Cross-Entropy-Based Sign-Selection Algorithms for Peak-to-Average Power Ratio Reduction of OFDM Systems

Luqing Wang and Chintha Tellambura

***Abstract*—Sign-selection uses a set of subcarrier signs to reduce the peak-to-average power ratio (PAR) of orthogonal-frequency-division multiplexing (OFDM). However, the computational complexity (worst-case) is exponential in N, the number of subcarriers. Suboptimal sign-selection algorithms, achieving different tradeoffs between the PAR reduction and complexity, have thus been developed. For example, the derandomization method achieves high PAR reduction of $\mathcal{O}(\log N)$ with relatively high complexity of $\mathcal{O}(N^2)$. On the other hand, selective mapping (SLM) and partial transmit sequences (PTS) sacrifice the achievable PAR reduction for lower complexity. In this paper, we develop two new cross-entropy (CE)-based sign-selection algorithms. Our algorithms simultaneously updates the probabilities of the signs of *all* subcarriers. As shown in Section II-D, the first algorithm obtains a PAR lower than the above methods with a complexity level of $\mathcal{O}(N^2)$. However, if the number of iterations is fixed, this algorithm obtains the same PAR reduction as derandomization, but with $\mathcal{O}(N \log N)$ complexity. Practical PAR reduction algorithms require that the extra cost of PAR reduction must be small. Therefore, we propose the second algorithm, which adaptively adjusts the probability of "elite" samples, and stops whenever a PAR threshold is reached. Our second algorithm achieves up to 95% complexity savings over the first (with only a 0.4-dB PAR reduction loss). The simulations confirm the complexity advantages of the proposed algorithms compared to SLM and derandomization.**

***Index Terms*—Cross-entropy (CE) method, importance sampling, orthogonal-frequency-division multiplexing (OFDM), peak-to-average power ratio (PAR) reduction.**

I. Introduction

Since orthogonal-frequency-division multiplexing (OFDM) suffers from the high peak-to-average power ratio (PAR) [1], many PAR-reduction techniques have been developed ([2] and references therein). Of these, probabilistic PAR reduction techniques are especially suitable when the number of subcarriers is large (≥ 64) [3], [4]. The sign-selection method [5], [6] is a probabilistic technique and the basic idea is to use a set of subcarrier signs to reduce the PAR. This gives a maximum of 2^N (N is the number of subcarriers) potential OFDM symbols for a given input. Clearly, the complexity of finding the optimal set of signs is exponential in N. Existing sign-selection algorithms can be divided into two groups: 1) algorithms with low-complexity, e.g., selected mapping (SLM) and partial transmit sequences (PTS) [3], [4], which, however, only obtain moderate PAR reductions; and 2) algorithms with large PAR reduction, e.g., the derandomization method [6], which limit the PAR growth to $\mathcal{O}(\log N)$, but have relatively high complexity of $\mathcal{O}(N^2)$.

The cross-entropy (CE) method is an iterative procedure for combinatorial optimization [7]. Each iteration involves generating a random sample according to a probability distribution and then updating the parameters of the probability distribution in order to produce better samples in the next iteration. In [8], we developed a CE-based sign-selection algorithm for PAR reduction. In this paper, we further develop two CE based PAR reduction algorithms. Large PAR reduction is obtained with lower complexity by simultaneously modifying the probabilities of the signs of *all* subcarriers via the CE method. In contrast, the derandomization method modifies the signs one by one. With a fixed number of samples, our algorithms obtain the same PAR reduction with an $\mathcal{O}(N \log N)$ complexity. Practical PAR reduction algorithms require that the extra cost of PAR reduction must be small. By adaptively adjusting the probability of "elite" samples, and stopping whenever a PAR threshold is reached, our second algorithm achieves up to 95% complexity savings over the first (with only a 0.4-dB PAR reduction loss). The simulations show that, for the same level of PAR reduction, our algorithms have lower complexity than the SLM and derandomization methods. To the best of our knowledge, our work is this first application of the CE method to PAR reduction.

This correspondence is organized as follows. Section II describes the general sign-selection method and develops two CE-based PAR reduction algorithms. The simulation results are provided in Section III comparing the new algorithms, the SLM and derandomization methods. The correspondence concludes in Section IV.

II. CE Method for PAR Reduction

A. PAR Definitions

OFDM maps a block of inputs bits to a set of N possibly complex symbols X_k chosen from an M-ary signal constellation $\mathcal{Q}_M$, e.g., pulse amplitude modulation (PAM), phase shift keying (PSK), or quadrature amplitude modulation (QAM). The set of N symbols is converted to time-domain samples via an Inverse discrete Fourier transform (IDFT). For an input OFDM block $\mathbf{X} = [X_0, \ldots, X_{N-1}]$, the discrete-time baseband equivalent signal $x(n)$ can thus be expressed as

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi nk/JN}, \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where J represents the oversampling factor, which must be $J \geq 4$ for sufficient accuracy [9], [10]. In this paper, we choose $J = 4$ for our simulations. The PAR is the ratio

$$\mathrm{PAR}(\mathbf{X}) \triangleq \frac{\max |x(n)|^2}{E\left[|x(n)|^2\right]}. \tag{2}$$

B. Sign-Selection Method

A sign sequence $\mathbf{c} = [c_0, \ldots, c_{N-1}]$ is pointwise multiplied with the input OFDM block $\mathbf{X}$ to give $\hat{\mathbf{X}} = [X_0 c_0, \ldots, X_{N-1} c_{N-1}]$. The time-domain OFDM signal of $\mathbf{X}$ is

$$\hat{x}(n; \mathbf{c}) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k X_k e^{j2\pi nk/JN}. \tag{3}$$

Note that the average power of $\hat{x}(n; \mathbf{c})$ is the same as that of $x(n)$ for all $\mathbf{c}$—i.e., the signs do not alter the average power. Since the signs are to be chosen to minimize the PAR, the resulting combinatorial optimization problem is

$$\begin{aligned} &\min_{\mathbf{c}} L(\mathbf{c}) \\ &\text{subject to: } \mathbf{c} \in \{1, -1\}^N \end{aligned} \tag{4}$$

where $L(\mathbf{c}) = \max_n |\hat{x}(n; \mathbf{c})|^2$ is the real-valued cost function, and $\{1, -1\}^N$ is the set of N-dimensional binary vectors. Note that depending on a specific implementation, some of the signs may be fixed, say, only $N_s \leq N$ variables signs are used.

Manuscript received June 15, 2007; revised March 13, 2008. First published June 20, 2008; current version published September 17, 2008. The associate editor coordinating the review of this paper and approving it for publication was Dr. Kostas Berberidis. This work has been supported in part by NSERC.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: wlq@ece.ualberta.ca; chintha@ece.ualberta.ca).

Digital Object Identifier 10.1109/TSP.2008.927484

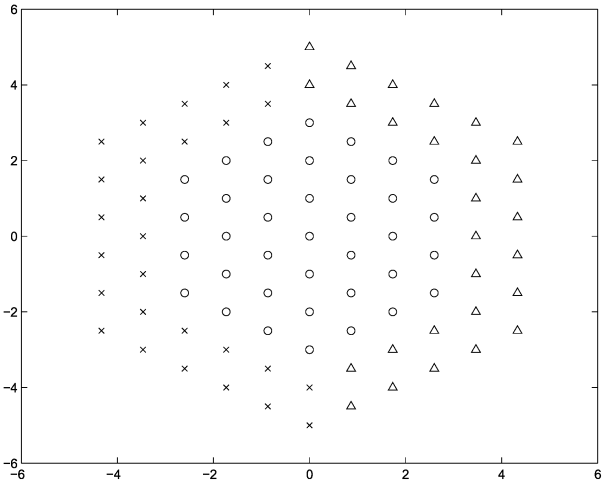

Fig. 1. The 91-point hexagonal constellation.

The optimal solution of (4), say, $\mathbf{c}^*$ may be transmitted to the receiver as side information for correctly decoding the transmit symbols. On the other hand, several techniques do, in fact, avoid side information [4], [5]. In this paper, we uses the hexagonal constellation to eliminate side information [11]. Fig. 1 illustrates an example of the 91-point hexagonal (91-Hex) constellation. The 64 points marked by $\circ$ and $\triangle$ are used to carry six information bits. Thus, the 91-Hex constellation has the same throughput and the same minimum Euclidian distance as the square 64QAM constellation. The signs of 27 outer points $\triangle$ can be modified (between $\triangle$ and $\times$) to minimize the PAR. Therefore, no side information is required at the receiver. Here, the mean of N_s is $\bar{N}_s = 27$.

To reduce the complexity of (4), suboptimal solutions are commonly used. For example, the SLM method randomly generates K sign sequences $\{\mathbf{c}_1, \ldots, \mathbf{c}_K\}$[3]. For a given OFDM block, a sign sequence $\mathbf{c}_i$ leading to the lowest PAR is selected from $\mathcal{C}$ as a suboptimal solution. This algorithm is fast when K is small. However, it does not reach highly accurate solutions because K is chosen to reduce complexity.

The derandomization method iteratively optimizes each sign by minimizing the probability that the PAR is larger than a threshold [6]. When all signs of the N subcarriers are allow to change, it limits the PAR to $\mathcal{O}(\log N)$ with $\mathcal{O}(N^2)$ complexity. Note that, to obtain the same level of PAR reduction, SLM requires prohibitively high complexity.

One may find a better approximation $\mathbf{c}^*$ by using the CE technique. The basic idea is to estimate a discrete probability distribution that will generate a solution close to $\mathbf{c}^*$.

C. The CE Sign-Selection (CESS) Algorithms for PAR Reduction

Without loss of generality, we consider the case that $N_s = N$. Let $\mathbf{c} = 1 - 2\mathbf{d}$, i.e., we generate the sign sequence $\mathbf{c} \in \{+1, -1\}^N$ from a binary vector $\mathbf{d} \in \{0, 1\}^N$. Let $L(\mathbf{d}) = \max_n |\hat{x}(n; 1 - 2\mathbf{d})|^2$. Each iteration of the CE method for the optimal solution $\mathbf{c}^*$ of (4) involves two phases [7]:

1) randomly generating a set of samples $\mathbf{d}_1, \ldots, \mathbf{d}_n$ with respect to a probability distribution $f(\mathbf{d}; \mathbf{p})$, where $\mathbf{p}$ is a probability parameter vector;
2) update $\mathbf{p}$ using $L(\mathbf{d}_1), \ldots, L(\mathbf{d}_n)$ so that $f(\mathbf{d}; \mathbf{p})$ generates "better" samples in the next iteration.

Each element of $\mathbf{d}$ is modeled an independent Bernoulli random variable with the probability distribution $P(d_k = 1) = 1 - P(d_k = 0) = p_k$, $k = 0, \ldots, N - 1$. The probability distribution is

$$f(\mathbf{d}; \mathbf{p}) = \prod_{k=0}^{N-1} p_k^{d_k} (1 - p_k)^{1 - d_k}. \tag{5}$$

The CE method optimizes $\mathbf{p} = [p_0, p_1, \ldots, p_{N-1}]$, which will generate a nearly optimal solution $\mathbf{d}^*$ that leads to low PAR.

However, $\mathbf{d}^*$ occurs with a very small probability. Estimating this probability using the Monte Carlo method requires a large number of samples of $\mathbf{d}$. Instead, the CE method estimates a relatively large probability $\Pr[L(\mathbf{d}) \leq \gamma]$, where γ is a relatively large threshold. Fewer samples are required to estimate this probability. The probability vector $\mathbf{p}$ is updated so that most samples generated by $f(\mathbf{d}; \mathbf{p})$ satisfy $L(\mathbf{d}) \leq \gamma$. The likelihood of $\mathbf{d}^*$ among these samples is increased. By iteratively letting $\gamma \to 0$, $f(\mathbf{d}; \mathbf{p})$ converges to an optimum pdf $f(\mathbf{d}; \mathbf{p}^*)$ that generates $\mathbf{d}^*$ with a minimum number of samples, and γ converges to $L(\mathbf{d}^*)$. In each iteration, $\mathbf{p}^*$ can be analytically found by solving [7]

$$\frac{1}{n} \sum_{i=1}^{n} I_{\{L(\mathbf{d}_i) \leq \gamma\}} \nabla \ln f(\mathbf{d}_i; \mathbf{p}) = \mathbf{0} \tag{6}$$

where $I_{\{x \leq \gamma\}} = 1$ when $x \leq \gamma$ and is 0 otherwise, and $\mathbf{d}_i$ are generated using $f(\mathbf{d}; \mathbf{p})$.

The partial derivative of (5) is given by

$$\frac{\partial \ln f(\mathbf{d}; \mathbf{p})}{\partial p_k} = \frac{d_k}{p_k} - \frac{1 - d_k}{1 - p_k} \tag{7}$$

where d_k is the kth element of $\mathbf{d}$. Substituting (7) into (6), the optimal p_k can be found as [7]

$$p_k^* = \frac{\sum_{i=1}^{n} I_{\{L(\mathbf{d}_i) \leq \gamma\}} d_{ik}}{\sum_{i=1}^{n} I_{\{L(\mathbf{d}_i) \leq \gamma\}}} \tag{8}$$

for a given γ, where d_{ik} is the kth element of $\mathbf{d}_i$.

Our sign-selection algorithm can thus be summarized as follows [8].

Algorithm 1 (CESS):

1) Let $p_0 \equiv 1$, $p_k = 0.5$, $k = 1, \ldots, N - 1$. Let $\rho = 0.1$. Define $n_s = \lceil \rho n \rceil$, where $\lceil a \rceil$ represents the smallest integer that is greater than or equal to a.
2) Generate n samples of $\mathbf{d}_i$ with respect to $\mathbf{p}$, and calculate their PAR (i.e., $L(\mathbf{d}_i)$).
3) Sort these samples in ascending order according to $L(\mathbf{d}_i)$. Denote the obtained PAR value sequence as $[L_0, \ldots L_{n-1}]$.
4) Find γ as $\gamma = L_{n_s}$, and update $\mathbf{p}$ using (8).
5) If $0 < p_k < 1$ for some k, go to Step 2 and repeat this procedure using the updated p_k. Otherwise.
6) Output the optimal sign sequence as[1] $\mathbf{c}^* = 1 - 2\mathbf{p}$.

The optimization will converge to $p_k = 0$ or 1 for all k. Alternatively, we may also stop the optimization after K iterations, and select the sample with lowest PAR for transmission.

Now, we propose a modified CESS algorithm with threshold (MCESST). This algorithm stops the optimization when the PAR of an OFDM signal is reduced to within the amplifier's linear range. Based on the full adaptive CE method (FACE) [7], MCESST uses

[1]If all $\boldsymbol{p}_k$ are either 1 or 0, $\mathbf{p}$ can only generate one sample $\mathbf{d} = \mathbf{p}$.

"elite" samples to adaptively adjust ρ. That is, instead of using ρ, we define the number of "elite" samples, n_e. After generating n samples $\mathbf{d}_1, \ldots, \mathbf{d}_n$, and sorting the corresponding $L(\mathbf{d}_i)$ in ascending order (denoting as $L_1, \ldots, L_n$), a threshold γ can be found such that $\gamma = L_{n_e}$. Then

$$\rho = P(L(\mathbf{d}) \le \gamma) \approx \hat{\rho} \tag{9}$$

where $\hat{\rho} = n_e/n$. Based on the central limit theory, $\hat{\rho}$ is a Gaussian random variable with mean ρ. Its variance decreases to 0 when n goes to ∞. We can also rewrite (9) as $P(L(\mathbf{d}) \le \hat{\gamma}) \approx n_e/n$, where $\hat{\gamma}$ is an estimation of γ. It is a random variable with mean γ, and its variance decreases to 0 when n goes to ∞.

The modified CESS algorithm starts with a relatively small number $n_{\min}$. Let the input back off (IBO) threshold be A. Our algorithm can be summarized as follows.

Algorithm 2 (Modified CESS With Threshold):

1) Initialize A, n_e, $n_{\min}$, $n_{\max}$, n_{inc}. Let $n = n_{\min}$.
2) Let $p_0 \equiv 1$, and initialize $p_k = 0.5$, $k = 0, \ldots, N-1$.
3) Generate n samples $\mathbf{d}_i$ with respect to $\mathbf{p}$, and calculate their PAR (i.e., $L(\mathbf{d}_i)$). Whenever a $\mathbf{d}_k$ leads to $\text{PAR} \le A$, output $\mathbf{c}^* = 1 - 2\mathbf{d}_k$ and stop the algorithm. Otherwise.
4) Sort these samples in ascending order according to $L(\mathbf{d}_i)$. Denote the obtained PAR value sequence as $[L_0, \ldots L_{N-1}]$.
5) Find γ as $\gamma = L_{n_e}$, and update $\mathbf{p}$ using (8).
6) If $\mathbf{p} \in \{0,1\}^N$, output $\mathbf{c}^* = 1 - 2\mathbf{p}$ and stop the algorithm. Otherwise.
7) If $n < n_{\max}$, let $n = n + n_{\text{inc}}$, then, go to Step 3.

We further exploit the idea of Tabu search [12] to lower the complexity. At the last several iterations where most p_k are 1 or 0, the n samples contain many duplicated samples. Therefore, when a sample is generated, it may be with previously generated samples. If it has been previously generated, its PAR needs not be computed again. Note that a list containing the samples of $\mathbf{d}$ and their PAR is also required in (8). We utilize this list for comparisons. Simulation results show that about 15% inverse fast Fourier transforms (IFFTs) can be saved.

D. Convergency and Complexity Discussion

1) Convergency of Our Algorithms: The CE convergence is proved in [13]. A basic assumption of the convergence is that the neighborhood of the optimum solution is also nearly optimum. Our optimization problem satisfies this assumption. The neighborhood in our case is defined in terms of the Hamming distance. The PAR difference between the optimal sign sequence $\mathbf{c}^*$ and another sign sequence $\mathbf{c}_1$ within Hamming distance D from $\mathbf{c}^*$ can be bounded as

$$\left| \max_n |\hat{x}(n;\mathbf{c}^*)|^2 - \max_n |\hat{x}(n;\mathbf{c}_1)|^2 \right| \le \frac{8}{N} \left(\sum_{i=1}^{D} |X_{k_i}| \right)^2$$

where k_i denote the position that $\mathbf{c}_1$ differs from $\mathbf{c}^*$. Consequently, when N is large, a small D always leads to near-optimal solutions.

Let the set $\mathcal{C} = \{\mathbf{c}_i : \mathbf{c}_i = 1 - 2\mathbf{d}_i, L(\mathbf{d}_i) \le \gamma\}$ which contains n_s "good" sign sequence samples that lead to small PAR. Intuitively, our algorithms count the number of 1's that appear in $\mathcal{C}$ to estimate how the signs will be in neighborhood of the optimal solution and use this information to update $\mathbf{p}$. If most good samples have a positive sign at the kth position, most probably, the neighborhood of the optimal solution will have a positive sign at this position, and p_k is then updated to close to 0. In this way, most of the samples generated in the next iteration have a small Hamming distance to $\mathcal{C}$. A local optimum that has a PAR no larger than those of $\mathcal{C}$ is thus ensured. On the other hand, even a p_k is close to 0, there is still a chance to generate some samples having a negative sign at the kth position. The samples may also have small PAR, which gives the possibility to escape from the local optimum and converge towards the global optimum.

2) Complexity of Our Algorithms for Finding a Near-Optimal Solution: We first analyze the complexity of our algorithms. Note that, to find a near optimal solution, no threshold should be used in Algorithm 2. in terms of the number of samples required to find a near optimal solution.

In the $(i-1)$th iteration, our algorithms estimate the probability $\rho = \Pr[L(\mathbf{d}) \le \gamma]$ and optimize $\mathbf{p}$ such that most probably $L(\mathbf{d}) \le \gamma$ in the ith iteration. The accuracy of estimating ρ is determined by n_e. Therefore, we consider the complexity of finding a near optimal solution when n_e is fixed. We assume that n_e is large and ρ is small such that the samples generated in each iteration describe $\mathbf{p}$ with negligible error.

Also assume that the optimization leads to negligible error. Then, in the ith iteration, no sample with $L(\mathbf{d}) > \gamma$ would be generated. In other words, the size of the search space at the ith iteration, denoted as S_i and[2] $S_1 = 2^{N_s}$, is reduced to $S_i \approx \rho S_{i-1}$. If a near optimal solution is found in the Kth iteration, we have

$$S_{\min} = S_K = \rho^{K-1} S_1 = \rho^{K-1} 2^{N_s}$$

where $S_{\min} \ge 1$ is the number of the near optimal solutions. Then

$$K = \frac{\log_2 S_{\min} - N_s}{\log_2 \rho} + 1.$$

The total number of samples generated in the optimization is nK. Simulation results show that n must be proportional to N in order to obtain the same PAR distribution for different N. Therefore, the total samples required to find a near optimal solution is of the order $\mathcal{O}(N^2)$. Exhaustive search may search 2^{N_s} samples to find the optimal solution.[3] SLM and derandomization could not find a near optimal solution.

3) Complexity Comparison for Finding Suboptimal Solutions: We may fix the number of samples or set up a threshold (as in Algorithm 2) to find suboptimal solutions with low complexity. Each sample is calculated using FFT. Thus, the complexity of our algorithms for finding suboptimal solutions is $\mathcal{O}(N \log N)$ multiplications, which is of the same order as SLM. However, simulations show that SLM requires more samples (i.e., higher complexity) to obtain the same PAR as CESS.

Derandomization iteratively calculates the signs as

$$c_k = -\text{sign}\left(\sum_{n=1}^{2JN} \sinh\left(\mu \sum_{r=1}^{k-1} c_r a_{nr} \right) \sinh(\mu a_{nk}) \times \prod_{r=k+1}^{N} \cosh(\mu a_{nr}) \right) \tag{10}$$

where $\text{sign}(x)$ is the sign of x, μ is a constant, and $\{a_{nk}\}$ is the set of the real and imaginary parts of $X_k e^{j2\pi nk/JN}$, which are calculated and stored before calculating c_k. Computing $\{a_{nk}\}$ requires $\mathcal{O}(N^2)$ multiplications. The memory requirement for storing $\{a_{nk}\}$ is $2JN^2$ float-point real numbers. Computing $\cosh(\mu a_{nr})$ requires $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ hyperbolic functions. Thus, calculating $\prod_{r=k+1}^{N} \cosh(\mu a_{nr})$ requires $\mathcal{O}(N^3)$ multiplications and $\mathcal{O}(N^3)$ hyperbolic functions. The rest of (10) requires $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ hyperbolic functions. Therefore, the complexity of derandomization is $\mathcal{O}(N^3)$ multiplications and $\mathcal{O}(N^3)$ hyperbolic

[2]The mean of $\boldsymbol{S}_1$ is $\boldsymbol{S}_1 = (\mathbf{1} + \bar{\boldsymbol{N}}_s/\boldsymbol{N})^N$.

[3]It is easy to see that, to find a near optimal solution, the average number of samples that exhaustive search needs to search is exponential in $\bar{\boldsymbol{N}}_s$.

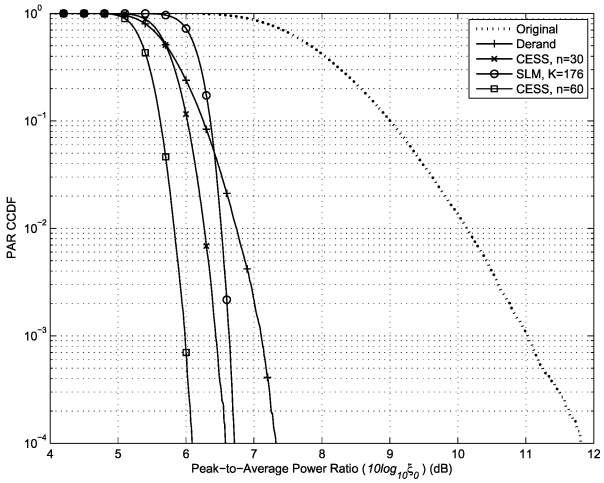

Fig. 2. PAR reduction comparison of CESS, SLM, and the derandomization method with $N = 128$ and the 91-Hex constellation.

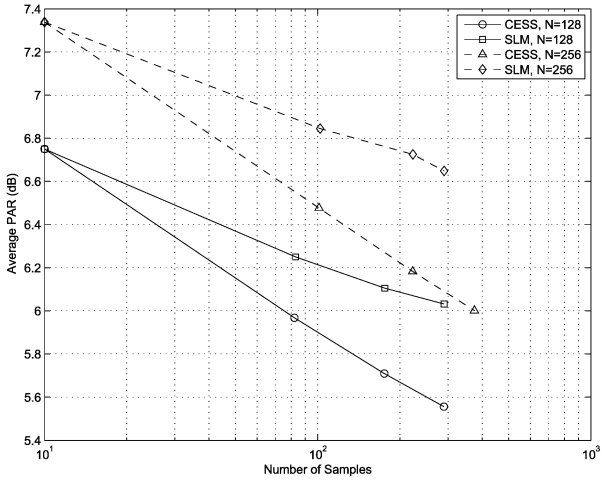

Fig. 3. PAR reduction comparison of CESS and SLM for the same complexity with the 91-Hex constellation.

functions. Note that the complexity of hyperbolic functions is much higher than that of multiplications.

By using more memory, we may calculate and store the results of $\prod_{r=k+1}^{N} \cosh(\mu a_{nr})$ for all k before the optimization. Then, the total memory requirement is $4JN^2$ float-point real numbers, but the complexity is reduced to $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ hyperbolic functions.

A greedy algorithm is proposed in [14] based on derandomization, with which the signs are iteratively calculated as

$$c_k = \arg \min_{c_k \in \{+1,-1\}} \left\| \sum_{i=1}^{k} \mathbf{a}_i c_i \right\|_p^p,$$

where $\mathbf{a}_i = [a_{ni}]$, $n = 1, \ldots, 2JN$ and $k = 2, \ldots, N$, are vectors each having $2JN$ items. It is proved that $p^{(\mathrm{opt.})} = \log 2JN$. Calculating $\mathbf{a}_i$ requires $\mathcal{O}(N^2)$ multiplications. Calculating c_k, $k = 2 \ldots N$, requires $4JN(N-1)$ power-p operations[4]. Therefore, the complexity of this algorithm is $\mathcal{O}(N^2)$. Its memory requirement is $2JN^2$ float-point real numbers (to store all $\mathbf{a}_i$). The Algorithm 3 in [14] also has a complexity of $\mathcal{O}(N^2)$ by noting that the $\cosh$ function used in this algorithm is defined on length-$(2JN)$ vectors.

In summary, the complexity of our algorithm is smaller than that of SLM and derandomization[5]. In Section III, we compare the PAR reduction of these algorithms by simulation.

III. Numerical Results

A. Performance Comparison of CESS, SLM and the Derandomization Method

We now compare CESS with SLM and the derandomization method [6], where $N = 128$ and the 91-Hex constellation is used. Fig. 2 plots the PAR complementary cumulative distribution function (CCDF), also called the clip probability, defined as $F(\xi) = \Pr[\mathrm{PAR}(\mathbf{X}) > \xi]$. The CESS algorithm is used with $\rho = 0.1$. At a 10^{-4} clip probability, CESS with $n = 30$ obtains a 5.3- dB PAR reduction, which is 0.6 dB larger than the derandomization method. The average number of samples generated by CESS in this case is 176. With the same complexity, the PAR reduction of SLM is 0.2 dB smaller than that of CESS at a 10^{-4} clip probability, and is 0.4 dB smaller than that of CESS at a 10^{-1} clip probability. By increasing the complexity, CESS with $n = 60$ obtains a 5.8-dB PAR reduction, which is about 1.2 dB larger than the derandomization method.

We next compare the PAR reduction of CESS and SLM for the same complexity. We consider two cases with $N = 128$ and $N = 256$, respectively. We choose $\rho = 0.1$ for CESS with $n = 10$, 20, 30 and 40, respectively. For each case of n, we count the average number of samples generated by CESS, and then use it as the number of candidates for SLM. Fig. 3 illustrates this comparison. Note that, when $\rho = 0.1$ and $n = 10$, CESS is the same as SLM since in this case only one sample leads to $L(\mathbf{d}_i) \leq \gamma$ in each iteration of CESS. We see that, with the same complexity, CESS leads to much smaller average PAR than SLM.

TABLE I
Simulation Parameters for Modified CESS With Threshold

A	$n_{\min}$	$n_{\max}$	n_{inc}	n_{e}	Bad Blocks	Worst PAR
6 dB	30	120	10	6	33	6.15 dB
6.5 dB	10	80	10	4	5	6.54 dB
7 dB	6	18	2	4	3	7.18 dB

B. Performance of the Modified CESS With Threshold

Now, we compare the modified CESS with threshold (MCESST) and SLM for a 128-subcarrier OFDM system. 10^5 OFDM blocks are simulated. Since the PAR of a typical single-carrier signal using square-root raised-cosine pulse shaping with a roll-off factor of 0.35 is about 6–7 dB [16], we compare MCESST and SLM with three thresholds $A = 6$, 6.5, and 7 dB (relative to the average power). When applying a threshold to SLM, we select a large candidate set containing 10^4 sign sequence samples, but stop the SLM optimization whenever a candidate that leads to a PAR below the threshold is found. The parameters for MCESST are selected as in Table I so that the PARs of virtually all input OFDM blocks are reduced to below A. We also include in this table, out of the 10^5 simulated OFDM blocks, the numbers of "bad" OFDM blocks (denoted as "bad blocks") whose PARs are larger than A after optimization, as well as the worst PAR.

With these settings, MCESST and SLM obtain the same PAR reduction for each threshold used. We thus compare their complexity in terms of the average number of IFFTs. Table II shows the complexity

[4]The optimal $\boldsymbol{p}$ is usually large. For example, $\boldsymbol{p}^{(\mathrm{opt.})} \approx \mathbf{6}$ or 7 when $\boldsymbol{N} = \mathbf{128}$ and $\boldsymbol{J} = 4$.

[5]Derandomization and the greedy algorithm also require a large number of additions, which may be a large burden in some implementations, e.g., using the multiplier-accumulator (MAC) [15], where the complexity of multiplications is comparable to additions.

TABLE II
AVERAGE NUMBER OF SAMPLES REQUIRED BY THE MODIFIED CESS WITH THRESHOLD AND SLM

A	MCESST	SLM	Ratio (MCESST/SLM)
6 dB	186.94	563.93	33.15%
6.5 dB	32.15	40.66	79.07%
7 dB	7.84	7.97	98.62%

comparison. MCESST has less complexity than SLM for all threshold settings.

IV. CONCLUSION

In this paper, we developed two cross-entropy sign-selection algorithms for OFDM PAR reduction. Our algorithms simultaneously updates the probabilities of the signs of *all* subcarriers. The first algorithm obtains near optimal solutions with $\mathcal{O}(N^2)$ complexity. If the number of iterations is fixed, this algorithm obtains the same PAR reduction as derandomization, but with $\mathcal{O}(N \log N)$ complexity. By adaptively adjusting the probability of "elite" samples, and by stopping whenever a PAR threshold is reached, our second algorithm achieves much lower complexity than the first. The simulations confirm the complexity advantages of the proposed algorithms compared to SLM and derandomization.

REFERENCES

[1] S. Litsyn, *Peak Power Control in Multicarriervarshne Communications*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[2] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun. Mag.*, vol. 12, no. 2, pp. 56–65, 2005.

[3] S. H. Müller and J. B. Huber, "A novel peak power reduction scheme for OFDM," in *Proc. 8th IEEE Int. Symp. Personal, Indoor Mobile Radio Commun. (PIMRC)*, Helsinki, Sep. 1–4, 1997, vol. 3, pp. 1090–1094.

[4] A. Jayalath and C. Tellambura, "SLM and PTS peak-power reduction of OFDM signals without side information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2006–2013, Sep. 2005.

[5] L. Wang and C. Tellambura, "A novel PAR reduction technique for OFDM systems using adaptive mapping," in *Proc. 15th Int. Conf. Wireless Commun.*, Calgary, AB, Canada, Jul. 7–9, 2003, pp. 270–277.

[6] M. Sharif and B. Hassibi, "Existence of codes with constant PMEPR and related design," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2836–2846, Oct. 2004.

[7] R. Rubinstein and D. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte Carlo Simulation, and Machine Learning*. New York: Springer-Verlag, 2004.

[8] L. Wang and C. Tellambura, "Peak-to-average power ratio reduction of OFDM systems using cross entropy method," in *Proc. 17th Int. Conf. Wireless Commun.*, Calgary, AB, Canada, Jul. 11–13, 2005, pp. 258–264.

[9] C. Tellambura, "Phase optimisation criterion for reducing peak-to-average power ratio in OFDM," *IEE Electron. Lett.*, vol. 34, no. 2, pp. 169–170, Jan. 1998.

[10] G. Wunder and H. Boche, "Peak value estimation of bandlimited signals from their samples, noise enhancement, and a local characterization in the neighborhood of an extremum," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 771–780, Mar. 2003.

[11] A. Pezeshk and B. H. Khalaj, "Extended hexagonal constellations as a means of multicarrier PAPR reduction," in *Proc. 1st EurAsian Conf. Information Communication Technology (EurAsia-ICT)*, Shiraz, Iran, Oct. 29–31, 2002, vol. 2510/2002, pp. 926–936.

[12] F. Glover, "Tabu search—Part I," *ORSA J. Comput.*, vol. 1, no. 3, pp. 190–206, 1989.

[13] L. Margolin, "On the convergence of the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 201–214, Feb. 2005.

[14] M. Sharif and B. Hassibi, "Towards reducing the gap between PMEPR of multicarrier and single carrier signals," in *Proc. 6th IEEE Workshop Signal Processing Advances in Wireless Commun.*, Jun. 5–8, 2005, pp. 380–384.

[15] F. Elguibaly, "A fast parallel multiplier-accumulator using the modified booth algorithm," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 902–908, Sep. 2000.

[16] S. Miller and R. O'Dea, "Peak power and bandwidth efficient linear modulation," *IEEE Trans. Commun.*, vol. 46, pp. 1639–1648, Dec. 1998.