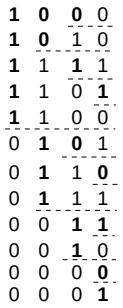

Fig. 12.1. Exposed bits are in boldface; a vector u *follows* vector v if u is below v.

from other members of B at the cost of exposing at most $2t$ additional bits in each of them. We call these vectors *good*. By (i) and (ii), at the cost of exposing at most $\log_2 m$ bits, each good vector v is already distinguished from all the vectors in A following it. On the other hand, all the vectors preceding v belong to B, and hence, v is distinguished also from them by at most $2t$ additional bits. Thus, we have at least $t^2 - t$ good vectors v and for each of them, $\mathrm{w}_A(v) \leqslant 2t + \log_2 m$. □

12.3 The isolation lemma

Let X be some set of n points, and $\mathcal{F}$ be a family of subsets of X. Let us assign a weight $w(x)$ to each point $x \in X$ and let us define the weight of a set E to be $w(E) = \sum_{x \in E} w(x)$. It may happen that several sets of $\mathcal{F}$ will have the minimal weight. If this is not the case, i.e., if $\min_{E \in \mathcal{F}} w(E)$ is achieved by a unique $E \in \mathcal{F}$, then we say that w is *isolating for* $\mathcal{F}$.

The following lemma, due to K. Mulmuley, U. Vazirani, and V. Vazirani (1987), says that – independent of what our family $\mathcal{F}$ actually is – a randomly chosen w is isolating for $\mathcal{F}$ with large probability.

Lemma 12.5. *Let $\mathcal{F}$ be a family of subsets of an n-element set X. Let $\mathbf{w} : X \to \{1, \ldots, N\}$ be a random function, each $\mathbf{w}(x)$ independently and uniformly chosen over the range. Then*

$$\mathrm{Prob}\,(\mathbf{w} \textit{ is isolating for } \mathcal{F}) \geqslant 1 - \frac{n}{N}.$$

Proof (Spencer 1995). For a point $x \in X$, set

$$\alpha(x) = \min_{E \in \mathcal{F};\, x \notin E} \mathbf{w}(E) - \min_{E \in \mathcal{F};\, x \in E} \mathbf{w}(E - \{x\}).$$

A crucial observation is that evaluation of $\alpha(x)$ does not require knowledge of $\mathbf{w}(x)$. As $\mathbf{w}(x)$ is selected uniformly from $\{1, \ldots, N\}$,

$$\mathrm{Prob}\,(\mathbf{w}(x) = \alpha(x)) \leqslant 1/N,$$

so that

$$\text{Prob}\,(\mathbf{w}(x) = \alpha(x) \text{ for some } x \in X) \leqslant n/N.$$

But if $\mathbf{w}$ had two minimal sets $A, B \in \mathcal{F}$ and $x \in A - B$, then

$$\min_{E \in \mathcal{F}; x \notin E} \mathbf{w}(E) = \mathbf{w}(B),$$

$$\min_{E \in \mathcal{F}; x \in E} \mathbf{w}(E - \{x\}) = \mathbf{w}(A) - \mathbf{w}(x),$$

so $\mathbf{w}(x) = \alpha(x)$. Thus, if $\mathbf{w}$ is *not* isolating for $\mathcal{F}$ then $\mathbf{w}(x) = \alpha(x)$ for some $x \in X$, and we have already established that the last event can happen with probability at most n/N. □

This lemma has many applications in the theory of computing. In particular, Mulmuley et al. (1987) used it to give an efficient randomized algorithm for finding a perfect matching in a graph. This result is a standard demonstration of the isolation lemma. Below we describe an application of different type: we use this lemma to show that, in the model of switching networks, counting is not weaker than nondeterminism. (Comparing the power of different modes of computation is one of the main problems in the theory of computing.)

A (switching-and-rectifier) *network* is a directed acyclic graph $G = (V, E)$ with two specified vertices $s, t \in V$, *some* of whose edges are labeled by variables x_i or their negations $\overline{x}_i$. The size of G is defined as the number of vertices. Each input $a = (a_1, \ldots, a_n) \in \{0,1\}^n$ defines a subgraph $G(a)$ of G obtained by deleting all edges whose labels are evaluated by a to 0, and removing the labels from the remaining edges. Let $|G(a)|$ denote the number of s-t paths in $G(a)$. A network G computes a boolean function in a natural way: it accepts the input a if and only if $|G(a)| > 0$. This is a *nondeterministic* mode of computation: we accept the input if and only if the labels of at least one s-t path in G are consistent with it. A *parity network* is a network with a *counting* mode of computation: we accept the input a if and only if the number of s-t paths consistent with a is odd, i.e., iff $|G(a)| = 1 \,(\text{mod}\, 2)$.

Using the isolation lemma one can show that, at the cost of a slight increase of size, every (nondeterministic) network may be simulated by a parity network.

Theorem 12.6 (Wigderson 1994)**.** *If a boolean function in n variables can be computed by a network of size L, then it can also be computed by a parity network of size at most $n \cdot L^c$, where $c \leqslant 10$.*

Proof. Given a graph $G = (V, E)$, a weight function $w : E \to \{1, \ldots, 2 \cdot |E|\}$ and an integer l, define the (unweighted, layered) version $G^l_w = (V', E')$ of G as follows. Replace every vertex $u \in V$ by $l + 1$ new vertices $u_0, u_1, \ldots, u_l$ in V' (i.e., V' consists of $l + 1$ copies of V, arranged in layers). For every edge (u, v) in E and every $0 \leqslant i \leqslant l - w(e)$ we put an edge $\left(u_i, v_{i+w(e)}\right)$

in E' (see Fig. 12.2). Let $d_w(G)$ denote the weight of the shortest s-t path in G (the weight of a path is the sum of weights of its edges; a path is *shortest* if its weight is minimal); hence, $d_w(G) \leqslant M \rightleftharpoons 2|V| \cdot |E| \leqslant |V|^3$ and $|V'| \leqslant (1+l)|V|$.

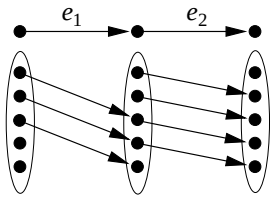

Fig. 12.2. $l = 4$, $w(e_1) = 2$ and $w(e_2) = 1$

It can be shown (do this!) that the graphs G^l_w have the following properties:

(i) if G has no s-t path, then for every w and l, G^l_w has no s_0-t_l path;

(ii) if G has an s-t path and $l = d_w(G)$, then G^l_w has an s_0-t_l path. Moreover, the later path is unique if the shortest s-t path in G is unique.

Now let $G = (V, E)$ be a network computing a given boolean function $f(x_1, \ldots, x_n)$. Say that a weight function w is *good* for an input $a \in \{0,1\}^n$ if either $G(a)$ has no s-t paths or the shortest s-t path in $G(a)$ is unique. For each input $a \in \{0,1\}^n$, taking the family $\mathcal{F}$ to be all s-t paths in the graph $G(a)$, the isolation lemma (Lemma 12.5) implies that at least one-half of all weight functions w are good for a. By a standard counting argument, there exists a set W of $|W| \leqslant \log_2(2^n) = n$ weight functions such that at least one $w \in W$ is good for every input. If w is good for a, then the graph $G^l_w(a)$ with $l = d_w(G(a))$ has the properties (i) and (ii). For different inputs a, the corresponding values of l may be different, but they all lie in the interval $1, \ldots, M$. Thus, there exist $m \leqslant n \cdot M$ networks $H_1, \ldots, H_m$ (with each $H_j = G^l_w$ for some $w \in W$ and $1 \leqslant l \leqslant M$) such that, for every input $a \in \{0,1\}^n$, the following holds:

(iii) if $|G(a)| = 0$, then $|H_j(a)| = 0$ for all j;

(iv) if $|G(a)| > 0$, then $|H_j(a)| = 1$ for at least one j.

Let s_j, t_j be the specified vertices in H_j, $j = 1, \ldots, m$. We construct the desired parity network H as follows: to each H_j add the unlabeled edge (s_j, t_j), identify t_j and s_{j+1} for every $j < m$, and add the unlabeled edge (s_1, t_m) (see Fig. 12.3).

It is easy to see that, for every input $a \in \{0,1\}^n$, $|H(a)| = 1 \pmod 2$ if and only if $|G(a)| > 0$. Indeed, if $|G(a)| = 0$, then by (iii), $H(a)$ has precisely

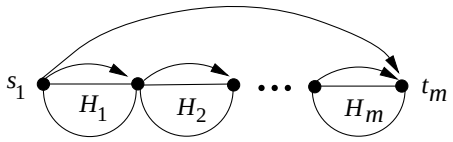

Fig. 12.3. Construction of the parity network H

two s_1-t_m paths (formed by added unlabeled edges). On the other hand, if $|G(a)| > 0$, then by (iv), at least one $H_j(a)$ has precisely one s_j-t_j path, implying that the total number of s_1-t_m paths in $H(a)$ is odd. Thus, H is a parity network computing the same boolean function f. □

For the sake of completeness, let us mention (without proof) the following interesting "parity-type" isolation lemma proved by Valiant and Vazirani (1986). View the cube $\{0,1\}^n$ as n-dimensional vector space $\mathbb{F}_2^n$, and let $\langle u, v\rangle = \sum_{i=1}^n u_i v_i \pmod 2$ denote the scalar product over $\mathbb{F}_2$.

Lemma 12.7. *Let $S \subseteq \{0,1\}^n$, $|S| \geqslant 2$. Let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be chosen independently from $\{0,1\}^n$ at random. Then, with probability at least $1/4$, there is an i such that $\langle v, \mathbf{w}_1\rangle = \ldots = \langle v, \mathbf{w}_i\rangle = 0$ for precisely one vector $v \in S$.*

12.4 Isolation in politics: the dictator paradox

One of the problems of politics involves averaging out individual preferences to reach decisions acceptable to society as a whole. In this section we will prove one isolation-type result due to Arrow (1950) which shows that, under some simple "democracy axioms" this is indeed a difficult task.

The simple process of voting can lead to surprisingly counterintuitive paradoxes. For example, if three people vote for three candidates, giving the rankings $x < y < z$, $y < z < x$, $z < x < y$, then a majority prefers y to x ($x < y$), x to z ($z < x$) but also z to y ($y < z$). In general, we have the following situation.

Suppose that $I = \{1, \ldots, n\}$ is a society consisting of a set of n individuals. These individuals are to be offered a choice among a set X of options, for example, by a referendum. We assume that each individual i has made her/his mind up about the relative worth of the options. We can describe this by a total order $<_i$ on X, for each $i \in I$, where $x <_i y$ means that the individual i prefers option y to option x. So, after a referendum we have a set $R = \{<_1, \ldots, <_n\}$ of total orders on X. A *social choice function* F takes such a set of total orders as input and comes up with a "social preference" on X, i.e., with some total order $<$ on X. Being total means, in particular, that the order $<$ is transitive: if $x < y$ and $y < z$ then $x < z$.

Given a social choice function F, a *dictator* is an individual $i_0 \in I$ such that for every referendum, the resulting social preference $<$ coincides with