Error and Quality Control Coding for DNA Microarrays

Olgica Milenkovic
Department of Electrical and Computer Engineering
University of Colorado
Boulder, CO, 80309, USA.
Email: olgica.milenkovic@colorado.edu

***Abstract*— DNA microarrays are one of the most important and widely-used systems for generating genomic data for disease diagnostics and monitoring. The technology supporting DNA microarray production represents a successful combination of ideas from molecular biology, biochemistry, VLSI design and mathematics. Nevertheless, there still exist many issues that have to be resolved in order to ensure very high reliability of microarray operation. Several carefully chosen methods for controlling the quality of microarrays have already made their way into the manufacturing process, but almost no schemes are known that can guarantee proper functionality of arrays in the presence of spot failure and drop-out events occurring after the manufacturing process. Classical error-control coding techniques can be used to ensure increased robustness under spot failure in terms of multiplexing different DNA sequences to multiple spots of the array. Such techniques necessarily lead to changes in the manufacturing process and its underlying quality control techniques. It is the goal of this work to propose an integrated framework for analyzing quality control and error-correction in DNA microarrays generated by photolitographic VLSIPS (Very Large Scale Immobilized Polymer Synthesis) methods. In this context, the issues of base scheduling, mask design and border-length minimization, construction of quality control arrays and good multiplexing strategies are addressed. The analysis is based on combining and extending results pertaining to the longest common subsequence problem, balanced and superimposed codes.**

I. Introduction

Recent advances in functional genomics have shown that genetic sequences, such as genes, RNA, proteins, and various regulatory elements represent parts of complicated large-scale networks. Properties of such networks can be inferred by measuring the concentration of various macromolecules within one or multiple cells in an organism. Determining these concentrations is a non-trivial process which is largely facilitated by DNA microarrays (Gene Chips) [1]. DNA microarrays are two-dimensional arrays of short single-stranded DNA sequences corresponding to gene segments. Messenger RNA molecules extracted from a cell are first converted into complementary DNA (cDNA), tagged with fluorescent markers and then distributed on a microarray substrate. Since Watson-Crick complementary DNA strands have a strong affinity to bond (hybridize) with each other, cDNA strands attach themselves to spots containing their complementary sequences. Which sequences succeeded in hybridizing with their complements can be detected by illuminating the chip by laser light of a given wavelength and measuring the intensity of fluorescence of the cDNA tags.

DNA microarrays can exhibit high failure rates during manufacturing or subsequent operation. Errors in the read-out signal can be attributed to missed steps in the production process, substrate defects, optical detection system malfunctioning, background illumination and other phenomena. In order to ensure high data integrity, Affymetrix [1] introduced in its production process a sophisticated technology that utilizes designated quality control spots for detecting erroneously synthesized DNA strands [8]. Although these procedures ensure that DNA arrays are properly created, they provide no built-in error-control mechanism that allows an array to recover from subsequent loss of data in one or more spots in the grid. This motivated the authors of [11] to propose a DNA strand multiplexing scheme with redundant spots that ensures reliable operation of the array in the presence of multiple spot failures. The results in [11] were validated in terms of fabricating chips of small size. Since the tests were performed on arrays with six, eight and ten spots only, the problems of quality control and cost of the production process were not addressed within the given context.

We propose to investigate a class of problems in combinatorics and error-control coding that arise in the process of fabricating DNA microarrays with multiplexed DNA strands. Many important results pertaining to the design of optimal masks, quality control spots and multiplexing schemes for error-correcting microarrays [6], [7], [8] have to be re-examined within this new setting. We address a class of such problems and describe future research directions in this area.

Sections II and III of the paper are mainly expository and structured in such a way that allows the interested reader to grasp both the technological and mathematical aspects underlying DNA microarray production and operation. New ideas regarding joint error-control coding and production quality control are described in Section IV of the manuscript. Detailed studies of algorithmic aspects of the underlying design process, code construction methods, and various optimization strategies will be provided in the full version of the paper.

[0]This work was supported by the NSF Grant CCF-0514921.

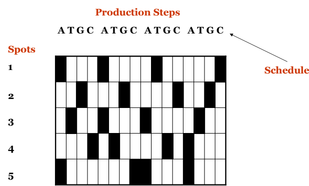

Fig. 1. Scheduling for photolitographic VLSIPS

II. DNA Microarrays Production and Quality Control

The process of DNA chip manufacturing pioneered by Affymetrix is based on photolitographic VLSIPS (Very Large Scale Immobilized Polymer Synthesis) methods which allow for simultaneous synthesis of multiple chips on one wafer. Each chip consists of a regular grid of spots at which a large number of short, customer selected single-stranded DNA sequences are placed. These DNA strands usually represent judiciously chosen subsequences of genes, or more precisely, DNA coding regions. Such strands have predefined structures which can in many cases be modeled by certain random processes.

The microarray fabrication process starts with creating a two-dimensional wafer on which certain molecules, termed *linkers*, are placed in a regular grid of spots. Linkers are equipped with photolabile protective groups that render them inactive in the absence of external light stimuli. At each step of the production process, a *mask* that specifies a subset of spots to be used during that production stage is designed. The mask is carefully imposed over the wafer and the system is illuminated. In the presence of light, the protective groups of the linkers dissolve, allowing the linkers to become operational. After this activation step, a solution containing a large number of the same DNA base (which is either adenin (A), guanin (G), thymin (T) or cytosin (C)) is dispensed on the wafer. This allows the nucleotides to bind to active linkers, creating the first base of the DNA sequences to be synthesized on the chip. Such DNA sequences are known as DNA probes. Each of the added nucleotides also contains a photolabile protective group which does not allow for multiple bases to attach to the linkers or to each other. The process is repeated an appropriate number of times by choosing at each step a different mask and a different nucleotide solution to be added to the wafer.

The order of the bases used in the production process is referred to as the *base schedule*. Usually, the schedule represents s periodic repetitions of the four DNA bases, which is denoted by $(ATGC)^s$. An example of a base schedule and masking process is shown in Figure 1. The masks for the five spots are represented by the columns of the array where a black rectangle corresponds to an exposed area while a white rectangle corresponds to a masked area. The DNA sequences generated at each of the five spots in the figure are $AATC$, $GGCG$, $TAAT$, $CTGA$, and $ACAA$, respectively.

The design process based on the periodic schedule shown in Figure 1 is *synchronous*, since each spot can be exposed to light only once during a period of length four. One can also use an *asynchronous* schedule, in which one spot can be exposed to light multiple times within a period [9].

Three important mathematical problems can be posed regarding the choice of the base schedule, the mask structures and the possibility of detecting production failures in a DNA microarray.

Base Scheduling: One problem of interest is to find the shortest possible asynchronous schedule that can produce a complete set of predefined probes. The shortest schedule ensures reductions in the cost of chip production and it also decreases the overall error-probability of probe formation. It is straightforward to see that such a schedule takes the form of a shortest common superstring of the probes, the computation of which is known to be NP-hard [9]. Furthermore, since the probes are usually selected in such a way that they poorly hybridize with each other, no large probe sequence overlaps are expected. Consequently, most currently available fabrication methods use periodic schedules.

Mask Design: Since spots are activated by illumination, it is possible that due to reflections, imprecisions in the optical system and the mask layout, spots neighboring the target spot become exposed to light as well. This unintended illumination can activate linkers of spots which were supposed to be masked, resulting in erroneous probe synthesis. One method to mitigate these effects is to minimize the total length of the borders of masks used in the VLSIPS process [6], [7]. The total border length of masks under a synchronous schedule corresponds to the sum of all pairwise Hamming distances d_H between probes at adjacent spots (adjacency in this context means that spots share at least one edge). If the set of probes consists of all possible DNA sequences of length N, the minimum border-length is achieved in terms of using two-dimensional Gray code mask designs [6]. When the probes are structurally constrained, some of the best known mask design methods rely on constructing *probe neighborhood graphs* [7]. Neighborhood graphs are complete graphs with vertices corresponding to the probes; the weight of an edge bridging two vertices equals the Hamming distance between the probes corresponding to the vertices. For such a graph, an approximate solution for the travelling salesman (TS) problem is sought, which specifies an ordering of the probes that is to be transferred to the array. This transfer is achieved in terms of *threading*, which is based on the use of certain space-filling curves that embed a one-dimensional string into a two-dimensional grid [7].

Quality Control: As described in the *Manufacturing Quality Control and Validation Studies of GeneChip Arrays* Manual in [1] the accuracy of the production steps is controlled in terms of a designated subset of *quality control spots* [8]. In this setting, m quality control spots contain $c < m$ different probes, and identical probes are synthesized using different steps of the manufacturing process. By performing hybridization tests on quality control spots one can identify if a faulty step occurred during probe synthesis. The work in [2], [3], [12] extends this testing framework one step further,

by proposing a coding scheme for identifying one erroneous production step. Information about erroneous production steps can be used to identify systematic problems in the production process.

From the perspective of coding theory the quality control problem can be formulated by referring to the following combinatorial objects.

Definition 2.1: [12] Let $\mathcal{A}$ be an $M \times N$ array of binary numbers such that the weight of each row equals r, the Hamming distance between any two columns is at least d, and the weight of each column is within the interval $[\omega_{min}, \omega_{max}]$, for $d \leq \omega_{min} < \omega_{max} \leq M - d$. Then $\mathcal{A}$ is called a *balanced binary code* with parameters $(M, N, r, d, \omega_{min}, \omega_{max})$.

A balanced binary code can be interpreted as an array of M quality control spots and N production steps. A "1" is placed at position (i, j) of the array if and only if the i-th quality control spot was active during the j-th step of the fabrication process. The constant row weight constraint is imposed in order to ensure that all quality control probes are of the same length, while the Hamming distance guarantees that two productions steps have distinguishable signatures even in the presence of d spot drop-outs. The restrictions on the weight of the columns guarantees that one can distinguish between a step failing and a step not being used in the production process, since one usually only measures relative rather than absolute fluorescence intensities.

The problem of designing balanced codes is well understood. To construct arrays corresponding to balanced codes it suffices to use combinatorial designs or subsets of codewords of codes with large minimum distance [2]. Far more interesting is the problem of detecting multiple production step failures. This question can be addressed in the framework of *superimposed designs* defined below [5], [10], [12].

Definition 2.2: [5] Let $\mathcal{A}$ be an $M \times N$ array of binary numbers with columns $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and rows of weight r. Such an array $\mathcal{A}$ is said to be a (M, N, r, s)-*superimposed design* with constraint r and strength s if all component-wise Boolean OR functions of not more than s columns of $\mathcal{A}$ are distinct.

Assume that the quality control matrix of a microarray corresponds to a superimposed design with constraint r. Then each probe has length r and multiple production step failures can be detected as follows. First, note that in order for a probe to be erroneously synthesized, *at least one* production step during which the probe was active has to fail. Consequently, the hybridization intensity profile of quality control probes contains information about the *component-wise Boolean OR function* of the columns of the control array. If each Boolean OR function of not more than s columns is unique, one can identify any set of not more than s failed production steps. It is important to observe that there is no guarantee that a quality control scheme based on a superimposed design can guarantee proper identifications of multiple production step failures in the presence of spot drop-outs in the control-array.

III. Microarray Error-Control Coding

Upon completion of the DNA microarray fabrication and testing process, the structure and properties of the array can change so that certain spots become nonfunctional. Spot failure is a common event which is very hard to detect since it manifests itself in terms of low fluorescence intensity during a hybridization experiment. But low fluorescence at a given spot can also be attributed to the gene corresponding to the synthesized probe being inactive in the tested cell. Consequently, there exists a strong need to design microarrays in such a way that even under spot failure, information about every single gene originally present in the grid is available. This can be achieved by multiplexing identical probes to different spots in the array. We provide next a sampling of ideas for strand multiplexing described in [11].

In what we will henceforth call a *multiplexed array*, every spot contains a mixture of a fixed number of different probes, and the number of spots exceeds the number of tested probes. For an array with the aforementioned properties, one seeks to design a binary *multiplexing matrix*, G, of dimension $M \times N$ where M denotes the number of spots, N denotes the number of distinct probes and $M > N$. The matrix G has to have full rank and the property that $G(i, j) = 1$ if and only if the ith spot contains the $j^{\,th}$ probe. Under the assumption that all spots have identical properties, that the system noise is additive and i.i.d, and that all probes show identical hybridization affinities with their complements, the optimal choice for G is the one that minimizes

$$\text{tr}\,(G^{*\,T}\, G^{*}), \tag{1}$$

where G^* denotes the well known pseudo-inverse of G defined as $G^* = (G^T\, G)^{-1}\, G^T$. An example of a multiplexing matrix with $M = 6$ and $N = 4$ was presented in [11] and is shown below.

$$G^T = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

For the above example, in the presence of any two spots failures the information about gene probes remains intact for any form of comparative study. Each spot contains two probes and each probe appears at three different spots. Note that the number of different probes placed at each spot should be small, and that the probes synthesized at the same spot should poorly hybridize with each other and their complementary DNA sequences.

IV. Joint Quality and Error-Control Coding

We describe next a new VLSIPS process for constructing multiplexed arrays based on a combination of ideas outlined in the previous two sections. The goal is to modify the VLSIPS process in such a way that multiple sequences can be synthesized at each spot, and that production failure can be tested under similar conditions as outlined in the Affymetrix standard.

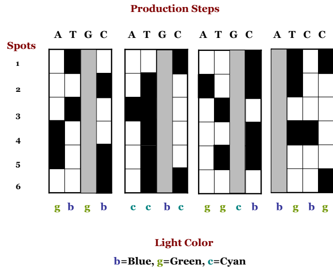

Fig. 2. A schedule for multiplexed array probe synthesis

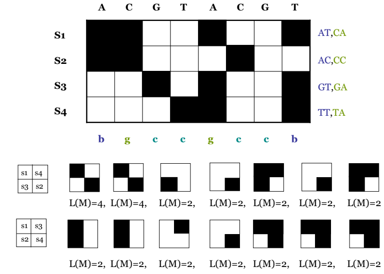

Fig. 3. Masks for generating multiple sequences per spot

Without loss of generality, assume that the goal is to synthesize two probes per spot. At the initial stage of production, two different types of linkers are placed at every spot. One type of linker is equipped with a photolabile group sensitive to green light, while the other linker contains a photolabile group susceptible to blue light. One of the probes at the spot will be grown from the green-tagged linkers, while the other will be generated starting from blue-tagged linkers. For simplicity, we will refer to the first probe as the *green probe* and to the second one as the *blue probe*. A controlled growth of the green and blue probes can be achieved by deactivating the photolabile groups through illumination by either green or blue light. At each step for which green light is used to deactivate the protective group (henceforth referred to as a *green step*), a solution of identical nucleotides tagged by green photolabile groups is applied to the wafer. A similar process is performed for all steps involving illumination by blue light. At first glance, this scheme seems to represent nothing more than a simple interleaving of two disjoint microarry production processes utilizing two different optical mechanisms. But with the proposed scheme one can also grow both probes present at a given spot during the same step of the production process. This can be achieved in several different ways. For simplicity, the reader may assume that *with the same mask*, the array is exposed first to green and then to blue light, successively deactivating the green and blue protective groups. In between these deactivation steps, solutions of appropriately tagged bases are added to the wafer. An example of the probe synthesis process for a multiplexed array is depicted in Figure 2. It is assumed that the schedule of nucleotides is a periodic repetition of $ATGC$, and each of the production steps is either blue b or green g, or a combination of both, denoted by c. For ease of visualization, a spot unmasked during manufacturing is indicated by a black square. The probes generated in Figure 2, along with their colors, are shown below.

$$\begin{array}{ll} Sp.1: TCC\,(\text{b}),\ CTC\,(\text{g}) & Sp.2: CTC\,(\text{b}),\ TAT\,(\text{g}) \\ Sp.3: TAT\,(\text{b}),\ ATT\,(\text{g}) & Sp.4: TCC\,(\text{b}),\ ATT\,(\text{g}) \\ Sp.5: CTC\,(\text{b}),\ ATT\,(\text{g}) & Sp.6: CTC\,(\text{b}),\ TCC\,(\text{g}). \end{array}$$

Several features of the multiplexed DNA probe design process can be deduced from Figure 2. First, in addition to a base schedule, for manufacturing purposes one also has to specify a *color schedule* (shown beneath the masks in Figure 2). The color schedule represents a sequence over the alphabet $\{b, g, c\}$ of length equal to the number of steps in the production process. If possible, both the base and color schedule should be designed so as to allow for fast assembly of all green and blue colored probes. As an example, observe that the fourth group of bases in Figure 2 contains the symbol C twice so that the base schedule is not periodic, but it reduces the overall number of steps in the production process. Note also that cyan colored steps allow for simultaneous extension of both probes, but they may be used only if the probes share the bases corresponding to these steps. Furthermore, sufficiently many blue/green steps have to be available for forming the prefixes of the probes bounded on the right by the cyan-colored base. Both the base and color schedules should facilitate the process of designing one given probe in terms of different subsets of steps. This feature ensures that not all replicas of a probe are formed by the same (and possibly faulty) production steps. Finally, it is to be noted that if the schedule is not properly chosen, many steps of the production process may remain unused, like the three gray-shaded blocks shown in Figure 2.

We discuss next some combinatorial and coding-theoretic questions arising in the study of sequence scheduling, mask design and quality control for multiplexed arrays.

Base Scheduling: As already pointed out, in addition to the base schedule, for multiplexed microarrays one also has to specify a color schedule. It is of interest to determine if the use of cyan colored steps can significantly reduce the number of productions stages for an asynchronous schedule. In the absence of cyan-colored steps, and for a synchronous schedule, the maximum number of steps needed to synthesize two probes per spot is $8\,L$, where L denotes the length of the probes.

Theorem 4.1: Let $\mathbf{U}$ and $\mathbf{W}$ be two sequences of length N over the alphabet $\mathcal{D} = \{A, T, G, C\}$ chosen uniformly and independently from each other. Assume that the length of the optimal asynchronous schedule for synthesizing $\mathbf{U}$ and $\mathbf{W}$ using only blue and green colors is $S_{b,g}(N)$, and that the length of the optimal asynchronous schedule for synthesizing $\mathbf{U}$ and $\mathbf{W}$ using blue, green and cyan colors is $S_{b,g,c}(N)$. Then $S_{b,g} = 2\,N$ and

$$\lim_{N\to\infty} \frac{E\left[S_{b,g,c}(N)\right]}{N} = \gamma,$$

where $1.29 \leq \gamma \leq 1.455$.
The proof of theorem 4.1 is based on some simple results from [4] and the references therein and is therefore omitted.

Conjecture 4.2.: Assume that two identical sets containing

M uniformly and independently chosen sequences of length $N > 3$ over the alphabet $\mathcal{D} = \{A, T, G, C\}$ are given. Let the sequences in the first set describe the blue probes, and let the sequences in the second set describe the green probes. Let the length of the optimal asynchronous schedule for synthesizing the $2M$ sequences by using blue, green and cyan colors be $S_{b,g,c}(N, M)$. Then

$$\lim_{M\to\infty} \lim_{N\to\infty} \frac{E\left[S_{b,g,c}(N, M)\right]}{N} \geq 2.35\,.$$

For most DNA microarrays probes have lengths not exceeding 30. For these lengths, the reduction in the number of production steps offered by a system with simultaneous probe extension features may not justify the high computational cost for finding a near-optimal schedule. But for certain non-medical applications of DNA microarrays, such as DNA cryptography, the probes are much longer. In this case, the production cost of a multiplexed array with steps of three colors may be much smaller than the one of its counterpart involving two colors only.

Mask Selection: For multiplexed arrays one also requires a set of carefully designed masks that can minimize the probability of erroneous probe synthesis. For the case that the green and blue probes are synthesized using a disjoint set of synchronous steps, the problem of mask design can be formulated in a similar manner as described in [7] (see Figure 3 for an illustration of the problem; there, $L(M)$ denotes the border length of a mask). A major modification in the procedure outlined in [7] is that each vertex Θ in the neighborhood graph corresponding to one spot is labeled by *two sequences*, specifying two differently colored probes $\mathbf{p}_1(\Theta)$ and $\mathbf{p}_2(\Theta)$. For two vertices in the neighborhood graph $\Theta_1 = \{\mathbf{p}_1(\Theta_1), \mathbf{p}_2(\Theta_1)\}$ and $\Theta_2 = \{\mathbf{p}_1(\Theta_2), \mathbf{p}_2(\Theta_2)\}$, the weight $\delta(\Theta_1, \Theta_2)$ of the edge connecting Θ_1 and Θ_2 equals $d_H(\mathbf{p}_1(\Theta_1), \mathbf{p}_1(\Theta_2)) + d_H(\mathbf{p}_2(\Theta_1), \mathbf{p}_2(\Theta_2))$. The sum of all distances $\delta(\Theta_1, \Theta_2)$ between vertices in the neighborhood graph corresponding to adjacent spots provides a measure of the total border-length of masks used during the production process. Note that in a multiplexed array, the mask border-length minimization problem has to be solved under a probe placement constraints imposed by the multiplexing matrix G. Furthermore, the probes at the same spot have to be at sufficiently large Hamming distance from each other in order to prevent erroneous hybridization. This makes the problem of placing similar probes at neighboring sites in the array quite challenging and poses new and interesting questions regarding optimal threading strategies.

The solution to the problem of minimizing border-length of masks with production steps of three colors strongly depends on the chosen base and color schedule. Due to space limitations, it will not be addressed in this paper.

Quality-Control: The process of testing a multiplexed array for faulty production steps can be partitioned into two stages. In the first stage, complements of green probes (appropriately tagged with green fluorescent tags) are applied to the wafer. At this stage, it is possible to infer if some of the steps using green or cyan light failed by designing quality control probe arrays based on superimposed codes. In the second stage of the testing process, the wafer is exposed to a solution of complementary sequences of blue probes. This stage allows for identifying erroneous blue steps in the production process, and it is facilitated by side-information provided from the previous stage. Notice that if no cyan steps are used, a quality control code appropriate for this setting reduces to an interleaved pair of superimposed codes. Otherwise, the quality control array should be constructed by using two superimposed codes that satisfy some simple joint constraints on the codewords specifying the cyan-colored production steps.

By referring to the results in [13] one can prove the following result regarding multiplexed arrays.

Theorem 4.2: Assume that there exists a linear error-control code with parameters $[n, k, d]$ containing the all-ones codeword. Then one can construct a quality control array for a multiplexed DNA chip with $2\cdot(2^k-2)$ disjoint blue and green production steps and n probes such that the length of each quality control probe is $2^{k-1} - 1$, and that the weights w of the columns in the quality control array satisfy $d \leq w \leq n-d$. Furthermore, with such an array any collection of less than $n/(n-d)$ failed blue or green steps, respectively, can be uniquely identified.

An interesting open problem is how to extend the result of theorem 4.2 in order to achieve the unique identification property in the presence of a given number of spot failures and involving cyan-colored production steps.

REFERENCES

[1] http://www.affymetrix.com

[2] N. Alon, C. Colbourn, A. Lings, and M. Tompa, "Equireplicate Balanced Binary Codes for Oligo Arrays," *SIAM J. Discrete Math*, vol. 14, no. 4, pp. 481-497, 2001.

[3] C. Colbourn, A. Ling, and M. Tompa, "Construction of Optimal Quality Control for Oligo Arrays," *Bioinformatics*, vol. 18 no. 4, pp. 529-535, 2002.

[4] V. Dančík, *Expected Length of Longest Common Subsequences*, Doctor of Philosophy Thesis, Department of Computer Science, University of Warwik, Sept. 1994.

[5] A. Dyachkov and V. Rykov, "A Survey of Superimposed Code Theory," *Problems in Control and Inform. Theory/Problemy Upravlen. Teor. Inform.*, vol. 12, no. 4, pp. 229242, 1983.

[6] W. Feldman and P. Pevzner, "Gray Code Masks for Sequencing by Hybridization," *Genomcis*, vol. 23, pp. 233-235, 1994.

[7] S. Hannehalli, E. Hubbell, R. Lipshutz, and P. Pevzner, "Combinatorial Algorithms for Design of DNA Arrays," *Advances in Biochemical Engineering/Biotechnology*, vol. 77, pp. 1-19, Springer Verlag, 2002.

[8] E. Hubbell and P. Pevzner, "Fidelity Probes for DNA Arrays," in *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology*, Heidelberg, Germany, pp. 113-117, August 1999.

[9] A. Kahng, I. Mandoui, S. Reda, X. Xu, A. Zelikovsky, "Design Flow Enhancement for DNA Arrays," *Proceedings of the 21st International Conference on Computer Design* (ICCD'03), 2003.

[10] W.H. Kautz and R.C. Singleton, "Non-random binary superimposed codes," *IEEE Transactions on Information Theory*, vol. 10, pp. 363-377, 1964.

[11] A. Khan, A. Ossadtchi, R. Leahy, and D. Smith, "Error-correcting Microarray Design," *Genomics*, vol. 81, pp. 157-165, 2003.

[12] R. Sengupta and M. Tompa, "Quality Control in Manufacturing Oligo Arrays: A Combinatorial Design Approach," in *Pacifc Symposium on Biocomputing*, Mauna Lani, Hawaii, pp. 348-359, January 2001.

[13] H. Vinck, S. Martirossian, and P. Gober, "On q-ary Superimposed Codes," *Proceedings of the IEEE Information Theory Workshop*, p. 118, South Africa, June 1999.