Deep-Sea Research, Vol. 26A, pp. 123 to 161.

0011-7471/79/0201-0123$02.00/0

Ocean acoustic tomography: a scheme for large scale monitoring

WALTER MUNK* and CARL WUNSCH†

(*Received* 4 *April* 1978; *in revised form* 3 *July* 1978; *accepted* 31 *July* 1978)

Abstract—We consider the problem of monitoring ocean basins for mesoscale fluctuations, using acoustic inverse techniques. The procedure, which has much in common with conventional seismology, consists of measuring perturbations in travel time between acoustic sources and receivers. Because the number of pieces of information is the *product* of the number of sources, receivers, and resolvable multipath arrivals, the economics of the system is enhanced over usual spot measurements. The temporal resolution required to distinguish multipath arrivals is estimated at 50 ms; the precision required to measure mesoscale perturbations is estimated at 25 ms. The required resolution and precision can be achieved by existing low-frequency (100- to 200-Hz) broadband (> 20-Hz) sources, but we are ultimately limited to 1000-km ranges by the variable ocean finestructure and associated micropaths. There appear to be no practical range limits imposed by micropaths if such broadband sources could be centered at 30 Hz.

Given the travel time measurements and their noise estimates, we show how actually to invert the system for the interior changes in sound speed and, by inference, for density. The method is analogous to the medical procedure called tomography (from the Greek 'slice'). Measures of the spatial resolution and of formal error bars are obtained. We conclude that such a system is achievable now and has potential for development in a number of directions.

NOTATION

r receivers, s sources, m moorings, p paths

$\theta, S, C,$	Temperature, salinity, sound speed.
$\bar{C}, \hat{C}, C_s, C_b,$	Values at axis, ray turning depths, surface and bottom.
$m, \alpha, \beta,$	Coefficients associated with sound speed [equation (1)].
$\delta\theta, \delta S, \delta C,$	Perturbations.
$r, z, t,$	Coordinates.
$T+\delta T,$	Travel time (reckoned from pulse transmission time).
$R, R_{cz},$	Range, range of convergence zone.
$l_v,$	Sound speed scale.
$p_u, p_l,$	Number of upper and lower ray turning points (reflections).
$\eta, B, \varepsilon,$	Parameters in canonical sound channel.
$n, n_0,$	Buoyancy frequency, n_0 is surface extrapolated value.
R, SR, BR,	Refracted, surface reflected, bottom reflected.
$S(f),$	Clock spectra of frequency.
$\varepsilon(t),$	Clock error.
$F F^*,$	Linear drift filter.
$\Delta t,$	Duration of transmitted pulse.
$\Delta f, (\Delta f)_T,$	Transmitted bandwidth.

* Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, U.S.A.
† Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

$\Delta\tau$,	Width of covariance peak.
N,	Noise density.
P,	Power.
Φ, Λ,	Fine scale parameters.
$(\Delta t)_M$,	Time interval over which measurements are averaged.
$(\Delta t)_0$,	Ocean decorrelation time.
$(\Delta f)_0$,	Ocean decorrelation frequency.
$i = 1, I$,	Ray path identification.
$j = 1, J$,	Horizontal grid area identification.
$k = 1, K$,	Vertical grid area identification.
$l = 1, L$,	Sequence of nonzero eigenvalues λ_l.
R_{ij},	Path length of ray i in grid area j.
E_{ij},	$R_{ij}/(C_j^0)^2$, [equation (17)].
U_l, V_l, λ_l,	Singular vectors and singular values.
δT_M, δT_N,	Measured and noise components of travel time perturbation δT.
A, B, C, D,	Configurations (Fig. 6).
A, B, D,	Matrices defined in equations (25) and (28).
σ^2, $\tilde{\sigma}^2 = \sigma^2/(C^0)^2$,	Noise variance in arrival time.

1. THE PROBLEM

RECENT observations have demonstrated that most of the kinetic energy of the ocean circulation is associated with variability on a relatively small scale (the mesoscale) of order 100 km (e.g. MODE GROUP, 1978). The general circulation (DC or time average) on the scale of the major ocean gyres probably contains only 1% of the kinetic energy. By analogy to the atmosphere, the general circulation may be regarded as climatology; the mesoscale variability is the ocean weather.

Measuring the ocean weather poses a formidable problem because of the small scale structure (100 km versus 1000 km in the atmosphere) compounded by the much longer ocean time scales (100 days versus 3 days in the atmosphere). Even for an area as small as 1 square megameter (1000 km × 1000 km), several full-time vessels, or several hundred fixed moorings, are required [airborne surveys using airborne expendable bathythermographs (AXBTs) could be a very helpful adjunct]. Satellite techniques (radiation and altimetry) hold some promise. In some regions, surface radiation patterns bear a relationship to features of the mixed layer and deeper (BERNSTEIN, BREAKER and WHRITNER, 1977). Altimetry, now at a speculative stage, could provide a close analogy to surface pressure maps. But surface fields are inadequate to represent the full ocean, just as surface pressure maps fail to fully represent atmospheric weather. There will always be a need for interior measurements of the ocean.

This paper examines the feasibility of studying and ultimately monitoring the oceans by measuring acoustic transmissions between moorings over large distances (Fig. 1). With m moorings, conventional techniques yield m spot measurements. But m moorings consisting of r receivers plus s sources yield $r \cdot s$ pieces of information, rather than $r+s = m$. There may be an additional advantage in working with spatially integrating measurements that filter out many of the small scales that contaminate conventional 'spot' measurements. Furthermore, for any given source–receiver pair there are p resolvable acoustic paths, where p is order 10 for the geometry we propose. These paths have distinct vertical

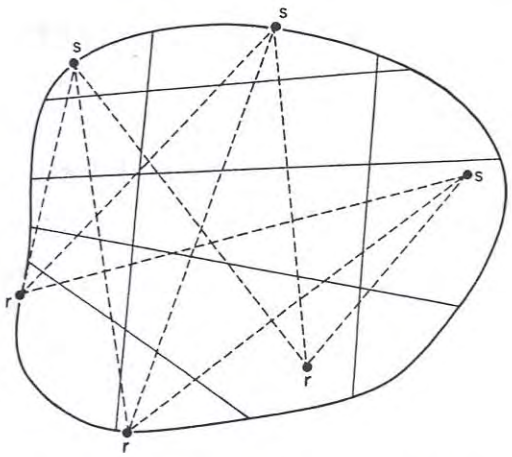

Fig. 1. Perturbations in acoustic travel time from any source s to any receiver r (dashed paths) are used to estimate sound speed perturbations in an arbitrary grid area (solid lines).

trajectories (Fig. 2), and so we obtain, for each source–receiver pair, information comparable with that from a vertical string of conventional instruments (MUNK and WORCESTER, 1976; MUNK and WILLIAMS, 1977).

We envisage the receivers as being predominantly shore-based. This permits real-time use of the data for an operational monitoring program. The spatial resolution could be greatly improved by well-positioned autonomous offshore listening stations, provided a method is devised for 'real-time' (e.g. daily or weekly) data transmissions.

Experimental requirements are severe, particularly with respect to the resolution of multipaths and the measurement of small anomalies in travel time. There is the additional requirement of disentangling the multiple slices through the oceans to produce something useful. The task is related to that of computerized tomography* (CT): multiple sectional X-rays from a source and receiver which are rotated around the patient, and are processed by computer to yield two-dimensional displays of the interior structure. The mathematical basis for a solution to this problem goes back to the work of the Austrian mathematician RADON in 1917, but the dramatic development towards a new diagnostic tool had to await extensive computer capacities.

There is some analogy with classical seismology, where properties of the Earth's interior are inferred from travel times of earthquake waves. Here the emphasis is on mesoscale space–time variability in sound speed.

A more useful presentation would be temperature or density. Suppose the local temperature–salinity relation is approximated by $S = S_0 + m\theta$, so that

$$\delta C/C = \alpha\delta\theta + \beta\delta S = \alpha\delta\theta(1 + m\beta/\alpha). \tag{1}$$

A typical value of m is $0.1‰\,(°C)^{-1}$. Then with $\alpha = 3.2 \times 10^{-3}\,(°C)^{-1}$ and $\beta = 1.0 \times 10^{-3}(‰)^{-1}$, the correction factor is $m\beta/\alpha = 0.03$. Thus the temperature perturbations are very nearly given by $\alpha^{-1}\delta C/C$; when necessary a correction can be made for salinity.

*'Tomo' is derived from the Greek word for cut or slice. For a recent discussion see SWINDELL and BARRETT (1977).

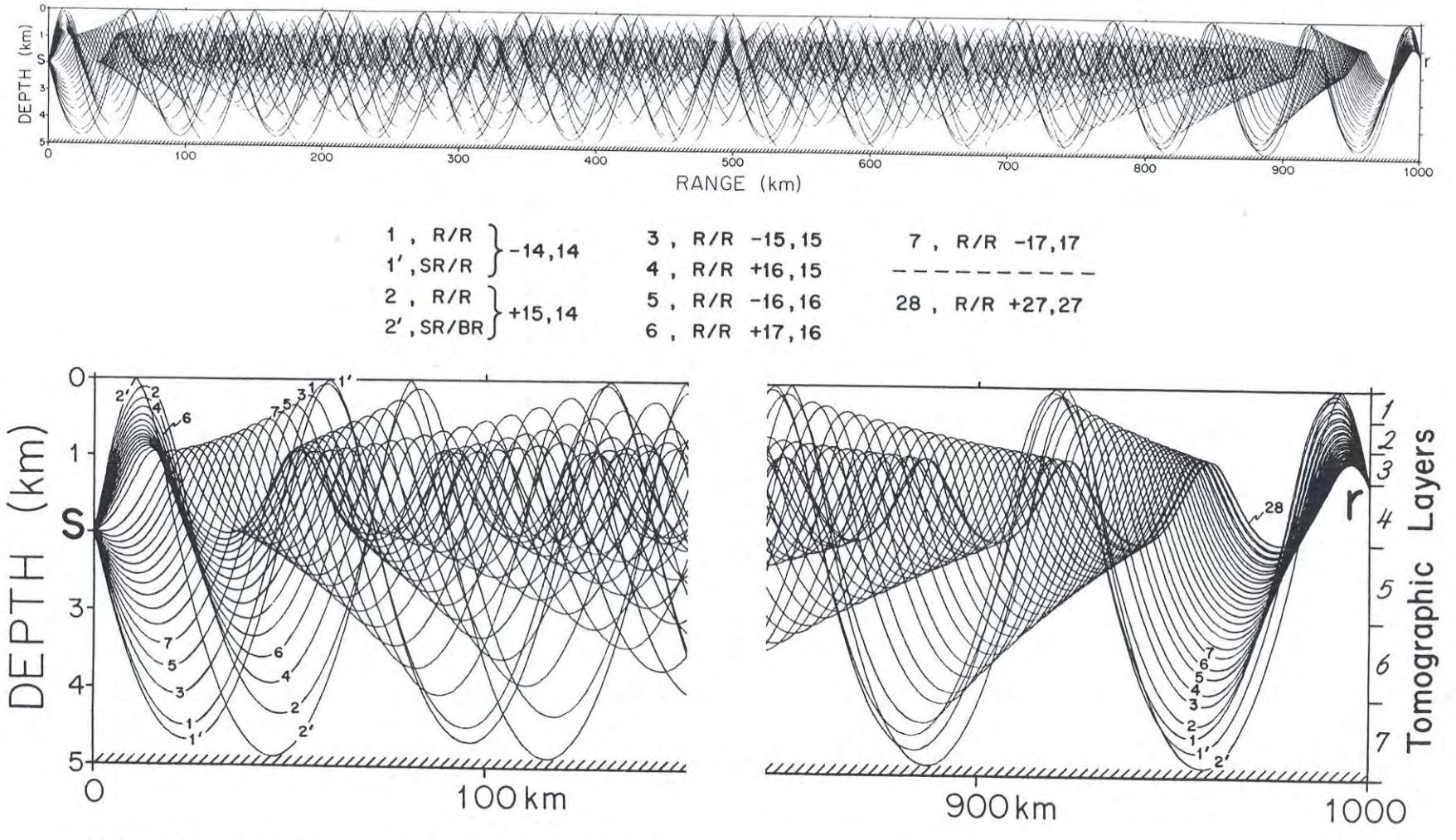

Fig. 2. Ray diagram for a 1000-km range, using the mean sound speed profile along a path between Eleuthera and Bermuda. The source is at 2000-m depth, the receiver is axial at 1280 m. For simplification only those rays that travel downward towards the receiver are plotted. The first and last 150 km are shown on an enlarged scale, with the earliest nine rays labeled in order of their arrival; the associated ray identifiers [equation (3)] are tabulated. The scale 1 to 7 on the bottom right identifies the layers used in the vertical tomographic slice (Table 2, Section 5). We are indebted to John Clark of the Institute for Acoustic Research for this figure.

If the temperature–salinity relation is reliably known, then density perturbations can be inferred as well. The accuracy of the densities will vary geographically with the degree of θ, S correlation.

The scheme of this paper is as follows. We first discuss the 'direct' problem of the behavior of sound transmission over distances of order 1 megameter. We consider the precision in measuring travel time along a resolved multipath; what is the requirement, and how can it be achieved? The second part of the paper deals with the 'inverse' problem—given observations of sound transmission in a region of ocean, how best to determine the interior structure. The final result is, we believe, a practical scheme for large scale ocean monitoring.

2. MESOSCALE VARIABILITY

In a typical midlatitude situation, the sound speed has a minimum at about 1 km depth; it increases upwards because of the increasing temperature, and it increases downwards because of the effect of increasing pressure. This configuration provides a wave guide. In the language of ray optics, the ray paths oscillate about the sound axis. A typical wavelength (upper plus lower loop) of the rays is 50 km. This length depends only weakly on ray inclination, and so rays from a source will come together at a 'convergence zone' range $R_{cz} \approx 50$ km.

To estimate perturbations in travel time associated with mesoscale variability, we take a typical temperature anomaly $\delta\theta = 3°\text{C}$ over a distance of 100 km, confined to the upper 500 m. This causes a fractional change in sound speed of $\delta C/C = \alpha\delta\theta = 10^{-2}$. A typical shallow upper loop extends for $\Delta r = 15$ km in the upper 500 m, and there will be two successive upper loops within the mesoscale disturbance. The associated perturbation in travel time is

$$\delta T = -\frac{\delta C}{C}\frac{2\Delta r}{C} = -0.2\,\text{s} \quad \text{for} \quad \delta\theta = +3°\text{C}.$$

A ray passing through several such disturbances (positive and negative) will show travel time changes of this order.

The corresponding perturbation by mesoscale currents is smaller. Even for the extreme case of a 3-knot current (1 knot $= 0.51\,\text{m s}^{-1}$) the fractional perturbation in propagation speed is only $u/C = 10^{-3}$. The geostrophic current is given by $\rho f u = \delta p/L_H$, where $f \approx 7 \times 10^{-5}\,\text{s}^{-1}$ is the Coriolis parameter, and $\delta p = g\delta\rho\ L_V$ is the mesoscale pressure perturbation, with L_H and L_V designating horizontal and vertical eddy scales. Write $\delta\rho/\rho = a\delta\theta$, $a \approx 2 \times 10^{-4}\,°\text{C}^{-1}$. The contributions from horizontal velocity and vertical displacement are then in the ratio

$$\frac{u/C}{\delta C/C} = \frac{ga\delta\theta L_V/(fCL_H)}{\alpha\delta\theta} = \frac{g}{fC}\frac{a}{\alpha}\frac{L_V}{L_H} = 93\frac{0.2\times10^{-3}}{3\times10^{-3}}\frac{1}{100} = 0.06. \tag{2}$$

Accordingly, the current plays a secondary role, except for truly barotropic eddies (if they exist). There is however, the intriguing possibility of measuring current components from the small *difference* in the travel times between opposite transmissions (WORCESTER, 1977).

The earliest indication of mesoscale variability in sound transmission comes from the Sound Channel Axis Velocity Experiment (SCAVE) conducted in 1961 to 1964

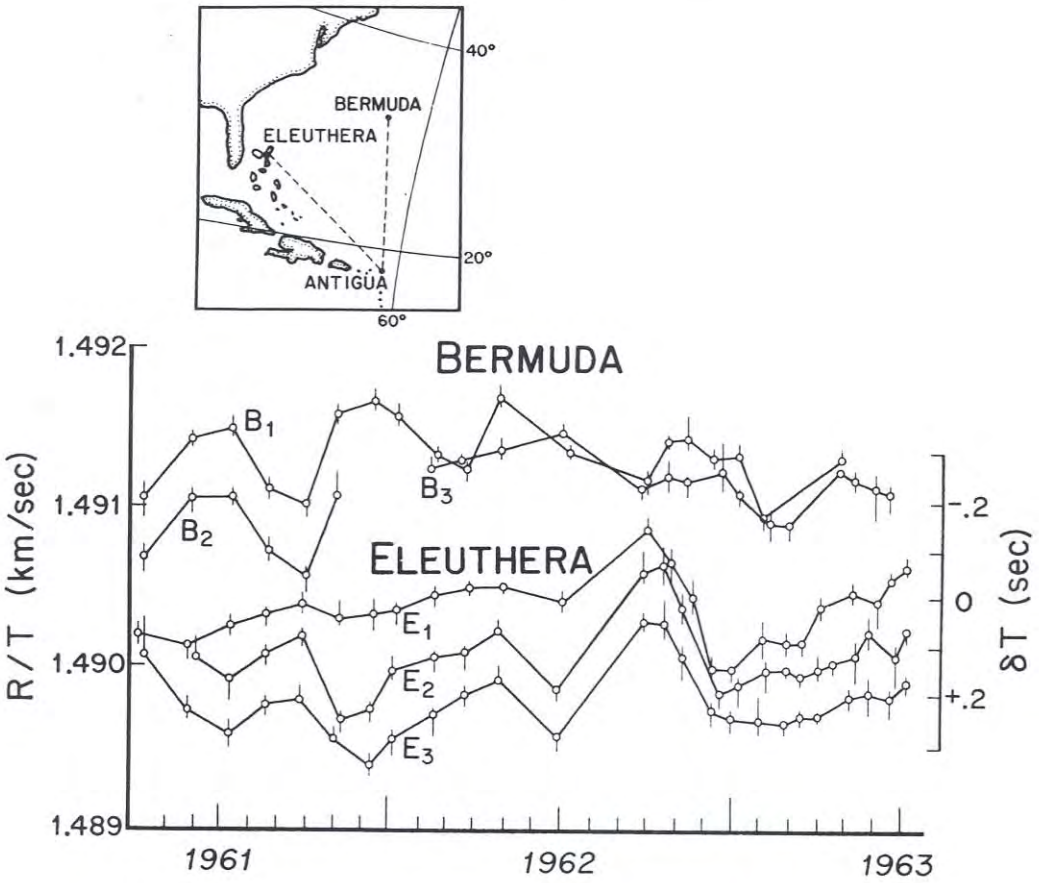

Fig. 3. Variations in the mean axial sound speed R/T and the corresponding perturbation in travel time δT (relative scale), according to HAMILTON (1977). Error bars give full spread of data. The source is at Antigua. Bermuda hydrophone B_1 is tethered 5000 ft above bottom, B_2 is in shallower water on bottom, both are at 875 m beneath the surface. Eleuthera hydrophones E_1, E_2, and E_3 are all on bottom at about 1000-m depth.

(HAMILTON, 1977). Precisely located and timed SOFAR charges were fired at axial depth off Antigua, using the hydrophone array of the Atlantic Missile Range. From the axial cutoff of the received signal at the Bermuda and Eleuthera field stations, the authors were able to ascertain travel times to within 30 ms. Over the period of 27 months the r.m.s. travel time variations are estimated at 200 ms, with time scales of a few months; the maximum change was by 500 ms in 3 months (Fig. 3). Hamilton remarks on the lack of correlation between the fluctuations for the Antigua–Bermuda and Antigua–Eleuthera paths (Bermuda and Eleuthera are separated by more than 1000 km). Fluctuations at the three individual Eleuthera hydrophones (60 km separations) are clearly correlated, but even here there are significant differences. All this is consistent with what we now know of mesoscale variability. Similar conclusions are reached from experiments in the North Pacific conducted by JOHNSON and NORRIS (1968).

Within a few years of Hamilton's pioneering work, STEINBERG and BIRDSALL (1966)

started the MIMI sound propagation experiments across the Straits of Florida, followed by a 1250-km transmission between Eleuthera and Bermuda (CLARK and KRONENGOLD, 1974). The sound source was a 406 Hz CW signal. The measured variations in 'extended phase' are consistent with variations in travel time of 200 ms per month, but it is not clear to what extent these are the direct result of ocean fluctuations, as compared to a random walk associated with interfering multipaths (DYSON, MUNK and ZETLER, 1976). Subsequently JOBST and DOMINIJANNI (1978) resolved a path by means of a pulse compression technique with arrival times measured to 0.1 ms precision; the measurements over a period of 3 days showed a drift in travel time (superimposed on tidal fluctuations) by 10 ms, consistent with a monthly variation by a few hundred ms.

3. MULTIPLE PATHS

In this section we estimate the number of possible acoustic paths (multipaths) between a source and receiver separated by a distance of order 1 megameter (Mm), and the different manner in which these various paths sample the oceans. We consider how the different paths can be identified in the rather complex multipath arrival sequence.

The language of ray optics is used throughout this paper. Let l_V designate the vertical scale of the sound-speed structure. Consider any refracted ray path from source to receiver, and a second path through the same endpoints but 'scattered' at an intermediate point which is separated from the direct path by a vertical distance l_V. Ray optics provides a useful approximation if the difference in travel time between scattered and direct rays exceeds one quarter cycle. For $l_V = 1000$ m (characteristic of sound channel) and 100-Hz frequency this holds to a range of several megameters. At lower frequencies one has to resort to more precise methods, such as an integration of the parabolic wave equation (FLATTÉ and TAPPERT, 1975), or an expansion into normal modes. In the case of statistical inhomogeneities all along the path (such as those due to internal waves) the range of applicability of ray optics is further reduced (FLATTÉ, DASHEN, MUNK, WATSON and ZACHARIASEN, 1978, Chapter 5).

Even when ray optics is applicable, it is still convenient to use normal mode expansions if the required number of modes is smaller than the required number of rays. Further, a normal mode expansion can produce additional information. For example, one can compare measured and computed group and phase velocities. The situation is like that in seismology where much of the early information about the Earth was gained from an examination of the arrival times along principal paths. In recent years, the use of normal modes has yielded much additional information.

Consider now the perturbation in travel time along any single path through a mesoscale disturbance $\delta C(x, y, z)$. One can separate two effects: a change in speed along the unperturbed path (early arrivals through a warm eddy), and a displacement of the path itself. The displacement effect is of second order because the arrival time is insensitive to path perturbations for a Fermat stationary path. This is an essential property of the eikonal (or ray optic) approximation.

Different ray paths give different weights of the water column, and this is crucial for getting at the vertical eddy structure. There are four classes of rays,

$$\text{R/R}, \quad \text{SR/R}, \quad \text{R/BR}, \quad \text{SR/BR}$$

according to whether they are purely refracted both above and beneath the axis, surface

reflected and refracted, refracted and bottom reflected, or surface and bottom reflected (Fig. 2). Bottom reflections lead to rapid attenuation.

For a 'normal' sound channel, 'steep' rays with turning points near the surface and bottom (where the sound speed is large) are the first refracted rays to arrive, even though their path is the longest; flat (or axial) rays come in last.

Consider a mesoscale perturbation extending horizontally over several convergence zones. A shallow perturbation alters early steep arrivals but not late axial arrivals; both are affected by a perturbation extending to axial depths.

The unscrambling of this information is discussed later; it depends on the condition that travel time is simply related to the way in which ray paths weight the water column, without the need for a precise knowledge of the ray orbits. The more straightforward (and ambitious) scheme is to identify each arrival with a known ray orbit $z(r)$ in the vertical plane between source and receiver.

We define a 'ray identifier'

$$\pm p^+, p^-, \tag{3}$$

where $\pm$ refers to upward–downward ray inclination at the source, p^+ is the number of upper turning points (surface reflections for SR/R) and p^- is the number of lower turning points (bottom reflections for R/BR). The three components of the identifier are not independent: the sign is always $(+)$ when $p^+ > p^-$ and always $(-)$ when $p^+ < p^-$. Figure 2 shows a few labeled ray paths.

The arrival structure is complicated. Rays come in groups: group 15, for example, contains $+15{,}15$; $-15{,}15$; $+16{,}15$; $-15{,}16$; to be followed by group 16, $+16{,}16\ldots$. The arrival sequence and spacing of the four constituents (to borrow a term from tidal literature) within a group depends on the depths of source and receiver, and the group number itself. Two constituents can arrive simultaneously, and this degeneracy causes trouble with resolution and identification. If necessary, we could use multiple receivers, which provide arrival angle as an additional discriminant, or one could place the transducers on flat-topped seamounts (guyots). For the sources, this would leave only constituents $+15{,}15$ and $+16{,}15$ in group 15. Similarly, for the receivers this leaves the constituents $-15{,}15$ and $+16{,}15$. If both sources and receivers are on guyots, then we are left with only $+16{,}15$, in group 15, with $+17{,}16$ in group 16, etc. Offhand, this may seem undesirable for having given away three fourths of the information. In fact, one sacrifices very little, for the constituents within a group do not carry independent information about mesoscale eddies (Section 6, Vertical Slice).

Aside from this ray degeneracy, there are problems with resolving and identifying any of the arrivals at long ranges, because of the spreading of the pulse by micropaths. We will now discuss the separation between (deterministic) multipath arrivals in some detail. The (probabilistic) spread is taken up in the next section.

Canonical sound channel

For definiteness we now consider the 'canonical sound channel' [Appendix A, equation (A1)]. A dimensionless sound speed excess is written

$$\phi^2(C) \equiv \frac{C-\bar{C}}{\varepsilon \bar{C}} = e^{\eta} - \eta - 1 = \frac{1}{2}\eta^2 + \frac{1}{6}\eta^3 + \ldots, \tag{4}$$

where

$$\eta = (z-\bar{z})/\tfrac{1}{2}B$$

is the dimensionless elevation above the sound axis $\bar{z}$ (order -1 km), 2ε (order 10^{-2}) is the fractional increase in sound speed in an adiabatic ocean (1.14%/km) over a scale depth B (order 1 km). The profile (4) follows directly from thermodynamic constants, a linear θ–S-relation, and the assumption of an exponential profile in buoyancy frequency

$$n = n_0 \, e^{z/B}. \tag{5}$$

Equation (4) describes perhaps the simplest sound channel that has built-in the fundamental asymmetry between the oceans above and beneath the sound axis. It will serve as an analytical guide to the subsequent discussion of path resolution and identification. However, the reader must be warned that, not unexpectedly, the application of this idealized sound channel to details of actual ocean measurements has left something to be desired (FLATTÉ *et al.*, 1978).

Set $\bar{z} = -1.3$ km as representative of the rather deep sound channel in the northwest Atlantic. (Off California $\bar{z} \approx -0.8$ km.) For a distance scale we choose $R_{cz} = 50$ km as a typical convergence zone range (this is twice the 'focal length' of the sound channel). All other scales, such as B, ε, and a time scale τ_0 then follow from the fundamental properties of a canonical ocean. The reader is referred to Appendix A for these relations and the derivation of the results illustrated in Figs. 4 and 5.

Suppose we have an axial receiver, and a source at a range of 20 convergence zones ($R \approx 1$ Mm). For an axial source ($z = z_s = -1.3$ km) the time of arrival is given by the intersection of the curves with the horizontal axis, and the first refracted arrival is the surface-limited doublet $\pm$ 15,15 (Fig. 4); its arrival time is $\tau = -2.75$ s before the final axial arrival. This is followed by + 16,15 at -2.33 s $-$ 15,16 at -1.98 s, the doublet + 16,16 at -1.66 s, etc. As time goes on the arrivals occur at ever closer intervals, with the axial doublet $\pm$ 20,20 at the finale.

As a second example, let the source be at $z = -2.5$ km, with the receiver remaining at the axis (or equivalently an axial source and a receiver at -2.5 km). The arrivals are given by the intersection of $z = -2.5$ km with the curves; the first arrival is + 15,15 at $\tau = -2.52$ s, then $-$ 15,16 at -2.16 s, followed very closely by + 16,15 at -2.15 s, $-$ 16,16 at 1.80 s, etc. But now the arrival sequence is terminated before the axial arrival. The lower turning point of $-$ 19,19 is at a depth $z = -2.4$ km and does not quite reach the source at -2.5 km. The final signal is associated with the closely spaced arrivals of + 18,18 and $-$ 18,19 at $\tau = -0.26$ and -0.25 s, respectively.

These illustrations are drawn for a canonical channel with a 50 km convergence zone. For different values of R_{cz} (but still with $R/R_{cz} = 20$) one needs to refer to the dimensionless coordinates τ/τ_0 and ϕ. For the general case one needs to go back to the analytical solutions, but many of the results are quite simple. Take for example the evolution from early steep ray paths to late axial paths. The cusps represent the upper and lower turning points, and these converge upon the axis as [equation (A8)]

$$\hat{\phi}^2 = \tfrac{1}{2}\omega + \tfrac{1}{2}\sqrt{(\omega^2 + 48\omega)}, \quad \omega = (-\tau/\tau_0)/(R/R_{cz}),$$

where $\hat{\phi}$ is the value of ϕ at the turning depth of the ray. Thus $\hat{\phi} \approx \omega^{1/2}$ for large ω and $(12\omega)^{1/2}$ for small ω. For an off-axis source (or receiver) the final arrival occurs when the cusps no longer reach the source distance from the axis.

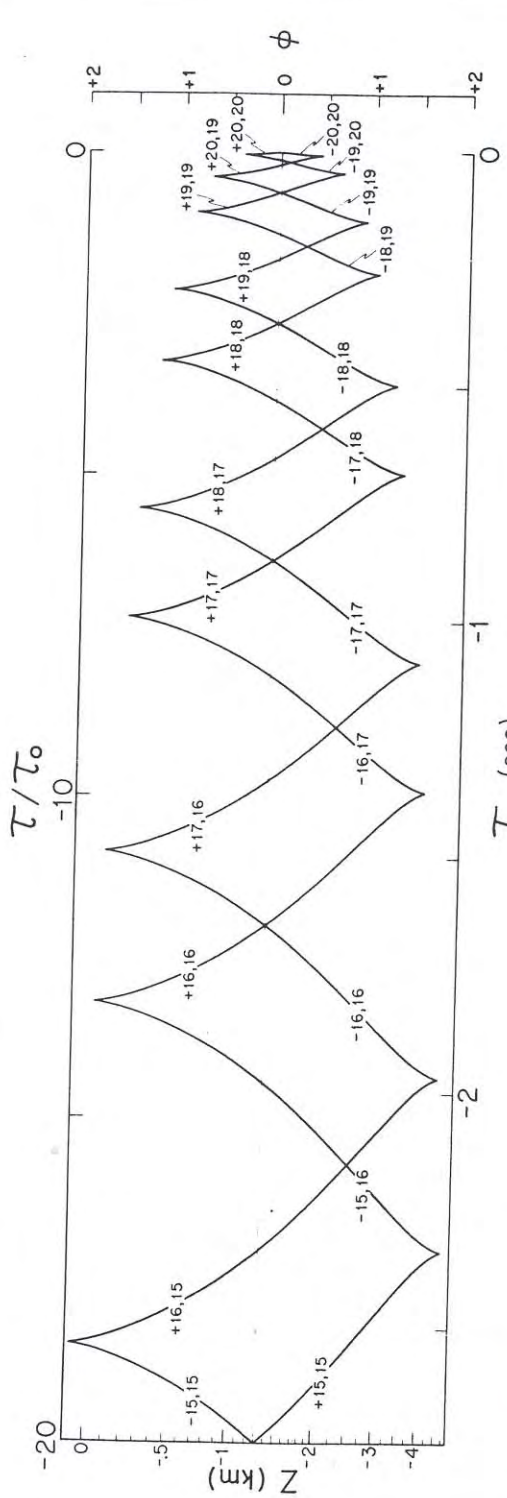

Fig. 4. R/R arrival structure at a range of 20 convergence zones in a canonical ocean. Numerical values correspond to the parameters $\bar{z} = -1.3$ km, $\bar{C} = 1.5$ km s^{-1}, $R_{cz} = 50$ km, hence $B = 1.44$ km, $\tau_0 = 0.137$ s, $\varepsilon = 0.0082$. Arrival times τ are reckoned from the final axial arrival and are thus negative. The vertical scale is associated with the source elevation: $\phi = 0$ corresponds to an axial source, $\phi = +2$ to a source near the surface or near the bottom. [The conversion to z-units is obtained from equation (4) using the stated numerical parameters.] For an axial receiver and a source at $z_s = z$ (or equivalently an axial source and a receiver at $z_r = z$), the arrival times are given by the intersections of $z = z_s$ with the constituent curves. For source and receiver both on the axis, the arrivals commence with group 15 (surface- or bottom-limited) and terminate with group $20 = R/R_{cz}$.

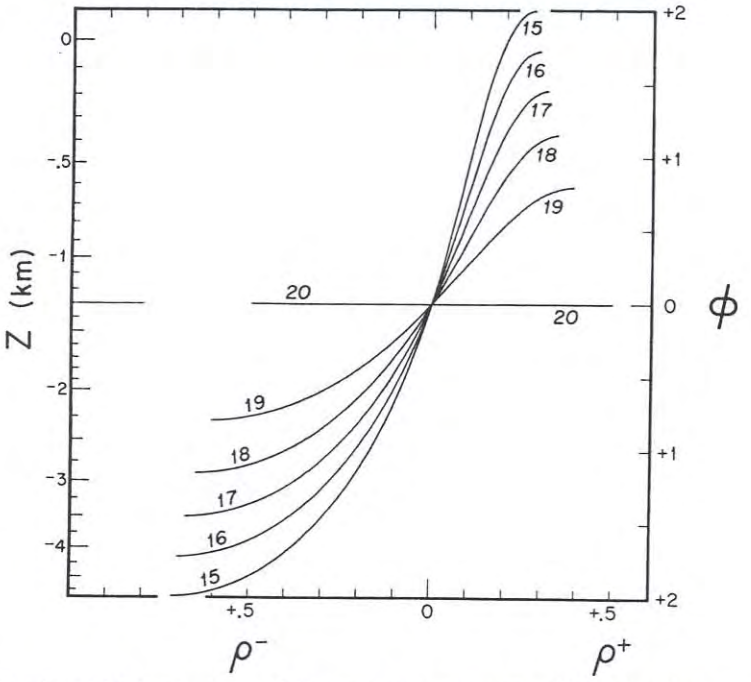

Fig. 5. Depth weighting at a range of 20 convergence zones ($R \approx 1\,Mm$) by subsequent group arrivals $\pm p,p$ with $p = 15, 16, \ldots 20$ ($\hat{\phi} = 2, 1.73, 1.46, 1.16, 0.80, 0$). $\rho(\phi,\hat{\phi})$ is the fractional time per double loop spent between the axis ($\phi = 0$) and ϕ [equation (A13)]. The depth scale (left) is drawn for a deep Atlantic sound channel, $\bar{z} = -1.3\,\text{km}$, $R_{cz} = 50\,\text{km}$.

The reader will have noticed three circumstances where two ray paths can arrive simultaneously:

(i) *Up–down degeneracy* $\pm\, p,p$ when source and receiver are at the same depth (axial in our first example) and the number of upper and lower loops are equal; $+\, p,p$ arrives before $-\, p,p$ when the receiver is above the axis, and vice versa;

(ii) *channel asymmetry degeneracy* $+p+1,p$ with $-p,p+1$. This degeneracy would be of type (i) in a symmetrical channel with equal travel time for an extra upper loop or an extra lower loop; the ocean assymetry places this degeneracy into the lower ocean. The arrival order is reversed as the sequence passes the depth of the degeneracy.

(iii) *Turning point degeneracies:* these occur as the turning points come near to the source, with $-p-1,p$ blending into $+\,p,p$ for upper turning points, and $+\,p,p-1$ blending into $-\,p,p$ for lower turning points. The inevitable effect is a reduction, with increasing arrival time, of the interval between the constituents. These degeneracies will cause difficulties with path resolution and identification. They call for a careful choice in the transmission geometry and the possible suppression of some of the constituents. (Hamilton just used the final cut-off, Fig. 3.)

The total number of ray groups is readily estimated from (A17) and (A1). The initial group number p_i is determined by the value of $\hat{\phi}_i = [(\hat{C}_i - \bar{C})/(\varepsilon\bar{C})]^{1/2}$, where

$$\begin{aligned} \hat{C}_i &= \text{lesser } (C_{\text{surface}}, C_{\text{bottom}}) && \text{for R/R} \\ &= C_{\text{bottom}} && \text{for SR/R} \\ &= C_{\text{surface}} && \text{for R/BR.} \end{aligned}$$

Taking $\hat{\phi}_i \approx 2$ for surface-limited R/R rays and $R/R_{cz} = 20$ gives $p_i = 15$. The final group number depends similarly on $\hat{\phi}_f$, with

$$\hat{C}_f = \text{larger } (C_{\text{source}}, C_{\text{receiver}}).$$

For an axial source and axial receiver, $\hat{\phi}_f = 0$ and accordingly $p_f = R/R_{cz} = 20$, with a total number of $p_f - p_i + 1 = 6$ RR groups.

The group interval is $2\tau_0\hat{\phi}^2$ [equation (A20)]; for a typical value $\hat{\phi} = 1$ this interval is 0.25 s. The final two groups are separated by $6\pi^2 B^2/(\bar{C}R) \approx 0.1$ s.

The relative weighting of different depth layers by the various group arrivals is plotted in Fig. 5. For example, take the rays $\pm$ 18,18 with turning points at $\hat{z}^+ = -0.37$ km and $\hat{z}^- = -2.90$ km ($\hat{\phi} = 1.16$). The fractional time spent within a 1-km layer centred at the axis (-0.8 to -1.8 km) is $\rho = \rho^+ + \rho^- = 0.13 + 0.14 = 0.27$.

Measured sound channel

For comparison we show the arrival sequence for typical sound speed profiles in the western North Atlantic and eastern North Pacific as obtained from numerical integration along the ray paths (Fig. 6). The coordinates can be interpreted as the phase (or trace) velocity and the group velocity associated with ray propagation: $C_p = C(z)\sec\theta(z)$ and this equals $\hat{C}$ from Snell's law and $C_g = R/T$.

The two ocean provinces differ considerably from one another (and from the canonical ocean). The Atlantic travel times are shorter by a few seconds than the Pacific travel times, and the duration of the arrivals is about twice as long. The first Atlantic R/R arrival is surface-limited, the final near-axial arrivals are limited by the source depth. Only two of four constituents are plotted, and these arrive increasingly closer in pairs as the turning depth approaches the source depth, in accordance with the display in Fig. 4. There is no chance to resolve these pairs in the late arrivals (nor is there a need to do so as they sample nearly the same ocean).

In the Pacific the earliest arrivals are also associated with the sound speed very near the surface, and the near-axial termination of the sequence is determined by source depth. But there is no law of nature that the axial arrivals come in last (PEDERSEN and WHITE, 1970). In the Pacific, the group velocity minimum is smaller than the axial sound speed by 1 m s^{-1}. Accordingly, $\hat{C}$ is double valued to both sides of the minimum, corresponding to distinct rays with the same ray identifier, but with different inclinations at the source and receiver. The identification of ray orbits is more difficult than for the monotonic case. One may have to resort to a combination of deep and near-axial sources (receivers) along the mooring, or to simple arrays that can discriminate between the two branches of the sequence on the basis of ray inclination.

Severe perturbation may create new ray paths that were absent in the unperturbed state. An example is provided by mesoscale eddies in the East Australia Current. These have been studied intensively during Project ANZUS (for Australia, New Zealand, United States) Eddy by satellite, aircraft and surface vessel (SCULLY-POWER, NYSEN, NILSSON, TWITCHELL, BROWNING, SWENSON, ANDREWS and BANNISTER, 1975), and by sound propagation (NYSEN, SCULLY-POWER and BROWNING, 1978). The eddies have warm, well-mixed cores measuring more than 100 km across and reach to depths of 300 m. In the Austral winter the cores extend to the surface where they give observable temperature anomalies and increased cloudiness. In the summer the cores appear to be confined to subsurface lenses. In one summer case studied in detail, the mixed core was associated with an eddy duct centred on a

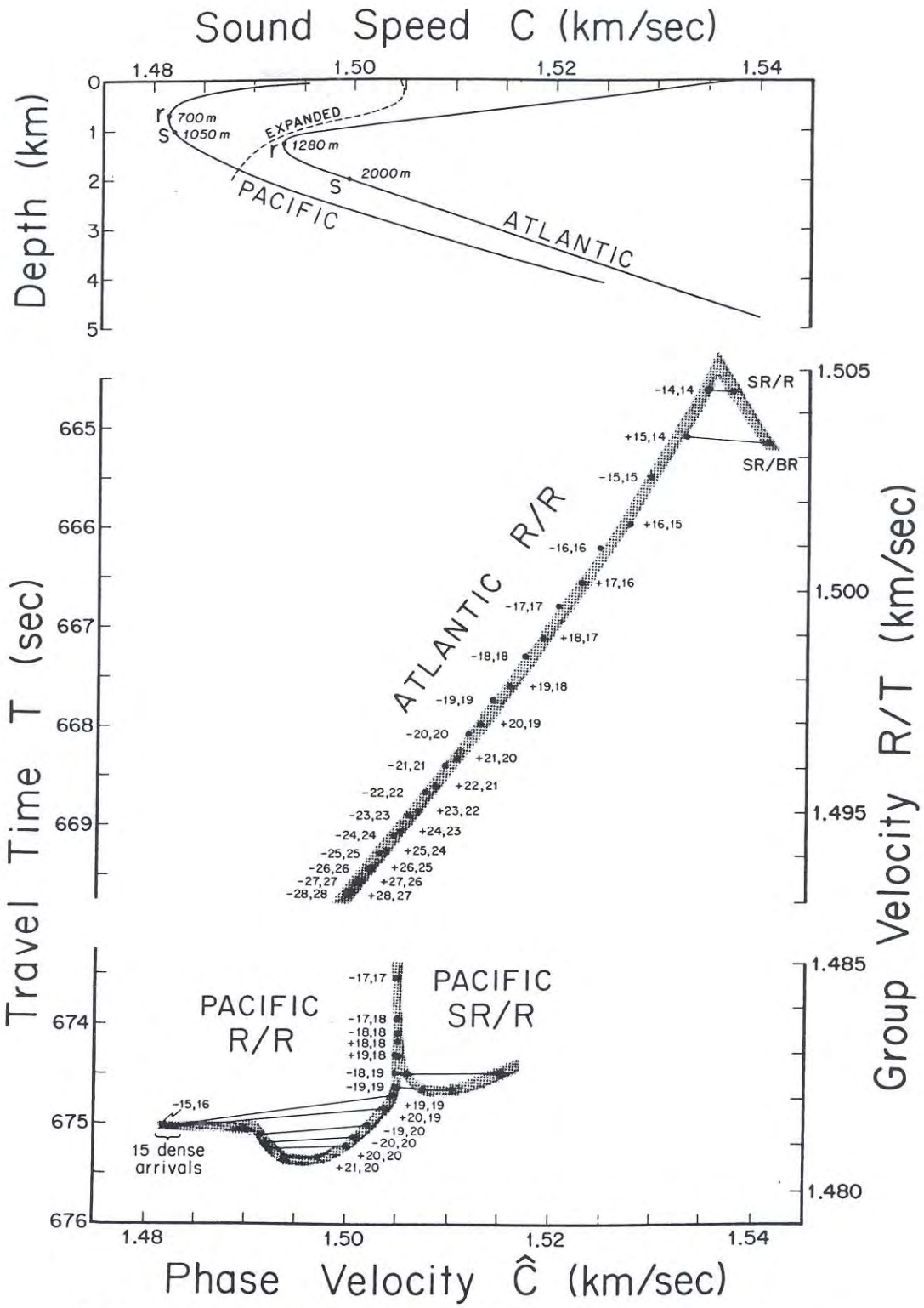

Fig. 6. Arrival sequence for Atlantic sound speed profile between Eleuthera and Bermuda (see Fig. 2) and for Pacific sound speed profile 220 miles west-southwest off San Diego (dashed curve on upper panel is Pacific profile in $10\times$ depth scale). Travel times T are for a range of 1 Mm, plotted against the sound speed $\hat{C}$ of the ray paths at the turning points, drawn to the same scale as the profiles in the upper panel. Alternately, the coordinates correspond to phase and group velocity. Each arrival is marked with a point; points with the same ray identifier $\pm p^{+}$, p^{-} are connected. For the Atlantic only those rays are plotted with travel downward toward the receiver $(-p, p; +p+1, p)$. We are indebted to John Clark of the Institute for Acoustic Research and Homer Bucker and Mel Pederson of NOSC for the ray calculation.

sound speed minimum at 210 m, in addition to a duct in the surface mixed layer, and, of course, the wave guide about the principal sound axis. Acoustic transmissions within the eddy structure over ranges of up to 85 km showed distinct surface and eddy duct arrivals, with leakage between the ducts (which is beyond the scope of ray optics). Even under such extreme conditions our procedure should still be applicable to those ray paths that were not trapped by the eddy duct, and this would apply to all ray paths if source and receiver lie above or beneath the duct.

The ANZUS Eddy Experiment points towards the difficulty in path stability and identification, and this is borne out in a numerical experiment by WEINBERG and ZABALGOGEAZCOA (1977). F. Fisher (private communication) has suggested that in such a situation one could calibrate the arrivals by aircraft-borne measurements.

4. RESOLUTION AND PRECISION

We have established the following design requirements:

(i) The resolution should be 50 ms, or less, to separate multipath arrivals.

(ii) The precision to which a (resolved) arrival is timed should be 25 ms, to avoid significant statistical error in the tomographic inversion (see numerical experiments, Fig. 7).

(iii) The depths of source and receiver should be near the axis, to give variety in ray paths for vertical discrimination.

In this section we will consider sources of error in the measurement of travel times. These errors arise for a number of reasons, including mooring motion, clock drift, ocean acoustic noise, and the finestructure in the sound speed profiles. Some of the analysis is intricate; we conclude that the resolution and precision requirements can be met. The reader who simply wishes to accept this conclusion can skip to Section 5 where we deal with the problem of deducing the ocean structure given sufficiently precise travel time measurements.

Mooring motion

For a 25 ms precision excursions need to be kept to within ± 35 m. It would, of course, be possible to allow much larger excursions, provided they can be measured and corrected for. There is no difficulty in monitoring mooring motion to within a fraction of a meter using bottom-based sonic transponders (PORTER, SPINDEL and JAFFEE, 1973; CHHABRA, 1977). The trouble is that the measurements have to be recorded and analyzed *in situ*, and somehow transmitted if one is to use the measured travel times in 'real time'. One would like to avoid all this.

Tri-moorings can be designed to stay within the required limits of motion. However, the setting of tri-moorings is awkward and time-consuming.

In some ocean areas (e.g. the eastern Pacific) a major fraction of the abyssal current is associated with predictable (barotropic) tidal components for which the mooring motion can be corrected. Much of the residual motion is at frequencies above the inertial frequency ω_i, and accordingly suppressed in long time averages. If the power spectrum of the mooring displacement is peaked at the inertial frequency and drops off as ω^{-2} for $\omega > \omega_i$ (as for internal waves), then averages formed over the measuring interval Δt_M are reduced by a factor $\omega_i \Delta t_M$; weekly averages reduce the time error by a factor of 10.

There remains the problem of mooring motion at low frequencies, such as those of the mesoscale perturbations themselves. We believe it is possible to achieve the requirements with very stiff one-point moorings (N. K. Chhabra, private communication). For a wire tension of 1500 lb (well within the breaking strength of 4000 lb for 3/16 in wire), a Webb SOFAR float moored 1 km above bottom in a uniform current of 5 cm s^{-1} can be held to within 2 m. We shall require 2.5 km moorings, so that the source is no deeper than 1 km beneath the sound axis; we estimate that with a determined effort it will be possible to keep the mooring motion within tolerable limits.

Timekeeping

The requirement is for a precision of 25 ms over 1 year. The abyssal environment is benign; temperature fluctuations will be within a few degrees.

Rubidium frequency standards retain their frequency to within 3×10^{-11} parts, or 1 ms in 1 year. The trouble is that they draw 13 W, more than the average power required for the acoustic transmission. Battery packages for 1 year's operation are extremely bulky.

Power consumption for quartz crystals can be kept to less than 1 W, and the performance of the best available crystals after a suitable warm-up time (1 month!) can meet our requirements. The timekeeping problem is crucial, and we shall present the evidence in some detail.

We have analyzed two 3-year records of mean daily frequencies of two carefully aged oscillators.* Frequencies were read at expedition start times t_0 and $t_1 = t_0 - 1$ month, and linearly extrapolated to $t = t_0 + 1$ month. The clock is set at t_0; integration of the linearly extrapolated frequencies gives computed clock time, integration of the measured frequencies gives measured clock time. Clock errors were computed for monthly intervals of t_0 (Table 1). The result is 0.14 ms r.m.s. error for both clocks. For 6-month and 12-month extrapolations, the r.m.s. errors are 4 and 18 ms, respectively. A 6-month mid-expedition interpolation gives smaller clock errors than a 6-month extrapolation.

Table 1. R.m.s. clock errors $\delta\varepsilon$ (ms) *for two F. E. precision crystals. The clock is 'set' at time* t_0 *(the start of the expeditions); frequencies are read at times* t_0 *and* t_1. *Clock errors at times* t *are based on a linear extrapolation (interpolation) of crystal frequency, by direct calculation based on the two* 3-year *time series of frequency, and by a statistical theory based on the power spectrum of the frequency record.*

Months			rms $\delta\varepsilon$ (ms)	
t_o	t_1	t	time domain	frequency domain
Extrapolation				
0	-1	1	0.14	0.13
0	-1	6	4.1	3.3
0	-1	12	18.2	12.6
Interpolation				
0	+12	6	2.8	1.1

A more satisfactory computation of expected errors is based on the spectrum of the crystal time series. We find that the power spectrum of the fractional variation in crystal

* Mr M. B. Bloch of Frequency Electronics, Inc., has kindly made these records available. We are indebted to B. Zetler for the calculations.

frequency can be tolerably fitted by

$$S(f) = \alpha_1 f^{-1} + \alpha_2 f^{-2}, \tag{8}$$

consisting of flicker noise plus random walk noise,* with

$$\alpha_1 = 2 \times 10^{-22}, \quad \alpha_2 = 8 \times 10^{-29}\ \text{Hz}.$$

(Between 1/3 and 1 cpy the measured spectrum lies above the fitted spectrum.) Set $t_0 = 0$; for $t > \alpha_1/\alpha_2 \approx 10$ days, the effect of random walk noise dominates, giving a mean-square clock error (Appendix B)

$$\langle(\delta\varepsilon)^2\rangle = \pi^2 \alpha_2 t^3 \left(\frac{2}{3} - \frac{1}{2}\frac{t}{t_1}\right).$$

The results (Table 1, last column) are generally consistent with the direct calculation.

For the case of the mid-experiment interpolation, $t/t_1 = +\frac{1}{2}$, a solution based on linear clock drift gives $\langle(\delta\varepsilon)^2\rangle = \frac{1}{3}\pi^2\alpha_2 t^3$, as compared to $(\frac{5}{12})\pi^2\alpha_2 t^3$ for the linear frequency drift. A quadratic frequency drift accommodating both frequency and time at start and end yields $(\frac{1}{12})\pi^2\alpha_2 t^3$. None of these procedures is optimum, of course; but we have found that optimum estimates do not improve the numerical results significantly.

The conclusion is that with care (and expense) we can meet our timekeeping requirements. There is always the chance of a 'catastrophic' frequency jump, but this would show up in the analysis by the fact that all travel times from one source would be anomalous. In this sense the tomographic data reduction can be made self-correcting to allow for clock errors, at the expense, of course, of the precision of the computed δC fields.

Pulse compression

We now consider the resolution and precision of ocean acoustic ranging.

Underwater explosions are the traditional source of acoustic pulses for measuring travel time. The shock front is followed by a series of bubble pulses at intervals of order 10 ms, but these can be handled by an appropriate deconvolution of the received signal, or they can be avoided entirely by the use of compressed steam†. For our purpose it is better to use broadband transmitters optimally coded (HELSTROM, 1968; BIRDSALL, 1976; WORCESTER, 1977) to yield a sharp autocorrelation peak with low sidebands. The procedure is to form the covariance of the received signal with a replica of the transmitted signal (with some modifications); for many purposes the covariance peak can be regarded as if it had in fact been the transmitted pulse.

The covariance peak width $\Delta\tau$ is roughly $(\Delta f)^{-1}$ s, where Δf is the transmitter bandwidth in Hz. Let P be the received signal power. For a duration Δt of the coded sequence, the signal energy $P\Delta t$ is compressed into an interval $\Delta\tau$, giving an effective power of $P\Delta t/\Delta\tau = P\Delta f \cdot \Delta t$. Thus pulse compression results in a processing gain of $\Delta f \cdot \Delta t$.

The resolution is $\Delta\tau = (\Delta f)^{-1}$. To resolve multipaths separated by 50 ms requires $\Delta f > 20$ Hz. There is a conflict here between efficient narrow band resonant transmitters and the requirement for a broad bandwidth.

For a resolved single path (no micropaths) the precision in measuring arrival times is determined by the ocean noise level (which produces a jitter in the zero crossings of the

* National Bureau of Standards Note 394, October 1970.

† VAPORCHOC built by Compagnie Générale de Géophysique.

received signal). The r.m.s. error in travel time is given by

$$\sigma = \Delta\tau(S/N)^{-1/2}, \quad \Delta\tau = 1/\Delta f \tag{9}$$

and can be reduced to any requireed limit by increasing the singal-to-noise ratio. The (compressed) signal peak power is $2P\Delta f \cdot \Delta t$. The noise power within the transmitted band is $N\Delta f$, where $N(f)$ is the noise power spectral density. Accordingly

$$\frac{S}{N} = \frac{2P\Delta f \cdot \Delta t}{N\Delta f} = \frac{2P\Delta t}{N}$$

is proportional to the energy transmitted in a coded sequence.

Micropaths

For 1-Mm transmissions at a few hundred Hz, resolution and precision are normally limited not by the transmitter bandwidth but by the bandwidth imposed by the variable ocean finestructure.

We can regard the sound speed profile $C(z)$ as an average taken over the transmission path. Then

(i) at any place and time there will be a finestructure superimposed on $C(z)$ due to vertical straining by internal waves and other sources of finestructure.

(ii) As a result, any (deterministic) multipath associated with the mean profile $C(z)$ will be split into a bundle of (sporadic) micropaths. (Each micropath obeys a Fermet condition of stationary travel time.)

(iii) Acoustic travel times differ from micropath to micropath, and accordingly the transmitted pulse is spread from the transmitted pulse width $(\Delta\tau)_T = (\Delta f)_T^{-1}$ to

$$\Delta\tau = (\Delta\tau)_T + (\Delta\tau)_O.$$

By analogy, we define $(\Delta f)_O = \pi^{-1}(\Delta\tau)_O^{-1}$ as the ocean coherent bandwidth,* or ocean decorrelation frequency. Experiments at frequencies separated by more than $(\Delta f)_O$ give independent estimates of arrival time.

(iv) The detailed structure of the received signal varies with the interference between micropaths, and decorrelates in a time $(\Delta t)_O$. Measurements separated in time by more than $(\Delta t)_O$ give independent estimates. [$(\Delta t)_O^{-1}$ is the frequency spread of a transmitted CW signal.]

The calculation of ocean decorrelation time and frequency is the object of a forthcoming monograph (FLATTÉ *et al.*, 1978). The decorrelation parameters are expressed as functions of a strength parameter Φ that depends on the variance of $\delta C/C$ along the acoustic path and the scale parameter Λ, which depends on the coherence scale of $\delta C/C$ relative to the Fresnel radius. Λ in particular is a sensitive function of the ocean environment. For the special case of a canonical sound channel perturbed by a GARRETT–MUNK (1975) model of the internal wave spectrum one obtains roughly

$$\Phi = 3.6 \times 10^{-2} f R^{1/2}, \quad \Lambda = 64 f^{-1} R, \quad \text{(steep path)}$$

$$\Phi = 4.8 \times 10^{-2} f R^{1/2}, \quad \Lambda = 19 f^{-1}, \quad \text{(flat path)}$$

* For a Gaussian pulse, $\Delta f \cdot \Delta\tau = \pi^{-1}$, where $\frac{1}{2}\Delta f$ and $\frac{1}{2}\Delta\tau$ are defined by the second central moments. For other pulse shapes $\Delta f \cdot \Delta\tau$ is larger than π^{-1} and approaches 1 in many practical cases. We use $\Delta f \cdot \Delta\tau = 1$ for transmitters, and the theoretical limit $(\Delta f)_O \cdot (\Delta\tau)_O = \pi^{-1}$ for the oceans (FLATTÉ *et al.*, 16.3, 1978).

with R in Mm and f in Hz. The ocean pulse spread is given by

$$(\Delta\tau)_O = 6\pi^{-3} f^{-1} \Phi^2 \Lambda \ln \Phi = 16 \ln \Phi \, R^2 \text{ ms}, \quad \text{(steep path)}$$
$$= 8 \ln \Phi \, R \text{ ms}, \quad \text{(flat path)} \tag{11}$$

and the ocean decorrelation time by

$$(\Delta t)_O = 2000^s/\Phi = 0.4 \times 10^5 f^{-1} R^{-1/2} \text{ s}, \quad \text{(steep path)}$$
$$= 5400^s/\Phi = 1.5 \times 10^5 f^{-1} R^{-1/2} \text{ s}, \quad \text{(flat path)}. \tag{12}$$

A requirement for resolving near-surface multipaths limits us to ranges of 1 Mm. Close doublets are difficult to resolve under any practical circumstances.

JOBST and DOMINIJANNI (1978) carried out some 206 ± 25 Hz transmissions over a steep path of 0.5 Mm. The measured spread is somewhat less than 10 ms. The computed spread (11) is 7 ms.

In the long run it is the ocean itself that imposes the limits to the precision to which it can be acoustically monitored. For a resolved pulse, the r.m.s. error in arrival time due to ocean micropaths is

$$\sigma = \frac{(\Delta\tau)_O}{\sqrt{r}}, \quad r = \frac{(\Delta t)_M \cdot (\Delta f)_T}{(\Delta t)_O \cdot (\Delta f)_O}. \tag{13}$$

$(\Delta t)_M$ is the time interval over which the arrival time T is to be averaged, and $(\Delta f)_T$ is the transmitter bandwidth. Accordingly r is the ratio of the time–bandwidth product of the measurement to that of the ocean, e.g. the number of independent estimates of T. For a weekly average with a 200-Hz transmission of 10-Hz bandwidth, we have

$$\sigma = 0.6 R^{3/4} \text{ ms}, \quad 0.8 R^{1/4} \text{ ms} \qquad (R \text{ in Mm})$$

for steep and flat paths, respectively. At long ranges the axial perturbations can be measured to greater precision, but mesoscale signals are smaller at this depth.

The number of independent samples in a week is very large, $r = 3000 R^{5/2}$ for the near-surface paths and $r = 460 R^{3/2}$ for near-axial paths (R in Mm). One is, of course, skeptical of such a large $\sqrt{r}$ gain; it is like obtaining the height of the emperor of China to the nearest micron by averaging the estimates of his 850 million loyal subjects.

One implicit assumption is that there are no intermediate time scales between the finestructure and the mesoscale. This assumption is supported by observed spectra of ocean variability (RICHMAN, WUNSCH and HOGG, 1977), which have a gap between the mesoscale and inertial frequencies.

The foregoing numerical values are only approximate, and it will take further work to give reliable estimates of the micropath pulse spread. The present conclusion is that the resolution of multipath arrivals can be accomplished. Once resolved, an average over successive measurements should yield a precision in arrival time that exceeds the requirement for mesoscale monitoring.

Ultra-low frequencies

There is an intriguing possibility of working at such low frequencies that arrival time differences between micropaths are less than $1/2\pi f$, so that there are no cancellations from opposite phases. Under these circumstances, micropaths no longer constitute the limiting factor, and the error in arrival time is given by

$$\sigma = (2\pi f)^{-1} (S/N)^{-1/2}. \tag{14}$$

Table 2. Acoustic power parameters for 1-Mm *range.*

Frequency band, $f \pm \frac{1}{2}(\Delta f)_T$, Hz	220 ± 20	30 ± 20
Pulse duration Δt, sec	240	240
Source level,[1] dB/μPa	178	170
watts acoustic	9	1
Transmission directivity gain,[2] dB	7	7
Spherical spreading,[3] dB	-120	-120
Attenuation, dB	- 10	- 2
Intensity at receiver dB/μPa	55	55
Noise: wind, dB/Hz	65 ± 10	40 ± 10
Shipping, dB/Hz	55 ± 10	65 ± 10
Total within transmitted band, dB	81	81
S/N at receiver, dB	- 26	- 26
Receiver gain, directivity[4] dB	13	6
pulse compression,[5] dB	33	40
S/N, dB	20	20

[1] The numbers refer to the *combined* power output of the transducer array.
[2] For five transducers at $\frac{1}{2}\lambda$ vertical separation the directivity gain is 10 log 5.
[3] 20 log R with $R = 10^6$ m. Cylindrical spreading (as commonly used beyond 10 km) applies only if multipaths are summed, and we need to resolve multipaths.
[4] These are estimated from 10 log 2 (array length/λ) with some practical upper limits.
[5] 10 log $(\Delta t \cdot \Delta f)$, where Δt is the pulse duration and Δf = lesser $[(\Delta f)_T, (\Delta f)_O]$. This gives $\Delta f = (\Delta f)_O = 8$ Hz ocean bandwidth for the 220-Hz transmission, and $\Delta f = (\Delta f)_T = 40$ Hz transmitter bandwidth for the 30-Hz transmission.

At large ranges the transition occurs at $\Phi = 1$. For the canonical ocean model with a GM internal wave spectrum, this corresponds to about 30 Hz at 1 Mm. For $S/N = 100$ (20 db), $\sigma = 0.5$ ms! Thus a *single* transmission measures travel time with far more than the required precision.

Some very low frequency sources have been built, but these are bulky, expensive, and transmit over a frequency band far too narrow for multipath resolution.

Table 2 gives an estimate of the power requirements and S/N requirements for the two frequency bands under consideration. The range is taken as 1 Mm. The 220 Hz transmission is for a steep path and fully saturated (many micropaths) whereas the 30-Hz transmission has no micropaths. Power requirements for the 220-Hz transmission are more severe on account of attenuation and of the limited processing gain from pulse compression in a saturated environment. The computations are for a coded transmission sequence over 4 min, thus taking full advantage of the ocean decorrelation time at 220 Hz

(there is no decorrelation limit at 30 Hz). If one sequence each hour should prove adequate, as may well be the case, the *average* power requirement would be 1/15 of the stated values.

5. HORIZONTAL SLICE

We now need to infer the ocean structure from the measurements of acoustic travel time along different ray paths. The procedure is called 'inverse theory', a subject that has been developed independently in a number of disciplines. The literature of geophysical inverse problems is vast (see BACKUS and GILBERT, 1967, 1968; PARKER, 1977; WIGGINS, 1972, COLIN, 1972, etc.), and there is a corresponding literature in tomography (SWINDELL and BARRETT, 1977). An account in terms perhaps more familiar to oceanographers is given by WUNSCH (1978), and it is worth pointing out that the procedure is a special case of what is normally called 'optimal estimation' (e.g. LIEBELT, 1967; GELB, 1974).

There are many ways to proceed and the methods outlined here have been adopted to be specific. Actual experience with field data and the search for computational efficiency will undoubtedly lead to improvements, but we believe that the principles will survive.

The time scales of interest are in the range of 1 week and longer so that signal-to-noise ratios can be improved by long-term averaging. For the moment we ignore any noise and assume that the measurements are perfect.

Perfect measurements

Suppose then we have a configuration of sources and receivers as displayed in Fig. 1. Figure 2 shows multipath arrivals in a vertical plane through a source and receiver. These paths sample the water column in different ways, so that the problem is fully three-dimensional. It is, however, convenient to specialize initially to the case in which sources and receivers are all at axial depth, and that the direct axial path is resolved and the only one used, as in Hamilton's SOFAR experiment (Fig. 3). This is the 'horizontal slice' model.

In this model we must then find, using the travel time information in the horizontal slice, the best estimate of the sound speed perturbations within the area. Ultimately these can be converted, with some determinable error, to the density perturbations, which are of more direct interest [equation (1)].

We divide the area of Fig. 1 into a series of sub-areas, $j = 1, J$. The grid can be arbitrary, but it might as well be an expression of geographic truths and oceanographic prejudices. (Here we use a square grid for analytical convenience.) The ray paths from all sources to all receivers are labelled $i = 1, I$. The travel time along path i is then given by

$$T_i = \sum_{j=1}^{J} R_{ij}/C_j, \tag{15}$$

where R_{ij} is the path length of ray i in grid area j (many R_{ij} are zero), and C_j is the average sound speed in grid area j. Ray curvature in the horizontal plane is neglected.

We can expand the sound speed

$$C_j = C^0 + \Delta C_j + \delta C_j,$$

where C^0 is an overall space–time average, ΔC_j are the mean spatial departures, and δC_j are the fluctuations. With the exception of extreme areas like the Gulf Stream both

departures are of order $10^{-2}C^0$. Call $C_j^0 = C^0 + \Delta C_j$; then

$$T_i = T_i^0 + \delta T_i = \sum \frac{R_{ij}}{C_j^0} - \sum \frac{R_{ij}}{C^0}\frac{\delta C_j}{C^0} + \sum \frac{R_{ij}}{C_j^0}\frac{\delta C_j(\delta C_j + 2\Delta C_j)}{(C^0)^2} + \ldots$$

The last term is of second order in the perturbations. Thus to first order,

$$T_i^0 = \sum_{j=1}^{J} R_{ij}/C_j^0, \tag{16}$$

$$\delta T_i = \sum_{j=1}^{J} E_{ij}\delta C_j, \quad E_{ij} = -R_{ij}/(C^0)^2, \quad i = 1, I. \tag{17}$$

If the initial perturbations ΔC_j^0 are sufficiently large we may wish to use a slightly different form of expression for δT_i, requiring only minor modification to what follows.

With I travel time perturbations and J grid areas, this system of simultaneous equations for δC_j can be over-determined, under-determined, or just-determined depending upon the ratio of I to J and the configuration of sources and receivers. The number of boxes necessary to describe the ocean is determined by the horizontal scale of the perturbations relative to the area of ocean under study. Grid area size should be chosen to be about the size of the expected perturbations. We believe that the under-determined case is likely to be the useful one.

The spatial resolving power of the tomographic procedure is a function of position within the area; it is preferable to have the system give maximum resolution where it can, rather than degrading it at the outset by choice of too coarse a grid. Choosing too fine a grid, can, however, lead to false oscillations in the solution, analogous to problems with super-resolution in time series analysis.

At this point we consider specific examples (Fig. 7A). The total area is taken as 1000 km $\times$ 1000 km; each box is 250 km $\times$ 250 km—somewhat large for mesoscale eddies. There are four sources and four receivers, with 16 rays ($i = 1, 16$) giving 16 equations to determine the sound speed perturbations δC_j in the 16 grid areas ($j = 1, 16$).

We write the system of equations

$$\mathbf{E}\,\delta C = \delta T, \quad \mathbf{E} = \{E_{ij}\}, \quad \delta C = \{\delta C_j\}, \quad \delta T = \{\delta T_i\}. \tag{18}$$

Superficially, configuration A is a fully determined system. But this is not the case if some of the eigenvalues of the matrix $\mathbf{E}$ vanish. In fact,

$$\delta T_6 = \delta T_5 - \delta T_1 + \text{constant} \cdot \delta T_2$$

so that δT_6 is not independent of the other travel times. It can be shown that configuration A is of rank 12, i.e. it has only 12 independent pieces of information. A slightly different configuration B with less symmetry has rank 13; a quite different layout C brings the rank up to 14. Configuration D has the full rank 16, but many of the 64 boxes are not covered at all and others are poorly resolved. Still, the smaller grid will give useful information in the central area.

We do not know if it is possible to orient the sources and receivers so as to make a rank 16 system with the 16 grid areas as shown; this is an optimization problem that needs to be examined. But redundancy is not necessarily a liability; it makes the configurations less sensitive to noise (as will be shown) than a system that is technically of full-rank but has small eigenvalues.

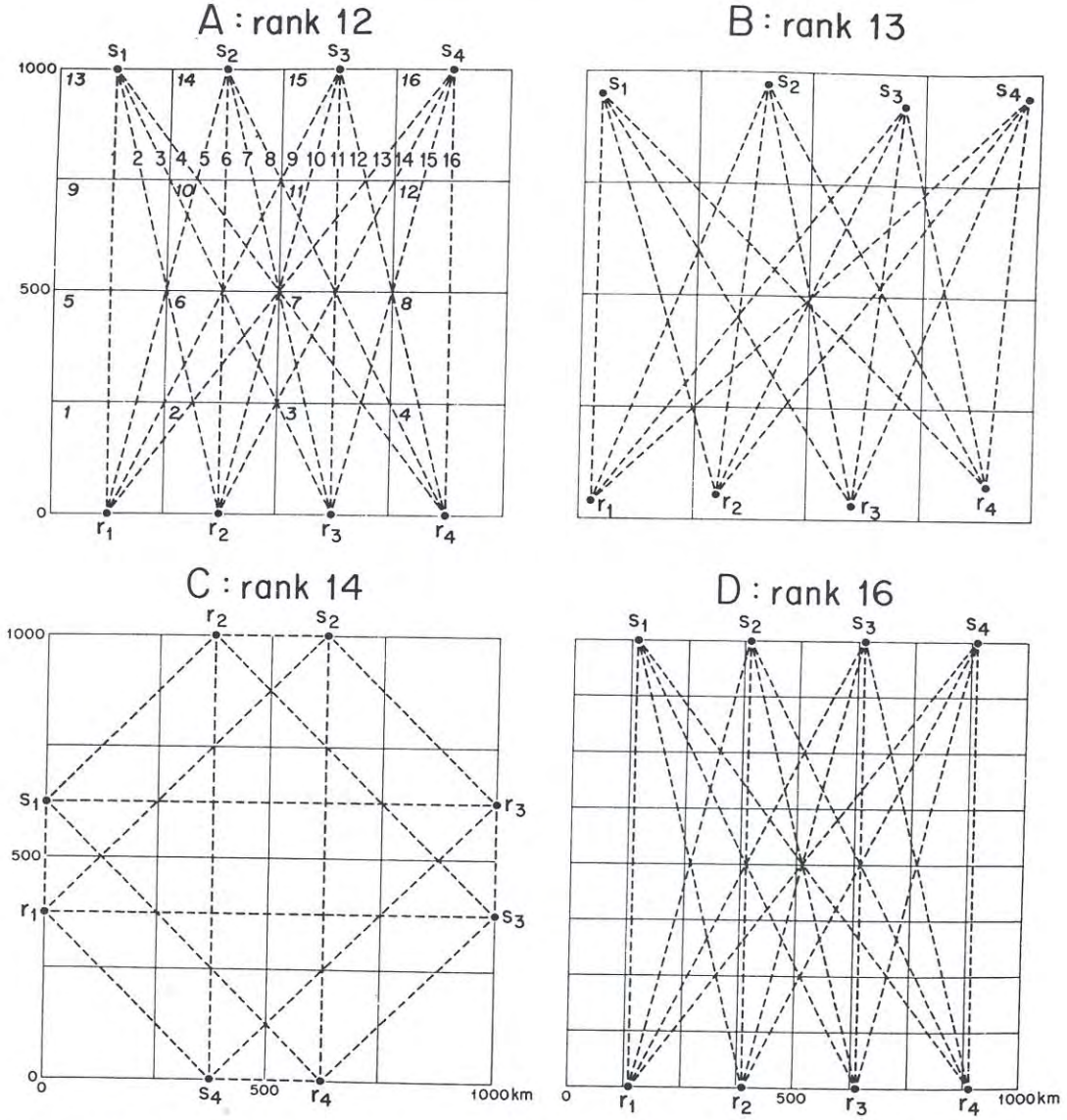

Fig. 7. Acoustic rays (dashed) are numbered $i = 1, 16$. The 1000 km × 1000 km area is divided into grid areas (bounded by solid lines) $j = 1, 16$ in configurations A, B, C, and $j = 1, 64$ in D.

We want to solve (18) as best we can. Because $\mathbf{E}$ is singular, we cannot simply invert it to find δC. Instead we use the singular value decomposition (LANCZOS, 1961; WIGGINS, 1972; WUNSCH, 1978) and write

$$\delta C = \sum_{l=1}^{L} \frac{(U_l \cdot \delta T)}{\lambda_l} V_l, \tag{19}$$

where

$$\mathbf{E}^T\mathbf{E}V_l = \lambda_l^2 V_l, \quad \mathbf{E}\mathbf{E}^T U_l = \lambda_l^2 U_l, \quad l = 1, L. \tag{20}$$

$\mathbf{E}^T$ is the transpose of $\mathbf{E}$. Equations (20) have nonzero solutions (termed the singular values) λ_l, $l = 1, L$, where L is the rank of $\mathbf{E}$. These solutions are ordered starting with the largest λ. For each singular value λ_l there corresponds a singular vector U_{il}, $i = 1, I$ and a singular vector V_{jl}, $j = 1, J$, so normalized that

$$\sum_{i=1}^{I} U_{il} U_{il'} = \delta_{ll'}, \quad \sum_{j=1}^{J} V_{jl} V_{jl'} = \delta_{ll'}.$$

U and V are normalized 'empirical orthogonal functions'. They are evaluated once and for all for a given configuration.

In equation (19), the expression

$$(U_l \cdot \delta T)/\lambda_l = \sum_{i=1}^{I} U_{il} \delta T_i/\lambda_l \equiv \delta T_l' \tag{21}$$

yields an ordered dataset $\delta T_l'$ with $L(\leqslant I)$ independent pieces of information, derived as a linear weighted sum of the measured dataset δT_i, $i = 1, I$. The desired perturbations δC_j are obtained as a sum of L orthogonal vectors V_{jl}, each weighted by $\delta T_l'$.

It can be shown (see references) that equation (19) is the particular solution to equation (18) with the minimum square perturbations $\delta C^T \delta C$; owing to the singular nature of $\mathbf{E}$, there are, of course, an infinite number of other solutions that differ from the one we have chosen by arbitrary multiples of all vectors satisfying $\mathbf{E} V_l = 0$, $l = L, J$, and about which we have no information. [These null space vectors normally contain higher wavenumbers than solution (19).] In principle, one can always obtain a unique answer by reducing the resolution to the point where the system becomes fully determined. The singular value decomposition explicitly shows which linear combinations of δC are determined.

Some numerical inversion experiments for configuration A are displayed in Fig. 8; the top left panel deals with the case of perfect measurements. The lower number in each grid area is δC_j, chosen at random and uncorrelated with the value in any other grid area. These values were used to compute the direct problem—the values of δT_i, $i = 1, 16$. The upper numbers are the result of inverting (18) by (19). The correlation between the 'true' and inferred value of δC is 0.81. A similar inversion (not shown) of a rank 13 system does better as one might expect; it gives a correlation of 0.91.

In some grid areas, the inversion has obviously done extremely well; in others the inferred value is rather far from the true value. We need some method of knowing where we can expect good results and where bad results. Notice that there is still no error in the measurements.

This question can be examined in terms of the response to a unit perturbation in box j_0 and zero elsewhere: $\delta C_j = \delta_{jj_0}$. We are dealing with a system of rank L in a vector space of dimension J; equation (19) tells us how to represent a vector in this space in terms of a set of expansion vectors V_l:

$$\delta_{jj_0} = \sum_{l=1}^{L} \alpha_{j_0 l} V_l + \varepsilon = \tilde{\delta}_{jj_0} + \varepsilon,$$

where $\alpha_{j_0 l}$ are constant coefficients. How well can δ_{jj_0} be represented by only L components of V_l in J-dimensional space? The error ε must be orthogonal to the expansion vectors V_l. This condition leads to

$$\tilde{\delta}_{jj_0} = \sum_{l=1}^{L} V_{jl} V_{j_0 l},$$

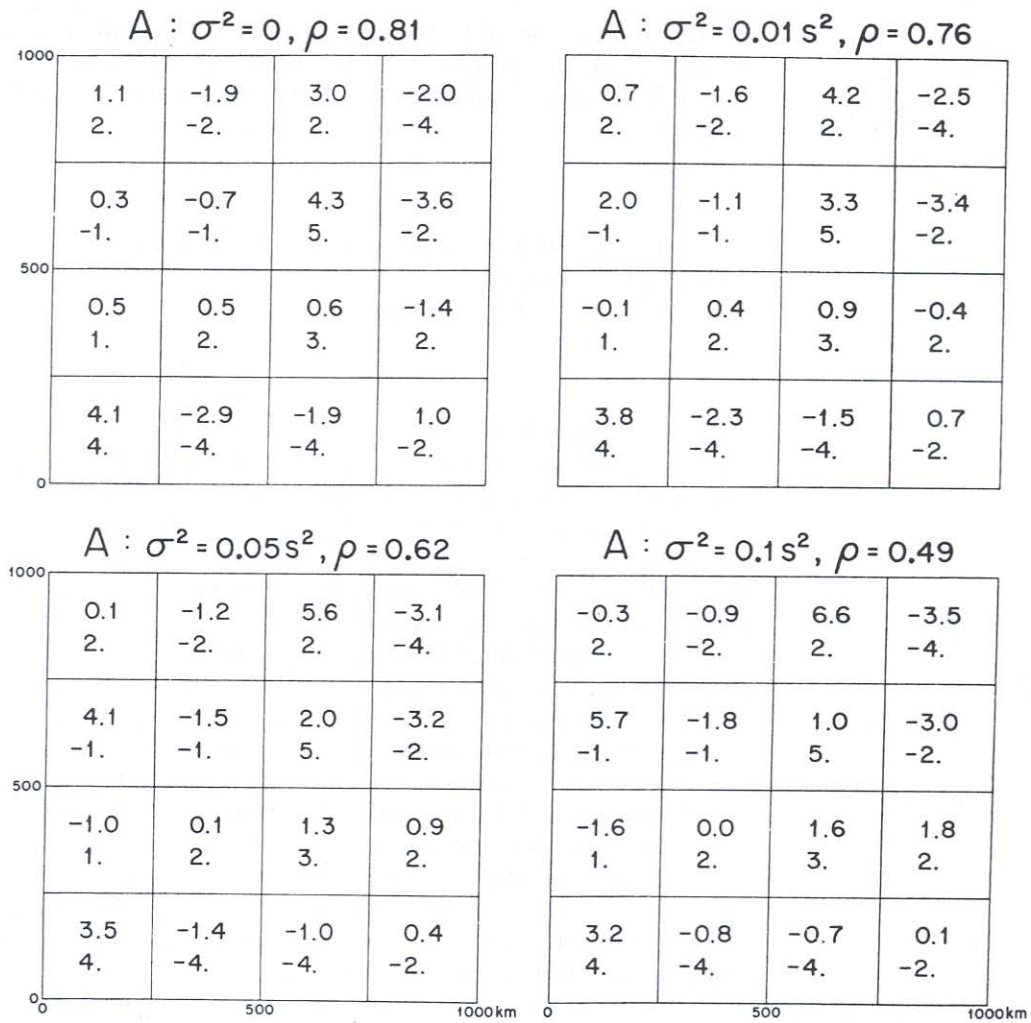

Fig. 8. Result of inversion of configuration A, Fig. 7. Lower numbers in each grid area are the 'true' values of δC_j in $\mathrm{m\,s^{-1}}$ (from which δT_i values were computed). Upper numbers are the values determined by inversion using equation (19). Mean-square errors in δT_i are taken as $\sigma^2 = 0, 0.01\,\mathrm{s^2}, 0.05\,\mathrm{s^2}$, and $0.1\,\mathrm{s^2}$, respectively. The correlation between true and inferred values of δC_j is $\rho = 0.81, 0.76, 0.62$ and 0.49, respectively.

which is the jth column of $\mathbf{V}\mathbf{V}^T$. ($\mathbf{V}$ is the $J \times L$ matrix whose columns are V_l.)

Figure 9 (left) is a plot of the 7th column of $\mathbf{V}\mathbf{V}^T$; e.g. the response to a unit input in grid area $j_0 = 7$. The inferred value δC_7 is only 0.8. The failure to come up to unity indicates an inadequate resolution, and this is apparent also from the sidelobes in neighboring grid areas. The diagonal element of $\mathbf{V}\mathbf{V}^T$ (Fig. 9, right) gives the inferred value in any grid area j_0 from a unit input in this same grid area j_0. It is evident that the value 0.8 is characteristic of the central ocean area; at the left and right margin the resolution is somewhat poorer. The nearly uniform values suggest that the underdetermined nature of configuration A is such that no single grid area is very well resolved nor very poorly resolved.

We have performed the inversion of the system by dividing the area with a grid and then solving the resulting algebraic equations for the unknowns in each grid area. There are many alternatives. A variety of methods exists (GORDON and HERMAN, 1974; BROOKS and DICHIRD, 1976) for obtaining the solution iteratively without having actually to find the singular vectors whose computation can be expensive, but it needs to be performed only once for any given configuration.

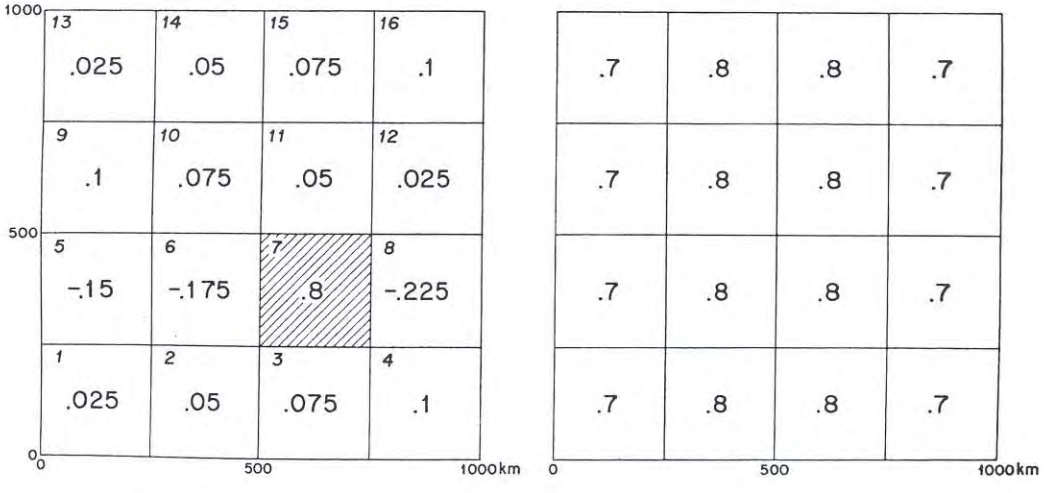

Fig. 9. Inversion bias for configuration A. Left: the 7th column of $\mathbf{V}\mathbf{V}^T$, giving the inferred values of δC_j for unity input in grid area $j_0 = 7$ and zero input elsewhere. Right: The diagonal element of $\mathbf{V}\mathbf{V}^T$, giving the inferred values of δC_{j_0} for unity input in grid area j_0 and zero input elsewhere. The amount by which values are short of unity is a measure of the failure to resolve a particular grid and of the spread of the solution into other grid areas.

The Fourier method of BRACEWELL and RIDDLE (1967) and its variants (SWEENEY and VEST, 1973) is equivalent, but avoids the necessity of defining the rectangular grid explicitly. There are fairly obvious versions of the technique which are analogous to super-resolution and maximum entropy-type methods. Alternatively, we can use some geophysically defined expansion functions, e.g. the low order representation of oceanographic dynamic modes (Rossby waves), and solve for their coefficients. These different methods have much to recommend them. The necessary changes in procedure are fairly straightforward and we will not explore them explicitly.

Noisy measurements

So far we have dealt with bias errors associated with a failure to resolve the assumed grid. We now discuss the statistical errors associated with uncertainty in the travel time measurements (mooring motion, clock error, pulse spread). These are two distinct sources of error.

The measured travel time is a sum of the true value plus noise:

$$\delta T_M = \delta T + \delta T_N.$$

For the present we assume uncorrelated noise, so that the expected product for two ray paths is given by

$$\langle (\delta T_N)_i (\delta T_N)_{i'} \rangle = \sigma^2 \delta_{ii'}.$$

Later we will consider correlated noise.

The system (18) of linear equations becomes now

$$\mathbf{E}\,\delta C + \delta T_N = \delta T_M. \tag{22}$$

This is a stochastic problem, and again there is a multiplicity of procedures.

If the system were of full rank, the best linear unbiased estimator (BLUE) is (SEBER, 1977)

$$\delta C = \mathbf{E}^T (\sigma^2 \mathbf{I})^{-1} [\mathbf{E} (\sigma^2 \mathbf{I})^{-1} \mathbf{E}^T]^{-1} \delta T_M = \mathbf{E}^T (\mathbf{E}\mathbf{E}^T)^{-1} \delta T_M, \tag{23}$$

which is equivalent to equation (19). $\mathbf{I}$ is the unit matrix. In the most useful case, however, the matrix $\mathbf{E}$ and hence $\mathbf{E}\,\mathbf{E}^T$ is singular and the inversion cannot be performed.

But the considerations that lead to (23) suggest that (19) is still the BLUE for δC for singular $\mathbf{E}$. Unfortunately, the variance of the result is very large; the presence of noise means that small errors in $(U_l \cdot \delta T)$ are amplified by the small singular values λ_l, and this makes estimates of δC highly uncertain.

The remedy is simple: allow some bias in the solution to reduce the variance. The following are equivalent minimum variance biased estimates of δC:

$$\delta C = \mathbf{E}^T (\mathbf{E}\mathbf{E}^T + \tilde{\sigma}^2 \mathbf{I})^{-1} \delta T_M \tag{24a}$$

$$= \sum_{l=1}^{L} \frac{\lambda_l (U_l \cdot \delta T_M)}{\lambda_l^2 + \tilde{\sigma}^2} V_l, \tag{24b}$$

where $\tilde{\sigma} = \sigma / C^0$. More generally (WIGGINS, 1972; LIELBELT, 1967) let

$$\mathbf{A} = \{\langle \delta C_j \delta C_{j'} \rangle\}, \quad \mathbf{B} = \{\langle (\delta T_N)_i (\delta T_N)_{i'} \rangle\} \tag{25}$$

be the covariance matrices of the solution and of the error, respectively. Then the minimum variance estimator is

$$\delta C = \mathbf{A}\,\mathbf{E}^T (\mathbf{E}\,\mathbf{A}\,\mathbf{E}^T + \mathbf{B})^{-1} \delta T_M. \tag{26}$$

If $\mathbf{A} \to \infty$ (no information about solution variance) then it can be shown that (26) approaches (23). Examination of (24b) shows that the singular vectors V_l corresponding to the smaller singular values λ_l do not contribute to the solution if $\lambda_l^2 \ll \tilde{\sigma}^2$. Thus failure of configuration A to be of full rank is unimportant if a configuration that is formally of higher rank contains singular values below the actual noise level. Figure 7 represents the inversion of configuration A with random noise added to the computed travel times prior to inversion. The inversion is not appreciably degraded for $\sigma^2 = 0.01\,\mathrm{s}^2$ (100 ms error). Our design requirement for a travel time precision of 25 ms is perhaps too demanding, particularly if one takes into account that the assumed δC magnitude in the numerical experiments is rather smaller than that expected for mesoscale fluctuations and that the assumed δC pattern is more irregular.

We can now compute the formal stastical error for the inversion. The mean-square

error in δC_j is given by

$$\sigma^2 \sum_{l=1}^{L} \frac{V_{il}}{\lambda_l^2}, \quad \sigma^2 \sum_{l=1}^{L} \frac{\lambda_l^2 V_{jl}^2}{(\sigma^2+\lambda_l^2)^2}, \tag{27a,b}$$

respectively, depending on whether we use the solution (19) or (24). The form (27a) makes explicit the increase in error as vectors corresponding to small values of λ_l are added—the conventional trade-off between high resolution (which requires many V_l) and statistical

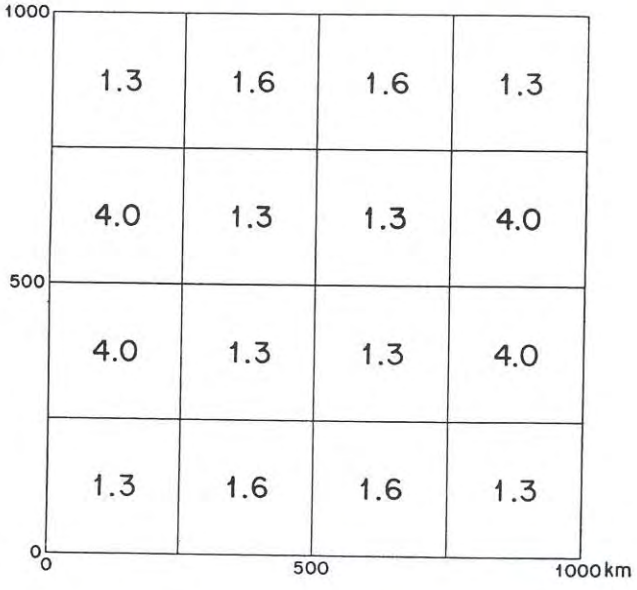

Fig. 10. Variance of the inferred δC_j in $(\mathrm{m\,s^{-1}})^2$ for configuration A, taking $\sigma^2 = 1\,\mathrm{s}^2$ for the mean square error in δT_i.

stability [which suggests truncating the sums in (19) and (27a)]. Figure 10 gives the mean-square error in δC according to equation (27a) for $1\,\mathrm{s}^2$ variance in δT.

We can anticipate that the errors in some of the travel times will be correlated for a variety of reasons. Movement of a source or receiver contributes a correlated error in all rays leading to it; a clock error correlates the errors in all rays from that source; many rays pass through neighboring water parcels and sense the same internal wave fluctuations. We can introduce a new coordinate system in which the noise correlation matrix is diagonalized and so reduce the system back to the diagonal case (see WIGGINS, 1972).

6. VERTICAL SLICE

We now consider the vertical plane (r, z) between any source and any receiver extending from the sea floor to the sea surface. Dimensions are typically 1000 km × 5 km. This plane is divided into vertical grid areas $k = 1, K$. There is a series of ray paths between source

and receiver as illustrated in Fig. 2. Travel time differences for about 10 of these should be resolvable (Section 3), and they are labeled $i = 1, I$ as previously.

Formally, the problem is the same as for the horizontal slice. In analogy with equation (18) we write

$$\mathbf{D}\,\delta C = \delta T, \quad \mathbf{D} = \{D_{ik}\}, \quad \delta C = \{\delta C_k\}, \quad \delta T = \{\delta T_i\}, \tag{28}$$

where

$$D_{ik} = -R_{ik}/(C^0)^2 \tag{29}$$

and R_{ik} is the path length of ray i in grid area k. (A good approximation can be obtained from the horizontal distance travelled in grid area k.)

In principle, we can treat horizontal and vertical planes together in a straightforward generalization to the three-dimensional problem. The ocean is then divided into volumes labeled $b = 1, B$; these are sampled by $r \cdot s \cdot p$ paths labeled $i = 1, I$. The formalism (28) is the same, provided we interpret $\mathbf{D} = \{D_{ib}\} = \{-R_{ib}/(C^0)^2\}$, with R_{ib} denoting the path length of ray i through volume b. For 10 resolvable paths in each vertical slice, a vertical division into three layers, and the horizontal division of configuration A (Fig. 6), we have 48 volumes, and there are 120 pieces of information in the 48 unknowns. But this overdeterminism depends upon the too optimistic assumption that 10 paths give 10 independent pieces of information. A more realistic estimate is that they will give three independent pieces of information for the equivalent of 36 equations in 48 unknowns, with about the same ratio of indeterminancy as for the horizontal slice alone.

To proceed with three-dimensional tomography in this fashion is an exercise in brute force we have not pursued. In some circumstances it may be the best method. AKI, CHRISTOFFERSON and HUSEBYE (1977) have applied it successfully to the equivalent seismic problem. It is undoubtedly the best way if the range R is only a few convergence zones ($R_{cz} \approx 50$ km) and if an isolated perturbation has horizontal scales of less than R_{cz}. The situation we have envisioned is the opposite of this case with a range of $20\,R_{cz}$ and disturbances typically of order 2 to $5\,R_{cz}$.

An examination of Fig. 2 will demonstrate that arrival times obtained in a single vertical slice contain virtually no information about the *range* of an isolated disturbance whose horizontal scale exceeds that of a convergence zone. A disturbance extending from 300 to 400 km (for example) affects the same rays as one extending from 600 to 700 km. On the other hand, steep rays sample the vertical slice from near-surface to near-bottom, whereas flat rays remain near the axis of the sound channel. Accordingly, there is considerable information about the vertical structure of the disturbance, and this is the basis of applying inverse techniques to the vertical slice.

To make this argument quantitative, we refer to Fig. 5, which gives the relative time R_{ik}/C^0 spent by each ray group i in various depth layers k of a canonical ocean. [Weighting by relative travel time and relative range agree to order $\gamma \approx 10^{-2}$, equation (A13).] For demonstration we have selected the North Atlantic ray paths shown in Fig. 2, and measured partial ranges R_{ik} of four selected rays in seven layers (Table 3). Values of R_{ik} are larger for the deep layer than the shallow layers, as a result of the asymmetry of the sound channel.

Table 4 gives the result of the inversion experiment. In the first column, values of δC_k are chosen to be roughly consistent with a mesoscale disturbance in the gravest dynamical

Table 3. Path lengths R_{ik} (km) *of rays* i *in layers* k. *The four ray numbers and seven layer numbers are identified in Fig. 2.*

		Rays			
		1	6	7	28
	Inclination at source	-12.8°	-11.3°	-9.4°	+2.3°
	Ray identifier	-14,14	-15,15	-17,17	+27,27
	Travel time T	664.65^s	665.51^s	668.25^s	669.58^s
Layers					
1	(0-400m)	154	120	0	0
2	(400-800m)	64	90	128	0
3	(800-1200m)	50	52	162	271
4	(1200-2000m)	85	112	68	606
5	(2000-3000m)	141	180	264	123
6	(3000-4000m)	211	314	332	0
7	(4000-5000m)	294	136	0	0

Table 4. Inverse problem for vertical slice. The seven layers are the same as in Table 3 and Fig. 2. FMVW designates 'First Mode Variance Weighted'.

	First Mode δC (m/s)			Second Mode δC (m/s)		Variance $(m/s)^2$
Layers	Assumed	"Naive" Inversion	FMVW Inversion	Assumed	FMVW Inversion	for σ^2 = 1 sec^2
1 (0-400m)	12.2	0.2	12.3	-8.7	-8.6	5
2 (400-800m)	15.0	3.6	15.6	-2.5	-2.0	13
3 (800-1200m)	8.4	5.5	8.3	2.0	1.9	80
4 (1200-2000m)	2.7	3.0	2.8	1.4	1.5	155
5 (2000-3000m)	0.4	5.4	0	0.2	0	0.4
6 (3000-4000m)	0	3.6	0	0	0	0.07
7 (4000-5000m)	0	4.3	0	0	0	0.00

mode, which has been observed to be dominant (RICHMAN *et al.*, 1977). We then solve the 'forward' problem of computing $\delta T_i = -R_{ik}\,\delta C_k/(C^0)^2$. Given δT_i, a straightforward inversion (called 'naive') does poorly, owing to the tendency of rays to dwell in the deeper layers; accordingly the inversion of four equations in seven unknowns by equation (19) automatically gives great weight to the deep layers.

If we knew nothing about the oceans, there would be no reason to reject this solution. But we know a great deal. To express our expectation that the mesoscale perturbations are largest in the upper layers, we modify the procedure by introducing the covariance matrix $\mathbf{A}_1$ of the first mode into the inversion scheme [equation (26) with $\mathbf{B} = 0$]. The results of this weighted inversion are nearly perfect.

A corresponding inversion for the second dynamic mode weighted (incorrectly) for the

covariance $\mathbf{A}_1$ of the first mode still gives excellent results. This demonstrates that the results are not sensitive to the imposed solution variance.

The last column in Table 4 gives the variance of the inferred δC_k for either the first or second mode ($\mathbf{A}_1$ weighting) owing to uncorrelated errors in δT_i with a mean square of $\sigma^2 = 1\,\mathrm{s}^2$. If we can achieve an r.m.s. time precision of 25 ms, as previously discussed, then $\sigma^2 = 6 \times 10^{-4}\,\mathrm{s}^2$, and the uncertainty in layer four (where it is largest) is still only $0.3\,\mathrm{m\,s}^{-1}$.

As we have given up on any horizontal range information from the vertical slice, we depend entirely upon the horizontal slice for horizontal localization. With δC determined in each layer for each vertical slice of the three-dimensional problem, we can then solve the horizontal problem for each layer separately. Suppose we have determined δC_k in layer $k = 1, K$. Then

$$\delta T_k = -[R/(C^0)^2]\,\delta C_k \tag{30}$$

is the equivalent perturbation in travel time along a horizontal straight ray path between a source and receiver in layer k. We determine $(\delta T_k)_i$ for each pair of source and receiver, and then proceed with the horizontal inversion problem as before. The horizontal slice problem is canonical, and the three-dimensional problem is solved as a successsion of K two-dimensional inversions.

7. OCEAN ACOUSTIC TOMOGRAPHY

An acoustic tomographic system appears to be both practical and useful. The limiting factor will probably always be the lack of spatial resolution owing to a limited number of sources and receivers, rather than to a lack of temporal resolution and precision.

Ordinarily there will be insufficient information to determine a unique solution; the presence of the null space vectors satisfying $\mathbf{E}V = 0$ means that there are an infinite number of other solutions that cannot be distinguished from the one chosen except insofar as they require additional structure (usually of high wavenumbers) not imposed by the measurements (Section 5). In principle, one can always obtain a unique answer by reducing the resolution required to the point where the system becomes fully determined. The singular value decomposition (19) explicitly computes those linear combinations of δC that are fully determined.

Resolution can be improved without adding to the number of acoustic sources and receivers. We can use any additional information at our disposal to reduce the null space and hence increase the number of non-zero singular values. Alternatively, we may attempt to reduce the solution variance in key locations. The inverse formalism makes it particularly convenient to accommodate additional constraints of a variety of types.

Direct ocean observations

One advantage of the oceanographic problem over the medical or seismological problem is that the interior of our physical space is accessible. (In the long run this may turn out to have been a liability, for it has delayed indirect methods which, in our prejudiced view, are made inevitable by the magnitude of the ocean monitoring task.)

Ships of opportunity with expendable bathythermographs (XBTs), planned expeditions, and moored buoys with conventional hydrographic instruments can provide *in*

situ measurements of δC_j. These can be trivially added to the equation set (18) in the form

$$\delta C_j = Q_j, \tag{31}$$

where Q_j is the measured value.

Direct observations can be aimed towards (i) improving the resolution and (ii) reducing the statistical uncertainty, particularly in important regions where the tomographic coverage is inadequate. One's confidence in the direct measurements relative to the acoustic observations can be expressed by augmenting the noise covariance matrix **B**. A study of the resolution of the acoustic measurements provides a quantitative measure of where such direct observations would be most useful and leads to the design of an objective observation strategy.

A priori *knowledge*

Previous experience can be useful in selecting a solution to the inverse problem. For example, if the Gulf Stream, or a Gulf Stream ring, is present in the grid, then it is likely that major changes in sound speed will tend to occur near these regions of intense motion rather than in more quiescent parts of the ocean. Satellite observations can help identify active regions. Such expectations are imposed upon the inversion by proper construction of the solution covariance matrix **A**. One simply makes **A** large in those areas where large changes are to be expected and small eslewhere. An attempt to force a major change in a particular grid area, if inconsistent with the measured travel times, distributes the resulting solution through many areas and leads to a solution covariance that is qualitatively inconsistent with the assumed covariance. More subtle deviations can be studied with the formal statistical methods of hypothesis testing.

In the absence of a clear idea of what form the solution should take we may have to explore a range of possibilities relative to different initializing assumptions. PARKER (1972) has discussed the problem of finding the envelope of all solutions acceptable under some set of bounds.

Constraints

In many cases we will have considerable insight into the expected dynamical behavior of the ocean. Various levels of sophistication are possible and useful.

If the motions are wavelike, then a requirement of conservation of (potential) sound speed

$$\sum_j \delta C_j = 0 \tag{32}$$

can be added to the constraint equations (18). Further, for any assumed $\theta(S)$ relations, density can be computed [equation (1)] and the conservation of mass over the entire volume, or separately in each horizontal slice, can be imposed as an additional constraint.

Mesoscale eddies are usually observed to move westward; Gulf Stream meanders tend to move eastward. Such kinematical expectations can be used to constrain the inversion as the system evolves in time. The simplest method is to use the **A** matrix for giving additional weight to those grid areas that lie west of observed mesoscale eddies and east of meanders. A more sophisticated approach is to introduce time as a new dimension and construct temporal as well as spatial constraints. A detailed discussion would take us too far afield; the formalism is that of optimal estimation [see for example LIEBELT's (1967,

Chapter 5) use of the Kalman filter and other *a priori* forms of optimal filters]. These same formalisms make it possible to use the subsequent evolution of the system for retrospectively refining previous inversions.

Future developments

The discussion thus far has been in terms of a basic tomographic system, using only acoustic travel time changes along a few resolvable multipaths.. There are a number of ways in which the tomography could be strengthened, at the price of some increase in complexity. Instead of one-way transmissions from sources to receivers, we can make reciprocal transmissions, and use the *differences* in opposite travel times to compute the mean current components along the path (WORCESTER, 1977). [This requires a high precision [equation (2)], 1 ms rather than 25 ms, but we believe such a precision can be obtained.] A straightforward extension of the basic tomography can determine the vector current field in three dimensions. Moreover, the density and velocity fields so derived are nearly in geostrophic balance, and so the estimate of each field can be improved by information about the other field.

But even without direct measurements of velocity, the geostrophic relationship can be used to deduce the velocity field by the conventional dynamic method, given the field in sound speed and adequate information about the $\theta(S)$ relations. As always, one would be missing a constant of integration (the reference velocity). But the constraints imposed by the conservation of water mass and of other properties, together with kinematical and dynamical conditions, can be used to deduce the reference velocity by yet another application of inverse theory. This is the problem of 'geostrophic inversion' discussed by WUNSCH (1978). Thus, the inversion of the acoustic system could provide the basic data for a second inversion to determine the reference velocity. Direct measurements of velocity by current meters or floats could be used to refine both the acoustic inversion and the geostrophic inversion.

We contemplate an eventual tomographic system for areas as large as the entire North Atlantic, using ultra low-frequency sources. The basic data would probably be the perturbations in phase and group velocities of acoustic normal modes, rather than the perturbations in travel time along ray paths. With sufficient resolution one might generate acoustic holograms of the ocean—perhaps the ultimate development of the tomographic system.

Conclusions

Our fundamental conclusion is that there are no barriers, theoretical or practical, to large scale acoustic monitoring of the ocean. At the present time we need field data to confirm in practice that we can identify and keep track of the different multipath arrivals and that our noise estimates are correct.

Tomographic systems may also be useful in regions of high currents (like the Gulf Stream), where conventional mooring techniques are unequal to the environment.

We are intrigued with the almost unlimited number of possibilities for future development.

Acknowledgements—We express our thanks to the Office of Naval Research for supporting this feasibility study, and the work leading up to it. For one of us (W.M.), the incentive came from an ongoing experimental program with ever-increasing ranges and decreasing frequencies, led by ROBERT SPINDEL, GORDON WILLIAMS, and PETER

WORCESTER. For the other (C.W.), the study is an outgrowth of applying inverse methods to standard oceanographic measurements. A preliminary report was prepared by the authors in collaboration with RICHARD L. GARWIN during the 1977 JASON Summer Study (JSS-77-8), supported by the Advanced Research Projects Agency.

REFERENCES

AKI K., A. CHRISTOFFERSSON and E. S. HUSEBYE (1977) Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, **82,** 277–296.

BACKUS G. E. and J. F. GILBERT (1967) Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society*, **13,** 247–276.

BACKUS G. and J. F. GILBERT (1968) The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, **16,** 169–205.

BERNSTEIN R. L., L. BREAKER and R. WHRITNER (1977) California Current eddy formation: ship, air, and satellite results. *Science*, **195,** 353–359.

BIRDSTALL T. G. (1976) On understanding the matched filter in the frequency domain. *IEEE Transactions on Education*, **19,** 168–169.

BRACEWELL R. N. and A. C. RIDDLE (1967) Inversion of fan-beam scans in radio astronomy. *Astrophysical Journal*, **150,** 427–434.

BROOKS R. A. and G. DICHIRO (1976) Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. *Physics in Medicine and Biology*, **21,** 689–732.

CHHABRA N. K. (1977) Correction of vector-averaging current meter records from the MODE-1 central mooring for the effects of low-frequency mooring line motion. *Deep-Sea Research*, **24,** 279–288.

CLARK J. G. and M. KRONENGOLD (1974) Long-period fluctuations of CW signals in deep and shallow water. *Journal of the Acoustical Society of America*, **56,** 1071–1083.

COLIN L., editor (1972) *Mathematics of profile inversion.* NASA Technical Memorandum TM X-62, **150,** 630 pp.

DYSON F., W. MUNK and B. ZETLER (1976) Interpretation of multipath scintillations Eleuthera to Bermuda in terms of internal waves and tides. *Journal of the Acoustical Society of America*, **59,** 1121–1133.

FLATTÉ S. M. editor, R. DASHEN, W. H. MUNK, K. M. WATSON and F. ZACHARIASEN (1978) *Sound transmission through a fluctuating ocean.* Cambridge University Press, in press.

FLATTÉ S. M. and F. D. TAPPERT (1975) Calculation of the effect of internal waves on oceanic sound transmission. *Journal of the Acoustical Society of America*, **58,** 1151–1159.

GARRETT C. and W. MUNK (1975) Space–time scales of internal waves: a progress report. *Journal of Geophysical Research*, **30,** 291–297.

GELB A., editor (1974) *Applied optimal estimation.* M.I.T. Press, 374 pp.

GORDON R. and G. T. HERMAN (1974) Three-dimensional reconstruction from projections: a review of algorithms. *International Review of Cytology*, **38,** 111–151.

HAMILTON G. R. (1977) Time variations of sound speed over long paths in the ocean. In: *International workshop on low-frequency propagation and noise*, Woods Hole, Massachusetts, Oct. 14–19, 1974, pp. 7–30.

HELSTROM C. W. (1968) *Statistical theory of signal detection*, second edition. Pergamon Press, 364 pp.

JOBST W. and L. DOMINIJANNI (1978) Measurements of the temporal, spatial and frequency stability of the underwater acoustic channel. *Journal of the Acoustical Society of America*, in press.

JOHNSON R. H. and R. A. NORRIS (1968) Geographic variation of SOFAR speed and axis depth in the Pacific Ocean. *Journal of Geophysical Research*, **73,** 4695–4700.

LANCZOS C. (1961) *Linear differential operators.* Van Nostrand, 564 pp.

LIEBELT, P. B. (1967) *An introduction to optimal estimation*, Addison Wesley, 273 pp.

MODE Group (1978) Mid-ocean dynamics experiment. *Deep-Sea Research*, **25,** 859–910.

MUNK W. H. (1974) Sound channel in an exponentially stratified ocean, with application to SOFAR. *Journal of the Acoustical Society of America*, **55,** 220–226.

MUNK W. H. and G. O. WILLIAMS (1977) Acoustic oceanography. *Nature*, **267,** 774–778.

MUNK W. H. and P. F. WORCESTER (1976) Monitoring the ocean acoustically. In: *Science, technology, and the modern navy—thirtieth anniversary* 1946–1976, (ONR-37), Office of Naval Research, Arlington, VA, pp. 497–508; also appears as: Weather and climate under the sea—the Navy's habitat. In: *Science and the future navy—a symposium*, 30th Anniversary Volume, Office of Naval Research, National Academy of Sciences, Washington, D.C., pp. 42–52.

NYSEN P. A., P. SCULLY-POWER and D. G. BROWNING (1978) Sound propagation through an East Australian Current eddy. *Journal of the Acoustical Society of America*, in press.

PARKER R. L. (1972) Inverse theory with grossly inadequate data. *Geophysical Journal of the Royal Astronomical Society*, **29,** 123–138.

PARKER R. (1977) Understanding inverse theory. *Annual Review of Earth and Planetary Sciences*, **5,** 35–64.

PEDERSEN M. A. and D. WHITE (1970) Ray theory for sources and receivers on an axis of minimum velocity. *Journal of the Acoustical Society of America*, **48,** 1219–1248.

PORTER R. P., R. C. SPINDEL and R. J. JAFFEE (1973) CW beacon system for hydrophone motion determination. *Journal of the Acoustical Society of America*, **53**, 1691–1699.
RADON J. (1917) Über die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten, Ber Saechs. *Akademie der Wissenschaften, Leipzig, Mathematisch-Physikalische Klasse*, **69**, 262–277.
RICHMAN J. G., C. WUNSCH and N. G. HOGG (1977) Space and time scales of mesoscale motion in the western North Atlantic. *Reviews of Geophysics and Space Physics*, **15**, 385–420.
SCULLY-POWER P. D., P. A. NYSEN, C. S. NILSSON, P. F. TWITCHELL, D. G. BROWNING, R. C. SWENSON, J. C. ANDREWS and R. W. BANNISTER (1975) A multisystem technique for the detection and measurement of warm core ocean eddies. *IEEE Oceans* **75**, 761–768.
SEBER G. A. (1977) *Linear regression analysis*. John Wiley, 465 pp.
STEINBERG J. C. and T. G. BIRDSALL (1966) Underwater sound propagation in the Straits of Florida. *Journal of the Acoustical Society of America*, **39**, 301–315.
SWEENEY D. W. and C. M. VEST (1973) Reconstruction of three-dimensional refractive index fields from multidirectional interferometric data. *Applied Optics*, **12**, 2649–2664.
SWINDELL W. and H. H. BARRETT (1977) Computerized tomography: taking sectional X-rays. *Physics Today*, 32–41.
WEINBERG N. L. and X. ZABALGOGEAZCOA (1977) Coherent ray propagation through a Gulf Stream ring. *Journal of the Acoustical Society of America*, **62**, 888–894.
WIGGINS R. A. (1972) The general linear inverse problem: implication of surface waves and free oscillations for earth structure. *Reviews of Geophysics and Space Physics*, **10**, 251–285.
WORCESTER P. F. (1977) Reciprocal acoustic transmission in a mid-ocean environment. *Journal of the Acoustical Society of America*, **62**, 895–905.
WUNSCH C. (1978) The general circulation of the North Atlantic west of 50°W determined from inverse methods. *Reviews of Geophysics and Space Physics*, **16**, 583–680.

APPENDIX A—RAY ARRIVAL STRUCTURE IN A CANONICAL SOUND CHANNEL

The canonical sound speed profile* is

$$\phi^2(C) \equiv \frac{C-\bar{C}}{\varepsilon \bar{C}} = e^{\eta} - \eta - 1 = \frac{1}{2}\eta^2 + \frac{1}{6}\eta^3 + \ldots \tag{A1}$$

where $\eta = (z-\bar{z})/\frac{1}{2}B$ is a dimensionless elevation above the sound axis, with $B(\approx 1\,\text{km})$ the ocean stratification scale; $2\varepsilon = B\gamma_A$, where $\gamma_A = -1.14\times 10^{-2}\,\text{km}^{-1}$ is the fractional sound speed gradient in an adiabatic ocean. For near-axial rays, $C\to\bar{C}$ and $\phi\to 0$. At the upper and lower turning points $\hat{\eta}^{\pm}$, we write $\hat{C} = C(\hat{\eta}^{\pm})$ and $\hat{\phi} = \phi(\hat{C})$. For steep rays with turning points near the surface and bottom, $\hat{\phi} \approx 1.75$.

The range and travel time are given by (Fig. A1)

$$p^+R^+ + p^-R^- - \text{sgn}(\eta_s\theta_s)\Delta R_s + \text{sgn}(\eta_r\theta_r)\Delta R_r = R, \tag{A2}$$

$$p^+T^+ + p^-T^- - \text{sgn}(\eta_s\theta_s)\Delta T_s + \text{sgn}(\eta_r\theta_r)\Delta T_r = T. \tag{A3}$$

The first two terms correspond to integer (axis to axis) loops associated with p^+ upper and p^- lower turning points, respectively. The last two terms are the contributions from fractional loops when the source and receiver are off axis. For the integer loops (CSC 27)

$$R^{\pm} = R_0\left[1 - \text{sgn}\,\eta\,\frac{2\sqrt{2}}{3\pi}\hat{\phi} + \frac{1}{12}\hat{\phi}^2 + \ldots\right] \tag{A4}$$

$$\tau^{\pm} = \tau_0\left[\text{sgn}\,\eta\,\frac{2\sqrt{2}}{9\pi}\hat{\phi}^3 - \frac{1}{24}\hat{\phi}^4 + \ldots\right]. \tag{A5}$$

* We follow the notation of MUNK's (1974) paper on the canonical sound channel (to be referred to as CSC) except that the vertical coordinate is now taken as positive upwards.

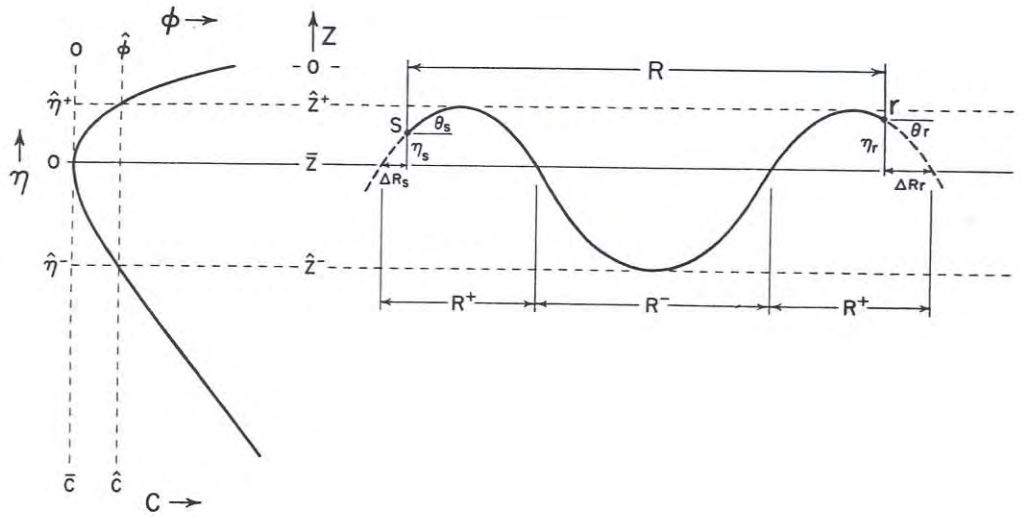

Fig. A1. Definition plot. Elevation z is positive upwards from sea surface, dimensionless elevation η is positive upwards from sound axis. The range R between source s and receiver r consists of a complete lower loop, R^-, and two partial upper loops: $R^+ - \Delta R_s$ near the source and $R^+ - \Delta R_r$ near the receiver; ΔR_s and ΔR_r are the distances (taken positive) of source and receiver from the nearest axial crossing. In the figure, source and receiver are both above the sound axis (η positive), ray inclination θ is positive at the source and negative at the receiver: $\operatorname{sgn}(\eta_s\theta_s) = +1$, $\operatorname{sgn}(\eta_r\theta_r) = -1$.

$\tau = T - R/\bar{C}$ is the time reckoned from the axial arrival, and so normally negative. Here, with $2\varepsilon = B\gamma_A$,

$$R_{cz} \equiv 2R_0 = \pi\varepsilon^{-1/2}B = \pi(2B/\gamma_A)^{1/2} \approx 50\,\text{km} \tag{A6}$$

is the range of the convergence zone, $B = 1.44\,\text{km}$, and $\varepsilon = 0.0082$. A convenient time scale is

$$\tau_0 = \frac{1}{2}\pi B\varepsilon^{1/2}/\bar{C} = \frac{1}{2}\pi^2 B^2/(\bar{C}R_{cz}) = \frac{\gamma_A^2}{8\pi^2}\frac{R_{cz}^3}{\bar{C}} \approx 0.137\,\text{s}. \tag{A7}$$

Equations (A4) and (A5) state that upper loops ($\operatorname{sgn}\eta = +1$) have shorter ranges and larger travel times than lower loops; further, that steep double loops (large $\hat{\phi}$) have larger ranges and shorter travel times than flat double loops.

Axial convergence

For an even number of integer loops, e.g. $p = p^+ = p^-$, $p(R^+ + R^-) = R$, $p(\tau^+ + \tau^-) = \tau$. Eliminating p and using (A4) and (A5) gives

$$\hat{\phi}^2 = \frac{1}{2}\omega + \frac{1}{2}\sqrt{(\omega^2 + 48\omega)} \qquad \omega = \frac{-\tau/\tau_0}{R/R_{cz}}, \tag{A8}$$

which, together with (A1), gives the convergence, with increasing arrival time, of turning points towards the sound axis (Fig. 4).

Fractional loops

We need the expressions corresponding to (A4) and (A5) for partial loops. Equation (A1) can be inverted from $\phi(\eta)$ to $\eta(\phi)$, and this yields a function (CSC17-19)

$$\Phi(\phi) = -\tfrac{1}{2}\sqrt{2}\,\mathrm{sgn}\,\eta\,\phi^{-1} + \tfrac{1}{3} - [\sqrt{(2)}/12]\,\mathrm{sgn}\,\eta\,\phi + \ldots$$

Writing

$$\sin\mu = \phi/\hat{\phi} \qquad 0 \leqslant \mu \leqslant \tfrac{1}{2}\pi \tag{A9}$$

the result is

$$\frac{\Delta R^{\pm}}{R_0} = -\sqrt{\left(\frac{2}{\pi}\right)}\,\mathrm{sgn}\,\eta\,\hat{\phi}\int_0^{\mu}\Phi(\mu)\sin\mu\,\mathrm{d}\mu = a + b\,\mathrm{sgn}\,\eta\,\hat{\phi} + c\hat{\phi}^2 \tag{A10}$$

$$\frac{\Delta\tau^{\pm}}{\tau_0} = -\sqrt{\left(\frac{2}{\pi}\right)}\,\mathrm{sgn}\,\eta\,\hat{\phi}^3\int_0^{\mu}\Phi(\mu)(1-2\sin^2\mu)\sin\mu\mathrm{d}\mu = d\hat{\phi}^2 + e\,\mathrm{sgn}\,\eta\,\hat{\phi}^3 + f\hat{\phi}^4, \tag{A11}$$

with

$$\begin{aligned} a &= \mu/\pi \\ b &= [\sqrt{(2)}/3\pi](\cos\mu - 1) \\ c &= (1/12\pi)(\mu - \cos\mu\sin\mu) \\ d &= \tfrac{1}{2}\cos\mu\sin\mu \\ e &= [\sqrt{(2)}/18](1 - 3\cos\mu + 2\cos^3\mu) \\ f &= (1/48)(-\mu + \cos\mu\sin\mu + 2\cos\mu\sin^3\mu). \end{aligned} \tag{A12}$$

It can be verified that for $\mu = \pi/2$, the expressions for the fractional loops equal half the integer loops: $\Delta R^{\pm} = \frac{1}{2}R^{\pm}$, $\Delta\tau^{\pm} = \frac{1}{2}\tau^{\pm}$.

Ray weighting

In any one loop, the time spent between the axis and some distance η is $2(\Delta T^{\pm} + \Delta R^{\pm}/\bar{C} + \Delta\tau^{\pm})$. (The factor 2 allows for upwards and downwards crossings.) The total time in a double loop is $T^{+} + T^{-} = R^{+}/\bar{C} + \tau^{+} + R^{-}/C + \tau^{-}$.

The fractional time between 0 and η is accordingly (Fig. 5)

$$\rho^{\pm}(\phi, \hat{\phi}) = \frac{(a + b\,\mathrm{sgn}\,\eta\,\hat{\phi} + c\hat{\phi}^2) + \gamma(d\hat{\phi}^2 + e\,\mathrm{sgn}\,\eta\,\hat{\phi}^3 + f\hat{\phi}^4)}{1 + (1/12)\hat{\phi}^2 - \gamma(\pi/48)\hat{\phi}^4} \tag{A13}$$

with

$$\gamma = 2\tau_0\bar{C}/\pi R_0 = 2\pi(B/R_{cz})^2 \approx 0.005.$$

The fractional time within a layer bounded by z_1 and z_2 is $\rho(\phi_2, \hat{\phi}) - \rho(\phi_1, \hat{\phi})$. For the case that z_1 is at the lower turning point and z_2 is at the upper turning point, we obtain $\mu = \frac{1}{2}\pi$, $a = \frac{1}{2}$, $c = 1/24$, $d = 0$, $f = -\pi/96$, and the result is

$$\rho^{+}(\hat{\phi}, \hat{\phi}) + \rho^{-}(\hat{\phi}, \hat{\phi}) = 1.$$

For small μ,

$$\rho = \frac{\mu/\pi}{1 - \gamma(\pi/48)\hat{\phi}^4}.$$

Because γ is small, equation (A13) reduces to $\rho^{\pm} \approx \Delta R^{\pm}/(R^{\pm}+R^{-})$, so that the weighting is nearly given by the fractional ranges.

Arrival structure

Equations (A2) and (A3) can now be written

$$A+B\hat{\phi}+C\hat{\phi}^2 = R/R_0 \tag{A14}$$

$$D+E\hat{\phi}+F\hat{\phi}^2 = \tfrac{1}{2}\pi\tau/(\tau_0\hat{\phi}^2), \tag{A15}$$

with

$$\begin{aligned}
A &= -\operatorname{sgn}\eta_s \operatorname{sgn}\theta_s a_s + \operatorname{sgn}\eta_r \operatorname{sgn}\theta_r a_r + p^+ + p^- \\
B &= -\operatorname{sgn}\theta_s b_s + \operatorname{sgn}\theta_r b_r - [2\sqrt{(2)}/3\pi](p^+ - p^-) \\
C &= -\operatorname{sgn}\eta_s \operatorname{sgn}\theta_s c_s + \operatorname{sgn}\eta_r \operatorname{sgn}\theta_r c_r + (1/12)(p^+ + p^-) \\
D &= -\operatorname{sgn}\eta_s \operatorname{sgn}\theta_s d_s + \operatorname{sgn}\eta_s \operatorname{sgn}\theta_r d_r \\
E &= -\operatorname{sgn}\theta_s e_s + \operatorname{sgn}\theta_r e_r + [\sqrt{(2)}/9](p^+ - p^-) \\
F &= -\operatorname{sgn}\eta_s \operatorname{sgn}\theta_s f_s + \operatorname{sgn}\eta_r \operatorname{sgn}\theta_r f_r - (\pi/48)(p^+ + p^-).
\end{aligned} \tag{A16}$$

The procedure is to solve (A14) for the turning parameter $\hat{\phi}$; the arrival structure is then obtained from (A15) for any specified set of source and receiver parameters (Fig. 6).

We note a few simple cases:

Number of groups. Set $p = p^+ = p^-$. Then

$$A = 2p,\ B = 0,\ C = p/6,\ D = 0,\ E = 0,\ F = -(\pi/24)p.$$

Hence

$$R/R_{cz} = \tfrac{1}{2}(A+C\hat{\phi}^2) = p(1+\tfrac{1}{12}\hat{\phi}^2) \tag{A17}$$

$$\tau/\tau_0 = 2\pi^{-1}F\hat{\phi}^4 = -\tfrac{1}{12}p\hat{\phi}^4. \tag{A18}$$

Group interval. Consider two neighboring groups p_1 and $p_2 = p_1+1$. The range equation gives $(p_1+1)\hat{\phi}_2^2 + p_1\hat{\phi}_1^2 = 12$, and it follows that

$$\frac{\tau_2/\tau_0}{\phi_2^2} - \frac{\tau_1/\tau_0}{\phi_1^2} = -\tfrac{1}{12}[(p_1+1)\hat{\phi}_2^2 - p_1\hat{\phi}_1^2] = +1. \tag{A19}$$

In dimensional units, we write $\delta p/p = -2\delta\hat{\phi}/\hat{\phi}$ from (A15), and so

$$\delta\tau = -(\tfrac{1}{12})\tau_0\delta(p\hat{\phi}^4) \approx 2\tau_0\hat{\phi}^2 \tag{A20}$$

for $\delta p = 1$. A typical group interval is 0.25 s. The final group interval is found by setting $\tau_2 = 0$, $p_2 = R/R_{cz}$, thus giving $\hat{\phi}_1^2 = 12\,R_{cz}/(R-R_{cz})$, and

$$\tau_1 = \tau_0\hat{\phi}_1^2 = \frac{6\pi^2 B^2}{\bar{C}(R-R_{cz})} = 0.086\,\text{s} \tag{A21}$$

for the penultimate group arrival.

APPENDIX B—CLOCK ERROR

The power spectrum in the fractional fluctuation of crystal frequency can be fitted tolerably by a sum of flicker noise plus random walk noise:

$$S(f) = \alpha_1 f^{-1} + \alpha_2 f^{-2}. \tag{B1}$$

The spectrum in clock error $\varepsilon(t)$ is then $(2\pi f)^{-2} S(f)$, and so $\langle \varepsilon^2 \rangle$ apparently diverges because of the terms in f^{-3} and f^{-4}. But this is not the case if we allow for a proper operational procedure.

Suppose the clock is set at the start of the experiment, $t = 0$, and the crystal frequency $\dot{\varepsilon}$ is measured at $t = 0$ and $t = t_1$. The simplest model is that of a linear frequency drift, with a predicted frequency and clock error of

$$\hat{\dot{\varepsilon}}(\tau) = \dot{\varepsilon}_0 + (\dot{\varepsilon}_1 - \dot{\varepsilon}_0)\tau, \quad \tau = t/t_1.$$

$$\hat{\varepsilon}(\tau) = \int_0^t \dot{\varepsilon}\, \mathrm{d}t = t_1[\dot{\varepsilon}_0 \tau + \tfrac{1}{2}(\dot{\varepsilon}_1 - \dot{\varepsilon}_0)\tau^2] + \varepsilon_0.$$

(It is now immaterial whether the 'systematic' crystal drift has been properly removed.) We consider only the departure $\delta\varepsilon$ from the predicted clock drift. This can be written in the form of a convolution

$$\delta\varepsilon(\tau) = \varepsilon(\tau) - \hat{\varepsilon}(\tau) = \int_{-\infty}^{\infty} \varepsilon(\xi)\, F(\xi)\mathrm{d}\xi$$

with

$$F(\xi) = \delta(\xi - \tau) - \delta(\xi) - (\tau - \tfrac{1}{2}\tau^2)\dot{\delta}(\xi) - \tfrac{1}{2}\tau^2\, \dot{\delta}(\xi - 1).$$

The mean-square clock error (corrected for linear frequency drift) is

$$\langle (\delta\varepsilon)^2 \rangle = \int_0^{\infty} (2\pi f)^{-2} S(f) \tilde{F}\tilde{F}^*\, \mathrm{d}f, \tag{B2}$$

where

$$\tilde{F}(f) = \int_{-\infty}^{\infty} F(\xi)\, \mathrm{e}^{-i\phi\xi}\, \mathrm{d}\xi, \quad \phi = 2\pi f t_1$$

and $\tilde{F}^*$ is its complex conjugate. The result is

$$\begin{aligned} FF^* &= 2 - 2\cos\phi\tau - \tau(2-\tau)\phi\sin\phi\tau + \tau^3(1-\tfrac{1}{2}\tau)\phi^2\cos\phi \\ &\quad - \tau^2\phi\sin\phi + \tau^2\phi\sin\phi(1-\tau) + \tau^2(1-\tau+\tfrac{1}{2}\tau^2)\phi^2 \\ &= \frac{1}{6!}[45\tau^4 - 60\tau^5 + 16\tau^6]\phi^6 + \text{order}\,\phi^8. \end{aligned}$$

FF^* is the filter associated with the removal of linear frequency drift. There are no errors at the start, $\tilde{F}\tilde{F}^* = 0$ for $\tau = 0$. $FF^* \to f^6$ as $f \to 0$, and so the $\langle (\delta\varepsilon)^2 \rangle$ does not diverge for flicker noise f^{-3} and random walk noise f^{-4}.

For times exceeding α_1/α_2, the clock error from random walk noise dominates the error from flicker noise. This is the case of interest.

Individual integrals in equation (B2) diverge at the lower limit. The procedure is to replace $\int_0^\infty d\phi$ by $\int_{\phi_l}^\infty d\phi$. Any sensible solution should be independent of ϕ_l as $\phi_l \to 0$, and this turns out to be the case. The final answer is

$$\langle(\delta\varepsilon)^2\rangle = \pi^2 \alpha_2 t^3 \left(\frac{2}{3} - \frac{1}{2}\frac{t}{t_1}\right). \tag{B3}$$