$$\frac{\varepsilon}{2}\sum a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum b_j(x)\frac{\partial u}{\partial x_j} = 0 \quad \text{in } G,$$

$$u(x) = f(x) \quad \text{on } \delta G$$

as $\varepsilon \to 0$. Of course the main interest is when the drift term is pointing inward at the boundary of G. These methods are also used to study the asymptotic behaviour of the first eigenvalue of the operator in the above equation in the region G as $\varepsilon \to 0$.

Chapters 16 and 17 again deal with the inhomogeneous case. They are concerned with some optimization and minimax problems. In Chapter 16 a suitable functional of stopping times is to be optimized over all stopping times. A description of the optimal stopping time is given in terms of solutions of certain variational inequalities. When it is a minimax type problem involving two stopping times, existence of a saddle point is proved and a description of it is given. In Chapter 17 the drift coefficients $\{b_j(t, x)\}$ depend on parameters which can be controlled by different players. The question of existence of a saddle point is studied for the resulting stochastic differential game.

The methods use the theory of quasilinear parabolic differential equations.

The volumes cover a wide variety of results proved under varying assumptions. It may sometimes be hard for a nonexpert to get a feeling for the conditions in terms of their role in establishing the results. There are some misprints, on occasion even in the statement of the theorems.

These volumes contain a considerable amount of detailed information that is quite important in the study of diffusion processes. The methods should prove useful in studying other problems as well. There are a lot of exercises to make the reader familiar with the ideas developed in the text. Complete references are given to other related works, where some of the material can be found. Most of the results have been obtained by the author himself in recent years.

S. R. S. VARADHAN

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 82, Number 6, November 1976

Sieve methods, by H. H. Halberstam and H.-E. Richert, Academic Press, London, 1974, xiii + 364 pp., $26.00.

The modern theory of sieve method has developed gradually, with fits and starts, over the past sixty years. From the outset the literature was hard to read because of the complicated nature of the arguments, while in recent times many of the most important results have remained unpublished, making it almost impossible to be well informed. Moreover, the literature has become tangled, and fragmented by a lack of unifying perspective. Expository accounts of the subject have usually been restricted to specific aspects, and in many cases even these have made difficult reading. An exception to this is found in Halberstam and Roth [**6**, Chapter 4], where the general nature and scope of sieve methods are lucidly exposed, although the more elaborate devices and applications are omitted.

Sieve methods is the first exhaustive account of this important topic. In the past, researchers have generally derived the sieve bounds required for an application, but now workers will find that usually an appeal to an appropriate theorem of *Sieve methods* will suffice. In addition to presenting the theoretical basis of various sifting techniques, many applications are described in detail. Each chapter concludes with detailed Notes describing the history of the results, the relevant literature, and the interrelationships between the various ideas. The book ends with a very useful Bibliography, in which papers are listed separately under the headings: Theoretical contributions, Surveys, Methodological developments, Applications, Extensions to algebraic domains, Variants, and Aids. Finally bibliographic references are provided for three hundred eightly-six papers.

When sifting, we start with a finite set $\mathcal{A}$ of integers, and then remove from $\mathcal{A}$ those elements which lie in one or more of a certain collection of arithmetic progressions. In the sieve of Eratosthenes, we begin with the integers in the interval $(1, x)$, and remove all those which are divisible by p for some prime $p \leqslant x^{1/2}$. Then the numbers remaining are precisely the primes in the interval $(x^{1/2}, x]$. The theoretical question is to estimate the number of integers remaining after a sifting process has been completed.

Consider the simple situation in which we have the integers n, $1 \leqslant n \leqslant x$, and we wish to know how many n have the property that $2\nmid n$, $3\nmid n$. To this end we subtract from $[x] = x + O(1)$ the number $x/2 + O(1)$ of integers n for which $2|n$, and we subtract the number $x/3 + O(1)$ of integers n for which $3|n$. The numbers n for which both $2|n$ and $3|n$ have been deleted twice, so we add back the number $x/6 + O(1)$ of such n, to obtain the answer, $x/3 + O(1)$. Obviously our procedure could be formulated in terms of Sylvester cross-classification, but the combinatorics can be rather simply effected by using the Möbius μ-function.

Let $\Omega(n)$ denote the number of primes dividing n, and put $\mu(n) = (-1)^{\Omega(n)}$ for square-free n, $\mu(n) = 0$ if n has a repeated prime factor. The property which characterizes μ is that

$$\sigma_0(n) =: \sum_{d|n} \mu(d) = \begin{cases} 1, & n = 1, \\ 0, & n > 1, \end{cases} \tag{1}$$

where the sum is extended over all positive divisors d of n. In the previous paragraph we counted n for which $(n, 6) = 1$. By (1),

$$\begin{aligned} \sum_{n \leqslant x;(n,6)=1} 1 &= \sum_{n \leqslant x} \sum_{d|n;d|6} \mu(d) = \sum_{d|6} \mu(d) \sum_{n \leqslant x; d|n} 1 \\ &= \sum_{d|6} \mu(d)\left(\frac{x}{d} + O(1)\right) = x\sum_{d|6} \frac{\mu(d)}{d} + O(1) \\ &= x/3 + O(1). \end{aligned}$$

Replacing 6 by a number k, we find by the same argument that

$$S =: \sum_{n \leqslant x;(n,k)=1} 1 = x \sum_{d|k} \frac{\mu(d)}{d} + O\Big(\sum_{d|k} 1\Big).$$

The first sum on the right is

$$= \prod_{p|k}\left(1 - \frac{1}{p}\right) = \frac{\phi(k)}{k},$$

while the sum in the error term is the divisor function $d(k)$. Thus

$$S = \phi(k)x/k + O(d(k)). \tag{2}$$

The sieve of Eratosthenes suggests setting $k = \prod_{p \leqslant x^{1/2}} p$, but then $d(k)$ in the error term is too large. In fact, if $k = P(z)$,

$$P(z) = \prod_{p<z} p, \tag{3}$$

then the error term in (2) is smaller than the main term only for

$$z < c(\log x)(\log \log x).$$

Thus (2), which is a modern embodiment of the sieve of Eratosthenes-Legendre, does not allow us to sift very far. Such was the situation until 1915, when Viggo Brun initiated the modern era of sieve methods. Brun's work was fraught with complicated details, but the spirit of the approach can be simply explained. The shortcoming of (2) is that it can be applied only when k has a small number of prime divisors; let us restrict our attention to divisors of k which have at most m prime factors. From the simple identity

$$\sum_{r=0}^{m} (-1)^r \binom{k}{r} = (-1)^m \binom{k-1}{m}$$

we see that $\sum_{d|n;\Omega(d) \leqslant m} \mu(d)$ is either smaller or larger than $\sigma_0(n)$, according as m is odd or even. Accordingly, for $i = 1, 2$ put

$$\sigma_i(n) = \sum_{d|n;\Omega(d) < 2m+i-1} \mu(d)$$

so that $\sigma_2(n) \leqslant \sigma_0(n) \leqslant \sigma_1(n)$ for all n. Arguing as we did to obtain (2), we find that

$$\begin{aligned} x \sum_{d|k;\Omega(d)<2m} \frac{\mu(d)}{d} + O\Big(\sum_{d|k;\Omega(d)<2m} 1\Big) &\leqslant S \\ \leqslant x \sum_{d|k;\Omega(d)<2m+1} \frac{\mu(d)}{d} &+ O\Big(\sum_{d|k;\Omega(d)<2m+1} 1\Big). \end{aligned} \tag{4}$$

This is the "pure Brun" sieve; to appreciate its sharpness we must (as frequently happens in sieve theory) estimate the various sums which arise. This done, we find that (4) is preferable to (2), for the error term in (4) is smaller, while the main term is essentially the same (pp. 46–52). In the case that $k = P(z)$, we can derive from (4) an asymptotic estimate for S provided $z \leqslant x^{1/(8 \log \log x)}$. An asymptotic estimate for S when z is of restricted size is known as a "fundamental lemma" (pp. 82–89, 204–213). In addition to being useful in its own right, a fundamental lemma can also serve as a starting point for more elaborate sifting arguments.

Although (4) is no longer considered very useful, the nature of the combinatorial sieve (Chapter 2) is now evident. For $i = 1, 2$ let $\chi_i(n)$ be the characteristic function of a (carefully chosen) set of integers; we put

$$\sigma_i(n) = \sum_{d|n} \mu(d)\chi_i(d), \tag{5}$$

and we require that

$$\sigma_2(n) \leqslant \sigma_0(n) \leqslant \sigma_1(n) \tag{6}$$

for all n. By our basic argument,

$$\begin{aligned} x\sum_{d|k} \frac{\mu(d)\chi_2(d)}{d} + O\Big(\sum_{d|k}\chi_2(d)\Big) &\leqslant S \\ &\leqslant x\sum_{d|k} \frac{\mu(d)\chi_1(d)}{d} + O\Big(\sum_{d|k}\chi_1(d)\Big); \end{aligned} \tag{7}$$

we wish to arrange that the error terms be small, while the sums in the main terms should be approximately

$$\sum_{d|k} \frac{\mu(d)}{d} = \prod_{p|k}\left(1 - \frac{1}{p}\right) = \frac{\phi(k)}{k}.$$

By further exploiting the parity idea, Brun constructed good sifting functions σ_i. In the 1950's J. Barkley Rosser developed another combinatorial sieve, which in some cases gives very sharp bounds. Although Rosser never published his work on sieves, the principles of his method are now known (Selberg [**13**, pp. 89–90]), and the details have been derived in certain cases Iwaniec [**8**], [**9**].

Once the form of the combinatorial sieve is realized, it is clear that our sifting functions $\sigma_i(n)$ need not be of the very special form (5). Rather, it would be enough to have

$$\sigma_i(n) = \sum_{d|n} \lambda_i(n) \tag{8}$$

for some function λ_i. Of course this observation is useful only if we can also arrange for (6) to hold, and to obtain useful estimates. In 1947 Atle Selberg had the beautiful idea of taking

$$\sigma_2(n) = \Big(\sum_{d|n} \Lambda_d\Big)^2,$$

with Λ_d real-valued, $\Lambda_1 = 1$. Then $\sigma_2(1) = 1$, and $\sigma_2(n) \geqslant 0$ for all n, so (6) holds. This corresponds to taking

$$\lambda_2(d) = \sum_{[r,s]=d} \Lambda_r\Lambda_s$$

in (8). In place of (7) we now have

$$S \leqslant x \sum_{d|k} \frac{\lambda_2(d)}{d} + \sum_{d|k} |\lambda_2(d)|. \tag{9}$$

The main term is a quadratic form in the Λ_r, subject to the linear constraint $\Lambda_1 = 1$. Supposing that $\Lambda_r = 0$ for $r \geqslant z$, this form can be minimized exactly, giving a value $\leqslant (\log z)^{-1}$ in the case that $k = P(z)$. Moreover, the extremal λ_2 satisfy $|\lambda_2(n)| \leqslant 1$ for all n, and $\lambda_2(n) = 0$ for $n \geqslant z^2$, so that

$$S \leqslant x/(\log z) + z^2.$$

Taking $z = x^{1/2}(\log x)^{-1}$, we find that

$$\pi(x) \leqslant (2 + \varepsilon)x/(\log x).$$

Of course the prime number theorem asserts more, $\pi(x) \sim x/(\log x)$, but the sieve argument (unlike analytic methods) applies equally well to any interval of length x, so that

$$\pi(x + y) - \pi(y) \leqslant (2 + \varepsilon)x/(\log x) \tag{10}$$

for $x \geqslant 2$, $y \geqslant 0$. This estimate, but with a large undetermined constant in place of $(2 + \varepsilon)$, was first derived by Hardy and Littlewood as an application of Brun's method. If we were to begin by considering consecutive integers $n \equiv a \pmod q$, we would find that

$$\pi(x + y; q, a) - \pi(y; q, a) \leqslant \frac{(2 + \varepsilon)x}{\phi(q)\log(x/q)}$$

for $(a, q) = 1$, $x \geqslant 2$, $y \geqslant 0$. This very important bound is known as the Brun-Titchmarsh inequality, although the sharp constant $2 + \varepsilon$ is obtained by Selberg's method (or the large sieve; see Montgomery and Vaughan **[10]**). Selberg's method readily yields good upper bounds in other, more general, sieve problems (Chapters 3, 4, 5).

As it stands, Selberg's method provides only upper bounds. Nevertheless, Selberg did obtain lower bounds (Chapters 6, 7, 8), by using his upper bound in conjunction with a further combinatorial argument. The combinatorial argument can again be couched in terms of Sylvester's cross-classification; but in sieve theory we appeal to "Buchstab's identity". The terminology is somewhat misleading, for in the first place several such identities exist, any one of which may be referred to as "Buchstab's identity"; and secondly Buchstab was not the first to discover them. The idea of these identities goes back to Legendre, and they are found in papers of Brun and of Rademacher. Nevertheless, the label is appropriate, for A. A. Buchstab, starting in 1937, was the first to use these identities systematically and effectively in sieve theory.

Let $l(n)$ denote the least prime dividing n. We may partition integers n according to the value of $l(n)$; the n for which $(n, P(z)) = 1$ are those for which $l(n) \geqslant z$. Then

$$S(x, z) = [x] - \sum_{p<z} |\{n \leqslant x : l(n) = p\}|,$$

where $S(x, z) = |\{n \leqslant x : (n, P(z)) = 1\}|$. As for the summand on the right, if $l(n) = p$ then $n = pm$, $l(m) \geqslant p$; that is, $(m, P(p)) = 1$. Hence,

$$S(x, z) = [x] - \sum_{p \leqslant z} S\left(\frac{x}{p}, p\right). \tag{11}$$

This is Buchstab's identity, and the path to a lower bound is now clear: We obtain an upper bound for $S(x/p, p)$, which, inserted in (11), gives a lower bound for $S(x, z)$. Such a lower bound can be inserted in (11) to produce a new upper bound, which in many cases is sharper than the original upper bound. This process of improving a sieve by iterative use of (11) is of great importance in the present state of sieve methods (Chapter 8).

For purposes of describing sieve methods, we have restricted our attention to the problem of estimating the size of $S = |\{n \leqslant x\colon (n, k) = 1\}|$. In fact the methods are much more general. The only use which we have made of the fact that $\mathcal{Q} = \{n\colon 1 \leqslant n \leqslant x\}$ is to say that

$$|\{n \in \mathcal{Q}\colon d|n\}| = x/d + O(1). \tag{12}$$

In fact, for purposes of sieving, it is enough to know that

$$|\{n \in \mathcal{Q}\colon d|n\}| = X\omega_0(d)/d + r_d \tag{13}$$

for square-free d, where $\omega_0(d)$ is multiplicative and satisfies certain regularity conditions, and $|r_d|$ is not too large on average. Moreover, rather than sift by each prime, we may choose a set $\mathcal{P}$ of sifting primes, and put

$$P(z) = \prod_{p < z; p \in \mathcal{P}} p, \tag{14}$$

generalizing (3). Then we require (12) only for d composed of primes in $\mathcal{P}$, and we may estimate $S(\mathcal{Q}, \mathcal{P}, z) = |\{n \in \mathcal{Q}\colon (n, P(z)) = 1\}|$. For example, if our interest is in twin primes, then we may take

$$\mathcal{Q} = \{n(n + 2)\colon 1 \leqslant n \leqslant X\},$$

$$\omega_0(p) = \begin{cases} 1, & p = 2, \\ 2, & p > 2; \end{cases}$$

we let $\mathcal{P}$ be the set of primes, and we find that $|r_d| \leqslant \omega_0(d)$.By Selberg's upper bound method we then find that the number of primes $p \leqslant X$ for which $p + 2$ is also prime does not exceed $(8 + \varepsilon)cX(\log x)^{-2}$, where $c = 2\Pi_{p>2}(1 - (p - 1)^{-2})$. The best known bound is with 4 in place of 8; the conjectured size is with 1 in place of 8. The weakness of our lower bound method is that we obtain positive lower bounds only when z is of restricted size, say $z = X^c$, $c < c_0 < \frac{1}{2}$. Nevertheless, in 1920 Brun showed that one can solve $P_9 + 2 = P_9$. That is, there are infinitely many integers n having at most 9 prime factors, for which $n + 2$ also has at most 9 prime factors. The subscripts have undergone many improvements. In particular, Alfred Rényi was the first to show that $p + 2 = P_k$ has solutions for some large fixed k; Bombieri's mean value theorem of 1965 allows one to take $k = 3$. In 1966, J. Chen announced that every large integer $2N$ may be written in the form $2N = p + P_2$. Little was thought of this until after the end of the cultural revolution, when Chen [4] published his proof (Chapter 11; see also Ross

[12]). In the main, *Sieve methods* develops the body of knowledge concerning sieves which has accumulated from approaches to the twin prime problem and related problems.

In order to sift, we need little information concerning $\mathcal{A}$ beyond (13) and some knowledge of ω_0. This makes sieve methods very general and flexible, but it also makes it possible to construct artificial examples which place limitations on the sharpness of our sieve bounds. Following Selberg (p. 239), instead of the natural choice $\mathcal{A} = \{n\colon 1 \leqslant n \leqslant x\}$ we consider the set $\mathcal{A}^\circ$ of integers n, $1 \leqslant n \leqslant x$, for which n has an odd number of prime factors. Pertaining to $\mathcal{A}$ we have (12), while for $\mathcal{A}^\circ$ we have

$$|\{n \in \mathcal{A}^\circ\colon d|n\}| = \frac{x}{2d} + O\left(\frac{x}{d}\exp\left(-\left(\log\frac{x}{d}\right)^{1/2}\right)\right);$$

this is (13) with $X = x/2$ instead of $X = x$. However, if (3) holds with $z > x^{1/3}$, then $S(\mathcal{A}^\circ, \mathcal{P}, z)$ counts precisely those p for which $z \leqslant p \leqslant x$. Thus

$$S(\mathcal{A}^\circ, \mathcal{P}, z) = \pi(x) - \pi(z^-) \sim x/(\log x),$$

so that the sieve bound

$$S(\mathcal{A}^\circ, \mathcal{P}, z) \leqslant (1 + o(1))X/(\log z)$$

is asymptotically best possible when $z = x^{1/2}$. Hence the *method* used to obtain (10) is best possible, although it is not known whether (10) itself might be improved. By similarly considering integers with an even number of prime factors it may be shown that the iterated lower bound method is best possible in certain cases. In light of this, it seems that if sieve methods are to succeed in treating Goldbach's problem, deeper use must be made of the properties of the set $\mathcal{A}$ which is being sifted.

In 1941, P. Kuhn discovered that sieve results may be improved by counting integers $n \in \mathcal{A}$ with certain weights (Chapters 9, 10). These weights, which often take the form $1 - \Sigma_{p|n} w_p$, tend to obviate the example of Selberg which we have just discussed. The determination of good values for the w_p gives rise to complicated extremal problems; in certain cases the choice

$$\omega_p = \begin{cases} \lambda(1 - (\log p)/(\log z_1)) & \text{if } z \leqslant p < z_1, \\ 0 & \text{otherwise,} \end{cases}$$

is indicated. Here λ and $z_1 > z$ are parameters whose values remain to be determined.

We have now outlined the main topics dealt with in *Sieve methods*. In closing we remark upon related topics, subjects of current interest, and recent additions to the literature. A fine discussion of the large sieve has been given by Bombieri [1], who also makes some interesting calculations concerning weighted sieves. The larger sieve of Gallagher [5] is not treated in *Sieve methods*, nor are questions of sifting by a thin set of primes (in this connection see the new book of Hooley [7]). The papers of Iwaniec [8], [9] and Selberg [13] concerning Rosser's method deserve further study. In particular, Iwaniec [9] has used the $\frac{1}{2}$-dimensional sieve to obtain sharp new results concerning numbers which are representable as a sum of two squares. Selberg **[14]** has shed further light on the relationship between the large sieve and the Selberg sieve. Diamond and Jurkat (unpublished) have extended the analysis of the iterated Selberg sieve to dimension $\kappa \neq 1$ (see also Porter **[11]**). Bombieri **[2]**, **[3]** has had some innovative ideas concerning weighted sieves. Vaughan **[15]** has given a simple proof of a sharp form of Bombieri's mean value theorem.

For years to come, *Sieve methods* will be vital to those seeking to work in the subject, and also to those seeking to make applications. The heavy notation in the book seems to be essential in formulating such general methods. Some parts of the book are much more difficult to read than others, but generally the text is lively and conversational. In concept and execution this is an excellent, long-needed work.

REFERENCES

1. E. Bombieri, *Le grand crible dans la théorie analytique des nombres*, Astérisque No. 18, Société Mathématique de France, Paris, 1974. MR **51** #8057.

2. ———, *On twin almost primes*, Acta Arith. **28** (1975), 177–193; corrigendum, ibid., **28** (1976), 760–763.

3. ———, *The asymptotic sieve* (to appear).

4. J. Chen, *On the representation of a large even integer as the sum of a prime and a product of two primes*, Sci. Sinica **16** (1973), 157–176.

5. P. X. Gallagher, *A larger sieve*, Acta Arith. **18** (1971), 77–81. MR **45** #214.

6. H. Halberstam and K. F. Roth, *Sequences*. I, Clarendon Press, Oxford, 1966. MR **35** #1565.

7. C. Hooley, *Applications of sieve methods to the theory of numbers*, Cambridge Tracts in Math., no. 70, Cambridge Univ. Press, London and New York, 1976.

8. H. Iwaniec, *On the error term in the linear sieve*, Acta Arith. **19** (1971), 1–30. MR **45** #5104.

9. ———, *The half dimensional sieve*, Acta Arith. **29** (1976), 69–95.

10. H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika **20** (1973), 119–134.

11. J. W. Porter, *An improvement of the upper and lower bound functions of Ankeny and Onishi*, Acta Arith. (to appear).

12. P. M. Ross, *On Chen's theorem that each large even number has the form $p_1 + p_2$ or $p_1 + p_2p_3$*, J. London Math. Soc. (2) **10** (1975), 500–506.

13. A. Selberg, *Sieve methods*, Proc. Sympos. Pure Math., vol 20, Amer. Math. Soc., Providence, R. I., 1971, pp. 311–351. MR **47** #3286.

14. ———, *Remarks on sieves*, Proc. Number Theory Conf. (Boulder, Colo., 1972), Univ. of Colorado, 1972, pp. 205–216. MR **50** #4457.

15. R. C. Vaughan, *Mean value theorems in prime number theory*, J. London Math. Soc. (2) **10** (1975), 153–162.

H. L. MONTGOMERY

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 82, Number 6, November 1976

Fourier series with respect to general orthogonal systems, by A. M. Olevskiĭ (translated by B. P. Marshall and H. J. Christoffers), Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 86, Springer-Verlag, Berlin, Heidelberg, New York, 1975, viii + 136 pp., $33.60.

Fourier series–the original Fourier series, that is, the ones using trigonometric functions–were the first series of orthogonal functions. They are either