


## Chapter 4

# The Scheduling Algorithm for Downlink Dedicated and Shared Channels in Multimedia CDMA Cellular Systems

---



*In this chapter, a cellular neural network and utility (CNU)-based scheduler is proposed for multimedia CDMA cellular networks supporting differentiated quality-of-service (QoS). The cellular neural network is powerful for complicated optimization problems and has been proved that it can rapidly converge to a desired equilibrium; the utility-based scheduling algorithm can efficiently utilize the radio resource for system and provide QoS requirements and fairness for connections. A relevant utility function for each connection is here defined as its radio resource function further weighted by both a QoS requirement deviation function and a fairness compensation function. The CNU-based scheduler determines a radio resource assignment vector for all connections so that the overall system utility is maximized and the system throughput can be achieved as high as possible. At the same time, the performance measures of all connections are kept closed to their QoS requirements in an efficient way. The simulation results exhibit that CNU-based scheduler has higher system throughput and larger QoS guaranteed region than other scheduling algorithm for environments with variant type of services.*

## 4.1 Notation List of Chapter 4

We summarize the important notations of this chapter in the following table. The temporary variables used in the proof which are also defined in the process of the derivation are not included in this table.

Table 4.1: Notation List of Chapter 4

Notation	Description
$W$	The bandwidth of the carrier of CDMA network
$SF_i$	The spreading factor used by connection $i$
$\alpha$	The downlink orthogonality factor for orthogonal channelization code
$N_0$	The AWGN noise power density
$\mathcal{I}_i$	The resulting interference of connection $i$ received at the base station in the downlink
$SIR_i(t)$	The received SIR connection $i$ at frame time $t$
	the required $SIR_q^*$ of voice connections
$P_{max}^*$	The required received power at base station for connection $i$
$\kappa_i$	The modulation scheme index adopted by connection $i$
$M_{\kappa_i}$	The modulation order with respect to modulation index $\kappa_i$
$\left(\frac{E_b}{N_0}\right)_{\kappa_i}^*$	The required $\frac{E_b}{N_0}$ for modulation scheme $\kappa_i$
$r_i(t)$	The allocated transmission rate of connection $i$ at frame time $t$
$N$	The number of connections in the system
$R_{m,i}^*$	The traffic parameter of mean rate for connection $i$
$P_{D,i}^*$	The QoS requirement of packet dropping probability for connection $i$
$D_i^*$	The QoS requirement of tolerable delay bound for connection $i$
$BER_i^*$	The BER requirement of connection $i$ derived from the required $E_b/N_0$
$d$	The distance from the new/handoff connection to the base station
$\zeta_i(t)$	The link gain model at frame time $t$
$\zeta_i^L(t)$	The log-normal shadowing part of the link gain model
$\zeta_i^S(t)$	The short-term fading part of the link gain model
$\bar{\zeta}_{NRT}$	The averaged mean link gain of all NRT connections
$Q_i(t)$	The queue size of connection $i$ at frame time $t$
$c_i(t)$	The ratio of assignment of radio resource for connection $i$ at frame $t$
$c_i^*(t)$	The ratio of assignment of radio resource for connection $i$ at frame $t$

Notation	Description
$R_i$	The transmission rate of connection $i$
$R_{s,i}$	The coded symbol transmission rate of connection $i$
$\mathcal{U}_i(t)$	The individual utility function for connection $i$
$\mathcal{U}(t)$	The system utility function for connection $i$
$\mathcal{R}_i(t)$	The radio resource function
$\mathcal{A}_i(t)$	The QoS requirement deviation function
$\mathcal{F}_i(t)$	
$w_i$	The target weighting factor of connection $i$
$\bar{w}_i$	The target weighting factor of connection $i$
$D_i(t)$	The head-of-line packet delay of connection $i$
$\hat{L}_i(t)$	The normalized measurement on the difference of guaranteed minimum transmission rate
$\beta_i$	The priority bias for RT connections
$\beta_0$	The basic reference value for NRT connections
$Y_{i,k}^{(1)}(t)$	The first external input for the $(i, k)$ -th neuron of CNN at frame time $t$
$Y_{i,k}^{(2)}(t)$	The second external input for the $(i, k)$ -th neuron of CNN at frame time $t$
$X_{i,k}(\tau)$	The output of the $(i, k)$ -th neuron of CNN at transient time $\tau$
$E(\tau)$	The energy function of CNN at transient time $\tau$
$H(\tau)$	The cost function of CNN at transient time $\tau$
$\Psi_1$	The system constraint 1 for CNN processor
$\Psi_2$	The system constraint 2 for CNN processor
$A_{i,k;j,m}$	The recurrent interconnection weight from $(i, k)$ -th neuron to $(j, m)$ -th neuron
$B_{i,k;j,m}^{(1)}$	The first external control weight from $(i, k)$ -th neuron to $(j, m)$ -th neuron
$B_{i,k;j,m}^{(2)}$	The second external control weight from $(i, k)$ -th neuron to $(j, m)$ -th neuron
$V_{i,k}$	The bias current for $(i, k)$ -th neuron

## 4.2 Introduction

In future wireless networks, heterogeneous and customized services with diverse traffic characteristics and QoS requirements are expected to be provided via a number of air interfaces. Also, multimedia applications are commonly accepted as enabling services, which are categorized into several classes [1]. To meet various traffic characteristics and QoS requirements of these potential applications, a sophisticated scheduling algorithm plays an essential role so that the system resource allocation is optimal, while retaining a

Notation	Description
----------	-------------

pre-defined QoS requirements and fairness among them.

Many scheduling algorithms have been widely studied for wireline networks [32]-[36]. For connections without hard delay and jitter bound requirements, weighted fairness among all connections is the reasonable criterion to share the system capacity, and many scheduling algorithms were developed based on the concept of generalized processor sharing (GPS) [32]. On the other hand, for those with explicit delay and jitter requirements, the scheduling algorithms considering the packet delay perform better than the algorithms in GPS class within the QoS guaranteed region.

The radio channel in wireless networks has quite different characteristics from that in wireline networks, and the available maximum transmission rate to each connection is location-dependent and time-varying due to link loss, shadowing, and multi-path fading. Scheduling for radio resource of wireless networks is to determine which connection and how long this connection can use the system resource so that the utilization of radio resource can be maximized while performance measures of each connection can be efficiently retained at its QoS requirements. We can categorize the design criteria into three main directions: the efficiency of the utilization of radio resource with respect to the link quality, the QoS requirement achievement, and the fairness among all connections. Several further constraints on terminal capability, power range, and some MAC control algorithms can be added in parallel to these scheduling criteria. Also, the multiple access scheme will impact on the link state and should be considered in the design of scheduling algorithms. Several literature studied the resource scheduling and allocation among connections in wireless networks with consideration of physical layer processing, power control range, and link conditions [38]-[39]. Bhargharvan, Lu, and Nandagopal [40] proposed a framework to

achieve long-term fairness in wireless networks. There are schemes considering either delay bound or minimum rate as its QoS requirements. For those schemes considering delay requirements, Varsou and Poor [42] studied the scheduling algorithm based on an EDF (earliest deadline first) concept adapted to wireless environments. They also proposed a simple analysis for the performance of generalized PEDF (powered earliest deadline first) and HOLPRO (head-of-line pseudo-probability assignment) scheduling schemes [43]. Stolyar and Ramanan studied a throughput-optimal scheduling algorithm for delay bounded system [44]; a variational scheduling algorithm for rate guarantee was also investigated. For non-real-time interactive connections, the rate guarantee is desirable. Kam and Siu considered the minimal rate guarantee with fairness in their proposed scheme [45]. Moreover, some schemes considered joint scheduling criteria to deal with complicated needs for systems. Shakkottai and Stolyar [46] considered both link quality and QoS requirements as the criteria and derived an exponential form of scheduling function via fluid Markovian techniques. Many of these scheduling algorithms above, [38]-[39], [44]-[47], were formulated in utility-based approaches. Generally, the utility function is defined as the benefit from receiving an amount of service for each connection so that the overall utility is maximized in addition to fulfilling the design criteria, such QoS requirements and fairness.

The utility-based scheduling algorithm over radio channels, is usually formulated as a complicated optimization problem with real time requirement. To solve the complicated constrained optimization problem, many intelligent techniques have been applied successfully, for example, genetic algorithm, feed-forward neural network, and generalized Hopfield neural networks (HNN). Among those intelligent techniques, the class of generalized HNN has been adopted for real-time tasks. However, the stability and the spurious response problems make the HNN ineffective in practical applications. A special type of

HNN, named *cellular neural network* (CNN) proposed in [66], has been proved that it can rapidly converge to desired equilibrium on vertex along the prescribed trajectories by applying a proper learning or design procedure [67]. The CNN has the architecture that all cells (neurons) have the same structure, i.e. the interconnection weights and bias current, but has much fewer number of inter-connection than that of HNN which is accomplished by locally recurrent inter-connection. The CNN was widely applied in image processing field and was suitable for VLSI implementation. However, to adopt the CNN technique for the scheduling optimization problem, modifications of its architecture and some basic assumptions are necessary, and therefore, its stability and non-spurious property are required to further investigate.

In the chapter, we propose a CNN and utility (CNNU)-based scheduler for downlink in multimedia CDMA cellular networks. The CNNU-based scheduler contains a utility function (UF) preprocessor, a radio-resource range (RR) decision maker, and a CNN processor. Noticeably, the utility function for each connection, adopted in the UF preprocessor, jointly considers radio resource efficiency, diverse QoS requirements, and fairness. It is a radio resource function weighted by both its QoS requirement deviation function and its fairness compensation function. The radio resource function indicates the radio resource efficiency of the connection, which is the maximum achievable transmission rate with its packet-level BER requirement guaranteed, in terms of link quality and the adaptive modulation scheme. The QoS requirement deviation function denotes the extent of the deviation from its call-level QoS requirements which are defined by delay bound, packet dropping ratio, and minimum transmission rate. For the fairness compensation function, a priority bias for each connection is set so that the real-time connections have relative priority over non-real-time connections. The fairness compensation function also makes each non-real-time connection share the radio resource proportional to its prede-

finer target weighting factor in terms of the source characteristics and the channel quality. The UF preprocessor generates a matrix of normalized utility functions of all connections. On the other hand, the RR decision maker determines a matrix showing the upper limit of radio resource assignment for each connection.

The CNN processor receives the matrix of normalized utility functions and the matrix of upper limits of radio resource assignment vector as inputs. At stable state, it determines an optimal normalized radio resource assignment vector for connections in multimedia CDMA cellular systems, by minimizing the system cost function which is in terms of the overall system utility function under system constraints of maximum transmission power, minimum spreading factor, and remaining queue length. The architecture of the CNN is constructed via the energy-based approach [68]-[69]. by mapping the system cost function to a proper energy function. It is designed in a two-layered configuration, which consists of a decision layer and an output layer, to reduce the number of inter-connections in the CNN. It can be shown that the stable equilibriums locate in the desired state space and the stability exists. The CNN is powerful for complicated optimization problem. The performance of the proposed CNNU-based scheduler is investigated by comparing with *Exponential Rule* [46] for systems using both dedicated and shared channel. The simulation results show that CNNU-based scheduler has higher system throughput by the amount of over 10% than that of exponential rule scheduling algorithm. It can be found that the CNNU-based scheduler can fully utilize the radio resource and make the performance measures of QoS requirements closed but below the desired level, and therefore achieve larger QoS guaranteed region than the exponential rule does. From the results of fairness comparison, CNNU-based scheduler can get more closed to the defined weighted fairness than the exponential rule algorithm does. We can conclude that the CNNU-based scheduler is efficient and effective for multimedia CDMA cellular networks.

The rest of the chapter is organized as follows. Section 4.3 presents the features and the operations of the considered system is presented. In section 4.4, an relevant utility function is then proposed, and a utility-based scheduling algorithm is proposed. In section 4.5, the design of CNNU-based scheduler is discussed, where preliminaries of CNN are firstly briefed, the design procedure based on the Lyapunov method, and the detailed configuration of the CNNU-scheduler are presented. The stability and convergence of the proposed CNN processor are discussed in section 4.6. Simulation results to examine the performance of the CNNU-based scheduler is conducted in section 4.7. Finally, some conclusions are summarized in section 4.8.

### 4.3 System Model

Assume that there are  $N$  real-time (RT) and non-real-time (NRT) connections (users) in the downlink transmissions of the multimedia CDMA cellular system with chip rate  $W$ . RT connections transmit on dedicated channels and NRT connections transmit on shared channels. For every active connection using either dedicated or shared channels, a fixed number of code channels with their corresponding spreading factors are given in the connection setup phase. A minimum spreading factor  $SF_i$  is therefore associated with the assigned code channels for connection  $i$ . The system radio resource is here defined to be the transmission power. It is limited by a maximum power budget denoted by  $P_{max}^*$  and scheduled to all connections every frame time period  $T_f$ .

For a downlink connection  $i$ , there are four QoS requirements defined in either the packet level, such as  $BER_i^*$ , or the call level, such as delay bound  $D_i^*$ , packet dropping ratio  $P_{D,i}^*$ , and minimum transmission rate  $R_{m,i}^*$ . For RT connections, hard delay bound  $D_i^*$  exists and  $P_{D,i}^*$  can be larger than zero; while for NRT connections, no explicit delay bound is imposed, but  $R_{m,i}^* > 0$  should be satisfied for interactive connections and  $R_{m,i}^* = 0$  be



set for best effort connections.

For a RT connection  $i$ , a transmission suspension in a soft fashion is carried out by allocating zero transmission power when its utility calculated by the scheduler is lower than those of NRT connections. At that moment, its link gain  $\zeta_i(t)$  is lower than the averaged mean link gains of all NRT connections  $\bar{\zeta}_{NRT}$  by a relative margin, and this relative margin should be considered to restrict the probability of transmission suspension below  $P_{D,i}^*$  due to the delay-sensitive nature. Denote by  $\zeta_i^*$  the suspension threshold of connection  $i$ , which is obtained by  $\mathbf{P}\{\zeta_i(t) \leq \zeta_i^*\} \leq P_{D,i}^*$ . Then the relative margin of  $\zeta_i(t)$  is a function of  $\bar{\zeta}_{NRT}$  and  $\zeta_i^*$ , and is dependent on the design of scheduling algorithm. For NRT connections, their transmissions are scheduled so that NRT connections will be allocated with proper radio resource to achieve high system utilization and keep the fairness and the QoS requirements fulfilled as much as possible.

Assume that the link-gain  $\zeta_i(t)$  and the interference  $\mathcal{I}_i(t)$  for connection  $i$  at time  $t$  can be measured at the user side and perfectly signaled to the base station. The  $\zeta_i(t)$  consists of the mean path loss, long-term fading, and short-term fading. It is given by

$$\zeta_i(t) = d_i^{-4} \cdot 10^{\frac{\zeta_i^L(t)}{10}} \cdot \zeta_i^S(t),$$

where  $d_i$  is the distance between the user  $i$  and its base station,  $\zeta_i^L(t)$  is the log-normal shadowing component, and  $\zeta_i^S(t)$  is the Rayleigh-fading component. The adaptive QAM modulation is adopted and the modulation order  $M_{\kappa_i}$  with index  $\kappa_i$  for connection  $i$  is determined according to the link gain quality and interference. The traffic source of connection  $i$  generates packets and packets are queued in its individual buffer. The buffer size is infinite. The source models are assumed to be on-off for RT connections, Pareto for NRT interactive connections, and batch Poisson with truncated geometrical batch size for NRT best effort connections.

The proposed CNNU-based scheduler determines an optimal normalized radio resource

assignment vector  $\vec{c}^*(t) = (c_1^*(t), \dots, c_N^*(t))$  to  $N$  connections via maximizing an overall system utility function so that the system throughput can be achieved as high as possible within the constraints of maximum power, minimum spreading factor, and waiting queue length. The transmission rate for connection  $i$  at  $t$ -th frame, denoted by  $r_i(t)$ , is then allocated according to  $c_i(t)$  of connection  $i$ .

## 4.4 Formulation of The Utility Function

The utility function for connection  $i$ ,  $\mathcal{U}_i(t)$ , is defined as the radio resource function of connection  $i$ ,  $\mathcal{R}_i(t)$ , weighted by its QoS requirement deviation function  $\mathcal{A}_i(t)$  and its fairness compensation function  $\mathcal{F}_i(t)$ . It can be expressed as

$$\mathcal{U}_i(t) = \mathcal{R}_i(t) \cdot \mathcal{A}_i(t) \cdot \mathcal{F}_i(t). \quad (4.1)$$

### 4.4.1 Radio Resource Function $\mathcal{R}_i(t)$

With the modulation order  $M_{\kappa_i}$  of the adaptive QAM modulation scheme and the corresponding  $(E_b/N_0)_{\kappa_i}^*$  to satisfy the  $BER_i^*$  requirement for connection  $i$ , the following inequality should hold

$$\frac{W}{R_{s,i}(t)} \cdot \frac{c_i(t) \cdot P_{max}^* \cdot \zeta_i(t)}{\mathcal{I}_i(t)} \geq \left( \frac{E_b}{N_0} \right)_{\kappa_i}^*, \quad (4.2)$$

where  $R_{s,i}(t)$  is its symbol rate and  $c_i(t)$  is its normalized radio resource assignment at time  $t$ . The  $\mathcal{I}_i(t)$  in Eq. (4.2) is given by  $[(1 - \alpha)P_{max}^* \cdot \zeta_i(t) + \sum_b P_{max}^* \cdot \zeta_{i,b}(t) + N_0W]$ , where  $\alpha$  is the orthogonality factor for downlink,  $b$  is the index referring to the adjacent base stations, and  $\zeta_{i,b}(t)$  is the link gain from base station  $b$  to connection  $i$ . The  $BER_i^*$  of connection  $i$  can be expressed by

$$BER_i^* = 0.2 \int_{\gamma} \exp \left\{ \frac{-1.5\gamma_i}{M_{\kappa_i} - 1} \right\} f_{\gamma_i}(\gamma) d\gamma,$$

where  $\gamma_i$  is the instantaneous  $(E_b/N_0)_{\kappa_i}$  received by connection  $i$ , and  $f_{\gamma_i}(\gamma)$  is the pdf of  $\gamma_i$  [70]. Since the channel state is assumed to be known and to remain constant during a frame time, the  $(E_b/N_0)_{\kappa_i}^*$  in Eq. (4.2) is given by

$$(E_b/N_0)_{\kappa_i}^* = \frac{-(M_{\kappa_i} - 1) \cdot \ln\{5BER_i^*\}}{1.5}. \quad (4.3)$$

We denote the *maximum achievable symbol rate* that can fulfill the  $(E_b/N_0)_{\kappa_i}^*$  at  $c_i(t) = 1$  by  $R_{s,i}^*(t)$ . Clearly,  $R_{s,i}^*(t) = \frac{W}{(E_b/N_0)_{\kappa_i}^*} \cdot \frac{P_{max}^* \cdot \zeta_i(t)}{\mathcal{I}_i(t)}$ . The  $R_{s,i}^*(t)$  is further limited by  $\frac{W}{SF_i}$  for a given spreading factor  $SF_i$  of the allocated code channel. Thus the  $R_{s,i}^*(t)$  can be obtained by

$$R_{s,i}^*(t) = \min \left\{ \frac{W}{(E_b/N_0)_{\kappa_i}^*} \cdot \frac{P_{max}^* \cdot \zeta_i(t)}{\mathcal{I}_i(t)}, \frac{W}{SF_i} \right\}. \quad (4.4)$$

According to (4.4), the most efficient modulation order  $M_{\kappa_i}$  is selected by the following inequality,

$$M_{\kappa_i} \leq \frac{SF_i \cdot P_{max}^* \cdot \zeta_i(t)}{\mathcal{I}_i(t) \cdot \left( \frac{-\ln\{5BER_i^*\}}{1.5} \right)} + 1 \leq M_{(\kappa_i+1)}. \quad (4.5)$$

Since the information bit of one symbol is  $\log_2 M_{\kappa_i}$ , consequently the radio resource function of connection  $i$ ,  $\mathcal{R}_i(t)$ , can be obtained by

$$\mathcal{R}_i(t) = \log_2 M_{\kappa_i} \cdot R_{s,i}^*(t) = \frac{1.5W \cdot \log_2 M_{\kappa_i}}{(M_{\kappa_i} - 1) \cdot \left[ \ln \frac{1}{BER_i^*} - \ln 5 \right]} \cdot \frac{P_{max}^* \cdot \zeta_i(t)}{\mathcal{I}_i(t)}. \quad (4.6)$$

Note that if the assignment of radio resource for connection  $i$ ,  $c_i(t)$ , is allocated, the transmission rate  $r_i(t)$  is therefore equal to  $c_i(t) \cdot \mathcal{R}_i(t)$ .

#### 4.4.2 The QoS Requirement Deviation Function $\mathcal{A}_i(t)$

The QoS requirement deviation function  $\mathcal{A}_i(t)$  is used to indicate how much extent the connection  $i$  deviates from its call-level QoS requirements. The higher extent the deviation from the QoS requirements is, the more resource allocated to the connection should be. For a RT connection  $i$ , a hard delay bound  $D_i^*$  is imposed on each packet.

Since QoS over wireless interface can be provided in a soft fashion, the QoS guarantee of packet dropping ratio due to excess delay is expressed by  $P_{rob} \{D_i(t) > D_i^*\} < P_{D,i}^*$ , where  $D_i(t)$  is the waiting time delay for head-of-line packet at time  $t$ . For an NRT interactive connection  $i$ , a different notion of QoS requirement is that a minimum transmission rate must be guaranteed by  $\mathbf{E}[r_i(t)] \geq R_{m,i}^*$ . As for an NRT best-effort connection  $i$ , no call level QoS requirements are guaranteed and the  $R_{m,i}^*$  is set to be 0.

From [72], the proposed *Modified Largest Weighted Delay First* (M-LWDF) algorithm suggests that an exponential rule [46] be the form with throughput optimal for the above call level QoS requirement constraints. Therefore, the *QoS requirement deviation function*  $\mathcal{A}_i(t)$  is defined as

$$\mathcal{A}_i(t) = \begin{cases} \exp \left\{ \frac{-\log(P_{D,i}^*) \cdot D_i(t) - \bar{D}(t)}{1 + [\bar{D}(t)]^{1/2}} \right\}, & \text{if } i \in \{\text{RT connections}\}, \\ \exp \left\{ \frac{\hat{L}_i(t) - \bar{L}(t)}{1 + [\bar{L}(t)]^{1/2}} \right\}, & \text{if } i \in \{\text{NRT interactive connections}\}, \\ 1, & \text{if } i \in \{\text{NRT best-effort connections}\}, \end{cases} \quad (4.7)$$

where  $\bar{D}(t) = \frac{1}{N} \sum_i \left( \frac{-\log(P_{D,i}^*)}{D_i^*} \right) \cdot D_i(t)$  is the average weighted delay,  $\hat{L}_i(t) = \hat{L}_i(t-1) + \left( \frac{R_{m,i}^* - r_i(t)}{R_{m,i}^*} \right)$  is the normalized measurement on the difference of guaranteed minimum transmission rate and the assigned rate, and  $\bar{L}(t) = \frac{1}{N} \sum_i \hat{L}_i(t)$ . For the RT connections, the delay of connection  $i$  is weighted by the log-scale packet dropping ratio and the inverse of the delay bound requirements [46]. If the weighted delay is more than the average weighted delay of all connections, the  $\mathcal{A}_i(t)$  will be exponentially increased, and more resource will be scheduled; on the other hand, if the weighted delay is less than the average weighted delay, the  $\mathcal{A}_i(t)$  will dramatically decayed, and less resource will be allocated. Similarly, for the NRT interactive connections, if the accumulated difference of the guaranteed minimum transmission rate and the assigned rate is greater than the average value, more resource is assigned. As to the NRT best-effort connections, this function is simply bypassed.

### 4.4.3 The Fairness Compensation Function $\mathcal{F}_i(t)$

The fairness compensation function is to ensure that RT connections using dedicated channels have the relative priority over NRT connections using shared channels. It is also the way that the radio resource shared by all NRT connections is assigned according to a predefined target weighting factor. With the pre-defined target weighting factor  $w_i$  for NRT connections  $i$  for  $1 \leq i \leq N$ , the radio resources are here expected to be allocated to any two any connections,  $i$  and  $k$ , so that their average assigned transmission rates,  $\mathbf{E}[r_i(t)]$  and  $\mathbf{E}[r_k(t)]$ , can be achieved by  $\frac{\mathbf{E}[r_i(t)]}{\mathbf{E}[r_k(t)]} = \frac{w_i}{w_k}$  [40]. The *fairness compensation function* for connection  $i$  till time  $t$ ,  $\mathcal{F}_i(t)$ , is defined by,

$$\mathcal{F}_i(t) = \begin{cases} \beta_i, & \text{if } i \in \{\text{RT connections}\}, \\ \beta_0 \cdot [(w_i - \bar{w}_i(t) - 1)^+ + 1], & \text{if } i \in \{\text{NRT connections}\}, \end{cases} \quad (4.8)$$

where  $\beta_i$  is the priority bias for RT connections to differentiate with NRT connections,  $\beta_0$  is the basic reference value set for NRT connections,  $(x)^+ = \max\{x, 0\}$ , and  $\bar{w}_i(t)$  is the moving-average of  $r_i(t)$ . For RT connections, the weighted fairness is not considered and only priority bias is set due to their QoS-driven nature. For NRT connections,  $[(w_i - \bar{w}_i(t) - 1)^+ + 1] = \max\{(w_i - \bar{w}_i(t)), 1\}$ , where  $(w_i - \bar{w}_i(t))$  indicates the unfairness of connections  $i$ , and the  $\mathcal{F}_i(t)$  is limited by  $\beta_0$  after  $\bar{w}_i(t)$  is above  $w_i$  and will make no further effects on the utility. The more extent of the unfairness of connection  $i$  is, the larger the  $\mathcal{F}_i(t)$  will be; then more resource will be scheduled to connection  $i$ , and the  $(w_i - \bar{w}_i(t))$  will be smaller afterwards. In the stationary situation, the unfairness of all NRT connections should be almost the same via the linear feedback control.

The target weighting factor  $w_i$  is defined as the target average transmission rate of NRT connection  $i$ . It is considered to be a function of its equivalent traffic source rate  $s_i^*$ , mean link gain  $\bar{\zeta}_i$ , mean interference level  $\bar{\mathcal{I}}_i$ , and its guaranteed minimum transmission

rate,  $R_{m,i}^*$ . The  $w_i$  is given by

$$w_i = \max \left\{ \frac{P_{max}^* \cdot \bar{\zeta}_i}{\bar{\mathcal{L}}_i \cdot \left(\frac{E_b}{N_0}\right)_i^*} \cdot \frac{s_i^*}{\sum_k s_k^*}, R_{m,i}^* \right\}, \quad (4.9)$$

where  $\left(\frac{E_b}{N_0}\right)_i^*$  is the required  $E_b/N_0$  to achieve  $BER_i^*$  of connection  $i$  using the least-order modulation scheme. The  $s_i^*$  can be given according to the effective bandwidth method proposed in [71]. The  $w_i$  is proportional to its mean maximum transmission rate,  $\frac{P_{max}^* \cdot \bar{\zeta}_i}{\bar{\mathcal{L}}_i \cdot \left(\frac{E_b}{N_0}\right)_i^*}$ , and its normalized effective bandwidth,  $\frac{s_i^*}{\sum_k s_k^*}$ . The lower bound  $R_{m,i}^*$  for  $w_i$  is to avoid the starvation problem of the connection  $i$  in bad link condition. Note that  $R_{m,i}^* = 0$  for the best-effort connection, and the target weighting factor of the best-effort connection is usually less than that of the interactive connection.

The priority bias  $\beta_i$  for RT connection  $i$  is a relative margin for  $\zeta_i(t)$  over the link gains of NRT connections, and is a function of its transmission suspension threshold  $\zeta_i^*$ , the average of the mean link gains of all NRT connections  $\bar{\zeta}_{NRT}$ , and the  $E_b/N_0$  requirements. The  $\beta_i$  is given by

$$\beta_i = \left( \frac{\bar{\zeta}_{NRT}}{\zeta_i^*} \cdot \frac{\left(\frac{E_b}{N_0}\right)_i^*}{\left(\frac{E_b}{N_0}\right)_{NRT}^*} \right) \cdot \beta_0. \quad (4.10)$$

The  $\beta_i$  makes the product  $\mathcal{R}_i(t) \cdot \mathcal{F}_i(t)$  of the RT connection  $i$  greater than the product of NRT connections till  $\zeta_i(t) \geq \zeta_i^*$ . Therefore, RT connections can share the radio resource with relatively higher priority over NRT connections via the setting of priority bias  $\beta_i$ .

## 4.5 Design of the CNNU-based Scheduler

Fig. 4.1 shows the block diagram of the CNNU-based scheduler. It contains a *utility function (UF) preprocessor*, a *radio-resource range (RR) decision maker*, and a *CNN processor*. The proposed CNNU-based scheduler takes the link information, interference, delay, queue length, and spreading factor of all connections as inputs, and finally outputs

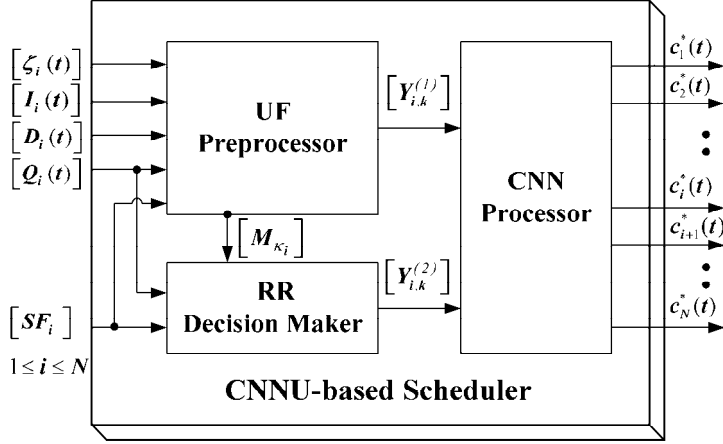


Figure 4.1: The block diagram of CNNU-based scheduler.

the optimal normalized radio resource assignment vector  $\vec{c}^*(t) = (c_1^*(t), \dots, c_N^*(t))$ , where  $c_i^*(t)$ ,  $1 \leq i \leq N$ , is expressed by  $K$  bits.

The *UF preprocessor* first calculates the utility function  $\mathcal{U}_i(t)$  given in Eq. (4.1),  $1 \leq i \leq N$ . Then it normalizes  $\mathcal{U}_i(t)$  by a compression function  $(1 - e^{-\sigma \mathcal{U}_i(t)})$ , expresses  $(1 - e^{-\sigma \mathcal{U}_i(t)})$  to be an  $1 \times K$  vector given by

$$\left( (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-1}, \dots, (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k}, \dots, (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-K} \right),$$

and finally constructs an  $N \times K$  input matrix  $[Y_{i,k}^{(1)}]$  for the CNN processor, where  $Y_{i,k}^{(1)} = (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k}$ . Notice that  $\sigma$  is a constant related to the slope and the linear region of the compression function. The compression function  $(1 - e^{-\sigma \mathcal{U}_i(t)})$  normalizes  $U_i(t) \in [0, \infty)$  into the unit range of  $[0, 1)$ . A good compression function is the one with broad linear range so that the individual utility function is normalized linearly within a reasonable range. The *UF preprocessor* also determines a vector of modulation order  $[M_{k_i}]$  for all connections and outputs to the *RR decision maker*. The *RR decision maker* determines the upper limit for the radio resource assignment for every connection  $i$  and expresses it by an  $1 \times K$  vector which is the upper limit multiplied by the bit-weighted vector  $(2^{-1}, \dots, 2^{-k}, \dots, 2^{-K})$ ,  $1 \leq i \leq N$ . Note that, with given spreading factor  $SF_i$  and

queue length  $Q_i(t)$  for connection  $i$ , its radio resource assignment  $c_i(t)$  is upper limited by  $\min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\}$ , which will be further discussed in section 4.5.2. Then the *RR decision maker* constructs the second input matrix  $[Y_{i,k}^{(2)}]$ ,  $1 \leq i \leq N$ ,  $1 \leq k \leq K$ , of which the element  $Y_{i,k}^{(2)}$  is given by  $\left( \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot 2^{-k}$ . The *CNN processor* receives input matrices,  $[Y_{i,k}^{(1)}]$  and  $[Y_{i,k}^{(2)}]$ , and determines the optimal radio resource assignment vector  $\vec{c}^*(t)$ . During the computation process, denote by  $\tau$  the instantaneous time index of the CNN and by  $\vec{c}(t, \tau)$  the instantaneous radio resource assignment vector at time  $\tau$  during the frame  $t$ . For each  $c_i(t, \tau)$ ,  $1 \leq i \leq N$ , it is represented by  $K$  bits,  $X_{i,k}(\tau)$ ,  $1 \leq k \leq K$ , and  $c_i(t, \tau)$  can be expressed by

$$c_i(t, \tau) \cong \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k}. \quad (4.11)$$

When the CNN processor arrives at an equilibrium, the output will converge to the optimal radio resource assignment vector, i.e.,  $\lim_{\tau \rightarrow \infty} \vec{c}(t, \tau) = \vec{c}^*(t)$ .

In the following, the design of the CNN processor for the CNNU-based scheduler is described. Characteristics of the original CNN proposed in [66] is first briefed, and a cost function corresponding to the system utility function with system constraints is formulated. A modified architecture for *CNN processor* is then presented, based on the Lyapunov method. The stability of the neural network is also verified.

#### 4.5.1 Preliminaries for Cellular Neural Networks

Consider a neural network with  $N \times K$  neurons arranged in a rectangular array, where neuron  $(i, k)$  is denoted by  $z_{i,k}$ . The output of  $z_{i,k}$  at time  $\tau$ , denoted by  $X_{i,k}(\tau)$ , can be expressed by  $X_{i,k}(\tau) = f(X_{i,k}^{(s)}(\tau))$ , where  $f(x) = \frac{1}{2} [|x| - |x - 1|] + \frac{1}{2}$  is the activation function of  $z_{i,k}$  and  $X_{i,k}^{(s)}(\tau)$  is the state variable of  $z_{i,k}$  at time  $\tau$ .  $X_{i,k}^{(s)}(\tau)$  consists of the recurrent inputs, external inputs, and a bias current. For each neuron  $z_{i,k}$ , it connects with all other neurons within its neighborhood, denoted by  $Z_n(i, k)$ . The area of  $Z_n(i, k)$



is determined according to the design of the neural network. Generally, the dynamics of the CNN at time  $\tau$  is represented by

$$\begin{aligned} \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} = & -\frac{X_{i,k}^{(s)}(\tau)}{\nu} + A_{i,k;i,k} \cdot X_{i,k}(\tau) + B_{i,k;i,k} \cdot Y_{i,k} \\ & + \sum_{z_{j,m} \in Z_n(i,k)} A_{i,k;j,m} \cdot X_{j,m}(\tau) + \sum_{z_{j,m} \in Z_n(i,k)} B_{i,k;j,m} \cdot Y_{j,m} + V_{i,k}, \end{aligned} \quad (4.12)$$

where  $\nu$  is a time constant for all neurons,  $A_{i,k;j,m}$  is the recurrent interconnection weight from neuron  $z_{j,m}$  to  $z_{i,k}$ ,  $B_{i,k;j,m}$  is the control weight of external input from  $z_{j,m}$  to  $z_{i,k}$ ,  $Y_{j,m}$  is the external input to the neuron  $z_{j,m}$ , and  $V_{i,k}$  is the bias current to  $z_{i,k}$  which is usually a fixed value  $V$ . It is worth mentioning that  $A_{i,k;i,k} > 1/\nu$  is hold so that the neuron  $z_{i,k}$  will eventually enter into a saturation region [66]. Also, the interconnection weights are assumed to be symmetric, that is,  $A_{i,k;j,m} = A_{j,m;i,k}$ , thus the CNN is stable [66].

An energy function at time  $\tau$  which decreases along the trajectories of Eq. (4.12) is generally expressed by [66]

$$\begin{aligned} E(\tau) = & -\frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K A_{i,k;i,k} X_{i,k}^2(\tau) - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^N \sum_{m=1}^K A_{i,k;j,m} X_{j,m}(\tau) X_{i,k}(\tau) \\ & - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^N \sum_{m=1}^K B_{i,k;j,m} Y_{j,m} X_{i,k}(\tau) - \sum_{i=1}^N \sum_{k=1}^K V_{i,k} \cdot X_{i,k}(\tau). \end{aligned} \quad (4.13)$$

At the stable state, outputs of neurons will arrive at an equilibrium of that the energy function is minimized. If the energy function is properly designed and acts as a cost function, such an optimization problem can be solved via the Lyapunov method [67], [68]-[69]. By the Lyapunov method, the CNN can be designed with a set of prescribed trajectories. The trajectories are described by the gradient of the Lyapunov function  $E(\tau)$  which is the energy of the CNN network at time  $\tau$ . With an appropriate energy function designed according to the cost function, the minimization of the cost can be achieved along the designed trajectories. In the mean time, it can be proved that the architecture

of the designed CNN can be related with the energy function by

$$\frac{dX_{i,k}^{(s)}(\tau)}{d\tau} = -\frac{X_{i,k}^{(s)}(\tau)}{\nu} - \frac{\partial E(\tau)}{\partial X_{i,k}(\tau)}. \quad (4.14)$$

Using Eq. (4.14), the desired system parameters of inter-connection weights, control weights, and bias currents can be found from the trajectories of energy function.

### 4.5.2 Formulation of Cost Function for CNN Processor

The cost function of CNN [68] for frame  $t$  at time  $\tau$ , denoted by  $\mathcal{H}(t, \tau)$ , consists of a cost function for the utility function, denoted by  $\mathcal{H}_u(t, \tau)$ , in conjunction with cost functions for system constraints  $\Psi_1$  and  $\Psi_2$ , denoted by  $\mathcal{H}_{\Psi_1}(t, \tau)$  and  $\mathcal{H}_{\Psi_2}(t, \tau)$ , respectively. The constraint  $\Psi_1 = \{\vec{c}(t) : \sum_n c_n(t) \leq 1\}$  is because the system transmission power is limited by a maximum power budget  $P_{max}^*$ . Notice that the assigned transmission rate  $r_i(t)$  of connection  $i$  for frame  $t$  is determined according to both the  $c_i(t)$  and the modulation order  $M_{\kappa_i}$ ; the  $r_i(t)$  is further limited by the minimum spreading factor  $SF_i$  and the waiting queue length  $Q_i(t)$ ; and too large  $c_i(t)$  with excess allocated power makes no effects on  $r_i(t)$ . Also the constraint  $\Psi_2 = \{\vec{c}(t) : c_i(t) \leq \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\}, \forall i\}$ , where  $\mathcal{R}_i(t)$  is the radio resource function indicating the maximum achievable transmission rate for connection  $i$  at time  $t$ . The constraint  $\Psi_2$  indicates no further utility can be gained if  $r_i(t)$  exceeds the supported rate which is the rate when the power ratio  $c_i(t)$  equals  $(\frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)})$ , or the necessary rate to transmit all remaining packets in  $Q_i(t)$ , where the necessary rate is the rate when  $c_i(t)$  equals  $(\frac{Q_i(t)/T_f}{\mathcal{R}_i(t)})$ . The  $\mathcal{H}(t, \tau)$  has the form of

$$\mathcal{H}(t, \tau) = \mathcal{H}_u(t, \tau) + \mathcal{H}_{\Psi_1}(t, \tau) + \mathcal{H}_{\Psi_2}(t, \tau). \quad (4.15)$$

The  $\mathcal{H}_u(t, \tau)$  is defined to be the difference between an overall normalized utility function and its maximum, and the overall normalized utility function is defined as  $\sum_{i=1}^N c_i(t, \tau) \cdot (1 - e^{-\sigma U_i(t)})$ . When  $\sum_{i=1}^N c_i(t, \tau) \leq 1$ ,  $\sum_{i=1}^N c_i(t, \tau) \cdot (1 - e^{-\sigma U_i(t)})$  is bounded

by 1. Thus the  $\mathcal{H}_u(t, \tau)$  is given by

$$\mathcal{H}_u(t, \tau) = \eta_0 \left[ 1 - \sum_{i=1}^N c_i(t, \tau) \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) \right], \quad (4.16)$$

where  $\eta_0$  is the coefficient for the  $\mathcal{H}_u(t, \tau)$ .

The  $\mathcal{H}_{\Psi_1}(t, \tau)$  is defined as

$$\mathcal{H}_{\Psi_1}(t, \tau) = \eta_1 \left[ \sum_{i=1}^N c_i(t, \tau) - 1 \right]^2, \quad (4.17)$$

where  $\eta_1 = \eta_1^+ \cdot u \left( \sum_{i=1}^N c_i(t, \tau) - 1 \right) + \eta_1^- \cdot \left( 1 - u \left( \sum_{i=1}^N c_i(t, \tau) - 1 \right) \right)$ ,  $u(\cdot)$  is the unit-step function,  $\eta_1^+$  is the slope constant for the cost increment when the total radio resource is greater than the maximum, and  $\eta_1^-$  is the slope constant for the cost increment otherwise. The ranges of  $\eta_1^+$  and  $\eta_1^-$  should be further investigated to ensure the stability and the desired output pattern of the *CNN processor*.

The  $\mathcal{H}_{\Psi_2}(t, \tau)$  is defined to be proportional to the difference  $\left[ c_i(t, \tau) - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right]$  if  $c_i(t, \tau) > \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\}$ ; otherwise, no cost will be incurred because the radio resource will be allocated to other connections for efficiency. It is given by

$$\mathcal{H}_{\Psi_2}(t, \tau) = \eta_2 \left[ \sum_{i=1}^N \left( \left( c_i(t, \tau) - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right)^+ \right)^2 \right], \quad (4.18)$$

where  $\eta_2$  is the coefficient for the  $\mathcal{H}_{\Psi_2}(t, \tau)$ .

Consequently, the cost function  $\mathcal{H}(t, \tau)$  is given by

$$\begin{aligned} \mathcal{H}(t, \tau) &= \eta_0 \left[ 1 - \sum_{i=1}^N c_i(t, \tau) \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) \right] + \eta_1 \cdot \left[ \sum_{i=1}^N c_i(t, \tau) - 1 \right]^2 \\ &+ \eta_2 \left[ \sum_{i=1}^N \left( \left( c_i(t, \tau) - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right)^+ \right)^2 \right]. \end{aligned} \quad (4.19)$$

### 4.5.3 The Architecture of CNN Processor

According to the cost function  $\mathcal{H}(t, \tau)$  at time  $t$ , the energy function  $E(\tau)$  can be designed for the CNN processor in the CNNU-based scheduler. However, some modifications

should be made to ensure the correctness of the desired output and the stability of the CNN processor. Firstly,  $c_i(t)$  is substituted by Eq. (4.11). The terms with only scalar are replaced by the terms consisted of state variable outputs; this just makes the resulting energy the same or a constant shift, which would be independent of the output pattern. Also, some additional auxiliary terms can be included to facilitate the convergence of CNN processor. Consequently,  $E(\tau)$  is given by

$$\begin{aligned}
E(\tau) = & -\eta_0 \left[ \sum_{i=1}^N \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) \cdot (1 - e^{-\sigma \mathcal{U}_i(t)}) \right] \\
& + \eta_1 \left[ \left( \sum_{i=1}^N \frac{1}{2} \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) - 1 \right) \cdot \left( \sum_{i=1}^N \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) \right) \right] \\
& + \eta_2 \left[ \sum_{i=1}^N \left( \left( \frac{1}{2} \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \right)^+ \cdot \right. \\
& \left. \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) \right] + \eta_3 \left[ \sum_{i=1}^N \left( \sum_{k=1}^K X_{i,k}(\tau) (1 - X_{i,k}(\tau)) \cdot 2^{-k} \right) \right]. \quad (4.20)
\end{aligned}$$

where  $\eta_3$  is a constant for additional auxiliary terms. The first item differs from the corresponding cost in Eq. (4.19) in that the scalar 1 is ignored. The  $\sum_{i=1}^N (\sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k}) \cdot (1 - e^{-\sigma \mathcal{U}_i(t)})$  is bounded above by 1 and has the same minimum as in the cost function. For the second and third terms, the quadratic forms in Eq. (4.19) are replaced by convex functions which merely contain state variable  $X_{i,k}(\tau)$  without any scalar. The local minimums would be the same; the resulting energy at any equilibrium would be shifted by a constant value, compared to the cost in Eq. (4.19) and independent of the inputs and the output pattern. The last term of Eq. (4.20) is an auxiliary factor which emerges to ensure the convergence because the energy function due to this auxiliary term approaches zero only when every state variable outputs approaches either one or zero.

By Eq. (4.14) and Eq. (4.20), the dynamics of each neuron in the proposed CNN for the CNNU-based scheduler can be expressed by

$$\frac{dX_{i,k}^{(s)}(\tau)}{d\tau} = -\frac{X_{i,k}^{(s)}(\tau)}{\nu_k} + \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k} - \eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1 \right) \cdot 2^{-k}$$

$$\begin{aligned}
& -\eta_2 \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right)^+ \cdot 2^{-k} \\
& -\eta_3 (1 - 2X_{i,k}(\tau)) \cdot 2^{-k} \\
= & -\frac{X_{i,k}^{(s)}(\tau)}{\nu_k} + \left[ 2\eta_3 \cdot 2^{-k} - \eta_1 \cdot 2^{-2k} - \eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \right. \right. \right. \\
& \left. \left. \left. \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot 2^{-2k} \right] \cdot X_{i,k}(\tau) + \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k} \\
& + \eta_2 \cdot u \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot \\
& \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \cdot 2^{-k} - \left( \sum_{j=1, j \neq i}^N \sum_{m=1}^K \eta_1 \cdot 2^{-(m+k)} \cdot X_{j,m}(\tau) \right) \\
& - \left( \sum_{m=1, m \neq k}^K \left[ \eta_1 - \eta_2 \cdot u \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \right. \right. \right. \right. \\
& \left. \left. \left. \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \right] \cdot 2^{-(m+k)} \cdot X_{i,m}(\tau) \right) + \eta_1 \cdot 2^{-k} - \eta_3 \cdot 2^{-k}, \tag{4.21}
\end{aligned}$$

where  $\nu_k$  is modified to  $2^k$  to retain the stability and desired output pattern of the designed CNN.

From Eq. (4.12) and Eq. (4.21), the recurrent interconnection weights, the external control weights, and the bias current can be determined by

$$\begin{cases}
A_{i,k;i;k} = -\eta_1 \cdot 2^{-2k} - \eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot 2^{-2k} + 2\eta_3 \cdot 2^{-k}, \\
B_{i,k;i;k}^1 = +\eta_0, \\
B_{i,k;i;k}^2 = +\eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right), \\
A_{i,k;j;m} = -\eta_1 \cdot 2^{-(k+m)} \cdot (1 - \delta_{k,m}) \delta_{i,j} - \eta_1 \cdot 2^{-(k+m)} \cdot (1 - \delta_{i,j}) \\
\quad - \eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot \delta_{i,j} \cdot 2^{-(k+m)}, \\
V_{i,k} = \eta_1 \cdot 2^{-k} - \eta_3 \cdot 2^{-k},
\end{cases} \tag{4.22}$$

where  $B_{i,k;i;k}^1$  and  $B_{i,k;i;k}^2$  are the external control weights for first and second external inputs,  $Y_{i,k}^{(1)} = (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k}$  and  $Y_{i,k}^{(2)} = \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \cdot 2^{-k}$ , respectively, and  $\delta_{x,y} = 1$  if  $x = y$ ,  $\delta_{x,y} = 0$  otherwise.

The range of coefficients  $\eta_0$ ,  $\eta_1$  ( $\eta_1^+$ ,  $\eta_1^-$ ),  $\eta_2$ , and  $\eta_3$  must be properly selected to ensure the stability and the desired response. For a tolerant error level  $\varepsilon$ , which is the maximum

difference between stable output  $\lim_{\tau \rightarrow \infty} \vec{c}(t, \tau)$  and the optimum  $\vec{c}^*(t)$ , the range of these coefficients are obtained as follows

$$\begin{cases} 0 < \eta_0 < \eta_3, \\ \eta_1^+ > 2^K, \\ \eta_1^- \geq \frac{\eta_0 \cdot 2^{-3}}{\varepsilon}, \\ \eta_2 \geq 2^K, \\ \eta_3 > \frac{1}{2} + \frac{\eta_1^-}{2}. \end{cases} \quad (4.23)$$

We have proved that with a matrix of given utility function and a matrix of radio resource assignment ratio upper limits, the proposed CNN architecture will converge to the neighborhood of the optimal pattern  $\vec{c}^*(t)$  within the difference  $\varepsilon$  with the range of coefficients given in Eq. (4.41). If  $\varepsilon \leq 2^{-K}$ , the CNN converges to  $\vec{c}^*(t)$  exactly shown in ?? .

However, the complexity of the interconnection is in the order of  $[N \times K]^{N \times K}$ , which is almost infeasible for practical implementation. In the next subsection, an equivalent two-layer structure for the CNN processor is proposed to efficiently reduce the complexity of interconnection.

To effectively reduce the complexity of interconnections, we propose an equivalent two-layer structure for the CNN processor which involves the first decision layer,  $[z_{i,k}^1]$ , with state variable output  $X_{i,k}(\tau)$ , and the second output layer,  $[z_{i,k}^2]$ , with state variable output  $c_i(t, \tau)$ . Fig. 4.2 shows the architecture of the two-layered CNN processor. The decision layer consists of  $N \times K$  neurons; the output layer is with an  $(N+1) \times 1$  array, where the output of the first neuron is the summation of all the others. The interconnections between the neurons of decision layer and those of output layer are defined by

- For the first decision layer to the second output layer, the connection weight between  $X_{i,k}(\tau)$  and  $c_j(t, \tau)$  is  $2^{-k}$ ,  $\forall k$  if  $j = i$ ; is zero if  $j \neq i$ .
- For the second output layer feedback to the first decision layer, only the first neuron output is connected to the  $X_{i,k}(\tau)$  of the decision layer with the interconnection weight  $\eta_1 \cdot 2^{-k}$  for  $\forall i$ .

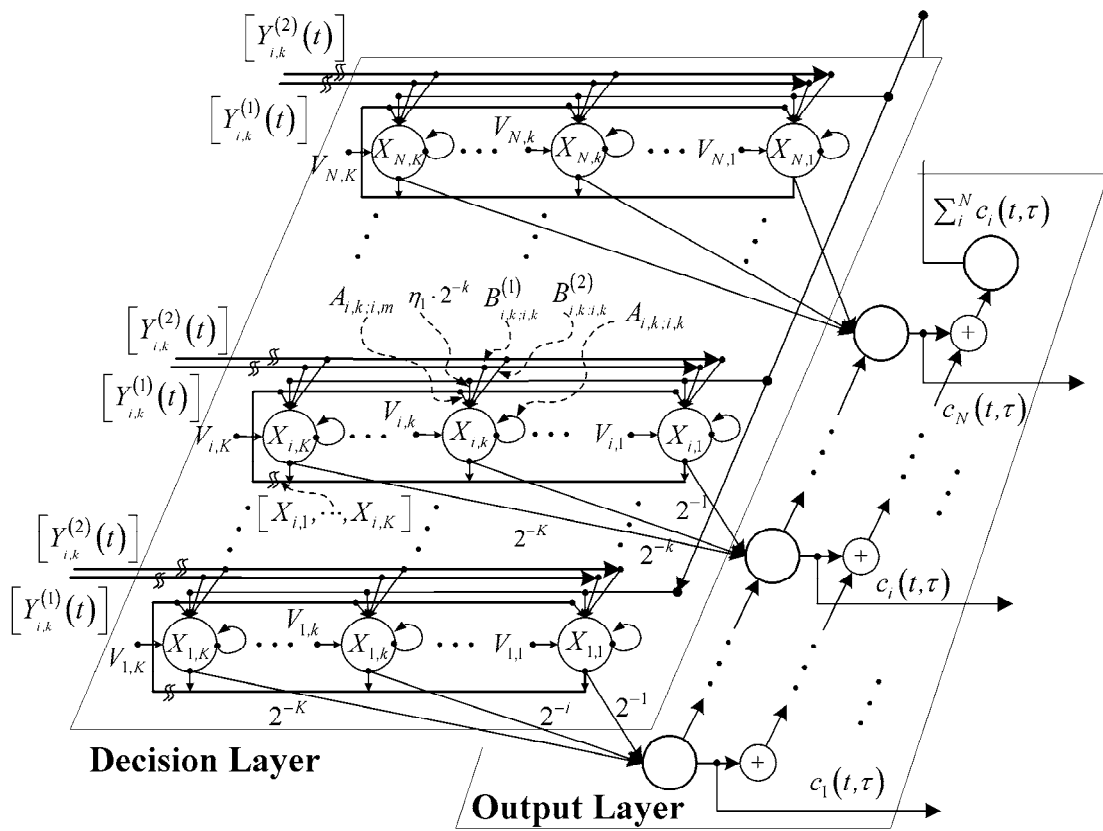


Figure 4.2: The two-layer structure of CNN processor.

## 4.5.4 The Two-Layer Structure for CNN Processor

The recurrent interconnection weights and the external control weights for the first decision layer defined in Eq. (4.22) are then modified to be

$$\begin{cases} A_{i,k;i;k} &= -\eta_1 \cdot 2^{-2k} - \eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot 2^{-2k} + 2\eta_3, \\ B_{i,k;i;k}^1 &= \eta_0, \\ B_{i,k;i;k}^2 &= \eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right), \\ A_{i,k;j;m} &= -\eta_2 \cdot u \left( \sum_{k=1}^K X_{i,k} - \min \left\{ \frac{W \cdot \log_2 M_{\kappa_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\} \right) \cdot \delta_{i,j} \cdot 2^{-(k+m)}, \\ V_{i,k} &= \eta_1 \cdot 2^{-k} - \eta_3, \end{cases} \quad (4.24)$$

For the second output layer, there are no external inputs, and only recurrent interconnection weights exist. The interconnection weight between  $c_i(t, \tau)$  and  $c_j(t, \tau)$  is given by  $\delta_{0,j}$  with  $i = 0$ .

It can be shown that the two-layer structured CNN processor has the same energy function and the local minimum as the single-layer one defined by Eq. (4.22). However, the complexity of interconnections in the former one is proportional to  $[3N \times K + N]$ , which is significantly lower than  $[N \times K]^{N \times K}$  in the latter one.

## 4.6 The Discussion for Stability and Convergence of CNN Processor

### 4.6.1 The relationship between the dynamics each cell and Energy function of CNN

*Lemma 1:*

With energy function defined in Eq. (4.13), the Eq. (4.12) can be expressed in terms of Lyapunov function by

$$\frac{dX_{i,k}(\tau)}{d\tau} = -\frac{X_{i,k}(\tau)}{\nu} - \frac{\partial E(\tau)}{\partial X_{i,k}(\tau)}. \quad (4.25)$$

*Proof:*



The assumptions on interconnection weights and external control weights are given by

$$A_{i,k;j,m} = \begin{cases} 1, & z_{j,m} \in Z_n(i,k) \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B_{i,k;j,m} = \begin{cases} 1, & z_{j,m} \in Z_n(i,k) \\ 0, & \text{otherwise} \end{cases}. \quad (4.26)$$

With a given Lyapunov function  $E(\tau)$ ,

$$\begin{aligned} E(\tau) &= -\frac{1}{2} \sum_i \sum_k A_{i,k;i,k} X_{i,k}(\tau)^2 - \frac{1}{2} \sum_i \sum_k \sum_j \sum_m A_{i,k;j,m} X_{i,m}(\tau) X_{i,k}(\tau) \\ &\quad - \frac{1}{2} \sum_i \sum_k \sum_j \sum_m B_{i,k;j,m} Y_{j,m} X_{i,k}(\tau) - \sum_i \sum_k V_{i,k} \cdot X_{i,k}(\tau). \end{aligned} \quad (4.27)$$

With condition Eq. (4.26), the partial derivative of Lyapunov function with respect to  $X_{i,k}(\tau)$  as follows:

$$\begin{aligned} \frac{\partial E(\tau)}{\partial X_{i,k}(\tau)} &= A_{i,k;i,k} X_{i,k}(\tau) - \frac{1}{2} \sum_j \sum_m A_{i,k;j,m} X_{i,m}(\tau) - \frac{1}{2} \sum_j \sum_m B_{i,k;j,m} Y_{j,m} - V_{i,k} \\ &= A_{i,k;i,k} X_{i,k}(\tau) - \frac{1}{2} \sum_{z_{j,m} \in Z_n(i,k)} A_{i,k;j,m} X_{i,m}(\tau) - \frac{1}{2} \sum_{z_{j,m} \in Z_n(i,k)} B_{i,k;j,m} Y_{j,m} \\ &\quad - V_{i,k}. \end{aligned} \quad (4.28)$$

Therefore, the dynamics of each neuron in Eq. (4.12) can be easily rewritten by Eq. (4.25). ■

With this relationship, we get some view further inside the dynamics of CNN model in the sense of Lyapunov function. The direction of the change rate (slope) of each neuron is proportional to the partial derivative of Lyapunov function. When the CNN with pre-defined trajectories are designed, Lyapunov method is a good approach.

## 4.6.2 The Stability of Designed CNN

*Lemma 2:*

The energy function defined in Eq. (4.13) will decrease along with time, namely,  $\frac{dE(\tau)}{d\tau} \leq 0$ .

*Proof:*

$$\begin{aligned}
\frac{dE(\tau)}{d\tau} &= -\eta_0 \sum_{i=1}^N \sum_{k=1}^K \frac{\partial \left[ X_{i,k}(\tau) \cdot 2^{-k} \cdot \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) \right]}{\partial X_{i,k}(\tau)} \cdot \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \\
&+ \eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) \cdot \sum_{i=1}^N \sum_{k=1}^K \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \\
&- \eta_1 \sum_{i=1}^N \sum_{k=1}^K \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \\
&+ \eta_2 \sum_{i=1}^N \phi_{2,i} \cdot \left[ \left( \sum_{k=1}^K \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \right) \cdot \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} \right) \right. \\
&\quad \left. - \psi_{2,i} \cdot \sum_{k=1}^K \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \right] \\
&+ \eta_3 \sum_{i=1}^N \sum_{k=1}^K (1 - 2X_{i,k}(\tau)) \cdot \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \\
&= -\eta_0 \sum_{i=1}^N \sum_{k=1}^K \left[ \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) \cdot 2^{-k} \right] \cdot \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \\
&+ \eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1 \right) \cdot \sum_{i=1}^N \sum_{k=1}^K \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \\
&+ \eta_2 \sum_{i=1}^N \phi_{2,i} \cdot \left[ \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \psi_{2,i} \right) \left( \sum_{k=1}^K \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \right) \right] \\
&+ \eta_3 \sum_{i=1}^N \sum_{k=1}^K (1 - 2X_{i,k}(\tau)) \frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau}, \tag{4.29}
\end{aligned}$$

where  $\psi_{2,i} = \min \left\{ \frac{W \cdot \log_2 M_{rs_i}}{SF_i \cdot \mathcal{R}_i(t)}, \frac{Q_i(t)/T_f}{\mathcal{R}_i(t)} \right\}$ , and  $\phi_{2,i} = \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - \psi_{2,i} \right)$ .

From the definition of activation of each neuron,  $\frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)}$  has following values:

$$\frac{dX_{i,k}(\tau)}{dX_{i,k}^{(s)}(\tau)} = \begin{cases} 1, & \text{if } 0 < X_{i,k}^{(s)}(\tau) < 1 \\ 0, & \text{if } 0 > X_{i,k}^{(s)}(\tau) \text{ or } X_{i,k}^{(s)}(\tau) > 1 \end{cases} \tag{4.30}$$

For the transient process of the CNN, that is,  $\{0 < X_{i,k}^{(s)}(\tau) < 1\}$ ,  $\forall(i, k)$ , the Eq. (4.29) can be rearranged by

$$\frac{dE(\tau)}{d\tau} = -\eta_0 \sum_{i=1}^N \sum_{k=1}^K \left[ \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) \cdot 2^{-k} \right] \cdot \frac{dX_{i,k}(\tau)}{d\tau}$$

$$\begin{aligned}
& +\eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1 \right) \cdot \sum_{i=1}^N \sum_{k=1}^K \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \\
& +\eta_2 \sum_{i=1}^N \phi_{2,i} \cdot \left[ \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \psi_{2,i} \right) \left( \sum_{k=1}^K \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \cdot 2^{-k} \right) \right] \\
& +\eta_3 \sum_{i=1}^N \sum_{k=1}^K (1 - 2X_{i,k}(\tau)) \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \\
= & \sum_{i=1}^N \sum_{k=1}^K \left\{ -\eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k} + \eta_1 \left( \sum_{j=1}^N \sum_{m=1}^K X_{j,m}(\tau) \cdot 2^{-m} - 1 \right) \cdot 2^{-k} \right. \\
& \left. + \eta_2 \phi_{2,i} \cdot \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \psi_{2,i} \right) + \eta_3 (1 - 2X_{i,k}(\tau)) \right\} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \\
\leq & \sum_{i=1}^N \sum_{k=1}^K \left\{ \frac{X_{i,k}^{(s)}(\tau)}{\nu} - \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k} + \eta_1 \left( \sum_{j=1}^N \sum_{m=1}^K X_{j,m}(\tau) \cdot 2^{-m} - 1 \right) \cdot 2^{-k} \right. \\
& \left. + \eta_2 \phi_{2,i} \cdot \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \psi_{2,i} \right) + \eta_3 (1 - 2X_{i,k}(\tau)) \right\} \cdot \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \\
= & - \sum_{i=1}^N \sum_{k=1}^K \left( \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} \right)^2 \\
\leq & 0.
\end{aligned} \tag{4.31}$$

When each neuron arrives at saturation region,  $\frac{dE(\tau)}{d\tau} = 0$ . ■

*Lemma 3:*

With any given external input vectors and any initial states, the energy function will approach a stable state. We have  $\lim_{\tau \rightarrow \infty} E(\tau) = \text{const}$  and  $\lim_{\tau \rightarrow \infty} \frac{dE(\tau)}{d\tau} = 0$ .

*Proof:*

We first prove that  $|E(\tau)|$  is bounded. With  $X_{i,k}(\tau) \in (0, 1)$ , and  $\sum_{i=1}^N \sum_{k=1}^K (X_{i,k}(\tau) 2^{-k}) \leq 1$  in any transient and steady states, Eq. (4.20) can be re-written as:

$$\begin{aligned}
\|E(\tau)\| & \leq \frac{\eta_0}{2} \left[ 1 - \sum_{i=1}^N (1 - e^{-\sigma \mathcal{U}_i(t)}) \right]^2 + \eta_1 \cdot (N - 1)^2 \\
& + \frac{\eta_2}{2} \cdot \sum_{i=1}^N (1 - \psi_{2,i})^2 + \eta_3 N^2
\end{aligned}$$

$$\leq \left(\frac{\eta_0}{2} + \eta_1\right) \cdot (N-1)^2 + \left(\frac{\eta_2}{2} + \eta_3\right) N^2. \quad (4.32)$$

Since  $N$  is finite, the therefore  $\|E(\tau)\|$  is bounded in any trajectory. For a more tight upper and lower bound for  $\|E(\tau)\|$ , we can further have

Since  $|E(\tau)| \leq E_{max}$  and  $\frac{dE(\tau)}{d\tau} \leq 0$ , we can easily have

$$\lim_{\tau \rightarrow \infty} E(\tau) = const.$$

■

Define the function  $\mathcal{J}_{i,k}(\vec{X}(\tau))$  of  $\vec{X}(\tau)$  and the input vector  $\vec{U}(t)$  for state variable  $X_{i,k}(\tau)$  by

$$\begin{aligned} \mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) &= -X_{i,k}(\tau) + \eta_0 \left(1 - e^{-\sigma U_i(t)}\right) \cdot 2^{-k} \\ &- \eta_1 \left(\sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1\right) \cdot 2^{-k} \\ &- \eta_2 \cdot \phi_{2,i} \cdot \left(\sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \psi_{2,i}\right) \cdot 2^{-k} - \eta_3 (1 - 2X_{i,k}(\tau)) \end{aligned} \quad (4.33)$$

And also define the function  $\mathcal{J}_{\nabla E}(\vec{X}(\tau))$  by

$$\mathcal{J}_{\nabla E}(\vec{X}(\tau), \vec{U}) = \prod_{i,k} \mathcal{J}_{i,k}(\vec{X}(\tau)) \cdot (2X_{i,k}(\tau) - 1). \quad (4.34)$$

*Lemma 4: Equilibrium Criteria*

For a output pattern  $\vec{X}(\tau) \in \{0, 1\}^{N \times K}$ , it is an equilibria for the network defined in Eq. (4.21) if the condition

$$\mathcal{J}_{\nabla E}(\vec{X}(\tau), \vec{U}(t)) > 0 \quad (4.35)$$

holds.

*Proof:*

According to the activation function  $f(X_{i,k}^{(s)}(\tau))$  of each neuron and Eq. (4.21), a state

variable output attains an equilibria when the state variable reaches the saturation region, that is,  $\frac{dX_{i,k}(\tau)}{d\tau} = 0$ , and

$$\begin{cases} X_{i,k}^{(s)}(\tau) > 1, & \text{as } X_{i,k}(\tau) = 1 \\ X_{i,k}^{(s)}(\tau) < 0, & \text{as } X_{i,k}(\tau) = 0 \end{cases} . \quad (4.36)$$

From  $\frac{dX_{i,k}^{(s)}(\tau)}{d\tau} = 0$  and the definition in Eq. (4.33), we have

$$\begin{aligned} \mathcal{J}_{i,k}(\vec{X}(\tau)) &= -X_{i,k}(\tau) + \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) \cdot 2^{-k} \\ &- \eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1 \right) \cdot 2^{-k} \\ &- \eta_2 \cdot \phi_{2,i} \cdot \left( \sum_{m=1}^K X_{i,m}(\tau) \cdot 2^{-m} - \psi_{2,i} \right) \cdot 2^{-k} - \eta_3 (1 - 2X_{i,k}(\tau)) \\ &= \mathcal{J}_{i,k}(\vec{X}(\tau), \vec{\mathcal{U}}(t)) + X_{i,k}(\tau). \end{aligned} \quad (4.37)$$

Based on Eq. (4.21), the criteria in Eq. (4.36) can be rewritten by

$$X_{i,k}^{(s)}(\tau) - \delta(X_{i,k}(\tau)) \begin{cases} > 0, & \text{as } X_{i,k}(\tau) = 1 \\ < 0, & \text{as } X_{i,k}(\tau) = 0 \end{cases} , \quad (4.38)$$

where  $\delta(X_{i,k}(\tau)) = X_{i,k}(\tau)$  is the diract-function with unit value if  $X_{i,k}(\tau) = 1$ . Then the criteria in Eq. (4.38) is equivalent to

$$\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{\mathcal{U}}) \begin{cases} > 0, & \text{as } X_{i,k}(\tau) = 1 \\ < 0, & \text{as } X_{i,k}(\tau) = 0 \end{cases} , \quad (4.39)$$

Therefore, the resulting criteria for a state variable is

$$\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{\mathcal{U}}) \cdot (2X_{i,k}(\tau) - 1) > 0. \quad (4.40)$$

With a given output pattern  $\vec{X}(\tau)$ , the intersection of equilibrium criteria for all state variables defined in Eq. (4.35) should be hold. ■

Before the lemma of desired output pattern property, the range of the coefficients of terms in the energy function is given to ensure the convergence of desired output pattern. For auxiliary term 1 only, it can be given by

$$\begin{cases} \eta_1^+ > 2^K, * \\ \eta^- \geq \frac{\eta_0 \cdot 2^{-3}}{\varepsilon}, \\ \eta_2 \geq 2^K, \\ \eta_0 < \eta_3 - 2, \\ \eta_1 \gg \eta_3 > \eta_0 > 0, \end{cases} \quad (4.41)$$

For auxiliary terms, it can be given by

$$\begin{cases} \eta_3 > \frac{3}{2} + \frac{\eta_2 - \eta_5}{2}, \\ \eta_1 \geq 2^{K-1}, \\ \eta_3 + 2 > \eta_0 > 2 - \eta_3, \\ \eta_5 \leq \eta_1, \end{cases} \quad \text{and } \eta_5 \cdot (1 - e^{-\sigma U_i(t)}) \leq 1 \quad (4.42)$$

*Lemma 5:*

As  $\lim_{\tau \rightarrow \infty} \frac{dE(\tau)}{d\tau} = 0$ ,  $\lim_{\tau \rightarrow \infty} X_{i,k}(\tau) = \text{const.}, \forall (i, k)$ . And the resulting output locates in the reasonable state space, that is,  $\vec{X}(\tau) \in \Psi_1 \cap \Psi_2$ .

*Proof:*

First, we show that the output of each neuron will eventually arrive at saturation region. According to Theorem 5 in [66], the sufficient and necessary condition to force the state variable into saturation region is that

$$A_{i,k;i,k} > \frac{1}{\nu_k}. \quad (4.43)$$

It can be further interpreted as

$$\begin{aligned} \frac{dX_{i,k}^{(s)}(\tau)}{d\tau} &= -\frac{X_{i,k}^{(s)}(\tau)}{\nu_k} + T_{i,k;i,k}^{(r)} X_{i,k}(\tau) + \Phi(\vec{X}') \\ &= \frac{1}{\nu_k} (\rho \cdot f(X_{i,k}^{(s)}(\tau)) - X_{i,k}^{(s)}(\tau)) + \Phi(\vec{X}') \\ &= \begin{cases} \frac{(\rho-1)}{\nu_k} \cdot X_{i,k}^{(s)}(\tau) + \Phi(\vec{X}'), & 0 < X_{i,k}^{(s)}(\tau) < 1 \\ 0, & X_{i,k}^{(s)}(\tau) \in \text{Saturation Region} \end{cases}, \end{aligned} \quad (4.44)$$

where  $\rho = \nu_k \cdot T_{i,k;i,k}^{(r)} > 1$ . Since  $\rho - 1 > 0$ , the differential equation has the form of positive feedback, and it will have  $X_{i,k}^{(s)}(\tau)$  enter the saturation region of which the value

is determined by  $\Phi(\vec{X}')$ . For  $\nu_k \cdot A_{i,k;i,k} = -\eta_1 - \eta_2 \phi_{2,i} + 2\eta_3$  with  $\eta_1 \gg 1$ , it can be seen that  $\nu_k \cdot A_{i,k;i,k} < 0$  when  $\phi_1 = 1$ . In this case,  $\frac{dX_{i,k}^{(s)}(\tau)}{d\tau} < 0$  for  $\forall z_{i,k}$  so that the constraint  $\Psi_1$  can be fulfilled. After  $\phi_1 = 0$ ,  $\nu_k \cdot A_{i,k;i,k} = -2\eta_1^0 - \eta_2 \phi_{2,i} + 2\eta_3$ . Since  $\eta_3 > \frac{3}{2} + \frac{\eta_2}{2}$ , it can be found that

$$\begin{aligned} \nu_k \cdot A_{i,k;i,k} &= -2\eta_1^0 - \eta_2 \phi_{2,i} + 2\eta_3 \\ &> 3 + 2 \cdot \frac{\eta_2}{2} - 2 - \eta_2 \phi_{2,i} \\ &> 1, \end{aligned} \tag{4.45}$$

the condition Eq. (4.43) holds, and every output of the state variable will eventually enter the saturation region under constraint  $\Psi_1$ .

It is left to be proved that the stable output patterns locate in the state space. To prove the completeness of the state space defined by  $\Psi_1 \cap \Psi_2$ , we first prove that all desired output patterns  $\{\vec{X}(\tau) \in \Psi_1 \cap \Psi_2\}$  are equilibria of the CNN. We then show that all  $\vec{X}(\tau)$  converges into  $\{\Psi_1 \cap \Psi_2\}$ .

For any given  $\vec{X}(\tau) \in \Psi_1 \cap \Psi_2$ , we have  $\sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \leq 1$  and  $\phi_1 = 0$ , and also  $\sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \leq \psi_{2,i}$  and  $\phi_{2,i} = 0$  for  $\forall i$ . For  $X_{i,k}(\tau) = 0$  for  $\forall(i, k)$ , the Eq. (4.33) is

$$\begin{aligned} \mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) &= \left[ \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) + 2 \left( 1 - \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) - \eta_3 \right] \cdot 2^{-k} \\ &< \left[ \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) + 2 \left( 1 - \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) - \eta_0 - 2 \right] \cdot 2^{-k} \\ &< - \left[ \eta_0 \cdot e^{-\sigma \mathcal{U}_i(t)} + 2 \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right] \cdot 2^{-k} \\ &\leq 0, \end{aligned} \tag{4.46}$$

where  $(\eta_0 < \eta_3 - 2)$  is given from Eq. (4.41). Then,

$$\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) \cdot (2X_{i,k}(\tau) - 1) = -\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) > 0$$

for all  $X_{i,k}(\tau) = 0$ .

For  $X_{i,k}(\tau) = 1$  for  $\forall(i, k)$ , the Eq. (4.33) would be

$$\begin{aligned} \mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) &= \left[ \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) + 2 \left( 1 - \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \right) + \eta_3 \right] \cdot 2^{-k} \\ &\geq \eta_3 \cdot 2^{-k} > 0 \end{aligned} \quad (4.47)$$

where  $\eta_3 > 0$  is given from Eq. (4.41). Then,

$$\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) \cdot (2X_{i,k}(\tau) - 1) = \mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) > 0$$

for all  $X_{i,k}(\tau) = 1$ . In summary, we can conclude that the elements  $\vec{X}$  of  $\Psi_1 \cap \Psi_2$  satisfies condition Eq. (4.35) and are all stable equilibria.

We now prove that  $\Psi_1 \cap \Psi_2$  contains all equilibria of the CNN. Define  $\Omega^*$  to be the collection of all equilibria for Eq. (4.35), that is,

$$\Omega^* = \left\{ \vec{X} \mid \mathcal{J}_{\nabla E}(\vec{X}) > 0 \right\}.$$

Then we claim that

$$\Omega^* \subseteq \Psi_1. \quad (4.48)$$

Define

$$\Lambda_1(i, k)(\tau) = \sum_{(j,m) \neq (i,k)} X_{j,m}(\tau) \cdot 2^m - 1, \quad (4.49)$$

and also

$$\Lambda_2(i, k)(\tau) = \left( \sum_{m \neq k} X_{i,m}(\tau) \cdot 2^m - \psi_{2,i} \right). \quad (4.50)$$

We first consider the case that the selected equilibrium  $\vec{X}$  satisfies constraint  $\Psi_2$  with  $\phi_{2,i} = 0$ . For  $X_{i,k}(\tau) = 1$ , Eq. (4.33) is greater than zero. Assume  $\Lambda_1(i, k) > -2^{-k} + 2^{K+1}$  and therefore  $\phi_1 = 1$ .

$$\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) = \left[ \eta_0 (1 - e^{-\sigma \mathcal{U}_i(t)}) - 2\eta_1 \cdot 2^{-k} - 2\eta_1 \Lambda_1(i, k)(\tau) + \eta_3 \right] \cdot 2^{-k}$$



$$\begin{aligned}
&< \left[ \eta_0 + \eta_3 - 2\eta_1 \cdot 2^{-k} \right] \cdot 2^{-k} \\
&< \left[ \eta_0 + \eta_3 - (\eta_0 + \eta_3) \cdot 2^{K-k} \right] \cdot 2^{-k} \\
&\leq 0 \qquad \qquad \qquad \forall k \leq K,
\end{aligned} \tag{4.51}$$

with the range of  $\eta_1 \geq (\eta_0 + \eta_3) \cdot 2^{K-1}$  given from Eq. (4.41). This contradicts Eq. (4.35) and  $\Lambda_1(i, k)$  must be equal to or less than zero. Therefore, if the output of the state variable  $X_{i,k}(\tau) = 1$  and constraint  $\Psi_2$  is satisfied,

Similarly, with  $\eta_2 \geq (\eta_0 + \eta_3) \cdot 2^{K-1}$ , we can also obtain the same conclusion.

On the other hand, for  $X_{i,k}(\tau) = 0$ , Eq. (4.33) is less than zero. For any time  $\tau$ , when  $\vec{X}(\tau) \notin \Psi_1 \cap \Psi_2$ , equivalently  $\phi_1 = 1$  or  $\phi_{2,i} = 1$ , and the output of the specific state variable  $X_{i,k}(\tau) = 0$ , it can be directly shown that

$$\begin{aligned}
\mathcal{J}_{i,k}(\vec{X}(\tau), \vec{U}) &= \left[ \eta_0 \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) - 2\eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1 \right) \right. \\
&\quad \left. - \eta_2 \cdot \phi_{2,i} \cdot \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - \psi_{2,i} \right) - \eta_3 \right] \cdot 2^{-k} \\
&< \left[ \eta_0 \left( 1 - e^{-\sigma \mathcal{U}_i(t)} \right) - 2\eta_1 \left( \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - 1 \right) \right. \\
&\quad \left. - \eta_2 \cdot \phi_{2,i} \cdot \left( \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} - \psi_{2,i} \right) - \eta_0 - 2 \right] \cdot 2^{-k} \\
&\leq 0.
\end{aligned} \tag{4.52}$$

This result indicates that when the constraints are not satisfied, the state variables will be inhibited to reach the stable equilibrium in the state space. As long as the output of a specific state variable reaches lower saturation region, it will then be stable, and the equilibrium will be dominant by other state variables in upper saturation region. From the both cases,  $X_{i,k}(\tau) = 0$  and  $X_{i,k}(\tau) = 1$ , we can conclude that  $\forall \vec{X} \in \Omega^*$  must belong to the state space, and  $\Omega^* \subseteq \Psi_1$ . ■

Denote by  $\nabla_\varepsilon \Psi_1$  the boundary surface of  $\Psi_1$  with width  $2^{-\varepsilon}$ , defined by

$$\nabla_\varepsilon \Psi_1 = \left\{ \vec{X}(\tau) \mid \left(1 - \|\vec{X}(\tau)\|\right) < 2^{-\varepsilon}, \vec{X} \in \Psi_1 \right\}.$$

When  $\varepsilon = K$ ,  $\nabla_K \Psi_1$  is the collection of the vertex of  $\Psi_1$ .

For a given desired level  $\varepsilon$ , the range of all coefficients are modified as follows

$$\begin{cases} \eta_3 > \frac{1}{2} + \frac{\eta_1^-}{2}, \\ \eta_1^- \geq 2^K, \\ \eta_1^- \geq \frac{2^{K+3}}{\eta_0}, \\ \eta_2 \geq 2^K, \\ \eta_0 < \eta_3, \\ \eta_1 \gg \eta_3 > \eta_0 > 0. \end{cases} \quad (4.53)$$

And denote  $\Omega_\varepsilon^*$  as the collection of all equilibrium for the CNN with Eq. (4.53) for a given  $\varepsilon$ .



*Lemma 6:*

For a given desired level  $\varepsilon$  and the ranges of all coefficients in Eq. (4.53), any equilibrium output  $\vec{X}$  of the proposed CNN locates in  $\nabla_\varepsilon \Psi_1 \cap \Psi_2$ .

*Proof:* With Eq. (4.53), it can be easily shown that  $\Omega_\varepsilon^*$  defined in Lemma 5 belongs to  $\Psi_1 \cap \Psi_2$ . We now show that with a given  $\varepsilon$ ,  $\Omega_\varepsilon^* \subseteq \nabla_\varepsilon \Psi_1 \cap \Psi_2$ . Since any  $\vec{X}$  of  $\Omega_\varepsilon^*$  belongs to  $\Psi_1 \cap \Psi_2$ , we have  $\sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \leq 1$  and  $\phi_1 = 0$ , and also  $\sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k} \leq \psi_{2,i}$  and  $\phi_{2,i} = 0$  for  $\forall i$ . Assume  $\Lambda_1(i, k)$  given in Eq. (4.49) is greater than or equal to  $2^{-\varepsilon}$ . For  $X_{i,k}(\tau) = 0$  for  $\forall(i, k)$ , the Eq. (4.33) is

$$\begin{aligned} \mathcal{J}_{i,k} \left( \vec{X}(\tau), \vec{U} \right) &= \left[ \eta_0 \left(1 - e^{-\sigma \mathcal{U}_i(t)}\right) + \eta_1^+ \left(1 - \sum_{i=1}^N \sum_{k=1}^K X_{i,k}(\tau) \cdot 2^{-k}\right) - \eta_3 \right] \cdot 2^{-k} \\ &> \left[ \eta_0 \left(1 - e^{-\sigma \mathcal{U}_i(t)}\right) - \eta_1^+ \Lambda_1(i, k) - \eta_0 \right] \cdot 2^{-k} \\ &> -\eta_0 \left[ e^{-\sigma \mathcal{U}_i(t)} - 2^\varepsilon \Lambda_1(i, k) \right] \cdot 2^{-k}, \end{aligned} \quad (4.54)$$

where  $\Lambda_1(i, k) = \left( \sum_{j=1}^N \sum_{m=1}^K X_{j,m}(\tau) \cdot 2^{-m} - 1 \right)$  as  $X_{i,k}(\tau) = 0$ . We can show that for the definition of  $\mathcal{U}_i(t)$  and the proper selection of  $\sigma$ , the following property holds,

$$\lim_{N \rightarrow \infty} \mathbf{P}_r \left\{ \bigcup_{i \leq N} \left\{ e^{-\sigma \mathcal{U}_i(t)} < 2^{-3} \right\} \right\} = 1.$$

Then, Eq. (4.54) becomes

$$\mathcal{J}_{i,k} \left( \vec{X}(\tau), \vec{\mathcal{U}} \right) > -\eta_0 \left[ 2^{-3} - 2^\varepsilon \Lambda_1(i, k) \right] \cdot 2^{-k}, \quad (4.55)$$

$$\mathcal{J}_{i,k} \left( \vec{X}(\tau), \vec{\mathcal{U}} \right) \cdot (2X_{i,k}(\tau) - 1) = -\mathcal{J}_{i,k} \left( \vec{X}(\tau), \vec{\mathcal{U}} \right) > 0$$

for all  $X_{i,k}(\tau) = 0$ . ■

Define  $2^{-\varepsilon}$ -neighborhood of  $\vec{X}^*$  as the set  $\left\{ \vec{X} \mid \|\vec{X} - \vec{X}^*\| < 2^{-\varepsilon}, \forall \vec{X} \in \nabla_\varepsilon \Psi_1 \cap \Psi_2 \right\}$ .

In the following, we prove that the energy function is absolutely monotonic in the region  $\nabla_\varepsilon \Psi_1 \cap \Psi_2$ , and the optimality is guaranteed within the radius  $2^{-\varepsilon}$  around the optimum.

*Lemma 7:*

With a given set of utility values  $\vec{\mathcal{U}}(t)$ , there exists an optimal output pattern, denoted by  $\vec{X}^*$ , with lowest cost. Then, the equilibrium of the CNN converges to the  $2^{-\varepsilon}$ -neighborhood of  $\vec{X}^*$ .

*Proof:*

Define by  $E \left( \vec{X}(\tau), \vec{\mathcal{U}}(t) \mid \Psi_2 \right)$  the energy function of the output pattern at equilibrium  $\vec{X}(\tau)$ , the given utility vector  $\vec{\mathcal{U}}(t)$  at time  $t$ , and the system constraint  $\Psi_2$ . Let  $\Delta(\vec{X})$  be the difference  $(1 - \|\vec{X}\|)$  with a given equilibrium  $\vec{X}$ . For  $\forall \vec{X} \in \nabla_\varepsilon \Psi_1 \cap \Psi_2$ , the function  $E \left( \vec{X}(\tau), \vec{\mathcal{U}}(t) \mid \Psi_2 \right)$  is shown to become

$$\begin{aligned} E \left( \vec{X}(\tau), \vec{\mathcal{U}}(t) \mid \Psi_2 \right) &= -\vec{X} \cdot \vec{Y}^{(1)} + \left( 1 - \Delta(\vec{X}) - 2 \right) \cdot \left( 1 - \Delta(\vec{X}) \right) \\ &= -\vec{X} \cdot \vec{Y}^{(1)} - \left( 1 - \Delta(\vec{X}) \right)^2 \\ &= - \left( 1 + \vec{X} \cdot \vec{Y}^{(1)} \right) + \Delta(\vec{X})^2, \end{aligned} \quad (4.56)$$

where  $\vec{Y}^{(1)} = (\dots, Y_{i,k}^{(1)}, \dots)$  is the vector of the first external input of all neurons. With given  $\varepsilon$ , the  $\Delta(\vec{X})$  is less than  $2^{-\varepsilon}$ . In Eq. (4.56), the first item  $\left(1 + \vec{X} \cdot \vec{Y}^{(1)}\right)$  is monotonic in every dimension, and has no local minimum other than  $\vec{X}^*$ . From *Lemma 1* and *Lemma 6*,  $\frac{dE(\tau)}{d\tau} < 0$  and  $-\left(1 + \vec{X} \cdot \vec{Y}^{(1)}\right)$  will eventually converge to  $-\left(1 + \vec{X}^* \cdot \vec{Y}^{(1)}\right)$  with the difference less than  $2^{-\varepsilon}$ . If we select  $\varepsilon > K/2$ , the second term is bounded by  $\Delta(\vec{X})^2 < 2^{-K}$ , and it will change the equilibrium point to another. Therefore, the equilibrium of the CNN converges to the  $2^{-\varepsilon}$ -neighborhood of  $\vec{X}^*$ . ■

### 4.6.3 Concluding Remarks of Designing Energy Function Based CNN by Lyapunov Method

Here are some conclusions on designing a CNN with desired trajectories:

1. Investigate the dynamics of each neuron, and the relationship between the energy function and each neuron, namely,  $\frac{\partial E(\vec{X}(\tau))}{\partial X_{i,k}(\tau)}$  and  $\frac{dX_{i,k}(\tau)}{d\tau}$
2. Identify the state variables and its desired output variable in terms of the considered problem.
3. Based on the objective function  $\mathcal{O}(\tau)$  and the system constraints  $\Psi_1, \dots$ , formulate the cost function in the form as  $\mathcal{H}(\tau) = \eta_0 \mathcal{O}(\tau) + \sum_k \eta_k \mathcal{C}_{\Psi_k}(\tau)$ .
4. Transform the cost function into energy function, and add some auxiliary terms to aid the convergence. Examine the physical meaning and the moving direction of each component of the energy function. Make sure that each component of the energy function is in the form of convex or decreasing function along with the trajectories or/and in time so that a local minimum exists.
5. Partially differentiate the energy function with respect to the state variable output, and find out the connection weights, including recurrent inter-connection and

external control weights, and also the bias current.

6. Verify that the equilibrium states exists, and that the state variables outputs in equilibrium state locate in the reasonable state space, and the output patterns has the lowest energy than any other pattern in the state space with given external inputs.

## Remarks on auxiliary terms in Lyapunov function

These are some experiences to design a stable circuit. In addition to the objective function and the system constraint function, we can add several auxiliary components with unchanged objective into the Lyapunov function. Notice that "unchanged objective" indicates that the auxiliary terms at equilibrium states are constant or zero so that the desired output pattern for the original objective function remains. The first one is to nudge the state variable outputs to the stable region. In this case, the cost of a state in different region with unchanged objective can be expressed as

$$\mathcal{C}_{Aux1}(\vec{X}(\tau)) = \begin{cases} 0, & \forall X_{i,k}(\tau) \in X^* = \{0, 1\} \\ \rho(X_{i,k}(\tau), X^*), & \forall X_{i,k}(\tau) \notin \{0, 1\} \end{cases}, \quad (4.57)$$

where  $\rho(X_{i,k}(\tau), X^*)$  denotes the distance between point Z and the saturation set  $X^*$ . Since at the equilibrium states this auxiliary term entails no further cost, this cost is only nonzero when the states are in transient situation if stable states are known. The second is in general to aid the Lemma 3 being hold where energy function is drawn to decrease along with time.

It is important for the designer to check whether the objective function is monotonically decreasing or partially (transiently) increasing to reach the equilibrium states. If it is monotonically decreasing, the convergence process has been well designed. Otherwise, some other auxiliary constraints might be added to ensure the convergence.

In this work, for example,  $c_i(0) = \sum_k X_{i,k}(0)$  is set higher than the eventual output value. A possibility of initial value could be  $\psi_{2,i}$ . Along with time, the  $c_i(\tau)$  will be decayed. There are two possible transient situations: one is intra  $c_i(\tau)$  inter-change between different bits when  $c_i(\tau)$  decreases; the other is that  $c_i(\tau)$  increases as  $c_j(\tau)$  decreases. In the first case, some MSB (more significant bit) becomes zero and other LSBs (less significant bit) become 1, that is,

$$\begin{aligned} X_{i,k}(\tau - 1) = 1 &\rightarrow, X_{i,k}(\tau) = 0, \\ X_{i,m}(\tau - 1) = 1 &\rightarrow, X_{i,m}(\tau) = 1, \text{ for } m < k \end{aligned} \quad (4.58)$$

In this situation, the cost (energy) related to  $c_i(\tau)$  decreases, however, we should make sure that the decrease of  $X_{i,k}(\tau)$  will force some  $\{X_{i,m}(\tau), \exists m < k\}$  to be increased rather than that  $X_{i,m}(\tau), \forall m < k$  merely remains the same when  $X_{i,k}(\tau)$  becomes zero.

In the second case, it involves that some  $c_j(\tau)$  releases the resource for other connections, and at the same time, some  $c_i(\tau)$  increases. The energy must be decreased even in this second case because that the release of resource of connection  $j$  for connection  $i$  must be owing to higher  $(1 - e^{-\sigma U_i(t)})$  than  $(1 - e^{-\sigma U_j(t)})$ . This transition of value of  $c_i(\tau)$  implies lower energy (higher system utility) achieved. However, two tasks must be considered. The transition between  $j$  and  $i$  makes lower energy (the correct direction), and the decrement of  $c_j(\tau)$  encourages the increment of  $c_i(\tau)$ . These auxiliary terms may facilitate the convergence process and the desired output states, however, it can be implicitly incorporated in other terms.

Note that in item 3, objective function usually relates to the external inputs, and those constraints consist of the interconnections between state variables. In item 4, for a formal Lyapunov function, the coefficients of quadratic terms (of state variable  $X_{i,k}^2$ ) are negative, those cross-product terms are usually negative ( $X_{i,k} \cdot X_{j,m}$ ), those products of state variables and static inputs are usually positive, and the products of state variables

and bias currents are usually positive.

## 4.7 Simulation Results and Discussion

In the simulation scenario, five types of services are assumed in three service classes. Type-1 service is assumed to be real-time class of traffic with peak rate 15kbps, activity factor 0.57,  $P_D^* = 0.05$ ,  $D^* = 40\text{ms}$ , and  $BER^* = 10^{-3}$ . Type-2 (type-3) service is assumed to be non-real-time interactive class of traffic with Pareto process [75] of which the mean rate is 8kbps (12kbps) ,  $R_{m,i}^*=7.2\text{kbps}$  ( $R_{m,i}^*=11\text{kbps}$ ), and  $BER^* = 10^{-5}$  ( $BER^* = 10^{-5}$ ). Moreover, type-4 (type-5) service is assumed to be non-real-time best effort class of traffic in batch Poisson distribution with mean rate 6kbps (15kbps) and mean batch size 1.2k bits (1.2k bits). Only  $BER^* = 10^{-5}$  QoS requirement is guaranteed. The proportion in the number of connections from type-1 to type-5 is kept at 1:1:1:1:1. Also, four modulation schemes, BPSK, QPSK, 16QAM, and 64QAM, are available for transmission as the BER requirement can be fulfilled and the remaining queue is enough.

We compare the proposed CNNU-based scheduler with the *exponential rule* (EXP) scheduling scheme. The performance measures are such as the average system throughput, the average packet dropping ratio of RT connections,  $\bar{P}_D$ , the average transmission rate of NRT interactive connections,  $\bar{R}_m$ , the ratio of RT connections in which their packet dropping ratio requirement is not guaranteed,  $\phi_{P_D}$ , the ratio of NRT interactive connections in which their minimum transmission rate requirement is not guaranteed,  $\phi_{R_m}$ , and the fairness variance index of NRT connections,  $F_v$ . The  $F_v$  is defined for measuring the variance of fairness to share the radio resource among all NRT connections. It is given by

$$F_v = \frac{1}{N_{NRT}} \sum_i^{N_{NRT}} \left| \frac{E[r_i(t)]}{\sum_j^{N_{NRT}} E[r_j(t)]} - \frac{w_i}{\sum_j^{N_{NRT}} w_j} \right|^2,$$

where  $N_{NRT}$  is the number of NRT connections. The fairness variance index shows the variance of the normalized radio resource allocated and the normalized proportion of

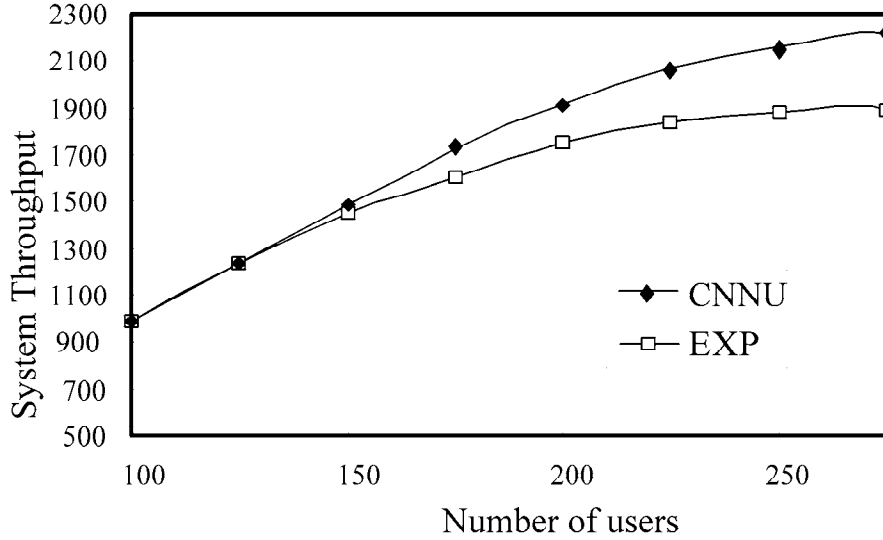


Figure 4.3: The average system throughput

resource desired to share.

Fig. 4.3 shows the average system throughput. It can be found that the CNNU-based scheduler has higher system throughput than the EXP scheduling scheme by more than 9% as the number of connections is greater than 200, and by higher than 15% as the number of connections increases up to 250. At around this point of 250 connections, the QoS requirements of RT connections begin to violate if the CNNU-based scheduler is adopted, while the exponential rule algorithm has begun to violate drastically the QoS requirement of minimum transmission rate,  $R_{m,i}^*$ ,  $i \in NRT$ , at the point of 175 connections. In summary, CNNU can always has a higher system capacity and a larger the QoS guaranteed region. This is because the CNN processor can quickly determine the optimal radio resource vector, and the CNNU-based scheduler can efficiently adapt to the link variation and effectively keep the QoS requirements guaranteed for different traffic classes. The priority bias makes the RT connections share the radio resource with relatively higher priority over NRT connections and, at the same time, ensures the average packet dropping ratio requirement of each connection according to its location dependent



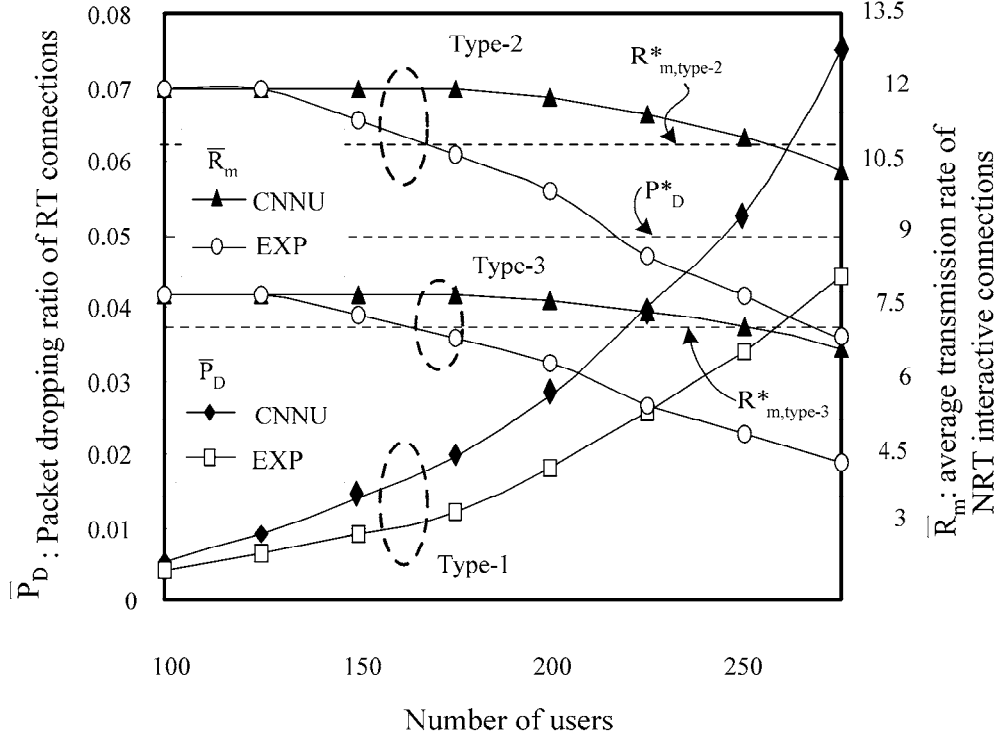


Figure 4.4: QoS performance measures of  $\bar{P}_D$  and  $\bar{R}_m$

information. Beyond the point of 250 connections, the throughput of EXP scheme is almost saturated, while the throughput of the CNNU-based scheduler continue to grow up with slightly lower slope. It is because the throughput of CNNU-based scheduler can achieve better utilization of multiuser diversity gain than that of the EXP scheduling scheme.

Fig. 4.4 depicts the comparison of performance measures, the average packet dropping ratio of type-1 RT connections  $\bar{P}_D$  and the average transmission rate of type-2 and type-3 NRT interaction connections  $\bar{R}_m$ . It can be found that the  $\bar{P}_D$  of the CNNU-based scheduler is always slightly greater than that of the EXP scheme; on the other hand, the  $\bar{R}_m$  of type-2 and type-3 connections of the CNNU-based scheduler are all greater than that of EXP scheme after the number of connections is greater than 125. It also shows that both of these QoS performance measures are guaranteed for the CNNU-based scheduler

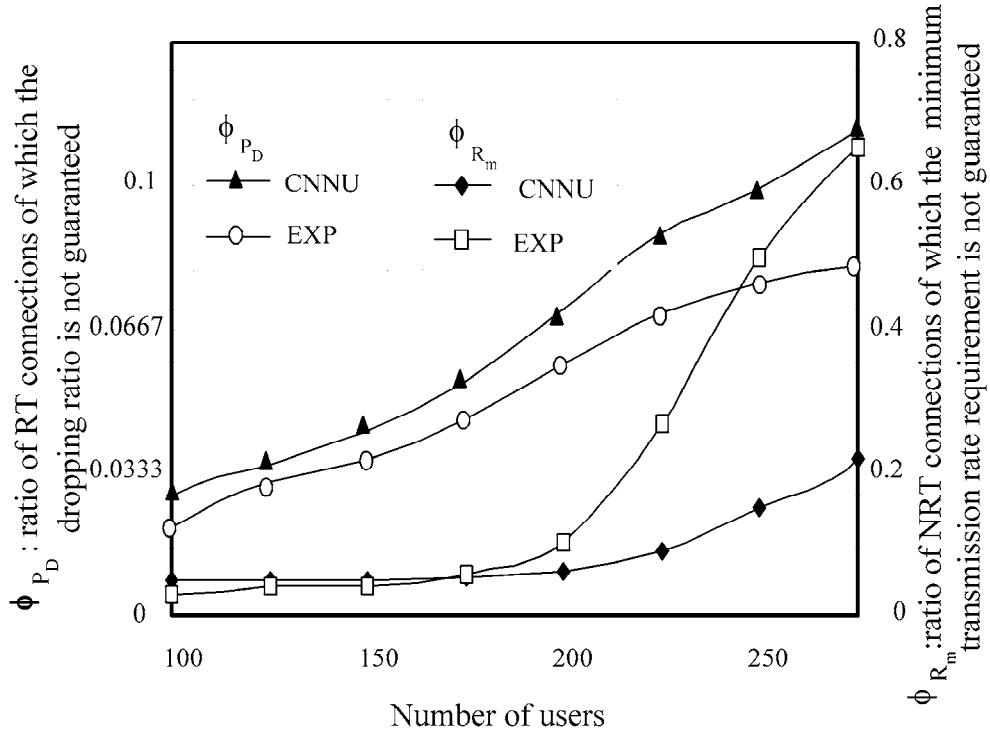


Figure 4.5: The ratio  $\phi_{P_D}$  for RT connections and the ratio  $\phi_{R_m}$  for NRT interactive connections

as the number of connections are less than 250, while these are guaranteed for the EXP scheme as the number of connections are less than 175. This shows that the CNNU-based scheduler has larger QoS guarantee region with differentiated QoS requirements and also higher average system throughput within this QoS guarantee region shown in Fig. 4.3. The maximum achievable throughput in QoS guarantee region of CNNU-based scheduler is higher than that of EXP scheme by 25%. The reason is that the CNNU-based scheduler can take advantages from the link adaptation and compensate the connections when their QoS requirements are about violated.

Fig. 4.5 shows the ratio  $\phi_{P_D}$  of RT connections and  $\phi_{R_m}$  of NRT interactive connections. The first glance of Fig. 4.5 indicates that the ratios of  $\phi_{P_D}$  and  $\phi_{R_m}$  are greater than zero at any traffic load conditions due to the existence of connections with very bad

link quality. It can be also found that within the light traffic load region ( less than 150 connections), both  $\phi_{P_D}$  and  $\phi_{R_m}$  of CNNU-based scheduler are slightly higher than that of the EXP scheme, but the two averaged QoS performance measures are still guaranteed, which have been described in Fig. 4.4. As the number of connections greater than 175,  $\phi_{R_m}$  of the EXP scheme becomes higher than that of CNNU-based scheduler, and increases drastically compared with CNNU-based scheduler after the number of connections greater than 200. The ratio  $\phi_{P_D}$  of the CNNU-based scheduler is always larger than that of the EXP scheme in this region. These results comes from that the EXP scheme mainly allocate the radio resource to compensate the connections with high likelihood to violate their QoS requirements, and the EXP scheme has lower ratios of  $\phi_{P_D}$  and  $\phi_{R_m}$  when the system is not overloaded. However, the capacity with guaranteed QoS requirements will be reduced. On the other hand, the CNNU-based scheduler jointly takes radio the resource efficiency and the deviation of QoS requirements measures into account, and its capacity is therefore enlarged, however, the resulting unguaranteed ratios,  $\phi_{P_D}$  and  $\phi_{R_m}$ , is getting larger. Moreover, the ratios  $\phi_{P_D}$  and  $\phi_{R_m}$  of CNNU-based scheduler are closed, while for exponential rule algorithm, its  $\phi_{R_m}$  is getting much higher than  $\phi_{P_D}$  in heavy load conditions. This shows that CNNU-based scheduler can effectively support the differentiated QoS requirements and fully utilize the link variation simultaneously, while exponential rule algorithm cannot adequately allocate the radio resource according to the extent of violation of requirements and also the individual link quality. Note that the results imply that the CNNU-based scheduler will not guarantee all the QoS requirements all the time, and a properly designed call admission control is required to reject the connections with very bad link quality in terms of the current traffic load conditions.

Fig. 4.6 shows the fairness variance index of NRT connections. It can be found that the fairness variance index of the CNNU-based scheduler retains within 1 in almost all

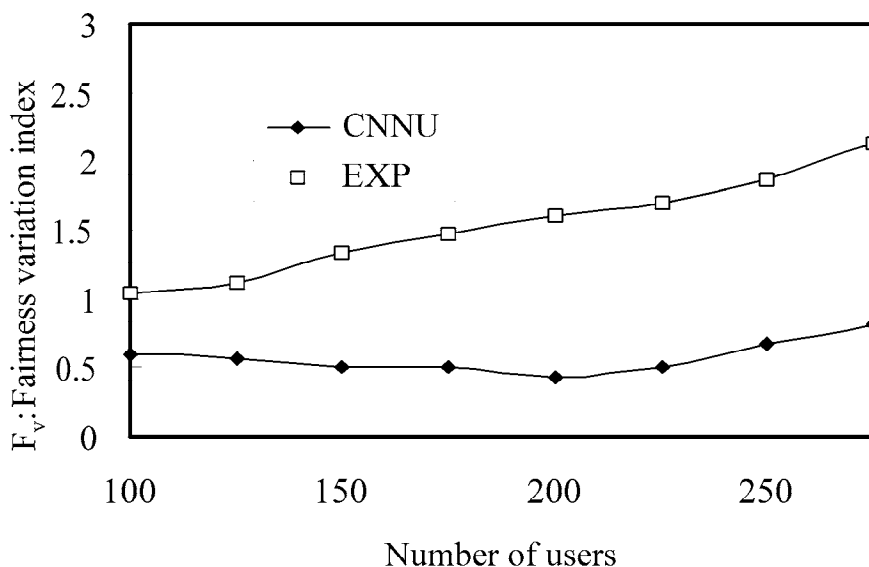


Figure 4.6: The fairness variation index for NRT connections

simulation cases, and grows up slowly as the number of connections increases; the fairness variance index of the EXP scheme, on the other hand, increases more quickly. This is because the fairness compensation function of the CNNU-based scheduler considers the location dependent information and aims to share the radio resource fairly as long as the minimum rate is guaranteed, while the design of the EXP scheme ignores the location dependent information and allocates rate fairly to all connections. The fairness compensation function, considering location dependent information, also facilitates the higher capacity for the CNNU-based scheduler shown in Fig. 4.3.

## 4.8 Concluding Remarks

This chapter presents a cellular neural network and utility-based (CNNU-based) scheduler, which jointly considers its radio resource efficiency, diverse QoS requirements, and fairness, to schedule the radio resource for connections in multimedia CDMA cellular systems. The utility function is defined to be the radio resource function properly weighted by the QoS requirement deviation function and the fairness compensation function. Also,

the cellular neural network (CNN) is adopted to solve the constrained optimization problem defined for the radio resource scheduling in a real time fashion. Simulation results show that the CNNU-based scheduler can efficiently allocate the radio resource to achieve higher throughput than the EXP scheduling scheme. It can also effectively support differentiated QoS requirements for connections with variant traffic characteristics. Moreover, the CNNU-based scheduler can enlarge the QoS guaranteed region under the complicated QoS requirements environments. The CNNU-based scheduler is effective for multimedia CDMA cellular systems with diverse of QoS requirements when both dedicated and shared channels are adopted. However, call admission control to accept/reject the good/bad user to ensure the operation of the CNNU-based scheduler should be further studied.

