The Diffusive Regime of Double-Diffusive Convection

D. E. Kelley

Department of Oceanography, Dalhousie University,
Halifax, NS, B3H 4J1, Canada

H. J. S. Fernando

Department of Mechanical & Aerospace Engineering, Arizona State University,
Tempe, AZ 85287, USA

A. E. Gargett

Ocean, Earth & Atmospheric Sciences, Old Dominion University,
Norfolk, VA 23529-0276, USA

J. Tanny

Holon Academic Institute of Technology
POB 305, Holon 58102, Israel.

E. Özsoy

Institute of Marine Sciences, Middle East Technical University,
P.O. Box 28, Erdemli, Icel 33731, Turkey

Abstract

The diffusive regime of double-diffusive convection is reviewed, with a particular focus on issues that are holding up the development of large-scale parameterizations. Some of these issues, such as interfacial transports and layer-interface interactions, may be studied in isolation. Laboratory work should help with these. However, we must also face more difficult matters that relate to oceanic phenomena that are not easily represented in the laboratory. These lie beneath some fundamental questions about how double-diffusive structures are formed in the ocean, and how they evolve in the competitive ocean environment.

Key words: Double Diffusion, Transport Processes, Mixing Processes, Exchange Coefficients, Trans-isopycnal Mixing, Discontinuity Layers

Email addresses:
1 Introduction

Of the two types of double-diffusive convection (DDC, henceforth) that may occur if large-scale vertical gradients of salinity $S$ and temperature $T$ have the same sign, the salt-finger (SF, henceforth) mode has been studied more extensively than the diffusive-layer (DL, henceforth) mode. There appear to be three broad reasons for this focus, not all of which are compelling.

First, there are historical reasons. DDC research began with the SF mode, in the guise of the “perpetual salt fountain” (Stommel et al., 1956). This may explain why the seminal and defining paper of Stern (1960) relegated the DL mode to a footnote, and why followup studies have dealt mainly with the SF mode.

Second, there are geographical reasons. SF is commonly found at locations that are easily accessed by research vessels, whereas DL is prevalent at high-latitude locations where the logistics of sampling are very demanding. Furthermore, the geographical expanse of SF-susceptible waters exceeds that of DL-susceptible waters. This in itself says little, however, since the high-latitude DL zones may be especially important to the global climate system. Besides, DL may be an important partner to SF in interleaving across the globe.

Third, there are reasons that stem from the difficulty of extrapolating laboratory measurements to the ocean. In the laboratory, the SF mode is more vigorous than the DL mode, e.g., causing larger buoyancy fluxes across interfaces with analogous $T$ and $S$ steps. However, this implies nothing about the relative strength of SF and DL fluxes in the ocean, since the $S$ and $T$ steps need not be analogous, even if the background gradients are analogous. (Indeed, relating $S$ and $T$ steps to background gradients is a key goal, discussed at some length below.) Another factor is the response of SF and DL to background shearing motions, which are difficult to set up in the laboratory but which are prevalent in the ocean. While SF transports are apparently inhibited by shear (Linden, 1974b; Kunze, 1994), it has been speculated that DL transports may not be (Padman, 1994). This may imply that the dominance of the SF mode in laboratory may not be relevant to the ocean.

Reasons such as these may explain the historical concentration on the SF mode, but they should not preclude future DL work. This case is underlined by the above-mentioned interest in the role of high-latitude oceanography in the climate system (e.g. Walsh and Crane, 1992). A key climate component is water mass formation in high-latitude seas, and it is possible that DL fluxes play a role in this process.

dan.kelley@dal.ca (D. E. Kelley),
fernando@asuvox.eas.asu.edu (H. J. S. Fernando),
gargett@ccpo.odu.edu (A. E. Gargett),
tanai@hait.ac.il (J. Tanny),
ozsoy@ims.metu.edu.tr (E. Özsoy).
Further evidence is provided by decades of observations of distinctive Arctic fine-structure signatures that are arguably of DL character. The Arctic signatures include both staircases (Neal et al., 1969) and intrusions. The role of DDC seems clear in the staircases, but for the intrusions there may be other agencies involved, at least at the initial stages of formation. These agencies might include differential mixing (Hebert, 1999; Merryfield, 2000a,b; Gargett, 2002) and baroclinic instability (May and Kelley, 2001a). Regardless of the formation mechanism, though, there are indications of the relevance of DDC to the observed phase of the interleaving, in the form of steppiness between interleaving water-masses (Perkin and Lewis, 1984). This suggests that the centimeter-scale DDC processes have links to remarkably large space and time scales, for the intrusions have been traced over hundreds of kilometers and they appear to be coherent even over decades (Perkin and Lewis, 1984; Carmack et al., 1997). This suggests a prime significance of DDC to the Arctic Ocean. Therefore, if the Arctic proves to be of significance to the global climate system, perhaps through effects on ice albedo, or on watermass interaction with nearby subpolar sites where deep convection occurs, then DDC may be relevant to the climate system.

With this potential importance in mind, we turn next to a discussion of DL signatures in the world ocean. After presenting an overview of susceptible ocean regions (section 2.1) and some thoughts on methods of detecting staircases (section 2.2), we offer an abbreviated list of some prominent DL examples that have been well sampled (section 2.3). Then, in the second half of the paper, we turn to matters of modelling DL. We start with the issue of parameterizing fluxes between a pair of DL layers in the laboratory (section 3.1) and then move on to the related, but more difficult, matter of parameterizing fluxes in oceanic DL staircases (section 3.2). This foundation is used for a cursory outline of some unresolved questions regarding oceanic staircase formation (section 3.3) and evolution (section 3.4).

Our focus is entirely on the DL case, and mostly on the staircase mode. Readers may note that aspects of the DL mode appear to be simpler than in the SF mode, but it is not clear whether this is because the DL mode is actually simpler or just because the DL has been so little studied that contradictory information has not been uncovered. We take up such issues at the end.

---

1 Readers with wider interests may consult other contributions to this volume, along with various reviews (Turner, 1974; Huppert and Turner, 1981; Turner, 1985; Schmitt, 1994; Fernando and Brandt, 1994), and themed collections (Brandt and Fernando, 1995), as well as recent GCM sensitivity studies exploring the role of DDC under a range of surface boundary conditions (Zhang et al., 1998; Merryfield et al., 1999; Zhang and Schmitt, 2000).
2 Observations

2.1 Susceptible regions

Laboratory studies indicate that DDC produces coherent structures, e.g., staircases and intrusions (together with mixed modes) that may be recognized from examination of the horizontal and vertical variations of $S$ and $T$. Unfortunately, the signatures have length scales (meters to tens of meters) that are below the resolution of ocean atlases. Therefore, we cannot construct easily maps of global staircase occurrence, and are left to estimate the world-wide prevalence of DDC by mapping susceptibility.

Building upon theoretical and laboratory studies, this may be done by comparing the patterns of $S$ and $T$ stratification at large scales. The DL mode of DDC is possible when $S$ and $T$ increase together with depth, and both theory and laboratory observations suggest that it is more intense when $S$ and $T$ nearly compensate in density terms. Given the thermal expansion rate $\alpha$ and the haline contraction rate $\beta$, then, DL is expected to be most intense when the density ratio

$$R_p = \frac{\beta \partial S/\partial z}{\alpha \partial T/\partial z}$$

is near the low end of the DL-susceptible range $1 < R_p < \infty$. Buoyancy fluxes across laboratory DL interfaces drop by an order of magnitude as $R_p$ is increased from 1.5 (a typical lower bound for the ocean) to 10 (see Figure 4 below), suggesting that the range $1 < R_p < 10$ provides a good criterion for ocean susceptibility to DL convection.

Figure 1 is a crude attempt to illustrate world-wide DL susceptibility according to this criterion. It is based on vertical gradients of $T$ and $S$ from global atlases (Levitus and Boyer, 1994; Levitus et al., 1994) for the ocean, and from other sources for the Black Sea. To prevent difficulties with gradient computation, waters above 50 m and below 2500 m depth were excluded from the analysis. It should be noted that sparse ocean sampling limits reliability in the map at high latitudes. Also, the horizontal averaging involved in the atlas generation may affect the vertical gradients used in Figure 1. For these reasons, this map is certainly not definitive. Despite such provisos, a pattern of enhanced susceptibility at high latitudes, clearly seen in numerous case studies, may be still be discerned on the diagram. There is also evidence of regions of susceptibility in the western portions of the northern subpolar gyres, as well as in an Atlantic region near South America.

Figure 1 indicates regions that are susceptible to DL at any depth in the water column. This is useful for a summary diagram such as this, since a map at any particular depth would miss some regions. For example, the northern polar regions...
Fig. 1. Regions that are susceptible to DL convection. The darker of the two gray tones indicates regions with the minimum density ratio in the water column lying between 1 and 3, with the lighter tone indicating the range 3 to 10. Areas outside this range, and areas in which the atlases report no values, are white on this diagram. All other things being equal, the darker tones would indicate greater susceptibility to DL convection. See text for data sources and processing details.

have $R_\rho < 3$ (i.e., highly unstable) through the depth range of 200 m to 300 m. The subpolar regions (e.g. the Labrador Sea and the Sea of Okhotsk) have similarly low $R_\rho$ values in a depth range that extends 100 m closer to the surface. The depth range is extended somewhat in the southern polar regions, covering 100 m to $\sim 400$ m. In addition to these high-latitude cases, there is also an DL-susceptible region to the east of Brazil. This is much deeper (1000 m to 1400 m) than in polar waters, presumably resulting more from juxtaposition of water masses (Bianchi et al., 2002) than from surface forcing. It also has higher density ratios, with $R_\rho \approx 5$. All other things being equal, these higher values might imply that DL is less likely to occur in the Brazil region than in other susceptible regions. However, double-diffusion may be disrupted by turbulence, and so $R_\rho$ may not be the sole determining factor. For example, the disruptive effect of baroclinic tides is likely to be stronger near Brazil than in the Arctic. Indeed, it is notable that DDC signatures are prominent in regions which have weak tidal stirring, and thus presumably weak tidal mixing. The Canada Basin of the Arctic Ocean provides a good example. Melling et al. (1984) have pointed out that the tides in this region are less than $< 0.01$ m/s. This may explain why DL signatures are prominent there, despite the relatively high density ratio of approximately 7.

Thus, while diagrams such as Figure 1 may provide a good indication of susceptibility, they should not be taken to indicate either the incidence of DDC or its importance in terms of fluxes. Rather, the purpose in constructing susceptibility maps is provide a basis for comparison with local observations of DDC signatures and fluxes. This might shed light on the interaction of DDC with the environ-
Fig. 2. Temperature profile at station 508 of the Eurasian Basin Experiment (Perkin and Lewis, 1984), sampled at quarter-meter resolution. Magnified in the inset is a region from 205 m to 220 m depth, which has several easily identified layers. The vertical and horizontal bars of the inset profile indicate the results of two staircase-detection methods discussed in the text.

2.2 Detection Methods

We lack an agreed-upon quantitative method of recognizing signatures of DDC structures in the ocean, in the sense that there is no parameter that is routinely calculated and compared from experiment to experiment. This stands in the way of global mapping exercises and the comparison of case studies.

For the construction of atlases as well as first-order descriptions of individual study regions, it would be useful to have a single measure of meter-scale “steppiness” for a whole pycnocline. Several possibilities for this come to mind. One is to examine the difference between observed profiles and artificial profiles constructed
by smoothing over large vertical scales (Galbraith and Kelley, 1996). Mixing an initially linearly stratified profile over a vertical distance $D$ increases the potential energy per unit mass by $D^2 N_0^2 / 12$, where $N_0$ is the initial buoyancy frequency. Therefore, computing potential energy anomalies gives information not just on the energy itself, but also on an integral estimate of layer thickness, perhaps estimating $N_0^2$ by smoothing the observed density profile. Distinguishing the regular steppiness of staircases from the irregular steppiness of turbulent mixing might require further steps, e.g. examination of the depth variation of the integral lengthscale.

For the identification of individual layers, e.g. for flux computation and for tracing staircase extents, other methods might be more useful. These could be based on qualitative signatures that have been used informally for decades, based on finescale signatures on $S$ and $T$ profiles and in $TS$ diagrams (Tait and Howe, 1968; Neal et al., 1969; Fedorov, 1970). Steppy profiles are often taken as evidence of thermohaline staircases of DDC origin, but some provisos are required.

First, the vertical $S$ and $T$ gradients should vary in intensity, but not in sign, over depth. Otherwise, interleaving could have caused the finestructure. Second, the $S$ and $T$ gradients should be of the DDC variety, since otherwise internal waves (Lazier, 1973; Lazier and Sandstrom, 1978) or variations in mixing rate (Phillips, 1972) could have caused the finestructure. Third, the patterns should be traceable laterally and temporally; $TS$ diagrams may be especially useful in this regard since tracing features in profiles can be difficult unless sampling is very rapid (Padman and Dillon, 1988). Furthermore, $TS$ diagrams are effective in distinguishing staircases from interleaving. This is important since sometimes the difference between signatures of staircases and interleaving is slight, if only profiles are examined: weak reversed gradients are difficult to distinguish from weak unreversed gradients. Since interleaving scales tend to be much larger than layer scales in staircases, misinterpreting intrusions for layers has the potential to confound the development of scaling laws for layer thickness and thus for large-scale diffusivities (see section 3.2 below).

Figure 2 provides an example, based on CTD profiling in the Arctic Eurasian Basin. The main pattern is that $T$ increases with depth (as does $S$, not shown here), owing to the presence of a relatively warm and salty layer that enters from the north Atlantic in the upper 500 m. Thus, the column is DL-susceptible. Steppiness is evident through much of the displayed depth range, especially in the lower half. Resolvable gradient reversals are rare in this portion of the water column, so the presence of a staircase seems likely. The inset shows a 20 m tall portion of the profile, along with two possible methods for determining the geometry of this presumed staircase.

The first method, displayed using vertical gray bars in the diagram inset, is a layer-detection scheme patterned on that used by Padman and Dillon (1988). It is based on the histogram of temperature. The basic parameters are the sampling interval $\delta z$ and the temperature gradient $\partial T / \partial z$ smoothed over the depth range of interest. The
product of these quantities provides a natural histogram bin width $\delta T$, in the sense that we would expect a linear profile to have one sample in each such bin. However, a layer within a steppy profile should have several samples in its histogram bin, the number depending on the ratio of the layer thickness to $\delta z$. In the present case $\delta z = 0.25 \text{ m}$ and $\partial T / \partial z = -0.02^\circ \text{C}$, so $\delta T = 0.007^\circ \text{C}$ was used for the $T$ histogram. The idea is to flag temperature bins with high occurrence rates. Therefore the main adjustable parameter in the method is the minimum number $n$ of elements that we demand in a bin before declaring a layer to have been detected. A natural scale for $n$ is $\bar{H} / \delta z$, where $\bar{H}$ is an initial estimate of the layer thickness (perhaps inferred visually or from the integral method sketched above). For the region of the inset $\bar{H} \approx 2 \text{ m}$ so $n \approx 10$. Writing $n = a_1 \bar{H} / \delta z$, we tested sensitivity with trials using $a_1 = 0.3$, 0.4, and 0.5, and found the results to be constant across this range. Figure 2, with $a_1 = 0.3$, shows the method agrees well with layers detected visually. Clearly, the task would have been impossible with CTD data sampled at $\delta z = 2 \text{ m}$, since the layers are only $2 \text{ m}$ thick.

The second method is based on gradients. One decision to be made is whether to focus on high-gradient zones (interfaces) or low-gradient zones (layers). For Figure 2 we searched for interfaces, flagging them as regions in which centered first-difference estimates of $\partial T / \partial z$ exceed the large-scale value $\partial T / \partial z$ by a particular factor, say denoted $a_2$. The horizontal gray bars in the inset of the figure show the results with $a_2 = 1.5$. The visual agreement is reasonably good in terms of interface spacing but not in terms of interface thickness. In this case, the results are quite sensitive to the choice of $a_2$, and a glance at the figure shows why (e.g. consider the second shallowest interface shown in the inset).

In these tests, the gradient technique proved less robust than the histogram technique, in terms of parameter sensitivity. This may not be a general result, but some general explanations come to mind. For one thing, gradient calculations are inherently sensitive to “noise.” For another, the profile (and others like it) seems to indicate that homogeneity of layers is more reproducible from layer to layer than is interface thickness. This may be because of poor resolution in the present example. For example, the second interface in the inset could be an example of a “split layer” (see section 3.2 below).

Tests with varied profiles should expose the relative power of these and other proposed techniques, and indicate choices for the parameters (e.g. $a_1$ and $a_2$ here) that are optimal for various profiling instruments and various physical regimes. The hope is that common techniques used in different studies will facilitate DDC incidence mapping, and that comparing these maps with maps of susceptibility and of

---

2 Herein lies a major difference between the SF and DL cases, since SF layers are often tens of meters thick, and are easily detected with archival CTD records having $\delta z = 2 \text{ m}$, whereas DL layers are often so thin as to be difficult to resolve with even with raw CTD records. This explains why the best examples of DL in the ocean are revealed by microstructure instrumentation, often from stable (ice) platforms that are immune to wave heaving.
forcing will shed light on the dynamics of DDC in the ocean. In the meantime, we are left to examine case studies. This task we take up in the next section.

2.3 Prominent DL Examples

2.3.1 Lakes and Trenches

The geophysical systems that seem most similar to the early laboratory experiments with bottom-heated salt gradients (see section 3.3.2 below) are geothermally-heated salty lakes and semi-isolated deep trenches.

In some lakes, the salt stratification is sufficient to partially stabilize the water column against the action of bottom heating. Several such cases have been studied. The first was Lake Vanda in Antarctica (Hoare, 1966; Shirtcliffe and Calhaem, 1968; Hoare, 1968; Huppert and Turner, 1972). Soon after, DL signatures were found in Powell Lake, a remnant fjord off British Columbia (Osborn, 1973), and in Lake Kivu in Africa (Newman, 1976). The signature in such cases can be striking; for example, Newman (1976) counted 150 regular layers in a single profile.

A similar situation is that of small pockets of saline water trapped near the ocean bottom in the trenches or “brine lakes”, where there is a supply of geothermal heat either from sediments or hot vents. These systems often display evidence of DL in the form of sharp interfaces between homogeneous deep water masses. Some of the earliest studies of this were done in Red Sea (Swallow and Crease, 1965; Voorhis and Dorson, 1975). The interfaces there are very stable, with $R_p \approx 9$ to 25 according to Turner (1969). Similar brine lakes have also been found in the deep eastern Mediterranean basin by Boldrin and Rabitti (1990), wherein the density ratio was even higher, $R_p \approx 600$, and the interfacial heat flux was estimated to be of the same order as the geothermal flux. These facts together with the isotopic composition of lake waters (Stenni and Longinelli, 1990) suggest that the water in the bottom layer is of fossilized origin, developed $\approx 4500$ years ago. Similar features were found in other Eastern Mediterranean brine lakes (MEDRIFF Consortium, 1995), which also showed the recent invasion of a new dense water mass of Aegean Sea origin. It appears that brine lakes can retain a memory of deep water history of a basin, which can be inferred from the water mass structure. McDougall (1984) investigated brine layers capped by a DL interface and supplied by hot saline plumes in a depression, as for the Red Sea brine layers. He found the system to converge to a stable fixed point in $TS$ space, perhaps explaining their observed stationary character.

2.3.2 High-latitude Zones

Foster and Carmack (1976) observed distinct DL staircases in the Weddell Sea, and further details of the spatial extent of layering in the central and western Weddell
Sea have been provided by Muench et al. (1990) and Robertson et al. (1995) respectively. The latter authors pointed out that the vertical fluxes are so low as to be near the detection limit of dissipation-rate profilers. This is reminiscent of the Melling et al. (1984) observation that tides are very weak in the Canadian Basin of the Arctic, where DL signatures are prominent. Indeed, Arctic sampling has provided many illustrations of layering in the ocean (Neal et al., 1969; Neshyba et al., 1971; Padman and Dillon, 1987, 1989; Rudels et al., 1999). As is the case for the Antarctic, Arctic staircase fluxes appear to be weak compared with mixing rates expected in the open ocean. For an example, Padman and Dillon (1987) estimated thermal fluxes to be just 10 times larger than molecular fluxes. One might compare this with a factor of 100 typically found for turbulent mixing in the open sea (Ledwell et al., 1993; Kelley and Van Scoy, 1999). Perhaps more relevant is a comparison with local heat balances. Padman and Dillon (1989) estimate that Arctic DL fluxes are roughly 2 times larger than turbulent fluxes. This argues that DL is important, but since the sum of these vertical fluxes is still much too small to account for historical estimates of Arctic heat losses, Padman and Dillon (1989) suggest that the balance may be determined by benthic mixing along the continental slope followed by isopycnal diffusion into the interior. Such a scenario is familiar in the wider context of studies of ocean mixing, in which vertical turbulent fluxes in the ocean interior are too small to explain inferred large-scale balances (Munk, 1966; Munk and Wunsch, 1998).

As for turbulent mixing in the ocean interior, then, it is not clear whether DL mixing is a rate-limiting step. One way it could still be important is by acting together with other phenomena. For example, it has been suggested that DL fluxes might alter the stratification enough to alter the rate of deep convection in the Greenland Sea (Carmack and Aagaard, 1973; McDougall, 1983). This result, which should be tested in high-resolution numerical models, might imply an indirect way in which DL could be relevant to the global ocean.

Much could be learned about DL in high-latitude waters by contrasting northern and southern cases. For example, the Southern Ocean has much lower density ratios than the northern waters of the Arctic. It would be helpful to know whether this difference in stratification explains differences in the staircase properties in the two regions, or whether other factors, such as local shearing rates, are more important; these topics are revisited below.

2.3.3 Black Sea

A compelling example of many aspects of DL, and one worthy of particular discussion given its potential as a future field site, is the Black Sea. This is the largest land-locked basin of the world with a positive surface water budget (Özsoy and Ünlüata, 1997, 1998). The region has no significant tidal mixing (Gregg and Özsoy, 1999), and this may explain the great variety of DDC processes found there.
The Bosphorus Strait connects the Black Sea to the Mediterranean Sea, and through it enter waters that are relatively warm, salty, and dense. These waters sink when entering the Black Sea (Özsoy et al., 2001), contributing to a stratification that is susceptible to DL throughout the water column with \( R_\rho \approx 2 \) at the top (below the mixed layer) and bottom (above the bottom convective layer), and with much higher values in intermediate depths (Özsoy et al., 1993). Two forms of DL activity result from the input. First, there are lateral intrusions at depths of 100 – 500 m, linked to the descending plume of Mediterranean waters cascading down the slope (Özsoy et al., 1993; Özsoy and Beşiktepe, 1995). The DL-susceptible interior may favor DL convection over SF convection (Turner, 1978), in the context of side-wall buoyancy source (heat and salt) introduced into the stratified environment (Tsinober et al., 1983; Tanny and Tsinober, 1988). Boundary mixing driven by the double-diffusive convection (Özsoy et al., 1993; Özsoy and Beşiktepe, 1995), may be the dominant ventilation mechanism across the permanent halocline of the Black Sea, given the inferred transport of shelf-derived materials, including dissolved nutrients, Chernobyl tracers and inorganic particulates between surface and deep layers (Buesseler et al., 1991; Codispoti et al., 1991; Özsoy et al., 1991, 1993; Buesseler and Livingston, 1997).

In addition to the intrusive mode, the Black Sea provides a remarkable example of the “one dimensional” layering mode. This is in the form of a convecting layer near the bottom. At a thickness of about 450 m, it is the largest known of its kind in the world’s oceans. A destabilizing geothermal heat flux of approximately 0.040 W/m² at the bottom (Zolotarev et al., 1979) acts against an otherwise stable salinity stratification. The transport between the bottom convective layer and the overlying waters occurs through a single DL interface. The observed homogenization of water properties within this layer, both vertically and across the basin, is inferred to occur in a timescale of 40 years, and this may have a bearing on the homogeneity in deposition/diagenesis of bottom sediments (Özsoy et al., 1991; Özsoy and Beşiktepe, 1995). Based on the comparison of flux ratios above the DL interface and in the water column, DDC is inferred as the main agency for heat and salt fluxes in the deep waters extending from the lower part of the pycnocline to the diffusive interface above this bottom convective layer (Özsoy et al., 1991; Murray et al., 1991).

This convecting layer itself may provide a useful test case, since it is thought to be far from the initial phase of convective layer growth identified in laboratory studies (Turner, 1968a; Fernando, 1987; Huppert and Linden, 1979), corresponding instead to a limiting case in which interfacial entrainment is minimal (Fernando, 1987, 1989a; Fernando and Ching, 1991; Özsoy et al., 1991; Özsoy and Beşiktepe, 1995). Further study might indicate whether this result carries over to other sites of interest. Another issue whose study in the Black Sea might prove useful elsewhere relates to the geothermal heat flux. The model of Huppert and Linden (1979) suggests that the DL heat flux is nearly an order of magnitude larger than the geothermal heat flux in this region. However, Fernando (1989a) has argued that this model overestimates fluxes for large convective Richardson numbers. Efforts to resolve
such contradictions may be aided by more work in the laboratory and in the field. The Black Sea, with its prominent DL signatures and its low disruptive shears, may provide an ideal site for the latter.

3 Modelling

3.1 Fluxes Between Two Convecting Layers

3.1.1 Do Layer Thermohaline Contrasts Determine Fluxes?

Early laboratory measurements of $S$ and $T$ fluxes across DDC interfaces separating well-mixed layers suggested that the fluxes $F^S$ and $F^T$ were determined by the inter-layer contrasts, $\Delta S$ and $\Delta T$, between the layers (Turner, 1965). It was hoped that laboratory measurement of the flux relationship would yield parameterizations for use in the ocean, since $\Delta S$ and $\Delta T$ are easy to measure, compared with fluxes.

For the salt-finger case, it has been shown that layer-based laboratory flux laws are unsuitable for extrapolation to the ocean (Gregg and Sanford, 1987; Lueck, 1987), probably because of the effects of shear on interface dynamics (Kunze, 1987, 1994). However, for the DL case, layer-based flux laws may yield more reliable predictions of ocean fluxes. For example, Melling et al. (1984) noted that the inferred conductive heat flux through the interfaces they observed in the western Canadian Arctic was in agreement with laboratory-based flux laws.
There may be several explanations for this important difference between the SF and DL cases in the ocean context. Perhaps SF and DL convective motions are affected differently by background shear (Kunze, 1987, 1994; Padman, 1994). Another possibility is that background motions in DL domains are too weak to have an appreciable affect on the convection. This issue needs resolution since it is fundamental to the understanding of double diffusion in the ocean, as opposed to the laboratory. Pending deeper understanding, though, we are left with a working hypothesis that DL fluxes in the ocean may be predicted in terms of layer properties alone. This holds out the possibility of parameterization of the ocean DL fluxes, if the flux laws can be determined and if the layer contrasts can be predicted. These issues are taken up in the next few sections.

3.1.2 Exponent in Flux Law

In the ground-breaking analysis of Turner (1965) it was suggested that a formula relating DDC fluxes to forcing parameters might be developed by analogy with single-component convection. For this, Turner used a convection model based on the assumption (called into question below) that the convective flux is independent of the thickness $H$ of the convecting layer. Therefore, if the $T$ flux is written in non-dimensional form as a Nusselt number $\Nu = F^T H / (\kappa \Delta T)$, where $\kappa$ is the molecular diffusivity, we have

$$\Nu \propto \Ra^{1/3},$$  (2)

where $\Ra = g \alpha \Delta T H^3 / (\nu \kappa)$ is the Rayleigh number. Here $g$ and $\nu$ are the gravitational acceleration and molecular viscosity, respectively. For thermal convection, dimensional analysis then suggests that the proportionality factor $^3$ may depend on nondimensional molecular properties such as the Prandtl number $\Pr = \nu / \kappa$. Since (2) may be rewritten as

$$F^T \propto \Delta T^{4/3},$$  (3)

flux formulations of this group have come to be called “$4/3$” flux laws.

Such laws have been used extensively in the DDC literature, to extrapolate laboratory measurements to the field. This has been done without extensive testing of the exponent, which is problematic since in the case of single-component convection, there is ample evidence that the exponent in (2) is smaller than the $1/3$ that is commonly assumed.

Figure 3 summarizes several studies of thermal convection in water. Fitted to the form $\Nu \propto \Ra^P$, these data suggest $P = 0.284$, and certainly not $P = 1/3$, given the statistical error bars. This is in good agreement with the prediction, $P = 0.27 \pm 0.02$, of a simple convection-cell model (Kelley, 1990), as well as the analysis

---

$^3$ Laboratory experiments suggest that the proportionality factor in (2) is of order 0.1, with only a weak dependence on the Prandtl number (Turner, 1973).
of Castaing et al. (1989), and the results of direct numerical simulations (Kerr, 1996). Thus, there is little support for the exponent hypothesis of Turner (1965).

Studies are still being made of this issue for thermal convection (e.g. Sommeria, 1999), but to date few tests have been done for the DDC case. For the SF case, Schmitt (1979, Table 3) reported equivalent $P$ values from 0.24 to 0.37, while McDougall and Taylor (1984) found $P = 0.23$ (although experimental scatter prevented them from rejecting an exponent of 1/3). In the DL case, we lack laboratory studies with enough statistical power to measure $P$ as well as desired. One exception would appear to be the Marmorino and Caldwell (1976) study, which reported $P = 0.33 \pm 0.01$. However, this study did not provide details of the error bar, and a re-analysis of the original data for that study (Marmorino, 1974, Table V) yields a 95% confidence interval of $0.27 < p < 0.47$ based on boot-strapped linear regression of logged values (Venables and Ripley, 1999). Clearly, this range is too great to distinguish between the alternative exponents discussed above. It seems that the exponent in the DL case may remain an open issue, until new laboratory experiments have been performed.

In the meantime, one might ask whether the use of an incorrect exponent could lead to large errors in extrapolating laboratory flux measurements to the ocean. The answer depends on the error in the exponent used in a calculation, as well as on the range of Ra values. A representative ocean value is $Ra = 3 \times 10^8$ using $\Delta T = 0.02^\circ C$ and $H = 2$ m for the Arctic (Padman, 1994, Table 1). Compared with the high end of the Ra range in the experiments of Marmorino and Caldwell (1976) and using $P = 0.28$, this implies that the 4/3 flux laws underestimate oceanic fluxes by roughly 40%. This is small compared with the variation in DDC flux at a given Ra, as discussed in the next section. Therefore, the use of 4/3 flux laws, while perhaps incorrect in principle, may be acceptable for rough oceanographic applications. This still leaves the issue of empirical calibration of flux laws for heat and salt, the subjects of the next two sections.

### 3.1.3 Coefficient in Heat-Flux Law

Assuming the validity of the 4/3 exponent for the time being, filling out terms in (3) yields

$$F^T = C \cdot \left( \frac{g \kappa^2}{\alpha^3 \nu} \right)^{1/3} (\alpha \Delta T)^{4/3}$$

where $C$ may depend on molecular properties such as Pr and the Lewis Number $Le = \tau_s/\tau$ (where $\tau_s$ is the haline diffusivity), as well as the density ratio $R_\rho$ and nondimensional measures of other factors such as shear and turbulence. This formula, sometimes with a multiplier of 0.086 that results from comparison with a similar law for thermal convection, has been used extensively in summaries of laboratory experiments. It has been common in the literature to ignore dependence on Le and Pr. This is valid if the temperatures of the laboratory conditions matched
those in the field, but this is not necessarily a good assumption for DL interfaces, which often occur in polar regions. Laboratory experiments are needed to determine the validity of this assumption. For now, Figure 4 summarizes the dependence of $C$ on $R_\rho$, revealed in the laboratory experiments of Turner (1965), Shirtcliffe (1973a), Crapper (1975), Marmorino and Caldwell (1976), Newell (1984), and Taylor (1988). The general trend is for $C$ to decrease with increasing density ratio, by one to two orders of magnitude over the oceanographically relevant range of $R_\rho = 1$ to 10. The rate of decrease is greatest for $R_\rho < 3$ or so, tailing off at higher values.

The observational scatter amounts to nearly a factor of 2 throughout the range. There are also a systematic differences between the results of different studies. These differences are unexplained as of yet. Some of the variation may result from
intercomparison of studies with ranging background conditions, if the exponent in the flux law is wrong, as discussed in the last section. Other factors may stem from varying experimental protocols.

Despite the scatter, there is a clear systematic tendency in Figure 4. Various formulations have been proposed linking $C$ and $R_\rho$. Most of these are purely empirical, but there is an important exception provided by Linden and Shirtcliffe (1978), who constructed a simple model of the diffusive interface. This yields the relationship

$$C_{LS78} = 0.086 \pi^{1/3} \left(1 - \text{Le}^{1/2} R_\rho \right)^{4/3} / \left(1 - \text{Le}^{1/2}\right)^{1/3}$$

(5)

The value $\text{Le} = 0.01$, appropriate to the roughly room temperature of the experiments, has been used in Figure 4; this value should be halved for application to the high-latitude ocean (Fofonoff, 1962). Within the range $2 < R_\rho < 7$, this equation predicts observed fluxes to within a factor of about 2, lending credibility to the elements of the theory. However, the Linden and Shirtcliffe (1978) theory underpredicts fluxes by more than an order of magnitude outside this middle range of density ratio. This limits the viability of the above formulation to oceanic applications.

Therefore, empirical formulations are of great interest for applications. Several such formulations have been proposed. The earliest, and simplest, was provided by Huppert (1971), who summarized the measurements of Turner (1965) using

$$C_{H71} = 0.32 / R_\rho^2.$$  

(6)

Based on new measurements, particularly in the low density-ratio regime, Taylor (1988), proposed a similar formulation,

$$C_{T88} = (0.27 \pm 0.01) / R_\rho^{2.1 \pm 0.2}.$$  

(7)

Through the range $1 < R_\rho < 10$, the latter formula predicts fluxes that are roughly 50% less than those of the Huppert (1971) formulation.

The log-log representation of Figure 4 illustrates that no power law can pass through the cloud of observations throughout this range. For this reason, more complicated empirical formulations have been proposed. For example, in considering their own laboratory experiments, Marmorino and Caldwell (1976) suggested the formulation

$$C_{MC76} = 0.00859 \exp(4.6 \exp[-0.54 \{ R_\rho - 1 \}]).$$  

(8)

This captures the dependence for $R_\rho > 4$, but it over-predicts fluxes by a factor of 2 or more at lower density ratios. To address this, Kelley (1990) took the Marmorino and Caldwell (1976) measurements, together with the other published values, and developed the empirical result

$$C_{K90} = 0.0032 \exp(4.8 / R_\rho^{0.72}).$$  

(9)
As Figure 4 illustrates, the various formulations produce results that are not much more scattered than the observations, at least for density ratios between 2 and 5. However, at higher density ratios, only the Marmorino and Caldwell (1976) and Kelley (1990) formulations lie within the experimental scatter. At density ratios below 2, the experimental scatter increases greatly. Although several of the formulations run through the main body of the observations in this $R_p$ range, it is worth noting that there are outliers with fluxes up to an order of magnitude greater than the rest of the observations. For this reason, caution is warranted for $R_p \lesssim 1.5$.

### 3.1.4 Coefficient in Salt-Flux Law

Historically, it has been common to express $F^S$ in terms of $F^T$, by the use of a nondimensional “flux ratio” defined by $\Gamma = \beta F^S / (\alpha F^T)$. Dimensional analysis suggests that $\Gamma$ should depend on nondimensional forcing parameters such as the density ratio, a Richardson number for background motions, etc., and also on molecular properties of seawater such as Le and Pr.

It has been suggested that $\Gamma$ should tend to unity as $R_p \rightarrow 1$, since then mixing might become more turbulent than diffusive. The reasoning is that the turbulent $T$ flux should be $w'T'$, where $w'$ and $T'$ are turbulent velocity and temperature fields, and similarly for the $S$ flux, and that the $S'$ and $T'$ anomalies should be proportional to $\Delta S$ and $\Delta T$. However, it is not clear whether this enhanced turbulence,
present in bottom-heated laboratory experiments, will also be present in oceanic staircases in which the convection is caused by the diffusion itself, as opposed to an externally supplied heat flux. It would be helpful to know whether run-down experiments or bottom-heated experiments are more relevant to the ocean, since the two setups may yield significantly different results. The former seem to have a less pronounced increase in $\Gamma$ at low density ratios (e.g. McDougall, 1981b). Beyond this disputed range of low $R_\rho$, say starting with $R_\rho = 2$, and ranging up to $R_\rho \approx 1/Le^{1/2} \approx 10$, it has been suggested the diffusive nature of interfaces should yield $\Gamma = Le^{1/2}$ (Veronis, 1968; Shirtcliffe, 1973b; Linden and Shirtcliffe, 1978). At still higher density ratios, theory suggests that $\Gamma \rightarrow Le \cdot R_\rho$, and this has been verified experimentally by Newell (1984).

Many laboratory studies have involved the measurement of how $\Gamma$ depends on various parameters (Turner, 1968a; Huppert, 1971; Linden, 1974a; Crapper, 1975, 1976; Linden and Shirtcliffe, 1978; Takao and Narusawa, 1980; Narusawa, 1986; Taylor, 1988; Fernando, 1989a; Kelley, 1990). A summary is presented in Figure 5. Although the observations generally support the limiting-case predictions, tending to about 0.1 at high density ratios, and perhaps increasing to 1 at low density ratios, it must be emphasized that there is considerable experimental scatter, especially for $R_\rho < 2$. There are also systematic inconsistencies from study to study, and not just at the low-$R_\rho$ regime discussed above; note in particular the factor of 2 disagreement in the two studies that cover the range $3 < R_\rho < 10$.

For use in modelling studies, it is useful to have a parameterization of the dependence of $\Gamma$ on $R_\rho$. Several published parameterizations are illustrated in Figure 5, but one is particularly worthy of note. It is based on a theory developed by Linden (1974a), and is appealing for its simplicity and its power to collapse laboratory observations. It states that

$$\Gamma = \frac{\xi R_\rho + Le^{1/2}(R_\rho - 1)^{3/2}}{\xi + (R_\rho - 1)^{3/2}}$$

(10)

where $\xi$ is a nondimensional parameter that relates to various somewhat general aspects of convective motion and turbulent entrainment. Linden (1974a) suggested the value $\xi = 0.5$ based on a variety of experimental results. The agreement between this prediction and the observations is reasonable (Figure 5). Indeed, the study of Kelley (1990), in which $\xi$ was taken to be an empirically adjustable parameter, yielded only slightly better agreement with observations, with $\xi = 0.7$.

As with the case of heat flux, though, the limit of low density ratio is problematic. For $R_\rho$ less than about 1.5, the scatter in $\Gamma$ measurements amounts to a factor of about 3. Also notable is the systematic disagreement between various experimental suites, amounting to a factor of about 2. Such discrepancies limit the reliability of applications to the ocean to perhaps within a factor of 2, for both heat and salt.
Fig. 6. Density-ratio dependence of nondimensional layer thickness $H/\left(\kappa/\sqrt{N}\right)^{1/2}$ in DL staircases in various ocean and lake regions, as reported by Kelley (1984), Fedorov (1988), and Padman and Dillon (1987). The curve is the prediction of a model of convectively-sheared DL interfaces (Kelley, 1984, 1988).

3.2 Fluxes Within Staircases

The interfacial steps $\Delta T$ and $\Delta S$ within a thermohaline staircase are given by the product of layer thickness $H$ and large-scale gradients of $T$ and $S$. Therefore, if the vertical fluxes of heat and salt in DL staircases are governed by layer-based laws as discussed above, then fluxes can be parameterized in terms of large-scale properties provided that the layer thickness $H$ can be parameterized in similar terms.

This issue was first taken up by Kelley (1984), who proposed on dimensional grounds that $H$ might scale as $H_K = \left(\kappa/\sqrt{N}\right)^{1/2}$, where $N$ is the large-scale buoyancy frequency. The ratio $G = H/H_K$ should thus depend only on molecular properties of the fluid and the large-scale gradient density ratio $R_{\rho}$. (Implicit in such a
formulation in terms of properties of the water column is that no length-scale is im-
posed by other agencies, whether during formation or during evolution; see sections
3.3 and 3.4 below.) If the molecular properties of seawater are similar from location
to location, then \( G \) should depend only on \( R_p \), and a test of the scaling would be to
see whether this scaling collapses oceanographic observations. The Kelley (1984)
study suggested that this is indeed the case for the DL mode of DDC. His later
study suggested that the resulting quasi-empirical relationship could be explained
in terms of a model of conditional layer splitting (Kelley, 1988).

Figure 6 shows the data used by Kelley (1984), together with the results of other
studies. There is uncertainty as to the general validity of the empirical relation-
ship between layer thickness and density ratio put forward by Kelley (1984). The
observations of Fedorov (1988) are in reasonable agreement. Some measure of con-

dence may be inferred from the fact that the observations compiled by Fedorov
However, Figure 6 also illustrates observations made in the field study of Pad-
man and Dillon (1987). These results are in systematic disagreement, by a factor
of \( \sim1/2 \) in layer thickness, with a comparable factor in predicted fluxes. This
disagreement exceeds the uncertainty in fluxes (recall Figure 4 above), providing
ample motivation for caution and for further study. Indeed, it would be helpful to
compare a wide variety of field observations on a diagram patterned on Figure 4.
Unfortunately, this is not a trivial exercise. For example, Muench et al. (1990) noted
layers in the Wedell Sea that are \( \sim10 \) times thicker than those in Figure 6. If this re-
sult is taken at face value, it might indicate a special feature of the very low density
ratios in that region, compared with the earlier results. Another possibility is that
it results from the enhanced nonlinearity of the equation of state under such ambi-
ent conditions (McDougall, 1981a,b). But a third possibility is also worth noting:
the comparison may be spurious. The steppy signals could represent intrusions, not
layers in a staircase. Evidence of this is provided by the coincidence of the steppy
signals and lobes on the \( TS \) diagram in the region, and the presence of strong lateral
variations in \( T \) and \( S \). Only a detailed re-analysis of the observations can resolve
this issue. This uncertainty is an illustration of the difficulties in distinguishing be-
tween intrusions and layers underline the need to develop reliable algorithms for
layer detection (see section 2.2 above).

It would also be helpful to explore more deeply the physical basis of the empirical
results relating layer thickness to background properties. For example, the Kelley
(1988) model suggested that \( H \) is controlled by a process of layer splitting, in
which new layers are formed from existing interfaces. Signatures of such a process
are common in oceanographic profiles (see Figure 8 below), but since there have
been no direct observations of time evolution in the ocean, laboratory work may be
the preferred next step. One advantage of such studies of staircase evolution is that
they may also shed light on the mechanisms by which staircases are formed. As
the next section illustrates in some detail, the topic of formation is problematic not
because of a lack of theories, but rather because of a surfeit.
3.3 Staircase Formation

3.3.1 Negative-Diffusivity Mechanism

In contrast to other mixing agencies, DDC transports density up-gradient. This may be expressed as a negative density diffusivity, \( i.e., K_\rho < 0 \) in an evolution equation of the form

\[
\frac{D\rho}{Dt} = \frac{\partial}{\partial z} \left( K_\rho \frac{\partial \rho}{\partial z} \right) + \ldots
\]  

so that perturbations of the form \( \rho' \sim \exp(at + 2\pi i z/\lambda) \) added to a linear profile will tend to grow, rather than decay as they would with down-gradient density transport. If \( K_\rho \) were a constant, the perturbation growth rate would increase without limit if the wavelength \( \lambda \) were to approach zero, since the timescale is proportional to \( \lambda^2/K_\rho \). However, DDC motions cannot extend to arbitrarily small length scales, and this could provide a method of scale selection. So too could variations of \( K_\rho \) with stratification. However, given our present lack of understanding of how to express DDC fluxes in terms of diffusivities, it does not seem useful to pursue these ideas much further in a diffusivity framework. More useful would be laboratory work and, eventually, direct numerical simulations (Kelley, 2001, Appendix). These could be analogous to the related problem of layers formed by flux divergences associated with stratification-dependent turbulent fluxes (Phillips, 1972; Posmentier, 1977; Ruddick et al., 1989; Balmforth et al., 1998). Indeed, interactions between DDC and turbulent mechanisms of finestructure generation seem particularly worthy of study.

3.3.2 Applied-Flux Mechanism

Turner and Stommel (1964) were the first to create a DL staircase in the laboratory, by heating a stable salinity gradient from below. They found that the initially smooth salinity gradient broke down into a series of convecting layers separated by relatively sharp interfaces. Figure 7, from the followup study of Fernando (1987), reveals the complicated nature of such systems. The focus in many studies has been on simple aspects of the response, and particularly on the thickness of layers that are created. For example, Turner (1968a) proposed that the bottom convective layer grows while maintaining marginal stability at the interface between the mixed layer and the stratified layer aloft, \( i.e. \) that the convection was non-penetrative. As this bottom-most mixed layer grows, a thermal boundary layer above it also grows, and Turner (1968a) argued that the time of creation of a second convecting layer is controlled by the evolution of the thermal boundary layer to a convectively critical state. Denoting by \( Ra_c \) the Rayleigh number of the thermal boundary layer at this time, Turner (1968a) suggested that the thickness of the bottom layer at the onset
Fig. 7. Side view of DL created by heating a salt gradient from below, after Figure 3 of Fernando (1987). Two interfaces are revealed by the striations: “A” is the first quasi-stationary density interface, while “B” is a secondary interface formed above it.

of the second layer is

\[ H_T = \left( \frac{1}{4} \text{Ra}_c \right)^\frac{1}{4} \left( \frac{\nu J_B^3}{\kappa^2 N_{SO}^3} \right)^\frac{1}{4} \]  

(12)

where \( N_{SO} \) is the buoyancy frequency of the initial salt stratification and \( J_B \) is the buoyancy flux applied at the bottom. Huppert and Linden (1979) have extended the analysis to describe the formation of a series of layers.

An issue needing further attention is the extent to which the convection is non-penetrative as assumed by Turner (1968a), or penetrative as assumed by Fernando (1987). Another issue of oceanic relevance is that the “initial” state may involve thermal as well as haline stratification, and is unknown in any case. Furthermore, there are very few situations in which an applied buoyancy flux has been measured (geothermally-heated regions being the exception). Some of these problems were addressed by the analysis of Fernando (1989b), which suggests that the thickness of convecting layers is given by

\[ H_F = c_1 \left( \frac{J}{N_S^3} \right)^{\frac{1}{2}} \left( \frac{1 - c_2 R_p}{(1 - R_{p}^{-1})^4} \right)^{\frac{1}{2}}, \]  

(13)

where \( c_1 \) and \( c_2 \) are constants to be determined empirically, and \( N_S \) is the haline component of stratification at the time of observation (not at some unobserved initial time), \( R_p \) is the instantaneous local density ratio, and \( J \) is the instantaneous local buoyancy flux. Molemaker and Dijkstra (1997) carried out a comparative direct numerical simulation study of the mechanisms proposed by Turner and Fernando and found support for the latter. However, their simulation was two dimensional, so further studies are needed to confirm their inferences.
While the Fernando (1989b) analysis of layer thickness circumvents several difficulties with the earlier theory, it is still difficult to apply to the ocean since the buoyancy flux $J$ is difficult to measure (Gregg, 1987; Gargett, 1997). Furthermore, testing (13) using $J$ calculated from layer-based flux laws is problematic since $\Delta S$ and $\Delta T$ depend on the observed $H$ and the observed stratification parameters, so the two sides of the equation are linked. The best possibility of testing the two theories cited above may be in geothermally heated domains. These should afford measurements of $J$ that can be trusted to within a factor of 5 (e.g. see Table 1 of (Zolotarev et al., 1979), for the Black Sea), so that if $H$ can be measured to within $\sqrt{5}$, hypotheses based on the theories could be tested.

More remains to be done in the laboratory as well. Kerpel et al. (1991) addressed the case in which the heat flux arises from a boundary of constant temperature, as opposed to constant flux. Several other cases come to mind, ranging widely in difficulty of execution in the laboratory. Examples include following the evolution if both salt and heat are supplied at a single boundary, and seeking an equilibrium when heat and salt are supplied at one boundary and removed at another boundary. Issues of the timescale of staircase creation and evolution are also particularly relevant for application to unsteady problems in the ocean, especially in rapidly developing interleaving systems.

### 3.3.3 Modified-Intrusion Mechanism

Recently a new idea has been presented for the formation of staircases, namely that they are modified intrusions (Merryfield, 2000b). This model, to date applied only to the SF case, produces reasonable predictions of measurable quantities such as layer thickness and the critical range of $R_\theta$ for staircase formation. As with many intrusion models, it is based on a diffusivity formulation of DDC fluxes, and as such it shares a dependence on parameters that are essentially unknown since we lack a theory for the DDC fluxes arising out of a given smooth stratification. Application to the DL case, e.g. for application to the Arctic, and including special Arctic features such as along-ridge flows of warm/salty water masses, would be most useful. So would more detailed studies of the formation of initial intrusions, e.g. accounting for baroclinicity in ocean fronts (Kuzmina and Rodionov, 1992; May and Kelley, 1997, 2001a,b), and the possibility of interleaving being set up by sloping boundaries even in the absence of initial lateral variation in water properties (Linden and Weber, 1977). Field work is needed to provide constraints on theories, and an ideal location for this may be the Black Sea, which is subject to both interleaving and staircase modes, perhaps linked (recall section 2.3.3 above), and to low levels of disruption by tidal motion.
3.4 Staircase Evolution

The staircase-formation analysis of Turner (1968a) was followed up by Huppert and Linden (1979), who added a set of ordinary differential equations representing the evolution of \((H, S, T)\) for a suite of layers coupled by layer-based flux formulas. Migration of interior interfaces was ignored, as was layer splitting, but layer merging was modeled crudely. They found reasonable agreement with laboratory observations, suggesting that such a model (perhaps with flux laws adjusted as discussed in section 3.1 above) could be useful in oceanographic applications. Still, this leaves out several oceanographic aspects that we might expect to be important.

Firstly, consider the lateral environment. Analyses of bottom-heated experiments typically regard the situation as one-dimensional, in the sense that lateral variations may be ignored in a statistical sense, even if they are present in instantaneous snapshots (Figure 7). This viewpoint is at odds with the Merryfield (2000b) model of staircase creation from interleaving fields. As for lateral variation in general, only one observational program, the C-SALT study of SF staircases (Schmitt et al., 1987), has sampled a thermohaline staircase at a wide range of lateral scales. This has revealed the expected intrusions near the edge of the staircase zone, but also isolated intrusions within the staircase interior. High-resolution sampling with towed thermistor chains (Marmorino, 1989, 1991) reveals that the lateral features have lengthscales of 1km. Similarly, Padman and Dillon (1988) found that station spacings of less than 1km were required to track individual layers in a DL staircase in the Canadian Basin of the Arctic. This is in line with the observations of Neshyba et al. (1971), an excerpt of whose observations is presented as Figure 8 here. Note in particular the occurrence of “split” layers in this diagram, whose variation in this graphical representation is probably the result of advection of horizontal variability, as opposed to temporal development (Kelley, 1987b). Perhaps the ideas of Stamp et al. (1998), regarding solitary varicose waves within interfaces, are relevant to the dynamics of split layers as well. For now, we must admit that the dynamics of these
split layers, and of lateral variability in general, are poorly understood. If these split layers are important to maintaining a critical layer thickness, as a counter to layer merging by convergent buoyancy fluxes (Kelley, 1988), then further study is definitely warranted. This is because they may hold the key to a usable parameterization of layer thickness, and thus vertical fluxes, in terms of large-scale properties (recall section 3.2 above).

Along with such external effects, there are possibilities for staircase evolution that arise from the properties of seawater. One is interface migration, which may occur due to the nonlinearity of the equation of state of seawater (McDougall, 1981a,b). Preferential interfacial migration may cause flux contributions (Kelley, 1987b; Rudels, 1991), which can be estimated to the first order by considering the entrainment velocities at the interface caused by each layer when the opposite layer is not entraining (Fernando, 1990). The entrainment is effective mainly at low stability ratios; if \( R_p \gtrsim 10 \) then entrainment essentially ceases and the interface becomes either stationary or detraining (Hull et al., 1989; Zangrando and Fernando, 1991).

Another general category of great interest is the effect on double-diffusion of classical oceanographic effects, such as planetary rotation, turbulent mixing, and background shear. For technical reasons, these issues are difficult to work with in the laboratory. Still, each merits examination.

The effect of planetary rotation is to reduce the strength of convection by preventing vortex stretching. However, Kelley (1987a) argues that planetary rotation is unlikely to inhibit DL convection given the small scales of typical oceanic staircases. This raises the question as to whether the scales of observed convecting layers are small because of the inhibition of larger-scale convection. The answer, which may affect ideas on layer scaling and thus on diffusivity formulation, may come from constructing better models of rotating double-diffusive convection, perhaps starting in the laboratory.

The effect of turbulence on SF and DL interfaces has been explored in the laboratory by Linden (1971) and Crapper (1976) respectively, but it is not clear how directly these relate to the ocean, in which mixing may have a different character (e.g. different length and time scales) than in these laboratory situations. A related topic, and one which needs more study, is the relationship between differential mixing and DDC. Differential mixing (DM, henceforth) results from the differing diffusivities for heat and salt acting on water parcels momentarily put into contact, e.g. by incompletely mixed breaking waves (Turner, 1968b; Gargett, 1988; Altman and Gargett, 1990). In contrast with the DDC case, DM may occur even if both \( S \) and \( T \) are stably stratified. Two-dimensional direct numerical simulations suggest that DM may be significant at low turbulent mixing rates (Merryfield et al., 1998), i.e., in the same realm in which DDC seems most prevalent. This raises questions about how DM and DDC might interact in the staircase mode. There is already good reason to study DM since it has recently been proposed as a cause of intru-
sions (Hebert, 1999; Merryfield, 2000a), something that has long been thought to have a double-diffusive cause.

The effect of shear has been alluded to in section 3.1 above. There are indications that the SF mode is inhibited greatly by ocean shears, but that the DL mode is not. Is this just because the DL observations are from areas of low shear? Perhaps not, since it has been proposed that background shear inhibits SF fluxes (Linden, 1974b; Kunze, 1994) but that shear may enhance DL momentum transport (Padman, 1994). Whether this proposed difference in SF and DL interface dynamics is verified or not, it illustrates the danger in relying on a direct analogy between the two modes of DDC.

4 Summary

In many ways, the DL case seems simpler than the SF case. For example, consider fluxes. For decades, oceanographers estimated SF fluxes in the ocean using layer-based flux laws based on laboratory experiments, but the C-SALT program (Gregg and Sanford, 1987; Lueck, 1987) revealed this to be in error by well over an order of magnitude. This may be the result of SF disruption by internal-wave shear (Linden, 1974b; Kunze, 1994). In the DL case, whether by virtue of weaker shears or because of differences in the dynamical nature of DL interfaces, empirical evidence suggests that it may be reasonable to extrapolate from the laboratory using layer-based flux laws. It remains to be seen whether this holds in general, but at least it provides a hypothesis worth testing, and a tentative basis for further computations, while awaiting further work on how DL fluxes are affected by turbulence, shear and other oceanic effects.

If we may express vertical DL fluxes of salt and heat in terms of layer height $H$ and the steps in $S$ and $T$ from layer to layer (without regard to internal-wave shear or turbulence levels, say) then the goal of parameterizing DL fluxes in terms of the large-scale stratification can be met if only we can parameterize $H$ in such terms. Such a parameterization has been proposed for the DL case (Kelley, 1984) but the evidence for it is divided enough to demand more empirical tests. The proposed mechanism behind the parameterization involves the splitting of layers when a critical thickness is achieved (Kelley, 1988). Because of this, and because layer splitting may provide a large component of lateral variability, this general topic seems greatly deserving of further study.

Any parameterization scheme based on an equilibrium assumption will be irrelevant if the ocean changes too quickly for staircases to be formed or to evolve to equilibrium. For this reason some of the most important questions about DDC in the ocean center on the mechanism of staircase formation. There may be several modes. The bottom-heating mode, easily set up in the laboratory, may also occur
in geophysical domains such as geophysically-heat lakes, and oceanic trenches and brine lakes. However, other possibilities exist. These include instabilities associated with up-gradient DDC density transports and a mode of formation from interleaving systems (Hebert, 1999; Merryfield, 2000a). Field programs may be needed to select and differentiate between such alternatives. The Black Sea and the Arctic are prime candidates for field sites. Each displays prominent signatures of DL staircases as well as intrusions. If the SF case provides any guide, an intensive field program using a variety of sampling protocols may be needed. The C-SALT field program re-invigorated research on the SF mode, motivating studies of the effects of shear on interfaces, of intrusions within staircases, of the linkage between vertical and lateral effects, etc. An intensive field program for the DL case might be as productive as C-SALT proved to be for the SF case, and go a long way to establishing whether DL is important in the ocean, globally or locally.

Acknowledgements

We are indebted to the reviewers of this manuscript, who pointed out weaknesses in the writing as well as important scientific issues, through each of the review stages. In depth and in breadth, these comments were unusual, and we appreciate them greatly.

References


