and the disease might have moved in the opposite direction. AIDS is not found in rural Haiti, Guerin notes. "It appears to us that the disease is an urban disease in a population that is in contact with tourists." The extent and nature of possible contacts between Haitians and Central Africans are currently unknown.

Haitian officials are also sensitive about suggestions that Haitians constitute a separate risk group simply because they are Haitians. Transmission among them is likely to occur by the standard routes, sexual contact and contaminated needles or blood products, Guerin says. The male patients, who constitute about 70 percent of the total in Haiti, rarely admit to homosexual practices, but these cannot be ruled out. "Homosexuality is a taboo subject in Haiti, and it is hard to get the information," Guerin explains.

A great many questions about AIDS, including the big one concerning the nature of the causative agent, remain unanswered. Nevertheless, the epidemiological studies are providing some interesting leads. As Quinn puts it, "The work in Zaire may give us some important clues about the course and spread of AIDS throughout the world." —**Jean L. Marx**

Factoring Gets Easier

Mathematicians are exploiting computer designs to factor large numbers in times that, as recently as 1 year ago, seemed inconceivable

Each fall for the past 15 years, a group of mathematicians has met in Winnipeg to discuss progress in factoring large numbers. They know which numbers are particularly hard to factor and they even have a "Ten Most Wanted List" of difficult numbers, as well as a longer list of numbers that they have simply designated "Wanted." The wanted and most wanted numbers, says Gus Simmons of Sandia National Laboratories in Albuquerque, New Mexico, are not just long numbers. They are numbers whose factors would be important to engineers, who use them to construct shift registers; to mathematicians, who use them in such algebraic topics as field theory; and to cryptographers, who use them in the design of codes. And they are numbers that are known to be inordinately difficult to factor.

Last year, the mathematicians decided that they had reached a point of diminishing returns. They could use powerful computers to factor 50-digit numbers, but 50 digits seemed to be the limit of computational feasibility. They had on their wanted and most wanted lists numbers of 60 digits or more. Reluctantly, they decided to go to press with a paper that is jokingly said to have taken 50 years to write because two of the contributors, Derek and Emma Lehmer of the University of California at Berkeley, worked on the project that long. The American Mathematical Society (AMS), which had sponsored the search for factors of these large numbers, agreed to publish what was then known and close off the project.

Recently, however, the whole factoring picture has changed. Mathematicians are polishing off 50-digit numbers in roughly an hour and are finding that 70-digit numbers, which they would have expected to take about 100 times as long as 50-digit ones to factor on a computer, are easily within reach. "In 1982, you could have collected money from anyone in that crowd if you bet that 60-digit numbers would be factored in the next year," says Simmons. "They had all worked on factoring for 12 to 15 years and they knew how difficult it was."

Factoring has long been of interest to mathematicians, but it recently has become something of a hot research topic because the ability to factor large numbers is related to the security of a newly developed cryptography system. The system, called RSA after the last initials of its inventors, uses large numbers that are the product of two primes as the heart of its algorithm. Anyone who can factor those numbers can break the code. When the RSA code was first proposed about 6 years ago, its inventors suggested using 80-digit composite numbers. Now they suggest using numbers of at least 200 digits which, for the time being, seem invulnerable.

"A few years ago, an interest in factoring was the hallmark of a proven eccentric."

John Brillhart of the University of Arizona, who is one of the old-time factorers, says he is somewhat disconcerted by the current interest in factoring. "A few years ago, an interest in factoring was the hallmark of a proven eccentric. Now it suddenly relates not only to the transfer of funds between banks but also to the national security. There's a certain amount of irony in this," he says.

The problem of factoring has fascinated mathematicians since the time of the ancient Greeks. But it is only comparatively recently that people have made progress. The beginning of the modern era of factoring was in the 1920's when the French mathematician Maurice Kraitchik developed some ideas that are now being implemented on large computers. Kraitchik's ideas, however, were not formulated in a particularly logical way. Yet Kraitchik made mathematicians realize that it might be possible to find clever ways to factor large numbers.

At about the same time that Kraitchik was developing his factoring methods, Derek and Richard Lehmer of the University of California at Berkeley began building mechanical devices to test for primes and to factor. In this precomputer era the Lehmers were able to factor 20-digit numbers, an extraordinary accomplishment since, for each additional three digits of a number, it seems that the factoring time is doubled.

About 15 years ago, the AMS decided to sponsor mathematicians in their search for the factors of large numbers. The idea was to make a table of all the known factors of numbers of the form $a^n \pm 1$, where a is a small whole number and n is a large number. These numbers have always been of enormous importance in number theory and algebra and are also used by engineers to generate random numbers.

The AMS table was to be called the Cunningham Project Table, in memory of a British colonel named A. J. Cunningham, who, around the turn of the century, compiled a table of factors of numbers of this sort. The reason for the AMS interest in the project was that, with the advent of large computers, mathematicians began to be able to fac-

What Does It Mean to Factor?

Every whole number is made up of prime numbers, which are only divisible by themselves and 1 (2, 3, 5, 7, and 11, for example). To factor a number, break it down into a product of prime numbers. For example, $15 = 3 \times 5$.

Of course, the bigger a number is, the harder it is to factor. The brute-force way to factor a number would be to simply try dividing it by all primes less than the square root of the number. But this method will not get you very far. It simply takes too long. As Hugh Williams of the University of Manitoba points out, if a computer could perform one division every billionth of a second, it would still take more than 35,000 years of computer time to factor the 58-digit number $2^{193} - 1$ with this method. But, with the faster methods of factoring, it takes Gus Simmons, James Davis, and Diane Holdridge of Sandia National Laboratories only 38.3 minutes to find that $2^{193} - 1 = 13821503 \times 61654440233248340616559 \times 14732265321145317331353282383$.—**G.K.**

tor numbers as large as 40 digits. It was the Cunningham Project mathematicians who gathered each year in Winnipeg and who compiled the list of wanted and most wanted numbers. "The numbers on that list have really been beaten with a very large stick," says Simmons.

So when the Cunningham Project was put to bed last year, mathematicians thought it was nearly impossible to do any more than they had already done. They knew the numbers in their list that they could not crack had factors, because it is relatively easy to test a number to see if it is or is not a prime. But it looked as if 50-digit numbers were about the largest they could factor. "It was thought that no one in the foreseeable future could whittle off much larger numbers," says Simmons.

What suddenly changed this picture was not faster computers. Instead, it was mathematicians' newfound interest in exploiting computer architecture—the "organization chart" for the machine which determines how information flows and the lines of communication. Currently, the world record for factoring is held by James Davis and Diane Holdridge of Simmons' group at Sandia. But two other groups are rapidly developing methods on other machines that seem likely to give the Sandia group some stiff competition.

The first real success in factoring large numbers on computers was reported in 1971 by Brillhart and John Morrison when they factored a famous 40-digit number on a computer. Their ideas are essentially the ones being used by mathematicians today as they tackle 60-digit numbers. If you have a number, m, that you want to factor, one way to do it would be to find two other numbers x and y so that $x^2 - y^2 = m$. So mathematicians set about looking for a set of two squares whose difference is m. Of course, if they used no special tricks to find x and y, it would be like looking for these numbers at random and would be completely infeasible. "You are really trying to find the difference of two squares, but you admit you don't have much hope of doing that," says Simmons. But as a next best approach, it may be possible to work in residue classes, dividing all of the equations by small primes and looking at remainders.

You lose some information when you work this way, but the payoff is that, with a powerful computer, it is possible to substitute what Simmons calls "sloppy arithmetic" for meticulous calculations so long as you are able to do perhaps hundreds of thousands of divisions. What mathematicians end up doing is taking the problem of factoring one large number and breaking it down into a problem of factoring hundreds of thousands of smaller numbers. There are potentially millions of these smaller numbers to choose from in any particular factoring problem, so the best strategy is not to waste time on any particular small number. If it does not factor easily, go on to another one. The two main versions of this strategy are called quadratic sieving and continued fractions.

For example, the problem of factoring a 60-digit number might be broken down into a problem of factoring hundreds of thousands of 30-digit numbers. But only about 6000 of these smaller numbers need to be completely factored. For the rest, the mathematicians only need to know if particular prime numbers do or do not divide them evenly. Millions of the 30-digit numbers can be generated for this problem.

The first real innovation that allowed mathematicians to tackle numbers beyond 50 digits occurred over a beer at Winnipeg in the fall of 1982. There had been a computing conference going on and the factoring group had attended as part of their annual meeting on the Cunningham Project. Simmons recalls that he, Marvin Wunderlich of Northern Illinois University and a member of the factoring group, and Tony Warnock, a Cray Research engineer, went out for a beer. "Marv and I were talking about why factoring is so computationally difficult. The thing that kills us is that we have vectors that are *very very* long. Several thousand components must be modified many thousands of times but only in a small number of places each time. Ordinarily, the time it takes to change a vector is proportional to the length of the vector. Tony said that the architecture of the Cray is such that you can change different positions in a vector and, so long as the positions are a constant length apart, the time it takes to make the changes is proportional to the number of changes you are making, not to the length of the vector. That's an incredible advance. Suddenly—and serendiptously—it appeared that the architecture of the Cray would allow us to do quadratic sieving enormously fast."

"At first," Simmons says, "we weren't even sure it was true. I'd never worked with a Cray myself before. When I came back from the conference, I got Jim Davis and Diane Holdridge together and we set to work. Serendipity was on our side. We took 52-, 53-, 54-digit numbers and did them in times thought to be a theoretical limit. We did a 52-digit number in 1.9 hours. This was a number that had well in excess of 100 hours of computing time devoted to it before we tackled it. We did a 58-digit most wanted number in 8.8 hours. But at that point we became alarmed. We knew that the length of time it takes to factor numbers grows exponentially. It looked like at 60 digits we would be up to 20 hours."

At that time, however, Davis made an important advance. The problem was that as the factoring algorithm progresses, the computer has to factor trial numbers using a collection of divisors known as a factor base where the trial numbers get bigger and bigger. Eventually, says Simmons, you get to the point where the trial numbers you are attempting to factor are almost as difficult to factor as the number you started with. "You end up running a million tests before you get one more trial number completely factored."

Davis, however, found a way to bring the whole factoring algorithm back to the beginning again when things get too large. "There is more ambiguity and sloppier arithmetic with his method. But the method is much much faster and it still can factor. It speeds up factorization by almost an order of magnitude," says Simmons. Now the Sandia group has refactored a 58-digit number that took them 8.8 hours with their old algorithm in 1.8 hours—a fivefold improvement in speed. They factored a 60-digit number in 2½ hours and a 63-digit "wanted" number in 5.18 hours. They hope to do even better next year when they get a Cray XMP—essentially two Crays hooked back to back. The new machine, they predict, should speed up their factoring by a factor of 4.

A key competitor for the Sandia factoring record is Wunderlich, who is exploiting a new computer at NASA's Goddard Space Flight Center to do the job. Wunderlich's plan is to do the countless trial divisions needed to factor a large number all at once. To do this, he needs to use a parallel processor computer—one that has a number of independent units to do this arithmetic. Wunderlich began planning this work in 1979, intending to use the Illiac IV, the first parallel processor ever built, and one with 64 separate units. But, Wunderlich says, "It was very hard to do anything sustained on that machine. It was an old generation machine and it was torn down shortly after I began. But my experience on the Illiac said to me that this is really the way to do factoring."

Wunderlich next got a grant to try factoring on a British machine called DAP, which has 4096 processors. There are only a few DAP's in the entire world because the machine was expensive and commercially unsuccessful. Wunderlich spent the summer of 1981 working on one of these machines at Queen Mary College in London, where he wrote parts of a factoring program but never got the program entirely running. "It takes a gigantic effort to put a new algorithm on a large machine," he remarks.

Soon afterward, Wunderlich heard from NSA, which asked him to submit a grant proposal. Wunderlich did and received, he says, "generous funding." He is using this grant money to try and factor large numbers on a very new machine called MPP, for Massively Parallel Processor. The computer, which belongs to NASA and is to be primarily used to analyze satellite data, has 16,384 parallel processors. For factoring, however, it seems ideal. "It's the right architecture, the right kinds of languages, and it has lots of support," says Wunderlich. But, he remarks, the machine is still so new that it is barely running. When the MPP is fully operational, however, Wunderlich thinks he will have "one of the fastest factoring programs in existence. I think I will be able to do a 60-digit number in 1 hour."

The third group of researchers in this factoring competition is building its own computer—one that will do only factoring. It is being built by Samuel Wagstaff at Purdue University together with Jeffrey Smith and Carl Pomerance at the University of Georgia, from parts that they order through the mail. They also plan to produce a special-purpose chip to do trial divisions. The investigators call their machine EPOC, for Extended Precision Operand Computer, or, more colloquially, The Georgia Cracker.

The EPOC computer has two features that will speed up factoring, Wagstaff says. First, it has a large word size. Normally, large computers can only handle word sizes of 32 bits. This machine handles 128 bits. "If you add two numbers of 128 bits, it takes ten operations on an IBM or CDC computer," says Wagstaff. "The EPOC does it in one operation." The second special feature of the EPOC is that it does some of the trial divisions in parallel and it does them in a separate part of the computer. Eventually, the EPOC builders plan to have their machine do several hundred of these divisions at the same time. They expect to be able to factor a 78-digit number in 1 day.

What is the future of factoring? Obviously, there must be a limit to the size of number that can be factored, but mathematicians no longer think that the limit depends on the speed of their computers alone. "I'm convinced now that large-scale computational problems such as factoring depend as much on the architecture of the machine as on its brute-force speed. If you can modify the architecture you can make enormous progress," Simmons says. "The exploitation of machine architecture is a whole new way of doing mathematics."

—GINA KOLATA

Specific Expression of Transferred Genes

Foreign genes, which were transferred into mice, appear to be expressed according to more normal patterns of tissue distribution

Two recent reports indicate that investigators may be on the verge of seeing virtually normal activity of foreign genes that have been transferred into mice. Introduction of new genes into mice has been accomplished many times during the past few years by injecting cloned genes into fertilized eggs. Although 20 to 30 percent of the recipient animals carry the transferred genetic material in their cells and can transmit it to their progeny, the genes have not been expressed normally.

The new results, suggesting that expression of an antibody gene and one coding for the protein transferrin may follow more normal patterns of expression, are therefore an important step forward. They follow closely on the heels of similar successes with gene transfer in fruit flies.

Ursula Storb of the University of Washington and Ralph Brinster of the School of Veterinary Medicine of the University of Pennsylvania and their colleagues injected 300 fertilized mouse eggs with the cloned gene coding for an antibody light chain of the kappa class. They eventually obtained six animals, all males, with the new gene. When these mice were mated with normal females, about half of the progeny carried the antibody gene, which is the expected result.

The investigators analyzed the expression of the gene, as indicated by its transcription into messenger RNA (mRNA), the first step of protein synthesis, in the progeny of three of the original animals. Transcription of the transferred gene "was high in the spleen and low in liver, and the mRNA is the size you would expect," Storb says. The spleen is rich in antibody-producing B cells; liver cells do not make antibodies.

"The result looks promising, but it still needs more work to show that [expression] is completely tissue-specific," Brinster cautions. Nevertheless, this is the first indication that a transferred gene might be expressed in mice, under