Number theory

Manfred R. Schroeder

The paradigm of abstract mathematics has many applications in science, engineering, even artistic design

Number theory is traditionally considered a rather abstract field, far removed from practical applications. In the recent past, however, the "higher arithmetic" has provided highly useful answers to numerous real-world problems. These include the design of new musical scales, powerful cryptographic systems, and diffraction gratings of acoustic and electromagnetic waves with unusually broad scatter, with applications in radar camouflage, laser speckle removal, noise abatement, and concert hall acoustics. Another prime domain of number theory is the construction of very effective error-correction codes, such as those used for picture transmission from space vehicles and in compact discs (CDs). Other new applications include schemes for spread-spectrum communication, "error-free" computing, fast computational algorithms, and precision measurements (of interplanetary distances, for example) at extremely high signal-to-noise ratios. In this manner the "fourth prediction" of general relativity (the slowing of electromagnetic radiation in gravitation fields, predicted by Einstein as early as 1907) has been fully confirmed. In contemporary physics the quasiperiodic route to chaos of nonlinear dynamical systems (the double-pendulum and the three-body problem, to mention two simple examples) are being analyzed in terms of such number-theoretic concepts as continued fractions, Fibonacci numbers, the golden mean, and Farey trees. Even the recently discovered new state of matter, christened quasicrystals, is most effectively described in terms of arithmetic principles. And last but not least, prime numbers, whose distribution combines predictable regularity and surprising randomness, are a rich source of pleasing artistic design.

The queen of math

According to Carl Friedrich Gauss, in the *Princeps Mathematicorum*, "mathematics is the queen of science—and number theory is the queen of mathematics." In fact, in Chinese the name of mathematics is *number science*. What could be more beautiful than a deep, satisfying relation between whole numbers? (One is almost tempted to call them *wholesome* numbers.) Indeed, it is hard to come up with a more appropriate designation than their learned name: the integers, meaning the "untouched ones."

Yet the theory of integers can provide totally unexpected answers to real-world problems. In fact, discrete mathematics is taking on an ever more important role. If nothing else, the advent of the digital computer and digital communication has seen to that. But even earlier, in physics, the emergence of quantum mechanics and discrete elementary particles put a premium on the methods and the spirit of discrete mathematics.

In mathematics proper, Hermann Minkowski, in the preface to his introductory book on number theory *Diophantische Approximationen*, published in 1907 (the year he gave special relativity its proper four-dimensional clothing in preparation for its journey into general covariance and cosmology), expressed his conviction that the "deepest interrelationships in analysis are of an arithmetical nature."

Yet much of our schooling concentrates on analysis and other branches of continuum mathematics to the virtual exclusion of number theory, group theory, combinatorics, and graph theory. As an illustration, at a recent symposium on information theory, the author met several young mathematicians, working in the field of *primality testing*, who, in all their studies up to the Ph.D., had not heard a lecture on number theory!

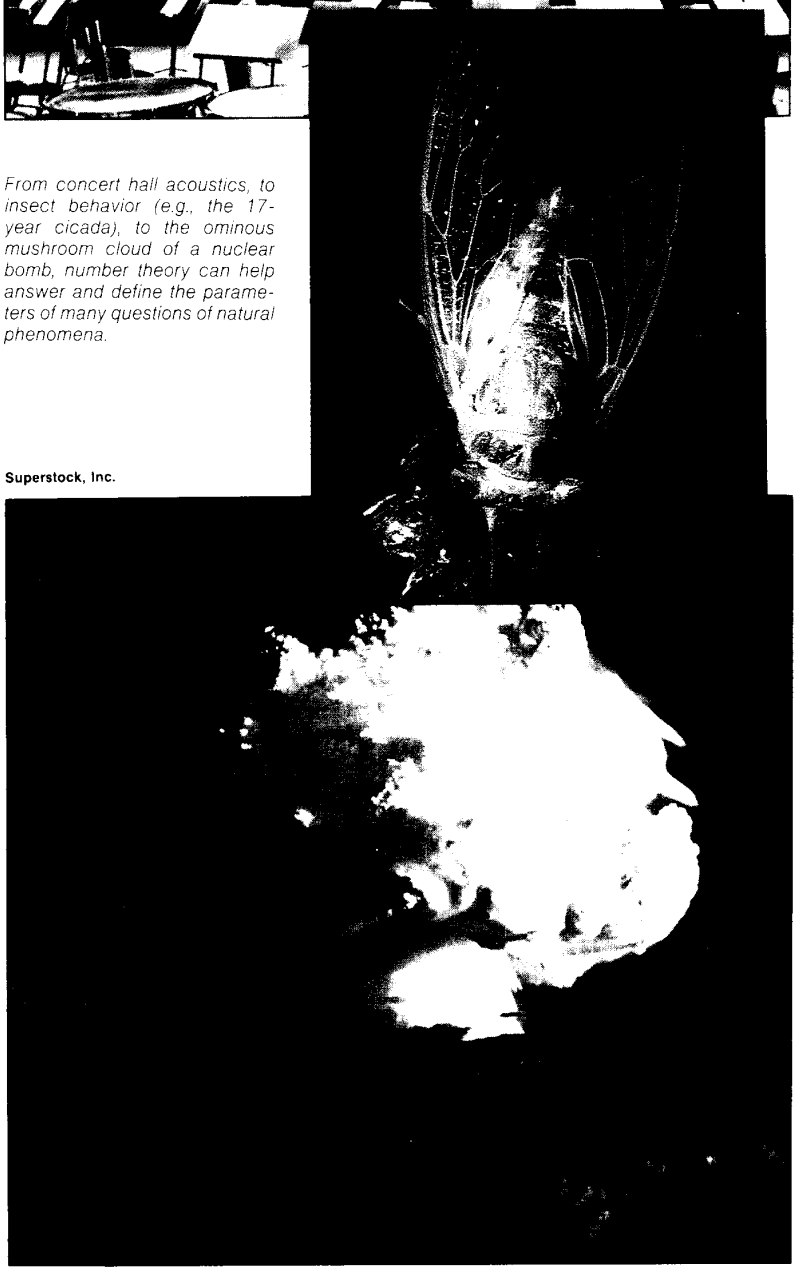
Or, to give an earlier example, when Werner Heisenberg discovered matrix mechanics in 1925, he didn't know what a "matrix" was (Max Born had to tell him), but neither Heisenberg nor Born knew what to make of the appearance of matrices in the context of the atom. (David Hilbert is reported to have told them to go look for a differential equation with the same eigenvalues, if that would make them happier. They did not follow Hilbert's well-meant advice and thereby may have missed discovering the Schrodinger wave equation.)

Integers have repeatedly played a crucial role in the evolution of the natural sciences. Thus, in the 18th century, Lavoisier discovered that chemical compounds are composed of fixed proportions of their constituents which, when expressed in proper weights, correspond to the ratios of *small integers*. This was one of the strongest hints to the existence of atoms; but chemists, for a long time, ignored the evidence and continued to treat atoms merely as a conceptual convenience devoid of physical meaning. Ironically, it was from the statistical laws of *large* numbers, in Einstein's analysis of Brownian motion at the beginning of our century, that the irrefutable reality of atoms and molecules finally emerged.

In the analysis of optical spectra, certain integer relationships between the wavelengths of spectral lines emitted by excited atoms gave early clues to the *structure of atoms*, culminating in the creation of matrix mechanics in 1925, an important year in the growth of integer physics. Later, the near-integer ra-

0278-6648/89/0010-0014$01.00 © 1989 IEEE

Carnegie Hall (photo by Steven J. Sherman)



From concert hall acoustics, to insect behavior (e.g., the 17-year cicada), to the ominous mushroom cloud of a nuclear bomb, number theory can help answer and define the parameters of many questions of natural phenomena.

Superstock, Inc.

tios of atomic weights suggested to physicists that the atomic nucleus must be made up of an *integer* number of similar nucleons. And the deviations from integer ratios led to the discovery of elemental isotopes.

And finally, small divergencies in the atomic weight of pure isotopes from exact integers constituted an early confirmation of Einstein's famous equation $E = mc^2$, long before the "mass defects" implied by these integer descrepancies blew up into the widely noticed and infamous mushroom clouds.

On a more harmonious theme, the role of integer ratios in musical scales has been appreciated ever since Pythagoras first pointed out their importance. The occurrence of integers in biology—from plant morphology to the genetic code—is pervasive. It has even been hypothesized that the North American 17-year cicada (type of insect) selected its life cycle because 17 is a prime number, prime cycles offering better protection from predators than non-prime cycles. The suggestion that the 17-year cicada "knows" that 17 is a *Fermat* prime (a prime of the form $2^{2^n} + 1$) has yet to be touted though.

Another reason for the resurrection of the integers is the penetration of our lives achieved by that 20th-century descendant of the abacus, the *digital computer.* An equally important reason for the recent revival of the integer is the congruence of *congruential arithmetic* with numerous modern developments in the natural sciences and digital communications—especially "secure" communication by cryptographic systems. Last, but not least, the proper protection and security of computer systems and data files rests largely on "keys" based on congruence relationships.

In congruential arithmetic, what counts is not a numerical value per se, but rather its remainder or *residue* after division by a *modulus.* Similarly, in wave interference (be it of ripples on a lake or of electromagnetic fields on a hologram plate) it is not path differences that determine the resulting interference pattern, but rather the residues after dividing by the wavelength. For perfectly periodic events, there is no difference between a path difference of half a wavelength or one-and-a-half wavelengths: in either case the interference will be destructive.

One of the most dramatic conse-

quences of congruential arithmetics is the existence of the chemical elements as we know them. In 1913, Niels Bohr postulated that certain integrals associated with electrons in "orbit" around the atomic nucleus should have integer values, a requirement that ten years later became comprehensible as a wave interference phenomenon of the newly discovered de Broglie *matter* waves: In essence, integer-valued integrals meant that path differences are divisible by the electron's wavelength without leaving a remainder.

Music and numbers

Ever since Pythagoras, small integers and their ratios have played a fundamental role in the construction of musical scales. There are good reasons for this preponderance of small integers both in the production and perception of music. String instruments, as abundant in antiquity as today, produce simple frequency ratios when their strings are shortened by integer fractions. Reducing the length of a string by one half produces the frequency ratio 1:2, the *octave;* and shortening by one third produces the frequency ratio 3:2, the perfect *fifth.*

For the human ear, ratios of small integers avoid unpleasant beats between harmonics. Apart from the frequency ratio 1:1 ("Unison"), the octave is the most easily perceived interval. Next in importance comes the perfect fifth. Unfortunately, as a consequence of the fundamental theorem of number theory, musical scales exactly congruent modulo the octave cannot be constructed from the fifth alone because there are no positive integers "k" and "m" such that

$$\left[\frac{3}{2}\right]^{m} = \left[\frac{2}{1}\right]^{k}$$

In fact, number theory tells us that $3^m = 2^n \pm 1$, except for m = 1, n = 1 and m = 2, n = 3. However, there are good *approximations:* if we write $3^m \approx 2^n$, or $\log_2 3 \approx n/m$. To find n/m, we expand $\log_2 3$ into a *continued fraction*

$$\log_2 3 = 1.58496\ldots = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \frac{1}{2}\ldots,}}}$$

or $\log_2 3 = [1; 1, 1, 2, 2, \ldots]$ in the usual notation.

Breaking the expansion off as shown after the second 2, yields the close approximation m = 12, n = 19. In other words, if we want to make a good fifth with an equal tempered (equal frequency ratio) scale, the basic interval $1:2^{1/12}$, the *semitone*, recommends itself. In fact, the semitone interval has come to dominate much of music. The equal tempered fifth comes out as $2^{7/12} = 1.498\ldots$, which equals 3:2 with an error of only one part in one thousand!

Another fortunate number-theoretic coincidence is the fact that the 7 in the exponent is coprime with 12. As a consequence, we can reach all 12 notes of the octave interval by repeating the fifth (modulo the octave). This is the famous Circle of Fifths.

Euler totients and cryptography

One of the most spectacular applications of number theory in recent times is *public-key cryptography* in which each potential recipient of a secret message *publishes* his encryption key, thereby avoiding the (often substantial) problems of secure secret-key distribution. But how can a key be public and yet produce secret messages? The answer is based on Euler's totient function (r) and the role it plays in inverting modular exponentiation. The public key consists of a modulus "r" and an exponent "s," coprime to (r). The message is represented by an integer "M" in the range $1 < M < r$, and the encrypted message "E" is given by a number in the same range, calculated as follows

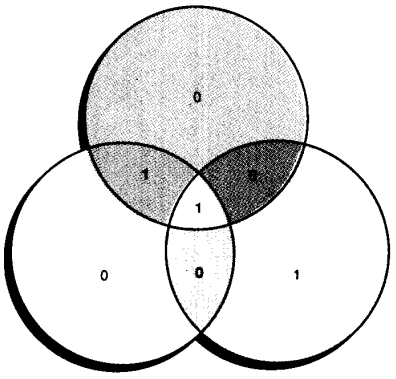


Fig. 1. A Venn diagram can illustrate the single-error correcting property of the original Hamming code, which has four information bits (bold 0s and 1s) and three check bits (light 0s and 1s). The parity in each of the three circles must be even. A single error, which entails one or several odd parities, can thus be uniquely detected and corrected.

$$E = M^s \bmod \phi r . \qquad (4)$$

Decrypting "E" is accomplished by calculating E^t mod "r," where the decrypting exponent "t" is given by

$$ts = 1 \bmod \phi(r) , \qquad (5)$$

i.e., ts = k (r) + 1 for some k. With such a "t,"

$$E^t = M^{st} = M^{k\phi(r)+1} \bmod r,$$

which, according to Euler's theorem, give the message "M" back.

So far, so good and—theoretically—trivial. The trick in public-key encryption is to choose "r" as the product of two very large primes, each say 200 digits long. (There is no paucity of such primes, and enough for all foreseeable purposes can be easily ferreted out from the jungle of composites in the 10^{200} neighborhood.)

Now, with a composite "r," prescription (5), so easily written down, becomes practically impossible to apply because (r) can be calculated only if the factors of "r" are known—and this knowledge is *not* published. In modern parlance, the mapping (4) is a *trap-door function.*

A trap-door function is, as the name implies, a (mathematical) function that is easy to calculate in one direction but very hard to calculate in the opposite direction. For example, it takes a modern computer only microseconds to multiply two 100-digit numbers. By contrast, to decompose a 200-digit number, having two 100-digit factors, into its factors can take "forever," even on the fastest computers available in 1989 and using the most efficient factoring algorithms known today.

On the other hand, knowing the factors of "r," as the legitimate recipient of the encrypted message E does, (r) can be easily calculated and decrypting becomes possible. The decrypting exponent "t" is obtained by Euclid's algorithm or by solving (5) directly:

$$t = s\phi^{(\phi(r))-1} \bmod \phi r .$$

Not so long ago, the most efficient factoring algorithms on a very fast computer were estimated to take trillions of years. But algorithms get more efficient by the month and computers become faster and faster every year. So there is no *guarantee* that one day a so-called "polynomial-time" algorithm will not emerge that will allow fast factoring of even 1000-digit numbers. Few mathematicians believe that a true polynomial-time algo-

Designing for better sound

Extensive physical tests and psychophysical evaluation of the acoustic qualities of concert halls around the world have established the importance of lateral sound waves. Such waves produce dissimilar signals of a listener's two ears, a kind of stereophonic condition that is widely preferred for music listening.

In order to convert sound waves travelling longitudinally (from the stage via the ceiling to the back wall) into lateral waves, the author has recommended ceiling structures that scatter sound waves, without absorption, into broad lateral patterns.

How should one go about designing such an ideal scatterer for sound (or light or radar) waves? Curiously, the answers come from number theory.

Consider a surface structure whose reflection coefficient r_n varies in equidistant steps along one axis according to

$$r_n = \exp(2\pi i n^2/p), \quad n = 0, \pm 1, \pm 2, \ldots \qquad (1)$$

where p is a prime and n^2 may be replaced by $(n^2) \bmod p$, its least nonnegative residue modulo p.

It is easy to show that the Discrete Fourier Transform (DFT) of r_n has constant magnitude. As a physical consequence, the intensities of the wavelets scattered into different directions from a surface with reflection coefficients (1) will be equal (in the customary Kirchhoff approximation of diffraction theory).

The different reflection coefficients are realized by "wells" of different depths

$$d_n = \frac{\lambda}{2p} (n^2) \bmod p \qquad (2)$$

as illustrated in Fig. A for p = 17. Such wells give a roundtrip phase change of $2d_n \cdot 2\pi/$ in accordance with the phase requirement of Equation (1). In Equation (2), is the longest wavelength to be scattered. For any integral submultiple of that wavelength, /m, the reflection coefficients (1) are changed to r_n^m, which for m = 0 (mod p) has the same flat Fourier property as r_n. For the width of the wells "w" one chooses typically half the smallest wavelength to be scattered over $\pm \pi/2$. Figure B shows the diffraction pattern of the grating in Fig. A. Note how well the broad scattering of sound is realized.

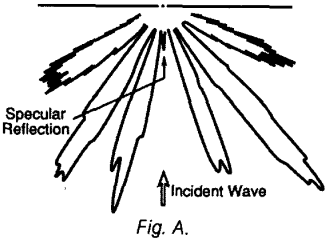


Fig. A.

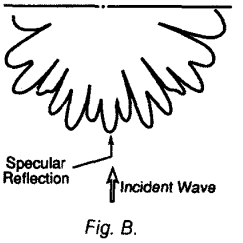


Fig. B.

rithm is just around the corner, but there also seems to be little prospect of proving that this will never occur. This is the Achilles heel of public-key cryptography.

Error correction codes from Galois fields

Galois sequences, with periods $n = p^m - 1$, are constructed with the help of an irreducible polynomial g(x) of degree "m" with coefficients from GF(p), such that g(x) is a factor of $x^n - 1$ but not a factor of $x^r - 1$ for $r < n$. Such polynomials are called *primitive.*

Binary Galois sequences (p = 2) with elements 0 and 1 (or, in certain other applications, 1 and −1) are the most important practical case. For p = 2 and m = 4, a primitive polynomial with the stated property is

$$g(x) = 1 + x + x^4,$$

from which the recursion

$$a_{k+4} = a_{k+1} + a_k \qquad (6)$$

is obtained. Beginning with the initial condition 1000 (or almost any other tuple, except the all-zero tuple), Equation (6) generates the binary Galois sequence of periodic length $2^4 - 1 = 15$:

1000, 10011010111; etc.
(repeated periodically).

The error correcting properties of codes based on such sequences result from the fact that the $2^4 = 16$ different initial conditions generate 16 different code words of length 15, that form a "linear code" (i.e., the sum of two code words is another code word). These code words define therefore a 4-dimensional linear subspace of the 15-dimensional space with coordinates 0 and 1. In fact, the 16 code words describe a simplex in that space (in three dimensions a simplex is a tetrahedron) and the resulting code is therefore called a *Simplex Code.*

Its outstanding property is that every pair of code words of length $n = 2^m - 1 = 15$ (for m = 4) bits (of which "m" bits are information bits and $n - m = 11$ are check bits) has the same Hamming distance (the number of 0, 1 disparities), namely $2^{m-1} = 8$. Thus, the code can recognize up to $2^{m-2} = 4$ errors and correct up to $2^{m-2} - 1 = 3$ errors. The price for this error correcting property is a reduced signaling efficiency, namely m/n = 4/15.

Several other codes can be derived from the simple Simplex Code. For example, the famous (and historically early) *Hamming Codes.* The code words of a Hamming Code are given by the orthogonal subspace of the Simplex Code of the same length. Hamming codes of length $n = 2^m - 1$ carry $n - m$ information bits and "m" check bits, and can correct precisely one error. The functioning of the Hamming Code for m = 3, n = 7, is illustrated in Fig. 1.

The $n - m = 4$ information bits, say 1001, are entered into the four inner areas of the Venn diagram (indicated by fat characters in Fig. 1). The m = 3 check bits (light characters) are entered into the three outer areas such that the sum in each circle is even.

The receiver of a code word, which may have been contaminated in transmission, checks the parity in each circle and marks all circles with odd sum. The intersection of these circles then specifies uniquely a single bit error (including in the check bits themselves). These three parity checks allow the receiver to distinguish between precisely $2^3 = 8$ different possibilities: a single error in any of the seven transmitted bits or *no* error. No wonder the Hamming Code is called a perfect code.

Epilog

Just a sprinkling of the numerous applications of number theory outside mathematics proper have been mentioned here. What riddle will be solved next by number theory? Is this effectiveness of the higher arithmetic completely unreasonable? Or are we witnessing here a "pre-established harmony" a la Leibniz (and his theory of monads) between mathematics and the real world?

About the author

Professor Manfred R. Schroeder, a Fellow of the IEEE, is Director of the Institute for "Schwingungs-Physik" of the University of Goettingen. The present article is based on his book *Number Theory in Science and Communications: With Applications in Cryptography, Physics, Digital Information, Computing and Self-Similarity,* 2nd enlarged edition (Springer-Verlag, Berlin, 1986). □