Configuration of VLSI Arrays in the Presence of Defects

JONATHAN W. GREENE AND ABBAS EL GAMAL

Stanford University, Stanford, California

Abstract. The penalties for configuring VLSI arrays for yield enhancement are assessed. Each element of the fabricated array is assumed to be defective with independent probability p. A fixed fraction R of the elements are to be connected into a prespecified defect-free configuration by means of switched interconnections. The probability that this can be done, known as the yield, must be bounded away from zero. The additional interconnections required increase the integrated circuit's area by the area overhead ratio AOR. Propagation delay is determined by the maximum connection length d. The following results are shown. Connection of RN fixed pins to distinct nondefective elements from an N-element linear array requires $d = \Theta(\log N)$, $AOR = \Theta(\log N)$. Connection of RN pairs of elements from two N-element linear arrays requires only constant d and AOR. Connection of a chain of RN^2 elements from an $N \times N$ array requires only constant d and AOR; this result is closely related to the percolation model of statistical physics. Connection of a $\sqrt{R}N \times \sqrt{R}N$ lattice from an $N \times N$ array requires $d = \Omega(\sqrt{\log N})$. Algorithms are presented that connect any fraction $R < 1 - p$ of the elements with yield approaching one as N increases.

Categories and Subject Descriptors: B.7.1 **[Integrated Circuits]**: Types and Design Styles—*VLSI (very large scale integration)*; B.7.3 **[Integrated Circuits]**: Reliability and Testing—*redundant design*; F.2.2 **[Analysis of Algorithms and Problem Complexity]**: Nonnumerical Algorithms and Problems—*routing and layout*

General Terms: Algorithms, Design, Reliability, Theory

Additional Key Words and Phrases: Circuit area, fault tolerance, percolation theory, probabilistic analysis, queuing processes, systolic arrays, wafer-scale integration, wire length

1. *Introduction*

As the size and complexity of very-large-scale integrated (VLSI) circuits grow, economic yields can be maintained by configuring each circuit around fabrication defects. Memories and processor arrays are particularly well suited to this technique because they have regular layouts and interchangeable elements. The technology for configurable circuits is well developed, having already found commercial application in 64K RAMs that include spare rows and columns [25]. Integrated switches that can be permanently opened or closed by a laser [17, 21, 25], and even nonvolatile electrically reprogrammable switches [14, 23] have been fabricated. These techniques are being used to develop memory systems that occupy an entire silicon wafer [14] and "restructurable" wafer-scale processor arrays [17].

The improvement in yield obtained comes at the expense of overhead area, occupied by the switches and extra interconnections, and an increase in signal

The work of the first author was partially supported by an NSF Graduate Fellowship and under Air Force Contract F 49620-79-C-0058. The work of the second author was partially supported under a DARPA contract MDA 903-79-C-0680 and Air Force contract F 49620-79-C-0058.

Authors' address: Information Systems Laboratory, Stanford University, Stanford, CA 94305.

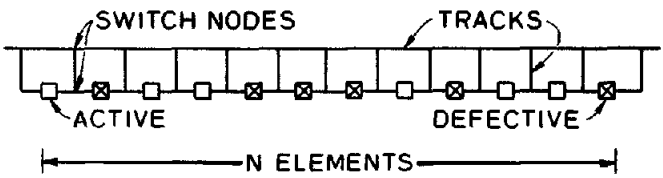

FIG. 1. A linear array of $N = 12$ elements after fabrication and testing. A switch node (omitted for clarity) is provided at every point where tracks meet.

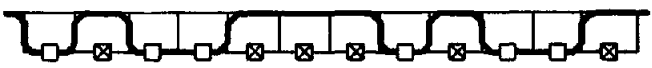

FIG. 2. Connection of a chain of $K = 6$ elements. The thick lines indicate the electrically connected segments of the wiring tracks.

propagation delay. Also, determining the switch settings may entail a nontrivial computation. In this paper, we investigate these penalties for several regular array configurations. Bounds on the severity of the penalties as a function of array size are derived, and algorithms for configuring the circuit are given. Our approach is best illustrated through the following simple example.

A linear array of K identical processors, connected in a chain, is to be implemented on a single integrated circuit. Assume that each processor, or more generally, circuit *element*, has an independent probability p of being *defective* and $1 - p$ of being *active*. Then the yield, or probability that a circuit is functional, is $(1 - p)^K$, approaching zero exponentially as K increases.

To prevent the yield from approaching zero, the number of elements on the circuit is increased to $N = K/R$ for some $R < 1 - p$, and switch nodes are provided to insert the elements in the chain. (See Figure 1.) After manufacture, the elements are tested. If the number of active elements is less than K, the circuit must be discarded. Otherwise, K elements can be connected as shown in Figure 2. Since $R < 1 - p$, the probability that the circuit has sufficient active elements approaches 1 exponentially as N increases. In this example, only one track is required in each wiring channel and so the overhead area is a constant times N.

Unfortunately, signals from one element to another can now encounter additional propagation delay since the connections between elements are longer than before. (The exact relationship between wire length and delay is not important here. See [2] for the details of this question.) Suppose that the maximum tolerable connection length is fixed at d, where the elements are spaced two units apart (to allow space for the vertical tracks). There are fewer than N elements at which to start a chain, probability $1 - p$ that the first element is active, and probability at most $1 - p^{d/2}$ that each connection can reach the next active element under the length constraint. Thus the probability that a chain can be connected is less than $N(1 - p)(1 - p^{d/2})^{K-1} < N \exp(-Kp^{d/2})$. This approaches zero as K increases unless the size of the circuit, N, grows exponentially with K—a very unsatisfactory situation.

We require here and throughout that the fraction of elements connected, R, must be held constant as N grows. Then the only way to maintain good yield is to permit d to grow with N. For fixed $R < 1 - p$ and arbitrarily large N, it is easily shown that $d = 2 + 4 \log N/(-\log p)$ permits connection of the chain with probability at least $1 - N^{-1}$.

Our analysis is based on two assumptions:

(1) The interconnect is defectfree. Since the interconnect requires fewer fabrication steps than the active circuitry and is not affected by variations in electrical parameters such as threshold voltage or leakage, defects in the interconnect are of secondary concern.
(2) Defects are independent and identically distributed. Though perhaps not completely accurate, this assumption does not seem to greatly limit the application of our results. It is unlikely that a single point defect would cause more than one element to fail, since the affected area is usually small compared with the size of an element.

For further discussion of these assumptions, see [9].

In the following section, we define four configuration problems for one- and two-dimensional arrays and state the main results of this paper. In Sections 3–6 we give the proofs of the results, including linear-time algorithms for programming the switches. Simulation data are also presented. In Section 7 we discuss some previous work, our results, and open problems.

2. *Problem Definitions and Results*

Except as noted, we assume the following wiring model. The elements are positioned in a linear or rectangular array. The region between two rows or columns of elements is known as a *channel*. Some number t of wiring *tracks* are positioned in each channel. For simplicity, we assume that tracks and elements are spaced at unit length intervals. (The model can be modified to accommodate elements of constant area more than 1 without affecting the order of growth of the bounds obtained.) Each track represents a physical signal path (or set of paths). At each point where tracks cross, a switch node is provided. After manufacture, the switch can be set so as to connect either or both of the crossing tracks in any way, including a crossover or two knockknees. (See Figures 3–6.) (Although crossovers are permitted, they are not necessary for any of our schemes.) Each resulting electrically connected path between elements is called a *connection*, and its length must be no greater than the *maximum connection length* d. The *area overhead ratio* (*AOR*) is defined as the total area of the circuit divided by the number of elements. It is determined by the number of tracks per channel and by the layout of the array.

A useful circuit results if the switches can be set to connect a fixed fraction ^1R of the total number of elements into some prespecified configuration, using only the tracks provided and without violating the length constraint d. The *yield* is the probability that the defects occur in such a way that this is possible. We are interested in finding the minimum order of growth of d and *AOR* needed to prevent the yield from approaching zero as the array size increases.

Let $O(\cdot)$ denote an upper bound, $\Theta(\cdot)$ an exact bound, and $\Omega(\cdot)$ a lower bound, all to within a constant factor. Our results are:

Section 3. The connection of a linear array of $K = RN$ fixed input/output pins (or ports) to distinct, active elements from a parallel N-element array is to be accomplished by means of a channel containing t wiring tracks between the pins

[1] To avoid divisibility problems, we assume that R is rational. This assumption incurs only a minor loss of generality since, to the authors' best knowledge, no circuit has ever contained an irrational number of components.

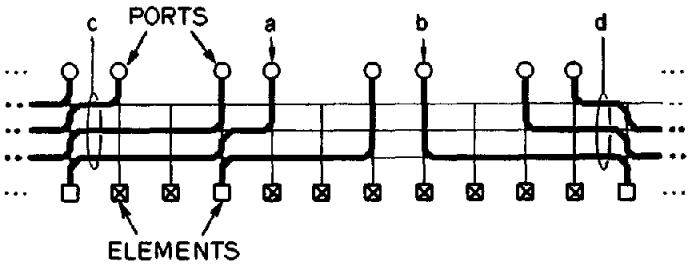

FIG. 3. A section of a selector. $K = (2/3)N$ and $t = 3$ wiring tracks are provided.

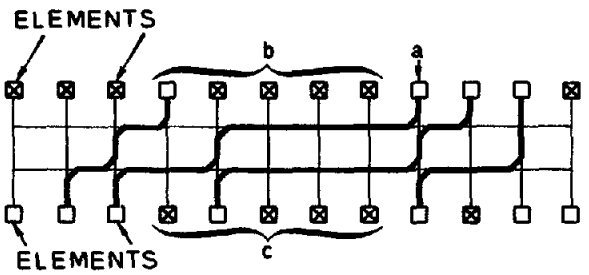

FIG. 4. Pairwise connection of two linear arrays of $N = 12$ elements.

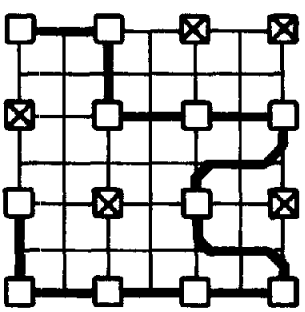

FIG. 5. Connection of a chain of $K = 11$ elements from a 4×4 array.

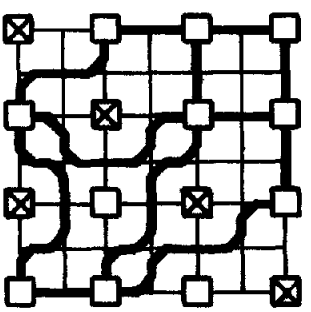

FIG. 6. Connection of a 3×3 square lattice from a 4×4 array.

and the array. We term this arrangement, shown in Figure 3, a *selector*. Theorem 1 states that unless $d = \Omega(\log N)$ and $t = \Omega(\log N)$, for an $AOR = \Omega(\log N)$, the yield approaches zero. This is due to the fact that the array contains a run of $\Theta(\log N)$ consecutive defects with probability $1 - O(N^{-1})$. The proof of Theorem 2 describes a queuing scheme that achieves these bounds for any $R < 1 - p$ with yield approaching 1.

Section 4. The connection of K pairs of active elements from two parallel N-element arrays, shown in Figure 4, is surprisingly easier than the task of the selector. A run of defects does not necessarily cause a problem here because there are no fixed pins—alternate pairs of elements can be connected from other parts of the array. The proof of Theorem 3 describes a scheme with constant d, t, and AOR that achieves yield $1 - O(N^{-1})$.

Section 5. The connection of a chain of $K = RN^2$ elements from an $N \times N$ array, as shown in Figure 5, can be achieved with no more than two wiring tracks between elements, constant AOR and d, and yield $1 - O(N^{-2})$ for any $R < 1 - p$. This is demonstrated by Theorem 4, which is based on certain results concerning the percolation model of statistical physics.

Section 6. The connection of a $K \times K$ square lattice from an $N \times N$ array of elements spaced at unit length, where $K^2 = RN^2$, is shown in Figure 6. This requires $d = \Omega(\sqrt{\log N})$ owing to the high probability that there is a block of $\Theta(\log N)$ defects that must be enclosed by a cycle of four connections. This result does not rely on the full wiring model above: Connections can be placed anywhere, even on top of elements. No lower bound is obtained on the required overhead area.

We use the following notations. For brevity, $1 - x$ is denoted by $\bar{x}$. The greatest integer less than or equal to x is denoted by $\lfloor x \rfloor$. The least integer greater than or equal to x is denoted by $\lceil x \rceil$. The cardinality of the set X is denoted by $||X||$. The indicator function $1(\cdot)$ is one if its argument is true, and zero otherwise. Natural logarithms are used throughout.

We also use several terms from graph theory, defined as follows. Two vertices v_1 and v_2 are *adjacent* if they share an edge (v_1, v_2). A *path* from v_1 to v_n is a sequence of distinct vertices and edges $v_1, (v_1, v_2), v_2, \ldots, (v_{n-1}, v_n), v_n$. A *cycle* is a path with $v_1 = v_n$. Two vertices are *connected* if there is a path from one to the other. A *component* is a maximal connected set of vertices; that is, every pair of vertices in the component is connected, but every proper superset contains a pair of vertices that is not connected. Where convenient, we use the terms *sites* or *elements* instead of *vertices*, and *connections* instead of *edges*.

3. *Selectors*

We begin by proving a lower bound on the maximum connection length d and number of tracks t required. This also gives a lower bound on the area overhead ratio since AOR is proportional to t in the selector.

Given a layout of selector, with pins placed in any way consistent with the wiring model, we show that the yield approaches zero unless d and t are $\Omega(\log N)$. For simplicity, d is taken to include only the horizontal length along the channel. The actual length is greater by the channel width, $t + 1$.

THEOREM 1. *For any $0 < \delta < 1$, the probability that $K = RN$ pins, aligned parallel to a linear array of N elements, can be connected to distinct active elements tends to zero as $\exp(-N^{1-\delta}\bar{p}/p)$ unless the number of tracks in the channel t and the maximum horizontal connection length d both satisfy*

$$d, t > \frac{\delta R \log N}{-2 \log p} - \frac{3}{2} = \Omega(\log N).$$

PROOF. For any $0 < \delta < 1$, let m be the largest odd integer less than or equal to $-\delta R \log(N)/\log(p)$. Divide the array and the pins by vertical cuts into $\lfloor (NR - m - 1)/m \rfloor$ blocks, each containing at least m pins, and a block at each end containing at least $(m + 1)/2$ pins. The total number of blocks is $B = \lfloor (NR - m - 1)/m \rfloor + 2 \geq NR/m$. Any block containing only one active element must have all but one of its pins connected to elements in other blocks. This requires that one side or the other of the block is crossed by at least $(m - 1)/2$ connections, and that at least one of these connections run a horizontal length of $(m - 1)/2$ before reaching an element in another block.

Suppose that $t < (m-1)/2$ or $d < (m-1)/2$. Then the selector can only be connected if all blocks have more than one active element. Let n_i be the total number of elements in block i. Since there must be more than one element in every block and a total of N elements in the array,

$$\sum_{i=1}^{B} n_i = N \qquad \text{and} \qquad n_i > 1 \qquad \text{for all} \quad i. \tag{3.1}$$

The yield is bounded by

$$\begin{aligned}
\text{yield} &\le \prod_{i=1}^{B} (1 - n_i p^{n_i - 1}\bar{p} - p^{n_i}) \\
&\le \left[1 - \left(\frac{N}{B}\right) p^{(N/B)-1}\bar{p} - p^{N/B}\right]^{B} \\
&\le \left[1 - \left(\frac{m}{R}\right) p^{(m/R)-1}\bar{p} - p^{m/R}\right]^{NR/m} \\
&< \exp(-Np^{(m/R)-1}\bar{p}) \\
&< \exp\left(\frac{-N^{1-\delta}\bar{p}}{p}\right).
\end{aligned}$$

The second inequality follows from constraints (3.1) and the fact that $\log(1 - xp^{x-1}\bar{p} - p^x)$ is convex $\cap$ in x for $x > 1$. (Convexity can be verified by finding that the second derivative is negative.) The fourth inequality follows from the relation $(1-x)^y \le \exp(-xy)$, and the last from the definition of m.

Substituting the definition of m into the assumptions on d and t yields the result. □

It is easily shown that $d, t = O(\log N)$ suffices to connect the selector with probability approaching 1. Simply divide the selector into $N/c \log N$ blocks of $c \log N$ elements and $Rc \log N$ pins for some properly chosen constant c. The Chernoff bound[2] can be applied to show that the probability that a given block has fewer than $Rc \log N$ active elements approaches zero exponentially in $c \log N$. Even when multiplied by the number of blocks, this value still approaches 0. Thus with probability approaching one, *all* blocks have at least enough active elements to connect their pins. Each block can then be connected separately using $t = Rc \log N$ tracks and maximum connection length $d = c \log N$.

Not surprisingly, the constant c is fairly large for this simple scheme. In the proof of the next theorem, we give a better scheme that is as easy to implement, though more difficult to analyze. The selector is constructed by distributing the pins nearly uniformly along the array. The scheme forms a queue of unconnected pins waiting for elements. The maximum queue size determines d and t.

THEOREM 2. *For any $R < \bar{p}$, let $z > 0$ be any constant such that*

$$\Phi(z) \triangleq p\ exp(zR) + \bar{p}\ exp(-z\bar{R}) < 1.$$

Then for any $\frac{1}{2} \le \delta < 1$ and arbitrarily large N it is possible to connect $K = RN$ pins to distinct active elements of a linear array with yield $1 - O(N^{1-1/\delta})$, using maximum horizontal connection length $d = \lceil \log(N)/2z\delta R \rceil = O(\log N)$ and a number of tracks $t = \lfloor Rd \rfloor = O(\log N)$.

[2] The Chernoff bound states that for a random variable X and any $z > 0$, $P(X \ge a) \le \exp(-za)E(\exp(zX))$ [8].

PROOF. We first describe the placement of the pins above the array. For $i \in \{1, 2, \ldots, N\}$, let $P_i = 1(\lfloor Ri \rfloor > \lfloor R(i-1) \rfloor)$. A pin is placed above element i if $P_i = 1$. Note that for any $0 \le i < j \le N$,

$$\lfloor R(j-i) \rfloor \le \sum_{k=i+1}^{j} P_k \le \lceil R(j-i) \rceil. \tag{3.2}$$

Let d and t be defined as in the statement of the theorem.

The connection procedure is phrased in terms of (+)-wires propagating from a pin rightward to an active element and (−)-wires propagating from an active element rightward to a pin. The procedure moves along the array from left to right, element by element. If an active element is encountered, the longest (+)-wire is terminated at it or else a new (−)-wire is started. If a pin is encountered, the longest (−)-wire is terminated at it or else a new (+)-wire is started. If more than t (−)-wires are stacked up, the longest one is removed and its element remains unused. Note that at any point along the array all wires must be of the same type.

For $1 \le i \le N$, let $W_i \in \{0, 1, \ldots\}$ equal t plus the number of (+)-wires less the number of (−)-wires passing to the right of element i. Define $W_0 = t$. Under the above procedure,

$$W_i = \max\{P_i - A_i + W_{i-1}, 0\}, \tag{3.3}$$

where $A_i = 1$ (element i is active). We can interpret W_i as the size of a queue.

The procedure can fail only in one of the following four circumstances:

(1) There are more than t (+)-wires at some point in the array.
(2) There are any (+)-wires left at the end of the array.
(3) A (+)-wire propagates from an as yet unconnected pin past d elements. Then, by (3.2), the wire passes at least $\lfloor Rd \rfloor$ pins connected to subsequent (+)-wires, implying that at least $\lfloor Rd \rfloor + 1$ (+)-wires are stacked up (e.g., in Figure 3 the (+)-wire leaving pin (b) passes three elements, so there are at least $\lfloor (\frac{2}{3}) \cdot 3 \rfloor + 1 = 3$ (+)-wires at (d)). Since $\lfloor Rd \rfloor + 1 > t$, this implies the occurrence of case (1) above.
(4) A (−)-wire propagates past d elements. Then, by (3.2), the wire passes at least $\lfloor Rd \rfloor$ pins connected to previous (−)-wires, implying that there are at least $\lfloor Rd \rfloor + 1$ (−)-wires at the element where the wire began (e.g., in Figure 3 the (−)-wire terminating at pin (a) passes three elements, so there are at least $\lfloor (\frac{2}{3}) \cdot 3 \rfloor + 1 = 3$ (−)-wires at (c)). But the procedure never stacks up more than $t < \lfloor Rd \rfloor + 1$ (−)-wires, so this cannot occur.

Thus the array is connected if the number of (+)-wires stacked up never exceeds t and no (+)-wires are left after the last element; that is, $W_i < 2t + 1$ for $1 \le i \le N - 1$ and $W_N < t + 1$.

We now show that this happens with probability approaching 1. Let $X_j = R - A_j$. By (3.2),

$$\sum_{j=k+1}^{i} P_j - A_j \le \lceil R(i-k) \rceil - \sum_{j=k+1}^{i} A_j < 1 + \sum_{j=k+1}^{i} X_j.$$

Using a standard trick of queuing theory [5], we apply (3.3) recursively to obtain

$$W_i = \max\left\{0, (P_i - A_i), \sum_{j=i-1}^{i} (P_j - A_j), \ldots, \sum_{j=2}^{i} (P_j - A_j), t + \sum_{j=1}^{i} (P_j - A_j)\right\}$$

$$< 1 + \max\left\{0, X_i, \sum_{j=i-1}^{i} X_j, \ldots, \sum_{j=2}^{i} X_j, t + \sum_{j=1}^{i} X_j\right\}. \tag{3.4}$$

Since the $\{X_i\}$ are independent and identically distributed, one can show by interchanging them that (3.4) has the same probability distribution as

$$1 + \max\{0, S_1, S_2, \ldots, S_{i-1}, t + S_i\}, \tag{3.5}$$

where $S_i = \sum_{j=1}^{i} X_j$. Applying the Chernoff bound, for any $z > 0$,

$$\begin{aligned} P(S_i \geq u) &\leq \exp(-zu)E(\exp(zS_i)) \\ &= \exp(-zu)\Phi^i(z), \end{aligned} \tag{3.6}$$

where $\Phi(z)$ is as defined above. This bound is useful only if $\Phi(z) < 1$. Fortunately, for $R < \bar{p}$ there are always $z > 0$ such that this is the case. (In particular, $z = \log(\bar{p}\bar{R}/pR)$ minimizes $\Phi(z)$.)

Applying (3.4)–(3.6), we can bound the probability that the procedure fails:

$$\begin{aligned} \text{1-yield} &\leq \sum_{i=1}^{N-1} P(W_i \geq 2t + 1) + P(W_N \geq t + 1) \\ &\leq \sum_{i=1}^{N-1} P(\max\{0, S_1, \ldots, S_{i-1}, t + S_i\} \geq 2t) \\ &\quad + P(\max\{0, S_1, \ldots, S_{N-1}, t + S_N\} \geq t) \\ &\leq \sum_{i=1}^{N-1} \left[\sum_{j=1}^{i-1} P(S_j \geq 2t) + P(S_i \geq t)\right] + \sum_{j=1}^{N-1} P(S_j \geq t) + P(S_N \geq 0) \\ &\leq \sum_{i=1}^{N-1} \left[\sum_{j=1}^{i-1} e^{-z2t}\Phi^j(z) + e^{-zt}\Phi^i(z)\right] + \sum_{j=1}^{N-1} e^{-zt}\Phi^j(z) + \Phi^N(z) \\ &= e^{-z2t}\frac{\Phi(z)}{1 - \Phi(z)}\left[N - 1 - \frac{1 - \Phi^{N-1}(z)}{1 - \Phi(z)}\right] \\ &\quad + 2e^{-zt}\Phi(z)\frac{1 - \Phi^{N-1}(z)}{1 - \Phi(z)} + \Phi^N(z) \\ &= O(N^{1-1/\delta}) + O(N^{-1/(2\delta)}) + O(\Phi^N(z)) \end{aligned}$$

since $t > Rd - 1 \geq \log(N)/2z\delta - 1$ and $\Phi(z) < 1$ is constant. For $\delta \geq 1/2$, $O(N^{1-1/\delta})$ is the dominant term. □

We have now demonstrated that the maximum connection length and number of tracks must grow asymptotically as $\Theta(\log N)$. For example, let $p = 0.5$ and $R = \frac{3}{7} \approx 0.4286$, so that about 85 percent of the expected number of active elements must be connected. Theorem 1 indicates that $t > (0.309)\log N$ is required to bound the yield away from zero. On the other hand, choosing z as large as possible in Theorem 2 shows that $t \approx (0.865)\log N$ achieves yield approaching 1.

Empirical results for the algorithm of Theorem 2, using the above values of p and R and four fixed values of N, appear in Figure 7. Note that each time N triples, an additional track is required to maintain about the same yield.

4. *Pairing of Two Parallel Arrays*

As in Section 3, simplicity leads us to consider only horizontal distance along the channel when measuring connection length.

It is easily shown that parallel connection of two linear arrays requires only constant d and t, independent of N. Construct the arrays of blocks that are b elements wide, as shown in Figure 8. Within each block it is possible to connect a number of pairs equal to the minimum of the number of active elements from the upper array and the number of active elements from the lower array. Connections of length $d = b - 1$ or less and $t = \lfloor b/2 \rfloor$ tracks suffice. As long as b is chosen so

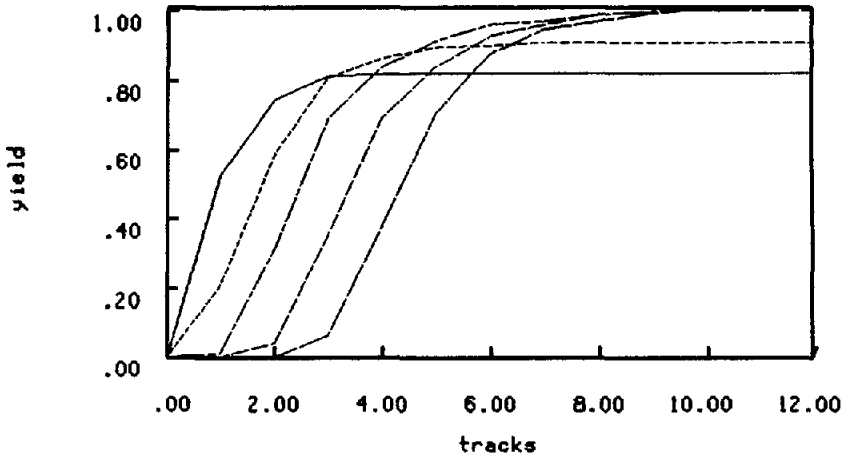

FIG. 7. Yield over 500 trials for an N element selector with $p = 0.5$, $R = 3/7$, t as indicated on the horizontal axis, and $d = 2t$. $N = 21$ (———), 63 (- - -), 189 (— · — · —), 567 (- - -), and 1701 (— —).

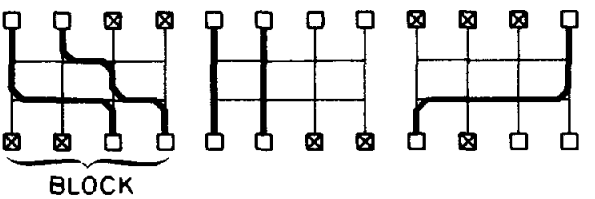

FIG. 8. A simple pairing scheme. In this example, the two linear arrays of $N = 12$ elements are divided into blocks that are $b = 4$ elements wide. The number of tracks is $t = \lfloor b/2 \rfloor = 2$.

that the expected value of this minimum is greater than Rb, the law of large numbers ensures that the total number of pairs connected is at least RN as $N \to \infty$. Only constant b is needed. However, the constant depends on the fraction of active elements that are connected: As R approaches $\bar{p}$, the constant becomes larger.

This simple scheme does not achieve a very good constant since it connects blocks independently. Fortunately, if we attempt to improve the constant by removing this restriction, the dependence between distant parts of the array remains negligible. This is exploited in the following theorem.

THEOREM 3. *For arbitrarily large N, and any $R < \bar{p}$, $\delta > 0$, $c < \bar{p}$ and integer $t > 0$ such that*

$$R \le \bar{p} - \frac{p\bar{p}}{2t+1} - \frac{\bar{p}^3 p^2 c^2}{(\bar{p}-c)^4 (t-1)^2} - 2\delta,$$

$K = RN$ pairs of active elements can be connected from two N-element arrays with $t = O(1)$ tracks, maximum horizontal connection length $d = \lfloor t/c \rfloor = O(1)$ and yield $1 - O(N^{-1})$.

PROOF. We use a Markov queuing model similar to that used for the selector in Theorem 2, but with two differences:

(1) In the selector, all pins must be connected. Here, some small fraction $\bar{p} - R$ of the active elements in each array need not be connected. Therefore an excess of (+)-wires from elements in the top array can be treated in the same way as an excess of (−)-wires: If $t + 1$ wires stack up, the oldest is dropped and its element left unused.
(2) Without the regularly spaced pins, adherence to the track constraint does not guarantee adherence to a length constraint. We therefore analyze the suboptimal scheme that first connects the array without regard to connection length and then removes any connections that exceed the length constraint.

The proof is in two steps. First, we show that with probability $1 - O(N^{-1})$ the scheme makes at least $[\bar{p} - p\bar{p}/(2t + 1) - \delta]N$ connections. Then we show that with probability $1 - O(N^{-1})$ at most $[\bar{p}^3p^2c^2/(\bar{p} - c)^4(t - 1)^2 + \delta]N$ of these connections need be removed.

Step 1. For $1 \le i \le N$, let $W_i \in \{-t, \ldots, 0, 1, \ldots, t\}$ equal the number of (+)-wires less the number of (−)-wires passing to the right of the ith elements in the arrays. Define $W_0 = 0$. Let $T_i = 1$(element i in the top array is active). Define B_i similarly for the bottom array. Under the above procedure,

$$W_i = \begin{cases} -t & W_{i-1} + T_i - B_i \le -t, \\ t & W_{i-1} + T_i - B_i \ge t, \\ W_{i-1} + T_i - B_i & \text{otherwise.} \end{cases}$$

It is easily shown that the Markov process $\{W_i\}$ has symmetric transition probabilities; that is, $P(W_i = y \mid W_{i-1} = x) = P(W_i = x \mid W_{i-1} = y)$. The process, therefore, has a uniform stationary distribution $P(W = x) = 1/(2t + 1)$ [4, p. 182]. The convergence to the stationary distribution is rapid. In fact, there exist $\alpha > 0$ and $0 < \beta < 1$ such that for any $j > i$ [4, p. 173],

$$\left| P(W_j = y \mid W_i = x) - \frac{1}{2t + 1} \right| \le \alpha\beta^{j-i}. \tag{4.1}$$

(This is a precise statement of the assertion above that the dependence between distant parts of the array is negligible.)

Let C_j denote the event that a connection is completed during the transition to W_j. (Clearly, only one connection can be completed per transition.) Then $C_j = \{W_{j-1} < 0, T_j = 1\} \cup \{W_{j-1} > 0, B_j = 1\} \cup \{W_{j-1} = 0, T_j = B_j = 1\}$. Since $P(B_j = 1) = P(T_j = 1) = \bar{p}$, independent of W_{j-1},

$$\begin{aligned} P(C_j) &= \bar{p}P(W_{j-1} < 0 \mid W_0 = 0) + \bar{p}P(W_{j-1} > 0 \mid W_0 = 0) \\ &\quad + \bar{p}^2P(W_{j-1} = 0 \mid W_0 = 0) \\ &= \bar{p} - p\bar{p}P(W_{j-1} = 0 \mid W_0 = 0) \\ &\ge \bar{p} - p\bar{p}\left(\frac{1}{2t + 1} + \alpha\beta^{j-1}\right). \end{aligned} \tag{4.2}$$

The last line follows from (4.1). Thus the expectation of the number of connections made, $S_C \triangleq \sum_{j=1}^{N} 1(C_j)$, satisfies

$$E(S_C) = \sum_{j=1}^{N} P(C_j) \ge N\left(\bar{p} - \frac{p\bar{p}}{2t + 1}\right) - O(1).$$

For $j > i + 1$, C_i and W_0 affect W_{j-1} only through W_i, so we can apply (4.1) and (4.2) to show that

$$\begin{aligned} P(C_j \mid C_i) &= \bar{p} - p\bar{p}P(W_{j-1} = 0 \mid C_i,\ W_0 = 0) \\ &\le \bar{p} - p\bar{p}\left(\frac{1}{2t+1} - \alpha\beta^{j-1-i}\right) \\ &\le P(C_j) + p\bar{p}\alpha(\beta^{j-1-i} + \beta^{j-1}). \end{aligned}$$

This permits us to upper bound the variance:

$$\begin{aligned} \operatorname{var}(S_C) &= E(S_C)^2 - E^2(S_C) \\ &= \sum_i \sum_j P(C_iC_j) - P(C_i)P(C_j) \\ &= \sum_{|j-i|\le 1} \sum_j P(C_i)[P(C_j \mid C_i) - P(C_j)] \\ &\quad + 2\sum_{j>i+1} \sum_j P(C_i)[P(C_j \mid C_i) - P(C_j)] \\ &\le \sum_{|j-i|\le 1} \sum_j 1 + 2\sum_{j>i+1} \sum_j p\bar{p}\alpha(\beta^{j-1-i} + \beta^{j-1}) \\ &= O(N). \end{aligned}$$

We apply Chebyshev's inequality [8] to bound the probability that fewer than the stated number of pairs are connected:

$$\begin{aligned} P\left(S_C < \left(\bar{p} - \frac{p\bar{p}}{2t+1} - \delta\right)N\right) &\le P(|E(S_C) - S_C| > \delta N - O(1)) \\ &\le \frac{\operatorname{var}(S_C)}{(\delta N - O(1))^2} \\ &= O(N^{-1}). \end{aligned}$$

Step 2. A connection of length exceeding d is completed during the transition to W_j only if: $j > d + 1$, the jth element in the proper array (top or bottom) is active, $\sum_{i=j-d}^{j-1} T_i \le t - 1$ and $\sum_{i=j-d}^{j-1} B_i \le t - 1$. (E.g., the wire terminating at element a in Figure 4 could not run length 6 if element a were not active, if there were $t = 2$ active elements among those five marked b, or if there were $t = 2$ active elements among those marked c.) Denote the conjunction of these events by D_j. Then the number of connections that must be removed is at most $S_D \triangleq \sum_{j=1}^{N} 1(D_j)$.

For $j \le d + 1$, $P(D_j) = 0$. For $j > d + 1$,

$$\begin{aligned} P(D_j) &= \bar{p}P\left(\sum_{i=j-d}^{j-1} T_i \le t - 1\right) P\left(\sum_{i=j-d}^{j-1} B_i \le t - 1\right) \\ &\le \frac{\bar{p}^3p^2c^2}{(\bar{p} - c)^4(t-1)^2}. \end{aligned} \tag{4.3}$$

(The last line follows from Chebyshev's inequality.) Thus we have

$$E(S_D) \le N\frac{\bar{p}^3p^2c^2}{(\bar{p} - c)^4(t-1)^2}.$$

For $j > i + d$, no T_k or B_k, $1 \le k \le N$, is involved in both D_i and D_j. Then $P(D_j \mid D_i) = P(D_j)$. Using techniques similar to those of Step 1, we can show that $\mathrm{var}(S_D) = O(N)$, and so the probability that more than a fraction $P(D_j) + \delta$ of the connections must be removed is $O(N^{-1})$. This completes Step 2 and the proof. □

As an example, for the values $p = 0.5$ and $R = \frac{3}{4}$, Theorem 3 indicates that $t = 2$ and $d = 5$ suffice to achieve yield approaching one as $N \to \infty$. (Here we bound the fraction of connections removed by (4.3), which can be evaluated using the tabulated cumulative binomial distribution.) This can be compared to the empirical results shown in Figure 9. These were obtained using an optimal scheme that proceeds from left to right along the array connecting each active element whenever the constraints permit.

5. *Chains Connected from a Two-Dimensional Array*

The problem of connecting a chain of active elements from a two-dimensional array is closely related to percolation theory. Percolation processes have been studied extensively since they were first defined by Broadbent and Hammersley [3]. A recent survey appears in [28].

The site percolation problem concerns an infinite lattice of sites and edges. Each site is independently *vacant* with some probability q or *occupied* with probability $\bar{q}$. We define a *cluster* to be a connected set of occupied sites, together with all adjacent (vacant) sites. A site is said to *percolate* if it is a member of an infinite cluster.[3] The probability that a site percolates is the same for any site and so can be expressed as a percolation probability function $R(\bar{q})$, which is monotonic increasing and attains the value 1 at $\bar{q} = 1$. Broadbent and Hammersley demonstrated that $R(\bar{q}) = 0$ for $\bar{q}$ less than some critical value characteristic of the lattice. Little else of an analytical nature is known about $R(\bar{q})$, although Monte Carlo estimates have established empirical curves for various lattices [6]. The curve for a square lattice is reproduced in Figure 10.

Our scheme for connecting a chain from an array can be analyzed using some results in percolation theory, contained in Lemmas 1–4 below. We restrict consideration to a square lattice, though the results readily generalize to other planar lattices.

Although in practice we can read $R(\bar{q})$ from Figure 10, an analytical lower bound is required for the task at hand. We obtain a weak bound using techniques similar to those of Hammersley [11]. A site percolates unless it is enclosed by a set of vacant sites. The set is minimal if it does not contain a proper subset that also encloses the given site. We call such a set an *enclosing walk* because it forms a closed self-avoiding walk, stepping from one vacant site to another; diagonal as well as horizontal and vertical steps are permitted since such a walk can enclose a site. Lemma 1 upper bounds the probability that a given site is inside an enclosing walk and thus lower bounds the percolation probability.

LEMMA 1. *For any $\delta > 0$, if $q \le \frac{1}{5} - \delta$, then $R(\bar{q}) \ge 1 - c_1 q^4$ for some constant c_1 dependent only on δ.*

PROOF. Every enclosing walk of length L surrounding a given site a must contain at least one of the first $\lfloor L/2 \rfloor$ sites directly above site a. Without loss of

[3] This definition of percolation is similar to that of [6]. It differs slightly from the more common one in [28] but is more suitable here.

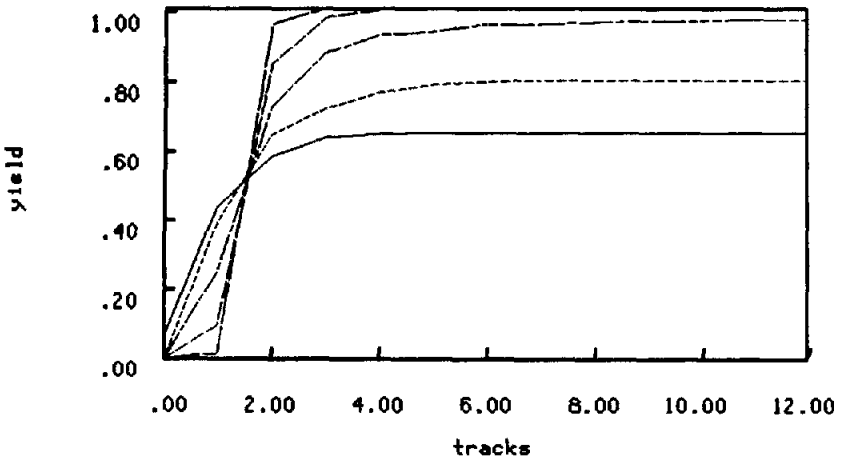

FIG. 9. Yield over 500 trials for pairing of two N element arrays with $p = 0.5$, $R = 3/7$, t as indicated on the horizontal axis, and $d = 2t$. $N = 21$ (———), 63 (- - -), 189 (— - — -), 567 (- - -), and 1701 (— —).

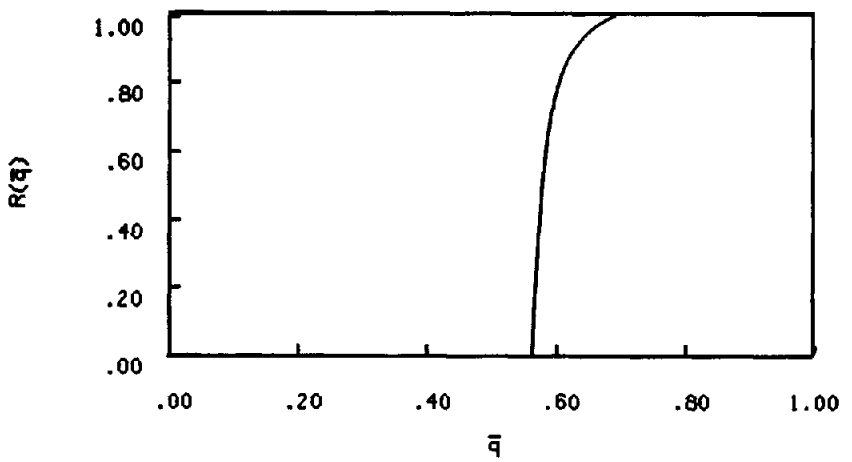

FIG. 10. The percolation function $R(\bar{q})$ for the square lattice as determined by Monte Carlo estimates. (Data taken from Figure 6 of [6].)

generality, consider some such site as the start of the walk. The enclosing walk makes a first step in one of eight directions to the second site. At each subsequent step, there are at most five choices for the next site that could not have been reached directly from the previous site. Thus there are no more than $(L/2)8(5)^{L-2}$ distinct enclosing walks of length L. Furthermore, each of the L sites in an enclosing walk must be vacant, which happens with probability q^L.

The probability that site a is inside an enclosing walk of length at least L_0 is, by the above arguments, at most

$$\sum_{L=L_0}^{\infty} \left(\frac{L}{2}\right) 8(5)^{L-2} q^L = \frac{4}{25} (5q)^{L_0} \left[\frac{L_0}{(1-5q)} + \frac{5q}{(1-5q)^2}\right]$$

$$\leq \left(\frac{4}{25}\right) (5q)^{L_0} \left[\frac{L_0}{5\delta} + \frac{1-5\delta}{25\delta^2}\right].$$

Since an enclosing walk must contain at least 4 sites, we substitute $L_0 = 4$ and obtain the result. □

We now prove that within large finite regions of the infinite lattice, the fraction of sites that percolate converges to $R(\bar{q})$.

LEMMA 2. *Let X be the number of sites in an $N \times N$ section of the infinite square lattice that percolate. If $q < \frac{1}{5}$, then for any $r < R(\bar{q})$, $P(X \le rN^2) \le O(N^{-2})$.*

PROOF. Let A and B denote, respectively, the events that sites a and b percolate. Let $d(a, b)$ be the Manhattan (rectilinear) distance between a and b. Let W denote the event that there is an enclosing walk surrounding a or b of length at least $d(a, b)/2$, and W^c its complement. Note that there can be no overlap between a walk of length less than $d(a, b)/2$ enclosing a and a walk of length less than $d(a, b)/2$ enclosing b. Thus A and B are conditionally independent given W^c, so that

$$\begin{aligned} P(A, B) &= P(A, B, W^c) \\ &= P(A \mid W^c)P(B \mid W^c)P(W^c) \\ &\le \frac{P(A)P(B)}{P(W^c)} \\ &= P(A)P(B)\left[1 + \frac{P(W)}{P(W^c)}\right]. \end{aligned}$$

In the proof of Lemma 1 it was demonstrated that if $q < \frac{1}{5}$, the probability that a given site is enclosed by a walk of length $L \ge L_0$ is at most $c_2 L_0 (5q)^{L_0}$, for some constant c_2 independent of L_0. We can therefore choose constants d_0 and $c_3 > 0$ such that, if $d(a, b) \ge d_0$, then $P(W) \le 2c_2(d(a, b)/2)(5q)^{d(a,b)/2} \le 1 - c_3$ and $P(W^c) \ge c_3$.

Finally, note that the number of sites at distance $d > 0$ from a given site on a square lattice is $4d$. We can now upper bound the variance of X as follows:

$$\begin{aligned} \operatorname{var}(X) &= E(X^2) - E^2(X) \\ &= \sum_a \sum_b [P(A, B) - P(A)P(B)] \\ &\le \sum_{\substack{a \\ d(a,b)<d_0}} \sum_b 1 + \sum_{\substack{a \\ d(a,b)\ge d_0}} \sum_b \left[P(A)P(B)\left[1 + \frac{P(W)}{P(W^c)}\right] - P(A)P(B)\right] \\ &\le \sum_a \left[1 + \sum_{d=1}^{d_0-1} (4d)\right] + \sum_a \sum_{d=d_0}^{2N} 4d \, \frac{2c_2(d/2)(5q)^{d/2}}{c_3} \\ &= O(N^2)O(1) + O(N^2)O(1) \\ &= O(N^2). \end{aligned}$$

The expectation of X is $EX = N^2 R(\bar{q})$. We now apply the Chebyshev inequality:

$$\begin{aligned} P(X \le rN^2) &\le P(|X - EX| \ge EX - rN^2) \\ &\le \frac{\operatorname{var}(X)}{(EX - rN^2)^2} \\ &= O(N^{-2}). \end{aligned}$$

□

LEMMA 3. *Suppose $q < \frac{1}{5}$. Consider those sites within an $N \times N$ section of the infinite lattice that belong to infinite clusters. Except for a negligible fraction*

$O(N^{-1}\log N)$, these sites form a single cluster within the $N \times N$ section, with probability $1 - O(N^{-2})$.

PROOF. When the $N \times N$ section is removed from the infinite lattice, the parts of the infinite cluster(s) that lie within the section may be disconnected into several components. Each component must be separated from the others by a self-avoiding walk on vacant sites starting and ending at the boundary of the section. By arguments similar to those in the proof of Lemma 1, it is easily shown that the probability of a self-avoiding walk on vacant sites starting on the boundary of the $N \times N$ section and having length at least $L_{\max} = -3 \log(N)/\log(5q)$ is $O(N^{-2})$. Thus with high probability, only sites within $L_{\max}$ of the boundary are cut off from the largest cluster. These sites account for a fraction $O(N^{-1}\log N)$ of the N^2 sites. □

The following intuitive lemma, known as the correlation inequality, is proved in [12, Lemma 4.1].

LEMMA 4. *For each $i \in \{1, \ldots, n\}$, let A_i be the event that every site in some finite nonempty set S_i is occupied. For each $j \in \{1, \ldots, m\}$, let B_j be the event that every site in some finite nonempty set T_j is occupied. No assumption about the exclusivity of the sets is made. Let $A = \cup A_i$ and $B = \cup B_j$. Then $P(A \mid B) \geq P(A)$.*

We are now ready for the main result of this section.

THEOREM 4. *For arbitrarily large N and any $R < \bar{p}$, a chain of length $K = RN^2$ can be connected from an $N \times N$ array with yield $1 - O(N^{-2})$ and maximum connection length*

$$d = \left\lceil \left(\frac{9 \log((\bar{p} - R)/c)}{\log(p)} \right)^{1/2} \right\rceil,$$

for some constant $c > 1$. No more than two tracks are required in any channel.

PROOF. The general idea is as follows. Construct the array of N^2/b square blocks of b elements each. Choose the constant b so that each block has high probability of containing at least four active elements.

Each block can be considered as corresponding to a site on a square lattice, and if the block has at least four active elements, consider the site occupied. Using Lemmas 1–3, we show that nearly all sites belong to a single large cluster. A tree of maximum degree 4 that spans the cluster, with all nonleaf sites occupied, can be constructed. This can also be considered as a spanning tree on the blocks. All the active elements in the cluster can be connected into a chain by looping around the tree, as shown in Figure 11. Since all "nonleaf" blocks have at least four active elements, it is never necessary to connect two elements from nonadjacent blocks. The construction of the spanning tree takes only $O(K)$ time. The connection of subchains in the blocks also takes $O(K)$ time since there are only a constant number of elements in each block.

Only two tracks are needed between adjacent blocks since only two connections are made between the blocks. The connection of all active elements within a block to the tracks between blocks requires only one track between elements. This is straightforward to prove by double induction from an $i \times j$ block to an $(i + 1) \times j$ or $i \times (j + 1)$ block. The maximum Manhattan connection distance required is

$$d = 6\sqrt{b} - 3. \tag{5.1}$$

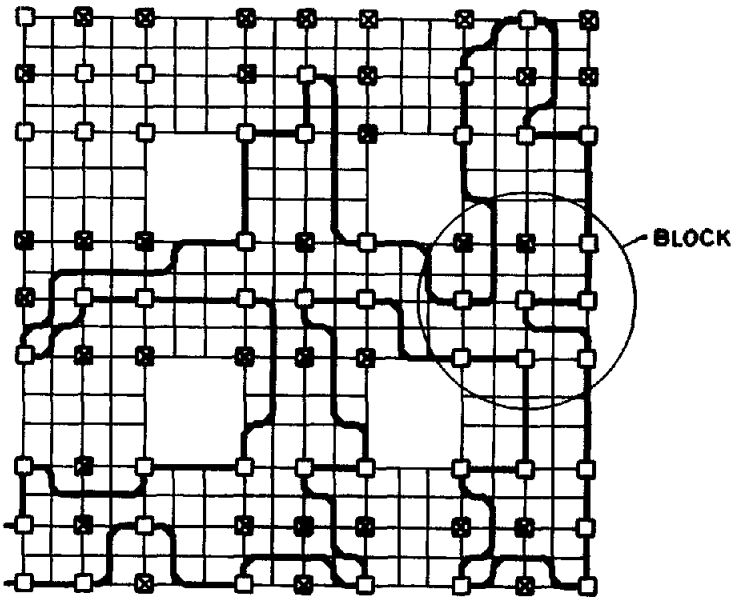

FIG. 11. A section of an array connected into a chain. Each block contains $b = 9$ elements. One track is provided between elements within a block, and two between blocks.

Now for the details. For any choice of $b \geq 4$, let q be the probability that a particular block contains fewer than four active elements. Then

$$q = \sum_{i=0}^{3} \binom{b}{i} \bar{p}^i p^{b-i} \leq \left(\frac{15}{64}\right) b^3 p^{b-3}. \tag{5.2}$$

Thus q can be made arbitrarily small by choice of b. As long as $q < \frac{1}{5}$, we can apply Lemmas 1–3 to percolation on an $N/\sqrt{b} \times N/\sqrt{b}$ lattice of sites corresponding to blocks, as described above.

By Lemmas 2 and 3, for any

$$r < R(\bar{q}), \tag{5.3}$$

at least rN^2/b blocks form a single cluster, with probability $1 - O(N^{-2})$.

We proceed to bound the number of active elements in any such rN^2/b blocks. Choose any

$$R < r\bar{p}. \tag{5.4}$$

By a simple application of Lemma 4, the probability that the number of active elements in the blocks is at least RN^2, given that all the blocks are in a cluster, is at least as great as the unconditional probability, where the elements of the blocks are considered independent. Since $R < r\bar{p}$, application of the Chebyshev inequality proves that the unconditional probability is $1 - O(N^{-2})$.

By Lemma 1 and (5.2),

$$\begin{aligned} R(\bar{q})\bar{p} &\geq [1 - c_1 q^4]\bar{p} \\ &\geq [1 - c_1((\tfrac{15}{64})b^3 p^{b-3})^4]\bar{p} \\ &\geq \bar{p} - cp^{d^2/9} \end{aligned} \tag{5.5}$$

for some constant $c > 1$ by (5.1).

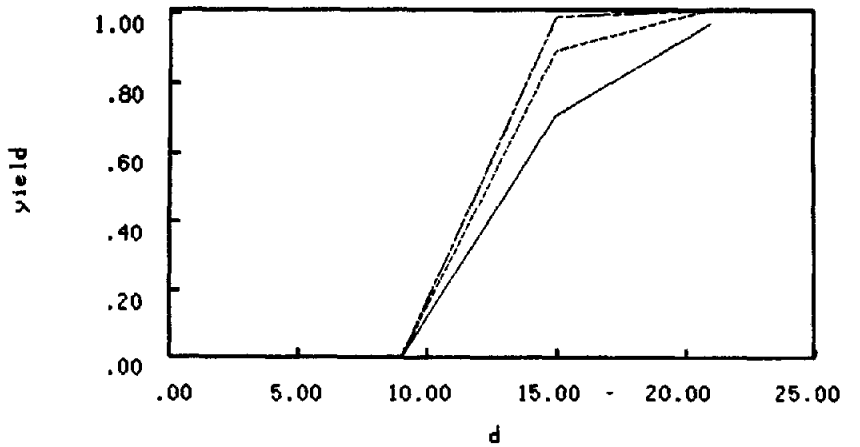

FIG. 12. Yield over 500 trials for connection of a chain from an $N \times N$ array with $p = 0.5$, $R = 3/7$, and d as indicated on the horizontal axis. $N = 12$ (———), 36 (- - -), and 60 (— · — · —).

Combining (5.3)–(5.5), we have shown that for any R such that $R \leq p - cp^{d^2/9}$, or equivalently for any

$$d \geq \left[\frac{9 \log((\bar{p} - R)/c)}{\log(p)}\right]^{1/2},$$

a chain of RN^2 elements can be connected with probability $1 - O(N^{-2})$. □

The following example illustrates how this scheme can be used in practice. Suppose each element is defective with probability $p = 0.5$. Choose $b = 9$ and $d = 6\sqrt{9} - 3 = 15$. The probability that a block of nine elements has at least four active elements is $\bar{q} \approx 0.7461$. From Figure 10 we see that for an infinite square lattice and this value of $\bar{q}$, any block is practically certain to belong to an infinite cluster. Even for finite arrays, $d = 15$ should suffice to achieve reasonable yields. This estimate is confirmed by the empirical results shown in Figure 12. These were obtained using the scheme of Theorem 4 with $p = 0.5$, $R = \frac{3}{7}$, and $d = 9$, 15, and 21, corresponding to $b = 4$, 9, and 16.

6. *Lattices Connected from Two-Dimensional Arrays*

Now we examine the connection of a square lattice from a two-dimensional array. Before proving a lower bound on the maximum connection length, we state the following lemma concerning the separability of the square lattice.

LEMMA 5. *Consider the graph corresponding to a $K \times K$ square lattice. Any partition of the K^2 vertices into three sets A, B, and C, such that $||A|| \leq ||B||$ and no vertex in A is adjacent to a vertex in B, must satisfy $||A|| \leq (||C||^2 - ||C||)/2$. The set C in such a partition is called a* (vertex) cutset.

In essence, the lemma bounds the size of the largest set that can be disconnected from the remaining, larger, part of the lattice by removing only a given number of vertices. It is readily apparent that choosing C, the cutset of removed vertices, to lie along a diagonal achieves the bound. The lemma follows from [27, Theorem 1].

The next theorem gives a lower bound on the required maximum connection length. Recall that the full wiring model need not be assumed. Only the Euclidean distance between the elements to be connected is considered, as a lower bound to connection length.

THEOREM 5. *Consider an $N \times N$ array with elements spaced unit distance apart. Let $K^2 = RN^2$. Then for any $0 < \delta < 1$, the probability that a $K \times K$ lattice can be connected tends to zero as*

$$O\left(N \exp\left[-\frac{2RN^{1/\delta}/3}{(-\delta \log N/\log p)^{3/2}}\right]\right)$$

unless the maximum connection length satisfies

$$d > \left(\frac{\delta \log N}{-2 \log p}\right)^{1/2} = \Omega(\sqrt{\log N}). \tag{6.1}$$

PROOF. The general idea is as follows. With high probability, there is a completely defective block of $\Theta(\sqrt{\log N}) \times \Theta(\sqrt{\log N})$ elements somewhere in the array. In fact, there are so many of these blocks distributed throughout the array that the lattice connections must enclose at least one of them. This is only possible if $d = \Omega(\sqrt{\log N})$ and, in particular, only if (6.1) is obeyed.

The proof proceeds in three steps. First we define sets of array elements called grids. Then we show that, with probability approaching one, there is a grid with all its elements defective. Finally, we assume the existence of a defective grid and use Lemma 5 to show that, if (6.1) is violated, it is not possible to connect a lattice.

Step 1. For any given N, K, p and δ, choose integer L such that

$$\frac{6}{R} \le L < \frac{6}{R} + 1.$$

Since d must clearly be at least 1, we need only consider the case when

$$m \triangleq \left\lfloor \left(\frac{\delta \log N}{-\log p}\right)^{1/2} \right\rfloor \ge 1. \tag{6.2}$$

A *grid* is defined as follows. (See Figure 13.) For each $j \in \{1, \ldots, L\}$, we define the jth *band* to be rows $j\lfloor N/L \rfloor - m + 1$ through $j\lfloor N/L \rfloor$ of the array. Consider a set of several $m \times m$ blocks of elements, positioned along the bands. In each band, the regions between the left side of the array and the leftmost block, between each pair of consecutive blocks, and between the rightmost block and the right side of the array are called *gaps*. If there is no gap wider than $2RN/(3m^2)$ elements in any band, the set of elements in the blocks is called a *grid*.

More formally, the $m \times m$ block at (x, y) is the set of elements

$$b(x, y) \triangleq \{e(i, j) : x - m + 1 \le i \le x,\ y - m + 1 \le j \le y\},$$

where $e(i, j)$ is the element in column i and row j of the array. A grid is a set of elements

$$G \triangleq \bigcup_{j=1}^{L} \bigcup_{i=1}^{n_j} b\left(x_{i,j}, j\left\lfloor \frac{N}{L} \right\rfloor\right)$$

where the $\{x_{i,j}\}$ are integers satisfying

$$m \le x_{1,j} \le \frac{2RN}{3m^2} + m,$$

$$x_{i,j} + 1 \le x_{i+1,j} \le x_{i,j} + m + \frac{2RN}{3m^2},$$

$$N - \frac{2RN}{3m^2} \le x_{n_j,j} \le N.$$

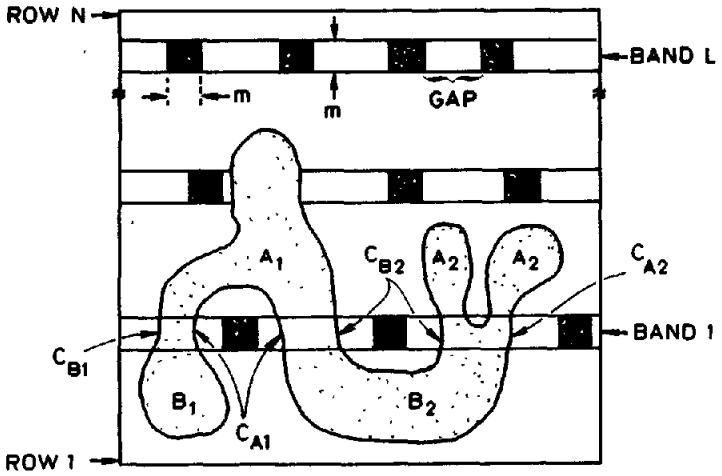

FIG. 13. The structure of a grid. The dark squares represent $m \times m$ blocks of elements arranged within the L bands. A typical gap is indicated. The blob represents the set of active elements connected around the grid.

Step 2. Consider the completely defective $m \times m$ blocks lying within the bands. The set of elements in these blocks constitutes a defective grid unless there is a gap at least $\lfloor 2RN/(3m^2) \rfloor + 1$ elements wide in any band. If there is such a gap, there must be a run of

$$\left\lfloor \frac{(\lfloor 2RN/(3m^2) \rfloor + 1) + (m-1)}{m} \right\rfloor > \frac{2RN}{3m^3}$$

adjacent nonoverlapping $m \times m$ blocks, none of which is completely defective. The probability that such a run exists, starting at any of the N columns in any of the L bands, is less than

$$\begin{aligned} LN(1-p^{m^2})^{2RN/(3m^3)} &\le LN \exp\left[\frac{-2RNp^{m^2}}{3m^3}\right] \\ &\le LN \exp\left[\frac{-2RN^{1-\delta}/3}{(-\delta \log N/\log p)^{3/2}}\right] \\ &= O\left(N \exp\left[-\frac{2RN^{1-\delta}/3}{(-\delta \log N/\log p)^{3/2}}\right]\right). \end{aligned}$$

The first inequality follows from the relation $(1-x)^y \le \exp(-xy)$, and the second from definition (6.2).

Step 3. We are given an arbitrarily large $N \times N$ array with a subset of its active elements connected into a $K \times K$ lattice. Suppose that the array contains a defective grid and that the maximum connection length d satisfies

$$d < \frac{m+1}{\sqrt{2}}. \tag{6.3}$$

We demonstrate a contradiction.

Note that a connection of length $d < m + 1$ is insufficient to cross a gap or a defective block. Furthermore, because the elements are connected in a square lattice, if any cycle of connections encloses a block of defective elements, there

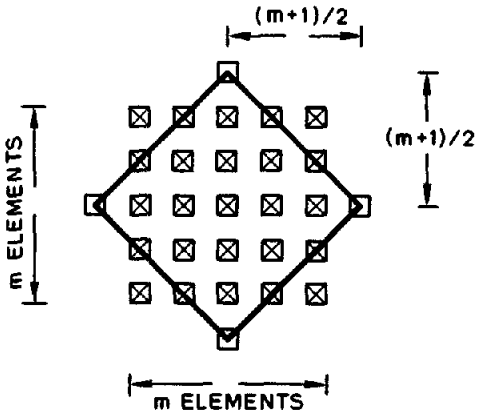

FIG. 14. The shortest connections enclosing an $m \times m$ block of defective elements. At least one connection must be of length $d \geq (m+1)/\sqrt{2}$.

must be a cycle of four connections that encloses the block.[4] However, as shown in Figure 14, enclosing a block with a cycle of four connections requires length greater than permitted by (6.3). Therefore the lattice connections cannot enclose a block but must instead extend around the blocks, as shown in Figure 13.

Since the bands are spaced at $\lfloor N/L \rfloor$-row intervals, there must be a band such that at least $K^2/2 - N\lfloor N/L \rfloor \geq K^2/3$ of the K^2 connected elements lie above it and at least this many lie below it, too. (See, e.g., band 1 in Figure 13.)

If all the elements *below* this band are removed, the set of remaining connected elements in and above the band is disconnected into some number n_A of components. For $i \in \{1, \ldots, n_A\}$, we partition the ith component into a set A_i of elements that lie strictly outside the band and a set C_{Ai} of elements within the band. Let n_B, $\{B_i\}$ and $\{C_{Bi}\}$ be defined similarly when the elements *above* the band are removed. (An example is shown in Figure 13, in which $n_A = n_B = 2$.) Note that the $\{C_{Ai}\}$ are disjoint, the $\{C_{Bi}\}$ are disjoint, and

$$\sum_{i=1}^{n_A} ||A_i|| \geq \frac{K^2}{3} \quad \text{and} \quad \sum_{i=1}^{n_B} ||B_i|| \geq \frac{K^2}{3}. \tag{6.4}$$

One of three cases applies.

Case 1. For all $i \in \{1, \ldots, n_A\}$,

$$||A_i|| \leq \frac{(2RN/(3m))^2}{2}. \tag{6.5}$$

Since no connection is long enough to cross the band, C_{Ai} constitutes a cutset disconnecting A_i from the remaining larger part of the lattice. By Lemma 5, $||A_i|| < ||C_{Ai}||^2/2$. Then

$$\begin{aligned} \sum_i ||C_{Ai}|| &> \sum_i \sqrt{2||A_i||} \\ &\geq \frac{K^2/3}{(2RN/(3m))^2/2} \frac{2RN}{(3m)} \quad \text{under constraints (6.4), (6.5)} \\ &= Nm. \end{aligned}$$

[4] *Note added in proof*: This can be shown rigorously using the concept of winding numbers [22]. Any directed cycle on the square lattice can be decomposed into directed cycles of four connections. The winding number of the original cycle or the number of times it wraps around the center of a block is the sum of the winding numbers of its constituent four-connection cycles. Thus the original cycle has a nonzero winding number only if at least one of its constituent cycles does. (This argument was suggested in [16].)

But since the $\{C_{Ai}\}$ are disjoint and all lie within the band, $\sum_i ||C_{Ai}|| \leq Nm$. This is a contradiction.

Case 2. For all $i \in \{1, \ldots, n_B\}$,

$$||B_i|| \leq \frac{(2RN/(3m))^2}{2}.$$

By symmetry with Case 1, this also implies a contradiction.

Case 3. For some i and j,

$$||A_i|| > \frac{(2RN/(3m))^2}{2} \quad \text{and} \quad ||B_j|| > \frac{(2RN/(3m))^2}{2}. \tag{6.6}$$

Since no connection is long enough to cross the band, every path connecting A_i to B_j includes an element in the band. Furthermore, the fact that defective blocks cannot be enclosed implies that there is a single gap in the band through which all such paths pass. Otherwise for some pair of gaps g_1 and g_2 there would be a path from A_i to B_j through gap g_1 but not g_2, and another path through gap g_2 but not g_1; connecting these paths through A_i, C_{Ai}, B_j, and C_{Bj} would form a cycle enclosing a defective block. Let C be the set of connected elements in this single gap. Since no gap is more than $2RN/(3m^2)$ elements wide, $||C|| \leq 2RN/(3m)$. C is a cutset disconnecting A_i from B_j, and so, by Lemma 5, either $||A_i||$ or $||B_j||$ must be less than $||C||^2/2 \leq (2RN/(3m))^2/2$. This contradicts (6.6).

Thus in every case a contradiction occurs. This completes Step 3. We conclude that the probability that the lattice can be connected goes to zero unless (6.3) is false; that is

$$d \geq \frac{m+1}{\sqrt{2}} > \sqrt{\frac{\delta \log N}{-2 \log p}},$$

which yields (6.1) and completes the proof. □

If the restriction that elements be spaced at unit distance is relaxed, the following Corollary can be demonstrated.

COROLLARY. *Consider an $N \times N$ array of rectangular elements of area A_e. Let $K^2 = RN^2$. Then for any $0 < \delta < 1$, the probability that a $K \times K$ lattice can be connected tends to zero as*

$$O\left(N \exp\left[-\frac{2RN^{1-\delta}/3}{(-\delta \log N/\log p)^2}\right]\right)$$

unless the maximum connection length satisfies

$$d > \sqrt{\frac{A_e \delta \log N}{-2 \log p}} = \Omega(\sqrt{A_e \log N}).$$

PROOF. The proof is basically the same as that of Theorem 5. One difference is that the m-row by m-column defective blocks must be replaced by m_r-row by m_c-column blocks so that each block is still approximately square in terms of physical distance. The denominator in the yield bound exponent changes because, if $m_r = \Theta(\log N)$ and $m_c = \Theta(1)$, we have $m_r^2 m_c = \Theta(\log^2 N)$ instead of $m^3 = \Theta(\log^{3/2} N)$.

The theorem and corollary readily extend to rectangular ($N_r \times N_c$) rather than square ($N \times N$) arrays, and to the connection of triangular and hexagonal lattices.

It is easily demonstrated that if no space is needed between elements to accommodate the connections, $d = O(\sqrt{\log N})$ is indeed all that is required. To connect a lattice, we simply divide the $N \times N$ array into square blocks containing $c \log N$ elements for some constant c. The Chernoff bound can be applied to show that for any $R < \bar{p}$ the probability that any particular block contains fewer than $Rc \log N$ active elements tends to zero exponentially in $c \log N$. Since there are only $N^2/c \log N$ blocks, the probability that *any* of them has fewer than $Rc \log N$ elements also tends to zero if c is properly chosen. We therefore suppose that a $\sqrt{Rc \log N} \times \sqrt{Rc \log N}$ sublattice can be connected in each block with maximum connection length $O(\sqrt{\log N})$. Each sublattice is then connected to those in the adjacent blocks to form the desired $\sqrt{R}\,N \times \sqrt{R}\,N$ lattice.

Unfortunately, wiring area *is* a consideration. Leighton and Leiserson [16] have shown that no more than $t = O(\log \log N)$ tracks are needed to connect the sublattices in the above scheme. Under the full wiring model, their method achieves $d = O(\sqrt{\log N} \log \log N)$ and $AOR = O((\log \log N)^2)$. In a forthcoming paper [10], we present a different scheme which achieves $d = O(\sqrt{\log N})$ and $AOR = O(1)$. This shows that Theorem 5 gives the best possible lower bound, to within a constant.

We conclude by noting that the bounds are increased if one assumes that the area of an element must increase linearly with the length of the longest wire it drives. Since the capacitance of a wire increases linearly with its length, the drive current needed to charge or discharge the wire in a given time, and hence the area of the driving transistor, must increase linearly as well [26]. We therefore suppose that the elements can be rectangular but must occupy area $A_e = \Omega(d)$. By the Corollary, $d = \Omega(\sqrt{A_e \log N})$. Therefore $A_e = \Omega(\log N)$, $AOR = \Omega(\log N)$ and $d = \Omega(\log N)$.

These bounds can be achieved using selectors in the following way. We make the elements of width $O(1)$ and height $O(\log N)$, and arrange them in K rows of N elements. Between each pair of adjacent rows, we place two selectors, one connected to the upper row of elements and the other connected to the lower row. The selectors share a common row of K ports positioned between them. The lattice can be connected if a chain can be formed in each of the rows of elements and if every row can be connected to the ports above and below it by the adjacent selectors. The yield can easily be shown to approach one using the arguments in Section 1 and Theorem 2.

7. *Discussion*

Previous schemes for connecting various configurations in faulty arrays include those of Manning [20], Aubusson and Catt [1], Koren [15], Fussell and Varman [7], Hedlund and Snyder [13], Lincoln Laboratory [17], and Lowry and Miller [18]. In most cases the schemes are studied empirically, but little analysis is provided. Rosenberg [24] gives a method for embedding chains, trees, pyramids, and lattices in linear arrays of faulty elements. Mangir and Avizienis [19] provide a detailed model for the variation of yield with interconnection complexity in fault tolerant circuits composed of interchangeable modules.

In this paper, we have given bounds on the maximum connection length d and the area overhead ratio AOR as a function of the array dimension N. Table I summarizes the order of growth of the bounds. Each upper bound has been demonstrated by presenting an algorithm for programming the switches to connect any fraction $R < 1 - p$ of the total number of elements. The algorithms all have

TABLE I. ORDER OF GROWTH OF BOUNDS

Problem	d	AOR	Reference
Chain in 1D array	$\Theta(\log N)$	$\Theta(1)$	—
Selector	$\Theta(\log N)$	$\Theta(\log N)$	Theorems 1, 2
Pairing	$O(1)$	$O(1)$	Theorem 3
Chain in 2D array	$O(1)$	$O(1)$	Theorem 4
Lattice in 2D array	$\Omega(\sqrt{\log N})$	—	Theorem 5
	$O(\sqrt{\log N} \log\log N)$	$O((\log\log N)^2)$	[16]
	$O(\sqrt{\log N})$	$O(1)$	[10]

running times linear in the number of array elements. A different algorithm achieving results similar to our Theorem 4 has been found independently by Leighton and Leiserson [16].

The table also lists the achievability results for the connection of a lattice obtained by Leighton and Leiserson and by the authors. The latter scheme, described in a forthcoming paper [10], attains the lower bounds for this problem and runs in linear time.

We conclude by mentioning several directions for future research.

Although experimental evidence shows that the laser-programmed interconnect is fairly reliable, it is not completely so [17]. It would therefore be useful to extend our results to accommodate defects in the switches and interconnect as well as the elements.

It is readily apparent that our algorithm for connecting a chain from a two-dimensional array can be distributed among several processors—perhaps even the elements themselves. Distributed algorithms might be found for other configurations.

The problem of dynamic fault tolerance, in which elements fail during use, is closely related to the static problem examined in this paper. Clearly, our lower bounds for static fault tolerance apply directly to the dynamic case. Our upper bounds are likely to be of interest in the dynamic case, too. The reprogrammable integrated switches mentioned in Section 1 could be used in this setting.

Finally, an alternate approach to configurable processor arrays is suggested by Fussell and Varman [7]. They propose making the task in question more accommodating as to the processor configuration on which it is run. For example, a priority queue can be implemented on an arbitrarily branched tree, which is usually more readily configured than a chain. A simplified version of the scheme in Theorem 4 can be used to connect such a tree.

ACKNOWLEDGMENTS. The authors would like to thank Rudolf Ahlswede for pointing out references related to Lemma 5, Scott Kirkpatrick and Harry Kesten for discussions on percolation, and the referees for suggesting improvements in our presentation. Discussions with Tom Leighton and Charles Leiserson are also gratefully acknowledged.

REFERENCES

1. AUBUSSON, R., AND CATT, I. Wafer-scale integration—A fault tolerant procedure. *IEEE J Solid-State Circuits SC-13*, 3 (June 1978), pp. 339–344.
2. BILARDI, G., PRACCHI, M., AND PREPARATA, F. A critique and appraisal of VLSI models of computation. In *VLSI Systems and Computation*, H.T. Kung, B. Sproull, and G. Steele, Eds. Computer Science Press, Rockville, Md., 1981, pp. 81–88.
3. BROADBENT, S.R., AND HAMMERSLEY, J.M. Percolation processes. I. *Proc Cambridge Phil. Soc 53* (1957), 629–641.

4. DOOB, J.L. *Stochastic Processes.* Wiley, New York, 1953.
5. FELLER, W. *An Introduction to Probability Theory and Its Applications.* Vol. II. 2nd ed. Wiley, New York, 1971, pp. 194–198.
6. FRISCH, H., HAMMERSLEY, J.M., AND WELSH, D. Monte Carlo estimates of percolation probabilities for various lattices. *Phys. Rev. 126,* 3 (May 1, 1962), 949–951.
7. FUSSELL, D., AND VARMAN, P. Fault tolerant wafer-scale architectures for VLSI. In *Proceedings of the 9th Annual Symposium on Computer Architecture* (Austin, Tex., Apr. 26–29). IEEE, New York, 1982, pp. 190–198.
8. GALLAGER, R.G. *Information Theory and Reliable Communication.* Wiley, New York, 1968, pp. 126–128.
9. GREENE, J.W. *Configuration of VLSI arrays in the presence of defects.* Ph.D. dissertation, Dept. of Electrical Engineering, Stanford Univ., Stanford, Calif., 1983.
10. GREENE, J.W., AND EL GAMAL, A. Configuration of VLSI arrays in the presence of defects—Part II. Submitted for publication. (See also [9]).
11. HAMMERSLEY, J.M. Bornes supérieures de la probabilité critique dans un processus de filtration. In *Le Calcul des Probabilités et ses Applications.* Centre National de la Recherche Scientifique, Paris, France, 1959, pp. 17–37.
12. HARRIS, T. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Phil Soc. 56* (1960), 13–20.
13. HEDLUND, K., AND SNYDER, L. Wafer scale integration of configurable, highly parallel (CHiP), processors (extended abstract). In *Proceedings of the International Conference on Parallel Processing* (Bellaire, Mich., Aug. 24–27). IEEE, New York, 1982, pp. 262–264.
14. HSIA, Y., CHANG, G., AND ERWIN, F. Adaptive wafer scale integration. *Jpn. J. App. Phys. 19,* Supp. 19-1 (1980), 193–202.
15. KOREN, I. A reconfigurable and fault tolerant VLSI multiprocessor array. In *Proceedings of the 8th Annual Symposium on Computer Architecture* (Minneapolis, Minn., May 12–14). IEEE, New York, 1981, pp. 425–442.
16. LEIGHTON, F.T., AND LEISERSON, C.E. Wafer scale integration of systolic arrays (extended abstract). In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science* (Chicago, Ill., Nov. 3–5). IEEE, New York, 1982, pp. 297–311.
17. LINCOLN LABORATORY. *Semiannual Technical Summaries. Restructurable VLSI Program.* Lincoln Laboratory, Lexington, Mass., March 1981, Sept. 1981, March 1982.
18. LOWRY, M., AND MILLER, A. Analysis of low-level computer vision algorithms for implementation on a VLSI processor array. In *Proceedings of the SPIE International Society of Optical Engineers (Robotics and Industrial Inspection Conference)* (San Diego, Calif., August 24–27). Vol. 360, 1982, pp. 143–150.
19. MANGIR, T., AND AVIZIENIS, A. Fault tolerant design for VLSI: Effect of interconnect requirements on yield improvement of VLSI designs. *IEEE Trans Comput C-31,* 7 (July 1982), 609–616.
20. MANNING, F. An approach to highly integrated computer-maintained cellular arrays. *IEEE Trans. Comput. C-26,* 6 (June 1977), 536–552.
21. MINATO, O., MASUHARA, T., SASAKI, T., SAKAI, Y., AND YOSHIZAKI, K. HI-CMOS II 4K static RAM. *Digest of Technical Papers, IEEE Solid State Circuits Conference.* IEEE, New York, 1981, pp. 14–15.
22. PENNISI, L. *Elements of Complex Variables,* 2nd ed. Holt, Rinehart & Winston, New York, 1976, pp. 177–178, 190.
23. RAFFEL, J. On the use of nonvolatile programmable links for restructurable VLSI. In *Proceedings of the Caltech Conference on Very Large Scale Integration* (Pasadena, Calif., Jan. 22–24). Caltech, Pasadena, Calif., 1979, pp. 95–104.
24. ROSENBERG, A. The Diogenes approach to testable fault-tolerant networks of processors. Rep. CS-1982-6.1, Computer Science Dept., Duke Univ., Durham, N.C., May 1982.
25. SMITH, R.T., CHLIPALA, J., BINDELS, J., NELSON, R., FISCHER, F., AND MANTZ, T. Laser programmable redundancy and yield improvement in a 64K DRAM. *IEEE J. Solid-State Circuits SC-16,* 5 (Oct. 1981), 506–513.
26. THOMPSON, C.D. *A complexity theory for VLSI* Ph.D. dissertation, Dept. of Computer Science, Carnegie-Mellon Univ., Pittsburgh, Pa., 1980, pp. 41–44.
27. WANG, D-L., AND WANG, P. Extremal configurations on a discrete torus and a generalization of the generalized Macaulay theorem. *SIAM J. Appl. Math. 33,* 1 (1977), 55–59.
28. WIERMAN, J.C. Percolation theory. *Ann. Prob. 10,* 3 (1982), 509–524.

RECEIVED MAY 1982; REVISED MARCH 1984; ACCEPTED FEBRUARY 1984