17 November 1961, Volume 134, Number 3490 SCIENCE

INSTRUMENTS AND TECHNIQUES

Stellar Scintillation

Stellar scintillation is shown to be a good indicator of the upper-air winds near the tropopause.

W. M. Protheroe

Except for dilution by the intervening space, the light from a star suffers little until it impinges upon the earth's atmosphere. In the short time required for the light to traverse the atmosphere, what was originally a plane-parallel wavefront becomes transformed into a corrugated wavefront by the refractive inhomogeneities which it encounters. When such a wavefront is sampled through the finite aperture of a telescope, several effects can be noted. The quality of the image formed by the telescope deteriorates from that predicted by physical optics; the image becomes enlarged and undergoes changes in size and position, since the direction of the normal of the wavefront is no longer constant, either in space or time. Furthermore, when the turbulence causing the refractive fluctuations is at a great distance from the telescope, of the order of kilometers, the intensity of the wavefront also becomes variable in both space and time.

This latter effect can be observed by viewing the illumination pattern from a bright star directly at the telescope aperture. The pattern appears to be crossed by a system of rapidly moving shadows and consequently is often called the shadow-band pattern. The total intensity of the telescopic image becomes variable, depending upon the sample of the shadow pattern selected by the telescope instant by instant. This variation in total intensity is called scintillation—or twinkling when the shadow pattern of a star is sampled directly with the eye. In the latter instance, the effective frequency of variation is cut off near 16 cycles per second, due to the time response of the eye. Scintillation may be studied either by measuring the fluctuations in image intensity for a given telescope aperture (*1*, *2*) or by measuring the shadow-band pattern directly (*3–5*). The results of both types of measurement are discussed here.

General Characteristics of Stellar Scintillation

While the observational characteristics of stellar scintillation have been discussed in detail elsewhere (*1*, *2*), it may be of interest to summarize some of these briefly. The more quantitative measurements have generally been made by means of photoelectric photometers. The fluctuating portion of the output current from the photocell, when corrected for shot noise, is directly proportional to the intensity variations of the shadow pattern integrated over the telescope aperture. This output signal is readily analyzed, with regard both to amplitude and to frequency distribution, by the techniques commonly applied to noise measurements.

The amplitude of the scintillation signal varies greatly from night to night and is strongly dependent upon both the size of the aperture and the altitude of the star above the horizon. For small apertures, say 1 to 3 inches, the peak-to-peak fluctuations of the signal as compared to the mean light level are of the order of 50 to 150 percent for stars near the zenith and can increase to several hundred percent for stars near the horizon. As larger apertures are used, the inherent fluctuations in the shadow pattern tend to average out; thus, the peak-to-peak amplitude for stars near the zenith may decrease to the order of 10 to 20 percent for apertures of 10 to 20 inches. The root mean square deviations of the signal are of the order of 30 percent for small apertures and fall off to the order of 5 to 10 percent for apertures measured in tens of inches. The strength of the scintillation also tends to increase whenever the wind field in the vicinity of the tropopause is strong, and hence, on the average, for mid-latitudes in the Northern Hemisphere, winter scintillation is stronger than summer scintillation.

The distribution of the scintillation signal with respect to frequency—that is, its Fourier spectrum—is another interesting parameter. This is likewise strongly influenced by size of aperture and by altitude and is even more strongly influenced by the wind field than is the total amplitude, or strength, of the scintillation signal. In general, for stars near the zenith, where small apertures are used the Fourier spectra tend to have a constant strength at frequencies from zero to around 100 cy/sec, with a decreasing strength from there to about 500 to 1000 cy/sec, where the amplitude becomes zero. On the other hand, when large apertures are used, the flat part of the spectrum extends to only 10 to 50 cy/sec, and the zero point is reached at anywhere from 100 to 500 cy/sec. The decrease in high-frequency components with increase in aperture size is readily explained as an aperture-smoothing effect.

As stars at lower altitudes are observed, the low-frequency components increase rapidly as the high-frequency components decline, and although the total bandwidth of the noise signal is quickly reduced, the low-frequency

The author is vice dean of the Graduate School of Arts and Sciences and associate professor of astronomy, University of Pennsylvania, Philadelphia.

components increase at such a rate that the total noise signal over all frequencies still increases.

A seasonal variation in the frequency distribution of the scintillation signals was first noted by Mikesell, who observed that the cutoff point of the Fourier spectrum (that is, the point at which the amplitude goes to zero) occurs at higher frequencies during the winter. Since the upper-air wind speeds are known to be higher during the winter season, the seasonal variation of the crossover led to a search for a correlation of wind speeds with stellar scintillation. Another clue to such a correlation was suggested by the observations of Mikesell, Hoag, and Hall (*6*), who showed by placing a slit over the telescope aperture, that a directional effect was associated with scintillation. When the slit was in one position, the spectrum of the signal was characteristic of the signal from a large aperture; when the slit was at right angles to this position, the spectrum was more characteristic of a spectrum associated with a small aperture. Hosfeld (*7*) showed that this directivity was in fact related to the upper-air winds.

In searching for a suitable correlation parameter of the scintillation signal, other than the directivity effect, it was found that neither the amplitude of the signal (either total or at specified frequencies) nor the cutoff frequency gave reliable correlations of a quantitative nature with the upper-air winds, although definite trends in the data were readily noted (*2*). The amplitude of the high-frequency components and the cutout frequency in general increased when the upper-air wind speeds increased. Unfortunately these parameters are strongly influenced by the total strength of the scintillation, and since the scintillation signal can show rather large fluctuations in magnitude over short intervals of time, it becomes quite difficult to say whether the particular values measured reflect a change in strength of the signal or a change in the upper-air winds.

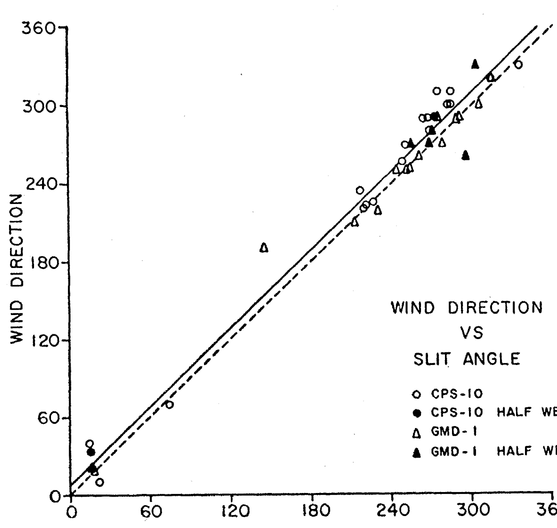
A quantitative description of the shape of the curve that was not strongly influenced by a variation of the signal amplitude was desired. In 1955 I proposed that the ratio of signal strength in a high-frequency band to signal strength in a low-frequency band be used as the correlate with wind speed, and I showed that this did indicate a relationship between scintillation and upper-air wind speeds in the vicinity of the tropopause that could be used as a measure of the wind speed (*2*). The wind speed was determined by taking geostrophic winds from the standard upper-air charts, the best correlation occurring for winds at the 200-millibar level. Since the geostrophic winds are, at best, only approximations to the true winds, it was decided to undertake a new program of observations which were to be made as nearly simultaneous with the upper-air wind soundings as possible, It was hoped that by such a procedure a definitive correlation might be made between the scintillation and the upper-air wind speeds.

Observational Program and Results

Preliminary tests, made at the Students' Observatory of the University of Pennsylvania, indicated that the directional effect was measurable with a slit as small as 1 by 4 inches. Consequently the scintillation measurements were made with a small-aperture telescope—one with a 4-inch aperture, as contrasted with the 12-inch aperture previously used.

The measuring equipment consisted of the 4-inch patrol camera of the Students' Observatory and a photoelectric photometer. A removable slit, 1 by 4 inches, which could be motor-driven through 190 degrees, was mounted over the objective. This mechanism was oriented in such a way that the zero fiducial of the slit was aligned with the direction of the true north when the telescope was in the meridian. The photocell output was amplified, and the noise-signal strength was measured in three pass bands, centered, respectively, at 10, 100, and 300 cy/sec. The ratio of the 100-cy/sec signal to the 10-cy/sec signal proved insensitive to the upper-air winds, especially at high wind speeds, and hence only the 300-cy/sec to 10-cy/sec ratios were used in the final correlations.

A normal observation consisted of two parts, a slit observation and a full-aperture observation. During the slit observation, the long axis of the 1- by 4-inch slit was rotated from the north to the south orientation and back. The output of the 300-cy/sec channel was observed, and the angular position



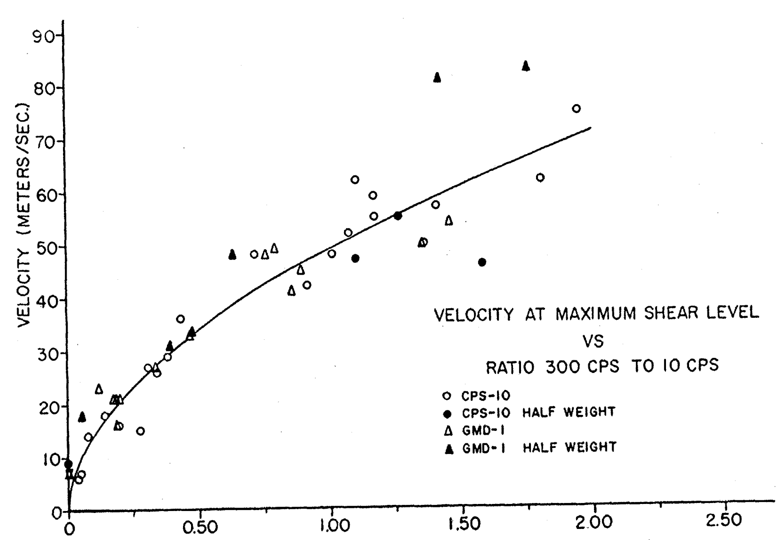


Fig. 1 (left). Correlation of slit angle with wind direction for the Bedford data. Fig. 2 (right). Correlation of scintillation ratios with wind speeds for the Bedford data.

of the slit corresponding to the minimum signal was noted when the slit was being driven in both the direct and the reverse modes. These two values were then averaged in order to remove the effect of a 10-second resistance-capacitance smoothing network at the channel output. This value was correlated with the upper-air wind directions.

The other part of the observational set consisted of a 5-minute reading with the full 4-inch aperture. The signals from both the 10 and the 300-cy/sec channels were corrected for instrumental noise by noting their values when the photocell was illuminated by a constant lamp adjusted to the mean illumination level of the star. The ratio of the high-frequency channel to the low-frequency channel was correlated with the magnitude of the wind velocities.

The scintillation observations were timed to occur as close as possible to 30 minutes after the normal release time of the balloons used for the 0300 Z upper-air soundings for the Philadelphia area. The balloons were released at a point approximately 10 miles due west of the observatory and were tracked by GMD-1 Rawinsonde equipment. The stars were always observed within 30 degrees of the zenith in order to insure that the layers causing the scintillation were being observed essentially normal to their stratification. This removed the necessity of making troublesome projection corrections.

Observations were made on 207 nights over 20 months, during which period radiosonde wind measurements were available for the Philadelphia area. Only 104 of the observations could be used, however, mostly because of inadequate meteorological data (the result, primarily, of failure of the balloon to reach sufficient altitude). Preliminary study of the data indicated that when the time differences between the balloon launchings and the scintillation observations were too great, discordant results were obtained. Consequently, only those observations in which the times of launching and of observation agreed to within 1 hour were used. This resulted in a final usable number of 67 observations.

In order to try to circumvent this problem of large attrition of the data and also to remove the objection of lack of control over the balloon soundings, it was decided to construct an instrument to be used at the Laurence G. Hanscom Field at Bedford, Massachusetts. Suitable meteorological equipment was in use at that field, and there was a further advantage in that a CPS-10 radar type wind-sounding instrument was available. This instrument gives reliable measures in high-velocity wind fields where the GMD-1 equipment becomes inaccurate. The scintillation equipment had the same aperture as that used at the University of Pennsylvania and was designed to be semiautomatic in its operation.

Table 1. Comparison of the results of the scintillation and wind correlations obtained with the GMD-1 equipment at the University of Pennsylvania and with the GMD-1 and CPS-10 equipment at Bedford.

Observations (No.)	Windspeed (m/sec) a	Windspeed (m/sec) b	Windspeed (m/sec) ρ	Probable error (m/sec)	Wind direction (deg) α	Wind direction (deg) β	Wind direction (deg) r	Probable error (deg)
	University of Pennsylvania (GMD-1 equipment)							
67	32.6 ± 0.7	0.49 ± 0.04	0.795	5.3	5.6	0.990	0.988	5.6
	Bedford (CPS-10 or GMD-1 equipment)							
46	59.6 ± 0.8	0.50 ± 0.02	0.957	4.1	5.3	0.998	0.994	6.4

A total of 46 observations were made with this equipment on 36 nights. Of these observations, 22 were made in conjunction with the GMD-1 equipment, and 24 were made with the CPS-10 equipment. Observations for which the time of scintillation measurement differed by more than 1 hour from a time 30 minutes after the balloon release are not included in the above number. Observations made at times differing from the desired time by ½ to 1 hour were given half weight in the reductions. It was also decided to give half weight to observations made in conjunction with GMD-1 soundings when the wind speeds were in excess of 60 meters per second. There were nine such half-weight observations in the 46 observations available for analysis.

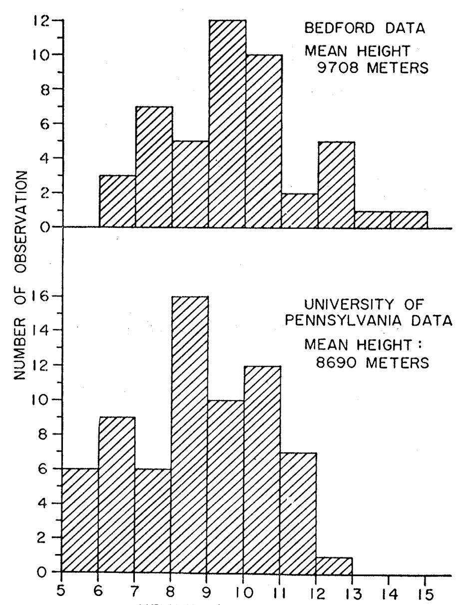


Fig. 3. Distribution of shear layer heights used in the wind correlations. The layers were chosen for each observation by selecting the layer at which the maximum vector gradient occurred in the vertical wind profile.

The two sets of data—University of Pennsylvania and Bedford—were reduced separately. While only the Bedford results are shown in Figs. 1 and 2, the final results for both locations will be discussed. Both sets of data gave the best correlation when values for winds at the maximum vector gradient were used.

The wind direction versus the slit angle (or the slit angle increased by 180 degrees since the sense of the motion of the pattern is indeterminable by this technique) was fitted by a linear relation

$$A = \alpha + \beta\theta$$

where A is the wind direction, α and β are constants, and θ is the slit angle.

The scatter diagram of the wind direction plotted against the slit angle for the Bedford data is presented in Fig. 1. The dashed line is the line of perfect correlation, while the solid line is the line fitted to the data by least squares. As given in Table 1, the resulting correlation coefficients r, for both the University of Pennsylvania and the Bedford data, are quite high. These values are beyond those expected from chance occurrence, as may be seen by noting that the odds are 100 to 1 that the correlation coefficient for the University of Pennsylvania data will not exceed 0.312 if the wind direction and slit angle are not related and, similarly that the correlation coefficient will not exceed 0.393 for the Bedford data. The probable errors of 5.6 and 6.4 degrees, respectively, are well within the limits of errors usually associated with the meteorological data.

The plot of wind speed against the scintillation ratio was fitted by a relation of the form

$$V = aR^b$$

where V is the wind speed in meters per second, a and b are constants, and R is the ratio of the 300-cy/sec component to the 10-cy/sec component.

The plot for the Bedford data is shown in Fig. 2, where the solid curve, of the form given above, was fitted by least squares. The indices of correlation ρ for the University of Pennsylvania and Bedford data, given in Table 1, are 0.795 and 0.957, respectively. The statistical significance of these results can be demonstrated by the fact that the odds are 1000 to 3 that the value for the University of Pennsylvania data will lie between 0.611 and 0.898 and that the value for the Bedford data will lie between 0.941 and 0.968. The probable errors of estimate for the two sets, 5.3 and 4.1 meters per second, respectively, are again well within the usual limits of uncertainty in the meteorological data. The difference in the a coefficients for the two locations arises from a difference in calibration of the two systems, which were somewhat different in design.

The distribution of the heights of the shear layers used in the correlation is presented in Fig. 3. The mean height for both observing sites is around 9000 meters, which is near the tropopause. That the turbulence causing stellar scintillation is expected to be at a considerable distance from the telescope has already been predicted by Keller (*8*) and Tatarsky (*9*). Image distortion is, however, caused by refractive inhomogeneities at all layers, even within the telescope itself.

While the correlations between the scintillation parameters and the wind speeds are quite good, it is not yet possible to state that there is a 1-to-1 relationship between the wind speed and the scintillation ratios. This is obvious from Fig. 2, where individual points may have relatively large deviations from the predicted values. That relatively large deviations may be found is, however, to be expected. In the first place, while an attempt was made to insure that the balloon soundings and the scintillation measurements would be as nearly simultaneous as possible, deviations as large as 1 hour had to be tolerated in order to have a statistically significant sample. It is well known that the upper-air wind fields can undergo large fluctuations in short periods of time. For example, on one night when multiple balloon soundings were made at Bedford, the wind underwent a vector change of 25 meters per second in the tropospheric region in 52 minutes. This is undoubtedly an extreme illustration, but it does indicate the relatively large effects encountered. Possibly more important than the time fluctuations of the wind field are its spatial fluctuations. These are particularly troublesome when high winds exist, since the sounding balloon may be transported tens of miles away from the observing site by the general wind movement while the telescope is looking through the wind field directly over the site. Another factor is the possibility of changes in the structure of the shadow-band pattern from night to night. The fact that the correlation is as good as it is indicates that the pattern structures on different nights can be expected to be quite similar. This relative constancy of structure was verified by the Optical Fourier Analyzer measurements described in the next section. Even though it is still not possible to state that the relationships given above are definitive and could be used for actual measurement of the upper winds, it does appear safe to conclude that stellar scintillation is an excellent indicator of the winds at the maximum-vector shear layer.

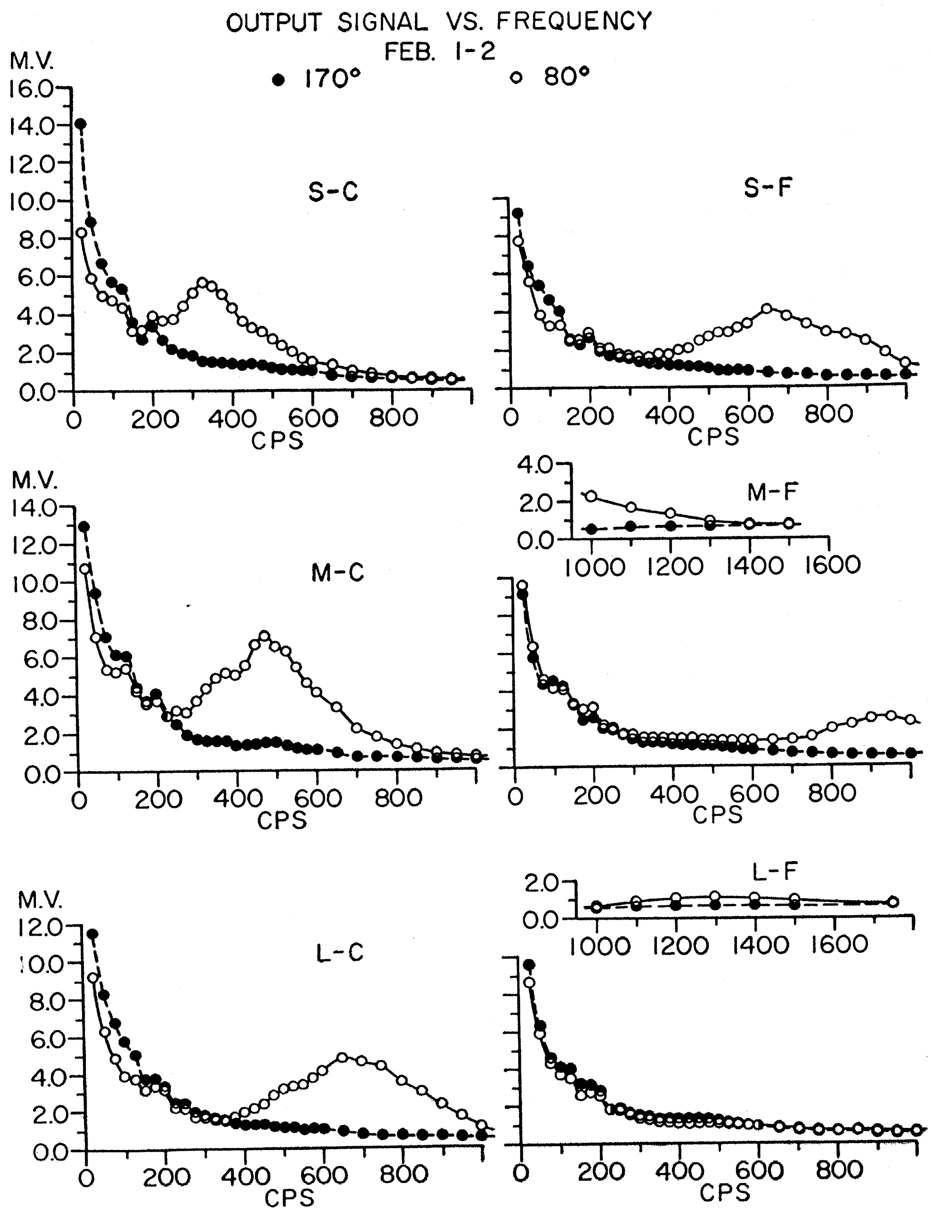


Fig. 4. Harmonic spectra of the photocell signal for the six lens-grating combinations on 1 and 2 February 1959.

Optical Fourier Analyzer

The observations just described do not give any direct information about the shadow-band pattern structure and hence about the atmospheric turbulence causing it. In order to make studies of the pattern structure, the Optical Fourier Analyzer was designed.

This is an instrument which permits measurement of the spatial power spectrum of the shadow-band pattern (*5*). It is essentially the optical analog of an electrical filter. Briefly, it functions by imaging the shadow-band pattern from a given star upon a transmission grating. Ideally, the grating should have a sinusoidal transmission function in one direction and a constant transmission function perpendicular to that direction. The grating actually used, however, was a Ronchi grating—that is, one that has a square-wave transmission function rather than a sinusoidal one. The total light transmitted by the grating is the product of the grating transmission function and the spatial intensity function of the shadow pattern. If this light is measured by a photocell, an electrical signal is generated which is proportional to the instantaneous product of the two functions. It is obvious that for a sinusoidal transmission function this signal is nothing more than the intergrand of the Fourier integral for that portion of the shadow pattern which is imaged on the grating, and that it corresponds to the wavelength associated with the grating spacing. When a square-wave transmission function is used, the odd harmonics of the principal spatial frequency of the grating are also introduced, but fortunately for the observations under discussion, it can be shown that the higher harmonics are of no measurable importance. Since the shadow pattern is time-dependent, the output of the photocell will also fluctuate with time. The time average of the fluctuating signal will be related to the strength of the spatial component of the shadow-band pattern, having a wave number corresponding to that of the transmission filter (*5*). The effective band pass of the spatial filter will be a function of the shape and size of the shadow pattern analyzed: the larger the sample, the narrower the effective band pass of the filter.

If the pattern is now imagined to be fixed—that is, to be without formation or decay of pattern elements—in translation along the direction of density variation of the grating, the output of the photocell will contain a constant frequency signal. The strength of the signal will be related to the strength of the shadow-band element size corresponding to the wave number of the grating. The frequency of the signal will be related to the element size and the velocity of translation according to the relation

$$V = lf$$

where V is the translational velocity of the pattern, l is the linear size of the pattern corresponding to the wave number of the grating, and f is the frequency of the photocell output signal in cycles per second.

In practice, the pattern undergoes a change with time as well as with translation, and hence the contributions to the fluctuating portion of the photocell output will be made over a wide frequency range. The strength of the pattern element will thus be related to the total strength of the fluctuating signal taken over all frequencies, while the predominating frequency will permit determination of the pattern velocity according to the expression just given. The half-width of the signal about this enhanced peak will depend upon the band pass of the optical filter and the rate of the growth and decay of the pattern.

In order to determine the complete spatial power spectrum of the shadow-band pattern it is necessary to have gratings corresponding to as many different wave numbers as are required to give the desired coverage. Those of small wave number—that is, of large element size—are limited by the pattern sample available. The practical limit is reached when the sample size is about three times the element size. The lowest possible spatial frequency in

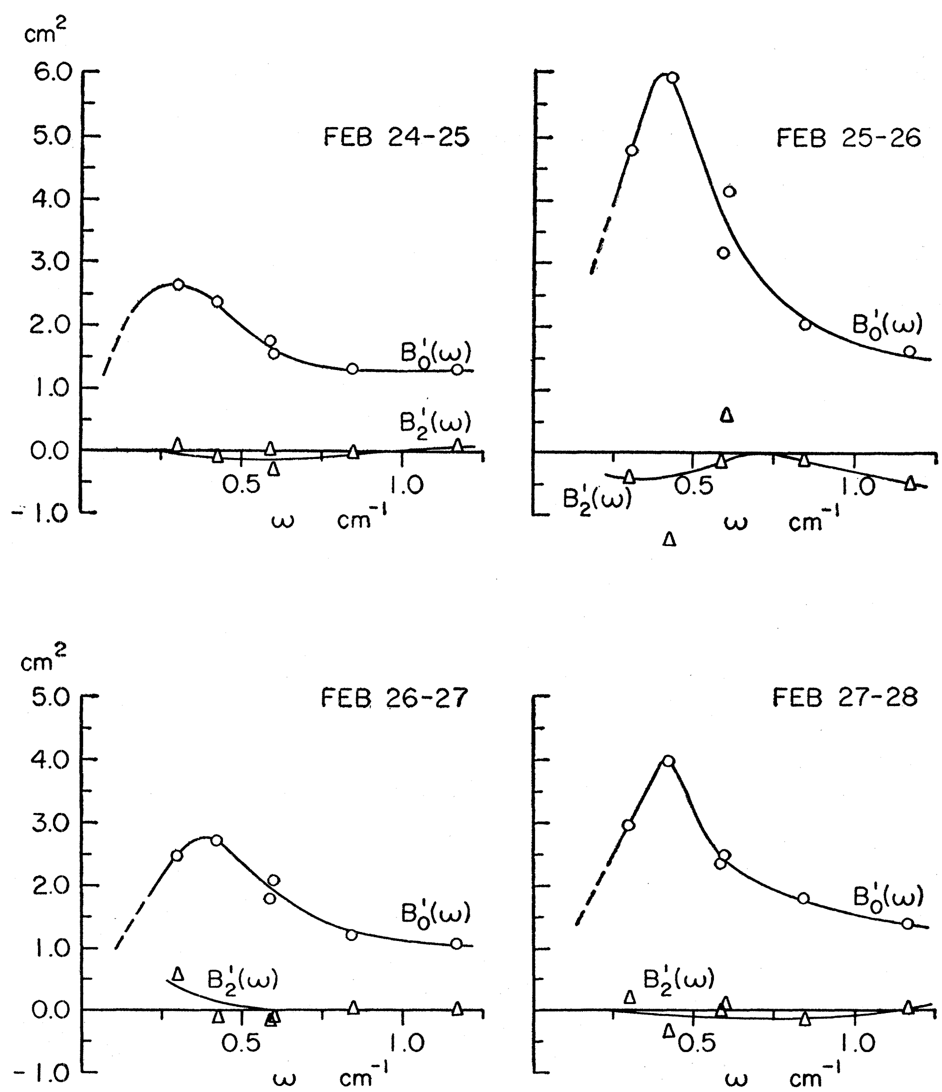


Fig. 5. Representative spatial power spectra of the shadow-band pattern on four nights in February 1959.

Table 2. Characteristic wavelengths and spatial frequencies of the available lens-grating combinations.

Lens*	Wavelength (cm)	Spatial frequency (cm^{-1})
Coarse grating (25 lines/inch)		
Short	21.08	0.298
Medium	14.94	0.420
Long	10.77	0.583
Fine grating (50 lines/inch)		
Short	10.54	0.596
Medium	7.47	0.840
Long	5.38	1.165

* Focal lengths (in inches) as follows: short, 2; medium, 3; long, 4.

the work described here was thus 0.298 cm^{-1} (or a wavelength of 21.08 cm), since the telescope used for the measurements was the 28-inch reflector of the Flower and Cook Observatory.

Six spatial frequencies were used by combining two gratings, designated fine and coarse, and three projection lenses, designated short, medium, and long. The known optical magnification of the projection lenses permitted relation of the image of the shadow-band pattern to the true dimensions of the pattern. The spatial frequencies and element sizes that were thus available are given in Table 2.

The observations were made with the analyzing gratings in two orientations: in one position, which gave the smallest high-frequency output from the photocell, the constant axis of the grating was aligned with the pattern motion; in the other position the axis was perpendicular to the pattern motion. The same procedure was followed for all six grating-lens combinations. As long as the pattern is not too asymmetric, the analysis with the gratings in these two positions is sufficient for measuring the two-dimensional power spectra.

The output from the photocell was recorded on magnetic tape and, subsequently, upon repetitive playback, a harmonic analysis of the 12 recorded signals, two signals for each of the six grating-lens combinations, was made. Examples of the spectra for the 12 signals recorded on the night of 1 February 1960 are shown in Fig. 4. The output of the harmonic analyzer used to measure the signal strength of the photocell is plotted against frequency. The open circles represent the case when the square-wave transmission function of the grating is aligned with the pattern motion; the solid circles, the case when the grating is at right angles to this orientation. The frequency of the signal peaks is directly proportional to the wave number, as would be predicted by the relation between pattern velocity and grating spacing. The calculated pattern velocity averaged over the combinations was 70.5 ± 1.07 meters per second at a direction of 80 (or 260) degrees measured clockwise from the north. Measurements of this type were made on 18 nights and gave values for the pattern speed and direction consistent with upper-air wind speeds. A direct comparison with the winds was not possible since radiosonde measurements were no longer being made in the Philadelphia area. On only one night was the pattern speed found to be indeterminable by this method. On eight nights, however, multiple peaks were found, indicating the possibility that several patterns were being superimposed. Multiple patterns of this sort tend to give rise to bad measurements when observed by the other technique described in the preceding section.

The spectra of the type represented in Fig. 4 were integrated numerically to determine the total signal strength. If the shadow pattern is not radically asymmetrical, the two-dimensional power spectrum may be written

$$B(\omega, \phi) = B_0(\omega) + B_2(\omega) \cos 2(\phi - \theta)$$

where B_0 is the symmetrical component of the pattern, B_2 is the asymmetrical component, ω is the spatial frequency, ϕ is the azimuthal angle of the pattern, and θ is the azimuth of the pattern motion.

Four representative spatial power spectra are shown in Fig. 5. It may be noted that the asymmetrical component $B_2(\omega)$ is quite small in amplitude and that the symmetrical component peaks near $\omega = 0.5$ cm^{-1}. This peaking was noted on all but four of the 18 nights; the element size corresponding to the average of the spatial frequencies at which the peaks occurred was 15.6 centimeters. The peak element sizes, when the peaks were measurable, ranged from 12.8 to 17.7 centimeters, and of course the peak element size must have been greater than 21.1 centimeters on the four nights on which the peak occurred outside the range of measurement.

Analysis of the half-widths of the frequency peaks, on the assumption that the pattern growth and decay is exponential in character, leads to the conclusion that the pattern elements have lifetimes measurable in milliseconds to tens of milliseconds. The shadow patterns are therefore quite transitory in nature, not only because of their rapid motion but also because of their rapid structural change. This introduces difficulties when other methods, such as that of determining the autocorrelation function of the pattern by means of two telescopes of variable separation, are attempted in order to extend the power spectrum into the long-wavelength region (*4*). If the longer-wavelength measurements are desired, it appears that the best way to obtain them is to use the Optical Fourier Analyzer with a large telescope aperture.

Summary and Conclusions

The ratio of the stellar scintillation signal in a frequency band centered at 300 cy/sec to the signal in a band centered at 10 cy/sec generated when a star near the zenith is observed with a 4-inch circular aperture has been shown to be dependent upon the upper-air winds. These ratios are capable of indicating the wind speeds at the level of maximum vector gradient with an accuracy comparable to that normally attained by conventional radiosonde measurements. Furthermore, when a 1- by 4-inch slit is placed over the telescope aperture, the wind direction, but not its sense, may be determined by noting the alignment of the long axis of the slit which gives the minimum 300-cy/sec signal.

The Optical Fourier Analyzer, a device which permits measurements of the two-dimensional spatial power spectrum of the shadow-band pattern, has given data upon the spatial and temporal structure of shadow-band patterns. The shadow-band pattern elements are characterized by a size of 15 to 16 centimeters and have a lifetime of the order of several to 10 milliseconds. A direct determination of the velocity of the pattern motion across the telescope is also possible and indicates that the pattern velocities are comparable to the upper-air wind velocities. The observational results on some nights can best be interpreted on the basis of two or more simultaneous patterns moving with different velocities (*10*).

References and Notes

1. A. H. Mikesell, "The scintillation of starlight," *U.S. Naval Observatory Publs. ser. 2, vol. 17, pt. 4* (1955).
2. W. M. Protheroe, "Preliminary report on stellar scintillation," *Contribs. Perkins Observatory, Ohio Wesleyan Univ. and Ohio State Univ.* **II**, No. 4 (1955).
3. ———, *J. Optical Soc. Am.* **45**, 171 (1955).
4. P. E. Barnhart, G. Keller, W. E. Mitchell, Jr., "Investigation of upper air turbulence by the method of analyzing stellar scintillation shadow patterns," *Air Force Cambridge Research Center, Bedford, Mass., Tech. Rept.* (1954).
5. W. M. Protheroe and K. Y. Chen, "The correlation of stellar shadow-band patterns with upper air winds and turbulence," *ibid.* (1960).
6. A. H. Mikesell, A. A. Hoag, J. S. Hall, *J. Optical Soc. Am.* **41**, 689 (1951).
7. R. Hosfeld, "Scintillation, stellar shadow bands, and winds aloft," paper presented before the American Astronomical Society, 20–23 June 1954.
8. G. E. Keller, *Astron. J.* **58** (1 July 1953).
9. V. I. Tatarsky, *Proc. Acad. Sci. U.S.S.R.* **120**, 289 (1958).
10. The work described in this article was supported in part by the Geophysics Research Directorate of the U.S. Air Force Research and Development Command.

Competitive Exclusion

The exclusion principle is recast in the context of a generalized scheme for interspecific interactions.

Bernard C. Patten

The colloquy in these pages (*1*) concerning the newly denominated but ancient theory of "competitive exclusion" has generated a controversy which appears to have resulted in a standoff. One wonders if the reason may not be implicit in A. N. Whitehead's remarks when he admonished his contemporaries for living off the intellectual capital accumulated in the 17th century, warning that any culture was doomed which could not throw off the inertia of habitual thinking and burst through the facade of its own concepts. Exclusion theory is controversial, it would seem, not so much because it isn't intuitionally reasonable or, for the most part, empirically expressed, but rather because it is couched in an archaic context of 19th-century dogma within which circular reasoning is the only alternative to progress (in terms acceptable to modern apprehension of scientific episteme). This circularity is reflected in the earliest and latest formulations of the exclusion principle:

> . . . it is the most closely-allied forms—varieties of the same species, and species of the same genus or related genera—which, from having nearly the same structure, constitution and habits, generally come into severest competition with each other; consequently, each new variety or species, during the progress of its formation, will generally press hardest on its nearest kindred, and tend to exterminate them. [Darwin, 1859]
>
> [Since] complete competitors cannot coexist . . . ecological differentiation is the necessary condition for coexistence. [Hardin, 1960]

Thus the total achievement of a century of thought amounts to providing the contrapositive of the original proposition.

The exclusion principle is regarded in this article as a legacy from the past, whose continued recognition at "law" status can only interfere with a healthy development of concepts whose further disquisition it tends to block. Therefore it should be relegated as prudently and expeditiously as possible to a de-emphasized position in a broader, more modern framework. A tentative step in this direction is provided, which casts exclusion in a context which includes also the cooperative aspects of interspecific phenomena. Strong reliance on cybernetic models as formulated by Ashby (*2*) is acknowledged.

Consider a universe Υ of entropy states, some at higher and others at lower levels of potential. Discrete enclaves of high potential (the sun for example) represent sources of unconstrained variety (information, negentropy) which transmit to low-potential sinks comprising states of maximally constrained variety (entropy). Let the subset υ represent the biological states in Υ at intermediate potentials. This collectivity, consisting of states of partially constrained variety, possesses the capacity to impose constraint upon information and so to generate entropy—an accomplishment, as will be shown, which requires an information store which υ seeks to maintain (and extend) contrary to the gradient of potential. The situation is analogous to a two-person von Neumann game of the non-zero-sum type (*3*) in which υ simultaneously seeks to gain information from Υ for use in blocking its gain of information from Υ. Solution of this paradox constitutes the fundamental problem of regulation. Two basic principles are involved: the law of entropy and the law of requisite variety. These laws are best discussed against a background of the nature of υ's organization.

Consider, as a functional element of υ, "species" A, which regulates a set of essential variables within a favorable range α beyond which A fails to survive. The subset α corresponds to a Hutchinsonian niche (*4*). Disturbances D, in the form of information from the environment Υ, threaten to drive the states of A outside of α. If D may be visualized as acting through some dynamic system P, a protocol characteristic of Υ, then the initial diagram of immediate effects takes the form

$$\boxed{D} \rightarrow \boxed{P} \rightarrow \boxed{A}$$

A forms another dynamic system R, a regulator which can be coupled to P to produce a machine capable of blocking the flow of variety to the essential variables:

$$\begin{tikzcd} & \boxed{P} \arrow[dr] & \\ \boxed{D} \arrow[ur] \arrow[dr] & & \boxed{A} \\ & \boxed{R} \arrow[uu] & \end{tikzcd}$$

The "game" takes the following sequence: (i) the environment Υ makes an arbitrary move D; (ii) R assumes a value determined by D's value; and

The author is associate professor of marine science, College of William and Mary, at the Virginia Institute of Marine Science, Gloucester Point.