[22] F. Göbel and A.A. Jagers, Random Walks on Graphs, *Stochastic Processes and their Applications*, **2** (1974), pp. 311-336.

[23] H. Halberstam and K.F. Roth, **Sequences**, Springer-Verlag (1983).

[24] W.H. Hayt and J.E. Kemmerly, **Engineering Circuit Analysis**, McGraw-Hill, 3rd ed. (1978).

[25] A. Israeli and M. Jalfon, Token management schemes and random walks yield self stabilizing mutual exclusion, *Proc. 9th Annual ACM Symposium on Principles of Distributed Computing*, Quebec City, Canada (1990), pp. 119-131.

[26] S. Janson, Poisson Approximation for Large Deviations, U.U.D.M. Report 1989:16 (Uppsala University) Oct. (1989).

[27] S. Janson, T. Luczak, and A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, *Proceedings of Random Graphs* '87 (Poźnan), to appear.

[28] J.G. Kemeny, J.L. Snell, and A.W. Knapp, **Denumerable Markov Chains**, Van Nostrand (1966).

[29] J. Spencer, **Ten Lectures on the Probabilistic method**, SIAM monograph (1986).

[30] J. Spencer and M. Klugerman, The Probabilistic Method, *Lecture Notes*, MIT, June (1990).

[31] P. Tetali and P. Winkler, Meeting Times for Random Walks on Graphs, *Technical Report*, Bellcore, October (1990).

[9] F.R.K. Chung, R.L. Graham and R.M. Wilson, Quasi-random graphs, *Combinatorica* (to appear).

[10] F.R.K. Chung and R.L. Graham, Quasi-random hypergraphs, *Random Structures and Algorithms*, **1** (1990), 105-124.

[11] F.R.K. Chung and R.L. Graham, Quasi-random set systems, preprint.

[12] F.R.K. Chung and P. Tetali, Communication Complexity and Quasi-randomness, *SIAM Journal on Discrete Mathematics*, to appear.

[13] P.G. Doyle and J.L. Snell, **Random Walks and Electrical Networks**, The Mathematical Association of America (1984).

[14] P. Erdős, On a Lemma of Littlewood and Offord, *Bulletin Amer. Math. Soc.* **51** (1945), pp. 898-902.

[15] P. Erdős, Problems and results in additive number theory, *Colloque sur la Théorie des Nombres (CBRM)* Bruxelles (1956), pp. 127-137.

[16] P. Erdős and R. Rado, Intersection theorems for systems of sets, *J. London Math. Soc.* **35** (1960), 85-90.

[17] P. Erdős and A. Rényi, Additive properties of random sequences of positive integers, *Acta arith.* **6** (1960), pp. 83-110.

[18] P. Erdős and V.T. Sós, On Ramsey-Turán type theorems for hypergraphs, *Combinatorica* **2** (1982), 289-295.

[19] P. Erdős and J. Spencer, Imbalances in k-colorations, *Networks* **1** (1971), 379-385.

[20] P. Erdős and P. Tetali, Representations of Integers as The Sum of k Terms, *J. on Random Structures and Algorithms*, **1** (1990), pp. 245-261.

[21] R.M. Foster, The Average Impedance of an Electrical Network, *Contributions to Applied Mechanics* (Reissner Anniversary Volume), Ann Arbor (1949), pp. 333-340.

Bibliography

[1] N. Alon, L. Babai, and A. Itai, A Fast and Simple Randomized Parallel Algorithm for the Maximal Independent Set Problem, *Journal of Algorithms*, **7** (1987), pp. 567-583.

[2] L. Babai, N. Nisan, and M. Szegedy, Multiparty protocols and logspace-hard pseudorandom sequences, *21st STOC* (1989), 1-11.

[3] B. Berger, The Fourth Moment Method, *2nd Annual ACM/SIAM Symposium on Discrete Algorithms*, San Francisco (1991).

[4] R.B. Boppana and J.H. Spencer, A Useful Elementary Correlation Inequality, *J. on Combinatorial Theory*, Series A, **50** (1989), pp. 305-307.

[5] G. Brightwell and P. Winkler, Maximum hitting time for random walks on graphs, *J. Random Structures and Algorithms*, **1** (1990), pp. 263-276.

[6] A.K. Chandra, M.L. Furst and R.J. Lipton, Multiparty protocols, *FOCS* (1983), 94-99.

[7] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari, The Electrical Resistance of a Graph Captures its Commute and Cover Times, *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, May (1989).

[8] F.R.K. Chung, Quasi-random classes of hypergraphs, *Random Structures and Algorithms*, to appear.

Thus $\Delta_k^{k-1}(f(m)) = \frac{1}{2^{mk}} disc_k^1(A_f)$

Conclusions and Open Problems. Note that Theorem 5.1 guarantees the *existence* of an array A such that $disc_k^i(A) \leq cn^{(2k-i)/2}$. Can we, in fact, find (effectively) or construct such an array? The question is open for $k > 2$. For $k = 2$ it is known that an $n \times n$ *Hadamard matrix* H works! That is,

$$disc_2^1(H) \leq n^{3/2}$$

However, it is not clear how to generalize the notion of Hadamard matrices for $k > 2$. Apart from being an interesting derandomization question by itself, this has the following implications. In view of Theorem 5.4, upper bounds on $disc_k^i$ yield, in turn, upper bounds on Δ_k^i. This is useful since "good" upper bounds on Δ_k^i imply "good" lower bounds on the communication complexity of multiparty protocols. Thus, making Theorem 5.1 constructive seems to be a very significant open problem.

refer to [12].)

In this section, we extend the equivalence by showing that $disc_k^i$ in the switching game is related to the "discrepancy" Δ_k^i in the context of multiparty communication complexity. Firstly, we associate with a given k–input function f, the k–dimensional array A_f of size $2^m \times \cdots \times 2^m$ where

$$A_f(i_1, \ldots, i_k) = f(x_1 = i_1, \ldots, x_k = i_k)$$

Thus we are assuming (without loss of generality) that each input x_j ranges from 1 to 2^m. Then we have

Theorem 5.4

$$\Delta_k^i(f(m)) = \frac{1}{2^{mk}} disc_k^{k-i}(A_f)$$

Proof. The intuition is going to be that the number of inputs each player knows corresponds to the number of coordinates required to specify a switch, and the possible bit sequences by the players correspond to the switch settings. We describe the proof for $i = k - 1$, and the general case is similar. By definition,

$$\Delta_k^{k-1}(f(m)) = \max_{\phi_1, \ldots, \phi_k} |E\left[f(x_1, \ldots, x_k)\phi_1(x_1, \ldots, x_k) \cdots \phi_k(x_1, \ldots, x_k)\right]|$$

where the maximum is taken over all functions $\phi_j : (\{0,1\}^m)^k \rightarrow \{0,1\}$ such that ϕ_j does not depend on x_j. And the expectation is over all possible 2^{mk} choices of $x_1, \ldots, x_k$. Thus

$$\Delta_k^{k-1}(f(m)) = \frac{1}{2^{mk}} \max_{\phi_1, \ldots, \phi_k} |\sum_{x_1} \cdots \sum_{x_k} [f(x_1, \ldots, x_k)\phi_1(x_1, \ldots, x_k) \cdots \phi_k(x_1, \ldots, x_k)]|$$

Whereas *discrepancy* of A_f in the switching game is defined as

$$disc_k^1(A_f) = \max_{s_{i_1}, \ldots, s_{i_k}} \sum_{i_1=1}^{2^m} \cdots \sum_{i_k=1}^{2^m} A(i_1, \ldots, i_k) s_{i_1} \cdots s_{i_k}$$

where the switch $s_{i_j} : \{1, \ldots, 2^m\}^k \rightarrow \{0,1\}$ depends on all but index i_j. It is now easy to see that the functions ϕ_j correspond to the switches s_{i_j}.

and Berger [3] gives an NC algorithm. We claim that our algorithms for the switching game extend to this problem, for $k > 2$, however we omit the proofs in this write-up.

5.6 Discrepancy in Other Contexts

Consider the following multiparty communication game first introduced by Chandra, Furst and Lipton [6]. There are k players, communicating via a common black board, trying to determine a Boolean function $f(x_1, \ldots, x_k)$ which accepts k inputs each m bits long. Player P_i knows all inputs except x_i, and the aim of the game is to minimize the number of bits $C_k^{k-1}(f)$ to be exchanged in order to compute f. More recently, Babai, Nisan and Szegedy showed [2] that the communication complexity in question is $C_k^{k-1}(f) \geq \log \frac{1}{\Delta_k^{k-1}(f(m))}$
where $\Delta_k^{k-1}(f(m))$ is "discrepancy" in the following precise sense:

$$\Delta_k^{k-1}(f(m)) = \max_{\phi_1,\ldots,\phi_k} |E\left[f(x_1, \ldots, x_k)\phi_1(x_1, \ldots, x_k) \cdots \phi_k(x_1, \ldots, x_k)\right]|$$

where the expectation is over all possible 2^{mk} choices of $x_1, \ldots, x_k$. And the maximum is taken over all functions $\phi_j : (\{0,1\}^m)^k \to \{0,1\}$ such that ϕ_j does not depend on x_j. (Intuitively, ϕ_j corresponds to possible messages communicated by player P_i.) Thus the "discrepancy" $\Delta_k^{k-1}(f(m))$ is of importance to give "good" lower bounds for the multiparty communication complexity $C_k^{k-1}(f(m))$.

In a recent work with Fan Chung [12], we generalized the multiparty communication game as follows. Given a k–input function f as before, let us assume each player knows i inputs, where i could be $1, \ldots, k-1$. (There could be $\binom{k}{i}$ potential players depending on which subset (of size i) inputs a player knows.) By extending the notion of "discrepancy" in a natural way to $\Delta_k^i(f(m))$, we were able to show the communication complexity of the new game to be $C_k^i(f) \geq \log \frac{1}{\Delta_k^i(f(m))}$. The motivation for the generalization was to demonstrate the relation between communication complexity and several hypergraph properties, categorized as *quasi-random* [10]. The connecting link was proving that *discrepancy* in the hypergraph context is the same as $\Delta_k^i(f(m))$. (For the details, please

$$
\begin{aligned}
&= \sum_i \Pr\left[Y_i = 1\right] \\
&= \sum_i \Pr\left[\mid \sum_j R_{ij} y_j \mid > \ c_4 n \right] \\
&= \sum_i \Pr\left[\sum_j \ \mid \frac{R_{ij}}{\sqrt{n}} y_j \mid > \ c_4 \sqrt{n} \right] \\
&> \sum_i c_3 \qquad\qquad \text{using Lemma 5.2} \\
&= c_3 n
\end{aligned}
$$

Once again we can find the desired setting for the y_j's, in parallel, so that at least $c_3 n$ plane sums are greater than $c_4 n$.

Step 3. It is easy to set the x_i's in parallel so that $P_i x_i = \mid P_i \mid$.

This completes the description of the parallel algorithm for the 3-dimensional case. As before, we briefly remark that the general case can be done analogously.

The (Hyper) Edge–Discrepancy Problem

In k–dimensions the *hyperedge–discrepancy* problem is stated as follows.

Consider the (complete) k–graph on a set V of n vertices defined by the k–edges $V^k = \{W : W \subseteq V, \mid W \mid = k\}$. Given a (red-blue) coloring, $\chi : V^k \rightarrow \{-1, +1\}$, of the k–edges define the *discrepancy* $\chi(W)$ of a subset W of the vertices to be the difference between the number of k–edges (induced) colored red and colored blue. We want to find a subset of the vertices that has the maximum such discrepancy. In general, we want to find

$$Disc_k(n) = \min_{\chi} \max_{W \subseteq V} \mid \chi(W) \mid$$

Erdős and Spencer showed [19] that $Disc_k(n) = O(n^{(k+1)/2})$. The proof is existential. As before, it remains open if the proof for the upper bound can be made constructive for $k > 2$. Thus it remains open if, for $k > 2$, there is an *explicit* edge–coloring of a k–graph such that $\chi(W) = O(n^{(k+1)/2})$, for every $W \subseteq V$. However, the lower bound that $Disc_k(n) = \Omega(n^{(k+1)/2})$ *is* constructive as follows. For the case $k = 2$, Spencer [29] gives a deterministic sequential algorithm by reducing it to the switching game problem

We now describe the parallel version for the 3-dimensional case.

Step 1. We choose the settings for the z–switches from a 4-wise independent distribution. Consider the random varibale for each row sum : $R_{ij} = \sum_k a_{ijk} z_k$.
By the *fourth moment lemma* (Lemma 5.2), we have

$$\Pr\left[\mid R_{ij} \mid > c_2\sqrt{n}\right] > c_1$$

or equivalently, in terms of the indicator random variable Z_{ij},

$$\Pr\left[Z_{ij} = 1\right] > c_1$$

Thus

$$\begin{aligned} E\left[Z\right] &= E\left[\sum_i \sum_j Z_{ij}\right] \\ &= \sum_i \sum_j E\left[Z_{ij}\right] \\ &= \sum_i \sum_j \Pr\left[Z_{ij} = 1\right] \\ &> c_1 n^2 \end{aligned}$$

If $E[Z] > c_1 n^2$ over all possible choices of z_k's from the 4-wise independent distribution, then there has to be at least one set of z_k's such that $Z > c_1 n^2$. This would mean there is at least one way of setting the z–switches such that at least $c_1 n^2$ row sums are greater than $c_2\sqrt{n}$. Now we appeal to the **FACT** in claiming that we can, in fact, find the desired setting by searching through the probability space of size $\Theta(n^2)$ in poly log time with polynomial number of processors.

Step 2. This step is very similar to Step 1. The y–switches are chosen 4-wise independently. And we deal with the plane sum variables $P_i = \sum_j R_{ij} y_j$.

$$E\left[Y\right] = E\left[\sum_i Y_i\right] = \sum_i E\left[Y_i\right]$$

For higher dimensions, we adopt a very similar strategy. The outline of the algorithm is as follows.

Derandomization in the General Case.

Given: An initial configuration $A(i_1, \ldots, i_k) = \pm 1$

Aim: To achieve $disc_k^{k-1}(A) \geq cn^{(k+1)/2}$, by setting the switches $s_{i_1}, \ldots, s_{i_k}$ to ± 1.

Step 1.	Pick $s_{i_k} = \pm 1$ so that at least $c_1 n^{k-1}$ *row sums* $\mid R[i_1, \ldots, i_{k-1}] \mid = \mid \sum_{i_k=1}^{n} A(i_1, \ldots, i_k) s_{i_k} \mid > c_2\sqrt{n}$
Step 2.	Pick $s_{i_{k-1}} = \pm 1$ so that at least $c_3 n^{k-2}$ *plane sums* $\mid \sum_{i_{k-1}=1}^{n} R[i_1, \ldots, i_{k-1}] s_{i_{k-1}} \mid > c_4 n$
....	. . . . .
Step k–1.	Pick $s_{i_2} = \pm 1$ so that at least $c'n$ $(k-1)$–dimensional *hyperplane sums* $\mid P_{i_1} \mid > c'' n^{(k-1)/2}$
Step k.	Pick $s_{i_1} = \pm 1$ so that $P_{i_1} s_{i_1} = \mid P_{i_1} \mid$.

Once again the last step of the algorithm is straight forward. We need to justify, however, that the conditional expectations that arise in a general Step J (setting s_{i_j}'s sequentially) can be effectively computed. In view of the fact that the j–dimensional *hyperplane sums* can only range from $-n^j$ to n^j, we can use a scheme similar to the one outlined in the 3–dimensional case.

The NC Algorithm

For the fourth moment lemma to be useful we further need the following FACT [1].

FACT. It is possible to construct, in polynomial time, a probability space Ω such that

$$\Omega = \{(X_1, \ldots, X_n)\} \subset \{-1, 1\}^n$$
$$\Omega \text{ is of size } \Theta(n^2)$$
$$\Pr[X_i = -1] = \Pr[X_i = +1] = \tfrac{1}{2}$$
$$X_i : \text{ 4-wise independent.}$$

where X ranges over all possible values a plane sum can take.

The important distinction to be made here is that the R_{ij}'s are not ± 1 any more. This is in view of the fact that the z–switches have already been set. Thus we couldn't apriori write down a simple formula for this probability as we could in Step 1. However, we salvage the situation by noting that we do have bounds on the range of integer values R_{ij}'s can take. For example, any row sum can range from $-n$ to n and a plane sum can range from $-n^2$ to n^2. Thus X can only be an integer between $-n^2$ to n^2. We compute the conditional probabilities using the following general scheme.

> Let $\phi_1, \ldots, \phi_n$ be fixed integers $\in [-n^2, n^2]$. And let $y_1, \ldots, y_n \in \{-1, +1\}$.
> Further, given X, we define functions
> $f_j(X) =$ number of solutions to $\quad y_j\phi_j + \cdots + y_n\phi_n = X$.
> The functions, $f_1, \ldots, f_n$, can be effectively calculated as follows:
> Initialization : $f_n(\phi_n) = f_n(-\phi_n) = 1$; and $f_n(\text{others}) = 0$
> Reverse Induction : $f_j(X) = f_{j+1}(X + \phi_j) + f_{j+1}(X - \phi_j)$.
> Now suppose y_1 through y_{l-1} have been set to -1 or $+1$; and y_l through y_n can be -1 or $+1$ with equal probability. Then
>
> $$\Pr\left[\sum_j y_j\phi_j = X \mid y_1, \ldots, y_{l-1}\right]$$
>
> $$\begin{aligned} &= \Pr\left[s_{l-1} + y_l\phi_l + \cdots + y_n\phi_n = X\right] \\ &= \Pr\left[y_l\phi_l + \cdots + y_n\phi_n = X - s_{l-1}\right] \\ &= f_l(X - s_{l-1}).2^{-(n-l+1)} \end{aligned}$$
>
> where $s_{l-1} = y_1\phi_1 + \cdots + y_{l-1}\phi_{l-1}$ is the fixed part.

Thus the conditional probabilities can be effectively calculated, allowing us to sequentially set the values for the y–switches. Once again it is not hard to see that Step 2 takes $O(n^3)$ time as well. Thus the whole algorithm takes $O(n^3)$ time. Note that this is optimal considering that the given input array is of size n^3.

This completes the 3–dimensional case.

Clearly, the factors in the above expressions can be efficiently computed. Thus beginning with z_1 we set the z's sequentially, each time choosing the values so that the conditional expectation is maximized. By Theorem 5.2, we are guaranteed to make $c_1 n^2$ row sums $> c_2\sqrt{n}$, at the end of this step. A careful analysis would show that this step takes $O(n^3)$ time.

Step 2. The situation here is more subtle. We want to set $y_l = \pm 1$ to maximize the number of *plane sums* that are $> c_4 n$. For each plane sum P_i, let

$$Y_i = \begin{cases} 1 & \text{if } \mid P_i \mid > c_4 n \\ 0 & \text{otherwise} \end{cases}$$

Once again, we want to maximize the expected number of such planes. Initially, before any switches are set, we know by Claim 2 that $E\left[\sum_i Y_i\right] = \Omega(n)$. Thus in a typical step we set $y_l = \pm 1$, maximizing the quantity

$$E\left[\sum_i Y_i \mid y_1, \ldots, y_l\right]$$

$$\begin{aligned} &= \sum_i E\left[Y_i \mid y_1, \ldots, y_l\right] \\ &= \sum_i \Pr\left[\mid P_i \mid > c_4 n \mid y_1, \ldots, y_l\right] \\ &= \sum_i \Pr\left[\mid \sum_j R_{ij} y_j \mid > c_4 n \mid y_1, \ldots, y_l\right] \\ &= \sum_i \Pr\left[\mid r_i + \sum_{j=l+1}^{n} R_{ij} y_j \mid > c_4 n\right] \end{aligned}$$

where r_i is the partial plane sum fixed by the choice of settings for the switches $y_1, \ldots, y_l$, and the (conditional) probability is over the space of ± 1 equally likely choices for $y_{l+1}, \ldots, y_n$. We can argue independently for each plane P_i, owing to the linearity of expectation. Thus consider i fixed. We need to be able to compute

$$\Pr\left[r_i + \sum_{j=l+1}^{n} R_{ij} y_j = X\right]$$

The Computation. In the above algorithm, Step 3 is straight forward and easy, and takes time $O(n^3)$. But we have to justify that Steps 1 and 2 can in fact be done deterministically (in time $O(n^3)$ as well). The hard part is in showing the appropriate conditional expectations can be computed efficiently:

Conditional Expectations.

Step 1. Let us associate the following *indicator random variable* with each ij-th row:

$$Z_{ij} = \begin{cases} 1 & \text{if } \mid R_{ij} \mid > c_2\sqrt{n} \\ 0 & \text{otherwise} \end{cases}$$

We want to maximize the expected number of rows with $Z_{ij} = 1$. When the z_k $(k = 1, \ldots, n)$ are randomly set, by Claim 1 we have $E\left[\sum_i \sum_j Z_{ij}\right] = \Omega(n^2)$. So starting with z_1, we want to set $z_k = \pm 1$ so as to maximize the number of *row sums* that are $> c_2\sqrt{n}$; i.e. in a typical step, we want to set $z_l = \pm 1$ so as to maximize the conditional expectation

$$E\left[\sum_i \sum_j Z_{ij} | z_1, \ldots, z_l\right]$$

$$\begin{aligned} &= \sum_i \sum_j E\left[Z_{ij} | z_1, \ldots, z_l\right] \\ &= \sum_i \sum_j \Pr\left[\mid R_{ij} \mid > c_2\sqrt{n} | z_1, \ldots, z_l\right] \\ &= \sum_i \sum_j \Pr\left[\mid r_{ij} + S_{n-l} \mid > c_2\sqrt{n}\right] \end{aligned}$$

where $r_{ij} = \sum_{k=1}^{l} a_{ijk} z_k$ is already fixed by the choices of $z_1, \ldots, z_l$, and $\sum_{k=l+1}^{n} a_{ijk} z_k = S_{n-l}$ is a well known random variable! Thus the conditional probability to be computed for each row is

$$\Pr\left[\mid R_{ij} \mid > c_2\sqrt{n} | z_1, \ldots, z_l\right] = \sum_{t \in T} \binom{n-l}{t} 2^{-(n-l)}$$

where

$$T = \left\{t : \mid r_{ij} + 2t - (n-l) \mid > c_2\sqrt{n}\right\}$$

Thus

$$\begin{aligned} E[Y] &= E[\sum_i Y_{ij}] \\ &= \sum_i E[Y_i] \\ &= \sum_i \Pr[Y_i] \\ &> c_3 n \end{aligned}$$

This implies that there are at least $c_3 n$ planes with $\mid P_i \mid > n$. This completes the proof of Claim 2.

Claim 3 is obvious. We merely go through each x_i setting to ± 1 so that $x_i P_i = \mid P_i \mid$.

This shows that there exist $\Omega(n)$ plane sums, each $\Omega(n)$ in value, thus yielding a discrepancy of $\Omega(n^2)$. This proves that there exists a way of setting x_i, y_j, and z_k such that the $disc_3^2 > cn^2$.

Note that the existential proof suggests itself a randomized algorithm. We now provide a deterministic algorithm, in the following section, with the details of derandomization.

5.5 Derandomization

The Sequential Algorithm

We describe the 3–dimensional case in detail, and give an outline of the more general case.

Given: Initial Configuration $A = \{a_{ijk}\} = \pm 1$

Aim: To achieve $disc_3^2(A) \geq cn^2$

Step 1.	Pick $z_k = \pm 1$ so that at least $c_1 n^2$ *row sums* $\mid R_{ij} \mid = \mid \sum_{k=1}^{n} a_{ijk} z_k \mid > c_2 \sqrt{n}$
Step 2.	Pick $y_j = \pm 1$ so that at least $c_3 n$ *plane sums* $\mid P_i \mid = \mid \sum_{j=1}^{n} R_{ij} y_j \mid \; > c_4 n$
Step 3.	Pick $x_i = \pm 1$ so that $P_i x_i = \mid P_i \mid$

Proof of Claim 1. Choose $z_k = \pm 1$ randomly, independent with equal probability. Let Z_{ij} denote the event that the row sum $\mid R_{ij} \mid > \sqrt{n}$. Further let the random variable Z denote the number of Z_{ij} which occur. By Corollary 5.1 (plugging $c = k = 1$), we have $\Pr[Z_{ij}] > c_1$, where $c_1 > 0$. Thus

$$\begin{aligned} E[Z] &= E[\sum_i \sum_j Z_{ij}] \\ &= \sum_i \sum_j E[Z_{ij}] \\ &= \sum_i \sum_j \Pr[Z_{ij}] \\ &> c_1 n^2 \end{aligned}$$

This implies that there **exists** a choice of settings for z_k's so that $Z > c_1 n^2$, meaning at least $c_1 n^2$ row sums $> \sqrt{n}$. This completes the proof of Claim 1.

Proof of Claim $1\frac{1}{2}$. We want to show that there exists a $c_2 > 0$ such that there exist at least $c_2 n$ planes each of which has at least $c_2 n$ row sums bigger than $\sqrt{n}$. Let p be the fraction of planes P_i for which $\mid \{j :\mid R_{ij} \mid > \sqrt{n}\} \mid \geq c_1 n/2$. Then the number of "good" pairs $\{i,j\}$ with $\mid R_{ij} \mid > \sqrt{n}$ is $\leq [pn]n + [(1-p)n]c_1 n/2 = (1-c_1)pn^2 + c_1 n^2/2 < pn^2 + c_1 n^2/2$

(This is because each of the pn planes might contribute up to n "good" pairs, and the rest ($[1-p]n$) of them can contribute no more than $c_1 n/2$ pairs.)

But by Claim 1, the number of "good" pairs is at least $c_1 n^2$. This yields that $p \geq c_1/2$. Thus we can choose $c_2 = c_1/2$ to complete the proof of Claim $1\frac{1}{2}$.

Proof of Claim 2. Once again choose $y_j = \pm 1$ at random. Let Y_i denote the event that $\mid P_i \mid > n$ and let Y denote the number of Y_i which occur. By Corollary 1 (plugging $c = c_2$, $k = 1$, and the real numbers being $R_{i1}/\sqrt{n}, \ldots, R_{in}/\sqrt{n}$) we have

$$\Pr[Y_i] = \Pr\left[\mid \sum_j \frac{R_{ij}}{\sqrt{n}} y_j \mid > \sqrt{n}\right] > c_3$$

For the parallel algorithm we need an analog of Corollary 1 under the weaker assumption of *limited independence*. Thus the following *fourth moment lemma* (due to B. Berger, [3] Theorem 3.6) is going to be useful for us to set the switches in parallel.

The Fourth Moment Lemma.
Let $\epsilon_1, \ldots, \epsilon_n$ be 4-wise independent random variables, each being 0 or 1 with some probability. i.e.

$$\Pr[\epsilon_i = a, \epsilon_j = b, \epsilon_k = c, \epsilon_l = d] = \Pr[\epsilon_i = a] \Pr[\epsilon_j = b] \Pr[\epsilon_k = c] \Pr[\epsilon_l = d]$$

for $1 \leq i < j < k < l \leq n$.

Lemma 5.2 *(B. Berger) For all $c, \alpha > 0$, with $\epsilon_i = \pm 1$ 4-wise independent, if $\mid \{i : \mid a_i \mid \geq 1\} \mid \geq cn$ then*

$$\Pr[\mid S \mid \geq \alpha\sqrt{cn}] \quad \geq \quad \frac{(1-\alpha^2)^2}{(1-\alpha^2)^2 + 2}$$

5.4 The Randomized Algorithm

We provide the proof of the 3–dimensional case in detail and the general case can be handled analogously.

Proof of $disc_3^2 > cn^2$.

Given an initial configuration $a_{ijk} = \pm 1$, we want to show that there *exists* a way to set the switches x_i, y_j, and z_k such that $\sum_{i,j,k=1}^{n} a_{ijk} x_i y_j z_k > cn^2$. The proof is in four steps :

Claim 1. There exists a setting of the $z_k = \pm 1$ so that $\Omega(n^2)$ *row sums* $R_{ij} = \sum_{k=1}^{n} a_{ijk} z_k = \Omega(\sqrt{n})$

Claim 1$\frac{1}{2}$. Once the z_k have been set according to Claim 1, there exist $\Omega(n)$ planes, each of which contains $\Omega(n)$ rows with row sums $= \Omega(\sqrt{n})$

Claim 2. There exists a setting of the $y_j = \pm 1$ so that $\Omega(n)$ of the *plane sums* $P_i = \sum_{j=1}^{n} R_{ij} y_j = \Omega(n)$

Claim 3. There exists a setting of the $x_i = \pm 1$ so that $P_i x_i = \mid P_i \mid$

Proof of Lemma. Let $disc^i_{k-1} = L$. A k–dimensional array of size n^k can be considered as n "slices" of $(k-1)$–dimensional arrays, each of size n^{k-1}. Thus in k–dimensions consider the problem of achieving a discrepancy of $disc^i_k$ on an *arbitrary* initial configuration with the help of $(i+1)n^{k-i}$ switches, each of which controls n^i lights. This means each $(k-1)$–dimensional "slice" has $(i+1)n^{k-1-i}$ switches, each of which is still in control of n^i lights. This is precisely the situation for $disc^i_{k-1}$. Assume we can solve the $k-1$–dimensional problem. Then the n independent "slices" provide a discrepancy of $n \times L = nL$. Thus

$$disc^i_k \geq nL = n \times disc^i_{k-1}$$

This proves that $disc^{k-1}_k$ is the only interesting case, the other cases follow inductively. We prove the probability lemmas first in the following section before setting out on the proof of Theorem 5.2.

5.3 The Probability Lemmas

The following is an old theorem due to Paul Erdős [14].

Theorem 5.3 *(Erdős) Let $I = [a,b)$ be an interval of length $b-a = 2t$. Let $x_1, \ldots, x_n$ be real numbers such that $\mid x_i \mid \geq 1$. Then the number of choices of $\pm$ signs such that*

$$\pm x_1 \pm x_2 \pm \cdots \pm x_n \ \in \ I$$

is less than or equal to the sum of the t largest binomial coefficients $\binom{n}{i}$.

Note that the total number of choices of signs is 2^n. Thus we have a good proportion of the choices for which the quantity of interest falls outside the interval I. For example, for the case $I = [-1,1)$, the theorem guarantees that, the number of choices of signs for which the sum falls in I is $\leq \binom{n}{n/2}$. The following not-so-hard-to-prove corollary of the theorem is especially useful for our purposes.

Corollary 5.1 *For all $c, k > 0$ there exists a $c' > 0$ such that if $x_1, \ldots, x_n$ satisfy $\mid \{i : \mid x_i \mid \geq 1\} \mid \geq cn$ and $\epsilon_i = \pm 1$, mutually independently then*

$$\Pr\left[\mid S = \epsilon_1 x_1 + \cdots + \epsilon_n x_n \mid > k\sqrt{n}\right] > c'.$$

The Upper Bound (the easy part).

Theorem 5.1 $\exists$ *array* A *s.t.* $\forall$ *settings* $disc_k^i(A) \leq cn^{(2k-i)/2}$

Proof. The *random* array works! The proof is straight forward using the probabilistic method of Paul Erdős. Consider a random array where, for each light,

$$\Pr[light = ON] = \Pr[light = OFF] = \frac{1}{2}$$

Then $disc_k^i = S_{n^k}$, the sum of n^k i.i.d.'s.
Thus the probability that a particular setting of switches is "BAD"

$$= \Pr[S_{n^k} > \lambda] < e^{-\lambda^2/2n^k}$$

The probability that all the settings are "BAD"

$$2^{(i+1)n^{k-i}}.e^{-\lambda^2/2n^k} \quad = 1,$$

if $\lambda = [2\log 2(i+1)]^{1/2} n^{(k+k-i)/2}$.

This implies there *exists* a "GOOD" array! □

Remark. For $k = 2$ it is known that an $n \times n$ *Hadamard matrix* H works —- $disc_2^1(H) \leq n^{3/2}$. But for $k > 2$, it remains open, if there is an *explicit* array which establishes Theorem 5.1.

The Lower Bound. One of the main results of the chapter is the following lower bound (the hard part).

Theorem 5.2 $\forall$ *arrays* A $\exists$ *setting s.t.* $disc_k^{k-1}(A) \geq c'n^{(k+1)/2}$

Notice that the theorem considers the special case $i = k-1$. However we show that, assuming Theorem 5.2, the other cases can be handled as well. The following lemma establishes the connection between the special case and the other cases.

Lemma 5.1 *For* $i = 1, \ldots, k-2$, $\quad disc_k^i \geq n \times disc_{k-1}^i$

dimensional hyperplane of n^i lights; i.e. each switch when thrown changes each light in the particular hyperplane from *off* to *on* or from *on* to *off*. There are $(i+1)n^{k-i}$ such switches and the aim is to maximize the difference between the number of lights *on* and *off*. We denote this by $disc_k^i$.

Thus in 3-dimensions we have two problems: $disc_3^2$ and $disc_3^1$. The distinction is that each switch controls a plane of lights in one case and a line of lights in the other. Intuitively, we would expect $disc_3^1$ to be higher than $disc_3^2$, and the intuition is right. The mathematical formulation of this case ($disc_3^2$) is as follows.

Let the array of n^3 lights be represented by $A(i,j,k) = \pm 1, i,j,k = 1,\ldots,n$. Thus $+1$ represents a light *on* and -1 a light *off*. Further we have x_i, y_j, z_k representing the $3n$ switches. "Throwing" a switch x_i corresponds to setting $x_i = -1$. Given an initial setting of $A(i,j,k) = \pm 1$ we define the *discrepancy* of A to be

$$disc(A) = \max_{x_i,y_j,z_k=\pm 1} A(i,j,k) \cdot x_i y_j z_k$$

i.e. the maximum difference between the number of lights *on* and *off* that one can obtain by throwing the switches. Further we define

$$disc_3^2 = \min_A \max_{x_i y_j z_k} A(i,j,k) x_i y_j z_k$$

to be the maximum possible regardless of the initial configuration for A. The general case $disc_k^i$ will have a similar mathematical formulation.

The Solution

The aim of the following section is to prove that $disc_k^i = \Theta(n^{(2k-i)/2})$. (We shall see in the following that the case $i = k-1$ is of special interest.)

What does $disc_k^i = \Theta(n^{(2k-i)/2})$ mean? In general, $disc_k^i \leq L$ means there exists a k–dimensional array A such that no matter how you throw the switches, the difference between the number of lights *on* and *off* is less than or equal to L. On the other hand, $disc_k^i \geq L$ means there exists a way of throwing the switches so that, no matter what the array A is, the difference between the number of lights *on* and *off* is at least L.

can obtain, given any initial configuration for the lights. A closely related problem is the *(Hyper) Edge-discrepancy* problem [19,29]. We make use of the proof techniques from [19] to show that *Discrepancy* = $\Theta\left(n^{(k+1)/2}\right)$. This result, however, is *existential* due to its critical use of *randomness*. We further provide an optimal deterministic sequential and an NC algorithm in this chapter. The basic idea is to use Raghavan–Spencer's method of conditional expectation to remove randomness. But the crucial part of the derandomization is in justifying the efficient computation of the conditional expectations. For the NC algorithm, we prove that Berger's *fourth moment method* works – 4-wise independent random setting for the switches works! Our algorithms also extend to the *(Hyper) Edge-discrepancy* problem.

To mention the previous work on the problems, Spencer [29] describes a polynomial time sequential algorithm for the 2–dimensional versions of the switching game and the edge-discrepancy problem. More recently Berger [3] has shown that by using 4-wise independence, a variety of discrepancy problems (including the 2-d switching game and the edge discrepancy) can be solved in NC. We prove that 4-wise independent random setting for the switches works in the higher dimensional cases as well.

Discrepancy naturally comes up in various other contexts. A notion of discrepancy was used in [8] to capture *quasi-randomness* properties in hypergraphs. Babai, Nisan, and Szegedy [2] used *discrepancy* to lower bound the communication complexity of certain multiparty protocols. In a recent paper [12], we showed that discrepancy in the context of *quasi-random classes* of hypergraphs is the same as the discrepancy in the context of the communication complexity in multiparty protocols. In this chapter (Section 7), we further relate discrepancy in the switching game to that in the context of the multiparty protocols.

5.2 The Generalization of the Switching Game

The Problem

Consider a k-dimensional array of n^k lights. Imagine each switch controlling an i-

Chapter 5

The Gale-Berlekamp Switching Game

5.1 Introduction

The history of the following *lights and switches* game (in 2–dimensions) goes back to the corner of a third-floor office at AT & T Bell Laboratories. Designed by Elwyn Berlekamp, the game consists of an 8×8 array of lights. There is a switch for each row and for each column which when thrown changes each light in its line from *off* to *on* or from *on* to *off*. Given an arbitrary initial configuration (of off's and on's) for the lights, the aim of the game is to turn as many lights *on* as possible. The 2–dimensional version is described in detail in [29]. In this paper we consider the following generalization of the game.

In k-dimensions, we formulate $k-1$ discrepancy problems associated with the switching game. However we show that there is only one non-trivial problem in each dimension, since the others follow "inductively" by solving the lower dimensional problems. In the non-trivial case, the game consists of an $n \times \cdots \times n$ array A (of size n^k) of lights. There are kn switches, each controlling a hyperplane of n^{k-1} lights. i.e. each switch depends only on one dimension of the array and we have k types for k dimensions. Each switch when thrown changes each light in its control from *on* to *off* or from *off* to *on*. *Discrepancy* is defined as the maximum difference between the number of lights on and off that one

There still remain the following open problems:
What is the adversary's optimal strategy on general graphs? Is there an efficient algorithmic way of deciding the meeting times between pairs of vertices in general graphs? What is the maximum meeting time on an n-vertex graph?

$$
\begin{aligned}
&= 1 + \frac{1}{d(x)} \sum_{z \in N(x)} \phi(z,y) + \beta \\
&> 1 + \frac{1}{d(x)} \sum_{z \in N(x)} M(z,y) \\
&= M(x,y)
\end{aligned}
$$

The inequality is strict , since at least one of the neighbors of x is closer to y than x is. The contradiction proves the theorem.

Corollary 4.5

$$M(x,y) \leq \frac{8}{27} n^3$$

Proof. Brightwell and Winkler [5] have shown that the "lollipop" graph (a clique on $2n/3$ vertices with a path on the remaining $n/3$ vertices attached at one end) is the unique extremal graph that has the maximum hitting time; namely, when the start vertex x is in the clique, and the end vertex y is at the far end of the path, the hitting time turns out to be $\frac{4}{27} n^3 + o(n^3)$. Thus, by the previous theorem, $M(x,y) \leq \frac{8}{27} n^3$.

Corollary 4.6 *For t a sink,* $\quad M(x,t) = H(x,t)$.

Proof. Immediate from the theorem, and since $M(x,y) \geq H(x,y)$.

Remarks and **Open Questions**.

Note that we tacitly assumed that the adversary's choice at each turn depends only on the current locations of the tokens (in particular, independent of the past history of moves). However, it is shown in [31] that this causes no loss of generality; i.e. the adversary always has an optimal strategy which is equivalent to a *tournament* on the vertices of the given graph. The general case of $k > 2$ tokens is also considered in [31]. The rule here being that whenever two tokens meet one is eliminated. It is shown that the expected number of steps before tokens on k of the n vertices reduce to one is at most $\frac{8}{27} k n^3$.

Here m is the number of edges in the graph G, and R denotes the effective resistance.

Lemma 4.2 *Every graph has a sink.*

Proof. Let us define the relation "$\leq$" on the vertices of a given graph G as follows. We say $x \leq y$ iff $H(x,y) \leq H(y,x)$. That this relation is *transitive* is easy to see from the first part of the previous lemma. Thus we have a preorder on the vertices of G, and thus there is a minimal vertex t such that $H(t,x) \leq H(x,t)$.

Theorem 4.8 *Let t be a sink. Then for any pair of vertices x and y in G,*

$$M(x,y) \leq H(x,y) + H(y,t) - H(t,y).$$

Proof. We use a potential function argument. Define

$$\begin{aligned} \phi(x,y) &= H(x,y) + H(y,t) - H(t,y) \\ &= H(y,x) + H(x,t) - H(t,x) \\ &= \phi(y,x) \end{aligned}$$

Thus ϕ is symmetric. And no matter which token the adversary decides to move, the value of ϕ goes down by 1:

$$\begin{aligned} \phi(x,y) &= 1 + \frac{1}{d(x)} \sum_{z \in N(x)} \phi(z,y) && \text{if token on } x \text{ is moved} \\ &= 1 + \frac{1}{d(y)} \sum_{z \in N(y)} \phi(x,z) && \text{if token on } y \text{ is moved} \end{aligned}$$

Moreover, t is a *sink* implies $H(y,t) - H(t,y)$ is always non-negative, and thus $\phi \geq 0$.

Now we claim that $M(x,y) \leq \phi(x,y)$. Suppose not. Consider $\beta = max_{x,y}\,[M(x,y) - \phi(x,y)]$, of all pairs realizing β choose a pair (x,y) of minimum distance. Let us assume the adversary moves the token on x. Then

$$M(x,y) = \phi(x,y) + \beta$$

On solving, we get $M(1,2) = M(2,3) = M(3,4) = 5$,
$M(1,3) = M(2,4) = 8$, and $M(1,4) = 9 = H(1,4)$.

Perhaps the most interesting question is what the adversary's strategy should be on general graphs. While we don't know the answer for this, we do show that the meeting time in general is at most twice the maximum hitting time between any two vertices. In [31], examples are given showing the *hitting time strategy* is not necessarily the adversary's optimal one, and also showing that $M(x,y)$ is not necessarily less than or equal to the $max_{u,v}\ H(u,v)$.

We define a vertex t to be a *sink*, if $H(t,y) \leq H(y,t), \quad \forall y \in G$. So it is at least as hard to get to a *sink* as it is the other way. For example, in a chain, the end-vertices would be *sinks*. It follows from our main theorem below that the adversary's optimal strategy never moves a token off a sink! The following lemma is a very general result relating hitting times in a cyclic fashion. We hit upon this result while trying to prove a *sink* exists.

Lemma 4.1 *Let x, y, and z be vertices of a connected, undirected graph G. Then*

$$H(x,y) + H(y,z) + H(z,x) = H(x,z) + H(z,y) + H(y,x).$$

More generally, let $x, w_1, w_2, \ldots, w_k, x \in G$, then
$H(x,w_1) + H(w_1,w_2) + \cdots + H(w_{k-1},w_k) + H(w_k,x) = H(x,w_k) + H(w_k,w_{k-1}) + \cdots +$
$H(w_2,w_1) + H(w_1,x)$

Proof. Making use of the formula for hitting times given by Theorem 4.5, it is straight forward to verify that

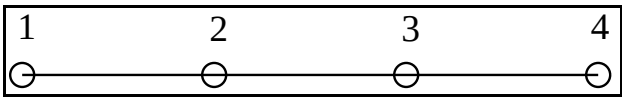
$$\begin{aligned} & H(x,w_1) + H(w_1,w_2) + \cdots + H(w_{k-1},w_k) + H(w_k,x) \\ &\quad = \quad H(x,w_k) + H(w_k,w_{k-1}) + \cdots + H(w_2,w_1) + H(w_1,x) \\ &\quad = \quad m[R_{x,w_1} + R_{w_1,w_2} + \cdots + R_{w_{k-1},w_k} + R_{w_k,x}] \end{aligned}$$

One way of reducing the number of tokens to one is passing them randomly to the neighbors and whenever two or more collide, collapsing them into single tokens. If the processors are synchronous this is a valid assumption. In this case, it is known (results from the literature cited) that, with high probability, the token count will be down to one in polynomial (in n) number of steps. However, in an asynchronous case, the worst case situation is characterized by the presence of an adversary whose aim is to delay stabilization. Accordingly, we define the *meeting time* $M(x, y)$ to be the expected number of moves before tokens placed initially at vertices x and y meet, given optimal play by the adversary in deciding at each step which token shall move.

Some examples should help.

Example 1. Note that the adversary is powerless when the given graph is a cycle (C_n) or a clique (K_n). i.e. it makes no difference whichever token is moved – the relative positioning of the two tokens remains the same.

Example 2. Consider a chain on 4 vertices.



The adversary's strategy is simple in this case – always move a token on the interior of the chain; when both the tokens are on the end-vertices, move either one. This gives rise to the following system of equations:

$$
\begin{aligned}
M(1,2) &= 1 + \frac{1}{2}M(1,3) \\
M(1,3) &= 1 + \frac{1}{2}[M(1,2) + M(1,4)] \\
M(1,4) &= 1 + M(1,3) \\
M(2,4) &= M(1,3) \\
M(2,3) &= 1 + max\{\frac{1}{2}M(1,3), \frac{1}{2}M(2,4)\} \\
&= 1 + \frac{1}{2}M(1,3) \\
M(3,4) &= M(1,2)
\end{aligned}
$$

Theorem 4.6

$$\sum_{\{x,y\}\in E} \frac{R_{xy}}{r_{xy}} = n - 1.$$

Proof. The proof is similar to the proof of Theorem 4.3, and begins with the reciprocity relation,

$$U_z^{xy}/c(z) = U_x^{zy}/c(x)$$

Multiplying both sides by c_{xy}, and taking sums on both sides,

$$\begin{aligned}
\sum_{x\in N(y)} U_z^{xy} c_{xy}/c(z) = \sum_{x\in N(y)} U_x^{zy} c_{xy}/c(x) &= 1 \text{ if } z \neq y \\
\Rightarrow \qquad \sum_{y\in G}\sum_{x\in N(y)} U_z^{xy} c_{xy}/c(z) &= n-1 \\
\Rightarrow \sum_{\{x,y\}\in G} \left(U_z^{xy} c_{xy}/c(z) + U_z^{yx} c_{yx}/c(z)\right) &= n-1
\end{aligned}$$

This clearly implies $\sum_{\{x,y\}\in E} \frac{R_{xy}}{r_{xy}} = n - 1$.

It is equally straight forward to generalize Theorem 5 to the following theorem; we omit the proof.

Theorem 4.7

$$H(x,y) = \frac{1}{2}\sum_z c(z)\,[R_{xy} + R_{yz} - R_{xz}]$$

4.3 The Meeting Time Problem

Self-stabilizing systems is a topic of recent interest in distributed computing. A protocol for a distributed computing network is said to be *self-stabilizing* if, regardless of the state it is begun in, it eventually enters a "legal" state and resumes normal operation. In a token management scheme, only one processor at a time is allowed to change state or perform some particular task. This processor is considered as "holding a token", the token being an abstract object. An obvious problem in designing such a scheme is in recovering from a situation when several tokens are present at the same time.

Proof. Obvious from Theorem 4.5.

We must mention here again that the analog of the above corollary, albeit in the random walk terminology, was proved in [22].

Thus the summary of this section of results is best expressed by

$$\boxed{\begin{aligned} H(x,y) &= mR_{xy} + \frac{1}{2}\sum_z d(z)\,[R_{yz} - R_{xz}] \\ H(y,x) &= mR_{xy} - \frac{1}{2}\sum_z d(z)\,[R_{yz} - R_{xz}] \end{aligned}}$$

4.2 General Networks

Let r_{xy} denote the resistance of branch xy of a general network. The underlying graph is now a *weighted* multigraph with *conductance* $c_{xy} = 1/r_{xy}$ as the weight on edge xy. And $c_{xz} = 0$ if x and z are non-adjacent. The weighted degree is $c(x) = \sum_{y \in N(x)} c_{xy}$, sum of the conductances of branches incident on node x. The transition probabilities p_{xz} of a random walk on G are now defined as follows.

$$\begin{aligned} Prob[\text{moving from x to z in one step}] = p_{xz} &= \frac{c_{xz}}{c(x)} \quad if\ \{x,z\} \in E, \\ &= 0 \qquad otherwise. \end{aligned}$$

It is not difficult to see that identity (*) is true in general networks provided we interpret the *weighted degree* $c(x)$ to be the degree of a vertex x. The proof of identity (**) below can be found in [13].

$$(**) \qquad V_{zy} = \frac{U_z^{xy}}{c(z)} \qquad \forall z$$

Using the above identity, it is straight forward to generalize all the results of the previous section. (i.e. just follow the same proofs.) Specifically, Theorem 4.3 holds true in a weighted sense yielding the following general version of Foster's theorem.

with Proposition 9-58 yields alternative proofs for Corollaries 4.1 and 4.3.

We now turn the above triangle inequality around deriving a result which deserves the status of a full theorem!

Theorem 4.5

$$H(x,y) = \frac{1}{2}\sum_z d(z)\,[R_{xy} + R_{yz} - R_{xz}]$$

Proof. By the triangle inequality (Corollary 4.3) we have

$$R_{xy} + R_{yz} - R_{xz} = 2\frac{U_z^{xy}}{d(z)}$$

Turning it around,

$$U_z^{xy} = (1/2)d(z)\,[R_{xy} + R_{yz} - R_{xz}]$$

Summing over z,

$$\sum_z U_z^{xy} = (1/2)\sum_z d(z)\,[R_{xy} + R_{yz} - R_{xz}]$$

But we have, $H(x,y) = \sum_z U_z^{xy}$
proving the theorem.

Recall that the commute time result gives a precise characterization of the "sum" of expected hitting times $(H(x,y)+H(y,x))$ between vertices x and y. While Theorem 4.5 can be used in giving yet another proof of the commute time result, we use it to derive a formula relating the "difference" of the expected hitting times $(H(y,x) - H(x,y))$ between x and y.

Corollary 4.4

$$H(y,x) - H(x,y) = \sum_z d(z)(R_{xz} - R_{yz})$$

Remark. From Theorem 4.3 it trivially follows that

$$\sum_{\{x,y\}\in E} C(x,y) = 2m(n-1).$$

Triangle Inequality for Effective Resistances. The following triangle inequality for effective resistances is well known in electrical network theory: (also mentioned as Lemma 14 in [7])

$$R_{xz} + R_{zy} \geq R_{xy}$$

We show that the inequality can be made precise with the help of traversals in a random walk.

Corollary 4.3 *Any three nodes x, y, z of a network satisfy*

$$R_{xz} + R_{zy} - R_{xy} = \frac{U_x^{yz}}{d(x)} + \frac{U_y^{xz}}{d(y)}$$

Proof. By Theorem 4.2 and Corollary 4.2,

$$\begin{aligned} R_{xz} + R_{zy} - R_{xy} &= \left(\frac{U_y^{xz}}{d(y)} + \frac{U_z^{yx}}{d(z)}\right) + \left(\frac{U_z^{xy}}{d(z)} + \frac{U_x^{yz}}{d(x)}\right) - \left(\frac{U_z^{xy}}{d(z)} + \frac{U_z^{yx}}{d(z)}\right) \\ &= \frac{U_x^{yz}}{d(x)} + \frac{U_y^{xz}}{d(y)}. \\ \text{also} \quad &= 2\frac{U_x^{yz}}{d(x)} = 2\frac{U_y^{xz}}{d(y)}. \end{aligned}$$

Notice that we have proved Corollaries 4.1 and 4.3 from first principles. Larry Ruzzo recently pointed out that Corollary 4.3 may be related to some already existing literature (Proposition 9-58 of [28]) on random walks.

Proposition 9-58.

$$\begin{aligned} H(x,y) + H(y,z) - H(x,z) &= \frac{U_z^{xy}}{\pi(z)} \\ H(x,y) + H(y,x) &= \frac{U_x^{xy}}{\pi(x)} \end{aligned}$$

where $\pi(x)$ = stationary probability of $x = d(x)/2m$.

Surely enough, our interpretation of effective resistance (Theorems 4.1 and 4.2) together

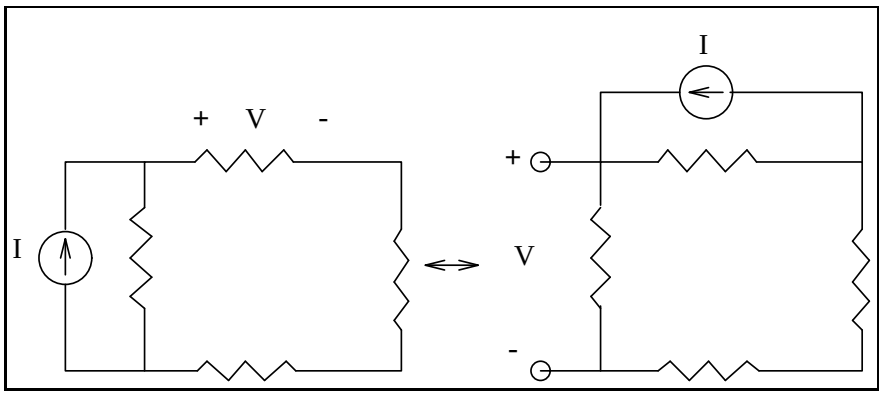


Figure 4.1: Reciprocity

Corollary 4.2 *The expected number of traversals out of w along a specific edge during a random walk from x to y is the same as the expected number of traversals out of x along a specific edge in a random walk from w to y.*

We are now ready to prove Theorem 4.3.

Proof of Theorem 4.3: By Corollary 4.2 we have that

$$U_z^{xy}/d(z) = U_x^{zy}/d(x).$$

where the l.h.s. refers to a random walk from x to y, and the r.h.s. refers to a random walk from z to y. Taking sums on both sides,

$$\sum_{x \in N(y)} U_z^{xy}/d(z) = \sum_{x \in N(y)} U_x^{zy}/d(x) = 1 \text{ if } z \neq y$$

This is because the expected number of times we reach y from one of its neighbors in a *nontrivial* ($z \neq y$) random walk from z to y is precisely 1. Therefore

$$\sum_{y \in G} \sum_{x \in N(y)} U_z^{xy}/d(z) = n - 1$$

$$\text{i.e.} \quad \sum_{\{x,y\} \in E} (U_z^{xy}/d(z) + U_z^{yx}/d(z)) = n - 1.$$

we are done by Theorem 4.2.

Corollary 4.1 *Expected commute time* $C(x,y) = H(x,y) + H(y,x) = 2mR_{xy}$.

Proof. Note that as long as we have not yet reached the "terminal" vertices x and y, we keep visiting other vertices of the graph. So by Theorems 4.1 and 4.2, we have

$$C(x,y) = \sum_{w \in G} U_w = R_{xy} \sum_{w} d(w)$$

$$\text{i.e. } C(x,y) = 2mR_{xy}.$$

It was pointed out to us by P.G. Doyle that the following theorem was originally due to R.M. Foster [21]. Foster proved a slightly more general theorem using network theory and we shall see in the next section that our proof can be extended to the general version.

Theorem 4.3

$$\sum_{\{x,y\} \in E} R_{xy} = n - 1.$$

Before proving this theorem, we interpret the Reciprocity Theorem of electrical networks in terms of random walks. The Reciprocity theorem is applicable only to single-source networks; we merely state the theorem, since the proof can be found in any standard circuit theory text book. (e.g. see [24] .)

Theorem 4.4 (Reciprocity) *The voltage V across any branch of a network, due to a single current source I anywhere else in the network, will equal the voltage across the branch at which the source was originally located if the source is placed at the branch across which the voltage V was originally measured.*

In other words, the location of the current source and the resulting voltage may be interchanged without a change in voltage.

The theorem requires that the direction of the current source have the same correspondence with the polarity of the branch voltage in each position. (The theorem also has a dual in terms of voltage source and current measured in a branch.)

Recall that we have interpreted the voltage V_{wy} when a unit current flows into x and out of y to be $U_w^{xy}/d(w)$. So by the reciprocity theorem, we have

Proof. Let U_w denote the expected number of visits to w in a random "roundtrip": random walk from x to y and back to x. Clearly, $U_w = U_w^{xy} + U_w^{yx}$. Using the above identity (*) and the *superposition principle* we have

$$\begin{aligned} U_w &= V_{wy}d(w) - V_{wx}d(w) \\ &= V_{xy}d(w) \\ &= R_{xy}d(w) \end{aligned}$$

We are done by noting that $U_w/d(w)$ is the expected number of traversals out of w along a specific edge in our random roundtrip.

Interpretation of Polya's recurrence theorem. The relation between *recurrent* random walks on lattices and resistance of infinite networks is immediate from our characterization. A random walk is said to be *recurrent* if the walker is certain to return to the starting point. When a positive probability exists that the walker never returns, the random walk is called *transient*. The famous theorem of Polya asserts that *simple random walk on a d-dimensional lattice is recurrent for* $d = 1, 2$ *and transient for* $d > 2$. From Theorem 4.1 it is clear that the random walk is recurrent iff the effective resistance between the starting point and "point at infinity" goes to infinity! Intuitively, infinite resistance at the boundary always forces the *random walker* to return home! Doyle and Snell discuss [13] the relation between recurrence and effective resistance, by relating effective resistance to *escape probability*, the probability that the walker never returns to the starting point. The subtle difference in the interpretations can be explained. They consider effective resistance to be the reciprocal of the current flow when a 1 volt source is applied between the points of interest. Our characterization is based on the alternative definition of effective resistance, namely, the voltage developed between the points of interest when the source is 1 Amp current.

Let $H(x,y)$ denote the expected *hitting time* for the random walk from x to y, and let $C(x,y)$ denote the expected *commute time* for the random roundtrip between x and y; i.e. $C(x,y) = H(x,y) + H(y,x)$. The following corollary gives a different way of proving one of the main results of [7].

Proof. Our characterization makes use of the well known analogy (see [13]) between random walks and electrical networks. Let U_z^{xy} be the expected number of visits to z (before reaching y) in a random walk from x to y. So $U_y^{xy} = 0$, and considering all possible ways of reaching z, we have

$$\begin{aligned} U_z^{xy} &= \sum_w U_w^{xy} p_{wz} \\ &= \sum_w \frac{U_w^{xy}}{d(w)} \quad (z \neq x, y) \end{aligned}$$

The quantity of interest to us is

$$\frac{U_z^{xy}}{d(z)} = \sum_w \frac{1}{d(z)} \frac{U_w^{xy}}{d(w)} \quad (z \neq x, y)$$

On the other hand, by Kirchoff's Current Law, it follows that

$$V_{zy} = \sum_w \frac{1}{d(z)} V_{wy} \qquad (z \neq x, y)$$

where V_{zy} is the voltage between z and y. Thus we get the following identity (*) by defining $V_{yy} = 0$ and $V_{xy} = U_x^{xy}/d(x)$:

$$(*) \qquad V_{zy} = \frac{U_z^{xy}}{d(z)} \qquad \forall z$$

The right hand side is clearly the expected number of traversals out of x along a specific edge. And *effective resistance* R_{xy} is, by definition, the voltage V_{xy} when a unit current enters x and leaves y. Hence we will be done by proving that a unit current indeed enters x and leaves y. The current in any branch wz is

$$\begin{aligned} i_{wz} &= (V_{wy} - V_{zy})/1 \\ &= \frac{U_w^{xy}}{d(w)} - \frac{U_z^{xy}}{d(z)} \end{aligned}$$

i.e. i_{wz} is the expected number of *net traversals* along wz. This being the interpretation of current, the current entering x is $\sum_w i_{xw} = 1 = \sum_z i_{zy}$ equals the current leaving y.

Theorem 4.2 *R_{xy} is also the expected number of traversals out of w ($\neq x, y$) along any specific edge (w, z) in a* simple *random walk from x to y and then back to x.*

Chapter 4

Hitting Times and the Meeting Time

4.1 Simple Networks

By a *simple network* we mean a (finite) electrical network with all resistances equal. In this section we assume, with no loss of generality, we are dealing with unit resistances. We discuss the general case of networks with unequal resistances in the next section.

Consider a simple network with n nodes and m branches. Thus the underlying graph is a multigraph $G = (V, E)$ with n vertices and m edges. Let $d(x)$ denote the degree of vertex x. The transition probability p_{xz} of a simple random walk on G is defined as follows.

$$\begin{aligned} Prob[\text{moving from x to z in one step}] = p_{xz} &= \frac{1}{d(x)} \quad if\ \{x,z\} \in E, \\ &= 0 \qquad otherwise. \end{aligned}$$

By a random walk *from x to y* we mean a random walk which begins at vertex x, goes around visiting vertices, and stops on reaching y.

Theorem 4.1 *The effective resistance R_{xy} between nodes x and y is exactly the expected number of traversals out of x along any specific edge (x, z) in a* simple *random walk starting at x and ending in y.*

$$
\begin{aligned}
&= e^{-\mu'+o(1)} \left[\frac{e}{(1+\epsilon)}\right]^{(1+\epsilon)\mu} \\
&\leq e^{h(n)\log n[\epsilon-(1+\epsilon)\log(1+\epsilon)]} \\
&< n^{-2+o(1)} \qquad \text{for } \epsilon > 0
\end{aligned}
$$

This completes the proof of Step 3. And further, by Borel-Cantelli lemma, we can conclude $r_k(n) \sim h(n)\log n$, completing the proof of Theorem 3.5.

before. i.e Theorems 3.3 and 3.4 extend without much modification: Let $\epsilon > 0$ be fixed.
(i) $\Pr[r_k(n) \leq (1-\epsilon)\mu] \leq n^{-2+o(1)}$

Proof. By Janson's second inequality, we have, for $0 \leq \epsilon \leq 1$,

$$\begin{aligned}
\Pr[r_k(n) \leq (1-\epsilon)\mu] &\leq e^{-\frac{1}{2(1+o(1))}\epsilon^2\mu} \\
&= e^{-\frac{1}{2(1+o(1))}\epsilon^2 h(n)\log n} \\
&= n^{-\frac{1}{2(1+o(1))}\epsilon^2 h(n)} \\
&< n^{-2+o(1)}, \quad \text{for } \epsilon > 0.
\end{aligned}$$

(ii) $\Pr[r_k(n) > (1+\epsilon)\mu] \leq n^{-2+o(1)}$

Proof. As before we first consider $r_k^*(n)$, the size of a maximal collection of pairwise disjoint representations of n. Since $r_k(n) \sim r_k^*(n)$ (by an argument identical to that in Theorem 3.4), it suffices to prove

$$\Pr[r_k^*(n) > (1+\epsilon)\mu] \leq n^{-2+o(1)}$$

Towards this, consider $\mathbf{s} = \{S_1, \ldots, S_t\}$, a collection of pairwise disjoint representations of n. Let $r_k(n;\mathbf{s}) =\mid \{\{x_1, \ldots, x_k\} : x_1 + \cdots + x_k = n; x_i \text{ distinct and } x_i \in \mathcal{S}\backslash\mathbf{s}\} \mid$ and $\mu' = E[r_k(n;\mathbf{s})]$. Using the proof of Lemma 2.12, we can show that $\mu' > (h(n) + o(1))\log n$, as long as $\mid \mathbf{s} \mid < c\log n$, where c is any positive constant. The rest is analogous to Lemma 3.3:

$$\begin{aligned}
\Pr[r_k^*(n) > (1+\epsilon)\mu] &= \sum_{i > (1+\epsilon)\mu}^{\infty} \Pr[\exists \mathbf{s} \text{ s.t. } \mid \mathbf{s} \mid = i \text{ and } \mathbf{s} \text{ is maximal }] \\
&= \sum_{i > (1+\epsilon)\mu}^{\infty} \Pr[\exists \mathbf{s} \text{ s.t. } \mid \mathbf{s} \mid = i] \times \Pr[\mathbf{s} \text{ is maximal } \Big| \mid \mathbf{s} \mid = i] \\
&\leq \sum_{i > (1+\epsilon)\mu}^{\infty} \frac{\mu^i}{i!} \times e^{-\mu' + o(1)} \\
&= e^{-\mu' + o(1)} \sum_{i > (1+\epsilon)\mu}^{\infty} \frac{\mu^i}{i!} \\
&\leq e^{-\mu' + o(1)} \left[\frac{e\mu}{(1+\epsilon)\mu}\right]^{(1+\epsilon)\mu}
\end{aligned}$$

We conclude the proof of Claim 4 by showing that

Claim 5.

$$\lim_{\delta \to 0} \lim_{n \to \infty} ADD' - ADD'_\delta = 0$$

The proof of Claim 5 is identical to that of Claim 3 (Section 3.2).

Thus Claims 4 and 5 together prove Step 1 :

$$\mu = E\left[r_k(n)\right] \sim h(n) \log n$$

Proof of Step 2. Recall that $r_l(n)$ was used to denote the number of representations of n as a sum of l distinct numbers from $\mathcal{S}$, and μ_l denoted the expectation of $r_l(n)$. It is easy to see that the proof of Lemma 2.8 goes through as long as $h(n) = n^{o(1)}$ (which we know by Condition 5), and thus

for $2 \leq l \leq (k-1)$,

$$\begin{aligned}
\mu_l &= \sum_{\substack{x_1 + \cdots + x_l = n \\ 1 \leq x_1 < \cdots < x_l < n}} \Pr\left[x_1\right] \cdots \Pr[x_l] \\
&= \sum C_k \left(\frac{h(x_1) \log x_1}{x_1^{k-1}} \right)^{1/k} \cdots C_k \left(\frac{h(x_k) \log x_l}{x_l^{k-1}} \right)^{1/k} \\
&= n^{o(1)} \sum \frac{1}{(x_1 \cdots x_l)^{(k-1)/k}} \\
&= n^{o(1)} SUM_l \\
&= n^{-1 + \frac{l}{k} + o(1)} \qquad \text{(by Lemma 2.8)}
\end{aligned}$$

This in turn can be used to bound the correlation as in Lemma 2.11. Let S_i and S_j be two valid representations of n as a sum of k distinct numbers from $\mathcal{S}$. The proof of Lemma 2.11 goes through, yielding

$$\Delta = \sum_{i \sim j} \Pr\left[S_i \wedge S_j\right] = o(1).$$

Proof of Step 3.

Once we establish Steps 1 and 2, the proof of Step 3 is identical to what we have seen

where $\sum^*$ is *short hand* for $\sum_{\substack{x_1+\cdots+x_k=n \\ 1\le x_1<\cdots<x_k<n}}$.

Consider $ADD' = \dfrac{\mu}{C_k^k h(n)\log n} = \dfrac{1}{C_k^k h(n)\log n}\sum^* \Pr[x_1]\cdots\Pr[x_k]$.

Let $x_i = \alpha_i n, \quad 0 < \alpha_i < 1$. Fixing $k, C_k > 0, z_0$, we claim that

Claim 4.

$$\lim_{n\to\infty} ADD' = \frac{1}{C_k^k} = \int_{\substack{0\le\alpha_1,\cdots,\alpha_{k-1}<1 \\ \alpha_1+\cdots+\alpha_{k-1}\le 1}} \frac{1}{[\alpha_1\cdots\alpha_{k-1}(1-\alpha_1-\ldots-\alpha_{k-1})]^{(k-1)/k}}$$

As before, we consider the following special summation:

$$ADD'_\delta = \frac{1}{C_k^k h(n)\log n}\sum\nolimits_\delta \Pr[x_1]\cdots\Pr[x_k]$$

where $\sum_\delta$ denotes $\sum_{\substack{x_1+\cdots+x_k=n \\ \delta n<x_1<\cdots<x_k<n}}$

Thus

$$ADD'_\delta = \frac{1}{C_k^k h(n)\log n}\sum\nolimits_\delta C_k\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C_k\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k}[h(x_1)\cdots h(x_k)]^{1/k}$$

Since $h(n)$ is monotone increasing,

$$\begin{aligned} ADD'_\delta &< \frac{1}{C_k^k h(n)\log n}\sum\nolimits_\delta C_k\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k}[h(n)\cdots h(n)]^{1/k} \\ &= \frac{1}{C_k^k h(n)\log n}h(n)\sum\nolimits_\delta C_k\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C_k\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k} \\ &= \frac{1}{C_k^k \log n}\sum\nolimits_\delta C_k\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C_k\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k} \\ &= ADD_\delta \text{ (from Section 3.2)} \end{aligned}$$

On the other hand, by Condition (3) above, we have for $\delta > 0, \quad h(n\delta) \sim h(n)$. Thus

$$\begin{aligned} ADD'_\delta &> \frac{1}{C_k^k h(n)\log n}[h(n)(1-o(1))]\sum\nolimits_\delta C_k\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C_k\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k} \\ &= \frac{1}{C_k^k \log n}[(1-o(1))]\sum\nolimits_\delta C_k\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C_k\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k} \\ &= (1-o(1))\,ADD_\delta \end{aligned}$$

(4) $h \to \infty$, and

(5) $h(n) = n^{o(1)}$

Note that the conditions are fairly natural. Functions such as $\log^2 n$ and $e^{\sqrt{\log n}}$ are some examples. We show that

Theorem 3.5 *For fixed k, $h(n)$ smooth, there exists an asymptotic basis such that*

$$r_k(n) \sim h(n)\log n$$

Outline of the Proof. The proof is very similar to the proof of the Main Theorem in Chapter 2. Let k be fixed. We choose a number z to belong to the desired sequence $\mathcal{S}$ with probability $p(z)$ given by

$$\begin{aligned} p(z) &= C_k \frac{(h(z).\log z)^{1/k}}{z^{(k-1)/k}} \quad \text{for} \quad z > z_0 \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

where z_0 is, as before, the smallest constant such that $p(z) \le 1/2$ and C_k is the following absolute constant, depending only on k:

$$C_k = \left[\int_{\substack{0 \le \alpha_1, \ldots, \alpha_{k-1} \le 1 \\ \alpha_1 + \cdots + \alpha_{k-1} \le 1}} \frac{1}{[\alpha_1 \cdots \alpha_{k-1}(1 - \alpha_1 - \ldots - \alpha_{k-1})]^{(k-1)/k}} \right]^{-1/k}$$

The proof that this construction works is in 3 steps:

Step 1. $\mu = E\left[r_k(n)\right] \sim h(n)\log n$

Step 2. The correlation term $\Delta = o(1)$

Step 3. Finally proving the inequalities, for $\epsilon > 0$,

$$(i) \quad \Pr[r_k(n) < (1-\epsilon)\mu] < n^{-2+o(1)}$$

$$(ii) \quad \Pr[r_k(n) > (1+\epsilon)\mu] < n^{-2+o(1)}$$

Proof of Step 1. By Definition,

$$\begin{aligned} \mu &= \sum_{\substack{x_1 + \cdots + x_k = n \\ 1 \le x_1 < \cdots < x_k < n}} \Pr[x_1] \cdots \Pr[x_k] \\ &= \sum\nolimits^{*} C_k \left(\frac{h(x_1)\log x_1}{x_1^{k-1}} \right)^{1/k} \cdots C_k \left(\frac{h(x_k)\log x_k}{x_k^{k-1}} \right)^{1/k} \end{aligned}$$

OP's. By Lemma 3.4, the size of a maximal matching $\leq L$. This in turn implies the independence number $> \mid V_S \mid - 2L$. ("Removal" of a maximal matching leaves an independent set!) Thus $r_k^*(n) \geq r_k(n) - 2L$.

(ii) Obvious, from Part (i) and Lemma 3.3.

Proof of Theorem 3.1. By Theorems 3.3 and 3.4 we have, for fixed k, for $\epsilon > (e-1)\sqrt{\frac{2}{b}}$,

$$(i) \quad \Pr[r_k(n) < (1-\epsilon)\mu] < n^{-2+o(1)}$$
$$(ii) \quad \Pr[r_k(n) > (1+\epsilon)\mu] < n^{-2+o(1)}$$

Given $\epsilon > 0$, we can make b large enough (by choosing C appropriately) such that the criterion $\epsilon > (e-1)\sqrt{\frac{2}{b}}$ holds. Then applying Borel-Cantelli lemma on both (i) and (ii) above yields the assertion of Theorem 3.1: For every fixed k, given $\epsilon > 0$, there exists an asymptotic basis $\mathcal{S}$ of order k such that, for n large,

$$(1-\epsilon)b\log n < r_k(n) < (1+\epsilon)b\log n$$

where b is an explicit constant.

3.4 Some Extensions

The alert reader would have noticed that the representation function $\Theta(\log n)$ comes up naturally in most of the proofs so far. While it seems crucial (at least with our techniques) to require that the representation function grow at least as fast as $\log n$, one might wonder how general a function the theorems extend to! We show that the same proof techniques work in proving that for fixed k, if $g(n) >> \log n$, monotone increasing, and "smooth" (in a certain precise sense) then, there exists an asymptotic basis such that $r_k(n) \sim g(n)$.

More precisely, let $h(n)$ be called *smooth* whenever the following conditions hold:

(1) $h \geq 1$

(2) h monotone increasing

(3) For any $\delta > 0$, $\quad h(n\delta) \sim h(n)$

However we would like to extend Lemma 3.3 and show $r_k(n) \sim r_k^*(n)$. Let S_i and S_j be two valid representations of n as a sum of k distinct numbers from S. By an *overlapping pair*, denoted OP, we mean a pair of representations of n which share at least one number (of the k numbers) in common.

Lemma 3.4 *a.a.* $\exists L$ *s.t. the number of disjoint OP's* $\leq L$.

Proof. Let us first compute the expected number μ_{op} of OP's

$$\begin{aligned} E[\#\text{ OP's }] &= \sum_{l=1}^{k-2} \sum_{|S_i \wedge S_j|=l} \Pr[S_i \wedge S_j] \\ \mu_{op} &= n^{-1+o(1)}, \quad \text{by Lemma 2.11.} \end{aligned}$$

By the *disjointness lemma* (Lemma 2.1), we have

$$\begin{aligned} \Pr[\text{ OP's } > L] &< \frac{\mu_{op}^L}{L!} \\ &< \frac{1}{L!}\left(n^{-1+o(1)}\right)^L \\ &< n^{-2+o(1)} \quad \text{for } L \text{ large enough.} \end{aligned}$$

By the Borel-Cantelli lemma, the above assertion implies that
a.a. $\exists L$ s.t. the number of disjoint OP's $\leq L$.

Theorem 3.4

$$\begin{aligned} &(i) \quad r_k(n) \leq r_k^*(n) + 2L \\ &(ii) \quad \Pr[r_k(n) > (1+\epsilon)\mu] < n^{-2+o(1)}, \qquad \textit{for } \epsilon > (e-1)\sqrt{\frac{2}{b}} \end{aligned}$$

Proof. (i) It helps to think in terms of graphs! Let $G = (V_S, E_S)$ represent the graph with a vertex v_i for each representation S_i of n; and $\{v_i, v_j\} \in E_S$ iff $S_i \cap S_j \neq \phi$. Note that the number of vertices $= r_k(n)$ and the size of a maximal independent set $= r_k^*(n)$. Further, the edges represent OP's and a *matching* represents a collection of disjoint

In the previous section we saw that $\mu \sim b \log n$. Therefore, for $\epsilon > \sqrt{\frac{2}{b}}$,

$$\Pr[r_k(n) < (1-\epsilon)\mu] \leq n^{-2+o(1)}.$$

To bound $r_k(n)$ from above we first consider $r_k^*(n)$, the size of a maximal collection of pairwise disjoint representations of n. We bound $r_k^*(n)$ from above and then show that $r_k(n) \sim r_k^*(n)$.

Let $\mathbf{s} = \{S_1, \ldots, S_t\}$ be a collection of pairwise disjoint representations of n. Further let $r_k(n;\mathbf{s}) = \mid \{\{x_1, \ldots, x_k\} : x_1 + \cdots + x_k = n; x_i \text{ distinct and } x_i \in \mathcal{S} \setminus \mathbf{s}\} \mid$. We want to estimate $\mu' = E[r_k(n;\mathbf{s})]$. Lemma 2.12 guarantees that $\mu' > (b + o(1)) \log n$, provided that $\mid \mathbf{s} \mid < C \log n$. Using this we can prove

Lemma 3.3 *For* $\epsilon > (e-1)\sqrt{\frac{2}{b}}$

$$\Pr[r_k^*(n) > (1+\epsilon)\mu] < n^{-2+o(1)}$$

Proof. $\Pr[r_k^*(n) > (1+\epsilon)\mu] = \sum_{i > (1+\epsilon)\mu}^{\infty} \Pr[\exists \mathbf{s} \text{ s.t. } \mid \mathbf{s} \mid = i \text{ and } \mathbf{s} \text{ is maximal }]$

$$\begin{aligned}
&= \sum_{i > (1+\epsilon)\mu}^{\infty} \Pr[\exists \mathbf{s} \text{ s.t. } \mid \mathbf{s} \mid = i] \times \Pr[\mathbf{s} \text{ is maximal } \Big| \mid \mathbf{s} \mid = i] \\
&\leq \sum_{i > (1+\epsilon)\mu}^{\infty} \frac{\mu^i}{i!} \times e^{-\mu' + o(1)} \\
&= e^{-\mu' + o(1)} \sum_{i > (1+\epsilon)\mu}^{\infty} \frac{\mu^i}{i!} \\
&\leq e^{-\mu' + o(1)} \left[\frac{e\mu}{(1+\epsilon)\mu}\right]^{(1+\epsilon)\mu} \\
&= e^{-\mu' + o(1)} \left[\frac{e}{(1+\epsilon)}\right]^{(1+\epsilon)\mu} \\
&\leq e^{b \log n [\epsilon - (1+\epsilon)\log(1+\epsilon)]} \\
&< n^{-2+o(1)} \qquad \text{for } \epsilon > (e-1)\sqrt{\frac{2}{b}}
\end{aligned}$$

By the **CLAIM**,

$$\begin{aligned} ADD^*_\delta &\leq \left(1+\frac{\epsilon}{10}\right) \sum_{(\frac{a_1}{m},\cdots,\frac{a_{k-1}}{m})\in D_{\delta''}} \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \\ &= \left(1+\frac{\epsilon}{10}\right)\frac{1}{m^{k-1}} \sum_{(\frac{a_1}{m},\cdots,\frac{a_{k-1}}{m})\in D_{\delta''}} \left[\frac{1}{\frac{a_1}{m}\cdots\frac{a_{k-1}}{m}(1-\frac{a_1}{m}-\cdots-\frac{a_{k-1}}{m})}\right]^{(k-1)/k} \\ &\leq \left(1+\frac{\epsilon}{10}\right)\left(1+\frac{\epsilon}{10}\right) TEMP'_m \end{aligned}$$

Once again we appeal to the *definition* of Integral, to assert that

$$\lim_{m\to\infty} TEMP'_m = INT_{\delta''}.$$

Thus, $\exists n''_0, \quad n > n''_0 \Rightarrow$

$$\begin{aligned} ADD^*_\delta &\leq \left(1+\frac{\epsilon}{10}\right)\left(1+\frac{\epsilon}{10}\right) INT_{\delta''} \\ &\leq (1+\epsilon) INT_{\delta''} \end{aligned}$$

completing the proof of Lemma 3.2.

3.3 The Second Part

The essence of this section is that $r_k(n)$ is, in fact, arbitrarily close to its expected value μ. We prove this by bounding $r_k(n)$ from above and below. The lower bound proof essentially appears at the end of the previous chapter.

Theorem 3.3 *For* $\epsilon > \sqrt{\frac{2}{b}}$

$$\Pr[r_k(n) < (1-\epsilon)\mu] \leq n^{-2+o(1)}$$

Proof. By Janson's second inequality (Theorem J, Ch.2) we have, for $0 \leq \epsilon \leq 1$,

$$\Pr[r_k(n) \leq (1-\epsilon)\mu] \leq e^{-\frac{1}{2(1+o(1))}\epsilon^2\mu}$$

Thus, $\exists n_0', \quad n > n_0' \Rightarrow$

$$\begin{aligned} ADD_\delta^* &\geq \left(1-\frac{\epsilon}{10}\right)\left(1-\frac{\epsilon}{10}\right) INT_{\delta'} \\ &\geq (1-\epsilon) INT_{\delta'} \end{aligned}$$

completing the proof of Lemma 3.1.

Proof of Lemma 3.2. Pick m so that

(i) $\delta'' + \frac{2}{m} < \delta$,

(ii) $TEMP'_m = \frac{1}{m^{k-1}} \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \in D_{\delta''}} \left[\frac{1}{\frac{a_1}{m}\cdots\frac{a_{k-1}}{m}(1-\frac{a_1}{m}-\cdots-\frac{a_{k-1}}{m})} \right]^{(k-1)/k} \leq (1+\frac{\epsilon}{10})\, INT_{\delta''}$,

(iii) and pick n so that **CLAIM** is applicable.

$$\begin{aligned} ADD_\delta^* &= \sum_{a_1,\ldots,a_{k-1}} \sum_{\substack{\frac{a_i}{m} \leq \frac{x_i}{n} < \frac{a_i+1}{m} \\ x_1,\ldots,x_{k-1} \geq \delta n \\ n-x_1-\cdots-x_{k-1} \geq \delta n}} TERM \\ &\leq \sum_{a_1,\ldots,a_{k-1}} \sum_{\substack{\frac{a_i}{m} \leq \frac{x_i}{n} < \frac{a_i+1}{m} \\ x_1,\ldots,x_{k-1} \geq (\delta-\frac{1}{m})n \\ n-x_1-\cdots-x_{k-1} \geq \delta n}} TERM \\ &\leq \sum_{\substack{a_1,\ldots,a_{k-1} \\ \frac{a_i}{m} \geq \delta'' \\ 1-\frac{a_1}{m}-\cdots-\frac{a_{k-1}}{m} \geq \delta''}} \sum_{\frac{a_i}{m} \leq \frac{x_i}{n} < \frac{a_i+1}{m}} TERM \end{aligned}$$

the last step is justified as follows:

The condition $\frac{a_i+1}{m} > \frac{x_i}{n} > \delta - \frac{1}{m} \quad \Rightarrow \quad \frac{a_i}{m} > \delta - \frac{2}{m} > \delta''$;

and the condition $1 - \frac{x_1}{n} - \cdots - \frac{x_{k-1}}{n} \geq \delta \quad \Rightarrow \quad 1 - \frac{a_1}{m} - \cdots - \frac{a_{k-1}}{m} \geq \delta \geq \delta''$.

Thus,

$$ADD_\delta^* \leq \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \in D_{\delta''}} \sum_{\frac{a_i}{m} \leq \frac{x_i}{n} < \frac{a_i+1}{m}} TERM$$

Consider,

$$\begin{aligned} ADD^*_\delta &= \sum_{a_1,\ldots,a_{k-1}} \sum_{\substack{\frac{a_i}{m} \le \frac{x_i}{n} < \frac{a_i+1}{m} \\ x_1,\ldots,x_{k-1} \ge \delta n \\ n-x_1-\cdots-x_{k-1} \ge \delta n}} TERM \\ &\ge \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \in D_{\delta'}} \sum_{\substack{\frac{a_i}{m} \le \frac{x_i}{n} < \frac{a_i+1}{m} \\ x_1,\ldots,x_{k-1} \ge \delta n \\ n-x_1-\cdots-x_{k-1} \ge \delta n}} TERM \\ &= \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \in D_{\delta'}} \sum_{\frac{a_i}{m} \le \frac{x_i}{n} < \frac{a_i+1}{m}} TERM \end{aligned}$$

where we dropped conditions (ii) and (iii) under the summation, as they became superfluous. Condition (ii) is implied, since $x_i/n \ge a_i/m \Rightarrow x_i \ge \delta' n \ge \delta n$. Condition (iii) is also implied as follows.

$$\begin{aligned} n - x_1 - \cdots - x_{k-1} &= n(1 - \frac{x_1}{n} - \cdots - \frac{x_{k-1}}{n}) \\ &\ge n\left(1 - \frac{a_1}{m} - \cdots - \frac{a_{k-1}}{m} - \frac{k-1}{m}\right) \\ &\ge n\left(\delta' - \frac{k-1}{m}\right) \\ &> \delta n \quad \text{by our choice of } \delta'. \end{aligned}$$

Using the above **Claim**, we now have

$$\begin{aligned} ADD^*_\delta &\ge \left(1 - \frac{\epsilon}{10}\right) \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \in D_{\delta'}} \left[\frac{1}{a_1 \cdots a_{k-1}(m - a_1 - \cdots - a_{k-1})}\right]^{(k-1)/k} \\ &= \left(1 - \frac{\epsilon}{10}\right) \frac{1}{m^{k-1}} \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \in D_{\delta'}} \left[\frac{1}{\frac{a_1}{m} \cdots \frac{a_{k-1}}{m}(1 - \frac{a_1}{m} - \cdots - \frac{a_{k-1}}{m})}\right]^{(k-1)/k} \\ &= \left(1 - \frac{\epsilon}{10}\right) TEMP_m \qquad \text{(say)} \end{aligned}$$

By the definition of (Riemann) Integral,

$$\lim_{m \to \infty} TEMP_m = INT_{\delta'}.$$

$$
\begin{aligned}
&= \frac{\left(1-\frac{m}{n}\right)^{k-1}}{\left[a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})\left(1+\frac{1}{a_1}\right)\cdots\left(1+\frac{1}{a_{k-1}}\right)\right]^{(k-1)/k}} \\
&\geq \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \frac{\left(1-\frac{m}{n}\right)^{k-1}}{\left[\left(1+\frac{1}{\delta m}\right)\cdots\left(1+\frac{1}{\delta m}\right)\right]^{(k-1)/k}} \\
&= \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \frac{\left(1-\frac{m}{n}\right)^{k-1}}{\left(1+\frac{1}{\delta m}\right)^{(k-1)^2/k}}
\end{aligned}
$$

Pick m_0' such that $m > m_0' \Rightarrow \frac{1}{\delta m} < \epsilon''$; and then pick n_0' such that $n > n_0' \Rightarrow \frac{m}{n} < \epsilon''$, so that

$$
\begin{aligned}
STUFF &> \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \frac{(1-\epsilon'')^{k-1}}{(1+\epsilon'')^{(k-1)^2/k}} \\
&> \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \left(1-\frac{\epsilon}{10}\right)
\end{aligned}
$$

This completes the proof of CLAIM.

We review here the basic definition of *Integral* which we shall find useful in proving Lemmas 3.1 and 3.2.

Riemann Integral. If $D \subseteq \Re^{k-1}$ is *bounded*, and $f(z_1,\ldots z_{k-1})$ is (Riemann) *integrable* then

$$
\int_D f = \lim_{m\to\infty} \sum_{\substack{a_1,\ldots,a_{k-1}\in Z \\ (\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m})\in D}} f(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m}) \times \frac{1}{m^{k-1}}
$$

Proof of Lemma 3.1. Pick m so that

(i) $\delta' - \frac{k-1}{m} > \delta$,

(ii) $TEMP_m = \frac{1}{m^{k-1}} \sum_{(\frac{a_1}{m},\ldots,\frac{a_{k-1}}{m})\in D_{\delta'}} \left[\frac{1}{\frac{a_1}{m}\cdots\frac{a_{k-1}}{m}(1-\frac{a_1}{m}-\cdots-\frac{a_{k-1}}{m})}\right]^{(k-1)/k} \geq (1-\frac{\epsilon}{10})\, INT_{\delta'}$,

(iii) and pick n so that **CLAIM** is applicable.

Let $TERM$ denote $\left[\frac{1}{(x_1x_2\cdots x_k)}\right]^{(k-1)/k}$.

$$
\begin{aligned}
&> n - \frac{n(a_1+1)}{m} - \cdots - \frac{n(a_{k-1}+1)}{m} \\
&= \frac{n}{m}\left[m - a_1 - \cdots - a_{k-1} - (k-1)\right]
\end{aligned}
$$

Finally, since each x_i can take atmost $\frac{n}{m}+1$ integer values, the number of terms in the summation in $STUFF$ is atmost $\left(\frac{n}{m}+1\right)^{k-1}$. Thus

$$
\begin{aligned}
STUFF &< \frac{\left(\frac{n}{m}+1\right)^{k-1}}{\left[\frac{na_1}{m}\cdots\frac{na_{k-1}}{m}\frac{n}{m}[m-a_1-\cdots-a_{k-1}-(k-1)]\right]^{(k-1)/k}} \\
&= \frac{\left(1+\frac{m}{n}\right)^{k-1}}{\left[a_1\cdots a_{k-1}[m-a_1-\cdots-a_{k-1}-(k-1)]\right]^{(k-1)/k}} \\
&= \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \frac{\left(1+\frac{m}{n}\right)^{k-1}}{\left[1-\frac{k-1}{m-a_1-\cdots-a_{k-1}}\right]^{(k-1)/k}}
\end{aligned}
$$

Since $(\frac{a_1}{m},\cdots,\frac{a_{k-1}}{m}) \in D_\delta$, we have $1-\frac{a_1}{m}-\cdots-\frac{a_{k-1}}{m} \geq \delta$. Or equivalently, $m-a_1-\cdots-a_{k-1} \geq \delta m$. Thus

$$
STUFF < \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \frac{\left(1+\frac{m}{n}\right)^{k-1}}{\left[1-\frac{k-1}{\delta m}\right]^{(k-1)/k}}
$$

Pick m_0 such that $m > m_0 \Rightarrow \frac{k-1}{\delta m} < \epsilon'$; and then pick n_0 such that $n > n_0 \Rightarrow \frac{m}{n} < \epsilon'$, so that

$$
\begin{aligned}
STUFF &< \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \frac{\left(1+\epsilon'\right)^{k-1}}{\left(1-\epsilon'\right)^{(k-1)/k}} \\
&< \left[\frac{1}{a_1\cdots a_{k-1}(m-a_1-\cdots-a_{k-1})}\right]^{(k-1)/k} \left(1+\frac{\epsilon}{10}\right)
\end{aligned}
$$

(ii) Lower estimate for $STUFF$:

We have for $1 \leq i \leq k-1$, $x_i < \frac{n(a_i+1)}{m}$, and since there are at least $(\frac{n}{m}-1)$ possible integer values for each x_i, we have

$$
STUFF > \frac{\left(\frac{n}{m}-1\right)^{k-1}}{\left[\frac{n(a_1+1)}{m}\cdots\frac{n(a_{k-1}+1)}{m}\frac{n}{m}[m-a_1-\cdots-a_{k-1}]\right]^{(k-1)/k}}
$$

where $b = C^k \int_{\substack{0 \le \alpha_1, \ldots, \alpha_{k-1} < 1 \\ \alpha_1 + \cdots + \alpha_{k-1} \le 1}} \frac{1}{[\alpha_1 \cdots \alpha_{k-1}(1 - \alpha_1 - \ldots - \alpha_{k-1})]^{(k-1)/k}}$
completing the proof of Theorem 3.2. This brings us to the end of the first part of the proof (*modulo* Lemmas 3.1 and 3.2)!

Proofs of Lemmas 3.1 and 3.2

Given $\delta > 0$, we define D_δ (D for *domain of integration*) as follows.

$$D_\delta = \{(\alpha_1, \ldots, \alpha_{k-1}) : \delta \le \alpha_1, \ldots, \alpha_{k-1} \text{ and } \delta \le 1 - \alpha_1 - \cdots - \alpha_{k-1}\}$$

A *representation* of n is of the form $x_1 + \cdots + x_k = n$. Further, we assume that $x_k = n - x_1 - \cdots - x_{k-1}$, and bother to specify only $x_1, \ldots, x_{k-1}$.

We need to establish the following estimates before we proceed with the proofs of Lemmas 3.1 and 3.2.

CLAIM. Fix $k, \delta, \epsilon' > 0$. Then
$\exists m_0, m > m_0 \Rightarrow \exists n_0, n > n_0 \Rightarrow$

$$(i) \quad \sum_{\substack{\frac{a_i}{m} \le \frac{x_i}{n} < \frac{a_i+1}{m}, \\ 1 \le i \le k-1}} \left[\frac{1}{(x_1 x_2 \cdots x_k)}\right]^{(k-1)/k} \le \left[\frac{1}{a_1 \cdots a_{k-1}(m - a_1 - \cdots - a_{k-1})}\right]^{(k-1)/k} \left(1 + \frac{\epsilon}{10}\right)$$

$$(ii) \quad \sum_{\substack{\frac{a_i}{m} \le \frac{x_i}{n} < \frac{a_i+1}{m}, \\ 1 \le i \le k-1}} \left[\frac{1}{(x_1 x_2 \cdots x_k)}\right]^{(k-1)/k} \ge \left[\frac{1}{a_1 \cdots a_{k-1}(m - a_1 - \cdots - a_{k-1})}\right]^{(k-1)/k} \left(1 - \frac{\epsilon}{10}\right)$$

for all $a_1, \ldots, a_{k-1}$ with $(\frac{a_1}{m}, \cdots, \frac{a_{k-1}}{m}) \in D_\delta$.

Proof. Let us denote

$$\sum_{\substack{\frac{a_i}{m} \le \frac{x_i}{n} < \frac{a_i+1}{m}, \\ 1 \le i \le k-1}} \left[\frac{1}{(x_1 x_2 \cdots x_k)}\right]^{(k-1)/k} \text{ by } STUFF.$$

(i) Upper estimate for $STUFF$:
We have, for $1 \le i \le k-1$, $x_i \ge \frac{n a_i}{m}$. And since $x_i < \frac{n(a_i+1)}{m}$, we have

$$x_k = n - x_1 - \cdots - x_{k-1}$$

Lemma 3.2 *Fix δ, δ'' with $\delta'' < \delta$. Fix $\epsilon > 0$. Then*

$$\exists n_0'', \quad n > n_0'' \;\Rightarrow\; ADD_\delta^* \leq INT_{\delta''}(1+\epsilon).$$

Now to prove Claim 2a, given δ, we simply pick

$$\delta', \text{ s.t. } \delta < \delta' \text{ and } INT_{\delta'} \geq INT_\delta(1-\epsilon)$$

and pick

$$\delta'', \text{ s.t. } \delta > \delta'' \text{ and } INT_{\delta''} \leq INT_\delta(1+\epsilon).$$

By Lemmas 3.1 and 3.2 then, this would imply

for $\epsilon > 0$ fixed , $\exists\ n_0\ \ (=\ \max\ (n_0', n_0''))$, $\quad n > n_0 \Rightarrow$

$$INT_\delta(1-\epsilon)^2 \leq ADD_\delta^* \leq INT_\delta(1+\epsilon)^2.$$

completing the proof of Claim 2a, and hence of Claim 2.

Proof of Claim 3. Consider $ADD - ADD_\delta$,

$$\frac{1}{C^k \log n} \sum_{\substack{x_1+\cdots+x_k=n \\ 1\leq x_1<\cdots<x_k<n}} \Pr[x_1]\cdots\Pr[x_k] \;-\; \frac{1}{C^k \log n} \sum_{\substack{x_1+\cdots+x_k=n \\ \delta n<x_1\cdots<x_k<n}} \Pr[x_1]\cdots\Pr[x_k]$$

$$= \frac{1}{C^k \log n} O\left(\sum_{\substack{x_1+\cdots+x_k=n \\ 1\leq x_i<\delta n \\ some\ \ i}} \Pr[x_1]\cdots\Pr[x_k] \right)$$

$$= O(sum)$$

where sum, defined and estimated in the previous section (Lemma 2.5), is as follows.

$$\begin{aligned} sum &= \sum_{\substack{x_1+\cdots+x_k=n \\ 1\leq x_1<\delta n}} \frac{1}{(x_1\cdots x_k)^{(k-1)/k}} \\ &= o(1) \end{aligned}$$

This completes the proof of Claim 3. As mentioned before, Claims 2 and 3 complete the proof of Claim 1. Thus we have

$$\mu \sim b \log n$$

Proof of Claim 2.

Let $\epsilon > 0$ be arbitrarily small. We need to show

$$\exists n_0, \quad n > n_0 \Rightarrow \qquad INT_\delta - \epsilon < ADD_\delta < INT_\delta + \epsilon$$

We want to further *preprocess* ADD_δ, i.e. get rid of the numerator part of it. For this purpose, pick n_0 so that for $\delta n_0 < z < n$, we have

$$(\log n)^{1/k}(1 - \frac{\epsilon}{2})^{1/k} < (\log z)^{1/k} < (\log n)^{1/k}$$

Thus

$$\begin{aligned}
ADD_\delta &= \frac{1}{C^k \log n} \sum\nolimits_\delta \Pr[x_1] \cdots \Pr[x_k] \\
&= \frac{1}{C^k \log n} \sum\nolimits_\delta C \left(\frac{\log x_1}{x_1^{k-1}} \right)^{1/k} \cdots C \left(\frac{\log x_k}{x_k^{k-1}} \right)^{1/k} \\
&< \frac{1}{C^k \log n} \sum\nolimits_\delta \frac{C^k \log n}{(x_1 \cdots x_k)^{(k-1)/k}} \\
&= \sum\nolimits_\delta \frac{1}{(x_1 \cdots x_k)^{(k-1)/k}} \\
&= ADD^*_\delta \quad \text{(say)}
\end{aligned}$$

Similarly we have $ADD_\delta > (1 - \frac{\epsilon}{2}) ADD^*_\delta$; thus

$$(1 - \frac{\epsilon}{2}) ADD^*_\delta < ADD_\delta < ADD^*_\delta$$

Now the consolation is that it is enough to prove Claim 2a below to complete the proof of Claim 2.

Claim 2a.

$$\lim_{n \to \infty} ADD^*_\delta = INT_\delta$$

Proof. Assume we have the following Lemmas 3.1 and 3.2, the proofs of which we defer to the end of this section.

Lemma 3.1 *Fix δ, δ' with $\delta < \delta'$. Fix $\epsilon > 0$. Then*

$$\exists n'_0, \quad n > n'_0 \;\Rightarrow\; ADD^*_\delta \geq INT_{\delta'}(1 - \epsilon).$$

By definition,

$$\begin{aligned}\mu &= \sum_{\substack{x_1+\cdots+x_k=n \\ 1\le x_1<\cdots<x_k<n}} \Pr[x_1]\cdots\Pr[x_k] \\ &= \sum{}^{*} C\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k}\end{aligned}$$

where $\sum^{*}$ is *short hand* for $\sum_{\substack{x_1+\cdots+x_k=n \\ 1\le x_1<\cdots<x_k<n}}$.

Consider $ADD = \dfrac{1}{C^k \log n}\sum^{*}\Pr[x_1]\cdots\Pr[x_k]$.
Let $x_i = \alpha_i n, \quad 0<\alpha_i<1$. Fixing $k, C>0, z_0$, we claim that
Claim 1.

$$\lim_{n\to\infty} ADD = \int_{\substack{0\le\alpha_1,\ldots,\alpha_{k-1}<1 \\ \alpha_1+\cdots+\alpha_{k-1}\le 1}} \frac{1}{[\alpha_1\cdots\alpha_{k-1}(1-\alpha_1-\ldots-\alpha_{k-1})]^{(k-1)/k}}$$

To avoid certain technical difficulties, we consider the following special summation:

$$ADD_\delta = \frac{1}{C^k \log n}\sum\nolimits_{\delta}\Pr[x_1]\cdots\Pr[x_k]$$

where $\sum_\delta$ denotes $\sum_{\substack{x_1+\cdots+x_k=n \\ \delta n<x_1<\cdots<x_k<n}}$

Proving Claim 1 now reduces to proving the following **Claims 2** and **3**:
Claim 2.

$$\lim_{n\to\infty} ADD_\delta = \int_{\substack{\delta\le\alpha_1,\ldots,\alpha_{k-1} \\ \delta\le 1-\alpha_1-\cdots-\alpha_{k-1}}} \frac{1}{[\alpha_1\cdots\alpha_{k-1}(1-\alpha_1-\ldots-\alpha_{k-1})]^{(k-1)/k}} = INT_\delta \quad \text{(say)}$$

Claim 3.

$$\lim_{\delta\to 0}\lim_{n\to\infty} ADD - ADD_\delta = 0.$$

We shall see that Claim 3 follows from computation already done in the previous chapter. But we do the proofs of both the claims in order.

for some constant c.

Remark 2. Our construction shows for $\epsilon \sim e\sqrt{\frac{2}{b}}$,

$$(1-\epsilon)b\log n < r_k(n) < (1+\epsilon)b\log n$$

The proof of the theorem is in two parts. In the first part, we prove that the expected number of representations, in the limit (as $n \to \infty$), approaches $b\log n$; i.e.

$$\mu = E[r_k(n)] \to b\log n \tag{1}$$

In the second part, we show that for $\epsilon > (e-1)\sqrt{\frac{2}{b}}$, and for n large,

$$(1-\epsilon)\mu < r_k(n) < (1+\epsilon)\mu \tag{2}$$

3.2 Getting The Expectation Right

Recall that $\mathcal{S}$ is a random sequence where the probability $p(z)$ of z occuring in $\mathcal{S}$ is given by

$$\begin{aligned} p(z) &= C\frac{(\log z)^{1/k}}{z^{(k-1)/k}} \quad \text{for} \quad z > z_0 \\ &= 0 \qquad\qquad\qquad\quad \text{otherwise.} \end{aligned}$$

where z_0 is the smallest constant such that $p(z) \leq 1/2$. Given ϵ, we choose C so that b is large enough to make $(e-1)\sqrt{\frac{2}{b}} < \epsilon$, where b comes naturally from the following theorem.

We wish to show that

Theorem 3.2

$$\mu \sim b\log n,$$

where $b = C^k \int_{\substack{0\leq\alpha_1,\ldots,\alpha_{k-1}\leq 1 \\ \alpha_1+\cdots+\alpha_{k-1}\leq 1}} \frac{1}{[\alpha_1\cdots\alpha_{k-1}(1-\alpha_1-\ldots-\alpha_{k-1})]^{(k-1)/k}}$

Chapter 3

Integer Sequences II

3.1 The Refinement

The aim of this chapter is to sharpen Theorem 2.1 in an attempt at settling an old conjecture of Erdős. We prove in what follows

Theorem 3.1 *For every fixed k, given $\epsilon > 0$, there exists an asymptotic basis $\mathcal{S}$ of order k such that, for n large,*

$$(1-\epsilon)b\log n < r_k(n) < (1+\epsilon)b\log n$$

where b is an explicit constant.

Erdős conjectured that much more should be true:

Conjecture 3.1 *For every fixed k, there exists an asymptotic basis $\mathcal{S}$ of order k such that, given $\epsilon > 0$, for n large,*

$$(1-\epsilon)c\log n < r_k(n) < (1+\epsilon)c\log n$$

for some constant c.

Remark 1. However, it is still open whether there exists an asymptotic basis such that

$$(c-1)\log n < r_k(n) < (c+1)\log n$$

where $\delta = \frac{1}{\mu}\sum_{i\sim j}\Pr[A_i \wedge A_j] = o(1)$, by Lemma 2.11. We also know, $\mu > (b_1 + o(1))\log n$. Thus

$$\Pr[r_k(n) \leq C_1\mu] \leq e^{-\frac{1}{2(1+o(1))}(1-C_1)^2(b_1+o(1))\log n}$$

Once again, we observe that C_1 can be chosen arbitrarily small, and b_1 can be made > 4, so that

$$\Pr[r_k(n) \leq C_1\mu] \leq n^{-2+o(1)}.$$

Theorem 2.3 now follows by simply applying the Borel-Cantelli lemma.

Recall that $b_1 = D_k C^k$. So by choosing $C > (\frac{3}{D_k})^{1/k}$, we can make $b_1 > 3$. (Note that this choice makes the condition $C > (\frac{1}{3D_k})^{1/k}$ in Lemma 2.7 superfluous.) Thus yielding

$$\Pr[\exists \mathbf{s} : |\ \mathbf{s}\ | \leq C_1 \log n, \mathbf{s} \text{ maximal }] < n^{-2+o(1)}.$$

But this implies

$$\sum_{n=1}^{\infty} \Pr[\exists \mathbf{s} : |\ \mathbf{s}\ | \leq C_1 \log n, \mathbf{s} \text{ maximal }] = \sum_{n=1}^{\infty} n^{-2+o(1)} < \infty$$

allowing us to conclude by the Borel-Cantelli lemma,

$$\text{A.a.} \qquad \exists n_1 \quad \text{s.t.} \qquad r_k(n) > C_1 \log n, \qquad \text{for } n > n_1.$$

Theorems 2.2 and 2.3 together imply the Main Theorem (Theorem 2.1).

Janson's Second Inequality and an alternative proof of Theorem 2.3.

We have realized Theorem 2.3 can be proved with much less effort with a very recent result of Svante Janson. (In particular, Lemma 2.12 becomes unnecessary.) Janson's second inequality [26] is a generalization of his original correlation inequality and can be described as follows. Let $A_1, \ldots, A_n$ be the events as defined in the Correlation Inequality lemma. Further let N denote the (random) number of events in the family $\{A_i\}$ which occur. Define

$$\mu = E[N] = \sum_i \Pr[A_i] \quad \text{and} \quad \delta = \frac{1}{\mu} \sum_{i \sim j} \Pr[A_i \wedge A_j]$$

Theorem J (S. Janson). If $0 \leq \epsilon \leq 1$, then

$$\Pr[N \leq (1-\epsilon)\mu] \leq e^{-\frac{1}{2(1+\delta)}\epsilon^2 \mu}.$$

(Note that Correlation Inequality is the special case when $\epsilon = 1$.)

To prove Theorem 2.3 we first identify A_i to be the event that a valid representation S_i of n is present in our random set $\mathcal{S}$. Then N is precisely $r_k(n)$. By Janson's theorem it would follow, for $0 \leq C_1 \leq 1$

$$\Pr[r_k(n) \leq C_1 \mu] \leq e^{-\frac{1}{2(1+\delta)}(1-C_1)^2 \mu}$$

$$
\begin{aligned}
&= \left[\prod_j (1 - \Pr[B_j])\right] e^{2\sum_{j\sim j'} \Pr[B_j \wedge B_{j'}]} \\
&\leq e^{-\sum_j \Pr[B_j]} e^{2\sum_{j\sim j'} \Pr[B_j \wedge B_{j'}]}
\end{aligned}
$$

With foresight, we have computed the two exponents in Lemmas 2.11 and 2.12 preceding this theorem! $r_k(n;\mathbf{s})$ is precisely the number of the probabilistic events B_j. Thus by Lemma 2.12,

$$\sum_j \Pr[B_j] = E[r_k(n;\mathbf{s})] = \mu' > (b_1 - o(1)) \log n.$$

By Lemma 2.11, the correlation

$$2 \sum_{j\sim j'} \Pr[B_j \wedge B_{j'}] \leq 2 \sum_{i\sim j} \Pr[S_i \wedge S_j] = o(1).$$

Together these facts imply

$$\Pr[\{S_1, \ldots, S_i\} \text{ is maximal }] < e^{-(b_1 - o(1)) \log n}.$$

Plugging this in (**) yields

$$
\begin{aligned}
\Pr[\exists \mathbf{s} : |\, \mathbf{s} \,| \leq C_1 \log n, \mathbf{s} \text{ maximal }] &< \sum_{i=0}^{C_1 \log n} \left(\frac{\mu^i}{i!}\right) e^{-(b_1 - o(1)) \log n} \\
&= e^{-(b_1 - o(1)) \log n} \sum_{i=0}^{C_1 \log n} \frac{\mu^i}{i!} \\
&< e^{-(b_1 - o(1)) \log n} \left(\frac{e}{C_1}\right)^{C_1 \mu}, \quad \text{provided } 0 < C_1 < 1 \\
&< e^{-(b_1 - o(1)) \log n} \left(\frac{e}{C_1}\right)^{C_1 (b_2 + o(1)) \log n}
\end{aligned}
$$

We choose C_1 small enough so that $(\frac{e}{C_1})^{C_1 (b_2 + o(1)) \log n} < e^{\log n}$. This is possible, since $(\frac{e}{C_1})^{C_1} \to 1$ as $C_1 \to 0^+$.

$$
\begin{aligned}
\Pr[\exists \mathbf{s} : |\, \mathbf{s} \,| \leq C_1 \log n, \mathbf{s} \text{ maximal }] &< e^{-(b_1 - o(1)) \log n} e^{\log n} \\
&= n^{-b_1 + 1 + o(1)}
\end{aligned}
$$

We are now set to complete the second part of the main theorem.

Theorem 2.3 *A.a.* $\exists n_1$ *s.t.* $r_k(n) > C_1 \log n, \quad n > n_1.$

Proof. We would first want to show the probability that $r_k^*(n) \leq C_1 \log n$ is *small*, for all n. Then we could complete the proof by using the Borel-Cantelli lemma. Recall that $\mathbf{s}$ denotes a collection of pairwise disjoint representations of n. We want to estimate

$$\begin{aligned}
&\Pr[\exists \mathbf{s} : |\ \mathbf{s}\ | \leq C_1 \log n, \quad \mathbf{s} \text{ maximal }] \\
&= \sum_{i=0}^{C_1 \log n} \Pr[\exists \mathbf{s} \ : |\ \mathbf{s}\ | = i, \quad \mathbf{s} \text{ maximal }] \\
&= \sum_{i=0}^{C_1 \log n} \sum_{\substack{S_1,\ldots,S_i \\ pairwise \\ disjoint}} (\Pr[S_1 \wedge \cdots \wedge S_i] \times \Pr[\{S_1, \ldots, S_i\} \text{ is maximal }]) \\
&= \sum_{i=0}^{C_1 \log n} \left(\sum_{\substack{S_1,\ldots,S_i \\ pairwise \\ disjoint}} \Pr[S_1 \wedge \cdots \wedge S_i] \right) \times \Pr[\{S_1, \ldots, S_i\} \text{ is maximal }] \qquad (**)
\end{aligned}$$

We already know how to estimate the first factor. By the *disjointness lemma*,

$$\sum_{\substack{S_1,\ldots,S_i \\ pairwise \\ disjoint}} \Pr[S_1 \wedge \cdots \wedge S_i] \leq \frac{\mu^i}{i!}$$

We estimate the second factor by the *correlation inequality*! Let B_j be the event that a set $S_j' \subset \mathcal{S} \setminus \mathbf{s}$ of k distinct numbers is a representation of n. Then,

$$\Pr[\{S_1, \ldots, S_i\} \text{ is maximal }] = \Pr[\wedge_j \bar{B}_j]$$

By the Correlation Inequality (Lemma 2.3),

$$\Pr[\wedge_j \bar{B}_j] \quad \leq \quad \left[\prod_j \Pr[\bar{B}_j] \right] e^{2 \sum_{j \sim j'} \Pr[B_j \wedge B_{j'}]}$$

where $\sum^{**}$ denotes the special summation (of Lemma 2.5) with the x_i's restricted to *strips* of width αn; i.e. the summation is over $x_1, \ldots, x_k$ such that

(i) $x_1 + \cdots + x_k = n$ and

(ii) $n\delta_n < x_1 < \alpha n,\ \alpha n < x_2 < 2\alpha n,\ \ldots\ , (k-2)\alpha n < x_{k-1} < (k-1)\alpha n.$

$$\begin{aligned} \mu' &\geq \sum^{**} \Pr[x_1] \cdots \Pr[x_k] - \sum_j \sum_{x_j \in \cup_i S_i}^{**} \Pr[x_1] \cdots \Pr[x_k] \\ &> (b_1 + o(1)) \log n - \sum_j \sum_{x_j \in \cup_i S_i}^{**} \Pr[x_1] \cdots \Pr[x_k] \end{aligned}$$

where the above estimate follows from the estimate (from below) for μ in Lemma 2.5. We would now like to show that for $j = 1, \ldots, k$

$$\sum_{x_j \in \cup_i S_i}^{**} \Pr[x_1] \cdots \Pr[x_k] = o(\log n).$$

Without loss of generality, we show this for x_k (i.e. $j = k$). As $S_1, \ldots, S_{C_1 \log n}$ are pairwise disjoint

$$\begin{aligned} \sum_{x_k \in \cup_i S_i}^{**} \Pr[x_1] \cdots \Pr[x_k] &= (C_1 \log n) \sum_{x_k \in S_1}^{**} \Pr[x_1] \cdots \Pr[x_k] \\ &< (C_1 \log n) \mu_{k-1}(n - x_k) \end{aligned}$$

But our choice of the range for $\sum^{**}$ was such that $x_k < n - (k-1)\alpha n$. Thus

$$\sum_{x_k \in \cup_i S_i}^{**} \Pr[x_1] \cdots \Pr[x_k] = (C_1 \log n)(n^{-1+(k-1)/k+o(1)}) = o(\log n).$$

(In the general case, the range for x_j ensures that $(n - x_j) > (1 - j\alpha)n$, allowing us the estimate $\mu_{k-1}(n - x_j) < n^{-1+(k-1)/k+o(1)}$.) Thus

$$\sum_j \sum_{x_j \in \cup_i S_i}^{**} \Pr[x_1] \cdots \Pr[x_k] = o(\log n)$$

and hence

$$\mu' > (b_1 + o(1)) \log n - o(\log n).$$

On the other hand we have, trivially, $\mu' < \mu < (b_2 + o(1)) \log n$ completing the proof of the lemma.

Similarly,

$$\begin{aligned}\Delta_3 &= \sum_{n/2<m\le n-m_0} \mu_l(m)\,(\mu_{k-l}(n-m))^2 \\ &< \left(n^{-1+l/k+o(1)}\right) \sum_{n/2<m\le n-m_0} \left((n-m)^{-1+\frac{k-l}{k}+\epsilon}\right)^2 \\ &= \left(n^{-1+l/k+o(1)}\right) \sum_{n/2<m\le n-m_0} (n-m)^{-2l/k+2\epsilon}\end{aligned}$$

Estimating by integration over the full range,

$$\begin{aligned}\Delta_3 &< n^{-1+l/k+o(1)} \left(\int_{m-0}^{n} (n-m)^{-2l/k+2\epsilon} + O(1)\right) \\ &< n^{-1+l/k+o(1)} \left(-(n-m)^{1-2l/k+2\epsilon}\ |_0^n + O(1)\right) \\ &= n^{-l/k+2\epsilon+o(1)} = o(1).\end{aligned}$$

Thus,

$$\sum_{|S_i\wedge S_j|=l} \Pr\left[S_i \wedge S_j\right] = \Delta_1+\Delta_2+\Delta_3+\Delta_4 = o(1)$$

and hence the claim,

$$\sum_{i\sim j} \Pr\left[S_i \wedge S_j\right] = \sum_{l=1}^{k-2} o(1) = o(1).$$

Let $\mathbf{s} = \{S_1,\ldots,S_t\}$ be a collection of pairwise disjoint representations of n. We would like to show that, when $|\ \mathbf{s}\ | < C_1 \log n$, deleting sets S_i from $\mathcal{S}$ does not hurt $r_k(n)$ much. For this purpose, set $r_k(n;\mathbf{s}) = |\ \{\{x_1,\ldots,x_k\} : x_1+\cdots+x_k = n; x_i \text{ distinct and } x_i \in \mathcal{S}\setminus \mathbf{s}\}\ |$ and define $\mu' = E[r_k(n;\mathbf{s})]$.

Lemma 2.12

$$|\ \mathbf{s}\ | < C_1 \log n \quad \Rightarrow \quad \mu' = \Theta(\log n).$$

Proof. By definition,

$$\begin{aligned}\mu' &= \sum_{\substack{x_1+\cdots+x_k=n \\ 1<x_1<\cdots<x_k<n \\ x_1,\ldots,x_k\notin\cup_i S_i}} \Pr[x_1]\cdots\Pr[x_k] \\ &> \sum_{x_1,\ldots,x_k\notin\cup_i S_i}^{**} \Pr[x_1]\cdots\Pr[x_k]\end{aligned}$$

Since m_0 is a constant, $\mu_l(m) < M$ (some constant), for $m \leq m_0$. We estimate the above summation in four parts ($\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$) : (i) $m \leq m_0$; (ii) $m_0 < m \leq n/2$; (iii) $n/2 < m \leq n - m_0$, and (iv) $n - m_0 < m$.

Two of the following four estimates are easy; we do the easy ones first.

$$\begin{aligned}
\Delta_1 &= \sum_{m \leq m_0} \mu_l(m) \left(\mu_{k-l}(n-m)\right)^2 \\
&< \left(n^{-1+\frac{k-l}{k}+o(1)}\right)^2 \sum_{m \leq m_0} M \\
&= n^{-2+\frac{2(k-l)}{k}+o(1)} = o(1)
\end{aligned}$$

The other extreme ($m > n - m_0$) can be estimated similarly:

$$\begin{aligned}
\Delta_4 &= \sum_{m > n-m_0} \mu_l(m) \left(\mu_{k-l}(n-m)\right)^2 \\
&< \left(n^{-1+\frac{l}{k}+o(1)}\right) \sum_{m > n-m_0} M^2 \\
&= n^{-1+\frac{l}{k}+o(1)} = o(1)
\end{aligned}$$

Second and third parts of the sum:

$$\begin{aligned}
\Delta_2 &= \sum_{m_0 < m \leq n/2} \mu_l(m) \left(\mu_{k-l}(n-m)\right)^2 \\
&< \left(n^{-1+\frac{k-l}{k}+o(1)}\right)^2 \sum_{m_0 < m \leq n/2} m^{-1+\frac{l}{k}+\epsilon} \\
&= n^{-2l/k+o(1)} \sum_{m_0 < m \leq n/2} m^{-1+\frac{l}{k}+\epsilon}
\end{aligned}$$

It now suffices to estimate the sum by an integral over the full range $0 \leq m \leq n$:

$$\begin{aligned}
\Delta_2 &< n^{-2l/k+o(1)} \left(\int_{m=0}^{n} m^{-1+\frac{l}{k}+\epsilon} + O(1)\right) \\
&= n^{-2l/k+o(1)} \left(m^{l/k+\epsilon} \mid_0^n + O(1)\right) \\
&= n^{-l/k+\epsilon+o(1)} = o(1).
\end{aligned}$$

Let S_i and S_j be two valid representations of n as a sum of k distinct numbers from $\mathcal{S}$. With abuse of notation, we let $\Pr[S_i]$ denote the probability that S_i is a valid representation. We would like to show that these events have *low correlation* in the following precise sense.

Lemma 2.11

$$\sum_{i \sim j} \Pr[S_i \wedge S_j] = o(1).$$

Proof. Note that $i \sim j$ implies S_i and S_j share at least 1 number and at most $k-2$ numbers.

$$\sum_{i \sim j} \Pr[S_i \wedge S_j] = \sum_{l=1}^{k-2} \sum_{|S_i \wedge S_j| = l} \Pr[S_i \wedge S_j]$$

Consider S_i, S_j such that $|\, S_i \wedge S_j \,| = l$. Say,

$$S_i = (z_1, \ldots, z_l, x_1, x_2, \ldots, x_{k-l}) \text{ and } S_j = (z_1, \ldots, z_l, y_1, y_2, \ldots, y_{k-l})$$

Let $\sum_i z_i$ be m. Then $\sum_i x_i = \sum_i y_i = n - m$. So

$$\begin{aligned}
&\sum_{|S_i \wedge S_j| = l} \Pr[S_i \wedge S_j] \\
&= \sum_m \sum_{\substack{z_1 + \cdots + z_l = m \\ x_1 + \cdots + x_{k-l} = n-m \\ y_1 + \cdots + y_{k-l} = n-m}} (\Pr[z_1] \cdots \Pr[z_l])\,(\Pr[x_1] \cdots \Pr[x_{k-l}])\,(\Pr[y_1] \cdots \Pr[y_{k-l}]) \\
&= \sum_m \left(\sum_{z_1 + \cdots + z_l = m} \Pr[z_1] \cdots \Pr[z_l] \right) \left(\sum_{x_1 + \cdots + x_{k-l} = n-m} \Pr[x_1] \cdots \Pr[x_{k-l}] \right)^2 \\
&= \sum_m \mu_l(m)\,(\mu_{k-l}(n-m))^2 .
\end{aligned}$$

Fortunately, we already made (in Lemma 2.8) the estimates $\mu_l(n) < n^{-1+\frac{l}{k}+o(1)}$, for $1 \le l \le k-1$. Fix $\epsilon < \frac{l}{2k}$. Then we can pick m_0 such that

$$\mu_l(m) < m^{-1+\frac{l}{k}+o(1)}, \quad \text{for } m > m_0.$$

Suppose not. i.e. the claim is false for some $n = n'$. Then by the Δ- *system lemma*, there exists a Δ- system $\{S_1^{k-1}, \ldots, S_\Delta^{k-1}\}$ of size $\Delta = c_{max} + 1$. Let the common intersection of the system be R. So $R = \{x_1, \ldots, x_r\}$, where $0 \leq r \leq k-2$. If $\sum_i x_i = m$, then removing the common intersection R from each set will yield $r^*_{k-1-r}(n'-m) \geq c_{max}+1$. This is impossible in view of (*) and the definition of c_{max}. This proves the lemma and, in fact, also shows that $c \leq c_{max}^{k-1}(k-1)!$

We are now ready to prove one side of the main theorem.

Theorem 2.2 *A.a.* $\exists c \exists n_0$ *s.t.* $r_k(n) < (6b_2ck + o(1))\log n, \quad n > n_0.$

Proof. Let $\hat{\mathbf{S}}$ be any maximal collection of pairwise disjoint representations of n as a sum of k distinct numbers. By our notation, $|\hat{\mathbf{S}}| = r_k^*(n)$. By Lemma 2.7a, we have

$$\text{A.a.} \quad \exists n_0 \ \text{ s.t.} \quad r_k^*(n) < 6\mu, \quad n > n_0.$$

This implies there are at most $(k) \times (6\mu)$ numbers in our collection $\hat{\mathbf{S}}$. (There are (almost always) at most 6μ sets in the collection, and each set has exactly k numbers.) As $\hat{\mathbf{S}}$ is maximal, any representation of n must use at least one number from the collection. However, the number of representations of n which use x is precisely $r_{k-1}(n-x)$. By Lemma 2.10 we know $r_{k-1}(n-x) < c$. Thus the total number of representations of n is at most $(6k\mu) \times c$. Hence the theorem follows by substituting the bound for μ.

2.3.2 Second part of the main theorem.

In the first part, we have proved that $r_k(n) < C_2 \log n$, for large n. We now would like to prove that $r_k(n) > C_1 \log n$, for large n. We prove the stronger inequality, that $r_k^*(n) > C_1 \log n$, for large n. Before we can prove this, we need two more lemmas.

Proof.
(i) Note that when S_i^l and S_j^l are disjoint, $\Pr[S_i^l]$ and $\Pr[S_j^l]$ are mutually independent probabilities. So the first part of the lemma follows immediately from the *disjointness lemma.*
(ii) From part (i) it follows that

$$\begin{aligned}\Pr[r_l^*(n) > 2k] &< \frac{\mu_l^{2k}}{(2k)!} \\ &< \frac{1}{(2k)!}\left(n^{-1+l/k+o(1)}\right)^{2k} \\ &= n^{-2k+2l+o(1)}\end{aligned}$$

As $l \leq k-1$, we can conclude

$$\Pr[r_l^*(n) > 2k] < n^{-2+o(1)}.$$

Once again, by the Borel-Cantelli lemma, the above assertion implies that

$$\text{a.a. for } 2 \leq l \leq (k-1) \qquad \exists n_l, \quad \text{s.t.} \quad n > n_l \Rightarrow r_l^*(n) < 2k.$$

But for any finite n_l, there are at most a finite (certainly $< \binom{n_l}{l}$) number of representations as a sum of l numbers. Therefore,

$$\text{a.a. for } 2 \leq l \leq (k-1) \qquad \exists c_l, \text{ s.t.} \quad \forall n, r_l^*(n) < c_l.$$

At this point we remind the reader that we set out to prove that $r_{k-1}(n)$ is at most a constant. And we make critical use of the Δ- *system lemma* (Lemma 2.2) to prove this.
Lemma 2.10

$$A.a. \quad \exists c, \ s.t. \ r_{k-1}(n) < c, \quad \forall n.$$

Proof. We know

$$\text{a.a.} \quad \text{for } 2 \leq l \leq (k-1) \qquad \exists c_l, \text{ s.t.} \quad \forall n, \quad r_l^*(n) < c_l. \tag{*}$$

Set $c_{max} = max_l\{c_l\}$. We claim (whenever c_l's exist),

$$\forall n \qquad r_{k-1}(n) \leq (c_{max})^{k-1}(k-1)!$$

$$
\begin{aligned}
&= \sum{}^{*} C\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k} \cdots C\left(\frac{\log x_l}{x_l^{k-1}}\right)^{1/k} \\
&= n^{o(1)} \sum{}^{*} \frac{1}{(x_1 \cdots x_l)^{(k-1)/k}} \\
&= n^{o(1)} SUM_l
\end{aligned}
$$

As $n/l < x_l$

$$
SUM_l = n^{-(k-1)/k+o(1)} \sum{}^{*} \frac{1}{(x_1 \cdots x_{l-1})^{(k-1)/k}}
$$

And once again we estimate crudely by the summation over all $(x_1, \ldots, x_{l-1})$, with $1 \le x_i \le n,\ i = 1, 2, \ldots, l-1$.

$$
\begin{aligned}
SUM_l &\le n^{-(k-1)/k+o(1)} \sum_{\substack{1 \le x_i \le n \\ i=1,\ldots,l-1}} \frac{1}{(x_1 \cdots x_{l-1})^{(k-1)/k}} \\
&= n^{-(k-1)/k+o(1)} \left(\sum_{1 \le x_1 \le n} \frac{1}{x_1^{(k-1)/k}} \right)^{l-1} \\
&= n^{-(k-1)/k+o(1)} \left(n^{1/k+o(1)} \right)^{l-1} \\
&= n^{-1+\frac{l}{k}+o(1)}
\end{aligned}
$$

Substituting this in the estimate for μ_l above, we get

$$
\begin{aligned}
\mu_l &\le n^{o(1)} (n^{-1+l/k+o(1)}) \\
&= n^{-1+l/k+o(1)}.
\end{aligned}
$$

Let S^l denote a representation of n as a sum of l distinct numbers. Then

Lemma 2.9 *For* $2 \le l \le (k-1)$,

(i)
$$
\sum_{\substack{\{S_1^l, \ldots, S_{2k}^l\} \\ pairwise \\ disjoint}} \Pr\left[S_1^l \wedge \ldots \wedge S_{2k}^l\right] < \frac{\mu_l^{2k}}{(2k)!}
$$

(ii)
$$
\Pr\left[r_l^*(n) > 2k\right] < n^{-2+o(1)}.
$$

Lemma 2.7a. A.a. $\exists n_0$ s.t. $r_k^*(n) \leq 6\mu$, for $n > n_0$.

Proof. Use Borel-Cantelli lemma! First, a few words about the *probability space* we are dealing with. Let Ω denote universal set of all subsequences of the sequence of natural numbers. Our *random* set $\mathcal{S}$ is the subsequence formed by choosing number z with probability $p(z)$. Thus the triple $(\Omega, \mathcal{S}, p(.))$ is our *probability space.* (For a rigorous proof that this is a *valid* probability space , ref. [23], pp. 141-142.)

Now let A_n denote the event $r_k^*(n) > 6\mu$. By Lemma 2.7, we have

$$\sum_{n=1}^{\infty} \Pr(A_n) = \sum_{n=1}^{\infty} n^{-2+o(1)} < \infty$$

By the Borel-Cantelli lemma, this implies that with probability 1, at most a *finite* number of A_n occur. Hence the lemma.

To prove the first part of the main theorem, we must extend Lemma 2.7a and show that $r_k(n)$ is $O(\log n)$. For this purpose we need a bound (of $O(1)$) on the number of representations of n as a sum of $k-1$ (distinct) numbers. (The puzzled reader is advised to skip the next few lemmas and read the proof of Theorem 2.2 for further motivation.) But we need to generalise some of our notation before we can proceed. Let $r_l(n)$ denote the number of representations of n as a sum of l distinct numbers from $\mathcal{S}$. And let $r_l^*(n)$ denote the size of a maximal collection of *pairwise disjoint* such representations. Further, let μ_l denote the expectation of $r_l(n)$. (note μ_k is simply μ)

Lemma 2.8 *For* $2 \leq l \leq (k-1)$,

$$\mu_l \leq n^{-1+\frac{l}{k}+o(1)}$$

Proof. The line of calculation is similar to that for μ, only simpler. Let l be fixed. Let $\sum^*$ denote $\sum_{\substack{x_1+\cdots+x_l=n \\ 1\leq x_1<\cdots<x_l<n}}$. Then by definition,

$$\mu_l = \sum_{\substack{x_1+\cdots+x_l=n \\ 1\leq x_1<\cdots<x_l<n}} \Pr[x_1]\cdots\Pr[x_l]$$

We say S_1 and S_2 are *disjoint* if they share no element in common. The following lemmas guarantee (with high probability) that a maximal collection of pairwise disjoint representations has size at most $O(\log n)$.

Lemma 2.6

$$\sum_{\substack{\{S_1,\ldots,S_{6\mu}\}\\ pairwise\\ disjoint}} \Pr[S_1 \wedge \cdots \wedge S_{6\mu}] \;\leq\; \frac{\mu^{6\mu}}{(6\mu)!}$$

Proof. Pairwise disjointness of the sets implies mutual independence of the associated events. The proof is hence a simple application of the *disjointness lemma* (Lemma 2.1) with $l = 6\mu$.

This in turn bounds the probability of having more than 6μ pairwise disjoint representations of n as follows. Let $r_k^*(n)$ denote the size of a maximal collection of pairwise disjoint representations. Then

Lemma 2.7 *For* $C > \left(\frac{1}{3D_k}\right)^{1/k}$,

$$\Pr\left[r_k^*(n) > 6\mu\right] \leq n^{-2+o(1)}.$$

Proof. Using Lemma 2.4 and the bound $m! > m^m e^{-m}$

$$\begin{aligned}
\Pr\left[r_k^*(n) > 6\mu\right] &< \frac{1}{(6\mu)!}(\mu)^{6\mu} \\
&< \frac{1}{(6\mu/e)^{6\mu}}(\mu)^{6\mu} \\
&= \left(\frac{e}{6}\right)^{6\mu} \\
&< \left(\frac{e}{6}\right)^{(6b_1+o(1))\log n} \\
&= n^{-6b_1+o(1)}.
\end{aligned}$$

To complete the proof of the lemma, we have to show that $b_1 \geq 1/3$. As $b_1 = D_k C^k$, we pick C large enough (so that $D_k C^k > 1/3$) to make $b_1 > 1/3$, yielding

$$\Pr\left[r_k^*(n) > 6\mu\right] \leq n^{-2+o(1)}.$$

$$
\begin{aligned}
sum &< \frac{1}{(n/k)^{(k-1)/k}} \sum_{1\le x_1\le n\delta_n}^{*} \frac{1}{(x_1\cdots x_{k-1})^{(k-1)/k}} \\
&< \frac{1}{(n/k)^{(k-1)/k}} \sum_{\substack{1\le x_1\le n\delta_n \\ 1\le x_i\le n \\ i=2,\ldots,k-1}} \frac{1}{(x_1\cdots x_{k-1})^{(k-1)/k}} \\
&= \left(\frac{k}{n}\right)^{(k-1)/k} \left(\sum_{1\le x_1\le n\delta_n} \frac{1}{(x_1)^{(k-1)/k}}\right) \left(\sum_{1\le x_2\le n} \frac{1}{(x_2)^{(k-1)/k}}\right)^{k-2} \\
&= \left(\frac{k}{n}\right)^{(k-1)/k} \left(\int_1^{n\delta_n} \frac{1}{x_1^{(k-1)/k}} dx_1 + O(1)\right) \left(\int_1^{n} \frac{1}{x_2^{(k-1)/k}} dx_2 + O(1)\right)^{k-2} \\
&= \left(\frac{k}{n}\right)^{(k-1)/k} \left(k x_1^{1/k} \mid_1^{n\delta_n} + O(1)\right) \left(k x_2^{1/k} \mid_1^{n} + O(1)\right)^{k-2} \\
&= \frac{k^{(k-1)(k+1)/k}}{n^{(k-1)/k}} \left(\delta_n^{1/k} n^{(k-1)/k} + o(\delta_n^{1/k} n^{(k-1)/k})\right) \\
sum &= O(\delta_n^{1/k})
\end{aligned}
$$

This shows

$$
\begin{aligned}
\mu_2 &= C^k (\log n) O(\delta_n^{1/k}) \\
&= o(\log n).
\end{aligned}
$$

Thus we have established,

$$
(b_1 + o(1)) \log n < \mu_1 < (b_2 + o(1)) \log n \quad \text{and } \mu_2 = o(\log n)
$$

and hence

$$
(b_1 + o(1)) \log n < \mu < (b_2 + o(1)) \log n
$$

where $b_1 = D_k C^k$ and $b_2 = C^k k^{(k-1)(k+1)/k}$.

2.3.1 First part of the main theorem.

Let $S_1 = \{a_1, \ldots, a_k\}$ and $S_2 = \{b_1, \ldots, b_k\}$ be two different representations of n in basis $\mathcal{S}$; that is, $S_1 \neq S_2, \;\; S_1, \; S_2 \subset \mathcal{S}$ and

$$
a_1 + \ldots + a_k = b_1 + \ldots + b_k = n.
$$

In the other direction, set $\alpha = 1/k(k-1)$. As $x_k < n$, we bound SUM from below by

$$SUM \quad > \quad \frac{1}{n^{(k-1)/k}} \sum_{n\delta_n < x_1 < n}^{*} \frac{1}{(x_1 \cdots x_{k-1})^{(k-1)/k}}$$

For a *lower bound* on this summation, our idea is to restrict the x_i's to *strips* of width αn. i.e. we bound the summation from below by summation over $(x_1, \ldots, x_{k-1})$, with restrictions $n\delta_n < x_1 < \alpha n$, $\alpha n < x_2 < 2\alpha n$, $\ldots$, $(k-2)\alpha n < x_{k-1} < (k-1)\alpha n$. (Our choice of α ensures that $x_{k-1} < x_k$.) Thus

$$\begin{aligned}
SUM \quad > \quad & \frac{1}{n^{(k-1)/k}} \sum_{n\delta_n < x_1 < \alpha n} \frac{1}{x_1^{(k-1)/k}} \quad \cdots \sum_{(k-2)\alpha n < x_{k-1} < (k-1)\alpha n} \frac{1}{x_{k-1}^{(k-1)/k}} \\
& \text{and once again, estimating by integrals} \\
= \quad & \frac{1}{n^{(k-1)/k}} \left(\int_{n\delta_n}^{\alpha n} \frac{1}{x_1^{(k-1)/k}} + O(1) \right) \cdots \left(\int_{(k-2)\alpha n}^{(k-1)\alpha n} \frac{1}{x_{k-1}^{(k-1)/k}} + O(1) \right) \\
= \quad & \frac{1}{n^{(k-1)/k}} \left(k x_1^{1/k} \, |_{n\delta_n}^{\alpha n} + O(1) \right) \cdots \left(k x_{k-1}^{1/k} \, |_{(k-2)\alpha n}^{(k-1)\alpha n} + O(1) \right) \\
= \quad & (1+o(1)) k^{k-1} \alpha^{(k-1)/k} \left[1 - (\alpha\delta_n)^{1/k} \right] \left[2^{1/k} - 1 \right] \cdots \left[(k-1)^{1/k} - (k-2)^{1/k} \right] \\
= \quad & D_k + o(1)
\end{aligned}$$

where $D_k = \left[2^{1/k} - 1\right] \left[3^{1/k} - 2^{1/k}\right] \cdots \left[(k-1)^{1/k} - (k-2)^{1/k}\right] \left(\frac{k^{k-1}}{k-1}\right)^{(k-1)/k}$, a positive absolute constant!

Consider now the second part:

$$\begin{aligned}
\mu_2 \quad = \quad & C'^k \sum_{1 < x_1 < n\delta_n}^{*} \frac{((\log x_1) \cdots (\log x_k))^{1/k}}{(x_1 \cdots x_k)^{(k-1)/k}} \\
< \quad & C'^k (\log n) \sum_{1 < x_1 < n\delta_n}^{*} \frac{1}{(x_1 \cdots x_k)^{(k-1)/k}} \\
= \quad & C'^k (\log n).(sum) \qquad \text{(say)}
\end{aligned}$$

It now suffices to show that $sum = o(1)$, since we are trying to prove $\mu_2 = o(\log n)$. The calculation is similar to what we did for SUM and proceeds as follows.

sufficiently slowly as $n \to \infty$.) We prove the lemma by showing that $\mu_1 = \Theta(\log n)$ and that $\mu_2 = o(\log n)$. For this purpose, let $\delta_n = 1/(\log n)$. Then

$$\begin{aligned}
\mu_1 &= C^k \sum_{n\delta_n < x_1 < n}^{*} \frac{((\log x_1)\cdots(\log x_k))^{1/k}}{(x_1 \cdots x_k)^{(k-1)/k}} \\
&= C^k(1+o(1))(\log n) \sum_{n\delta_n < x_1 < n}^{*} \frac{1}{(x_1 \cdots x_k)^{(k-1)/k}} \\
&= C^k(1+o(1))(\log n).(SUM) \quad \text{(say)}
\end{aligned}$$

where the numerator comes out as a common term, in view of the range of summation. It now suffices to show that $SUM = \Theta(1)$. Note that $n/k < x_k < n$, since x_i's form an increasing sequence. As $n/k < x_k$, we bound SUM from above by

$$SUM < \frac{1}{(n/k)^{(k-1)/k}} \sum_{n\delta_n < x_1 < n}^{*} \frac{1}{(x_1 \cdots x_{k-1})^{(k-1)/k}}$$

We can bound this tricky summation from above by summation over all $(x_1, \ldots, x_{k-1})$, with $1 \le x_i \le n$. That is,

$$\begin{aligned}
SUM &< \frac{1}{(n/k)^{(k-1)/k}} \sum_{\substack{1 \le x_i \le n \\ i=1,\ldots,k-1}} \frac{1}{(x_1 \cdots x_{k-1})^{(k-1)/k}} \\
&= \left(\frac{k}{n}\right)^{(k-1)/k} \left(\sum_{1 \le x_1 \le n} \frac{1}{(x_1)^{(k-1)/k}} \right)^{k-1}
\end{aligned}$$

We now estimate this discrete sum by integrating over the same range:

$$\begin{aligned}
SUM &< \left(\frac{k}{n}\right)^{(k-1)/k} \left(\int_1^n \frac{1}{x_1^{(k-1)/k}} dx_1 + O(1) \right)^{k-1} \\
&= \left(\frac{k}{n}\right)^{(k-1)/k} \left(k x_1^{1/k} \,|_1^n + O(1) \right)^{k-1} \\
&= \frac{k^{(k-1)(k+1)/k}}{n^{(k-1)/k}} \left(n^{(k-1)/k} + o(n^{(k-1)/k}) \right) \\
&= (1+o(1))k^{(k-1)(k+1)/k}.
\end{aligned}$$

2.3 Proof of the Main Theorem

Let k be fixed. Our aim is to show that there exists an asymptotic basis of order k and that the number of representations is roughly $\log n$ for n *large*. Consider a sequence $\mathcal{S}$ of natural numbers where the probability $p(z)$ of z occuring in $\mathcal{S}$ is given by

$$\begin{aligned} p(z) &= C\frac{(\log z)^{1/k}}{z^{(k-1)/k}} \qquad \text{for } z > z_0 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

where z_0 is the smallest constant such that $p(z) \leq 1/2$. We shall choose C so that $C^k D_k > 3$, where
$D_k = \left[2^{1/k}-1\right]\left[3^{1/k}-2^{1/k}\right]\cdots\left[(k-1)^{1/k}-(k-2)^{1/k}\right]\left(\frac{k^k-1}{k-1}\right)^{(k-1)/k}$, a *not-so-nice* positive absolute constant!

Let $r_k(n)$ be the random variable denoting the number of representations of n as a sum of k distinct numbers from $\mathcal{S}$. Then

Lemma 2.5

$$\mu = E\left[r_k(n)\right] = \Theta(\log n)$$

Proof.

$$\begin{aligned} \mu &= \sum_{\substack{x_1+\cdots+x_k=n \\ 1\leq x_1<\cdots<x_k<n}} \Pr[x_1]\cdots\Pr[x_k] \\ &= \sum{}^{*} C\left(\frac{\log x_1}{x_1^{k-1}}\right)^{1/k}\cdots C\left(\frac{\log x_k}{x_k^{k-1}}\right)^{1/k} \end{aligned}$$

where $\sum^*$ is *short hand* for $\displaystyle\sum_{\substack{x_1+\cdots+x_k=n \\ 1\leq x_1<\cdots<x_k<n}}$. We estimate the above sum by breaking it into two parts $(\mu_1+\mu_2)$, depending on whether $x_1 > n/(\log n)$ or not. (It may be noted that "$\log n$" here may be replaced by any positive function that approaches infinity

The proof and a more general version of the theorem (for infinite set systems) can be found in [16].

Correlation Inequality.
In a probability space, we know that the probability of a (finite) conjunction of events is exactly the product of the individual probabilities, provided that the events are *mutually independent*. In a similar spirit, the following *Correlation Inequality* allows us to estimate the probability of a conjunction by the product of the probabilities, provided that the events have *low correlation*! We need some notation to state the lemma precisely.

Let Ω be a finite set of elements, and S be a random subset of Ω given by $\Pr[x \in S] = p_x$, where we assume these events to be mutually independent over $x \in \Omega$. Let A_i denote the event $X_i \subset S$, where $X_1, \ldots, X_n$ are subsets of the universal set Ω. Write $i \sim j$ if $i \neq j$ and $X_i \cap X_j \neq \phi$. Further let all $\Pr[A_i] \leq 1/2$, then

Lemma 2.3 (Correlation Inequality)

$$\prod \Pr[\bar{A}_i] \leq \Pr\left[\bigwedge \bar{A}_i\right] \leq \left[\prod \Pr[\bar{A}_i]\right] e^{2\sum_{i\sim j} \Pr[A_i \cap A_j]}.$$

The proof requires only elementary probability theory and appears in [4]. The original proof of the result was given by Svante Janson and appears in [27].

Borel-Cantelli Lemma.
The last tool we make use of in this paper is one of great importance in probability theory. We merely state the well known Borel-Cantelli lemma in the following; the proof can be found in [23] and also in any standard book on probability theory.

Lemma 2.4 (Borel-Cantelli). *Let $\{A_j\}$ be a sequence of events in a probability space. If*

$$\sum_{j=1}^{\infty} \Pr(A_j) < \infty,$$

then with probability 1, at most a finite number of the events A_j can occur.

2.2 Probabilistic Tools

Disjointness Lemma.

Let $A_1, A_2, \ldots$ be events in a probability space. The following elementary lemma is often useful in proving the probability of occurrence of "large" subsets of mutually independent events is "small". We refer to this as *disjointness lemma* in view of the nature of its application in this paper.

Lemma 2.1 (disjointness lemma) *If* $\sum_i \Pr[A_i] \leq \mu$, *then*

$$\sum_{\substack{\{A_1,\ldots,A_l\} \\ mutually \\ independent}} \Pr[A_1 \wedge A_2 \wedge \cdots \wedge A_l] \leq \frac{\mu^l}{l!}$$

Proof. Let $\sum'$ denote the summation in the statement of the lemma. Then by mutual independence of the probabilities,

$$\begin{aligned}
\sum{}' \Pr[A_1 \wedge \cdots \wedge A_l] &= \sum{}' \Pr[A_1] \cdots \Pr[A_l] \\
&\leq \sum_{all\ \{A_1,\ldots,A_l\}} \Pr[A_1] \cdots \Pr[A_l] \\
&= \frac{1}{l!} \sum_{(A_1,\ldots,A_l)} \Pr[A_1] \cdots \Pr[A_l] \\
&= \frac{1}{l!} \left(\sum_i \Pr[A_i] \right)^l \\
&= \frac{1}{l!} \mu^l .
\end{aligned}$$

Erdős - Rado Δ -system.

Let n and l be positive integers, $l \geq 3$. We say that l sets form a $\Delta - system$ (of size l), if they have pairwise the same intersection. Let $\mathcal{F}$ be a family of sets, each of size n. Then

Lemma 2.2 (Δ- system lemma) $\mathcal{F}$ *contains a* Δ- *system of size* l *whenever*

$$| \mathcal{F} | \geq (l-1)^n n!$$

Chapter 2

Integer Sequences I

2.1 Introduction

Let $\mathcal{S}$ be a sequence of natural numbers, and let $r_k(n)$ denote the number of representations of n in the form

$$n = a_1 + a_2 + \cdots + a_k, \quad 0 < a_1 < \ldots < a_k \ \ (a_i \in \mathcal{S})$$

$\mathcal{S}$ is said to be an *asymptotic basis* of order k if there exists a natural number n_0 such that $r_k(n) > 0$ for $n > n_0$.

In this chapter we prove the following Main Theorem.

Theorem 2.1 *For every fixed k there exists an asymptotic basis of order k such that $r_k(n) = \Theta(\log n)$.*

We shall actually conclude that in a suitable probability (measure) space almost all sequences are of *the right kind.*

In the following section we review some basic techniques we make use of from probability theory.

known exponential bound due to Israeli and Jalfon [25]. The second part of Chapter 4 describes the solution for the meeting time problem.

Discrepancy is a combinatorial problem that arises in various contexts such as communication complexity in multi-party protocols, quasi-randomness in hypergraphs, tournament ranking, edge-discrepancy and the Gale-Berlekamp switching game. In Chapter 5, we introduce a very natural and interesting generalization (to higher dimensions) of the switching game.

In k-dimensions, the game consists of an $n \times \cdots \times n$ array A (of size n^k) of lights. There are kn switches, each controlling a hyperplane of n^{k-1} lights. i.e. each switch depends only on one dimension of the array and we have k types for k dimensions. Each switch when thrown changes each light in its control from *on* to *off* or from *off* to *on*. *Discrepancy* is defined as the maximum difference between the number of lights on and off that one can obtain, given any initial configuration for the lights. It follows from a result of Erdős and Spencer that discrepancy $= \Theta\left(n^{(k+1)/2}\right)$. This result, however, is *existential* due to its critical use of *randomness*. An optimal deterministic sequential and an NC algorithm are given in this paper. The basic idea is to use Spencer's method of conditional expectation to remove randomness. But the crucial part of the derandomization is in justifying the efficient computation of the conditional expectations. For the NC algorithm, we prove that 4-wise independent random setting for the switches works!

The second part of the thesis describes results on random walks on finite graphs. Random walks in graphs have been found to have interesting analogies in resistor networks. Doyle and Snell made an excellent development of this topic in [13]. A *simple* random walk on a graph is one in which from any vertex of the graph there is equal probability of moving to a neighboring vertex. Given a graph, the underlying electrical network is the network obtained by replacing vertices by nodes and edges by electrical resistances. The *effective resistance* between any two nodes x and y can be defined as the voltage that develops between x and y when a unit current is maintained through them (i.e. enters one and leaves the other).

In Chapter 4 we present precise characterization of effective resistance in electrical networks in terms of random walks on the underlying graphs. Our interpretation of effective resistance enables us in giving new results and interesting proofs for some known results. In particular, the interpretation of Polya's recurrence theorem in terms of electrical networks becomes obvious. We also obtain a precise version of the triangle inequality for effective resistances. This, in turn, is used in characterizing the *one-way hitting times* between two vertices in terms of effective resistances of the network. (much in the spirit of the *commute time* result of [7]). This settles the open question – whether it is possible to capture the one-way hitting times *exactly* in terms of resistive networks!

The hitting time result is further used in solving the *meeting time* problem, that arose in the study of *self-stabilizing systems*. This is joint work with Peter Winkler. The *meeting time* problem, posed by Israeli and Jalfon [25], can be stated as follows: Suppose two (or more, in general) tokens are placed on an n-vertex undirected graph G, and at each time-step a certain adversary decides which of them is to make the next move. Each move is one step of a simple random walk. The adversary is trying to keep the tokens apart as long as possible. What is the expected time before the two tokens meet? We show that the meeting time is bounded by $\frac{8}{27}n^3$, improving the previously

constants c_1 and c_2 such that

$$c_1 \log n \leq r_2(n) \leq c_2 \log n \qquad \text{for large } n.$$

which more than answered the question. For a thorough exposition of the above result and several other results of similar nature, the reader is strongly recommended to refer to [23].

In Chapter 1 we prove a more general version of Sidon's original question using probabilistic techniques, old and new. We prove that for every fixed k, there exists an asymptotic basis of order k such that the number of representations of n is $\Theta(\log n)$. The main theorem of Chapter 1 was jointly done with Paul Erdős [20].

In Chapter 2 we consider a refinement and some extensions of our results in Chapter 1. A long standing conjecture (worth \$ 1000) of Paul Erdős asserts that

$\forall k\ \exists$ constant $C\ \exists$ asymp. basis $\mathcal{S}\ \forall \epsilon\ \exists n_0$,

$$n > n_0 \Rightarrow \qquad (1-\epsilon)C \log n < r_k(n) < (1+\epsilon)C \log n$$

We succeed in proving the following weaker version of the conjecture

$\forall k\ \forall \epsilon$ constant $\exists C\ \exists$ asymp. basis $\mathcal{S}\ \exists n_0$,

$$n > n_0 \Rightarrow \qquad (1-\epsilon)C \log n < r_k(n) < (1+\epsilon)C \log n$$

The construction is (as usual) *random*, but the proofs are (unusually) simple due to the use of *correlation inequalities*.

Moreover, we show that the Main Theorem can be extended in the following sense. It is possible to construct asymptotic bases with the representation function $r_k(n) >> \log n$. We show that the same proof techniques work in proving that
for fixed k, if $g(n) >> \log n$, monotone increasing, and "smooth" then there exists an asymptotic basis such that $r_k(n) \sim g(n)$ (where the notion of "smoothness" is made precise in Chapter 2).

Chapter 1

Introduction

The first part of the thesis concerns the construction of integer sequences with certain additive properties. Where (deterministic) construction seems impossible, the standard alternative approach is to try techniques from probability theory (i.e. resort to *randomness*) and show the *existence* of such sequences. In the study of sequences, it is often of interest to arrive at possible relationships between the behaviour of the representation function and the rate of growth of the sequences in question. Many theorems of this flavor on the behaviour of the representation function in the addition of sequences have been obtained by Erdős [15] and Erdős–Rényi [17]. The main idea here being to impose a probability measure on the space of integer sequences, such that almost all integer sequences have some prescribed rate of growth. This makes the representation function easier to investigate without further reference to the actual rate of growth of the sequences. In addition, the proofs require only basic results from probability theory.

A set of natural numbers is called an *asymptotic basis* of order k if every number (sufficiently large) can be expressed as a sum of k distinct numbers from the set. The requirement is that every number n (sufficiently large) have at least one representation. We denote by, $r_k(n)$, the number of such representations of n. In 1932 S. Sidon raised the following question: Does there exist an asymptotic basis of order 2 such that, for every $\epsilon > 0$, $\lim_{n\to\infty} \frac{r_2(n)}{n^\epsilon} = 0$? In 1956 Erdős [15] proved that there exist positive

List of Figures

4.1 Reciprocity 48

5.1 Introduction 58
5.2 The Generalization of the Switching Game 59
5.3 The Probability Lemmas 62
5.4 The Randomized Algorithm 63
5.5 Derandomization 65
5.6 Discrepancy in Other Contexts 72

Contents

1 Introduction **1**

2 Integer Sequences I **5**

2.1 Introduction 5

2.2 Probabilistic Tools 6

2.3 Proof of the Main Theorem 8

2.3.1 First part of the main theorem. 11

2.3.2 Second part of the main theorem. 16

3 Integer Sequences II **25**

3.1 The Refinement 25

3.2 Getting The Expectation Right 26

3.3 The Second Part 35

3.4 Some Extensions 38

4 Hitting Times and the Meeting Time **44**

4.1 Simple Networks 44

4.2 General Networks 51

4.3 The Meeting Time Problem 52

5 The Gale-Berlekamp Switching Game **58**

Acknowledgments

Joel Spencer got me interested in the probabilistic method with an excellent course in the spring of 1988. My thesis work would not have been possible without his insight, inspiration and guidance. I would like thank Bud Mishra for the encouragement and the support he gave me during my initial *aimless* years.

I am grateful to Paul Erdős for sharing his ideas with me. The first chapter is joint work with him and essentially got me started on my dissertation. The hitting time problem discussed in the fourth chapter has its roots in some conversations with Prabhakar Raghavan. Thanks to Peter Winkler for "sharing" the meeting time problem with me. It is always a lot of fun working with Peter. I am also thankful to Fan Chung and Ravi Boppana for many useful discussions.

I thank all my friends for continuing to be my friends despite having to put up with me during the not-so-easy times. Special thanks are due Sundareswaran for helping me out with printing the thesis. Finally, I am especially grateful to my wonderful parents, to whom I owe the most.

known results. The main result here is a formula for the *one-way transit times* (otherwise known as "hitting times") between two vertices in terms of effective resistances in the network. This settles the open question– whether it is possible to capture hitting times *exactly* in terms of resistive networks!

Results on Discrepancy Problems.

Discrepancy is a combinatorial problem that arises in various contexts such as communication complexity in multi-party protocols, *quasi-randomness* in graphs and hypergraphs, hypergraph coloring, and the Gale-Berlekamp switching game. In k–dimensions, the game consists of an $n \times \cdots \times n$ array A (of size n^k) of lights. There are kn switches, each controlling a hyperplane of n^{k-1} lights. Each switch when thrown changes each light in its control from *on* to *off* or from *off* to *on*. *Discrepancy* is defined as the maximum difference between the number of lights on and off that one can obtain, given any initial configuration for the lights. It follows from a result of Erdős and Spencer that discrepancy $= \Theta\left(n^{(k+1)/2}\right)$. This result, however, is *existential* due to its critical use of *randomness*. A deterministic sequential and an NC algorithm are given in this thesis. The basic idea is to use Spencer's method of conditional expectation to remove randomness. But the crucial part of the derandomization is in justifying the efficient computation of the conditional expectations. For the NC algorithm, we prove that 4-wise independent random setting for the switches works!

Applications and Analysis of Probabilistic Techniques

Prasad V. Tetali

Abstract

The topic of the thesis falls under the scope of *Randomness, Quasi-randomness, and Derandomization.*

Results on Integer Sequences.

The first part of the thesis is concerned with the construction of integer sequences with certain additive properties. Where (deterministic) construction seems impossible, the standard alternative approach is to try techniques from probability theory (i.e. resort to *randomness*) and show the existence of such sequences.

Janson's *Correlation Inequalities* allow us to estimate the probability of a conjunction of events (not necessarily independent) by the product of the probabilities, provided that the events have certain *quasi-independence*. In a joint work with Paul Erdős, I have used these and combinatorial techniques in proving the following number-theoretic result. Let us call a set of natural numbers an *asymptotic basis* of order k, if every number (sufficiently large) can be expressed as a sum of k distinct numbers from the set. We proved that for every fixed k, there exists an asymptotic basis S of order k such that the number of representations $r_k(n)$ of n is $\Theta(\log n)$.

Results on Random Walks on Graphs.

Random walks on graphs have been known to have interesting analogies in resistor networks. A precise characterization of *effective resistance* in electrical networks is provided in this thesis in terms of random walks on the underlying graphs. The interpretation of effective resistance yields interesting new results and new proofs for some

Author's Current Address

Department of Computer Science
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street, New York, NY 10012 USA
E-mail Address : tetali@spunky.cs.nyu.edu

© Copyright by Prasad V. Tetali 1991
All Rights Reserved

Applications and Analysis of Probabilistic Techniques

Prasad V. Tetali

October 1991

A dissertation in the Department of Computer Science submitted to the faculty of the Graduate School of Arts and Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University.

Approved

(Signed) _____

Joel Spencer

Research Advisor

© Copyright by Prasad V. Tetali 1991
All Rights Reserved