BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 40, Number 1, Pages 109–119
S 0273-0979(02)00957-6
Article electronically published on October 17, 2002

Sieves in number theory, by George Greaves, Springer-Verlag, New York, 2001, xii+304 pp., 94.95 euros, ISBN 3-540-41647-1

The book under review gives a comprehensive account of the Rosser-Iwaniec method, the most important development in the construction of number sieves since the advent in 1947 of Selberg's λ-method. Sieve literature has grown prodigiously since the publication of [HR], and Dr. Greaves has had to make some difficult decisions on what to include and what to omit. On the whole his choices have been wise; students and experts alike will have much to learn from his careful presentation. If the book does not always make for easy reading, that is due largely to the nature of the subject: not only does sieve architecture rest on complicated combinatorial foundations, but these culminate nowadays in none too easy linear differential delay boundary value problems and also link up with results and techniques from modern analytic number theory.

We begin with a brief description of sieve methods by setting the stage: Let $\mathcal{P}$ be a finite set of primes—usually this is an infinite, increasing sequence of primes truncated at some number $z > 2$—and refer to $\mathcal{P}$ as a 'sieve'. Let P denote the product of all the primes in $\mathcal{P}$. In Selberg's terminology, $\mathcal{P}$ is said to 'sift out' an integer n if n is divisible by some prime p in $\mathcal{P}$. Then, writing (n, P) for the highest common factor of n and P, $\mathcal{P}$ sifts out n if and only if $(n, P) > 1$; by the same token, the indicator function of all integers n that are *not* sifted out by $\mathcal{P}$ is

$$\sum_{d|(n,P)} \mu(d) = \begin{cases} 1 & \text{when } (n,P) = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(Here the Moebius function $\mu(d)$ is 1 at $d = 1$, $(-1)^\nu$ when d is squarefree and the product of ν primes, and 0 otherwise. Also, $d \mid m$ means that d is a factor of m.)

Next, let $\mathcal{A}$ be a finite integer sequence and suppose it is such that the 'probability' that a prime p from $\mathcal{P}$ divides an element a of $\mathcal{A}$ is $\rho(p)/p$, where $0 \le \rho(p) < p$ and $\rho(p) = 0$ when p is not in $\mathcal{P}$. Writing $\mathcal{A}_d = \{a \in \mathcal{A} : d \mid a\}$ and $\rho(d) = \prod_{p|d} \rho(p)$, we make this a little more precise by postulating the existence of an approximation X to the cardinality $|\mathcal{A}|$ of $\mathcal{A}$ such that the quantities

$$r_{\mathcal{A}}(d) := |\mathcal{A}_d| - \frac{\rho(d)}{d} X, \quad d \mid P,$$

are small, perhaps only on average (of some kind) over the divisors d of P not exceeding some parameter D. (The reader might find it helpful to think of $\mathcal{A}$ as the sequence of values taken on an interval of length X by an irreducible polynomial $h(\bullet)$ with integer coefficients and having no fixed prime divisors; in this case $\rho(p)$ is the number of $\bmod p$ incongruent solutions of the congruence $h(x) \equiv 0 \bmod p$.)

The simplest problem in any sieve method is to estimate as accurately as possible the number $S(\mathcal{A}, P)$ of elements in $\mathcal{A}$ that are not sifted out by $\mathcal{P}$. From what has

2000 *Mathematics Subject Classification.* Primary 11N35, 11N36.

been said,

$$\begin{aligned} S(\mathcal{A}, P) &= |\{a \in \mathcal{A} : (a, P) = 1\}| = \sum_{a\in\mathcal{A}} \Bigg(\sum_{d|(a,P)} \mu(d) \Bigg) \\ &= \sum_{d|P} \mu(d)|\mathcal{A}_d| = X \sum_{d|P} \mu(d)\frac{\rho(d)}{d} + \sum_{d|P} \mu(d) r_{\mathcal{A}}(d) \\ &= XV(P) + R \qquad (2) \end{aligned}$$

say, where

$$V(P) = \sum_{d|P} \mu(d)\rho(d)/d = \prod_{p\in\mathcal{P}} (1 - \rho(p)/p).$$

This beguiling result, often referred to as the Eratosthenes-Legendre formula, appears at first sight to be the end of the story; actually, it is barely the beginning. To see this, consider the classical case first studied by Eratosthenes in the third century BC and reformulated two millennia later by Legendre, of $\mathcal{A} = \{n \in \mathbb{N} : 1 \le n \le X\}$ and $\mathcal{P} = \{p : p \le \sqrt{X}\}$. Here $|\mathcal{A}_d| = \llcorner X/d \lrcorner = X/d + (\llcorner X/d \lrcorner - X/d)$, so that

$$\pi(X) - \pi(\sqrt{X}) + 1 = S(\mathcal{A}, P) = X \prod_{p\le\sqrt{X}} \left(1 - \frac{1}{p}\right) + \sum_{d|P} \mu(d)(\llcorner X/d \lrcorner - X/d),$$

where $\pi(y)$ denotes, as usual, the number of primes not exceeding y. The leading term on the right is not asymptotic to $X/\log X$ as $X \to \infty$, as one might have hoped from knowledge of the Prime Number Theorem, but to a quantity larger by a constant factor. Thus the 'remainder' sum R is actually of the same order of magnitude as the leading term. Indeed, a little reflection shows the formula of Eratasthenes-Legendre to be a direct application of the inclusion-exclusion principle, which is known to be of practical use only when the number of 'properties' to be excluded is very small—here this is the number $\pi(\sqrt{X})$ of primes not exceeding $\sqrt{X}$, and it is much too large.

In the literature one finds mostly the need for three kinds of estimates of $S(\mathcal{A}, P)$:

$$\text{(I)} \qquad S(\mathcal{A}, P) \ll XV(P);$$

this, the simplest and also the most frequent application, occurs in auxiliary situations, as when one wants to show that some subset consisting of integers that have no small prime factors is of a relatively negligible size.

$$\text{(II)} \qquad S(\mathcal{A}, P) \sim XV(P);$$

such an asymptotic result, consonant with what naive probabilistic reasoning suggests, is valid only when $\mathcal{P}$ is very sparse or when $(\log|\mathcal{P}|)/\log X \to 0$ as $X \to \infty$. Nevertheless, it too has many applications and goes under the name of 'a fundamental lemma', coined long ago by Kubilius in connection with his study of the distribution of values of additive arithmetic functions.

$$\text{(III)} \qquad S(\mathcal{A}, P) > 0 \quad \text{or} \quad S(\mathcal{A}, P) \to \infty \quad \text{as} \quad X \to \infty;$$

such statements assert that $\mathcal{A}$ contains an element, or many elements, having no prime factors from $\mathcal{P}$ and, therefore, subject to a proper choice of $\mathcal{P}$, only very few large ones. Most of the approximations we have to the classical conjectures of prime number theory come under this heading, although for the best approximations currently known one needs additional optimising procedures. For example, in the particular case to which we referred earlier, it is known that if g is the degree of

$h(\bullet)$, then $h(n)$ is infinitely often an 'almost-prime' P_{g+1}—a number having at most $g+1$ prime factors [R].

So, to devise a *practical* sieve method one looks for an effective approximation to the indicator function (1). Let

$$\sum_{d|(n,P)} \chi(d)\mu(d)$$

be such an approximation, where $\chi(1) = 1$, but otherwise the values taken by the 'modifying' factor $\chi(\bullet)$, all real, remain to be chosen. Writing $q(d)$ for the least prime factor of an integer $d > 1$ and $q(1) = \infty$, associate with χ the 'complementary' function $\bar{\chi}(\bullet)$ given by

$$\bar{\chi}(1) = 0, \quad \bar{\chi}(d) = \chi(d/q(d)) - \chi(d) \quad (d > 1)$$

(there is a typo in this statement on p. 76). To link χ and $\bar{\chi}$ we have

The Fundamental Sieve Identity. *Let $\psi(\bullet)$ be any arithmetic function and n a squarefree natural number. Then*

$$\sum_{d|n} \mu(d)\psi(d) = \sum_{d|n} \mu(d)\psi(d)\chi(d) + \sum_{d|n} \mu(d)\bar{\chi}(d) \sum_{\substack{t|n \\ Q(t)<q(d)}} \mu(t)\psi(dt),$$

where $Q(t)$ denotes the largest prime factor of $t > 1$ and $Q(1) = 1$. In particular, when ψ is multiplicative we have

$$\sum_{d|n} \mu(d)\psi(d) = \sum_{d|n} \mu(d)\psi(d)\chi(d) + \sum_{d|n} \mu(d)\psi(d)\bar{\chi}(d) \prod_{\substack{p|n \\ p<q(d)}} (1 - \psi(p)),$$

and hence, when $\psi(\bullet) = 1$ identically,

$$\sum_{d|n} \mu(d) = \sum_{d|n} \mu(d)\chi(d) + \sum_{\substack{d|n \\ q(d)=q(n)}} \mu(d)\bar{\chi}(d). \tag{3}$$

Let $\nu(d)$ denote the number of prime factors of a squarefree number d. Brun's first choice of χ was

$$\chi(d) = \begin{cases} 1, & \nu(d) \le k \\ 0 & \text{otherwise,} \end{cases}$$

where k is a positive integer. Here $\bar{\chi}(d) = 1$ if and only if $\nu(d) = k+1$, so that, by (3)

$$\sum_{d|n} \mu(d) = \sum_{\substack{d|n \\ \nu(d)\le k}} \mu(d) + (-1)^{k+1} \sum_{\substack{d|n, \nu(d)=k+1 \\ q(d)=q(n)}} 1; \tag{3$'$}$$

in particular, if k is even and ℓ odd, we have

$$\sum_{\substack{d|n \\ \nu(d)\le \ell}} \mu(d) \le \sum_{d|n} \mu(d) \le \sum_{\substack{d|n \\ \nu(d)\le k}} \mu(d). \tag{4}$$

(In probability theory, these are known as Bonferroni's inequalities.) Brun used the right-hand inequality, with a suitable choice of k, to obtain the first non-trivial piece of information about prime twins, namely that

$$\sum_{\substack{p \\ p+2=p'}} 1/p < \infty;$$

he accomplished this by estimating $\pi_2(x) := |\{p \le x : p+2 = p'\}|$ from above, but his estimate, though good enough for this purpose, contained an unwanted factor $(\log\log x)^2$. A few years ago Hooley ([H1]; see also [FH]) showed how to remove this blemish by what seems with hindsight to have been the natural generalization of Brun's so-called 'pure' sieve: Let

$$\mathcal{P} = \bigcup_{j=1}^{r} \mathcal{P}_j \quad (r \ge 2), \quad P_j = \prod_{p \in \mathcal{P}_j} p,$$

be any partition of $\mathcal{P}$ (so that $\mathcal{P}_i \cap \mathcal{P}_j = \phi$ when $i \ne j$). Then by the right-hand inequality in (4),

$$\sum_{d|(a,P)} \mu(d) = \prod_{j=1}^{r} \sum_{d|(a,P_j)} \mu(d) \le \prod_{j=1}^{r} \sum_{\substack{d|(a,P_j) \\ \nu(d) \le k_j}} \mu(d) \tag{5}$$

for any choice of r *even* integers k_j. Perhaps Brun did not miss this idea but did not follow it up because it seemed not to possess a natural analogue of the left-hand inequality in (4); instead he devised a more complicated set of constraints on the prime decompositions of the divisors d_j of P_j, also in terms of the counting numbers $\nu(d_j)$, that, while it lacks the multiplicative structure of (5), does possess a symmetric lower counterpart, as in (4). This, the famous Brun Sieve method held the stage for more than thirty years and, after a slow start, played a role in many interesting investigations. It is now, as Greaves remarks, largely of historical interest (for a full account see e.g. [HR] or [S]). What Greaves does instead is to present in Chapter 3 a preliminary version of the Rosser-Iwaniec method that is weaker but simpler and constitutes a good introduction to it. He gives a reference to Hooley's method but no details. There is, in fact, a lower counterpart to (5) that rests on the almost trivial inequalities

$$y_1 \cdots y_r - \sum_{\ell=1}^{r} (y_\ell - x_\ell) \prod_{\substack{j=1 \\ j \ne \ell}}^{r} y_j \le x_1 \cdots x_r \le y_1 \cdots y_r$$
$$\text{whenever} \quad 0 \le x_j \le y_j \ (j = 1, \ldots, r);$$

just take

$$x_j = \sum_{d|(a,P_j)} \mu(d) \quad \text{and} \quad y_j = \sum_{\substack{d|(a,P_j) \\ \nu(d) \le k_j}} \mu(d) \quad (j = 1, \ldots, r)$$

and apply (3′) to the terms $y_\ell - x_\ell$. This, the Brun-Hooley method, is technically much simpler than Brun's, and indeed than any other known method—which, as Hooley has shown [H2], can be important in applications—and it is about as accurate as Brun's. It is a versatile and accessible tool that will see much use in the future and could simplify accounts of many past applications.

For most practical purposes Brun's method was superceded by Selberg's powerful and elegant λ-method, that saw the light of day in a short note dated 1947 and rests on the daringly simple observation that

$$\sum_{d|n} \mu(d) \leq \left(\sum_{d|n} \lambda_d \right)^2$$

for any choice of real numbers λ_d so long as $\lambda_1 = 1$. There are now many books that give fine presentations of Selberg's method, but Greaves' account of it in Chapters 2 and 7 contains some interesting and original refinements.

The above inequality leads to an upper estimate of $S(\mathcal{A}, P)$ only, but here too there is a readily available lower counterpart. To describe it we first make the slight change of notation mentioned earlier—from now on, and for the rest of this review, think of $\mathcal{P}$ as an infinite increasing sequence of primes and write $P(z) = \prod_{p \in P, p < z} p$. Now we go back to (3) and make the apparently trivial choice $\chi(d) = 1$ at $d = 1$ and $\chi(d) = 0$ otherwise; then $\bar{\chi}(d) = 1$ when d is prime and is otherwise also 0, and (3) leads directly to Buchstab's identity

$$S(\mathcal{A}, P(z)) = |\mathcal{A}| - \sum_{\substack{p \in \mathcal{P} \\ p < z}} S(\mathcal{A}_p, P(p)). \tag{6}$$

Obviously, applying *upper* bounds in the sum on the right will lead to a *lower* bound for $S(\mathcal{A}, P(z))$. There are a good many technical difficulties on the way, but these were successfully overcome in an important paper [AO] in 1964 by Ankeny and Onishi, who gave there, among other things, a first general lower bound method based on Selberg's sieve method and (6).

Hard upon the heels of [AO] there appeared the paper [JR] of Jurkat and Richert in which Selberg's sieve is combined with an ingeniously chosen infinite iteration of (6) to obtain upper and lower estimates of $S(\mathcal{A}, P(z))$ for all problems in which $\rho(p)$ is, in a very weak sense, about equal to 1 on average. (This is the case in the problem involving the irreducible polynomial $h(\cdot)$ cited earlier, where $\rho(p)$ is 1 on average by the prime ideal theorem.) For this class of 'linear' problems a pair of important examples of Selberg (see Chapter 4.5.1) show the estimates of [JR] to be best possible, but it should be said that [AO] is only very slightly weaker than [JR] in the case of linear problems. A full account of the methods of [AO] and [JR], with many applications, are to be found in [HR].

But [JR] gains added importance from the fact that it inspired Iwaniec, then a young research worker in Warsaw, to develop an alternative procedure, slightly sharper and also more flexible than that in [JR], and it is a careful, accurate and systematic account of Iwaniec's method that is the principal achievement of this treatise. It would transpire later that Rosser had discovered much the same method much earlier but had published no details—I know of only two research announcements, in volumes 43 (p. 173) and 47 (p. 383) of the AMS *Bulletin* (the latter jointly with W. J. Harrington), and Le Veque quoted a 'fundamental lemma' version in one of his early papers. So one speaks nowadays of the Rosser-Iwaniec sieve (R-I, for short), but it is Iwaniec who has developed, refined and applied it over the years, alone or in partnership with others, in a veritable stream of important papers.

To sketch the combinatorial basis of the Rosser-Iwaniec method we go back to (3) and look for a choice of χ that is superior to Brun's and incorporates the outcome of the iterations of (6) in [JR]—Davenport (and probably others, too) remarked once in conversation that whenever iteration works there is usually at bottom a better approach. Substituting from (3) in (2) we have

$$\begin{aligned} S(\mathcal{A}, P(z)) &= \sum_{a\in\mathcal{A}} \Bigg(\sum_{d|(a,P(z))} \mu(d)\chi(d) + \sum_{\substack{d|(a,P(z))\\ q(d)=q((a,P(z)))}} \mu(d)\bar{\chi}(d) \Bigg) \\ &= \sum_{d|P(z)} \mu(d)\chi(d)|\mathcal{A}_d| + \sum_{d|P(z)} \mu(d)\bar{\chi}(d) S(\mathcal{A}_d, P(q(d)) \\ &= S_1 + S_2, \end{aligned} \tag{7}$$

say. Here

$$\begin{aligned} S_1 &= X \sum_{d|P(z)} \mu(d)\chi(d)\rho(d)/d + \sum_{d|P(z)} \mu(d)\chi(d) r_{\mathcal{A}}(d) \\ &= X \Bigg(V(P(z)) - \sum_{d|P(z)} \mu(d)\bar{\chi}(d) \frac{\rho(d)}{d} V(P(q(d))\Bigg) + \sum_{d|P(z)} \mu(d)\chi(d) r_{\mathcal{A}}(d) \end{aligned} \tag{8}$$

by the Fundamental Sieve Identity with $\psi(d) = \rho(d)/d$; and

$$S_2 = \sum_{\substack{d|P(z)\\ \mu(d)=1}} \bar{\chi}(d) S(\mathcal{A}_d, P(q(d))) - \sum_{\substack{d|P(z)\\ \mu(d)=-1}} \bar{\chi}(d) S(\mathcal{A}_d, P(q(d))) = S_{21} - S_{22}\,. \tag{9}$$

We suppose from now on that $\chi(\bullet)$ assumes only the values 0 and 1 and is *divisor-closed* in the sense that when $\chi(d) = 1$ and $t \mid d$, then also $\chi(t) = 1$. It follows at once that $\bar{\chi}(d)$ also takes only the values 0 and 1, so that χ and $\bar{\chi}$ may be viewed as indicator functions of disjoint subsets of divisors of $P(z)$. Now we are ready: suppose we want to construct a $\chi = \chi^+$ that leads to the upper bound

$$S(\mathcal{A}, P(z)) \le S_1^+, \tag{10}$$

where S_1^+ is the sum S_1 in (8) with χ^+ in place of χ (with a similar interpretation of S_2^+ in (9)). Such a bound holds if (see (7)) $S_2^+ \le 0$. By (9), $S_2^+ \le S_{21}^+$ anyway, so if we require of χ^+ that

$$\bar{\chi}^+(d) = 0 \quad \text{when} \quad \mu(d) = 1,$$

then indeed (10) does hold. (We shall come back to this apparently drastic condition later.) Next, we have simply dropped S_{22}^+ because each term in that sum is non-negative, but we want $\bar{\chi}^+(d)$ when $\mu(d) = -1$ to be such that little is lost by leaving out each $S(\mathcal{A}_d, P(q(d)))$ when $\mu(d) = -1$. Finally, if S_1^+ is to be dominated by the first expression on the right of (8) (with $\chi = \chi^+$, of course), then $\chi^+(d)$ needs to be 0 when d is too large, say $\chi^+(d) = 0$ when $d \mid P(z)$ and $d \ge D$; the parameter D is to be such that the second expression on the right of (8) is an error term. To express these three conditions in a more tangible form we borrow from the insights of R-I and give χ a quasi-multiplicative structure: Let $d > 1$ and squarefree have canonical prime decomposition

$$d = p_1 p_2 \cdots p_r, \quad p_1 > p_2 > \cdots > p_r = q(d), \tag{11}$$

and suppose that[1]

$$\chi(d) = \eta(p_1)\eta(p_1p_2)\cdots\eta(p_1\cdots p_r)$$

[1] Greaves does not use this notation, but his B-notation is equivalent.

where $\eta(\bullet)$ is an arithmetical function to be determined that takes only the values 0 and 1. We deduce at once that

$$\bar{\chi}(d) = \chi(d/q(d))(1 - \eta(d)).$$

In the case of χ^+ it is evident that $\bar{\chi}^+(d) = 0$ when $\mu(d) = 1$ if we choose $\eta^+(t) = 1$ whenever $\mu(t) = 1$, that is, when $\nu(t)$ is even; then

$$\chi^+(d) = \eta^+(p_1)\eta^+(p_1p_2p_3)\cdots .$$

To find $\eta^+(t)$ when $\mu(t) = -1$, we observe that $\bar{\chi}^+(d)$ can take the value 1 only if $\eta^+(d) = 0$; therefore we shall require that $\eta^+(t) = 0$ when $\bar{\chi}^+(t) = 1$ and $\mu(t) = -1$. From what we said earlier, when $\mu(t) = -1$ and $\bar{\chi}^+(t) = 1$ we want the best lower estimate of $S(\mathcal{A}_t, P(q(t)))$ to be small. Now one learns from [JR] and, to a lesser extent, from [AO] that one may expect the lower bound for $S(\mathcal{A}, P(z))$ to take the form

$$S(\mathcal{A}, P(z)) \geq XV(P(z))\{f(u) + o(1)\}, \quad u = \frac{\log D}{\log z}, \tag{12}$$

where $f(u)$ is a continuous, monotone increasing function that is non-negative and approaches 1 as $u \to \infty$ at an exponential rate or even faster. In the important case of linear problems (and for a large class of other problems) there exists a number $\beta \geq 1$ such that $f(u) = 0$ when $u \leq \beta$, that is when $z \geq D^{1/\beta}$. Thus, when we turn to $S(\mathcal{A}_t, P(q(t)))$ and replace z by $q(t)$ and D by D/t (because of the summation over t) we see that if

$$\frac{\log(D/t)}{\log q(t)} \leq \beta, \quad \text{or} \quad q(t) \geq (D/t)^{1/\beta},$$

then indeed the best available lower bound for $S(\mathcal{A}_t, P(q(t)))$ is essentially negligible. Thus we are led to define

$$\eta^+(t) = \begin{cases} 1, & q(t) < (D/t)^{1/\beta} \\ 0, & q(t) \geq (D/t)^{1/\beta} \end{cases} \quad \text{when } \mu(t) = -1,$$

and we now have χ^+ fully specified, apart from the value of β, and satisfying the conditions set out for it earlier; when $d \mid P(z)$ and is as in (11), $\chi^+(d) = 1$ if and only if the $\llcorner\frac{1}{2}(r-1)\lrcorner$ inequalities

$$p_{2j+1}^{\beta+1}p_{2j}p_{2j-1}\cdots p_2p_1 < D, \quad j = 0, 1, \ldots, \llcorner\frac{1}{2}(r-1)\lrcorner$$

are satisfied. Note in particular that $\chi^+(d) = 0$ when $d \geq D$. To summarize, we have by (10) and (8) that

$$S(\mathcal{A}, P(z)) \leq X\{V(P(z)) + T^+(P(z))\} + \sum_{\substack{d \mid P(z) \\ d < D}} \mu(d)\chi^+(d)r_{\mathcal{A}}(d) \tag{13}$$

where

$$T^+(P(z)) = \sum_{d \mid P(z), \mu(d) = -1} \bar{\chi}^+(d)\frac{\rho(d)}{d}V(P(q(d))).$$

The expression

$$\begin{aligned} &V(P(z)) + T^+(P(z)) \\ &= V(P(z))\Bigg\{1 + \sum_{\substack{d|P(z) \\ \mu(d)=-1}} \bar{\chi}^+(d)\frac{\rho(d)}{d} \prod_{q(d)\le p<z} \left(1 - \frac{\rho(p)}{p}\right)^{-1}\Bigg\}, \end{aligned} \tag{14}$$

and the coefficient of $V(P(z))$ on the right, obviously at least as large is 1, is closely approximable by a continuous function $F(u)$ (with $u = (\log D)/\log z$ as before) that decreases to 1 monotonically as $u \to \infty$ at a rate that is exponential or faster. Thus, as a comparison to (12), we have (essentially) that

$$S(\mathcal{A}, P(z)) \le XV(P(z))\{F(u) + o(1)\}. \tag{15}$$

When u is very large—in other words, when z is small—(12) and (15) lead to asymptotics for $S(\mathcal{A}, P(z))$ of the fundamental lemma type. (We have to add that the simplicity of (12) and (15) is deceptive; we have assumed implicitly that there has been enough information about the averages of $r_{\mathcal{A}}(d)$ over $d \mid P(z)$ and $d < D$ to make them genuine error terms.)

To identify the functions F and f, and to determine the number β, requires the introduction of a parameter $\kappa > 0$ associated with the average behavior of $\rho(p)$ over the primes of $\mathcal{P}$. Following Iwaniec, κ is a *sifting density* for ρ if there exists a constant $L = L(\kappa) > 1$ such that

$$\prod_{w\le p<z} \left(1 - \frac{\rho(p)}{p}\right)^{-1} \le \left(\frac{\log z}{\log w}\right)^{\kappa} \left(1 + \frac{L}{\log w}\right), \quad 2 \le w < z. \tag{16}$$

Clearly κ is not uniquely defined here—if κ is admissible in (16), so is any number larger than κ. Nevertheless, in any given sieve problem the optimal κ is usually known; for example, in all linear problems $\kappa = 1$. (This is also a good place to remark that the combinatorial analysis I have sketched above is correct in all particulars for problems with density $\kappa \ge \frac{1}{2}$.) Chapter 4 shows that F and f have to satisfy a certain simultaneous linear differential-delay boundary value problem (see Chapter 4.2.1) involving κ and $\beta = \beta_\kappa$ that is soluble under the several stated conditions on F and f only if β is the unique largest positive solution of an equation that is algebraic when 2κ is an integer and otherwise is transcendental.

The Rossen-Iwaniec sieve method is remarkable in several ways: it effects a successful marriage between a complicated combinatorial schema and some quite hard analysis, under no more than the weak, one-sided condition (16); it is valid for all densities $\kappa > 0$ (actually for $\kappa = 0$ too) and is best possible not only for $\kappa = 1$ (as is [JR]) but also for $\kappa = \frac{1}{2}$; and the remainder sums have so flexible a structure that, in the linear case, they are accessible to estimation by exponential sum procedures that allow one to sift beyond the natural limit (see Chapter 6).

It transpires, however, that for problems with κ beyond 1, R-I is weaker even than [AO]—as Iwaniec himself has remarked, this illustrates the power of Selberg's method. In fact, a hybrid of Selberg's method and R-I exists (see [DHR] and the papers cited there) that improves on [AO], especially for smaller κ—say κ between 1 and 10. This is mentioned in the book but not described in detail (see Chapter 7). It is easy to point out at least where the improvement is made: in (9) interpreted for S_2^+, S_{21}^+ was disposed of by choosing $\bar{\chi}^+(d) = 0$ when $\mu(d) = 1$, whereas we can apply Selberg's sieve to at least some of the terms in S_{21}^+. The resulting combinatorics are of the R-I kind and not much more complicated, but the boundary value problem in which they culminate becomes considerably harder, except when 2κ is an integer. Concrete examples like Selberg's when $\kappa = 1$ and Iwaniec's when $\kappa = \frac{1}{2}$ are not known and greatly to be desired. We have $\beta_{1/2} = 1$ and $\beta_1 = 2$, and Selberg has given a convincing argument that suggests that $\beta_\kappa \sim 2\kappa$ as $\kappa \to \infty$ (see [S] and Chapter 7); existing methods yield only $\beta_\kappa \sim 2.44\ldots\kappa$, so probably much remains to be done. It may be that a radically different combinatorial approach is needed when $\kappa > 1$. The so-called vector sieve of Brüdern and Fouvry [BF1], [BF2] is an interesting step in this direction.

What about applications? Where problems in categories I and II are concerned, Brun's or the Brun-Hooley method usually suffices; it is in the context of such applications that Erdős remarked in 1965 that "Brun's method is perhaps our must powerful elementary tool in number theory." Numerous standard applications are presented systematically in [HR], and Greaves has not attempted to repeat them in this book. It is instructive to look at the many papers of Erdős where the sieve is used; one sees that while the sieve method that is applied may be simple, the way in which it is introduced is often highly insightful. To illustrate this remark, recall the classical result of Schnirelmann that combines his theory of the addition of integer sequences with Brun's sieve to prove that there exists an absolute constant k such that every natural number greater than 1 is the sum of at most k primes (from much more recent work, $k \leq 17$ or 18 using only elementary techniques).

In problems of category III one uses R-I for good results, but invariably one does better when combining it with other ideas, some of these set out expertly in Chapter 5. For example, the lower bound sieve alone applied to the numbers $\{p + 2 : p \leq x\}$ shows, when the Bombieri-Vinogradov theorem is used to estimate remainder sums, that $p + 2$ is infinitely often a P_4. A few years ago Professor Vaughan pointed out to me that the primes p for which $p + 2$ is the product of exactly four primes make a negligible contribution, so that actually $p + 2$ is infinitely often a P_3. But years earlier J. R. Chen had used the lowerbound sieve in combination with a certain weighting procedure *and* with an idea like Vaughan's (that had actually inspired Vaughan) to prove that $p + 2$ is infinitely often a P_2.

A weighted sieve method seems, by setting out to achieve less, actually to achieve more. The idea was pioneered by P. Kuhn in 1941; weighting procedures can become technically very complicated, but Kuhn's original method was rather simple: Let $\mathcal{A}$ be the sequence to be sifted by some truncation of $\mathcal{P}$, and suppose that $a \leq D^g$ for every a in $\mathcal{A}$. Sift out from $\mathcal{A}$ all those a's that have small prime factors from $\mathcal{P}$, say prime factors $p \leq D^{1/s}$ where s is a large positive number, and then eliminate the negligibly small number of the remaining a's that are not squarefree with respect to larger primes. Now form the sum

$$M = \sideset{}{'}\sum_{\substack{a \in \mathcal{A} \\ (a, P(D^{1/s}))=1}} m(a),$$

where the dash signifies that no a counted in the sum is divisible by the square of a prime exceeding $D^{1/s}$, and the weight $m(a)$ is constructed to be positive only if a has very few prime factors from $\mathcal{P}$ beyond $D^{1/s}$. Kuhn's simple choice of $m(a)$ illustrates the weighting procedure well. Let t be a number such that $1 < t < s$

and b a positive integer, and let

$$m(a) = 1 - \frac{1}{b+1} \sum_{\substack{p|a \\ p<D^{1/t}}} 1.$$

On the one hand, it is easy to see that M can be expressed in terms of S-functions, and on the other, showing that M is positive for some choice of parameters b, s and t implies that $\mathcal{A}$ contains elements a with $m(a) > 0$, and for these, a has at most b prime divisors between $D^{1/s}$ and $D^{1/t}$. Hence, for each such a,

$$D^g \geq a \geq D^{b/s+(\nu(a)-b)/t}$$

or

$$\nu(a) \leq gt - (t/s - 1)b.$$

Chen used Kuhn's method as part of his proof that $p + 2 = P_2$ infinitely often, and Richert [R] used the slightly more sophisticated logarithmic weights from [AO] to obtain his result about almost-prime values of irreducible polynomials (quoted earlier). Chapter 5 describes several much more complicated weighted sieves, some due to Greaves himself, which hold the promise of improving many known results. Selberg constructed another highly ingenious weighted sieve many years ago that is described in [S] and has been generalized elegantly in [H-B1] to yield novel information.

Sieve methods and sieve ideas are now well-established in the arsenal of number theory. The methods are very general, and this seemed at one time to detract from their efficacy in tackling special problems, but they have combined so well with other methods that this generality need no longer be viewed as a handicap. One has only to think of the role they have played in narrowing the gaps in which primes and almost-primes occur; or in Iwaniec's proof [I] that $n^2 + 1 = P_2$ infinitely often, which uses Linnik's dispersion method; or in the recent spectacular diophantine results of Friedlander and Iwaniec [FI] showing that infinitely many primes p have representation $p = m^2 + n^4$, and of Heath-Brown [H-B2] that $p = m^3 + 2n^3$ for infinitely many primes. An application that has been eclipsed by Wiles' proof of FLT but still deserves honorable mention in that context because of its (relative) simplicity is the proof of the first case of FLT for infinitely many prime exponents by Adelman and Heath-Brown based on a sieve theorem of Fouvry (see [Ri] for a fine account). There is a host of other remarkable applications, some of which are mentioned and listed in the bibliography.

As for sieve ideas, they too abound in current literature. One might describe the use of the generalized von Mangoldt's function $\Lambda_k(\bullet)$ as a 'local' sieve, another way of isolating primes and almost-primes. As illustrations we have Selberg's famous formula that led to an elementary proof of the Prime Number Theorem, or of Vaughan's identity that is so effective in dealing with exponential sums over prime arguments, or Bombieri's 'asymptotic' sieve. Then there is the 'enveloping' sieve of Hooley (Greaves' teacher) that has an important role in his classical "Applications of sieve methods to the theory of numbers" that might well have been cited in Greaves' bibliography, but wasn't. There is the use of smooth ('friable') numbers—numbers having only small prime factors—which has revolutionised work on Waring's problem in the many papers of Vaughan and Wooley. And there is the 'large' sieve of Linnik, Renyi and Roth that, in the hands of Bombieri, Montgomery, Huxley and others has had a profound influence on analytic techniques in number theory. Although the large sieve is not obviously a device of the kind we have been describing, Montgomery, in one of his earliest papers, used its arithmetical formulation to derive an upper bound for sifting an interval by not just one, but several, residue classes mod$p$ for each p of a truncated sieve $\mathcal{P}$ that has the form characteristic of Selberg's method. Finally, I mention a curious recent use of a sieve [FK] where almost-primes that are *not* primes play a crucial role.

The notion of 'sifting'—of isolating and counting one kind of sequence as a sub-sequence of another—of primes or square-free numbers in intervals and/or in arithmetical progressions, or of arithmetical progressions among the primes, is now so common an occurrence in the higher arithmetic that it has come to stay. I hope that the study of Dr. Greaves' book will inspire the next generation of sifters and that its reception will encourage him to produce a second volume more directed towards applications, which are after all the objectives of sieve methods.

Addendum. Dr. Greaves has informed me that condition (2.2) in the statement of Theorem 2 on p. 55 needs to be strengthened to condition (2.1) on p. 265 for the result to remain true. He plans to post a list of corrigenda on his university web-page in due course.

References

[AO] N. C. Ankeny and H. Onishi, *The general sieve*, Acta Arith. **X** (1964), 31–62. MR **29:**4740

[BF1] J. Brüdern and E. Fouvry, *Lagrange's Four Squares Theorem with almost prime variables*, J. Reine Angew. Math. **454** (1994), 59–96. MR **96e:**11125

[BF2] ———, *Le crible à vecteurs*, Compositio math. **102** (1996), 337–355. MR **97f:**11079

[DHR] H. Diamond, H. Halberstam and H.-E. Richert, *Combinatorial Sieves of Dimension exceeding one, II*, Analytic Number Theory, Proc. of a Conference in Honor of Heini Halberstam, Birkhäuser, 1996, pp. 265–308. MR **97e:**11112

[FH] K. Ford and H. Halberstam, *The Brun-Hooley Sieve*, J. Number Theory **81** (2000), 335–350. MR **2001d:**11095

[FI] J. Friedlander and H. Iwaniec, *The polynomial $X^2 + Y^4$ captures its primes*, Annals of Math. **148** (1998), 945–1040. MR **2000c:**11150a

[FK] K. Ford and S. Konyagin, *On two conjectures of Sierpinski concerning the arithmetic functions σ and ϕ*, Proceedings of Conference in Honor of A Schinzel, de Gruyter, Berlin (1999), 795–803. MR **2000d:**11120

[H-B1] D. R. Heath-Brown, *Almost-prime k-tuples*, Mathematika **44** (1997), 245–266. MR **99a:**11106

[H-B2] ———, *Primes represented by $x^3 + 2y^3$*, Acta Math. **186** (2001), 1–84.

[HR] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, New York/London, 1974. MR **54:**12689

[H1] C. Hooley, *An almost-pure sieve*, Acta Arith. **LXVI** (1994), 359–368. MR **95g:**11092

[H2] ———, *On the intervals between numbers that are sums of two squares: IV*, J. Reine Angew. Math. **452** (1994), 79–109. MR **95j:**11090

[I] H. Iwaniec, *Almost-primes represented by quadratic polynomials*, Invent. Math. **47** (1978), 171–188. MR **58:**5553

[JR] W. B. Jurkat and H.-E. Richert, *An improvement of Selberg's sieve method*, Acta Arith. **XI** (1965), 217–240. MR **34:**2540

[Ri] P. Ribenboim, *Recent results about Fermat's Last Theorem*, Expositiones Math. **5** (1987), 75–96. MR **89c:**11046

[R] H.-E. Richert, *Selberg's sieve with weights*, Mathematica **16** (1969), 1–22. MR **40:**119

[S] A. Selberg, *Collected Papers*, vol. II, Springer Verlag, 1991. MR **95g:**01032

Heini Halberstam
University of Illinois
E-mail address: heini@math.uiuc.edu