IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 10, 2018, accepted September 3, 2018, date of publication September 19, 2018,
date of current version October 17, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871026

Approximate Perturbation Aided Lattice Encoding (APPLE) for G.fast and Beyond

YANGYISHI ZHANG[1], RONG ZHANG[1], (Senior Member, IEEE), ANAS F. AL RAWI[2], (Member, IEEE), AND LAJOS HANZO[1], (Fellow, IEEE)

[1]Department of Electronics and Computer Science, The University of Southampton, Southampton SO17 1BJ, U.K.
[2]BT Group Plc, Ipswich IP5 3RE, U.K.

Corresponding author: Lajos Hanzo (lh@ecs.soton.ac.uk)

This work was supported in part by the EPSRC Projects under Grants EP/Noo4558/1 and EP/PO34284/1, in part by the European Research Council's Advanced Fellow Grant QuantCom, and in part by the Royal Society under Grant IF170002.

ABSTRACT G.fast suffers from strong far-end crosstalk at high frequencies in cable binders containing a large number of twisted copper pairs. For the 212-MHz G.fast spectrum, the power penalty incurred by the conventional zero-forcing precoding-based linear vectoring (LV) scheme is far more substantial than it was over the 30-MHz VDSL2 spectrum. In this paper, we propose a novel non-LV (NLV) scheme based on Babai's nearest plane approximation of the closest lattice point problem on the reduced lattice basis. Similar to the conventional Tomlinson–Harashima precoding (THP)-based NLV, the proposed approximate perturbation aided lattice encoding (APPLE) scheme closely approaches the dirty paper coding capacity which provided that the system employs a fully rate-adaptive power allocation policy per tone per pair. However, if the system employs a scalar power policy that is only rate-adaptive with respect to each tone, APPLE becomes capable of achieving a higher throughput per binder than THP. APPLE's transmitter complexity is considerably lower than that of the conventional lattice encoding schemes (e.g., vector perturbation) and comparable to that of THP.

INDEX TERMS G.fast, far-end crosstalk, vectoring, transmit precoding, lattice reduction, closest lattice point, nonconvex optimization.

I. INTRODUCTION

G.fast has emerged as the new standard of the copper based digital subscriber line (DSL) technology. The motivation of G.fast is to achieve a substantial throughput boost, while preserving the fiber-to-the-cabinet (FTTC) or fiber-to-the-distribution-point (FTTdp) network infrastructure, which aims for fiber-level performance without the excessive cost of fiber laying over the subscribers' *last mile* coverage. Since the DSL standard relies on discrete multitone (DMT) transmission, the research community generally resorted to spectrum expansion to satisfy the need for ever-increasing data rates. The operational G.fast profile only exploits the first half (up to 106 MHz) of the proposed 212 MHz spectrum, which is claimed to support Gigabit throughput rates.

However, future broadband access anticipates multi-Gigabit throughputs, which necessitates the exploitation of the full 212 MHz spectrum and possibly an even further spectrum expansion. However, since DSL relies on pairs of twisted copper wires, its performance is detrimentally affected by the inevitable electromagnetic coupling, i.e. crosstalk, between closely spaced pairs. The DSL termination unit (DTU), which is deployed in either a street-corner cabinet for FTTC or at a distribution point for FTTdp, serves a number of customer premise equipment (CPE) within the same street. The pairs connecting these CPEs are usually bundled together into a single large-diameter DSL binder. As a DSL binder is coated with a screening surface, it is the crosstalk within the binder, which is the dominant form of impairment. The impact of crosstalk increases rapidly with frequency, which dictates that the high-frequency tones exhibit considerably worse channel quality than those of low frequency. Far-end crosstalk (FEXT) refers to the interference imposed on the distant receiver. As demonstrated by the measurements of Fig. 1, the channel frequency response of both

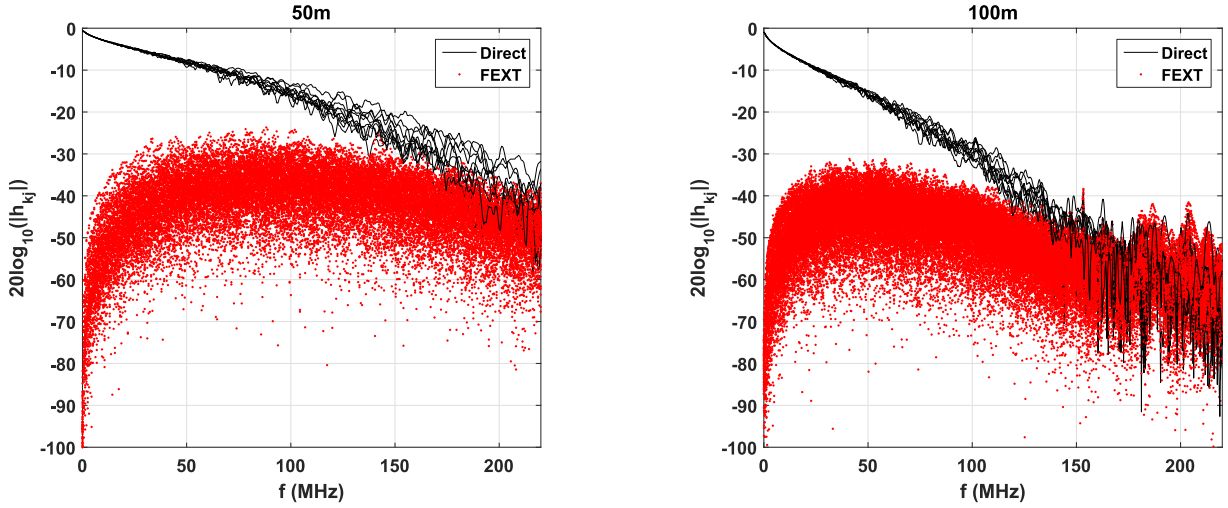


FIGURE 1. Direct and FEXT path channel frequency responses over the 212 MHz bandwidth for 10-pair cables of length 50m and 100m, respectively. For both the direct and the FEXT paths, the channel gain decreases at a much faster rate for longer cables. In order to have non-negative signal-to-interference ratio at the receiver end, the channel gain of the direct path must be at least 10 dB higher than the average of FEXT because the number of FEXT paths dominates within the cable.

the direct paths and the FEXT paths decreases over frequency. However, since the *number* of FEXT paths dominates, the total FEXT level experienced per direct path actually increases over frequency. The increase of the relative FEXT exhibits a much faster rate for long cables than for short ones.

In general, FEXT is quasi-static over time and can be mitigated by vectoring [1]. When a DSL binder is exclusively managed by a single Internet service provider (ISP), vectoring is more or less capable of eliminating FEXT. The FEXT-cancelation effect of vectoring is achieved by transmit precoding (TPC) for downstream communication. For frequencies beyond 106 MHz, the channel inversion type of linear vectoring (LV), including both the regularized channel inversion of [2] and the diagonal precoding of [3], would impose significantly higher signal-to-noise-ratio (SNR) penalty than below 106 MHz of the G.fast spectrum, subject to a specific per-pair *transmit power spectral density* (TxPSD) mask and to a particular per-pair *aggregate transmit power* (ATP) limit [4]. On one hand, the TxPSD mask is lower for G.fast frequencies beyond 106 MHz. On the other, inverting the ill-conditioned high-frequency channels results in higher PSD improvement than it does for low-frequency bands.

The SNR penalty caused by LV at high frequencies results in a significantly higher information loss. Modulo type non-linear vecoring (NLV) schemes may be used to mitigate this loss. The optimal vectoring scheme is expected to achieve the throughput given by the dirty paper coding (DPC) scheme of [5]. The Tomlinson-Harashima precoding (THP) [6], [7] based NLV reduces the SNR penalty by successive interference cancelation (SIC) at the transmitter side. Despite its near-optimality regarding the throughput per binder under a fully adaptive power policy, poor user fairness is an issue commonly associated with THP. On the other hand, since in practice DSL systems are susceptible to non-stationary impairments such as the impulsive noise, usually suboptimal power policies are employed in favor of agile reconfigurability. Under a scalar power policy, where the power allocation is uniform for all customers at each tone, the overall performance of THP is predominantly influenced by that of the worst-case customer on a per-tone basis.

In this paper, we aim for solving the problems associated with the THP-based NLV scheme. *We propose a novel vectoring scheme, which is termed as approximate perturbation aided lattice encoding (APPLE). Our scheme achieves a higher throughput per binder than NLV under a scalar power allocation policy at a modest complexity increase. For time-invariant channels, APPLE is near-optimal, just like NLV under the fully optimized power allocation policy.* The remainder of the paper is structured as follows. Section II introduces the main components of a vectored multi-user DSL system, including the basics of the conventional LV and NLV schemes as well as the related lattice theory preliminaries. Section III details the operation of APPLE with respect to its main components. Section IV characterizes the performance of APPLE and its advantage over conventional vectoring schemes. Finally, Section V concludes the paper.

II. SYSTEM MODEL

A DMT-based multi-user G.fast system consists of a group of multiple-input-multiple-output (MIMO) subsystems of identical dimensionality. Considering a binder having K twisted pairs modulated with $\mathcal{I}$ DMT tones, the received signal becomes:

$$\boldsymbol{y}^i = \boldsymbol{H}^i\boldsymbol{x}^i + \boldsymbol{n}^i \quad (i = 1, 2, \ldots, \mathcal{I}), \tag{1}$$

where the $(K \times 1)$-element vectors $\boldsymbol{y}^i$, $\boldsymbol{x}^i$ and $\boldsymbol{n}^i$ represent the received symbol vector, transmitted symbol vector and additive white Gaussian noise (AWGN) imposed on the tone indexed by i and $E\{\|\boldsymbol{n}^i\|^2\} = \sigma_n^2$. The $(K \times K)$-element

transfer matrix $\boldsymbol{H}^i$ is the DMT channel of the ith tone. When the cyclic prefix used for DMT is sufficiently long, these channels are considered independent and thus do not inflict any inter-tone interference. The diagonal entries of $\boldsymbol{H}^i$ are the direct link channel coefficients corresponding to each individual pair, while the off-diagonal entries are the FEXT channel coefficients. On the frequency band below 106 MHz, $\boldsymbol{H}^i$ is approximately diagonal and therefore near-orthogonal. However, for the freuqency band beyond 106 MHz, $\boldsymbol{H}^i$ no longer exhibits diagonal dominance. In G.fast, since the symbols are all drawn from quadrature amplitude modulated (QAM) constellations [8], we will consider (1) as a complex-valued system.

The main purpose of downstream vectoring is to maximize the throughput of a DSL binder via a *FEXT canceller* and a *power controller*. Additionally, modulo type vectoring schemes also utilize back end *signal coolant*.[1] When relying on an optimized choice of the coolant-controller-canceller trio, each CPE's received symbol y_k^i becomes mutually independent. Therefore the equalized symbol vector $\boldsymbol{z}^i$ can be readily obtained from $\boldsymbol{y}^i$ without multi-stream equalization. The general structure of a vectored DTU transmitter is given by a cascade of the three modules formulated as:

$$\boldsymbol{x}^i = \boldsymbol{\Omega}^i \boldsymbol{A}^i \tilde{\boldsymbol{u}}^i. \tag{2}$$

1) **Signal Coolant**. Applying the back-end signal coolant $\boldsymbol{l}^i$ to the users' message symbol vector $\boldsymbol{u}^i$ will reduce the power boost incurred by the FEXT canceller $\boldsymbol{\Omega}^i$. The coolant $\boldsymbol{l}^i$ results in a remapping of $\boldsymbol{u}^i$ to the expanded signal space. The above remapping transforms $\boldsymbol{u}^i$ into the perturbed symbol vector $\tilde{\boldsymbol{u}}^i$. The distributed CPE receivers can remove the effect of $\boldsymbol{l}^i$ and detect each received symbol y_k^i independently without the full knowledge of $\boldsymbol{l}^i$, which is analogous to the physical phenomenon of 'vaporization' of liquid coolant.
2) **Power Controller**. Represented by the non-negative real-valued diagonal matrix $\boldsymbol{A}^i$, it is used for coordinating the transmit power allocated to each message symbol u_k^i 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 for tone i corresponding to the controller $\boldsymbol{A}^i$ is given by $\boldsymbol{P}^i = (\boldsymbol{A}^i)^T \boldsymbol{A}^i$.
3) **FEXT Canceller**. The front-end FEXT canceller $\boldsymbol{\Omega}^i$ 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 channel $\boldsymbol{G}^i = (\boldsymbol{H}^i)^{-1}$. Hence the equivalent channel between the message symbol vector $\boldsymbol{u}^i$ and the received symbol vector $\boldsymbol{y}^i$ is interference-free. In certain vectoring schemes, such as THP, the CPE receivers may require an additional equalizer to compensate for the uncorrelated direct channels.

The initialization of G.fast systems, which is used for informing the DTU of the downstream *channel state information* (CSI), currently relies on the training-based vectoring feedback mechanism of [9] defined for the *frequency division duplexing* based VDSL2 deployment. However, the *time division duplexing* mode used in G.fast exhibits bi-directional channel symmetry, which suggests that the downstream CSI can be inferred from the upstream CSI. Since DSL channels are essentially time invariant, we will assume that the DTU has perfect knowledge of the downstream CSI, which is acquired during the initialization process.

A. PRELIMINARIES

The vectoring scheme of (2) invoked for the system given in (1) can be viewed as a lattice-based mapping spanning $\mathcal{I}$ independent Euclidean spaces, where each Euclidean space contains K complex-valued dimensions. Within one instance of the Euclidean space $\mathbb{C}^K$, a K-dimensional lattice is a subspace isomorphic to the subspace of Gaussian integers $\mathbb{G}^K$, therefore $\mathbb{G}^K$ itself also constitutes a lattice. Lattices are frequently used as tools for investigating correlated signaling either across multiple spatial locations (as in our scenario), or time instances or alternatively frequency bands, all of which are characterized by the *dimensions* of the corresponding lattice. In fact, lattice-based coding schemes generally have provable optimality [10], [11]. A comprehensive introduction to the fundamentals of lattice theory and its general applications may be found in [12]. Without loss of generality, a lattice is defined as follows:

- **Lattice**. A $(K \times K)$-element matrix $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_K]$ having linearly independent columns generates a K-dimensional lattice $\mathcal{L}(\boldsymbol{G})$. The columns of the **generator matrix** $\boldsymbol{G}$ constitute a **basis** of $\mathcal{L}(\boldsymbol{G})$. The basis vectors constitute a parallelotope of volume $V(\boldsymbol{G}) = \sqrt{\det(\boldsymbol{G}^H \boldsymbol{G})}$.

Since integer-operand algorithms are mostly based on real-valued integers, the K-dimensional complex-valued Euclidean space $\mathbb{C}^K$ must be decoupled into the $2K$-dimensional real-valued Euclidean space $\mathbb{R}^{2K}$. The subspace $\mathbb{G}^K$ is subsequently transformed into the subspace of real integers $\mathbb{Z}^{2K}$. This is accomplished by the following transformation of matrices and vectors:

$$\boldsymbol{G} \rightarrow \begin{bmatrix} \Re(\boldsymbol{G}) & -\Im(\boldsymbol{G}) \\ \Im(\boldsymbol{G}) & \Re(\boldsymbol{G}) \end{bmatrix}, \quad \boldsymbol{u} \rightarrow \begin{bmatrix} \Re(\boldsymbol{u}) \\ \Im(\boldsymbol{u}) \end{bmatrix}. \tag{3}$$

The real-valued system characterizes the geometric properties of lattices more explicitly, but it is equivalent to the original system. It is important to note that any lattice $\mathcal{L}(\boldsymbol{G})$ of more than two real dimensions has an infinite set of legitimate basis, but the corresponding parallelotopes all have identical volume. However, a basis constituted by short vectors is in general considered to be of better quality than one containing long vectors. The commonly accepted *orthogonality defect* criterion for assessing the quality of a lattice

[1] The back end modulo operation, as will be discussed later, typically serves the purpose of reducing precoded signal power at the transmitter's output, akin to how coolant prevents excessive temperature increase in motors and reactors.

basis $\boldsymbol{G}$ is defined as:

$$\delta(\boldsymbol{G}) = \frac{\prod_{k=1}^{K} \|\boldsymbol{g}_k\|}{V(\boldsymbol{G})}, \tag{4}$$

where $\delta(\boldsymbol{G}) \geq 1$ and the equality is true if and only if $\boldsymbol{G}$ is perfectly orthogonal. Since $V(\boldsymbol{G})$ is constant for all basis of the same lattice $\mathcal{L}(\boldsymbol{G})$, (4) shows that a shorter basis is also more orthogonal.

For the convenience of further analysis, we shall employ the uniformly distributed input (UDI) as the general message symbol alphabet, which is representative of the classic square-shaped and cross-shaped QAM constellations, complying with the choice of constellations in the G.fast standard [8].

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

$$\mathcal{U} = \{\boldsymbol{u} : |u_k| < \frac{1}{2}, k = 1, 2, \ldots, 2K\}, \tag{5}$$

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

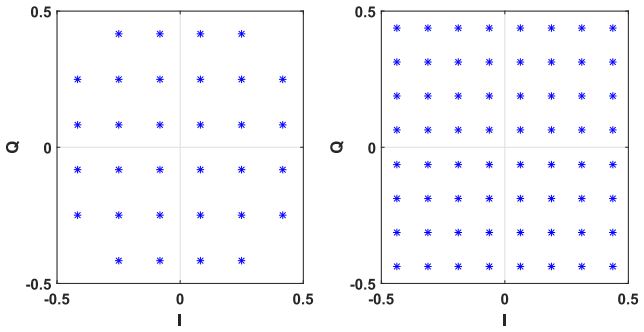


FIGURE 2. Uniformly Distributed Input (UDI) approximation of 32 QAM (left) and 64 QAM (right). Both the cross-shaped and the square-shaped constellations can be squeezed into the *unit square*. The average constellation energy increase due to the UDI approximation is uniquely determined by the QAM scheme itself. This is sometimes known as the power loss [13] due to modulo encoders.

The 2D UDI depicted in Fig. 2 approximates the conventional QAM constellations of any given modulation order. The power characteristics of a given QAM constellation, such as its peak-to-average power ratio and maximum symbol energy can be readily inferred from the UDI approximation associated with amplitude scaling. The required scaling is mostly a result of the convention that QAM constellations have an average energy constraint, rather than obeying the maximum amplitude constraint exhibited by UDI. However, it is also a consequence of the power loss [13] incurred by applying the signal coolant in THP and in other modulo-based encoders. For a real-valued vector $\boldsymbol{y}$, the (entry-wise) base-a *modulo reduction* $\boldsymbol{\Gamma}_a[\boldsymbol{y}]$, commonly associated with the signal coolant, is defined as:

$$\boldsymbol{\Gamma}_a[\boldsymbol{y}] = \boldsymbol{y} - a\lfloor \frac{\boldsymbol{y}}{a} + \frac{1}{2} \rfloor, \tag{6}$$

where $\lfloor \cdot \rfloor$ represents the (entry-wise) floor function. As seen in [14], modulo encoders require a common modulo base for each user's input alphabet. If $\boldsymbol{y}$ is drawn from the UDI, then each entry y_k naturally shares the common modulo base of $a = 1$. All modulo reductions in this paper are given in base-1, unless explicitly stated otherwise.

B. CONVENTIONAL VECTORING

Prior to the ratification of VDSL2, it was maintained that LV is near-optimal due to the diagonal-dominant structure of the CSI matrix, as well as owing to the low insertion loss of the direct channels. Under the common definition of vectoring, the signal coolant and FEXT canceller correlate only within the same DMT tone. Without loss of generality, let us limit the discussion within this section to a single DMT modulated tone. In LV based downstream transmission, the transmitted symbol vector $\boldsymbol{x}_{\text{LV}}$ and the equalized symbol vector $\boldsymbol{z}_{\text{LV}}$ are:

$$\boldsymbol{x}_{\text{LV}} = \boldsymbol{G}\boldsymbol{A}\boldsymbol{u}, \quad \boldsymbol{z}_{\text{LV}} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{n}, \tag{7}$$

where the FEXT canceller is represented by $\boldsymbol{G} = \boldsymbol{H}^{-1}$. Due to the power constraints represented by the TxPSD mask and the ATP budget, the legitimate transmit power policy is limited by $\boldsymbol{G}$.

In contrast to the 30 MHz VDSL2 spectrum, G.fast will occupy a bandwidth of 212 MHz. The channel quality of a high frequency tone suffers from the significant increase of the direct channel insertion loss and from the FEXT. The NLV scheme uses modulo-type signal cooling and therefore incurs a lower power penalty than LV. In NLV, the FEXT canceller $\boldsymbol{Q}$ of Fig. 3, which is a unitary matrix, is obtained via the triangular decomposition of $\boldsymbol{H}^H = \boldsymbol{Q}\boldsymbol{R}$. The diagonal matrix $\boldsymbol{D}$ is constructed using the diagonal entries of the *lower* triangular matrix $\boldsymbol{R}^H$. With feedback aided signal cooling, the transmitted vector of a NLV based DTU is given by:

$$\boldsymbol{x}_{\text{NLV}} = \boldsymbol{Q}\boldsymbol{A}\tilde{\boldsymbol{u}}, \quad \text{where } \tilde{\boldsymbol{u}} = \boldsymbol{\Gamma}[\boldsymbol{u} + (\boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{R}^H)\tilde{\boldsymbol{u}}], \tag{8}$$

and the corresponding equalized symbol vector is formulated as:

$$\boldsymbol{z}_{\text{NLV}} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{\Gamma}[\boldsymbol{D}^{-1}\boldsymbol{n}]. \tag{9}$$

Comparing the expressions in (7) and (9), given the same non-orthogonal channel matrix $\boldsymbol{H}$, the same noise power and power constraint, it is readily seen that (9) results in a higher SNR, since the power enhancement achieved by a *full channel inversion* $\boldsymbol{G}$ is higher than that of a *partial inversion* $\boldsymbol{D}^{-1}$ of the diagonal/direct channels.

It has already been shown that using THP as a TPC technique for UDI message symbols can in theory approach the Shannon capacity of parallel Gaussian channels upto a small discrepancy of 1.53 dB within the high SNR regime. This SNR-loss is owing to the shaping loss [13]. However, under practical circumstances, the performance of vectored DSL heavily depends on the validity of the assumption that the DTU has perfect non-causal knowledge of the downstream CSI. Additionally, impulsive noise and/or alien FEXT

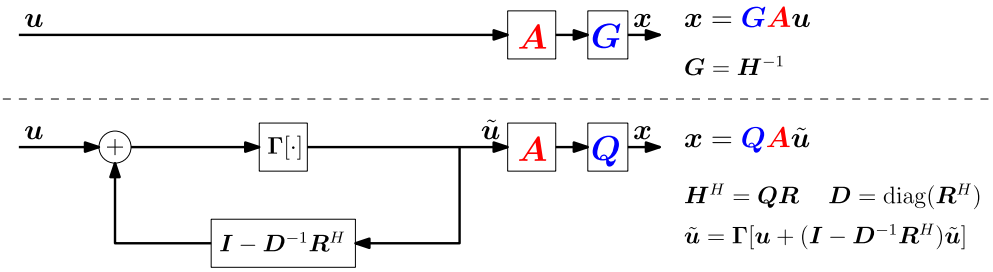


FIGURE 3. Conventional LV based DTU (top) and conventional NLV based DTU (bottom) operating on a single tone. The three components of a vectored transmitter are: front-end linear FEXT canceller (blue), real-valued power controller (red) and back-end signal coolant.

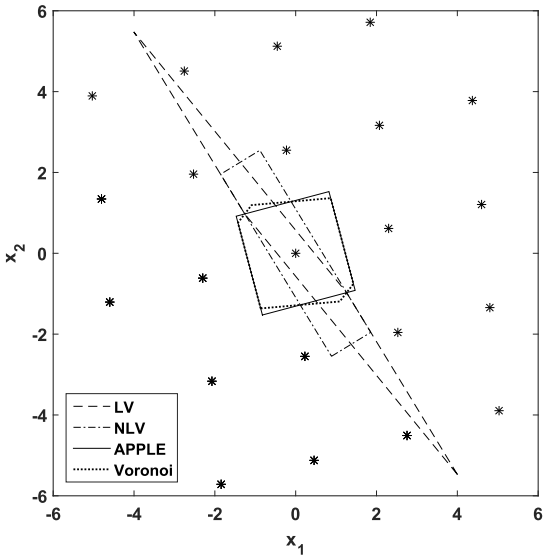


FIGURE 4. The Fundamental Mapping Region of each vectoring scheme and the Voronoi Cell for a two-user system in $\mathbb{R}^2$. If the special configuration of $\boldsymbol{A} = \boldsymbol{I}$ is invoked for the power controller, then each vectoring mapping region is respectively congruent to the fundamental mapping region.

caused by other interfering sources operating within the same frequency band will also impact the performance of vectoring. Therefore fully vectored transmission with (information-theoretic) optimal power policy is not used in practice.

C. VECTORING MAPPING REGION

The mapping region representation intuitively characterizes the most significant power transfer characteristics of a vectoring scheme, such as the TxPSD exhibited by the vectored transmitter. By assuming the power controller of Eq. (7) to be an identity matrix $\boldsymbol{A} = \boldsymbol{I}$, it is plausible that LV maps a given vector $\boldsymbol{u}$ of UDI symbols, as a point in $\mathbb{R}^{2K}$, onto an origin-centered *fundamental* parallelotope *mapping region* $\boldsymbol{G}\mathcal{U}$, 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 NLV is geometrically the orthotope $\boldsymbol{Q}\boldsymbol{D}^{-1}\mathcal{U}$ obtained by a shear transformation of the parallelotope $\boldsymbol{G}\mathcal{U}$. An example in $\mathbb{R}^2$ is illustrated in Fig. 4. 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 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 vectoring mapping region be the (*origin-centered*) K-dimensional polytope $\mathcal{P}$. Then we can list some of these relevant geometric properties as follows.

1) SECOND CENTRAL MOMENT

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

$$\sigma_{\mathcal{P}}^2 = \int_{\mathcal{P}} \frac{\|\boldsymbol{x}\|^2}{V(\mathcal{P})}\, d\boldsymbol{x} = \int_{\mathcal{P}} \sum_{k=1}^{K} \frac{|x_k|^2}{V(\mathcal{P})}\, d\boldsymbol{x}, \tag{10}$$

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:

$$\xi_{\mathcal{P}} = \int_{\mathcal{P}} \max_{1\le k\le K} \frac{|x_k|^2}{V(\mathcal{P})}\, d\boldsymbol{x}. \tag{11}$$

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.

2) MAXIMUM EUCLIDEAN SPREAD

The TxPSD mask and ATP limit in [4] 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 $\boldsymbol{v} = [v_1, v_2, \ldots, v_K]$, 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:

$$v_k = \max_{\boldsymbol{x}\in\mathcal{P}} \|\boldsymbol{x}\cos\theta_k\| \quad (k = 1, 2, \ldots, K), \tag{12}$$

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 (12) 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. 5 demonstrates the difference between the two cases.

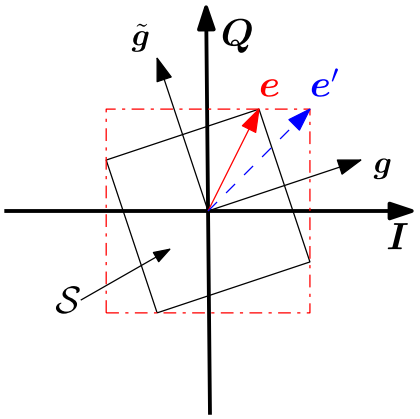


FIGURE 5. The Maximum Euclidean Spread of a system $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{u}$ in $\mathbb{R}^2$ decoupled from a system $\boldsymbol{x} = \boldsymbol{g}\boldsymbol{u}$ in $\mathbb{C}$, where we have $\boldsymbol{G} = [\boldsymbol{g}, \tilde{\boldsymbol{g}}]$, $\|\boldsymbol{g}\| = \|\tilde{\boldsymbol{g}}\|$ 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 review the two-user example of Fig. 4 in more detail, where both users transmit a 1D signal, i.e. PAM symbols. Observe that the MES per real dimension is characterized by a minimum-sized bounding box (e.g. the dot dashed boundary of Fig. 5) confining the vectoring mapping region $\mathcal{S}$. 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. 5). 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. 1.

Algorithm 1 Maximum Euclidean Spread per Complex Dimension

1 **Input**: An origin-centered N-dimensional parallelotope $\mathcal{P}$ whose edges are given by the columns of the $N \times N$ complex-valued matrix $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_N]$;
2 **Output**: N-vector $\boldsymbol{v} = [v_1, v_2, \ldots, v_N]$ indicating the MES per complex dimension;
3 $\bar{\boldsymbol{G}} = \{[\Re(\boldsymbol{G})^T, \Im(\boldsymbol{G})^T]^T, [-\Im(\boldsymbol{G})^T, \Re(\boldsymbol{G})^T]^T\}$;
4 **for** $m = 1, 2, \ldots, 2N$ **do**
5 **for** $n = 1, 2, \ldots, 2N$ **do**
6 **if** $\bar{\boldsymbol{g}}_n(m) < 0$ **then**
7 $\boldsymbol{s}_n \leftarrow -\bar{\boldsymbol{g}}_n/2$;
8 **else**
9 $\boldsymbol{s}_n \leftarrow \bar{\boldsymbol{g}}_n/2$;
10 $\boldsymbol{q}_m \leftarrow \sum_{n=1}^{2N} \boldsymbol{s}_n$;
11 **if** $m \leq N$ **then**
12 $v_m \leftarrow \sqrt{\boldsymbol{q}_m^2(m) + \boldsymbol{q}_m^2(m+N)}$
13 **else if** $v_{m-N} < \sqrt{\boldsymbol{q}_m^2(m) + \boldsymbol{q}_m^2(m-N)}$ **then**
14 $v_{m-N} \leftarrow \sqrt{\boldsymbol{q}_m^2(m) + \boldsymbol{q}_m^2(m-N)}$

Line 3 of Alg. 1 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. 1 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 for both LV and NLV.

III. APPROXIMATE PERTURBATION AIDED LATTICE ENCODING

Lattice-based encoding constitutes an important category of near-capacity transmission techniques. The success of lattice-based encoding as a vectoring scheme will depend on the specific geometry of the lattice generated by the inverse channel $\boldsymbol{G}$. Each tone operates independently and therefore the DMT based system of (1) spawns $\mathcal{I}$ lattices in independent Euclidean spaces. Based on the coolant-controller-canceller vectoring model of (2), we will consider the basic operations of APPLE for a single DMT channel. Let the FEXT canceller be represented by $\boldsymbol{\Omega} = \boldsymbol{G}$ and the power controller be of a designated configuration $\boldsymbol{A}$ (configuring the power controller will be the main concern of Section III-D). Then the integer-valued signal coolant $\boldsymbol{l}_{\text{opt}}$ is produced according to the following cost function:

$$\boldsymbol{l}_{\text{opt}} = \arg\min_{\boldsymbol{l}} \{ \max_{1\leq k\leq K} |x_k|^2 \}, \tag{13}$$

which may be relaxed to the following *integer least squares* problem, given that $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{A}(\boldsymbol{u} + \boldsymbol{l})$:

$$\boldsymbol{l}'_{\text{opt}} = \arg\min_{\boldsymbol{l}} \|\boldsymbol{G}\boldsymbol{A}(\boldsymbol{u} + \boldsymbol{l})\|^2. \tag{14}$$

The exact solution to the optimization problem of (14) is given by enumeration algorithms such as the classic sphere encoder of [15], which results in the *vector perturbation* based TPC scheme proposed in [16]. In $\mathbb{R}^{2K}$, the sphere encoder finds the lattice point on $\mathcal{L}(\boldsymbol{GA})$ that is the closest to the given point $\boldsymbol{GAu}$. The vectoring mapping region of the VP encoder of (14) is the Voronoi cell of the lattice $\mathcal{L}(\boldsymbol{GA})$ centered at the origin. The Voronoi cell is in general not a parallelotope, if $\mathcal{L}(\boldsymbol{GA})$ does not have any orthogonal basis (Fig. 4). Therefore no mapping matrix exists for VP, if no orthogonal basis exists for $\mathcal{L}(\boldsymbol{GA})$. For lattices of moderate dimensions, the second central moment of a $2K$-dimensional Voronoi cell is tightly lower-bounded by that of a $2K$-dimensional sphere of the same volume [17].

A conventional sphere encoder consists of the *lattice reduction* preprocessing followed by the search-tree based enumeration. The enumeration result is not affected by lattice reduction, but the enumeration complexity is. Known enumeration algorithms that work on lattices having unknown statistics tend to have an exponentially increased complexity, as demonstrated in [18]. Given a DSL binder encapsulating a large number of twisted pairs, the computational cost of VP becomes excessive. As a consequence, we will find an approximate solution of (14) based on the *nearest plane* method of Babai [19] on short lattice basis. We will first characterize the performance of lattice reduction, and then we present the coice of the additive signal coolant $\boldsymbol{l}$ based on the lattice reduction aided TPC scheme of [20]. Finally we present the novel power control scheme and the bit loading algorithm for achieving the claimed near-optimal performance of APPLE.

A. LATTICE REDUCTION

Babai's *nearest plane approximation of the closest lattice point* [19] depends heavily on the specific geometric properties of the reduced lattice basis. The objective of lattice reduction is to find the particular basis containing the shortest vectors out of all legitimate basis. The notion has since been relaxed to that of finding a basis containing *reasonably short* vectors in favor of a reduced computation time, using for example the Lenstra-Lenstra-Lovász (LLL) method [21]. A value of $1 - 10^{-6}$ is used for the LLL constant defined in [21] in order to find a shorter basis than the one found with the aid of the usual LLL constant of 0.75.

For comparison, we consider a set of channel measurements based on a 100 meter DSL binder encapsulating 10 twisted pairs of 0.5 mm diameter each. The measurements span the entire 212 MHz bandwidth of the G.fast spectrum. The performance of the LLL reduction and that of the optimal Minkowski reduction [22] are compared in Fig. 6, where the orthogonality defect of the LLL-reduced basis $\delta(\boldsymbol{G}_L)$ and that of the Minkowski reduced basis $\delta(\boldsymbol{G}_M)$ are plotted. On the left of Fig. 6, it is shown that the orthogonality defect of the basis associated with the inverse channels increases over frequency. The quality of these basis starts to degrade at below 50 MHz, while the degradation becomes dramatic beyond 100 MHz

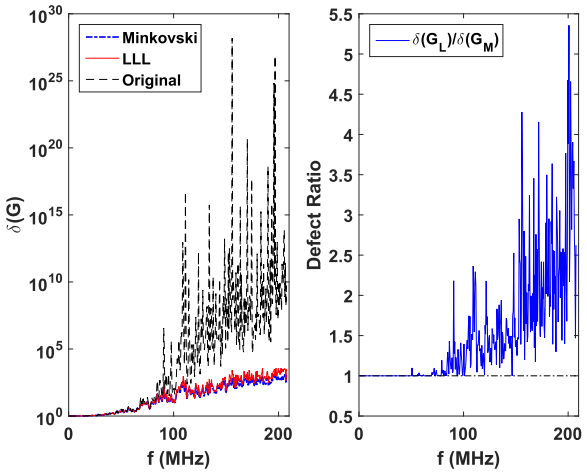


FIGURE 6. Orthogonality defect of the inverse DMT channels and the reduced basis. For comparison purpose, we assume that the Minkowski reduction algorithm in [22] does indeed produce the *shortest possible basis* for a given lattice. The basis given by the inverse channel loses orthogonality before 50 MHz, but LLL-reducible and Minkowski-reducible basis only occur after 90-100 MHz.

without lattice reduction, which brings a major challenge to the 212 MHz G.fast profile in comparison to the current operational profile of 106 MHz. Despite $\delta(\boldsymbol{G})$ exceeds 10^{10} for a handful of inverse channel matrices $\boldsymbol{G}$, both $\delta(\boldsymbol{G}_L)$ and $\delta(\boldsymbol{G}_M)$ manage to stay below 10^5 for the entire 212 MHz bandwidth. Based on the right side of Fig. 6, the relative gap between $\delta(\boldsymbol{G}_M)$ and $\delta(\boldsymbol{G}_L)$ is in fact negligible compared to the gap between $\delta(\boldsymbol{G}_L)$ and $\delta(\boldsymbol{G})$, therefore the employed LLL reduction is near-optimal.

The orthogonality defect, as well as several other significant geometric measures of an LLL-reduced basis, such as the lengths of the basis vectors, can be theoretically (upper and lower) bounded for any lattice associated with a particular channel matrix. However, the known bounds are loose, hence they do not reflect the average performance sufficiently accurately. Therefore we do not incorporate these bounds into our investigations.

B. OPERATION OF APPLE

Let us consider the complex-valued DTU-side encoder of (2) as a decoupled real-valued system for one of the $\mathcal{I}$ DMT tones. Based on the DTU transmitter structure of Fig. 7, we employ the following triangular decomposition of the matrix $\boldsymbol{GA}$:

$$\boldsymbol{GA} = \boldsymbol{WBT}^{-1}, \tag{15}$$

where $\boldsymbol{W}$ has orthogonal columns and $\boldsymbol{B}$ is an upper triangular matrix with $\mathrm{diag}(\boldsymbol{B}) = \mathrm{diag}(\boldsymbol{I})$. The integer-valued unimodular matrix $\boldsymbol{T}$ transforms the given lattice basis $\boldsymbol{GA}$ to the LLL-reduced basis $\boldsymbol{WB}$. Following the decomposition of Eq. (15), the nearest-plane solution $\boldsymbol{l}_{\mathrm{np}}$ of the optimization problem of (14) is given by:

$$\boldsymbol{l}_{\mathrm{np}} = -\boldsymbol{T}\,\tilde{\boldsymbol{l}}, \quad \text{where } \tilde{\boldsymbol{l}} = \lceil \boldsymbol{W}^{-1}\boldsymbol{GAu} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}} \rfloor, \tag{16}$$

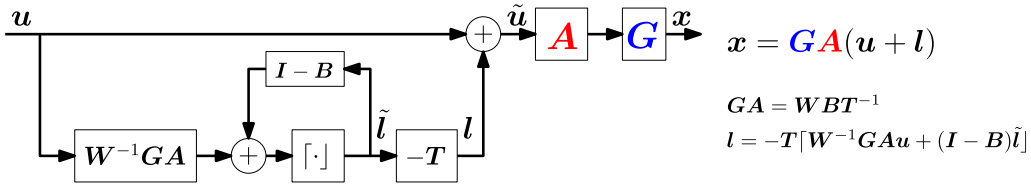


FIGURE 7. APPLE based non-linear DTU for a single tone.

where $\lceil \cdot \rfloor$ denotes the element-wise rounding-off operation to the nearest integer. The symbol vector $\boldsymbol{x}_{\text{APPLE}}$ transmitted by the APPLE encoder is then given by:

$$\begin{aligned}\boldsymbol{x}_{\text{APPLE}} &= \boldsymbol{GA}(\boldsymbol{u} + \boldsymbol{l}_{\text{np}}) \\ &= \boldsymbol{GA}(\boldsymbol{u} - \boldsymbol{T}\lceil \boldsymbol{W}^{-1}\boldsymbol{GAu} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}}\rfloor) \\ &= \boldsymbol{W}(\boldsymbol{m} - \boldsymbol{B}\lceil \boldsymbol{m} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}}\rfloor) \\ &= \boldsymbol{W}\hat{\boldsymbol{u}}, \end{aligned} \tag{17}$$

where we have $\boldsymbol{m} = \boldsymbol{BT}^{-1}\boldsymbol{u}$ and $\hat{\boldsymbol{u}} = \boldsymbol{m} - \boldsymbol{B}\lceil \boldsymbol{m} + (\boldsymbol{I} - \boldsymbol{B})\tilde{\boldsymbol{l}}\rfloor$. The coolant is removed by a modulo operation at the CPE side. Therefore the equalized symbol vector is given by:

$$\boldsymbol{z}_{\text{APPLE}} = \boldsymbol{Au} + \boldsymbol{\Gamma}[\boldsymbol{n}]. \tag{18}$$

Since $\boldsymbol{B}$ is an upper triangular matrix with diagonal entries of $b_{kk} = 1\ \forall k$, the feedback encoder in Fig. 7 is carried out in reverse order, commencing from $k = 2K$. In this case $\hat{u}_k$ is given by:

$$\begin{aligned}\hat{u}_k &= m_k - \sum_{j=k}^{2K} b_{kj}\lceil m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j\rfloor \\ &= m_k - (\sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j + \lceil m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j\rfloor) \\ &= (m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j) - \lceil m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j\rfloor \\ &= \boldsymbol{\Gamma}[m_k - \sum_{j=k+1}^{2K} b_{kj}\tilde{l}_j]. \end{aligned} \tag{19}$$

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 (17) and (19), the mapping matrix of APPLE is $\boldsymbol{W}$ and the vectoring mapping region of APPLE is characterized by the set $\boldsymbol{W}\mathcal{U}$.

Let the power controller be $\boldsymbol{A} = \boldsymbol{I}$ and the inverse channel $\boldsymbol{G}$ be LLL-reducible, i.e. $\boldsymbol{T} \neq \boldsymbol{I}$. Then the fundamental mapping region of APPLE has a lower second central moment and a lower MES per dimension than those of LV and NLV. However, since the mapping matrix $\boldsymbol{W}$ of APPLE and the mapping matrix $\boldsymbol{QD}^{-1}$ of NLV both have orthogonal columns, the corresponding fundamental mapping regions are orthotopes. This is graphically demonstrated in Fig.4.

C. COMPLEXITY OF APPLE

In this section, we consider the complexity of the APPLE transmitter and compare it to that of the conventional LV and NLV. Since the system requires initialization, we will briefly investigate the complexity of initialization (excluding the configuration of the power controller) and the standard symbol encoding independently.

The initialization of APPLE mainly involves the decomposition of (15), i.e. a combination of LLL-reduction having a complexity order of $\mathcal{O}[(2K)^4 \log(\max \|\boldsymbol{g}_k\|)]$ and a triangular decomposition having a complexity order of $\mathcal{O}[(2K)^3]$, while the initialization of LV and NLV requires matrix inversion and triangular decomposition (both imposing a complexity order of $\mathcal{O}(K^3)$ for complex-valued matrices), respectively. APPLE incurs the highest initialization overhead, but its polynomial complexity order is similar to that of the other aforementioned matrix operations. Bearing in mind that the DSL channels are quasi-static, the complexity of initialization can be ignored for practical purposes.

During the steady-state operation, both the power controller and the FEXT canceller (for all vectoring schemes) are linear matrix filters of size K, therefore they have a complexity order of $\mathcal{O}(K^2)$. The decision feedback loop in THP has a complexity order of $\mathcal{O}(K^2)$, whereas the same loop in APPLE has a complexity order of $\mathcal{O}[(2K)^2]$. Meanwhile, APPLE includes two additional matrix filters, both of which have a complexity order of $\mathcal{O}[(2K)^2]$. Hence the overall complexity of APPLE is $\mathcal{O}[(2K)^2]$, while the complexity of both LV and NLV is $\mathcal{O}(K^2)$. If the matrix filters rely on parallel computing, then LV has a complexity order of $\mathcal{O}(K)$, NLV has $\mathcal{O}(K^2)$ and APPLE has $\mathcal{O}[(2K)^2]$.

It is worth noting that an extension of the LLL algorithm has been proposed in [23] for operating directly in $\mathbb{C}^K$ without using the decoupling transformation of (3). The results of [23] demonstrate that the complex-valued LLL (cLLL) algorithm performs almost identically to the original real-valued LLL algorithm. In this case, the complexity of APPLE is also $\mathcal{O}(K^2)$. Therefore the complexity of APPLE is comparable to that of NLV, which is significantly lower than that of the sphere encoder. We summarize our findings in Table. 1.

D. POWER CONTROLLER AND POWER ALLOCATION POLICIES

The power controller enforces a designated power allocation policy for the K twisted pairs of the DSL binder with respect to each of the $\mathcal{I}$ DMT channels in the system

TABLE 1. Complexity order of the vectoring schemes with parallel computing.

	Coolant	Controller	Canceller
LV	N/A	$\mathcal{O}(K)$	$\mathcal{O}(K)$
NLV	$\mathcal{O}(K^2)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$
APPLE	$\mathcal{O}(2K)+\mathcal{O}[(2K)^2]+\mathcal{O}(2K)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$
APPLE-cLLL	$\mathcal{O}(K)+\mathcal{O}(K^2)+\mathcal{O}(K)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$

characterized by (1). The main functionality of the power controller is to coordinate the amount of power allocated to the message-coolant mixture signal so that the per-pair TxPSD mask $\mathbb{P}^i$ and per-pair ATP limit $\mathbb{A}$ defined in [4] are both satisfied, which are:

$$\max_{\boldsymbol{x}^i\in\mathcal{P}^i} |x_k^i|^2 \leq \mathbb{P}^i \quad \forall i,k \tag{20}$$

$$\sum_{i=1}^{\mathcal{I}} \max_{\boldsymbol{x}^i\in\mathcal{P}^i} |x_k^i|^2 \leq \mathbb{A} \quad \forall k, \tag{21}$$

where $\mathcal{P}^i$ is the vectoring mapping region corresponding to the ith tone. The power controller's allocation policies may be given in two categories, including the *scalar power policy* and the *optimized power policy*.

Under a scalar power allocation policy, the power controller sets the scaling matrix $\boldsymbol{A}^i$ for each tone i as a scalar matrix, and the difference amongst each user's channel quality and FEXT condition is ignored. The scalar power allocation policy can be efficiently reconfigured in case of impulsive noise. Typically, the power controller's allocation policy is configured and determined during initialization, when the DTU-side symbol encoder acquires downstream CSI. However, if the seamless rate adaptation (SRA) protocol such as [24] is invoked (e.g. when high-power impulsive noise occurs), then the power allocation policy must be modified without re-initialization in order to sustain continuous connection. In this case, the flexible reconfiguration of a scalar power allocation policy is advantageous.

Alternatively, the power controller may employ a specifically formulated optimized power allocation policy, which maximizes the sum rate of the binder with respect to a particular SER requirement. Configuring the power controller for this specific criterion requires two-dimensional optimization. Due to the discrete nature of *bit allocation* in a practical rate adaptive system, computing the power controller's optimal power allocation policy is a non-convex optimization problem. Furthermore, the constraint of the TxPSD mask and the ATP limit are imposed with respect to each twisted pair. The context is akin to a wireless scenario, where the power constraint is imposed with respect to each transmit antenna rather than the whole antenna group. Related work about the optimized power allocation policy for DSL systems utilizing LV/NLV under per-pair power constraints can be found in [25].

1) SCALAR POWER ALLOCATION POLICY

Let the scalar power allocation policy be given by $\boldsymbol{P}^i = \gamma^i \boldsymbol{I}$, in which case the scaling matrix is $\boldsymbol{A}^i = \boldsymbol{I}\sqrt{\gamma^i}$. In general, if we denote the vectoring mapping region of APPLE for tone i as $\mathcal{P}^i$, we may find that $E\{\mathbb{P}^i/\gamma^i\} = \xi_{\mathcal{P}^i}$ based on the definition of (11). Given the orthotope shape of $\mathcal{P}^i$ based on (17), (18) and (19), 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 1: For square QAM (i.e. 2^b-QAM for $b = 2, 4, 6, \ldots$) constellations, the average SER ζ^i_{APPLE} of tone i achieved by APPLE under a scalar power allocation policy is lower bounded by:

$$\zeta^i_{APPLE} \geq 1 - \left[1 - erfc\left(\sqrt{\frac{2\mathbb{P}^i}{\sigma_n^2 \det[\boldsymbol{H}^i(\boldsymbol{H}^i)^H]^{-1/K}}}\right)\right]^2, \tag{22}$$

where $erfc(\cdot)$ represents the complementary error function. Furthermore, $\boldsymbol{H}^i$ is the channel matrix of tone i and σ_n^2 is the noise variance, as in (1).

The SNR of each equalized symbol z_k^i of the same tone i is characterized by γ^i/σ_n^2 based on (18) and on the scalar power policy we employed, which is dependent on the second moment of APPLE's vectoring mapping region according to the relationship between (10) and (11). 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 [12], the second moment of a $2K$-dimensional hypercube $\mathcal{C}$ having volume[2] $V(\mathcal{C}) = V(\mathcal{P}^i) = \det[(\boldsymbol{G}^i)^H\boldsymbol{G}^i] = \det[\boldsymbol{H}^i(\boldsymbol{H}^i)^H]^{-1}$ is formulated as:

$$\sigma_{\mathcal{C}}^2 = dim(\mathcal{C})G(\mathcal{C})V(\mathcal{C})^{2/dim(\mathcal{C})} = \frac{K\det[\boldsymbol{H}^i(\boldsymbol{H}^i)^H]^{-1/K}}{6}, \tag{23}$$

where $dim(\mathcal{C}) = 2K$ and the dimensionless second moment of a hypercube is $G(\mathcal{C}) = 1/12$. Upon revisiting the difference between (10) and (11), we may identify that $K\xi_{\mathcal{P}^i}/\sigma^2_{\mathcal{P}^i}$ actually represents the peak-to-average ratio (PAR) of the per-pair TxPSD. Since the transmitted signal $\boldsymbol{x}^i$ 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 PAR, which is given by that of the $\mathcal{U}$-constellation of (5), i.e. $(1/2)^2/(1/12) = 3$. Therefore the following holds regarding the geometry of APPLE's vectoring mapping region $\mathcal{P}^i$:

$$\xi_{\mathcal{P}^i} \geq \frac{3\sigma^2_{\mathcal{P}^i}}{K} \geq \frac{3\sigma^2_{\mathcal{C}}}{K} = \frac{\det[\boldsymbol{H}^i(\boldsymbol{H}^i)^H]^{-1/K}}{2}. \tag{24}$$

[2]If the transformation of (3) is applied, then the fundamental volume is $\det(\boldsymbol{G}^i)$ for the real-valued basis $\boldsymbol{G}^i$.

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

$$E\{\gamma^i/\sigma_n^2\} \leq \frac{2\mathbb{P}^i}{\sigma_n^2 \det[\boldsymbol{H}^i(\boldsymbol{H}^i)^H]^{-1/K}}. \tag{25}$$

As a consequence of using the additive coolant $\boldsymbol{l}_{\text{np}}^i$ of (16), the mixture $\tilde{\boldsymbol{u}}^i$ of Fig. 7 (rather than the $\hat{\boldsymbol{u}}^i$ of (17)) is drawn from an infinite-sized periodically-expanded constellation. Thus for each user, $\tilde{u}_k^i$ is distributed over the infinite-sized QAM constellation.[3] For infinite-QAM, every point on the constellation diagram has exactly four nearest neighbors of identical distance. Hence the exact SER performance of APPLE is identical to that of an infinite-QAM constellation transmitted over an AWGN channel, and the proof of Lemma 1 is completed.

As demonstrated in Fig. 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 neighbors, since there will be 'holes' in the constellation. This typically results in an overestimation of the encoded PSD $\xi_{\mathcal{P}^i}$ 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 [26].

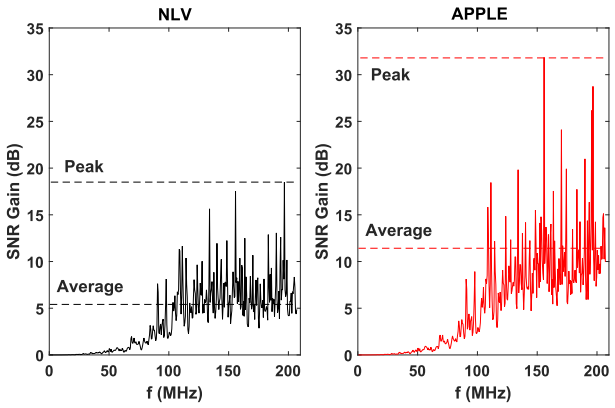


FIGURE 8. SNR gain of NLV and that of APPLE in comparison to the SNR penalty of LV.

On the other hand, given the mapping matrix $\boldsymbol{W}^i$, the exact second moment of the corresponding parallelotope is given by the trace function, $\sigma_{\mathcal{P}^i}^2 = tr[(\boldsymbol{W}^i)^H\boldsymbol{W}^i]$. We may apply similar analysis to the geometry of LV and NLV for finding the relative SNR and SER performance of each vectoring scheme. For simplicity, we assume the minimum PAR $K\xi_{\mathcal{P}^i}/\sigma_{\mathcal{P}^i}^2 = 3$ for LV and NLV as well, even though the actual PAR is higher in both cases. In Fig. 8, we characterize the SNR gain of APPLE over LV and compare it to the SNR gain of NLV over LV for the 100-meter cable characterized by Fig. 6. It is shown that APPLE has a *peak* SNR gain in excess of 10 dB over NLV for the worst-case channel, while the advantage of APPLE is more than 5 dB over NLV in *average*.

[3]This is easily obtained if we splice infinite amount of square QAM constellations, e.g. the boxed 64 QAM of Fig. 2, side by side.

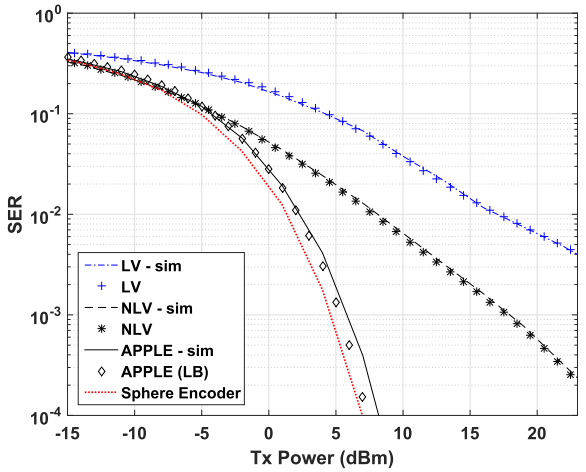


FIGURE 9. SER performance of 16 QAM vs ATP limit per pair under scalar power allocation policy over the full 212 MHz bandwidth. The lower bound of APPLE is given by Lemma 1, while the theoretical performance of LV and NLV is calculated with the trace approximation.

In Fig. 9, the simulated SER of each vectoring scheme using 16 QAM as the input constellation, the lower bound of Lemma 1, as well as the theoretical SER performance of LV and NLV using the trace approximation are compared. It is shown that Lemma 1 has a 0.5 dB difference compared to our simulation result, while the gap is negligible for LV and NLV 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 VP scheme, both of which significantly outperform the conventional NLV and LV, by over 15 dB SNR margin. Due to the ATP constraint of (21), whose historical maximum is 8 dBm in the G.fast 106b profile [8], transmitting a full set of 16 QAM symbols over all tones at an SER of below 10^{-4} is impossible for LV and NLV.

2) OPTIMIZED POWER ALLOCATION POLICY

Assuming a given symbol error rate requirement, configuring the optimized power allocation policy for a rate-adaptive system is also often referred to as the *bit loading* process. For an M_k^i-QAM based system, bit loading is an optimization process, which we formulate as follows:

$$\max \sum_{i=1}^{\mathcal{I}} \sum_{k=1}^{K} b_k^i \quad \text{where } b_k^i = \log_2(M_k^i), \tag{26}$$

subject to the constraints of (20) and (21). Based on the characterization of the equalized symbol vector $\boldsymbol{z}^i$ in (7), (9) and (18), the diagonal element A_{kk}^i of the controller $\boldsymbol{A}^i$ defines the SNR of the equalized symbol. For a given PSD σ_n^2 of the AWGN, A_{kk}^i also defines the legitimate power allocation P_{kk}^i. For QAM symbols, the legitimate bit load b of an AWGN

channel with respect to a given SNR η and a coding gain α is given by:

$$b = \log_2 M = \log_2(1 + \frac{\alpha\eta}{\delta}), \tag{27}$$

where δ characterizes the gap between the M-QAM scheme and the capacity of an AWGN channel with respect to the SER target ζ. The SNR gap δ is defined as [27]:

$$\delta = \frac{1}{3}\left[\sqrt{2}\, erfc^{-1}\left(\frac{\zeta}{2}\right)\right]^2. \tag{28}$$

It is apparent that $\eta_k^i = P_{kk}^i/\sigma_n^2$. However, since b_k^i assumes integer values upto a bit cap of $b_{\max}$, the minimum required power P_{kk}^i also takes discrete values. Therefore the optimization problem of (26) is non-convex.

On the other hand, the power constraints (20) and (21) impose instantaneous limits on the peak TxPSD and peak ATP per pair. Given the vectoring mapping region $\mathcal{P}^i$ (which depends on $\boldsymbol{A}^i$) of the transmitted symbol vector $\boldsymbol{x}^i$ and the MES $\boldsymbol{v}^i$ of the polytope $\mathcal{P}^i$, the peak TxPSD of the ith tone of the kth pair is characterized by $|v_k^i|^2$, while the peak ATP of the kth pair is characterized by $\sum_{i=1}^{\mathcal{I}} |v_k^i|^2$ accordingly. We note that the power allocation P_{kk}^i calculated using (27) and (28) refers to the average constellation energy, which must be up scaled by the peak-to-average ratio in order to find the peak value of the TxPSD and that of the ATP.

E. BIT-LOADING ALGORITHM

Let us consider the common scenario, where the power controller invokes the optimized power allocation policy. The combinatorial optimization problem (26) 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 $\mathcal{I}$ independent Euclidean spaces.

The LLL reduction algorithm involved in (15) presents many analytical challenges. If $\boldsymbol{A}^i$ is not a scalar matrix, then decomposing $\boldsymbol{G}^i\boldsymbol{A}^i$ results in a completely different unimodular matrix from the one obtained by decomposing $\boldsymbol{G}^i$. 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 [28] and [29]. The algorithm is given in two consecutive parts, namely in Alg. 2 and Alg. 3, respectively.

Let $f(b_k^i)$ 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 i pair k.[4] Given our power allocation policy $\boldsymbol{P}^i$ and the time-invariant FEXT canceller $\boldsymbol{G}^i$, $F(\boldsymbol{P}^i)$ represents the mapping matrix $\boldsymbol{W}^i$ of APPLE following the decomposition $\boldsymbol{G}^i\sqrt{\boldsymbol{P}^i} = \boldsymbol{W}^i\boldsymbol{B}^i(\boldsymbol{T}^i)^{-1}$ akin to that of (15). Alg. 2 and Alg. 3

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

Algorithm 2 TxPSD-Constrained Bit Loading

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

Algorithm 3 ATP-Constrained Bit Loading

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

are the two parts of the loading algorithm that operates under the constraint of (20) and (21), 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 [30].

In Alg. 2, the system is initialized by assigning the maximum admissible bit allocation $b_k^i = b_{\max}$. 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 i, the TxPSD of each pair is compared against the mask $\mathbb{P}^i$. If the highest TxPSD exceeds the mask, then the corresponding pair is denoted as $k_{\max}$ and its TxPSD is given by $|v_{k_{\max}}^i|^2$. Using the greedy principle, the specific pair k^* where subtracting a single bit would have caused the largest reduction of $|v_{k_{\max}}^i|^2$ is selected. The bit load $b_{k^*}^i$ is then updated to have one less bit. The TxPSD

characterized by the new bit allocation of tone i is then compared against the mask. Alg. 2 terminates, when the TxPSD mask is fully complied with for all tones right across the entire bandwidth of the system.

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

If the power controller invokes a scalar power allocation policy, then the efficiency of Alg. 2 (as well as of Alg. 3) 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 vectoring mapping region 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 (15) from scratch for each update of the power policy, as seen in line 8 of Alg. 2.

IV. RESULTS

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 [4] 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 LV and NLV, 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. we have $\|\boldsymbol{x}^i\|^2 \leq K\mathbb{P}^i$.

In Fig. 10 and Fig. 11, 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 the optimized and for the scalar power allocation policy. The performance of APPLE is almost identical to that of NLV, when the controller employs the optimized power allocation policy. By contrast, under the scalar power allocation policy, the sum-rate of APPLE exceeds that of both LV and NLV.

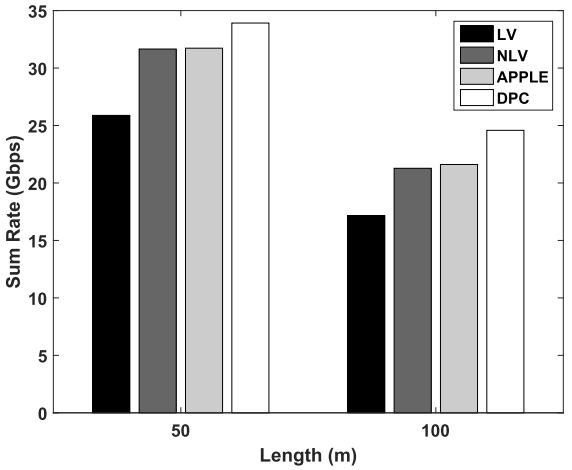


FIGURE 10. Sum-rate of 10-Pair 0.5mm short DSL binders configured with *optimized power allocation policy*. An extended baseband spectrum of upto 300 MHz is used with noise floor at -150 dBm/Hz.

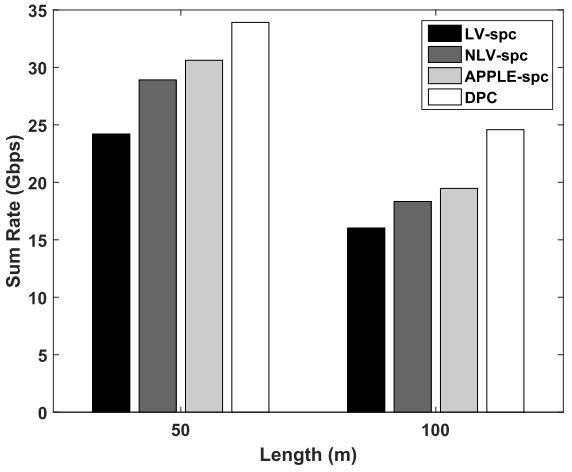


FIGURE 11. Sum-rate of 10-Pair 0.5mm short DSL binders configured with *scalar power allocation policy*. An extended baseband spectrum of upto 300 MHz is used with noise floor at -150 dBm/Hz.

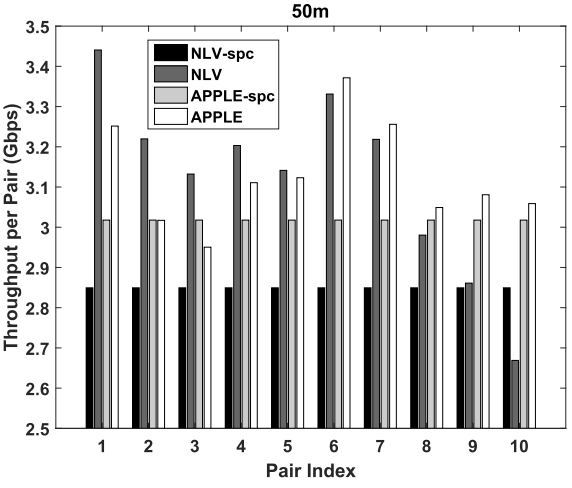


FIGURE 12. Per-pair throughput of a 10-Pair 0.5mm DSL binder of length *50m*.

The throughput of APPLE and that of NLV are compared in more details in Fig. 12 and Fig. 13. It is explicitly shown that APPLE can achieve an improved user-fairness and a

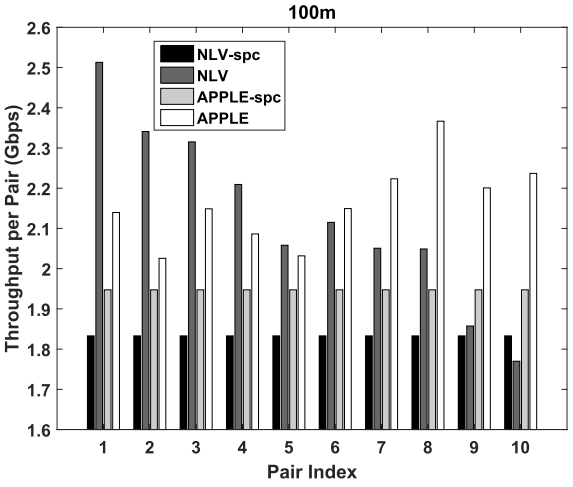


FIGURE 13. Per-pair throughput of a 10-Pair 0.5mm DSL binder of length *100m*.

better worst-case performance than NLV under the optimized power allocation policy, while retaining an identical sum-rate. On the other hand, APPLE can also achieve a higher sum-rate than NLV when both have identical fairness under the scalar power allocation policy.

V. CONCLUSIONS

Given the expanded spectrum of G.fast, the extra high-frequency tones no longer have diagonally-dominant channels. The crosstalk has dramatically increased, especially beyond 106 MHz. To satisfy the peak TxPSD per pair constraint and the peak ATP per pair constraint, we proposed a novel vectoring scheme, referred to as APPLE, by exploiting the geometric properties of lattices and their short-length basis. The improved signal coolant of APPLE results in significant SNR gains over both LV and NLV, hence it achieves a better SER performance. For the family of modulo type vectoring schemes, APPLE is shown to achieve a higher sum-rate than NLV under guaranteed fairness. Viewed from a different perspective, APPLE achieves improved fairness over NLV, while maintaining a similar sum-rate. Since DSL channels are quasi-static, the advantage of APPLE is realized at minor complexity sacrifice over the state-of-the-art NLV bench marker.

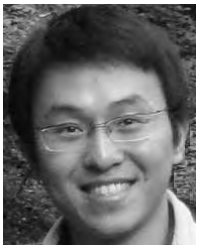
ACKNOWLEDGMENT

The research data of this paper can be found at https://doi.org/10.5258/SOTON/D0647

REFERENCES

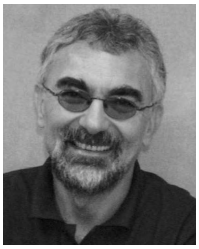
[1] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1085–1104, Jun. 2002.
[2] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[3] R. Cendrillon, M. Moonen, J. Verlinden, T. Bostoen, and G. Ginis, "Improved linear crosstalk precompensation for DSL," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, May 2004, pp. IV-1053–IV-1056.
[4] *Fast Access to Subscriber Terminals (G.fast)—Power Spectral Density Specification*, document G.9700, 2014.
[5] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
[6] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, nos. 5–6, pp. 138–139, 1971.
[7] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. COMM-20, no. 4, pp. 774–780, Aug. 1972.
[8] *Fast Access to Subscriber Terminals (G.fast)—Physical Layer Specification*, document G.9701, 2014.
[9] *Self-FEXT Cancellation (Vectoring) for Use With VDSL2 Transceivers*, document G.993.5, 2015.
[10] C. Ling and J.-C. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.
[11] T. Wang, S. C. Liew, and L. Shi, "Optimal rate-diverse wireless network coding," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2411–2426, Jun. 2017.
[12] J. Conway and N. Sloane, "Sphere packings, lattices and groups," in *Grundlehren Der Mathematischen Wissenschaften*. Berlin, Germany: Springer-Verlag, 1993.
[13] M. L. Honig, *Capacity-Approaching Multiuser Communications over Multiple Input/Multiple Output Broadcast Channels*. Hoboken, NJ, USA: Wiley, 2009, p. 384.
[14] 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, Jun. 2015.
[15] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, no. 1, pp. 181–199, Aug. 1994.
[16] 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.
[17] D. J. Ryan, I. B. Collings, I. V. L. Clarkson, and R. W. Heath, Jr., "Performance of vector perturbation multiuser MIMO systems with limited feedback," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2633–2644, Sep. 2009.
[18] 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.
[19] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.
[20] 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.
[21] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematische Annalen*, vol. 261, pp. 515–534, 1982.
[22] 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.
[23] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2701–2710, Jul. 2009.
[24] V. Oksman and B. Heise, "Method for seamless bit rate adaptation for multicarrier DSL," U.S. 7 519 124 B2, Apr. 14, 2009.
[25] 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, Mar. 2017.
[26] X.-C. Zhang, H. Yu, and G. Wei, "Exact symbol error probability of cross-QAM in AWGN and fading channels," *J. Wireless Commun. Netw.*, vol. 2010, no. 1, p. 917954, Dec. 2010.
[27] J. M. Ciofli, "A multicarrier primer," *ANSI T1E1*, vol. 4, pp. 91–157, Nov. 1991.
[28] D. Z. Filho, R. R. Lopes, R. Ferrari, R. Suyama, and B. Dortschy, "Bit loading for precoded DSL systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 3, Apr. 2007, pp. III-353–III-356.
[29] J. Neckebroek *et al.*, "Novel bitloading algorithms for coded G.fast DSL transmission with linear and nonlinear precoding," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 945–951.
[30] J. Campello, "Practical bit loading for DMT," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, Jun. 1999, pp. 801–805.

YANGYISHI ZHANG received the B.Sc. degree in electronics and communication engineering from the University of Liverpool, U.K., in 2014, and the M.Sc. degree (Hons.) in wireless communications from The University of Southampton, U.K., in 2015, where he is currently pursuing the Ph.D. degree with the Southampton Wireless Group. In 2017, he was an Invited Researcher with British Telecom in Adastral Park, Martlesham, U.K. His research interests include multi-user communication, geometry of numbers, and lattice theory.



RONG ZHANG (M'09–SM'16) received the Ph.D. degree in wireless communications from The University of Southampton (UoS) in 2009. He was a Research Assistant with the Mobile Virtual Centre of Excellence, which is one of U.K.'s largest industrial-academic partnership in ICT. He was with the School of Electronics and Computer Science, UoS, where he contributed to a number of international projects as a UoS Lead Researcher. He took his industrial consulting leave from Huawei EU R&D as a System Algorithms Expert. He is currently an Assistant Professor with the Southampton Wireless Group, UoS. He has a total of 80+ IEEE/OSA publications, including 55+ journals (20+ of which as the first author). He is a RAEng Industrial Fellow and a member of the OSA. He was a recipient of the prestigious RAEng Industrial Fellowship. Owing to his outstanding academic achievements, he received the prestigious Dean's Publication Award from the Faculty of Physical Sciences and Engineering, UoS. He has been a TPC member /invited session chair of major conferences for several times. He regularly serves as a reviewer for IEEE/OSA journals and funding bodies.

ANAS F. AL RAWI (S'08–M'13) received the M.Sc. degree (Hons.) in communications and signal processing and the Ph.D. degree in communications and signal processing from Newcastle University, U.K., in 2007 and 2011, respectively. From 2011 to 2012, he was a Post-Doctoral Researcher with Swansea University and Queen's University Belfast, where he focused on ray-tracing algorithms for positioning and self-organized network design. Since 2013, he has been heavily involved in developing novel transmission techniques in attempt to push the information capacity of copper beyond current limits. His role involves modeling current and future generations of broadband technologies. He is currently a Senior Scientist with BT Labs and a Royal Society Industry Fellow with the Cavendish Laboratory, University of Cambridge. His research interests include computational electromagnetics, transmission cross-layer optimization, and cooperative and multiple-input-multiple-output systems.



LAJOS HANZO (M'91–SM'92–F'04) received the master's degree in electronics in 1976, the Ph.D. degree in 1983, the Honorary Doctorate degree from the Technical University of Budapest in 2009, the Honorary Doctorate degree from The University of Edinburgh in 2015, and the D.Sc. degree in 2004. In 2016, he was admitted to the Hungarian Academy of Sciences. During his 40-year career in telecommunications, he held various research and academic positions in Hungary, Germany, and U.K. During 2008–2012, he was a Chaired Professor with Tsinghua University, Beijing. Since 1986, he has been with the School of Electronics and Computer Science, The University of Southampton, U.K., where he holds the Chair in telecommunications. He is an enthusiastic supporter of industrial and academic liaison, and he offers a range of industrial courses. He has successfully supervised 111 Ph.D. students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totaling in excess of 10 000 pages, and published 1768 research contributions at IEEE Xplore. He is a fellow of the Royal Academy of Engineering, IET, and EURASIP. He acted both as a TPC Chair and the General Chair of the IEEE conferences, presented keynote lectures, and has been awarded a number of distinctions. He is currently directing an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council, U.K., the European Research Council's Advanced Fellow Grant, and the Royal Society's Wolfson Research Merit Award. He is also a Governor of the IEEE VTS. During 2008–2012, he was the Editor-in-Chief of the IEEE Press. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk. He has 38 000+ Google Scholar citations and an H-index of 78.