AN APPLICATION OF DIGITAL FILTERING

An optimal adaptive digital filter based on an exact relationship between homogeneous differential and difference equations locates and then tracks the "instantaneous frequency" of a signal in noise. An example of the performance of a programmed model of this filter is given.

R. P. Rich and H. Shaw, Jr.

An analog filter is an electrical circuit so designed that a certain selected range of frequencies is transmitted while other frequencies are almost entirely suppressed. Such filters are continuous in the sense that corresponding to each instant of time there is an input value and an output value. The lumped parameter model is, for most engineering purposes, an adequate representation of an analog filter. When the lumped parameters can be considered constant, the analog filter may be described by a linear ordinary differential equation with constant coefficients.

Upon comparing the classical solutions of difference and differential equations (homogeneous with constant coefficients), it is clear that either such equation may be substituted for the other without truncation error or change of order. That is, any solution of the one agrees exactly with a solution of the other at those time points (sample points) where the solution of the difference equation is defined. This substitution is effected by means of a simple equivalence relation between the roots of the respective auxiliary equations and by proper choice of initial conditions.

All this suggests the possibility of replacing analog filters by digital filters. A digital filter is a digital computer either programmed or especially built to transform sampled (discrete) input, via nonhomogeneous difference equations, into discrete output. The chief advantage of this replacement is the ease of changing numbers in a digital computer compared to changing components in an electrical circuit. The chief disadvantage is the introduction of an additional parameter, that is, the sampling interval or time between successive sample points. The consequences of sampling are, however, well understood[1] and present no difficulty if the sampling rate is properly chosen.

In recent years a number of applications of digital filters has appeared.[e.g. 2, 3] It is our purpose to present another application, an adaptive

[1] J. W. Tukey and R. B. Blackman: *The Measurement of Power Spectra,* Dover Publications, Inc., New York, 1958, p. 120.

[2] M. J. Levin, "Generation of a Sampled Gaussian Time Series Having a Specified Correlation Function," *I.R.E. Trans.* IT-6, 1960, 545–548.

[3] E. J. Gauss, "Estimation of Power Spectral Density by Filters," *J. Assoc. Comp. Mach.,* 11, 1964, 98–103.
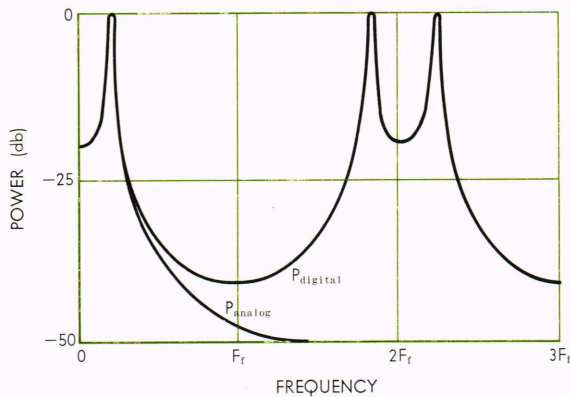
Fig. 1 — Comparison of power characteristics for analog and digital second order filters. $Q = 10$ and $F_c = F_f/5$, where F_c is the filter center frequency and F_f is the folding frequency.
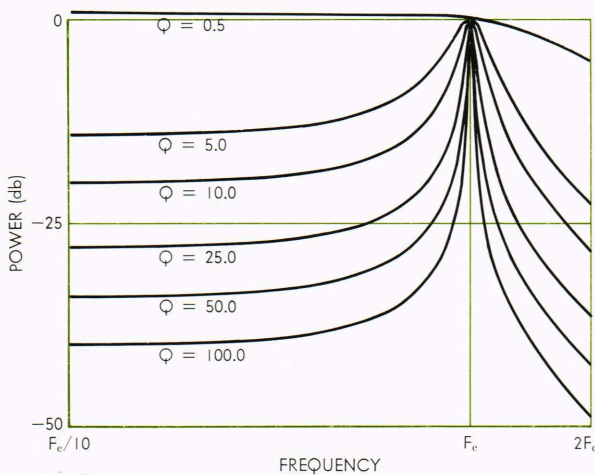

Fig. 2—Power characteristics of simple second order digital band-pass filters for various values of Q. It is assumed that the sampling rate is such that $F_c\Delta T \leq 0.1$, where ΔT is the sampling interval.

second order band-pass filter that tracks the instantaneous frequency of a signal in noise. The center frequency and bandwidth of the filter adjust according to a short-time average and standard deviation, respectively, of the time between successive zero crossings of the filter output. The filter is optimal in the sense that the signal-to-noise transfer for a sinusoid in white noise is maximized.

We first describe a class of digital filters suggested by the equivalence relation. We then specialize to a second order band-pass filter and maximize the signal-to-noise transfer. Finally we present the adaptive filter, a programmed model of the filter, and a simple example of its performance.

A Class of Digital Filters

A class of k-th order digital filters or, more precisely, nonhomogeneous difference equations, has the form

$$R_n - \sum_{m=1}^{k} A_m R_{n-m} = A S_n.$$

Here S_n and R_n are the discrete input and output (signal and response), respectively, at time $T_n = n \cdot \Delta T$, where ΔT is the sampling interval. It is worth noting that this form can be handled very economically by a digital computer. The response at T_n is just a linear combination of the signal at T_n and the responses at $T_{n-1}, T_{n-2}, \ldots, T_{n-k}$. In view of the equivalence relation, it is natural to expect a similarity between these digital filters and the corresponding k-th order analog filters.

Filters of this class are characterized by their frequency response functions,[4]

$$F(f\Delta T) = A \bigg/ \left(1 + \sum_{m=1}^{k} A_m e^{-2\pi jmf\Delta T}\right).$$

The power characteristic of a filter, that is, the square of the amplitude of the steady-state response to a sinusoid of unit amplitude and frequency f sampled at intervals of length ΔT, is just the square of the modulus of F; that is,

$$P(f\Delta T) = |F(f\Delta T)|^2.$$

It is important to observe that P is periodic in $f\Delta T$ with period 1 and displays half-wave symmetry. Thus, for a given ΔT, the plot of P against f can be folded back to exactly coincide with itself along any of the ordinate lines through integral multiples of $F_f = 1/(2\Delta T)$ (see Fig. 1). For this reason F_f is called the folding frequency.[1]

Clearly the problem of determining, by digital means, what frequency in a signal produces a given response becomes ambiguous if the signal contains frequencies higher than the folding frequency. This difficulty, innate in sampling, is readily overcome by filtering out the offending frequencies prior to sampling.

A Simple Band-Pass Filter

The simplest digital band-pass filter is given by the second order difference equation

$$R_n = A_1 R_{n-1} + A_2 R_{n-2} + A S_n.$$

The equivalence relation between homogeneous differential and difference equations suggests taking

[4] J. R. Ragazzini and G. F. Franklin, *Sampled-Data Control Systems*, McGraw-Hill Book Co., Inc., New York, 1958, pp. 72 and 111.

and

$$A_1 = 2e^{-K\Delta T} \cos(2\pi F_c \Delta T)$$

and

$$A_2 = -e^{-2K\Delta T}.$$

The homogeneous equation, properly initialized, will then generate a damped sinusoid with time constant $1/K$ and frequency F_c. The power characteristic has a single peak between zero and the folding frequency and is nearly symmetric near F_c, the center frequency of the filter.

It is somewhat tidier to take ΔT as the unit of time, that is, to measure time in number of sample points. Frequency is then in units of cycles per point, and we can rewrite

$$A_1 = 2e^{-1/B_{tc}} \cos(\pi/B_{chp})$$

and

$$A_2 = -e^{-2/B_{tc}},$$

where B_{tc} is the band-pass time constant in points and B_{chp} is the band-pass center half period in points.

The filter can be normalized so that the power characteristic is unity at the center frequency by taking

$$A = (1 + A_2) \sqrt{1 - (e^{1/B_{tc}} A_1{}^2)/(1 + e^{-1/B_{tc}})^2}.$$

Figure 2 is a plot of power characteristics for several values of Q. The parameter

$$Q = \frac{\pi}{2} \frac{B_{tc}}{B_{chp}},$$

well known from analog filtering, is approximately the ratio of the center frequency of the filter to the width of the power characteristic at half power.

Optimization

The input of a filter consists of a desirable and an undesirable component. The desirable component is generally called signal while the undesirable component, if more or less random, is called noise. Signal power is usually concentrated around one frequency (or several frequencies) while noise power tends to be spread out more or less uniformly over a range of frequencies.

An extreme example is a sinusoid plus white noise, that is, noise with a flat power spectrum. In this case, the band-pass filter output signal-to-noise power ratio can be made as large as desired by centering the filter at the sinusoid frequency, assuming this known, and taking Q sufficiently large. If, however, the filter is centered at some frequency other than the sinusoid frequency, Q must not be taken so large as to lose the signal altogether. This suggests maximizing the signal-to-noise transfer, $T_{s/n}$, as a function of Q, given the signal frequency, F_s, and F_c.
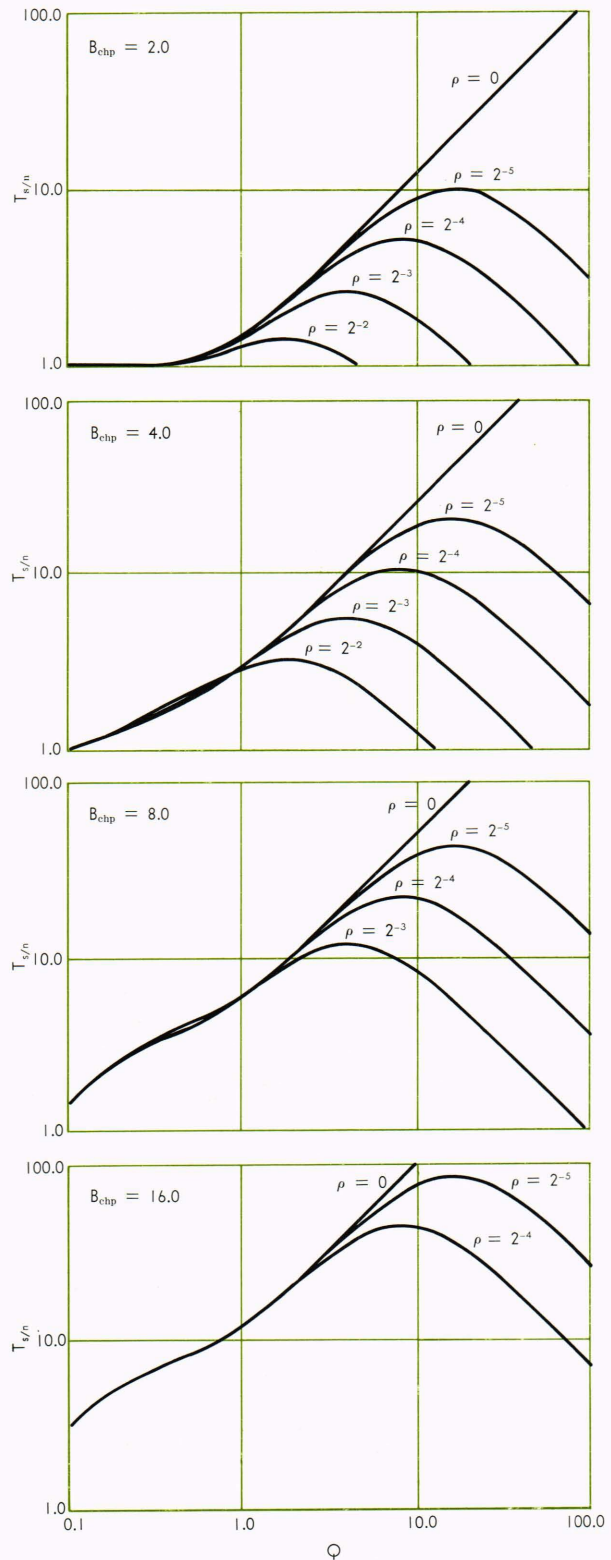

Fig. 3—Signal-to-noise transfer, $T_{s/n}$, as a function of Q for various values of $\rho = (F_c - F_s)/F_c$ and filter centerings B_{chp}. F_s is the sinusoidal signal frequency.

Now, by the signal-to-noise transfer of a filter we mean the ratio of the output signal-to-noise power ratio to the input signal-to-noise power ratio. Thus, for our second-order digital filters,

$$T_{s/n} = \frac{P(B_{chp}, B_{tc}, 1/S_{hp})}{\int_0^1 P(B_{chp}, B_{tc}, \xi)d\xi},$$

where S_{hp} is the signal half period in points. The integration can be carried out by partial fractions and the results simplified somewhat to give

$$T_{s/n} = \tanh\alpha\left[\frac{(\cosh\alpha\,\sin\beta)^2 + (\sinh\alpha\,\cos\beta)^2}{\cos^2\gamma - 2\cosh\alpha\,\cos\beta\,\cos\gamma + \sinh^2\alpha + \cos^2\beta}\right],$$

where $\alpha = 1/B_{tc}$, $\beta = \pi/B_{chp}$, and $\gamma = \pi/S_{hp}$.

Figure 3 consists of plots of $T_{s/n}$ versus Q for various filter centerings and values of

$$\rho = (F_c - F_s)/F_c,$$

where ρ is relative displacement of the sinusoid frequency from the filter center frequency.

Referring to Fig. 3 and noting that

$$F_c = 1/(2\,B_{chp}\,\Delta T)$$

and

$$F_s = 1/(2\,S_{hp}\,\Delta T),$$

it is readily verifiable that

$$Q_{optimum} \doteq \frac{\frac{1}{2}\,S_{hp}}{|S_{hp} - B_{chp}|}.$$

An Adaptive Digital Filter

It is now easy to describe an optimal adaptive digital filter for tracking the instantaneous frequency of a signal in noise. Such a filter would adjust its centering, B_{chp}, according to the best available estimate of signal half period, S_{hp}, and its Q optimally according to half the ratio of S_{hp} to the best available error estimate.

By instantaneous frequency we mean half the reciprocal of a short-time average of the times between successive zero crossings of the filter output. This interpretation of the useful though somewhat self-contradictory term "instantaneous frequency" is consistent with the phase derivative of frequency modulation and with the time between compression nodes of the doppler effect. Zero crossings are
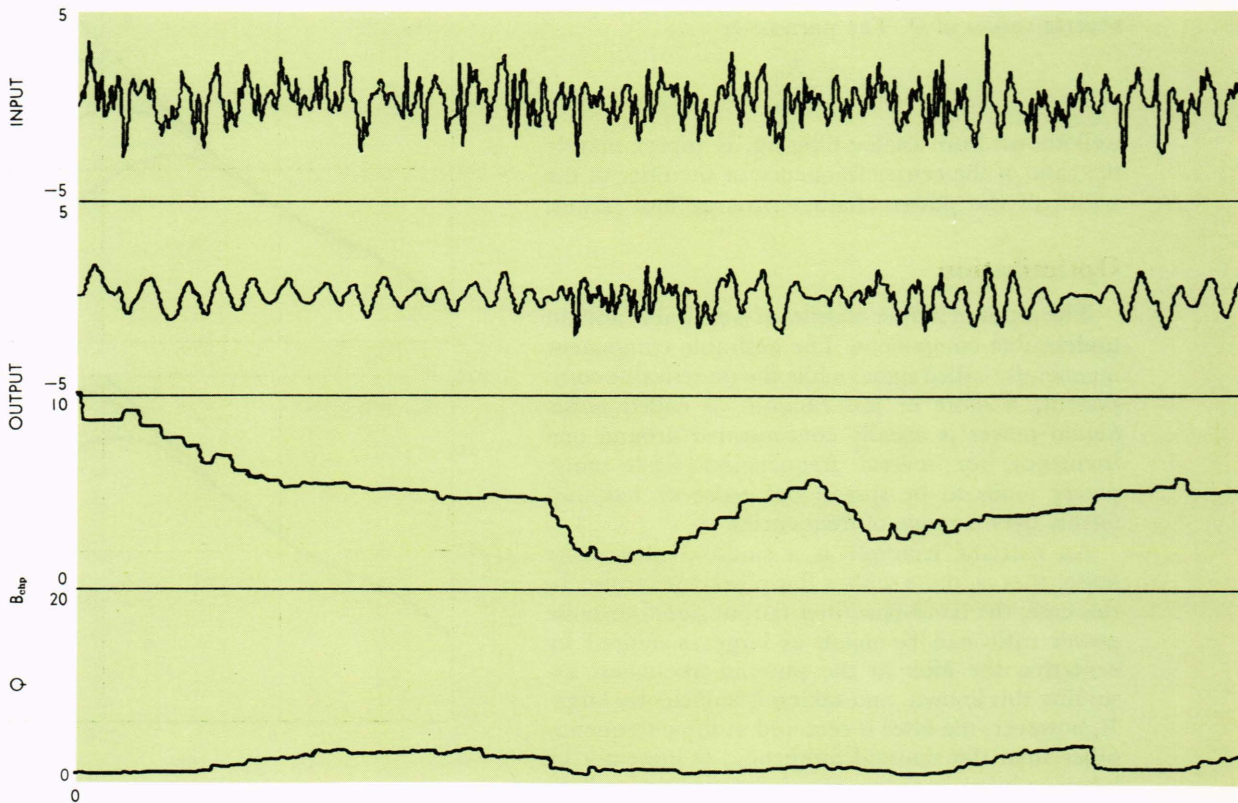

Fig. 4—Example of filter performance for an input consisting of a sinusoid sample

readily estimated by linear interpolation between successive sign changes of the filter output.

Thus, the filter may be centered initially at an estimate of the signal frequency. The width of the filter may be set initially to correspond to the judged quality of the estimate. Thereafter each time the filter output crosses zero, the filter parameters adjust to take account of the new information.

Particular refinements of the filter will probably occur to the reader. Prefiltering based on the noise spectrum, for example, is generally desirable. It is obviously essential to filter out DC bias.

Example of an Adaptive Filter

A model of the adaptive filter has been coded in FORTRAN II for the IBM 7094 computer. This model consists of subroutines that perform the basic operations of averaging, band-pass filtering, zero detecting and half period computation, and logical sequencing. Since each subroutine performs its own initialization and resetting, logical sequencing amounts to advancing time by unity (time is measured in points), plotting, and either terminating or proceeding to process the next point.

Each example is completely defined by a brief FORTRAN control program that specifies the subroutine parameters, defines the output (which quantities are to be plotted against time and what scaling is to be used), and calls the subroutines in their proper order. A machine run produces a listing of the control program and a plot. Plotting is performed via library subroutines that produce a tape that controls a CALCOMP plotter.

This program structure is especially convenient for running a large number of experimental examples. A listing for each subroutine, together with a brief control program listing and a plot for each example, furnishes complete and automatic documentation. Since, for the purpose of this paper, complete documentation is unnecessary, we substitute a description of the method of averaging, an indication of the organization of a typical control program, and an example of the performance of the model of the adaptive filter.

Both short-time half period and standard deviation averages are performed in the same way by combining the arithmetic mean of the more recent values with an exponential average of the preceding values. The object of this combined average is to

400   800
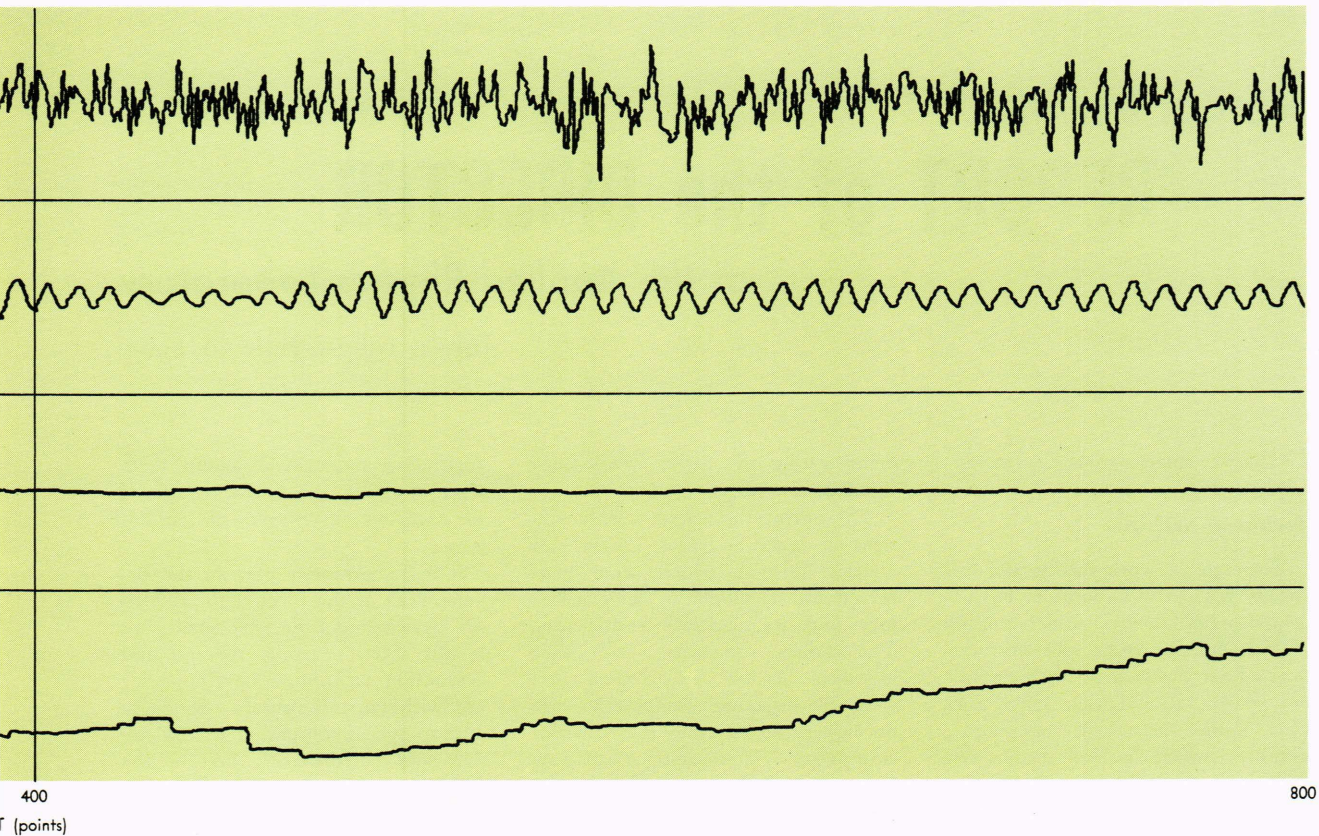
T (points)

5 points per half period plus Gaussian white noise. Signal-to-noise ratio is 3/16.

produce some stability and yet to retard ringing (too much stability). An obvious choice for the time constant of the exponential average (a first order low-pass digital filter) is the time constant of the band-pass filter in half periods. While a five-point arithmetic mean produces good results, this choice does not appear to be at all critical.

A typical control program is organized as follows:

1. INITIALIZE
2. DO ONE POINT
   a. AVERAGE
   b. COMPUTE BAND-PASS PARAMETERS
   c. INJECT SIGNAL VALUE
   d. PREFILTER
   e. BAND-PASS FILTER
   f. COMPUTE HALF PERIOD
   g. DEFINE OUTPUT
3. SEQUENCE

As an example consider a sine wave with a 5-point half period, $S_{hp} = 5$, plus white Gaussian noise with mean zero and variance (total average power) one. The noise power spectrum is thus flat between zero and the folding frequency. No pre-filter is used. The signal-to-noise power ratio is $3/16$, that is, the signal is 7.3 db below the noise. Input amplitude, output amplitude, B_{chp}, and Q are plotted against time in Fig. 4. The filter moves from 10 to 5 points per half period and Q increases while the output becomes appropriately sinusoidal.

Conclusion

By combining the analog technique of filtering with the flexibility of digital computing we have been able to present an adaptive digital filter that could prove to be a useful new tool in the study of signals in noise.

We feel that the residual qualitativeness of our presentation may be more instructive to digital computer programmers than detrimental to engineers. It is, perhaps, as R. W. Hamming writes throughout his book,[5] "The purpose of computation is insight, not numbers."

[5] R. W. Hamming, *Numerical Methods for Scientists and Engineers,* McGraw-Hill Book Co., Inc., New York, 1962, pp. vii, 276, and 400.

Excerpts from the

REPORT of the DIRECTOR

of the Applied Physics Laboratory

July 1, 1963—June 30, 1964

TO THE PRESIDENT OF THE UNIVERSITY:

Technical Activities

Through its work during the last twenty years, the Laboratory has acquired a fundamental understanding of the characteristics and performance of three classes of Navy systems now in operational use. These are: (1) surface-launched, shipborne, anti-air warfare, guided missile systems, (2) fleet ballistic missile systems, and (3) satellite navigation systems. It is interesting to note that all of these systems have come into being within the fifteen years since the Korean War.

The surface-launched missile systems are based on Talos, Tartar, and Terrier missiles which were conceived and developed by the Laboratory and its associate contractors. The radars, computers, and other elements of the fire control and weapons direction equipment were developed for the Navy by industrial contractors. These systems are now deployed on 56 cruisers, frigates, and destroyers. The Tartar and Terrier missile systems are also installed on ships of the French, Italian, and Australian navies who chose them after studying the performances of the world's inventory of similar items.

The Laboratory's part in the development of the Fleet Ballistic Missile System has been very small, but it was asked to apply the test and evaluation techniques developed for use with its anti-aircraft systems to a thorough investigation of the design and performance, first of the Polaris prototype systems and subsystems and finally of the operational system itself.

The Satellite Navigation System