Baroclinic Instability: An Oceanic Wavemaker for Interdecadal Variability

ALAIN COLIN DE VERDIÈRE AND THIERRY HUCK

Laboratoire de Physique des Océans, Université de Bretagne Occidentale, Brest, France

(Manuscript received 20 August 1997, in final form 3 June 1998)

ABSTRACT

Numerical simulations of coarse-resolution, idealized ocean basins under constant surface heat flux are analyzed to show that the interdecadal oscillations that emerge naturally in such configurations are driven by baroclinic instability of the mean state and damped by horizontal diffusion. When the surface heat fluxes are diagnosed from a spinup in which surface temperatures are strongly restored to apparent atmospheric temperatures, the most unstable regions diagnosed by large downgradient eddy heat fluxes are located in the basin northwest corner where the surface heat losses are largest. The long-wave limit of the baroclinic instability of idealized mean flows in a three-layer model with vertical shears as observed in the GCMs demonstrates that growth rates of order one cycle per year can be produced locally, large enough to amplify thermal anomalies in the face of lateral diffusion. The proposed instability mechanism that favors surface-intensified perturbations also explains the lack of oscillations if the restoring to a surface climatology is too strong. To assess whether this instability process of oceanic origin is robust enough to cause interdecadal variability of coupled ocean–atmosphere models, a four-box ocean–atmosphere model is constructed. Given the large heat capacity of the ocean as compared to the atmosphere, the dynamical system that governs the model evolution is reduced to only two degrees of freedom, the oceanic overturning thermohaline circulation and the interior north–south temperature gradient. The authors show that, when the baroclinic instability growth rate exceeds the overall dissipation caused by turbulent eddy diffusion in the atmosphere and ocean and infrared back radiation, the dynamical system undergoes a Hopf bifurcation, and interdecadal oscillations emerge through a limit cycle.

1. Introduction

Interdecadal variability of the temperature of the surface layers of the ocean has been identified very early by Bjerknes (1964) and later by Deser and Blackmon (1993), Kushnir (1994), and Hansen and Bezdek (1996) among others. The variability takes the form of large-scale surface-intensified anomalies. The EOF analysis of Deser and Blackmon for the North Atlantic reveals the dominance of a dipole, clearly intensified near the western boundary roughly situated off Newfoundland. Reverdin et al. (1997) showed that the anomalies are surface intensified and that the salinity signal in the North Atlantic is coherent with the temperature signal. They point out that the region of the slope current is a likely origin for the long-period fluctuations. Following Stommel's (1961) idea that different boundary conditions for temperature and salinity could lead to multiple steady states of the thermohaline circulation (THC), Marotzke (1990) and Weaver and Sarachik (1991) noted that oceanic GCMs run under mixed boundary conditions (restoring for temperature, flux for salinity) can also exhibit decadal oscillations with clear advective origins. The study of the oscillations was then pursued in simpler contexts with only one active variable, forced by surface constant flux; see Huang and Chou (1994), Greatbatch and Zhang (1995), Cai et al. (1995), and Chen and Ghil (1995). Greatbatch and Zhang pointed out the strong similarity between the oscillations observed in such a square-box ocean model forced solely by constant heat flux and those of the fully coupled GFDL ocean–atmosphere model described by Delworth et al. (1993). The oscillations obtained under constant flux were shown to persist in a similar form by Chen and Ghil (1996) when a simple ocean model was coupled to an energy balance model (EBM) of the atmosphere, removing the assumptions of constant flux at the air–sea interface. What was demonstrated in that study is that, at the very low frequencies of concern here, the ocean sees almost a constant heat flux with time variations much smaller than mean values. A sizable fraction of the constant solar flux at the top of the atmosphere drives the ocean below. Furthermore, a number of sensitivity runs allowed them to suggest that the transition from a steady THC to an oscillatory one occurred through a Hopf bifurcation as either the atmospheric turbulent heat diffusivity or the ocean–atmosphere coupling coefficient decreased.

Corresponding author address: A. Colin de Verdière, Laboratoire de Physique des Océans, Université de Bretagne Occidentale, 6 Avenue le Gorgeu, BP 809, 29285 Brest cedex, France.
E-mail: acolindv@univ-brest.fr

© 1999 American Meteorological Society

Although the advective origin of the interdecadal oscillations under mixed boundary conditions was recognized early by Weaver and Sarachik (1991), the precise mechanisms under which the oscillations proceed have remained elusive. Winton (1996) demonstrated clearly the three-dimensional character of the oscillations by comparison with two-dimensional simulations that did not exhibit this type of interdecadal variability. Winton (1996) and Greatbatch and Peterson (1996) suggested that key to the oscillatory nature of the THC is the existence of boundary-trapped waves that propagate in the Kelvin wave sense. Winton proposed that the oscillation is triggered by thermal wind currents impinging normal to the eastern boundary, the resulting anomaly propagating in an uninterrupted manner around the basin. Greatbatch and Peterson further observed that only the western boundary was crucial to the existence of the oscillations and suggested that southward propagating boundary waves perturb the western boundary current that in turn generate perturbations that are advected to the northeast corner and play a role in reinitiating the wave propagation. If this region plays an important role, the demonstration of the absence of sensitivity to relaxation along the northern–eastern boundary is problematic. Huck et al. (1999a, hereafter HCW) carried out a specific f-plane experiment with symmetric forcing in latitude (the northern boundary being removed) and showed that interdecadal oscillations proceeded in much the same way without boundary wave propagation. A long series of experiments under constant flux showed further that forcing amplitudes and rotation were conducive to the oscillatory state (stronger amplitudes), while the mixing processes such as convection, horizontal mixing and dissipation were of a damping nature (smaller amplitudes). Runs without the β effect showed that the variation of the Coriolis parameter with latitudes was not crucial as noted earlier by Winton. Although the character of the oscillations is essentially three-dimensional, HCW showed that the oscillation can be described using only two active variables: the strength of the overturning ψ and the meridional south–north temperature difference anomaly ΔT. Through heat conservation, the rate of change $\Delta\dot{T}$ relates to the temperature advection as $(-\psi\overline{\Delta T})$, while it is observed in the numerical experiments that ψ itself does not relate to ΔT as in Stommel's (1961) model but instead that the derivative $\dot{\psi}$ relates to ΔT—the existence of such a phase lag between overturning and temperature gradient was already discussed by Greatbatch and Peterson. It is the existence of this quadrature between the ''interior forcing'' ΔT and the western boundary current (which is the dominant contribution to the overturning ψ) that sustains the oscillations. The origin of this time delay is really three-dimensional through a combination of advection and wave propagation. It is reminiscent of the spinup of an ocean basin driven by a wind stress curl (Anderson and Gill 1975) in which the western boundary current builds up and narrows on the timescale it takes for Rossby waves to cross the basin. While Winton (1996) and Greatbatch and Peterson (1996) advocate the importance of viscous Kelvin waves in the adjustment, interior potential vorticity (PV) waves exist in the presence of an underlying mean circulation and stratification. They can be observed, even in pure f-plane cases, to play a role in establishing the period of the oscillations. The main point that we want to discuss in this paper is concerned, however, with *the energy sources of the oscillations* and we want to draw the attention to ''*the wavemaker*'' that must be present to sustain the oscillations in the face of dissipation because the coarse-resolution models that we are using are indeed very dissipative! Given that the oscillations that are observed are truly three-dimensional, it is very difficult to find out ''what is driving what'' unless a systematic search for the energy sources is made. This is essentially what we report in this paper. In the absence of explicit momentum advection, no conversion from mean kinetic energy is possible and the only remaining possibility involves the conversion from mean potential energy through baroclinic instability. This process has been identified in the western boundary current region that plays the role of a wavemaker to excite whatever waves are possible in the interior. To demonstrate this proposition a number of planetary geostrophic (PG) simulations are carried out in a situation where the ocean is first put to steady equilibrium under restoring boundary conditions at the surface and then allowed to depart from this initial state under the diagnosed flux kept constant from there on. This procedure generates the familiar western intensification of buoyancy flux that is a key feature of large-scale ocean–atmosphere interaction. Once in this configuration of constant flux, classical stability analyses of the numerical solutions are possible and the dominant destabilizing factors identified (section 2).

If our conjecture that baroclinic instability is the dominant process driving the oscillations is correct, then the simplest model exhibiting baroclinic instability at scales beyond the Rossby radius of deformation is a three-layer model. A third layer is needed because on these long interdecadal timescales, the barotropic mode is always in equilibrium with its forcing and is then unlikely to play an important role. It is, in any case, absent from our flat-bottom, buoyancy-driven experiments with no bottom friction. The growth rates of unstable perturbations to idealized mean flows whose vertical shears are calibrated from the numerical model solutions are discussed in section 3.

Finally we continue Chen and Ghil's (1996) lines of thought in section 4 to construct a coupled ocean–atmosphere box model that shows that their important suggestion of a transition from a steady to an oscillatory state via a Hopf bifurcation can be interpreted in a somewhat different way: we propose that the limit cycles of the oscillations arise when the growth rate of perturbations (due to the baroclinically unstable western

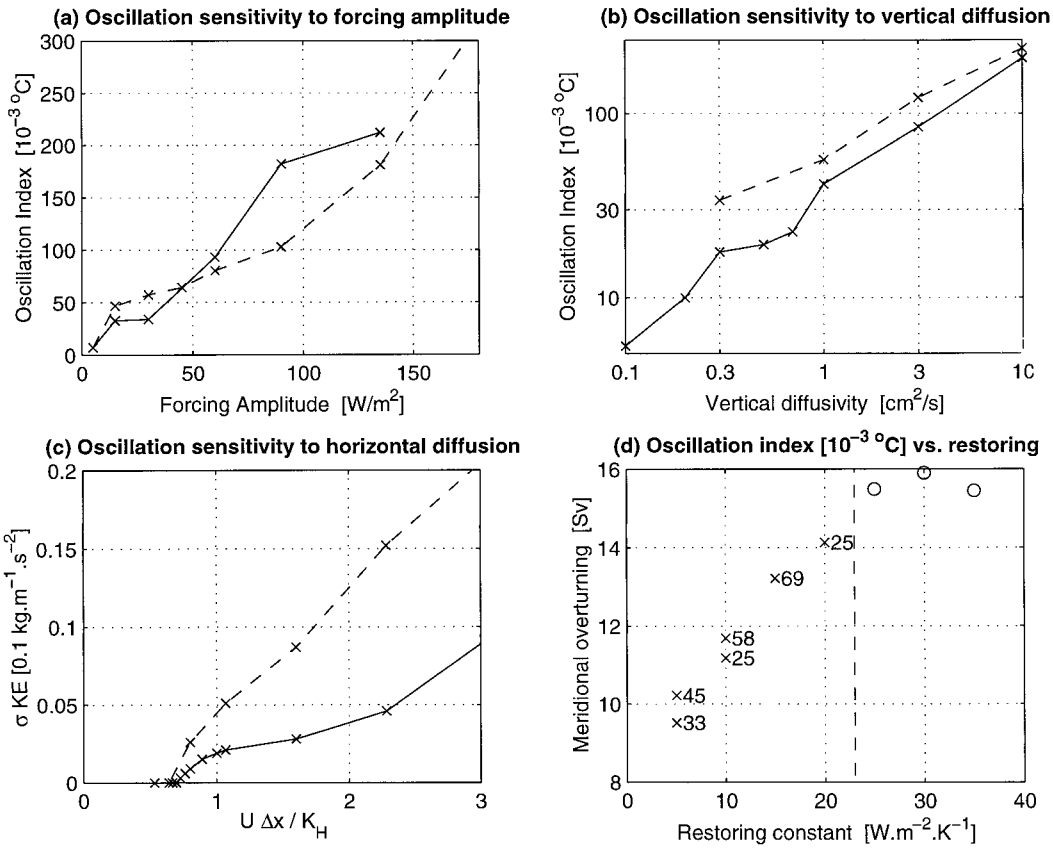

FIG. 1. Sensitivity of the interdecadal oscillations to (a) variations of the amplitude of the meridional distribution of surface heat flux; (b) variations of the vertical diffusivity coefficient (log–log plot); (c) variations of the horizontal diffusivity coefficient. Abcissa is the horizontal Peclet number $U\Delta x/K_h$ with $U = 1$ cm s^{-1} and $\Delta x = 160$ km and ordinate is the square root of the kinetic energy. Note that no oscillations are found for Peclet numbers less than 0.64 (associated with diffusivity coefficients larger than 2500 m^2 s^{-1}); and (d) variations of the restoring constant (when the surface temperature is restored to a linear meridional temperature distribution). The presence (×) or absence (○) of the oscillations is indicated along with the oscillation index in the former case. The mean overturning strength is plotted on the vertical axis, since the restoring atmospheric temperatures are changed along with the restoring constant. Note that oscillations disappear for values between 20 and 25 W m^{-2} K^{-1}. All of these results have been obtained with the HR configuration (Table 1) for the solid line, but with a purely geostrophic model (no momentum dissipation but no-slip boundary conditions imposed) for the dashed line. The oscillation index is the basin average of temperature standard deviations over a period.

boundary current region) *exceeds* the diffusive timescale resulting from the cumulative damping action of turbulent oceanic–atmospheric diffusivities and infrared back radiation. On these interdecadal timescales, the atmosphere is restricted to a damping role with its variables enslaved to the active oceanic variables.

2. Evidences for baroclinic instability

a. Critical damping terms

HCW's experiments have identified the parameters that govern interdecadal oscillations, with essentially the amplitude of the surface flux driving the variability and all mixing processes but vertical mixing acting as a brake. Figure 1 summarizes some of these results that demonstrate the geostrophic/inviscid character of the oscillations. The reader is referred to HCW's paper for details of the experiments that led to the construction of these figures. Essential parameters of these medium resolution (HR) runs are recalled in Table 1. We first define an oscillation index as the basin-mean standard deviation of temperature over one period of the oscillation and consider the influence of an imposed surface heat flux zonally uniform and linearly varying in latitude. As the amplitude of the forcing increases, so does the oscillation index (Fig. 1a). If instead the forcing is kept constant but the vertical mixing increases, the oscillation index again increases (Fig. 1b). Since higher vertical mixing implies a stronger overturning of the mean circulation, both results stress that the more energetic the mean circulation, the larger the amplitude of the oscillations.

As shown in HCW, the horizontal diffusion appears as the main damping for controlling the oscillation am-

TABLE 1. Configurations of the low-resolution (LR) and the higher-resolution (HR) runs discussed in the text. The models are described in details in Colin de Verdière (1988) and Huck et al. (1999b), respectively. They are based on the planetary geostrophic equations with Laplacian friction closure for a flat-bottomed Cartesian β plane centered at 40°N and extending from 20° to 60°N. K_H (K_V) is the horizontal (vertical) diffusivity, while A_H is the horizontal viscosity (compared to which vertical viscosity is safely negligible).

Model	Domain Lat (km)	Domain Long (km)	Domain Depth (m)	Grid points (long × lat × vert)	Grid spacing (km)	K_H ($m^2\ s^{-1}$)	K_V ($m^2\ s^{-1}$)	A_H ($m^2\ s^{-1}$)
LR	4500	6000	4500	12 × 16 × 15	375	2000	10^{-4}	1.0×10^6
HR	4480	5120	4500	32 × 28 × 15	160	700	10^{-4}	1.5×10^5

plitude (Fig. 1c). For values beyond 2500 $m^2\ s^{-1}$, no oscillation is observed. The relevant nondimensional parameter is the horizontal Peclet number defined at the resolution scale as $U\Delta x/K_h$ whose value at the critical threshold is 0.64. We observe that the amplitudes of the oscillations (as measured by the square root of the kinetic energy) increase as the square root of the deviation of the Peclet number from its critical value, suggesting a supercritical Hopf bifurcation. Around the bifurcation the oscillation period is nearly constant (29 yr) and similar to the damped oscillations in the nonoscillating regime. A similar transition with respect to variations of the coupling coefficient was observed by Chen and Ghil (1996) in an ocean model coupled to an energy balance model. Dispersion by oceanic mesoscale eddies at a diffusivity rate of 1000 $m^2\ s^{-1}$ gives a Peclet number of 1.6, suggesting that the real ocean might well operate within the regime of oscillations. From this critical value of the damping, we can infer the actual growth rate of the instability that sustains the oscillations. At the horizontal resolution, $\Delta x = 160$ km, of these sensitivity experiments (HR runs), a timescale $\Delta x^2/K_H$ of 120 days emerges. In comparison the vertical mixing plays a very weak role to damp the anomalies: The diffusive timescale $\Delta z^2/K_V$ reaches 3 yr for a 100-m depth interval at the pivot value of $10^{-4}\ m^2\ s^{-1}$. This very weak direct damping effect of vertical mixing is, in fact, more than counterbalanced by the increase in the mean THC that follows an increase in K_V. Of course, the convective adjustment acting as a very large vertical diffusion where static instability occurs plays a significant damping role on the oscillations as shown in HCW.

If the surface forcing is now a relaxation toward an apparent atmospheric temperature, we can again find a critical value for the relaxation coefficient [$\lambda = d(\text{heatflux})/dT$] that separates the oscillatory solutions from the steady ones (Fig. 1d). The critical value so obtained $\lambda \approx 22\ W\ m^{-2}\ K^{-1}$ corresponds to a timescale of $\rho C_p \Delta z/\lambda = 105$ days for the HR run mixed layer depth (50 m), which agrees well with the one inferred above from variations of the horizontal diffusion. This is a valuable result in view of the analysis of Seager et al. (1995), who showed through modeling of the response of the lower atmosphere to SST anomalies that the sensitivity of heat flux with respect to SST was of the order of $15\ W\ m^{-2}\ K^{-1}$, significantly less than values currently used in ocean models. Such a value translates to a restoring coefficient of order one cycle/year for 100-m vertical resolution, and the present results suggests that Seager et al.'s inferred values are low enough to allow interdecadal oscillations to occur.

To summarize the damping nature of horizontal diffusivity and the active nature of the mean-state advection, the Peclet number UL/K_h appears as a key parameter that controls the strength of the oscillations while the sensitivity of heat flux to SST, which is also very important, could be measured against advection by a number such as $\rho C_p U/\lambda$.

b. Description of the oscillation

We describe here the spontaneous oscillations that arise after the model has been spun up for thousands of years to steady equilibrium under restoring boundary conditions, the surface fluxes diagnosed and kept constant from there on. Having many runs of different resolution at our disposal, we have decided to illustrate the oscillation with a low-resolution run that can be easily reproduced with modest computing equipment (LR run, see Table 1). Of course, most of the analysis that follows has been reproduced at both low and medium resolution with no essential differences.

After a transient phase of 10–15 yr (described later) the perturbations grow and evolve into an oscillatory state (Fig. 2) that is independent of the initial perturbation that triggers the instability. The thermal anomalies are surface intensified and prominent in the northwest quadrant where the cooling is greatest. The current anomalies circulate along the contours of the temperature anomalies leaving little net eddy heat flux except near the western boundaries. The motions are geostrophic with little interior divergence. In much the same way as for the mean field, the anomalies of the divergence field are concentrated along the western and northern boundaries, upwelling and downwelling occurring there to connect the surface current anomalies to the deep current anomalies (of the opposite sense since no net barotropic transport is permitted).

In Fig. 2, the oscillation is shown initially in a state of weak thermal structure. On the other hand, the western boundary current (WBC) anomaly is strongly positive, being fed through upwelling along the coast. At $t = 3y$ the overturning is at its maximum and induces by $t = 6y$ a western intensified positive thermal anomaly

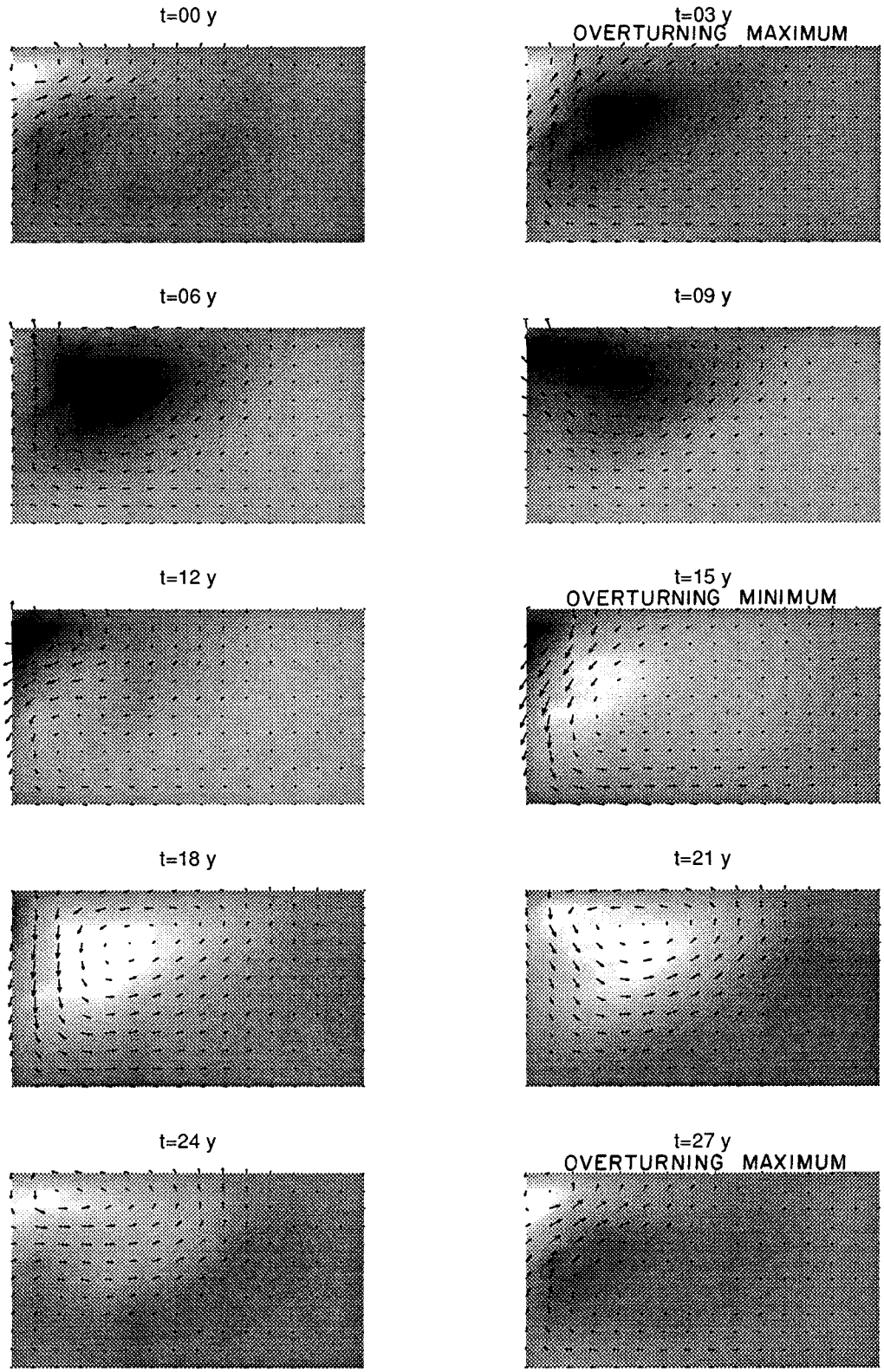

FIG. 2. The anomalies of surface temperature and surface currents during a full cycle of the oscillations (LR run). Basin size is 6000 km by 4500 km. Frames are 3 years apart. The overturning maximum (minimum) are indicated. The largest temperature and velocity anomalies correspond roughly to ½°C and ½ cm s^{-1}.

that covers half of the basin. The WBC anomaly is now oriented in the northward direction, part of the transport being supplied by the recirculation from the southern branch of the warm anomaly. The net overturning is now weaker and the situation has evolved from a prevailing vertical recirculation to a horizontal recirculation. From $t = 6y$ to $t = 15y$, a westward propagation of the warm anomaly is readily apparent. As the anomaly reaches the western boundary, the positive WBC anomaly disappears and reverses when the southward moving branch of the warm anomaly has reached the western wall. At this point, the vertical recirculation phase has resumed, but in the opposite direction with little horizontal interior recirculation, and the overturning is at its minimum. Then the negative WBC anomaly induces an offshore interior cold anomaly and the second half of the period of oscillation proceeds similarly. Maxima of domain-averaged potential and kinetic energy are associated respectively with the phase of the cold surface temperature anomaly and the overturning maximum. At $t = 3y$ the kinetic energy is maximum and the potential energy small (warm anomaly at surface), the situation reversing at $t = 15y$, about one-quarter of a period later.

Characteristic phase diagrams (Figs. 3a and 3b) in the x–t plane show a partition of the domain between the western third where the temperature oscillations are large and stationary and the remaining interior in which weaker thermal anomalies propagate westward against the mean eastward circulation in the northern part of the basin. In the meridional plane the anomalies of the WBC present also the character of stationary oscillations, intensified in the northern part of the domain (Fig. 3c). We next show that the distinct character of the oscillations (stationary versus propagative) in the different subregions is associated with vastly different heat transports. To concentrate on the mechanisms that drive the oscillations, we have examined the various terms of the equation governing the evolution of eddy temperature variance, the ''mean'' (overbar) being defined as the initial state of the constant flux experiment and the ''eddy part'' (prime) as departure from that initial state. The terms that dominate the growth of temperature variance are of the form $(-\overline{u_i'T'}\,\partial\overline{T}/\partial x_i)$. We observed that the terms containing vertical velocities are at least one order of magnitude less than the horizontal terms so that only the latter are illustrated in Fig. 4. The largest positive term appears to be the downgradient north–south eddy velocity temperature fluxes $(-\overline{v'T'}\partial\overline{T}/\partial y)$ that dominates in the northern third of the western boundary region. Since the mean $\partial\overline{T}/\partial y$ is negative, positive $\overline{v'T'}$ eddy fluxes at the western boundary appear to be at the heart of the existence of the oscillations. Fortunately, in these experiments, the regions with positive values (enhancing the anomalies) remain in the same location such that the time-averaged pattern is similar to instantaneous situations (on the contrary, under zonally uniform flux, temperature anomalies do travel around the

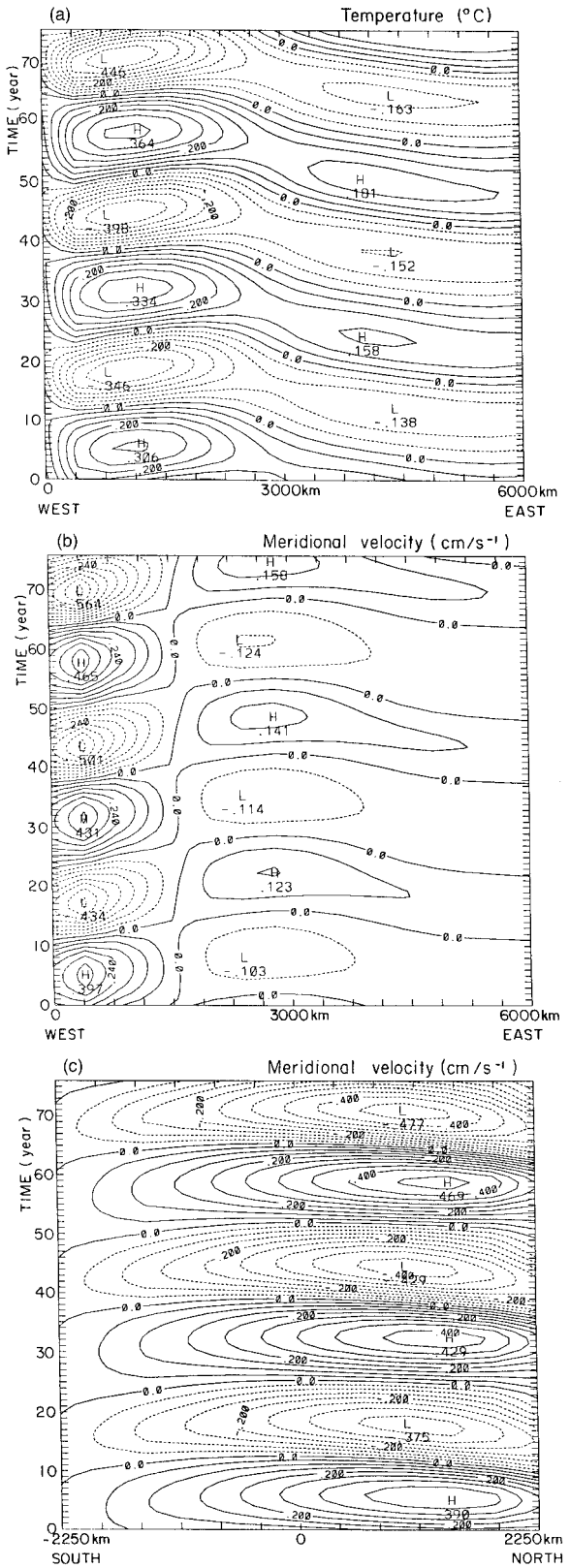

FIG. 3. Characteristic diagram in the x–t plane at a central latitude (a) and (b) in the y–t plane averaged over the western boundary current region (c).

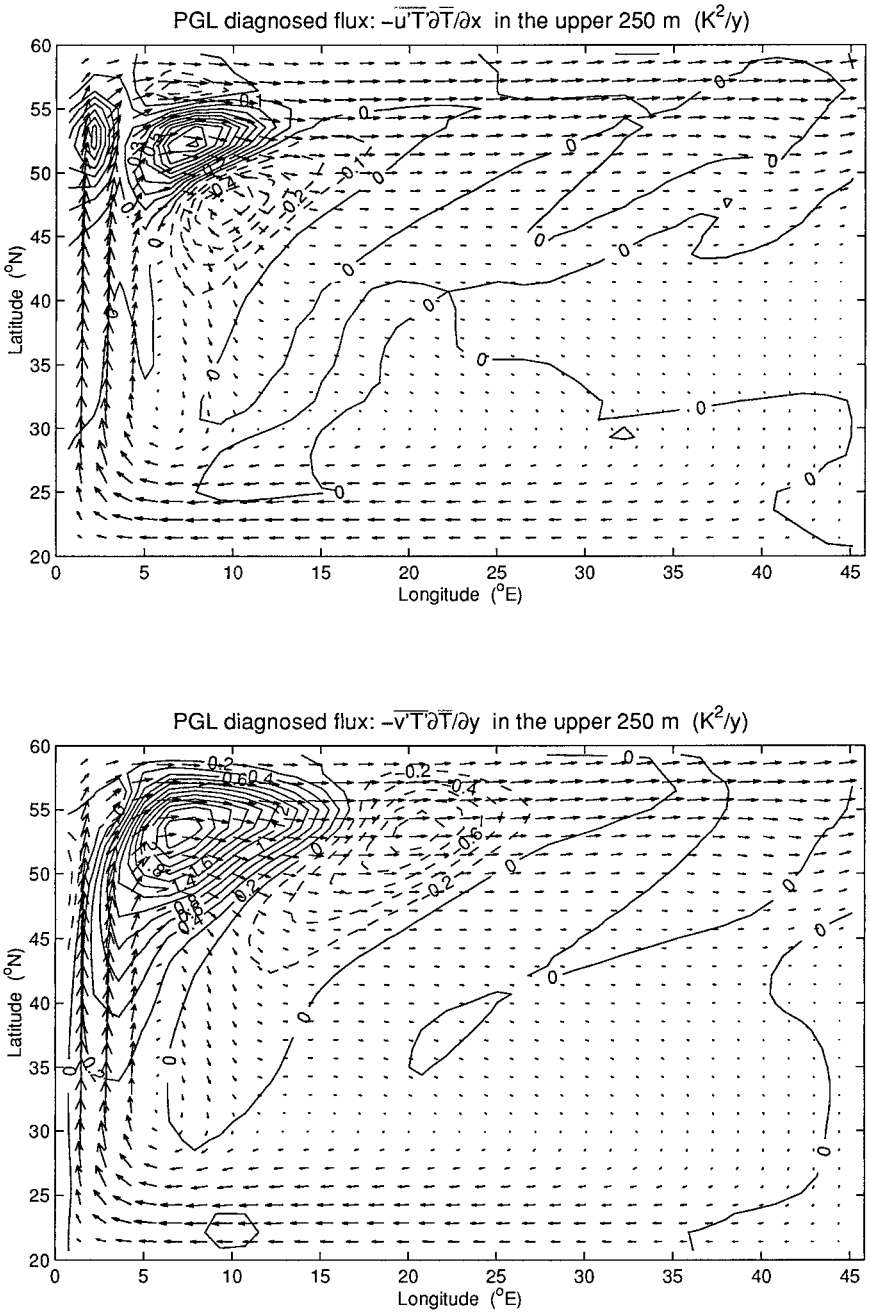

FIG. 4. The driving terms of the equation for temperature variance (HR run). Largest values occur for meridional fluxes in the northwest corner. The mean surface circulation that is superimposed shows that the most unstable region is located where the western boundary current turns eastward.

northern half of the basin and the region where the terms are positive do vary a lot along a period: the time-averaged pattern is then almost an order of magnitude smaller than the instantaneous snapshots and there is no well-defined driving area). The vertical structure of the $\overline{v'T'}$ eddy fluxes (Fig. 5) shows a surface and western intensification. The source region of the temperature variance coincides with the region of highest variance, in agreement with the stationary character of the oscillation in the northwest quadrant. Note, on the other hand, that the region of active interior propagation is associated with neutral or damping conditions. The existence of the near-surface positive $\overline{v'T'}$ is obvious from a comparison of the velocity and temperature time series in Figs. 3a and 3b. Near the western boundary meridional velocity and temperature correlate positively to a

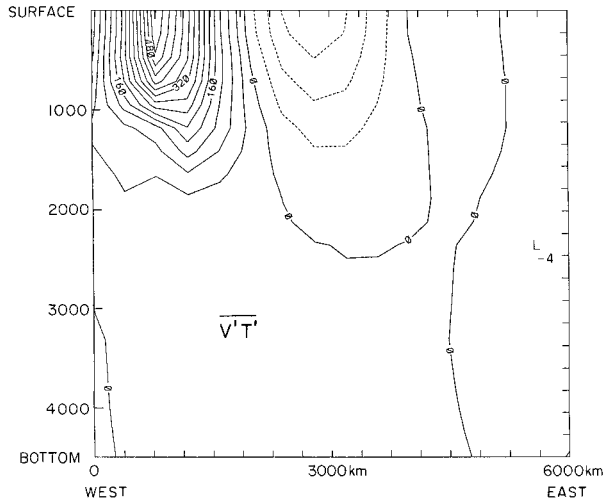

FIG. 5. Longitude–depth section of the meridional eddy fluxes $\overline{v'T'}$ at a central latitude (LR run). Values must be multiplied by 10^{-6} to obtain cm s^{-1} K units.

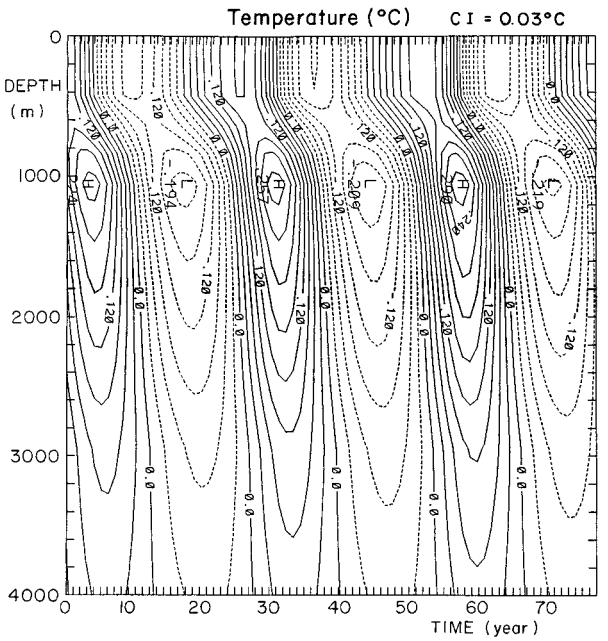

FIG. 6. Characteristic z–t diagram of the temperature field at a point situated in the unstable northwest corner region (LR run). The phase shift of the temperature distribution appears below the mixed layer.

high degree. The alongshore velocity being nearly geostrophic, the pressure extrema must lie to the east of temperature extrema. But since the determination of the hydrostatic pressure amounts to a simple integration of the temperature field, we are led to conclude that phase lags exist necessarily between temperatures at various depths, a confirmation being provided by Fig. 6. A significant vertical phase lag appears between the upper levels where convection dominates and the lower levels that lag by a quarter period. In contrast no such vertical shifts have been observed in the stable interior regions of the basin. This reminds one of the classical three-dimensional organization of baroclinically unstable perturbations of a zonal mean flow found under quasi-geostrophy. This familiar index of vertical phase lag is indeed necessary to allow downgradient eddy heat fluxes and release of potential energy when the flow is in approximate geostrophic and hydrostatic balance. These observations strongly suggest that the basic driving mechanism of the interdecadal oscillations is a local baroclinic instability of the western boundary region. This is in agreement with the previously mentioned nearly perfect phase opposition between the total potential energy and the kinetic energy (whose most part comes from the western boundary current, which is the largest contribution to the overturning). This association of vertical phase shift of the temperature distributions (required for the existence of downgradient eddy heat fluxes) and baroclinic instability helps to understand why the oscillations are observed to be damped by the convection scheme adjustment (HCW) since the latter acts to remove all vertical structure and phase lags necessary for the instability.

Given the horizontal boundary layer structure of the mean velocity profile near the western wall, and the complicated vertical distribution of the mean currents necessary to equilibrate the mean buoyancy loss at the surface, the basic state that we must deal with is far more complicated than the idealized zonal flows commonly used in baroclinic instability studies. However, the observed signature of the unstable motions are telling us that similar mechanisms operate albeit in a more complicated geometrical setting. The theoretical stability analysis of a mean state such as ours is a daunting perspective but we believe that we can increase our confidence and judge the efficiency of this baroclinic instability process by neglecting entirely the horizontal boundary layer structure of the mean state and obtaining growth rates of perturbations of a mean flow with realistic vertical shears only. This neglect of the horizontal structure is justified in part because no transfer from the mean kinetic energy is possible in our model equations (see section 3).

c. The initial instability

Since we allude to an instability mechanism as a way to sustain the oscillation against dissipation, it is worth looking at the transient phase immediately after the constant buoyancy flux has been switched on. When the previous run (with restoring surface boundary conditions) has been integrated for a long time (4000 yr), it appears that such a state is stable under a switch to flux conditions when initial external perturbations are absent. In some cases, Cai et al. (1995) and Greatbatch and Peterson (1996) have triggered the oscillations by modifying the forcing through a zonal redistribution of the surface heat flux. If an instability is at the heart of their

existences, such a procedure mixes the instability process and the spinup toward a new equilibrium with the modified forcing. A more traditional hydrodynamic stability practice is simply to add a small amplitude temperature perturbation to the initial state. A large range of initial perturbations (uniform temperature anomaly, different Fourier modes) have been added in the surface layers to show that over a 10–15-yr period the model state converges to the same previously described oscillation pattern (Fig. 2) independent of initial conditions, its full amplitude being obtained after several periods of oscillations. The transient state itself is patterned after the initial perturbations, and for instance with a uniform surface temperature anomaly, a front appears along the boundary of the convection region, but the anomaly progressively builds up in the northwestern quadrant to take the organized structure in Fig. 2. We have checked that the initial growth rates of a given perturbation during the first few years (as measured from the rms velocity or rms temperature) is independent of initial amplitudes to show that the instability is not of a finite-amplitude nature. We have also observed that, if the run with restoring boundary condition is further from equilibrium due to a shorter time integration, there is no need for external perturbations to trigger the oscillations following a switch to flux conditions. After only a few months, a perturbation builds up in the western boundary region where the buoyancy loss (negative heat flux) is largest. This is in the unstable region that was pointed out previously. We may expect that any imbalances in that region will amplify rapidly because of the baroclinic instability mechanism, which is strongest where mean vertical shears are largest. To equilibrate heat losses at the surface by horizontal heat advection of nearly geostrophic currents requires precisely the existence of such large mean vertical shears.

d. The structure of the interior temperature anomalies

The surface-intensified temperature anomalies that emerge under flux boundary conditions after a transient phase of several oscillation periods have a well-defined spatial and temporal structure that needs to be rationalized because it has already been shown in HCW that the observed quadrature between the western boundary current and the interior temperature gradients is central to the oscillation mechanism. One may expect then that the interdecadal periods found in GCMs are somehow related to the propagation time of the anomalies across the basin. To account for the wave propagation observed in Figs. 3a–c, we point out a very simple mechanism that relies on the existence of the underlying mean flow. The first obvious comment is that, away from the western boundary current region, the interior perturbations are geostrophic with fluid circulating along the isotherms in agreement with the weakness of the eddy fluxes (Fig. 4). In the present simulations there is a

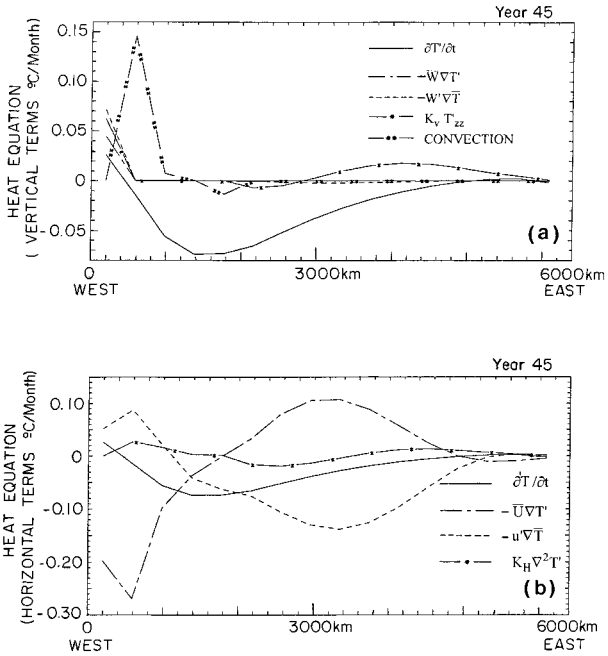

FIG. 7. The terms of the temperature equations at a given time (45 years after switch to flux condition, LR run) and at midbasin latitude as a function of zonal distance across the basin: (a) vertical terms, and (b) horizontal terms.

distinct boundary between the stable and unstable regions that disappears in experiments with zonally uniform flux for which the instability processes are not confined near the western boundary. To understand why the anomalies are not simply passively advected by the dominant eastward mean flow in the northern part of the basin, let us consider a warm perturbation embedded in a mean temperature gradient with temperature decreasing northward: west (east) of its center, northward (southward) surface-intensified perturbation velocities bring warm (cold) water that propagates the anomaly toward the west. This is exactly the classical argument for Rossby wave propagation with the mean meridional temperature gradient taking the role of the β effect. However, there is a mean eastward flow that is also associated with the mean temperature gradient so that a competition arises between mean advection and propagation. In some of our experiments, the westward propagation dominates while in others, the two effects nearly equilibrate. It is possible to gain some insight into such dynamics by considering the various terms of the temperature anomaly evolution equation. The leading terms (Fig. 7) governing the evolution of thermal anomalies are horizontal mean advection of anomalies, horizontal eddy advection of the mean temperature, and horizontal diffusion. What is worth remarking is that the vertical terms do not seem to play an important role in the interior, and horizontal terms dominate over vertical ones (for both advection and diffusion). The two advective

terms that dominate the interior heat balance are, therefore,

$$\overline{U}\frac{\partial T'}{\partial x} + v'\frac{\partial \overline{T}}{\partial y}.$$

Because both mean flows and anomalies are surface trapped, we look for separation of variables and assume exponential dependence in the vertical as

$$\left.\begin{aligned} \overline{U}_z &= \overline{\kappa}\,\overline{U} \\ v'_z &= \kappa' v' \end{aligned}\right\} \quad \text{with } z \text{ increasing upward.}$$

Using the thermal wind equation, the above transforms into

$$\frac{g\alpha}{f}(\kappa'^{-1} - \overline{\kappa}^{-1})\frac{\partial \overline{T}}{\partial y}\frac{\partial T'}{\partial x},$$

where α is the thermal expansion coefficient. The interior temperature anomalies obey to leading order

$$\frac{\partial T'}{\partial t} - c\frac{\partial T'}{\partial x} = K_H \nabla^2 T',$$

where

$$c = \frac{g\alpha}{f}(\overline{\kappa}^{-1} - \kappa'^{-1})\frac{\partial \overline{T}}{\partial y}$$

and the diffusion term has been added for completeness. This relation describes a zonal propagation, the y and z dependence appearing parametrically. Since the mean temperatures decrease northward, the observations of the LR runs (Fig. 3a) that indicate dominant westward propagation in the interior ($c > 0$) correspond to the case for which eddy advection of mean temperature dominates over mean zonal advection of the anomalies, a result consistent with what is observed in the heat balance. The above relation implies that this particular regime occurs because the waves are shallower than the mean flow ($\kappa' > \overline{\kappa}$). These waves represent a particularly simple subset of nondivergent surface-trapped PV waves. Note that the restoring force in this example comes from the gradient of potential vorticity provided by the mean meridional temperature gradient, and not by β, so that such waves are allowed on f planes as well. The effect of lateral diffusion is to allow for a second mode, which in the case of positive c is propagating eastward. When the Peclet number (based on c) is large, the mode has a short eastward decay scale K_H/c so that it is expected to play a role on the western side of a basin.

3. Baroclinic instability in simplified analytical and numerical models

Motivated by the previous idea that baroclinic instability of the western boundary current region provides the energy source for the oscillations to be maintained against dissipation, we wish to evaluate its strength as gauged by the growth rate of unstable perturbations on a realistic mean flow. Although baroclinic instability is of central importance to eddy production in midlatitude oceans and atmospheres at the scale of the Rossby radius of deformation, the question is really whether it may be active at scales much beyond the Rossby radius. The Phillips (1954) two-layer model of a zonal flow shows this not to be the case because of the existence of a low wavenumber cutoff that suppresses the instability of the largest scales. With such a crude vertical resolution, the barotropic and baroclinic modes are incapable of strong interactions essentially because their timescales become too unequal, the barotropic timescale becoming very much smaller than the baroclinic one for large spatial scales. Models with continuous vertical structure, such as Eady's or Charney's, do not have such large-scale cutoff essentially because higher baroclinic modes are involved in the interaction. Studies of the THC at coarse resolution in PG models neglect entirely relative vorticity, an assumption that is justified by the excellent comparison with full PE models. The consequence is that the barotropic mode becomes entirely diagnostic and hence unable to participate in the baroclinic instability (in fact, in our box geometry, buoyancy-driven simulations devoid of bottom friction, it is exactly zero). Therefore, we are led to think that large-scale baroclinic instability is allowed if interactions between higher baroclinic modes are possible and one may expect models of the THC under constant flux to behave in a rather different way if the number of layers in the vertical is smaller or larger than two, a point to which we will come back. The existence of baroclinic instability at planetary scales has already been shown theoretically by Colin de Verdière (1986) and Cavallini et al. (1988). However none of the thermohaline circulation simulations reported so far in the literature ever mentioned active baroclinic instability. Given that the unstable perturbations are surface intensified, we believe that the reason lies with the presence of the strong restoring boundary conditions on surface temperature and/or salinity. When the surface fields are restored on a timescale shorter than the growth rate of the baroclinically unstable perturbations, there is no way that the instability may amplify. The situation is vastly different, however, under flux boundary conditions (or mixed boundary conditions) because no such external controls exist to damp the unstable waves. At this point what is needed to confirm this idea is to find out the growth rates for large-scale flows with realistic vertical shears and judge the vigor of the instability by comparing them with the damping timescales associated with the restoring boundary conditions or lateral diffusivity present in numerical calculations. Although dissipation is small, we are looking for a mechanism that applies for β planes as well as f planes. In the absence of relative vorticity advection, friction is needed to break the geostrophic constraint. Given that the instability appears in the simulations in regions such as near the western boundary

where friction, although small, is not negligible, leads to study the problem under PG dynamics with Laplacian friction included. Of course, the experiments show the instability of a very complicated basic state, that of a western boundary current hugging a wall with vertical as well as horizontal shear—a situation far too complicated to study analytically. Consequently, we simplify the picture and consider the stability conditions in an unbounded domain.

a. Linear instability calculations

It is simpler to study the instability using constant density layers of varying depth in the vertical. When the advection of momentum is neglected, it is easy to show that the only remaining nonlinearity, the advection of layer thickness, vanishes identically in the two-layer case. Consequently, as argued previously, the interaction of higher baroclinic modes is a necessary condition for baroclinic instability and a three-layer model provides the essentials. The three layers are of density ρ_i and thickness h_i with $i = 1$–3 numbering the layers from the top. A rigid lid imposed at the surface and a flat bottom gives

$$\sum_{i=1}^{3} h_i = H,$$

where H is the uniform fluid depth. Since the barotropic mode is zero, the condition $\Sigma\, h_i\mathbf{u}_i = 0$ reduces the problem from three to two degrees of freedom in the vertical, that is, the first and second baroclinic modes. The pressure being hydrostatic, the geostrophic velocity can be expressed in terms of the layer thickness after use of the two above conditions (here $\mathbf{j}$ is a vertical unit vector):

$$f\mathbf{j} \times \mathbf{u}_{g1} = g_1'\boldsymbol{\nabla}h_1 + g_3'\boldsymbol{\nabla}h_2$$
$$f\mathbf{j} \times \mathbf{u}_{g2} = g_3'\boldsymbol{\nabla}(h_1 + h_2). \qquad (1)$$

Because the two upper layers are supposed to represent the main thermocline, h_1 and h_2 have been assumed small compared to H. The reduced gravity g_1' is $g(\rho_3 - \rho_1)/\rho_0$, and g_3' is $g(\rho_3 - \rho_2)/\rho_0$. Although most studies of baroclinic instability choose basic states that consist of zonal flows, the experiments do not suggest this to be a particularly good choice since the predominantly meridional flows along the western boundaries appear to be potentially the most unstable. On a β plane departing from the zonal flow assumptions implies the instability of forced flows. However, when the buoyancy forcing is stationary, the case of interest here, the mass conservation equations for the perturbations do not contain the forcing and hence for each layer the linearized perturbations (with primes) obey

$$\left(\frac{\partial}{\partial t} + \mathbf{U}_i \cdot \boldsymbol{\nabla}\right)h_i' + \mathbf{u}_i' \cdot \boldsymbol{\nabla}H_i + H_i\boldsymbol{\nabla} \cdot \mathbf{u}_i' + h_i'\boldsymbol{\nabla} \cdot \mathbf{U}_i = 0, \qquad (2)$$

where $\mathbf{U}_i$ and H_i represent the velocity and layer thickness of the mean state.

A first remark concerns the last term $h_i'\boldsymbol{\nabla} \cdot \mathbf{U}_i$, which is not present in "free" baroclinic instability calculations, the mean flow being assumed divergenceless. However, suppose that the mean flow is horizontally convergent (as would happen in the neighborhood of solid boundaries in our simulations for instance), then we can expect exponential growth of the layer i at a rate that is just $-\boldsymbol{\nabla} \cdot \mathbf{U}_i$. The example of the surface layer in a cooling region illustrates the effect of that sole term: if the full depth h_1 becomes larger than the mean depth H_1, then the convergence of mass flux exceeds the constant diapycnal mass loss to the lower layer (the buoyancy forcing), and the lower interface will continue to deepen in an unstable fashion. Although the divergence of the mean flow in the numerical experiments is nonzero, the previous examination of the leading terms in the heat balance has shown that the terms containing vertical velocities are small, and therefore we simply assume that the term associated with mean flow divergence does not play a major role in the instabilities that are observed. We assume a constant mean flow in the following local analysis and neglect entirely the mean flow divergence from now on.

In the inviscid case (i.e., with geostrophic velocities) it is not difficult to show that a necessary condition for linear perturbations (varying as $e^{i(\mathbf{K} \cdot \mathbf{x} - \omega t)}$) to be unstable is that the quantity $\mathbf{K} \times \boldsymbol{\nabla}(f/H_i)$ changes signs between layers one and two, where f/H_i is simply the mean PV. If the mean PV gradients are parallel, then instability requires them to be opposite in each layer, while, if they rotate from one layer to the next, it is always possible to find a direction of a wavenumber vector that allows the quantity to change sign. This necessary condition appropriate to the inviscid case is not very useful in the present discussion. Although Colin de Verdière (1986) has shown that β was essential to the existence of the instability in the inviscid case, the numerical experiments show without doubt that interdecadal oscillations exist on an f plane (see HCW). If the claim is made that large-scale baroclinic instability drives these oscillations, we need to demonstrate that it works on f planes as well as on β planes. There can be no conversion from potential to kinetic energy without vertical velocities that a strict adherence to f-plane geostrophy precludes. So, paradoxically, the consideration of friction by allowing divergence of the perturbations actually broadens the conditions for the existence of baroclinic instability at large scales. It will, in particular, allow for nonzero growth rates in f-plane cases. Adding a Laplacian dissipation in the horizontal momentum equations makes a correction to the geostrophic perturbation velocities introduced in (1) as

$$\mathbf{u}_i = \frac{\mathbf{u}_{gi} + E\mathbf{j} \times \mathbf{u}_{gi}}{1 + E^2}. \qquad (3)$$

The velocity amplitudes above are function of time only, and $E = A|\mathbf{K}|^2/f$ is a horizontal wavenumber-dependent Ekman number. Knowing for each wavenumber how the perturbation thicknesses relate to velocities allows one to cast Eq. (2) in terms of thickness only, and it is then an algebraic task to calculate eigenvalues. To carry out the calculation, a mean flow profile must be chosen. The most unstable region delineated in our experiments by the maximum downgradient eddy heat fluxes in Fig. 4 appears to be slightly north of the midbasin position in the western boundary region. The surface-trapped current has a northeastward direction (LR run) that veers sharply to the left going deeper, a veering that is entirely consistent with the warm water transport that the WBC must carry out to equilibrate surface cooling. The thermocline deepens to the east and even more to the south due to the large zonal shear that is observed. To represent the full vertical structure with a three-layer model, we averaged mean velocities and densities over the thicknesses chosen for the layers. Figures 8a and 8b show the growth rates as a function of horizontal wavenumbers amplitude for the LR and HR cases, respectively. Typical growth rates of one to two cycle per year are obtained for realistic values of the friction coefficient and mean vertical shear. These values have to be compared with the damping rates associated with lateral diffusion that increase with the square of wavenumber. Taking it into account, it is readily seen from the figures that baroclinic instability is allowed in a window that opens at the low wavenumber end. The horizontal scales of the perturbations (half wavelengths say) that maximize the growth rates are typically of the order of ten Rossby radius of deformation (the Rossby radius is 70 km in these examples).

How do these growth rates compare with the observations in the numerical model? First, we observe qualitative agreement between these values and the numerically derived damping rate thresholds (section 2a), beyond which the interdecadal oscillations are not observed. Second, we have added some very small perturbations (0.01°C) to the mean state of the model (LR run) to satisfy the conditions of linearization that underlie the above analytic calculations. The model is then integrated for about 20 years and an appropriate measure of the growth rate over that period is obtained by dividing the temperature variance rate of change by the variance. This quantity is then averaged in the vertical and over time. Irrespective of the initial shape of the surface perturbations, positive growth rates for the temperature variance of 1–2 cycles per year are observed to develop invariably in the northwest quadrant (Fig. 9). To compare with the analytically derived growth rates, such values would have to be divided by 2. We consider this favorable comparison between the growth rates observed in the numerical experiments and those from the three-layer model as another piece of evidence that adds to our case. It is tempting to use these types of calculations to try to predict when interdecadal oscillations should appear in a GCM simulation. In principle, the computed growth rates that depend on friction and shear could be compared with the timescales associated with the damping processes (lateral diffusion and restoring boundary conditions). The effect of the intensity of the shear is straightforward. For a given friction coefficient, Fig. 8d shows that there is a linear relation between the growth rate and the upper vertical shear. However, the dependence on the friction coefficient is considerably more complex. In the particular example of Fig. 8c, it is observed that the growth rates go to zero as the friction increases but there is a second cutoff if the friction is too low (outside the range of values explored in the figure) because, as argued previously, the perturbations must have some vertical velocities, hence divergence for the instability to proceed. Because of this difficulty and because of the calibration of the vertical structure that is inherent in this kind of comparison, it is difficult to provide a simple criterion that would offer quantitative predictive skills for the presence of interdecadal oscillations.

b. Layered-model numerical calculations

To further assess the efficiency of the proposed instability process, a number of exploratory runs have been carried out with the same layered formulations as above, the diabatic forcing being introduced through mass exchange between the layers. To do so, the mass conservation equations for each layer are advanced in time and the velocities are then computed from the height fields assuming PG dynamics (Rayleigh friction was used for simplicity). We selected rather small forcing amplitudes to prevent surfacing of the layers so that the solutions are quasi linear and remain close to the regime of the previous analytical calculations. When no wind forcing is applied, the solution is internal, so that a two-layer case (with a shallow upper layer) becomes equivalent to a $1\frac{1}{2}$-layer model. We started experimenting with such a model and found that we could never reproduce an oscillatory solution. Although it is possible in principle to cause an instability of a different nature through the divergence of the mean flow, the last term in (2) already discussed, its growth rate O(1 cycle/16 years) is too small and does not exceed the critical value imposed by the horizontal mixing. The next step was then to add an intermediate layer, necessary for baroclinic instability to occur, as recalled previously. With this $2\frac{1}{2}$-layer configuration, we managed to reproduce unsteady behavior with typical variability periods around 25 yr. The important point that we want to stress lies with this fundamental difference between $1\frac{1}{2}$- and $2\frac{1}{2}$-layer models regarding the variability under constant flux: while the $1\frac{1}{2}$-layer model never induced any variability after a few hundred years of integration, the $2\frac{1}{2}$-layer model driven by exactly the same forcing within the same geometry exhibited irregular decadal fluctuations of significant amplitudes over thousands of years.

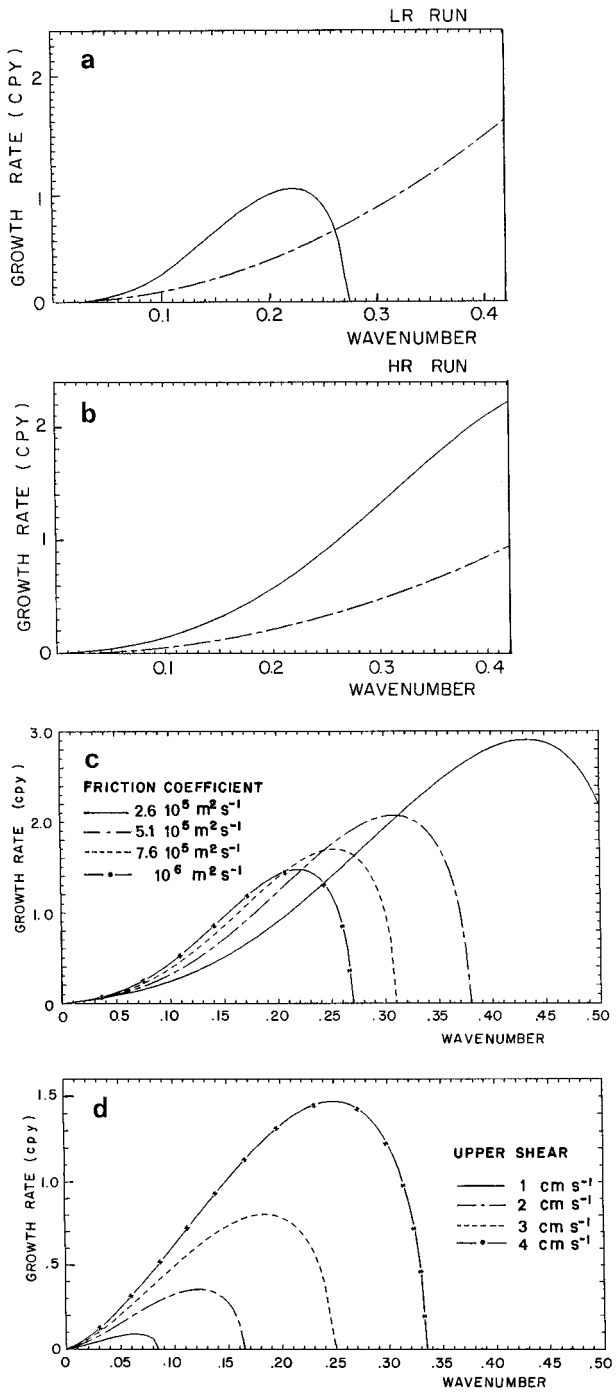

FIG. 8. The growth rate (cycle/year) in the three-layer model (solid line) as a function of wavenumber scaled by the inverse Rossby radius of deformation $(g_3'H)^{1/2}/f$. This must be compared with the inverse timescale associated with horizontal diffusion (dashed line). Common parameter values are $h_1 = 100$ m, $h_2 = 200$ m, and $h_3 = 4200$ m. (a) LR case: $U_1 = 1.5$ cm s^{-1}, $V_1 = 4.5$ cm s^{-1}, $U_2 = 0$, $V_2 = 1.5$ cm s^{-1}, $g_3' = 0.9 \times 10^{-2}$ m s^{-2}, $g_1' = 1.8 \times 10^{-2}$ m s^{-2}. (b) HR case: $U_1 = 2.8$ cm s^{-1}, $V_1 = -1.2$ cm s^{-1}, $U_2 = 1.9$ cm s^{-1}, $V_2 = -0.5$ cm s^{-1}, $g_3' = 0.8 \times 10^{-2}$ m s^{-2}, $g_1' = 0.84 \times 10^{-2}$ m s^{-2}. (c) The LR case for different values of the Laplacian friction coefficient (2.6×10^5, 5.1×10^5, 7.6×10^5, 10^6 m^2 s^{-1}). Growth rate variations with respect to the friction coefficient is monotonic with small growth rate associated with large friction coefficient at a given wavenumber. (d) for different values of the upper-layer meridional velocity (1, 2, 3, and 4 cm s^{-1}). All other velocity components are zero. The friction coefficient is set at 10^5 m^2 s^{-1}. The growth rate increases monotonically with the shear at a given wavenumber.

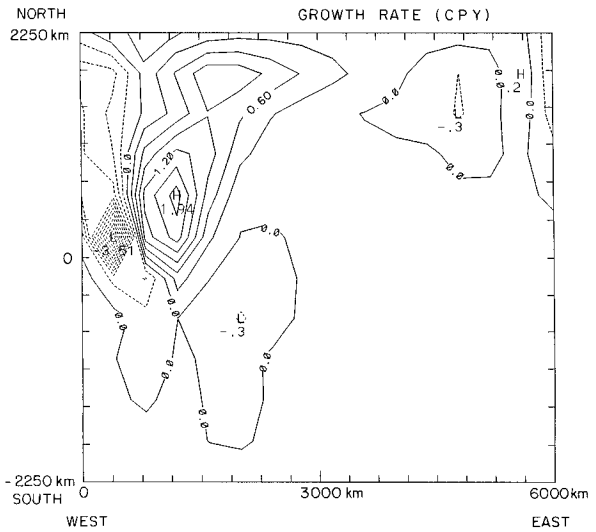

FIG. 9. The growth rates inferred from a 20-year integration after a switch to constant flux and an addition of a small (0.01°C) temperature perturbation of the form $\sin(4x) \times \sin(y)$ (LR configuration). Positive growth rates between 1 and 2 cycle/year stand out in the northwest quadrant. The contour interval is 0.3 cycle/year.

The results of these numerical experiments again support our proposal that long-wave baroclinic instability drives the decadal variability in ocean models forced by quasi-steady surface buoyancy fluxes.

4. A coupled ocean–atmosphere box model

Given, first, that the ocean model is observed to oscillate on interdecadal timescales under constant heat flux and, second, that the driving of the oscillation is linked to the instability of the western boundary current region, we propose to go a step further in discussing the implications of this mechanism for coupled ocean–atmosphere models. The coupled model study of Delworth et al. (1993) shows variability on a 50-yr timescale whose origin is associated with variations in the intensity of the THC, resulting in western-intensified large-scale SST anomalies that bear encouraging similarities with oceanic observations. Indeed, Greatbatch and Zhang (1995) and Greatbatch and Peterson (1996) made the case that what Delworth et al. saw in their model was oceanically driven with atmospheric perturbations following the oceanic perturbations generated by a mechanism similar to what is found in coarse-resolution ocean-only models. Intermediate in complexity, Chen and Ghil's (1996) study confirmed that

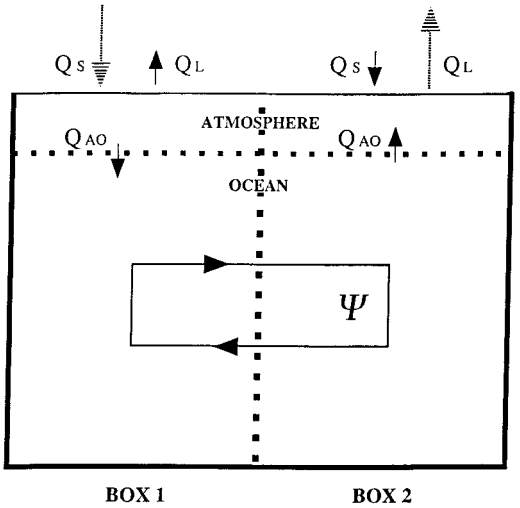

FIG. 10. The geometry of the four-box ocean–atmosphere model.

the oscillations persist almost unaltered in the situation of an ocean model coupled to an atmospheric EBM. The objective pursued in this section is to use our knowledge of the phase relations between the oceanic meridional temperature gradient and the overturning of the THC to explore, through a coupled ocean–atmosphere box model, the conditions under which what has been observed in the ocean model can persist with an atmosphere overhead. Of course, the box models have no quantitative ambitions but allow instead to account for interactions of large-scale feedbacks that may be pursued later in GCMs. This particular box model has already been considered by Marotzke and Stone (1995), who used it to discuss the stability of equilibrium solutions and the proper way of making flux corrections that preserve the correct sensitivity of model climates. The emphasis here is not on the equilibrium solutions but on oscillatory states and on the dynamical choices that allow them.

Consider then the situation in Fig. 10, where two atmospheric boxes are coupled to two oceanic boxes. The atmospheric boxes exchange heat externally through incident solar flux Q_S and infrared back radiation flux Q_L to outer space. We assume that the heat flux Q_{AO} between ocean and atmospheric boxes is equal to $\lambda(T_A - T)$, that is, proportional to the difference between ocean and atmosphere temperatures. The characteristic response time of the atmosphere to an oceanic thermal anomaly is $\rho C_p h/\lambda$, ratio of atmospheric thermal inertia to the coupling coefficient. While a value of 40 W m^{-2} K^{-1} for the coupling coefficient is customary in ocean modeling (following Haney 1971), Seager et al. (1995) have suggested that this is too large and that values of 15 W m^{-2} K^{-1} or less are more appropriate. With such values, a height of the tropopause, $h = 10$ km, and a heat capacity $(\rho C_p)_A = 10^3$ J m^{-3} K^{-1}, the response time of the atmosphere turns out to be of the order of one week, so small compared to the characteristic oceanic interdecadal timescales that it is an excellent approximation to assume that the atmosphere is always in equilibrium balance with its fluxes. This assumption is likely to be robust against any known uncertainties in the coupling coefficient λ, so that for each of the two atmospheric box i we write

$$Q_S^i - Q_L^i - Q_{AO}^i + K_A(T_A^j - T_A^i) = 0, \qquad (4)$$

with $j = 3 - i$ and where K_A, the turbulent heat diffusion coefficient (in W m^{-2} K^{-1}), parameterizes the turbulent exchange by the large-scale atmospheric eddies. The heat conservation equation for the oceanic box i is

$$(\rho C_p h)_0 \frac{\partial T^i}{\partial t} = Q_{AO}^i + \frac{(\rho C_p)_0}{A_i}\psi(T^j - T^i) + K_0(T^j - T^i), \qquad (5)$$

where A_i is the horizontal area of box i, ψ (m^3 s^{-1}) and K_0 (W m^{-2} K^{-1}) are, respectively, the THC overturning strength and the oceanic heat diffusivity, the two processes that strive to homogenize the heat content between the two oceanic boxes.

Since our main interest is not in the mean state but in the interdecadal oscillations, we assume, in the following, a known mean state and concentrate our attention on the *variability* of the *meridional temperature differences,* respectively $x = T^1 - T^2$ for the ocean and $y = T_A^1 - T_A^2$ for the atmosphere. The choice of a timescale τ (=1 yr) allows us to rewrite Eqs. (4) and (5) for the temperature differences as

$$2Q_S - By - \lambda(y - x) - 2K_A y = 0 \qquad (6)$$

and

$$\dot{x} = \alpha\lambda(y - x) - 2\psi x - 2\alpha K_0 x, \qquad (7)$$

where $2Q_S$ is the difference of solar flux between the tropical and polar box, the infrared flux Q_L has been linearized around a mean state as $A + BT_A$, α is $\tau/(\rho C_p h)_0$, and ψ is now scaled by hA_i/τ (we have also assumed for simplicity equal areas for the two boxes). From (6), one obtains immediately

$$y = \frac{\lambda x + 2Q_S}{\lambda + B + 2K_A}, \qquad (8)$$

showing that the atmosphere temperature anomalies depend linearly on oceanic temperature anomalies, being reduced by a factor $\lambda/(\lambda + B + 2K_A)$, which illustrates the dissipative roles of large-scale atmospheric turbulence and infrared back radiation. Values of B and K_A are indeed significantly smaller than λ, yielding a reduction of atmospheric anomalies of about 10% compared to the oceanic anomalies. The ocean–atmosphere heat flux difference between the boxes ΔQ_{AO} can then be expressed entirely in terms of the oceanic variables x as

$$\Delta Q_{AO} = \frac{\lambda}{\lambda + B + 2K_A}[2Q_S - x(B + 2K_A)]. \qquad (9)$$

Quite naturally the flux driving the ocean is made of

two parts, a constant part that is slightly reduced compared to the solar flux at the top of the atmosphere and a variable part that summarizes the dissipative effects of atmospheric turbulence and infrared back radiation to damp the oceanic anomalies. When y is eliminated in (7), a single equation for x is obtained:

$$\dot{x} = \frac{2\alpha\lambda Q_S}{\lambda + B + 2K_A} - 2\psi x - \delta x, \tag{10}$$

where the last term of this equation δ is

$$\delta = \alpha\left[2K_0 + \lambda\frac{B + 2K_A}{\lambda + B + 2K_A}\right].$$

The parameter δ summarizes the dissipation of the oceanic temperature anomalies through both oceanic and atmospheric eddies and infrared back radiation. Marotzke and Stone's values of B and K_A of 1.7 and 1.3 W m^{-2} K^{-1}, respectively, induce a damping timescale of about 15 yr, showing the weakness of the temperature dissipation that the ocean box sees. Remarkably enough, each of the three contributors to the dissipation are observed to have the same order of magnitude[1] in the present state of the climate system. Linearizing (10), the sought after thermodynamic equation for the deviation x' from the temporal mean state (denoted by an upper bar) is

$$\dot{x}' = -2\overline{x}\psi' - \delta x'. \tag{11}$$

Consistent with the remark of HCW that in the numerical experiments $x'/\overline{x} \ll \psi'/\overline{\psi}$, the term $\overline{\psi}x'$ has been neglected. It would be desirable to have a box model that would correctly represent both the mean and the time variable part, but unfortunately, we have not yet been able to do this and therefore leave this assumption as a necessary adjustment of the box model to the 3D model results. We take the mean as obeying these assumptions and restrict our focus to a discussion of the temporal variability around a prescribed mean state. This difficulty should not be too surprising in view of the observation of Winton (1996) that interdecadal oscillations do not occur in 2D models.

To close the system we need another equation for the dynamics. Originating from Stommel's (1961) study, the box models that have attempted to describe the mean state usually assume a linear diagnostic relation between the overturning and the meridional density gradient. As shown previously, the observed behavior of the 3D models at interdecadal timescales shows that a *time lag* exists between the *overturning* and the *interior meridional temperature gradient* so that the relation between the overturning and the density gradient becomes prognostic (HCW):

$$\dot{\psi}' = Kx' + \mu\psi' - \gamma\psi'^3. \tag{12}$$

The second term that has been added on the rhs of (12) represents, in a rather ad hoc way, the linear growth rate due to baroclinic instability in the western boundary current region, while the last is a saturation amplitude limiting term[2] that drives the system back to stability at large amplitudes. The combination of Eq. (11) and (12) forms a dynamical system in the plane of the two active variables y'–x' that we argue produces the qualitative physics of interdecadal oscillations in the coupled ocean–atmosphere system. We assume that K and $\overline{x}$ are given and study this simple system as a function of the dissipation δ (typically much smaller than 1) and of the overall linear growth rate μ of the instabilities. The last parameter γ is simply there to parameterize the stabilizing effect of the perturbations at large amplitudes.

The Poincaré–Bendixon theorem (see, e.g., Nayfeh and Balachandran 1995) indicates that a necessary condition for the existence of periodic solutions is that the divergence takes both signs in the (x', ψ') plane:

$$\frac{\partial\dot{x}'}{\partial x'} + \frac{\partial\dot{\psi}'}{\partial\psi'} = \mu - \delta - 3\gamma\psi'^2.$$

So that periodic solutions can exist only if μ is greater than δ. Now when the product $\mu\delta$ is smaller than unity, there is no mean state other than zero in these perturbations equations, a desirable situation since we study deviations from the mean. The linear stability properties near the origin can be studied assuming perturbations varying as e^{st} and we obtain the eigenvalues equation

$$s^2 + s(\delta - \mu) + \omega_0^2 - \mu\delta = 0$$

in which the notation $\omega_0 = (2K\overline{x})^{1/2}$ introduces the oscillation frequency of the system when μ and δ are zero. When $\delta + \mu < 2\omega_0$, a condition satisfied in the weakly dissipative, weakly unstable case that we study, it is readily seen that μ must be larger than δ for the origin to be unstable. In this case the roots have an imaginary part so that the origin is a spiral source. A finite-amplitude limit cycle appears and Hopf bifurcation occurs in this parameter range of small dissipation as μ becomes larger than δ. As the state system separates from the origin, the $\gamma\psi'^3$ limiting term becomes important and brings the system back toward the origin. Choosing a mean-state meridional temperature contrast $\overline{x}$ of 10°C, $\lambda = 15$ W m^{-2} K^{-1}, $B = 1.7$ W m^{-2} K^{-1}, $K_A = 1.3$ W m^{-2} K^{-1}, and an oceanic turbulent diffusivity of 10^3 m^2 s^{-1} leads to an overall coefficient of dissipation $\delta = 6.8 \times 10^{-2}$ yr^{-1}. Once the mean state is chosen, the constant K fixes the period. The amplitude of the oscillation is plotted on Fig. 11c as μ increases from zero while keeping all other parameters constant. The familiar $\sqrt{\mu}$ be-

[1] Oceanic mesoscale diffusivity observed to be $O(10^3\ \mathrm{m^2\ s^{-1}})$ translates to a heat conductivity $10^3 \times (\rho c_p h)_o/A$ of 1 W m^{-2} K^{-1} for an area A of 2000 km by 2000 km.

[2] The precise form of this last term is at this point arbitrary and other alternatives exist such as $-\gamma x^2\psi$, in which case the dynamical system would be a Van der Pol oscillator in the limit $\delta \to 0$.

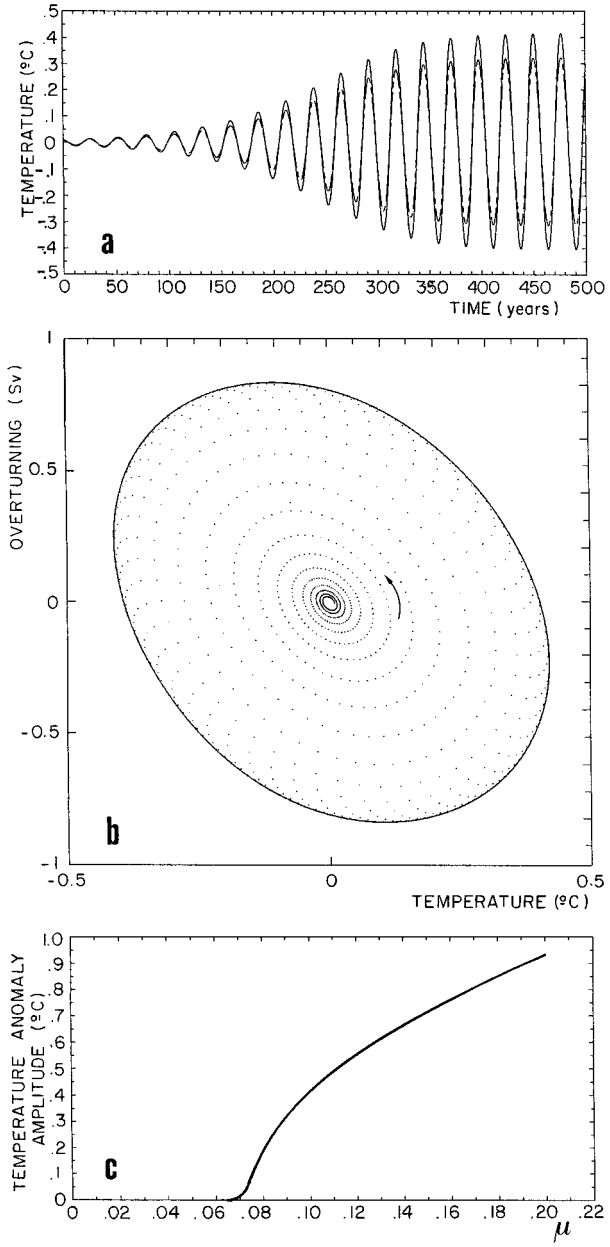

FIG. 11. Results of the ocean–atmosphere four box model for the following conditions above critical: $\mu = 0.1$, $\delta = 6.8 \times 10^{-2}$, $\gamma = 1305$. (a) Oceanic (solid) and atmospheric (dashed) temperature anomalies as a function of time. (b) The limit cycle in the phase plane of meridional temperature difference and overturning streamfunction. (c) Bifurcation diagram of the amplitudes of the oscillations against the growth rate of baroclinic instability.

havior indicates that a genuine Hopf bifurcation occurs at $\mu = \delta$. With the above parameters, the limit cycle of the solution is shown in Figs. 11a and 11b for $\mu = 0.1$ yr^{-1}. A small initial thermal anomaly of 0.01°C grows to reach a limit cycle over several oscillation periods. Such values of μ for the box model represent growth rates averaged over the 3D domain in the numerical experiments, and it is much less than what baroclinic instability can produce locally (that was seen to be in the range of one cycle per year). It is this *low value* of the overall *growth rate*[3] needed to sustain a finite-amplitude oscillation in the box model that gives us confidence that the mechanism that we propose might play a role in the climate system. The instability mechanism observed in the northwest quadrant of the idealized models is the candidate to generate SST anomalies on a timescale short compared to the cumulative dissipation effects by atmospheric–oceanic turbulence and infrared back radiation. On these interdecadal timescales the atmosphere reacts passively, enslaved to the periodicities of the oceanic temperatures (Fig. 11a). The transition at the bifurcation of this four-box model is very similar to the transition that has been observed in the numerical experiments with respect to the horizontal Peclet numbers described in section 2a, Fig. 1c. Similarly, Chen and Ghil conjectured on such a Hopf bifurcation from the square root dependency of the THC amplitude in their numerical model when the atmospheric diffusivity K_A and the coupling coefficient λ were varied. The present four-box model shows that the transition from a steady to an oscillating regime depends ultimately from a single parameter δ that integrates the various damping processes of temperature anomalies, atmospheric and oceanic turbulence, infrared back radiation, and air–sea coupling coefficients. The decrease of either of λ, K_A, or K_O coefficients lowers the overall dissipation parameter δ. When δ is below the growth rate μ of the large-scale baroclinically unstable modes of the THC circulation, spontaneous oscillations emerge. What has been added to Chen and Ghil's picture is that the properties of the limit cycle are governed by the instability of the mean state of the THC as measured by this instability parameter μ and by the dissipation as measured by this parameter δ.

5. Concluding remarks

The present analysis of coarse-resolution ocean models suggests that, in the real ocean, western boundary current regions may be baroclinically unstable at scales beyond the Rossby radius and drive interdecadal oscillations. In ocean models driven by constant buoyancy fluxes with sufficiently low subgrid-scale diffusivity, phase lags appear (i) between temperature and pressure in the horizontal direction and (ii) between temperatures at different depths. Such phase lags are precisely what is required to produce downgradient eddy heat fluxes under nearly geostrophic dynamics. These fluxes release the mean potential energy in the outflow regions of the surface western boundary current. We associate this powerful energy source to the *wavemaker* of the inter-

[3] With the value of dissipation provided by the box model, the critical amplification factor δ^{-1} is about 15 yr.

decadal oscillations. Excited by this source, surface-intensified potential vorticity waves of interdecadal period appear in the more stable interior. We explored a nondivergent limit of large-scale mean flow perturbations that shows that the sense of propagation of these PV waves relative to the mean flow depends on their degree of surface trapping. The required underlying potential vorticity gradient is due to the mean temperature gradient associated with the mean flow and not so much to β so that these waves are allowed on f planes. With the help of a three-layer model, analytical baroclinic instability calculations show that the growth rate of the unstable perturbations in western boundary current regions is compatible with the stability boundaries that delineate the presence of oscillations in the numerical runs with respect to the horizontal diffusivity or the restoring time constant (for runs with restoring boundary conditions instead of constant flux). Furthermore, very small temperature perturbations seeded in the numerical models amplify at rates that are comparable with those computed from the three-layer model. To decide whether such a process is at work in complex coupled GCM simulations, we suggest using the standard diagnostic tools that have been used in the past to identify the energy sources of geostrophic turbulence, namely, observations of downgradient eddy heat transport, of vertical phase lags of temperature (density) anomalies, and of periodic conversion from mean potential energy. Another way, of course, is to test the sensitivity of the observed variability to a reduction of lateral diffusivity or air–sea coupling coefficient. We have advanced the conjecture from the results of an idealized ocean–atmosphere box model that these free oscillations, generated through an instability process in the ocean (the high heat capacity fluid), may easily force oscillations in the atmosphere (the low heat capacity fluid). It appears that the present levels of turbulent eddy activities in both fluids (whose effect is to damp the oscillations) are consistent (sufficiently low) with a persistence of the oscillations. Although we hope that we have described the processes in a convincing way at the large scale, we believe that the decisive step will be to show that similar physics persists when the turbulence at the Rossby radius of deformation is explicitly resolved (Spall 1996). We know that such turbulence has a role in the ocean that goes far beyond what a simple parameterization through a diffusion law can produce. A further complexity concerns the presence of bottom topography: Winton (1997) observes that its inclusion has a damping effect on interdecadal oscillations. We suggest that the geography of the marginal stability boundaries of the process proposed herein is probably altered.

Similarly nonlinear interactions of unstable waves and zonal flows in the atmosphere cannot be excluded as a source of low-frequency variability of their own as numerical experiments (James and James 1989; James et al. 1994) or thermally driven experiments in rotating annulus geometries (Fruh and Read 1997) tend to indicate. There are indeed some studies that take such low-frequency atmospheric variability as granted to propose that the interdecadal oceanic response amounts to a simple integration of atmospheric white noise (Frankignoul et al. 1997).

Of course, there are several other mechanisms that have been put forward to rationalize the presence of interdecadal variability. Weaver et al. (1993) have emphasized that such variability can exist under mixed boundary conditions if the $E-P$ flux has sufficient amplitude. In the context of the present study, this may be understood from the viewpoint that, if the $E-P$ amplitudes are large enough, buoyancy anomalies could be generated through an instability process for salinity similar to that we have discussed for temperature. Equally important will be to judge the efficiency of the present mechanism when mechanical forcing at the air–sea interface is included. Strong additional feedbacks exist, either local as between the oceanic mixed layer and the wind stress and/or remote as between the western boundary current transport and the wind stress curl. Such processes have been considered initially by Bjerknes (1964) from analysis of observations and extended more recently by Latif and Barnett (1994) from analysis of coupled GCMs. We feel confident that the mechanism that we have identified in our idealized ocean models might play a role in more realistic contexts (see, e.g., Capotondi and Holland 1997).

Acknowledgments. We acknowledge many constructive comments from two reviewers and from O. Thual on a first version of this paper. The work was carried out within the Laboratoire de Physique des Océans, UMR 6523 supported by CNRS, Jfremer, and Université de Bretagne Occidentale.

REFERENCES

Anderson, D. L. T., and A. E. Gill, 1975: Spin-up of a stratified ocean with applications to upwelling. *Deep-Sea Res.,* **22,** 583–596.

Bjerknes, J., 1964: North Atlantic air–sea interactions. *Advances in Geophysics,* Vol. 10, Academic Press, 1–82.

Cai, W., R. J. Greatbatch, and S. Zhang, 1995: Interdecadal variability in an ocean model driven by a small, zonal redistribution of the surface buoyancy flux. *J. Phys. Oceanogr.,* **25,** 1998–2010.

Capotondi, A., and W. R. Holland, 1997: Decadal variability in an idealized ocean model and its sensitivity to surface boundary conditions. *J. Phys. Oceanogr.,* **27,** 1072–1093.

Cavallini, F., F. Grisciani, and R. Mosetti, 1988: Bounds on the eigenvalues of the planetary scale baroclinic instability problem. *Dyn. Atmos. Oceans,* **12,** 71–80.

Chen, F., and M. Ghil, 1995: Interdecadal variability of the thermohaline circulation and high-latitude surface fluxes. *J. Phys. Oceanogr.,* **25,** 2547–2568.

——, and ——, 1996: Interdecadal variability in a hybrid coupled ocean–atmosphere model. *J. Phys. Oceanogr.,* **26,** 1561–1578.

Colin de Verdière, A., 1986: On mean flow instabilities within planetary geostrophic equations. *J. Phys. Oceanogr.,* **16,** 1981–1984.

——, 1988: Buoyancy driven planetary flows. *J. Mar. Res.,* **46,** 215–265.

Delworth, T., S. Manabe, and R. J. Stouffer, 1993: Interdecadal var-

iations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Climate,* **6,** 1993–2011.

Deser, C., and M. L. Blackmon, 1993: Surface climate variations over the North Atlantic Ocean during winter: 1900–1989. *J. Climate,* **6,** 1743–1753.

Frankignoul, C., P. Muller, and E. Zorita, 1997: A simple model of the decadal response of the ocean to stochastic wind forcing. *J. Phys. Oceanogr.,* **27,** 1533–1546.

Fruh, W. G., and P. L. Read, 1997: Wave interactions and the transition to chaos of baroclinic waves in a thermally driven rotating annulus. *Philos. Trans. Roy. Soc. London,* **355,** 101–153.

Greatbatch, R. J., and S. Zhang, 1995: An interdecadal oscillation in an idealized ocean basin forced by constant heat flux. *J. Climate,* **8,** 81–91.

——, and K. A. Peterson, 1996: Interdecadal variability and oceanic thermohaline adjustment. *J. Geophys. Res.* [*Oceans*], **101,** 20 467–20 482.

Hansen, D. V., and H. F. Bezdek, 1996: On the nature of decadal anomalies in North Atlantic sea surface temperature. *J. Geophys. Res.,* **101,** 8749–8758.

Huang, R. X., and R. L. Chou, 1994: Parameter sensitivity of the saline circulation. *Climate Dyn.,* **9,** 391–409.

Huck, T., A. Colin de Verdière, and A. J. Weaver, 1999a: Decadal variability of the thermohaline circulation in ocean models. *J. Phys. Oceanogr.,* **29,** 865–892.

——, ——, and ——, 1999b: The effect of different parameterization and boundary conditions applied to the momentum equation in coarse resolution thermohaline circulation model. *J. Mar. Res.*

James, I. N., and P. M. James, 1989: Ultra-low-frequency variability in a simple atmospheric circulation model. *Nature,* **342,** 53–55.

James, P. M., K. Fraedrich, and I. N. James, 1994: Wave–zonal-flow interaction and ultra-low frequency variability in a simplified global circulation model. *Quart. J. Roy. Meteor. Soc.,* **120,** 1045–1067.

Kushnir, Y., 1994: Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Climate,* **7,** 141–157.

Latif, M., and T. P. Barnett, 1994: Causes of decadal climate variability over the North Pacific and North America. *Science,* **266,** 634–637.

Marotzke, J., 1990: Instabilities and multiple equilibria of the thermohaline circulation. Ph.D. dissertation, Institut für Meereskunde, Christian-Albrechts Universität, 126 pp.

——, and P. H. Stone, 1995: Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. *J. Phys. Oceanogr.,* **25,** 1350–1364.

Nayfeh, A. H., and B. Balachandran, 1995: *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* Wiley Series in Nonlinear Science, 685 pp.

Phillips, N. A., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two level quasigeostrophic model. *Tellus,* **6,** 273–286.

Reverdin, G., D. Cayan, and Y. Kushnir, 1997: Decadal variability of hydrography in the upper northern North Atlantic in 1948–1990. *J. Geophys. Res.,* **102,** 8505–8531.

Seager, R., Y. Kushnir, and M. A. Cane, 1995: On heat flux boundary conditions for ocean models. *J. Phys. Oceanogr.,* **25,** 3219–3230.

Spall, M. A., 1996: Dynamics of the Gulf Stream/deep western boundary current crossover. Part II: Low-frequency internal oscillations. *J. Phys. Oceanogr.,* **26,** 2169–2182.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus,* **13,** 224–230.

Weaver, A. J., and E. S. Sarachik, 1991: Evidence for decadal variability in an ocean general circulation model: An advective mechanism. *Atmos.–Ocean,* **29,** 197–231.

——, J. Marotzke, P. F. Cummins, and E. S. Sarachik, 1993: Stability and variability of the thermohaline circulation. *J. Phys. Oceanogr.,* **23,** 39–60.

Winton, M., 1996: The role of horizontal boundaries in parameter sensitivity and decadal-scale variability of coarse-resolution ocean general circulation models. *J. Phys. Oceanogr.,* **26,** 289–304.

——, 1997: The damping effect of bottom topography on internal decadal-scale oscillations of the thermohaline circulation. *J. Phys. Oceanogr.,* **27,** 203–208.