Analysis of Inverse Crosstalk Channel Estimation
Using SNR Feedback

P. Whiting, G. Kramer, Fellow, IEEE, C. Nuzman, Member, IEEE, A. Ashikhmin, Senior Member, IEEE,
A.J. van Wijngaarden, Senior Member, IEEE, and M. Živković

Abstract—Digital subscriber line (DSL) data rates for short loops are typically limited by crosstalk between adjacent lines rather than by background noise. Precoding can reduce crosstalk in the downstream from the access node to the customer premises equipment significantly if an accurate estimate of the inverse crosstalk channel is provided. Recently, a backward-compatible method has been proposed for estimating downstream crosstalk channels using standardized signal-to-noise ratio (SNR) reports. This paper develops a probabilistic model of the estimation process and, within this model, provides conditions under which successive updates of the precoder are guaranteed to converge to the ideal inverse precoder. Bounds on estimator variance and convergence times are obtained and optimized with respect to system parameters. The analysis can be applied to the situation in which a new line is being activated and added to a group of precoded lines seamlessly, that is, with controlled impact on the SNR of the active lines. Two phases are proposed to achieve seamless activation; the protection phase is used to let the active lines learn the crosstalk from the activating line and the acquisition phase is used to let the activating line learn the crosstalk from the active lines. Results of the analysis are illustrated by numerical simulations.

Index Terms—Crosstalk, digital subscriber line, perturbation methods, channel estimation, error compensation, spectrum management, vectoring, stochastic optimal control, convergence.

I. INTRODUCTION

DIGITAL subscriber line (DSL) networks are currently being transformed to better support the deployment of triple-play services, which in particular concentrate on high-definition video services as well as data and voice. Operators are migrating to very high speed digital subscriber line (VDSL) technology which offers significantly higher rates for short distances. The performance of such VDSL systems is typically limited by crosstalk from other active lines in the same group of cables or a binder. Crosstalk is typically dominant at distances below 1500 meters and may be orders of magnitude larger than any other type of impairment in a DSL system. Crosstalk usually varies significantly across tones within the binder, as well as from binder to binder.

A number of approaches to mitigate the impact of crosstalk have been recently proposed [1]–[8]. A promising approach to suppress crosstalk uses precoding [3]. The precoder adds a controlled amount of the signals transmitted over interfering DSL lines (disturbers) to the victim line, canceling the crosstalk and so increasing the victim line’s signal-to-noise ratio (SNR) and transmission rate. Results in [8] for example show that significant increases in rate are achieved in this way. However, the performance is strongly dependent on the accuracy of the crosstalk estimates made available to the precoder.

Several crosstalk estimation algorithms for DSL channels have been proposed. Practical considerations dictate that the algorithms for estimating downstream crosstalk channels must be run at the access node (operator side), using some form of feedback provided by the customer premise equipment (CPE) modems. Most crosstalk estimation proposals require new CPE modems capable of measuring and feeding back quantized versions of the received signals or receiver error signals [9], [10]. The use of receiver error feedback for crosstalk estimation forms the basis of a newly approved ITU recommendation, referred to as G.vector or G.993.5 [17]. The error-feedback approach can be fast and effective, but does not provide a way to estimate the crosstalk into DSL lines terminated by currently deployed CPEs. This limitation may hinder the introduction of crosstalk cancellation solutions within the DSL access network.

An alternative approach for crosstalk estimation uses a feedback mechanism already available in CPE modems today [12]–[16]. In particular, current standards specify that an access node can periodically request SNR measurements from CPE modems [13], [16]. The SNR feedback methods of [14], [15], described in an appendix to G.vector [17], learn the crosstalk channel by perturbing the precoder coefficients and observing the resulting changes in SNR. The precoder is updated after each estimation step, to drive the residual crosstalk channel to zero. A key advantage of these methods is that they work with “off-the-shelf” CPEs. As indicated in [14], [15], the accuracy of the estimates produced by SNR feedback methods can depend critically on implementation choices. The present work extends and refines the analysis sketched out in [14] with the goal of rigorously characterizing the performance of SNR feedback estimation as a function of system parameters and algorithm settings. The analysis allows us to answer questions such as “Under what conditions will residual crosstalk levels converge to zero within prescribed tolerances?” and “How long will it take the algorithm to converge?” and “How can the impact of the method on neighboring lines be controlled?” The stochastic analysis of convergence properties provided is complementary to the deterministic analysis found in [15].

This paper is organized as follows. Section II describes
models for precoding with estimation errors and for the statistics of the SNR feedback. A tuning-precoder is presented that comprises a difference estimator and a precoder. The difference estimator takes as input SNR reports that are fed back from the CPEs and outputs estimates of the current residual crosstalk coefficients. These estimates are computed from changes in the SNR reports which result when the residual crosstalk coefficients themselves are perturbed. It should be noted that the accuracy of the residual crosstalk estimates increases the more we reduce interference using the precoder. This fact motivates an iterative procedure for updating precoder coefficients where we exchange a process of measurement for one of stochastic control.

In Section III, we use the theoretical constructions to derive a bound on the distribution of mean residual crosstalk interference following an update to the precoder matrix, in terms of the mean crosstalk interference prior to the update. In this way we can identify sufficient conditions for convergence of the crosstalk interference to below a target level. More specifically we identify bounds for the mean rate of convergence as well as the tail probability of the time to converge. Related analysis also shows that the precoder coefficients pertaining to a joining line can be learned whilst controlling its impact on other lines by a combination of estimation and gradually raising the power.

Section IV applies the analysis to the operational problem of updating the precoder to accommodate activation of a new DSL line. The activation occurs in two phases. The purpose of the first phase (the protection phase) is to determine precoder coefficients that will protect active lines from crosstalk from the activating line, so that the presence of the new line does not degrade their (already high) data rates. The second activation phase (the acquisition phase) is focused on learning the precoder coefficients needed to suppress crosstalk from the active lines into the new line, in order to increase the attainable data rate on the new line. These two phases are analogous to the “O-P-VECTOR-1” and “O-P-VECTOR-2” initialization phases defined in the G.vector recommendation [17].

In Section V, the performance of the estimation procedures is evaluated by simulation and compared with analytical predictions. The analysis and simulations confirm that residual crosstalk can be driven to below a target threshold using a small number of SNR reports, while carefully controlling the system impact. The parameters of the algorithm must be correctly chosen in order to meet these goals simultaneously. Finally, in Section VI we present our conclusions.

II. SYSTEM MODEL

Consider a DSL system with \( L \) active lines in a bundle or cable which uses inverse precoding [2], [3] to suppress crosstalk. The overall channel for a given tone, with precoding, is given by

\[
\tilde{y} = HCx + \tilde{u},
\]

where \( H \) is the \( L \times L \) channel matrix, \( C \) is the \( L \times L \) precoding matrix, and where \( x, \tilde{y}, \) and \( \tilde{u} \) are the channel input vector, channel output vector, and noise vector, respectively. The entries of the matrices and vectors are complex numbers; the entries of \( H \) are denoted \( h_{v,k}, 1 \leq v, k \leq L, \) and similarly for other matrices; the entries of \( x \) are \( x_k, 1 \leq k \leq L, \) and similarly for other vectors; \( \Re(x) \) and \( \Im(x) \) denote the real and imaginary parts of \( x, \) respectively; we write \( j = \sqrt{-1}. \)

The transmit powers across lines are collected in the vector \( P \) with \( P_k = E[|x_k|^2] \), where \( E[\cdot] \) denotes expectation. We model \( u \) as a complex, zero mean Gaussian random variable, with power \( V[u_k] = N_k, \) where \( V[\cdot] \) denotes variance, and with the real and imaginary parts of \( u_k \) being independent and identically distributed (i.i.d.). We write \( P[E|F] \) for the probability of event \( E \) conditioned on event \( F. \)

It is convenient to normalize for channel attenuation with

\[
H = \text{diag}(H)(I + G), \tag{2}
\]

where \( \text{diag}(H) \) is the diagonal matrix of direct gains \( h_{v,v}, 1 \leq v \leq L, \) \( I \) is the identity matrix, and \( G \) is the relative crosstalk coefficient matrix, which has zeros on the main diagonal. In typical DSL environments, each diagonal element of \( H \) dominates the off-diagonal elements in a given row, i.e., \( h_{v,k} \ll h_{v,v} \) for \( k \neq v, \) so the magnitudes of the coefficients of \( G \) satisfy \( |g_{v,k}| = |h_{v,k}/h_{v,v}| \ll 1 \) for \( k \neq v. \) Standardized DSL crosstalk models predict that relative crosstalk values are approximately proportional to frequency and to the square root of line length (see [11] and the references therein). For example, at line lengths of less than 1 km and frequencies below 17 MHz (where crosstalk cancellation would typically be applied in practice), these models predict that 99% of normalized crosstalk coefficient magnitudes are less than 0.09.

The normalized received vector is

\[
y = \text{diag}(H)^{-1}\tilde{y} = (I + G)Cx + \tilde{u} = Rx + u, \tag{3}
\]

where \( R = (I + G) \) is called the resultant matrix and \( u = \text{diag}(H)^{-1}\tilde{u} \) is called the relative noise vector. We denote the relative noise power on line \( k \) by \( V[u_k] = N_k. \)

A. Controlling the Residual Matrix

Our goal is eventually to determine an ideal inverse precoder \( C = (I + G)^{-1}, \) such that the resultant matrix is identity, \( R = I. \) Equivalently, we want to drive the residual matrix \( \Theta = R - I \) to zero. When this is achieved, the precoder implicitly provides an estimate for the crosstalk channel via the expression \( \tilde{G} = C^{-1} - I. \)

The normalized receiver error signal can be expressed by

\[
e = y - x = \Theta x + u.
\]

For a given line \( v, \) a key performance measurement obtained via standardized mechanisms is an SNR estimate \( \gamma_v = P_v/Z_v, \) where

\[
Z_v := E[|e_v|^2] = \sum_k |\theta_{v,k}|^2 P_k + N_v. \tag{5}
\]

Since the transmit power \( P_v \) is known precisely, an estimate of \( \gamma_v \) is equivalent to an estimate of \( Z_v. \)

In order to drive the residual matrix \( \Theta \) to zero, we must manipulate the precoder matrix \( C, \) without a priori knowledge of the channel matrix \( H. \) A basic operation needed is to add
a complex value \( w \) to the residual channel from line \( k \) to line \( v \), obtaining the new residual \( \hat{\theta}_{v,k} = \theta_{v,k} + w \). This is done by adding \( w \) to precoder element \( c_{v,k} \); in matrix notation \( \hat{C} = C + w e_v e_k^T \), where \( e_v \) represents column \( v \) of an identity matrix. The new residual is \( \hat{\Theta} = \Theta + w (I + G) e_v e_k^T \). In addition to the direct effect of the perturbation, \( \hat{\theta}_{v,k} = \theta_{v,k} + w \), there are, in addition, second order side effects in the way that line \( k \) crosstalks into all other lines. Namely, we have \( \hat{\theta}_{m,k} = \theta_{m,k} + w y_{m,v} \) for all \( m \neq v \). Because the second order effects do not affect the SNR of the victim line whose channel is being estimated, these effects do not play a role in the convergence analysis. In typical cases, \( G \) is small enough that the second order effects are negligible. In extreme cases, it might be necessary to limit the perturbation magnitude \( |w| \) to avoid unwanted SNR fluctuations on neighboring lines.

**B. Inverse Crosstalk Channel Estimation by Precoder Tuning**

We use the term “tuning precoder” to refer to a system in which a precoder works in concert with SNR feedback and a difference estimator, as described in [14], [15]. We briefly describe this system to set up the analysis that follows.

The interaction of the tuning precoder with the CPEs as it would operate for a single tone is shown in Fig. 1. The precoder combines the signals from the lines according to the current precoder setting \( C \); the residual \( \Theta \) determines the error power \( Z_v \), which is reported back to the difference estimator via the SNR report \( \gamma_v \). The difference estimator sends commands to modify the precoder coefficients.

The precoder is updated in order to (i) reduce the residual after an estimation step, or (ii) perturb the current residual to gain data for a new estimate. The goal of the difference estimator at step \( n \) is to estimate the complex residual matrix \( \Theta[n] \) and then eliminate the estimated component, as in

\[
\Theta^{[n+1]} = \Theta^{[n]} - \hat{\Theta}^{[n]},
\]

where the error estimate \( \hat{\Theta}^{[n]} \) is computed as described below.

After the new error estimate \( \hat{\Theta}^{[n]} \) has been obtained, the estimate of the inverse precoder is updated as \( \hat{C}^{[n+1]} = \hat{C}^{[n]} - \hat{\Theta}^{[n+1]} \). Note that the rows of the precoder in (6) can be updated independently of each other.

The difference estimator works by examining the change in error power that results from a change in the current residual \( \Theta[n] \), as described above. We will use complex perturbations with magnitude \( \Delta \) and phase angle \( \phi \). After making a perturbation \( w = \Delta e^{-j\phi} \) to the precoder coefficient \( c_{v,k} \), the unknown residual \( \theta_{v,k} \) changes to \( \theta_{v,k} + w \). Denote by \( Z_v(\Delta e^{-j\phi}, k) \) the error power observed in this case, and note that the difference in error power

\[
Z_v(\Delta e^{-j\phi}, k) - Z_v = |\theta_{v,k} + \Delta e^{-j\phi}|^2 P_k - |\theta_{v,k}|^2 P_k = \Delta^2 P_k + 2\Delta \text{Re}(\theta_{v,k} e^{-j\phi}) P_k
\]

is a simple function of the unknown value \( \theta_{v,k} \) and known quantities \( \phi \), \( \Delta \), and \( P_k \).

Given unbiased estimates \( \hat{Z} \) of the error power \( Z \), we obtain that

\[
d_{v,k}(\Delta, \phi) := \frac{Z_v(\Delta e^{-j\phi}, k) - Z_v - \Delta^2 P_k}{2\Delta P_k}
\]

is an unbiased estimate of \( \text{Re}(\theta_{v,k} e^{-j\phi}) \). In a two-dimensional representation of the complex plane, this corresponds to the projection of \( \theta_{v,k} \) onto the ray through the origin at angle \( \phi \) from the x-axis (real axis). Similarly, one can obtain, for example, the projection in the direction \( \phi = \pi/2 \).

It follows that

\[
\hat{\theta}_{v,k}(\Delta, \phi) := (d_{v,k}(\Delta, \phi) + j d_{v,k}(\Delta, \phi + \pi/2)) e^{j\phi}
\]

is an unbiased estimate of the complex residual \( \theta_{v,k} \).

**A1.** \( Z_v \) is obtained by averaging \( T \) independent realizations of the normalized error \( |e_v|^2 = |y_v - \hat{x}_v|^2 \), where \( \hat{x}_v \) denotes the receiver’s estimate of the transmitted value \( x_v \).

**A2.** The DSL line is operating at a low error rate, such that \( \hat{x}_v = x_v \) with high probability.

**A3.** Estimates obtained at different times are based on different realizations of \( |e_v|^2 \).

**A4.** The number of realizations \( T \) is large enough so that \( \hat{Z}_v \) is well approximated by a Gaussian distribution, with mean \( Z_v \) and variance \( \sigma^2 \).

**A5.** The reported SNR level \( \gamma_v \) is obtained from the normalized error magnitude estimate as \( \gamma_v = 1/\hat{Z}_v \)."
can be uniformly bounded. This is a key to the convergence analysis, as it provides a way to bound the estimation error as a function of the current error power \(Z_v\).

**Theorem 1**: Suppose that the transmitted signals \(x_i\) are complex, zero mean, random variables, each with variance \(P_i\), that there is \(\kappa \geq 1\) such that
\[
\mathbb{V}[|x_k|^2] \leq \kappa P_k^2,
\]
for all \(k\)
and that the real and imaginary parts of \(x_i\) are uncorrelated, with equal variance. Suppose further that the noise \(u_v\) is complex Gaussian with variance \(N_v\) and having i.i.d. real and imaginary components. Then we have
\[
\sigma_v^2 \leq \kappa Z_v^2.
\]

**Proof**: See Appendix A.

The conditions that we require on the signals \(x_k\) in this result are satisfied for typical discrete multitone (DMT) systems (e.g., where \(x_k\) is chosen uniformly over a square grid). For such signals, the value \(\kappa = 1\) is typically sufficient. Moreover, when the error signal is approximately Gaussian (as when consisting of numerous commensurate components), then the bound \(\kappa = 1\) is tight.

The variance of \(d_{v,k}\) is given by
\[
\mathbb{V}[d_{v,k}(\Delta, \phi)] = \frac{\mathbb{V}[\hat{Z}_v(\Delta e^{-j\phi}, k)] + \mathbb{V}[\hat{Z}_v]}{4\Delta^2 P_k^2} \leq \frac{\kappa}{4MT\Delta^2 P_k^2} \left[Z_v(\Delta e^{-j\phi}, k)^2 + Z_v^2\right],
\]
where we have used Theorem 1 and the fact that \(\hat{Z}_v\) is based on averaging \(MT\) independent samples. The variance of the residual crosstalk estimate is then
\[
\mathbb{V}[\theta_{v,k}] = \mathbb{V}[d_{v,k}(\Delta, \phi)] + \mathbb{V}[d_{v,k}(\Delta, \phi + \pi/2)].
\]

From these expressions, it is clear that the accuracy of the crosstalk estimate improves as the actual error power \(Z_v\) goes down. Thus there is an incentive to quickly update the precoder with the aim of making further estimation more accurate. On the other hand, if the precoder is updated too quickly, with estimates of poor quality, further estimation becomes less accurate. The principles of stochastic control can be used to guide the estimation process to converge as quickly and reliably as possible.

Examining (11), the difference estimator offers four parameters that can be used to control estimation performance and convergence. Each of these are briefly discussed below.

**Repetition factor** \(M\). This factor provides a means to trade-off the conflicting goals of (a) obtaining crosstalk estimates using the fewest measurements possible, and (b) ensuring that crosstalk estimates are accurate before using them. The convergence analysis of Section III shows how to optimize bounds on the rate of convergence in terms of \(M\).

**Power** \(P_k\). If the disturbing line is involved in an active DSL session, the transmit power is fixed and cannot be used as a control parameter. However, for a line that is preparing to initiate a new DSL session, it may be possible to send signals with variable power. Such a scenario is discussed in Section IV-B.

**Perturbation angle** \(\phi\). Throughout this paper, we advocate and analyze using uniformly random perturbation angles. Analytically, this simplifies the analysis and practically it improves predictability of the distribution of the interference following a crosstalk estimation step.

**Perturbation magnitude** \(\Delta\). The perturbation magnitude is important for two reasons. First, it directly impacts the SNR of the victim line during the perturbation, and secondly, it affects the accuracy of estimation via (11).

It may easily be verified that the direct impact of perturbation on SNR is bounded by the expression \(Z_v(\Delta e^{-j\phi}, k) \leq (\sqrt{Z_v} + \Delta P_k)^2\). Thus, to guarantee that the multiplicative increase in interference due to perturbation, namely \(Z_v(\Delta e^{-j\phi}, k)/Z_v\), does not exceed \(F\), it suffices to take
\[
\Delta \leq \left(\sqrt{F} - 1\right) \frac{\sqrt{Z_v}}{\sqrt{P_k}}.
\]

For example, using \(F = 2\) guarantees that the SNR is not reduced by more than 3 dB during perturbation.

The bound (13) suggests the desirability of using adaptive perturbations that decrease in proportion to the square root of the interference in successive iterations. We will see in later sections that such an adaptive scheme is optimal for maximizing lower bounds on the rate of convergence. If perturbations are too small, their effect is lost in measurement noise; and if they are too big, their direct impact introduces additional measurement noise. The choice \(\Delta \sim \sqrt{Z_v}\) finds the happy middle ground.

### III. Convergence Analysis

As discussed previously, efficient crosstalk estimation via SNR feedback requires multiple estimation iterations. Because measurement noise is unavoidable, the sequence of interference levels observed after each iteration is a random process that must be controlled statistically. Denote by \(\xi_{v,k}^{[n]} = \beta \xi_{v,k}^{[n]} P_k/N_v\) the normalized interference from line \(k\) into line \(v\) in iteration \(n\). The key questions are:

- **Will the interference process converge to below a given target level** \(\xi_{v,k}^{[n]} \leq \beta?\) **What parameter settings are required to achieve convergence?**
- **How quickly will the process converge?**
- **What is the risk of observing an unacceptably high interference level** \(\xi_{v,k}^{[n]}\) along the way?**

This section provides analytical tools needed to answer these questions. We first formulate the desired interference process as a geometric super-martingale, resulting in sufficient conditions for convergence and a bound for the expected number of iterations required. This analysis shows that the key object to model is the distribution of the new interference \(\xi_{v,k}^{[n+1]}\), conditioned on the previous interference \(\xi_{v,k}^{[n]}\). The remaining sections give explicit bounds on this and related distributions in terms of problem data. The analytical results are applied to specific application scenarios in Section IV.
A. Convergence Criterion

The task of the tuning precoder is to ensure that the underlying interference power $\xi_v^{[n]}$ from each disturber rapidly becomes and then stays small. To analyze $\xi_v^{[n]}$, we employ the theory of martingales [18]. We show that the interference level in successive iterations, when updated by properly configured estimation steps, becomes an almost geometric super-martingale, as defined below. This allows us to show that the interference will converge to the target level, and allows us to bound the mean convergence time.

Definition 1: Let $V_n$ be a sequence of bounded non-negative random variables, that is $0 \leq V_n \leq V$ for some real $V$. Let $V_n$ be adapted with respect to a filtration $\{\mathcal{F}_n\}$. The sequence $V_n$ is said to be a geometric super-martingale provided there is a constant $\alpha$, where $0 < \alpha < 1$, such that

$$E[V_n|\mathcal{F}_{n-1}] \leq \alpha V_{n-1}. \quad (14)$$

Loosely speaking, (14) means that the next value of the process, conditioned on its entire past history, is less than the current value by at least a constant factor. For practical purposes we require only that (14) holds for $V_{n-1} > \beta > 0$, and we refer to a sequence satisfying this condition as a $\beta$-almost geometric super-martingale ($\beta$-ags). A $\beta$-ags is said to have converged if there has been a first entrance $V_n \leq \beta$. The following result establishes that there will be such a first entrance, almost surely.

Theorem 2: Let $V_n$ be a $\beta$-ags with mean initial value $E[V_0] = \bar{V}$, and let $T_\beta$ be the first entrance time into $[0,\beta]$. Then almost surely $T_\beta < \infty$ and

$$E[T_\beta] \leq \frac{\log(V_0/\beta)}{\log(1/\alpha)} + \frac{1}{1-\alpha} < \infty. \quad (15)$$

Proof: See Appendix B.

This result shows that, unless $\alpha$ is close to one, the mean convergence time will not be much larger than the number of steps that would be required to deterministically decrease from $V_0$ to $\beta$ geometrically in steps of $\alpha$.

In the next section, we will show that the crosstalk estimation process can be designed to be a $\beta$-ags, where the convergence parameter $\alpha$ is a function of algorithmic parameters.

B. Normalized Notation

Before proceeding it is useful to introduce several notational conventions. As mentioned previously, the normalized interference from disturber $k$ is defined to be

$$\xi_v^{[n]} := \frac{\theta_v^{[n]}P_k}{N_v}. \quad (16)$$

The total interference is denoted by

$$\xi_v := \sum_k \xi_v^{[n]} \quad (17)$$

A typical goal is to drive $\xi_v$ down to below $\beta$. The normalized unperturbed error power is denoted by

$$z_v := \frac{Z_v}{N_v} = 1 + \xi_v. \quad (18)$$

From (7), the normalized perturbed error power is given by

$$\frac{Z_v(\Delta e^{-j\phi_v^{[n]}k})}{N_v} = z_v + \eta_v^{[n]} + 2\sqrt{\eta_v^{[n]}P_k}\xi_v^{[n]}\cos(\psi - \phi), \quad (19)$$

where $\psi$ is the phase angle of $\theta_v^{[n]}$, and where

$$\eta_v^{[n]} := \frac{\Delta^2 P_k}{N_v}. \quad (20)$$

is the normalized perturbation magnitude for the coefficient from line $k$ to line $v$. Using (11) together with the coefficient of variation bound (10), we obtain a bound on the conditional variance of the difference estimator

$$V[d_v^{[n]}(\Delta, \phi)|\phi, \psi] \leq \frac{N_v}{P_k}v(z_v, \eta_v^{[n]}, \xi_v^{[n]}, \psi - \phi), \quad (21)$$

where

$$V(z, \eta, \xi, \omega) := \frac{\kappa}{4MT\eta} \left[z^2 + \left(z + \eta + 2\sqrt{\eta \xi \omega}\right)^2 \right]. \quad (22)$$

This result provides a starting point from which to compute bounds on the distribution of the interference $\xi_v^{[n+1]}$ after the $n$-th iteration, relative to the interference at the start of that step. We first consider expectations, then tail probabilities.

C. Bounds on Expected Interference

If we choose the perturbation angle $\phi$ uniformly at random in each step, then $\psi - \phi$ is also uniform. Taking the expectation over (21) gives

$$V[d_v^{[n]}(\Delta, \phi)] \leq \frac{\kappa N_v}{4MT\eta_v^{[n]}P_k} \left[z_v^2 + (z_v + \eta_v^{[n]})^2 + 2\eta_v^{[n]}\xi_v^{[n]} \right] = \frac{\kappa N_v}{2MT P_k} \left[z_v^2 + \xi_v^{[n]}\eta_v^{[n]} + (z_v + \xi_v^{[n]}P_k)^2 \right] \quad (23)$$

The right-hand side of the last expression is a convex function of the perturbation size $\xi_v^{[n]}$, which is minimized by the choice $\xi_v^{[n]} = \sqrt{z_v}$. This means that optimal perturbation size, from the point of view of minimizing the bound on expected estimation error, is proportional to the current squared error signal (i.e., that $\Delta^2$ should be proportional to $Z_v$).

In what follows, we thus take adaptive perturbations of the form $\eta_v^{[n]} = \lambda z_v$. Then the mean interference after the $n$-th estimation step is

$$\mathbb{E}[\xi_v^{[n+1]}] := \mathbb{E}[\theta_v^{[n+1]}|\xi_v^{[n]}]$$

$$= \frac{\mathbb{E}[[\theta_v^{[n+1]}^2]P_k^{[n+1]}]}{N_v}$$

$$\leq \frac{P_k^{[n+1]}\kappa}{P_k^{[n+1]}\lambda} \left[\frac{1}{\lambda} + \frac{\lambda}{2}\right] z_v^{[n]} + \xi_v^{[n]} \right]. \quad (23)$$

From the bound (13) and the definition of $\eta$, it follows that the SNR impact of perturbations will not exceed a factor $F$ as long as the constant of proportionality satisfies $\lambda \leq (\sqrt{F} - 1)^2$. It is desirable to take $\lambda$ as close to $\sqrt{2}$ value is possible, subject to the impact constraint. For example, we can use $\lambda = 0.17$ and $\lambda = 1$ if the impact factors are $F = 2$ or $F = 4$, respectively.
respectively. Note that we attempt to control the impact too tightly using small values of λ, the bound on the expected interference blows up.

Suppose that in one iteration of a control algorithm the difference estimator is used to estimate and reduce all L residual components affecting a given victim line, and that the power levels are constant, i.e., \( P_k^{n+1} = P_k^n \). Then we obtain

\[
E[\xi_v^{n+1}] = \sum_{k \neq v} E[\xi_{v,k}^{n+1}] \leq \alpha \left( M, \lambda, \xi_v^n \right) \xi_v^n, \tag{24}
\]

where

\[
\alpha(M, \lambda, \xi) = \frac{\kappa}{MT} \left[ \frac{1}{\lambda} + 1 + \frac{\lambda^2}{2} \right] \left( 1 + \frac{1}{\xi} \right) L + 1, \tag{25}
\]

which is suggestive of a super-martingale. Indeed, since the function \( \alpha \) is a decreasing function of \( \xi \), \( \alpha(M, \lambda, \beta) \) is an upper bound on the \( \alpha \) in (14) whenever \( \xi > \beta \). We summarize our insights in the following theorem.

Theorem 3: The interference process \( \xi_v^n \) is a \( \beta \)-ags when using the difference estimator of Sec. II-B with adaptive perturbation size \( \Delta = \sqrt{\lambda} z_{v,k}/P_k \) and with \( \alpha(M, \lambda, \beta) < 1 \).

The expected time for first entrance into \([0, \beta]\) is given by (15) with \( V_0 = E[\xi_v^0] \) and \( \alpha(M, \lambda, \beta) \). Furthermore, the perturbation size that minimizes the right-hand side of (23) is given by choosing \( \lambda = \sqrt{2} \).

We term \( E[\xi_v^{n+1}]/\xi_v^n \) the expected improvement ratio.

If \( \beta = 1 \), meaning that the goal is to reduce crosstalk interference to the same level as background noise, then for typical values \( M = 1, T = 256, \kappa = 1, \lambda = 1 \), the bound (25) says that the interference process will converge for up to 50 disturbing lines. When there are 10 disturbing lines with these parameters, \( \alpha \) is approximately equal to 0.2.

D. Bounds on Interference Distribution

For some applications, we are concerned not only with the mean behavior of interference estimates, but on the tail probabilities. In other words, for a given interference level \( \xi_v^n \), we are interested in the distribution of the new interference level \( \xi_v^{n+1} \). Many different bounding expressions are possible, and in general there are trade-offs between the complexity of the analysis and the tightness of the bound.

Recall that, conditioned on an (unknown) residual value \( \theta_{v,k} = |\theta_{v,k}| e^{i\phi} \), the difference estimator \( d_{v,k}(\Delta, \phi) \) is Gaussian with variance bounded by (21). Since \( \theta_{v,k}^{n+1} = \theta_{v,k} - \theta_{v,k}^{n} \), it follows that the distribution of the normalized interference after update is bounded in distribution by the weighted sum

\[
\xi_{v,k}^{n+1} \leq \frac{P_k^{n+1}}{P_k^n} \left[ w_{v,k}(\psi - \phi) X_1 + w_{v,k}(\psi - \phi - \pi/2) X_2 \right], \tag{26}
\]

where

\[
w_{v,k}(\omega) = v(\xi_v^n, \xi_v^n, \xi_v^n, \xi_v^n, \omega),
\]

and where \( X_1 \) and \( X_2 \) are chi-squared variables with one degree of freedom.

When the same base measurement \( z_v \) is used to form the two difference estimators \( d_{v,k}(\Delta, \phi) \) and \( d_{v,k}(\Delta, \phi + \pi/2) \), these two estimates are not independent, which means that \( X_1 \) and \( X_2 \) are not independent in (26). On the other hand, these two variables are independent if each difference estimator uses an independently generated base measurement; operationally this requires four measurements per estimation step rather than three. For simplicity, we analyze the four measurement case in this section, taking \( X_1 \) and \( X_2 \) to be independent. In our experience, the performance with and without independent base measurements is similar.

1) Distribution of Total Interference Ratio \( \xi_v^{n+1}/\xi_v^n \): We sacrifice some tightness for simplicity. We take the worst case angle, \( \phi = \psi \) to obtain the bounds

\[
v(z, \eta, \xi, \omega) \leq v(z, \eta, \xi, 0) = \frac{\kappa}{4MT\eta} \left[ z^2 + \left( z - \xi + \left( \sqrt{\eta} + \sqrt{\xi} \right)^2 \right)^2 \right]\]

\[
\leq \frac{\kappa}{4MT\eta} \left[ z^2 + (z - \xi + (1 + \eta/z) (z + \xi))^2 \right]
\]

\[
= \frac{\kappa}{4MT\eta} \left[ 1 + \left( 2 + \frac{2}{z} \right) \left( 1 + \frac{\xi}{z} \right)^2 \right]
\]

\[
:= \tilde{v}(z, \eta, \xi),
\]

where we used the Cauchy-Schwarz inequality and \( \eta/z \geq 0 \) to obtain the second inequality, and further algebraic manipulations to obtain the final expression. From (26), we then have

\[
\xi_{v,k}^{n+1} \leq 2 \frac{P_k^{n+1}}{P_k^n} \tilde{v}(z_v^n, \xi_v^n, \xi_v^n, \xi_v^n) \frac{[X_1 + X_2]}{2}. \tag{27}
\]

Now using the bound \( \xi_{v,k} \leq \xi_v \), assuming that power levels \( P_k^n \) do not vary in the time index \( n \), i.e., that bit swap is disabled, and using adaptive perturbations of the form \( \theta_{v,k} = \lambda z_v^n \), we bound the total interference from \( L \) residual terms as

\[
\xi_v^{n+1} = \sum_k \xi_v^{n+1} \leq \tilde{v}(z_v^n, \lambda^2 z_v^n, \xi_v^n) X_S,
\]

where \( X_S \) is a chi-squared variable with \( 2L \) degrees of freedom. It follows that the ratio \( \xi_v^{n+1}/\xi_v^n \) is bounded in distribution by a gamma distribution with shape parameter \( L \) and with mean

\[
s = 2L \bar{\xi}(z_v^n, \lambda^2 z_v^n, \xi_v^n) / \xi_v^n \leq \frac{\kappa L}{2MT\xi_v^n} \left[ 1 + \left( 2 + \lambda \left( 1 + \frac{\xi_v^n}{z_v^n} \right) \right)^2 \right]
\]

\[
\leq \frac{\kappa L}{2MT} \left[ 1 + \frac{1}{\xi_v^n} \right] \left[ \frac{5}{\lambda} + 8 + 4\lambda \right],
\]

where the last step follows by using \( \xi_v^n \leq z_v^n \). As long as the interference has not yet reached the target level, \( \xi_v^n \geq \beta \), the mean is bounded by

\[
s \leq \frac{\kappa L}{2MT} \left( 1 + \beta^{-1} \right) \left[ \frac{5}{\lambda} + 8 + 4\lambda \right]. \tag{28}
\]
2) Refined Bound of Interference Component $\xi^{(n+1)}_{v,k}$: In the protection phase of the seamless joining procedure described in Section IV, we will be interested in bounding the tail probability of $\xi^{(n+1)}_{v,k}$, given that $\xi^{(n)}_{v,k} \leq \beta$. As described in that section, there is only a single disturber of interest, and so $\xi^{(n)}_{v,k} = \xi^{(n+1)}_{v,k}$. We choose to keep a constant normalized perturbation level $\eta_{v,k} = \eta$.

Observe from (27) that $\mathbb{E} (X_1^2 + X_2^2)/2$, being the average of two chi-square variables with one degree of freedom, is exponential with unit mean. This means that $\xi^{(n+1)}_{v,k}$ can be bounded by an exponential random variable. To obtain a tighter bound, we will need to consider the weighted sum of two chi-squared variables. The refined bound can be obtained in three steps. First, the function $\nu(\cdot)$ in (22) is maximized over the range of $\xi^{(n)}_{v,k}$ to obtain

$$\bar{\nu}(\beta, \phi) := \max_{0 \leq \xi \leq \beta} \nu (1 + \xi, \eta, \xi, \psi - \phi)$$

where the last expression follows from observing that $\nu (1 + \xi, \eta, \xi, \psi - \phi)$ is a convex function of $\sqrt{\xi}$. Next, we have

$$\mathbb{P} \left[ \xi^{(n+1)}_{v,k} > \beta | \xi^{(n)}_{v,k} \leq \beta, \phi \right] \leq \mathbb{P} \left[ Y(\beta, \phi) \frac{P_k^{(n+1)}}{P_k^{(n)}} \geq \beta \right],$$

where $Y$ is the weighted sum

$$Y(\beta, \phi) = \bar{\nu}(\beta, \phi) X_1 + \bar{\nu}(\beta, \phi + \pi/2) X_2$$

of independent chi-squared variables $X_1, X_2$. A power series for the tail probability of $Y$ is given in Appendix C. The final step is to average over the uniformly selected perturbation direction to obtain

$$\mathbb{P} \left[ \xi^{(n+1)}_{v,k} > \beta | \xi^{(n)}_{v,k} \leq \beta \right] \leq \frac{1}{2\pi} \int_0^{2\pi} \mathbb{P} \left[ Y \geq \beta P_k^{(n)} / P_k^{(n+1)} \right] d\phi$$

$$= f(\beta, r),$$

where $r = P_k^{(n)} / P_k^{(n+1)}$. As shown in Section IV, the power ratio in the right-hand side of inequality (29) can be used to control the risk that the interference at any stage will exceed a desired threshold $\beta$.

IV. ACTIVATION PHASES

We apply the results of Section III to a two-phase procedure for activating a new line, when several precoded lines are already active. First, the protection phase protects active lines from interference from the joining line, and secondly the acquisition phase acquires the precoder coefficients needed to suppress crosstalk from the active lines into the joining line. In the protection phase, the transmit power of the joining line is gradually increased as the precoder is adapted; the key question is how quickly can the line be brought to full power, while controlling the risk of affecting active lines. In the second phase, power levels are held fixed, and the key question is how quickly can the data rate of the new line be brought up to its maximum level. Because the analysis of the protection phase is more complex, we first analyze the acquisition phase.

A. Acquire Coefficients from Active Lines into a Joining Line

The goal of the acquisition phase is to reduce the interference from a group of $L - 1$ active lines into a joining line $v$ to an acceptable, predetermined level. As described in Section III, the normalized interference $\xi^{(n)}_v$ should satisfy $\xi^{(n)}_v \leq \beta$ after convergence. We will suppose that the powers are fixed, that we use adaptive perturbations, and that the perturbation angle $\phi$ is chosen uniformly at random. We will also suppose that the following information is available:

1) The powers $P_k$ for the joining and the disturbing lines;
2) The initial crosstalk estimate;
3) auxiliary parameters, $\kappa, T, M$.

In the absence of other information, zero is the most natural choice for initial crosstalk estimates. If more accurate estimates are available, then the convergence time will be reduced. A poor estimate would have the opposite effect.

The pseudo-code for the acquisition algorithm is as follows:

**Acquisition algorithm:**

**input** : victim line index $v$, disturber line index set $K$ (which may also include $v$), initial precoder matrix $C^{(0)}$.

**output** : improved precoder $C^{(n+1)}$.

**Step 1.** Set $n = 0$.

**Step 2.** Use $M$ SNR reports to estimate baseline interference $Z_v^{(n)}$.

**Step 3.** Select next disturber $k$ in $K$.

**Step 3.1** Set perturbation magnitude $\Delta_k^{(n)} = \sqrt{\lambda Z_v^{(n)}/P_k}$ and choose perturbation angle $\phi_k^{(n)}$ uniformly at random.

**Step 3.2** Use $M$ SNR reports to estimate each of the perturbed interference levels $Z_v(\Delta_k^{(n)} e^{-j\phi_k^{(n)}} k)$ and $Z_v(\Delta_k^{(n)} e^{-j\phi_k^{(n)}} k)$.

**Step 3.3** Estimate current residual $\hat{\theta}_{v,k}(\Delta_k^{(n)} e^{-j\phi_k^{(n)}})$ via (9).

**Step 3.4** If not all disturbers in $K$ have been perturbed, go to Step 3. Otherwise go to Step 4.

**Step 4.** Select next disturber $k$ in $K$.

**Step 4.1** Update precoder to obtain $c_{v,k}^{(n+1)} = c_{v,k}^{(n)} - \hat{\theta}_{v,k}^{(n)}$.

**Step 4.2** If not all disturbers in $K$ have been updated, go to Step 4. Otherwise go to Step 5.

**Step 5.** If all residual interference levels $\hat{\theta}_{v,k}^{(n)}$ are sufficiently small, or if $n$ is sufficiently large, stop. Otherwise, set $n \leftarrow n + 1$ and go to Step 2.

The steps above may be done on all tones in parallel for which SNR reports are available.

1) **Expected Convergence Time:** As described in Section III-C, the algorithm will converge almost surely if $\alpha(M, \lambda, \beta) < 1$. This condition can be checked in advance. The value for $\alpha$ can also be used to bound the mean time to converge via Theorem 2. For example, suppose that we wish to raise the SNR by 30 dB and have a target of $\beta = 1$. Substituting into the bound we find that $\mathbb{E}[T_\beta] \leq 4$, so that on average no more than four iteration steps are required. Figure 2 illustrates the relation between the bound on the mean convergence time and the number of disturbers.
2) Tail distribution of convergence time: It is also of interest to estimate or bound the probability that the procedure does not converge in a given number of iterations. Using (28) we may construct a random walk \( \{W_n\} \) which bounds \( \xi_K^n \) in distribution. Let \( W_0 = \xi_0 \) and
\[
W_{n+1} = F_n W_n, \quad n \geq 0,
\]
where \( F_n \) is a gamma random variable, written as \( F_n \sim \Gamma(L-1, \tau/(L-1)) \), with mean \( \tau \) given by the right side of (28). Note that the Gamma distribution \( \Gamma(a, b) \) has two parameters: \( a \) is the scale and \( b \) is the shape. The mean is \( ab \) and the variance is \( ab^2 \). The variable \( F_n \) provides a bound, in distribution, on the factor by which the interference level decreases after the \( n \)-th estimation step. It follows that \( \xi_K^n \) is bounded by \( W_n \) up to the step when \( \xi_K^n \) falls below \( \beta \), and that we can bound the probability that \( \xi_K^n \) converges by the probability that \( W_n \) converges.

Define \( Y_t := 10 \log_{10} F_t - 10 \log_{10} \alpha \), where \( \alpha \) is chosen such that \( Y_t \) has negative mean. Let \( D_K \) be the event that the sequence \( W_n \) is larger than \( \beta \) for all \( n \leq K \) and let
\[
A_K := \left\{ \omega : \frac{1}{K} \sum_{t=1}^{K} Y_t(\omega) > 10 \log_{10}(\beta/\xi_0) - 10 \log_{10} \alpha \right\}
\]
be the event that \( W_K \) is larger than \( \beta \). Clearly \( D_K \) is a subset of \( A_K \), and so \( R_K := P[A_K] \) is an upper bound on \( Q_K := P[D_K] \), which in turn is an upper bound on the probability that the interference \( \xi_K^n \) does not converge within \( K \) steps.

The probability \( R_K \) can be evaluated using the Chernoff bound for a sum of independent, identical random variables:
\[
Q_K \leq R_K \leq \exp \left( -\frac{a}{K} I(\beta/K - 10 \log_{10} \alpha) \right),
\]
where \( a = 10 \log_{10} \epsilon \), and where
\[
I(x) = \sup_{t>0} \left( tx - \log \frac{\Gamma(\nu+t)}{\Gamma(\nu)} \right)
\]
and \( \Gamma(x) \) is the Gamma function of \( x \).

Table I shows this Chernoff upper bound (computed numerically), as well as \( R_K, Q_K \) (estimated by Monte Carlo simulation) for the case \( 10 \log_{10} \alpha = -16.12 \, \text{dB} \) and a required gain of \( 30 \, \text{dB} \) in SNR. The results show that \( Q_K \) is very close to \( R_K \) and we may deduce that it is very unusual for the random walk to converge to below \( \beta \) and then later climb above this level. It follows that we can obtain tight bounds by sharpening the Chernoff bound on \( R_K \) using well known refined estimates [19].

In practice the SNR algorithm takes even fewer steps to converge than the bounding random walk does because the initial step of bounding of \( \xi_K^n \) by \( W_n \) is not tight. Further simulation results are given in Section V.

### Table I: Probability of Convergence Within \( K \) Steps

<table>
<thead>
<tr>
<th>( K )</th>
<th>( Q_K )</th>
<th>( R_K )</th>
<th>Chernoff Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.577</td>
<td>0.577</td>
<td>0.9977</td>
</tr>
<tr>
<td>5</td>
<td>2.18 \times 10^{-2}</td>
<td>2.18 \times 10^{-2}</td>
<td>0.1249</td>
</tr>
<tr>
<td>6</td>
<td>3 \times 10^{-5}</td>
<td>3 \times 10^{-5}</td>
<td>2.993 \times 10^{-4}</td>
</tr>
</tbody>
</table>

B. Protecting Active Lines from Joining Line Interference

The steps taken during the protection phase guard active lines from excessive crosstalk interference from the joining line. This phase takes place before the acquisition phase discussed above. The active lines are vulnerable to interference from the joining line because they are operating at high SNR with commensurate bit loadings. To avoid causing decoding errors, the (normalized) interference from the joining line must be kept below a controlled predetermined threshold \( \beta \) throughout the joining process. This can be accomplished by having the joining line start at some sufficiently low power \( P_k^0 \) and then gradually raising power as the required precoder coefficients are more and more accurately estimated [12]. Estimation of \( \theta_{e,k} \) can proceed in parallel on multiple victim lines \( v \), as long as the second-order side effects mentioned in Section II-A are negligible.

Consider in particular protecting a victim line \( v \) from crosstalk from a joining line \( k \). To make the method useful, the power should be ramped up quickly while controlling the risk of causing high interference. For one iteration, (29) gives a trade-off between power increase \( r = P_k^{(n+1)}/P_k^n \) and the risk of exceeding the interference threshold \( \beta \). In particular, given a probability limit \( \delta \), one can compute the maximum value of \( r \) such that \( f(\beta, r) \leq \delta \). It is then possible to increase the power by a factor of \( r^K \) in \( K \) steps, with the probability of exceeding the threshold at some stage being bounded by \( K\delta \). If desired, the simpler but looser bound (27) can be used instead of (29) to bound the maximal ratio \( r \).

In general, the desired overall power increase \( R = P_k^{[K]}/P_k^0 \) and an overall probability threshold \( \epsilon \) are given. In this case (29) can be used to find the smallest number of steps \( K \) such that \( f(\beta, R^{1/K}) \leq \epsilon/K \). If there is no such solution using \( M \) SNR reports per iteration, then the parameter \( M \) can be increased until a viable solution is found. Large values of \( \epsilon \) permit a more aggressive algorithm in which \( K \) and \( M \) can be smaller.

As before, the impact of the perturbation on the SNR is controlled by the normalized perturbation parameter \( \eta^{[n]}_{k} \). In the previous section, we found that making the perturbation scale adaptively with the total interference \( z_v \) is an efficient
choice. Since in the present case the total interference is to remain within a small interval \([1, 1 + \beta]\), it makes sense to use a fixed normalized perturbation \(\eta\). That is, the perturbation magnitude should be set as \((\Delta \theta_{v,k}^{[n]})^2 = \eta N_v / P_k^{[n]}\), which is inversely proportional to the joining line power, and proportional to the background noise level of the active line \(v\).

The pseudo-code for the protection algorithm is as follows:

**Protection algorithm:**

**input**: victim line index \(v\), joining line index \(k\), initial precoder matrix \(C^{[0]}\), safe initial power level \(P_k^{[0]}\), desired final power level \(P_k\), interference bound \(\beta\), and risk threshold \(\varepsilon\).

**output**: improved precoder \(C^{[K]}\) such that the residual interference satisfies \(\theta_{v,k}^2 P_k \leq \beta N_v\).

**Step 1.** Pre-compute the smallest number of steps \(K\) (and associated smallest repetition parameter \(M\)) such that the power increment \(r = (P_k/P_k^{[0]})^{1/K}\) has risk below \(\varepsilon/K\).

**Step 2.** Set the joining line transmission power to \(P_k^{[0]}\) and set time parameter \(n = 0\).

**Step 3.** Form the estimate \(\hat{\theta}_{v,k}^{[n]}\) using SNR difference estimation with perturbation magnitude \((\Delta \theta_{v,k}^{[n]})^2 = \eta N_v / P_k^{[n]}\) and random perturbation angle \(\phi\).

**Step 4.** Update precoder to obtain \(c_{v,k}^{[n+1]} \leftarrow c_{v,k}^{[n]} - \hat{\theta}_{v,k}^{[n]}\).

**Step 5.** Increment the power as \(P_k^{[n+1]} = r P_k^{[n]}\). If \(P_k^{[n+1]} \geq P_k\), stop. Otherwise set \(n \leftarrow n + 1\) and repeat from Step 3.

V. NUMERICAL RESULTS

This section contains further numerical illustration of the application of the analysis of Section III to the seamless joining procedures described in Section IV.

In the simulations, the SNR reports \(\gamma_v\) were generated as inverses of Gaussian error power measurements \(Z_v\), in keeping with Assumption A5. The mean level \(Z_v\) and variance \(\sigma_v^2 / T\) were calculated using Lemma 2 of Appendix A. In this calculation the random variables \(a_i\) were complex data signals scaled by the residual crosstalk magnitude. The complex data signals were chosen uniformly from a square centered at the origin of the complex plane. Relative crosstalk coupling coefficients had randomly chosen phases, and their magnitudes were selected to achieve a desired gap between performance with and without crosstalk cancellation.

A. Acquisition Phase

We first report on a set of experiments related to the acquisition phase of Section IV. In the experiments \(L = 5\), i.e., there are \(L - 1 = 4\) active lines and there is one joining line. The normalized background noise is \(N_v = 10^{-5}\) on each line, so that the SNR in the absence of crosstalk is 50 dB. We set the relative crosstalk \(g_{v}\) into the given victim line \(v\) such that the SNR without precoding was 15.97 dB. Rough initial estimates were used to determine precoder coefficients that raised the SNR to 18.10 dB. A total of nine SNR measurements were used per update: an unperturbed measurement, and then measurements for two perturbations on each of \(L_d = 4\) lines. The target SNR was 3 dB below ideal \((\beta = 1)\).

The magnitudes of the elements of the relative crosstalk channel \((I + G)\) are

\[
\begin{bmatrix}
1.0000 & 0.14142 & 0.07071 & 0.01414 & 0.00707 \\
0.01088 & 1.00000 & 0.1608 & 0.01489 & 0.01409 \\
0.03026 & 0.00888 & 1.00000 & 0.01241 & 0.03089 \\
0.00118 & 0.00504 & 0.02248 & 1.00000 & 0.00383 \\
0.00063 & 0.00294 & 0.00062 & 0.00423 & 1.00000
\end{bmatrix}
\]

The complex crosstalk coefficients \(g_{1,k}\) and initial precoder settings \(c_{1,k}\) into line 1, for disturbers \(k = 2, \ldots, 5\), are given by

\[
\begin{align*}
0.100 + 0.100j & & \text{and} & & -0.1753 - 0.0267j \\
-0.050 + 0.050j & & & & -0.0151 - 0.0598j \\
-0.010 - 0.010j & & & & 0.0163 + 0.0049j \\
0.005 - 0.005j & & & & -0.0101 + 0.0034j
\end{align*}
\]

respectively.

1) **Fixed Perturbations:** To emphasize the importance of using adaptive perturbations, we first performed simulations using fixed perturbations. That is, we used the algorithm described in Section IV-A, except that the same perturbation magnitude \(\Delta\) was used in each iteration. Illustrative results for the fixed case are shown in Table II. The first column gives the magnitude of the perturbation normalized by the square root of the noise. The second column is the number of SNR measurements per SNR estimate \(M\), and the third column is the time required, in a single trial, to bring the normalized interference below \(\beta = 1\), assuming that each SNR report requires 11 seconds to obtain. The values of \(M\) were chosen empirically to roughly minimize convergence time. The examples in Table II illustrate that sufficiently large values of \(M\) can ensure convergence with fixed perturbations, but that the convergence time is very long.

2) **Adaptive Perturbations:** We next report on a set of experiments done using adaptive perturbations with magnitude \(\Delta_t = \lambda Z_v / P_k\). Figure 3 shows the progression of the SNR of the joining line, as the precoder is updated with successively better estimates of the inverse channel, using the acquisition algorithm presented in Section IV-A. The base SNR measured in Step 2 of each successive iteration is depicted. The five vertical bars associated with each iteration are obtained using different values of the adaptive perturbation parameter \(\lambda\). For all cases, the initial SNR in iteration zero was approximately 18 dB. Values of \(\lambda\) near 1 converge quickly, while much lower and much higher values of \(\lambda\) take longer to converge, as was anticipated in Section IV.

<table>
<thead>
<tr>
<th>(\Delta / \sqrt{N_v})</th>
<th>(M)</th>
<th>convergence time [hours]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>20</td>
<td>3.9</td>
</tr>
<tr>
<td>0.7</td>
<td>45</td>
<td>6.7</td>
</tr>
<tr>
<td>0.4</td>
<td>85</td>
<td>48.3</td>
</tr>
</tbody>
</table>
Although Figure 3 does not show the SNR values observed during perturbation measurements, it is worth noting that very large values of $\lambda$ cause excessive SNR fluctuations during perturbation, in addition to slow convergence. For example, the choices $\lambda = 16$ and $\lambda = 36$ cause perturbation measurements to be about 14 dB and 17 dB below base measurements.

![SNR vs Number of Iterations](image)

**Fig. 3.** SNR of a joining line with successive iterations of the acquisition algorithm. For a given iteration number, the five vertical bars represent performance when using one of five different values of the adaptive perturbation parameter $\lambda$. From left to right, the values are $\lambda = \{0.04, 0.25, 1.0, 16, 36\}$.

To examine these results in more detail, repeated trials were performed to estimate the probability $P_S$ that the SNR level would come within 3 dB of ideal after at most five iterations of the acquisition algorithm.

![SNR vs Number of SNR Reports](image)

**Fig. 4.** Reported SNR for a joining line and an active line during several iterations of the acquisition phase.

![Mean Iterations and Bound on Gain and Simulated Mean](image)

**Table IV**

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$\alpha$ (dB)</th>
<th>Actual mean gain (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1585</td>
<td>-13.67</td>
<td>-15.00</td>
</tr>
<tr>
<td>328</td>
<td>-13.66</td>
<td>-15.26</td>
</tr>
<tr>
<td>76.6</td>
<td>-13.62</td>
<td>-15.19</td>
</tr>
<tr>
<td>1.07</td>
<td>-10.99</td>
<td>-12.33</td>
</tr>
</tbody>
</table>

Table III shows the estimated probability of success for several values of $\lambda$ along with the average number of steps required to converge. For comparison, the upper bound on the mean number of steps to converge, based on Theorem 2 in conjunction with (25), is also given. In all cases 100,000 trials were used.

The table shows that the likelihood of failure and the expected number of steps required grows as $\lambda$ moves away from the value $\sqrt{2}$ that minimizes the bound. This is consistent with the results given in Section IV.

As a general remark, $M = 1$ is typically optimal when $\beta$ is relatively large, as might be the case when initially acquiring crosstalk coefficients for a joining line. On the other hand, larger values of $M$ may become useful when the goal is to obtain $\beta \ll 1$, as might be the case when tracking slow changes to crosstalk coefficients in “showtime”.

Figure 4 shows a typical, complete set of SNR measurements using adaptive perturbation with $\lambda = 1$. The plot depicts the SNR levels experienced by the joining line and by one of the four active lines each time the precoder coefficients are changed during the acquisition phase. In each iteration, there is a base measurement, followed by eight perturbation measurements. These nine measurements correspond to one algorithm iteration. For the joining line, the subsequent base measurement reflects the higher SNR obtained using an improved inverse channel estimate. The figure shows that the joining line can experience SNR drops of around 4 dB relative to the base measurement during the perturbation steps. Due to the second-order effects mentioned in Section II-A, the SNR on the neighboring line also experiences fluctuations, particularly during the first iteration. When needed, these fluctuations can be reduced by limiting the perturbation magnitude.

Finally, for these trials we compared the value (25) with the experimental results. In each row of Table IV, the initial level of interference $\xi^{[n]}$ is fixed and we report the empirical mean improvement ratio $\mathbb{E}^{[n+1]} / \mathbb{E}^{[n]}$ after one iteration, averaging over 10,000 random trials. The empirical values are compared with the upper bound $\alpha(M, \lambda, \xi)$ from (25). Other parameters of the simulation included the background noise $N_0 = 10^{-2}$, perturbation scale $\lambda = 1$, repetition factor $M = 1$, number of lines $L = 5$, and symbols per report $T = 256$. When the normalized interference $\xi$ is significantly larger than one, the reduction in interference achieved in one iteration is about 15 dB. When $\xi$ is close to one, the improvement is usually reduced somewhat. In all cases, the bound is fairly tight, within 1.5 dB of the empirical values. This gap is because the actual error variances are somewhat smaller than implied by the coefficient of variation bound $\kappa = 1$.

### B. Protection Phase

We also considered the scenario where a single active line acquires the crosstalk coefficient of a single joining line using the algorithm proposed in Section IV. In this situation, we set $\beta = 1$, meaning that the base SNR on an active line should fall by no more than 3 dB during the protection phase.
There can be an additional loss in SNR from the perturbations themselves and, as $\zeta = 1$, these will also be roughly 3 dB. We set $N = 10^{-5}$, $\kappa = 1$, and chose the overall reliability to be $\epsilon = 10^{-2}$. At start up we took an initial power of -45 dB relative to full power as sufficient to ensure that the additional normalized crosstalk interference is within $\beta$, taking into account the initial error in the estimate of $g$.

For each trial, the phase of $g$ was chosen uniformly at random, and the squared magnitude was $|g|^2 = 0.1$. Thus, if the joining line were to start at full power, the victim line SNR would be immediately drop to 10 dB.

Figure 5 plots the probability of excessive mean interference $E_K$ against the number of power increments $K$. Using the simple exponential bound (27) shows that the target probability can be satisfied with less than $K = 10$ iterations, whereas the refined bound (29) gives a bound of $K = 7$ iterations. The simulations demonstrate that in fact $K = 5$ is feasible.

Fixing $K = 7$, a joining experiment was then performed in which the same initial estimate was used giving 48.8 dB as starting SNR. Figure 6 shows a histogram for the actual SNR values observed on active lines at all stages of the algorithm. The right-most mode in the histogram corresponds to the unperturbed SNR, and the other modes correspond to various perturbation impacts. The additional losses from perturbations can be seen. Using (13) we determined the maximum total SNR loss should be 8.3 dB with the above parameter values, neglecting estimation error. It can be seen that this bound is approached, but only rarely, in the course of the experiment. Closer inspection of the results showed that the active line SNR fell rarely below the target value of 47 dB, with only seven out of 10,000 update sequences falling more than 3 dB below target during joining. Our results also showed that even the perturbed SNR rarely fell below 46 dB.

VI. CONCLUSIONS

Our results confirm that the use of SNR feedback can form the basis of a robust scheme to acquire precoder coefficients. As we show, iterating between residual estimation and precoder updates while adaptively adjusting the perturbation magnitude results in efficient and accurate coefficient acquisition. Principal amongst the considerations for any scheme are the time to acquire the precoder coefficients and the impact on the other lines. Our theoretical framework allows the trade-off between these to be struck. The common feasibility requirement for these schemes is that the residual crosstalk interference be reliably reduced to a predetermined target level.

Our numerical results show that crosstalk coefficient acquisition can be accomplished in only a small number of algorithm steps and in such a way that the number of steps needed, as well as the SNR loss during measurement, can be predicted in advance. Additional speed is nonetheless desirable and this can be accomplished in various ways which have not been described here. For instance, while the estimation procedures presented here produce independent estimates on all tones, the estimates on multiple tones can be jointly processed, thereby taking advantage of structure in crosstalk channels as a function of frequency.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their insightful comments and suggestions.

APPENDIX A

MEASUREMENT VARIANCE

The goal of this section is to prove Theorem 1. The following lemma gives a special property of complex variables with a certain symmetry.

Lemma 1: Let $a = c(x + jy)$ be a complex random variable where $c$ is a complex constant and where $x$ and $y$ are real random variables. Suppose that $x$ and $y$ are each zero mean, mutually uncorrelated, and have identical variance $E[x^2] = E[y^2] = \sigma^2/2$. Then $E[a^2] = 0$.

The proof is straightforward:

$$E[a^2] = c^2 (E[x^2] - E[y^2] + 2jE[x]E[y]) = 0.$$  

Under the conditions of Theorem 1, each of the interference signals $\theta_{i,j}, x_j$ and the noise $u_0$ satisfy Lemma 1. The next lemma expresses the variance of the squared magnitude of sums of such variables.

Lemma 2: Let $S = \sum_{i=1}^{N} a_i$ where the variables $a_i$, $1 \leq i \leq N$, are mutually independent, complex valued random
variables with bounded fourth moments. Suppose further that 
\( E[a_i] = 0 \), \( E[|a_i|^2] = \sigma_i^2 \), and \( E[a_i^2] = 0 \). Then

\[
\mathbb{V}[|S|^2] = \sum_{i=1}^{N} \mathbb{V}[|a_i|^2] + \sum_{i=1}^{N} \sum_{k \neq i} \sigma_i^2 \sigma_k^2.
\]

**Proof:** The fourth moment of \( S \) can be expressed by

\[
\mathbb{E}[|S|^4] = \mathbb{E}[SSSS] = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} E[a_i a_j a_k a_l].
\]

By the independence properties and first and second order expectations, we have

\[
E[a_i a_j a_k a_l] = \begin{cases} 
E[|a_i|^4], & \text{for } i = j = k = l, \\
\sigma_i^2 \sigma_j^2 \sigma_k^2, & \text{for } i = j, k = l, i \neq k, \\
\sigma_i^2 \sigma_j^2 \sigma_k^2, & \text{for } i = l, k = j, i \neq k, \\
0, & \text{otherwise}.
\end{cases}
\]

so that

\[
\mathbb{E}[|S|^4] = \sum_{i=1}^{N} E[|a_i|^4] + 2 \sum_{i=1}^{N} \sum_{k \neq i} \sigma_i^2 \sigma_k^2.
\]

Since \( \mathbb{E}[|S|^2] = \sum_i \sigma_i^2 \), we have

\[
\mathbb{E}[|S|^2]^2 = \sum_{i=1}^{N} \sigma_i^4 + \sum_{i=1}^{N} \sum_{k \neq i} \sigma_i^2 \sigma_k^2
\]

and the lemma follows from \( \mathbb{V}[|S|^2] = \mathbb{E}[|S|^4] - \mathbb{E}[|S|^2]^2 \). \( \square \)

The following lemma can be used to bound the coefficient of variation of the interference.

**Lemma 3:** For the variables \( a_i \) defined in Lemma 2, suppose that there is \( \kappa > 0 \) such that \( \mathbb{V}[|a_i|^2] \leq \kappa \mathbb{E}[|a_i|^2]^2 \) for each \( i = 1, \ldots, M \). Then we have

\[
\mathbb{V}[|S|^2] \leq \max(\kappa, 1) \mathbb{E}[|S|^2]^2.
\]

**Proof:** We have

\[
\mathbb{V}[|S|^2] = \sum_{i=1}^{N} \mathbb{V}[|a_i|^2] + \sum_{i=1}^{N} \sum_{k \neq i} \sigma_i^2 \sigma_k^2
\]

\[
\leq \max(\kappa, 1) \left( \sum_{i=1}^{N} \sigma_i^4 + \sum_{i=1}^{N} \sum_{k \neq i} \sigma_i^2 \sigma_k^2 \right)
\]

\[
= \max(\kappa, 1) \left( \sum_{i=1}^{N} \sigma_i^4 \right)^{1/2}.
\]

\( \square \)

Theorem 1 follows from the preceding three lemmas, on taking \( S = \left( \sum_{k \neq i} \sigma_i^2 \right) + u_k \). Because the noise \( u_k \) is complex Gaussian, it satisfies the conditions of Lemma 3 with \( \kappa = 1 \).

**Appendix B**

**Proof of Theorem 2**

Let the stopping time \( T_\beta \) be the smallest \( n \) such that \( V_n \leq \beta \).

Let \( M_n \) be the process \( V_n \) stopped when the interval \([0, \beta]\) is entered, i.e. \( M_n = V_n \) for \( n < T_\beta \) and \( M_n = 0 \) for \( n \geq T_\beta \).

For the stopped process, we have

\[
E[M_n | F_{n-1}] \leq M_{n-1},
\]

as the inequality trivially holds beyond the stopping time \( T_\beta \). It follows immediately that the limit \( M_\infty \) exists almost surely since the process is a positive super-martingale.

For a random variable \( X \) and event \( A \), the expectation of \( X \) times the indicator \( 1_A \) of \( A \) is denoted \( E[X; A] := E[X1_A] \).

In this notation, the tail probability of the stopping time can be expressed by

\[
\mathbb{P}[T_\beta > n] = \mathbb{P}[M_n > \beta] = \frac{\beta}{\mathbb{E}[M_n]} \leq \frac{\beta}{M_\infty}.
\]

To bound the latter expression, note that for any \( n \geq 1 \), we have

\[
E[M_n; M_n > \beta] \leq E[M_n; M_{n-1} > \beta] \leq \alpha E[V_{n-1}; V_{n-1} > \beta] = \alpha E[M_{n-1}; M_{n-1} > \beta],
\]

using the definition of \( M_n \) and the martingale property of \( V_n \). Then by induction, we have

\[
E[M_n; M_n > \beta] \leq \alpha^n E[M_0; M_0 > \beta] \leq \alpha^n V_0.
\]

As a result, we have

\[
\mathbb{P}[T_\beta > n] \leq \min(1, \alpha^n V_0 / \beta).
\]

Finally, let

\[
M = \left\lfloor \frac{\log(V_0 / \beta)}{\log(1/\alpha)} \right\rfloor
\]

be the smallest index such that \( \alpha^M V_0 / \beta < 1 \). The expected stopping time can be bounded as

\[
E[T_\beta] = \sum_{n=0}^{\infty} \mathbb{P}[T_\beta > n] \leq M + \frac{\alpha^M V_0}{\beta(1-\alpha)} \leq M + \frac{1}{1-\alpha},
\]

as required.

**Appendix C**

**Weighted Sum of Chi-Squared Variates**

The following Lemma shows how to compute the tail probability of a weighted sum of chi-squared random variables with one degree of freedom, as needed for (29). Let \( B(a, b) = (\Gamma(a)\Gamma(b)) / \Gamma(a+b) \) be the beta function and define

\[
\mathbb{P}_h(v) = \begin{cases} 
1 + v + \ldots + \frac{v^h}{h!}, & \text{exp}(-v).
\end{cases}
\]

**Lemma 4:** If \( X_\alpha \sim \Gamma(\frac{1}{2}, \frac{1}{\alpha}) \), \( X_\beta \sim \Gamma(\frac{1}{2}, \frac{1}{\beta}) \), \( \beta \leq \alpha \) and set

\[
H(u) = \mathbb{P}[X_\alpha + X_\beta \geq u]
\]

(31)
then
\[ H(u) = C(\alpha, \beta) \sum_{k=0}^{\infty} \left( \frac{\alpha - \beta}{\alpha} \right)^k \frac{1}{\alpha} P_k(\alpha u) B(\frac{1}{2}, \frac{1}{2} + k), \]

where \( C(\alpha, \beta) = \frac{1}{\Gamma(\alpha) / (\Gamma(1/2))^2} \).

Proof: Let \( X_\alpha \) be a gamma random variable with scale \( 1/\alpha \) and with shape \( \nu \), so that the density of \( X_\alpha \) is
\[ f_\alpha(x) = \frac{\alpha^\nu x^{\nu-1} e^{-\nu x}}{\Gamma(\nu)}, \]
for \( x \geq 0 \).

Let \( X_{1/2} \) be another such random variable but with scale \( 1/\beta \) instead of \( 1/\alpha \) and with density \( f_{1/2}(x) \).

The density \( h(u) \) of the sum \( X_\alpha + X_{1/2} \) satisfies
\[ h(u) = \int_0^u f_{\beta}(y)f_{1/2}(u-y) dy \]
\[ = \frac{(\alpha \beta)^\nu}{\Gamma(\nu)} \int_0^u e^{-\alpha(u-y)} y^{\nu-1} (u-y)^{\nu-1} e^{-\beta y} dy. \]

Set \( y = ut \) and \( C(\alpha, \beta) := (\alpha \beta)^\nu / \{\Gamma(\nu)\}^2 \). It follows that
\[ h(u) = C e^{-\alpha u} u^{2\nu-2} \int_0^1 e^{-(\beta-\alpha)at} t^{\nu-1}(1-t)^{\nu-1} dt, \]

If we set \( \nu = 1/2 \) and expand the exponential in the integrand, terms involving \( t \) can be integrated out to obtain
\[ h(u) = C e^{-\alpha u} \sum_{k=0}^{\infty} (\alpha - \beta)^k \frac{u^k}{k!} B(\frac{1}{2}, \frac{1}{2} + k). \]

By integrating we obtain the desired survivor function \( H(u) \) in (32). \( \square \)

REFERENCES

Gerhard Kramer (S’91–M’94–SM’08–F’10) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada in 1991 and 1992, respectively, and the Dr. sc. techn. degree from the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, in 1998. From 1998 to 2000, he was with Endora Tech AG, Basel, Switzerland. From 2000 to 2008 he was with Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ. From 2009-2010 he was a faculty member of USC in Los Angeles, CA. He joined the Technische Universität München in 2010 as Professor for Communications Engineering. Prof. Kramer is a member of the Board of Governors of the IEEE Information Theory Society since 2009. He has served as Associate Editor, Guest Editor, and Publications Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. He served as co-chair of the technical program committee of the 2008 IEEE International Symposium on Information Theory, and as founding co-chair of the first, second, and third Annual Schools of Information Theory in 2008-2010. Prof. Kramer is serving as a member of the emerging technologies committee of the IEEE Communications Society. He is a co-recipient of the IEEE Communications Society 2005 Stephen O. Rice Prize paper award, a Bell Labs President’s Gold Award in 2003, and a recipient of an ETH Medal in 1998. He was awarded the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research in 2010.

Carl J. Nuzman (M’00) received Bachelors degrees in Electrical Engineering and Mathematics at the University in Maryland College Park in 1996, and a Ph.D. in Electrical Engineering from Princeton University in 2000. He is currently a Distinguished Member of Technical Staff in the Mathematics of Networks and Communications Department at Bell Labs, Alcatel-Lucent, in Murray Hill, New Jersey. Since joining Bell Labs in 2000, he has performed basic research in the fields of probability, optimization, and dynamical systems. He also has performed research with applications to optical networking, optical devices, micro electro-mechanical systems, power control, traffic modeling, and digital subscriber line systems.

Alexei Ashikhmin (M’00, SM’08) is with the Mathematics of Networks and Communications Research Department, Bell Laboratories, Alcatel-Lucent, in Murray Hill, NJ. He received a Ph.D. degree in Electrical Engineering from the Institute of Information Transmission Problems, Russian Academy of Science, Moscow, Russia, in 1994. In 1995 and 1996, he was with the Mathematics and Computer Science Department, Delft University of Technology, The Netherlands. From 1997 to 1999 he was a Postdoctoral Fellow at the Modeling, Algorithms, and Informatics Group of Los Alamos National Laboratory. Since 1999, he has been with Bell Laboratories. His research interests include communications theory, the theory of error correcting codes, and classical and quantum information theory. From 2003 to 2006 Dr. Ashikhmin served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. Dr. Ashikhmin is a co-recipient of the Bell Labs President’s Gold Award in 2003 and the IEEE Communications Society 2005 Stephen O. Rice Prize paper award.

Adriaan J. de Lind van Wijngaarden (S’87–M’98–SM’03) received an engineering degree in electrical engineering from Eindhoven University of Technology, The Netherlands, in 1992, and a doctorate in engineering from the University of Essen, Germany, in 1998. From 1992-1998, he was a Research Engineer with the Digital Communications Group at the Institute for Experimental Mathematics, University of Essen, Germany. He has been with Bell Laboratories, Murray Hill, NJ, since 1998. Dr. De Lind van Wijngaarden has been deeply engaged in both theoretical and application-driven research in communications, information theory and coding. He provided key contributions to recording systems, high-speed optical systems and broadband access. He is currently an Associate Editor for Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY and he has served as a Publication Editor for the same journal from 2005 until 2008. He has co-organized several international conferences and workshops.

Miroslav Živković holds an engineering degree in electronics and telecommunications from the Faculty of Electrical Engineering at the University of Belgrade, Serbia. He was with Bell Labs in The Netherlands, where he contributed significantly to the area of dynamic spectrum management for DSL systems. He is currently with TNO-ICT, Delft, The Netherlands. His current research interests include performance analysis of service-oriented architecture systems and composite web services.