Isoneutral Diffusion in a *z*-Coordinate Ocean Model

STEPHEN M. GRIFFIES

Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey

ANAND GNANADESIKAN

Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey

RONALD C. PACANOWSKI

Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey

VITALY D. LARICHEV

Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey

JOHN K. DUKOWICZ AND RICHARD D. SMITH

Los Alamos National Laboratory, Theoretical Division, Los Alamos, New Mexico

(Manuscript received 3 March 1997, in final form 6 August 1997)

ABSTRACT

This paper considers the requirements that must be satisfied in order to provide a stable and physically based isoneutral tracer diffusion scheme in a *z*-coordinate ocean model. Two properties are emphasized: 1) downgradient orientation of the diffusive fluxes along the neutral directions and 2) zero isoneutral diffusive flux of locally referenced potential density. It is shown that the Cox diffusion scheme does not respect either of these properties, which provides an explanation for the necessity to add a nontrivial background horizontal diffusion to that scheme. A new isoneutral diffusion scheme is proposed that aims to satisfy the stated properties and is found to require no horizontal background diffusion.

1. Introduction

The mixing of ocean tracers occurs predominantly along directions tangent to the locally referenced potential density surface (Iselin 1939; Montgomery 1940; Solomon 1971; Redi 1982; Olbers et al. 1985; McDougall 1987a; Gent and McWilliams 1990; Ledwell et al. 1993; Kunze and Sanford 1996). The analyses from Ledwell et al. (1993) and Kunze and Sanford (1996) establish the large degree to which tracer mixing along these *neutral directions* dominates the cross or *dianeutral* mixing. Their results substantiate the long standing hypothesis that the ocean interior is highly adiabatic in the sense that tracer properties are only very slowly mixed across the neutral directions. In turn, the measurements support the relevance of layer models for simulating the adiabatic aspects of ocean dynamics (e.g., Bleck et al. 1992; Hallberg 1995). Furthermore, they provide a severe constraint that the *z*-coordinate models must satisfy in order to provide physically realistic simulations of the ocean interior.

As argued by Redi (1982), a large component of ocean tracer mixing can be parameterized as downgradient diffusion along neutral directions, referred to in the following as *isoneutral diffusion.* Redi's work provides a conceptual framework allowing the *z* models to move away from the physically unrealistic horizontal/vertical diffusion in which the diffusive fluxes are oriented according to the local geopotential direction. What is necessary, therefore, is a straightforward rotation of the diffusion tensor to align the diffusive fluxes along the neutral directions. Cox (1987, hereafter C87) implemented a diffusion scheme in the Cox (1984) version of the GFDL ocean model, which attempted to numerically realize Redi's isoneutral/dianeutral diffusion. The C87 diffusion scheme improved certain aspects of the *z* model's fidelity when compared to the traditional horizontal diffusion simulations. In particular, C87 pro-

Corresponding author address: Dr. Stephen M. Griffies, Geophysical Fluid Dynamics Laboratory, Princeton University, Route 1, Forrestal Campus, Princeton, NJ 08542.
E-mail: smg@gfdl.gov

duced a more realistic thermocline and mitigated the Veronis effect (Veronis 1975; Gough and Lin 1995; GFDL Ocean and Climate Groups 1996, personal communication). Cox's work forms the foundations upon which numerous other models have grown over the years, and the original C87 discretization of isoneutral diffusion has remained fundamentally unchanged.

Over the course of long-term climate integrations (order $\geq$ 100 years), the effects from how tracer mixing is parameterized become especially visible in a numerical simulation. It is on such timescales that the positive affects from the C87 scheme have been diagnosed. Unfortunately, it is on such timescales that the problems with this scheme are also most apparent. Namely, C87 contains a numerical instability, whose characteristics are described in this paper, that prevents it from being run without an added nonneglible amount of horizontally aligned *background diffusion.* This background diffusion contributes to the general overly diffused nature of the coarse resolution *z* models. One result is that tracer properties are not well preserved over the thousands of kilometers seen in observations; a problem intimately related to the models having too much dianeutral mixing. Large-scale preservation of tracer properties is essential in order to utilize the *z* models for climate simulations. This paper focuses on the question: Is it possible to realize isoneutral diffusion in a *z*-coordinate ocean model in a physically based and numerically stable fashion so that no added background diffusion is required? The new diffusion scheme documented in this paper indicates that it is possible to do so within certain limitations.

The problem with background horizontal diffusion is most apparent when gauged in terms of the smallness of the measured dianeutral diffusivity. Background diffusion can create a tracer flux that dominates the flux parameterized with dianeutral diffusivity in regions where the isoneutral slope S is larger than $(A_D/A_{\text{back}})^{1/2}$, where A_{back} is the horizontal background diffusivity and A_D is the dianeutral diffusivity. The typical background diffusivity used with the C87 scheme is normally larger than 10% of the isoneutral diffusivity, which means $A_{\text{back}} \approx 10^6$ cm^2 s^{-1}. With an observed $A_D \approx 0.1$ cm^2 s^{-1} (Ledwell et al. 1993; Kunze and Sanford 1996), background horizontal diffusion contributes to a larger dianeutral flux than that explicitly parameterized by A_D when the isoneutral slopes are greater than 3×10^{-4}, which is a modest slope commonly realized in the ocean. The problems associated with overly large dianeutral diffusion have been pointed out in various studies (Veronis 1975; McDougall and Church 1986; Gough and Welch 1994; Böning et al. 1995; Gough and Lin 1995; Hirst et al. 1996). Some of these references point to the problems with traditional horizontal/vertical diffusion, for which the C87 scheme was meant to remedy. As mentioned previously, C87 indeed mitigated the problems, yet because of the need for background diffusion, it did not solve them completely [see especially Hirst et al. (1996) for focus on problems with horizontal background diffusion]. Additionally, the recent confidence taken in the $A_D \approx 0.1$ cm^2 s^{-1} measurements places a high priority on finding the means to eliminate essentially all background diffusion in the *z* models.

In addition to diffusive tracer mixing, Gent and McWilliams (1990, hereafter GM90) argued for the presence of a quasi-adiabatic stirring mechanism that conserves all tracer moments between isoneutral layers, yet systematically reduces the isoneutral slopes and so acts as a sink for available potential energy. This transport can be parameterized by a divergence-free advective velocity (Gent et al. 1995), or equivalently with an antisymmetric stirring tensor (Plumb and Mahlman 1987), which gives rise to a skew-diffusive flux (Griffies 1998). The GM90 stirring therefore complements that mixing obtained with the symmetric Redi diffusion tensor. Many ocean modelers have implemented the GM90 scheme in *z* models, in addition to C87 diffusion, and have found the simulations to have an added amount of realism over models without GM90 (e.g., Danabasoglu and McWilliams 1995; Large et al. 1996). As a result, the GM90 ideas, along with Redi diffusion, may provide a useful, albeit incomplete, framework for parameterizing ocean tracer mixing. It follows that in order to evaluate the affects of isoneutral diffusion and GM90 eddy-induced advection (or skew diffusion) in *z* models, it is necessary to provide a clean numerical realization of both processes. This paper focuses on the Redi diffusion process, and the paper by Griffies (1998) focuses on GM90 skew diffusion.

Before entering the main body of this paper, it is useful to summarize the two central results that provide a foundation for our new isoneutral diffusion scheme: First, we find that it is necessary to build the property of downgradient tracer diffusion, and the consequent variance reduction, into a numerical isoneutral diffusion scheme. Otherwise, the scheme can produce uncontrollable upgradient fluxes and result in model blowups. Second, it is necessary to balance the isoneutral fluxes of the active tracers in order to realize a zero isoneutral flux of locally referenced potential density. Otherwise, even for a variance reducing scheme, grid-scale noise can result. Both of these properties are fundamental to what is meant by isoneutral diffusion, both in the continuum and on the numerical lattice.

The plan of this paper is the following. Section 2 presents numerical experiments that illustrate the problems with the C87 diffusion scheme. In section 3, kinematical properties of isoneutral/dianeutral diffusion are discussed. In section 4, problems of the C87 scheme are interpreted in terms of these kinematical properties. Section 5 describes a new discretization of isoneutral diffusion based on a functional formalism. Section 6 presents results from model simulations using this scheme. Section 7 presents summary and conclusions. There are five appendices: appendix A critiques a linear stability analysis that incorrectly concludes that the C87 scheme

is unstable for large isoneutral slopes; appendix B discusses basic mathematical properties of cabbeling and thermobaricity; appendix C considers issues related to realizing steep isoneutral slopes and provides a critique of the commonly used *slope clipping* scheme of C87; appendix D discusses what is meant by ''downgradient diffusion'' in the new isoneutral diffusion scheme; and appendix E provides details for discretizing the full Redi tensor.

Some conventions should be noted. The word ''isoneutral'' is preferred to the more commonly employed ''isopycnal,'' by which we mean processes occuring in the neutral directions. ''Temperature'' refers to potential temperature θ, which is temperature referenced to the ocean surface. Tensorial notation is employed, in which the labels m, n, which represent the coordinates x, y, z, are assumed to be summed if repeated. Vectorial notation is also employed for added clarification. The labels i, j, k refer to discrete spatial grid points in the x, y, z directions used for expressing equations on the lattice, and no summation will be assumed for these labels. The diffusive flux discretization is lagged one time step in order to ensure stability of the time discretization (Haltiner and Williams 1980); the time step is omitted for brevity. Model results in this paper are computed with the GFDL MOM 2 Ocean Model (Pacanowski 1996).

2. Model experiments

For the numerical experiments presented in this paper, a sector domain is employed that extends from 5°N to 65°N latitude and 60° wide in longitude. The horizontal grid resolution is 2.4° × 2.4° and there are 18 unevenly spaced vertical levels with a flat bottom at 4000 m. A rigid-lid boundary condition is applied to the surface. The external mode is solved using the streamfunction method with a conjugate gradient algorithm and 9-point numerics for the Laplacian. For the cases in which both temperature and salinity are active, the surface level is forced by restoring to the zonally averaged annual mean Levitus (1982) climatology. When salinity is uniform and constant, the linear temperature profile of Cox and Bryan (1984) is used. Both cases use a 50-day timescale for the restoring, defined over the 35-m top level. The model diagnoses density by using the Bryan and Cox (1972) equation of state, which consists of separate cubic approximations to the UNESCO equation of state (Gill 1982) for each vertical model level. The dianeutral diffusivity is constant with depth. Momentum is dissipated with constant viscosity of 10^9 cm^2 s^{-1} in the horizontal and 10 cm^2 s^{-1} in the vertical and no wind forcing is used. The scheme of Rahmstorf (1993) is employed to completely gravitationally stabilize the vertical columns at each time step. In regions where the slopes of the neutral directions steepen, the scaling of the isoneutral diffusion coefficient as described by Gerdes et al. (1991) is used in order to maintain linear stability

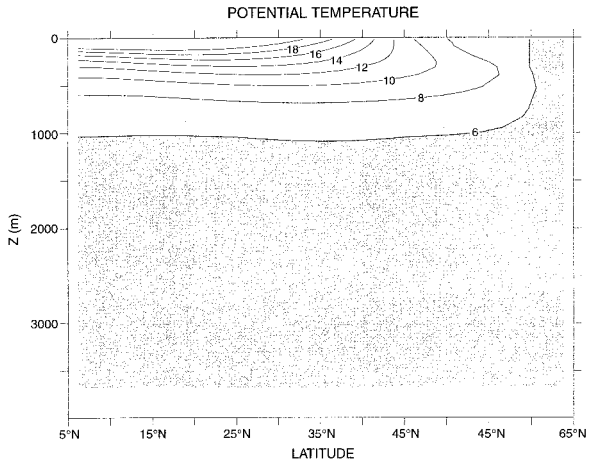

FIG. 1. Temperature after 3400 years of integration with horizontal (A_H = 10^7 cm^2 s^{-1}) and vertical (A_V = 0.5 cm^2 s^{-1}) diffusion. This is the initial condition for the experiments in which both temperature and salinity are active. The initial condition for the cases with only temperature active is similar, except that the temperature is stably stratified everywhere.

of the diffusion equation. Further discussion of this scaling is provided in appendix C. The time steps are 1 h for the momentum and external mode, 16 h for the tracers, and no other acceleration is used. The basic steady-state thermohaline circulation seen in such a model was discussed by Bryan (1975).

Using this model configuration, various test problems are carried out based on the following approach. First, the model is spun up to a pseudoequilibrium using a vertical diffusivity of 0.5 cm^2 s^{-1} and horizontal diffusivity[1] of 10^7 cm^2 s^{-1}. The integration is long enough to establish the thermocline and outcropping in the northern part of the basin associated with free convection. Figure 1 provides a meridional snapshot of the temperature after 3400 years for the case with temperature and salinity both active. After such a spinup, horizontal diffusion is switched to various isoneutral diffusion schemes. These ''switching'' experiments provide an inexpensive means to evaluate the integrity of the numerics. Model years will refer to years subsequent to this switch.

a. Experiments with C87

The special case of an ocean model with buoyancy dependent on a single variable, such as a single active tracer, is a useful place to begin testing the C87 diffusion scheme. In the continuum, the isoneutral portion of the diffusion tensor will not act on this tracer since the

[1] With a 10^7 cm^2 s^{-1} diffusion coefficient, the horizontal diffusion equation grid CFL number $A_H \Delta t/(\Delta x)^2$ is 2 × 10^{-4}, which is well within the constraints of stability for the linear diffusion equation (see appendix C).

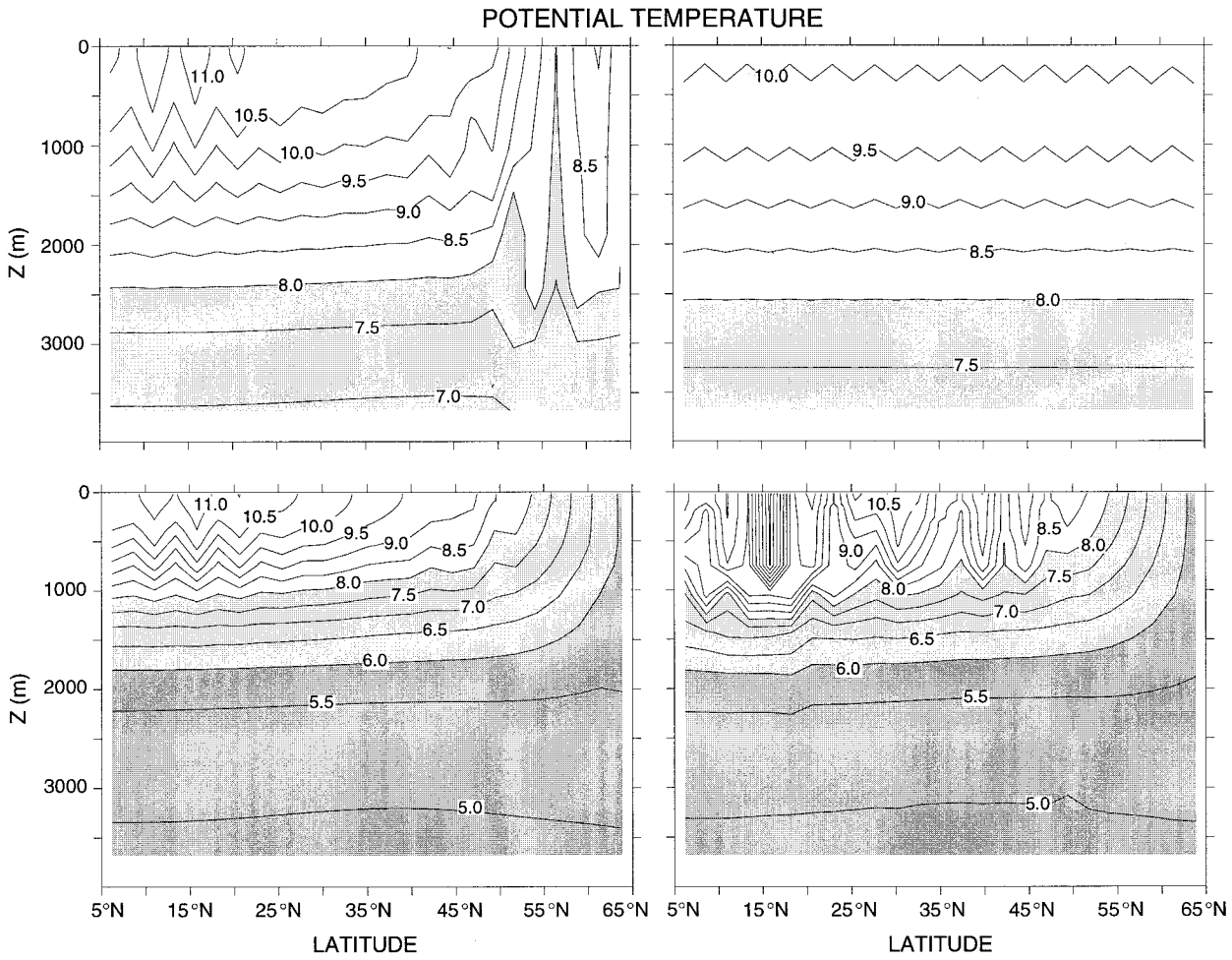

FIG. 2. Potential temperature from four different ''switching'' experiments, each of which used the C87 diffusion scheme. (a) Upper left panel: 600 years after the switch was made using a single active tracer (salinity held constant in space and time), 10^7 cm^2 s^{-1} isoneutral diffusivity, 0.5 cm^2 s^{-1} dianeutral diffusivity, and no advective transport. (b) Lower left panel: 300 years after the switch was made using two active tracers with 10^7 cm^2 s^{-1} isoneutral diffusivity, 0.5 cm^2 s^{-1} dianeutral diffusivity, and no advective transport. (c) Upper right panel: 900 years after the switch was made using a single active tracer, 10^7 cm^2 s^{-1} isoneutral diffusivity, 0.5 cm^2 s^{-1} dianeutral diffusivity, and FCT advective transport. (d) Lower right panel: 200 years after the switch was made using two active tracers, 10^7 cm^2 s^{-1} isoneutral diffusivity, 0.5 cm^2 s^{-1} dianeutral diffusivity, 10^7 cm^2 s^{-1} thickness diffusivity with the GM90 advective flux. The domain-averaged tracer remained constant for each of these experiments, thus indicating the absence of false sources in the C87 scheme.

neutral directions are parallel to the isotracer surfaces. This property is true for both linear and nonlinear equations of state (section 3 and appendix B provide further discussion). To test the ability of the C87 scheme to respect this property, we ran the sector model using horizontal diffusion with salinity fixed throughout the domain at 35 psu, and allowed temperature to be active with the nonlinear equation of state. We then switched from horizontal to C87 diffusion using 10^7 cm^2 s^{-1} for the isoneutral diffusion coefficient, removed all surface forcing, and turned off tracer advection.

Figure 2a shows the solution 600 years after the switch. As expected, vertical gradients of isotherms are weakened due to the nonzero dianeutral diffusivity. However, there is a near complete loss of numerical integrity. In the north, temperature has undergone a vigorous amount of vertical mixing due to convection acting on the unstable solution. In the south, where convection is absent, the presence of unstable grid waves is clear. These waves grew unbounded in subsequent years of integration, which caused the solution to eventually blow up. Figure 2b shows the temperature at year 300 for the case in which salinity is also allowed to change. This solution for two active tracers is consistent with the single active tracer experiment. For both cases, setting the dianeutral diffusivity to zero increased the growth rate of the instability by roughly one order of magnitude.

Dispersion errors associated with grid Peclet number violations encountered with centered advection (Bryan et al. 1975; Weaver and Sarachik 1990) have traditionally been considered one of the main reasons for em-

ploying horizontal background diffusion with C87 (see comment at the end of C87). Since tracer advection is removed from these experiments, Peclet grid noise has nothing to do with this unstable behavior.

Another example of the problems with the diffusion scheme can be seen when performing the switching experiment while maintaining an Eulerian advective transport. For centered differenced advection, the solution blows up (not shown) even sooner than without advection. This behavior might be expected based on grid Peclet arguments. A reasonable question to ask is whether a completely monotonic advection scheme, such as flux corrected transport (FCT) (see Gerdes et al. 1991), is sufficient to suppress the C87 instability. Although FCT acts only on the advective fluxes, it might provide enough dissipation to stabilize the C87 scheme. Figure 2c shows that this possibility is not completely realized, even with a single active tracer for which the diffusive fluxes should vanish. After switching the diffusion from horizontal to C87, switching the advection from centered to FCT, and releasing the surface forcing the model undergoes a spindown in which the isotherms eventually flatten and spread apart due to the dianeutral diffusion and lack of surface forcing. However, by year 900, there is a nontrivial grid wave whose growth is able to overcome the stabilizing aspects of the FCT scheme. The amplitude of this grid wave continued to grow in subsequent years. Turning off all diffusion and using FCT alone resulted in smooth and flat isotherms, with no sign of grid noise.

The implementation of GM90 eddy-induced advective transport in z-coordinate models has been associated with the ability to remove the background horizontal diffusion otherwise necessary with the C87 scheme. It is noted that GM90 acts to reduce the isoneutral slopes, thus producing more horizontally aligned diffusive fluxes. Such a reduction in slopes may provide for increased numerical stability according to the linear stability analysis discussed in appendix A. However, that analysis is in error and, so, is not relevant for isoneutral diffusion. A different reasoning for why GM90 stabilizes certain experiments is provided in Griffies (1998). Unfortunately, as seen in Fig. 2d, the switching experiment with GM90 eddy-induced advection, C87 diffusion, two active tracers, and 0.5 $\text{cm}^2\ \text{s}^{-1}$ dianeutral diffusivity is not stable. Indeed, when compared to Fig. 2b, the solution looks even worse, and it blows up sooner than without GM90 advection. Such behavior is consistent with the problems encountered when implementing GM90 eddy-induced advection as described by Weaver and Eby (1997). As discussed by Griffies (1998), the problems with this GM90 experiment are related to the method of implementing *both* the GM90 closure as well as isoneutral diffusion.

b. Comments on the switching experiments

We performed numerous other switching experiments in this and other model configurations with various permutations of subgrid-scale parameterizations. Consistently, the only way to eliminate the unstable grid waves was to add at least 10% background horizontal diffusion to the C87 diffusion scheme. As discussed in the introduction, such background diffusion is not a viable choice for realistic climate modeling. It should be noted that when running the C87 scheme with centered advection from the start of a spinup from uniform tracer fields, the sector model remained stable, with only modest grid noise. Therefore, the switching experiments provide substantially stronger tests of the numerical integrity than the spinup experiments. The reason is that the diffusive fluxes of the active tracers have the potential to be much larger in the period after switching between diffusion processes than during a spin up from rest. Within the spirit of evaluating the performance of a scheme under various conditions, some of which could be realized in more realistic models run under time-varying forcing and with bottom topography, we conclude that the C87 scheme is unsound. The remainder of this paper is devoted to providing the physical and numerical understanding necessary to interpret these problems with C87 and thereafter to derive a new scheme that aims to rectify them.

3. Kinematics of isoneutral diffusion

a. The diffusion tensor and the diffusion operator

A diffusion tensor $\mathbf{K}$ is a symmetric, positive semi-definite second-order tensor. The components $F^m(T)$ of the diffusive flux are related to the diffusion tensor through $F^m(T) = -K^{mn}\partial_n T$, where T is any tracer and K^{mn} are components to the diffusion tensor $\mathbf{K}$. The diffusion tensor therefore acts as a matrix operator that orients the tracer gradient in the process of defining the tracer diffusive flux. The tracer will evolve due to divergences of the flux, resulting in the diffusion equation $\partial_t T = -\partial_m F^m(T) \equiv R(T)$, where $R(T)$ is termed the *diffusion operator*. The simplest type of diffusion is isotropic diffusion, for which $K^{mn} = A\delta^{mn}$, $A > 0$ is a positive diffusion coefficient, and δ^{mn} is the Kronecker delta, which equals unity when $m = n$ and vanishes otherwise. For horizontal/vertical diffusion, the tracer flux in the vertical is distinguished from that in the horizontal which means the diffusion tensor is anisotropic and has components $K^{mn} = A_H(\delta^{mn} - \hat{z}^m\hat{z}^n) + A_V\hat{z}^m\hat{z}^n$, where A_H, A_V are nonnegative horizontal and vertical diffusion coefficients, respectively, and $\hat{z}^n$ is the nth component to the unit vector (0, 0, 1) in the vertical direction.

Isoneutral/dianeutral diffusion is completely analogous to horizontal/vertical diffusion, only now it is the dianeutral direction, defined by the dianeutral unit vector $\hat{\gamma}$, that is distinguished from the isoneutral directions, defined by two unit vectors $\hat{e}_1$ and $\hat{e}_2$. The diffusion tensor therefore takes the form

$$K^{mn} = A_I(\delta^{mn} - \hat{\gamma}^m\hat{\gamma}^n) + A_D\hat{\gamma}^m\hat{\gamma}^n, \qquad (1)$$

where A_I, A_D are the isoneutral and dianeutral diffusion coefficients, respectively. Writing the diffusion tensor in this form provides for a simple geometric interpretation. Namely, the isoneutral piece $A_I(\delta^{mn} - \hat{\gamma}^m\hat{\gamma}^n)$ acts as a projection operator that projects out that component $(\nabla T)_{\text{iso}}$ of the tracer gradient within the tangent plane defined by the two neutral directions. The dianeutral piece $A_D\hat{\gamma}^m\hat{\gamma}^n$ projects out that component $(\nabla T)_{\text{diap}}$ parallel to the dianeutral unit vector $\hat{\gamma}$. Written as a matrix, the isoneutral/dianeutral (Redi) diffusion tensor takes the form (Redi 1982)

$$\mathsf{K} = \frac{A_I}{(1 + S^2)} \times \begin{pmatrix} 1 + S_y^2 + \epsilon S_x^2 & (\epsilon - 1)S_xS_y & (1 - \epsilon)S_x \\ (\epsilon - 1)S_xS_y & 1 + S_x^2 + \epsilon S_y^2 & (1 - \epsilon)S_y \\ (1 - \epsilon)S_x & (1 - \epsilon)S_y & \epsilon + S^2 \end{pmatrix}, \tag{2}$$

and the small slope approximation to this tensor (GM90) is

$$\mathsf{K}^{\text{small}} = A_I\begin{pmatrix} 1 & 0 & S_x \\ 0 & 1 & S_y \\ S_x & S_y & \epsilon + S^2 \end{pmatrix}. \tag{3}$$

In these expressions, $\mathbf{S} = (S_x, S_y, 0) = (-\partial_x\rho/\partial_z\rho, -\partial_y\rho/\partial_z\rho, 0)$ is the isoneutral slope vector with magnitude S, and $\epsilon = A_D/A_I \approx 10^{-7}$ to 10^{-8} is the ratio of the dianeutral to isoneutral diffusion coefficients. It is important to note that the kinematical properties discussed in the remainder of the section apply to both the full and small angle diffusion tensors.

b. Tracer variance

Integrating the diffusion equation $\partial_t T = -\nabla\cdot\mathbf{F}$ over a source-free domain with insulating boundaries (i.e., Neumann boundary conditions) indicates that diffusion will not change the total amount of the tracer. Integrating the tracer squared $\partial_t T^2 = -2\nabla\cdot(T\mathbf{F}) + 2\nabla T\cdot\mathbf{F}$ over the same domain indicates that diffusion will not increase the tracer variance

$$\begin{aligned} \partial_t \int d\mathbf{x}\ T^2 &= 2\int d\mathbf{x}\ \nabla T\cdot\mathbf{F}, \\ &= -2\int d\mathbf{x}\ \partial_m T K^{mn}\partial_n T, \\ &\le 0. \end{aligned} \tag{4}$$

In this expression, $d\mathbf{x}$ is the volume element $dx\,dy\,dz$, and the inequality follows since the diffusion tensor is symmetric and positive semidefinite or, equivalently, the diffusive flux is directed down the tracer gradient ($\nabla T\cdot\mathbf{F} \le 0$). Note that the diffusion tensor could have a zero determinant, which is the case for zero dianeutral diffusion with the Redi tensor. Downgradient diffusion, and the resulting reduction of tracer variance, is the first of two fundamental properties that we aim to realize in a numerical diffusion scheme. It will be referred to as Property I in the subsequent development.

It is useful to explicitly consider the case of diffusion with the isoneutral/dianeutral diffusion tensor. In this case, the variance equation takes the form

$$\begin{aligned} \partial_t \int T^2\,d\mathbf{x} &= -2\int d\mathbf{x}\ [A_I(|\hat{e}_1\cdot\nabla T|^2 + |\hat{e}_2\cdot\nabla T|^2) \\ &\qquad + A_D|\hat{\gamma}\cdot\nabla T|^2], \\ &= -2\int d\mathbf{x}\ [A_I(|\nabla T|^2 - |\hat{\gamma}\cdot\nabla T|^2) \\ &\qquad + A_D|\hat{\gamma}\cdot\nabla T|^2]. \end{aligned} \tag{5}$$

The first form is analogous to that resulting from horizontal/vertical diffusion, in which the unit vectors $\hat{e}_1$, $\hat{e}_2$, $\hat{\gamma}$ replace the unit vectors $\hat{x}$, $\hat{y}$, $\hat{z}$, respectively. The second form suggests the following interpretation. The term $-A_I|\nabla T|^2 \le 0$ represents an isotropic term that acts to dissipate all gradients, and hence all curvature, just as that occurring in isotropic diffusion. The second term $A_I|\hat{\gamma}\cdot\nabla T|^2 \ge 0$ represents the effects of isoneutral diffusion acting to align the tracer isolines parallel to the neutral directions. Through this alignment process, structure is added to the tracer field, which therefore acts to *increase* the tracer variance; hence the positive sign for this term. When the tracer is perfectly aligned along the neutral direction, cancellation occurs between the two components of the isoneutral diffusion term. In the competition between the dissipative and alignment components, the dissipative component wins since $-|\nabla T|^2 + |\hat{\gamma}\cdot\nabla T|^2 = -|\nabla T\times\hat{\gamma}|^2 = -|\hat{e}_1\cdot\nabla T|^2 - |\hat{e}_2\cdot\nabla T|^2 \le 0$, thus ensuring that the total tracer variance will not increase.

The previous discussion suggests the following example in order to illustrate an important point. Consider a tracer field with its power concentrated in the long wavelengths; that is, it is a smooth yet nonuniform tracer field. Allow this tracer to diffuse downgradient along static neutral directions that have a lot of spatial power in high wavenumbers: for example, the density field has a grid noise structure. Isoneutral diffusion will result in the tracer becoming aligned along the neutral directions, which means that it will have power transferred to the high wavenumbers. In the process, the total tracer variance will reduce due to the dissipative nature of diffusion. Therefore, even though tracer variance will not increase with isoneutral diffusion, it is possible to introduce grid noise into the tracer field if such noise is in the density field. This point will prove fundamental in the development of section 4.

c. Balance of the active tracer isoneutral diffusive fluxes

The isoneutral diffusion operator is a nonlinear function of active tracers through the dependence of the

diffusion tensor on temperature and salinity. In this way, isoneutral diffusion differs fundamentally from more traditional forms of linear diffusion in which the diffusion tensor is independent of tracers. Since the isoneutral/dianeutral diffusion tensor is constructed locally according to the dianeutral unit vector $\hat{\gamma}$, it is necessary to discuss some ideas about neutral directions and how to compute $\hat{\gamma}$.

In the ocean, neutral directions are those for which an adiabatic displacement of a parcel is not affected by buoyancy forces. As a familiar example, note that after a free convective event, a vertical column is typically neutrally buoyant and the vertical is correspondingly a neutral direction. McDougall (1987a) formalized this idea to all three directions, and argued that because mixing can act unopposed by buoyancy forces, neutral directions are relevant for orienting the tracer diffusive fluxes.

As discussed by McDougall and Jackett (1988), it is not possible to define a coordinate γ that can globally describe the envelope of neutral directions, unless one neglects the affects of cabbeling and thermobaricity through linearizing the equation of state (see appendix B for further details). However, all that is necessary for our purposes is a local description, which is available for the general case of two active tracers with a nonlinear equation of state. For this description, start by noting that through any point in the ocean with pressure p, there passes a potential density surface $\rho_p(\theta, s)$, which is referenced to the same pressure p. McDougall (1987a) showed that the neutral directions at this point lie within the tangent plane to the ρ_p surface at this point. It follows that the plane's unit normal vector $\nabla\rho_p|\nabla\rho_p|^{-1}$ at this point is perpendicular to the neutral directions within this plane, so it can be identified with the dianeutral unit vector $\hat{\gamma}$.

It is useful to relate $\nabla\rho_p$ to the active tracer gradients when computing $\hat{\gamma}$. For this purpose, note that the log-gradient of the locally referenced potential density can be written in terms of the gradients of the active tracers through $\nabla \ln\rho_p(\theta, s) = (\partial \ln\rho_p/\partial\theta)\nabla\theta + (\partial \ln\rho_p/\partial s)\nabla s$. Note the absence of pressure gradients due to the local referencing. It follows that, since this density gradient is of interest only when evaluated at the in situ pressure where the potential density is referenced, the potential density partial derivatives are equivalent to the thermal and saline expansion coefficients $-\partial \ln\rho_p/\partial\theta = \alpha_p(\theta, s) \equiv \alpha(\theta, s, p)$ and $\partial \ln\rho_p/\partial s = \beta_p(\theta, s) \equiv \beta(\theta, s, p)$ (section 3.7.4 of Gill 1982; McDougall 1987a). For example, the vertical component of this gradient yields the familiar buoyancy frequency $N^2 = -gd \ln\rho_p/dz = g(\alpha\partial_z\theta - \beta\partial_z s)$.

Since it is sufficient to evaluate all quantities in this section at the local pressure p, the p label can be unambiguously dropped from both the potential density and the expansion coefficients in order to reduce clutter. It is important to remember, however, the local referencing used for all subsequent equations. With these conventions, the log-gradient of the locally referenced potential density ρ is written:

$$\nabla \ln\rho = -\alpha\nabla\theta + \beta\nabla s, \tag{6}$$

which yields the expression for the dianeutral unit vector:

$$\hat{\gamma} = -\frac{\alpha\nabla\theta - \beta\nabla s}{|\alpha\nabla\theta - \beta\nabla s|}. \tag{7}$$

A very important consequence of Eqs. (6) and (7) is that the isoneutral diffusive flux of the locally referenced potential density vanishes. It follows that the isoneutral diffusive fluxes of the active tracers are coupled since

$$\rho^{-1}\mathbf{F}_I(\rho) = -\alpha\mathbf{F}_I(\theta) + \beta\mathbf{F}_I(s) = 0. \tag{8}$$

This relation is taken as the second of two fundamental properties that we aim to realize in a numerical isoneutral diffusion scheme. It will be referred to as Property II in the subsequent development.

4. Analysis of the C87 diffusion scheme

a. Grid-scale computational modes

For the small angle approximated diffusion tensor, the C87 discretization of the isoneutral diffusion flux of a tracer in a two-dimensional x–z model is

$$-F^x_{i,k} = A_I\left[\delta_x T_{i,k} - \left(\frac{\delta_x\rho_{i,k}}{\delta_z\overline{\rho}^{\,x,z}_{i,k-1}}\right)\delta_z\overline{T}^{\,x,z}_{i,k-1}\right], \tag{9}$$

$$-F^z_{i,k} = A_I\left(\frac{\delta_x\overline{\rho}^{\,x,z}_{i-1,k}}{\delta_z\rho_{i,k}}\right)^2\left[\delta_z T_{i,k} - \left(\frac{\delta_z\rho_{i,k}}{\delta_x\overline{\rho}^{\,x,z}_{i-1,k}}\right)\delta_x\overline{T}^{\,x,z}_{i-1,k}\right]. \tag{10}$$

The notation is standard for the MOM 2 model (Pacanowski 1996) and is defined in Fig. 3. In order to define the diffusive fluxes consistently on the model's grid, the x flux must be placed at the east face of T-cell (i, k) and the z flux at the bottom of this same cell. For off-diagonal terms, a double spatial average $\overline{(\,)}^{x,z}$ brings the z-derivative term appearing in the x flux onto the east face of a T-cell, and the x-derivative term appearing in the z flux onto the bottom face of the T-cell. The grid stencil used for computing the x component of the flux is given in Fig. 4. Similar six point stencils are used for computing the other components of the flux.

The practical difficulty encountered when discretizing isoneutral diffusion in the z models concerns how to handle the off-diagonal terms. This issue is not unique to the B grid used in the GFDL model since the physical process (tracer diffusion) concerns only tracer points and tracer cells. The discretization given by C87 considers tracer gradients and the corresponding isoneutral slopes to be independent, each of which needs to be spatially averaged in order to bring it onto the relevant face of the T-cell. In addition, the horizontal and vertical density gradients appearing in the isoneutral slopes are handled independently of one another. Therefore, the

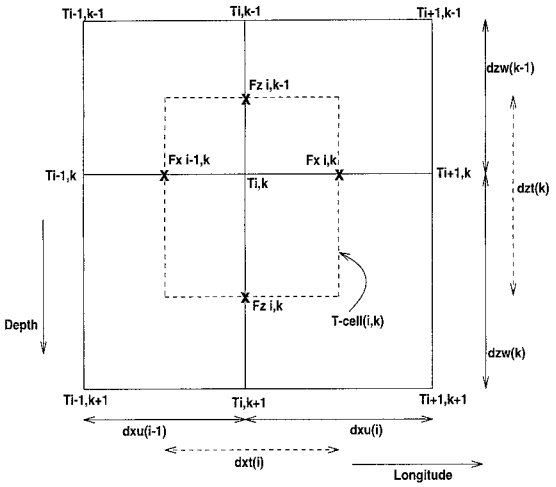

FIG. 3. Grid arrangement in the longitudinal-vertical plane for the ocean model. The dashed lines represent the boundary of the *T*-cell, with $T_{i,k}$ at the center. The discretized horizontal component of the diffusive flux F^x is located on the vertical boundaries, and the vertical component F^z on the horizontal boundaries. The grid dimensions are indicated. Note that the *i, k* label is used for the flux F^x located on the east boundary of the tracer cell, whose center is the tracer point $T_{i,k}$, and likewise for the flux F^z at the bottom of the *T*-cell. The difference operators acting on a tracer are given by $\delta_x T_{i,k} = (T_{i+1,k} - T_{i,k})/dxu_i$ and $\delta_z T_{i,k} = (T_{k,i} - T_{k+1,i})/dzw_k$. On a flux, they are $\delta_x F^x_{i,k} = (F^x_{i+1,k} - F^x_{i,k})/dxt_i$ and $\delta_z F^z_{i,k} = (F^z_{i,k} - F^z_{i,k+1})/dzt_k$. The divergence of the diffusive flux yields the diffusion operator $R_{i,k} = -(\delta_x F^x_{i-1,k} + \delta_z F^z_{i,k-1})$ centered at the tracer point $T_{i,k}$.

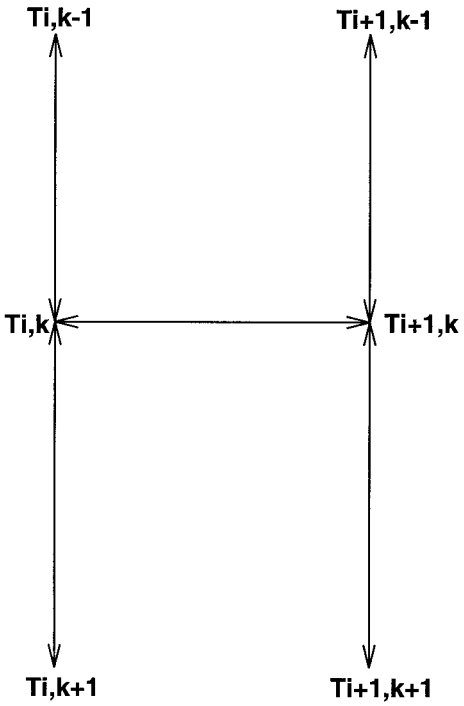

FIG. 4. Grid stencil for computing the *x* component of the flux $F^x_{i,k}$ using the C87 scheme: $F^x_{i,k}$ is located in between the two tracer points $T_{i,k}$ and $T_{i+1,k}$. There are six grid points necessary since, in addition to the horizontal gradient, the average of four vertical gradients is used to define a *z* gradient on the east face of the *T*-cell. Densities are referenced to level k. The arrows denote the pairs of points used for computing the horizontal gradient and the four vertical gradients.

average operator appears individually on only the numerator or denominator. There is nothing fundamental about this particular discretization. Rather, it merely represents a series of convenient choices based on details of the model grid.

To start our critique of the C87 discretization, consider the *z* derivative of the tracer appearing in the *x* component to the flux

$$\delta_z \overline{T}^{x,z}_{i,k-1} = \frac{T_{i,k-1} - T_{i,k+1} + T_{i+1,k-1} - T_{i+1,k+1}}{4dzt_k}. \quad (11)$$

The same form appears for the *z* derivative of the density appearing in the calculation of the slope. It is apparent that the combination of a *z* average and a *z* derivative allows for the presence of $2\Delta z$ structures $T_{i,k-1} = T_{i,k+1}$ and $\rho_{i,k-1} = \rho_{i,k+1}$ for which the discretized *z* derivative on the east face will vanish. Therefore, this wave, or *computational mode*, will be invisible to the *x* component of the computed isoneutral diffusive flux. Likewise for the *z* flux, $2\Delta x$ computational modes exist due to the combination of an *x* average with an *x* derivative. Indeed, for the $2\Delta x$ modes in the density field, the *z* component to the isoneutral flux vanishes identically! This property of the C87 scheme will prove to be crucial to understanding why the scheme is unstable.

In general, when working on the B grid and acting on a single field, such combinations of an average in one direction combined with a derivative in the same direction introduce computational modes. The potential for grid splitting and possible amplification of the modes must be addressed when such modes exist. Namely, if these modes are either completely invisible to the dissipation, or worse if they are amplified, then they can be very harmful to the integrity of the numerical solution. In particular, when appearing in the active tracer fields, and therefore in the density field, certain computational modes can lead to model instabilities, as we now show.

b. Increasing tracer variance

Consider a two-dimensional configuration with density containing the $2\Delta x$ form $\rho_{i+1,k} = \rho_{i-1,k}$. For this field configuration or mode, the previous discussion showed that the *z* component of the discretized isoneutral flux is identically zero. Therefore, the projection of the diffusive flux onto the tracer gradient is entirely zonal and consists of the product

$$F^x \delta_x T = -A_I (\delta_x T)^2 \left(1 - \frac{\delta_z \overline{T}^{x,z} \delta_x \rho}{\delta_x T \delta_z \overline{\rho}^{x,z}}\right). \quad (12)$$

Hence, if slopes of the isoneutrals S_ρ and tracer S_T satisfy the inequality

$$|S_\rho| \equiv \left| \frac{\delta_x \rho}{\delta_z \overline{\rho}^{x,z}} \right| > \left| \frac{\delta_x T}{\delta_z \overline{T}^{x,z}} \right| \equiv |S_T|, \tag{13}$$

then the flux on the east face will be directed up the tracer gradient.

In general, an upgradient component to the diffusive flux vector can be the result of projecting the downgradient isoneutral flux vector onto orthogonal Cartesian axes. As such, it is important to distinguish between particular flux components, which may be upgradient, and the diffusive flux vector, which *should* be downgradient with respect to the tracer gradient (Property I in section 3b). The above discrete configuration is an example of how, in the C87 numerical scheme, an upgradient component in the x direction is not compensated by a downgradient component in the z direction, since the z component identically vanishes. Therefore, the numerically realized diffusive flux vector is upgradient. Furthermore, this upgradient flux can in general be distributed over the extent of the model domain, thus ensuring the increase in tracer variance. The rate of increasing variance is directly proportional to the difference between the slopes, and to the value of the diffusion coefficient. Since the strength of the growth in variance is proportional to A_I, the larger the isoneutral diffusivity, the larger the background horizontal diffusivity needed to suppress the variance increase.

Figure 5 shows a simple realization of this discussion in which a passive tracer field is diffused in the background of a density held constant in time with a $2\Delta y$ structure ($\rho_{i,j+1,k} = \rho_{i,j-1,k}$), constant in longitude, and stratified in depth. The passive tracer is initialized with unity on the surface level and zero below. There is no surface forcing, advection is turned off, and there is zero dianeutral diffusivity. As anticipated, the result is a uniform increase in tracer variance.

c. Active tracer fluxes and the basic problem with C87

Consider a single active tracer for which the isoneutral diffusive flux of this tracer *should* vanish. For example, the x component of the flux in the small angle limit is given by

$$\begin{aligned} F_I^x(\theta) &= -A_I \left(\partial_x \theta - \frac{\partial_x \rho}{\partial_z \rho} \partial_z \theta \right) \\ &= -A_I \left(\partial_x \theta - \frac{\alpha \partial_x \theta}{\alpha \partial_z \theta} \partial_z \theta \right) = 0. \end{aligned} \tag{14}$$

The diffusive flux vanishes because the thermal expansion coefficient α cancels between the numerator and denominator, thus providing an equivalence between the slope $S_\rho = -\partial_x \rho / \partial_z \rho$ of the neutral direction and the slope $S_\theta = -\partial_x \theta / \partial_z \theta$ of the isotherms. Importantly, this result is valid for both linear and nonlinear equations of state.

In C87, the density gradients are computed by referencing each density to the same depth level. For example, the x slope of the neutral direction is computed on the east face of a T-cell through

$$-Sx_{i,k} = \frac{\delta_x \rho_{i-1,k}}{\delta_z \overline{\rho}^{x,z}_{i,k-1}} = \frac{\rho^{(k)}_{i,k} - \rho^{(k)}_{i-1,k}}{dxu_{i-1}} \times \frac{4dzt_k}{\rho^{(k)}_{i,k-1} - \rho^{(k)}_{i,k+1} + \rho^{(k)}_{i+1,k-1} - \rho^{(k)}_{i+1,k+1}}, \tag{15}$$

where the superscript on ρ symbolizes the depth reference level. This expression does not provide for cancellation of the thermal expansion coefficient as occurs in the continuum, unless the equation of state is linear. The reason is that density gradients are not computed explicitly in terms of the expansion coefficients, which would allow for the proper alignment of the temperature isolines (which define the correct neutral directions for this example) and the computed neutral directions. Hence, the computed neutral directions, as determined by the slope calculation, are not aligned with the isolines of locally referenced potential density. The effect produces a spurious flux of locally referenced potential density along the misaligned neutral directions. Note that it is common to check the integrity of isoneutral diffusion schemes by inserting the symbol "ρ" into the discretized expression for the isoneutral flux. Such an assessment would lead to the *incorrect* conclusion that C87 will not flux locally referenced potential density. The present analysis therefore points to the problem with this approach to deducing the self-consistency of the discretization.

The misalignment of the computed neutral directions might be harmless were it not for the following property of seawater. For a stably stratified column, and especially for a stretched grid in the vertical for which level $k + 1$ is further below level k than level $k - 1$ is above it, the C87 computation of the neutral directions can produce an effective expansion coefficient in the numerator that is greater than that in the denominator. This behavior is typical since α is an increasing function of temperature (Gill 1982). Therefore, the discretized slope of the neutral direction has a tendency to be steeper than the slope of temperature

$$|S_\rho| \approx \left| S_\theta \frac{\alpha^{\text{warm}}}{\alpha^{\text{cold}}} \right| > |S_\theta|. \tag{16}$$

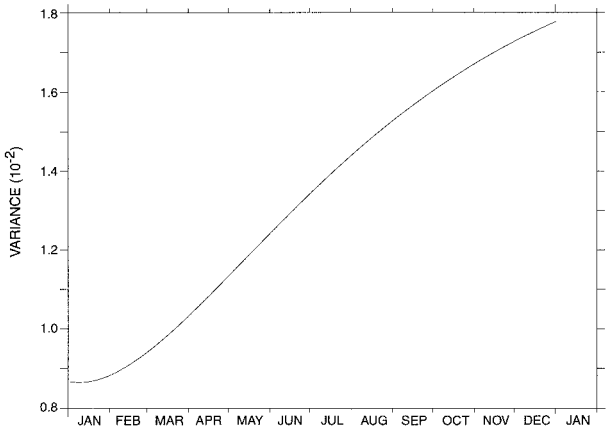

FIG. 5. Variance $[V^{-1} \int \theta^2 \, dV - (V^{-1} \int \theta \, dV)^2]$ for a passive tracer diffused with C87 in the background of a constant density field with a $2\Delta y$ structure ($\rho_{i,j+1,k} = \rho_{i,j-1,k}$) and with zero dianeutral diffusivity. Shown here is the result for a single year of integration.

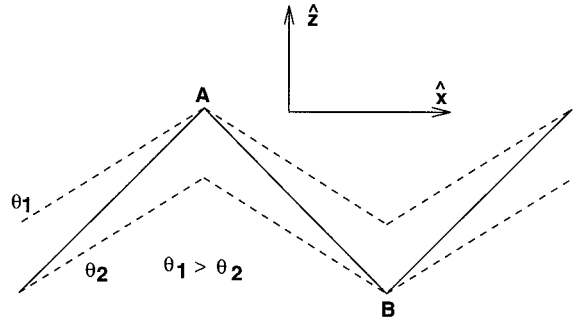

FIG. 6. Unstable profile associated with not properly orienting the isoneutral diffusive fluxes of the active tracers. The dashed lines are temperature isolines, which define the proper neutral directions for the case of an equation of state dependent only on temperature. The solid line is the surface determined by the neutral direction computed using the C87 scheme. For the case of two active tracers, the dashed lines denote isolines of locally referenced potential density. For both cases, a physically correct orientation produces no difference between the solid and dashed lines. *These lines should be parallel.*

This inequality is identical to Eq. (13) for which the C87 scheme will produce upgradient diffusion if the $2\Delta x$ computational mode $\rho_{i+1,k} = \rho_{i-1,k}$ is present.

Figure 6 illustrates an unstable profile corresponding to the above inequality combined with a $2\Delta x$ computational mode. With stable stratification, $\theta_1 > \theta_2$, and so point A is a local minimum with respect to the horizontal direction $\hat{x}$. Upgradient fluxes along $\hat{x}$, with zero compensating vertical flux (recall that the vertical flux vanishes for this $2\Delta x$ mode), will cause point A to cool. The converse occurs at point B. The tendency is to increase the amplitude of the wave thus providing for its instability. The upgradient fluxes induce an unbounded increase in temperature variance so long as inequality (16) is maintained. The situation for two active tracers is similar, only now the dashed lines in Fig. 6 denote isolines of locally referenced potential density. In effect, the nonlinear equation of state, combined with the upgradient fluxes and computational modes, feeds ''energy'' into the numerical solution, which causes the tracer variance to grow unbounded. It is this instability that must be stabilized by the introduction of $A_{\text{back}} \approx A_I$ background horizontal diffusion.

d. A nonlinear instability due to spurious density fluxes

Upgradient diffusive fluxes are generally destabilizing in a numerical diffusion scheme. Furthermore, they do not correspond to the physical process of interest. Therefore, in developing the new isoneutral diffusion scheme, much effort is focused toward correcting this problem (see section 5). As discussed here, however, reducing tracer variance is not sufficient to stabilize the profile shown in Fig. 6, or any wave structure in which the inequality (16) is respected. Stability is realized only if there is exact alignment between the locally referenced potential density surfaces (dashed lines) and the computed neutral directions (solid line).

To see why this profile is generally unstable, consider any profile in which the inequality (16) is satisfied, for which Fig. 6 provides a particular case. Point **A** is a local temperature maximum with respect to the solid line. A downgradient diffusive flux of temperature aligned along the solid line causes point **A** to cool, and the converse occurs at point **B**. This downgradient temperature flux amplifies the initial wave since as the temperature tries to align with the neutral direction, the neutral direction in turn steepens. This steepening is due to the nonlinear nature of isoneutral diffusion, for which the diffusion tensor is a function of the active tracers, which are themselves being diffused. Therefore, the profile is nonlinearly unstable to downgradient diffusion aligned along the spurious neutral directions. The instability will occur regardlesss of the linearity or nonlinearity of the equation of state as long as inequality (16) is satisfied. For two active tracers, the instability generalizes with the dashed lines representing locally referenced potential density isolines.

The instability represented by Fig. 6 is equivalent to the following analytical discussion. Consider a temperature profile that in some local interior region of the ocean takes the form $\theta = \theta_o + bz + B(t) \cos(2\pi x/L)$ and let the directions along which θ is diffused downgradient be determined by $\rho = \rho_o + az + A(t) \cos(2\pi x/L)$. Here ρ surfaces equal neutral surfaces when the slope ratio $S_\rho/S_\theta = (Ab)/(Ba)$ is unity. As suggested by Fig. 6, let ρ be directly coupled to θ, yet let it be an imperfect approximation to locally referenced potential density. The misalignment introduces an unphysical degree of freedom. The zonal wavelength L corresponds to the grid scale over which the slopes are computed in the numerical case, and the constants a, b determine a stable linear vertical stratification. Assume the isoneutral flux-

es are computed with the small slope tensor and a constant diffusivity. The downgradient diffusion of temperature along the ρ surfaces induces the evolution $\partial_t B(t) = A_I(2\pi/L)^2(Ab/a - B)$. A normal diffusive adjustment to the misaligned slopes causes the temperature wave amplitude to change. However, because the neutral directions are directly dependent on temperature, as the temperature wave changes, so does the ρ wave. The amplitude of both waves grows as long as $S_\rho/S_\theta > 1$. Even if the slope misalignment is small, the diffusivity $A_I \approx 10^7$ cm^2 s^{-1} can provide a nontrivial growth. This growth is largest when the wave is short, as for the case when L corresponds to the grid scale in a numerical model. Variance is reduced if the diffusion is downgradient and the boundaries are either insulated or held with a fixed tracer value (see section 5). Therefore, all waves in a model will not grow; only those for which the slope misalignment is relevant. What occurs, therefore, is a reduction in variance with some of the original spectral density of variance being transferred preferentially to the grid scale. This process must saturate since variance is bounded from below (i.e., it is ≥ 0). At saturation, the amount of small-scale spectral density is bounded above by the variance in the initial condition.

To illustrate the importance of balancing the isoneutral fluxes of the active tracers, consider two methods that act to squelch the instability. Prescription A employs the relation $\nabla \ln\rho = -\alpha\nabla\theta + \beta\nabla s$ [see Eq. (6)] for the computation of the neutral direction slope in the x direction

$$Sx_{i,k} = \frac{\alpha^{(i,k)}\delta_x\theta_{i-1,k} - \beta^{(i,k)}\delta_x s_{i-1,k}}{\alpha^{(i,k)}\delta_z\overline{\theta}^{x,z}_{i,k-1} - \beta^{(i,k)}\delta_z\overline{s}^{x,z}_{i,k-1}}, \qquad (17)$$

where $\alpha^{(i,k)}$ and $\beta^{(i,k)}$ are evaluated using the temperature, salinity, and pressure values at grid point (i, k) in *both* the numerator and denominator. Referring to Fig. 6, this prescription corrects the alignment of the computed neutral directions (solid line) so that it is now parallel to the locally referenced potential density isolines (dashed lines). Prescription B employs the C87 discretization of the salinity diffusive flux, including the C87 calculation of the slopes, yet the temperature flux is diagnosed through imposing the constraint $\alpha\mathbf{F}_I(\theta) = \beta\mathbf{F}(s)$ [see Eq. (8)]. This prescription performs an alignment complementary to that done with prescription A. For a single active tracer, both prescriptions trivially stabilize the solution since all active tracer diffusive fluxes vanish. Figures 7a and 7b show the solution for the two active tracer switching experiments, which should be compared to that obtained with C87 in Fig. 2b. Almost all of the unstable behavior has been eliminated.

Ideally, both prescriptions A and B eliminate the numerical instability by zeroing out the isoneutral diffusive flux of locally referenced potential density along the computed neutral directions; that is, they balance the active tracer fluxes and hence provide a self-consistent discretization of isoneutral diffusion. In practice, prescription A performs slightly less satisfactorily than B perhaps due to truncation errors allowing for slight imbalances. Additionally, as discussed in appendix C, in order to solve the diffusion equation in regions of steep isoneutral slopes it is necessary to solve the vertical piece of the diffusion equation implicitly. The effect is to split the vertical flux $F^z(T) = K^{zn}\partial_n T$ into a piece $(K^{zz}\partial_z T)$ solved implicitly and another $(K^{zx}\partial_x T + K^{zy}\partial_y T)$ solved explicitly. The split introduces the possibility of numerical mismatch between the two parts and, so, cannot in general ensure $\alpha F^z_I(\theta) = \beta F^z_I(s)$. This result explains the localization to the far north of problems seen using prescription A since it is in this region that there are stronger vertical isoneutral fluxes, thus allowing more opportunity for truncation errors to produce this mismatch. Closer analysis of the solution produced with prescription B (not shown) also indicates some residual instability in the far north, thus pointing to the generality of this residual instability encountered when splitting the vertical flux.

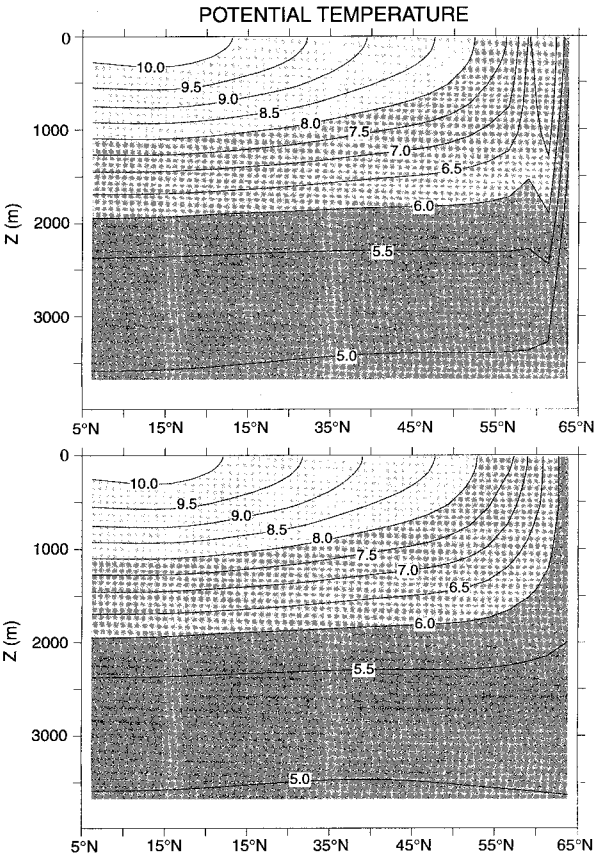

FIG. 7. (a) Upper panel: Meridional-depth snapshot after 300 years from a switching experiment in which the slopes of the neutral directions were computed using the relation $\nabla \ln\rho = -\alpha\nabla\theta + \beta\nabla s$ [see Eqs. (6) and (17)]. (b) Lower panel: Same slice for the case in which the salinity flux is computed just as in C87, yet the temperature flux is diagnosed through $\alpha\mathbf{F}(\theta) = \beta\mathbf{F}(s)$.

5. The new isoneutral diffusion scheme

a. Functional formalism

This section presents the mathematical framework allowing for a systematic incorporation of the downgradient and variance nonincreasing properties of diffusion into a discretization of the isoneutral diffusion operator (Property I of section 3). This framework is based on the property that, for any linear self-adjoint operator, it is possible to associate a functional, whose functional derivative is equal to that operator [Courant and Hilbert (1953); see also Goloviznin et al. (1977), Tishkin et al. (1979), Korshiya et al. (1980) for examples similar to the following]. The functional corresponding to the diffusion operator $R(T)$ is given by

$$\mathcal{F} = -\frac{1}{2}\int d\mathbf{x}\ \partial_m T K^{mn}\partial_n T = \frac{1}{2}\int d\mathbf{x}\ \boldsymbol{\nabla} T\cdot\mathbf{F}. \quad (18)$$

Since the diffusion tensor is symmetric and positive semidefinite at every point in the ocean, the functional defined over any arbitrary volume is negative semidefinite ($\mathcal{F} \leq 0$). Correspondingly, the downgradient property ($\boldsymbol{\nabla} T\cdot\mathbf{F} \leq 0$) of diffusive mixing is equivalent to a negative semidefinite functional. This is an important equivalence that will hold, within a finite volume interpretation (clarified in the subsequent development), in the discrete case as well.

The total ocean is an important special volume of interest for defining the functional. Assuming no sources and using insulating boundaries,

$$\partial_t \int d\mathbf{x}\ T^2 = 4\mathcal{F} \leq 0, \quad (19)$$

where the tracer variance given by Eq. (4) was employed. Hence, for this particular volume, $4\mathcal{F}$ can be interpreted as the sink of tracer variance arising from the effects of downgradient diffusion.

To motivate the form for the functional derivative relating $\mathcal{F}$ to $R(T)$, consider an infinitesimal perturbation, or *variation*, of the tracer field $T \to T + \delta T$. This tracer variation induces a variation in the functional, which is given by [see Courant and Hilbert (1953) for more discussion]

$$\begin{aligned}\delta\mathcal{F} &= \int d\mathbf{x}\left[\frac{\delta\mathcal{L}}{\delta T}\delta T + \partial_m\left(\frac{\delta\mathcal{L}}{\delta T_m}\delta T\right) - \partial_m\left(\frac{\delta\mathcal{L}}{\delta T_m}\right)\delta T\right]\\ &= \int d\mathbf{x}\left[\frac{\delta\mathcal{L}}{\partial T} - \partial_m\left(\frac{\delta\mathcal{L}}{\delta T_m}\right)\right]\delta T, \quad (20)\end{aligned}$$

where $2\mathcal{L} = \boldsymbol{\nabla} T\cdot\mathbf{F}(T) = -\partial_m T K^{mn}\partial_n T$ is a negative semidefinite quadratic form, and $T_m = \partial_m T$. Dropping the total divergence term $\partial_m(\delta T\delta\mathcal{L}/\delta T_m)$ requires the application of either one of the natural boundary conditions: 1) $\delta T(\mathbf{x}, t) = 0$ on the boundaries of the domain or 2) $\hat{N}^m\delta\mathcal{L}/\delta T_m = 0$, where $\hat{N}$ is the boundary's normal. Here $\delta T(\mathbf{x}, t) = 0$ corresponds to taking a Dirichlet condition for the tracer (tracer specified on the boundaries), whereas $\hat{N}^m\delta\mathcal{L}/\delta T_m = 0$ corresponds to a Neumann or no-flux boundary condition, where $\delta\mathcal{L}/\delta T_m$ are the components to a generalized flux. These results are valid for any functional $\mathcal{F}$, which can be written as $\int d\mathbf{x}\ \mathcal{L}$. Specializing now to the case of the diffusion functional with a diffusion tensor independent of the tracer—that is, linear diffusion of passive tracers—yields $\delta\mathcal{L}/\delta T = 0$ and $\delta\mathcal{L}/\delta T_m = -K^{mn}\partial_n T = F^m(T)$. Therefore, using $\delta T(\mathbf{x})/\delta T(\mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$, the desired relation between the functional and the diffusion operator is given by the compact expression

$$\frac{\delta\mathcal{F}}{\delta T} = R(T). \quad (21)$$

This continuum relation has a natural finite volume generalization

$$\frac{\delta\mathcal{F}}{\delta T_{i,j,k}} = \int_{V_{i,j,k}} \boldsymbol{\nabla}\cdot\mathbf{F}(T)\ dV_{i,j,k}, \quad (22)$$

where $V_{i,j,k}$ is a finite cell volume associated with tracer $T_{i,j,k}$. It is over this finite volume that the discretized operator possesses the downgradient properties. We elaborate on this important point in the subsequent development and in appendix D.

As previously mentioned, the functional formalism is strictly useful for the case of linear diffusion of the passive tracers. Since the active tracers are diffused with the same operator as the passive tracers, this restriction is of no substantial limitation. Yet, in order to prevent the nonlinear computational instability described in section 4, it is crucial to add to the functional formalism the constraint that the slopes be computed so that the active tracers closely approximate the self-consistency or balance condition on the isoneutral diffusive fluxes $\alpha\mathbf{F}_I(\theta) = \beta\mathbf{F}_I(s)$. Otherwise, even though the scheme will not increase tracer variance, it will be subject to grid noise.

In summary, for any symmetric and positive semidefinite diffusion tensor, the functional formalism allows for a straightforward incorporation of the variance reducing properties implied by such a mixing tensor to be readily built into a discretization scheme. The procedure is to first discretize the functional $\mathcal{F}$ and then to take the discrete version of the functional derivative given by Eq. (22). The power of the formalism is that for any consistent discretization of $\mathcal{F}$, the corresponding discretization of the diffusion operator inherits the desired downgradient properties over the corresponding finite volume. Additionally, this result means that the discrete diffusion operator $R(T)$ will not increase the tracer variance, the eigenvalues of $R(T)$ will all be positive, and the scheme is ensured to be at least conditionally stable in a linear sense. It follows that because the C87 diffusion operator can increase tracer variance, it does not correspond to a semidefinite functional.

The expressions for the case of isoneutral/dianeutral

diffusion will prove necessary for the subsequent discussion. Writing the components of the diffusion tensor given in Eq. (1) as $K^{mn} = (A_I - A_D)(\delta^{mn} - \hat{\gamma}^m\hat{\gamma}^n) + A_D\delta^{mn}$ brings the functional for isoneutral/dianeutral diffusion into the convenient form

$$\mathcal{F} = -\frac{1}{2}\int d\mathbf{x}\ (A_I - A_D)|\hat{\gamma}\times\boldsymbol{\nabla}T|^2 - \frac{1}{2}\int d\mathbf{x}\ A_D|\boldsymbol{\nabla}T|^2$$
$$= -\frac{1}{2}\int d\mathbf{x}\ \boldsymbol{\nabla}T\cdot[(A_I - A_D)(\hat{\gamma}\times\boldsymbol{\nabla}T)\times\hat{\gamma} + A_D\boldsymbol{\nabla}T], \tag{23}$$

which allows for the identification of the diffusive flux

$$\mathbf{F}(T) = -(A_I - A_D)(\hat{\gamma}\times\boldsymbol{\nabla}T)\times\hat{\gamma} - A_D\boldsymbol{\nabla}T. \tag{24}$$

In the small slope approximation, the functional is given by

$$\mathcal{F}^{\text{small}}$$
$$= -\frac{1}{2}\int d\mathbf{x}A_I(\partial_x T + S_x\partial_z T)^2 + (\partial_y T + S_y\partial_z T)^2$$
$$-\frac{1}{2}\int d\mathbf{x}A_D(\partial_z T)^2$$
$$= -\frac{1}{2}\int d\mathbf{x}\boldsymbol{\nabla}T\cdot A_I[\hat{x}(\partial_x T + S_x\partial_z T) + \hat{y}(\partial_y T + S_y\partial_z T)$$
$$+\hat{z}(S_x\partial_x T + S_y\partial_y T + (\epsilon + S^2)\partial_z T)], \tag{25}$$

and the corresponding small angle flux components are

$$\mathbf{F}_h(T) = -A_I(\boldsymbol{\nabla}_h + \mathbf{S}\partial_z)T \tag{26}$$

$$F^z(T) = \mathbf{S}\cdot\mathbf{F}_h(T) - A_D\partial_z T, \tag{27}$$

where $\mathbf{F}_h = (F^x, F^y, 0)$ is the horizontal diffusive flux vector, $\mathbf{S}$ is the isoneutral slope vector, $\epsilon = A_D/A_I$, and $\boldsymbol{\nabla}_h = (\partial_x, \partial_y, 0)$ is the horizontal gradient operator.

b. Discretizing the functional

The functional given by either Eq. (23) for the full tensor, or (25) for the small angle approximation, consists of quadratic terms that take the form $(\partial_m T\partial_n\rho - \partial_n T\partial_m\rho)^2$, with $m \neq n$. Their discretization defines grid stencils in the corresponding two-dimensional (m, n) plane. This observation motivates a discretization of the functional where its different pieces are discretized separately within their respective two-dimensional plane. The exception to this two-dimensionality arises from the term $|\boldsymbol{\nabla}\rho|^{-2}$, which occurs due to the two factors of the dianeutral unit vector $\hat{\gamma} = \boldsymbol{\nabla}\rho|\boldsymbol{\nabla}\rho|^{-1}$ appearing in the full tensor. This gradient contains all three differential operators, which means that in discretizing the full tensor we must consider extending the stencil to points off a given plane. How we handle this detail will be discussed in appendix E.

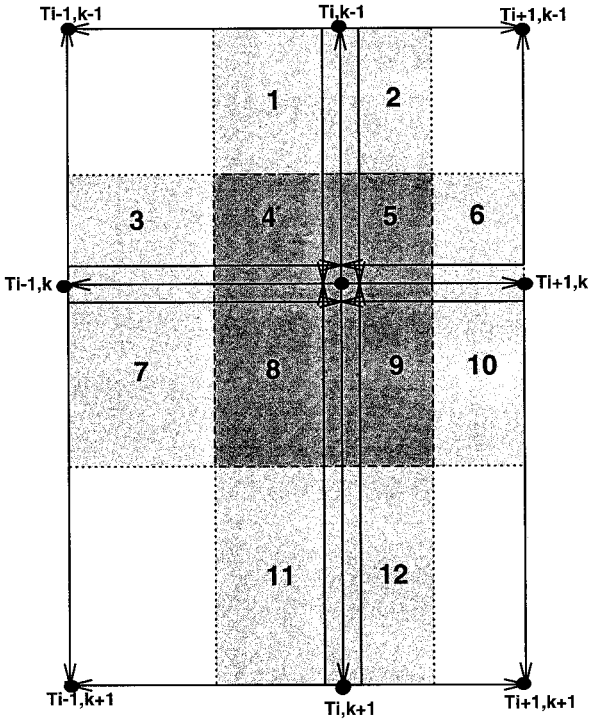

FIG. 8. The grid stencil for discretizing the $\mathcal{F}_{i,k}^{(x-z)}$ piece of the functional. There are a total of 12 triads, with 12 corresponding quarter-cells (shaded regions), to which the central point $T_{i,k}$ contributes. The slightly darker shading is used for the four central quarter-cells (4, 5, 8, 9), which make up the T-cell whose center is the tracer point $T_{i,k}$. Each of the 12 triads is indicated by a pair of lines with arrows on the end extending outward from the vertex of the triad. Four of the triads have $T_{i,k}$ as a vertex.

In the longitudinal–depth plane, consider the term $(\partial_x T\partial_z\rho - \partial_z T\partial_x\rho)^2$. A simple discretization is to use nearest neighbor tracer grid points for constructing the discrete differential operators. Since the model grid is staggered, such a choice employs second-order accurate difference operators in the functional, and will likewise be the case for the corresponding diffusion operator. Higher order in accuracy schemes can be derived by extending the stencil outward to include more points. Using the nearest neighbors, a typical component of the discretizaton of $(\partial_x T\partial_z\rho - \partial_z T\partial_x\rho)^2$ consists of *triads* of tracer and density values. For example, one such triad contains contributions from the grid points $(i - 1, k)$, (i, k), and $(i, k - 1)$, where the corresponding term in the functional is $(\delta_x T_{i-1,k}\delta_z\rho_{i,k-1} - \delta_z T_{i,k-1}\delta_x\rho_{i-1,k})^2$. As seen in Fig. 8, this triad is one of 12 that contain contributions from the central tracer point $T_{i,k}$.

The triads partition the area of the longitudinal–depth plane into a series of quarter-cells (see Fig. 8). There is a single unique quarter-cell for each of the triads. For example, the triad $(i - 1, k)$, (i, k), and $(i, k - 1)$ is associated with the quarter-cell 4, and triad (i, k), $(i, k + 1)$, and $(i - 1, k + 1)$ with quarter-cell 11. The areas (actually, the volume when considering the third dimension) of each quarter-cell define the volume element

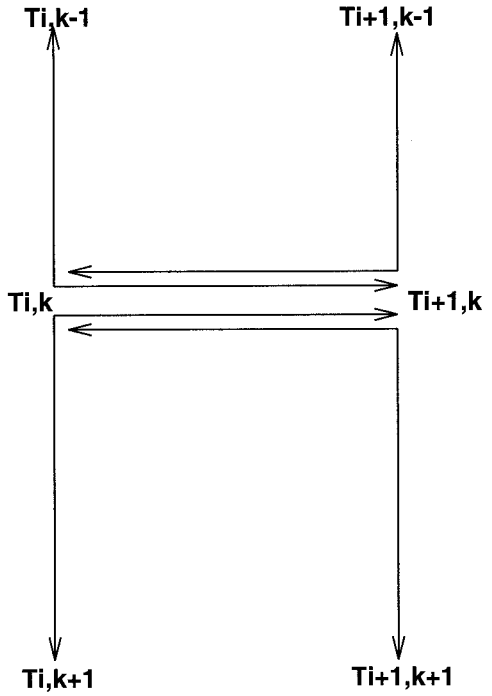

FIG. 9. The grid stencil for the x component of the diffusive flux as computed using the new scheme. Four density triads are drawn, with reference points taken at their corners located at the tracer points $T_{i,k}$ and $T_{i+1,k}$. The flux component $F^x_{i,k}$ is defined in between these two tracer points, which is also the center of the east face of the T-cell $T_{i,k}$.

associated with the triad to be used in discretizing the functional. Therefore, the discretization of the functional corresponding to that contribution from the x–z plane is given by

$$\mathcal{F}^{(x-z)} = -\frac{1}{2}\sum_{i,k}\sum_{n=1}^{12} A(n)V(n) \times \frac{(\partial_x T^{(n)}\partial_z \rho^{(n)} - \partial_z T^{(n)}\partial_x \rho^{(n)})^2}{|\nabla \rho^{(n)}|^2}, \quad (28)$$

where $V(n)$ is the volume of the nth quarter cell and $A(n)$ is the corresponding nonnegative diffusion coefficient. The superscript (n) on the tracer and density refers to the particular finite-difference discretization of the gradient for the nth quarter-cell. For example, in quarter-cell 1 shown in Fig. 8, $\partial_x T^{(1)}$ symbolizes the discrete derivative $\delta_x T_{i-1,k-1}$, and $\partial_z T^{(1)} = \delta_z T_{i,k-1}$. Equation (28) means that the discretization of the functional $\mathcal{F}^{(x-z)} = \Sigma_{i,k}\, \mathcal{F}^{(x-z)}_{i,k}$ is built by summing over the tracer points $T_{i,k}$ on the lattice and, for each tracer point, summing over the 12 triads/quarter-cells that contain some contribution from the tracer point $T_{i,k}$. For purposes of discretizing the diffusion operator $R(T)_{i,k}$ at a particular tracer point, it is only necessary to consider the terms appearing inside the $\Sigma_{n=1}^{12}$ sum over the triads since the derivative $\partial \mathcal{F}^{(x-z)}/\partial T_{i,k}$ will eliminate the $\Sigma_{i,k}$ sum over all tracer points. Note that each of the 12 contributions to the functional $\mathcal{F}^{(x-z)}_{i,k}$ vanishes individually when the tracer T is replaced by locally referenced potential density ρ. This is an important property that must be respected when the detailed discretization of the density derivatives is specified (next subsection).

These details about quarter-cells, triads, and their respective volumes are important since they make precise the notion about downgradient diffusion on the lattice. Namely, these 12 quarter-cells define the finite size volume, mentioned in the discussion of Eq. (22), for which the new scheme provides downgradient fluxes of the tracer. Since the finite volume encompasses many grid cells, the locally defined diffusive fluxes discretized on the faces of the T-cell (next subsection) will not generally satisfy the downgradient property individually. In other words, the new scheme is not ''positive definite'' for each cell, rather it is positive-definite only over the semilocal finite volume defined by the 12 quarter-cells. This is an important qualification that must be kept in mind when interpreting results from this scheme. Note, however, that the new scheme will not increase tracer variance since variance reduction depends only on the negative semidefiniteness of the globally defined functional. The finite volume interpretation of downgradient diffusion in the new scheme is made more mathematically precise in appendix D.

c. Flux discretization with density triads

After discretizing the functional, it is necessary to take its derivative with respect to the tracer in order to obtain the discrete diffusion operator [Eq. (22)]

$$\frac{1}{V_{T_{i,k}}}\frac{\delta \mathcal{F}^{(x-z)}}{\delta T_{i,k}} = R(T)_{i,k}, \quad (29)$$

where $V_{T_{i,k}}$ is the volume of the T-cell whose center is the tracer point $T_{i,k}$. The effect from this derivative is to break or separate the tracer triads. However, it preserves the integrity of the density triads. The preservation of the density triads suggests a heuristic approach to directly discretizing the individual diffusive fluxes on the cell faces. Namely, construct the pieces of this flux using density triads as fundamental building blocks. The result of this derivation, presented in the following, is identical to that obtained when the functional derivative is directly computed and the resulting terms are combined into the divergence of a tracer flux across the T-cells [details placed in an appendix to Pacanowski (1996)]. One thing to note immediately is that the use of unbroken density triads, when weighted by their respective tracer gradients (described below), removes the computational modes. The reason is that there is no splitting of the grid, which is a characteristic of the computational modes. As discussed in section 4, density computational modes are fundamentally related to the problems with the C87 scheme. Therefore,

by eliminating these modes, we already see the crucial role the density triads play in the new scheme.

For the purpose of describing the discretization, it is sufficient to discretize the diffusion flux from the small angle tensor since the basic ideas are the same for both the small angle and full slope diffusion tensors. It is also sufficient to continue considering the two-dimensional longitudinal–depth geometry. We start by discretizing the x component $F^x(T) = -A_I(\partial_x T + S_x \partial_z T)$ on the east face of the T-cell (i, k) (see Fig. 3). Recall that in the C87 discretization discussed in section 4, the central practical difficulty is how to handle the off-diagonal term $A_I S_x \partial_z T = -A_I(\partial_x \rho / \partial_z \rho)\partial_z T$. When employing density triads as fundamental units, notice that there are four triads surrounding the east face of each T-cell. Each defines a discretization of the isoneutral slope S_x, so each has an associated diffusion coefficient A_I. The diffusion coefficients are selected according to one of the chosen slope constraints discussed in appendix C in order to satisfy the requirements of linear stability for steep isoneutral sloped regions. In addition, for each triad it is necessary to choose a reference point for calculating the locally referenced potential densities. This point can be chosen anywhere as long as there is a unique point for each of the triads. Since the equation of state is already computed for each tracer point, it is convenient to choose the corner of the triad as the reference point. By constructing $A_I S_x \partial_z T$ as an average over these four triads, each multiplied by their respective vertical tracer gradient, the discretized diffusive flux component F^x is correctly placed at the east face of the T-cell. Additionally, in order to account for nonuniform grids, it is important to weight each term in the average by its associated vertical grid spacing. Figure 9 summarizes this discussion by showing the stencil for the new scheme. Importantly, it requires the *same* six densities and tracers as for the C87 scheme (see the C87 stencil in Fig. 4), so the grid ''footprint'' is identical to that of C87.

Putting the pieces together yields the discretized x component of the small angle tensor isoneutral diffusion flux:

$$-F^{x\,\mathrm{small}}_{i,k}(T) = \frac{1}{4dzt_k} \sum_{kr=0}^{1} dzw_{k-1+kr} \sum_{ip=0}^{1} A^{(i+ip,k)}_{(i,k|i+ip,k-1+kr)} (\delta_x T_{i,k} + Sx^{(i+ip,k)}_{(i,k|i+ip,k-1+kr)} \delta_z T_{i+ip,k-1+kr}), \tag{30}$$

(refer to Fig. 3 for definitions of the grid spacing factors *dzt* and *dzw* and difference operators). The sum over *ip* and *kr* represents the sum over the four triads shown in Fig. 9. The isoneutral slope is computed through the relation

$$Sx^{(i1,k1)}_{(i2,k2|i3,k3)} = -\left(\frac{\delta_x \rho^{(i1,k1)}_{i2,k2}}{\delta_z \rho^{(i1,k1)}_{i3,k3}}\right). \tag{31}$$

The corresponding diffusion coefficient $A^{(i1,k1)}_{(i2,k2|i3,k3)}$ is set according to the relevant slope criteria discussed in appendix C. The superscripts $(i1, k1)$ refer to the corner point of the triad, which is used for determining the reference points in calculating the densities. It is important to note the use of the *same* reference point for density gradients in both the numerator and denominator. Accordingly, the density gradients are computed as

$$\delta_m \rho^{(i1,k1)}_{i2,k2} = (\rho_\theta)_{i1,k1} \delta_m \theta_{i2,k2} + (\rho_s)_{i1,k1} \delta_m s_{i2,k2}, \tag{32}$$

where the labels on the density partial derivatives indicate the grid point for which the temperature, salinity, and pressure are used in evaluating these terms. In the model, the coefficients $\rho_\theta = (\partial\rho/\partial\theta)$ and $\rho_s = (\partial\rho/\partial s)$ are computed by analytically differentiating the cubic polynomial used for approximating the UNESCO equation of state (Bryan and Cox 1972), thus producing quadratic expressions which are diagnosed.

It is important to highlight the direct computation of the expansion coefficients ρ_θ and ρ_s for the purposes of computing the density gradients and, hence, for computing the isoneutral slopes. As discussed in section 4, a proper alignment of the computed neutral directions will eliminate spurious and unstable fluxing of locally referenced potential density. Otherwise, the new scheme, even though it cannot increase tracer variance, would be exposed to a form of the nonlinear instability described in section 4d and hence subject to grid noise. A viable alternative to computing the density gradients as given in Eq. (32) is to directly diagnose one of the active tracer fluxes from the other (i.e., prescription B discussed in section 4d). Such a scheme was tested, but it did not alter the solution substantially. The choice of computing the slopes in terms of the expansion coefficients represents a decision to emulate as closely as possible the kinematics discussed in section 3, and in particular Eq. (7) for the dianeutral unit vector. The direct calculation of ρ_θ and ρ_s is therefore a fundamental and novel aspect of the new scheme.

The diagonal piece of the flux consists of the horizontal gradient of the tracer weighted by the four diffusion coefficients, which correspond to the four triads used for constructing the off-diagonal term. Should the slopes for the four triads all lie within the stability range and therefore not require scaling, then the sum collapses and the diagonal piece becomes $A_I \delta_x T_{i,k}$, where A_I is the unscaled diffusion coefficient.

The resulting diffusive flux can be thought of as an average of four ''subfluxes'' associated with each of the four triads. Each of the subfluxes correctly vanishes individually when locally referenced potential density is substituted for the tracer. In this manner, the new scheme provides for a completely symmetric or democratic sampling of the diffusive fluxes associated with each of the four vertical tracer gradients surrounding the east face of the *T*-cell. Should any one of these subfluxes lie adjacent to a solid boundary, its contribution to the average over the four triads is eliminated. Additionally, the vertical tracer gradient and the corresponding density triad are fundamentally coupled. Therefore, it is not possible to identify a discretization of the K^{xz} off-diagonal diffusion tensor component independently of the particular tracer, whereas it was possible to do so with the C87 scheme. The resulting discretization of the diffusion operator recovers the traditional five-point Laplacian in the limit of flat isoneutral directions. This operator has well-known stability properties; that is, it fluxes tracer downgradient and it has no computational modes.

The *z* component of the small angle tensor flux is given by

$$-F^{\text{zsmall}}_{i,k}(T) = \frac{1}{4dxt_i}\sum_{ip=0}^{1} dxu_{i-1+ip}\sum_{kr=0}^{1} A^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)}Sx^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)}(Sx^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)}\delta_z T_{i,k} + \delta_x T_{i-1+ip,k+kr}). \quad (33)$$

The construction of the *z* component is based on the same arguments as the *x* component, through the use of the density triads. The new scheme does allow for the diagonal terms in the diffusion tensor to be identified independently from the tracers. Most importantly,

$$K^{\text{zzsmall}}_{i,k} = \frac{1}{4dxt_i}\sum_{ip=0}^{1} dxu_{i-1+ip}\sum_{kr=0}^{1} A^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)}(Sx^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)})^2, \quad (34)$$

which means the implicit in time algorithm for solving the vertical diffusion equation is identical to C87.

6. Numerical tests of the new diffusion scheme

As described in section 4 and illustrated in Fig. 7, the switching experiments to isoneutral diffusion are greatly stabilized upon providing a self-consistent balance of the isoneutral diffusive fluxes of the active tracers so that $\alpha\mathbf{F}_I(\theta) = \beta\mathbf{F}_I(s)$. This stabilization was achieved independently of discretizing with the density triads. Analogous tests with the new scheme show similar results (not shown). Furthermore, they indicate that it is possible to remove all explicit dianeutral diffusivity when solving the isoneutral diffusion problem, even in the especially difficult case of strongly evolving active tracers.[2] The ability to do so in a *z*-coordinate model opens up interesting possibilities for exploring mixing effects due to the nonlinear equation of state (see appendix B), which have heretofore been swamped by the horizontal background diffusion. This section briefly illustrates further aspects of the new scheme. All experiments are conducted with zero horizontal background diffusivity.

a. Effects due to the triad discretization

Consider the effects in a two-dimensional model with a $2\Delta x$ profile in salinity yet vertical stratification in temperature and allow the equation of state to be linear.[3] The $2\Delta x$ salinity structure induces a $2\Delta x$ density structure. With this structure, C87 produces a zero vertical flux [Eq. (12)] and an upgradient horizontal flux, which means that the variance created in the *x* direction is not removed in the *z* direction. The effect is to move water parcels along neutral directions in θ–*s* space until there is a 1–1 relationship between θ and ρ. The initial spread in salinity is replaced by a spread in temperature, with the densest water colder than any water initially in the domain, and the lightest water warmer than any initially in the domain. The final range is roughly $\alpha\Delta\theta \approx \beta\Delta s$. In cold water, α is small so that small horizontal vari-

[2] Dispersion errors from advective fluxes may qualify this statement.

[3] This is the only model result in this paper not computed with MOM 2.

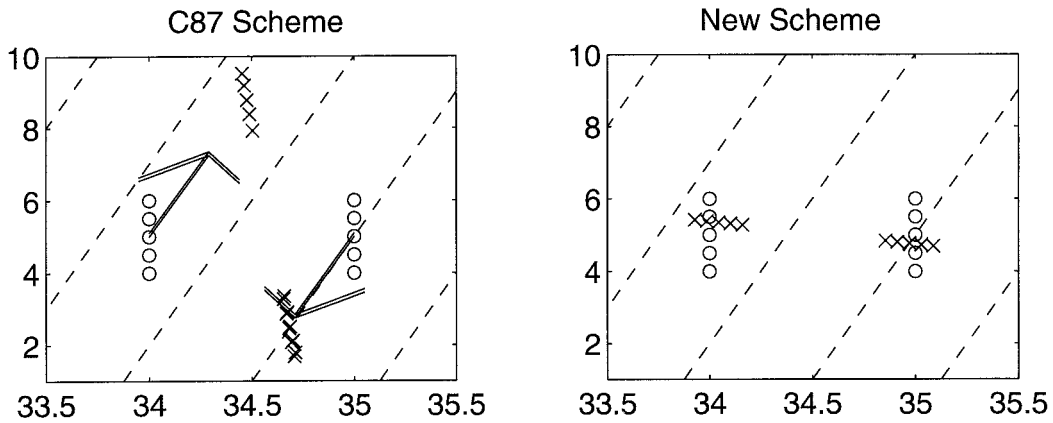

FIG. 10. Results in the temperature (vertical)–salinity (horizontal) diagram of diffusing an initially $2\Delta x$ salinity profile and vertically stratified temperature profile. The initial condition (○) and 1-yr (×) results are shown for a two-dimensional diffusion model with horizontal spacing of 100 km and vertical spacing of 100 m: (a) Left panel: C87 scheme. Note that the point cloud on the right divides into two point clouds because points on the boundaries feel one column, while those in the interior feel two. (b) Right panel: New diffusion scheme.

ations in salinity will drive large horizontal variations in temperature, thus creating very cold water. By contrast, the new scheme has no computational mode and so diffuses temperature and salinity in both the vertical and horizontal. The result is the more reasonable temperature–salinity distribution shown in Fig. 10b. All the temperatures now lie within the initial range. There are a few points, however, where the salinity is slightly outside the initial range. These unphysical results may be attributed to small amounts of upgradient diffusion occurring for the individual cell faces (see discussion of the finite volume interpretation of downgradient fluxes in section 5 and appendix D).

Now reconsider the passive tracer experiment shown in Fig. 5 using the new scheme. The tracer variance for this experiment, shown in Fig. 11a, reduces immediately. Figure 11b shows a meridional slice at the surface in the middle of the basin for both the C87 and new results. This figure shows unphysical values (those values outside the range [0, 1]) for the tracer associated with the increasing variance in the C87 scheme. The new scheme produces physically realistic values.

b. Diffusion and GM90 advection

The unsuccessful relaxation experiment with GM90 advection and C87 diffusion shown in Fig. 2d, which used 0.5 $cm^2\ s^{-1}$ dianeutral diffusivity, was cleanly simulated with the new diffusion scheme (not shown). A more difficult problem to solve with the C87 scheme is the case with identically zero dianeutral diffusivity. Such an experiment is possible with the new diffusion scheme (as seen in Fig. 12). The solution relaxes toward a horizontally uniform state, as expected since the GM90 scheme acts to reduce the available potential energy of the system. We found a similar smooth relaxation when allowing the Eulerian velocity to act on the tracers through the addition of a centered difference advective flux, in addition to the GM90 advection and isoneutral diffusion (not shown).

c. Computational timing requirements

The computational requirements of the new diffusion scheme are more than the C87 scheme. With the idealized sector model used in this study, the new diffusion scheme using the small angle tensor took roughly 10%–20% more computational time than the MOM 2 implementation of the C87 scheme. These numbers represent the time for the model as a whole. This result is consistent with comparisons made in a realistic coarse-resolution global ocean model (not shown). The discretization provided in appendix E for the full isoneutral diffusion tensor took roughly six times longer than the new small angle tensor. If further justification is provided for employing the full tensor (see appendix C for more discussion), then some sort of approximation to the scheme provided in appendix E should be considered in order to reduce the large time requirements for using this tensor. In general, for the small tensor, the 10%–20% increased time requirements over the C87 scheme appear modest considering the substantially improved numerical representation of the isoneutral diffusion process.

7. Summary and conclusions

The main purpose of this paper was to present the physical and numerical properties that are necessary in order to stably and accurately realize isoneutral/dianeutral diffusion of tracers in *z*-coordinate ocean models. Model tests with the C87 (Cox 1987) diffusion scheme were seen to result in unphysical and unstable solutions. In addition to showing problems when diffusing with

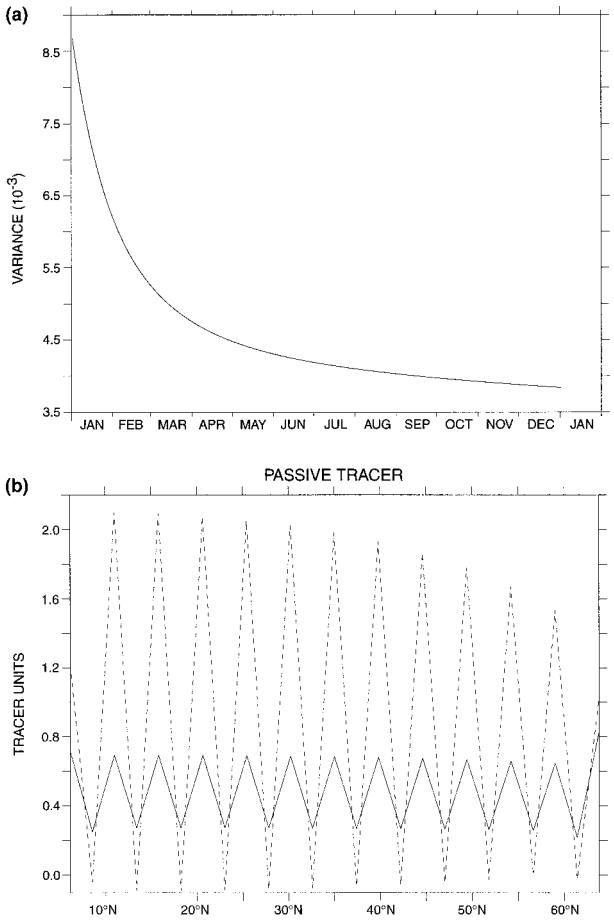

FIG. 11. (a) Passive tracer variance for an experiment in which the density field is held constant with a $2\Delta y$ structure ($\rho_{i,j+1,k} = \rho_{i,j-1,k}$), there is zero dianeutral diffusivity, and the isoneutral flux is computed using the new diffusion scheme. Shown here is the result for a single year of integration. (b) Surface values in the middle of the basin for the passive tracer experiment corresponding to (a) at the end of one year. Isoneutral diffusion will introduce the $2\Delta y$ structure into the passive tracer field since this profile is fixed in the density field. However, diffusion ideally should not introduce unphysical values, which for this case are those with values outside the range [0, 1]. Shown here are the results from the C87 scheme (dashed line) and the new diffusion scheme (solid line). Subsequent integration did not change the solutions substantially.

the C87 scheme, problems occured when C87 diffusion was run with either FCT Eulerian advection or the GM90 eddy-induced advective transport. Although exceptions to these negative results may occur in other flow regimes, the goal of our study was to provide a diffusion scheme that performs well under a broad spectrum of flows, including those undergoing large perturbations such as might occur in realistic climate variability and climate change experiments. Therefore, our evaluation of the C87 scheme leads us to conclude that it does not generally provide a numerically sound or physically realistic simulation of isoneutral diffusion in z-coordinate ocean models.

After the model tests with the C87 scheme, some basic kinematical properties of isoneutral/dianeutral diffusion were discussed. From this analysis, two central properties desired from a numerical diffusion scheme were identified: (I) It should flux tracers downgradient, which implies that it will not increase tracer variance. (II) It should provide for a proper orientation of the neutral directions so that there is a balance between the isoneutral fluxes of the active tracers, thus providing for a zero isoneutral diffusive flux of locally referenced potential density. Property II is a self-consistency condition for isoneutral diffusive fluxes. This property distinguishes isoneutral diffusion of active tracers, which is a nonlinear process due to the dependence of the diffusion tensor on the active tracers, from linear diffusion processes such as that for passive tracers. Based on this discussion, the following core problems with the C87 scheme were identified: (i) The presence of computational grid-scale modes and (ii) the ability to increase tracer variance when (a) these computational modes appear in the density field and (b) the equation of state is nonlinear, which results in a computed neutral direction that is steeper than the isolines of locally referenced potential density. The nonlinear equation of state, combined with the upgradient flux associated with the computational mode, feeds the growth in the numerical instability that causes the active tracer variance to grow unbounded.

Besides upgradient fluxes, there is another problem with the unstable density profile associated with the C87 instability (Fig. 6). The problem is fundamentally associated with the misalignment of the computed neutral directions and isolines of locally referenced potential density. A result of the misalignment is that the isoneutral diffusive flux of the active tracers is not constrained to satisfy the balance $\alpha \mathbf{F}_I(\theta) = \beta \mathbf{F}_I(s)$. The imbalance of active tracer fluxes in C87 effectively provides for an extra computational degree of freedom. This extra degree of freedom will manifest as an unstable mode if the computed neutral directions are steeper than the isolines of locally referenced potential density. Any scheme satisfying this misalignment, even if it reduces variance, will produce an accumulation of power at high wavenumbers; that is, grid noise. We have seen this result in tests with the variance reducing scheme in which we purposefully computed the neutral directions as in C87 (not shown). What occurs amounts to an unstable form of self-diffusion, where the scheme attempts to diffuse locally referenced potential density downgradient along spuriously computed neutral directions. The process will saturate since variance reduces, yet variance is nonnegative. Depending on the initial variance, the result could be a substantial amount of power at the grid scale: an effective ripping apart of the solution. The instability depends on the nonlinear nature of active tracer diffusion with the Redi diffusion tensor. The reason is that, if there were no feedback, locally referenced potential density isolines would simply align with the spurious neutral directions, as will occur for a passive

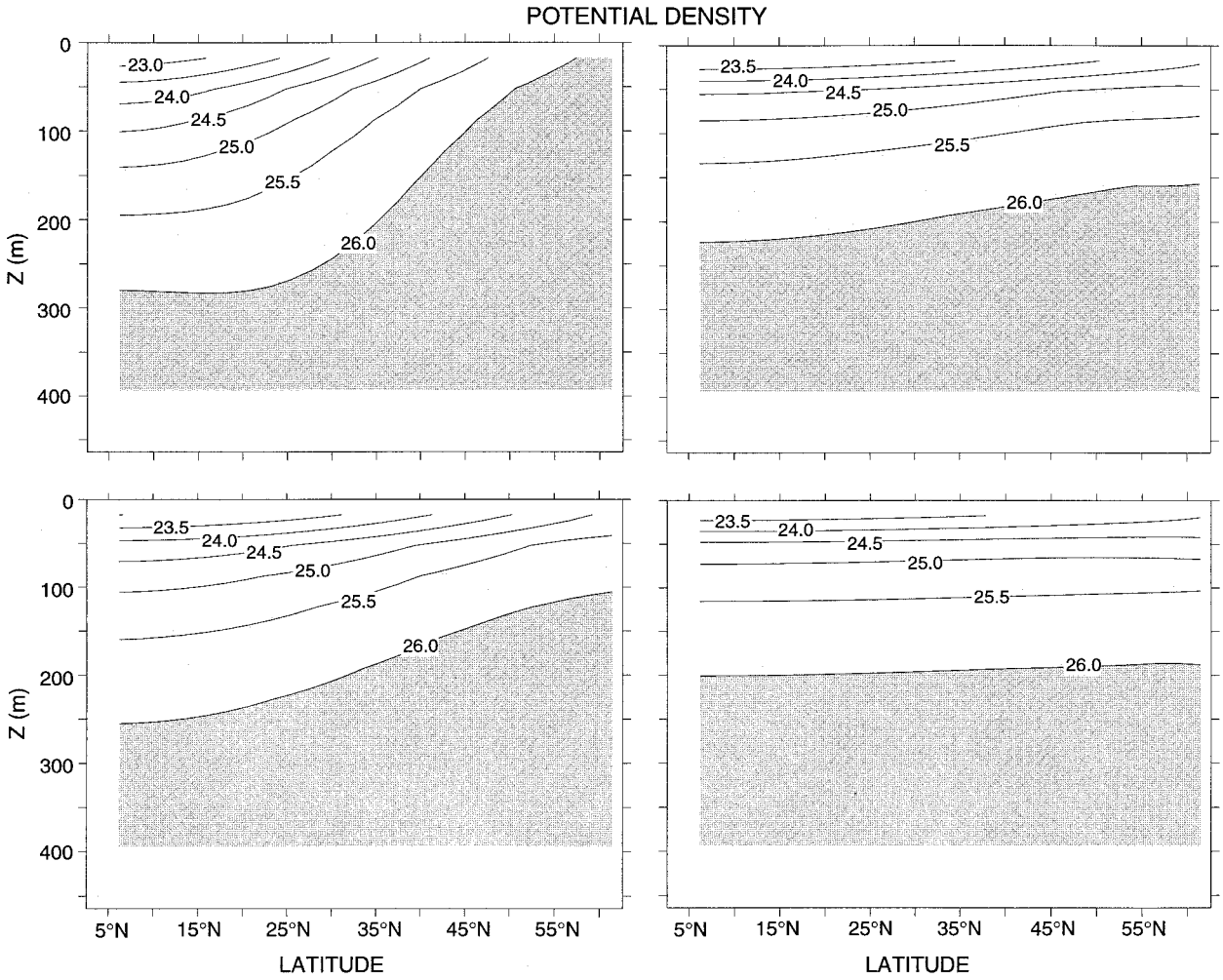

FIG. 12. Potential density, referenced to the surface, shown in the middle of the basin over the upper 400 m of the model. These plots were obtained from a switching experiment in which both the new isoneutral diffusion and GM90 advection act on two active tracers. There is zero dianeutral diffusivity and zero Eulerian advection. The panels show the solution after 20 years (upper-left panel), 40 years (lower-left panel), 100 years (upper-right panel), and 200 years (lower-right panel) of integration.

tracer in the background of a static density field. Instead, as the locally referenced potential density tries to align, the spurious neutral directions are themselves altered, thus leading toward a continued growth in grid-scale waves until the process finally saturates.

As a corollary, this nonlinear instability does not occur in isopycnal layer models since, when they employ two active tracers, the evolution of the second tracer is diagnosed from the first in such a way to preserve the integrity of the potential density layers. The z models can either perform an analogous diagnosis (prescription B in section 4c), or properly compute the neutral directions (prescription A in section 4c) in order to eliminate the instability.

An appropriate framework for building the downgradient property of diffusive mixing into the discretized isoneutral diffusion operator is a functional formalism. Use of this formalism produces a scheme that does not increase tracer variance. Furthermore, it will diffuse tracers downgradient when considering the net effects over a finite volume region of the grid (see Fig. 8 and appendix D). A central strength of the formalism is that the functional is negative semidefinite, which means that any consistent discretization will also be negative semidefinite. Since the negative semidefiniteness is the crucial property that ensures the downgradient nature of the resulting diffusive flux, any consistent discretization starting from the functional will yield a conditionally stable diffusion operator.

We derived the discretized diffusion operator both through performing the discrete functional derivative [details given in Pacanowski (1996)], as well as more heuristically by identifying triads of density points as fundamental building blocks for use in constructing the isoneutral slopes. The triads, when weighted by their respective tracer gradients, eliminate computational

modes since there is no longer any grid splitting. By building in the constraint that the scheme properly orients the neutral directions so that the isoneutral fluxes of the active tracers are balanced, the formalism readily incorporates both the active and passive tracers. The result is a diffusion operator that is consistent, stable, and aims to satisy physical properties that are fundamental to isoneutral diffusion.

There are five appendices that clarify certain details not central to the discussion in the main text. We point to appendix C as being quite important as it provides arguments for employing more sensible and physically based numerical techniques for maintaining linear stability in regions of steep isoneutral slopes. Most notably, the commonly used *slope clipping* scheme introduced by C87 was shown to be especially unphysical in convective regions. This analysis motivated slope scaling, as suggested by Gerdes et al. (1991), Danabasoglu and McWilliams (1995), or the full slope isoneutral diffusion tensor, as a sensible alternative to slope clipping. Currently, we have no arguments motivating one form of scaling over another [however, see Large et al. (1996) for one argument].

Model tests with the new diffusion scheme indicate that it effectively eliminates the problems with the C87 scheme and, hence, allows for removal of horizontal background diffusivity. The effects from the finite volume downgradient fluxing of tracers, as rendered by use of density triads, provides for a more physically reasonable simulation of passive tracer diffusion as well as a more sensible action on water mass properties. Tests of the new scheme alone or combined with the GM90 advection allow for a clean separation of the effects due to isoneutral diffusion and GM90.

In closing, it is important to highlight some of the shortcomings of the new diffusion scheme. Most notably, the two fundamental physical properties described in this paper that we aimed for in the numerical diffusion scheme are not completely realized. First, the scheme does not satisfy a ''downgradient'' property at each of the tracer cells. Rather it does so only over a finite volume that incorporates more than one cell (see section 5a and appendix D). As such, for a pure diffusion problem, it is possible to realize tracer values within a particular tracer cell that are outside a physically realistic range due to spurious upgradient fluxes at the cell face. Preliminary experiments suggest that this problem may be most severe for passive tracers (Gnanadesikan 1998, manuscript submitted to *J. Mar. Res.*) since they are not constrained in a manner similar to the active tracers. The results of Beckers et al. (1997, personal communication) suggest that it is not possible to do better within the constraints of a consistent, linear, numerical isoneutral diffusion operator whose grid stencil extends over the nine points (for a two-dimensional small angle tensor) used here. Second, the new diffusion scheme does not precisely balance the active tracer fluxes in the vertical due to the splitting of the vertical flux into a temporally explicit and implicit piece. The degree to which each of these problems arise will depend on the particulars of the experimental configuration. Some preliminary results reported in Griffies (1998) and Gnanadesikan (in preparation 1997) indicate that in simulations with isoneutral diffusion combined with eddy-advection processes, the solutions are less problematical than with only isoneutral diffusion. Given these caveats, it remains our conclusion that the new diffusion scheme provides a sound tool for systematically addressing various subgrid-scale parameterization theories, including isoneutral diffusion, in *z* coordinate ocean climate models.

Acknowledgments. We wish to thank Jeff Anderson, Jean-Marie Beckers, Kirk Bryan, Bob Hallberg, Isaac Held, Jerry Mahlman, Trevor McDougall, Tony Rosati, Robbie Toggweiler, Eli Tziperman, and Martin Visbeck for discussions and comments. JKD and RDS wish to acknowledge the support of the DOE CHAMMP program. We dedicate this paper to the memory of Mike Cox.

APPENDIX A

Clarification of a Linear Stability Analysis

Bell (1994) provides a linear numerical stability analysis of the equation $\partial_t T = A_I(\partial_{xx} T + S\partial_{zx} T)$, where S is spatially constant. His analysis, assuming an explicit time step and the usual discretized spatial derivatives, indicates the presence of linear modes whose amplitude grows when $|S| > 2\Delta z/\Delta x$. Assuming that this equation is a relevant idealization of isoneutral diffusion, Bell concluded that the discretization of isoneutral diffusion as given by C87 is linearly unstable if the given slope condition is satisfied. However, the assumption that this equation is relevant for isoneutral diffusion in a *z* model is incorrect since it neglects an essential diagonal term. Namely, the simplest form of isoneutral diffusion is given by $\partial_t T = A_I(\partial_{xx} T + S^2\partial_{zz} T + 2S\partial_{zx} T) = A_I(\partial_x + S\partial_z)^2 T$, where S is some constant slope. With the $\partial_{zz} T$ term added, a linear analysis of the discrete equation allows for a conditionally stable scheme. Another way to understand this result is to note that with the extra term, it is possible to produce a variance reducing scheme as shown in this paper. Through the usual energy arguments, such a scheme is at least conditionally stable.

APPENDIX B

Cabbeling and Thermobaricity

It is useful to present the explicit form for the mixing effects induced by the nonlinear equation of state. It is worth noting that these effects are trivially implemented within a *z*-coordinate ocean model by simply employing a nonlinear equation of state when diagnosing density. These effects provide irreversible changes to the locally

referenced potential density. An alternative formulation of the following, in much more detail, can be found in McDougall (1991).

Using the notation from section 3, the time tendency of locally referenced potential density, when considering just the effects due to isoneutral diffusion of the active tracers, is given by

$$\begin{aligned}\partial_t \rho &= \rho(-\alpha \partial_t \theta + \beta \partial_t s) \\ &= \rho[\alpha \nabla \cdot \mathbf{F}_I(\theta) - \beta \nabla \cdot \mathbf{F}_I(s)] \\ &= -\nabla(\rho\alpha) \cdot \mathbf{F}_I(\theta) + \nabla(\rho\beta) \cdot \mathbf{F}_I(s), \end{aligned} \tag{B1}$$

where $\alpha \mathbf{F}_I(\theta) = \beta \mathbf{F}_I(s)$ was used to reach the last equality. The forcing terms, which vanish for a linear equation of state and which cannot be written as the divergence of a flux, represent *cabbeling* and *thermobaricity*. These processes provide irreversible, nondiffusive forms of mixing (McDougall 1987b). For the special case of a single active tracer, the neutral directions are aligned parallel to the isotracer surfaces, thus providing for a zero isoneutral diffusive flux of the single active tracer and therefore an absence of cabbeling and thermobaricity. As a corollary, an *adiabatic* ocean model, in which locally referenced potential density is materially conserved, necessarily employs a linear equation of state or a single active tracer with either a linear or nonlinear equation of state.

To isolate the cabbeling and thermobaricity processes, use the identities

$$\nabla s \cdot \mathbf{F}_I(\theta) = \nabla \theta \cdot \mathbf{F}_I(s),$$

$$\nabla(-\rho\alpha) = \rho_{\theta\theta} \nabla\theta + \rho_{\theta s} \nabla s + \rho_{\theta p} \nabla p$$

and likewise for $\nabla(\rho\beta)$. Note the presence of pressure gradients. These terms represent the effects of probing different pressure surfaces, and hence different potential density surfaces; that is, these are the thermobaric terms. Such a probing of different pressure surfaces is necessary when computing the spatial gradients of the thermal and saline expansion coefficients.[4] With these substitutions, the evolution of locally referenced potential density takes the form

$$\begin{aligned}\partial_t \rho &= \rho_{\theta\theta} \nabla\theta \cdot \mathbf{F}_I(\theta) + \rho_{ss} \nabla s \cdot \mathbf{F}_I(s) + 2\rho_{\theta s} \nabla s \cdot \mathbf{F}_I(\theta) \\ &\quad + \nabla p \cdot (\rho_{\theta p} \mathbf{F}_I(\theta) + \rho_{sp} \mathbf{F}_I(s)) \\ &= \nabla\theta \cdot \mathbf{F}_I(\theta)[\rho_{\theta\theta} - 2\rho_{\theta s}(\rho_\theta/\rho_s) + \rho_{ss}(\rho_\theta/\rho_s)^2] \\ &\quad + \nabla p \cdot (\rho_{\theta p} \mathbf{F}_I(\theta) + \rho_{sp} \mathbf{F}_I(s)). \end{aligned} \tag{B2}$$

The first part of this expression can be written as the product of two quadradic forms by introducing a vector $\mathbf{V} = (1, \alpha/\beta)$ and a metric $\rho_{ab} = \partial_a \partial_b \rho$, where the labels a, b represent the two tracer fields θ, s. This definition renders

$$\partial_t \rho = \nabla\theta \cdot \mathbf{F}_I(\theta) * \mathbf{V} \cdot \mathbf{V} + (\rho_{\theta p} \mathbf{F}_I(\theta) + \rho_{sp} \mathbf{F}_I(s)) \cdot \nabla p, \tag{B3}$$

where the inner product

$$\begin{aligned}\mathbf{V} \cdot \mathbf{V} &= \rho_{ab} V^a V^b \\ &= \rho_{\theta\theta} - 2\rho_{s\theta}(\rho_\theta/\rho_s) + \rho_{ss}(\rho_\theta/\rho_s)^2 \\ &= -\alpha_\theta - 2(\alpha/\beta)\alpha_s + (\alpha/\beta)^2 \beta_s \end{aligned} \tag{B4}$$

represents the squared length of the vector $\mathbf{V}$ on the curved potential density surface characterized locally by the metric ρ_{ab}.

The term

$$\text{cabbeling} = \nabla\theta \cdot \mathbf{F}_I(\theta) * \mathbf{V} \cdot \mathbf{V} \tag{B5}$$

represents the effects from cabbeling. As written here, cabbeling is seen to be the product of a piece associated with the downgradient isoneutral flux of temperature ($\nabla\theta \cdot \mathbf{F}_I(\theta) \le 0$) and a piece associated with the local geometric properties intrinsic to the potential density surface ($\mathbf{V} \cdot \mathbf{V}$). In the ocean, the total or Gaussian curvature $\det(\rho_{ab})(1 + \rho_\theta^2 + \rho_s^2)^{-1}$ is negative. This negative curvature renders $\mathbf{V} \cdot \mathbf{V} \le 0$, which, when combined with downgradient isoneutral diffusion of temperature, always results in a nonnegative tendency for ρ and a consequent downward dianeutral advection (McDougall 1987b).

The terms proportional to the pressure gradient

$$\text{thermobaricity} = (\rho_{\theta p} \mathbf{F}_I(\theta) + \rho_{sp} \mathbf{F}_I(s)) \cdot \nabla p \tag{B6}$$

represent the effects from thermobaricity, which depend on the pressure dependence of the equation of state for seawater. Note that the *thermobaric* term $\rho_{\theta p} \mathbf{F}_I(\theta) \cdot \nabla p$ dominates the *halobaric* term $\rho_{sp} \mathbf{F}_I(s) \cdot \nabla p$; hence the common name *thermobaricity* used for the sum of both terms (McDougall 1987b). In contrast to cabbeling, thermobaricity does not provide a sign-definite source for ρ.

In the manner written here, both cabbeling and thermobaricity effects can be readily diagnosed in z-level models. Use of the new isoneutral diffusion scheme described in this paper, which does not rely on background horizontal diffusion, allows for a potentially clean manifestation of these sometimes subtle effects on the formation and transformation properties of water masses.

APPENDIX C

Linear Stability Constraints Related to Steep Slopes

Traditionally, the implementation of isoneutral diffusion in the z-coordinate models has assumed that slopes of the neutral directions are small (C87, GM90). Additionally, this assumption is implicit in potential density (isopycnal) layer models of the type described by Bleck et al. (1992) or Hallberg (1995), which employ a diagonal diffusion tensor (see GM90 for the full, nondiagonal diffusion tensor in isopycnal coordinates). The small slope approximation is valid over a good part of the World Ocean where slopes larger than 1/100 are

[4] We thank T. McDougall for emphasizing this point.

uncommon. Computationally, it allows fewer terms to be computed in the diffusion operator and so results in significant computational savings. However, in order to satisfy a linear numerical stability condition described below, it is necessary to incorporate ad hoc prescriptions in order to admit steep slopes. The importance of steep slope regions in the ocean are related to their largely determining air–sea exchange processes especially relevant for climate modeling. In addition, recent studies indicate that effects of mesoscale eddies are important in regions near areas of active deep convection where isoneutral slopes are essentially infinite (e.g., Send and Marshall 1995). Therefore, it could be that parameterizing dynamics in steep sloped regions will include quasi-adiabatic stirring, such as GM90, and isoneutral diffusion. For example, Visbeck et al. (1997) have employed the full slope tensor in their simulations of mixing within convection regions since this tensor is valid for all slopes. The purpose of this extended appendix is to provide a linear stability analysis for the full isoneutral tensor. In the process, we provide a summary of the various means available for enforcing the constraints implied from the linear stability analysis relevant for steep isoneutral sloped regions.

a. Linear stability analysis

Numerically realizing isoneutral diffusion along steep slopes is difficult partly because of the small vertical to horizontal aspect ratio in the ocean and hence the ocean model grid. As slopes steepen, the projection of diffusive fluxes onto the vertical become stronger, pushing up against the limits of the linear stability criteria for the diffusion equation. This issue is relevant for discretizing both the small and full tensors. In particular, the linear numerical constraint from the diffusion equation, as discussed in C87, indicates that an explicit numerical scheme with a leapfrog time step will be stable if the grid CFL number satisfies

$$\frac{|K^{mn}|\Delta t}{\Delta x_m \Delta x_n} \leq \frac{1}{4}, \tag{C1}$$

where K^{mn} are the components of the diffusion tensor $\mathbf{K}$ [Eqs. (2) and (3)], and there is no sum implied in this expression. Assuming a geophysically relevant vertical to horizontal aspect ratio for the grid ($\Delta z/\Delta x \leq 1/1000$), the two-dimensional horizontal subsystem is stable when the diffusion equation in the horizontal is stable. In general, satisfying this stability constraint in the horizontal is trivial and so will not be considered further. Solving the vertical K^{zz} diagonal piece implicitly, as done by C87, points to the K^{xz} and K^{yz} cross terms as setting the most restrictive constraint. From these terms, the diffusion equation using the fluxes from the full tensor will be linearly stable when, for each grid cell,

$$\frac{|S_a|}{1 + S_x^2 + S_y^2} \leq \frac{\Delta a \Delta z}{4 A_I \Delta t} \equiv \delta, \tag{C2}$$

where a is either x or y. The small tensor's stability is determined with the $1 + S_x^2 + S_y^2$ denominator set to unity. The slope check parameter δ is the same as used in the stability analysis of C87. For the small angle tensor, δ represents the maximum allowable slope that can be used before some prescription must be employed to ensure numerical stability. For many large-scale ocean model configurations, this slope check parameter is roughly 1/100, thus providing for the self-consistency of the discretization of the small slope tensor. For the full tensor, consider the case where $|S_x| \geq |S_y|$ at some point. If the constraint $|S_x|(1 + S_x^2)^{-1} \leq \delta$ is satisfied, then the constraint in Eq. (C2) is satisfied. In order to derive a conservative stability constraint, drop the extra term and introduce the notation $\Sigma \equiv \max(|S_x|, |S_y|)$.

Just as for the small angle tensor, the stability constraint for the full tensor is satisfied for small isoneutral slopes. Additionally, the $(1 + \Sigma^2)^{-1}$ self-regulating factor allows the numerical constraint to also be satisfied for large angles. Thus, the full tensor has two domains of numerical stability: $0 \leq \Sigma \leq S_{(-)}$ and $S_{(+)} \leq \Sigma < \infty$. The slopes $S_{(\pm)}$ are determined from solutions to $\Sigma = \delta(1 + \Sigma^2)$. This equation has two roots $S_{(\pm)} = [1 \pm (1 - 4\delta^2)^{1/2}]/(2\delta)$. Note that $S_{(+)}S_{(-)} = 1$, and so $S_{(-)} = S_{(+)}^{-1}$. The special case when $\delta \geq 1/2$ means that the slope range vanishes; that is, the full tensor is stable over all slopes $0 \leq \Sigma \leq \infty$, no numerical slope check need be applied, and all diffusion coefficients are set to the unscaled value A_I. Written as a constraint on the isoneutral diffusion coefficient, satisfying $\delta \geq 1/2$ requires $A_I \leq \min(\Delta x \Delta z/2\Delta t)$.

The above analysis is based on the conservative assumption that, if all components to the diffusion tensor produce linearly stable diffusion, then the scheme is linearly stable. Although conservative, our experience has shown that violation of these constraints can result in unacceptably large numerical inaccuracies. These inaccuracies are of special concern since they make it more difficult to realize the balance $\alpha F_I^z(\theta) = \beta F_I^z(s)$, thus exposing the solution to the instability discussed in section 4.

As slopes steepen to greater than the grid aspect ratio, the slope that is estimated by the scheme becomes less accurate due to the need to extrapolate outside of the differencing scheme's stencil. However, in so far as the scheme is respecting the balance in the active tracer fluxes along the computed neutral directions, the scheme is respecting the integrity of the neutral directions. In other words, the new scheme provides a self-consistent discretization of isoneutral diffusion regardless of the neutral direction slopes. In this sense, the errors from this extrapolation are no different in principle than the errors incurred with isopycnal layer models in regions where the neutral directions depart from the isopycnal layers. In addition, as seen in the discussion of section 4, these extrapolation errors are not the cause of the instability in the C87 scheme.

b. Comparing the slope checking prescriptions

There are various methods employed for preserving linear numerical stability in non–small slope regions when

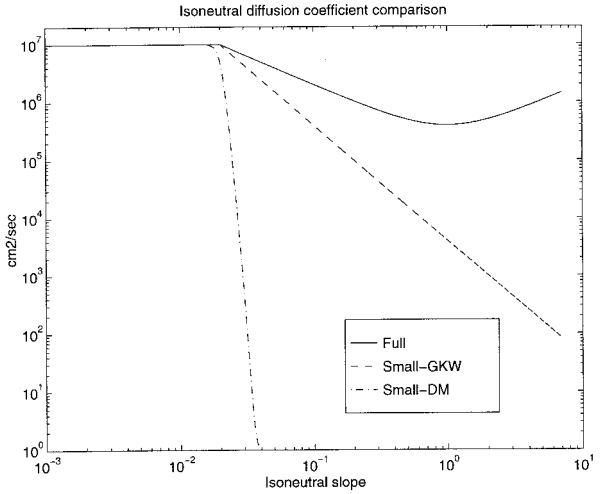

FIG. C1. Comparison of the effective isoneutral diffusion coefficient used in the full and small slope isoneutral tensors for the sector model described in section 2. The region in which 10^7 cm^2 s^{-1} is available is only for slopes $< \delta = 0.02$ for the small angle tensor and $S < 0.02$, $S > 1/0.02$ for the full Redi tensor. Other regions require the scaling as indicated in this figure. Shown is the quadratic scaling (dashed) $A_I(\delta/S)^2$ of Gerdes et al. (1991) (GKW), the exponential scaling (dot–dashed) $0.5A_I(1 - \tanh[(|S|-\delta)/S_d]$ of Danabasoglu and McWilliams (1995) (DM) with $S_d = 0.002$, and the $\delta(S + 1/S)$ scaling for the full tensor (solid). We emphasize that for a slope of 0.1, for example, the components of the small angle and full tensor do not differ by more than 0.01. However, as seen in this figure, the difference in the isoneutral diffusion coefficients that are used for the different scaling functions range over many orders of magnitude.

using the small slope tensor. The first, and perhaps most widely employed, is known as *slope clipping* and was introduced in C87. Slope clipping limits the slope along which isoneutral diffusion occurs. Namely, when the slope $|S_x| > S_{\text{clip}}$, with S_{clip} some maximum slope usually taken as a fraction of the grid parameter δ given in Eq. (C2), then $\delta_z\rho$ is replaced by ρ_{zc} so that $|\delta_x\rho/\rho_{zc}| = S_{\text{clip}}$. The same sort of clipping is applied independently to the y component of the slope. Limiting the slopes along which the fluxes are directed introduces significant false dianeutral fluxes in those regions where the actual slopes exceed the clipped slope. There are arguments supporting the presence of increased dianeutral fluxes in regions of steep slopes where there is the potential for enhanced energy release at all scales. Therefore, one may guess that slope clipping is a reasonable parameterization of these effects. However, the dianeutral fluxes implied through slope clipping can be rather large and unphysical. For example, consider the case where density is a linear function only of temperature and there is no vertical stratification. The vertical component to the clipped isoneutral flux of temperature becomes

$$F^{z\text{clip}}_{i,k} = A_I|\delta_x \overline{\theta}^{x,z}_{i-1,k}|S_{\text{clip}} > 0. \tag{C3}$$

For a case with $A_I = 10^7$ cm^2 s^{-1}, $S_{\text{clip}} = 0.01$, and $\partial_x\theta = 10^{-7}$ K cm^{-1}, the vertical flux of temperature is 0.01 K cm s^{-1}, which, when multiplied by $\rho_o C_p \approx 4 \times 10^6$ J (m^3 K)$^{-1}$, corresponds to a heat flux of roughly 400 W m^{-2}. This is a substantial upward flux, which is unphysical since it occurs in an unstratified column. This effect provides strong motivation to seek other means of satisfying the linear stability constraints in steeply sloping regions.

Rather than slope clipping, Gerdes et al. (1991) proposed a quadratic scaling of the isoneutral diffusion coefficient: $A_I \to A_I(\delta/S)^2$, as the slope reaches values greater than δ. In this manner, it is possible to preserve the isoneutral orientation of the diffusive flux while maintaining linear numerical stability. The quadratic dependence of the scaling is suggested by the $(1 + S^2)^{-1}$ factor, which appears in the full tensor. For this prescription, isoneutral diffusion in the limit of $|S| \gg \delta$ reduces to $\partial_z(\delta^2 A_I \partial_z T)$, that is, just vertical diffusion. This result is reasonable as it yields a zero isoneutral diffusive flux in the limit of vertically unstratified columns, and so completely decouples convection and isoneutral diffusion. Danabasoglu and McWilliams (1995) proposed an exponential scaling of the form

$$0.5A_I\left[1 - \tanh\left(\frac{|S| - \delta}{S_d}\right)\right],$$

where S_d prescribes the half-width over which the transition from a diffusion coefficient near A_I to a near-zero diffusion coefficient occurs. This scaling exponentially shuts off all the isoneutral diffusion above a critical slope. Such scaling, as discussed by (Large et al. 1997), appears to be necessary for smoothly coupling isoneutral diffusion with the GM90 scheme to a mixed layer scheme.

For the case when the full tensor is employed with the grid parameter $\delta < 1/2$, a prescription for stably realizing intermediate slopes must be given. For this purpose, we seek some function $\mu(\Sigma)$ that preserves the condition $\mu(\Sigma)(\Sigma^{-1} + \Sigma)^{-1} \le \delta$. A function that satisfies the equality in a smooth fashion is given by

$$\mu(\Sigma) = \begin{cases} 1 & \text{if} \quad \Sigma \le S_{(-)} \\ \delta(\Sigma + \Sigma^{-1}) & \text{if} \quad S_{(-)} \le \Sigma \le S_{(+)} \\ 1 & \text{if} \quad S_{(+)} \le \Sigma < \infty. \end{cases} \tag{C4}$$

Note that $\mu(\Sigma)$ is smooth since $\delta(S_{(+)} + S^{-1}_{(+)}) = \delta(S_{(-)} + S^{-1}_{(-)}) = 1$. The scaling function $\mu(\Sigma)$ decreases from unity to a minimum of 2δ at $\Sigma = 1$. As seen by the comparison in Fig. C1, this scaling is significantly more mild than that necessary for stabilizing the small angle tensor. Again, such a mild scaling is allowed since the full tensor has an intrinsic scaling through the $(1 + S^2)^{-1}$ factor.

Although perhaps cleaner from a numerical perspective due to its natural rescaling properties, it is not clear whether the full tensor is a physically relevant alternative to the small slope tensor. Further studies are necessary to clarify this issue.

APPENDIX D

Downgradient Diffusion over Finite Volumes

The discussion in section 5 argued for an interpretation of ''downgradient diffusion'' in the new scheme in a finite

volume context. The relevant finite sized volume is given by the shaded region in Fig. 8 for the x–z plane. It is useful to be more mathematical in what is actually meant by such an interpretation. For this purpose, consider the discretized functional for isoneutral diffusion

$$\mathcal{F} = -\frac{1}{2}\sum_{i,j,k}\sum_{n=1}^{36} A(n)V(n)|\nabla T^{(n)} \times \hat{\gamma}^{(n)}|^2$$
$$= -\frac{1}{2}\sum_{i,j,k}\sum_{n=1}^{36} \nabla T^{(n)}\cdot A(n)V(n)[(\hat{\gamma}^{(n)} \times \nabla T^{(n)}) \times \hat{\gamma}^{(n)}]. \quad \text{(D1)}$$

The notation corresponds to that used in Eq. (28). Note that each term in the sum over the 36 quarter-cells is negative semidefinite. This expression is completely analogous to that provided for the continuous case in section 5a. Correspondingly, the discretized scalar product of the tracer gradient and tracer flux associated with the finite volume $V_{i,j,k} = \Sigma_{n=1}^{36} V(n)$ is given by

$$[\nabla T\cdot\mathbf{F}]_{V_{i,j,k}} \equiv -\left(\sum_{n=1}^{36} \nabla T^{(n)}\cdot A(n)V(n)[(\hat{\gamma}^{(n)}\times\nabla T^{(n)})\times\hat{\gamma}^{(n)}]\right) \le 0. \quad \text{(D2)}$$

This expression provides an explicit form for what is meant by ''downgradient diffusion'' in the new scheme. It is only in such a coarse-grained sense that the new scheme can be said to provide a downgradient isoneutral diffusive flux. The corresponding result for the small tensor diffusive flux is (restricting to the x–z plane for the sake of brevity)

$$\mathcal{F}^{\text{small}} = -\frac{1}{2}\sum_{i,k}\sum_{n=1}^{12} A(n)V(n)(\partial_x T^{(n)} + S_x^{(n)}\partial_z T^{(n)})^2$$
$$= -\frac{1}{2}\sum_{i,k}\sum_{n=1}^{12} A(n)V(n)[\partial_x T^{(n)}(\partial_x T^{(n)} + S_x^{(n)}\partial_z T^{(n)}) + \partial_z T^{(n)}(S_x^{(n)}\partial_x T^{(n)} + (S^2)^{(n)}\partial_z T^{(n)})], \quad \text{(D3)}$$

and the corresponding finite volume ''downgradient'' interpretation is given by

$$[\nabla T\cdot\mathbf{F}]^{\text{small}}_{V_{i,k}} \equiv -\left\{\sum_{n=1}^{12} A(n)V(n)[\partial_x T^{(n)}(\partial_x T^{(n)} + S_x^{(n)}\partial_z T^{(n)}) + \partial_z T^{(n)}(S_x^{(n)}\partial_x T^{(n)} + (S^2)^{(n)}S_x^{(n)})]\right\} \le 0, \quad \text{(D4)}$$

where $V_{i,k} = \Sigma_{n=1}^{12} V(n)$.

APPENDIX E

Full Tensor Diffusive Flux

The basic grid stencil that was discussed in the main text for the small tensor is also employed for the full tensor. The only added step requires the discretization of the $|\nabla\rho|^2$ term occurring in the denominator of the functional. It is sufficient to discuss only the results for the x component since the y and z components follow similarly. This component takes the discretized form

$$-F^x_{i,k,j} = \frac{1}{4dyt_j\cos\phi^T_j}\sum_{jq=0}^{1} dyu_{j-1+jq}\cos\phi^U_{j-1+jq}\sum_{ip=0}^{1} A_I Sy^{(i+ip,k,j)}_{(i+ip,k,j-1+jq\,|\,i+ip,k,j)}$$
$$\times\frac{Sy^{(i+ip,k,j)}_{(i+ip,k,j-1+jq\,|\,i+ip,k,j)}\delta_x T_{i,k,j} - Sx^{(i+ip,k,j)}_{(i,k,j\,|\,i+ip,k,j)}\delta_y T_{i+ip,k,j-1+jq}}{(Sx^{(i+ip,k,j)}_{(i,k,j\,|\,i+ip,k,j)})^2 + (Sy^{(i+ip,k,j)}_{(i+ip,k,j-1+jq\,|\,i+ip,k,j)})^2 + 0.5[1 + (\delta_z\rho^{(i+ip,k,j)}_{i+ip,k-1,j})^2/(\delta_z\rho^{(i+ip,k,j)}_{i+ip,k,j})^2]}$$
$$+\frac{1}{4dzt_k}\sum_{kr=0}^{1} dzw_{k-1+kr}\sum_{ip=0}^{1} A^{(i+ip,k,j)}_{(i,k,j\,|\,i+ip,k-1+kr,j)}$$
$$\times\frac{\delta_x T_{i,k,j} + Sx^{(i+ip,k,j)}_{(i,k,j\,|\,i+ip,k-1+kr,j)}\delta_z T_{i+ip,k-1+kr,j}}{1 + (Sx^{(i+ip,k,j)}_{(i,k,j\,|\,i+ip,k-1+kr,j)})^2 + 0.5[(Sy^{(i+ip,k,j)}_{(i+ip,k,j-1\,|\,i+ip,k-1+kr,j)})^2 + (Sy^{(i+ip,k,j)}_{(i+ip,k,j\,|\,i+ip,k-1+kr,j)})^2]}.$$

The x–y cross term (the first term) vanishes in the small slope limit. Additionally, the diffusion coefficient for the x–y cross term has been set to the unscaled value A_I based on the stability analysis given in appendix C. In the denominators, the discretization for the density gradient extending in the direction orthogonal to the gradients in the numerator are constructed as an average of squares. This averaging brings the discretized gradients back onto the plane corresponding to the other part of the denominator. Performing the average after taking the square avoids introducing computational modes. In the denominator of the second term, $1 \gg |Sx|, |Sy|$ in the small slope limit, thus reducing this term to the result given by Eq. (30). Further details can be found in Pacanowski (1996).

Note added in proof. Soon after this paper was accepted, the work of Kershaw (1981) and that of Beckers et al. (1998) came to our attention. The functional approach employed in this paper was also used by Kershaw for his discretization of a diffusion operator. Beckers et al. provide a general proof that the resulting linear rotated diffusion scheme cannot maintain positive definiteness of a tracer field.

REFERENCES

Beckers, J.-M., H. Burchard, J. M. Campin, E. Deleersnijder, and P. P. Mathieu, 1998: Another reason why simple discretizations of rotated diffusion operators cause problems in ocean models: Comments on ''Isoneutral diffusion in a *z*-coordinate ocean model.'' *J. Phys. Oceanogr.,* in press.

Bell, M. J., 1994: Notes on how to choose parameter values for the Cox numerical ocean circulation model. U.K. Met. Office Forecasting Research Tech. Rep. No. 135, 43 pp.

Bleck, R., C. Rooth, D. Hu, and L. T. Smith, 1992: Salinity-driven thermocline transients in a wind and thermohaline forced isopycnic coordinate model of the North Atlantic. *J. Phys. Oceanogr.,* **22,** 1486–1505.

Böning, C., W. R. Holland, F. O. Bryan, G. Danabasoglu, and J. C. McWilliams, 1995: An overlooked problem in model simulations of the thermohaline circulation and heat transport in the Atlantic Ocean. *J. Climate,* **8,** 515–523.

Bryan, K., 1975: Three-dimensional numerical models of the ocean circulation. *Numerical Models of Ocean Circulation,* Natl. Acad. Sci., 94–105.

——, and M. D. Cox, 1972: An approximate equation of state for numerical models of the ocean circulation. *J. Phys. Oceanogr.,* **2,** 510–514.

——, S. Manabe, and R. C. Pacanowski, 1975: A global ocean–atmosphere climate model. Part II: The oceanic circulation. *J. Phys. Oceanogr.,* **5,** 30–46.

Courant, R., and D. Hilbert, 1953: *Methods of Mathematical Physics.* Vols. 1 and 2. Wiley, 560 and 830 pp.

Cox, M. D., 1984: A primitive equation, 3-dimensional model of the ocean. GFDL Ocean Group Tech. Rep. 1, GFDL. [Available from GFDL, Route 1, Box 308, Princeton, NJ 08542.]

——, 1987: Isopycnal diffusion in a *z*-coordinate ocean model. *Ocean Modelling,* (unpublished manuscripts), **74,** 1–5.

——, and K. Bryan, 1984: A numerical model of the ventilated thermocline. *J. Phys. Oceanogr.,* **14,** 674–687.

Danabasoglu, G., and J. C. McWilliams, 1995: Sensitivity of the global ocean circulation to parameterizations of mesoscale tracer transports. *J. Climate,* **8,** 2967–2987.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.,* **20,** 150–155.

——, J. Willegrand, T. J. McDougall, and J. C. McWilliams, 1995: Parameterizing eddy-induced tracer transports in ocean circulation models. *J. Phys. Oceanogr.,* **25,** 463–474.

Gerdes, R., C. Köberle, and J. Willebrand, 1991: The influence of numerical advection schemes on the results of ocean general circulation models. *Climate Dyn.,* **5,** 211–226.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics.* Academic Press, 662 pp.

Goloviznin, V. M., A. A. Samarskii, and A. P. Favorskii, 1977: A variational approach to constructing finite-difference mathematical models in hydrodynamics. *Sov. Phys. Dokl.,* **22,** 432–434.

Gough, W. A., and W. J. Welch, 1994: Parameter space exploration of an ocean general circulation model using an isopycnal mixing parameterization. *J. Mar. Res.,* **52,** 773–796.

——, and C. A. Lin, 1995: Isopycnal mixing and the Veronis effect in an ocean general circulation model. *J. Mar. Res.,* **53,** 189–199.

Griffies, S. M., 1998: The Gent–McWilliams skew flux. *J. Phys. Oceanogr.* **28,** 831–841.

Hallberg, R., 1995: Some aspects of the circulation in ocean basins with isopycnals intersecting the sloping boundaries. Ph.D. dissertation, University of Washington, 244 pp.

Haltiner, G. J., and R. T. Williams, 1981: *Numerical Prediction and Dynamic Meteorology.* 2d ed. Wiley, 477 pp.

Hirst, A. C., D. R. Jackett, and T. J. McDougall, 1996: The meridional overturning cells of a World Ocean model in neutral density coordinates. *J. Phys. Oceanogr.,* **26,** 775–791.

Iselin, C. O., 1939: The influence of vertical and lateral turbulence on the characteristics of the waters at mid-depth. *Trans. Amer. Geophys. Union,* **20,** 414–417.

Kershaw, D. S., 1981: Differencing of the diffusion equation in Lagrangian hydrodynamical codes. *J. Comput. Phys.,* **39,** 375–395.

Korshiya, T. K., V. F. Tishkin, A. P. Favorskii, and M. Y. Shashkov, 1980: Flow-variational difference schemes for calculating the diffusion of a magnetic field. *Sov. Phys. Dokl.,* **25,** 832–834.

Kunze, E., and T. B. Sandford, 1996: Abyssal mixing: Where it is not. *J. Phys. Oceanogr.,* **26,** 2286–2296.

Large, W. G., G. Danabasoglu, S. C. Doney, and J. C. McWilliams, 1997: Sensitivity to surface forcing and boundary layer mixing in a global ocean model: Annual-mean climatology. *J. Phys. Oceanogr.,* **27,** 2418–2447.

Ledwell, J. R., A. J. Watson, and C. S. Law, 1993: Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. *Nature,* **364,** 701–703.

Levitus, S., 1982: *Climatological Atlas of the World Ocean.* NOAA Prof. Paper No. 13, U.S. Govt. Printing Office, Washington, DC, 173 pp.

McDougall, T. J., 1987a: Neutral surfaces. *J. Phys. Oceanogr.,* **17,** 1950–1967.

——, 1987b: Thermobaricity, cabbeling, and water-mass conversion. *J. Geophys. Res.,* **92,** 5448–5464.

——, 1991: Parameterizing mixing in inverse models. *Dynamics of Oceanic Internal Gravity Waves, Proc. Sixth 'Aha Huliko 'a Hawaiian Winter Workshop,* P. Müller and D. Henderson, Eds., Manoa, HI, University of Hawaii at Manoa, 355–386.

——, and J. A. Church, 1986: Pitfalls with numerical representations of isopycnal and diapycnal mixing. *J. Phys. Oceanogr.,* **16,** 196–199.

——, and D. R. Jackett, 1988: On the helical nature of neutral trajectories in the ocean. *Progress in Oceanography,* Vol. 20, Pergamon, 153–183.

Montgomery, R. B., 1940: The present evidence on the importance of lateral mixing processes in the ocean. *Bull. Amer. Meteor. Soc.,* **21,** 87–94.

Olbers, D. J., M. Wenzel, and J. Willebrand, 1985: The inference of

North Atlantic circulation patterns from climatological hydrographic data. *Rev. Geophys.,* **23,** 313–356.

Pacanowski, R. C., 1996: MOM 2 documentation, user's guide and reference manual. GFDL Ocean Tech. Rep. 3.1, Geophysical Fluid Dynamics Laboratory/NOAA. [Available from GFDL, Princeton University, Princeton, NJ 08542.]

Plumb, R. A., and J. D. Mahlman, 1987: The zonally averaged transport characteristics of the GFDL general circulation/transport model. *J. Atmos. Sci.,* **44,** 298–327.

Rahmstorf, S., 1993: A fast and complete convection scheme for ocean models. *Ocean Modelling,* (unpublished manuscripts), **101,** 9–11.

Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation. *J. Phys. Oceanogr.,* **12,** 1154–1158.

Send, U., and J. Marshall, 1995: Integral effects of deep convection. *J. Phys. Oceanogr.,* **25,** 855–872.

Solomon, H., 1971: On the representation of isentropic mixing in ocean models. *J. Phys. Oceanogr.,* **1,** 233–234.

Tishkin, V. F., A. P. Favorskii, and M. Y. Shashkov, 1979: Variational-difference schemes for the heat-conduction equation on nonregular grids. *Sov. Phys. Dokl.,* **24,** 446–448.

Veronis, G., 1975: The role of models in tracer studies. *Numerical Models of Ocean Circulation,* Natl. Acad. Sci., 133–146 pp.

Visbeck, M., J. Marshall, T. Haine, and M. Spall, 1997: Specification of eddy transfer coefficients in coarse resolution ocean circulation models. *J. Phys. Oceanogr.,* **27,** 381–402.

Weaver, A. J., and E. S. Sarachik, 1990: On the importance of vertical resolution in certain ocean general circulation models. *J. Phys. Oceanogr.,* **20,** 600–609.

——, and M. Eby, 1997: On the numerical implementation of advection schemes for use in conjunction with various mixing parameterizations in the GFDL ocean model. *J. Phys. Oceanogr.,* **27,** 369–377.