PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Cascaded neural networks for sequenced propagation estimation, multiuser detection, and adaptive radio resource control of third-

William S. Hortos

SPIE.

Cascaded neural networks for sequenced propagation estimation, multiuser detection, and adaptive radio resource control of third-generation wireless networks for multimedia services

William S. Hortos

Florida Institute of Technology, Orlando Graduate Center, 3165 McCrory Place, Suite 161, Orlando, FL 32803

ABSTRACT

A hybrid neural network approach is presented to estimate radio propagation characteristics and multiuser interference and to evaluate their combined impact on throughput, latency and information loss in third-generation (3G) wireless networks. The latter three performance parameters influence the quality of service (QoS) for multimedia services under consideration for 3G networks. These networks, based on a hierarchical architecture of overlaying macrocells on top of micro- and picoells, are planned to operate in mobile urban and indoor environments with service demands emanating from circuit-switched, packet-switched and satellite-based traffic sources. Candidate radio interfaces for these networks employ a form of wideband CDMA in 5-MHz and wider-bandwidth channels, with possible asynchronous operation of the mobile subscribers.

The proposed neural network (NN) architecture allocates network resources to optimize QoS metrics. Parameters of the radio propagation channel are estimated, followed by control of an adaptive antenna array at the base station to minimize interference, and then joint multiuser detection is performed at the base station receiver. These adaptive processing stages are implemented as a sequence of NN techniques that provide their estimates as inputs to a final-stage Kohonen self-organizing feature map (SOFM). The SOFM optimizes the allocation of available network resources to satisfy QoS requirements for variable-rate voice, data and video services. As the first stage of the sequence, a modified feed-forward multilayer perceptron NN is trained on the pilot signals of the mobile subscribers to estimate the parameters of shadowing, multipath fading and delays on the uplinks. A recurrent NN (RNN) forms the second stage to control base stations' adaptive antenna arrays to minimize intra-cell interference. The third stage is based on a Hopfield NN (HNN), modified to detect multiple users on the uplink radio channels to mitigate multiaccess interference, control carrier-sense multiple-access (CSMA) protocols, and refine call handoff procedures. In the final stage, the Kohonen SOFM, operating in a hybrid continuous and discrete space, adaptively allocates the resources of antenna-based cell sectorization, activity monitoring, variable-rate coding, power control, handoff and caller admission to meet user demands for various multimedia services at minimum QoS levels.

The performance of the NN cascade is evaluated through simulation of a candidate 3G wireless network using W-CDMA parameters in a small-cell environment. The simulated network consists of a representative number of cells. Mobile users with typical movement patterns are assumed. QoS requirements for different classes of multimedia services are considered. The proposed method is shown to provide relatively low probability of new call blocking and handoff dropping, while maintaining efficient use of the network's radio resources.

Keywords: Propagation channel estimation, multimedia services, multiuser detection, wireless communication networks, Kohonen self-organizing feature maps, Hopfield neural networks, multilayer perceptron neural networks

1. INTRODUCTION

The International Telecommunications Union (ITU) has developed requirements for the next generation of mobile communication networks to provide anywhere, anytime, bandwidth-on-demand multimedia services to users. These services include toll-quality voice, variable-rate video, and high-speed data of 144 and 384 kilobits per second (kbps) for high-mobility users with wide-area coverage and 2 megabits per second (Mbps) for low-mobility users with small-cell coverage. The set of ITU requirements are called IMT-2000. Since current systems of cellular and PCS digital services are considered second-generation networks, the IMT-2000 requirements have been formulated for third-generation (3G) wireless networks.

Part of the SPIE Conference on Applications and Science of Computational Intelligence II
Orlando, Florida • April 1999 SPIE Vol. 3722 • 0277-786X/99/$10.00

Prominent in the radio interface design of leading IMT-2000 proposals is wideband direct-sequence (DS), code division multiple access (CDMA), which is based on a coded spread spectrum technique. One leading proposal, called cdma2000, has been submitted by the CDMA Development Group (CDG) and the Telecommunications Industry Association (TIA) in North America. The other leading proposal, W-CDMA, is promoted jointly by ARIB in Japan and the European Telecommunications Standards Institute (ETSI) in Europe.

The main objectives of this paper are the estimation and enhancement of the system performance in proposed 3G DS-CDMA wireless networks for integrated multimedia services. The approach to these objectives is to use radio resource allocation (RRA) to effect interference diversity to reduce variance, thereby averaging out the fluctuations in order to increase channel capacity subject to quality of service (QoS) requirements. There is an opportunity to use neural network (NN) techniques in interference-cancelling receiver design as conventional matched filters are often inefficient and offer suboptimal performance in multiuser detection (MUD) and other near-far resistant receivers are too complex to implement. Since a DS-CDMA system is interference limited, properly designed interference cancellation methods will improve capacity. As a direct consequence, adaptive power control and other interference mitigation techniques based on NN techniques in DS-CDMA cellular radio systems are applied to improve signal-to-interference ratio (SIR). The effect of interference variation on QoS of integrated systems for services with different rates and powers has only recently been considered, but not in great depth.

2. FEATURES OF THIRD-GENERATION DS-CDMA NETWORKS

The 3G air interface proposals based on CDMA focus on two main types, network asynchronous and synchronous. In the former type, the base stations (BSs) are not synchronized, while in the latter they are synchronized within a few microseconds. The W-CDMA system proposed by ARIB and ETSI is an asynchronous network. The limited scope of this work restricts the focus to salient features of W-CDMA uplinks and to NN techniques that can enhance their performance.

The W-CDMA radio link offers variable bandwidths of 1.25, 5.0 MHz and higher multiples of 10 and 20 MHz in later extensions. At the time of this writing, chip rates are set at 1.024, 3.840 Mcps, 2×3.840 Mcps, and 4×3.840 Mcps. W-CDMA employs long spreading codes.[1] On the forward and reverse links use is made of variable-length orthogonal sequences as channelization codes. A short variable-length Kasami code is proposed for the uplink to facilitate implementation of the MUD. On the uplink, W-CDMA employs time-multiplexed pilot symbols for coherent detection. Since the pilot symbols are user-dedicated, they can be used for channel estimation with adaptive antennas as well.

The W-CDMA traffic channel structure is based on a single-code transmission for low data rates and multicodes for higher data rates. Multiple services belonging to the same connection are, in normal cases, time-multiplexed in stages. Time multiplexing takes place after both outer coding and inner coding, the multiservice data stream is mapped to one or more dedicated physical data channels. In the case of multicode transmission, every other data channel is mapped into the quadrature (Q) channel and every other into the in-phase (I) channel.

W-CDMA has two different types of packet data transmission possibilities. Short data packets can be appended directly to a random access burst. This method, called *common channel packet transmission*, is used for short, infrequent packets, where the link maintenance for a dedicated channel (DCH) would cause an unacceptable overhead. Longer, more frequently transmitted packets are sent on DCHs. A large single packet is transmitted using a single-packet scheme, where the DCH is released immediately after the packet has been transmitted. In a multipacket scheme the DCH is maintained by transmitting power control and synchronization information between subsequent packets. The W-CDMA random access burst is 10 ms long and transmitted with fixed power, and the access principle is based on the slotted Aloha scheme. Data arrives on the transport channel in the form of *transport blocks*. A variable number of transport blocks arrive on each transport channel at each transmission time instant. The *transmission time interval* is restricted to the set {10, 20, 40, 80 ms}.

A key feature of the W-CDMA air interface is the ability to transport multiple parallel services (transport channels) with different QoS requirements on one connection. Parallel transport channels are separately channel-coded and interleaved. The coded transport channels are then time-multiplexed into a coded composite transport channel. Different coding and interleaving schemes can be applied to a transport channel depending on the specific QoS requirements for error rates, delay, etc. Rate matching is applied in order to match the bit rate of the coded composite transport channel to one of the limited set of bit rates of the physical channels. Static rate matching is distributed between the parallel transport channels so that the transport channels fulfill their QoS requirements at approximately the same channel signal-to-interference ratio (SIR). For example, rate 1/3 convolutional coding is typically applied for low-delay services with moderate bit-error rate requirements (BER = 10^{-3}), while a concatenation of rate 1/3 convolutional coding and outer Reed-Solomon coding plus interleaving can be applied for high-quality services (BER = 10^{-6}).

3. QUALITY OF SERVICE REQUIREMENTS FOR WIRELESS MULTIMEDIA SERVICES

3G multimedia services can be classified into two main categories: real-time mode and packet data mode. Real-time services can be variable-rate, such as, the 8 kbps and 13 kbps voice codecs used in IS-95. In real-time mode, a large amount of digitized information is transmitted over a relatively long duration, whereas packet-mode services are provided to bursty information sources characterized as on-off processes. For packet data services, transmission stops at the end of the data burst, since no information is generated during the unpredictable off intervals. Real-time services may be selected as constant bit-rate (CBR) or variable bit-rate (VBR), and transmission is continuously maintained during the call. Packet data services are provided to users with demand for high transmission rates, but short service times. BER and delay requirements are necessary to ensure QoS.

3.1. W-CDMA operation for multimedia services

In W-CDMA, a number of options are available to integrate multi-rate services: (1) trade off processing gain for an increased information rate in the same spread bandwidth and (2) pair up basic data channels until the required information rate is obtained. The phrase "basic channel" refers to the CBR transmission with the highest processing gain. The radio resource controller fully controls the choice of appropriate coding scheme, interleaving parameters, and rate-matching parameters.

The media access controller (MAC) must be able to support a mixture of services. The MAC protocol controls the data stream delivered to the physical layer over the transport channels. If an MS wants to transmit data of different services, e.g., a real-time service and packet data, it is assigned two sets of transport formats, one for the real-time service and one for the packet service. As for a single service, the MS may use any transport format assigned for real-time services, whereas it may only use the transport formats for the data service. The MS is assigned a specific output power/rate threshold. The aggregate output power/rate will never exceed the threshold. Thus, the transport formats used for data service fluctuate adaptively to the used transport formats of the speech service.[2] One proposed handoff approach dynamically adapts the amount of the RRs based on the current network conditions, that is, on the average connection-dropping probability and the utilization of RR reserves, to improve the RR resource utilization and the blocking probability.[3]

3.2. Models of interference, multimedia service demand, and radio resource allocation

The following conditions express the interference and QoS constraints in the operation of wireless multimedia networks. The dimensions of the model have grown from the exposition in [4] due to the increase in the number of radio resource categories available in W-CDMA to support the demand for simultaneous multiple services.

1. The co-channel constraint (CCC) is that the same transport (physical-layer) channel cannot be assigned simultaneously to certain pairs of mobile users in the radio cells. The CCC is determined by the co-channel interference (CCI), which, in turn, depends on the interference control applied at the N base stations (BSs) of the network.
2. The adjacent channel constraint (ACC) is that channels adjacent in their domain's distance metric (frequency, time slot or PN code) cannot be assigned to adjacent radio cells simultaneously.
3. The co-site channel constraint (CSC) is that any pair of channels assigned to a radio cell must be at a minimum distance in their domain. In W-CDMA, this distance depends on the interference level produced by the adaptive antenna selection, activity monitoring, and power control and service classes active at each BS's coverage area.

The constraints have customarily been described by an $N \times N$ symmetric matrix, called the *interference matrix* $\boldsymbol{C}$ for single-service cellular networks. Each off-diagonal element c_{ij} in $\boldsymbol{C}$ represents the minimum separation distance between a channel assigned to radio cell (or sector) i and a channel assigned to radio cell (or sector) j. The CCC is represented by $c_{ij} = 1$, while the ACC is represented by $c_{ij} = 2$. Setting $c_{ij} = 0$ indicates that BSs i and j are allowed to assign the same channel to users in their service areas. Each diagonal element c_{ii} in $\boldsymbol{C}$ represents the minimum separation distance between any two channels assigned to cell (or sector) i. This is the CSC and $c_{ii} \geq 1$ is always satisfied, provided that, in sectored networks, sectors are equivalent to cells. In 3G wireless networks, the dimensions of the matrix increase to accommodate the number of physical-layer channels available to each BS to support multiple real-time and packet data services to each active MS in the system. Let C be the maximum number of physical-layer channels that can be supported at any BS. Then, the 3G W-CDMA interference matrix is an $N{\cdot}C \times N{\cdot}C$ symmetric matrix.

Since DS-CDMA capacity can only be increased by reducing other-user interference, this suggests a departure from the model of a two-dimensional interference matrix for N BSs assigning fixed M channels. In a W-CDMA network, each BS i is assumed to be able support $m_i^c(I)$ common channels and $m_i^d(I)$ dedicated channels, where $m_i^c(I) + m_i^d(I) = m_i(I)$ and

$m_i(I) \le M(I)$, the total number of channels in the network, and $m_i(I)$ represents the local channel capacity of the service area I and $M(I) \le N \cdot C$. As discussed by Gilhousen, *et. al.*[5], using adaptive antenna-array beamforming, voice activity monitoring, selectable spreading factor and coding rates, and uplink power control can regulate interference, I, in a W-CDMA network. Interference regulation determines the number of available channels. Viewing the adaptive elements as the RR controls of the network suggests representing their effect on network channel capacity as a composite mapping, $\Psi = \Phi \bullet \Gamma$, on a $4N$-dimensional lattice, $\Re : \Phi : \Re = (S \times V \times B \times P)^N \rightarrow I^N$; $\Gamma : I^N \rightarrow \boldsymbol{M} = \{\boldsymbol{m} : \boldsymbol{m} = (m_1, m_2, \ldots, m_N)$; m_i = # channels in the service area of BS $i\}$, where S represents the set of antenna-array cell sector values in any BS coverage area; V is the set of states of voice activity monitoring in each area; B is the set of spreading factors (4 – 256) available and coding rates (1, 2, 4, 8, 13, 32 kbps); P is the set of discrete power control levels to the mobile subscribers (0 – 10 dB, in 0.25 dB steps); I is the interval on the real line bounding interference levels, while $\boldsymbol{M}$ is the subset of the set of N-dimensional vectors of non-negative integers, whose i'th component is the total channel capacity at BS i. The composite mapping relates the RRA to channel capacity, through the interference level that the assignment generates in the network cells.

For a single-service DS-CDMA network, the traffic demand for physical-layer channels in each BS coverage area, in a network of N BSs, is represented by an N-vector called the *traffic demand vector* $\boldsymbol{T}$.[4] Simultaneous multimedia services increase the dimensions of traffic demand to form an array $\boldsymbol{T}$, where each row vector $\boldsymbol{t}_i = (t_{i1}^{rt}, t_{i2}^{rt}, \cdots, t_{iK_i}^{rt}; t_{i1}^{p}, t_{i2}^{p}, \cdots, t_{iL_i}^{p})$, with t_{ij}^{rt} (t_{ij}^{p}) the number of units of real-time (packet-data) service class j from both new calls and handoffs that are assigned to cell i. The t_{ij}^{rt} and t_{ij}^{p} are nonnegative with a zero value indicating no service demands of class j at cell i. Let q_{ikl} denote the physical-layer channels to support service class l units of the k'th active call assigned to cell i. Then, the interfering channel constraints described by the interference matrix can be expressed

$$|q_{ikl} - q_{jmn}| \ge c_{ij}, \text{ for } i \ne j,\ 1 \le i, j \le N \cdot C;\ k \ne m,\ 1 \le k \le t_{il}^{u},\ 1 \le m \le t_{jn}^{u}, \text{ where } u = rt \text{ or } p. \quad (1)$$

Since the same physical-layer channel cannot be used simultaneously in two interfering cells (intercell) or by two interfering users (intracell), interference conditions have been considered hard constraints. With the introduction of adaptive RR, the constraints can be considered "soft" to the limits imposed by bounds on the resource sets. When service demands cannot be satisfied due to a constraint, the corresponding request for a new call or a call handoff is blocked. For this reason, the joint probability of call blocking or handoff dropping is a useful performance metric for an adaptive RRA algorithm.

4. THE CASCADE OF NEURAL NETWORKS

NN estimation techniques can be applied to multipath fading, imperfect power control, and non-uniform traffic. Multipath fading is a major impairment to CDMA operation, since each additional path adds extra interference to degrade system performance.[6] To support integrated multimedia services, the multiaccess interference (MAI) at the BS requires mitigation to meet QoS requirements.[7] Reduction in MAI can greatly increase link capacity. A candidate technique is the interference canceller; another is the adaptive antenna array, viewed as adaptive cell sectorization. A third is MUD, based upon NN techniques, such as reduced-complexity radial basis functions (RBFs) or Hopfield NNs (HNNs). Interference cancellers can be essentially classified as either single-user or multiuser. The former reduces MAI using a linear filter in one instance, and is simpler to implement than the latter. W-CDMA uses long random spreading code sequences on the uplink. The time-varying nature of the spreading code sequence, when observed over every symbol period, excludes adoption of single-user interference cancellers. One MUD receiver is the decorrelating receiver. This requires very complex computation of the inverse correlation matrices among different user's spreading codes and is considered impractical. On the other hand, the multi-stage IC version of the non-linear replica generation type is attractive, since interference replica generation and subtraction is performed successively for different users.[8] An issue for MUD receivers is the knowledge required on users' parameters, e.g., time delays, signal strength, etc. Accurate channel estimation is needed to generate the interference replica of each user. The proposed multi-stage interference canceller extends the structure of the coherent multi-stage IC (COSMIC).[9]

Feedforward multilayer perceptron (MLP) neural networks (NNs) are proposed to estimate signal strength, fade rate, shadowing standard deviation, and principal path coefficients on the uplinks, based on the received pilot signals from the mobiles. Outputs from the MLP-NN together with the service demands of the active calls and handoffs are input to a second-stage recurrent neural network (RNN) for adaptive antenna array control. The RNN produces estimates of the number and type of beams or sectors, denoted $\hat{s}_i$, and the directed gain of each active element at BS i, $\hat{g}_{ik}$, $k = 1, \ldots, \hat{s}_i$. These estimates along with service demands of calls and handoffs are input to the third stage, an MUD receiver based on a discrete-form HNN to reduce MAI. The reduced interference then eases the performance requirements for stringent power control, antenna

control, activity monitoring, and code and spreading factor assignments to support the QoS of multimedia service demands. Reduced interference increases the reserve of available RRs, thereby reducing the possibility of new call blocking; unnecessary handoffs, both soft and hard; and dropped handoffs to other BSs. Excessive call handoffs are a major component of end-to-end connection latency that impacts the QoS of the real-time services. The reduced MAI estimates, antenna-array sector indices and gains, generated in the first three NN stages, and the multimedia service demands, are collectively input to a modified Kohonen self-organizing feature map (SOFM). The SOFM determines the best RRA array, mapping through the estimated residual MAI profile to meet the current service demands. This is shown in Figure 1.

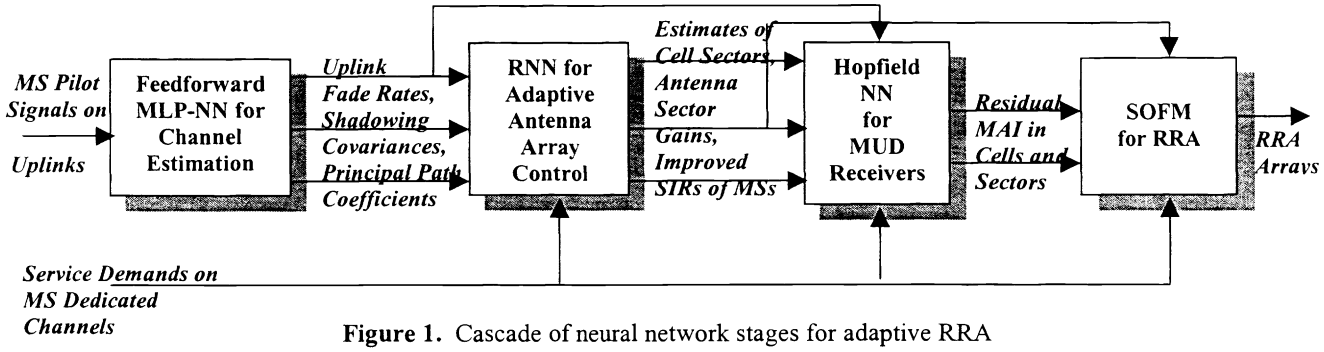

Figure 1. Cascade of neural network stages for adaptive RRA

4.1. Multilayer perceptron neural network for propagation channel estimation

The first stage of the cascade is a fast NN method for uplink propagation channel estimation in dense urban environments. The method is an extension based on the work in [10]. The performance of the technique is compared to that of COST 231 models [11], Walfisch-Bertoni model [12], and the Saunders-Bonar model [13] in terms of estimation error and computation time.

Propagation losses between two points in an urban environment can expressed as a sum of free-space path losses L_0, which depend on frequency f and distance d and an attenuation term which accounts for the effect of shadowing from obstructions such as buildings. The factor ρ is fade rate determined by the multipath fading in the given operating environment and σ is the excess attenuation term due to shadowing from environmental obstructions:

$$L = L_0 + \sigma_{obs} = 32.4 + 20\log[f(\text{MHz})] + 2\rho\log[d(\text{km})] + \sigma_{obs} \quad (2)$$

The attenuation term is a function of the heights and spatial distribution of the buildings and other man-made obstructions between the MS transmitter and BS receivers. Traditionally, this function is calculated using electromagnetic (EM) theory.[12, 13] EM models provide accurate predictions, but suffer from long computation times. The model of the uplink between MS j to BS i is given by the transfer function, $\lambda_{ji}(z) = \sum_{l=0}^{U} \lambda_{jil} z^{-l}$, where U, the number of principal paths, is typically less than 8.

The MLP-NN has been trained and tested with actual BS site measurements taken in Munich as well as simulated pilot signals varied over W-CDMA design parameter values. MLP-NN predictions reported by [10] reveal a mean error in the test sets of –2.1dB and 0.44 dB and a standard deviation of error of 6.3 dB and 6.6 dB, respectively. These results have at least 50% lower standard deviation compared to the predictions given in the COST 231 final report.[14] Continuous operation of the BSs allow frequent updates of the MLP-NN estimates to improve accuracy and to allow adaptation to changing conditions.

The MLP-NN approach also reduces the computation time of channel predictions in cellular network planning. Propagation losses have been predicted for a 3392 × 2400 m^2 urban area with a resolution of 4 × 4 m^2 using the three models cited above and the MLP-NN.[10] The results indicate for this area size the MLP-NN is at least four times faster than any cited methods. Further speed improvements can be achieved by implementing the NN in parallel analog hardware and by scaling its operation over smaller 90 × 90 m^2 picocells with a resolution of 1 × 1 m^2 to achieve at least 62 times reduction

4.2. Recurrent neural network for adaptive antenna control

The second NN stage implements space- and time-diversity combining of the mobile signals on the uplink via adaptive antenna arrays using a recurrent NN (RNN) technique to offer better performance with lower computational complexity than adaptive arrays based on square-root recursive least squares (RLS) schemes.[15]

The RNN method is evaluated for a QPSK/W-CDMA system with L receiving antenna elements, over time-varying multipath channels. Following QPSK demodulation in the BS receiver, the mobile signals are despread with matched filters to obtain complex signals $\hat{x}_1(n),\ldots,\hat{x}_L(n)$. Since the RNN accepts only real inputs, complex signals are formatted into in-phase (I) and quadrature (Q) components for input to the RNN. Thus, for L receiving antennas, the RNN has $L_0 = 2L$ external inputs and $L_1 = 2$ fully interconnected neurons, providing I and Q output signals. The neuron output at time $n+1$ depends on the external inputs $u_l(n)$ at the previous time instant and the previous outputs of the neurons $v_l(n)$, described by the following:

$$s_k(n+1) = \sum_{l=1}^{L_1} w_{kl}(n) v_l(n) + \sum_{l=1}^{L_0} w_{k,l+N_1}(n) u_l(n) \tag{3}$$

$$v_k(n+1) = f(s_k(n+1)) \tag{4}$$

where $w_{kl}(n)$ is the weight of the connection from the l'th input to the k'th neuron, and $f(\cdot)$ is the sigmoid function.

The most widely known algorithm for training of the RNNs is the real-time recurrent learning (RTRL) algorithm, which updates the RNN weights according to the following rule.[16] For $i, k = 1, \ldots, L_1$ and $j = 1, \ldots, L_0+L_1$,

$$w_{ij}(n+1) = w_{ij}(n) + \alpha \sum_{k=1}^{L_1} e_k(n+1) p_{ij}^k(n+1) \tag{5}$$

$$p_{ij}^k(n+1) = f'(s_k(n+1)) \left[\sum_{l=1}^{L_1} w_{kl} p_{ij}^l(n) + \delta_{ik} u_j(n) \right] \tag{6}$$

where α is the learning gain and $e_k(n) = d_k(n) - v_k(n)$ is the error at the k'th neuron, d_k is the desired output, δ_k is the Kronecker delta , and $f'(\cdot)$ the derivative of the sigmoid $f(\cdot)$. Algorithm "sensitivity" is defined as $p_{ij}^k = \Delta v_k / \Delta w_{ij}$.

The RNN is trained with random pilot symbols, with weights w_{ij} initialized to random values satisfying $|w_{ij}| < 10^{-2}$.[15] The learning gain α values for the RTRL algorithm are selected between 0.04 and 0.1 after heuristic optimization.[15] Following training, the RNN is set to decision-directed mode to track the channel variations and correct for distortions on the transmitted pilot signals.

The performance of the RNN array has been compared to that of the linear adaptive array structure trained with the RLS algorithm in an IS-95 CDMA network.[15] The linear structure has a two-tap FIR filter with complex coefficients in each of its antenna branches, and a training period of 200 pilot symbols. With seven co-channel interferers, at a BER of 10^{-3}, the RNN structure with two receiving antennas performs 3 dB better than the RLS technique. With four receiving antennas the improvement increases to 4 dB. A six-element RNN array performs slightly better than the RLS technique, giving a maximum improvement of 2 dB at a BER of 10^{-5}. The comparative BER performance of the RNN and RLS techniques is compared as the number of mobile users varies.[15] With SNR fixed at 14 dB in an IS-95 system and four receiving antennas, the RNN performs about seven orders of magnitude better than the RLS technique for a single user. As the number of users increases, the relative advantage of the RNN over the RLS decreases to about 2.5 orders of magnitude for 16 users. The result shows that, while RNN arrays have an advantage for channels dominated by multipath fading, only smaller improvements are obtained in interference-dominated channels. This is the rationale for high-level interference mitigation in the third-stage MUD NN of the cascade.

4.3. Hopfield neural networks for joint multiuser detection

When using DS-CDMA each transmitter modulates a different signature signal waveform, known to the receiver. The received signal is the superposition of the signals transmitted by each individual MS. As shown by Verdu, in both the synchronous and the asynchronous transmission cases, optimal multiuser detection (OMD) is an NP-hard problem that is equivalent to the maximization of an integer quadratic objective function.[17] Research efforts have focused on the development of suboptimal receivers, that are near-far resistant, have reasonable computational complexity, with performance comparable to the OMD receiver. Both Mitra and Poor in [18, 19] and Mulgrew in [20] have proposed receivers based on radial basis functions (RBFs) whose output is a linear combination of nonlinear functions, each of which is applied to the vector input data. These RBF receivers are useful for decentralized detection of a single-user in multiuser channels. They perform well for a small number of synchronous users, but training time is exponential in the number of users.

Kechriotis and Manolakos have introduced the design of a single-layer feedback NN receiver with O(K) neurons capable of demodulating the information transmitted by K synchronous or asynchronous users, sending CDMA packets over the same nearly Gaussian channel. OMD can be formulated as an energy minimization problem and thus be solved in practically constant time using an analog VLSI-implemented HNN.[21, 22] Simulation suggests the HNN detector outperforms the conventional matched-filter detector to attain near-optimal BER performance with lower complexity than the RBF detectors.

If the coded waveforms assigned to each user are orthogonal and the transmitted signals are antipodal ($\{+1,-1\}$), then the conventional detector (CD) can recover information bits by first passing the received signal through a bank of filters matched to the users' signature waveforms, and next deciding on the information bits based on the sign of the output. However, the major limitation of the CD is that its performance degrades severely when powers of the transmitting users are dissimilar.

Assume K active transmitters share the same channel at a given time. A signature waveform $s_k(t)$, limited to $t \in [0,T)$, is assigned to each transmitter. Denote the i'th information bit of the k'th user as $b_k^{(i)} \in \{+1,-1\}$. In a DS-CDMA system, the signal at a receiver is the superposition of K transmitted signals and additive noise. Each $s_k(t)$ is the convolution of the transmitted MS traffic channels and the mulitpath-channel transfer function, estimated in the first MLP-NN stage.

$$r(t) = \sum_{i=-P}^{P} \sum_{k=1}^{K} b_k^{(i)} s_k(t - iT - \tau_k) + n(t),\ t \in \boldsymbol{R} \tag{7}$$

In (7) $\tau_k \in [0,T)$ are the relative time delays between the users and $2P+1$ is the packet or frame size. In those systems where the BSs cooperate to maintain synchronism, $\tau_k = 0,\ k = 1,\ldots,K$. In a CD a simple thresholding device produces an estimate $\hat{b}_k^{(i)}$ for the i'th information bit of the k'th user based on the sign of the i'th output of the k'th matched filter:

$$y_k^{(i)} = \int_{iT-\tau_k}^{(i+1)T-\tau_k} r(t) s_k(t - iT - \tau_k)dt,\ \ \bar{b}_{CD}^{(i)} = sign(\bar{y}^{(i)}) \tag{8}$$

where $\bar{y}^{(i)} = [y_0^{(i)}, y_1^{(i)}, \cdots, y_{K-1}^{(i)}]$. In the OMD, an estimate is produced for the information vector transmitted at the discrete time instant i, based on the maximization of the logarithm of the likelihood function. In the synchronous case, it holds that:[23]

$$\bar{b}_{OMD}^{(i)} = \arg \max_{\bar{b} \in \{+1,-1\}^K} \{2\bar{y}^{(i)T}\bar{b} - \bar{b}^T H \bar{b}\} \tag{9}$$

where $H \in \boldsymbol{R}^{K \times K}$ is the symmetric matrix of signal cross-correlations:

$$h_{kl} = \int_0^T s_k(t) s_l(t)dt \tag{10}$$

A detection scheme, suboptimal to the OMD, with low computational complexity called the multistage detector (MSD) has been proposed.[24] The MSD consists of a sequence of stages m = 1, 2, ..., each producing an estimate $\bar{b}_{MSD}^{(i)}(m)$ given as:

$$\bar{b}_{MSD}^{(i)}(m+1) = sign\left(\bar{y}^{(i)} - (H - E)\bar{b}_{MSD}^{(i)}(m)\right) \tag{11}$$

where E is a diagonal matrix with elements $e_{ii} = \int_0^T s_i^2 dt$ (signal energies). The output of the first stage (m=1) is initialized to the estimate of the CD. The MSD is insensitive to the near-far problem. In the *asynchronous* case, the OMD problem is written in the form of (9), defining matrices $H(j) \in \boldsymbol{R}^{K \times K}$, $j = -1,\ 0,\ 1$ as

$$h_{kl}(j) = \int_{-\infty}^{\infty} s_k(t - \tau_k) s_l(t + jT - \tau_l)dt \tag{12}$$

and the matrix $\tilde{H} \in \boldsymbol{R}^{(2P+1)K \times (2P+1)K}$ as

$$\tilde{H} = \left[\begin{array}{c|ccccc} H(0) & H(-1) & 0 & \cdots & 0 \\ H(1) & H(0) & H(-1) & & \vdots \\ 0 & H(1) & H(0) & \ddots & 0 \\ \vdots & & \ddots & \ddots & H(-1) \\ 0 & \cdots & 0 & H(1) & H(0) \end{array}\right]. \tag{13}$$

Then, the optimum receiver for the asynchronous case is viewed as a larger combinatorial optimization problem of the form

$$\tilde{\bar{b}}_{OMD}^{(m)} = \arg \max_{\bar{b} \in \{+1,-1\}^{(2P+1)K}} \{2\tilde{\bar{y}}^{(im)^T} \tilde{\bar{b}} - \tilde{\bar{b}}^T \tilde{H} \tilde{\bar{b}}\} \tag{14}$$

where $\tilde{\bar{y}}^{(m)^T}$ is the row vector consisting of the sampled outputs of the matched-filter bank corresponding to the m'th packet. If the packet length is relatively large, even a small number of users causes a restrictive computational effort to solve (14).

HNNs are single-layer networks with output feedback consisting of simple processors (neurons) where the connection between two processors is established through a conductance T_{ij} that transforms the voltage outputs of amplifier j to a current input for amplifier i. Externally supplied bias currents I_i are also input to every neuron j. Each neuron i receives a weighted sum of the activations of other neurons in the network, and updates its activation according to the rule: [25]

$$V_i = g(U_i) = g\left(\sum_{j \neq i} T_{ij} V_j + I_i \right) \tag{15}$$

where $g(U_i)$ can be an antipodal thresholding function resulting in $V_i = g(U_i) = sign(U_i)$. Hopfield has shown that, for symmetric connections ($T_{ij} = T_{jj}$), the equations (5) for the activation of the neurons always lead to convergence to a stable state.[25] Moreover, when the T_{ii} are zero and $g(\cdot)$ approaches the antipodal thresholding function, the stable states of the network of N neurons are local minima of the energy function given as:

$$E = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} V_i V_j - \sum_{i=1}^{N} V_i I_i \tag{16}$$

The cross-correlation matrix H is symmetric. Moreover, equation (9) can be rewritten as

$$\begin{aligned} \bar{b}_{OMD}^{(i)} &= \arg \min_{\bar{b} \in \{+1,-1\}^K} \{-\bar{y}^{(i)^T} \bar{b} + \tfrac{1}{2} \bar{b}^T H \bar{b}\} = \arg \min_{\bar{b} \in \{+1,-1\}^K} \{-\bar{y}^{(i)^T} \bar{b} + \tfrac{1}{2} \bar{b}^T (H - E) \bar{b} + \tfrac{1}{2} \bar{b}^T E \bar{b}\} \\ &= \arg \min_{\bar{b} \in \{+1,-1\}^K} \{-\bar{y}^{(i)^T} \bar{b} + \tfrac{1}{2} \bar{b}^T (H - E) \bar{b}\} \end{aligned} \tag{17}$$

since $\bar{b}^T E \bar{b}$ is always a positive number. The matrix $-(H - E)$ is symmetric and has zero diagonal elements. Therefore, the OMD objective function can be translated into the HNN energy in (16) with weight matrix $\boldsymbol{T} = -(H - E)$ and biases $\boldsymbol{I} = \bar{y}^{(i)}$.

The initial state of the HNN MUD coincides with the initial state of the CD. The energies of the active users are assumed to vary relatively slowly and can be estimated by the MLP-NN and RNN stages. The HNN weights can be preset according to the energies of the users and the known values of the cross-correlations of their signature waveforms. The discrete-time approximation of the equation of motion of the i'th neuron of the HNN is given by:

$$\Delta U_i = -\frac{U_i}{\tau} + \sum_{i \neq j} T_{ij} V_j + I_i \tag{18}$$

If $g(\cdot) = sign(\cdot)$, the dynamics of the i'th neuron at the instant $t = m + 1$, are described by the following:

$$V_i(m+1) = sign(U_i(m) - \frac{U_i(m)}{\tau} + \sum_{i \neq j} T_{ij} V_i(m) + I_i) \tag{19}$$

Setting $\tau = 1$ and substituting in equation (19) for the values of $\boldsymbol{T}$ and $\boldsymbol{I}$ for the proposed HNN detector, (19) becomes $V_i(m+1) = sign\left(y_i - \sum_{i \neq j} h_{ij} V_j(m)\right)$ which can be written in matrix form:

$$V(m+1) = sign(\bar{y} - (H - E) V(m)) \tag{20}$$

Computing (20) and (11) for the case where the RC constant $\tau = 1$ and $g(\cdot) = sign(\cdot)$, the estimate of the (m+1)'th stage of the MSD coincides with the output of the discrete-time approximation of this HNN at $t = m + 1$. Since the update of the estimate of each MSD stage is being performed synchronously, an infinite number of stages MSD is essentially equivalent to a discrete HNN operating in synchronous, fully parallel updating mode.[26] Under certain conditions, the HNN energy function can be shown to have a unique local minimum that coincides with the global minimum of the OMD problem.[27]

In the asynchronous W-CDMA transmission case, the dimension of the optimization problem grows dramatically with packet size and the number K of users. If K users transmit packets of length $2P + 1$, the corresponding HNN receiver will have $K \cdot (2P + 1)$ neurons. Due to sparsity of the matrix $\tilde{H}$, the number of interconnections required for the HNN is reduced

to $(3(2P+1)-2)\cdot K^2$. When the packet size is relatively small and K is small to moderate, an extended version of the HNN detector used in the synchronous case can be used. Computer simulations for $K = 3$ asynchronous users transmitting packets of length 31 bits have been performed.[21] Due to the large OMD detector size, computer simulations require long processing time. Comparisons are thus reported only with respect to the CD and HNN detector with $g(\cdot) = sign(\cdot)$ and RC = 1.0.

BER vs. SNR is compared for $K = 3$ asynchronous users using optimized Gold sequences of length $L = 127$. The energy of one of the users is 10 times larger than the energy of each of the other two users, so that the maximum near-far ratio is 10. Packet length is equal to 31 bits. The cumulative BER has been computed by simulating both the CD and HNN MUD for 10^7 transmitted sets of symbols for each value of SNR, drawn randomly from a uniform distribution.[21] Results for this case show the HNN detector to have uniform improvement in SNR over the BER range from 10^{-2} to 10^{-4} of 1 to 1.5 dB compared to the CD performance. During simulations, values of the delays τ_k, $k = 1, 2, 3$, are changed randomly every 500 symbols, so that the BER values represented the performance of the detectors averaged over all possible delays.

4.4. The self-organizing feature map for RRA

The idea of radio resources (RRs), e.g., transmit power, antenna array control, spreading factors, coding rates, transport channels, in a W-CDMA network, "competing" to be assigned calls suggests application of the SOFM approach. The approach modifies Kohonen's SOFM to solve discrete-space optimization problems among the lattice of RRs.[28] Development of a static RRA (SRRA) and extensions to dynamic RRA (DRRA) problems are discussed in previous work by the author.[4] All feasible solutions to the SRRA problem lie at the vertices of an n-dimensional hypercube, where $n = N{\cdot}R$, N is the number of BSs and $R = S{\cdot}V{\cdot}B{\cdot}P$ available RR combinations of antenna beams or sectors, activity monitoring, coding rates/spreading factors, and power control. Note that n is the dimension of $\Re$, the domain of Ψ. The image of the vertices also intersects the constraint hyperplane defined by the interference matrix, traffic demand array and channel constraints (1) due to the RRAs. Since each entry t_{ij}, of the traffic demand array $\boldsymbol{T}$ can be assumed integer-valued for all i, j, the image of the RR constraints set can be shown to form an integral polytope. Consider the neurons on this hypercube, defined as

$$X_{j,\boldsymbol{r}} = \begin{cases} 1, \text{ if coverage area } j \text{ is assigned radio resource vector } \boldsymbol{r} = (s, v, \beta, p) \\ 0, \text{ otherwise} \end{cases}$$

for j = 1, $\cdots$, N; and $\boldsymbol{r} \in S \times V \times B \times P$. Let $\boldsymbol{X}$ denote the n-dimensional array of these variables. Normalizing the range of values for each RR and the interference bounds to the interval [0,1], the set of RRs and its Γ-image in the interference range are each contained in unit hypercubes. A vertex is approached continuously from within the unit hypercube, starting from a point on the constraint hyperplane and inside the hypercube. This represents a feasible, non-integer solution to the RRA problem. The continuous variable approach in the interior of the hypercube is denoted by $w_{r,j}$, so that, for a quality metric Q, $Q(\boldsymbol{W}) = Q(\boldsymbol{X})$ at the vertices. The value $w_{r,j}$ represents the *probability* that the variable in the $\boldsymbol{r}, j$ position of the array $\boldsymbol{X}$ is activated. The vector $\boldsymbol{r}$ is integer-valued, an index into the lattice of allowable RRAs. Kohonen's self-organization is applied to the array of *synaptic weights*, $\boldsymbol{W}$. This modification permits the SOFM to solve discrete-space optimization problems.

The structure of the discrete-space SOFM consists of an input layer of N nodes, and an $R \times N$-dimensional array of output nodes. The output nodes correspond to the solution array of discrete-valued RRAs, while the input layer represents the N BS coverage areas in the W-CDMA network. The weight connecting input node j to node $\boldsymbol{r}$ of the output array of nodes is given by $w_{r,j}$. A cell in which an assignment of $\boldsymbol{r}$ is required is presented to the network through the input layer at node j. Physically, an incoming call or handoff is presented to the network at BS j. The nodes of the output layer compete with each other to determine which subarrray of the solution array to meet the QoS requirements of the input with minimal impact on the cost potential. The synaptic weights are then *adapted* to indicate the RRA decision using the neighborhood topology.

Consider the case where RRA $\boldsymbol{r}$ is required at BS j^*. An input vector $\boldsymbol{x}$ is presented to the network with a "1" in position j^* and 0 elsewhere. For each node $\boldsymbol{r} = (i_1, i_2, \cdots, i_R)$ of the outer layer, the value V_{r,j^*}, the cost to the objective function of RRA $\boldsymbol{r}$ to BS j^*, is computed. The *cost potential* V_{r,j^*} of node $\boldsymbol{r}$ for a given input vector $\boldsymbol{x}$ ($x_j = 0$, $\forall j \neq j^*$, $x_{j^*} = 1$) is defined by

$$V_{\boldsymbol{r},j^*} \equiv \sum_{i=1}^{N} \sum_{s \in \Re} P_{j^*,i,(\|\Psi(\boldsymbol{r})-\Psi(\boldsymbol{s})\|+1)} w_{i,s} \tag{21}$$

where the interference caused by the RRA is represented by the weight $P_{j,i,d+1}$, termed the proximity indicator where $d = \|\Psi(\boldsymbol{r}) - \Psi(\boldsymbol{s})\|$ is the distance in the service-channel capacity range between the images of RRAs $\boldsymbol{r}$ and $\boldsymbol{s}$. If

$\Psi(\boldsymbol{r}) = \Psi(\boldsymbol{s})$, then the interference cost should be at a maximum, with cost decreasing until the two active channels are sufficiently separated, so that interference, or contention for resources, is below a threshold value. The array $\boldsymbol{P}$ is defined as

$$P_{j,\,i,\,d+1} = \max\,(0,\, P_{j,\,i,\,d-1}),\; d = 1,\, \cdots,\, M-1; \quad P_{j,\,i,\,1} = c_{ji},\; \forall\, j,\; i \neq j; \quad P_{j,j,1} = 0,\; \forall\, j. \tag{22}$$

The *dominant node*, $\boldsymbol{m}_0$, of the outer layer is the node with minimum cost potential for a particular input vector. In the terminology of this representation $V_{\boldsymbol{m}_0 j^*} \leq V_{\boldsymbol{r},j^*}$ for all nodes $\boldsymbol{r}$ and fixed j^*. The *neighborhood* of the dominant node $\boldsymbol{m}_0$, is the set of nodes $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\cdots$, $\boldsymbol{m}_{\eta_{j^*}}$, ordered according to the values of the cost potentials, i.e., $V_{\boldsymbol{m}_0 j^*} \leq V_{\boldsymbol{m}_1 j^*} \leq V_{\boldsymbol{m}_2 j^*} \leq \cdots \leq V_{\boldsymbol{m}_{\eta_{j^*}} j^*}$, where η_{j^*} is the size of the neighborhood in the SOFM network for BS j^*. Thus, dominant nodes and their neighborhoods are determined by competition according to the objective function, and the weights are modified according to Kohonen's weight adaptation rules within the dominant neighborhood. The size of the dominant neighborhood depends upon which BS coverage area is receiving input as well as the variety and level of services that the input requires from the network.

When weight updating is complete, the array $\boldsymbol{W}$ has been moved in a direction that may be away from the constraint hyperplane, resulting in an infeasible solution. In the next step, the weights of the nodes outside the dominant neighborhood organize themselves around the modified weights, so that $\boldsymbol{W}$ remains a feasible solution to the RRA problem during the update. This step can be performed by a hill-climbing HNN or HC-HNN. Representing the weight matrix $\boldsymbol{W}$ as a vector $\boldsymbol{w}$, $\boldsymbol{w}$ is considered to the vector of states of a continuous HNN. The HNN performs random and asynchronous updates on $\boldsymbol{w}$, *excluding* the weights in the dominant neighborhood, to minimize the energy function:

$$E \equiv \left\| \boldsymbol{w} - (\wp\, \boldsymbol{w} + \tau) \right\|^2 \tag{23}$$

where $\wp$ is the projection onto the constraint hyperplane given by

$$\sum_{\boldsymbol{r} \in \Re} \boldsymbol{X}_{j,\boldsymbol{r}} = \boldsymbol{t}_j\,, \; \forall j = 1,\, \cdots,\, N. \tag{24}$$

and $\tau = (\mathbf{I} - \wp)\boldsymbol{T}$, where $\mathbf{I}$ is the identity operator. The energy function (23) is expressed in terms of a solution vector $\boldsymbol{x}$, constructed from the solution array $\boldsymbol{X}$, by ordering the elements $x_{i,k,\boldsymbol{r}}$ where $\boldsymbol{r} = (s, v, \beta, p)$, according to the ordering of four-integer indices and an ordering on the number of service classes k. The antenna-beam/cell-sector and power-control values, s and p, are initialized from the estimates output by the first and second stages of the NN cascade, then optimized in the SOFM based the residual interference levels produced by third-stage HNN MUD. In terms of $\boldsymbol{x}$, the demand constraints can be expressed as $\boldsymbol{Dx} = \boldsymbol{T}$, where $\boldsymbol{T}$ is the demand array and array $\boldsymbol{D}$ consists of N subarrays of 1's and 0's. The next random call and corresponding service requirements input to the SOFM network commences a new update period of the Kohonen algorithm, which determines a new dominant node and its neighborhood of nodes and modifies their weights. This procedure is repeated until the SOFM weights stabilize to a feasible 0-1 solution which is a local minimum of the optimal RRA problem. This process can be very slow.

As the algorithm converges, the magnitude of weight modifications and the size of the neighborhoods are decreased. Initially, the size of the neighborhood for each subarray of $\boldsymbol{W}$, given by $\eta = (\eta_1, \eta_2, \cdots, \eta_N)$ is large, but is decreased incrementally until $\eta_j = \left\| \boldsymbol{t}_j \right\|$, the total level of service demands at BS j, for all N stations. Since the weight modifications depend on the order in which the calls are input, the SOFM approach is inherently stochastic. In this form, the SOFM network must be run repeatedly to arrive at different local minima.

The following procedure for the SOFM algorithm can be applied to the SRRA problem in W-CDMA networks.

1. Initialize the weight vectors of the network as $\boldsymbol{w}_{j,\boldsymbol{r}} = \boldsymbol{t}_j / R$, which gives an initial feasible, possibly non-integer solution.
2. Randomly select a new call (with service demand k) for a BS. Represent this requirement as the input array $\boldsymbol{x}$. Find the position j^* (BS coverage area) which is active, i.e., $x_{j^*,k} = 1$.
3. Calculate the *quality or cost potential* $V_{\boldsymbol{r},j^*}$ for each index $\boldsymbol{r}$ in the output layer array according to (21).
4. Determine the dominant node, $\boldsymbol{m}_0$, by competition such that $V_{\boldsymbol{m}_0 j^*} = \min V_{\boldsymbol{r},j^*}$, $\forall \boldsymbol{r} \in \Re$, and identify its neighboring nodes $\boldsymbol{m}_1, \boldsymbol{m}_2, \cdots, \boldsymbol{m}_{\eta^*}$, where $\eta_{j^*} \geq t_{j^*}$ is the size of the neighborhood for input resource requirement at j^* for service class k.
5. Update the weights in neighborhood of dominant node according to the rule

$$\Delta \boldsymbol{w}_{j^*,r} = \alpha(\eta, \mathrm{n})[\boldsymbol{e} - \boldsymbol{w}_{j^*,r}] \quad \forall \boldsymbol{r} \ni V_{\boldsymbol{r},j^*} < V_{m_{\eta j^*} j^*} \text{ where}$$

$$\alpha(\eta, \mathrm{n}) = \frac{\alpha(\mathrm{n})\gamma_{j^*}}{\left\| \boldsymbol{t}_{j^*} \right\|} \exp\left[\frac{-\left| V_{m_0,j^*} - V_{r,j^*} \right|}{\sigma(\mathrm{n})} \right]$$

which is a modified version of Kohonen's SOFM *slow* updating rule, where α and σ are monotonically decreasing and positive functions of sampled time, γ is a normalized weighting vector used in tie-breaking for a network node. For all other weight vectors, outside the neighborhood being updated, $\Delta \boldsymbol{w}_{j^*,k} = 0$. The weights are updated as $\boldsymbol{w}_{j,r} \leftarrow \boldsymbol{w}_{j,r} + \Delta \boldsymbol{w}_{j,r}$.

6. The weights will no longer lie on the constraint hyperplane, so a hill-climbing HNN is applied to return to a feasible solution. The array $\boldsymbol{w}$ is modified around the weight adaptations of the SOFM into order that $\boldsymbol{D}\boldsymbol{w} = \boldsymbol{T}$.
7. Repeat Step 2 until RR requirements in all cells have been selected as input vectors to the SOFM network. This forms one period of the algorithm. The procedure is repeated for K periods. In each subsequent period, α and σ are decreased geometrically or according to any monotonically decreasing function.
8. Repeat Step 2 until $\left\| \Delta \boldsymbol{w}_{r,j} \right\| \cong 0, \forall \boldsymbol{r}, j$, this condition is considered stable convergence of the weights for a given neighborhood size. Decrease the neighborhood sizes η_j linearly for all j.
9. Repeat Step 8 until $\eta_j = \left\| \boldsymbol{t}_j \right\|$, for each BS coverage area j, $j = 1, \cdots, N$.

For the multiservice demand array $\boldsymbol{T}$, the normalized weighting vector γ, a heuristic used to damp oscillations in the algorithm updates is modified from the form used in [30] as follows

$$\gamma_i = \left| \sum_{j=1}^{N} \left(\sum_{k=1}^{K_i} t_{ik}^{rt} c_{ij} + \sum_{l=1}^{L_i} t_{il}^{p} c_{ij} \right) \right| - c_{ii} \ , \ i = 1, \ldots, N. \tag{25}$$

Each element in the vector γ is then normalized. Following [31], SOFM parameters can be selected heuristically, with $K = 10$,

$$\alpha(0) = \min_{1 \le i \le N, 1 \le k \le K_i, 1 \le l \le L_i} (t_{ik}^{rt}, t_{il}^{p}), \ \alpha(\mathrm{n}+1) = 0.9\alpha(\mathrm{n}),$$

$$\sigma(0) = 9, \ \sigma(\mathrm{n}+1) = 0.9\sigma(\mathrm{n}),$$

$$\eta_j(0) = \left\| \boldsymbol{t}_j \right\| + \lfloor N/K \rfloor, \quad \eta_j(\mathrm{n}+1) = \eta_j(\mathrm{n}) - 1.$$

The average probability of new call blocking, average probability of dropped handoffs and the total network capacity for each service class can be used alternately to evaluate the NN cascade as a SRRA algorithm in 3G networks. The evaluation criteria can be augmented with the terms $\kappa_j \cdot \boldsymbol{r}$, representing infrastructure costs of using RRA $\boldsymbol{r}$ in coverage area j.

The initial state of the DRRA problem is the stable SRRA solution, where a new call or handoff with multiservice demands cannot be assigned to a BS without a rearrangement of existing RRAs. A time-varying multiservice traffic demand array $\boldsymbol{T}(\mathrm{n})$ and the resource constraint relations are satisfied when $\boldsymbol{D}(\mathrm{n})\boldsymbol{x}(\mathrm{n}) = \boldsymbol{T}(\mathrm{n})$ at sample time n. Each epoch n represents the arrival of a single or multiple new calls to or handoffs between the cells of the network. During that period, input vectors, corresponding to the cell in which the call is placed or handoff requested, are presented to the SOFM network at a rate determined by the distribution of multiservice demands in the network at that time. Since feasibility is always restored during the second stage of the SOFM, any rearrangement of the existing calls to enable a new call is automatic. If no rearrangement is possible, either the SOFM cannot converge to a feasible set of RRAs, or a feasible rearrangement may be found by allowing the interference levels to increase above acceptable QoS levels. Higher interference levels, in turn, diminish capacity. In either outcome, the call can be blocked and the previous state of the system reinstated.

For more robust convergence, step 6 of the algorithm uses Abe's approach [32] to ensure that a HC-HNN only leads to feasible RRAs that are stable points of the system of update weights that now satisfy $\boldsymbol{D}(\mathrm{n})\boldsymbol{x}(\mathrm{n}) = \boldsymbol{T}(\mathrm{n})$ at sample time n. A piecewise-linear saturation function replaces the exponential of distance used in the weight update rule in step 5. In step 7, faster updating is accomplished based on Abe's convergence acceleration for HC-HNNs to optimize integration step sizes, now applied in only $K = 1$ period..[30] After the SRRA is completed to initiate the DRRA algorithm, step 8 is omitted. At the

start of each epoch in the DRRA SOFM, the neighborhood function is initially set to the row vector of greatest length in the multiservice demand array $\boldsymbol{T}(n)$ at sample time n.

5. SIMULATION RESULTS FOR MULTISERVICE RRA PROBLEMS

The performance of the NN cascade for RRA in W-CDMA networks can be evaluated, based on simulations of multimedia extensions of the cellular network models considered earlier by Kunz.[33] The interference matrices and traffic demand vector for a 25-cell network are used to represent a wireless multimedia network with non-homogenous traffic, by decomposing the number of calls at BS j in the demand vector into a row of the service demands in those calls from each real-time and packet-data subclass. Thus, the traffic demand vector becomes a multiservice traffic demand array. Time-varying traffic loading is approximated through cyclic rotation of the rows of the arrays $\boldsymbol{T}$ or periodic replacement of a selected row with a new traffic vector during the simulation run.

The computation time for the W-CDMA RRA simulations grows rapidly with the number of distinct service classes and the number of possible radio resource vector selections. For this reason, the simulation models are limited to four real-time service classes: digitized voice encoded with 8-kbps, 13-kbps, and 32-kbps codecs, together with 64-kbps compressed video; and four packet service classes with transmission rates of 64 kbps, 128 kbps, 384 kbps and 768 kbps. The radio resources form a finite set of vectors. The first entry in each vector is the selection of adaptive-antenna array sectorization or beamforming as any number of omni-, 180°-, 120°-, 45°-, 72°-, or 60°-sectors at each BS, such that the total coverage of the sectors equals 360°. The second entry is the selection of channel activity monitoring with 0 for "off" and 1 for "on". The third entry is the selection of spreading factor, based on 1.92-Mcps and 3.84-Mcps spread channels, of 4, 16, 64 and 256 with the use of rate matching assumed to align digitized-voice and packet-data rates with the chip rates. The fourth entry is the selection of power control at the mobile users, with 0-level or no power control, 4-level control, 20-level control, 40-level control over a 10-dB reference transmit power range. Each resource vector $\boldsymbol{r} = (s, v, \beta, p)$ is mapped to an estimated interference value I_r based on MAI statistics collected in microcellular networks and the minimum SIR values equivalent to the QoS required by the service classes active in the row corresponding to each BS in the traffic demand array. The interference establishes the actual CDMA frequency reuse performance, and thus determines the number of common and dedicated channels available to meet multiservice demands. Added to the objective function is the total infrastructure cost of the current network-wide RRA. The cost coefficient vector for each assignment $\boldsymbol{r}$ is $\boldsymbol{c} = (1, 1, 1, 1)$ and the individual RRA cost terms $\boldsymbol{r} \cdot \boldsymbol{c}$ are summed over the number of active BSs in the network and the number active service classes at those stations.

In order to exercise the three stages of MLP-NN channel estimation, RNN antenna array control, and HNN MUD, additional network features are given. The BS transceivers are assumed to use rate 1/3, constraint length $k = 9$ convolutional encoders. The multipath delay model is a 2- to 8-path profile, with the principal paths ordered by magnitude according to the recommended IMT-2000 channel propagation model. Each path is subjected to independent Rayleigh fading with power scaled to the IMT-2000 model and Doppler frequency of $f_D = 100$ Hz. The BER performance for the HNN MUD has been measured, in laboratory experiments using a channel fading emulator, as a function of the average E_b/N_0 as the number of active users varies. The average BER monotonically falls as the average E_b/N_0 increases, while that of the CD receiver approaches an error floor that depends on K the number of active users. As K increases, the E_b/N_0 loss from the single-user case increases due to residual MAI. When $K = 8$, however, E_b/N_0 loss at BER = 10^{-3} is only about 2.5–3 dB. The operation of the RNN antenna array control (AAC) is unlike a fixed multibeam antenna, considered earlier.[4] The RNN adaptive array can change the beam direction finely, initiates coverage from the omnibeam pattern and forms the optimum beam pattern adaptively, and can be designed to direct the beam toward the resolved path of each user and to realize coherent Rake combining even though their arrival angles are quite different. The values used for the RNN AAC are based on laboratory experiments using a channel fading simulator, assuming K users, one desired and $K - 1$ interfering users. For a comparison of RNN AAC and two-antenna diversity reception, the measured average BERs of the RNN AAC were determined as a function of received power ratio of interfering users to desired user. The data rate and chip rate are 64 kbps and 4.096 Mcps, respectively. The simulated paths of all mobile subscribers arrive from the same direction. The average E_b/N_0 and the arrival angle of the desired user were set to 11.7 dB and 62°, respectively. The average BER reduction was observed to be about one order of magnitude compared to the antenna diversity case, even if the worst-case interferer's power is 10 dB higher, corresponding to the situation of 10 interferers or a single user with 10 higher data rate. The improvement offered by the RNN AAC diminishes as the arrival angle of the interferers' signals approach that of the desired user's signal. When two users are within the beamwidth, either one could be blocked. These empirical results are used in the network simulations.

Simulations were performed on a PC, based on the adaptive learning models included in MATLAB's Neural Network Toolbox, with custom C routines to implement the network models as well as the MLP-NN, RNN, HNN and SOFM of the

NN cascade. Due to limitations of run time, once presented with the traffic demand array at each iteration of the simulation, the fours NN stages of RRA algorithm are run sequentially, with the outputs of the preceding stage used as inputs to the succeeding stages. Algorithm performance is measured alternately on the basis of average probability of call blocking, average probability of dropped handoffs, and the total active service classes in the network (channel capacity), together with the average number of iterations (ANIs) required for asymptotic convergence, based on a prescribed error value.

The traffic demand vector $[10, 11, 9, 5, 9, 4, 5, 7, 4, 8, 8, 9, 10, 7, 7, 6, 4, 5, 5, 7, 6, 4, 5, 7, 5]^T$, introduced by Kunz [33] for a single-service voice network of 25 base stations, is expanded to an 8×25 array of multiservice demands, where, for simplicity, it is assumed that each call carries a requirement for only one service class. This array is given by

$$\boldsymbol{T} = \begin{bmatrix} 3&3&0&0&0&1&0&3&0&2&0&1&5&0&3&0&0&0&0&0&1&0&0&0&0 \\ 2&2&2&0&2&1&0&2&0&2&0&1&5&0&2&0&0&2&0&2&1&4&0&2&0 \\ 0&1&2&0&3&1&0&2&1&4&0&1&0&0&2&6&0&3&0&2&1&0&0&2&0 \\ 1&1&1&1&0&1&1&0&0&0&0&1&1&1&0&0&2&0&0&3&1&0&1&0&0 \\ 2&0&1&1&2&0&1&0&1&0&2&1&2&2&0&0&2&0&2&0&1&0&1&3&0 \\ 0&0&1&1&0&0&1&0&1&0&2&2&0&1&0&0&0&0&2&0&1&0&1&0&3 \\ 2&2&1&1&2&0&2&0&1&0&2&1&2&2&0&0&0&0&1&0&0&0&1&0&2 \\ 0&2&1&1&0&0&0&0&0&0&2&0&0&1&0&0&0&0&0&0&0&0&1&0&0 \end{bmatrix}^T \begin{matrix} \text{8- kbps voice} \\ \text{13- kbps voice} \\ \text{32- kbps voice} \\ \text{64- kbps video} \\ \text{64 - kbps packet data} \\ \text{128 - kbps packet data} \\ \text{384 - kbps packet data} \\ \text{768 - kbps packet data} \end{matrix}$$

The same 25×25 interference matrix introduced for Kunz' Helsinki model is used in the simulations.

The SOFM for the SRRA problem for this network is simulated, with $K = 100$ initial states and the RRAs all initialized to (1, 0, 64, 0) in each BS coverage area. The average combined blocking and dropped handoff probability is 0.087 for the real-time classes and 0.173 for the packet-data classes, with the ANI equal to 1874.2 for the final SOFM stage. Even though there is an eight-fold increase in the complexity of the multimedia network example over Kunz' network model, the convergence of NN-cascade RRA algorithm occurs in fewer iterations than the convergence of HNN algorithm in 2450 iterations reported by Kunz for his voice network model. The overall cascade simulation is very slow due to the sequential operation of the stages in generating estimates. The SOFM is slower than the single-service SOFM evaluated in [4], since it must "learn" the correct RRA iteratively over a larger search space to meet a vector of demands. The algorithm produces higher blocking probabilities for the packet-data demands, since only a limited number of DCHs are allowed allocation to this class, while real-time services have access to all channels supported by the RRA. The algorithm slowly increments the sectorization values to 6, sets the activity monitoring "on", while the power control assignments vary over the 25 BS areas according to the number of active high-rate packet data requests.

To evaluate the SOFM for the DRRA problem, the 25 rows of the matrix $\boldsymbol{T}$ above are cyclically shifted 5 positions down every Π periods, with Π = 10, 20, 50, and 100, to represent dynamic local traffic demands at the BSs. In order to examine the sensitivity of the algorithm to initial RRAs, as it responds to demand shifts, the following three patterns in each coverage area are initially used in this simulation: (1) RRAs are set to (1, 0, 64, 0); (2) RRAs are set to (3, 1, 64, 20); and (3) the RRAs are set to the final values at the completion of the SRRA after $K = 100$ periods. In response to these cyclic demand shifts, the average blocking/dropped handoff probabilities for the DRRA with initial RRA pattern (1) increase from 0.087 to 0.291 for the real-time services and from increase from 0.173 to 0.426 for the packet-data services, as the number of periods Π decreases from 100 to 10, respectively. For initial RRA pattern (2), the average blocking/dropped handoff probabilities increase from 0.057 to 0.206 for real-time services and from 0.098 to 0.323 for packet-data services, as Π decreases from 100 to 10, respectively. Lastly, using the final RR patterns from the SRRA problem to initiate DRRA algorithm results in the average blocking/dropped handoff probabilities from 0.034 to 0.157 for real-time services and from 0.078 to 0.299 for packet-data services, as Π decreases from 100 to 10, respectively.

6. CONCLUSIONS

In this paper, a cascade model of an MLP-NN for channel estimation, an RNN for adaptive antenna control, a discrete-form HNN for joint multiuser detection, and a discrete-space Kohonen SOFM has been proposed for the problem of allocating radio resources to meet the QoS requirements of multimedia service demands in 3G wireless networks. W-CDMA network parameters on the uplinks have been assumed to model the resources available to support the diverse SIR and delay requirements for variable-rate audio, high-rate packet data, and real-time video. Simulation results for the performance of each of the first three NN stages have been presented for representative W-CDMA scenarios. Finally, both the static and dynamic versions of the complete NN cascade algorithm have been simulated for the RRA of multimedia extensions of

published cellular network models. The simulation results have been informally compared to earlier published results for single-stage HNN and SOFM techniques applied to resource allocation in single-service voice networks.

ACKNOWLEDGMENTS

The author wishes to thank his colleagues at FIT for their support and to commend the Mobile and Portable Radio Research Group at Virginia Tech for their continuing insights into the behavior of emerging mobile radio networks.

REFERENCES

1. T. Ojanperä and R. Prasad, "An overview of air interface multiple access for IMT-2000/UMTS," *IEEE Commun. Mag.*, vol. 36, no. 9, pp. 82–95, Sept. 1998.
2. E. Dahlman, B. Gudmundson, M. Nilsson, and J. Sköld, "UMTS/IMT-2000 based on wideband CDMA," *IEEE Commun. Mag.*, vol. 36, no. 9, pp. 82–95, Sept. 1998.
3. K. Das and S. D. Morgera, "Interference and SIR in integrated voice/data wireless DS-CDMA networks – a simulation study," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1527–1538, Oct. 1997.
4. W. Hortos, "Self-organizing feature maps for dynamic control of radio resources in CDMA microcellular networks," *Appl. and Sci. of Artificial Neur. Net. II, Proc. of SPIE*, vol. 3093, Orlando, FL, Apr. 1998.
5. K. Gilhousen, I. Jacobs, R. Padovani, A. Viterbi, L. Weaver, and C. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol.50, no. 2., pp. 303-312, 1991.
6. G. L. Turin, "The effects of multipath and fading on the performance of direct-sequence CDMA systems," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 507–603, July 1984.
7. A. Jalali and P. Mermelstein, "On the bandwidth efficiency of CDMA systems," in *Proc. IEEE ICC'94.*, May 1994, pp. 515–519.
8. Y. C. Yoon, *et. al.*, "A spread-spectrum multiaccess system with cochannel interference cancellation for multipath fading channels," *IEEE J. Sel. Areas of Commun.*, vol. SAC-11, No. 7, pp. 1067-1075, Sept. 1993.
9. M. Sawahashi, et. al,, "Pilot symbol-aided coherent multistage interference canceller using recursive channel estimation for DS-CDMA mobile radio," *IEICE Trans. Commun.*, vol. E79-B, pp. 1262-1270, Sept. 1996.
10. R. Fraile and N. Cardona, "Fast neural network method for propagation loss prediction in urban environments," *Electron. Lett.*, vol. 33, no. 24, pp. 2056–2058, 1997.
11. J. Walfisch and H. L. Bertoni, "A theoretical model of UHF propagation in urban environments," *IEEE. Trans. Antennas Propag.*, vol. 36, no. 12, pp. 1788 –1796, 1988.
12. S. R. Saunders and F. R. Bonar, "Explicit multiple building diffraction attenuation function for mobile radiowave propagation," *Electron. Lett.*, vol. 27, no. 14, pp. 1276–1277, 1991.
13. B. E. Gschwendtner and F. M. Landstorfer, "Adaptive propagation modelling using a hybrid neural network technique," *Electron. Lett.*, vol. 32, no. 3, pp. 162–164, 1996.
14. *EUROCOST: European cooperation in the field of scientific and technical research. COST 231, Final Report*, 1996.
15. M. Benson and R. A. Carrasco, "Recurrent neural network array for CDMA mobile communication systems," *Electron. Lett.*, vol. 33, no. 25, pp. 2105–2106, 1997.
16. R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, pp. 270-280, 1989.
17. S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. on Info. Theory*, vol. 32, pp. 85–86, Jan. 1986.
18. U. Mitra and H. V. Poor, "Neural network techniques for adaptive multi-user demodulation," *IEEE J. Select. Areas Commun.*, vol. 12, no. 9, Dec. 1994.
19. U. Mitra and H. V. Poor, "Adaptive receiver algorithms for near-far resistant CDMA," *IEEE Trans. Commun.*, vol. 43, pp. 1713–1724, 1995.
20. B. Mulgrew, "Applying radial basis functions," *IEEE Signal Proc. Mag.*, vol. 13, pp. 50–65, Mar. 1996.
21. G. I. Kechriotis and E. S. Manolakos; "Hopfield neural network implementation of the optimal CDMA multiuser detector," *IEEE Trans. on Neural Networks*, vol. 7, pp. 131–141, Jan. 1996.

22. G. I. Kechroitis and E. S. Manolakos, "A hybrid digital computer – Hopfield neural network spread-spectrum CDMA detector for real-time multi-user demodulation," *Proc. 1994 IEEE-SP Int. Workshop on Neur. Networks for Signal Processing*, Ermoni, Greece, pp. 545–554, Sept. 1994.

23. S. Verdu, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.

24. M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple access communications," *IEEE Trans. on Comm.*, vol. 38, pp. 509–519, Apr. 1990.

25. J. J. Hopfield and D. W. Tank, " Neural computation of decisions in optimization problems," *Biological Cybern.*, vol. 52, pp. 141-152, 1985.

26. J. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, New York, 1991.

27. Z. Xie, C. K. Rushforth, and R. T. Short, "Multiuser signal detection using sequential decoding," *IEEE Trans. on Comm.*, vol. 38, pp. 578–582, May 1990.

28. T. Kohonen, "Self-organized formation of topologically correct feature maps," *Bio. Cybern.*, vol. 43, no. 1, pp. 59-69, 1982.

29. K. N. Sivarajan, R. J. McEliece, and J. W. Ketchum, "Channel assignment in mobile radio," in *Proc. 39th IEEE Veh. Technol. Soc. Conf.,* pp. 846-850, San Francisco, CA, 1989.

30. S. Abe, "Convergence acceleration of the Hopfield neural network by optimizing integration step sizes," *IEEE Trans. Syst. Man and Cybern.-Pt.B*, vol. 26, no. 1, pp. 194-201, 1996.

31. K. Smith and M. Palaniswami, "Static and dynamic channel assignment using neural networks," *IEEE J. Selected Areas Comm.*, vol. 15, no. 2, pp. 238-249, 1997.

32. S. Abe, "Global convergence and suppression of spurious states of the Hopfield neural nets," *IEEE Trans. Circuits Syst.*, vol. CAS-40, no. 4, pp. 246-257, 1993.

33. D. Kunz, "Channel assignment for cellular radio using neural networks," *IEEE Trans. Veh. Technol.*, vol. VT-40, pp. 188-193, 1991.