A Near-Optimal Linear Crosstalk Precoder for VDSL

Raphael Cendrillon, George Ginis, Etienne Van den Bogaert, Marc Moonen

Abstract

Crosstalk is the major source of performance degradation in VDSL. In downstream transmission crosstalk precoding can be applied. The transmitted signal is predistorted, such that the predistortion annihilates with the crosstalk introduced in the binder. Several crosstalk precoders have been proposed. Unfortunately they either give poor performance or require non-linear operations, which results in a high complexity. In this paper we present a simple, linear diagonalizing crosstalk precoder with low run-time complexity. A lower bound on the performance of the DP is derived. This allows performance to be predicted without explicit knowledge of the crosstalk channels, which simplifies service provisioning considerably. This bound shows that the DP operates close to the single-user bound. So the DP is a low complexity design with predictable, near-optimal performance. The combination of spectra optimization and crosstalk precoding is also considered. Spectra optimization in a broadcast channel generally involves a highly complex optimization problem. Since the DP decouples transmission on each line, the spectrum on each modem can be optimized through a dual decomposition, leading to a significant reduction in complexity.

Part of this work has been presented at the IEEE Global Telecommun. Conf. (GLOBECOM) , Nov. 2004 and the IEEE Int. Conf. on Acoust., Speech and Sig. Processing (ICASSP), May 2004. The material in the Appendix also forms part of the paper "A Near-Optimal Linear Crosstalk Canceler for VDSL," submitted to the IEEE Transactions on Signal Processing.

Raphael Cendrillon and Marc Moonen are with the Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, Heverlee, 3001 Belgium. Email: {raphael.cendrillon, marc.moonen}@esat.kuleuven.ac.be. George Ginis is with the Broadband Communications Group, Texas Instruments, 2043 Samaritan Drive, San Jose, CA 95124. Email: gginis@ti.com. Etienne Van den Bogaert is with the DSL Research and Innovation group, Alcatel Bell, Francis Wellesplein 1, Antwerp 2018, Belgium. Email: etienne.van_den_bogaert@alcatel.be

This work was carried out in the frame of IUAP P5/22, *Dynamical Systems and Control: Computation, Identification and Modelling* and P5/11, *Mobile multimedia communication systems and networks*; the Concerted Research Action GOA-MEFISTO-666, *Mathematical Engineering for Information and Communication Systems Technology*; IWT BANITS Project, *Broadband Access Networks Integrated Telecommunication System* sand was partially sponsored by Alcatel-Bell.

Index Terms

Crosstalk precoding, diagonal dominance, digital subscriber lines, dynamic spectrum management, linear, reduced complexity, vectoring

EDICS— 3-TDSL

I. Introduction

Next generation DSL systems such as VDSL aim at providing extremely high data-rates, up to 52 Mbps in the downstream. Such high data-rates are supported by operating over short loop lengths and transmitting in frequencies up to 12 MHz. Unfortunately, the use of such high frequency ranges causes significant electromagnetic coupling between neighbouring twisted pairs within a binder. This coupling creates interference, referred to as crosstalk, between the DSLs within a network. Over short loop lengths crosstalk is typically 10-15 dB larger than the background noise and is *the* dominant source of performance degradation.

In upstream communication the receiving modems are co-located at the *central office* (CO) or at an *optical network unit* (ONU) located at the end of the street. This allows joint reception of the signals transmitted on the different lines, thereby enabling *crosstalk cancellation*[1].

In *downstream* (DS) communication the receiving modems reside within different *customer premises* (CP). The receiving modems are not co-located, so joint reception and crosstalk cancellation is impossible. Fortunately, in DS communication the *transmitting* modems are co-located at the CO. So joint transmission is possible. Predistortion is introduced into each signal before transmission. The predistortion is chosen such that it annihilates with the crosstalk introduced in the binder, a technique known as crosstalk precoding.

Several crosstalk precoder designs have been proposed. The simplest is a linear structure based on the *zero-forcing* (ZF) criterion[2] and is described in more detail in Sec. IV. Unfortunately, with this design all modems experience the channel of the weakest line in the binder as will be demonstrated. When the channels of the different lines vary significantly, due to varying line lengths or bridged taps, this design gives poor performance.

A decision feedback structure, based on the Tomlinson-Harashima precoder, was shown to operate close to the single-user bound[1] and is described in more detail in Sec. V. Unfortunately this structure relies on non-linear operations, which lead to a high run-time complexity.

Other techniques use joint linear processing at both the transmit and receive side of the link[3][4]. This requires co-location of both CO and CP modems, which is typically not possible since different customers are situated at different locations.

A companion paper [5] investigates the design of linear crosstalk cancelers for upstream VDSL.

In upstream transmission the VDSL crosstalk channel matrix is *column-wise diagonal dominant* (CWDD). This property was used to show the near-optimality of the linear ZF crosstalk canceler[5]. In downstream transmission the crosstalk channel matrix is not CWDD, but *row-wise diagonal dominant* (RWDD). The result is that *zero-forcing precoder* (ZFP) designs are highly sub-optimal as will be demonstrated.

This paper presents an alternative linear precoder based on a channel diagonalizing criterion. This technique has a low complexity. This paper analyzes the performance of the *diagonalizing precoder* (DP) in a VDSL environment. It is shown that due to the RWDD of the VDSL channel matrix the diagonalizing design leads to negligible increase in transmit power. As a result, this simple linear structure achieves near-optimal performance. We develop bounds to show that the DP operates close to the single-user bound in VDSL channels. These bounds allow performance of the DP to be predicted without explicit knowledge of the crosstalk channels, which simplifies service provisioning significantly.

In this paper, the combination of spectra optimization and crosstalk precoding is also considered. In upstream VDSL, application of the linear ZF canceler allowed the transmit spectra to be optimized independently for each line, reducing complexity considerably[5]. Unfortunately, in downstream VDSL, the total power constraints for each line must be satisfied after application of the crosstalk precoder. This couples the spectra optimization problem between lines, increasing complexity. This paper addresses this problem through an algorithm based on dual decomposition. The algorithm efficiently optimizes the transmit spectra for all lines, and achieves a significant reduction in complexity over existing spectra optimization algorithms for the general broadcast channel.

The rest of this paper is organized as follows. The system model for a network of VDSL modems transmitting from a single CO/ONU to a multitude of CPs is given in Sec. II. A property of the downstream VDSL channel, known as RWDD, is explored. As described in Sec. III, from an information theoretical perspective the downstream VDSL channel is a *broadcast channel*

(BC). This allows the single-user bound to be applied to upper bound the capacity of the channel. Sec. IV describes the ZFP and the problems it has with transmit power enhancement. Sec. V describes the multi-user *Tomlinson-Harashima precoder* (THP) as a state-of-the-art solution, and shows that it requires a high run-time complexity.

To address these problems, Sec. VI describes a much simpler linear design, the DP, which has a low complexity. Sec. VI uses the RWDD property to formulate a lower bound on the performance of the DP. This bound shows that the DP operates close to the single-user bound. Sec. VII describes power loading algorithms for use with the DP. Existing power loading algorithms for the BC are extremely complex, having a polynomial complexity in the number of lines and tones. Application of the DP allows the power loading problem to be solved efficiently through the use of a dual decomposition. Sec. VIII compares the performance of the different precoders through simulation.

II. SYSTEM MODEL

Assuming that the modems are synchronized and *discrete multi-tone* (DMT) modulation is employed we can model transmission independently on each tone

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k. \tag{1}$$

Synchronization is straight-forward to implement when the transmitting modems are co-located, which is the assumption we make here. We assume perfect knowledge of the crosstalk channels. In practice these must be identified using MIMO channel identification techniques[6], [7]. The vector $\mathbf{x}_k \triangleq \left[x_k^1, \cdots, x_k^N\right]^T$ contains transmitted signals on tone k, where the tone index k lies in the range $1 \ldots K$. There are N lines in the binder and x_k^n is the signal transmitted onto line n at tone k. The vectors $\mathbf{y}_k$ and $\mathbf{z}_k$ have similar structures. The vector $\mathbf{y}_k$ contains the received signals on tone k. The vector $\mathbf{z}_k$ contains the additive noise on tone k and is comprised of thermal noise, alien crosstalk, RFI etc. The $N \times N$ matrix $\mathbf{H}_k$ is the crosstalk channel matrix on tone k. The element $h_k^{n,m} \triangleq [\mathbf{H}_k]_{n,m}$ is the channel from transmitter m to receiver n on tone k. The diagonal elements of $\mathbf{H}_k$ contain the direct-channels whilst the off-diagonal elements contain the crosstalk channels. The transmit correlation on tone k is defined $\mathbf{S}_k \triangleq \mathcal{E}\left\{\mathbf{x}_k \mathbf{x}_k^H\right\}$. We denote the transmit PSD of user n on tone k as $s_k^n \triangleq \mathcal{E}\left\{|x_k^n|^2\right\}$. For the time being we assume that the transmit PSD on each line must obey a spectral mask constraint. That is

$$s_k^n \leq s_k^{\mathrm{mask}},\ \forall k, n. \tag{2}$$

In Section VII we will extend upon this and consider DSL modems operating under a total power constraint. The noise power experienced by receiver n on tone k is defined $\sigma_{k,n} \triangleq \mathcal{E}\left\{|z_k^n|^2\right\}$. Since the transmitting modems are co-located, the crosstalk signal transmitted from a disturber into a victim must propagate through the full length of the victim's line. This is depicted in Fig. 1, where CO2 is the disturber and CP1 is the victim. The insulation between twisted pairs increases the attenuation. As a result, the crosstalk channel matrix $\mathbf{H}_k$ is *row-wise diagonally dominant* (RWDD), since on each row of $\mathbf{H}_k$ the diagonal element has the largest magnitude

$$|h_k^{n,m}| \ll |h_k^{n,n}|,\ \forall m \neq n. \tag{3}$$

RWDD implies that the crosstalk channel $h_k^{n,m}$ from a disturber m into a victim n is always weaker than $h_k^{n,n}$, which is the direct channel of the victim[1]. The degree of RWDD can be characterized with the parameter α_k

$$|h_k^{n,m}| \leq \alpha_k\,|h_k^{n,n}|,\ \forall m \neq n. \tag{4}$$

Note that crosstalk precoding is based on joint transmission. As such it requires the co-location of transmitting modems. So in all channels where crosstalk precoding can be applied, the RWDD property holds. RWDD has been verified through extensive measurement campaigns of real binders. In 99% of lines α_k is bounded

$$\alpha_k \leq K_{\mathrm{xf}} \cdot f_k \cdot \sqrt{d_{\mathrm{coupling}}},$$

where $K_{\mathrm{xf}} = -22.5$ dB and f_k is the frequency on tone k in MHz[8]. Here d_{coupling} is the coupling length between the disturber and the victim in kilometers. The coupling length can be upper bounded by the longest line length in the binder. Hence

$$\alpha_k \leq K_{\mathrm{xf}} \cdot f_k \cdot \sqrt{l_{\mathrm{max}}}, \tag{5}$$

where l_{max} denotes the length of the longest line in the binder. To find a value for α_k that is independent of the particular binder configuration, l_{max} can be set to 1.2 km, which is the

[1]Contrast this with the CWDD experienced in upstream transmission, where the crosstalk channel $h_k^{n,m}$ from a disturber into a victim is always weaker than the direct channel of the disturber $h_k^{m,m}$.

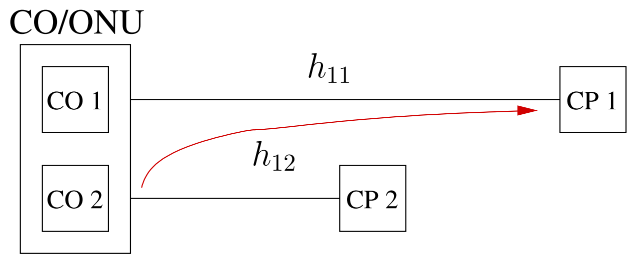

Fig. 1. Row-wise Diagonal Dominance $|h_{11}| \gg |h_{12}|$

maximum deployment length for VDSL[2]. The following sections show that RWDD ensures a well-conditioned crosstalk channel matrix. This results in the near-optimality of the DP.

III. Theoretical Capacity

We start with a bound on the capacity of the downstream VDSL channel with coordinated transmitters operating under spectral mask constraints. This will prove useful in evaluating crosstalk precoder performance since it provides an upper bound on the achievable data-rate with any possible crosstalk precoding scheme.

Theorem 1: When all transmitters are subject to a spectral mask (2) the achievable data-rate for user n, denoted by R_n, is upper bounded

$$R_n \leq \sum_k b_{k,\text{mask}}^n \tag{6}$$

where the data-rate of user n on tone k is upper bounded by

$$b_{k,\text{mask}}^n \triangleq \Delta_f \log_2 \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} s_k^{\text{mask}} \left| h_k^{n,n} \right|^2 \left[1 + (N-1)\, \alpha_k \right]^2 \right), \tag{7}$$

and Δ_f denotes the tone-spacing.

Proof of Theorem 1: CO modems are co-located and do transmission in a joint fashion, so from an information theoretical perspective this is a broadcast channel[10]. We start by considering the so-called *single-user bound*, which is the capacity achieved when all transmitters (CO modems)

[2]Standardization groups are currently considering the deployment of VDSL2 at lengths greater than 1.2 km[9]. However at such distances far-end crosstalk is no longer the dominant source of noise, and the benefits of far-end crosstalk cancellation and precoding are reduced considerably.

are used to communicate to a single receiver (CP modem). In this case the received signal on the CP modem is

$$y_k^n = \overline{\mathbf{h}}_k^n \mathbf{x}_k + z_k^n,$$

where $\overline{\mathbf{h}}_k^n \triangleq [\mathbf{H}_k]_{\mathrm{row}\, n}$. Using the single-user bound the achievable data-rate of user n on tone k is limited to

$$\begin{aligned} b_k^n &\leq \Delta_f I(\mathbf{x}_k; y_k^n), \\ &= \Delta_f \log_2 \left(1 + \sigma_{k,n}^{-1} \overline{\mathbf{h}}_k^n \mathbf{S}_k \left(\overline{\mathbf{h}}_k^n \right)^H \right), \end{aligned} \tag{8}$$

where $I(a;b)$ denotes the mutual information between a and b. To account for the sub-optimality of practical coding schemes, we include the SNR-gap to capacity Γ[11]. This results in the following achievable data-rate for user n on tone k

$$b_k^n = \Delta_f \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \overline{\mathbf{h}}_k^n \mathbf{S}_k \left(\overline{\mathbf{h}}_k^n \right)^H \right). \tag{9}$$

In the single-user case the single-user bound can be achieved with a matched transmit filter. In the multi-user case the single-user bound can be achieved through dirty paper coding[12], [13].

Define the elements of the correlation matrix $s_k^{n,m} \triangleq [\mathbf{S}_k]_{n,m}$, and the diagonal elements $s_k^n \triangleq [\mathbf{S}_k]_{n,n}$. Since $\mathbf{S}_k$ is positive semi-definite it must be true that

$$s_k^{n,m} \leq \sqrt{s_k^n s_k^m}. \tag{10}$$

Hence

$$\begin{aligned} \overline{\mathbf{h}}_k^n \mathbf{S}_k \, \overline{\mathbf{h}}_k^{n\,H} &= \sum_i h_k^{n,i} \sum_j s_k^{i,j} \left(h_k^{n,j} \right)^*, \\ &\leq \sum_i \left| h_k^{n,i} \right| \sqrt{s_k^i} \sum_j \left| h_k^{n,j} \right| \sqrt{s_k^j}, \\ &= \left(\sum_i \left| h_k^{n,i} \right| \sqrt{s_k^i} \right)^2, \end{aligned}$$

where (10) is applied in the second line. Combining this with (9) and (4) yields

$$b_k^n \leq b_{k,\mathrm{bnd}}^n,$$

where the data-rate of user n on tone k is bounded by

$$b_{k,\mathrm{bnd}}^n(s_k^1, \ldots, s_k^N) \triangleq \log_2 \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \left| h_k^{n,n} \right|^2 \left(\sqrt{s_k^n} + \alpha_k \sum_{m \neq n} \sqrt{s_k^m} \right)^2 \right). \tag{11}$$

Combining this with (2) leads to (7), which completes the proof. ∎

Multi-user and multi-antenna techniques are often used in wireless systems and lead to large increases in the *signal power* at the receiver. The observation is that if the path from transmit antenna n to receive antenna n is weak, then the path from transmit antenna m to receive antenna n might be strong. The result is a statistical averaging across spatial dimensions, an effect known as spatial diversity, which leads to large improvements in performance[14].

In VDSL channels there is, unfortunately, no equivalent to spatial diversity. This can be seen in equation (7). Here the RWDD of $\mathbf{H}_k$ implies that very little increase can be made in the signal power through the use of multiple VDSL transmitters. This is the case since, when transmitters are co-located, the crosstalk channel from transmitter m to receiver n is always much weaker than the direct channel from transmitter n to receiver n. Note that the benefit, although small, increases with the crosstalk channel strength α_k and the number of crosstalkers N.

Although spatial diversity is negligible, the use of co-ordinated transmission is by no means fruitless. Instead of benefiting through spatial diversity, the primary benefit in VDSL channels is crosstalk precoding. That is, co-ordinated transmission does not increase signal power in VDSL, but instead decreases *interference power*.

IV. Zero Forcing Precoder

The simplest precoder design is the *zero forcing precoder* (ZFP). Define the vector

$$\widetilde{\mathbf{x}}_k \triangleq [\widetilde{x}_k^1, \ldots, \widetilde{x}_k^N]^T,$$

which contains the symbols intended for each user on tone k. The ZFP multiplies the true symbols $\widetilde{\mathbf{x}}_k$ with a precoding matrix $\mathbf{P}_{k,\text{zf}}$ prior to transmission. The transmitted symbols are then

$$\mathbf{x}_k = \mathbf{P}_{k,\text{zf}} \widetilde{\mathbf{x}}_k.$$

The ZFP is based on a zero-forcing criterion, which leads to the following precoding matrix

$$\mathbf{P}_{k,\text{zf}} \triangleq \beta_{k,\text{zf}}^{-1} \mathbf{H}_k^{-1},$$

where the scaling factor is defined

$$\beta_{k,\text{zf}} \triangleq \max_n \left\| \left[\mathbf{H}_k^{-1}\right]_{\text{row}\, n} \right\| . \tag{12}$$

The power of the true symbols are set to obey the spectral mask

$$\widetilde{s}_k^n \leq s_k^{\text{mask}},\ \forall k, n; \tag{13}$$

where $\widetilde{s}_k^n \triangleq \mathcal{E}\left\{|\widetilde{x}_k^n|^2\right\}$. The scaling factor $\beta_{k,\text{zf}}$ ensures that compliance with the spectral masks, a described by (2), is maintained after precoding. Consider

$$\begin{aligned} x_k^n &= [\mathbf{P}_{k,\text{zf}}]_{\text{row}\, n}\, \widetilde{\mathbf{x}}_k, \\ &= \beta_{k,\text{zf}}^{-1} \sum_m \left[\mathbf{H}_k^{-1}\right]_{n,m} \widetilde{x}_k^m. \end{aligned} \tag{14}$$

Under this condition

$$\begin{aligned} s_k^n &= \beta_{k,\text{zf}}^{-2} \mathcal{E}\left\{\left|\sum_m \left[\mathbf{H}_k^{-1}\right]_{n,m} \widetilde{x}_k^m\right|^2\right\} \\ &= \beta_{k,\text{zf}}^{-2} \sum_m \left|\left[\mathbf{H}_k^{-1}\right]_{n,m}\right|^2 \widetilde{s}_k^m, \\ &\leq \beta_{k,\text{zf}}^{-2} \left\|\left[\mathbf{H}_k^{-1}\right]_{\text{row}\, n}\right\|^2 s_k^{\text{mask}}, \\ &\leq s_k^{\text{mask}}, \end{aligned}$$

where the second line uses the fact that the true symbols are uncorrelated, and the last lines uses (12). Hence the ZFP maintains compliance with the spectral mask constraints.[3]

During transmission the predistortion introduced by the ZFP annihilates with the crosstalk. The received vector

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{P}_{k,\text{zf}} \widetilde{\mathbf{x}}_k + \mathbf{z}_k, \\ &= \beta_{k,\text{zf}}^{-1} \widetilde{\mathbf{x}}_k + \mathbf{z}_k, \end{aligned}$$

and each user experiences a crosstalk free channel. All users experience the same direct channel gain $\beta_{k,\text{zf}}^{-1}$. Unfortunately, this causes all users to experience the worst channel in the binder. To see this consider the case when all users transmit at the spectral mask; $s_k^m = s_k^{\text{mask}},\ \forall m$. Each user then achieves a data-rate of

$$b_{k,\text{zfp}}^m = \Delta_f \log_2\left(1 + \Gamma^{-1} \sigma_{k,m}^{-1} s_k^{\text{mask}} \beta_{k,\text{zf}}^{-2}\right),\ \forall m.$$

[3]Contrast this with the design of linear crosstalk cancelers for upstream VDSL, where the problem was to design a crosstalk canceler that avoids noise enhancement[5]. Here the problem is to design a crosstalk precoder that avoids transmit power enhancement.

Now consider the upper bound on the achievable data-rate (7). Since $b_{k,\mathrm{zfp}}^m \le b_{k,\mathrm{opt}}^m,\ \forall m$; the scaling factor is bounded

$$\beta_{k,\mathrm{zf}}^{-2} \le |h_k^{m,m}|^2 \left[1 + (N-1)\,\alpha_k\right]^2,\ \forall m.$$

This implies that

$$\beta_{k,\mathrm{zf}}^{-2} \le \left[1 + (N-1)\,\alpha_k\right]^2 \min_m |h_k^{m,m}|^2 .$$

Hence

$$b_{k,\mathrm{zfp}}^n \le \Delta_f \log_2\left(1 + \Gamma^{-1}\sigma_{k,n}^{-1} s_k^{\mathrm{mask}} \left[1 + (N-1)\,\alpha_k\right]^2 \min_m |h_k^{m,m}|^2\right),\ \forall n.$$

Since $[1 + (N-1)\alpha_k]^2 \simeq 1$, all users in the binder will experience a direct channel gain of approximately $\min_m |h_k^{m,m}|$. When the line lengths vary significantly, or if one of the lines in the binder contains a bridged tap, the weakest channel in the binder will be significantly weaker than the other channels. In this case the ZFP gives extremely sub-optimal performance. For example consider a scenario with ten 300 m lines and one 1200 m line. With the ZFP all lines will experience the direct channel of the 1200 m line. In many cases the ZFP leads to even worse performance than without crosstalk precoding as will be shown in Sec. VIII.

V. Tomlinson-Harashima Precoder

In downstream transmission a decision feedback structure can be used for crosstalk precoding[1]. This can be seen as the multi-user extension of the Tomlinson-Harashima precoder, which is commonly used for precoding against ISI in single-user channels[15], [16]. The structure of the *Tomlinson-Harashima Precoder* (THP) is now described. Consider the QR decomposition of the crosstalk channel matrix

$$\mathbf{H}_k^T \overset{\mathrm{qr}}{=} \mathbf{Q}_k \mathbf{R}_k, \tag{15}$$

where $\mathbf{Q}_k$ is a unitary matrix and $\mathbf{R}_k$ is upper triangular. Here $(.)^T$ is used to denote the transpose operation. Now

$$\mathbf{H}_k = \mathbf{R}_k^T \mathbf{Q}_k^T.$$

Prior to transmission, the signal is pre-multiplied with $\mathbf{Q}_k$ such that

$$\mathbf{x}_k = \mathbf{Q}_k \acute{\mathbf{x}}_k, \tag{16}$$

where the vector $\acute{\mathbf{x}}_k \triangleq [\acute{x}_k^1, \ldots, \acute{x}_k^N]^T$, whose meaning is yet to be defined. The received vector is then

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{Q}_k \acute{\mathbf{x}}_k + \mathbf{z}_k, \\ &= \mathbf{R}_k^T \acute{\mathbf{x}}_k + \mathbf{z}_k. \end{aligned} \tag{17}$$

The power of $\acute{x}_k^n$ is set to obey the spectral mask

$$\acute{s}_k^n \leq s_k^{\text{mask}},\ \forall n; \tag{18}$$

where $\acute{s}_k^n \triangleq \mathcal{E}\left\{|\acute{x}_k^n|^2\right\}$. Since $\mathbf{Q}_k$ is unitary, compliance with the spectral masks (2) is maintained after the precoding operation. To see this, we first assume that the signal $\acute{x}_k^n$ of each user is independent, which is approximately true[1]. Under this assumption

$$\begin{aligned} s_k^n &= \mathcal{E}\left\{\left|\sum_m [\mathbf{Q}_k]_{n,m} \acute{x}_k^m\right|^2\right\}, \\ &= \sum_m \left|[\mathbf{Q}_k]_{n,m}\right|^2 \acute{s}_k^m, \\ &\leq \sum_m \left|[\mathbf{Q}_k]_{n,m}\right|^2 s_k^{\text{mask}}, \\ &= s_k^{\text{mask}},\ \forall n; \end{aligned}$$

where (16) is used in the first line, (18) is used in the third line, and the unitarity of $\mathbf{Q}_k$ is used in the last line[1].

From (17) it is clear that the transmission channel has been transformed into a lower triangular channel $\mathbf{R}_k^T$. This channel is causal in the sense that there is an order in the crosstalk of the users. User 1 experiences crosstalk from no-one; user 2 experiences crosstalk only from user 1; user 3 experiences crosstalk only from users 1 and 2; and so on.

This causal structure admits the use of the Tomlinson-Harashima precoder to precompensate for the effects of crosstalk[1]. User 1 experiences no crosstalk. Hence the signal of user 1 can be transmitted directly; that is

$$\acute{x}_k^1 = \widetilde{x}_k^1,$$

where $\widetilde{x}_k^n$ denotes the true symbol intended for user n. At this point the signal transmitted by user 1 is known. This allows the remaining users to predistort their signals, and annihilate the crosstalk introduced by user 1. User 2 then operates free from crosstalk. The signal transmitted by user 2 is then also known, which allows the remaining users to predistort their signals, and annihilate the crosstalk introduced by user 2. This procedure iterates until all users have predistorted their signals to annihilate all crosstalk introduced in the channel. Each user is then received free from crosstalk. The symbol transmitted by user n on tone k after Tomlinson-Harashima precoding is thus

$$\acute{x}_k^n = \mathrm{mod}_{s_k^{\mathrm{mask}}} \left[\widetilde{x}_k^n - \sum_{m=1}^{n-1} \frac{r_k^{m,n}}{r_k^{n,n}} \acute{x}_k^m \right], \tag{19}$$

where the modulo operation is defined

$$\mathrm{mod}_M[x] \triangleq x - \sqrt{M} \left\lfloor \frac{x + \sqrt{M}/2}{\sqrt{M}} \right\rfloor .$$

Here $\lfloor \cdot \rfloor$ denotes the rounding-down operation[1], and $r_k^{n,m} \triangleq [\mathbf{R}_k]_{n,m}$. The modulo ensures that spectral mask compliance is maintained after precoding. That is, if $\widetilde{\mathbf{x}}_k$ obeys the spectral masks, then $\acute{\mathbf{x}}_k$ does as well.

At the receiver a second modulo operation is applied to estimate the transmitted symbol

$$\begin{aligned}
\hat{x}_k^n &= \mathrm{mod}_{s_k^{\mathrm{mask}}} \left[\frac{y_k^n}{r_k^{n,n}} \right], \\
&= \mathrm{mod}_{s_k^{\mathrm{mask}}} \left[\acute{x}_k^n + \sum_{m \neq n} \frac{r_k^{m,n}}{r_k^{n,n}} \acute{x}_k^m + \frac{z_k^n}{r_k^{n,n}} \right], \\
&= \mathrm{mod}_{s_k^{\mathrm{mask}}} \left[\widetilde{x}_k^n + \frac{z_k^n}{r_k^{n,n}} \right], \\
&= \widetilde{x}_k^n + \frac{\widetilde{z}_k^n}{r_k^{n,n}}.
\end{aligned}$$

The second line makes use of (17). The third line makes use of (19) and the property

$$\mathrm{mod}\left(\mathrm{mod}(a) + \mathrm{mod}(b)\right) = \mathrm{mod}\left(a + b\right).$$

The effective noise $\widetilde{z}_k^n$ in the fourth line is similar to the original noise z_k^n except that it exhibits a wrap-around effect on the edges of the QAM constellation. If the QAM constellation has many symbols, the wrap-around effect is rare and has negligible impact on performance[1]. Under this assumption the data-rate achieved by user n on tone k is

$$b_{k,\mathrm{th}}^n = \Delta_f \log_2(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \widetilde{s}_k^n \, |r_k^{n,n}|^2).$$

The RWDD of the channel matrix can be used to show that $|r_k^{n,n}| \simeq |h_k^{n,n}|$ [1]. As a result, for small α_k, the THP operates very close to the single-user bound

$$b_{k,\mathrm{th}}^n \simeq b_{k,\mathrm{mask}}^n.$$

So the THP gives near-optimal performance. Unfortunately it requires a modulo operation at the receiver and redesign of the CP modem. It is often preferable to place the computation burden of crosstalk precoding on the transmit side, since the *digital subscriber line access multiplexer* (DSLAM) often has a high level of computing power at its disposal. CP modems, on the other hand, are a commodity and must be produced as inexpensively as possible. Unfortunately with the THP, a modulo operation must be applied at both the transmit and receive side of the link. This modulo operation can increase CP modem complexity considerably, since it must be adapted to match the specific constellation used on each particular tone. In addition to this, CO and CP modems are typically manufactured by different hardware vendors, making joint design difficult.

VI. Diagonalizing Precoder

This section presents a simple linear precoder, the DP, that requires transmitter side operations only. Like the THP, this precoder operates close to the single-user bound. The DP multiplies the true symbols $\widetilde{\mathbf{x}}_k$ with a precoding matrix $\mathbf{P}_{k,\mathrm{dp}}$ prior to transmission. The transmitted symbols are

$$\mathbf{x}_k = \mathbf{P}_{k,\mathrm{dp}} \widetilde{\mathbf{x}}_k. \tag{20}$$

The DP is based on a channel diagonalizing criterion. After precoding, each user should see their own direct channel free from crosstalk. Contrast this with the ZFP where after precoding each user experiences a channel gain of unity, scaled by $\beta_{k,\mathrm{zf}}^{-1}$. The DP precoding matrix is defined

$$\mathbf{P}_{k,\mathrm{dp}} \triangleq \beta_{k,\mathrm{dp}}^{-1} \mathbf{H}_k^{-1} \mathrm{diag}\left\{ h_k^{1,1}, \ldots, h_k^{N,N} \right\},$$

where $\mathrm{diag}\{\gamma_1, \ldots, \gamma_N\}$ denotes the diagonal matrix with elements $\gamma_1, \ldots, \gamma_N$ along the main diagonal. Here the scaling factor is defined

$$\beta_{k,\mathrm{dp}} \triangleq \max_n \left\| \left[\mathbf{H}_k^{-1} \mathrm{diag}\left\{ h_k^{1,1}, \ldots, h_k^{N,N} \right\} \right]_{\mathrm{row}\, n} \right\|. \tag{21}$$

As with the ZFP, the scaling factor $\beta_{k,\mathrm{dp}}$ ensures that compliance with the spectral masks is maintained after precoding. That is, if $\widetilde{\mathbf{x}}_k$ obeys the spectral masks, then $\mathbf{x}_k$ will as well.

During transmission the predistortion introduced by the DP annihilates with the crosstalk. The received vector is then

$$\begin{aligned}\mathbf{y}_k &= \mathbf{H}_k\mathbf{P}_{k,\mathrm{dp}}\widetilde{\mathbf{x}}_k + \mathbf{z}_k, \\ &= \beta_{k,\mathrm{dp}}^{-1}\mathrm{diag}\left\{h_k^{1,1},\ldots,h_k^{N,N}\right\}\widetilde{\mathbf{x}}_k + \mathbf{z}_k. \end{aligned} \tag{22}$$

Application of the DP diagonalizes the channel matrix. Each user now experiences its direct channel, scaled by $\beta_{k,\mathrm{dp}}$ and completely free from interference. RWDD in the crosstalk channel matrix implies that $\beta_{k,\mathrm{dp}} \simeq 1$. As a result, each user operates close to its single-user bound, and the DP is near-optimal. To see this consider the *singular value decomposition* (SVD) of $\mathbf{H}_k$

$$\mathbf{H}_k \stackrel{\mathrm{svd}}{=} \mathbf{U}_k\Lambda_k\mathbf{V}_k^H. \tag{23}$$

The RWDD of $\mathbf{H}_k$, as described in (3), ensures that its rows are approximately orthogonal. As a result

$$\mathbf{H}_k\mathbf{H}_k^H \simeq \mathrm{diag}\left\{\left|h_k^{1,1}\right|^2,\ldots,\left|h_k^{N,N}\right|^2\right\}.$$

Combining this with (23) leads to the following approximation

$$\mathbf{U}_k\Lambda_k\Lambda_k^H\mathbf{U}_k^H \simeq \mathrm{diag}\left\{\left|h_k^{1,1}\right|^2,\ldots,\left|h_k^{N,N}\right|^2\right\}.$$

This implies that the left singular vectors can be closely approximated as

$$\mathbf{U}_k \simeq \mathbf{I}_N. \tag{24}$$

Furthermore, the singular values can be approximated

$$\left|[\Lambda_k]_{n,n}\right| \simeq \left|h_k^{n,n}\right|. \tag{25}$$

Using (21), the scaling factor can then be approximated as

$$\begin{aligned}\beta_{k,\mathrm{dp}} &= \max_n\left\|\left[\mathbf{V}_k\Lambda_k^{-1}\mathbf{U}_k^H\mathrm{diag}\left\{h_k^{1,1},\ldots,h_k^{N,N}\right\}\right]_{\mathrm{row}\,n}\right\|, \\ &\simeq \max_n\left\|[\mathbf{V}_k]_{\mathrm{row}\,n}\right\|, \\ &= 1, \end{aligned}$$

where (24) and (25) are applied in the second line. The motivation behind the DP design is now clear. Since

$$\mathbf{H}_k^{-1}\mathrm{diag}\left\{h_k^{1,1},\ldots,h_k^{N,N}\right\} \simeq \mathbf{V}_k,$$

the DP precoding matrix is close to unitary. As a result, the scaling factor $\beta_{k,\mathrm{dp}}$ will be close to unity. The DP then achieves near-optimal performance, operating close to the single-user bound. This observation is made rigorous through the following theorem.

Theorem 2: If $A_{\text{min}}^{(m)} \geq \alpha_k m B_{\text{max}}^{(m)},\ m = 1 \ldots N-1$; the data-rate achieved by the DP can be lower bounded

$$R_n \geq \sum_k b_{k,\mathrm{dp-bnd}}^n, \tag{26}$$

where

$$b_{k,\mathrm{dp-bnd}}^n \triangleq \Delta_f \log_2 \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \widetilde{s}_k^n \left|h_k^{n,n}\right|^2 f^{-1}(N, \alpha_k)\right), \tag{27}$$

$$f(N, \alpha_k) \triangleq \left(\frac{A_{\text{max}}^{(N-1)}}{A_{\text{min}}^{(N)}}\right)^2 + (N-1)\left(\frac{B_{\text{max}}^{(N-1)}}{A_{\text{min}}^{(N)}}\right)^2, \tag{28}$$

$$\begin{bmatrix} A_{\text{max}}^{(m)} \\ B_{\text{max}}^{(m)} \end{bmatrix} \triangleq \left(\prod_{i=1}^{m} \begin{bmatrix} 1 & (i-1)\,\alpha_k \\ \alpha_k & (i-1)\,\alpha_k \end{bmatrix}\right) \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{29}$$

and

$$A_{\text{min}}^{(m)} \triangleq 1 - \sum_{i=1}^{m} \alpha_k \,(i-1)\, B_{\text{max}}^{(i-1)}. \tag{30}$$

Proof of Theorem 2: Equation (22) implies that after application of the DP the signal at receiver n is

$$y_k^n = \beta_{k,\mathrm{dp}}^{-1} h_k^{n,n} \widetilde{x}_k^n + z_k^n.$$

Hence the received signal power for user n on tone k is $\beta_{k,\mathrm{dp}}^{-2} \widetilde{s}_k^n \left|h_k^{n,n}\right|^2$, the received interference power is zero, and the received noise power is $\sigma_{k,n}$. So the data-rate achieved by the diagonalizing precoder is

$$b_{k,\mathrm{dp}}^n(\widetilde{s}_k^n) = \log_2 \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \beta_{k,\mathrm{dp}}^{-2} \widetilde{s}_k^n \left|h_k^{n,n}\right|^2\right). \tag{31}$$

Define the matrix $\overline{\mathbf{G}}_k \triangleq [\overline{g}_k^{n,m}]$, where $\overline{g}_k^{n,m} \triangleq h_k^{n,m}/h_k^{n,n}$. Now

$$\mathbf{H}_k = \mathrm{diag}\left\{h_k^{1,1}, \ldots, h_k^{N,N}\right\} \overline{\mathbf{G}}_k,$$

hence

$$\mathbf{H}_k^{-1} \mathrm{diag}\left\{h_k^{1,1}, \ldots, h_k^{N,N}\right\} = \overline{\mathbf{G}}_k^{-1}. \tag{32}$$

Since the transmitters are co-located at the CO, the DS channel is RWDD (4). This implies that $\overline{\mathbf{G}}_k \in \mathbb{A}^{(N)}$, where $\mathbb{A}_n^{(N)}$ denotes the set of $N \times N$ diagonally dominant matrices, as defined in the Appendix. Theorem 5 from the Appendix can be applied to bound the elements of $\overline{\mathbf{G}}_k^{-1}$ as follows

$$\left| \left[\overline{\mathbf{G}}_k^{-1} \right]_{n,m} \right| \leq \begin{cases} A_{\text{max}}^{(N-1)} / A_{\text{min}}^{(N)}, & n = m; \\ B_{\text{max}}^{(N-1)} / A_{\text{min}}^{(N)}, & n \neq m. \end{cases} \tag{33}$$

Combining (21) and (32) implies $\beta_{k,\text{dp}}^2 \leq f(N, \alpha_{k,})$. Combining this with (31) leads to (27), which concludes the proof. ■

In practice we have found that $A_{\text{min}}^{(m)} \geq \alpha_k m B_{\text{max}}^{(m)},\ m = 1 \ldots N-1$; holds for N up to 25 and for frequencies up to 12 MHz, so the bound applies in most practical VDSL scenarios.

The function $f(N, \alpha_k)$ can be interpreted as an upper bound on the scaling factor $\beta_{k,\text{dp}}^2$. Recall that the scaling factor is the increase in transmit power that results from precoding with $\mathbf{H}_k^{-1} \text{diag} \left\{ h_k^{1,1}, \ldots, h_k^{N,N} \right\}$. It is included to ensure that the DP does not increase the transmit power on any line. In RWDD channels it is found that $f(N, \alpha_k) \simeq 1$. As a result each modem operates at a rate

$$b_{k,\text{zf}}^n \simeq \Delta_f \log_2 \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \widetilde{s}_k^n \, |h_k^{n,n}|^2 \right).$$

So the DP completely removes crosstalk. A scaling factor close to unity can be chosen, and each user operates close to their single-user bound.

Note that the bound (26) can be used to guarantee a data-rate without explicit knowledge of the crosstalk channels. This is because the bound only depends on the binder size, direct channel gain, and background noise power. Good models for these characteristics exist based on extensive measurement campaigns. Crosstalk channels on the other hand are poorly understood and actual channels can deviate significantly from the few empirical models that exist, see for example Fig. 2. This can make provisioning of services difficult.

Using the bound (26) allows us to overcome this problem. The bound tells us that the actual crosstalk channel gain is not important as long as RWDD is observed. RWDD is a well understood and modeled phenomenon. The value for α_k from (5) is based on worst 1% case models. Hence for 99% of lines α_k will be smaller and a data-rate above the bound (26) is achieved. So the bound is a useful tool not just for theoretical analysis, but for provisioning of services as well. Simulations in Sec. VIII will use this bound to show that the DP operates close to the single-user bound, and hence is a near-optimal design.

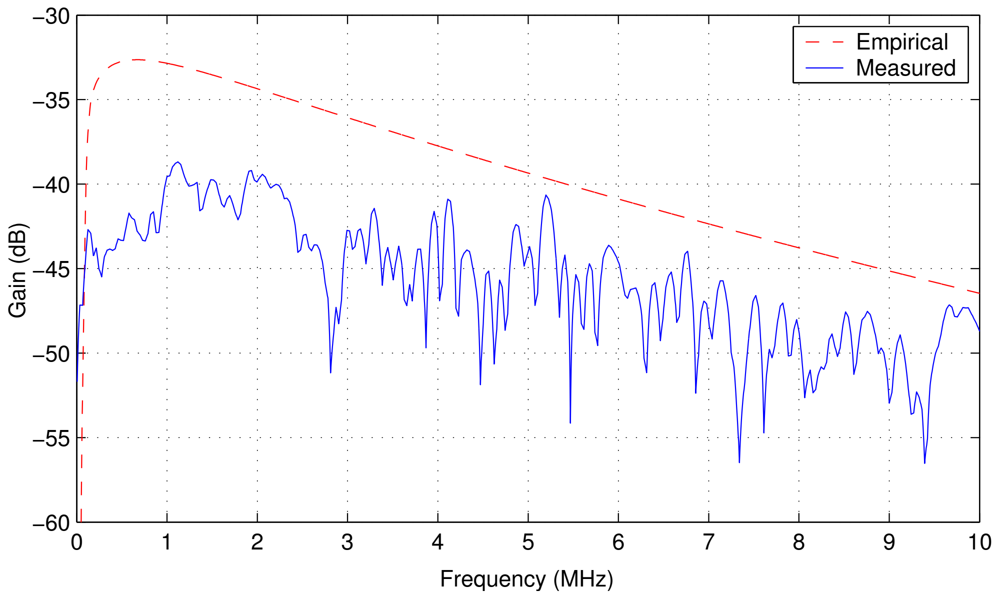

Fig. 2. Crosstalk Channel Transfer Functions (1 km cable, 0.5 mm pairs)

It is interesting to compare this to upstream VDSL where the channel matrix is CWDD. As a result the right, rather than the left, singular vectors are close to the identity matrix. The upstream channel can then be approximated $\mathbf{H}_k \simeq \mathbf{U}_k\Lambda_k$. The linear zero-forcing canceler can be applied at the receiver side, and is well approximated by $\mathbf{H}_k^{-1} \simeq \Lambda_k^{-1}\mathbf{U}_k^H$. Note that $\mathbf{U}_k^H$ is unitary and has no effect on the noise statistics if the background noise is spatially white. The diagonal matrix Λ_k scales the noise and signal powers equally and hence does not cause noise enhancement. This was used in previous work to show that the linear ZF canceler is near-optimal in upstream VDSL[5]. Here we have examined the dual problem and used the RWDD of $\mathbf{H}_k$ to show that the DP is near-optimal for downstream VDSL.

VII. Spectra Optimization

Whilst current VDSL standards require the use of spectral masks, there is growing interest in the use of adaptive transmit spectra, a technique known as dynamic spectrum management[17]. This section investigates the optimization of transmit spectra for use with the DP. Each transmitter is then subject to a total power constraint

$$\sum_k s_k^n \leq P_n, \ \forall n; \tag{34}$$

which replaces the spectral mask constraint (2). The goal is to maximize a weighted sum of the data-rates of the modems within the network

$$\max_{\mathbf{s}_1,\ldots,\mathbf{s}_N} \sum_n w_n R_n \quad \text{s.t. } \sum_k s_k^n \leq P_n, \; \forall n; \tag{35}$$

where the vector $\mathbf{s}_n \triangleq [s_1^n, \ldots, s_K^n]$ contains the PSD of user n on all tones. The weights $w_1, \ldots, w_N$ are used to ensure that each modem achieves its target data-rate. The data-rate R_n is a function of the transmit PSDs $\mathbf{s}_1, \ldots, \mathbf{s}_N$, and also depends on the type of crosstalk precoder used.

If an optimal precoder is used, the objective function becomes convex[12][18][19]. Solving (35) then requires the solution of a KN-dimensional convex optimization. Although the cost function is convex, no closed form solution is known[19]. Conventional convex optimization techniques, such as interior point methods, have a polynomial complexity in the dimensionality of the search space. In ADSL $K = 256$, whilst in VDSL $K = 4096$. The binder size N can be anything between 2 and 100. The resulting search has a high dimensionality, for which these techniques are prohibitively complex. A low complexity, iterative algorithm has been proposed for the special case where an unweighted rate-sum is maximized, that is $w_n = 1$ for all n[12]. Unfortunately, since this algorithm cannot optimize a weighted rate-sum, it cannot ensure that the target rates are achieved on each line. These target rates are essential to ensure that each customer achieves their desired service.

In this section a spectra optimization algorithm is developed for use with the DP. Since the DP removes all crosstalk, the spectrum optimization can be solved efficiently through a dual decomposition. This reduces complexity considerably. Furthermore, Theorem 2 ensures that this approach operates close to the single-user bound.

A. Theoretical Capacity

We start by extending the single-user bound from Sec. III to VDSL modems that may vary their transmit spectra under a total power constraint (34). The resulting upper bound is useful for evaluating crosstalk precoder performance with optimized spectra.

Theorem 3: When the transmit PSD s_k^n is allowed to vary under a total power constraint (34), the achievable data-rate for user n can be upper bounded

$$R_n \leq \max_{\mathbf{s}_n} \sum_k b_{k,\text{bnd}}^n(s_k^1, \ldots, s_k^N) \quad \text{s.t. } \sum_k s_k^m \leq P_m, \; \forall m; \tag{36}$$

where $b^n_{k,\text{bnd}}$ is defined in (11).

Proof of Theorem 3: Follows from inspection of (11) and (34). ■

The optimization problem described by (36) is convex, and can be solved efficiently using interior-point methods. The important observation is that this bound does not depend on the crosstalk channels. Hence an upper bound on performance can be obtained without explicit knowledge of the binder configuration.

B. Near-optimal Linear Precoder

Transmit spectra optimization with the DP is now considered. Since the transmit spectra are being optimized under a total power constraint, spectral masks do not apply. As a result, the scaling factor $\beta_{k,\text{dp}}$ can be discarded by setting

$$\beta_{k,\text{dp}} = 1, \ \forall k; \tag{37}$$

and so the DP simplifies to

$$\mathbf{P}_{k,\text{dp}} \triangleq \mathbf{H}_k^{-1} \text{diag}\left\{ h_k^{1,1}, \ldots, h_k^{N,N} \right\}.$$

From (20), the signal sent by transmitter n on tone k is

$$x_k^n = \sum_m p_{k,\text{dp}}^{n,m} \widetilde{x}_k^m,$$

where the elements of the precoding matrix are defined $p_{k,\text{dp}}^{n,m} \triangleq [\mathbf{P}_{k,\text{dp}}]_{n,m}$. Hence the power on line n is

$$s_k^n = \sum_m \left| p_{k,\text{dp}}^{n,m} \right|^2 \widetilde{s}_k^m. \tag{38}$$

Combining (31), (37) and (38), the original optimization problem (35) becomes

$$\max_{\widetilde{\mathbf{s}}_1, \ldots, \widetilde{\mathbf{s}}_N} \sum_n \sum_k w_n \log_2 \left(1 + \Gamma^{-1} \sigma_{k,n}^{-1} \left| h_k^{n,n} \right|^2 \widetilde{s}_k^n \right) \qquad \text{s.t.} \ \sum_k \sum_m \left| p_{k,\text{dp}}^{n,m} \right|^2 \widetilde{s}_k^m \le P_n, \ \forall n.$$

Observe that, when using the DP, the data-rate of each user depends only on its own transmit PSD. It is independent of the PSDs of the other users since all interference will be pre-filtered. Unfortunately, the optimization is still coupled between users. This is the case since the total power constraint on each modem must be satisfied *after* the precoding operation. As a result the PSD sent by a particular user $\widetilde{s}_k^n$ is not equal to the transmit PSD of the corresponding modem

Algorithm 1 Optimal Power Allocation with the DP

repeat

for each n: $\widetilde{s}_k^n = \left[w_n \left(\sum_m \lambda_m \left|p_{k,\mathrm{dp}}^{m,n}\right|^2\right)^{-1} - \Gamma\sigma_{k,n}\left|h_k^{n,n}\right|^{-2}\right]^+, \ \forall k$

for each n: $\lambda_n = \left[\lambda_n + \epsilon\left(\sum_k \sum_m \left|p_{\mathrm{dp},k}^{n,m}\right|^2 \widetilde{s}_k^m - P_n\right)\right]^+$

until convergence

s_k^n. These PSDs are coupled through the precoding matrix $\mathbf{P}_{k,\mathrm{dp}}$ and as a result the optimization must be done jointly across all users. Nevertheless, it is still possible to optimize the transmit PSDs efficiently through the use of a dual objective. First note that the objective function is concave and the constraints form a convex set. As a result the KKT conditions are sufficient for optimality. Examining these leads to

$$\widetilde{s}_{k,\mathrm{dp}}^n = \left[\frac{w_n}{\sum_m \lambda_m \left|p_{k,\mathrm{dp}}^{m,n}\right|^2} - \frac{\Gamma\sigma_{k,n}}{\left|h_k^{n,n}\right|^2}\right]^+,$$

where $[x]^+ \triangleq \max(0, x)$. This is effectively a waterfilling solution, with a waterfilling level that implicitly takes into account the power constraints on each line through the Lagrangian multipliers $\lambda_1, \ldots, \lambda_N$. The power allocation of each user is coupled through the Lagrangian multipliers. The Lagrangian multipliers must be chosen such that for each line the power constraint is tight

$$\sum_k \sum_m p_{k,\mathrm{dp}}^{n,m} \widetilde{s}_{k,\mathrm{dp}}^m = P_n, \ \forall n;$$

or the corresponding Lagrangian multiplier λ_n is zero. An efficient solution can be found with Alg. 1, which uses sub-gradient descent in λ-space[20]. The algorithm is found to work well with a step-size ϵ equal to $1/N$. Alg. 1 has polynomial complexity in N and linear complexity in K. This is a significant reduction compared to existing power allocation algorithms for the broadcast channel, which have polynomial complexity in KN[19].

VIII. Performance

This section evaluates the performance of the DP in a binder of 8 VDSL lines. The line lengths range from 150 m to 1200 m in 150 m increments as shown in Fig. 3. For all simulations the line diameter is 0.5 mm (24-AWG). Direct and crosstalk channel transfer functions are generated

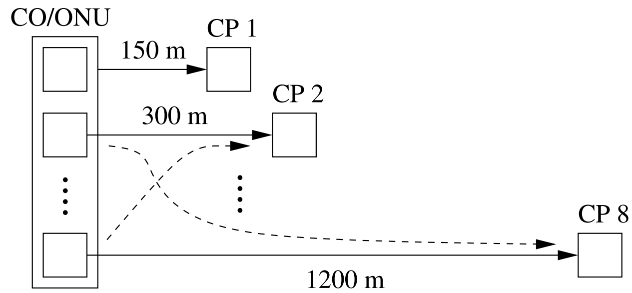

Fig. 3. Downstream VDSL scenario

using semi-empirical models[8]. The target symbol error probability is 10^{-7} or less. The coding gain is set to 3 dB and the noise margin to 6 dB. As per the VDSL standards the tone-spacing Δ_f is set to 4.3125 kHz[21][8]. The modems use 4096 tones, and the 998 FDD bandplan. Background noise is generated using ETSI noise model A[8]. Performance is compared with the ZFP, THP and the single-user bound.

A. Fixed Transmit Spectra

Current VDSL standards require that modems transmit under a spectral mask of -60 dBm/Hz[21][8]. This section evaluates the performance of the DP when all modems are operating at this mask.

Fig. 4 shows the data-rate achieved on each of the lines with the different crosstalk precoding schemes. As predicted, the ZFP gives quite poor performance, with all lines forced to operate at the rate of the weakest line in the binder, which in this case is the 1200 m line. In fact, for all of the lines shorter than 1200 m, the ZFP results in worse performance than with no precoding at all.

The DP avoids the problems of the ZFP, and achieves substantial gains, typically 30 Mbps or more, over conventional systems with no crosstalk precoding. As can be seen in Fig. 4 the DP achieves near-optimal performance, operating close to the single-user bound. This is a direct result of the RWDD of $\mathbf{H}_k$, which ensures that the scaling parameter $\beta_{k,\mathrm{dp}}$ is always close to unity. The scaling parameters of the ZFP $\beta_{k,\mathrm{zf}}$ and the DP $\beta_{k,\mathrm{dp}}$ are plotted for each tone in Fig. 5. The scaling parameter in the ZFP is quite large, which results in poor performance on the shorter lines. On the other hand the scaling parameter in the DP is typical close to unity, and hence it has negligible effect on performance.

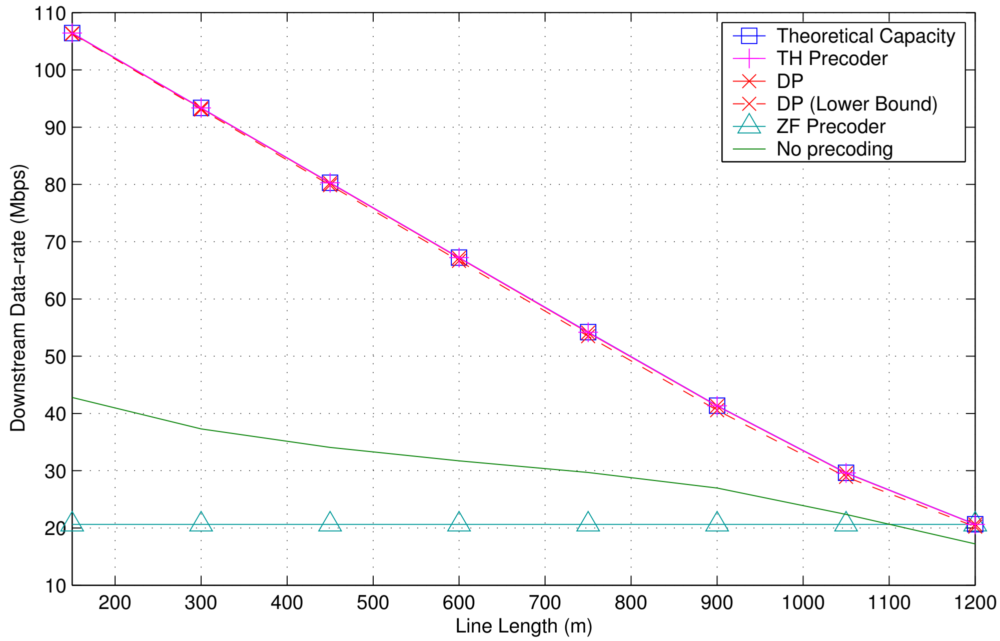

Fig. 4. Data-rate with Different Precoders

Fig. 6 shows the data-rate of the DP as a percentage of the single-user bound. Performance does not drop below 99% of the single-user bound. The lower bound on the performance of the DP (26) is also included for comparison. As can be seen the bound is quite tight and guarantees that the DP will achieve at least 97% of the single-user bound.

B. Optimized Transmit Spectra

This section investigates the performance of the DP with spectra optimized by Alg. 1. A total power constraint of 11.5 dBm is applied to each modem as per the VDSL standards[21][8]. Spectral mask constraints are not applied. Fig. 7 shows the data-rates achieved on each line. The use of optimized spectra yields a gain of 5-8 Mbps. The benefit is more substantial on the longer lines, where a 5 Mbps gain can double the achievable data-rate.

Fig. 7 shows that spectra optimization gives maximum benefit on long lines. This is to be expected since on long lines the direct channel gain decreases more rapidly with frequency. Note that the benefit of adaptive spectra, when crosstalk has already been cancelled, comes primarily from the modem loading power in the best parts of the channel, which are typically in the lower

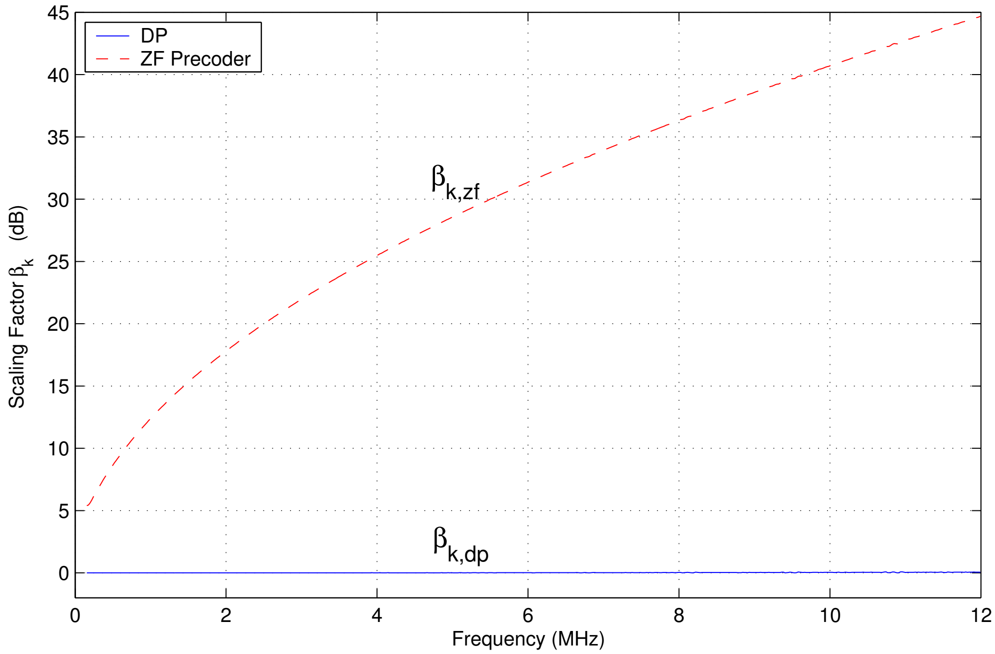

Fig. 5. Scaling Factor β_k

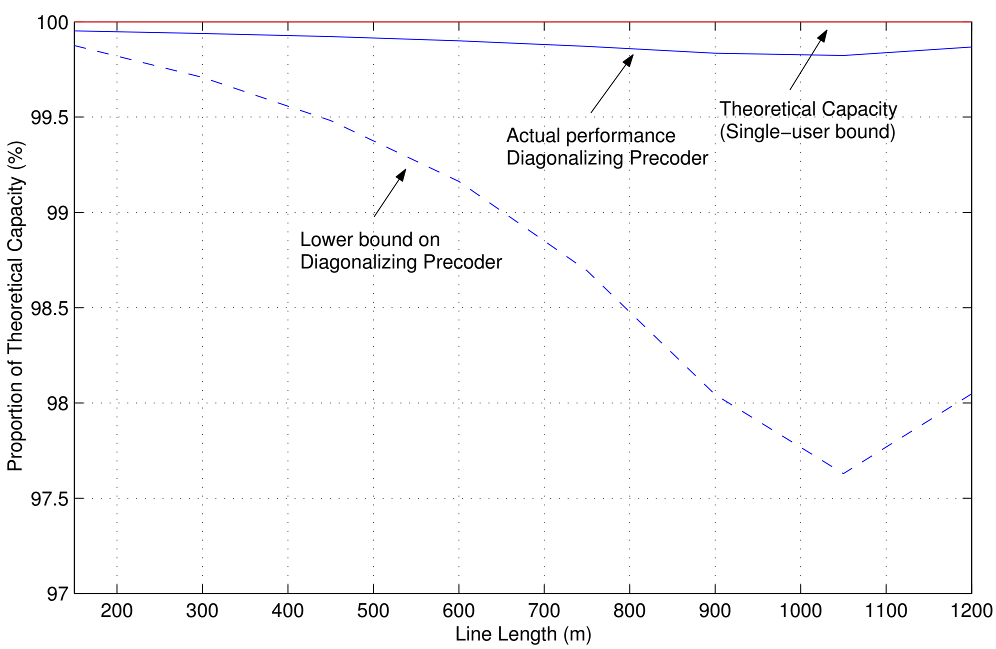

Fig. 6. Proportion of Theoretical Capacity Achieved by DP

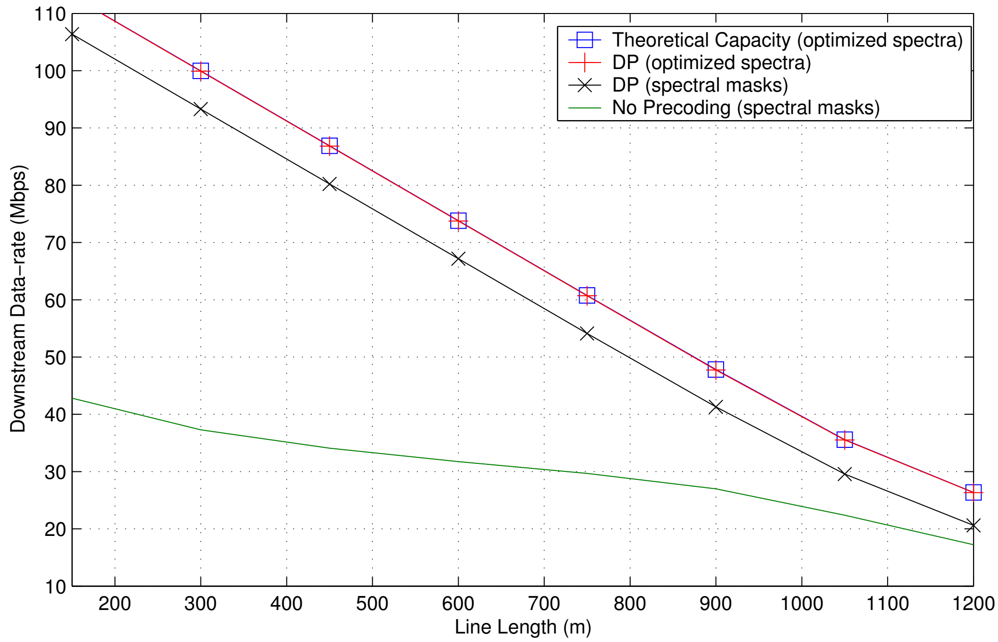

Fig. 7. Data-rate Achieved with Optimized Spectra

frequencies.

IX. Conclusions

This paper investigated the design of crosstalk precoders for downstream VDSL. Existing designs suffer from poor performance or have a high run-time complexity. A novel linear precoder, based on a channel diagonalizing criterion, is proposed. This design has a low complexity does not require receiver-side operations. This is important since it helps to keep CP modem complexity as low as possible.

Any linear crosstalk precoder must include a scaling factor to ensure that spectral masks are maintained after precoding. In some cases this scaling factor decreases performance by forcing certain modems to operate below their transmit mask. Fortunately VDSL channels with co-located transmitters are row-wise diagonal dominant. This ensures that the scaling factor of the diagonalizing precoder (DP) is close to unity, and hence has negligible impact on performance.

An upper bound on the capacity of the multi-user VDSL channel was derived. This single-user bound shows that spatial diversity in the VDSL environment is negligible. Therefore the best outcome that a crosstalk precoder can achieve is the complete pre-filtering of crosstalk without increasing transmit power.

A lower bound on the performance of the DP was derived. This bound depends only on the binder size, direct channel gain and background noise for which reliable models and statistical data exist. As a result the performance of the DP can be accurately predicted, which simplifies service provisioning considerably. This bound shows that the DP operates close to the single-user bound. So the DP is a low complexity design with guaranteed near-optimal performance.

The combination of spectra optimization and crosstalk precoding was also considered. The bounds were extended to VDSL systems with optimized spectra. Spectra optimization in a broadcast channel generally involves a highly complex optimization problem. Since the DP decouples transmission on each line, the spectrum on each modem can be optimized through a dual decomposition, leading to a significant reduction in complexity.

Appendix

Define the set $\mathbb{A}^{(N)}$ of $N \times N$ matrices, such that for any $\mathbf{A}^{(N)} \in \mathbb{A}^{(N)}$, it holds that

$$\begin{aligned} |a_{n,n}| &= 1; \\ |a_{n,m}| &\leq \alpha_k, \ \forall n \neq m; \end{aligned}$$

where $a_{n,m} \triangleq \left[\mathbf{A}^{(N)}\right]_{n,m}$. Define the set $\mathbb{B}^{(N)}$ of $N \times N$ matrices, such that for any $\mathbf{B}^{(N)} \in \mathbb{B}^{(N)}$, it holds that

$$\begin{aligned} |b_{n,n}| &= 1, \ \forall n < N; \\ |b_{N,N}| &\leq \alpha_k; \\ |b_{n,m}| &\leq \alpha_k, \ \forall n \neq m; \end{aligned}$$

where $b_{n,m} \triangleq \left[\mathbf{B}^{(N)}\right]_{n,m}$.

Theorem 4: Consider any $\mathbf{A}^{(N)} \in \mathbb{A}^{(N)}$ and $\mathbf{B}^{(N)} \in \mathbb{B}^{(N)}$. The magnitude of the determinants of $\mathbf{A}^{(N)}$ and $\mathbf{B}^{(N)}$ can be bounded as follows

$$\left|\det\left(\mathbf{A}^{(N)}\right)\right| \leq A_{\max}^{(N)}, \tag{39}$$

$$\left|\det\left(\mathbf{B}^{(N)}\right)\right| \leq B_{\max}^{(N)}, \tag{40}$$

where $A_{\max}^{(N)}$, $B_{\max}^{(N)}$ and $A_{\min}^{(N)}$ are defined in (29) and (30). Furthermore, if

$$A_{\min}^{(m)} \geq \alpha_k m B_{\max}^{(m)},\ m = 1 \ldots N-1; \tag{41}$$

then the following bound also holds

$$\left|\det\left(\mathbf{A}^{(N)}\right)\right| \geq A_{\min}^{(N)}. \tag{42}$$

Note that $|\cdot|$ denotes the absolute value operator, whilst $\det(\cdot)$ denotes the determinant operator.

Proof of Theorem 4: The proof is based on induction. Begin by assuming that the bounds (39), (40) and (42) hold for any $N \times N$ matrices of the form $\mathbf{A}^{(N)}$ and $\mathbf{B}^{(N)}$ for some specific value of N. Now consider any matrix $\mathbf{A}^{(N+1)} \in \mathbb{A}^{(N+1)}$. Decompose $\mathbf{A}^{(N+1)}$ as

$$\mathbf{A}^{(N+1)} = \left[\begin{array}{ccc|c} & & & a_{1,N+1} \\ & \mathbf{A}^{(N)} & & \vdots \\ & & & a_{N,N+1} \\ a_{N+1,1} & \cdots & a_{N+1,N} & 1 \end{array}\right],$$

where $a_{n,m} \triangleq \left[\mathbf{A}^{(N+1)}\right]_{n,m}$ and $\mathbf{A}^{(N)}$ is the submatrix containing the first N rows and columns of $\mathbf{A}^{(N+1)}$. By expanding the determinant along the last row of $\mathbf{A}^{(N+1)}$ it can be seen that

$$\begin{aligned} \left|\det\left(\mathbf{A}^{(N+1)}\right)\right| &= \left|\det\left(\mathbf{A}^{(N)}\right) + \sum_{m=1}^{N} (-1)^{N+1-m} a_{N+1,m} \det\left(\left[\begin{array}{cc} \overline{\mathbf{A}}_m^{(N)} & \mathbf{a}_{N+1} \end{array}\right]\right)\right|, & (43) \\ &\leq \left|\det\left(\mathbf{A}^{(N)}\right)\right| + \sum_{m=1}^{N} \alpha_k \left|\det\left(\left[\begin{array}{cc} \overline{\mathbf{A}}_m^{(N)} & \mathbf{a}_{N+1} \end{array}\right]\right)\right|, & (44) \end{aligned}$$

where $\overline{\mathbf{A}}_m^{(N)}$ is the sub-matrix formed by removing column m from $\mathbf{A}^{(N)}$ and $\mathbf{a}_{N+1} \triangleq [a_{1,N+1} \ldots a_{N,N+1}]^T$. The second line exploits the fact that row permutation does not affect the magnitude of a determinant. Define the permutation matrix

$$\Pi_m \triangleq [\mathbf{e}_1 \cdots \mathbf{e}_{m-1}\ \mathbf{e}_{m+1} \ldots \mathbf{e}_N\ \mathbf{e}_m],$$

where $\mathbf{e}_m$ is defined as the mth column of the $N \times N$ identity matrix. Note that $\Pi_m^T [\overline{\mathbf{A}}_m^{(N)}\ \mathbf{a}_{N+1}] \in \mathbb{B}^{(N)}$. Using the fact that row permutations have no effect on the magnitude of a determinant, together with (40) and (44) now implies

$$\sum_{m=1}^{N} \alpha_k \left|\det\left(\left[\begin{array}{cc} \overline{\mathbf{A}}_m^{(N)} & \mathbf{a}_{N+1} \end{array}\right]\right)\right| \leq \alpha_k N B_{\max}^{(N)}. \tag{45}$$

Combining this with (44) and (39) yields

$$\left|\det\left(\mathbf{A}^{(N+1)}\right)\right| \leq A_{\max}^{(N)} + \alpha_k N B_{\max}^{(N)}.$$

Note that by definition

$$A_{\max}^{(N+1)} = A_{\max}^{(N)} + \alpha_k N B_{\max}^{(N)}, \tag{46}$$

hence

$$\left|\det\left(\mathbf{A}^{(N+1)}\right)\right| \leq A_{\max}^{(N+1)}. \tag{47}$$

Now consider any matrix $\mathbf{B}^{(N+1)} \in \mathbb{B}^{(N+1)}$. Decompose $\mathbf{B}^{(N+1)}$ as

$$\mathbf{B}^{(N+1)} = \left[\begin{array}{ccc|c} & & & b_{1,N+1} \\ & \mathbf{C}^{(N)} & & \vdots \\ & & & b_{N,N+1} \\ b_{N+1,1} & \cdots & b_{N+1,N} & b_{N+1,N+1} \end{array}\right],$$

where $b_{n,m} \triangleq \left[\mathbf{B}^{(N+1)}\right]_{n,m}$ and $\mathbf{C}^{(N)}$ is the submatrix containing the first N rows and columns of $\mathbf{B}^{(N+1)}$. By expanding the determinant along the last row of $\mathbf{B}^{(N+1)}$ it can be seen that

$$\left|\det\left(\mathbf{B}^{(N+1)}\right)\right| = \left|b_{N+1,N+1}\det\left(\mathbf{C}^{(N)}\right) + \sum_{m=1}^{N}(-1)^{N+1-m}\, b_{N+1,m}\det\left(\left[\begin{array}{cc} \overline{\mathbf{C}}_m^{(N)} & \mathbf{b}_{N+1}\end{array}\right]\right)\right|, \tag{48}$$

where $\overline{\mathbf{C}}_m^{(N)}$ is the sub-matrix formed by removing column m from $\mathbf{C}^{(N)}$ and $\mathbf{b}_{N+1} \triangleq [b_{1,N+1} \ldots b_{N,N+1}]^T$. Note that $\mathbf{C}^{(N)} \in \mathbb{A}^{(N)}$ and

$$\Pi_m^T \left[\begin{array}{cc} \overline{\mathbf{C}}_m^{(N)} & \mathbf{b}_{N+1}\end{array}\right] \in \mathbb{B}^{(N)}.$$

Using the fact that row permutations have no effect on the magnitude of a determinant, together with (39), (40), and (48) now yields

$$\left|\det(\mathbf{B}^{(N+1)})\right| \leq \alpha_k A_{\max}^{(N)} + \alpha_k N B_{\max}^{(N)}.$$

Note that by definition

$$B_{\max}^{(N+1)} = \alpha_k A_{\max}^{(N)} + \alpha_k N B_{\max}^{(N)}, \tag{49}$$

hence

$$\left|\det(\mathbf{B}^{(N+1)})\right| \leq B_{\max}^{(N+1)}. \tag{50}$$

Combining (46) and (49) in matrix form yields

$$\begin{bmatrix} A_{\max}^{(N+1)} \\ B_{\max}^{(N+1)} \end{bmatrix} = \begin{bmatrix} 1 & \alpha_k N \\ \alpha_k & \alpha_k N \end{bmatrix} \begin{bmatrix} A_{\max}^{(N)} \\ B_{\max}^{(N)} \end{bmatrix}. \tag{51}$$

We now proceed with the inductive proof. First note that $\left|\mathbf{A}^{(1)}\right| = 1$ and $\left|\mathbf{B}^{(1)}\right| \leq \alpha_k$, so (39) and (40) hold for $N = 1$. Hence through induction, (47) and (50) imply that (39) and (40) must hold for all N. This concludes the proof for the upper bounds (39) and (40). We now turn our attention to the lower bound (42). First note that from (43)

$$\left|\det\left(\mathbf{A}^{(N+1)}\right)\right| \geq \left|\left|\det\left(\mathbf{A}^{(N)}\right)\right| - \sum_{m=1}^{N} \alpha_k \left|\det\left(\left[\begin{array}{cc} \overline{\mathbf{A}}_m^{(N)} & \mathbf{a}_{N+1} \end{array}\right]\right)\right|\right|. \tag{52}$$

We assume that (42) holds for some specific value of N. Hence

$$\left|\det(\mathbf{A}^{(N)})\right| \geq A_{\min}^{(N)}. \tag{53}$$

Combining this with (41) and (45) implies

$$\left|\det\left(\mathbf{A}^{(N)}\right)\right| \geq \sum_{m=1}^{N} \alpha_k \left|\det\left(\left[\begin{array}{cc} \overline{\mathbf{A}}_m^{(N)} & \mathbf{a}_{N+1} \end{array}\right]\right)\right|.$$

So from (52)

$$\left|\det\left(\mathbf{A}^{(N+1)}\right)\right| \geq \left|\det\left(\mathbf{A}^{(N)}\right)\right| - \sum_{m=1}^{N} \alpha_k \left|\det\left(\left[\begin{array}{cc} \overline{\mathbf{A}}_m^{(N)} & \mathbf{a}_{N+1} \end{array}\right]\right)\right|.$$

Combining this with (53) and (45) leads to the bound

$$\left|\det(\mathbf{A}^{(N+1)})\right| \geq A_{\min}^{(N)} - \alpha_k N B_{\max}^{(N)}.$$

Note that by definition

$$A_{\min}^{(N+1)} \triangleq A_{\min}^{(N)} - \alpha_k N B_{\max}^{(N)},$$

hence

$$\left|\det(\mathbf{A}^{(N+1)})\right| \geq A_{\min}^{(N+1)} \tag{54}$$

Now note that $\left|\mathbf{A}^{(1)}\right| = 1$ and $A_{\min}^{(1)} = 1$, so (42) holds for $N = 1$. Hence through induction, (54) implies that (42) holds for all N. This concludes the proof for the lower bound (42). ■

Theorem 5: If $\mathbf{G} \in \mathbb{A}^{(N)}$ and $A_{\min}^{(m)} \geq \alpha_k m B_{\max}^{(m)},\ m = 1 \ldots N-1$; then the magnitude of the elements of $\mathbf{G}^{-1}$ can be bounded

$$\left| \left[\mathbf{G}^{-1} \right]_{n,m} \right| \leq \begin{cases} A_{\max}^{(N-1)} / A_{\min}^{(N)}, & n = m; \\ B_{\max}^{(N-1)} / A_{\min}^{(N)}, & n \neq m. \end{cases} \tag{55}$$

Proof of Theorem 5: By definition of the matrix inverse

$$\left| \left[\mathbf{G}^{-1} \right]_{n,m} \right| = \left| \det \left(\overline{\mathbf{G}}^{m,n} \right) \right| / \left| \det \left(\mathbf{G} \right) \right|, \tag{56}$$

where $\overline{\mathbf{G}}^{m,n}$ is the sub-matrix formed by removing row m and column n from $\mathbf{G}$. Now $\mathbf{G} \in \mathbb{A}^{(N)}$ so from Theorem 4

$$|\det(\mathbf{G})| \geq A_{\min}^{(N)}. \tag{57}$$

If $m = n$ then $\overline{\mathbf{G}}^{m,n} \in \mathbb{A}^{(N-1)}$ and from Theorem 4

$$\left| \det(\overline{\mathbf{G}}^{m,m}) \right| \leq A_{\max}^{(N-1)},\ \forall m. \tag{58}$$

If $m \neq n$ then $\Pi_n^T \overline{\mathbf{G}}^{m,n} \Pi_m \in \mathbb{B}^{(N-1)}$ and from Theorem 4

$$\left| \det \left(\overline{\mathbf{G}}^{m,n} \right) \right| = \left| \det \left(\Pi_n^T \overline{\mathbf{G}}^{m,n} \Pi_m \right) \right| \leq B_{\max}^{(N-1)},\ \forall m \neq n. \tag{59}$$

Combining (56), (57), (58) and (59) yields (55), which concludes the proof. ■

References

[1] G. Ginis and J. Cioffi, "Vectored Transmission for Digital Subscriber Line Systems," *IEEE J. Select. Areas Commun.*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[2] M. Tsatsanis, *DSL Handbook*. Auerbach, 2004, ch. 13 - Multi-line DSL System Architectures.

[3] M. Honig, P. Crespo, and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 614–629, Apr 1992.

[4] G. Taubock and W. Henkel, "MIMO Systems in the Subscriber-Line Network," in *Proc. of the 5th Int. OFDM-Workshop*, 2000, pp. 18.1–18.3.

[5] R. Cendrillon, G. Ginis, M. Moonen, and E. Van den Bogaert, "A Near-Optimal Linear Crosstalk Canceler for VDSL," submitted to *IEEE Trans. Signal Processing*, 2004.

[6] C. Zeng, C. Aldana, A. Salvekar, and J. Cioffi, "Crosstalk identification in xDSL systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1488–1496, Aug 2001.

[7] S. Galli, C. Valenti, and K. Kerpez, "A frequency-domain approach to crosstalk identification in xDSL systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1497–1506, Aug 2001.

[8] *Very high speed Digital Subscriber Line (VDSL); Functional Requirements*, ETSI Std. TS 101 270-1, Rev. V.1.3.1, 2003.

[9] T. Starr, *VDSL2 Proposed Profiles*, ANSI Std. Contrib. T1E1.4/2003-623, 2004.

[10] T. Cover, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.

[11] T. Starr, J. Cioffi, and P. Silverman, *Understanding Digital Subscriber Line Technology*. Prentice Hall, 1999.

[12] W. Yu and J. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, September 2004.

[13] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian MIMO Broadcast channel," submitted to *IEEE Trans. Inform. Theory*, 2004.

[14] G. Foschini and M. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[15] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780, Aug 1972.

[16] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *IEEE Electron. Lett.*, vol. 7, pp. 138–139, Mar 1971.

[17] *Dynamic Spectrum Management*, ANSI Draft Std. T1E1.4/2003-018, Rev. 15, 2004.

[18] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna broadcast channel," in *Proc. Int. Symp. Inform. Theory (ISIT)*, 2002.

[19] A. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 211–219, Jan. 2001.

[20] B. Bertsekas, *Non-linear programming*. Athena Scientific, 1999.

[21] *Very-high bit-rate Digital Subscriber Lines (VDSL) Metallic Interface*, ANSI Std. T1.424, 2004.