linear matrix filter of size K, therefore it has a complexity order of $\mathcal{O}(K^2)$. For LR-ZFP and LR-THP using the *real-valued* LLL algorithm, the corresponding complexity order of the FEXT canceller is $\mathcal{O}[(2K)^2]$, which is comparable to $\mathcal{O}(K^2)$ for $K >> 2$. Furthermore, both schemes can employ the complex-valued LLL algorithm to fit into a canceller of size K. Secondly, the power controller is another matrix filter of identical size to that of the FEXT canceller, therefore it also incurs a complexity order of $\mathcal{O}(K^2)$. Finally, regarding the coolant of each nonlinear MUP, the decision feedback loop in THP has a complexity order of $\mathcal{O}(K^2)$. The similar loop in the LR-THP encoder has a complexity order of $\mathcal{O}[(2K)^2]$, which also becomes $\mathcal{O}(K^2)$ for $K >> 2$. Again, for LR-THP, the additional matrix filters presented during cooling have a complexity order of $\mathcal{O}[(2K)^2]$, whereas the coolant of LR-ZFP only has a complexity order of $\mathcal{O}[(2K)^2]$ in total.

For the VP-related schemes, it is widely known that all algorithms conceived for solving the CVP within a constant distance amount to a complexity order growing exponentially in $2K$ [188]. Thus the cooling stage of VP-related schemes incurs the highest complexity for $K >> 2$, which is significantly higher than that of the polynomial complexity order obeyed by all other schemes covered in this section. On the other hand, it is worth noting that the matrix filters are usually applied using parallel computing for improved hardware efficiency. Hence the practical complexity order of a matrix filter of size K becomes $\mathcal{O}(K)$. We summarize our findings in Table. 2.1.

Table 2.1: Complexity Order of the Vectoring Schemes Using Parallel Computing

	Coolant	Controller	Canceller
ZFP	N/A	$\mathcal{O}(K)$	$\mathcal{O}(K)$
THP	$\mathcal{O}(K^2)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$
LR-ZFP	$\mathcal{O}(K)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$
LR-THP	$\mathcal{O}(K)+\mathcal{O}(K^2)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$

An empirical complexity comparison between the classic DFSE and IDEALE is presented in Fig. 2.19 for VP-based cooling, where the transmit CSI is measured from a 20-pair 50 m DSL binder over the 212-424 MHz bandwidth, based on the additional future range under consideration for G.mgfast. The complexity measurement is based on the number of Euclidean distance calculations in both algorithms. Even though there is no specific limit on the maximum number of generations allowed during the sieving stage of IDEALE, the simulations characterized in Fig. 2.19 show that the complexity of IDEALE remains relatively constant in comparison to the DFSE. By contrast, the $\mathcal{K}$-best sphere encoder achieves a moderately improved success probability over IDEALE, as shown in Fig. 2.18, at the expense of an increased complexity order. However, owing to the sorting operations involved in the $\mathcal{K}$-best sphere encoder during each enumeration step, as well as to the higher level of flexibility of IDEALE in terms of the population size N_p and of the number of generations g, the latency incurred by IDEALE would still be lower when relying on parallel computing.

2.10 Chapter Summary

The (quasi-)static channel states of DSL binders are not sufficiently exploited even by the state-of-the-art non-linear MUD/MUP techniques conceived in the industry. Their slowly-varying propagation characteristics make it possible to incorporate the more advanced algorithms of Sec. 2.5, 2.6 and 2.7 having modest complexity during both initialization and run time. The ambition of closer-to-capacity operation inspires the conception of the optimal MUP techniques of Sec. 2.7 for practical DSL systems. However,

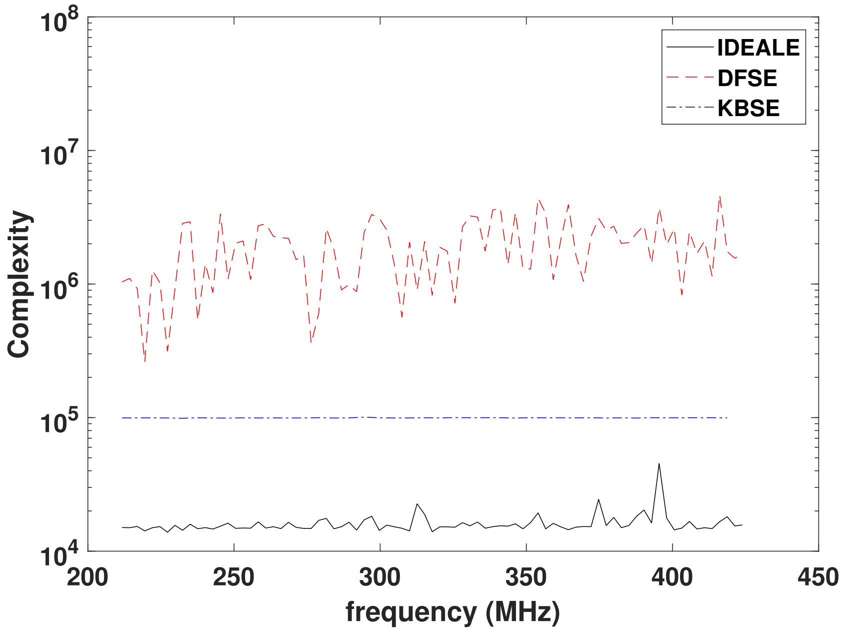


FIGURE 2.19: Complexity comparison between the DFSE, the $\mathcal{K}$-best variant (KBSE) and IDEALE by the number of Euclidean distance calculations, averaged over 500 simulations. The average-case complexity of IDEALE is less than 1% of the complexity of DFSE, while the complexity of KBSE given the success probability of Fig. 2.18 is around 10% of the complexity of DFSE.

their reliability is somewhat dependent upon the accuracy of transmit CSI acquisition at the VCE given the standard CSI feedback mechanism of Fig. 2.2 due to the changing physics associated with the excitation of adjacent twisted copper wires, which requires further investigations beyond the scope of this thesis.

In the comparative study of the state-of-the-art linear and non-linear MUPs of Sec. 2.3.1 and 2.3.2 in terms of their limitations, we focus our attention on exploiting the quasi-static nature of DSL channels in order to achieve improved performance by employing lattice reduction based preprocessing as demonstrated later in Sec. 2.5. Relying on the duality between broadcast and multiple access systems, as discussed in Sec. 2.2 and further demonstrated in Sec. 2.4.2, we critically appraised the MUD/MUP duality from a graphical and a lattice theoretical perspective. As a result, the concept of transmitter-side *vectoring mapping regions* as shown in Fig. 2.12 was proposed as the dual pair of the classic receiver-side *decision regions*. Using the proposed semi-analytical model, we focussed our attention on the family of MUPs approaching a vectoring mapping region in the shape of the associated lattice's Voronoi region in Sec. 2.7, which is theoretically optimal for any given channel matrix. However, the complexity associated with finding the Voronoi region increases exponentially with the scale (dimension) of the multi-user system. Consequently, we carried out the critical appraisal of the low-complexity sphere encoder variants and proposed the low-complexity IDEALE algorithm. The complexity comparison of the MUPs is summarized in Tab. 2.1 and Fig. 2.19.

Chapter 3

Spectrum Balancing for Vectored Networks

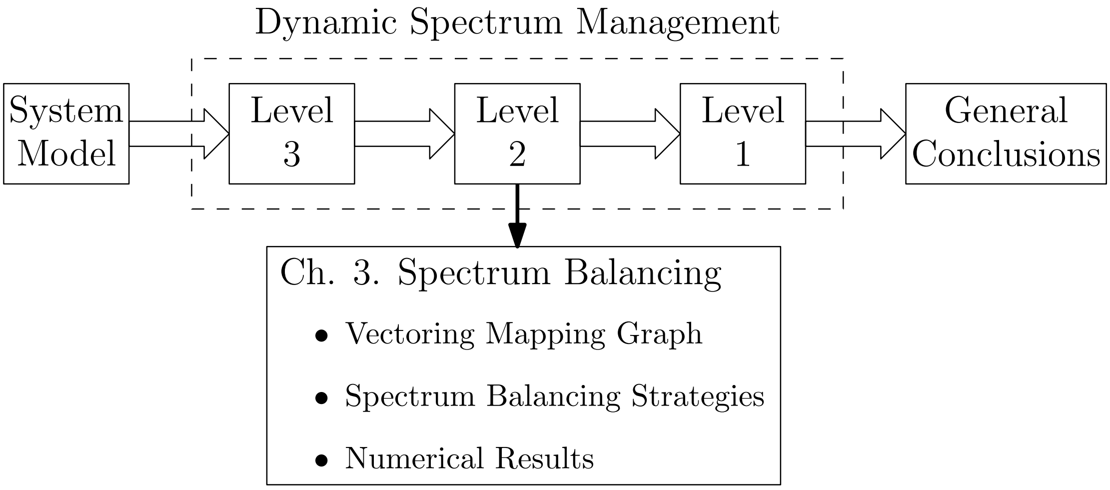


FIGURE 3.1: Outline of Ch. 3.

DSM is a prominent research subject in the history of wireline communications. Based on the historic definition of [71] (Table 1.1), past research of DSM was mainly concentrating on independent *bit-loading* based level-2, i.e. DSB, as shown in the landmark contributions of [209, 210, 211, 212, 150]. However, with the introduction of ITU-T G.993.5, MUP-based level-3 DSM became mandatory. As a consequence, a joint spectrum management strategy must be utilized, whose objective is to simultaneously optimize both the power and constellation assignment for maintaining the same bit error rate (BER) for all tones and all CPEs, while cancelling all known interferences as well as maximizing the sum rate of the vectored DSL binder. We note that depending on the applications, the BER requirement may rarely be unequal within a vectored group [213]. In general, DSB strategies relying on the proverbial 'water-filling' criterion are commonly employed to avoid the 'worst-case dominance' of plain FEXT cancellation dispensing with DSB.

In our context, the total transmit power is partitioned across two dimensions, namely the frequency and the signal space. On the other hand, the power constraints are valid for each individual pair of a binder, imposed on both a PSD and a total power basis. Despite the knowledge of the theoretical rate region for the majority of the MUP schemes determined under the usual sum-power constraint or per-antenna power contraint in wireless systems, e.g. [123][214], achieving the *optimal multi-level DSM* remains a critical research challenge in the developing of wireline access networks in the literature [132, 215, 216, 217]. This is particularly true for the hostile crosstalk-intensive environment, operating under the radical power constraints of both G.fast and the forthcoming G.mgfast. In this section, we will conduct an empirical case study on the performance of multi-level DSM, i.e. joint DSB-vectoring, employing both the conventional MUPs and the LRMUPs, in addition to a semi-analytical characterization of the achievable bounds.

Given the DMT modulated multi-user DSL system of (1.4), we have to find a set of $T \times K$ appropriate M-QAM constellations, which results in the maximum *sum rate* whilst meeting the BER target, which is formulated as:

$$\max \sum_{t=1}^{T} \sum_{k=1}^{K} b_k^t, \text{ where } b_k^t = \log_2(M_k^t) \text{ and } b_k^t \leq b_{\max}, \tag{3.1}$$

subject to the bit cap $b_{\max}$ and to the per-pair TxPSD mask $\mathbb{P}^t$ as well as to the per-pair ATP budget $\mathbb{A}$ [63]:

$$\max_{\boldsymbol{x}^t \in \mathcal{P}^t} |x_k^t|^2 \leq \mathbb{P}^t \qquad \forall t, k \tag{3.2}$$

$$\sum_{t=1}^{T} \max_{\boldsymbol{x}^t \in \mathcal{P}^t} |x_k^t|^2 \leq \mathbb{A} \qquad \forall k. \tag{3.3}$$

It is worth noting that (3.2) and (3.3) are *instantaneous power limits* that regulate the maximum TxPSD (known as the limiting PSD mask in [63]) and ATP occurrence. In contrast to the Shannon-Hartley theorem, the actual achievable rate under the constraints of both (3.2) and (3.3) is bounded away from the channel capacity. In this chapter, we will characterize the performance of the multi-pair copper binders relying on multi-tone modulation with the aid of our novel semi-analytical model, the *vectoring mapping graph*, against both the information-theoretical and the practical limits.

3.1 Vectoring Mapping Graph

For a vectoring scheme, its mapping region representation intuitively characterizes the most significant power transfer characteristics, such as the TxPSD exhibited by the transmitter. Recall in Section 2.3.2 that the output of a base-ϕ modulo operation $\Gamma_\phi[\boldsymbol{a}]$ (2.14) is distributed over the square region $\hat{\mathcal{U}} = (-\phi/2, \phi/2] + j(-\phi/2, \phi/2]$ in $\mathbb{R}^2$. For a square- or cross-shaped QAM constellation of the same modulo base, its convex hull has a marginally smaller area than $\hat{\mathcal{U}}$ especially when the constellation is large. Therefore it is beneficial to use the scaled continuous set $\mathcal{U} = \hat{\mathcal{U}}/\phi$ which approximates the power characteristics of both the output and the input of the modulo operator. In order to characterize the general behavior of square constellations, we now formalize the concept of the *uniformly distributed input* from (2.26) as follows:

Definition 3.1 (Uniformly Distributed Input)**.** The UDI refers to the *continuous* region $\mathcal{U}$ represented by the $2K$-dimensional unit hypercube centred at the origin obeying:

$$\mathcal{U} = \{\boldsymbol{u} : |u_k| < \frac{1}{2}, \quad k = 1, 2, ..., 2K\}, \tag{3.4}$$

which represents the union of K independent users' QAM constellations or the union of $2K$ independent users' pulse amplitude modulated (PAM) constellations.

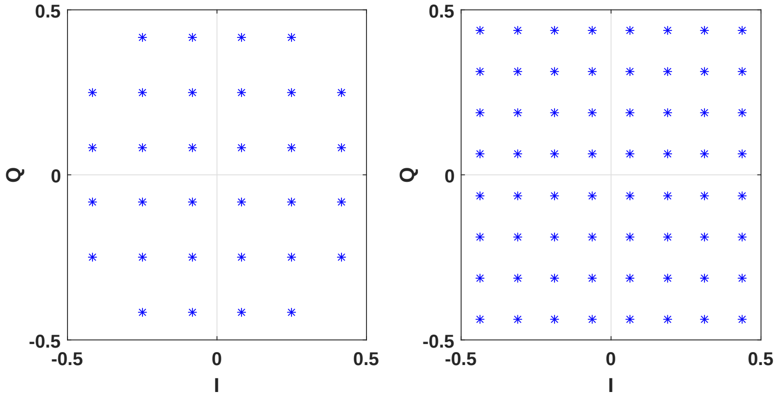


FIGURE 3.2: Uniformly Distributed Input (UDI) approximation of 32 QAM (left) and 64 QAM (right). Both the cross-shaped and the square-shaped constellations can be squeezed into the *unit square*. The average constellation energy increase due to the UDI approximation is uniquely determined by the QAM scheme itself. This is referred to as the precoding loss (cf. Sec. 2.3.2), which is caused by the transmit signal cooling during non-linear MUP.

As seen in [22], all the legitimate choices of the constellations of modulation order higher than $2^4 = 16$ follow the UDI approximation. Therefore we may utilize the UDI as a generalized symbol alphabet for any bit loading configuration. The power characteristics associated with the original constellations (with unit average energy) will be handled together with the desired power allocation by the power controller. In order to quantitatively characterize the performance of the vectoring schemes in conjunction with spectrum load balancing, we will first introduce the general concept of VMR, whose geometric properties assist the configuration of the power controller. With the aid of the VMR model, we then propose a joint vectoring and spectrum balancing algorithm, followed by the performance comparison of multi-level DSM arrangements relying on the MUP schemes of Ch. 2.

3.1.1 A Unified Perspective

Before discussing in detail the characteristics of a VMR model in the signal space and its applications in joint DSB and vectoring, we first have to justify, why VMR is an important model for the practical study of both vectoring and spectrum balancing. For the convenience of the reader, the system model conceived in Shannon's noisy channel coding theorem is reproduced in Fig. 3.3. Let us now consider the equivalence between Fig. 3.3 and all system models presented in Ch. 2 by introducing the (amplified) AWGN before the FEXT channel, as well as by considering the FEXT channel and the distributed CPEs as a single decoder. For a given channel $\boldsymbol{H}^t$ and its inverse $\boldsymbol{G}^t = (\boldsymbol{H}^t)^{-1}$, the equivalent decoder structure is identical for most of the non-linear MUPs. We will now discuss the upper bound on the achievable sum rate of these MUPs in more details in the context of FEXT channels.

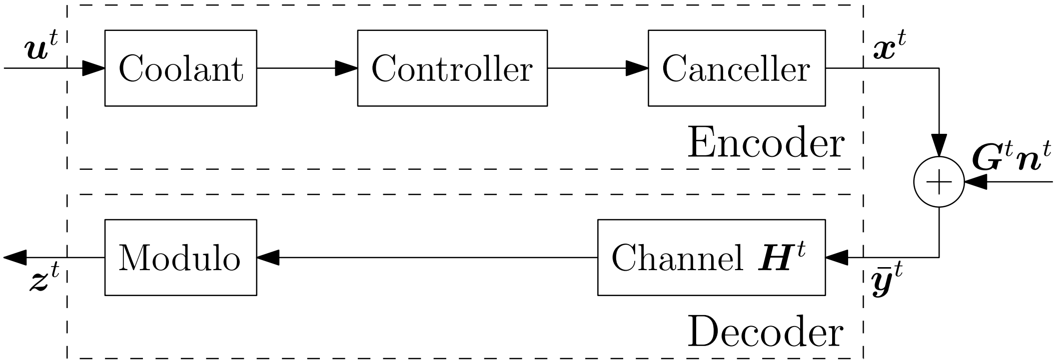


FIGURE 3.3: System model for the equivalent AWGN channel. The position of the FEXT channel $\boldsymbol{H}^t$ and the AWGN is swapped. Since $\boldsymbol{H}^t$ is time-invariant, we combine it with the deterministic receivers into a single equivalent decoder.

We will now quantify the information rate $I(\boldsymbol{u}^t, \boldsymbol{z}^t)$ supported by this unified equivalent system. Let $E\{\boldsymbol{n}^t(\boldsymbol{n}^t)^H\} = (\sigma_n^t)^2\boldsymbol{I}_K$ and $E\{\boldsymbol{u}^t(\boldsymbol{u}^t)^H\} = (\sigma_s^t)^2\boldsymbol{I}_K$. Following the data processing inequality of [146], we have:

$$\begin{aligned}
I(\boldsymbol{u}^t, \boldsymbol{z}^t) &\leq I(\boldsymbol{x}^t, \bar{\boldsymbol{y}}^t) \\
&= h(\bar{\boldsymbol{y}}^t) - h(\boldsymbol{n}^t) - \log_2|\det(\boldsymbol{G}^t)| \\
&\leq \log_2[\det(\pi e E\{\bar{\boldsymbol{y}}^t(\bar{\boldsymbol{y}}^t)^H\})] - \log_2[\det(\pi e(\sigma_n^t)^2\boldsymbol{I}_K)] - \log_2|\det(\boldsymbol{G}^t)| \\
&= \log_2\left[\det\left(\boldsymbol{I}_K + \frac{1}{K(\sigma_n^t)^2}\boldsymbol{H}^t E\{\boldsymbol{x}^t(\boldsymbol{x}^t)^H\}(\boldsymbol{H}^t)^H\right)\right],
\end{aligned} \tag{3.5}$$

where the last equality is the mutual information upper bound of the DMT channel $\boldsymbol{H}^t$ given $\boldsymbol{x}^t$. The first inequality indicates that $\bar{\boldsymbol{y}}^t$ uniquely and unambiguously determines $\boldsymbol{z}^t$, but not vice versa due to the modulo receivers, i.e. we have $h(\bar{\boldsymbol{y}}^t) = h(\boldsymbol{z}^t) + h(\bar{\boldsymbol{y}}^t|\boldsymbol{z}^t)$, where $h(\bar{\boldsymbol{y}}^t|\boldsymbol{z}^t)$ is termed as the modulo loss (cf. Sec. 2.3.2). Therefore the equality is only satisfied when the receivers are linear or when the non-linear encoder follows the MMSE criterion. In practice, it may also be ignored if the system's SNR is sufficiently high.

The second inequality in (3.5) follows the fact that the Gaussian distribution maximizes the (differential) entropy for a random variable having fixed mean and variance. More specifically, equality is only possible for the second inequality, when the input $\boldsymbol{x}^t$ follows the Gaussian distribution, which achieves the capacity of the inner AWGN channel in Fig. 3.3 dispensing with the equivalent codecs. From the entire system's perspective, $\boldsymbol{u}^t$ must also follow the Gaussian distribution in order to achieve the outer channel's capacity under the constraint $E\{\boldsymbol{u}^t(\boldsymbol{u}^t)^H\} = (\sigma_s^t)^2\boldsymbol{I}_K$. Conversely, the achievable information rate of (3.5) is bounded away from the channel's capacity due to the shaping loss associated with the suboptimal distributions of both $\boldsymbol{u}^t$ and $\boldsymbol{x}^t$. When the FEXT channel $\boldsymbol{H}^t$ is ill-conditioned, the shaping loss due to the suboptimal distribution of $\boldsymbol{x}^t$ becomes significant, which has not been fully accounted for in the similar previous analysis of [142, 145]. We should also note that (3.5) is investigated under the sum power constraint rather than under the realistic constraints under both (3.2) and (3.3). However, [132] showed that these constraints are convertible under the MAC-BC duality.

In the following sections, we will focus our attention on studying the distribution of $\boldsymbol{x}^t$, while dispensing with further discussions about the shaping loss due to $\boldsymbol{u}^t$, since the latter is a limitation inherent in all the ITU-T recommendations. However, it should be noted that this limitation has been addressed in both [142] and [206], where two different lattice coding techniques were proposed for recovering the shaping loss due to $\boldsymbol{u}^t$. By inference, the IFP scheme discussed in Sec. 2.6 could also be customized for achieving the same purpose, which requires joint signal processing at the receiver instead of relying on time-domain lattice coding. Nonetheless, these schemes only become capable of fully recovering the shaping loss due to $\boldsymbol{u}^t$ under the infinite codeword length idealization.

3.1.2 Vectoring Mapping Region

The geometric nature of the non-linear, modulo-based algorithms involved in lattice reduction allows for more rigorous yet intuitive characterization of the precoders' power transfer behaviour. By assuming the power controller to be an identity matrix $\boldsymbol{A} = \boldsymbol{I}$, it is plausible that ZFP maps a given vector $\boldsymbol{u}$ of UDI symbols, as a point in $\mathbb{R}^{2K}$, onto an origin-centred *fundamental* parallelotope *mapping region* $\boldsymbol{G}\mathcal{U}$[1], which is a translation of the basis of the lattice $\mathcal{L}(\boldsymbol{G})$. Since the distribution region of $\tilde{\boldsymbol{u}}$ and that of $\boldsymbol{u}$ are identical, the fundamental mapping region of the THP scheme is geometrically the orthotope $\boldsymbol{Q}\boldsymbol{D}^{-1}\mathcal{U}$ obtained by a shear transformation of the parallelotope $\boldsymbol{G}\mathcal{U}$. With any other designated choice of the power controller $\boldsymbol{A}$, the *vectoring mapping region* illustrates the specific distribution of the precoded symbol vector $\boldsymbol{x}$ in the signal space. The underlying transformation of any given vectoring scheme corresponds to mapping the elements of the UDI set $\mathcal{U}$ onto the vectoring/fundamental mapping region, usually by (implicitly) invoking a *mapping matrix*. A mapping matrix exists if the corresponding vectoring

---

[1] We use the set $\mathcal{U}$ to interchangeably denote its convex hull in this thesis.

mapping region is a parallelotope, which includes the special case of an orthotope. The power transfer characteristics exhibited by vectoring are worth investigating in consideration of the TxPSD and ATP constraints originally conceived for the DSL standards. Let the VMR be the (*origin-centred*) K-dimensional polytope $\mathcal{P}$. Then we can list some of these relevant geometric properties as follows.

3.1.2.1 Second Central Moment

The (normalized) second central moment of an origin-centred K-dimensional polytope $\mathcal{P}$ is given by:

$$\sigma_{\mathcal{P}}^2 = \int_{\mathcal{P}} \frac{\|\boldsymbol{x}\|^2}{V(\mathcal{P})}\, d\boldsymbol{x} = \int_{\mathcal{P}} \sum_{k=1}^{K} \frac{|x_k|^2}{V(\mathcal{P})}\, d\boldsymbol{x}, \tag{3.6}$$

where $\sigma_{\mathcal{P}}^2$ represents the total average power $E\{\|\boldsymbol{x}\|^2\}$ of the signal $\boldsymbol{x}$. When the power constraints are imposed on each individual element x_k, it is more insightful to consider the following relative of $\sigma_{\mathcal{P}}^2$, which we define as:

$$\xi_{\mathcal{P}} = \int_{\mathcal{P}} \max_{1\le k\le K} \frac{|x_k|^2}{V(\mathcal{P})}\, d\boldsymbol{x}. \tag{3.7}$$

It is apparent that $\xi_{\mathcal{P}}$ is the average peak power per pair of the binder. Calculating either $\sigma_{\mathcal{P}}^2$ or $\xi_{\mathcal{P}}$ for a general convex polytope $\mathcal{P}$ remains an open problem. However, for well-behaved shapes such as parallelotopes, $\sigma_{\mathcal{P}}^2$ can be readily obtained, which is most commonly used as a performance measure for systems employing the scalar power allocation policy. On the other hand, for the allocation policy optimized for individual pairs, we seek to find the second moment per (complex-valued) dimension of $\mathcal{P}$, which is apparently related to both the geometry and the positioning of $\mathcal{P}$. By definition, the Voronoi cell is the optimal achievable VMR satisfying the nearest Euclidean neighbour criterion (2.24). The optimality of any other VMR is then decided by the volume ratio of its intersection with the Voronoi cell to its remainder. In particular, this ratio may be matched to the strict definition of the success probability in Sec. 2.7.3 under a proximity factor of 1.

3.1.2.2 Maximum Euclidean Spread

The TxPSD mask and ATP limit in [63] are proposed with respect to the maximum TxPSD per-tone-per-pair and the maximum power per-pair. The absolute power limit (rather than the average power) per-pair exhibited by the encoded signal $\boldsymbol{x}$ is characterized by a set of K scalars $\boldsymbol{v} = [v_1, v_2, ..., v_K]$, which we refer to as the maximum Euclidean spread (MES) of $\mathcal{P}$. The kth element v_k of the MES vector $\boldsymbol{v}$ is given by:

$$v_k = \max_{\boldsymbol{x}\in\mathcal{P}} \|\boldsymbol{x}\cos\theta_k\| \quad (k = 1, 2, ..., K), \tag{3.8}$$

where θ_k refers to the angle between $\boldsymbol{x}$ and the 2D Cartesian plane spawned by the kth pair in the signal space. It may be readily recognized that $\mathcal{P}$ actually belongs to $\mathbb{R}^{2K}$ and each pair (that transmits QAM symbols) occupies two of the $2K$ real dimensions. The two real dimensions constitute a 2D Cartesian plane. If we denote the projection region of $\mathcal{P}$ onto the Cartesian plane spawned by the kth pair as $J_{\mathcal{P}|k}$, then (3.8) may be intuitively interpreted as the length of the longest vector in $J_{\mathcal{P}|k}$. The elements of $\boldsymbol{v}$ are referred to as the MES per (complex) dimension. We note that if a PAM-based system is considered, then the MES per real dimension is the projection of $\mathcal{P}$ onto each Cartesian axis. Fig. 3.4 demonstrates the difference between the two cases.

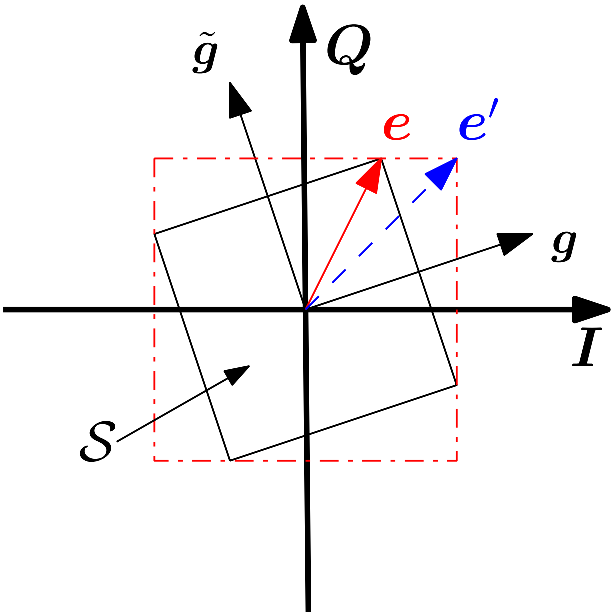


FIGURE 3.4: The Maximum Euclidean Spread of a system $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{u}$ in $\mathbb{R}^2$ decoupled from the equivalent system in $\mathbb{C}$, where we have $\boldsymbol{G} = [\boldsymbol{g}, \tilde{\boldsymbol{g}}]$, $\|\boldsymbol{g}\| = \|\tilde{\boldsymbol{g}}\|$ and $\boldsymbol{g} \perp \tilde{\boldsymbol{g}}$. Mapping the UDI $\mathcal{U}$ onto $\boldsymbol{G}$ results in the solid square fundamental mapping region $\mathcal{S}$. The minimum bounding box of the fundamental mapping region is the dashed square, which shows the MES of $\mathcal{S}$ in $\mathbb{R}^2$. However, it is apparent that $\|\boldsymbol{e}\| = \max_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{x}\|$ and therefore $\|\boldsymbol{e}\|$ should be the MES of $\mathcal{S}$ in $\mathbb{C}^1$.

Example. Let us now consider a two-user system in more details, where both users transmit a 1D signal, i.e. PAM symbols. Observe that the MES per real dimension is characterized by a bounding box (e.g. the dot dashed boundary of Fig. 3.4) confining the VMR $\mathcal{S}$. The minimum-sized bounding box of the parallelogram $\mathcal{S}$ contains four vertices of $\mathcal{S}$. By exploiting the symmetry of parallelograms, we only have to find two of the four vertices that are located the farthest along the Cartesian axis. The symmetry of a parallelogram can be readily extended to that of a $2K$-dimensional parallelotope. For a $2K$-dimensional parallelotope, we have to find exactly $2K$ vertices out of all 2^{2K} vertices of the parallelotope.

The analysis of the above example of MES in $\mathbb{R}^{2K}$ may be extended to the vectoring mapping region $\mathcal{P}$ of a QAM-based system in $\mathbb{C}^K$. The MES of $\mathcal{P}$ in the kth Cartesian plane, namely v_k, is upper bounded by the vector sum of $\mathcal{P}$ on the two Cartesian axis of the kth plane (as demonstrated by $\boldsymbol{e}$ and $\boldsymbol{e}'$ of Fig. 3.4). Due to the symmetry of $\mathcal{P}$, the K vertices corresponding to the K MES values v_k constitute a subset of the $2K$ vertices corresponding to the MES per real dimension. The MES $\boldsymbol{v}$ can be found using Alg. 2.

Line 3 of Alg. 2 may be ignored, if the input is a real-valued matrix that had already been decoupled. Line 10 computes the vector sum and finds the farthest vertex $\boldsymbol{q}_m$ of $\mathcal{P}$ in the real dimension m. Line 11-14 calculates and compares the projection of $\boldsymbol{q}_m$ in the given Cartesian plane. Since Alg. 2 calculates the MES of a parallelotope, it can be used for determining the maximum amplitude of the transmitted symbols and therefore also the peak TxPSD per pair.

```
Input: An origin-centred $N$-dimensional parallelotope $\mathcal{P}$ whose edges are given by
  the columns of the $N \times N$ complex-valued matrix $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, ..., \boldsymbol{g}_N]$;
Output: $N$-vector $\boldsymbol{v} = [v_1, v_2, ..., v_N]$ indicating the MES per complex dimension;
$\bar{\boldsymbol{G}} = \{[\Re(\boldsymbol{G})^T, \Im(\boldsymbol{G})^T]^T, [-\Im(\boldsymbol{G})^T, \Re(\boldsymbol{G})^T]^T\}$;
for $m = 1, 2, ..., 2N$ do
    for $n = 1, 2, ..., 2N$ do
        if $\bar{\boldsymbol{g}}_n(m) < 0$ then
            $\boldsymbol{s}_n \leftarrow -\bar{\boldsymbol{g}}_n/2$;
        else
            $\boldsymbol{s}_n \leftarrow \bar{\boldsymbol{g}}_n/2$;
        end
    end
    $\boldsymbol{q}_m \leftarrow \sum_{n=1}^{2N} \boldsymbol{s}_n$ ;
    if $m \leq N$ then
        $v_m \leftarrow \sqrt{\boldsymbol{q}_m^2(m) + \boldsymbol{q}_m^2(m+N)}$
    else if $v_{m-N} < \sqrt{\boldsymbol{q}_m^2(m) + \boldsymbol{q}_m^2(m-N)}$ then
        $v_{m-N} \leftarrow \sqrt{\boldsymbol{q}_m^2(m) + \boldsymbol{q}_m^2(m-N)}$
    end
end
```

Algorithm 2: Maximum Euclidean Spread per Complex Dimension

3.1.3 Babai's Approximations

Let us consider the complex-valued general system model of (1.5) as a real-valued system with respect to a single discrete tone. We will characterize the operations of the nearest-plane algorithm in details, followed by the simpler rounding-off algorithm. Based on the transmitter structure of Fig. 3.5, we employ the following triangular decomposition of the matrix $\boldsymbol{GA}$ similar to that of (2.35):

$$\boldsymbol{GA} = \boldsymbol{WBZ}^{-1}, \tag{3.9}$$

where $\boldsymbol{W}$ has orthogonal columns and $\boldsymbol{B}$ is an upper triangular matrix with $\text{diag}(\boldsymbol{B}) = \text{diag}(\boldsymbol{I})$. The integer-valued unimodular matrix $\boldsymbol{Z}$ transforms the given lattice basis $\boldsymbol{GA}$ to the LLL-reduced basis $\boldsymbol{WB}$. Following the decomposition of Eq. (3.9), the nearest-plane solution $\boldsymbol{l}_{\text{np}}$ of the optimization problem of (2.25) is given by:

$$\boldsymbol{l}_{\text{np}} = -\boldsymbol{Z}\tilde{\boldsymbol{l}}, \text{ where } \tilde{\boldsymbol{l}} = \lceil \boldsymbol{W}^{-1}\boldsymbol{GAu} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}} \rfloor, \tag{3.10}$$

where $\lceil\cdot\rfloor$ denotes the element-wise rounding-off operation to the nearest integer. The symbol vector $\boldsymbol{x}_{\text{NP}}$ produced by the nearest-plane cooling algorithm is formulated as:

$$\begin{aligned}\boldsymbol{x}_{\text{NP}} &= \boldsymbol{GA}(\boldsymbol{u} + \boldsymbol{l}_{\text{np}}) \\ &= \boldsymbol{GA}(\boldsymbol{u} - \boldsymbol{Z}\lceil \boldsymbol{W}^{-1}\boldsymbol{GAu} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}}\rfloor) \\ &= \boldsymbol{W}(\boldsymbol{m} - \boldsymbol{B}\lceil \boldsymbol{m} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}}\rfloor) \\ &= \boldsymbol{W}\hat{\boldsymbol{u}}, \end{aligned} \tag{3.11}$$

where we have $\boldsymbol{m} = \boldsymbol{BZ}^{-1}\boldsymbol{u}$. Since $\boldsymbol{B}$ is an upper triangular matrix with diagonal entries of $b_{kk} = 1 \ \forall k$, the feedback encoding operation of Fig. 3.5 is carried out in reverse order, commencing from $k = 2K$. In this case $\hat{u}_k$ is given by:

$$\begin{aligned}\hat{u}_k &= m_k - \sum_{j=k}^{2K} b_{kj}\lceil m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j\rfloor \\ &= m_k - (\sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j + \lceil m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j\rfloor) \\ &= (m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j) - \lceil m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j\rfloor \\ &= \Gamma[m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j]. \end{aligned} \tag{3.12}$$

The last equality results from the fact that the difference between any real number and its nearest integer belongs to the UDI interval of $(-1/2, 1/2]$. As a consequence, $\hat{\boldsymbol{u}}$ and the message symbol vector $\boldsymbol{u}$ are both distributed over the UDI region $\mathcal{U}$. Therefore, according to (3.11) and (3.12), the mapping matrix of the nearest-plane scheme is $\boldsymbol{W}$ and its corresponding VMR is characterized by the set $\boldsymbol{W}\mathcal{U}$. Similarly, the decomposition $\boldsymbol{GA} = \boldsymbol{FZ}^{-1}$ performed in the rounding-off scheme results in the VMR $\boldsymbol{F}\mathcal{U}$. Let the power controller be $\boldsymbol{A} = \boldsymbol{I}$ and the inverse channel $\boldsymbol{G}$ be LLL-reducible, i.e. $\boldsymbol{Z} \neq \boldsymbol{I}$. It is now trivial to show that the rounding-off and the nearest-plane schemes are characterized by the same VMRs of LRZFP and LRTHP as shown in Fig. 2.12, respectively. Since both approximations are based on the system design of VP, we will term them as the approximate perturbation aided lattice encoders [96] (APPLE).

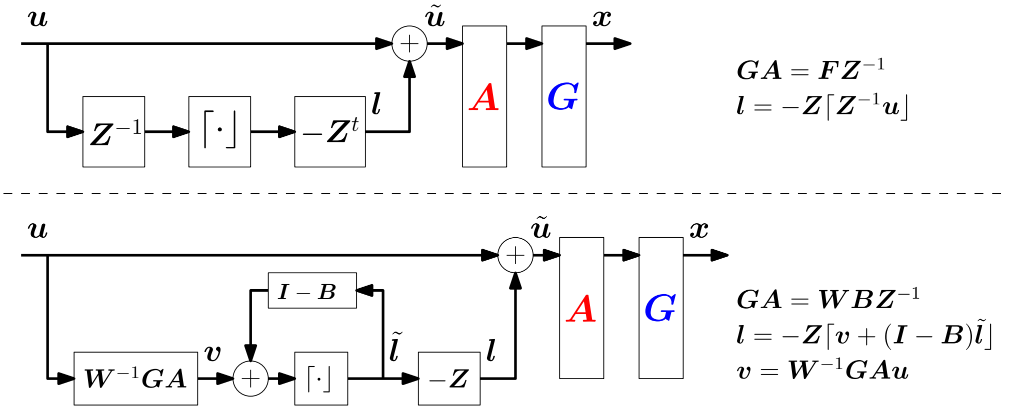


FIGURE 3.5: Approximate perturbation aided lattice encoders [96] relying on Babai's rounding-off (top) and nearest-plane (bottom) methods, respectively.

3.2 Transmit Output Analysis

Due to the paucity of specific studies on applying the ACM protocol over LRMUPs, employing lattice reduction aided multi-level DSM requires further theoretical clarifications. In this section, we characterize both the LRMUP's power controller configuration and the MUP's general I/O relationships. Despite the fact that the characterization here is primarily based on the conventional information-theoretical formulation, the specific scenario obeying the instantaneous power constraints of both (3.2) and (3.3) may also be characterized by the VMR model of Sec. 3.1.

3.2.1 Power Controller Configuration

Let the *real-valued non-negative* diagonal matrix $\boldsymbol{P}^t = \text{diag}\{P_1^t, P_2^t, \cdots, P_K^t\}$ determine the power assigned to tone t. Then the corresponding amplitude scaling matrix is $\boldsymbol{A}^t = \text{diag}\{\sqrt{P_1^t}, \sqrt{P_2^t}, \cdots, \sqrt{P_K^t}\}$. Since diagonal matrices are generally not commutative in multiplication, we cannot apply $\boldsymbol{P}^t$ arbitrarily within the DP. For ZFP and the THP, the equalized symbol vectors are formulated as:

$$\boldsymbol{z}^t = \boldsymbol{A}^t\boldsymbol{u}^t + \boldsymbol{n}^t \text{ for ZFP,} \tag{3.13}$$

$$\boldsymbol{z}^t = \boldsymbol{A}^t\boldsymbol{u}^t + \Gamma_{\phi^t\boldsymbol{A}^t}\left[(\boldsymbol{\Sigma}^t)^{-1}\boldsymbol{n}^t\right] \text{ for THP,} \tag{3.14}$$

where $\boldsymbol{\Sigma}^t$ follows the same definition of (2.11) for tone t. Since we have $E\{|u_k^t|^2\} = 1$, the detection SNR experienced by the QAM demapper can be expressed as:

$$\text{SNR}_{k,\text{ZFP}}^t = \frac{P_k^t}{E\{|n_k^t|^2\}}, \tag{3.15}$$

$$\text{SNR}_{k,\text{THP}}^t = \frac{P_k^t|r_{k,k}^t|^2}{E\{|\bar{n}_k^t|^2\}}, \tag{3.16}$$

where $\bar{n}_k^t$ represents the AWGN term which corresponds to n_k^t enhanced by the modulo operator on the right side of Fig. 2.7. This is exactly the 'modulo loss' presented in Section 2.3.2.3. Near the typical operating point at the BER target, of say 10^{-7}, the detection SNR is sufficiently high for ensuring that the noise-enhancement effect of the modulo receiver becomes negligible. Therefore we may assume that $\bar{n}_k^t$ follows the Gaussian distribution, and consequently, the SNR gap of (3.33) is applicable both to the ZFP based linear vectoring, as well as to all other modulo receiver based vectoring schemes. However, for low modulation orders, the modulo loss is not negligible and the exact value may be determined via simulation, as done in [218].

Generally, using a common modulo base ϕ is mandatory, if we apply the family of lattice reduction aided precoding schemes to a vector of symbols drawn from different constellations. The enumeration process of VP discussed in Section 2.7 also requires a common modulo base for all message symbols [219]. Therefore, the power allocation conceived for lattice reduction aided precoding has to be applied after the modulo operations. Thus, we can define the power allocation schemes for the LRMUPs as follows:

$$\boldsymbol{x}^t = \boldsymbol{F}^t\boldsymbol{A}^t\Gamma_{\phi^t}\left[(\boldsymbol{Z}^t)^{-1}\boldsymbol{u}^t\right] \text{ for LR-ZFP,} \tag{3.17}$$

$$\boldsymbol{x}^t = (\boldsymbol{S}^t)^{-H}\boldsymbol{A}^t\,(\boldsymbol{x}^t)' \text{ for LR-THP,} \tag{3.18}$$

$$\boldsymbol{x}^t = \boldsymbol{G}^t\boldsymbol{A}^t(\boldsymbol{u}^t + \boldsymbol{l}^t) \text{ for VP,} \tag{3.19}$$

where

$$(\boldsymbol{x}^t)' = \Gamma_{\phi^t}\left[\Gamma_{\phi^t}\left[(\boldsymbol{Z}^t)^H\boldsymbol{u}^t\right] + (\boldsymbol{I} - (\boldsymbol{B}^t)^H)(\boldsymbol{x}^t)'\right].$$

However, if $\boldsymbol{A}^t$ is a non-scalar matrix, then the operations in (3.17) and (3.18) no longer fulfil the zero forcing signal reception criterion. Therefore we have to use a scalar matrix $\boldsymbol{A}^t$ for these approximate lattice precoding schemes, which corresponds to SSB.

On the other hand, for sphere encoder based VP, the power assignment policy of (3.19) affects the choice of the coolant $\boldsymbol{l}^t$. If the scaling matrix $\boldsymbol{A}^t$ of (3.19) is non-scalar, then the optimization of (2.25) may no longer necessarily produce the optimal coolant with respect to the power-controlled encoder of (3.19). The main reason is that the two lattices $\mathcal{L}(\boldsymbol{G}^t\boldsymbol{A}^t)$ and $\mathcal{L}(\boldsymbol{G}^t)$ are not normally isomorphic. More specifically, their Voronoi cells are only congruent if $\boldsymbol{A}^t$ is a scalar matrix. For the sake of analytical tractability, we shall restrict $\boldsymbol{A}^t$ to be a scalar matrix for sphere encoder based VP, in which case $\mathcal{L}(\boldsymbol{G}^t\boldsymbol{A}^t)$ and $\mathcal{L}(\boldsymbol{G}^t)$ are isomorphic lattices. However, when employing the rounding-off or the nearest plane approximations in VP, the corresponding VMRs have been shown to be parallelotopes which have trivial characterization. Therefore, the pair of Babai's approximations can circumvent the power controller's limitation to SSB policies. Nonetheless, the equalized symbol vector has the same common form for the lattice-based MUPs, expressed as:

$$\boldsymbol{z}^t = \boldsymbol{A}^t \boldsymbol{u}^t + \Gamma_{\phi^t \boldsymbol{A}^t} \left[\boldsymbol{n}^t \right], \tag{3.20}$$

which results in the detection SNR formulated as:

$$\mathrm{SNR}^t_{k,\mathrm{LR}} = \frac{P^t_k}{E\{|\bar{n}^t_k|^2\}}. \tag{3.21}$$

3.2.2 TxPSD Characterization

Since both (3.2) and (3.3) are defined on a per-pair basis, we have to evaluate TxPSD for each active pair, given the power assignment policy and the precoder. Based on [218] and the general assumption where $E\{|u^t_k|^2\} = 1$, in ZFP and the THP, $E\{|x^t_k|^2\}$ is related to $\boldsymbol{P}^t$ by:

$$E\{|x^t_k|^2\}|_{\mathrm{ZFP}} = \sum_{j=1}^{K} |g^t_{k,j}|^2 P^t_j \qquad \text{with } g^t_{k,j} \text{ drawn from } \boldsymbol{G}^t, \tag{3.22}$$

$$E\{|x^t_k|^2\}|_{\mathrm{THP}} = \sum_{j=1}^{K} |Q^t_{k,j}|^2 \rho^t_j P^t_j \qquad \text{with } Q^t_{k,j} \text{ drawn from } \boldsymbol{Q}^t. \tag{3.23}$$

The modulo operator found in the THP encoder on the left side of Fig. 2.7 causes a slight increase of the average constellation energy characterized by ρ^t_j, whose value is uniquely determined by the constellation b^t_j. We have seen in Section 2.3.2.3 that this is the 'precoding loss'. For square M^t_j-QAM constellations, $\rho^t_j = M/(M-1)$ converges quickly to one upon increasing $b^t_j = \log_2 M^t_j$ and can be safely ignored for large b^t_j values. This is exactly the precoding loss previously defined in Section 2.3.2.3.

The output of both the LR-ZFP-based and of the LR-THP-based transmitters can be characterized in a format similar to (3.23) based on the power assignment policy of (3.17) and (3.18). The pair of approximate lattice precoding schemes follows the same encoding steps constituted by the modulo operations, power allocation and linear filtering, as implied by Eq. (3.17) and (3.18). Since the result of modulo operation exhibits the same PSD in the approximate lattice precoding schemes as well as in the THP, we can characterize the TxPSD of the LR-ZFP and the LR-THP as follows:

$$E\{|x^t_k|^2\}|_{\mathrm{LRZFP}} = \sum_{j=1}^{K} |f^t_{k,j}|^2 \rho^t P^t \text{ with } f^t_{k,j} \text{ drawn from } \boldsymbol{F}^t, \tag{3.24}$$

$$E\{|x^t_k|^2\}|_{\mathrm{LRTHP}} = \sum_{j=1}^{K} |s^t_{k,j}|^2 \rho^t P^t \text{ with } s^t_{k,j} \text{ drawn from } (\boldsymbol{S}^t)^{-H}. \tag{3.25}$$

Because the choice of constellation is shared by all pairs on the same tone for the LR-

ZFP and the LR-THP, we emphasize this fact by neglecting the pair index j for the terms ρ^t and P^t in Eq. (3.24) and (3.25). These average TxPSD definitions trivially follow their respective VMR characterizations in addition to the fundamental loss (cf. Sec. 2.3.2) associated with non-linear MUPs in general. On the other hand, calculating the instantaneous TxPSD per tone per pair is more straightforward, since only the MES of the corresponding VMR has to be calculated.

3.2.3 A Fundamental Limit

By contrasting (3.19) to both (3.17) and (3.18), it may be readily seen that the VP-based encoder does not share the aforementioned power transfer characteristics of the other modulo encoders. Our previous investigations have established that $\boldsymbol{x}^t$ is distributed over the origin-centred Voronoi cell of the lattice $\mathcal{L}(\boldsymbol{G}^t)$ or stretched lattice $\mathcal{L}(\boldsymbol{G}^t\boldsymbol{A}^t)$. By definition (2.24), the Voronoi cell has the lowest normalized second central moment among all tessellating shapes (e.g. VMRs) associated with the same lattice. However, the computation of Voronoi cells in high dimensional Euclidean spaces requires algorithms having prohibitively high complexity, such as those of [220] or [221]. This dilemma naturally raises the question of how close we may approximate the Voronoi cell and how high the maximum achievable shaping gain may be attained by optimal DSM.

Ryan *et al.* had shown [163] that for lattices spawned by the inverse of wireless channels, the geometric properties of the resultant Voronoi cell resemble those of a hypersphere. In particular, it is demonstrated in [163] that the second moment of a Voronoi cell is closely lower bounded by that of the hypersphere having the same dimension and volume. The second moment of the Voronoi cell of $\mathcal{L}(\boldsymbol{G}^t)$ represents the average total TxPSD of the VP-based encoder having the input alphabet of $\mathcal{U}/\phi$ and no amplitude scaling, which was lower-bounded in [163] as follows:

$$E\{\|\boldsymbol{G}^t(\boldsymbol{u}^t+\boldsymbol{l}^t)\|^2\} \geq \frac{K\Gamma(K+1)^{1/K}}{(K+1)\pi}\det[(\boldsymbol{H}^t)^H\boldsymbol{H}^t]^{-1/K}. \tag{3.26}$$

We extend this result by hypothesizing that the mapping region of VP *is* a hypersphere[2], in which case the equality holds in (3.26). Since the hypersphere is perfectly isotropic, each VP-encoded symbol x_k^t of the same tone t will have an identical share of the total TxPSD $E\{\|\boldsymbol{x}^t\|^2\}|_{\mathrm{VP}}$. Given that $\boldsymbol{A}^t$ is a scalar matrix in (3.19), the per-pair TxPSD of the VP encoded symbol vector may be characterized by:

$$E\{|x_k^t|^2\}|_{\mathrm{VP}} = \frac{E\{\|\boldsymbol{x}^t\|^2\}|_{\mathrm{VP}}}{K}P^t. \tag{3.27}$$

---

[2] Since the hypersphere does not tessellate the Euclidean space $\mathbb{R}^{2K}$, a Voronoi cell cannot be an exact hypersphere. However, using this approximation will allow us to investigate the absolute upper bound, albeit unachievable.

3.3 Static Spectrum Balancing

Let the scalar power allocation policy be given by $\boldsymbol{P}^t = \gamma^t \boldsymbol{I}$, in which case the scaling matrix is $\boldsymbol{A}^t = \boldsymbol{I}\sqrt{\gamma^t}$. In general, if we denote the VMR of the nearest plane based VP for tone t as $\mathcal{P}^t$, we may find that $E\{\mathbb{P}^t/\gamma^t\} = \xi_{\mathcal{P}^t}$ based on the definition of (3.7). Given the orthotope shape of $\mathcal{P}^t$ based on similar derivation to those of (3.11) and (3.12), as well as the improved orthogonality of reduced lattice basis, we may characterize the average achievable SNR of each tone and therefore the SER performance of APPLE.

Lemma 3.2. *For square QAM (i.e. 2^b-QAM for $b = 2, 4, 6, ...$) constellations, the average SER ζ^t_{NP} of tone t achieved by the nearest plane algorithm under SSB is lower bounded by:*

$$\zeta^t_{NP} \geq 1 - \left[1 - erfc\left(\sqrt{\frac{2\mathbb{P}^t}{\sigma_n^2 \det[\boldsymbol{H}^t(\boldsymbol{H}^t)^H]^{-1/K}}}\right)\right]^2, \tag{3.28}$$

where $erfc(\cdot)$ represents the complementary error function. Furthermore, $\boldsymbol{H}^t$ is the channel matrix of tone t and σ_n^2 is the noise variance, as in (1.4).

The SNR of each equalized symbol z^t_k of the same tone t is characterized by γ^t/σ_n^2 based on the scalar power policy we employed, which is dependent on the second moment of the nearest plane algorithm's VMR according to the relationship between (3.6) and (3.7). The second moment of a $2K$-dimensional parallelotope is known to be lowered bounded by that of a hypercube of the same volume and dimension. Based on the definition in [158], the second moment of a $2K$-dimensional hypercube $\mathcal{C}$ having volume $V(\mathcal{C}) = V(\mathcal{P}^t) = \det[(\boldsymbol{G}^t)^H\boldsymbol{G}^t] = \det[\boldsymbol{H}^t(\boldsymbol{H}^t)^H]^{-1}$ is formulated as:

$$\sigma^2_{\mathcal{C}} = dim(\mathcal{C})G(\mathcal{C})V(\mathcal{C})^{2/dim(\mathcal{C})} = \frac{K\det[\boldsymbol{H}^t(\boldsymbol{H}^t)^H]^{-1/K}}{6}, \tag{3.29}$$

where $dim(\mathcal{C}) = 2K$ and the dimensionless second moment of a hypercube is $G(\mathcal{C}) = 1/12$. Upon revisiting the difference between (3.6) and (3.7), we may identify that $K\xi_{\mathcal{P}^t}/\sigma^2_{\mathcal{P}^t}$ actually represents the PAPR of the per-pair TxPSD. Since the transmitted signal $\boldsymbol{x}^t$ is assumed to be distributed over the hypercube $\mathcal{C}$, we may further hypothesis that $\mathcal{C}$ is positioned in such a way that its 2D facets are parallel to the Cartesian planes. In this case, each pair achieves a common minimum PAPR, which is given by that of the $\mathcal{U}$-constellation of (3.4), i.e. $(1/2)^2/(1/12) = 3$. Therefore the following holds regarding the VMR of the nearest plane algorithm $\mathcal{P}^t$:

$$\xi_{\mathcal{P}^t} \geq \frac{3\sigma^2_{\mathcal{P}^t}}{K} \geq \frac{3\sigma^2_{\mathcal{C}}}{K} = \frac{\det[\boldsymbol{H}^t(\boldsymbol{H}^t)^H]^{-1/K}}{2}. \tag{3.30}$$

Hence the following SNR bound exists for the nearest plane algorithm independent from the optimality of the employed lattice reduction criterion:

$$E\{\gamma^t/\sigma_n^2\} \leq \frac{2\mathbb{P}^t}{\sigma_n^2 \det[\boldsymbol{H}^t(\boldsymbol{H}^t)^H]^{-1/K}}. \tag{3.31}$$

As a consequence of using the additive coolant $\boldsymbol{l}_{\text{np}}^t$ of (3.10), the mixture $\tilde{\boldsymbol{u}}^t$ of Fig. 3.5 for a given tone t (rather than the $\hat{\boldsymbol{u}}^t$ of (3.11)) is drawn from an infinite-sized periodically-expanded constellation. Thus for each user, $\tilde{u}_k^t$ is distributed over the infinite-sized QAM constellation (Fig. 1.11). For infinite-QAM, every point on the constellation diagram has exactly four nearest neighbours of identical distance. Hence the exact SER performance of the nearest plane algorithm is identical to that of an infinite-QAM constellation transmitted over an AWGN channel, and the proof of Lemma 3.2 is completed.

As demonstrated in Fig. 3.2, both even-bit square-QAM and odd-bit cross-QAM constellations may be approximated by the UDI. However, due to the 'missing' corner points of a cross-QAM constellation, some points of the infinitely-spliced cross-QAM will no longer have four equi-distance neighbours, since there will be 'holes' in the constellation. This typically results in an overestimation of the encoded PSD $\xi_{\mathcal{P}^t}$ and therefore an overestimation of SER. A study of the exact effect of the missing points as well as the error rate performance of the cross-QAM in linear environments may be found in [222]. We will now consider the performance of the nearest plane *approximate perturbation aided lattice encoder* (APPLE) in comparison to other known algorithms.

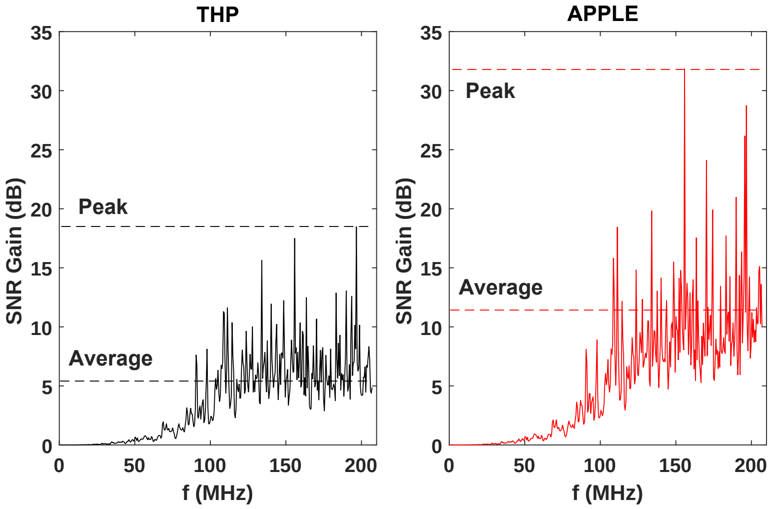


FIGURE 3.6: SNR gain of THP and that of APPLE in comparison to the SNR penalty of ZFP.

Given the mapping matrix $\boldsymbol{S}^t$, the exact second moment of the corresponding parallelotope is given by the trace function, $\sigma_{\mathcal{P}^t}^2 = tr[(\boldsymbol{S}^t)^H \boldsymbol{S}^t]$. We may apply similar analysis to the geometry of ZFP and THP for finding the relative SNR and SER performance of each vectoring scheme. For simplicity, we assume the minimum PAR $K\xi_{\mathcal{P}^t}/\sigma_{\mathcal{P}^t}^2 = 3$ for ZFP and THP as well, even though the actual PAR is higher in both cases. In Fig. 3.6, we characterize the SNR gain of APPLE over ZFP and compare it to the SNR gain of THP over ZFP for the 100-meter cable characterized by Fig. 1.5. It is shown that APPLE has a *peak* SNR gain in excess of 10 dB over THP for the worst-case channel, while the advantage of APPLE is more than 5 dB over THP in *average*.

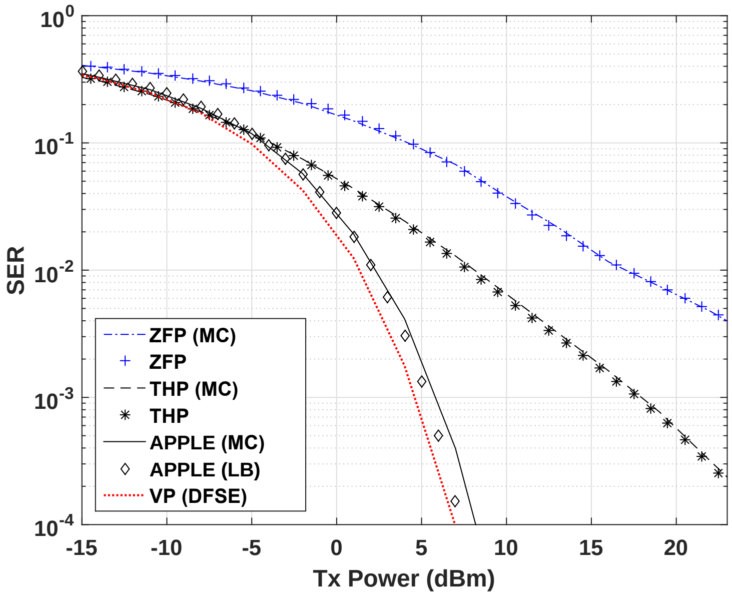


Figure 3.7: SER performance of 16 QAM vs ATP limit per pair under scalar power allocation policy over the full 212 MHz bandwidth. Simulation results are given by the Monte Carlo (MC) method. The DFSE signal cooling scheme is employed for vector perturbation. The lower bound of APPLE is given by Lemma 3.2, while the theoretical performance of ZFP and THP is calculated with the trace approximation.

In Fig. 3.7, the simulated SER of each vectoring scheme using 16 QAM as the input constellation, the lower bound of Lemma 3.2, as well as the theoretical SER performance of ZFP and THP using the trace approximation are compared. It is shown that Lemma 3.2 has a 0.5 dB difference compared to our simulation result, while the gap is negligible for ZFP and THP between simulation results and the theoretical performance. At high SNR regime, the SER performance of APPLE is close to that of the sphere-encoder-based optimal VP scheme, both of which significantly outperform the conventional THP and ZFP, by over 15 dB SNR margin. Due to the ATP constraint of (3.3), whose historical maximum is 8 dBm in the G.fast 106b profile [22], transmitting a full set of 16 QAM symbols over all tones at an SER of below 10^{-4} is impossible for ZFP and THP.

3.4 Dynamic Spectrum Balancing

For vectored DSL systems, the effective channel between the equalized symbol vector $\boldsymbol{z}^t$ and the message symbol vector $\boldsymbol{u}^t$ constitutes a diagonal matrix. Hence the average constellation energy $E\{|u_k^t|^2\}$ of each message symbol's alphabet, i.e. the power allocated to each message symbol, is determined only by the detection SNR requirement of the equalized symbol z_k^t. On the other hand, the choice of constellation is restricted by the detection SNR in the form of the standard capacity expression:

$$b = \log_2 M = \log_2(1 + \frac{\eta}{\sigma}), \tag{3.32}$$

where b is the bandwidth efficiency, i.e. the number of bits per message symbol, and η is the detection SNR. For a given square (i.e. even-bit) QAM constellation, its SNR gap σ [223] towards the Shannon limit of Gaussian channel is defined in terms of the corresponding symbol error rate (SER) target:

$$\sigma = \frac{1}{3}\left[\sqrt{2}\,\mathrm{erfc}^{-1}\left(\frac{SER}{2}\right)\right]^2, \tag{3.33}$$

The SER target can be trivially converted to the BER target which we aim for, based on the bit mapping scheme of the constellation. As demonstrated in Fig. 3.8, each QAM scheme has a specific operating point with respect to the given SER target, which may be calculated individually by the exact SER expression of the QAM constellation relying on either odd or even number of bits per symbol[222][105].

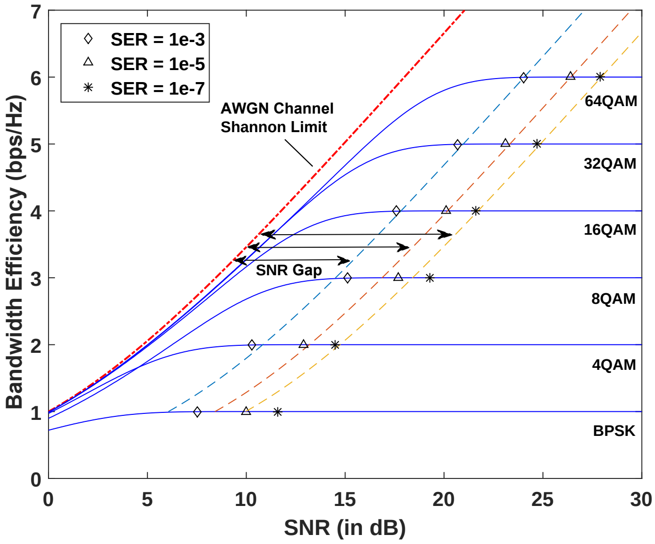


Figure 3.8: The SNR gap towards the Shannon limit of AWGN Channel. For large and even-bit QAM constellations, the SNR gap approximation locates the minimum required SNR for a given symbol error rate target. The suboptimal odd-bit QAM constellations, e.g. BPSK and 8QAM, demonstrate (slightly) wider gaps towards the Shannon limit.

DSB Algorithm

Conventional spectrum balancing strategies are conceived for isolated level-2 DSM in legacy DSL systems. These strategies are largely based on approximations of the convex optimization approaches such as water-filling. In fact, [224] and later [225] established general duality principles between non-convex optimization problems in multi-carrier systems and their convex counterparts. More broadly, DSB approaches can be classified into two main categories:

- **Centralized Algorithms**: Relying on complete CSI knowledge, the DP can employ centralized DSB strategies to achieve optimal spectrum balancing (OSB [226]). Due to the high complexity of OSB in the face of high-dimensional systems, subgroup-based [227] and iteration-based [228][229] OSB were proposed. The family of centralized algorithms will constitute competitive candidates for the next generation, because they can achieve optimal or near-optimal sum rate. However, they have the main disadvantage of being complicated to reconfigure, if the spectral load has to be adjusted due to unexpected IN strikes.

- **Distributed Algorithms**: Mainly used in legacy systems dispensing with vectoring prior to VDSL2, distributed DSB algorithms are typically outperformed by centralized ones, because each transmitting modem of the distributed regime can only optimize itself. The best known distributed algorithm is the iterative water-filling (IWF) scheme of [230][209] and by extension the selective IWF scheme of [210]. Distributed algorithms tend to have lower complexity (e.g. the distributed DSB scheme of [231]) than centralized ones, but the complexity of SSB is still the lowest. Additionally, autonomous algorithms [211][212] relying on a hybrid of centralized and iterative distributed approaches to OSB were shown to have comparable performance to the centralized algorithms in legacy DSL systems.

With respect to vectored DSL systems such G.mgfast and/or G.fast, the DSB paradigm becomes slightly more complicated. In essence, multi-level DSM employs both complex-valued (coordinated QAM signalling, i.e. vectoring) and real-valued (coordinated gain control, i.e. DSB) spectrum management strategies for achieving the optimal sum-rate of a multi-pair channel. From a holistic perspective, the optimal multi-level DSM scheme should ideally aim for jointly optimizing all operational layers defined in Table 1.1, and for all known interferences [232]. The general multi-level DSM paradigm and its algorithms may be considered as follows depending on the transmission link direction:

- **Upstream**. In the upstream, a multi-pair DSL channel is effectively reduced to a diagonal interference channel, i.e. K independent single-pair channels with only background noise and no crosstalk, when either the linear ZF or the ZF-DFE MUD is invoked. In this case, DSB reduces to a trivial, water-filling-like power allocation. When (weighted-)MMSE crosstalk cancellation is used, the MUD itself becomes coupled with the inner spectrum balancing policy. In this case, a joint optimization is necessary.

- **Downstream**. In accordance with our discussion in Ch. 2 and unlike the case with the upstream, the MUPs can decouple a downstream multi-pair channel to independent single-pair channels, but the power controller's deployment retains the cross-correlation among all of the copper pairs (cf. Sec. 3.2.2). In this case we may choose to fix the MUP's configuration and subsequently optimize the inner spectrum balancing policy (for the ZF criterion, cf. Sec. 3.2.1). Alternatively, we may also jointly optimize the MUP and the spectrum balancing policy (for the MMSE criterion).

For the general system model presented in Sec. 3.1.1, each transmitted QAM symbol exhibits an average TxPSD $E\{|x_k^t|^2\}$ that depends on the power allocation P_k^t for all $k = 1, 2, ..., K$ of the same tone t. Therefore, we will mainly consider centralized DSB algorithms. Furthermore, due both to the non-convexity of VMR computation (which is required for per pair TxPSD and per pair ATP characterization) and to the paucity of literature for joint LRMUP-DSB, we shall use the greedy heuristic bit loading algorithm based on [233][218] as an extension of the provably optimal single-pair case of [234] for fairly assessing the performances of the vectoring schemes. At the time of writing, we have not found successful application of the duality principle of [224][225] to overcome the non-convexity associated with lattice reduction. As shown in [235], using the result of the convex optimization as an initial solution is capable of improving the efficiency of the subsequent heuristic algorithm. However, the application is limited to one-dimensional optimization with respect to a single-pair multitone scenario, and extending the approach to 2D remains an open problem. Motivated readers are encouraged to consult [150] for an in-depth survey of the DSB algorithms conceived for DSL transceivers. Let us consider the common scenario, where the power controller invokes the optimized power allocation policy. Since the power constraints (3.2) and (3.3) are instantaneous, we must evaluate the achievable rates of the system by the MES of the corresponding vectoring mapping region. As a result, we would get a lower bound for the bit loading performance with regard to the (information theoretical) capacity. The combinatorial optimization problem (3.1) may now be considered as fitting the largest parallelotope within the duocylinder defined by the TxPSD mask of each pair, with the additional constraint of the ATP limit spanning T independent Euclidean spaces.

The LLL reduction algorithm involved in (3.9) presents many analytical challenges. If $\boldsymbol{A}^t$ is not a scalar matrix, then decomposing $\boldsymbol{G}^t\boldsymbol{A}^t$ results in a completely different unimodular matrix from the one obtained by decomposing $\boldsymbol{G}^t$. The two unimodular matrices are not known to be related linearly, hence there is no known technique of efficiently deriving APPLE's vectoring mapping region from its power-policy-agnostic fundamental mapping region. Therefore we employ a heuristic bit loading algorithm appropriately adapted from [233] and [218]. The algorithm is given in two consecutive parts, namely in Alg. 3 and Alg. 4, respectively.

Let $f(b_k^t)$ denote the minimum required power allocation with respect to the given SER target, coding gain and AWGN PSD, when transmitting at b bits/symbol on tone t pair k.[3] Given our power allocation policy $\boldsymbol{P}^t$ and the time-invariant FEXT canceller $\boldsymbol{G}^t$, $F(\boldsymbol{P}^t)$ represents the mapping matrix $\boldsymbol{S}^t$ of APPLE following the decomposition $\boldsymbol{G}^t\sqrt{\boldsymbol{P}^t} = \boldsymbol{S}^t\boldsymbol{B}^t(\boldsymbol{Z}^t)^{-1}$ akin to that of (3.9). Alg. 3 and Alg. 4 are the two parts of

---

[3] If the multi-pair channel of a given tone t is partially vacant, i.e. $b_k^t = 0$ for some choices of k, then we define $f(0) = \epsilon$ for some complementary signal with negative SNR.

```
Initialization: $b_k^t \leftarrow b_{\max}\ \forall t, k$; $\boldsymbol{P}^t \leftarrow f(b_{\max})\boldsymbol{I}_K\ \forall t$; $\boldsymbol{S}^t \leftarrow F(\boldsymbol{P}^t)$; $\boldsymbol{v}^t \leftarrow \mathrm{MES}(\boldsymbol{S}^t)$;
for all tones $t = 1, ..., T$ do
    while $\max_{1 \le k \le K} |v_k^t|^2 > \mathbb{P}^t$ do
        $k_{\max} \leftarrow \arg\max_k (\max_{1 \le k \le K} |v_k^t|^2)$;
        for all candidate pairs $k = 1, ..., K$ do
            $\bar{b}_k^t \leftarrow b_k^t - 1$; $\bar{P}_k^t \leftarrow f(\bar{b}_k^t)$;
            $\bar{\boldsymbol{P}}^t \leftarrow \mathrm{diag}[P_1^t, ..., P_{k-1}^t, \bar{P}_k^t, P_{k+1}^t, ..., P_K^t]$;
            $\bar{\boldsymbol{S}}^t \leftarrow F(\bar{\boldsymbol{P}}^t)$; $\bar{\boldsymbol{v}}^t \leftarrow \mathrm{MES}(\bar{\boldsymbol{S}}^t)$;
            $\Delta |v_{k_{\max}}^t|^2 |_k \leftarrow |v_{k_{\max}}^t|^2 - |\bar{v}_{k_{\max}}^t|^2$;
        end
        $k^* \leftarrow \arg\max_k (\Delta |v_{k_{\max}}^t|^2 |_k)$;
        $b_{k^*}^t \leftarrow b_{k^*}^t - 1$;
        $\boldsymbol{P}^t \leftarrow \mathrm{diag}[P_1^t, ..., P_{k^*-1}^t, f(b_{k^*}^t), P_{k^*+1}^t, ..., P_K^t]$;
        $\boldsymbol{S}^t \leftarrow F(\boldsymbol{P}^t)$; $\boldsymbol{v}^t \leftarrow \mathrm{MES}(\boldsymbol{S}^t)$;
    end
end
```

Algorithm 3: TxPSD-Constrained Bit Loading

```
Initialization: $b_k^t$, $\boldsymbol{P}^t$, $\boldsymbol{S}^t$ and $\boldsymbol{v}^t$ from Alg. 3;
while $\max_{1 \le k \le K} (\sum_{t=1}^{T} |v_k^t|^2) > \mathbb{A}$ do
    $k_{\max} \leftarrow \arg\max_k [\max_{1 \le k \le K} (\sum_{t=1}^{T} |v_k^t|^2)]$;
    for all tones $t = 1, ..., T$ and lines $k = 1, ..., K$ do
        $\bar{b}_k^t \leftarrow b_k^t - 1$; $\bar{P}_k^t \leftarrow f(\bar{b}_k^t)$;
        $\bar{\boldsymbol{P}}^t \leftarrow \mathrm{diag}[P_1^t, ..., P_{k-1}^t, \bar{P}_k^t, P_{k+1}^t, ..., P_K^t]$;
        $\bar{\boldsymbol{S}}^t \leftarrow F(\bar{\boldsymbol{P}}^t)$; $\bar{\boldsymbol{v}}^t \leftarrow \mathrm{MES}(\bar{\boldsymbol{S}}^t)$;
        $\Delta |v_{k_{\max}}^t|^2 |_{t,k} \leftarrow |v_{k_{\max}}^t|^2 - |\bar{v}_{k_{\max}}^t|^2$;
    end
    Find $(t^*, k^*) \leftarrow \arg\max_{t,k} [\Delta |v_{k_{\max}}^t|^2 |_{t,k}]$;
    $b_{k^*}^{t^*} \leftarrow b_{k^*}^{t^*} - 1$;
    $\boldsymbol{P}^{t^*} \leftarrow \mathrm{diag}[P_1^{t^*}, ..., P_{k^*-1}^{t^*}, f(b_{k^*}^{t^*}), P_{k^*+1}^{t^*}, ..., P_K^{t^*}]$;
    $\boldsymbol{S}^{t^*} \leftarrow F(\boldsymbol{P}^{t^*})$; $\boldsymbol{v}^{t^*} \leftarrow \mathrm{MES}(\boldsymbol{S}^{t^*})$;
end
```

Algorithm 4: ATP-Constrained Bit Loading

the loading algorithm that operates under the constraint of (3.2) and (3.3), respectively. The entire loading algorithm operates under the greedy bit-removal concept, which has been proven to be optimal for a single-pair DMT scenario [234].

In Alg. 3, the system is initialized by assigning the maximum admissible bit allocation $b_k^t = b_{\max}$. The allocation policy of the power controller, the vectoring mapping regions and the MES are determined for all tones and all pairs thereafter. With respect to each tone t, the TxPSD of each pair is compared against the mask $\mathbb{P}^t$. If the highest TxPSD exceeds the mask, then the corresponding pair is denoted as $k_{\max}$ and its TxPSD is given by $|v_{k_{\max}}^t|^2$. Using the greedy principle, the specific pair k^* where subtracting a single bit would have caused the largest reduction of $|v_{k_{\max}}^t|^2$ is selected. The bit load $b_{k^*}^t$ is then updated to have one less bit. The TxPSD characterized by the new bit allocation of tone t is then compared against the mask. Alg. 3 terminates, when the TxPSD mask is fully complied with for all tones right across the entire bandwidth of the system.

Alg. 4 seeks to comply with the ATP requirement using the results obtained in Alg. 3. If the highest ATP exceeds the limit $\mathbb{A}$, then the pair $k_{\max}$ associated with the highest ATP $\sum_{t=1}^{T}|v_{k_{\max}}^{t}|^{2}$ is identified first. The algorithm then determines the load b_{k*}^{t*}, where subtracting a single bit would have caused the largest reduction of the ATP of the pair $k_{\max}$. Then b_{k*}^{t*} is updated to have one less bit and the new ATP of each pair is compared against the limit $\mathbb{A}$. Alg. 4 terminates, when the maximum ATP has been reduced below the limit.

It is worth noting that for all vectoring schemes except APPLE, there is no need to compute the LLL-related factorization for each intermediate step of searching for the optimal bit reduction index. If the power controller invokes a scalar power allocation policy, then the efficiency of Alg. 3 (as well as of Alg. 4) is boosted in two ways. On one hand, the scalar power allocation policy does not differentiate between each pair, hence searching for the optimal pair in line 5-9 for the case of achieving a bit reduction is unnecessary. On the other hand, the VMR and the mapping matrix of APPLE can be efficiently obtained from the fundamental mapping region with the aid of simple scaling. Therefore it is unnecessary to carry out the lattice reduction of (3.9) from scratch for each update of the power policy, as seen in line 8 of Alg. 3. Finally, we should note that both algorithms may be trivially applied for conventional sum rate optimization by substituting the MES computation with the average TxPSD formulation presented in Sec. 3.2.2.

3.5 Performance Comparisons

In this section, we present comparative simulation results for the benchmark MUP algorithms present in Fig. 2.4. Their performance will be characterized in terms of the SER and the sum rate. It is worth noting that the results presented in this section only characterize the performances under the particular set of channel measurements portrayed in Fig. 1.5. However, for channel measurements taken with other DSL binders of the same type and physical parameters, the performance fluctuations should be minimal.

3.5.1 Level-3 MUP Performance

In order to compare the performance of each MUP for transmission over the DSL binder having frequency domain channels characterized by Fig. 1.5, we simulate the average SER of the multi-pair system having SSB for level-2 DSM, versus the average ATP per pair, using the system configuration of Table 3.1. A power constraint is invoked by normalizing the TxPSD to the peak value of the elements of $\boldsymbol{x}^{t}$ for ensuring that the constraint is satisfied for all pairs and for each transmission. A quantitative discussion of optimized joint-level DSM will be presented in the next section regarding the sum rate achieved.

Table 3.1: Default Vectoring Configurations

Parameter	Value
Constellation	16QAM
Modulation	DMT
Channel Coding	N/A
Lower Spectral Bound	517.5 kHz
Upper Spectral Bound	212 MHz
Tone Spacing	517.5 kHz
Number of Pairs	10
AWGN Floor	-150 dBm/Hz
Binder Length	100 m
$B_{\max}$	12
$N_{\max}$	12

Fig. 3.9 demonstrates the SER performance of the vectoring schemes over the expanded 212 MHz G.fast channel profile assuming that the DP as the downstream transmitter has perfect CSI knowledge. The performance of linear precoding and the THP, as well as of the LRMUPs is compared. It is clearly seen in Fig. 3.9 that the best SER achieved by the conventional precoding schemes (THP) is approximatley 8 times higher than the worst-case performance of LRMUP (LR-ZFP) at the recommended operating point [63] of 4 dBm per-pair ATP. At the relaxed 8 dBm per-pair ATP operating point, the SER of conventional precoding is ten times higher than that of the LRMUP. The gain of the linear MMSE precoding over ZFP becomes most prominent at low to medium SNRs, which exceeds that of the LR-ZFP at the ATP of -8 dBm or lower. The sorted THP scheme is seen to be the most advantageous one at medium SNRs, outperforming the conventional THP. However, the gain of these linearly improved schemes is insignificant compared to that of lattice reduction.

Fig. 3.10 compares the robustness of the conventional precoding schemes and their lattice reduction aided counterparts in the face of CSI estimation errors, when the vectoring mechanism of Fig. 1.10 is invoked. The approximate lattice precoding schemes significantly outperform their counterparts operating without lattice reduction. Additionally, it is also apparent that the LR-THP and VP both exhibit high robustness against imperfect CSI estimation, whilst all the other precoding schemes suffer from a substantial SNR loss, as well as from a high SER floor above 10^{-3}. However, the LR-ZFP achieves the same SER performance as the THP at the 4 dBm operating point, even if the DP has access to perfect transmit CSI knowledge in the case of the THP.

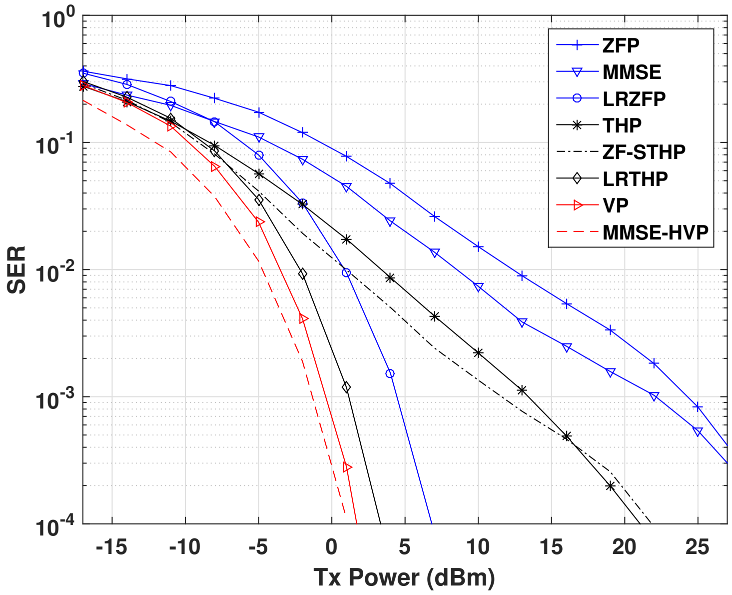


FIGURE 3.9: Average SER for transmission over the 212 MHz G.fast channel profile versus the average ATP per pair. The message symbol alphabet is 16QAM and a scalar power assignment policy is enforced. The (linear) MMSE scheme is based on the regularized ZFP of [121].

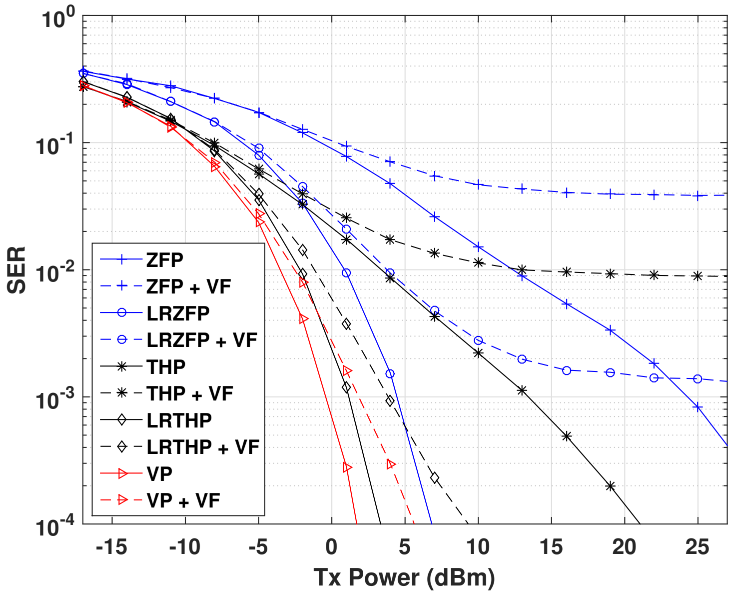


FIGURE 3.10: Average SER for transmission over the 212 MHz G.fast channel profile with respect to the robustness against imperfect CSI knowledge. The case where the ONU has access to perfect non-causal knowledge of the downstream CSI is compared against the case, where the DP acquires the downstream CSI with the aid of the vectoring feedback loop of Fig. 1.10.

It has been shown in [133] that more advanced linear MUP schemes, such as the one proposed and investigated in [132], may become capable of *outperforming* the plain THP at certain operating points associated with moderate degrees of CSI estimation error, subject to an optimized level-2 DSB policy. This is not observed for the operating point defined in this section based on the standard operations described in both Section 1.4.2 and [22], due both to the potential difference in the measured channel and to the lack of optimized multi-level DSM in this section. Additionally, we also recommend further investigations of the practical operating point regarding the tolerable degree of CSI imperfection in G.mgfast systems. On the other hand, the seemingly surprising result showing the superiority of linear MUP over the classic THP reported in [133] was considered to be due to the fact that the plain THP as a greedy scheme (i.e. first user gets the best performance) is susceptible to instability, therefore it is more sensitive to imperfect CSI in the face of ill-conditioned multi-pair channel. This observation is consistent with our comparisons and it is explained in the beginning of Section 2.3.2.2. In particular, this has lead to improved THP-like schemes such as both the sorted THP and the LR-THP.

3.5.2 Multi-Level DSM Performance

3.5.2.1 General Achievable Rate

Perfect DP-side CSI knowledge is assumed for the performance characterization of multi-level (joint level-2 and level-3) spectrum management. In Fig. 3.11, the throughput per pair is shown for each MUP. The greedy bit loading technique of Alg. 3 and Alg. 4 is invoked under the bit cap of $b_{\text{max}} = 14$ and 4 dBm ATP limit per pair, under the standard TxPSD mask defined in [63]. The SSB policy employed by the LRMUPs constrains their *degree of freedom*. Thus their performance is compromised as a result of the associated worst-case dominance. However, using the LR-ZFP under the SSB policy will still increase the binder's total sum rate by 6% over that of ZFP using greedy bit loading. Given that the DSL channel is quasi-static, the long-term complexity of the LR-ZFP will be identical to that of ZFP, since the additional complexity of initialization can be ignored.

Fig. 3.12 quantifies the average bit-loading over all pairs per tone for each MUP. As indicated by the channel quality degradation characterized in Fig. 1.5 and Fig. 2.11, the number of supported bits drops at high frequencies for all MUPs. We note that the sum rate of the idealized VP, where the VMR is a perfect hypersphere, is slightly higher than that of the DSB-aided THP, even though the former does not rely on DSB policy. Even though the hypersphere VMR and subsequently the ideal performance of VP is not achievable in reality, it may be practically achievable using the optimal lattice coding strategy for IFP.

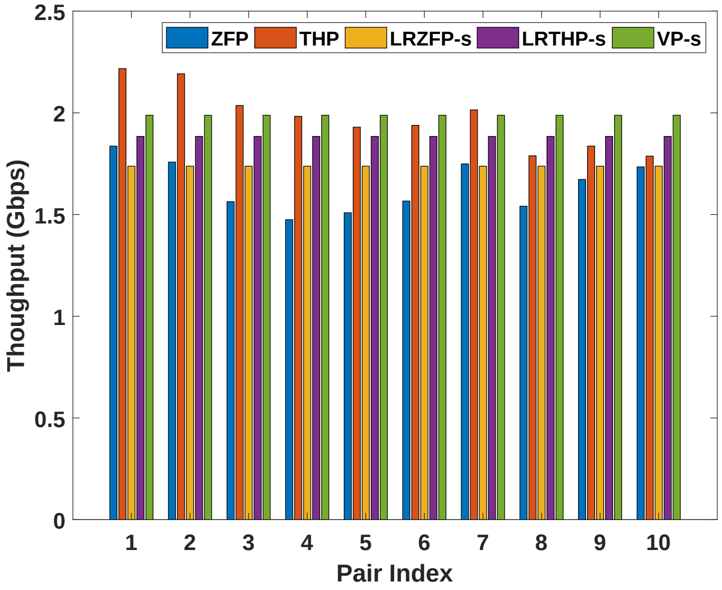


FIGURE 3.11: Throughput per pair for the vectoring schemes using the 100-meter 10-pair cable characterized by Fig. 1.4. The TxPSD mask of [63] is invoked and the ATP limit is 4 dBm per pair. The simulation uses an AWGN floor of $N_0 = -150$ dBm/Hz and the bit cap is $b_{\max} = 14$ bits per pair per tone. The power policies for ZFP and the THP are optimized with the EZF bit loading algorithm of [218], while SSB policies (-s) are employed by the LR-ZFP, the LR-THP and VP.

Additionally, it has been identified in Section 2.3.2.3 that the only loss of the optimized THP is the 0.255 bps/Hz 'shaping loss' at high SNRs. However, the influence of the 'modulo loss' becomes significant particularly for the low SNR range as depicted in Fig. 2.8 and also reported in [218]. At low SNRs, the precoding loss is also considerably higher due to the lower admissible constellation size whose effect has been characterized in Fig. 3.2. Under these considerations, the sum rate achieved by the LR-THP relying on SSB is 4.5% lower than that of the THP. However, it was discovered in [96] that an alternative MUP whose VMR overlaps with that of the LR-THP achieves the same near-optimality as the optimized THP. More interestingly, as portrayed in Tab. 3.2, if all MUPs employ the SSB policy, then the achievable sum rate of the LR-ZFP becomes marginally higher than that of the THP, despite the fact that the former has lower run-time complexity. However, we should note that this observation heavily relies on the goodness of the (reduced-)lattice basis in the multi-dimensional signal space, which is in practice dependent on the channel's profile.

Finally, the influence of SDN-aided cross-ISP vectoring is demonstrated in Tab. 3.2, given that the minimum-complexity SSB policy is used. Assuming that the two vectored groups of five subscribers each from two individual ISPs treat each other as alien FEXT in the case of LLU, we may observe a sum rate boost of upto 86% from employing cross-ISP vectoring. Moreover, both approximate lattice precoding schemes attain a higher net gain in sum rate than their standard counterparts operating without lattice reduction.

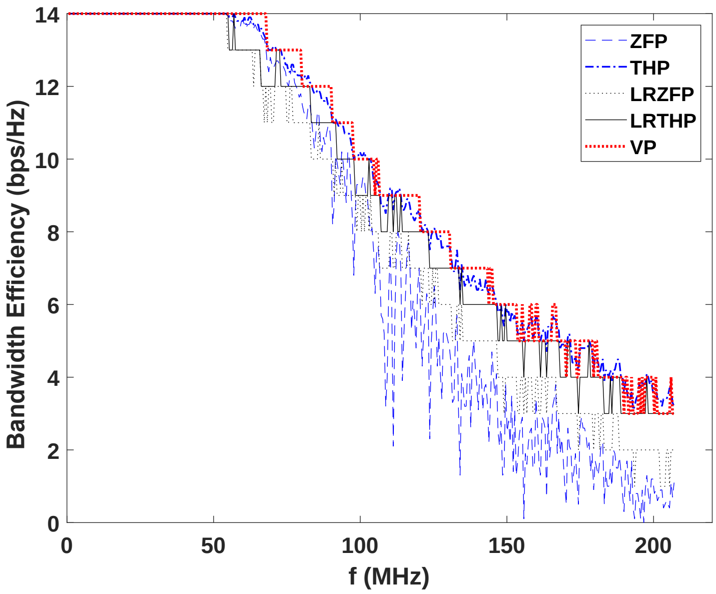


FIGURE 3.12: Average bit loading per tone over all pairs. The simulation configuration from Fig. 3.11 is used. The spectral load for ZFP and the THP are optimized with the greedy bit loading algorithm of Alg. 3 and 4, while SSB policies (-s) are employed by the LR-ZFP, the LR-THP and VP.

TABLE 3.2: Sum Rate Performance Comparison (Gbps) for 10-pair 100-metre DSL binder

	ZFP	THP	LR-ZFP	LR-THP	VP
LLU-SSB	8.7	9.2	9.7	10.5	N/A
SSB	14.5	17.2	17.4	18.8	19.9
DSB	16.4	19.7	17.4	18.8	19.9

3.5.2.2 Empirical Lower Bound

We consider an ambitious futuristic configuration, which we expect to herald the next-generation metallic access network beyond G.fast. The following results are obtained using channel measurements characterizing a 100-meter 10-pair DSL binder as well as a 50-meter 10-pair one, both of which occupy the expanded baseband spectrum of frequencies spanning all the way upto 300 MHz. The TxPSD mask of [63] is imposed for the below 212 MHz range and a flat TxPSD mask extended from the 212 MHz point onwards is imposed for frequencies over 212 MHz. The ATP limit remains at 8 dBm per pair as in the 106b G.fast profile, while a -150 dBm/Hz AWGN PSD is assumed. Without loss of generality, let the bit loading be capped at 15 bits corresponding to 32768-QAM as in VDSL2. The sum-rate achieved by APPLE is compared against those achieved by the conventional ZFP and THP, as well as against the capacity given by the sum-rate of DPC. The DPC capacity is computed under the relaxed constraint of the total *average* (instead of peak) TxPSD per binder akin to the context of the original DPC scheme, i.e.

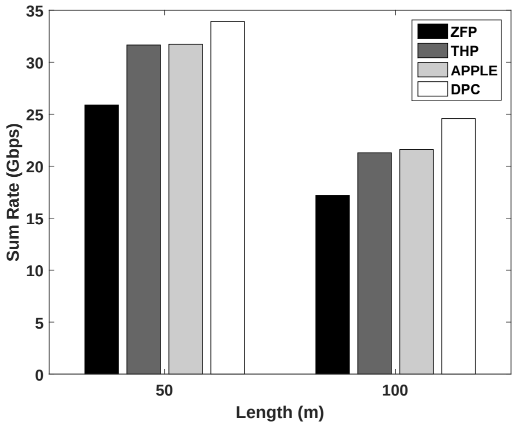


FIGURE 3.13: Sum-rate of 10-Pair 0.5mm short DSL binders under *DSB*. An extended baseband spectrum of upto 300 MHz is used with noise floor at -150 dBm/Hz.

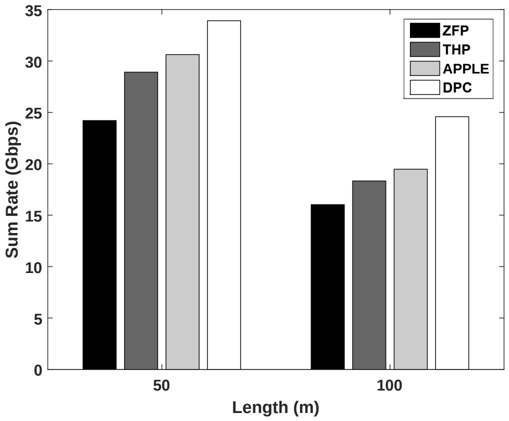


FIGURE 3.14: Sum-rate of 10-Pair 0.5mm short DSL binders under *SSB*. An extended baseband spectrum of upto 300 MHz is used with noise floor at -150 dBm/Hz.

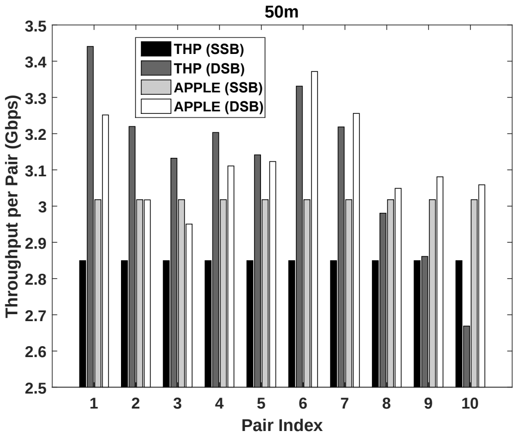


FIGURE 3.15: Per-pair throughput of a 10-Pair 0.5mm DSL binder of length *50m*.

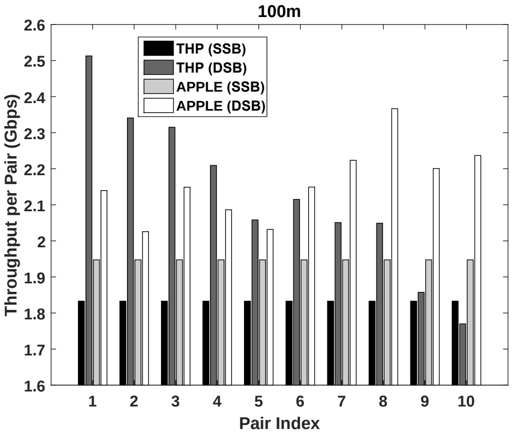


FIGURE 3.16: Per-pair throughput of a 10-Pair 0.5mm DSL binder of length *100m*.

we have $\|\boldsymbol{x}^i\|^2 \leq K\mathbb{P}^i$.

In Fig. 3.13 and Fig. 3.14, the sum-rate lower bound of each vectoring scheme, obeying the constraints of the peak TxPSD and peak ATP is shown for both the 50m and 100m cable length and for both DSB and for SSB. The performance of APPLE is almost identical to that of THP, when the power controller employs DSB. By contrast, under SSB, the sum-rate of APPLE exceeds that of both ZFP and THP. The throughput of APPLE and that of THP are compared in more details in Fig. 3.15 and Fig. 3.16. It is explicitly shown that APPLE can achieve an improved user-fairness and a better worst-case performance than THP under DSB, while retaining an identical sum-rate. On the other hand, APPLE can also achieve a higher sum-rate than THP when both have identical fairness under SSB.

3.6 Chapter Summary

In this chapter, we conducted a performance study of multi-level DSM based on joint vectoring and spectrum balancing, under practical power requirements defined by the recently published metallic access network standards. More specifically, the problem of adaptive rate and power allocation is investigated in conjunction with MUP, subject to a pair of peak power constraints per copper pair. Following the unified system model of Fig. 3.3, we expanded the concept of *shaping loss* discussed earlier in Sec. 2.3.2.3 associated with a standard bit-to-symbol mapping scheme in the single-stream domain to that of the vectoring mapping region in the multi-stream domain. Subsequently, we formulated the power transfer characteristics of the MUPs of Ch. 2 in Sec. 3.2 based on the geometric properties of the corresponding vectoring mapping regions, in order to define the operations of the multi-level DSM algorithms Alg. 3 and 4.

In order to fairly compare the performances of multi-level DSM associated with different MUPs, a reference evaluation framework was proposed. Firstly, we related the peak-power-constrained spectrum balancing problem to the novel concept of maximum Euclidean spread of Sec. 3.1.2.2 based on the vectoring mapping region characterizations. Additionally, Alg. 2 was proposed for solving the maximum Euclidean spread for all parallelotopes at low complexity. Moreover, we revisited the fundamental spherical limit of the vectoring mapping region associated with any given MUP scheme in Sec. 3.2.3. By extension, a lower SNR-loss bound (Lemma 3.2) was derived for the class of lattice mapping based MUPs in Ch. 2 associated with a parallelotope-shaped vectoring mapping region, which also constitutes a tight approximation for the performances both of LRTHP and of APPLE as discussed in Sec. 3.1.3. Furthermore, we generalized the optimal heuristic DSM algorithm conceived for single-user multi-tone systems dispensing with MUP to the pair of bit loading algorithms applicable to other types of DSL systems in Sec. 3.4 (Alg. 3 and 4), the latter of which is also the first discrete adaptive power allocation algorithm conceived for LR-aided non-linear multi-user systems.

Based on our vectoring mapping region modelling, we comparatively investigated the performances of both SSB-based DSM with respect to the SER and DSB-based multi-level DSM in terms of the sum rate. It was observed in Fig. 3.10 that LR-aided techniques are capable of significantly boosting the SER performances in comparison to their standard counterparts without LR, even in the face of practical transmit CSI estimation errors. The sphere encoder was shown in Fig. 3.9 and 3.10 to achieve the best SER performance both with and without perfect transmit CSI. However, when comparing the sum rate performances obtained via the bit loading algorithms of Sec. 3.4, we observe in Fig. 3.11 and 3.12 that the advantage of LR becomes less prominent. The observation follows the fact that the standard non-linear THP is near-optimal in terms of its sum rate, when relying on optimal water-filling-based DSB. Conversely, we observed in Sec. 3.5.2.2 that the sum rate performance of THP degrades considerably when employing the straightforward SSB strategy, due to the poor user fairness. In this regard, employing the LR-based preprocessing strikes a better trade-off between fairness and sum rate, which is also characterized by the comparison of THP and APPLE seen in Sec. 3.5.2.2.

Chapter 4

Error Control for DSL Networks[1]

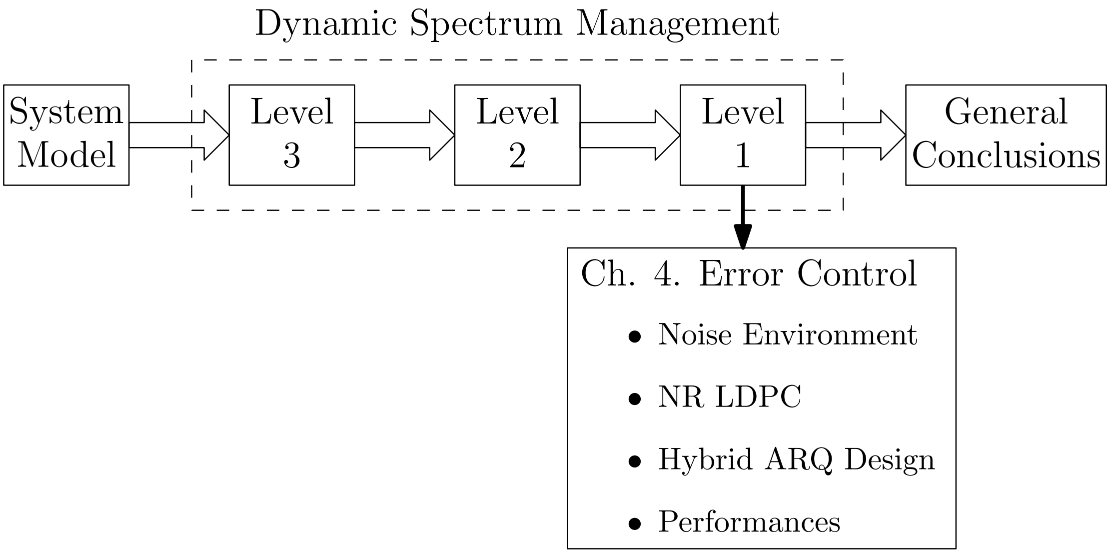


FIGURE 4.1: Outline of Ch. 4.

4.1 Noise-Contaminated Metallic Networks

As we have discussed earlier in Sec. 1.4.4, the metallic access networks including both DSL and power line systems constitute hostile noisy environments. In contrast to crosstalk, which is commonly treated as *interference*, we define *noise* as the stochastic contamination of the desired signal. For the consistency of this treatise and for agreement with the typical definition of SINR, apart from crosstalk, any other sources of interference will also be treated as noise. Furthermore, noise is present even in single-pair or interference-free wireline networks. In this treatise, we shall shift our attention from the multi-pair-based level 2 and level 3 DSM (cf. Ch. 3) to the single-pair-based level

---

[1] This chapter is based on collaborative work with Shuai Shao, whose contribution to the EXIT-chart analysis is gratefully acknowledged.

1 DSM. Without loss of generality, the noise sources of metallic access networks mainly consist of three categories listed as follows:

- **Stationary Noise**. In DSL networks, the stationary noise has a constant average PSD of σ_n^2 over the entire operational bandwidth of the spectrum. Furthermore, the amplitude of the stationary noise obeys the Gaussian distribution in $\mathbb{C}$ according to the central limit theorem. As a result, the complex-domain AWGN model $\mathcal{CN}(0, \sigma_n^2)$ is commonly used for characterizing the stationary noise environment in DSL and other communication systems, where both the real and the imaginary part has a PSD of $0.5\sigma_n^2$. Since the Gaussian distribution is transformed into another Gaussian distribution by the DFT and IDFT, the AWGN model is applicable in both the time-domain (TD) and the frequency-domain (FD) analysis. A reference PSD value of around -150 dBm/Hz is commonly used for configuring the power controller, as discussed in Ch. 3. However, this particular value of the AWGN PSD is generally a worst-case parameter, which reflects the combined effect of both the DSL cable and of the measuring equipment, since lower AWGN PSD values have also been achieved in practical experiments.

- **Impulsive noise**. The antenna-like structure of some DSL cables such as the overhead drop wires and of the power lines typically suffer from IN caused by the transient-state responses of other electrical apparatus in close proximity. In the DSL standard, IN events are classified as repetitive electrical impulse noise (REIN) and single high impulse noise events (SHINE) [82], depending on whether the impulses occur at a quasi-constant inter-arrival time (IAT). More specifically, a realistic IN model is characterized by the joint probability distribution of its amplitude, impulse duration, IAT and FD correlation [83]. However, simplified models are also widely used in practical performance assessment [67]. For analytical tractability, we will consider the Bernoulli-Gaussian IN (BGIN) model of [85] as a reasonable simplification. The BGIN model characterizes the total background noise samples using the following Gaussian mixture PDF:

$$P_n(p, \sigma_n^2, \alpha) = (1-p)\,\mathcal{CN}(0, \sigma_n^2) + p\,\mathcal{CN}(0, \alpha\,\sigma_n^2), \tag{4.1}$$

  where p is the Bernoulli parameter, σ_n^2 is the AWGN PSD and α is the PSD ratio of the expected impulses over the AWGN. The main problem of (4.1) is the inability to characterize the IN's bursty nature, particularly its duration and FD correlation. Specifically, the properties of DFT suggest that a TD impulse having a short duration spreads its concentrated energy over a wide bandwidth in the FD and vice versa. Therefore, in the FD analysis, a number of consecutive tones become contaminated, whenever an impulse occurs.

- **Radio-Frequency Interference**. This is sometimes also referred to in parlence as noise, because it is not measured during the initialization of the DSL transceiver units. The majority of wireless access networks have been operating in the passband, at multi-GHz carrier frequencies. However, the baseband operation of metallic access networks is still susceptible to RFI from TV broadcast services and amateur radio. The spectral conflicts escalate even further upon the impending arrival of the G.mgfast standard, whose designated 848 MHz baseband bandwidth overlaps with the entire family of TV services. Since the amateur radio band is not used for standardized services, there is a paucity of literature on the distribution of RFI. Therefore, RFI suppression techniques typically rely on practical measurements [84], which may be improved by the massive AI integration in the holistic next-generation access network. Based on the central limit theorem, we will dispense with further discussions about RFI by considering it as a part of the white noise.

4.2 General Procedure

In practice, due to the limited accuracy of impulsive noise prediction, identification and removal, a worst-case noise margin has to be observed upon configuring the power controller for DSB. However, such an error control approach does not exploit any potential *a posteriori* knowledge about the IN event, which will inevitably lead to pessimistic sum rate performance estimates. On the other hand, interleaving and ARQ-enabled retransmission schemes are also employed for mitigating the bursty errors caused by IN events. In the next generation metallic access networks, the hybrid ARQ (HARQ) aided level 1 DSM [71] (cf. Sec. 1.4.4) will have a more prominent role in impulsive noise protection (INP), since powerful capacity-approaching LDPC-based FEC schemes will replace the classic Reed-Solomon codes. Since the updated INP standard is yet to be fully documented, we will rely on the recent LDPC design recommended by the 5G New Radio (NR) specifications. Based on the NR LDPC design, we propose a novel reduced-latency HARQ protocol incorporating both carefully conceived strategic decoder activation and early decoding-iteration termination, which were originally designed for turbo codes [236].

The main framework of HARQ-aided error control is depicted in Fig. 4.2. Compared to the multi-pair multi-tone DSL network's physical layer architecture portrayed in Fig. 1.10 of Ch. 1, we will mainly consider the operation of the FD symbol encoders and decoders of Fig. 1.10, while considering a single-pair topology. Specifically, we will concentrate our attention on the general procedures of the LDPC-based FEC coding scheme and the CRC-aided retransmission protocol. For simplicity of discussion, we may also view the entire section of the transmission link from the output of the FEC encoder to the input of the FEC decoder as the *equivalent forward channel.* It should be noted that some of the building blocks portrayed in Fig. 4.2 do not necessarily operate

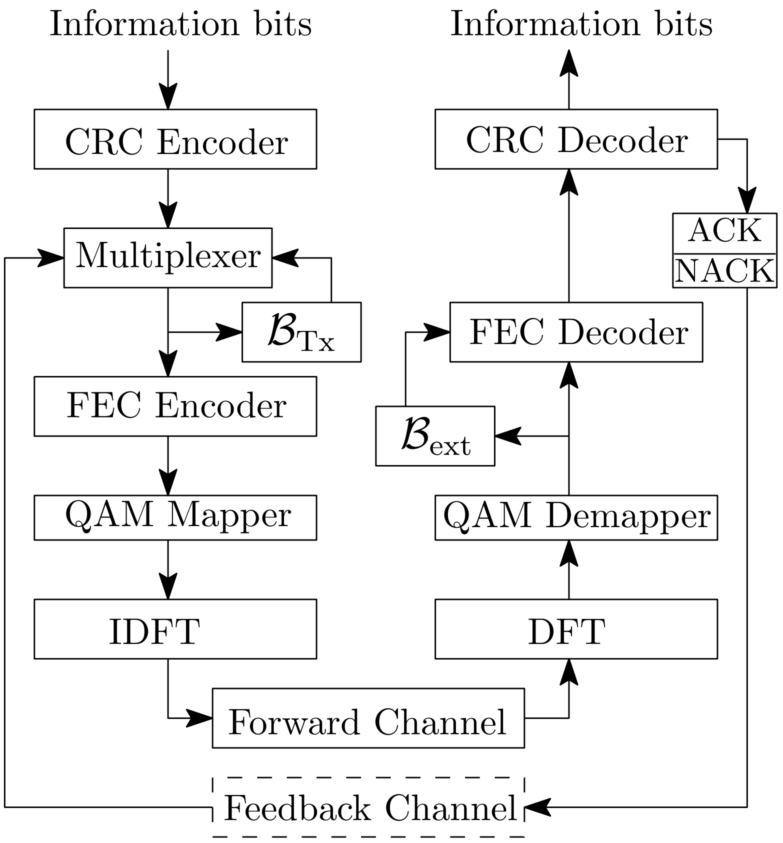


FIGURE 4.2: Hybrid ARQ aided error control for a single-pair multi-tone DSL link (adopted from [82]).

in a strictly sequential order, such as the FEC encoder and the QAM mapping in the trellis coded modulation (TCM) scheme employed in G.fast.

The basic operations of error control relying on both error detection and error correction codes are described as follows. The procedure is symmetric in both upstream and downstream transmission, since we do not consider the multi-user context. At the information source, a block of message-carrying information bits are fed into the CRC error detection encoder of Fig. 4.2, where the CR prefix is attached. During the first transmission attempt, the packet is directly fed into the FEC encoder, which attaches parity bits to the block. Simultaneously, the block is also stored in a transmission queue in the buffer $\mathcal{B}_{\text{Tx}}$ of Fig. 4.2. The fully encoded information block is then transmitted over the equivalent forward channel to the FEC decoder, which will attempt to correct the erroneously received information bits based on the parity bits. Meanwhile a copy of the received encoded block is stored in the external buffer $\mathcal{B}_{\text{ext}}$ of Fig. 4.2 (as opposed to the FEC decoder's internal buffer $\mathcal{B}_{\text{int}}$ shown later in Fig. 4.11). If the number of contaminated bits in the received information block does not exceed the error correction capability of the FEC code, then all bit errors can indeed be corrected. The FEC decoder of Fig. 4.2 subsequently passes the decoded information block to the error detection decoder, which determines whether the information block is received free from errors or not. If the information block is received correctly, a positive acknowledgement (ACK) flag is sent to the transmission multiplexer of Fig. 4.2 via the feedback channel, otherwise a negative acknowledgement (NACK) flag is sent. When the transmission multiplexer receives a NACK or has waited for an agreed amount of time without receiving either type of acknowledgement, it will suspend transmitting any new information blocks and initiates a retransmission of the information block currently stored in $\mathcal{B}_{\text{Tx}}$. When a retransmitted code block is received, the demodulator (DFT+QAM) of Fig. 4.2 will instruct $\mathcal{B}_{\text{ext}}$ to combine this newly received code block with the current memory state of $\mathcal{B}_{\text{ext}}$ via soft bit (cf. Sec. 4.3.2) addition. Alternatively, a sequence of increasingly stronger FEC codewords associated with gradually increased parity bit density may be transmitted for attaining an increased coding gain throughout the consecutive retransmission attempts. Moreover, typically hybrid ARQ schemes incorporating both strategies are employed in industrial standards [237], as we will discuss in Sec. 4.4. The retransmission cycle is repeated until the multiplexer receives a positive ACK or the maximum number of retransmissions has been exhausted. Subsequently, the multiplexer resumes transmitting new information blocks, while both $\mathcal{B}_{\text{Tx}}$ and $\mathcal{B}_{\text{ext}}$ of Fig. 4.2 are reset.

4.3 New Radio LDPC

Since the 90s, coding theory and practice have been relying on the class of iteratively decoded near-capacity graph codes such as turbo [238] and LDPC codes. However, the basic operations of LDPC coding were formally proposed as early as 1963 in the PhD thesis of Robert Gallager [239]. However, LDPC codes had been neglected for several decades, since their complexity was deemed excessive in that era. Nonetheless, with the advent of powerful DSP hardware, graph codes have become prevalent in both wireline and wireless systems. The family of LDPC codes has shown superiority in the 5G case studies [240]. Subsequently, LDPC has gained dominance both in the 5G NR wireless standard and in the G.mgfast wireline standard. In this section, we will discuss the basic principles of the NR LDPC code and its asymptotic performance.

4.3.1 Code Structure

The family of LDPC codes has a pair of equivalent characterizations. Specifically, the parity check matrix (PCM) of LDPC codes is related to the generator matrix conceived for other FEC codes, while the *Tanner graph* (also referred to as the factor graph) [241] is the graphical dual counterpart of the PCM. From an information theoretical perspective, the family of LDPC codebooks sharing the same asymptotic performances over a large number of channel use events is termed as an *ensemble*, which is characterized by either a *base PCM* or a *base graph* [240]. In general, the base graph has sufficient statistics of the corresponding full Tanner graph, while facilitating a more compact characterization. In the 5G NR standard [237], the process of mapping the base graph onto the full Tanner graph is termed as *lifting*, which essentially expands each code bit in the base PCM into a square matrix of a designated lifting size Z_c. To facilitate flexible design, multiple lifting sizes are conceived in NR for supporting many coding rates and code block lengths relying on only two base graphs. Specifically, 51 values of Z_c ranging from 2 to 384 are supported [237]. Since the lifting sizes do not affect the analytical average performances of the NR LDPC codes, we will discuss the PCMs and Tanner graphs in their respective base forms for the remainder of this treatise, while noting that each non-zero entry carries a $(Z_c \times Z_c)$-element square matrix.

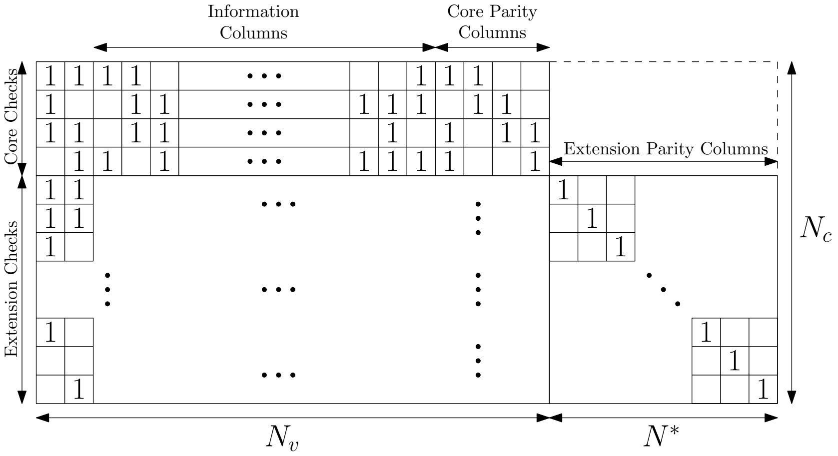


FIGURE 4.3: The second variant of the NR LDPC base PCM (Base Graph 2). The core LDPC region has $4Z_c$ rows, due to the degree limit of its VNs conceived in [240].

For a given LDPC code, the columns of the PCM represent the LDPC encoded codeword, which corresponds to the variable nodes (VN) on one side of the Tanner graph. On the other side, the rows of the PCM characterize the parity checks associated with the codeword, which are termed as the check nodes (CN). The connections in the Tanner graph are known as the *edges*, which correspond to the non-zero entries in the PCM. Moreover, the *degree* of a given node represents the number of edges connected to it. Due to the duality, the most widely employed family of LDPC decoding algorithms may be interpreted as a message-passing process between the set of VNs and CNs along the edges in the Tanner graph. The most powerful algorithm of this category is termed as the sum-product (SP) or the belief propagation algorithm [242].

The NR standard facilitates a hybrid FEC design in favor of coding rate flexibility, which is particularly important for incremental redundancy aided (IR) HARQ (IR-HARQ). Specifically, the NR LDPC code is constituted by the concatenation of a conventional LDPC and a *low density generator matrix* (LDGM) based code variant [243]. Consequently, the PCM of the NR LDPC code of Fig. 4.3 consists of three regions, including the $(N_c - N^*) \times N_v \times Z_c^2$ core LDPC region, the $N^* \times N_v \times Z_c^2$ LDGM region directly below the core LDPC region, and an $N^* \times N^* \times Z_c^2$ information transfer region to the right of the LDGM region. The information transfer region is an identity matrix, which essentially appends degree-Z_c *extension parity* to the existing LDPC-LDGM block. The remainder (top right) of this PCM is left blank or padded with an all-zero matrix. For the first base graph BG1 we have $N_v = 26, N_c = 46$ and $N^* = 42$, whilst for BG2 we have $N_v = 14, N_c = 42$ and $N^* = 38$. In terms of the holistic design, the information columns correspond to all the message-carrying information bits where the first two columns are punctured during transmission as seen in Fig. 4.10. The right-most four columns of the LDPC-LDGM region correspond to the core parity bits of the codewords. In contrast to the core LDPC parity used for error correction, the degree-Z_c extension parity is mainly used for the convenience of codeword puncturing during IR-HARQ. On the other hand, the rows of this L-shaped PCM consist of the core parity checks at the top and the extension parity checks at the bottom. In general, the extension check rows belonging to the LDGM region have much lower density of '1's than the core check rows belonging to the core LDPC region.

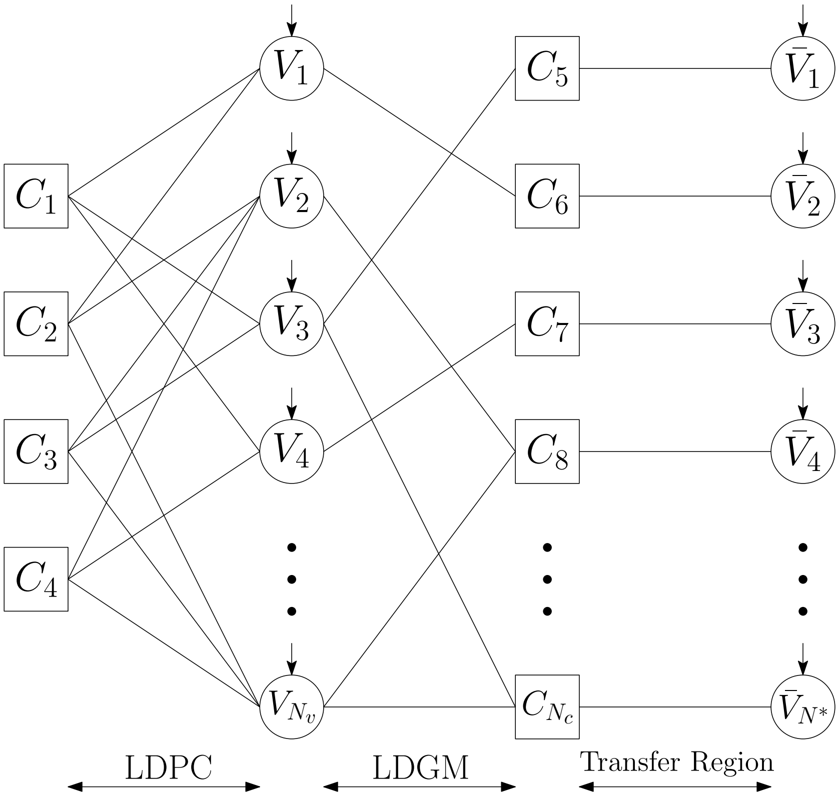


FIGURE 4.4: An example of the base Tanner graph of NR LDPC. The number of CNs in the core LDPC region is limited to $4Z_c$.

The three regions of the NR LDPC PCM naturally lead to the expanded Tanner graph characterization of Fig. 4.4. Hence, the NR LDPC decoding process may be viewed as three interconnected message-passing processes between four sets of nodes. More specifically, the core LDPC region and the LDGM region share the same set of VNs, whilst the LDGM region has the same set of CNs as the information transfer region. The set of VNs $\{\bar{V}_i\}$ associated with the information transfer region directly receives a portion of the noisy raw information from the channel and transfers it to the CNs associated with the LDGM region. The set of VNs associated with the core LDPC (and with the LDGM) region receives the remaining noisy raw information and exchanges the processed mutual information (MI) both with the CNs of the core LDPC region and with those of the LDGM region. Again, the LDGM region is sparser than the LDPC region. Therefore, the nodes associated with the core LDPC have higher average degree than those of the LDGM. In the next section, we will characterize the asymptotic performance of the NR LDPC code based on Fig. 4.4 as well as on the classic SP decoding algorithm.

4.3.2 Extrinsic Information Transfer Analysis

4.3.2.1 Soft Information

The optimal LDPC decoding algorithm satisfies the *maximum a posteriori* decision criterion, which relies on the *a posteriori probability* (APP) information associated with the (demodulated) soft bits. For convenience, many message-passing-aided soft decoders accept the log likelihood ratio (LLR) of the received code bits as input *message*, rather than using their APPs. In particular, the LLR associated with a specific code bit b belonging to a received symbol y is formulated as:

$$L(b) = \ln\left[\frac{P(b=0|y)}{P(b=1|y)}\right]. \tag{4.2}$$

By considering the generic BGIN model of (4.1), while ignoring the influence of the *constant* DSL channel for the moment, we may naturally formulate the bit-level LLRs in the form of:

$$L(b) = \ln\left[\frac{(1-p)\sum\limits_{x\in\mathcal{U}_0} e^{-|y-x|^2/\sigma_n^2} + \frac{p}{\sqrt{\alpha}}\sum\limits_{x\in\mathcal{U}_0} e^{-|y-x|^2/(\alpha\sigma_n^2)}}{(1-p)\sum\limits_{x\in\mathcal{U}_1} e^{-|y-x|^2/\sigma_n^2} + \frac{p}{\sqrt{\alpha}}\sum\limits_{x\in\mathcal{U}_1} e^{-|y-x|^2/(\alpha\sigma_n^2)}}\right], \tag{4.3}$$

where $\mathcal{U}_0$ represents the half constellation[2] having a '0' in the bit position of b, while $\mathcal{U}_1$ is defined similarly for '1's in the bit position of b. It should be noted that the exponent in (4.3) is calculated in the complex domain $\mathbb{C}$, hence the noise variance is $\sigma_n^2/2$. In floating point calculations, (4.3) overflows quickly when the variance of AWGN tends to zero. In [244], an approximation was proposed for reducing the complexity of LLR calculation and for preserving numerical stability. More specifically, the summation over the respective half constellation is replaced by taking the minimum. For large constellations, such as 4096 QAM [245], such an approximation facilitates near-optimal low-complexity soft demodulation, since the likelihood functions are now only calculated for the closest pair of symbols rather than for the entire constellation. Another reasonable approximation of (4.3) is to calculate the first few terms of the geometric series expansion of the exponential functions, which also preserves numerical stability in floating point calculations.

---

[2] Within the scope of this chapter, $\mathcal{U}$ represents an actual discrete QAM constellation rather than some continuous approximation of it.

When the parameter set (p, σ_n^2, α) of the BGIN model is perfectly known to the demodulator, the soft information extracted from (4.3) is indeed reliable. More specifically, the confidence level of a soft bit b is at a local maximum only when the corresponding received symbol y overlaps with one of the constellation points. This avoids a specific problem of the conventional AWGN-based LLR measure, where the confidence level falsely grows when a received symbol is located far away from the constellation's outer boundary, since in reality such an observation is more likely a consequence of IN contamination. The comparison is depicted in Fig. 4.5, where we observe that the more distinguishable the impulse is from the AWGN, the better our LLR modelling is. However, (4.3) becomes much less accurate at high AWGN variance (low background SNR), since we observe that the local maximum shifts away from the actual constellation point's position.

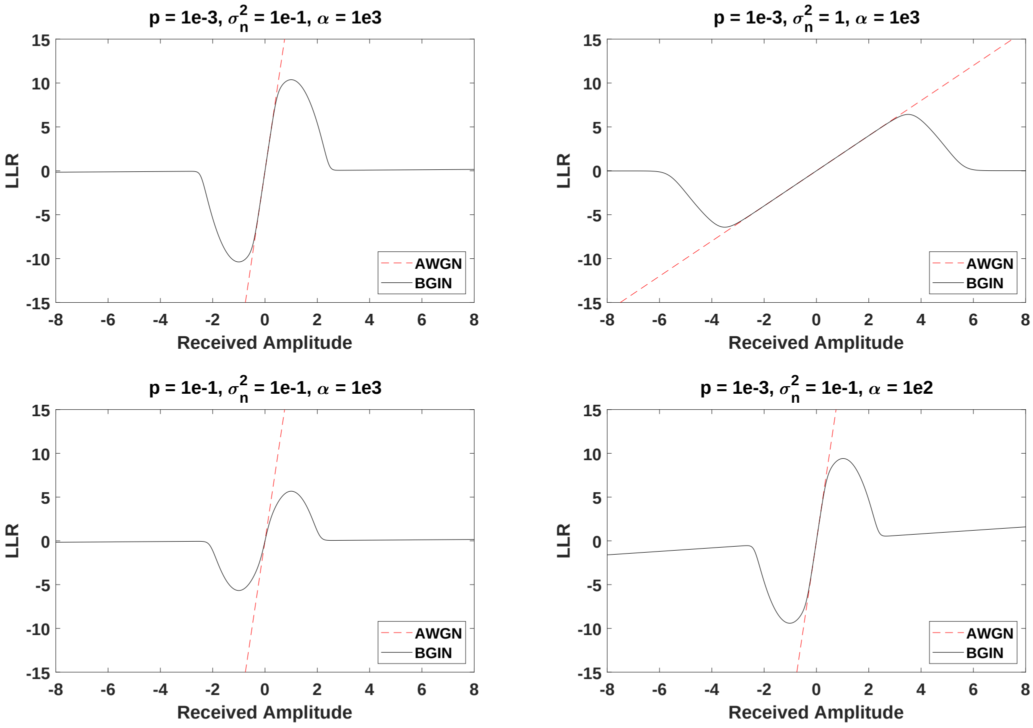


FIGURE 4.5: BPSK LLR legitimacy comparison for different BGIN parameter sets (p, σ_n^2, α).

4.3.2.2 Sum-Product Decoding

Based on the turbo coding terminology [246], the message-passing subroutine of the SP decoding algorithm may be interpreted as an information reinforcing and exchanging process. More specifically, the CND (VND) reads the *a priori* LLR inputs from the VND (CND) and produces *a posteriori* LLRs. The *newly gained* soft information characterized by the difference between the *a posteriori* LLRs and the *a priori* inputs corresponds to the *extrinsic* LLR outputs, which are fed into the VND (CND). To summarize, if all edges of the Tanner graph are simultaneously active for message passing (referred to as the flooding scheme in [240]), then the following LLR conversions may be defined for the VND and the CND, respectively:

$$L_{i,e} = \begin{cases} L_{\text{ch}} + \sum\limits_{i \neq j} L_{j,a} & \text{for VND,} \\ \sum\limits_{i \neq j} \boxplus L_{j,a} & \text{for CND,} \end{cases} \tag{4.4}$$

where $L_{i,e}$ and $L_{j,a}$ refer respectively to the ith extrinsic output LLR and the jth *a priori* input LLR associated with the specific node. Furthermore, L_{ch} refers to the demodulated channel LLRs, which are fed into the VND for each decoding iteration. The $\boxplus$ operator is defined according to [247] as follows:

$$\sum_{i \neq j} \boxplus L_{j,a} = \ln \left[\frac{1 - \prod\limits_{i \neq j} (1 - e^{L_{j,a}})/(1 + e^{L_{j,a}})}{1 + \prod\limits_{i \neq j} (1 - e^{L_{j,a}})/(1 + e^{L_{j,a}})} \right]. \tag{4.5}$$

In general, SP decoding approximates the MAP criterion, since the APP of a received codeword is formulated as *the sum of bitwise LLR products*. Since message-passing is a *local* operation, relying only on the edges that are directly connected to a particular VN (CN), the SP decoding algorithm may approach the MAP optimality, if and only if the specific LDPC code can be characterized by a special type of Tanner graph whose edges do not constitute cycles[3]. However, for the general class of LDPC codes including the NR LDPC code, SP decoding is capable of achieving good performance as long as the mutual information associated with the received signal exceeds the *decoding threshold*, a lower bound uniquely determined by the specific choice of the codebook and the decoding algorithm [249]. The same criterion also applies to the average-case SP decoding performance based on the *density evolution* [250] analysis of LDPC code ensembles.

4.3.2.3 Mutual Information Evolution

In [251], the classic semi-analytical tool known as the *extrinsic information transfer* (EXIT) chart was proposed for characterizing the average-case convergence behaviour of iteratively decoded graph codes. More specifically, the evolution of the *extrinsic* LLRs' mutual information (MI) I_E is recorded as a function of the input *a priori* MI I_A for a particular component decoder after each iteration of message-passing. For general LDPC codes, the EXIT chart may be portrayed from either the VND's perspective or the CND's. Based on (4.4) and on the base graph shown in Fig. 4.4, we may formulate the respective EXIT functions associated with the VND and with the CND as follows

---

[3] These proverbial 'cycle-free' Tanner graphs guarantee that the result of SP message-passing decoding is accurate and unambiguous, as discussed in [248]

for the family of classic LDPC codes [252]:

$$I_{E,\text{VND}} = J\Big\{\Big[(d_v-1)[J^{-1}(I_{A,\text{VND}})]^2 + [J^{-1}(I_{\text{ch}})]^2\Big]^{1/2}\Big\}, \tag{4.6}$$

$$I_{E,\text{CND}} = 1 - J\Big[(d_c-1)^{1/2}J^{-1}(1-I_{A,\text{CND}})\Big], \tag{4.7}$$

where I_{ch} is the MI associated with the sequence of demodulated channel LLRs, and the empirical J-function characterizes the mapping from the variance of LLRs to the corresponding MI, which is formulated according to [251] as follows:

$$I_\sigma = J(\sigma) = 1 - \int_{-\infty}^{\infty} \frac{e^{-(\xi-\sigma^2/2)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \log_2(1+e^{-\xi})\mathrm{d}\xi. \tag{4.8}$$

In practice, calculating (4.8) and its inverse function $J^{-1}(I_\sigma)$ is non-trivial. Therefore, it is more common to approximate the dual pair via a piecewise curve fitting technique, as conceived in the appendix of [252].

As we shall see in Sec. 4.3.2.4, Eq. (4.6) and (4.7) do not accurately characterize the concurrent message-passing processes associated with the NR LDPC Tanner graph portrayed in Fig. 4.4. In particular, it may be observed from Fig. 4.4 that the shared set of VNs (of the core LDPC and LDGM region) carries information across the adjacent decoding regions. Since the pair of message processing operations defined in (4.4) has to treat all incoming messages (*a priori* LLRs) as the list of input arguments, simply applying (4.6) and (4.7) to the core LDPC and LDGM code region respectively will not yield the correct amount of *extrinsic* MI for either region. On the other hand, if we represent the decoding process as a single Tanner graph by viewing the concatenated PCMs of Fig. 4.3 as a single PCM, both (4.6) and (4.7) will become invalid, since the presence of the all-zero region and the information transfer region results in the undesired consequence that the core LDPC and LDGM code regions lose their designated codebook properties. In fact, the information transfer region indirectly assigns channel information to the LDGM CNs. Based on these observations, we circumvent the above undesired consequences by proposing the following group of equations for characterizing the message-passing process of the NR LDPC decoder:

$$\begin{aligned} I_{E,\text{VND,LDPC}} = J\Big\{\Big[&(d_{v,\text{LDPC}}-1)[J^{-1}(I_{A,\text{VND,LDPC}})]^2 + \\ &d_{v,\text{LDGM}}[J^{-1}(I_{A,\text{VND,LDGM}})]^2 + [J^{-1}(I_{\text{ch}})]^2\Big]^{1/2}\Big\}, \end{aligned} \tag{4.9}$$

$$I_{E,\text{CND,LDPC}} = 1 - J\Big[(d_{c,\text{LDPC}}-1)^{1/2}J^{-1}(1-I_{A,\text{CND,LDPC}})\Big], \tag{4.10}$$

$$I_{E,\text{VND,LDGM}} = J\Big\{\Big[(d_{v,\text{LDGM}} - 1)[J^{-1}(I_{A,\text{VND,LDGM}})]^2 + d_{v,\text{LDPC}}[J^{-1}(I_{A,\text{VND,LDPC}})]^2 + [J^{-1}(I_{\text{ch}})]^2\Big]^{1/2}\Big\}, \tag{4.11}$$

$$I_{E,\text{CND,LDGM}} = 1 - J\Big\{\Big[(d_{c,\text{LDGM}} - 1)[J^{-1}(1 - I_{A,\text{CND,LDGM}})]^2 + [J^{-1}(1 - I_{\text{ch}})]^2\Big]^{1/2}\Big\}. \tag{4.12}$$

Considering the shared set of VNs, both (4.9) and (4.11) indicate a pair of mirrored message-passing processes associated with the same three sources of *a priori* input, i.e. the pair of *extrinsic* outputs from both sets of CNs and the output from the demodulator. On the other hand, the message-passing behaviors of the CNs are rather distinct. The core LDPC CNs' behaviors (4.10) follow the classic formulation of (4.7), while the LDGM CNs take the *extrinsic* output both from the LDGM VNs and from the VNs associated with the information transfer region. Therefore, the LDGM CNs should additionally account for the demodulated channel information as formulated in (4.12). The *a priori* and *extrinsic* MI associated with the extension VNs of the transfer region is neglected here, since the extension parity bits have no direct influence on the recovery of the information bits, unlike the core parity bits. Given the revised set of MI evolution functions (4.9) to (4.12), we may now graphically characterize the standard decoding of the NR LDPC codes with the aid of a more accurate portray of the concurrent message-passing processes, as shown in Fig. 4.6. Furthermore, we complement the analysis of the NR LDPC codes with our novel 3D EXIT chart design.

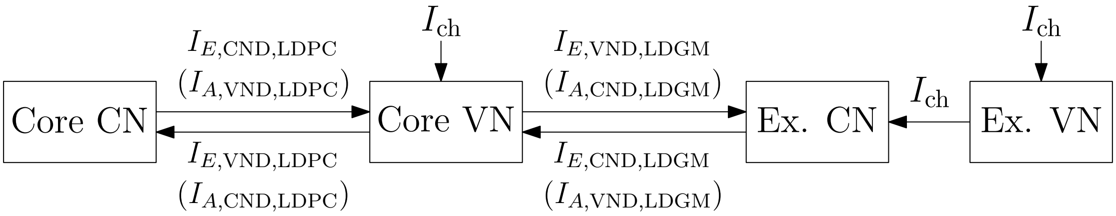


FIGURE 4.6: The concurrent message-passing processes associated with the SP decoding of the NR LDPC codes. The notations correspond to those of Eq. (4.9) to (4.12) and to the core checks and extension checks of Fig. 4.3

4.3.2.4 EXIT Chart Analysis of the NR LDPC Code

For a conventional LDPC coding scheme, since I_E and I_A chain each other in the iterative decoding process, the EXIT functions associated both with the CND and with the VND may be plotted in a single EXIT chart [251]. The area between the pair of EXIT curves constitutes a *tunnel*. It was shown in [252] that, provided the pair of EXIT curves closely matches each other and hence has a small open-tunnel area, then the corresponding LDPC code ensemble is of capacity-approaching nature, regardless of the type of channel. Furthermore, as long as the open EXIT tunnel extends to the $(1, 1)$ point of the EXIT chart, we will have an infinitesimally low probability of bit errors. This is known as the *open tunnel condition* of EXIT chart analysis [253]. Intuitively, the EXIT curves characterize the average-case MI evolution trajectory, hence the per-iteration MI approximately evolves in a staircase shape between the EXIT curves. Therefore, an open EXIT tunnel allows the MI associated both with the VND and with the CND to approach the $(1, 1)$ point. By contrast, if the pair of EXIT curves cross over before reaching the $(1, 1)$ point, the tunnel is said to be closed, which results in an error-floor.

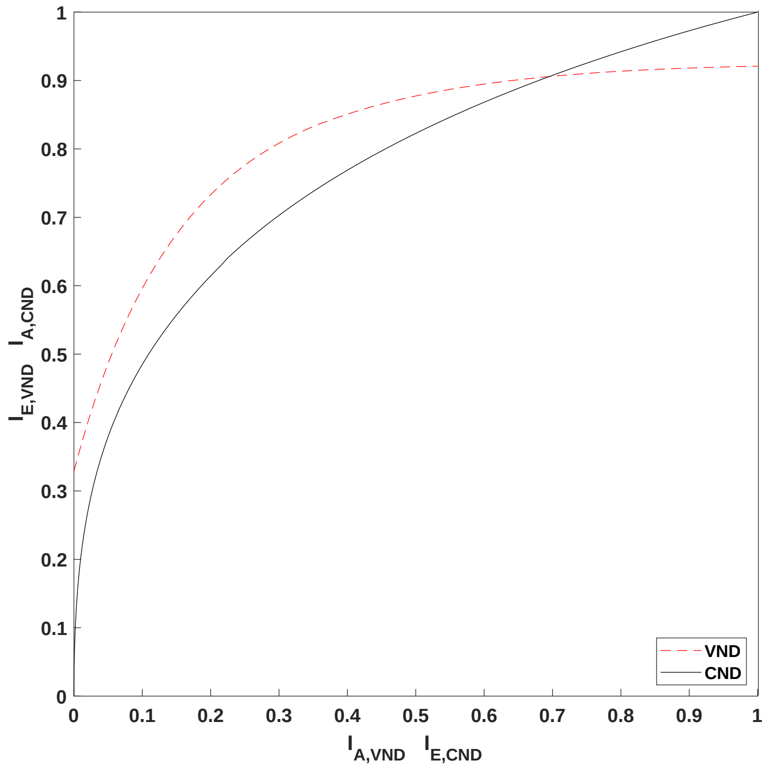


FIGURE 4.7: Conventional 2D EXIT chart of the NRLDPC, treating the entire PCM as a cohesive codeword. At SNR = 0 dB in an AWGN channel, the SP decoder can readily achieve an arbitrarily small probability of error with a large number of decoding iterations, yet the EXIT tunnel remains closed.

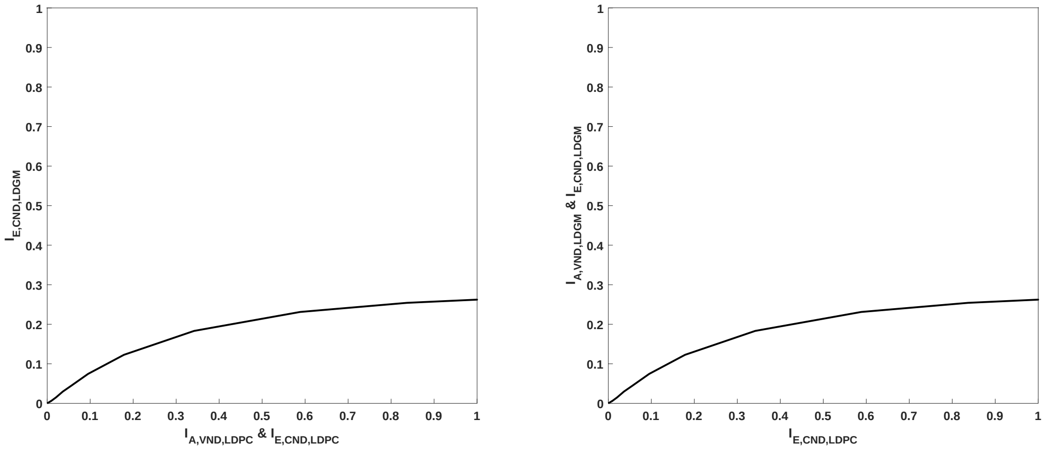
Due to the particular shape of the PCM of the NR LDPC code seen in Fig. 4.3, the conventional EXIT chart is not capable of characterizing the concurrent message-passing processes shown in the Tanner graph of Fig. 4.4. Specifically, the VND and CND EXIT curves cross over before reaching the $(1, 1)$ point, hence no open tunnel is observed in the conventional EXIT chart [251], as seen in Fig. 4.7. Therefore, special types of EXIT charts are required for accurately characterizing the average MI evolution trajectories associated with the decoding of the NR LDPC codes.

According to the information exchange diagram of Fig. 4.6, the pair of message-passing blocks associated with the core LDPC region and with the LDGM region actually affect each other. Naturally, if we treat both concurrent subroutines jointly, then the evolution of four MI variables formulated in (4.9) to (4.12) constitutes an EXIT chart in 4D. To illustrate this graphically, we characterize their pairwise relationships in a group of six decoupled 2D EXIT charts first, by considering the respective *predicted* decoding MI evolution trajectories, as shown in Fig. 4.8, where the chosen NR LDPC code has a coding rate of 1/3 and a block length of 450. In particular, at SNR$=0$ dB, Fig. 4.8(c) suggest that both the VNs and the CNs associated with the core LDPC region may output the maximum *extrinsic* MI within a given number of decoding iterations, thus reaching the $(1,1)$ point of the EXIT chart. By contrast, the *extrinsic* MI $I_{\text{E,CND,LDGM}}$ gleaned from the LDGM CNs does not reach 1, which is due to the influence of the VNs associated with the extension parity columns seen in Fig. 4.3. Since the information-bearing section of a NR LDPC code block only exists in the information columns shared by the core LDPC and LDGM region, having the *extrinsic* MI of (4.9), (4.10) and (4.11) growing to $(1,1,1)$ constitutes sufficient condition for achieving an arbitrarily small block error rate (BLER).

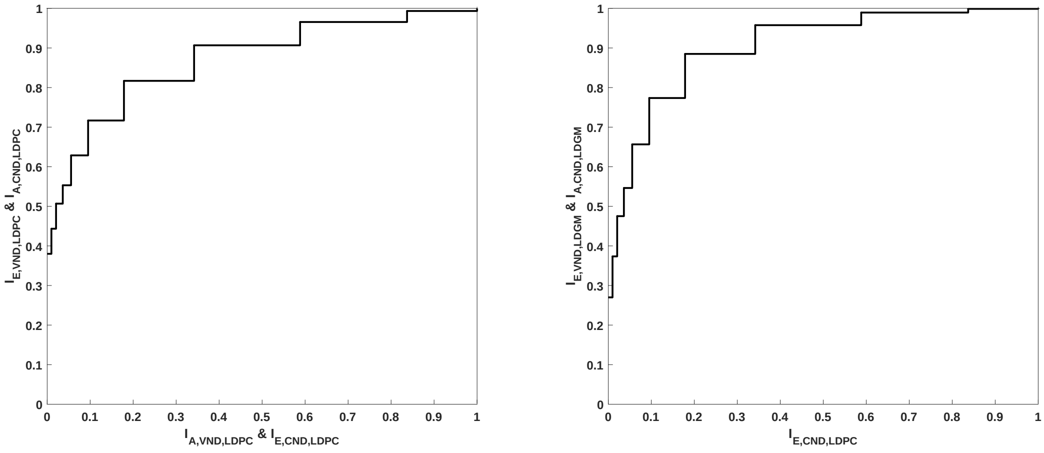
To facilitate a compact representation of the holistic decoding process, we lift the 2D projections portrayed in Fig. 4.8 into a pair of 3D EXIT charts shown in Fig. 4.9, incorporating the internal message-passing processes associated with both the core LDPC region and with the LDGM region of the NR LDPC code, as well as the additional *extrinsic* MI gleaned from each other. In Fig. 4.9(a) and 4.9(b), we portray the condition of SNR$=-2$ dB for the same coding parameters, where the pair of 3D EXIT tunnels is jointly closed. In particular, we note that the *extrinsic* MI is still increased obeying a staircase shaped function, approximately bounded by the pair of EXIT surfaces. Each of the four EXIT surfaces corresponds respectively to the average *extrinsic* MI evolution associated with the formulations of (4.9) to (4.12). However, the predicted decoding MI evolution trajectories associated with both regions terminate prematurely. As discussed earlier, $I_{\text{E,CND,LDGM}}$ is the only exception amongst the four MI variables of the NR LDPC decoder that fails to reach 1, and thus it fails to satisfy the open tunnel condition. Therefore, as we can see in the case of a jointly-open EXIT tunnel shown in Fig. 4.9(c) and 4.9(d) at SNR$=-2$ dB, the predicted MI evolution trajectories are capable of approaching the respective $(1,1)$ point, when observing them along the axis corresponding to $I_{\text{E,CND,LDGM}}$.

4.4 NR LDPC-Coded HARQ

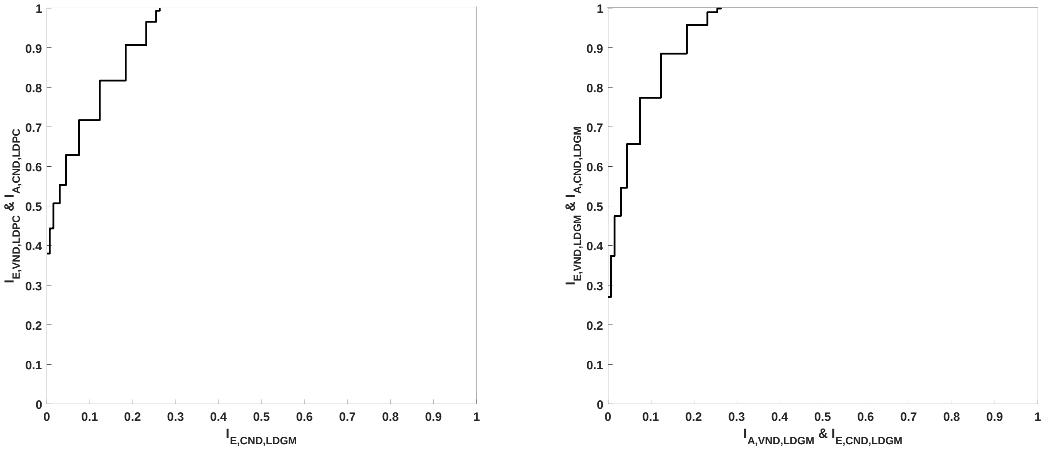
The practical design of our NR LDPC-coded HARQ will have to strike an attractive trade-off between the achievable BLER, decoding latency and bandwidth efficiency. In particular, HARQ may jointly optimize both the NR LDPC decoding operations and the retransmission mechanism in order to maximize the coding gain of a particular NR LDPC codebook under fixed delay, bandwidth and power consumption requirements.



(a) LDPC (b) LDGM

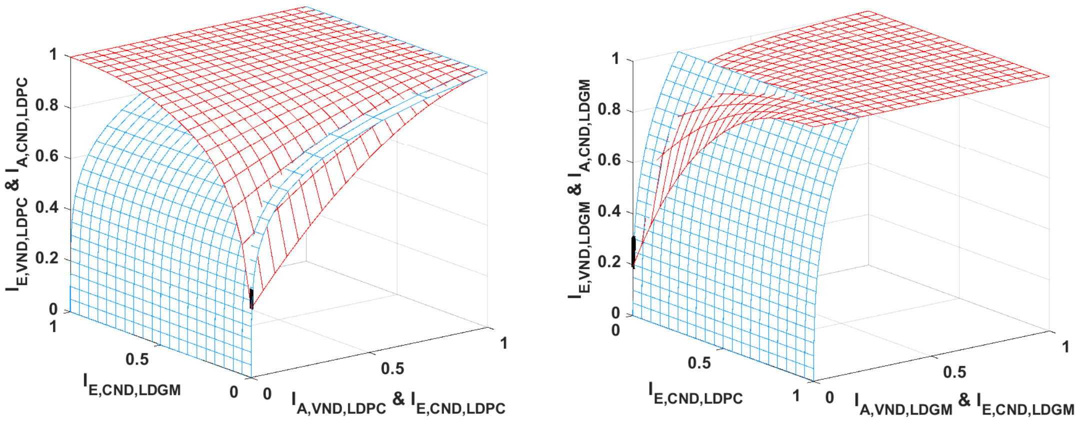


(c) LDPC (d) LDGM

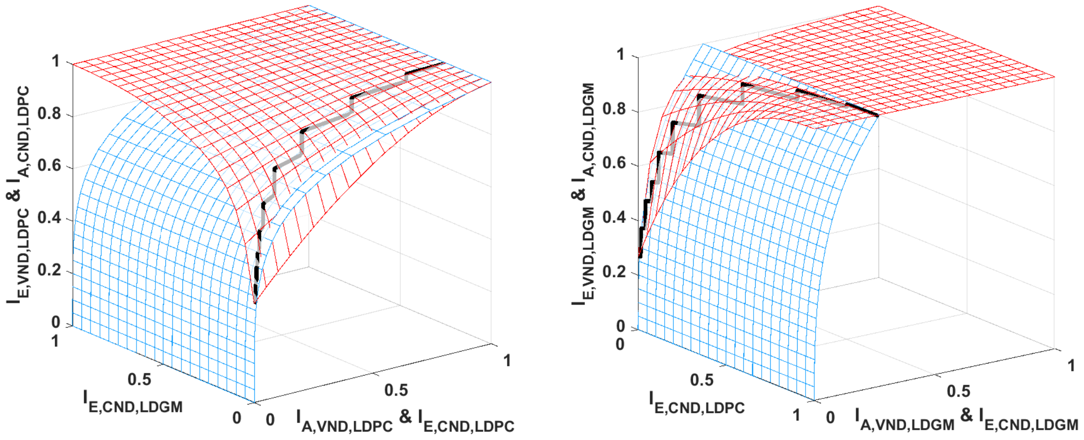


(e) LDPC (f) LDGM

FIGURE 4.8: Projections of the 3D MI evolution of Fig. 4.9 on the 2D planes. Observe that the *extrinsic* MI from the LDGM CNs is the only case of not approaching 1, due to the influence from the VNs associated with the extension parity columns of Fig. 4.3



(a) LDPC (SNR = −2 dB, tunnel closed) (b) LDGM (SNR = −2 dB, tunnel closed)



(c) LDPC (SNR = 0 dB, tunnel open) (d) LDGM (SNR = 0 dB, tunnel open)

FIGURE 4.9: 3D EXIT charts of the core LDPC and of the LDGM region. The top pair shows a jointly-closed 3D EXIT tunnel in terms both of the core LDPC region and of the LDGM region, whilst the bottom pair shows a jointly-open 3D EXIT tunnel. Generally, the open tunnel condition for the NR LDPC codes, which is of 4D nature in our EXIT chart analysis, has to be fulfilled for the joint EXIT tunnel of both regions.

In the following sections, we recall the classic approaches used in HARQ protocols and then propose a novel low-complexity HARQ design for our NR LDPC-coded wireline systems. The complexity reduction strategies follow the design principles conceived for turbo-coded LTE systems [236]. The benefit of the strategy employed is two-fold. If the combined code block stored in the receiver buffer $\mathcal{B}_{\text{ext}}$ after a particular (re)transmission is deemed by the 3D EXIT chart to result in a closed EXIT tunnel, then we defer the iterative decoding process until either the tunnel becomes open as a benefit of receiving more redundancy or the final retransmission is received, so that the complexity associated with unsuccessful decoding attempts is mitigated. On the other hand, a pair of MI threshold conditions is checked during each iteration of the message-passing process in order to curtail a decoding process at the earliest instant that is likely to fail. Additionally, an internal buffer $\mathcal{B}_{\text{int}}$ is assigned to the NR LDPC decoder for storing the internal states of the VND and of the CND, when a NACK is triggered, so that the decoding of future retransmissions may be beneficially accelerated.

4.4.1 HARQ Classification

The goal of HARQ is to exceed the performance limit of using either FEC or ARQ in isolation. As a further benefit, the optimal HARQ strategy is also capable of adapting to the time-variant channel environment. Similar to turbo codes, the NR LDPC codes have three operational regions: 1) the low-SNR region, where the errors are unlikely to be corrected; 2) the medium-SNR *waterfall* region, where the success probability of FEC decoding drastically increases as a function of SNR, and 3) the high-SNR region, where the BLER becomes vanishingly low. As a result, the best HARQ strategy allows the embedded NR LDPC code to operate close to the waterfall region, i.e. at the lowest possible SNR, while also eliminating the error floor in the high-SNR region. This design philosophy resulted in three classic HARQ categories:

- **Conventional HARQ**: More commonly seen in the early development of HARQ in the 1970s, where the goal has been to simply keep retransmitting identical copies of the original code block containing the same information plus FEC redundancy, until either the CRC test passes or the maximum retransmission limit is reached [254]. In contrast to the HARQ depicted in Fig. 4.2, the receiver conceived for a conventional HARQ protocol operates in a memoryless manner, without writing to or reading from an external buffer $\mathcal{B}_{\text{ext}}$. Therefore, each received code block is decoded independently. This approach was shown to achieve only modestly higher throughput than the following more advanced approaches according to [236].

- **Chase-combining** aided HARQ (CC-HARQ)[255]: In order to glean useful information from failed retransmissions, CC-HARQ introduced $\mathcal{B}_{\text{ext}}$ for storing past retransmissions of the same code block. For soft-output FEC decoders, the LLRs associated with each code block received during retransmissions are added to $\mathcal{B}_{\text{ext}}$. This particular strategy resembles the on-demand repetition coding philosophy, which essentially improves the channel SNR under the MRC principle. As a benefit, CC-HARQ avoids some of the coding loss associated with conventional repetition coding, since extra 'redundancy' is only incorporated during retransmissions as and when needed. However, the redundancy associated with the embedded FEC is still required. As a result, CC-HARQ is capable of boosting the performance of FEC at low channel SNRs. Owing to its similarity to Chase-combining aided repetition coding, this specific HARQ strategy also relies on the *repetition of redundancy* [256].

- **Incremental-redundancy** aided HARQ (IR-HARQ)[257]: Under high channel SNR conditions, FEC is no longer necessary for passing the CRC test. In order to minimize the potential loss of bandwidth efficiency, a minimum amount of FEC-related redundancy (or no FEC at all as proposed in [258]) is attached to the first transmission. If the initial attempt failed, then a new code block containing *additional redundancy* is sent. The redundancy sent during each retransmission is selected from a deterministic sequence of punctured sub-blocks of the full-length code block, as discussed earlier in Sec. 4.3.1. These *indexed sub-blocks* are termed as *redundancy versions* in the NR standard [237]. The sequence of redundancy versions is known at both sides of the channel. IR-HARQ always attempts to use the minimum amount of FEC redundancy first, therefore its benefits become most prevalent in the high channel SNR region.

In recent industrial applications such as in NR LDPC-coded systems, CC and IR are jointly implemented via puncturing the PCM of Fig. 4.3 and combining the received sub-blocks. More specifically, a CRC-encoded information block is first padded with zeros so that the length of the resultant bit sequence matches the PCM's input size requirement. This process is termed as *filler bit insertion* in the NR standard [237]. Based on the full-length code block at the output of the NR LDPC encoder, the transmitter selects the current redundancy version to be transmitted with the aid of a circular buffer according to the predefined sequence. The size of the circular buffer is chosen to be $N = (N_v - 2) \times Z_c$ based on the values in Fig. 4.3. In particular, the first transmission of length E_r typically contains all the information bits (minus those associated with the first two information columns [240]), the CRC and core LDPC parity bits. For each subsequent retransmission, the circular buffer starts from the current redundancy version's starting position (Fig. 4.10) and extends around the buffer by E_r positions. Whenever the right side of the buffer shifts out of bound, the left side of the PCM is recycled, again dispensing with the two starting information columns. The four redundancy versions specified in the NR standard are transmitted periodically following the default order of $0, 2, 1, 3, 0, ...$, with RV_0 being the self-decodable code block. We reproduce a demonstration of the circular buffer's operation from the NR standard in Fig. 4.10, where the four RV markers denote the starting bit position of each indexed redundancy version. b_0 represents the starting bit position of the full-length code block, leaving a $2Z_c$-bit gap from the starting bit position of RV_0.

4.4.2 Complexity Reduction for HARQ

The HARQ protocol has a major impact on the throughput-centric *enhanced mobile broadband* (eMBB) operating mode of 5G NR. Naturally, metallic access networks contaminated by hostile IN also require a reliable and efficient HARQ protocol for error control. In this section, we will design efficient complexity reduction techniques for the NR LDPC-coded HARQ protocol (Fig. 4.11). More specifically, since the asymptotic average performance of the NR LDPC codes may be accurately characterized by EXIT

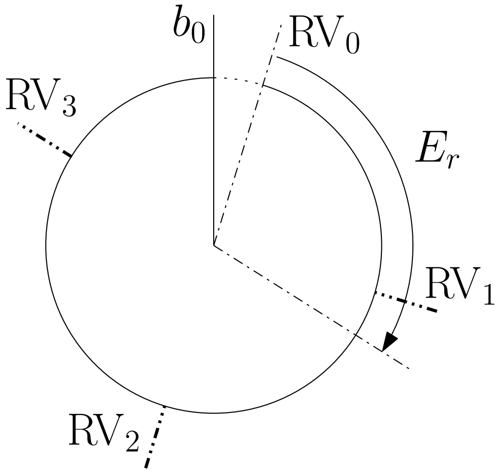


Figure 4.10: Circular buffer aided redundancy version (RV) selection in NR LDPC coding [237]. The circle represents the entire codeword, which starts at b_0. The first RV of length E_r starts two information columns after b_0 on the base PCM of Fig. 4.3. Subsequent retransmissions follow a default order of $\text{RV}_0, \text{RV}_2, \text{RV}_1, \text{RV}_3, \text{RV}_0, \ldots$, while always skipping the first two information columns in the original full-length codeword.

analysis, we may reduce the overall complexity of HARQ by (probabilistically) avoiding redundant or futile decoding operations.

4.4.2.1 Deferred Iteration Based HARQ

The performance of HARQ is typically quantified in terms of the SNR gain associated with the retransmissions of the same information together with its parity[4]. As a result, the SP decoder may theoretically achieve arbitrarily low BLER at any channel SNRs, as long as the total number of permitted retransmissions per information block is unbounded. With the aid of such an idealized *lossless HARQ* [259] protocol, the channel SNR may be boosted indefinitely until the combined SNR falls in the decoder's waterfall region. Since the NR LDPC-coded HARQ protocol relies on a hybrid of CC and IR, the benefit of retransmission comes both from channel LLR accumulation (CC) and from the adaptation of the concatenated coding rate (IR), the latter of which may be interpreted as follows. Due to the punctured code structure associated with each redundancy version, the average degree of the (punctured) PCM is an increasing function of the number of transmitted redundancy versions, *up to* a full cycle of four (Fig. 4.10). From this point onwards, the degree distribution becomes fixed for subsequent retransmissions, if the length E_r of each redundancy version remains constant.

Attempting to decode a single code block at SNRs below the waterfall region will have a high probability of failing. Hence, for the sake of overall complexity reduction, decoding attempts below the waterfall-SNR should generally be avoided, which may be achieved

---

[4] In contrast to wireless fading channels, combining the retransmitted code blocks sent over static metallic channels does not acquire diversity gain when relying on the fixed interleaving strategy defined in [237], but it may succeed in avoiding particularly hostile instances of the IN.

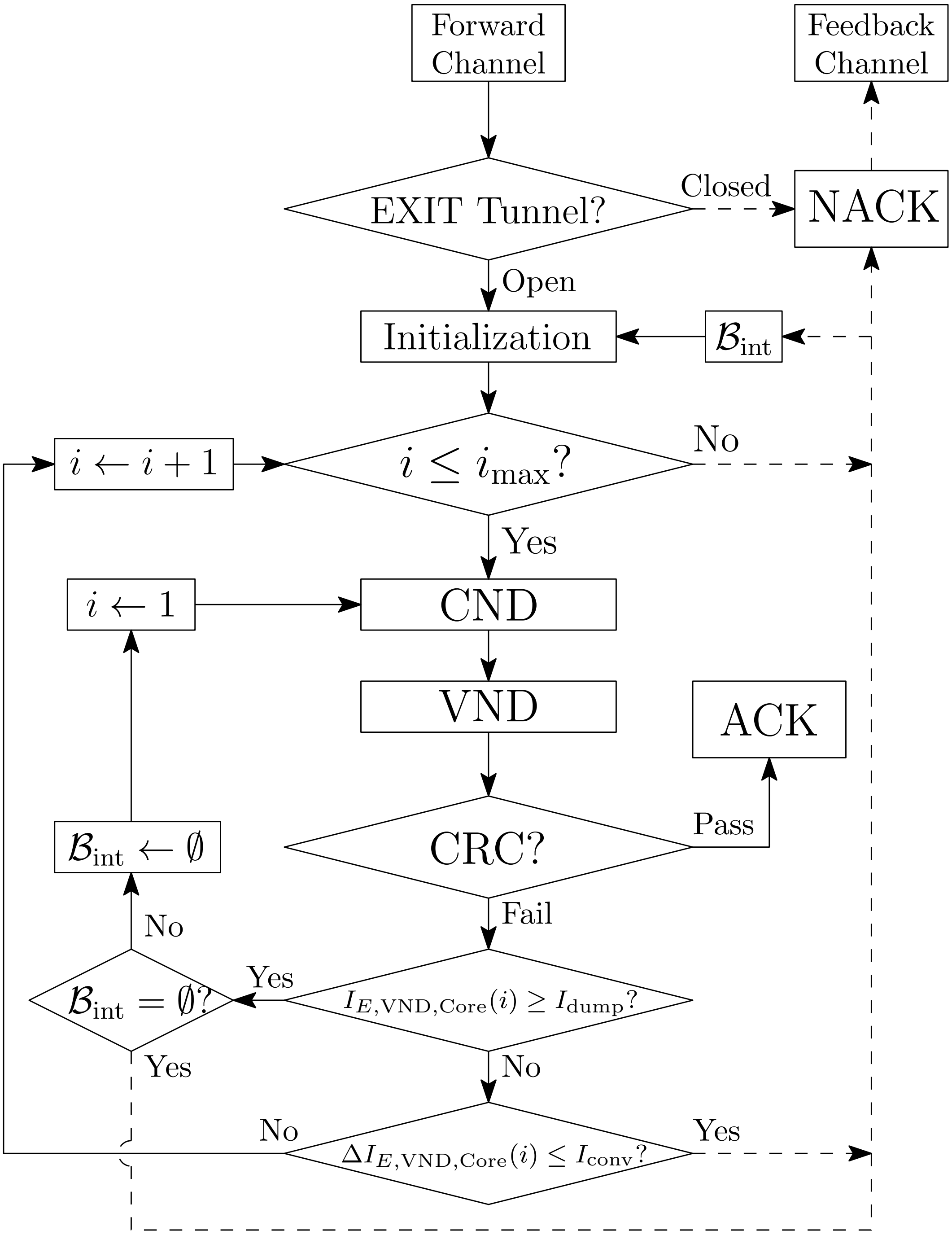


FIGURE 4.11: Single pair error control scheme based on the NR LDPC-coded HARQ protocol. The complexity-QoS trade-off is optimized with the aid of both deferred iteration (DI) and curtailed iteration (CI) strategies. Maximum iteration count is $i_{\max}$ while $\mathcal{B}_{\text{int}}$ denotes the *internal* buffer of the NR LDPC decoder. The *external* buffer $\mathcal{B}_{\text{ext}}$ portrayed in Fig. 4.2 is not shown explicitly here.

with the aid of a 'switch' controlling the SP decoder's activation status. Motivated by the EXIT analysis of the NR LDPC codes in Sec. 4.3.2.4, it was established that the waterfall-SNR essentially corresponds to the SNR, where an open tunnel in the core LDPC EXIT chart emerges (Fig. 4.8(c)). Let us now consider the operation of the *deferred iteration* (DI) aided decoder switch as follows. Upon the reception of each redundancy version, we check whether the accumulated SNR calculated by taking into account all previous redundancy versions stored in $\mathcal{B}_{\text{ext}}$ is deemed sufficient for meeting the open tunnel condition of the NR LDPC code. If this is the case, the SP decoder is activated immediately and remains activated for all future retransmissions of the current information block. Otherwise, namely if the accumulated SNR of $\mathcal{B}_{\text{ext}}$ is deemed to be insufficient for satisfying the open tunnel condition, then the SP decoder is deactivated and *the decoding iterations are deferred* until either the EXIT tunnel becomes open or the maximum retransmission limit is reached.

Since the EXIT chart only characterizes the average decoding performance, having the DI switch may degrade the attainable goodput. More particularly, when a marginally closed EXIT tunnel is observed after receiving a specific redundancy version, SP decoding may still have a moderate probability of success for this particular code block. Therefore, in order to prevent the DI switch from triggering redundant retransmissions when the EXIT tunnel is only marginally closed, a MI calibration step is introduced because the average SNR is below the tunnel-opening SNR, but the new instantaneous SNR may be above it. Specifically, since the DI switch assesses the tunnel-opening condition based on the iterative calculation of the *extrinsic* MI using the initial input MI associated with the channel LLRs, we calibrate the requirement for the minimum initial input MI by a simple remapping, as follows.

Specifically, the input MI calibration is carried out by matching the theoretical input MI threshold to the thresholds found by Monte Carlo simulation for each number of retransmissions. As a result, the decoder may become activated in the presence of a marginally open EXIT tunnel. As shown later in Fig. 4.12, such a calibration step essentially finds the minimum required input MI for an exact integer-valued number of transmission attempts. Subsequently, by EXIT analysis, we find the amount of channel input MI associated with the same *estimated* number of transmission attempts. Subtracting the offset between the pair of input MI will guarantee that the calibrated prediction of the number of transmission attempts $\tilde{N}_r$ satisfies $\tilde{N}_r = \lfloor N_r \rfloor$, which is the closest integer of the simulated average number of retransmissions N_r from below.

4.4.2.2 Early Termination and Curtailed Iteration

The DI strategy we have discussed so far mainly aims for reducing the computational complexity averaged over all (re)transmissions from an ARQ perspective. When viewed from an FEC perspective, the complexity of the SP message-passing decoding of LDPC codes is a function of both the PCM degree distribution and of the number of decoding iterations. As we have presented at the beginning of Sec. 4.4.2.1, the PCM of the NR LDPC code has a variable degree distribution following the first couple of (re)transmissions. Therefore, this complexity contribution is fixed under a particular set of codec and redundancy version specifications, regardless of the near-instantaneous channel noise condition of the HARQ system.

On the other hand, optimizing the number of decoding iterations has also been a predominant research topic in the context of graph codes [236], particularly for the family of turbo codes due to its extensive use in the 3G and 4G cellular standards. Without loss of generality, the most promising complexity-reduction and power-saving contributions mainly accrue from having an adaptive number of decoding iterations [260, 261], MI thresholding [262] and iterative CRC test [263]. These strategies may be classified into two broader categories, depending on whether the decoder is failing or not. More specifically, we shall use early termination (ET) to refer to the scenarios, where the decoder succeeds earlier than expected. By contrast, the curtailed iteration (CI) terminology will refer to the scenarios, where the decoder curtails a futile decoding attempt.

Following these criteria, the adaptive number of iteration based strategies may actually be interpreted as generalizations of DI. For these strategies, the number of decoding iterations required is associated with a list of channel SNRs [260, 261]. The MI-thresholding based strategies also aim for CI by actively monitoring the MI evolution status per iteration, and subsequently ending the decoding process if the *incremental MI* between two consecutive iterations falls below a given threshold [262]. On the other hand, the iterative CRC test strategy typically belongs to the ET category, which is highlighted as follows. The iterative decoder produces hard-decision outputs from the *a posteriori* LLRs for each iteration, which is immediately subjected to the CRC test [263][5]. Naturally, the iterative decoding process is terminated as soon as the CRC test passes. In the following sections, we shall consider ET as the decoder's default operational mode.

4.4.2.3 Curtailed Iteration Design

In NR LDPC coding, the main decoder output of interest is constituted by the *a posteriori* LLRs from the VNs corresponding to the information and core parity columns of Fig. 4.3. Correspondingly, the MI convergence threshold I_{conv} shall be based on the *a posteriori* MI associated with the core VNs. For the convenience of implementation, I_{conv} will be configured according to the corresponding *extrinsic* MI $I_{\text{E,VND,Core}}$. Since the decoding switch requires an offline training in order to acquire the EXIT tunnel status associated with the reception of each redundancy version, the MI convergence threshold may be initialized concurrently based on the predicted MI evolution in the EXIT anlysis (e.g.

---

[5] Alternatively, the intermediate hard-decision result may be re-encoded and compared to the received code block for an alternative way of information consistency verification [264].

Fig. 4.9(c)). Specifically, when an open EXIT tunnel is detected upon the reception of a particular redundancy version under the given channel conditions, the increment of $I_{\mathrm{E,VND,Core}}$ between the two most recent iterations is recorded as I_{conv}.

In order to further optimize the efficiency of the NR LDPC-coded HARQ protocol, an internal buffer $\mathcal{B}_{\mathrm{int}}$ of Fig. 4.11 is assigned to the NR LDPC decoder. Following each unsuccessful decoding attempt, a NACK flag is sent to the transmitter, while all current *extrinsic* LLRs are written into $\mathcal{B}_{\mathrm{int}}$. At the start of the next decoding attempt, the decoder proceeds from the decoding state stored in $\mathcal{B}_{\mathrm{int}}$ while reading new channel LLRs from $\mathcal{B}_{\mathrm{ext}}$ of Fig. 4.2. Since the MI of $\mathcal{B}_{\mathrm{int}}$ is higher than the zero MI of a fresh initial state, decoding from $\mathcal{B}_{\mathrm{int}}$ facilitates a 'boosted start'.

Generally, the $\mathcal{B}_{\mathrm{int}}$ of Fig. 4.11 only becomes useful when the DI switch mistakenly activates the NR LDPC decoder. This happens when a marginally closed EXIT tunnel is detected as actually being open or when the received redundancy version is contaminated by IN. In this case, the decoder operates in the low-SNR region, which may lead to 'false belief', which is an inherent limitation of the SP (i.e. belief propagation) decoding algorithm, when invoked by HARQ. In particular, since the MI is only an indicator of the *confidence level* regarding the immediate polarity (i.e. hard-decision output) of the *a posteriori* LLRs, it is susceptible to making confident, yet incorrect, decisions. This is readily understood, if we consider the entire HARQ process as a single long iterative decoding session spanning multiple retransmissions. In particular, the role of $\mathcal{B}_{\mathrm{int}}$ in Fig. 4.11 may essentially be viewed as connecting the disjoint decoding attempts of different retransmissions. Moreover, this holistic decoding process is implemented with the aid of gradually improved channel LLR injection to the VND following the reception of each additional redundancy version. However, when the initial channel LLRs are received either under poor SNR conditions or in the face of IN contamination, they may prevent the SP decoder from correctly decoding the original information bits. Specifically, the *extrinsic* LLRs may now converge to an incorrect bit sequence output, despite potentially having improved channel LLR injections during later decoding iterations.

In order to prevent false beliefs, another MI thresholding strategy is employed for clearing the $\mathcal{B}_{\mathrm{int}}$ of Fig. 4.11. More specifically, the dumping threshold I_{dump} is compared to $I_{\mathrm{E,VND,Core}}$ following a failed CRC test, and the SP decoder declares encountering a false belief, when $I_{\mathrm{E,VND,Core}} \geq I_{\mathrm{dump}}$. If $\mathcal{B}_{\mathrm{int}}$ is not empty, its stored *extrinsic* LLRs are dumped and a fresh decoding attempt is made. If $I_{\mathrm{E,VND,Core}} \geq I_{\mathrm{dump}}$ while the contents of $\mathcal{B}_{\mathrm{int}}$ have already been dumped, dumping again will not provide any benefits for the subsequent decoding attempts, until the channel LLRs are updated following the next retransmission. As a result, CI is triggered and the new decoding state is written into $\mathcal{B}_{\mathrm{int}}$.

4.5 Performance Evaluation

The performance of the NR LDPC-coded low-complexity HARQ scheme is characterized in Fig. 4.11 in terms of the BLER, goodput and decoding complexity. For the sake of consistency, we compare multiple block lengths and coding rates transmitted over the same single-pair DSL line. The exact simulation parameters are listed in Tab. 4.1 following the conventions of the NR standard [237], where applicable.

TABLE 4.1: System parameters of the NRLDPC-coded HARQ system

Parameter	Short length	Medium length	Low coding rate
K'	200	400	200
E_r	240	480	300
Base Graph	2		
Q_m	2		
CRC	$z^{24}+z^{23}+z^{6}+z^{5}+z+1$		
$i_{\max}$	10		
I_{conv}	0.002 (Default)		
I_{dump}	0.99 (Default)		
Retransmission limit	8		
Bandwidth	420 MHz		
AWGN floor	-150 dBm/Hz		
Cable length	50 m		
IN parameter (p, α)	$(0.01, 10^3)$		

4.5.1 Goodput

For any combination of FEC, ARQ and channel/SNR condition, the *Goodput* η of the corresponding HARQ system may be formulated as:

$$\eta = \frac{K'Q_m}{E_r N_r}(1 - \text{BLER}), \tag{4.13}$$

where K' and E_r are the total number of information (and CRC) bits and the number of coded bits per redundancy version, respectively. Their ratio K'/E_r may be interpreted as the conventional coding rate. Furthermore, N_r represents the average number of retransmissions required per code block, thus accounting for the *effective* coding rate, while Q_m is defined as the number of bits per symbol. In contrast to an idealized system, where N_r is unbounded, the goodput of a typical HARQ system having a limited number of retransmissions is reduced owing to having a non-zero BLER.

Fig. 4.12(a), 4.12(c) and 4.12(e) depict the goodput of the three coding configurations of Tab. 4.1, based on the combination of CI and DI strategies. Due to the time invariant nature of the metallic channel, HARQ relying on interleaving is unable to glean diversity gain. As a result, the corresponding goodput increases following a staircase-shaped function, resulting in a limited range of preferable operating SNRs, since the increase of SNR does not necessarily improve goodput. Otherwise, the four combinations of CI and DI strategies have a negligible effect on the goodput, except at high SNRs, where CI occasionally prevents extra iterations, hence sometimes causing an extra retransmission. When employing the standard AWGN-based soft demodulation rather than the IN-based model of (4.1), we may observe that the respective goodput performances under the 'Gaussian' condition in Fig. 4.12, dispensing with DI and CI, become significantly worse at low to medium SNRs. Due to the strict ATP and TxPSD constraints investigated earlier in Ch. 3, we may see that employing IN-based soft-demodulation is more preferable in practical systems.

4.5.2 Complexity

The complexity of SP decoding is quantified in terms of the number of add-compare-select (ACS) operations per iteration and the number of decoding iterations. Moreover, decoding the retransmitted redundancy versions constitutes additional complexity, which should also be averaged over the number of retransmissions. Therefore, the average complexity of HARQ-aided standard SP decoding may be formulated as follows according to the operations of Eq. (4.4):

$$\mathcal{O} = \frac{1}{N_r} \sum_{r=1}^{N_r} i_r \left[\sum_{d_v=1}^{V} N_{d_v}^{(r)} d_v^{(r)} + \sum_{d_c=1}^{C} N_{d_c}^{(r)} (d_c^{(r)} - 1) \right], \tag{4.14}$$

where i_r is the number of decoding iterations for the rth retransmission, while V and C represent the highest observable degree of the VNs and the CNs, respectively. For conventional LDPC code designs outside the NR family, the degree distribution d_v/d_c and the number of degree-d_v/d_c nodes N_{d_v}/N_{d_c} constitute a pair of constant parameters for each retransmission for CC-HARQ based systems. As mentioned in Sec. 4.4.2.1, due to the hybrid implementation of CC and IR strategies, NR LDPC-coded HARQ protocols actually have variable degree distribution for the first few retransmissions. Specifically, when decoding the accumulated information stored in $\mathcal{B}_{\text{ext}}$ of Fig. 4.2 from the first few retransmissions, some VNs and CNs associated with the full-sized PCM are punctured, thus reducing the number of ACS operations per iteration. This may be more intuitively understood from the circular buffer's perspective in Fig. 4.10, where the full-length codeword is constructed in (four) overlapping segments characterized by each redundancy version.

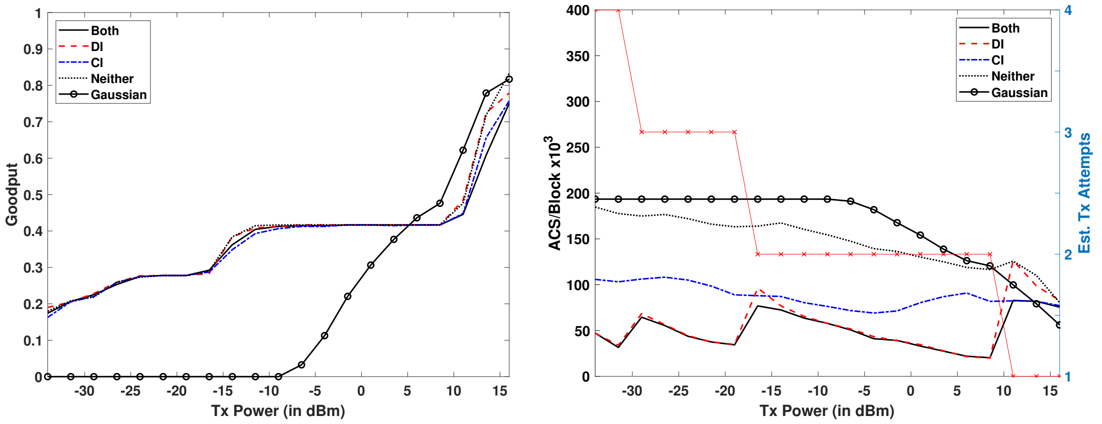
It may be observed in Fig. 4.12(b), 4.12(d) and 4.12(f) that the average complexity generally decreases with the SNR, but relatively sharply increases at the specific SNRs where the estimated number of retransmissions reduces. This result naturally follows the fact that when trying to achieve an identical amount of *extrinsic* output MI using a low amount of accumulated *a priori* MI from $\mathcal{B}_{\text{ext}}$, the corresponding EXIT tunnel becomes narrower and therefore more decoding iterations have to be performed in order to reach the $(1, 1)$ point of the core LDPC EXIT chart. When comparing to the AWGN-based soft demodulation aided 'Gaussian' references, we may observe that the complexity of standard HARQ ('Neither') relying on IN-based soft demodulation is lower at low to medium SNRs, given that both schemes employ CRC-based ET strategy.

With respect to the pair of complexity reduction strategies, CI is capable of reducing the decoding complexity by approximately 50% compared to the CRC-only early termination strategy. The DI philosophy may reduce the complexity by a further 75% in the best case at low SNRs, while having no effect when the number of transmission attempts estimated by EXIT-chart analysis is lower than 2. Furthermore, CI has minimal effect when jointly implemented with DI. For metallic channels, acquiring EXIT tunnel related data does not incur significant complexity overhead. Therefore, it is straightforward to construct a look-up table for the estimated number of transmission attempts of a code block, given the knowledge of the CSI and of the operating SNR. Since EXIT-chart analysis accepts MI as its input, the statistics of IN do not have to be known for the DI switch, as long as the soft demodulator is capable of producing reliable LLR estimates based on the formulation of (4.3).

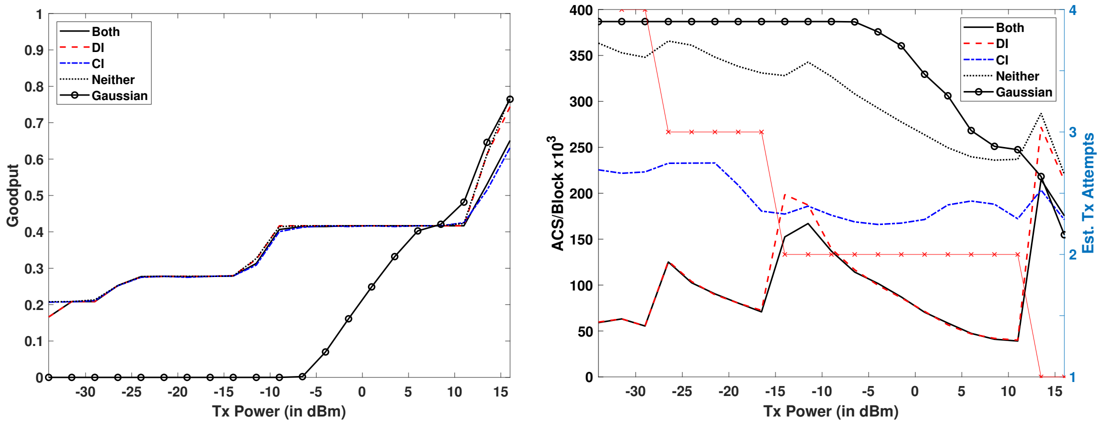
4.6 Chapter Summary

In this chapter, the performance of the class of NR LDPC codes as well as of the NR LDPC-coded HARQ protocol were investigated in the context of next-generation DSL systems in the face of IN. The (base) PCM of the NR LDPC codes shown in Fig. 4.3 facilitates the construction of a concatenated 'L'-shaped structure consisting of the standard LDPC PCM and the LDGM PCM, as well as an additional identity matrix for the convenience of HARQ. As a consequence, the standard SP decoding algorithm described in Sec. 4.3.2.2 now becomes associated with a hybrid Tanner graph portrayed in Fig. 4.4, consisting of three interacting message-passing blocks, corresponding to the three PCM sub-regions. Bearing in mind the unique structure of the NR LDPC codes, we analysed their EXIT characteristics in the presence of IN-based soft information in Sec. 4.3.2. When relying on the standard SP decoding, the performance of the NR LDPC codes may no longer be characterized by the classic 2D EXIT charts. Instead, a pair of 3D EXIT charts as shown in Fig. 4.9 was conceived for representing the average convergence behaviour of the NR LDPC codes.

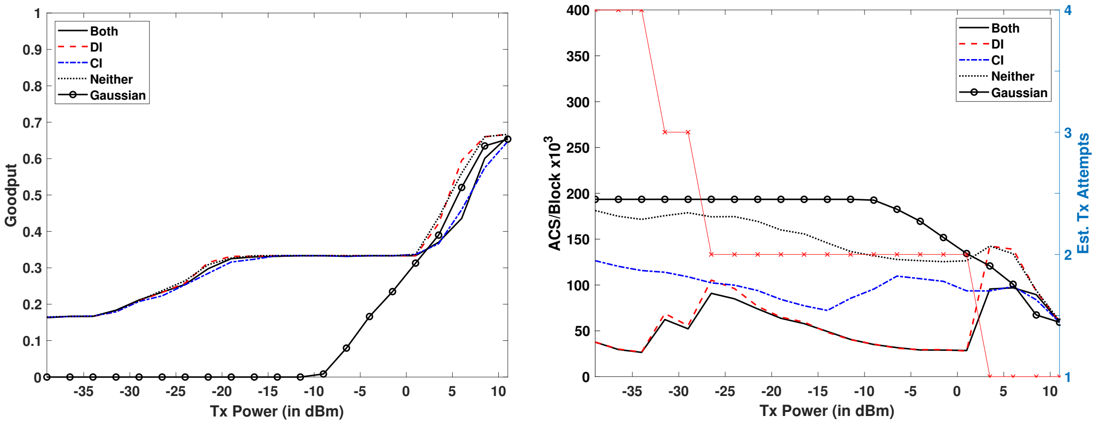
Relying on the 3D EXIT chart characterizations, we conceived a novel low-complexity design for the NR LDPC-coded hybrid CC-IR-HARQ protocol as detailed in Fig. 4.11. In the proposed protocol, the redundant operations of the SP decoder are probabilis-



(a) $(K' = 200, E = 240)$ (b) $(K' = 200, E = 240)$



(c) $(K' = 400, E = 480)$ (d) $(K' = 400, E = 480)$



(e) $(K' = 200, E = 300)$ (f) $(K' = 200, E = 300)$

FIGURE 4.12: Goodput and complexity vs Total Transmit Power (Tx Power) of the proposed NR LDPC-coded HARQ protocol. The Tx Power is defined with respect to an AWGN floor of -150 dBm/Hz and 420 MHz bandwidth (Tab. 4.1), while the IN parameters are defined as in (4.1). The Gaussian reference cases are obtained by employing the standard AWGN-based soft demodulation. The default operation of the NR LDPC decoder employs CRC-based ET strategy.

tically avoided with the aid of a sophisticated hybrid combination of three strategies. Specifically, for the DI strategy of Sec. 4.4.2.1, multiple 3D EXIT charts are constructed offline for predicting the decoding behaviour after receiving each redundancy version, in order to guarantee that the decoder only becomes activated, once the receiver is likely to have accumulated sufficient MI for a successful decoding attempt.

Additionally, a pair of MI thresholds for the decoder's *extrinsic* output and an internal decoder buffer are introduced for the CI of failing decoding attempts and for boosting the start of decoding retransmitted redundancy versions, as discussed in Sec. 4.4.2.3. By comparing the goodput and complexity associated with different NR LDPC coding specifications, we observed in Fig. 4.12 that the reduced-complexity protocol suffers no substantial goodput reduction compared to the typical design having no DI or CI. Furthermore, the complexity reduction is significant at lower SNRs, where a high number of retransmissions are typically required.

Chapter 5

Conclusions and Future Research

5.1 Practical Problems

For the forthcoming metallic wireline broadband access network standard G.mgfast, and the converged wireline-wireless network paradigm beyond 5G, the ultimate optimality of both the wireless and the wireline access networks must be achieved. This vision imposes challenges on the wireline communications community, because wireline access networks are used for the ultra high speed and URLLC as the backhaul of the next generation wireless access network.

In this treatise, we presented an overview of the state-of-the-art DSL technologies, as well as of the emerging solutions for future wireline network architectures. More specifically, we investigated the dominant challenge of FEXT precancellation in DSL wireline access networks. For the enhanced vectoring approach, we found that lattice reduction significantly improves the performance of conventional MUPs at a modest extra complexity during run-time, due to the time-invariant nature of DSL channels. Furthermore, our performance assessment of the MUPs indicates that the gain achieved by LRMUP does not necessarily rely on complex DSB strategies. For particular channel conditions, the performance of a low-complexity MUP having lattice reduction may be better than a higher complexity one having no lattice reduction. This phenomenon is observed in our performance assessment for the LR-ZFP and the plain THP. However, as we have mentioned, the optimal multi-user algorithm is always a potential solver for the NP-hard *exact CVP* problem. Finding the most efficient exact CVP solver, which should preferably have an average case sub-exponential (or lower) complexity order, is still an open problem in active research. Solving this will be crucial for large-scale (e.g. 100-pair binder) implementation of the phantom mode DSL. However, relying on recent development of lattice-based cryptography, we found that a hybrid of sieving and enumeration strategies, termed as IDEALE in our study, is capable of striking considerably better performance-complexity trade-off than the classic CVP-solving approaches.

Moreover, under the general trend of network unification, we found that (SDN-aided) cross-ISP vectoring allows more efficient exploitation of the multi-dimensional signal space and it nearly doubles the sum rate in a two-ISP LLU scenario. Furthermore, SDN and NFV also support *low-cost* realization of *fibre-level* performance over copper, thanks to the recent discovery of 'hidden' signalling modes using the existing telephony-based DSL binders. However, employing these promising new signalling modes requires considerable amount of physical modelling in future research, and the associated CPE modification cost should be carefully compared against that of FTTH for different geographical regions.

Finally, the error control scheme conceived for next generation cellular wireless networks is studied for the metallic access networks in the absence of a particular set of error control specifications. In particular, we verified the effectiveness of IN-based soft-input-soft-output LDPC decoding based on the channel coding and retransmission techniques defined in the NR standard, with the aid of our novel 3D EXIT chart analysis. Furthermore, low-latency HARQ designs were investigated for the NR LDPC codes, which result in significantly lower decoding complexity at the expense of a negligible reduction of goodput.

5.1.1 Complexity vs. Performance

Shannon's channel capacity quantifies the maximum mutual information associated with *a single channel use*. When the sum rate associated with multiple DMT symbol durations is considered, the effective channel capacity must be evaluated with respect to the *processing delay* of the MUP algorithms as well. Firstly, the initialization overhead associated with the LLL lattice reduction algorithm and other channel matrix factorization operations such the QR decomposition does not affect the processing delay of the MUP during run-time, and the overhead itself may be deemed affordable on average (i.e. polynomial[1]). Secondly, their operations may be expedited by invoking parallel algorithms, such as the parallel sphere encoder of [191]. This is also conceptually the approach taken both by the K-best [184] and by the fixed complexity sphere decoding (FSD) [266] algorithm, as well as by the parallel THP algorithms of [267][268]. Additionally, parallel computing can also be used for pipelining the lattice reduction algorithm [269][270]. However, we should note that these reduced-complexity variants are generally suboptimal compared to their original sequential counterparts, because the former typically ignores a sizeable part of the solution set that has a low probability of containing the global optimum. Nonetheless, even though for low-dimensional systems the performances of low-complexity algorithms match those of their original counterparts sufficiently well, the trade-off must be reinvestigated for large-scale systems.

---

[1] However, the worst-case complexity of the LLL algorithm has been shown to be infinite in [265].

5.1.2 The Near-Far Problem

Since the telecommunications industry has developed according to a demand-driven model, the design philosophy of access networks is gradually shifting from the network-centric paradigm to user-centric [271][75], where the main focus becomes quality of experience (QoE) rather than the conventional QoS [272]. From a user-centric perspective, the access network must be capable of providing satisfactory data rate and high reliability for each user, subject to their individual QoE requirement. However, due to the different wire-length of each user in the wireline access network, providing good QoE for all users is a challenging task in the face of the *near-far problem*.

The detrimental effect of the near-far problem in DSL networks mainly results in the violation of the *user fairness*. Recall from Fig. 1.2 that each customer premise and hence the CPE is generally located at a different distance from the G.fast DP. Given the propagation characteristics we discussed in Section 1.3, the signals of the users who are far away from the DP are often overwhelmed by those of the users that are closer. The overall performance associated with mixed binder length is typically much worse than that of fixed-length binder due to the worst-case dominance effect. Furthermore, as demonstrated by the results of [134], the performance gap between different MUPs is sometimes also affected by the mixed binder length. The typical solution to the near-far problem in wireless communications is to eliminate the interference with the aid of SIC. However, this solution cannot be readily applied in DSL networks because unlike the mobile terminals in wireless networks, the CPEs cannot be relocated to improve user fairness.

It was shown in [96] that lattice reduction based preprocessing can be used for improving the fairness guarantee of the THP-like approach. In particular, this may be interpreted as the equalization of the channel matrix eigenvalues, representing the balanced CIR of each user. However, the drawback of the LR-aided approach is that the users having good CIRs must sacrifice their performance for the sake of fairness. From a user-centric perspective, multiple lattice basis associated with the channel matrix may be stored so that the precoding matrix may be adjusted based on the predicted QoE requirement.

5.2 Future Research

Given the shared theoretical modelling approach for both the wireline and wireless communication systems, the main difference between them is the impairment imposed by the physical properties of the respective channels. Compared to the typically (quasi-)static, frequency-selective wireline channel model characterized in Sec. 1.3 as well as to the types of coloured noise discussed in Ch. 4, wireless channels vary significantly faster than wireline ones, while the noise spectrum is normally white. Such a difference encourages the exploitation of the time domain diversity of wireless channels, as well as straightforward channel coding design and optimization for systems contaminated by white noise only. By contrast, the time diversity of wireline systems has to be exploited differently based on practical traffic patterns, while the channel coding design has to consider the information-theoretical implications of coloured noise. These research challenges present excellent opportunities for future research.

5.2.1 System Imperfections

Approaching the multi-pair multi-tone channel capacity characterized by DPC requires instantaneous and non-causal knowledge of the communication environment, including both the channel matrices and the highly-coloured noise. Practical DSL transceiver units suffer from the following problems:

- *Imperfect MUP Design.* Firstly, the imperfect transmit CSI obtained via the standard vectoring feedback loop of Fig. 2.2 in Sec. 2.1 causes violation of the ZF signal reception criterion of both the linear ZFP of Sec. 2.3.1.1 and the ZF-THP of Sec. 2.3.2.1, and therefore it results in residual crosstalk. The same impairment applies to other criteria such as MMSE, which also causes mismatch between the MUP and the actual downstream channel. For these signal reception criteria, we need to update the vectoring mapping graphs of Sec. 3.1 slightly to reflect the change of the linear FEXT-canceller front end. Moreover, the classic THP-aided non-linear vectoring [68] considered by the DSL community is susceptible to instability in the face of imperfect transmit CSI. Hence the THP-based multi-level DSM was outperformed by linear multi-level DSM in the case of [133]. On the other hand, even though the family of lattice-based techniques of Sec. 2.5, 2.6 and 2.7 are generally less susceptible to CSI error, the performance of LR techniques is dependent upon the numerical precision of the LR algorithm, since our comparisons have shown that perfect LR may improve the robustness of conventional MUPs as seen in Sec. 2.5. In particular, the numerical stability of certain versions of the LLL algorithm was studied in [168]. Characterizing the realistic performance of non-linear MUPs, including that of the optimal sphere-encoder of Sec. 2.7.1, as well as of its low-complexity variants such as those of Sec. 2.7.2, requires further research. Practical solutions may be obtained via AI-related techniques such as the DE algorithm derived from Sec. 2.7.3 and with the aid of other machine learning techniques.
- *Imperfect Noise Estimation.* Secondly, the sum rate achieved by vectored transmission is strongly influenced by the noise level and distribution. The majority of the existing multi-level DSM research, as discussed in Sec. 3.4, assumes the noise to be white, which consists of the typical AWGN plus a noise margin reserved for the worst-case noise bursts. Naturally, a conservative design philosophy will lead to

suboptimal performance, as shown later by the goodput comparisons of Fig. 4.12. Specifically, the capacity of the multi-channel system is underestimated due to the overestimated average noise power. Since all residual interference may be considered as noise, e.g. alien crosstalk, RFI and IN, it is challenging to construct an accurate practical noise model accounting for all types of sources, since we would like to exploit the *a posteriori* knowledge of IN events and its benefit, as investigated in Sec. 4.3.2.1. Generally, bursty noise constrains the potential adoption of multi-level DSM because of the associated retrain cost. Consequently, learning the noise statistics *accurately* may improve the overall performance of DSL wireline networks considerably. In particular, we have shown that HARQ-based level-1 DSM is capable of adapting to the noise environment if the noise statistics is perfectly known to the demodulator. In contrast to the type of multi-level DSM we have investigated in Sec. 3.4, a holistic design that combines all three levels may result in significantly better performance than our current observations.

5.2.2 Retrain Cost

So far we have assumed that the vectoring control protocol of Section 2.1 only has to be invoked once at the initialization stage to train the VCE, which will then continue to operate for multiple DMT symbol durations. This is in general not a strong assumption concerning the time-invariant nature of DSL channels. However, there exist other factors, which can substantially change the frequency response of a particular DSL binder, such as physically moving or bending the binder at some midpoint. Since DSL binders are normally placed overhead as drop wires or buried underground, the probability of these events are slim. Therefore, the cost associated with retraining the VCE for updating the CSI knowledge does not generally constitute a performance bottleneck.

However, for the 424 MHz G.mgfast profile, we should note that the number of tones is over 8,000 [5]. In this case, the total initialization overhead associated with LLL or other lattice reduction techniques of Sec. 2.4.3 may potentially exceed the acceptable processing delay for initialization. The case study for the NR LDPC codes in Sec. 4.3.1 have also assumed small-scaled channel coding, whose complexity escalates quickly in terms of both the EXIT tunnel acquisition and of the actual decoding operations, when the codeword approach a length of 8,000. However, as implied by Fig. 2.11, LLL reduction is only required for frequencies above 90 MHz. Furthermore, the average complexity of the LLL algorithm and that of the QR decomposition are both of a polynomial order as seen in Sec. 2.9, the latter of which is the mandatory preprocessing for the THP. Therefore, the practicality of the LRMUP is comparable to that of the THP as non-linear MUP candidates.

For NR LDPC-coded HARQ schemes, the interleaving strategy and the sequence of redundancy versions defined in the standard may be adapted to the channel and data traffic conditions for achieving improved power efficiency as well as for realizing the otherwise non-existent diversity gain in metallic networks. On the other hand, we may observe from Fig. 4.3 and 4.4 that treating the entire PCM holistically is somewhat inefficient due to the concurrent message-passing structure discussed in Sec. 4.3.2.3 and 4.3.2.4. As a result, the number of ACS operations per decoding iteration may be reduced with the aid of an optimized decoding schedule switching between the sub-regions of the PCM.

On the other hand, due to the DSL's susceptibility to the stochastic IN and RFI, the DSB policy that only specifies the AWGN PSD and a static noise margin has to be frequently updated in practice, due to the limitation in acquiring the statistics of IN. The SRA protocol is an existing solution, which is capable of providing a real-time DSB policy update without requiring VCE retraining. However, the complexity associated with the optimization of DSB may become a performance bottleneck in the face of IN and RFI. For this reason, the low complexity of the SSB policy is favourable. As shown in Section 3.5.2, the idealized VP relying on the SSB policy has a similar performance to that of the THP relying on optimized DSB. Therefore, using an optimally-tuned low-complexity sphere encoder aided VP may potentially become the capacity-achieving solution for vectored DSL systems.

5.2.3 Compatibility

We will consider both the backward compatibility with current DSL standards such as G.fast and VDSL2, as well as the forward compatibility with future standards following our vision for the wireline access network. In general, the compatibility problem occurs as a consequence of the multi-standard operation of DSL. As investigated in [144], using the THP in a system mixed with 20% linear receivers does not significantly downgrade the performance compared to the ideal THP transceiver structure of Fig. 2.7. Since the modulo receivers are commonly used by both the THP and the LRMUPs, the negative impact of legacy linear receivers on the performance of large-scale deployment of non-linear MUP is modest. For the new hardware requirement at the DP side, we have justified that the non-linear optimization block of VP can be incorporated as a simple attachment into the widely deployed ZFP-based architecture. Therefore, the operational expenditure associated with VP should be moderate compared to the alternative non-linear MUP architectures. This is due to the fact that the VP encoder is fully compatible with the existing linear MUP, hence its linear front end and the non-linear optimizer can be maintained or replaced independently.

To overcome the bandwidth efficiency limit of the state-of-the-art DSL deployment in preparation for the next generation access network paradigm, the architecture of the current wireline access network has to be fundamentally refined. In this case, the forward compatibility issue results in a two-fold CAPEX trade-off. Firstly, based on the investigation of [20], fibre placement should be prioritized in areas where the CAPEX associated with FTTdp and that of FTTH is comparable. This route requires the corresponding deployment of fibre-based CPEs. Secondly, if FTTH is significantly more expensive than FTTdp, then the unexplored signalling modes of DSL binders should be employed. In general, utilizing the TDSL transmission mode (Section 1.3) hidden within the existing DSL binders requires modifications of the critical components of the state-of-the-art CPE hardware, such as adding THz antennas and RF down converters. However, the phantom mode signalling is at the moment a well-established technology, even though it is not widely exploited in the industry yet. For these reasons, the CAPEX of CPE modifications and last mile fibre placement should be carefully assessed and compared.

Appendix A

NR LDPC Base Graphs

The base graphs of the NR LDPC codes are intended to be a more compact characterization of the PCM as well as of its average performance. Both base graphs share the same general structure as follows:

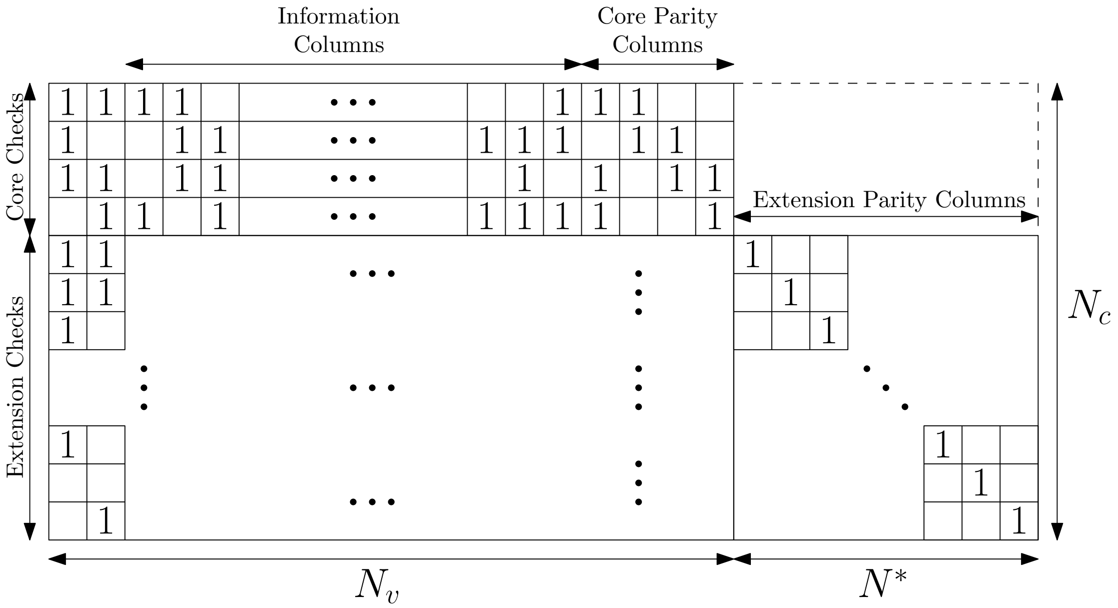


Figure A.1: Concatenated structure of the NR LDPC base graphs. The three base graph regions correspond respectively to: 1). standard LDPC (top left); 2) LDGM (bottom left); 3) identity matrix (bottom right). All information and CRC bits, as well as the self-decodable FEC parity bits are contained in the information and core parity columns.

More specifically, the two versions defined in the current NR standard revision are listed as follows, which should be implemented in conjunction with the lifting (i.e. expanding) specifications from the same revision:

1	1	1	1	0	1	1	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	1	1	1	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure A.2: NR LDPC: Base Graph 1 (reproduced from [237]).

1	1	1	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	1	0	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

FIGURE A.3: NR LDPC: Base Graph 2 (reproduced from [237]).

Bibliography

[1] S. Gunn, "Pdflatex instructions," 2001.

[2] C. J. Lovell, "Updated templates," 2011.

[3] S. Gunn and C. J. Lovell, "Updated templates reference 2," 2011.

[4] M. Ruffini, "Multidimensional convergence in future 5G networks," *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 535–549, Feb 2017.

[5] R. Strobel, "Physical layer framework for the multi-Gigabit copper-fiber access networks," *Fiber and Integrated Optics*, vol. 37, no. 6, pp. 291–313, 2018.

[6] F. van Lingen, M. Yannuzzi, A. Jain, R. Irons-Mclean, O. Lluch, D. Carrera, J. L. Perez, A. Gutierrez, D. Montero, J. Marti, R. Maso, and a. J. P. Rodriguez, "The unavoidable convergence of NFV, 5G, and Fog: A model-driven approach to bridge cloud and edge," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 28–35, Aug 2017.

[7] A. Tzanakaki, M. Anastasopoulos, I. Berberana, D. Syrivelis, P. Flegkas, T. Korakis, D. C. Mur, I. Demirkol, J. Gutierrez, E. Grass, Q. Wei, E. Pateromichelakis, N. Vucic, A. Fehske, M. Grieger, M. Eiselt, J. Bartelt, G. Fettweis, G. Lyberopoulos, E. Theodoropoulou, and D. Simeonidou, "Wireless-optical network convergence: Enabling the 5G architecture to support operational and end-user services," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 184–192, Oct 2017.

[8] P. T. Dat, A. Kanno, N. Yamamoto, and T. Kawanishi, "Seamless convergence of fiber and wireless systems for 5G and beyond networks," *Journal of Lightwave Technology*, pp. 1–1, 2018.

[9] S. Gosselin, A. Pizzinat, X. Grall, D. Breuer, E. Bogenfeld, S. Krauss, J. A. T. Gijon, A. Hamidian, N. Fonseca, and B. Skubic, "Fixed and mobile convergence: which role for optical networks?" *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 11, pp. 1075–1083, November 2015.

[10] K. Kanonakis, I. Tomkos, H. Krimmel, F. Schaich, C. Lange, E. Weis, J. Leuthold, M. Winter, S. Romero, P. Kourtessis, M. Milosavljevic, I. N. Cano, and O. Prat,

"An OFDMA-based optical access network architecture exhibiting ultra-high capacity and wireline-wireless convergence," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 71–78, August 2012.

[11] Z. Li, *Telecommunication 4.0: Reinvention of the communication network.* Springer Singapore, 2017.

[12] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu, and Y. Liu, "A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges," *IEEE Commun. Surveys Tuts.*, pp. 1–1, 2018.

[13] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, Fourthquarter 2018.

[14] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surveys Tuts.*, pp. 1–1, 2018.

[15] S. Bemby, H. Lu, K. H. Zadeh, H. Bannazadeh, and A. Leon-Garcia, "ViNO: SDN overlay to allow seamless migration across heterogeneous infrastructure," in *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, May 2015, pp. 782–785.

[16] K. Kerpez and G. Ginis, "Software-defined access network (SDAN)," in *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, March 2014, pp. 1–6.

[17] K. J. Kerpez, J. M. Cioffi, G. Ginis, M. Goldburg, S. Galli, and P. Silverman, "Software-defined access networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 152–159, Sep. 2014.

[18] R. Zheng, W. Yang, and J. Zhou, "Future access architecture: Software-defined accesss networking," in *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Jan 2014, pp. 881–886.

[19] Cisco, "Global mobile data traffic forecast update," Feb 2017.

[20] J. R. Schneir and Y. Xiong, "Cost assessment of FTTdp networks with G.fast," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 144–152, August 2016.

[21] ITU-T, "G.993.5 : Self-FEXT cancellation (vectoring) for use with VDSL2 transceivers."

[22] ——, "G.9701 : Fast access to subscriber terminals (G.fast) - Physical layer specification."

[23] ——, "G.9807.1 : 10-Gigabit-capable symmetric passive optical network (XGS-PON)."

[24] ITU-BCSD, "The state of broadband 2018: Broadband catalyzing sustainable development," Septmber 2018.

[25] ITU-T, "G.992.5 : Asymmetric digital subscriber line 2 (ADSL2) transceivers - Extended bandwidth ADSL2 (ADSL2plus)."

[26] ——, "H.265 : High efficiency video coding."

[27] L. Dake, C. Zhaoyun, and W. Wei, "Trends of communication processors," *China Communications*, vol. 13, no. 1, pp. 1–16, Jan 2016.

[28] M. T. Raza, S. Lu, M. Gerla, and X. Li, "Refactoring network functions modules to reduce latencies and improve fault tolerance in NFV," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 10, pp. 2275–2287, Oct 2018.

[29] J. M. Cioffi, K. J. Kerpez, C. S. Hwang, and I. Kanellakopoulos, "Terabit DSLs," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 152–159, November 2018.

[30] S. Jagannathan, V. Pourahmad, K. Seong, J. M. Cioffi, M. Ouzzif, and R. Tarafi, "Common-mode data transmission using the binder sheath in digital subscriber lines," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 831–840, March 2009.

[31] W. Foubert, C. Neus, L. V. Biesen, and Y. Rolain, "Exploiting the phantom-mode signal in DSL applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 4, pp. 896–902, April 2012.

[32] L. F. Alloin, P. Biyani, L. Pierrugues, S. Zafaruddin, and C. V. Kumar, "DM-CM diversity receiver for a wireline communication system," May 26 2015, uS Patent 9,042,498.

[33] ITU-T, "G.995.2 : Enhanced common mode limits and measurement methods for customer premises equipment operating on copper pairs."

[34] C. Leung, S. Huberman, K. Ho-Van, and T. Le-Ngoc, "Vectored DSL: Potential, implementation issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1907–1923, Fourth 2013.

[35] D. Acatauassu, S. Höst, C. Lu, M. Berg, A. Klautau, and P. O. Börjesson, "Simple and causal copper cable model suitable for G.fast frequencies," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4040–4051, Nov 2014.

[36] D. Gomes, E. Medeiros, A. Klautau, and E. Pelaes, "Mitigation of alien crosstalk for downstream DSL impaired by multiple interferers," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2380–2383, Nov 2017.

[37] A. A. Amayreh, J. L. Masson, M. Hélard, and M. Ouzzif, "Alien crosstalk elimination in digital subscriber line systems," *IET Communications*, vol. 8, no. 10, pp. 1714–1723, July 2014.

[38] G. Ginis and C.-N. Peng, "Alien crosstalk cancellation for multipair digital subscriber line systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 016828, Feb 2006.

[39] K. M. Ali, G. G. Messier, and S. W. Lai, "DSL and PLC co-existence: An interference cancellation approach," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3336–3350, Sep. 2014.

[40] S. Galli, K. J. Kerpez, H. Mariotte, and F. Moulin, "PLC-to-DSL interference: statistical model and impact on VDSL2, vectoring, and G.fast," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 7, pp. 1992–2005, July 2016.

[41] M. Sorbara, P. Duvaut, F. Shmulyian, and S. Singh, "Construction of a DSL-MIMO channel model for evaluation of FEXT cancellation systems in VDSL2," in *Sarnoff Symposium 2007*. Nassau Inn, Princeton, NJ: IEEE, Apr 2007, pp. 1 – 6.

[42] J. Maes, M. Guenach, and M. Peeters, "Statistical MIMO channel model for gain quantification of DSL crosstalk mitigation techniques," in *2009 IEEE International Conference on Communications*, June 2009, pp. 1–5.

[43] T. Starr, J. M. Cioffi, and P. Silverman, *Understanding digital subscriber line technology*, ser. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 1999, no. v. 1.

[44] R. F. M. van den Brink, "Modeling the dual-slope behavior of in-quad EL-FEXT in twisted pair quad cables," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2153–2163, May 2017.

[45] O. Ogundapo, A. Duffy, and C. Nche, "Parameter for near end crosstalk prediction in twisted pair cables," in *2016 IEEE International Symposium on Electromagnetic Compatibility (EMC)*, July 2016, pp. 485–490.

[46] P. Lafata, "Realistic modeling of far-end crosstalk in metallic cables," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 435–438, March 2013.

[47] V. Belevitch, "Theory of proximity effect in multiwire cables. 2." *Philips Research Reports*, vol. 32, no. 2, pp. 96–117, 1977.

[48] E. Oswald, "Application of a phantom circuit for xDSL communications," in *Proc. 13th International OFDM-Workshop*, Aug 2008.

[49] ITU-T, "G.998.1: ATM-based multi-pair bonding."

[50] ——, "G.998.2: Ethernet-based multi-pair bonding."

[51] ——, “G.998.3: Multi-pair bonding using time-division inverse multiplexing.”

[52] W. Coomans, R. B. Moraes, K. Hooghe, A. Duque, J. Galaro, M. Timmers, A. J. van Wijngaarden, M. Guenach, and J. Maes, “XG-FAST: Towards 10 Gb/s copper access,” in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 630–635.

[53] K. Wang and D. M. Mittleman, “Metal wires for terahertz wave guiding,” *Nature*, vol. 432, no. 7015, p. 376, 2004.

[54] D. M. Mittleman, “Frontiers in terahertz sources and plasmonics,” *Nature Photonics*, vol. 7, no. 9, p. 666, 2013.

[55] T. Brulé, “Spectral and temporal distribution of biomolecules by dynamic SERS,” Ph.D. dissertation, Dijon, 2014.

[56] T.-I. Jeon, J. Zhang, and D. Grischkowsky, “THz Sommerfeld wave propagation on a single metal wire,” *Applied Physics Letters*, vol. 86, no. 16, p. 161904, 2005.

[57] J. C. Wiltse, “Surface-wave propagation on a single metal wire or rod at millimeter-wave and Terahertz frequencies,” in *2006 IEEE MTT-S International Microwave Symposium Digest*, June 2006, pp. 970–973.

[58] A. Markov, S. Gorgutsa, H. Qu, and M. Skorobogatiy, “Practical metal-wire THz waveguides,” *arXiv preprint arXiv:1206.2984*, 2012.

[59] O. Lamparter, L. Fang, J. Bischoff, M. Reitmann, R. Schwendener, T. Zasowski, and X. Zhang, “Multi-Gigabit over copper access networks: Architectural evolution and techno-economic analysis,” *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 22–27, August 2019.

[60] V. Oksman, R. Strobel, T. Starr, J. Maes, W. Coomans, M. Kuipers, E. B. Tovim, and D. Wei, “MGFAST: A new generation of copper broadband access,” *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 14–21, August 2019.

[61] ITU-T, “G.993.2 : Very high speed digital subscriber line transceivers 2 (VDSL2).”

[62] W. Shieh, “OFDM for flexible high-speed optical networks,” *Journal of Lightwave Technology*, vol. 29, no. 10, pp. 1560–1577, May 2011.

[63] ITU-T, “G.9700 : Fast access to subscriber terminals (G.fast) - Power spectral density specification.”

[64] G. Fettweis, M. Krondorf, and S. Bittner, “GFDM - Generalized frequency division multiplexing,” in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, April 2009, pp. 1–4.

[65] B. Farhang-Boroujeny and H. Moradi, “Derivation of GFDM based on OFDM principles,” in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 2680–2685.

[66] C.-L. Tai, B. Su, and C. Jia, "Frequency-domain decoupling for MIMO-GFDM spatial multiplexing," *arXiv preprint arXiv:1803.06448*, 2018.

[67] J. Zhang, S. Chen, R. Zhang, A. F. A. Rawi, and L. Hanzo, "Differential evolution algorithm aided turbo channel estimation and multi-user detection for G.fast systems in the presence of FEXT," *IEEE Access*, vol. 6, pp. 33 111–33 128, 2018.

[68] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1085 – 1104, June 2002.

[69] D. Hincapie, G. Maierbacher, and M. Leibiger, "Rate and reach gains of vectored DSL in the current access network," in *Broadband Coverage in Germany. 9th ITG Symposium. Proceedings*, April 2015, pp. 1–6.

[70] J. Rückert, R. Bifulco, M. Rizwan-Ul-Haq, H. Kolbe, and D. Hausheer, "Flexible traffic management in broadband access networks using software defined networking," in *2014 IEEE Network Operations and Management Symposium (NOMS)*, May 2014, pp. 1–8.

[71] J. M. Cioffi, M. Brady, V. Pourahmad, S. Jagannathan, W. Lee, Y. Kim, C. Chen, K. Seong, D. Yu, M. Ouzzif, H. Mariotte, R. Tarafi, G. Ginis, B. Lee, T. Chung, and P. Silverman, "Vectored DSLs with DSM: The road to ubiquitous Gigabit DSLs," in *Proc. World Telecommun. Conf.*, May 2006.

[72] T. Bai, H. Zhang, J. Zhang, C. Xu, A. F. A. Rawi, and L. Hanzo, "Impulsive noise mitigation in digital subscriber lines: The state-of-the-art and research opportunities," *IEEE Commun. Mag.*, vol. 57, no. 5, pp. 145–151, May 2019.

[73] F. V. der Putten, "Overview of ITU-T SG15 Q4 xDSL and G.(mg)fast," May 2017.

[74] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2017–2046, Fourthquarter 2015.

[75] V. Wong, R. Schober, D. Ng, and L. Wang, *Key technologies for 5G wireless systems.* Cambridge University Press, 2017.

[76] S. K. Sharma, T. E. Bogale, L. B. Le, S. Chatzinotas, X. Wang, and B. Ottersten, "Dynamic spectrum sharing in 5G wireless networks with full-duplex technology: Recent advances and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 674–707, Firstquarter 2018.

[77] X. Wang, H. Huang, and T. Hwang, "On the capacity gain from full duplex communications in a large scale wireless network," *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2290–2303, Sep. 2016.

[78] X. Xie and X. Zhang, “Does full-duplex double the capacity of wireless networks?” in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 253–261.

[79] N. Ehtiati and B. Champagne, “A general framework for mixed-domain echo cancellation in discrete multitone systems,” *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 769–780, February 2013.

[80] R. Nongpiur, D. J. Shpak, and A. Antoniou, “An analysis of a near-end crosstalk cancelation system that uses adaptive filters,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 10, pp. 3306–3316, Nov 2008.

[81] P. Tsiaflakis, Y. Lefevre, W. Coomans, and J. Maes, “Friendly full duplex: A multiuser full duplex method for mgfast in coexistence with G.fast,” in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–6.

[82] ITU-T, “G.998.4 : Improved impulse noise protection for digital subscriber line (DSL) transceivers.”

[83] T. Bai, H. Zhang, R. Zhang, L. L. Yang, A. F. A. Rawi, J. Zhang, and L. Hanzo, “Discrete multi-tone digital subscriber loop performance in the face of impulsive noise,” *IEEE Access*, vol. 5, pp. 10 478–10 495, 2017.

[84] F. Sjoberg, R. Nilsson, P. O. Borjesson, P. Odling, B. Wiese, and J. A. C. Bingham, “Digital RFI suppression in DMT-based VDSL systems,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 11, pp. 2300–2312, Nov 2004.

[85] M. Mirahmadi, A. Al-Dweik, and A. Shami, “BER reduction of OFDM based broadband communication systems over multipath channels with impulsive noise,” *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4602–4615, November 2013.

[86] J. Neckebroek, M. Moeneclaey, M. Guenach, M. Timmers, and J. Maes, “Comparison of error-control schemes for high-rate communication over short DSL loops affected by impulsive noise,” in *2013 IEEE International Conference on Communications (ICC)*, June 2013, pp. 4014–4019.

[87] T. Bai, C. Xu, R. Zhang, A. F. Al Rawi, and L. Hanzo, “Performance of HARQ-assisted OFDM systems contaminated by impulsive noise: Finite-length LDPC code analysis,” *IEEE Access*, vol. 7, pp. 14 112–14 123, 2019.

[88] U. Challita, L. Dong, and W. Saad, “Proactive resource management for LTE in unlicensed spectrum: A deep learning perspective,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4674–4689, July 2018.

[89] S. Sendra, A. Rego, J. Lloret, J. M. Jimenez, and O. Romero, “Including artificial intelligence in a routing protocol using software defined networks,” in *2017 IEEE*

International Conference on Communications Workshops (ICC Workshops), May 2017, pp. 670–674.

[90] R. Alvizu, S. Troia, G. Maier, and A. Pattavina, “Matheuristic with machine-learning-based prediction for software-defined mobile metro-core networks,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 9, pp. D19–D30, Sep. 2017.

[91] J. Zhang, S. Chen, X. Mu, and L. Hanzo, “Evolutionary-algorithm-assisted joint channel estimation and turbo multiuser detection/decoding for OFDM/SDMA,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1204–1222, March 2014.

[92] N. Samuel, T. Diskin, and A. Wiesel, “Deep MIMO detection,” in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–5.

[93] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec 2017.

[94] S. Yang and L. Hanzo, “Fifty years of MIMO detection: The road to large-scale MIMOs,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, Fourthquarter 2015.

[95] M. R. Albrecht, L. Ducas, G. Herold, E. Kirshanova, E. W. Postlethwaite, and M. Stevens, “The general sieve kernel and new records in lattice reduction,” in *Advances in Cryptology – EUROCRYPT 2019*. Cham: Springer International Publishing, 2019, pp. 717–746.

[96] Y. Zhang, R. Zhang, A. F. A. Rawi, and L. Hanzo, “Approximate perturbation aided lattice encoding (APPLE) for G.fast and beyond,” *IEEE Access*, vol. 6, pp. 53 438–53 451, 2018.

[97] Y. Zhang, R. Zhang, J. Zhang, T. Bai, A. F. Al Rawi, M. Moonen, and L. Hanzo, “Far-end crosstalk mitigation for future wireline networks beyond G.mgfast: A survey and an outlook,” *IEEE Access*, vol. 8, pp. 9998–10 039, 2020.

[98] S. Shao, Y. Zhang, R. G. Maunder, and L. Hanzo, “3D EXIT chart analysis and scheduling of 3GPP New Radio LDPC decoder,” *IEEE Trans. Veh. Technol. (to be submitted)*, 2020.

[99] H. Cramér, *Mathematical methods of statistics (PMS-9)*. Princeton university press, 1946, vol. 9.

[100] C. Rao, “Information and accuracy attainable in the estimation of statistical parameters,” *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.

[101] J. Louveaux and A. . V. D. Veen, “Adaptive DSL crosstalk precancellation design using low-rate feedback from end users,” *IEEE Signal Processing Letters*, vol. 13, no. 11, pp. 665–668, Nov 2006.

[102] I. Binyamini and I. Bergel, “Adaptive precoder using sign error feedback for FEXT cancellation in multichannel downstream VDSL,” *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2383–2393, May 2013.

[103] Z. Shi, Q. Wang, J. Jin, D. Jiang, and G. Liu, “Achievability of the channel reciprocity and its benefit in TDD system,” in *2010 5th International ICST Conference on Communications and Networking in China*, Aug 2010, pp. 1–4.

[104] L. Yang, “Multiuser transmission via multiuser detection: Altruistic-optimization and egocentric-optimization,” in *2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, April 2007, pp. 1921–1925.

[105] Goldsmith, A., *Wireless communications.* Cambridge University Press, 2005.

[106] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439 – 441, May 1983.

[107] A. Wyner, “Recent results in the Shannon theory,” *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, January 1974.

[108] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[109] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[110] L. L. Yang, *Multicarrier communications.* Wiley, 2009.

[111] L. Wang, L. Xu, S. Chen, and L. Hanzo, “Three-stage irregular convolutional coded iterative center-shifting K-best sphere detection for soft-decision SDMA–OFDM,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 2103–2109, May 2009.

[112] M. Tomlinson, “New automatic equaliser employing modulo arithmetic,” *Electronics Letters*, vol. 7, no. 5, pp. 138 – 139, Mar 1971.

[113] H. Harashima and H. Miyakawa, “Matched-transmission technique for channels with intersymbol interference,” *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774 – 780, Aug 1972.

[114] S. M. Zafaruddin, I. Bergel, and A. Leshem, “Signal processing for Gigabit-rate wireline communications: An overview of the state of the art and research challenges,” *IEEE Signal Processing Magazine*, vol. 34, no. 5, pp. 141–164, Sep. 2017.

[115] V. Oksman and B. Heise, "Method for seamless bit rate adaptation for multicarrier DSL," Apr. 14 2009, US Patent 7,519,124.

[116] R. Cendrillon, M. Moonen, T. Bostoen, and G. Ginis, "The linear zero-forcing crosstalk canceler is near-optimal in DSL channels," in *Global Telecommunications Conference*, vol. 4. IEEE, Nov 2004, pp. 2334 – 2338.

[117] R. Cendrillon, G. Ginis, E. V. den Bogaert, and M. Moonen, "A near-optimal linear crosstalk precoder for downstream VDSL," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 860–863, May 2007.

[118] R. Cendrillon, M. Moonen, J. Verlinden, T. Bostoen, and G. Ginis, "Improved linear crosstalk precompensation for DSL," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, May 2004, pp. iv–1053–6 vol.4.

[119] R. Cendrillon, G. Ginis, M. Moonen, and K. Van Acker, "Partial crosstalk precompensation in downstream VDSL," *Signal Processing*, vol. 84, no. 11, pp. 2005–2019, 2004.

[120] A. Leshem and L. Youming, "A low complexity linear precoding technique for next generation VDSL downstream transmission over copper," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5527–5534, Nov 2007.

[121] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195 – 202, Jan 2005.

[122] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec 2001.

[123] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[124] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[125] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711 – 1721, May 2007.

[126] P. A. Lopes and J. A. Gerald, "Leakage-based precoding algorithms for multiple streams per terminal MU-MIMO systems," *Digital Signal Processing*, vol. 75, pp. 38–44, 2018.

[127] P. Patcharamaneepakorn, S. Armour, and A. Doufexi, "On the equivalence between SLNR and MMSE precoding schemes with single-antenna receivers," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1034 – 1037, July 2012.

[128] J. Louveaux and A. van der Veen, "Adaptive precoding for downstream crosstalk precancelation in DSL systems using sign-error feedback," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3173–3179, June 2010.

[129] I. Bergel, S. Zafaruddin, and A. Leshem, "Large matrix asymptotic analysis of ZF and MMSE crosstalk cancelers for wireline channels," *arXiv preprint arXiv:1803.06019*, 2018.

[130] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2008–2026, April 2011.

[131] K. R. Kumar, G. Caire, and A. L. Moustakas, "Asymptotic performance of linear receivers in MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4398–4418, Oct 2009.

[132] W. Lanneer, P. Tsiaflakis, J. Maes, and M. Moonen, "Linear and nonlinear precoding based dynamic spectrum management for downstream vectored G.fast transmission," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1247–1259, March 2017.

[133] J. Maes, C. Nuzman, and P. Tsiaflakis, "Sensitivity of nonlinear precoding to imperfect channel state information in G.fast," in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 290–294.

[134] L. Stigant, R. Zhang, L. D. Humphrey, T. P. Linney, T. Morsman, and A. F. Al Rawi, "Spatially adaptive linear precoding for systems beyond G.fast," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–6.

[135] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *34th Asilomar Conference on Signals, Systems and Computers*, vol. 2. Pacific Grove, CA, USA: IEEE, Oct 2000, pp. 1627 – 1631.

[136] M. Hekrdla, A. Matera, W. Wang, D. Wei, and U. Spagnolini, "Ordered Tomlinson-Harashima precoding in G.fast downstream," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[137] K. Zu, F. C. B. F. Müller, C. Lu, P. Eriksson, and A. Klautau, "Rate balancing based Tomlinson–Harashima precoding for G.fast systems," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1519–1522, Aug 2016.

[138] R. Habendorf and G. Fettweis, "On ordering optimization for MIMO systems with decentralized receivers," in *International Symposium on Intelligent Signal Processing and Communications Systems*, vol. 4. Melbourne: IEEE, May 2006, pp. 1844 – 1848.

[139] S. Kinjo and S. Ohno, "A Tomlinson-Harashima precoding for correlated MIMO channels," in *International Symposium on Intelligent Signal Processing and Communications Systems.* Naha: IEEE, Nov 2013, pp. 257 – 262.

[140] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *URSI International Symposium on Signals, Systems, and Electronics.* Pisa: IEEE, Oct 1998, pp. 295 – 300.

[141] D. Wubben, J. Rinas, R. Bohnke, V. Kuhn, and K. D. Kammeyer, "Efficient algorithm for detecting layered space-time codes," in *International ITG Conference on Source and Channel Coding*, Berlin, Jan 2002, pp. 399 – 405.

[142] M. L. Honig, *Capacity-approaching multiuser communications over multiple input/multiple output broadcast channels.* Wiley-IEEE Press, 2009, pp. 384–.

[143] J. Maes, "Making G.fast faster," in *2016 G.fast Summit*, May 2016.

[144] R. Strobel, A. Barthelme, and W. Utschick, "Implementation aspects of nonlinear precoding for G.fast — coding and legacy receivers," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 111–115.

[145] R. D. Wesel and J. M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 824–831, March 1998.

[146] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[147] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *2014 IEEE International Symposium on Information Theory.* IEEE, 2014, pp. 2794–2798.

[148] A. Barthelme, R. Strobel, M. Joham, and W. Utschick, "Weighted MMSE Tomlinson-Harashima precoding for G.fast," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[149] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[150] S. Huberman, C. Leung, and T. Le-Ngoc, "Dynamic spectrum management (DSM) algorithms for multi-user xDSL," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 109–130, First 2012.

[151] J. W. Bos, C. Costello, M. Naehrig, and D. Stebila, "Post-quantum key exchange for the TLS protocol from the ring learning with errors problem," in *2015 IEEE Symposium on Security and Privacy*, May 2015, pp. 553–570.

[152] C. Ling and J. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct 2014.

[153] A. Lenstra, H. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.

[154] P. Nguyen and B. Vallée, *The LLL algorithm: Survey and applications*, ser. Information Security and Cryptography. Springer Berlin Heidelberg, 2009.

[155] D. Wubben, D. Seethaler, J. Jalden, and G. Matz, "Lattice reduction," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 70–91, May 2011.

[156] H. Minkowski, "Ueber die positiven quadratischen Formen und über kettenbruchähnliche Algorithmen." *Journal für die reine und angewandte Mathematik*, vol. 107, pp. 278–297, 1891.

[157] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.

[158] J. Conway and N. Sloane, *Sphere packings, lattices, and groups*, ser. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1993.

[159] G. Hanrot, X. Pujol, and D. Stehlé, "Algorithms for the shortest and closest lattice vector problems," in *International Conference on Coding and Cryptology.* Springer, 2011, pp. 159–190.

[160] C. Ling, "Towards characterizing the performance of approximate lattice decoding in MIMO communications," in *6th International ITG-Conference on Source and Channel Coding*, April 2006, pp. 1–6.

[161] R. Zhang, A. F. A. Rawi, L. D. Humphrey, and L. Hanzo, "Expanded constellation mapping for enhanced far-end-cross-talk cancellation in G.fast," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 56–59, Jan 2017.

[162] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537 – 544, Mar 2005.

[163] D. J. Ryan, I. B. Collings, I. V. L. Clarkson, and R. W. Heath, "Performance of vector perturbation multiuser MIMO systems with limited feedback," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2633 – 2644, Oct 2009.

[164] B. Helfrich, "Algorithms to construct Minkowski reduced and Hermite reduced lattice bases," *Theoretical Computer Science*, vol. 41, pp. 125–139, 1985.

[165] R. Kannan, "Improved algorithms for integer programming and related lattice problems," in *Proceedings of the fifteenth annual ACM symposium on Theory of computing.* ACM, 1983, pp. 193–206.

[166] W. Zhang, S. Qiao, and Y. Wei, "HKZ and Minkowski reduction algorithms for lattice-reduction-aided MIMO detection," *IEEE Trans. Signal Process*, vol. 60, no. 11, pp. 5963–5976, Nov 2012.

[167] A. Akhavi, "The optimal LLL algorithm is still polynomial in fixed dimension," *Theoretical Computer Science*, vol. 297, no. 1-3, pp. 3–23, 2003.

[168] X. Xie, X.-W. Chang, and M. A. Borno, "Partial LLL reduction," in *IEEE GlobalComm*, 2011.

[169] C. Stierstorfer and R. F. Fischer, "Lattice-reduction-aided Tomlinson–Harashima precoding for point-to-multipoint transmission," *AEU - International Journal of Electronics and Communications*, vol. 60, no. 4, pp. 328 – 330, 2006.

[170] X. W. Chang, J. Wen, and X. Xie, "Effects of the LLL reduction on the success probability of the Babai point and on the complexity of sphere decoding," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4915–4926, Aug 2013.

[171] D. Silva, G. Pivaro, G. Fraidenraich, and B. Aazhang, "On integer-forcing precoding for the Gaussian MIMO broadcast channel," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4476–4488, July 2017.

[172] S. Hong and G. Caire, "Reverse compute and forward: A low-complexity architecture for downlink distributed antenna systems," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 1147–1151.

[173] D. Micciancio, "Efficient reductions among lattice problems," in *Proceedings of the nineteenth annual ACM-SIAM symposium on discrete algorithms.* Society for Industrial and Applied Mathematics, 2008, pp. 84–93.

[174] S. Stern and R. F. H. Fischer, "Optimal factorization in lattice-reduction-aided and integer-forcing linear equalization," in *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, Feb 2017, pp. 1–6.

[175] R. Zamir, *Lattice coding for signals and networks: A structured coding approach to quantization, modulation, and multiuser information theory.* Cambridge University Press, 2014.

[176] U. Erez and R. Zamir, "Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct 2004.

[177] W. Kositwattanarerk, S. S. Ong, and F. Oggier, “Construction A of lattices over number fields and block fading (wiretap) coding,” *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2273–2282, May 2015.

[178] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, “Lattice-reduction-aided broadcast precoding,” *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057–2060, Dec 2004.

[179] U. Fincke and M. Pohst, “Improved methods for calculating vectors of short length in a lattice, including a complexity analysis,” *Mathematics of Computation*, vol. 44, no. 170, pp. 463–471, 1985.

[180] C. P. Schnorr and M. Euchner, “Lattice basis reduction: Improved practical algorithms and solving subset sum problems,” *Mathematical Programming*, vol. 66, no. 1, pp. 181–199, 1994.

[181] M. O. Damen, H. E. Gamal, and G. Caire, “On maximum-likelihood detection and the search for the closest lattice point,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct 2003.

[182] Z. Xie, C. K. Rushforth, R. T. Short, and T. K. Moon, “Joint signal detection and parameter estimation in multiuser communications,” *IEEE Trans. Commun.*, vol. 41, no. 8, pp. 1208–1216, Aug 1993.

[183] K. wai Wong, C. ying Tsui, R. S. . Cheng, and W. ho Mow, “A VLSI architecture of a K-best lattice decoding algorithm for MIMO channels,” in *2002 IEEE International Symposium on Circuits and Systems. Proceedings*, vol. 3, May 2002, pp. III–III.

[184] Z. Guo and P. Nilsson, “Algorithm and implementation of the K-best sphere decoding for MIMO detection,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491–503, March 2006.

[185] T. Fukatani, R. Matsumoto, and T. Uyematsu, “Two methods for decreasing the computational complexity of the MIMO ML decoder,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 87, no. 10, pp. 2571–2576, 2004.

[186] K. Lee and J. Chun, “ML symbol detection based on the shortest path algorithm for MIMO systems,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5477–5484, Nov 2007.

[187] R. Y. Chang and W. Chung, “Best-first tree search with probabilistic node ordering for MIMO detection: Generalization and performance-complexity tradeoff,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 780–789, February 2012.

[188] B. Hassibi and H. Vikalo, “On the sphere-decoding algorithm I. Expected complexity,” *IEEE Trans. Signal Process*, vol. 53, no. 8, pp. 2806 – 2818, Aug 2005.

[189] V. Guruswami, D. Micciancio, and O. Regev, “The complexity of the covering radius problem,” *Computational Complexity*, vol. 14, no. 2, pp. 90–121, 2005.

[190] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point search in lattices,” *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201 – 2214, Aug 2002.

[191] Ö. Dagdelen and M. Schneider, *Parallel enumeration of shortest lattice vectors.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 211–222.

[192] K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: A practical approach to global optimization.* Springer Science & Business Media, 2006.

[193] C. Studer, D. Seethaler, and H. Bölcskei, “Finite lattice-size effects in MIMO detection,” in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, Oct 2008, pp. 2032–2037.

[194] H. Vetter and Y. Sun, “Legacy user support in multiuser MIMO with vector perturbation precoding,” in *Proc. Indoor and Mobile Radio Communications 21st Annual IEEE Int. Symp. Personal*, Sep. 2010, pp. 887–892.

[195] C. Masouros, M. Sellathurai, and T. Ratnarajah, “Vector perturbation based on symbol scaling for limited feedback MISO downlinks,” *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 562–571, Feb. 2014.

[196] J. Jalden and B. Ottersten, “On the complexity of sphere decoding in digital communications,” *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, April 2005.

[197] M. Mohaisen, B. Hui, K. Chang, S. Ji, and J. Joung, “Fixed-complexity vector perturbation with block diagonalization for MU-MIMO systems,” in *2009 IEEE 9th Malaysia International Conference on Communications (MICC)*, Dec 2009, pp. 238–243.

[198] C. Zheng, X. Chu, J. McAllister, and R. Woods, “Real-valued fixed-complexity sphere decoder for high dimensional QAM-MIMO systems,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4493–4499, Sept 2011.

[199] X. Chen, G. He, and J. Ma, “VLSI implementation of a high-throughput iterative fixed-complexity sphere decoder,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 60, no. 5, pp. 272–276, May 2013.

[200] M. Guenach, “Novel lattice reduction algorithms: Precoder reduction and vector perturbation tradeoffs,” *IEEE Commun. Lett.*, p. 1, 2018.

[201] D. Micciancio and P. Voulgaris, “A deterministic single exponential time algorithm for most lattice problems based on Voronoi cell computations,” *SIAM Journal on Computing*, vol. 42, no. 3, pp. 1364–1391, 2013.

[202] A. Becker, N. Gama, and A. Joux, “A sieve algorithm based on overlattices,” *LMS Journal of Computation and Mathematics*, vol. 17, no. A, pp. 49–70, 2014.

[203] W. S. Chua, C. Yuen, and F. Chin, “A continuous vector-perturbation for multi-antenna multi-user communication,” in *VTC2007-Spring*. Dublin: IEEE, Apr 2007, pp. 1806 – 1810.

[204] E. Y. Kim and J. Chun, “Optimum vector perturbation minimizing total MSE in multiuser MIMO Downlink,” in *Proc. IEEE Int. Conf. Communications*, vol. 9, Jun. 2006, pp. 4242–4247.

[205] M. Mazrouei-Sebdani and W. A. Krzymień, “On MMSE vector-perturbation precoding for MIMO broadcast channels with per-antenna-group power constraints,” *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3745–3751, Aug. 2013.

[206] Y. Avner, B. M. Zaidel, and S. S. Shitz, “On vector perturbation precoding for the MIMO Gaussian broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5999–6027, Nov. 2015.

[207] D. J. Ryan, “Space-time vector perturbation precoding,” in *2010 IEEE International Conference on Communications*, May 2010, pp. 1–6.

[208] D. A. Karpuk, A. Barreal, O. W. Gnilke, and C. Hollanti, “Nested lattice codes for vector perturbation systems,” *arXiv preprint arXiv:1604.07048*, 2016.

[209] W. Yu, G. Ginis, and J. M. Cioffi, “Distributed multiuser power control for digital subscriber lines,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, June 2002.

[210] Y. Xu, T. Le-Ngoc, and S. Panigrahi, “Selective iterative waterfilling for digital subscriber lines,” *EURASIP J. Adv. Signal Process*, no. 59068, 2007.

[211] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, “Autonomous spectrum balancing for digital subscriber lines,” *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4241–4257, Aug 2007.

[212] P. Tsiaflakis, M. Diehl, and M. Moonen, “Distributed spectrum management algorithms for multiuser DSL networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4825–4843, Oct 2008.

[213] J. Verdyck and M. Moonen, “Dynamic spectrum management in digital subscriber line networks with unequal error protection requirements,” *IEEE Access*, vol. 5, pp. 18 107–18 120, 2017.

[214] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, June 2007.

[215] A. R. Forouzan, M. Moonen, J. Maes, and M. Guenach, "Joint level 2 and 3 dynamic spectrum management for downstream DSL," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3111–3122, October 2012.

[216] P. Tsiaflakis, J. Vangorp, J. Verlinden, and M. Moonen, "Multiple access channel optimal spectrum balancing for upstream DSL transmission," *IEEE Commun. Lett.*, vol. 11, no. 4, pp. 398–300, April 2007.

[217] V. Le Nir, M. Moonen, J. Verlinden, and M. Guenach, "Optimal power allocation for downstream xDSL with per-modem total power constraints: broadcast channel optimal spectrum balancing (BC-OSB)," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 690–697, Feb 2009.

[218] J. Neckebroek, M. Moeneclaey, W. Coomans, and M. Guenach, "Novel bitloading algorithms for coded G.fast DSL transmission with linear and nonlinear precoding," in *IEEE International Conference on Communications.* London: IEEE, June 2015, pp. 945 – 951.

[219] A. Li and C. Masouros, "A constellation scaling approach to vector perturbation for adaptive modulation in MU-MIMO," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 289–292, June 2015.

[220] E. Viterbo and E. Biglieri, "Computing the Voronoi cell of a lattice: the diamond-cutting algorithm," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 161–171, Jan 1996.

[221] M. Dutour Sikirić, A. Schürmann, and F. Vallentin, "Complexity and algorithms for computing Voronoi cells of lattices," *Math. comp.*, vol. 78, no. 267, pp. 1713–1731, 2009.

[222] X. chun Zhang, H. Yu, and G. Wei, "Exact symbol error probability of cross-QAM in AWGN and fading channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, no. 1, p. 917954, Nov 2010.

[223] J. M. Cioffi, "A multicarrier primer," *ANSI T1E1*, vol. 4, pp. 91–157, 1991.

[224] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.

[225] P. Tsiaflakis, I. Necoara, J. A. K. Suykens, and M. Moonen, "Improved dual decomposition based optimization for DSL dynamic spectrum management," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2230–2245, April 2010.

[226] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, “Optimal multiuser spectrum balancing for digital subscriber lines,” *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, May 2006.

[227] S. Wei, L. Youming, and Y. Miaoliang, “Low-complexity grouping spectrum management in multi-user DSL networks,” in *2009 WRI International Conference on Communications and Mobile Computing*, vol. 1, Jan 2009, pp. 381–385.

[228] R. Cendrillon and M. Moonen, “Iterative spectrum balancing for digital subscriber lines,” in *IEEE International Conference on Communications, 2005. ICC 2005. 2005*, vol. 3, May 2005, pp. 1937–1941.

[229] R. Lui and W. Yu, “Low-complexity near-optimal spectrum balancing for digital subscriber lines,” in *IEEE International Conference on Communications, 2005. ICC 2005. 2005*, vol. 3, May 2005, pp. 1947–1951.

[230] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, “Iterative water-filling for Gaussian vector multiple access channels,” in *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No.01CH37252)*, June 2001, pp. 322–.

[231] J. Papandriopoulos and J. S. Evans, “Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks,” in *2006 IEEE International Conference on Communications*, vol. 7, June 2006, pp. 3270–3275.

[232] W. Lanneer, J. Verdyck, P. Tsiaflakis, J. Maes, and M. Moonen, “Vectoring-based dynamic spectrum management for G.fast multi-user full-duplex transmission,” in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–5.

[233] D. Z. Filho, R. R. Lopes, R. Ferrari, R. Suyama, and B. Dortschy, “Bit loading for precoded DSL systems,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07*, vol. 3, April 2007, pp. III–353–III–356.

[234] J. Campello, “Practical bit loading for DMT,” in *1999 IEEE International Conference on Communications (Cat. No. 99CH36311)*, vol. 2, 1999, pp. 801–805 vol.2.

[235] T. N. Vo, K. Amis, T. Chonavel, and P. Siohan, “A computationally efficient discrete bit-loading algorithm for OFDM systems subject to spectral-compatibility limits,” *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2261 – 2272, June 2015.

[236] H. Chen, R. G. Maunder, and L. Hanzo, “A survey and tutorial on low-complexity turbo coding techniques and a holistic hybrid ARQ design example,” *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1546–1566, Fourth 2013.

[237] *NR; Multiplexing and channel coding*, 3GPP, June 2019, ver. 15.6.0.

[238] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1,” in *Proceedings of ICC ’93 - IEEE International Conference on Communications*, vol. 2, May 1993, pp. 1064–1070.

[239] R. G. Gallager, “Low density parity check codes (Ph. D. dissertation),” *Massachusetts Institute of Technology, Cambridge, Mass, USA*, 1963.

[240] T. Richardson and S. Kudekar, “Design of low-density parity check codes for 5G New Radio,” *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 28–34, March 2018.

[241] R. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.

[242] F. R. Kschischang, B. J. Frey, and H. . Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[243] W. Zhong and J. Garcia-Frias, “LDGM codes for channel coding and joint source-channel coding of correlated sources,” *EURASIP J. Adv. Signal Process*, pp. 942–953, Jan 2005.

[244] A. J. Viterbi, “An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb 1998.

[245] L. Hanzo, S. X. Ng, W. Webb, and T. Keller, *Quadrature amplitude modulation: From basics to adaptive trellis-coded, turbo-equalised and space-time coded OFDM, CDMA and MC-CDMA systems.* IEEE Press-John Wiley, 2004.

[246] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo coding, turbo equalisation, and space-time coding.* Wiley Online Library, 2002.

[247] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 429–445, March 1996.

[248] T. Etzion, A. Trachtenberg, and A. Vardy, “Which codes have cycle-free Tanner graphs?” *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2173–2181, Sep. 1999.

[249] R. Yazdani and M. Ardakani, “Waterfall performance analysis of finite-length LDPC codes on symmetric channels,” *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3183–3187, Nov 2009.

[250] T. J. Richardson and R. L. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb 2001.

[251] S. ten Brink, “Convergence behavior of iteratively decoded parallel concatenated codes,” *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727 – 1737, Oct 2001.

[252] S. ten Brink, G. Kramer, and A. Ashikhmin, “Design of low-density parity-check codes for modulation and detection,” *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, April 2004.

[253] M. El-Hajjar and L. Hanzo, “EXIT charts for system design and analysis,” *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 127–153, First 2014.

[254] E. Y. Rocher and R. L. Pickholtz, “An analysis of the effectiveness of hybrid transmission schemes,” *IBM Journal of Research and Development*, vol. 14, no. 4, pp. 426–433, July 1970.

[255] D. Chase, “A combined coding and modulation approach for communication over dispersive channels,” *IEEE Trans. Commun.*, vol. 21, no. 3, pp. 159–174, March 1973.

[256] G. Benelli, “An ARQ scheme with memory and soft error detectors,” *IEEE Trans. Commun.*, vol. 33, no. 3, pp. 285–288, March 1985.

[257] D. Mandelbaum, “An adaptive-feedback coding scheme using incremental redundancy (Corresp.),” *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 388–389, May 1974.

[258] S. Lin and P. Yu, “A hybrid ARQ scheme with parity retransmission for error control of satellite channels,” *IEEE Trans. Commun.*, vol. 30, no. 7, pp. 1701–1719, July 1982.

[259] P. Larsson, L. K. Rasmussen, and M. Skoglund, “Throughput analysis of ARQ schemes in Gaussian block fading channels,” *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2569–2588, July 2014.

[260] Soonyoung Kim, Jinsu Chang, and Moon Ho Lee, “Simple iterative decoding stop criterion for wireless packet transmission,” *Electronics Letters*, vol. 36, no. 24, pp. 2026–2027, Nov 2000.

[261] F. Li and A. Wu, “On the new stopping criteria of iterative turbo decoding by using decoding threshold,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5506–5516, Nov 2007.

[262] J. Y. Chen, L. Zhang, and J. Qin, “Average-entropy variation in iterative decoding of turbo codes and its application,” *Electronics Letters*, vol. 44, no. 22, pp. 1314–1315, October 2008.

[263] Zheng Ma, Wai Ho Mow, and Pingzhi Fan, “On the complexity reduction of turbo decoding for wideband CDMA,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 353–356, March 2005.

[264] M. Rovini and A. Martinez, “Efficient stopping rule for turbo decoders,” *Electronics Letters*, vol. 42, no. 4, pp. 235–236, Feb 2006.

[265] J. Jalden, D. Seethaler, and G. Matz, “Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 2685–2688.

[266] L. G. Barbero and J. S. Thompson, “Fixing the complexity of the sphere decoder for MIMO detection,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2131–2142, June 2008.

[267] Y. Gu and K. K. Parhi, “High-speed architecture design of Tomlinson–Harashima precoders,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 9, pp. 1929–1937, Sep. 2007.

[268] L. Zhang, Y. Cai, R. C. de Lamare, and M. Zhao, “Robust multibranch Tomlinson–Harashima precoding design in amplify-and-forward MIMO relay systems,” *IEEE Trans. Commun.*, vol. 62, no. 10, pp. 3476–3490, Oct 2014.

[269] Y. Luo and S. Qiao, “A parallel LLL algorithm.” in *C3S2E.* Citeseer, 2011, pp. 93–101.

[270] C. Ling, W. H. Mow, and N. Howgrave-Graham, “Reduced and fixed-complexity variants of the LLL algorithm for communications,” *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1040–1050, March 2013.

[271] C. I, S. Han, Z. Xu, S. Wang, Q. Sun, and Y. Chen, “New paradigm of 5G wireless Internet,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 474–482, March 2016.

[272] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 65–75, Nov 2014.

Index

Author Index