Energy-Efficient MAC-PHY Resource Management with Guaranteed QoS in Wireless OFDM Networks

Ying Jun Zhang[†], *Member, IEEE* and Khaled Ben Letaief[‡], *Fellow, IEEE*

[†]The Department of Information Engineering
The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
Email: yjzhang@ie.cuhk.edu.hk

[‡]Center for Wireless Information Technology
The Dept. Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
Email: eekhaled@ee.ust.hk

***Abstract*–Most of the current resource management algorithms are confined to a single layer of the network protocol stack, which leads to an inferior system performance. In this paper, we propose a joint MAC-PHY layer resource allocation algorithm. The proposed algorithm jointly optimizes the bandwidth and power allocation through an integrated design of packet scheduling, subcarrier allocation, and power control. Analytical and numerical results will show that the proposed algorithm is able to provide the same QoS and fairness guarantee as fair queueing systems do in a wired channel. Meanwhile, the power efficiency and system performance are significantly improved compared to the traditional systems where resources are allocated based on a strict layering architecture.**

***Key Words*–Cross layer design, OFDM, Adaptive resource allocation, Packet scheduling.**

I. INTRODUCTION

Adaptive and intelligent resource management has become one of the key techniques for future wireless communications to ensure QoS (Quality of Services) guarantees given the limited resource availability [1-3]. In this paper, cross MAC-PHY layer resource management is studied for OFDM packet networks to provide QoS differentiation and guarantees, user fairness, and efficient power and spectrum utilization.

OFDM (Orthogonal Frequency Division Multiplexing) is one of the leading candidates for the high-rate transmission for future wireless communications due to its excellent performance over multipath fading channels [1]. Its inherent multicarrier nature provides enormous opportunities for adaptive resource management, such as subcarrier allocation, adaptive modulation, and power distribution [1-3]. Packet scheduling at the MAC (medium access control) or network layer is another critical component in future wireless packet networks. It allocates the bandwidth resources at a packet level and provides QoS guarantees as well as fairness among the users [4-6].

Most of the existing resource allocation methodologies are developed based on a strict layering structure, in which each layer in the protocol stack is designed and operated independently. However, the isolated component design of the network results in an inefficient utilization of the resources. For example, consider the adaptive subcarrier-bit-and-power allocation algorithms proposed for multiuser OFDM systems in [2, 3]. Mainly confined to the PHY layer, these algorithms fail to optimally allocate resources in a packet-switched network, since the randomness in the queueing behavior and the traffic arrival are not considered. Similarly, traditional packet scheduling algorithms in the MAC layer are designed without considering the characteristics of the wireless physical channels, such as user mobility, multipath fading, high error rate, and time-varying channel capacity [4-6]. Therefore, these algorithms fail to provide QoS and fairness guarantees in wireless networks.

In this paper, we are motivated to propose a cross MAC-PHY layer resource management algorithm for the downlink transmission of wireless packet networks with OFDM signaling. In contrast to [2, 3, 4-6], the bandwidth and power allocation at the MAC and PHY layers including packet scheduling, subcarrier allocation, and power distribution are jointly designed within an integrated framework to take advantage of the inter-dependencies between the two layers. The ultimate goal is to minimize the overall power consumption, satisfy the QoS requirements, and ensure the fairness between users.

We formulate the objective of the cross MAC-PHY layer resource allocation into a constrained optimization problem, which can be solved by the use of linear integer programming (LIP). Analytical and numerical results will demonstrate that the proposed algorithm is able to provide QoS and fairness guarantees for data services over wireless media. Meanwhile, power and spectrum efficiency is greatly improved compared to traditional systems. Furthermore, we can also make an easy tradeoff between QoS guarantees and power efficiency by adopting different design parameters.

The remainder of this paper is organized as follows. The system model of the proposed algorithm is described in the next section. In Section III, we briefly introduce the packet-scheduling criterion that is adopted in this paper. In Section IV, the cross-layer resource allocation algorithm is presented. The performance in terms of delay, throughput, and fairness is analyzed in Section V. In Section VI, we investigate the system performance through numerical results. Finally, the paper is concluded in Section VII.

This work is supported in part by the Hong Kong Research Grant Council under grant # HKUST6250/04E

II. SYSTEM MODEL

This paper investigates a downlink OFDM system with N subcarriers and K users. The time axis is divided info frames. A frame is further divided into a control phase and a data transmission phase that contains S OFDM symbols. On arriving at the BS, the packets from different users are buffered in separate queues, which are assumed to have infinite lengths. The BS then estimates the downlink channel, allocates bandwidth and power resources according to CSI (Channel state information), and informs the mobiles of the resource allocation decisions in the control phase. In the data transmission phase, the selected packets are encoded by a rate r convolutional code and modulated into M-ary symbols. They are then transmitted on the assigned subcarriers with an appropriate transmit power. Since the BS schedules transmission at the beginning of each frame, packets that arrive during the current frame will not be scheduled for transmission until the next frame.

Let us define $c_u(n,s,k)$ to be the subcarrier allocation indicator. That is, $c_u(n,s,k)=1$ means that the n^{th} subcarrier of the s^{th} symbol of frame u is allocated to user k for packet transmission, and $c_u(n,s,k)=0$, otherwise. In this paper, we do not allow more than one user to share a subcarrier. It follows that

$$\sum_{k=1}^{K} c_u(n,s,k) \le 1 \; \forall u,n,s \,. \tag{1}$$

Assume that the packets are of fixed length and each contains d information bits. If $g_u(k)$ packets are scheduled for transmission from user k's queue during frame u, then a total of $g_u(k)d$ information bits will be sent to user k in this frame. As each subcarrier is able to transmit $\log_2 M$ coded bits, the number of subcarriers allocated to user k should satisfy the following equation:

$$r\log_2 M \sum_{n=1}^{N}\sum_{s=1}^{S} c_u(n,s,k) = g_u(k)d \;\; \forall u,k \,. \tag{2}$$

Meanwhile, the total number of packet transmission in a frame is limited by the capacity of a frame:

$$\sum_{k=1}^{K} g_u(k) \le \frac{NSr\log_2 M}{d} \; \forall u \,. \tag{3}$$

The above transmission yields a total transmission power of

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{s=1}^{S} p_u(n,s,k)c_u(n,s,k) \,. \tag{4}$$

during frame u, where $p_u(n,s,k)$ denotes the amount of power that is transmitted to user k on the n^{th} subcarrier of symbol s.

In this paper, we concentrate on data services that are delay insensitive but not tolerable of packet errors. If a packet is received in error, ARQ (Automatic repeat request) is applied to recover the packet errors. There is no constraint on the maximum number of retransmission. For such data traffics, system performance is usually measured by the packet delay, throughput, as well as user fairness. Therefore, we will mainly consider these three aspects for performance evaluation.

III PACKET SCHEDULING

In wired networks, packet-scheduling algorithms have been proposed to provide fair services to users [4-6]. Most packet scheduling algorithms endeavor to serve the backlogged users in proportion to the share of bandwidth reserved by the users, with the objective function being

$$\min\left|\frac{W_{n_1}(t_1,t_2)}{\phi_{n_1}} - \frac{W_{n_2}(t_1,t_2)}{\phi_{n_2}}\right| , \tag{5}$$

for any pair of users n_1 and n_2 that are continuously backlogged during an arbitrary time interval (t_1, t_2). ϕ_n represents the share of bandwidth reserved by user n and $W_n(t_1,t_2)$ denotes the amount of traffic served for user n during the time period (t_1, t_2).

In this paper, we seek to serve packets in a way that approximates a fair queueing scheme in a wired environment, which is referred to as the *reference system*. Specifically, the proposed scheme explicitly limits the *leading* or *lagging* of the flows compared to the reference system. This is achieved by putting the following constraint when allocating bandwidth resources:

$$\left|Q_u(k) - Q_u^r(k)\right| \le \eta_k \;\; \forall k \tag{6}$$

where $Q_u(k)$ is the queue length (in terms of the number of packets) of user k at the end of frame u, if the selected packets from user k's buffer are transmitted successfully, and $Q_u^r(k)$ is the queue length of user k at the end of frame u in the reference system. η_k is the maximum allowable leading or lagging (also in terms of the number of packets) for user k.

IV. OPTIMAL CROSS-LAYER RESOURCE ALLOCATION

The objective of the resource allocation is to minimize the overall power consumption while guaranteeing the provision of QoS and fairness. In other words, we endeavor to find $p_u(n,s,k)$, $c_u(n,s,k)$, and $g_u(k)$ that minimize the total transmission power given by (4) subject to the constraints described in Eqns. (1, 2, 3, 6). The objective is formulated into the following constrained optimization problem:

$$\underset{c_u(n,s,k),\,g_u(k)}{\arg\min} \sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{s=1}^{S} p_u(n,s,k)c_u(n,s,k) \tag{7}$$

subject to:

C1: $$\frac{\left|H_u(n,s,k)\right|^2 p_u(n,s,k)}{N_o B} = \gamma_k \;\; \forall u,n,s,k$$

C2: $$\sum_{k=1}^{K} c_u(n,s,k) - 1 \le 0 \ \forall n,s$$

C3: $$\sum_{n=1}^{N}\sum_{s=1}^{S} c_u(n,s,k) - g_u(k)\frac{d}{r\log_2 M} = 0 \ \forall k$$

C4: $$g_u(k) \le \eta_k + w_u(k) \ \forall k$$

C5: $$g_u(k) \ge -\eta_k + w_u(k) \ \forall k$$

C6: $$\sum_{k=1}^{K} g_u(k) = \min\left(\sum_{k=1}^{K} Q_{u-1}(k), \frac{NSr\log_2 M}{d}\right)$$

C7: $$c_u(n,s,k) \in \{0,1\} \ \forall n,s,k$$

C8: $$g_u(k) \in \{0,\cdots,Q_{u-1}(k)\} \ \forall k$$

Constraint C1 corresponds to the power control, which is to smooth the variance in the wireless channel capacity and maintain a sufficiently low PER (packet error rate). By doing so, the wireless channel appears to be static and almost error free to the MAC layer and the packet scheduling algorithm designed for wired channels can therefore be applied. In the equation, γ_k is the SNR required by user k for achieving QoS requirements, $H_u(n,s,k)$ denotes the channel coefficient of user k on the n^{th} subcarrier of the s^{th} OFDM symbol during the u^{th} frame, N_o is the power spectral density of the white noise, and B is the bandwidth of one subcarrier.

Constraints C2 and C3 correspond to Eqns. (1) and (2). C4 and C5 are derived from the QoS constraint described in (6), with $w_u(k)$ being

$$w_u(k) = Q_{u-1}(k) - Q_u^r(k) + a_u(k)\,. \tag{8}$$

where $a_u(k)$ is the number of packet arrivals during frame u. Likewise, C6 is to ensure a work-conserving service discipline. Finally, C7 and C8 correspond to the fact that $c_u(n,s,k)$ is a 0-1 variable while $g_u(k)$ is a nonnegative integer variable that is no larger than the user's queue length at the beginning of the frame, or equivalently, at the end of the previous frame.

The above optimization problem can be solved by LIP method [7] by substituting C1 into the objective function (8). The solution to the problem provides an optimal resource allocation including packet scheduling specified by $g_u(k)$, subcarrier allocation specified by $c_u(n,s,k)$, and power allocation, which is obtained by

$$p_u(n,s,k) = \frac{\gamma_k N_o B}{|H_u(n,s,k)|^2} c_u(n,s,k)\,. \tag{9}$$

It can be easily seen that the larger the value of η_k is, the larger the feasible region is, and consequently the smaller the objective function (i.e., the total transmit power) becomes. However, a larger η_k results in a less stringent QoS and fairness guarantees, which will be proved later. Therefore, η_k is a parameter that allows trade-off between the QoS provisioning and the power efficiency.

V. PERFORMANCE ANALYSIS

In this section, the performance of the proposed system is investigated. The results are achieved under the assumption that the PER, which is controlled by power allocation, is sufficiently low. The following lemmas should be obvious from the proposed algorithm described in Section IV.

Lemma 1: Let $t_{k,i}$ denote the departure time of the ith packet of session k in the proposed system, and $t_{k,i}^r$ denote the departure time of the ith packet of session k in the reference system. Then,

$$t_{k,i} \le t_{k,i+\eta_k}^r\,. \tag{10}$$

Lemma 2 (Throughput guarantee): Let $W_k(t_1,t_2)$ be the number of packets transmitted by session k during the time period (t_1, t_2) in the proposed system, and $W_k^r(t_1,t_2)$ be the counterpart in the reference system. Then for an arbitrary time interval (t_1, t_2) the following inequality holds

$$\left|W_k(t_1,t_2) - W_k^r(t_1,t_2)\right| \le 2\eta_k\,. \tag{11}$$

Based on Lemma 1 and 2, we can derive the following results.

Theorem 1: (Delay bound guarantee): If the delay experienced by a packet of session k in the reference system is bounded by $D_{\max,k}^r$, then the delay experienced by a packet of session k in the proposed system, denoted by $D_{k,i}$, is bounded by

$$D_{k,i} \le D_{\max,k}^r + t_{k,i+\eta_k}^r - t_{k,i}^r\,. \tag{12}$$

Proof: Let $\tau_{k,i}$ denote the arrival time of the ith packet of session k. Note that this value is the same for both the proposed and the reference system. Then, from Lemma 1, the delay of packet $i$of session k becomes

$$\begin{aligned} D_{k,i} &= t_{k,i} - \tau_{k,i} \\ &\le t_{k,i+\eta_k}^r - \tau_{k,i} + t_{k,i}^r - t_{k,i}^r \\ &= t_{k,i}^r - \tau_{k,i} + t_{k,i+\eta_k}^r - t_{k,i}^r \\ &\le D_{\max,k}^r + t_{k,i+\eta_k}^r - t_{k,i}^r \end{aligned}$$

Theorem 2: (Fairness guarantee): If in the reference system,

$$\left|\frac{W_{k_1}^r(t_1,t_2)}{\phi_{k_1}} - \frac{W_{k_2}^r(t_1,t_2)}{\phi_{k_2}}\right| \le \beta$$

then given an arbitrary interval (t_1, t_2), the following inequality holds for the proposed system

$$\left|\frac{W_{k_1}(t_1,t_2)}{\phi_{k_1}} - \frac{W_{k_2}(t_1,t_2)}{\phi_{k_2}}\right| \le \beta + \frac{2\eta_{k_1}}{\phi_{k_1}} + \frac{2\eta_{k_2}}{\phi_{k_2}} \tag{13}$$

for any pair of users n_1 and n_2 that are continuously backlogged during the interval.

Proof: From Lemma 2, we can let

$$W_k(t_1,t_2) = W_k^r(t_1,t_2) + \delta_k \ \forall k$$

where $-2\eta_k \le \delta_k \le 2\eta_k$. Then,

$$\left|\frac{W_{k_1}(t_1,t_2)}{\phi_{k_1}} - \frac{W_{k_2}(t_1,t_2)}{\phi_{k_2}}\right| = \left|\frac{W_{k_1}^r(t_1,t_2)}{\phi_{k_1}} - \frac{W_{k_2}^r(t_1,t_2)}{\phi_{k_2}} + \frac{\delta_{k_1}}{\phi_{k_1}} - \frac{\delta_{k_2}}{\phi_{k_2}}\right|$$

$$\le \left|\frac{W_{k_1}^r(t_1,t_2)}{\phi_{k_1}} - \frac{W_{k_2}^r(t_1,t_2)}{\phi_{k_2}}\right| + \left|\frac{\delta_{k_1}}{\phi_{k_1}}\right| + \left|\frac{\delta_{k_2}}{\phi_{k_2}}\right|$$

$$\le \beta + \frac{2\eta_{k_1}}{\phi_{k_1}} + \frac{2\eta_{k_2}}{\phi_{k_2}}$$

The above lemmas and theorems indicate that the proposed system, operated over a wireless media, is able to provide similar performance to a fair queueing system in a wired environment with a slight difference that is determined by η_k. When η_k is set to be 0, the performance in terms of throughput, delay, fairness, and etc. is the same as that in the reference system.

VI. NUMERICAL RESULTS

In this section, the performance of the proposed system is investigated by simulation. An OFDM system with 256 subcarriers and a symbol duration T_s=200μs is considered. Assume that each session generates data streams at a rate of 100*k*bps and each packet contains 128 bits. Assume that a rate ½ convolutional code with a constraint length 3 is applied, and QPSK modulation is used on every subcarrier.

Depending on the requirements on the QoS, complexity, and efficiency, we can select any of the fair queueing schemes that are designed for the wired environment as the reference system. In the simulations, Virtual Clock (VC) service discipline is selected due to its easy implementation. In order to demonstrate the advantages of the proposed system, we will compare the performance with traditional systems, which include a SMSP (Static MAC and static PHY) system, a DMSP (Dynamic MAC and static PHY) system, and a DMDP (Dynamic MAC and dynamic PHY) system. Unlike the proposed system, the traditional systems are based on a strict layering architecture. In the notations, static MAC corresponds to a fixed multiplexing system and dynamic MAC stands for statistical multiplexing. Similarly, static PHY implies that there is no adaptation in the physical layer while dynamic PHY means that the power and subcarrier are dynamically allocated.

Fig. 1 compares the average packet delay of the proposed algorithm with the traditional systems mentioned above. In this simulation, there are 12 active users. Each frame contains 1 OFDM symbol for data transmission and $\eta_k = 1$ for all users. From the figure, we can see that by introducing statistical multiplexing, the average packet delays of the DMSP, DMDP, and the proposed cross-layer algorithm are reduced to less than 1/5 of the SMSP algorithm when the system is stable. Moreover, the proposed algorithm significantly outperforms the algorithms based on static transmission or a strict layering architecture. It can be seen that the proposed algorithm requires 8dB and 5dB less E_b/N_o to achieve a stable performance, compared with the DMSP and DMDP systems, respectively.

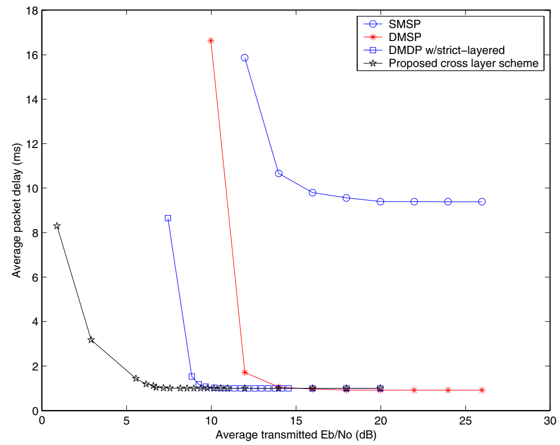

Fig. 1: Average packet delay of the proposed cross-layer design algorithm and the traditional systems when *K*=12 and *S*=1.

The throughput of the proposed algorithm, which is defined to be the total number of successfully transmitted packets in a time unit, is compared with that of the conventional systems in Fig. 2. The system throughput in an AWGN (additive white Gaussian noise) channel is also plotted for illustration. The simulation parameters are the same as those in Fig. 1. From the figure, it can be seen that by dynamically allocating the subcarriers and power (i.e., the DMDP and the proposed scheme), we are able to achieve the same diversity order as in an AWGN channel, which is significantly higher than that without PHY layer adaptation (i.e., DMSP in the figure). However, based on a strict layering structure, the DMDP algorithm suffers more than 2dB performance loss compared to that in the AWGN environment. In contrast, the proposed algorithm, benefiting from the joint design of the MAC and PHY layer resource allocation, is able to outperform the AWGN benchmark by 2dB. This is because the proposed algorithm enjoys large dimensions of freedom for channel allocation. As a result, it is very likely that the users might transmit on the channels with channel gains higher than unity. When the system throughput is 1.8, for instance, the proposed cross-layer resource allocation algorithm outperforms the DMSP scheme and the DMDP scheme by around 8dB and 5dB, respectively.

Fig. 3 and Table 1 investigate the tradeoff between the power efficiency and QoS guarantee by choosing different values of

η_k . In Fig. 3, throughput is plotted as a function of SNR required at the receiver The figure shows that the power efficiency can be greatly improved if a slightly less-stringent QoS is acceptable. More than 3dB power gain can be achieved when η_k increases from 0 to 1. A further 1dB gain is observed when η_k increases to 2.

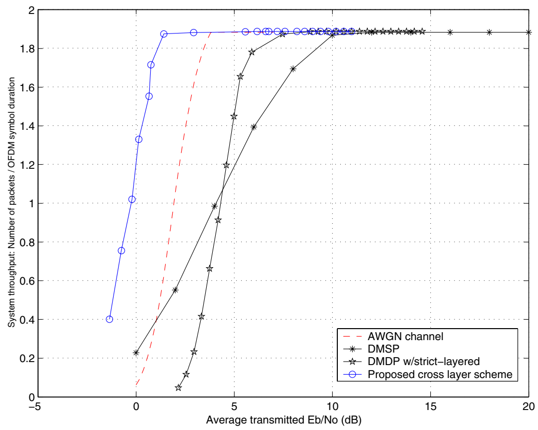

Fig. 2: System throughput of the proposed cross-layer design algorithm and the traditional systems when K=12 and S=1.

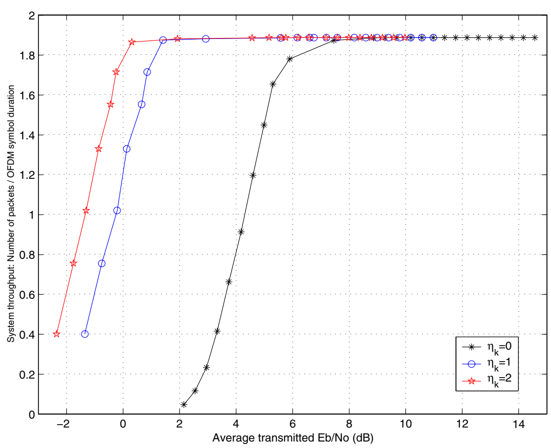

Fig. 3: Comparison of the throughput for different values of η_k, where K=12 and S=1

Table 1: Average packet delay for different values of η_k, where K=12 and S=1, average transmitted $E_b/N_o = 12dB$

	η_k=0	η_k=1	η_k=2
Average Packet delay (ms)	1.0151	1.0188	1.0214

Although the power efficiency is significantly improved by increasing the value of η_k , a sacrifice the QoS guarantee is expected in the meantime. The average packet delay for different values of η_k is studied in Table 1. It can be seen that the packet delay increases with the increase of η_k . This result is consistent with Theorem 1.

VII. CONCLUSIONS

In this paper, a cross MAC-PHY layer resource allocation algorithm was proposed for wireless packet networks with OFDM signaling. In the proposed algorithm, the resource allocation in the MAC and PHY layers were designed within an integrated framework to take advantages of the inter-dependencies between the two layers. Specifically, packet scheduling, subcarrier allocation, and power control were jointly optimized to maximize the overall power efficiency and provide QoS guarantees. It was proved that the proposed system is able to provide similar QoS and fairness guarantees as the fair queueing systems do in a wired environment It was also demonstrated that the proposed cross-layer resource allocation algorithm significantly improves the power efficiency and the system performance in comparison with the traditional systems with a static resource allocation or with an adaptive resource allocation based on a strict layered architecture.

REFERENCES:

[1] T. Keller and L. Hanzo, “Adaptive multicarrier modulation: a convenient framework for time frequency professing in wireless communications,” *IEEE Proc.*, vol. 88, pp. 611-640, May 2000.

[2] Y. J. Zhang and K. B. Letaief, “Multiuser subcarrier and bit allocation along with adaptive cell selection for OFDM transmission,” *IEEE Proc. ICC’02*, vol. 2, pp. 861-865, 2002.

[3] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch “Multiuser OFDM with adaptive subcarrier, bit, and power allocation,” *IEEE J. Select. Areas Commun.*, Vol. 17, No. 10, pp. 1747-1758, Oct. 1999.

[4] H. Zhang, “Service disciplines for guaranteed performance services in packet-switching networks,” *IEEE Proc.*, vol. 83, no. 10, pp. 1374-1396, Oct. 1995.

[5] A. K. Parekh and R. G. Gallager, “A generalized processor sharing approach to flow control in integrated services networks: the single-node case,” *IEEE/ACM Trans. Network.*, vol. 1, no. 3, pp. 344-357, June 1993.

[6] L. Zhang, “VirtualClock: a new traffic control algorithm for packet switching networks,” *ACM Trans. Computer Systems*, vol. 9, pp. 101-124, May 1991.

[7] G. V. Reklaitis, A. Rakvindran, K. M. Ragsdell, *Engineering optimization: Methods and applications*, Wiley, 1983