30 September 1960, Volume 132, Number 3431 SCIENCE

CURRENT PROBLEMS IN RESEARCH

Random Processes in Control and Communications

The application of some very pure mathematics to some very practical problems has led to new insights.

R. F. Drenick

In the last 15 years or so, some remarkable changes have taken place in the kind of mathematics used in applications and also in the applications which utilize mathematics. A field in which this change is particularly striking is that of communications and automatic control. The upheaval in the thinking of the engineers in this field is profound and apparently far from finished. They have had to adopt a new viewpoint and to learn to use a new set of mathematical tools. In return, they have succeeded in solving new problems and in finding new and more elegant solutions for some old problems. In some cases, areas have been uncovered whose existence was unsuspected until a short time ago, and new realizations have been obtained which have excited attention and interest well beyond the immediate field of control and communications.

An interesting feature of this upheaval is that it draws heavily on what is often called "pure" mathematics, and sometimes called, also, "abstract" or "modern" or even "useless." In any case, this is certainly the kind of mathematics which was developed originally for its own intrinsic interest. Nothing was further from the minds of the people who created it than the idea that it might ever be applied in practice, least of all to something as worldly and practical as communications and control engineering.

This article is a brief, sketchy (and therefore highly inaccurate) story of how theory and application got together in this field and of what came of the alliance.

The "Signal Space"

An engineer in the control or communications field deals fundamentally with two concepts, one which he calls signals and another which he calls transducers. The first are variable voltages, or currents, or similar physical quantities, such as, for instance, the current in some wire in a home television set. The second are devices which change the character of signals. The television set as a whole is such a transducer, and a relatively simple one at that. A digital computer is another and much more complicated one.

What must an engineer do who has decided to reform and to start dealing with his problems in the "new and modern" way? The first thing he should do is to go and see a mathematician and explain his problems. It will then invariably develop that an altogether exasperating language barrier exists between the two. What, the mathematician will ask blankly, is a signal? The engineer will then go to the blackboard and begin rhapsodizing on the subject.

He will say that signals are in a way both the raw materials and the finished products of his technology. He is used to visualizing them as plots of some quantity—voltage, say, versus time, as in Fig. 1*a*. Sometimes he thinks of them in a more stylized form, as what he calls a sampled signal. In this form, shown in Fig. 1*b*, the signal voltage is specified only at certain instants of time, called the sampling instants. This is much simpler than the complete version of Fig. 1*a* and almost as good. In either version, the time marked t is the "present time" beyond which the signal has not yet progressed. The rest the engineer often calls the "past" or the "history" (and sometimes, somewhat redundantly, the "past history") of the signal. He usually visualizes a history as very long, with its beginning way back in dim and unremembered ages, so that it can for practical purposes be taken to be infinitely long.

In some applications, the signal voltage can assume only a discrete number of values, as opposed to the preceding example where any voltage was possible. Such a signal the engineer calls "quantized," and the allowed signal levels are the "quantum levels." These levels can be labeled with numbers, or with letters, as, for instance, the signal in Fig. 2 (which employs a six-letter alphabet).

The engineer has now explained his mental picture of a signal, and he rests his case. The mathematician has begun to see the outlines of the problem. He tells the engineer that he strongly suspects a signal to be a case of what he and his colleagues call a random (or stochastic) process. (Now it is the engineer's turn to look blank.) The mathematician then calmly proceeds to give the engineer's mental picture its first severe jar by telling him that he ought to visualize a signal as a point in an infinite-dimensional space. Now, the engineer vaguely remembers from his college days that spaces come in various dimensions, usually one- or two- or three-dimensional. Four-dimensional spaces are rare but sometimes fashion-

The author is affiliated with the Bell Telephone Laboratories, Murray Hill, N.J. This article is based on a paper presented 28 December 1959 at a symposium on "Trends in the Application of Mathematics," sponsored jointly by Section A of the AAAS and the Society of Industrial and Applied Mathematics, at the Chicago meeting of the AAAS.

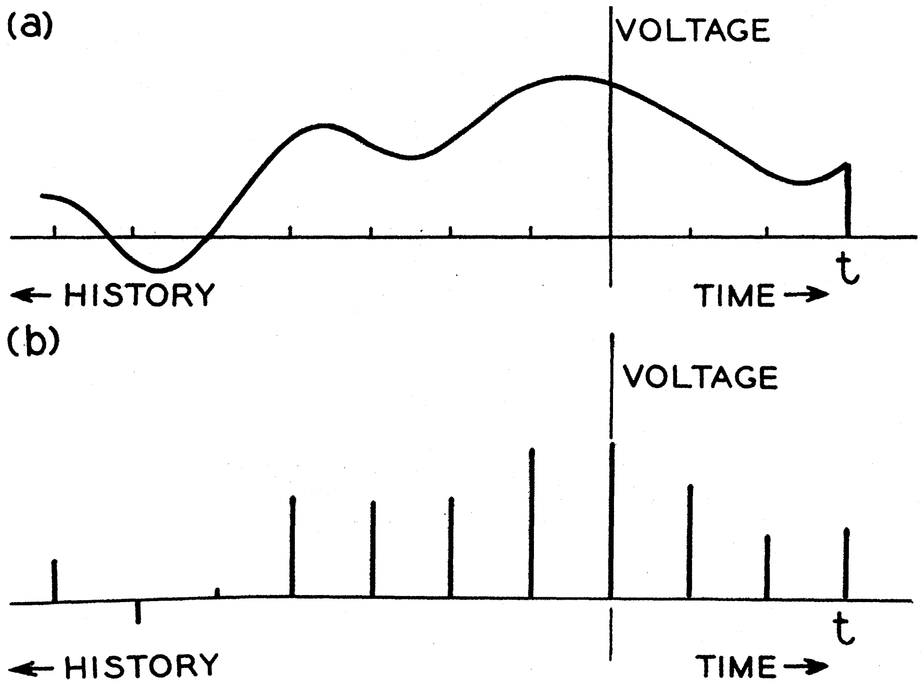

Fig. 1. (*a*) Plot of signal voltage versus time; (*b*) plot of a sampled signal, in which the signal voltage is specified only at certain instants of time.

able, and n-dimensional ones are possible if nearly inconceivable. Infinite-dimensional spaces, however, strike him as clearly preposterous. The mathematician, trying hard to hide his condescension, reminds the engineer that the dimensionality of a space is always indicated by the number of coordinates needed to fix the location of a point in it. The engineer concedes. The history of a sampled signal trailing away into the dim dark past is specified (for all practical purposes) by infinitely many voltage samples. The engineer concedes again. Each one of these samples may be considered a coordinate; hence, the signal is specified by infinitely many coordinates and can be thought of as a point in an infinite-dimensional space. Q.E.D., concludes the mathematician. The engineer gives in.

He does so reluctantly at first because he knows only too well from bitter experience that infinities must be approached with the utmost tact and respect, and he feels decidedly uncomfortable at first in his new "signal space." But he is soon engaged by the exhilarating simplicity of it. Every signal which can possibly appear in, for instance, a television receiver, westerns, commercials, dramas, whodunits, quiz shows, experts' panels, puppet shows—each is reduced (perhaps quite appropriately) to the insignificance of a point in a space.

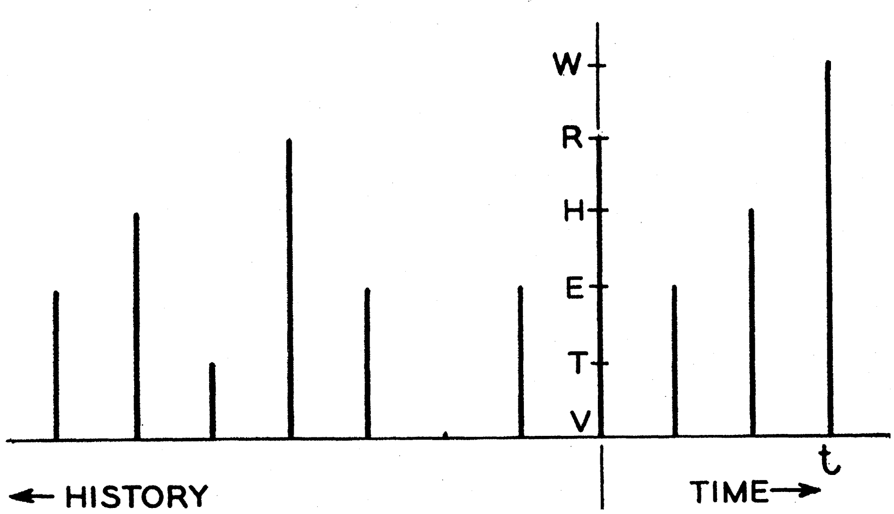

Fig. 2. Plot of a "quantized" signal, in which a six-letter alphabet is used to label the signal levels.

Probability Measures

However, the picture so far proposed by the mathematician lacks some important ingredients. One of its defects is precisely the relative insignificance of the various points of the signal space. For realism, the picture should allow for the fact that some kinds of signals tend to come up more often than others. Whodunits are more likely to appear than dramas, and westerns more likely than either.

The mathematician has just the right remedy for this difficulty: he proceeds to cover the whole signal space, with calm disregard for its magnitude, with what he calls a "probability measure." Like a tremendous blanket, this measure covers every set of points in the space and assigns a certain weight to it—namely, the probability of its occurrence. It assigns a definite probability to all westerns, and among these a definite, though smaller, probability to all psychological westerns, and a smaller one yet to all psychological westerns containing barroom brawls in which a chair is thrown through a great mirror.

Probability measures come in many forms and textures. Some are nice and smooth, as blankets ought to be, but others are lumpy or full of holes (quantized signals, such as the one in Fig. 2, have measures that consist of nothing but lumps). There is one class of measures, however, which are great favorites, partly because of their smoothness and partly because of several other engaging properties. These are the so-called "Gaussian" measures. They are straightforward generalizations of the Gaussian, or normal, probability distribution which is widely used in statistics and which has the well-known bell shape shown in Fig. 3. In communications problems, this curve shows the probability of occurrence of a single non-quantized signal voltage, plotted against that voltage itself.

One can ask next what the joint Gaussian probability of occurrence of the voltages in two adjacent signal samples looks like, and the answer is, a two-dimensional bell shape like the one in Fig. 4. It, too, is a concept much used in statistics. The exact shape of the bell varies from case to case. Typically, it is elliptical. It can be circular, however, as in Fig. 4, and this is a particularly useful case—namely, one in which the two adjacent signal voltages are statistically independent.

One can now go on to three-dimensional bell shapes and four-dimensional

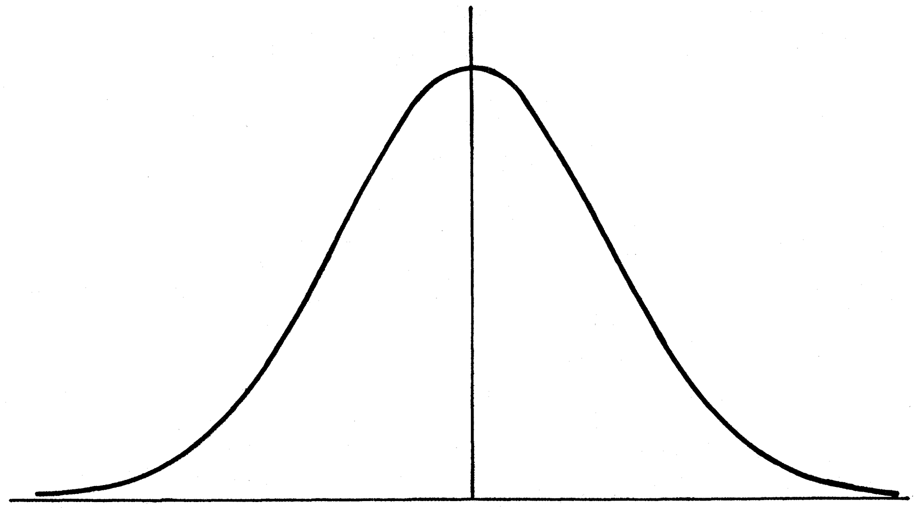

Fig. 3. Plot of the Gaussian probability distribution for a single nonquantized signal voltage.

ones, and so on. The Gaussian probability measure is, in a way, an infinite-dimensional bell shape. It, too, can be elliptical or circular, and in the latter case all signal voltages are statistically independent. Engineers have come to call this the "white" case (for a reason which would be hard to explain here and which is not very good anyway).

Gaussian measures in general, and white ones in particular (Gaussian and otherwise), will figure prominently in our story.

Borel Fields

The idea of a probability measure usually fits rather nicely into the intuitive notion an engineer has of a signal space, and there is usually little trouble in persuading him to adopt it. Little does he know that he has let himself be talked into an upheaval in his way of thinking.

The cue he missed lies in the word *set*, which was slipped in unobtrusively at the start of the preceding section. As it happens, single points in signal spaces are usually quite uninteresting. A point is, for instance, a very long television program, specified to the minute detail of every one of its signal-voltage samples. Two such programs which are exact copies of each other except for a small difference in one of their signal voltages are two different points in signal space. It is clear that a single point is too insignificant an item to have any appreciable probability of occurrence. It is only sets of many points that are worth talking about.

As far as the engineer is concerned, this means that he must learn to think in terms of point sets in signal space. This will be a new and possibly traumatic experience for him. The mathematician will tell him that sets can be operated on by forming unions between pairs of sets, and also intersections, and that in fact not only pairs but finitely and infinitely many sets can be involved in these operations.

The mathematician will further reveal that he and his colleagues like to work with families of sets, which they call Borel fields. These families are so large to begin with that the forming of unions and intersections, no matter how often it is repeated, can never produce extrafamilial offspring. This is a safety precaution rather than some form of mathematical morality, for it allows the mathematician to perform set operations ad infinitum without having to worry about straying outside the family.

To a mathematician, it is the combination of the three concepts—the signal space, the probability measure, and the Borel field—which makes up what he calls a "random process." Thus, if the measure is Gaussian, he will speak of a Gaussian random process, and if the sample voltages are statistically independent, he will often call it a white random process. The engineer, by now totally overwhelmed by the flood of new concepts, can think of no reason for not doing the same. He now calls his signals random processes, and is proud of them.

The feeling of comfort and togetherness, however, which pervades a mathematician when he is working in Borel fields usually fails to envelop the engineer. On the contrary, he finds Borel fields irritatingly useless and completely unmanageable. They have turned his neat signal space into a nightmare of blobs, called sets, which are continuously and uncontrollably splitting up and melting together again. The solid ground to which he is accustomed, and on which he is widely known to have both feet, has been replaced with some shifty quicksand-like material. Borel fields, he feels, may be all right for mathematicians but they definitely are not the stuff on which to build a technology.

On closer inspection, however, they often reveal some surprising structure which is as permanent and firm as structures can be. This comes from the

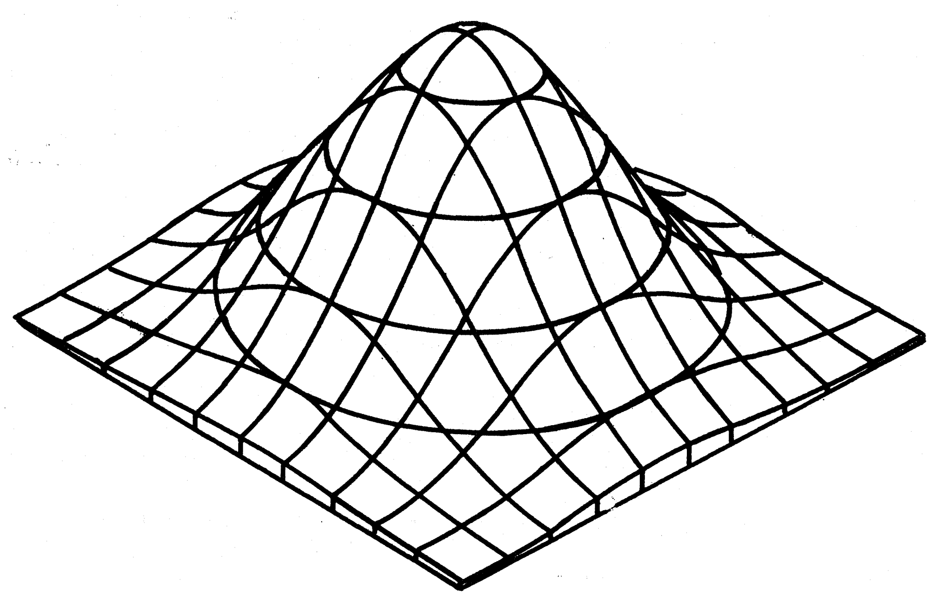

Fig. 4. Plot of the joint Gaussian probability distribution for nonquantized signal voltages in two adjacent signal samples.

fact that a Borel field consists of sets which consist of histories which, in turn, consist of many samples. It is these samples which are in the end responsible for the structure. Every history contains a most recent sample, which is the most recent event in the past of that particular signal. Events such as this one, skimmed collectively from all signals in the space, form a layer of sets which lies across the Borel field of all histories like a thin sediment deposited by time on a huge bed of earlier ones. A similar layer, just beneath the most recent one, corresponds to the sample immediately prior to the most recent one; and so on. What results from this is a distinct stratification of the Borel field into infinitely many layers, one layer per sample of history.

The mathematician, or engineer, can now engage in what can only be called "mathematical archeology." An archeologist digging into the site of some ancient dwelling place often removes layer after layer of successive cultures, each layer telling through its artifacts and ashes some of the history of the locale. Similarly, the mathematician can peel layer after layer from his Borel fields and work deeper and deeper into the history buried in the signal space. Archeologist and mathematician alike must work carefully to be sure not to disturb an earlier layer whenever they remove a later one. To the mathematician, this means that he must make sure of the statistical independence of a layer before he dare remove it. Other than that, however, he has few troubles. He need not worry about sandstorms, theft, or local revolutions. He can dig without obstacle further and further into the past, until in the limit he comes to the beginning of time.

At this point, however, there is an awesome surprise awaiting him. He expects, no doubt, that when he reaches it, history will be exhausted, the Borel fields will be empty. But this need not be so. It can happen, and often does happen, that there remains a residual Borel field, remote and unapproachable, which has some almost apocalyptic properties. It is quite unlike the layers that he has removed previously. The signals which issue from it are perfectly predictable. If they ever existed at all, they continue to exist and evolve in a precise and foreordained fashion through all the layers of history up to the present and on into the future. Nothing that can happen in the course of history can in any way affect them. Once sounded, they ring forever true. This prehistoric Borel field is often called the "remote past" of a random process.

Transducers

Engineer and mathematician have now come to see eye to eye on the subject of signals, the raw materials and finished products of the technology. They must next get together on the machines that work on the materials. As we have said earlier, these are the transducers; they are the devices which convert one signal into another.

A transducer, the engineer will explain, has one signal going into it, called the input, and another one coming out of it, called the output. (Sometimes there are more than one of either.) Each possible input is a point in a signal space, which is now, understandably, called the input space; the outputs are points in an output space.

Many transducers have two outstanding properties. (i) They are "retrospective" (also called "causal"). That is, the present output cannot depend on the future of the input. This is quite an essential property. It says in effect that no physical device can respond to a stimulus that has yet to be applied. (ii) They are one-one, which means that every point (signal) in input space corresponds to one and only one point (signal) in output space. This is not really as essential a property as the first and is actually violated quite often in practice.

To the mathematician, a transducer is a transformation and a rather special one at that. It transforms input space into output space and, in the process, transforms the probability measure and the Borel field of input histories into their counterparts for the output. He can truly say that much is known in the mathematical theory of transformations, but he will also have to concede that much more could be known.

Great favorites among engineers are the so-called linear transducers, more often referred to as "linear filters." They are the easiest to design and the cheapest to build, relatively speaking, and the control and communications systems now in existence use them lavishly. Linear filters have the pleasant theoretical property that the sum of two input signals generates the sum of the corresponding output signals. This is called the superposition principle. It may not seem essential, but the fact is that it makes linear filters particularly easy to deal with. The standard procedure used by engineers for analyzing filters consists of resolving the input signal into many sine waves and tracing it through the filter, sine wave by sine wave.

The mathematician's equivalent to a linear filter is a linear transformation, obviously, and the mathematical theory of linear transformations has seen a vast amount of research, from some humble beginnings 200 years ago and more, up to some highly sophisticated and abstract work in recent decades. It is not overly surprising, therefore, that the first contact between the engineer and the mathematician took place in this area of random process theory.

Wiener's Prediction Theory

The contact was established by a small booklet in yellow paper cover (promptly dubbed the "Yellow Peril" by engineers), written in 1942 by N. Wiener. He was the first mathematician to diagnose some of the engineer's problems, or at any rate the first one to suggest a treatment along the lines I have described above. He prescribed a healthy dose of signal spaces, complete with Borel fields and probability measures, and some gentle applications of transformations. From among the latter he chose the engineer's favorite, the linear filters. Under these circumstances, as he further pointed out, the most appropriate probability measure is the Gaussian, which I mentioned above. The reason for this lies essentially in one fact. It so happens that linear filters preserve the Gaussian character of a measure: if it is Gaussian over input space then it will be Gaussian also over output space.

To the engineer, this would have been a poor argument indeed, and it would never have persuaded him to abandon his sine wave approach to linear-filter problems for the statistical one. But Wiener further showed him that a Gaussian random process could be specified by a quantity (the so-called "spectral density") which is very closely related to his cherished sine waves. What suspicion remained in the engineer's mind was swept away by a final demonstration—namely, that he could calculate by a very simple formula the spectral density of the output of a linear filter from that of the input. In other words, the engineer was shown that, by adopting Wiener's approach, he could do many of the things he had been doing before and that he could do them in a conceptually much more satisfying way.

Beyond that, however, the new theory provided him with some novel and most revealing insights into the foundations of his technology. It showed him that there are some things which he can hope to accomplish with linear filters, disregarding perhaps certain practical complications, but that there are others he cannot under any circumstances hope to achieve.

He cannot, for instance, hope to ever build a filter which will generate an output with a remote past from an input that lacks one. In order to do that he would have to construct a filter that anticipates the future (that is, one which violates the basic requirement of causality), and this is no more possible than building a perpetual-motion machine.

Beyond generating remote pasts, however, there is little that linear filters cannot do to Gaussian random processes. In fact, one of the central theorems is that, given any two such processes (both without remote pasts), one can always find a linear filter which will accept one of them as input and generate the other as output, and vice versa. One can put this slightly differently by saying that all Gaussian processes form a class whose members are freely transformable into each other by linear filters. It is then customary to let one member of the class represent all others, and a particularly distinguished senior member is picked for the purpose—namely, a white Gaussian random process.

The main objective of Wiener's theory was the problem of predicting the future of a Gaussian random process from knowledge of its past. Wiener showed first of all that the best predictions could always be made by a linear filter. It is useless to try to improve on that by any devices, no matter how ingeniously they are contrived—and this includes tea leaves, gypsy women, and astrologers. A properly designed linear filter simply cannot be beaten for quality and consistency of performance when it comes to predicting Gaussian processes.

The predicting filter accepts as input the process that is to be predicted and generates as output the best guess of what the future will be. The guess will usually be in error by some amount, but on the average it will be as good as or better than any other forecasting procedure for Gaussian processes. In one case, however, the predicting filter will be dead right every time—namely, when the process originates completely in the remote past. For in this case, as we have said before, the process is perfectly predictable to begin with.

In either case, the filter goes about its guessing business in such a clever way that gypsy women might have trouble imitating it. It first transforms the input into a white process, the senior member of the class—and this is always possible, as we know. In the white form, a novel feature appears: part of the original input turns out to be perfectly predictable and part of it completely unpredictable. The predicting filter now proceeds with cold logic. It bases its forecast on the perfectly predictable part exclusively and ignores the rest. Since the rest is totally unpredictable, it is useless anyway, and the prediction error caused by its omission is unavoidable.

All this is concerned with Gaussian processes and linear filters only. The extent to which Wiener's theorems apply to more general processes and transducers is not known. Many people suspect that most of them carry over, but no satisfactory theory exists which demonstrates this.

Shannon's Information Theory

There is, however, a handsome body of theory which holds for very general processes and transducers. This is C. E. Shannon's information theory.

The questions which led to the development of information theory are different from those dealt with by Wiener. One starts out by assuming again that a certain type of signal is given, or, more precisely, a random process complete with signal space, probability measure, and Borel field. It need not be Gaussian. In fact, it will be convenient in what follows to assume that the signal is a quantized one, looking, for example, like the one in Fig. 2. It is true, as I mentioned above, that this forces on us an extremely lumpy probability measure and some attendant discomfort, but it does make explanations here a little easier.

Information theory was developed to answer questions such as these: Is it possible to define some quantity, some single number, which would measure somehow the information conveyed on the average per sample of a signal? What properties should this quantity have and in what units should it be measured? How is it affected when the signal is passed through a transducer? How is it affected if it is passed through a communications channel with noise—that is, a channel in which the signal is distorted, mutilated, or drowned in interference? Are there ways of combating the effect of noise?

The answer to the first question is affirmative. Intuitively we know that the information conveyed by a signal lies in its surprise value. The message "Fire!" in a movie house is more informative than "Coming Attractions." The more surprised one is over receiving a particular signal, the greater, one can say, is the amount of information conveyed by it. On the other hand, the more certain the future value of a signal sample is, on the basis of the history that has preceded it, the less surprising it is, and the less information it conveys when it actually arrives. A signal, for instance, with seven quantum levels labeled "Monday," "Tuesday," and so on, conveys no information at all when it tells us that today is Tuesday if we knew already that yesterday was Monday.

These intuitive notions of the information content of a signal can be put in a more definite form. We now surmise that a signal will be most informative if, for one thing, all of its samples are statistically independent or, to use our present terminology, if it is white. The known history will then in no way compromise the surprise value of the sample scheduled to arrive next. For another, the probability of occurrence for the quantum levels of an individual sample should be as nearly equal as possible. On the other hand, if the history of a signal completely determines its future, if nothing can possibly happen that could not be anticipated, or in still other words, if the signal is perfectly predictable, then it conveys no information at all.

Surprisingly, these two requirements (the conditions under which it should be a maximum and those under which it should be zero), plus a third one of somewhat more involved nature, suffice to define uniquely a quantity which measures the mean amount of information per sample of a random process. As it happens, it is almost exactly what physicists since the days of Boltzmann have been calling the "entropy."

The main achievement of information theory, however, lies not in merely defining a quantity of information but in showing that it has some most remarkable properties. To begin with, the theory shows that this quantity remains unaffected when a signal is passed through a transducer. This is most satisfying because one feels intuitively that a one-to-one transformation, no matter what it does to a signal, should not affect the information content.

The second important result is this: given two random processes (signals) with the same information content per sample, it is always possible to find a transducer which transforms one signal into the other.

The practical difficulties of constructing the necessary transducer may be quite formidable. But it is clear that, in principle at least, all signals with equal information content form a class whose members are freely negotiable among themselves through the use of suitable transducers.

This theorem is very reminiscent of one of Wiener's which I have mentioned, and the similarity goes in fact even further. For, among all these negotiable class members, there is one particularly distinguished senior member, and this is the one which is white and which has quantum levels of equal probability of occurrence. This is the one which, as I said above, carries the most information per sample. We can now say, further, that it is the one member in the class which can get along with the fewest samples per second or with the fewest quantum levels per sample. It is the most efficient member of the class. Where the others are "redundant," in the engineer's vernacular, this one wastes no words and gives its messages in the most laconic and abrupt of telegraphic styles.

The usefulness of this class concept lies in the following fact: Many communications channels, in practice, can accept only certain kinds of signals, and only at certain limited sample rates. In other words, a channel can be characterized by the kind of signal which it is designed to carry and by the sample rate which it can accept. However, along with this design signal the channel can be forced to carry any other signal, provided only that it belongs to the same class as the design signal. For, as we now know, any signal from this class can be freely transformed (or "encoded," as engineers are apt to say) into the design signal by a suitable transducer. This assertion is the essence of one of Shannon's famous coding theorems—namely, the one for the so-called noiseless channel.

Another, even more famous, coding theorem applies to noisy channels—that is, to channels in which the signal is distorted by some interference. It stands to reason that in such channels some information will always be lost because of the interfering noise, and this is actually so. One can, and in fact does, express the signal degradation by a single number—namely the loss of information content, and calls it the "equivocation" due to noise.

It is tempting at this point to speculate as follows. Since a noisy channel will inevitably lead to some loss of information, some equivocation, in the signal, is it perhaps possible to "enrich" the signal artificially prior to its transmission? In other words, can one somehow raise its initial information content by injecting into it some extra but irrelevant information and then let the channel degrade it? It is conceivable that the artificially enriched signal would release to the channel only that information which it did not need in the first place and thus would emerge on the other end as clean and undefiled as it started out. Such a scheme is in fact possible, and this is the assertion of the second coding theorem.

Unfortunately, it is easier to say that this can be done than to find how to do it. The trouble lies with the preparation of the signal prior to transmission. The problem is not only one of how much extra information to inject into the signal prior to its ordeal by transmission. What matters more is that the enriched signal must be fortified against all possible adversities which it may encounter during its passage through the channel. This is somewhat like the speeches of highly placed government officials which ought to be so written that they cannot be misunderstood, no matter how far out of context they are quoted. These problems are very difficult, by all indications, and no altogether satisfactory solutions have been found for them, either by mathematicians working in information theory or by ghost writers working on official speeches.

Conclusion

Our story started out with a main character, the engineer, in search of a solution to his problems. In the first few chapters he played a prominent, if somewhat helpless, role, but toward the end he faded into the background and left the stage to the mathematician. This is no accident but a part of the story's plot. The engineer has in fact returned to his laboratory. He is now engaged in overcoming the countless vicissitudes and irritating complications which must be overcome whenever the grand and sweeping theories of the mathematician are reduced to practice. In the process, he has had to develop many variations on the original theories, some of them similar to their prototypes but others having only the vaguest of resemblances. His patient (and largely anonymous) labor, however, has found the most varied realizations, in automatic pilots and guidance systems, in oil refineries and nuclear reactors, in radars and navigation systems, in telephone and television, and in many others.

The engineer has been joined in his labors by some unexpected and unprecedented company. The theories which had been precipitated mainly by his problems have attracted the interest and activities of physicists, biologists, neurologists, psychologists, and linguists, not to even mention mathematicians. In all these fields the theories have raised at least as many questions as they have answered, and there are now, it seems, more frontiers that are waiting to be conquered than there are conquerors.