Cross-Layer Modelling for Efficient Transmission of Non-realtime Data Traffic over Downlink DS-CDMA Heterogenous Networks

Keivan Navaie, Shahrokh Valaee and Elvino S. Sousa

Abstract— **In this paper, we develop a cross-layer model for downlink interference in heterogenous DS-CDMA wireless cellular networks. In this model, interference is described as a function of application layer parameters (traffic characteristics) and physical layer variations (channel characteristics). We show that for a heterogenous service DS-CDMA network, downlink interference is a second-order self-similar process and thus has long-range dependence. We then use the predictive structure of total downlink interference to maximize non-realtime data throughput. We use fractional Gaussian noise (fGn) to model the self-similarity of downlink interference. In the proposed method, the base-station uses an optimal linear predictor, based on the fGn model, to estimate the level of interference. The estimated interference is then used to allocate power to users. To maximize data throughput, we use time domain scheduling. The simulation studies confirm the self-similarity of downlink interference and validate the fGn model. The simulation results also show a substantial performance improvement using the proposed predictive-adaptive scheme and confirm that the interference model is still valid after applying the proposed method.**

Index Terms— **Cross-Layer modelling, downlink interference, DS-CDMA networks, fractional Gaussian noise, self-similar process, time domain scheduling.**

I. INTRODUCTION

A PROMISING air interface technology for future wireless communications is the *direct sequence code division multiple access* (DS-CDMA), which has been shown to be interference-limited [1]. Hence, exploiting the fluctuations of the total interference for system performance improvement is a major challenge of the next generation heterogenous wireless networks. Several parameters operating in different network layers beget interference fluctuation. They include, for instance, the total number of active users, their call duration, the allocated power to each call in the corresponding base-station, channel variation and user mobility. In this paper, we use a cross-layer approach [2] to model downlink interference as a time-series. In the cross-layer modelling, for a given time-scale of interest, the physical layer knowledge of the wireless medium is shared with higher layers to develop comprehensive models in the corresponding time scales.

This work was partly supported by Bell University Laboratories.

K. Navaie was with the Bell University Laboratories, University of Toronto. He is now with the Broadband Communications and Wireless Systems (BCWS) Center, System and Computer Engineering Department, Carleton University, Ottawa, Ontario, Canada, K1S 5B6. (email: keivan@sce.carleton.ca)

S. Valaee and E. S. Sousa are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada, M5S 3G4. (email: {valaee,sousa}@comm.utoronto.ca)

The time variations of the interference has been studied in [3], where it has been observed that the total interference in a data-centric DS-CDMA system is a self-similar process. It has been shown in [3] and [4] that the source of self-similarity of the total interference is the user traffic characteristics. The predictive structure of the self-similar interference is then used in [3] and [4] to develop resource management algorithms; they use a predictive mechanism to adjust transmission rate with the variations of total downlink interference.

In [5], we have shown that under certain conditions, the total downlink interference for a heterogenous cellular network is a self-similar process and thus has long-range dependence. In our heterogenous model the network serves a mixed traffic of real-time services and non-realtime services with non-Poisson traffic characteristics. User traffic characteristics and channel fading process are the two parameters that create self-similarity in downlink interference. We have also shown in [5] that the conditions for the self-similarity of downlink interference are more general than those proposed in [3] and [4]. The proposed model is cross-layer since both physical characteristics of the wireless channel and user traffic parameters are used in the model. The long-range dependence of downlink interference is valid in time-scales of the order of that used in rate control and admission control.

The present paper utilizes the predictive nature of self-similar model of interference to develop a novel cross-layer adaptive-predictive radio resource controller. In this paper, we model the self-similarity of the total downlink interference with the fractional Gaussian noise (fGn) [6] and then design an optimal predictor. Note that a substantial simplification is possible for Gaussian self-similar processes; i.e. an infinite family of distributions can only be represented by three parameters over the entire scaling region. The three parameters are the mean, the variance, and the self-similarity index. The optimal predictor is then used in a time domain scheduler to maximize non-real time data throughput. We show that the proposed fGn model for the downlink interference is valid even after using the proposed scheduler. The application of optimal linear predictor on the fGn model is simple and can easily be realized in either mobile terminal, base-station, or radio network controller.

In the proposed method, the mobile terminal measures the received downlink interference in each time slot. A linear predictor is then employed to estimate the interference level in the next time-slot based on this measured value and the stored interference levels in the past. The model parameters are also adaptively adjusted based on the stored interference values. The estimated interference is then used in base-station or radio network controller to allocate power to users over the next time-slot. First, power is allocated to real-time traffics and then the remaining power is assigned to users with best-effort service. We use a time domain scheduler [7] to maximize the throughput of non-realtime data traffic. In time domain scheduling, over each time interval, the total power is allocated to a single user with the rest of the users kept silent. The temporal extent of the interval depends on the user data rate, the average channel conditions, and the available power.

We simulate a heterogenous DS-CDMA network based on the Universal Mobile Telecommunication System (UMTS) standard [8]. We validate the model and observe self-similarity in total downlink interference when the call duration—of at least one service—has a heavy tail distribution. The self-similarity of downlink interference persists even after the time domain scheduler is applied. Simulation results show that fGn is an appropriate model for downlink interference. We also study the non-realtime data throughput of this scheme by locating the call admission region. The simulation results show that our proposed method provides a higher throughput than other competing schemes.

Organization of this paper is as follows. The interference model and the results on the self-similarity of the interference are presented in Section II. In Section III, we discuss the proposed method. The simulation results are presented in Section IV. The paper is concluded in Section V.

II. Interference Model

In a DS-CDMA cellular network, the total downlink interference, $I(n)$, is a weighted sum of the transmitted power of base-stations (BSs), $P^c(n)$, for $n \in \mathbb{Z}$, $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ and $c = 1, \ldots, N_C$, where N_C is the number of cells in the network,

$$I(n) = \sum_{c=1}^{N_C} \xi^c(n) P^c(n) g^c(n). \quad (1)$$

The weight coefficients are the corresponding channel gains, $g^c(n)$, and the average normalized cross-correlation, $\xi^c(n)$, between the user's signal and the received signals of other users. The two processes $g^c(n)$ and $\xi^c(n)$ are stationary processes independent of $P^c(n)$. Each sample of $I(n)$ is valid over a window of length T_w seconds, the *modeling time scale*. We assume that $T_w \gg T_c$, where $1/T_c$ is the spreading bandwidth of the cellular CDMA network. Without loss of generality, let the user under study be located in cell 1. In (1), we assume that the power allocated to that user is not included in $P^1(n)$.

To study $I(n)$, we assume that a *regular power transmission* regime is applied network-wide, in which the transmitted power by any BS is not substantially higher than the transmitted power by other BSs. This assumption is practically valid if a load balancing mechanism is applied in the cellular network.

In time-slot n, each BS serves a set of active users (calls) in its coverage area, therefore the transmitted power by the BS c, $P^c(n)$, is the sum of allocated powers to all calls in the corresponding coverage area,

$$P^c(n) = \sum_{j=1}^{J} \sum_{i \in \mathbb{N}} p^c_{ji}(n - \upsilon^c_{ji} + 1), \quad (2)$$

where J is the number of services provided by the network, $\mathbb{N} = \{1, 2, \ldots\}$, $p^c_{ji}(.)$ is the allocated power of call i of service j of cell c, and $\upsilon^c_{ji} \in \mathbb{Z}$ is the start time of the ith call in cell c that receives service j. Calls are enumerated by i in the order of their arrival, such that in each cell c, $\upsilon^c_{ji} \leq \upsilon^c_{ji+1}$. For the ith call of service j in cell c with a call duration of $\tau^c_{ji} \in \mathbb{N}$ seconds, $p^c_{ji}(.)$ is the allocated power in its call duration, and is equal to zero otherwise.

To characterize $I(n)$, we first need to obtain the characteristics of $P^c(n)$ and $g^c(n)$. We assume that for each given cell c and service j, the call duration sequence process $\{\tau^c_{ji}, i \in \mathbb{N}\}$, the new call arrival rates sequence process $\{\mu^c_{ji}(.), i \in \mathbb{N}\}$, and the allocated power sequence process $\{p^c_{ji}(.), i \in \mathbb{N}\}$ are independent and identically distributed (i.i.d.) random processes. We denote τ^c_{ji}, $\mu^c_{ji}(n)$ and $p^c_{ji}(n)$ by the generic random variables τ^c_j, $\mu^c_j(n)$ and $p^c_j(n)$, respectively.

In this model, the *traffic characteristics* of a user of service j is specified by three processes, $\mu^c_j(n)$, τ^c_j and $p^c_j(n)$, where $p^c_j(n)$ is a function of the service type j, the bit-rate, and the power allocation strategy in the network. In [5], we show that the downlink interference can be *completely specified* by *traffic characteristics* corresponding to different services provided by the network and *channel processes*, $g^c(n)$, for all c. Here, we briefly review the models we use in this paper for new call arrival process, call duration process, allocated power process, and wireless channel process.

1) New Call Arrivals: Assuming that the arrival rate of new calls for each service type is less than the value for which the network was designed, we show in [5] that, using a regular interface based call admission control, the Poisson distribution is an appropriate model for call arrival $\mu^c_j(n)$,

$$\Pr\{\mu^c_j(n) = \nu\} = \frac{(\lambda^c_j)^\nu e^{-\lambda^c_j}}{\nu!},$$

where λ^c_j is the arrival rate.

2) Call Durations: Here, we denote both packet duration (for packet-oriented transmission) and call duration (for connection oriented transmission) as "call duration". For voice service, an exponentially distributed call duration is assumed [9]. For non-voice traffic, a general heavy-tail distribution is considered (see e.g. [10]). A random variable X is said to be heavy-tailed with infinite variance if for $0 < \kappa < 2$, there exist a slowly varying function $L(x)$ such that as $x \to \infty$,

$$P(|X| \geq x) \sim L(x) x^{-\kappa},$$

where the symbol '$\sim$' means *behaves asymptotically as* (i.e., $\phi(k) \sim \varphi(k)$ means: $\lim_{k\to\infty} \frac{\phi(k)}{\varphi(k)} = 1$). A function $f(x) > 0$, $x \in \mathbb{R}$ is called a *slowly varying function* if for all $u \in \mathbb{R}_+$, $\frac{f(ux)}{f(x)} \to 1$, as $x \to \infty$. An example of a heavy tail distribution is Pareto distribution:

$$Pr\{\tau = l\} = \eta_0 l^{-\alpha-1},$$

where $l \in \mathbb{N}$ and

$$\eta_0 \triangleq \frac{1}{\sum_{l=1}^{\infty} l^{-\alpha-1}}.$$

Pareto distribution has been used to model call durations (see e.g. [10]).

3) Allocated Power to Each Call: For $p_j^c(n)$, we note that for a given channel, the allocated power to a given user at time-slot n is a concave function of its bit-rate [9].

4) Wireless Channel: We assume that the channel gain, $g^c(n)$ is a second-order stationary process for $c = 1, \ldots, N_C$. To obtain the channel gain $g^c(n)$, we assume a deterministic distance-dependent path loss and two fading effects: fast fading and shadowing. Note that fast fading (e.g., Rayleigh or Rician) affects $P^c(n)$ in (1) in smaller time-scales than the shadowing. Fast fading is also partly cancelled by the fast power control. Moreover, the short-range effect of fast fading is averaged out in longer time-scales such as T_w. Therefore, we neglect fast fading and assume that the channel gain $g^c(n)$ is given by

$$g^c(n) = L_c d_c^{-\gamma_c} \theta^c(n), \tag{3}$$

where d_c is the distance between the base-station c and the user for which the downlink interference is measured, γ_c is the path loss exponent which is a function of the antenna height and the signal propagation environment, L_c is an environmental constant, and $\theta^c(n)$ is the shadowing process. The shadowing process $\theta^c(n)$, has a log-normal distribution with standard deviation σ_c. The Gudmundson correlation model [11] is used for log-normal shadowing as

$$\Theta^c(n+1) = \eta^c \Theta^c(n) + (1-\eta^c)\nu^c(n), \tag{4}$$

where the time-scale is T_f (fading period), $T_f \geq T_w$, $\Theta^c(n) = \log \theta^c(n)$ is the log-normal fading in dB, $\nu^c(n)$ is a zero-mean white Gaussian noise with variance $\sigma_c^2(1+\eta^c)/(1-\eta^c)$, and $0 < \eta^c < 1$ is the channel correlation coefficient.

The auto-covariance function of $g^c(n)$ is denoted by $C_g^c(k)$. We further assume that

$$C_g^c(k) \sim L_g^c(k) k^{-\beta_g^c}, \qquad k \to \infty, \tag{5}$$

where k denotes time with a temporal resolution T_w, $L_g^c(k)$ is a slow varying function, and $\beta_g^c > 0$ is the channel auto-covariance decay exponent. In [5], we show that (5) is consistent with shadowing models in the corresponding time scales.

A. Self-similarity of Downlink Interference

Here we restate the results in [5] where we show that the total downlink interference in multi-service wireless CDMA networks is an asymptotically self-similar process where the self-similarity emanates from user traffic characteristics. First, we present the definition of an asymptotically self-similar process:

Definition [12]: A real-valued second-order stationary random process $I = \{\ldots, I(-1), I(0), I(1), \ldots\}$ is called an *asymptotically self-similar process* (as-s), with self-similarity index $H = 1 - \beta/2$, $0 < \beta < 1$, if

$$\lim_{m\to\infty} C_{(m)}(k) = \frac{C_{(m)}(0)}{2}\left((k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}\right), \tag{6}$$

where $k \in \mathbb{Z}_+$, $C_m(k)$ is the auto-covariance function of I^m that is the average process of I over blocks of length m.

In other words, a process I is as-s if the correlation coefficients of the average process of block length m as $m \to \infty$ are identical to those of a self-similar process. A sufficient condition for a second-order stationary process I being asymptotically self-similar is that for $k \in \mathbb{Z}_+$, $k \to \infty$, the auto-covariance function of I, $C(k)$, behaves asymptotically as $L(k)k^{-\beta}$, (i.e. $C(k) \sim L(k)k^{-\beta}$), in which $0 < \beta < 1$, and $L(k)$ is a slowly varying function [12].

Suppose that the downlink interference process, $I = \{\ldots, I(-1), I(0), I(1), \ldots\}$, is a finite-mean, finite-variance second-order stationary process. In the following proposition, we derive the necessary conditions for the self-similarity of downlink interference.

Proposition 1 [5]: Consider the downlink interference process, I, and let β_P^c, $c = 1, \ldots, C$ satisfy

$$\sum_{j=1}^{J} \lambda_j^c \Pr\{\tau_j^c = k\} r_{j(k)}^c(k) \sim L_P^c(k) k^{-\beta_P^c - 2}, \ k \to \infty, \tag{7}$$

where $L_P^c(k)$ is a slowly varying function. Now, I is an as-s process with self-similarity index $H = 1 - \beta^*/2$ if there exists at least one c such that $0 < \beta_P^c < 1$ or $0 < \beta_g^c < 1$, and

$$\beta^* = \min_c \min\{\beta_P^c, \beta_g^c\}. \tag{8}$$

Proposition 1 gives the sufficient condition as a combination of the service call arrival rate, λ_j, the service call duration distribution, $\Pr\{\tau_j = k\}$ for $k \to \infty$, and the asymptote of the correlation function of the allocated power, $r_{j(k)}^c(k)$, for $k \to \infty$.

III. Throughput Maximization of non-realtime Data in The Downlink

In this section, we introduce a fractional Gaussian noise (fGn) model for the total downlink interference and use this model to devise an optimal downlink interference predictor. The predicted downlink interference is then used to allocate the available power to non-realtime users. Finally, we utilize *time domain scheduling* to maximize the throughput of non-real time users with best-effort quality-of-service (QoS). The block diagram of the proposed scheme is illustrated in Fig. 1. In the following subsections, we discuss the sub-blocks of the proposed system.

A. Model-based Optimal Downlink Interference Prediction

There are a number of well-known models for self-similar processes [6]. If the variance and the self-similarity index, H, of a zero-mean self-similar process are known—subject to assuming an idealized Gaussian setting—the process can be modelled by a *fractional Gaussian noise* (fGn) [6]. fGn is a self-similar Gaussian process with the auto-covariance function

$$\gamma(k) = \frac{\sigma_0^2}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right), k \in \mathbb{Z}, \tag{9}$$

where σ_0^2 is the variance and H is the self-similarity index.

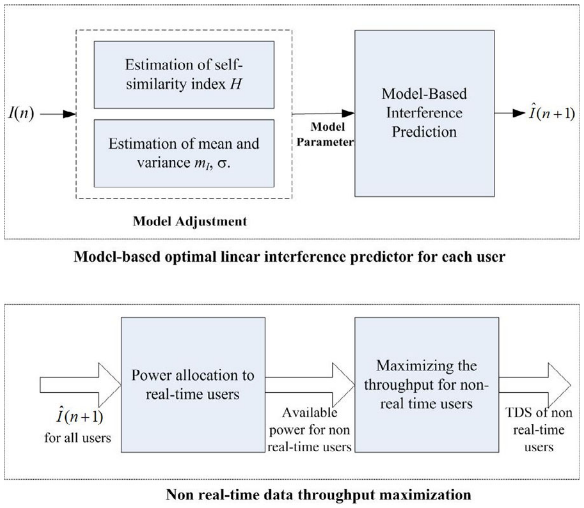

Fig. 1. Model-based predictive-adaptive throughput maximization of non-realtime data in the downlink.

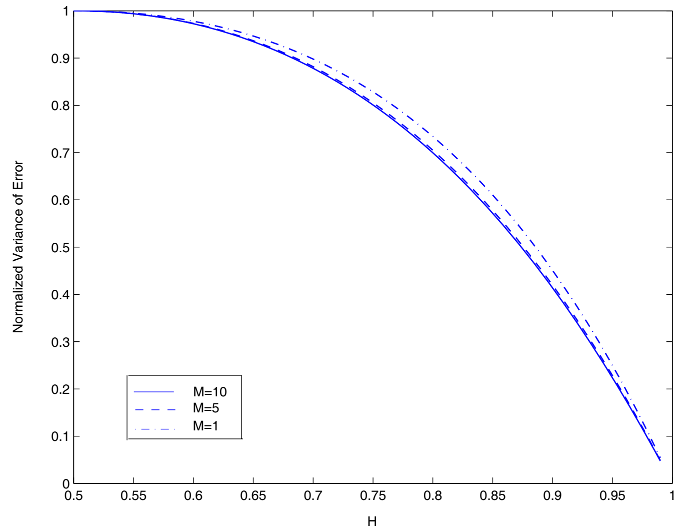

Fig. 2. Variance of the prediction error versus self-similarity index, H, for different values of M, for a fGn model with $\sigma_0 = 1$.

We assume a large number of users in the coverage area of the network. We also select an appropriate time-scale and assume that the resource control mechanism does not alter the total downlink interference model. The interference is then modelled using a fGn process as follows:

$$I(n) = m_I(n) + z(n), \tag{10}$$

where $m_I(n)$ is the average interference of the process $I(n)$ measured over a large window of size K (i.e. $(n-K, n-1]$), and $z(n)$ is fGn. In our proposed scheme, the parameters of the interference model in (11) are adjusted based on the past interference measurements.

We use the correlation structure of the total received interference and the auto-covariance of fGn, (9), to propose the following optimal linear predictor for the total interference [13]

$$\hat{I}(k+1) = m_I(k) + (\mathbf{\Gamma}^{-1}\underline{\gamma})^T(\mathbf{I}(k) - m_I(k)\mathbf{1}_M), \tag{11}$$

where $\hat{I}(k+1)$ is the predictor of $I(k+1)$, $\mathbf{I}(k) \triangleq [I(k), I(k-1), \ldots, I(k-M+1)]^T$ is the $M \times 1$ vector of stored interference measurements, $M \leq K$ is the memory length of the predictor, $\mathbf{\Gamma}$ is the covariance matrix with entities $\Gamma_{ij} \triangleq \gamma(i-j)$, $\underline{\gamma} \triangleq [\gamma(1), \ldots, \gamma(M)]^T$ and $\mathbf{1}_M$ is the $M \times 1$ identity vector; the superscript T denotes vector transposition. The variance of the prediction error is

$$\varepsilon = \gamma(0) - \gamma_1^T \mathbf{\Gamma}^{-1} \gamma_1. \tag{12}$$

In Fig. 2, we illustrate the variance of the prediction error versus self-similarity index H for a fGn process with $\sigma_0 = 1$ with three values of M. It is seen that, the variance of error decreases with H. Also the variance of error is not sensitive to the value of M for $M \geq 5$. The figure shows that the optimal linear predictor can be implemented using a small number of measured interference samples.

1) Model Adjustment: Since self-similarity of downlink interference emanates from user traffic characteristics, the application of the proposed scheme—and possibly other radio resource control mechanisms—would not alter the self-similarity of the downlink interference; nevertheless it may alter the model parameters. In this paper, this claim is validated with simulations (see Section IV). To compensate the effect of model parameter variations, we adaptively adjust the model parameters in appropriate time scales. We assume that K samples of the stored measured interference in the past are available in the predictor.

For the estimation of the mean of the interference, a simple weighted summation of measured values of total interference is used for mean estimation as follows:

$$m_I(k) = \frac{1}{K} \sum_{m=k-K}^{k-1} I(m). \tag{13}$$

The interference variance is also estimated as follows

$$\sigma_0 = \frac{1}{K-1} \sum_{m=k-K}^{k-1} \big(I(m) - m_I(k)\big)^2. \tag{14}$$

To estimate the self-similarity index, H, we use an on-line version of Abry-Veitch wavelet-based estimator in [14]. This method is based on performing a weighted least squares fitting between the different octaves, j_1, j_2, $j_2 > j_1$, in a multi-resolution platform using wavelet transformation. It is straightforward to show that the computational complexity for this method is in the order of $O(K)$ [14]. The confidence interval of the estimated self-similarity index, $\widehat{H}$, by Abry-Veitch method is given in [14]:

$$\widehat{H} - \sigma_{\widehat{H}} z_\beta \leq H \leq \widehat{H} + \sigma_{\widehat{H}} z_\beta, \tag{15}$$

where z_β is the $(1-\beta)$ quantile of the standard Gaussian distribution, (i.e. $P(z \geq z_\beta) = \beta$) and

$$\sigma_{\widehat{H}} = \Big(\frac{2}{(\ln 2)^2 2^{-j_1} K}\Big)\Big(\frac{1-2^J}{1-2^{-(J+1)}(J^2+4)+2^{-2J}}\Big), \tag{16}$$

where $J = j_2 - j_1$.

Note that in our proposed scheme the interference predictor in Fig. 1, is implemented for all users including realtime and non-realtime users. This interference predictor can be implemented in mobile station, base-station or radio network controller. In each case, the measured interference level should be provided to the predictor in appropriated time-scales. The base-station or radio network controller then uses the predicted interference levels to evaluate and then allocate the available transmit power in the next control window.

B. Downlink Power Allocation

Let $\mathcal{G}$ be the set of real-time calls (such as voice and multimedia) served with *guaranteed* delay requirement and $\mathcal{B}$ be the set of delay-tolerant calls served under *best-effort* service category. For all users, the value of the total interference is predicted in the next control window using the model in (10). The parameters of the model, $m_I(n)$, σ_0, and H, are evaluated and adjusted based on the received interference, utilizing the methods presented in Section III.A. Using an offset shift to the required received bit energy to the interference-plus-noise spectral density, E_b/I_0, threshold, the power allocation to the users in set $\mathcal{G}$ is performed conservatively. Therefore, after allocating the remaining available transmission power to non-realtime users, their E_b/I_0 will be larger than or equal to the actual required threshold. This offset is then adaptively adjusted to satisfy the initial required E_b/I_0 for the users in $\mathcal{G}$.

We assume that the total transmit power of the base-station at time instant n is $P_{max}(n)$. This value may be set either permanently in the network dimensioning phase or adaptively by the radio network controller. Therefore, the available power for non-realtime users in time-slot n, $P_A(n)$, is

$$P_A(n) = P_{max}(n) - \sum_{i \in \mathcal{G}} p_i(n), \tag{17}$$

where $p_i(n)$ is the allocated power to call i; we have dropped the cell and service indexes for the brevity of discussion.

Here, our main objective is to maximize throughput of non-realtime traffic. Thus, we consider the problem of optimal allocation of $P_A(n)$ in the nth control window with length T_w. Let $t \in [nT_w, (n+1)T_w)$ be the time variable and let $p_i(t)$ and $r_i(t)$ denote the allocated power and the instantaneous rate of the ith user, respectively. Define the average rate of user i over time-slot n by

$$R_i(n) \triangleq \frac{1}{T_w} \int_{nT_w}^{(n+1)T_w} r_i(t) dt. \tag{18}$$

The total number of data bits transmitted from traffic i over time-slot n is $R_i(n)T_w$.

For a given value of $P_A(n)$ and N non-realtime traffics, we assume there exist average rates $R_i(n)$ such that

$$R_i(n) = \frac{W}{\rho_i} \mathrm{SIR}_i, \qquad i = 1, \ldots, N, \tag{19}$$

where ρ_i is the required E_b/I_0 for user i, SIR_i is the received signal-to-interference ratio of user i and $W = 1/T_c$ is the spreading bandwidth. The objective is to provide the average rate of $R_i(n)$ for user i in T_w. If such rates can be found for all users, we say that the rate vector $\mathbf{R}(n) \triangleq (R_1(n), \ldots, R_N(n))$ is *feasible*. For a feasible rate vector, the total number of transmitted bits over the nth sampling interval is given by $\sum_{i=1}^{N} T_w R_i(n)$. For time-slot n, we define, $\mathbf{p}_n(t) \triangleq (p_1(t), \ldots, p_N(t))$ and $\mathbf{r}_n(t) \triangleq (r_1(t), \ldots, r_N(t))$. The power and rate are allocated so as to maximize the total throughput defined as $\frac{1}{T} \sum_{i=1}^{N} T_w R_i(n)$ where T is the time required to transmit the traffic of all users. Note that since we have assumed a feasible rate vector, we will always have $T \le T_w$.

We select rate vector $\mathbf{r}_n(t)$ and allocate power $\mathbf{p}_n(t)$ such that the total throughput is maximized. Let $\mathbf{R}(n)$ be a feasible average rate vector. If we assume that the total number of transmitted bits is constant i.e. $\sum_{i=1}^{N} T_w R_i(n) = B$, then the maximum throughput problem can be formulated in terms of the total transmit time T:

$$\min_{\mathbf{p_n(t)}, \mathbf{r_n(t)}} \quad T \tag{20}$$

$$\text{s.t.} \quad \frac{1}{T_w} \int_0^T r_i(t) dt = R_i(n), \;\; i = 1, \ldots, N \tag{21}$$

$$\sum_{i=1}^{N} p_i(t) \le P_A(n), \qquad 0 \le t \le T \tag{22}$$

$$\sum_{i=1}^{N} T_w R_i(n) = B. \tag{23}$$

In this optimization problem, the first constraint, (21), selects the rates $r_i(t)$ such that $\mathbf{R}(n)$ is feasible. We note that to reduce T, it is beneficial to allocate the maximum available power over the window $[0, T]$. Therefore, the second constraint, (22), indeed holds with equality. Note that the power should only be assigned to user i when it is larger than a detection threshold; the users with the allocated power less than the threshold are not selected for transmission in the current control window. In other words, when the assigned power is smaller than the detection threshold, the signal cannot be detected and the allocated power will be wasted. The last constraint shows that the total number of transmitted bits is constant.

In the following, we show that the solution of (20)-(23) belongs to the set of *time domain schedulers*. Time domain scheduling is a scheme in which the total power is allocated to a single user over the normalized (with respect to T_w) time extent of ϕ_i seconds, where ϕ_i is the fraction of T_W that is allocated to user i. During this time extent, base-station transmits only to user i; the rest of the users are kept inactive. We assume that there exists a scheduling scheme $\mathcal{T}$, that can transmit $R_i(n)T_w$ bits to user i, $i = 1, \ldots, N$, within $T_{\mathcal{T}} \le T_w$ seconds. Let $T \le T_{\mathcal{T}}$ be a time extent in which $\mathcal{T}$ serves more than one non-real time users (namely Q) simultaneously. It is straightforward to show that equal amount of data can be transmitted within $T_{\text{TDS}} < T$ for any $\mathcal{T}$, where T_{TDS} is the time required in the time domain scheduling. Therefore, time domain scheduling is throughput optimal. The basic idea is that in time slot n, the feasible rates, $R_i(n)$, are related to

TABLE I

SIMULATION PARAMETERS

Parameter	Value
Number of BSs	19
Cell Radius	100 m
BSs Transmit Power	10 W
Physical Layer	Based on UMTS
Power Control	Fast Power Control 1500/s
T_w	10 ms
Standard Deviation of Fading	8 dB
Loss Exponent	-4
Thermal noise density	-174.0 dBm/Hz
T_f	100 msec
$E\xi^1$	0.5
Services	12.2 kbps voice, 32 and 64 kbps data
12.2 kbps voice	$E_b/I_0 = 5$ dB, 5 Erlangs
32 kbps data	$E_b/I_0 = 3$ dB Pareto Dist., $\alpha_1 = 1.5, E\tau_1 = 2$ s
64 kbps data	$E_b/I_0 = 2$ dB Pareto Dist., $\alpha_2 = 1.8, E\tau_2 = 1.5$ s

the corresponding allocated power (see (19)). In $\mathcal{T}$, the total available power, P_A, is simultaneously allocated to Q users. In time domain scheduling, the total power is allocated to user i over $\phi_i T$ seconds, which removes intra-cell interference received from other non-real time users. We use this fact and then through a straightforward mathematical derivation, we show that $\sum_{i=1}^{Q} \phi_i < 1$. Therefore, $T_{\text{TDS}} < T$, this shows that the solution of the optimization problem in (20) belongs to a set of time domain schedulers.

Note that, if $R_i(n)$ is a feasible rate, ϕ_i should be lower bounded by $R_i(n)/R_{max}$ where R_{max} is the maximum bit-rate of the system.

IV. SIMULATION

To study the system performance, we consider a two-tier hexagonal cell configuration with a wrap-around technique [15]. A UMTS [8] cellular network, with a fast power controller running at 1500 updates per second, is simulated. Cross-correlation between the codes is assumed to be equal to 0.5. Three traffic types are used: 12.2 kbps voice (with the required $E_b/I_0 = 5$ dB), 32 kbps data (with $E_b/I_0 = 3$ dB) and 64 kbps data (with $E_b/I_0 = 2$ dB). We assume 5 Erlangs of voice traffic. For data traffics, we assume Pareto call duration with $\alpha_1 = 1.5, E\tau_1 = 2$ minutes and $\alpha_2 = 1.8, E\tau_2 = 1.5$ minutes. Control window is $T_w = 10$ ms. Both data traffics have Poisson distributions with average rate of 10 arrivals per second. Channel fading is based on the Gudmundson [16] model with $\sigma_c = 8$ dB and $T_f = 100$ ms. A distance-dependent channel loss with path exponent $\gamma = -4$ is considered. Users are distributed uniformly with different service types. The details of the simulation setting are given in Table I.

A. Model Validation

The heavy-tail call durations of data traffics satisfy the conditions of Proposition 1. This proposition also gives the self-similarity index $H = 0.75$. We study the time trace of the received downlink interference measured at different locations. Measurements were taken at an arbitrary location in cell $c = 1$.

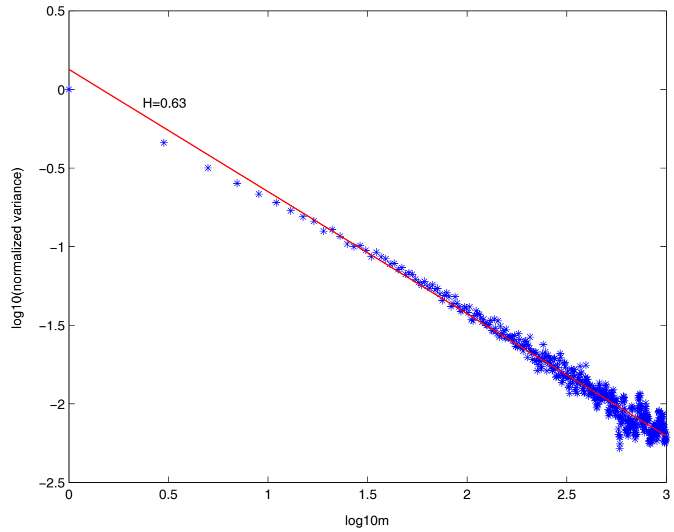

Fig. 3. The Normalized variance plot for total received downlink interference in the simulated network.

High variations are seen in traces with different time scales. A bursty behavior is observed in all time scales and as noticed, the total interference is self-similar.

To estimate the self-similarity index H, we use the variance plot method [6]. In this method, we divide $I(n)$ into non-overlapping blocks, each with m samples. For an as-s process with the self-similarity index H, the variance of the mean processes for $m \to \infty$ in a logarithmic scale, is a straight line. We estimate H by using the slope of this line. Fig. 3 illustrates the variance in the logarithmic scale. Using a linear curve fitting, we have $H = 0.63$. We also estimate the value of the H parameter using the Whittle estimator. This method is shown to be more accurate than the variance plot [6]. The Whittle estimator gives $H = 0.65$. The discrepancy between the estimated value of H by the Whittle estimator and that obtained from Proposition 1 is mainly due to the fact that, unlike the computation in Proposition 1 the interference threshold crossing does not allow the network to accept all call requests.

We now show that fGn is an appropriate model for $I(n)$. We use the quantile-quantile (Q-Q) plot [6] to show that the total interference is a Gaussian process. In a Q-Q plot, a Gaussian distributed process appears as the 45-degree reference line. Fig. 4 shows that the received total interference for the above configuration can be closely approximated by the Gaussian distribution.

B. Performance of the proposed adaptive-predictive method

To study the performance of the time domain scheduling scheme, we add a fourth non-realtime traffic. This traffic constitutes fixed-size packets with Poisson arrival with rate λ. For each λ, the bit rate is varied by changing the packet length. We set $N = 100$ and use a linear predictor with $M = 5$ taps to predict the interference level; we have found that the predictor error is not sensitive to the value of M for $M \geq 5$. We also assume that the cross-correlation between DS-CDMA codes in cell $c = 1$ is 0.5.

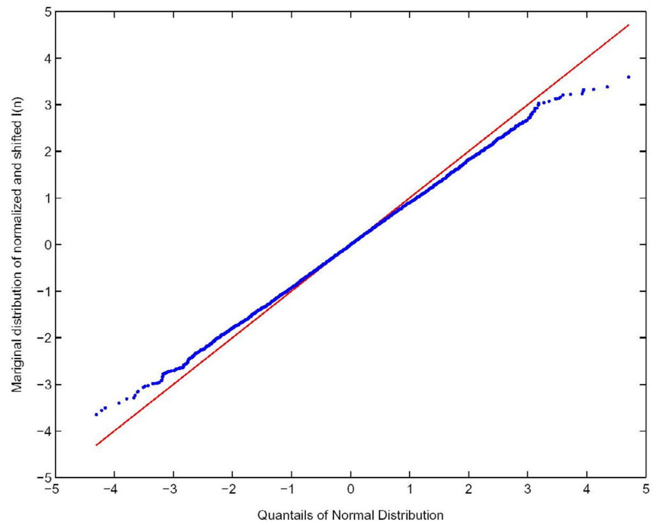

Fig. 4. The Q-Q plot for the marginal distribution of total interference.

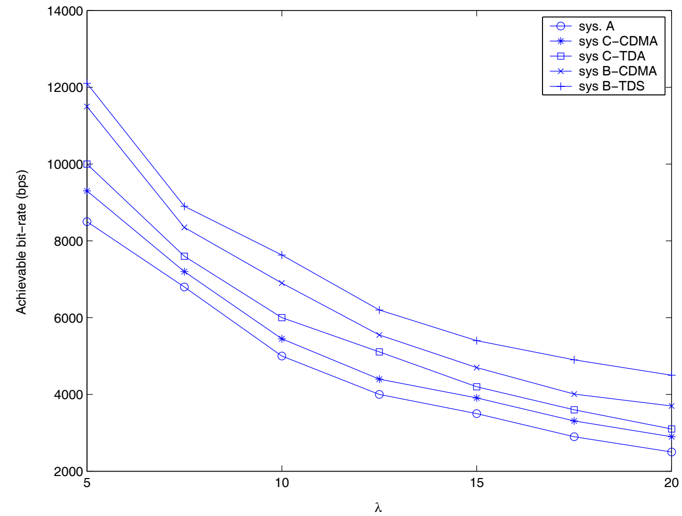

Fig. 5. The admission region versus packet arrival rate λ.

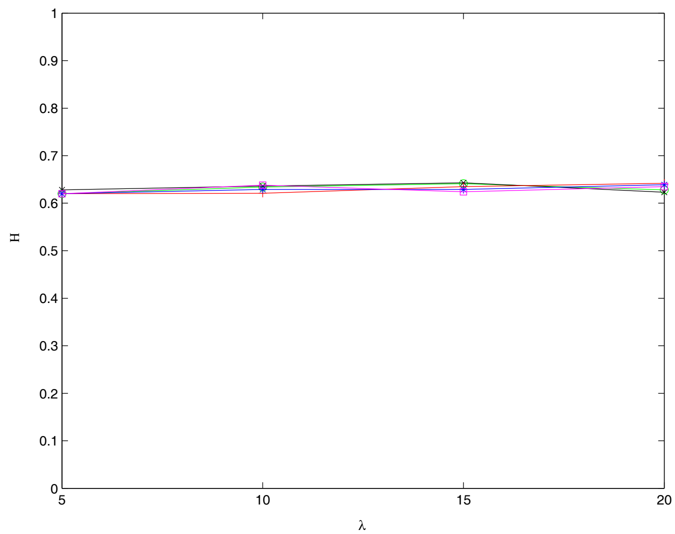

Fig. 6. The estimated values for H for five different runs versus non-realtime packets arrival rate for a given throughput $R = 5000$bps.

We have run 50 independent trials with users uniformly dispersed in the cell. In the sequel, we report the average of these runs. We consider three systems for comparison. The first system (System A) uses a pre-assigned share of 20% of the maximum downlink transmit power for non-realtime users. This share of power is then used to simultaneously transmit to non-realtime users (i.e. pure CDMA). The second system (System B) uses the proposed method in this paper. We first consider the system using time domain scheduling (System B-TDS) and then without time domain scheduling (System B-CDMA). A confidence interval of 95% for the estimate of the self-similarity index, H, is considered. The third system uses the average value of 5 last samples of measured interference as the predicted value (System C). Here, we also consider two cases with using time domain scheduling (System C-TDS) and without it (System C-CDMA).

We compare the admission region for the best-effort services of the systems A, B and C. The admission region is defined as the average achievable bit-rate for a given arrival rate of best-effort data services in the presence of a given set of realtime traffics. A significant improvement of admission region is seen in Fig. 5 using our proposed method (System B-TDS). In Fig. 5, it is also interesting to note that there are two different gains. The first one is due to using the predictor and the second one is due to the time domain scheduler. The first gain is captured when we move from System C to System B.

A very important question is: "What is the impact of applying time domain scheduler on the self-similarity of interference?" Fig. 6 illustrates the estimated values of H after applying the proposed method. The figure confirms that self-similarity of the downlink interference in not affected by time domain scheduler. This is due to the fact that the self-similarity of the downlink interference emanates from the traffic characteristics of the realtime calls that have heavy tail call durations.

We study the effect of utilizing the proposed method on the packet drop ratio (PDR) of data traffic. Fig. 7 illustrates the PDR versus packet arrival rate of best-effort data traffic. In each case, we consider two traffic patterns: the first traffic is the one we have used earlier in this section, and the second traffic corresponds to a system with higher self-similarity index ($H = 0.7$) in downlink interference. As it is seen, the PDR of the system using the proposed method is smaller than the system using the average value of measured interference. Fig. 7 also shows the improvement in PDR with the higher level of self-similarity in the downlink interference. This observation shows the ability of our proposed method to exploit temporal correlation in the downlink interference to improve system performance.

V. Conclusions

In this paper, we developed a cross-layer model for downlink interference in heterogenous DS-CDMA wireless cellular network. In this model, interference is described as a function of application layer parameters (traffic characteristics) and physical layer variations (channel characteristics). We showed that for a heterogenous service DS-CDMA network downlink interference is a second-order self-similar process and has long-range dependence. We then used the predictive structure

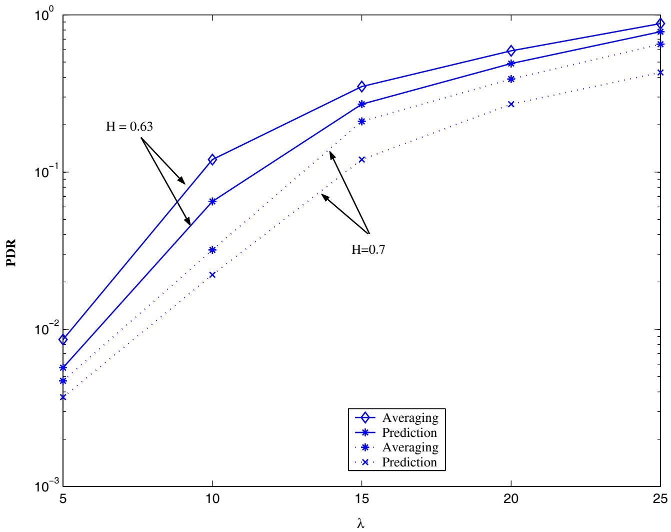

Fig. 7. Packet drop ratio of best-effort data traffic versus packet arrival rate λ for the proposed technique and a system that in each time instance uses the average value of 5 last samples of measured interference and pure CDMA.

of total downlink interference to maximize non-realtime data throughput. We used fGn to model the self-similarity of downlink interference. In the proposed method, the base-station uses an optimal linear predictor, based on the fGn model, to estimate the level of interference. The estimated interference is then used to allocate power to users. To maximize data throughput, we use time domain scheduling. The simulation studies showed the self-similarity of the total interference and validated the fGn model of downlink interference. The simulation results also showed a substantial performance improvement using the proposed predictive-adaptive scheme and confirmed that after applying the proposed method the interference model is still valid.

REFERENCES

[1] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. J. Weaver, and C. E. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 303–312, May 1991.

[2] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine*, pp. 78–80, Oct. 2003.

[3] J. Zhang, M. Hu, and N. B. Shroff, "Bursty data over CDMA: MAI self similarity rate control and admission control," *in Proceedings of IEEE INFOCOM'02*, pp. 391–399, 2002.

[4] J. Zhang and T. Konstantopoulos, "Multi-access interference process is self-similar in multimedia CDMA cellular networks," *preprint*, 2002. [Online]. Available: http://www.eas.asu.edu/∼junshan/pub/

[5] K. Navaie, S. Valaee, A. R. Sharafat, and E. S. Sousa, "On the downlink interference in heterogeneous wireless DS-CDMA networks," *to Appear in IEEE Transactions on Wireless Communications*.

[6] J. Beran, *Statistics for Long-Memory Processes*. NewYork: Chapman & Hall, 1994.

[7] F. Berggren, S. L. Kim, R. Jäntti, and J. Zander, "Joint power control and intra-cell scheduling of DS-CDMA non-real time data," *IEEE Journal on Selected Area in Communications*, vol. 19, pp. 1860–1870, Oct. 2001.

[8] H. Holma and A. Toskala, *WCDMA for UMTS: Radio access for third generation mobile communications*. John Willy and Sons, 2000.

[9] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Addison-Wesley, 1995.

[10] A. Erramilli, M. Roughan, D. Veitch, and W. Willinger, "Self-similar traffic and network dynamics,," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 800–819, May 2002.

[11] G. L. Stuber, *Principles of Mobile Communication*. Kluwer Academic Publishers, 1996.

[12] B. Tsybakov and N. D. Georganas, "Self-similar processes in communications networks," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1713–1725, Sep. 1998.

[13] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, 1991.

[14] M. Roughan, P. Abry, and D. Veitch, "Real-time estimation of the parameters of long-range dependence," *IEEE/ACM Transactions on Networking*, vol. 8, pp. 467–476, Aug. 2000.

[15] H. Harada and R. Prasad, *Simulation and software radio for mobile communications*, ser. The Artech House universal personal communications series. Artech House, 2002.

[16] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.