Preamble

The World Ocean, considered as an active dynamical system, is in permanent motion. Most of its manifestations can be related to wave phenomena. Besides the well known surface waves, there are also waves of other nature; among these, internal gravity waves are particularly important. They exist due to the presence of vertical fluid stratification, they are permanently generated, and they evolve and are destroyed again in the deep ocean. The amplitudes of internal waves are usually much larger than those of surface waves, due to the weak returning force, and their amplitudes can sometimes reach values of 100 m and more (see refs. [5], [23], [120], [128], [180], [192], and [193]).

Numerous *in situ* measurements, carried out in all regions of the World Ocean (see, for instance, refs. [63], [119], [123], and [157]), have shown that internal gravity waves exist wherever a stable vertical stratification of a fluid is observed. They were discovered more than 100 years ago, and were understood by the scientists of the day to be a disappointing obstacle disturbing the "correct" structure and dynamics of the oceanic water masses. More than half a century passed before the importance of internal waves to the global dynamics of the ocean was realized. Because they penetrate water thicknesses from the free surface to the bottom, internal waves play an important, and sometimes a decisive, role in many hydrophysical processes occurring in the ocean.

Recent specific interest in internal waves has emerged because they influence the horizontal and vertical exchange processes in the ocean, their generation of small-scale turbulence, and their formation of both fine structures and oceanic stratification [67], [68]. According to the idea of global meridional oceanic circulation first formulated by Sandström almost 100 years ago [209], oceanic water masses are cooled in the polar areas and are downwelled into the abyss; they then propagate further with abyssal fluxes to the equator. In the equatorial regions they upwell to the surface, are heated, and are carried poleward with the surface currents. Munk and Wunsch [167] have concluded that it is not yet clear where and how the abyssal waters return to the surface. However, quite powerful sources of water mixing must exist in the World Ocean to maintain 2000 TW (1 TW $= 10^{12}$ W) of the pole–equator heat flux associated with the global circulation.

In this mechanism of meridional overturning circulation, internal waves play a crucial role in the larger oceanographic context since they are thought to be associated closely with diapycnal mixing processes in the deep ocean, so providing a comparatively smooth observed vertical fluid stratification. In the absence of effective abyssal mixing (which is basically due to internal waves), the vertical thermohaline structure of the ocean would be represented by a thin (a few tens of meters), warm surface layer of a fluid, below which the main part of the stagnant, cold waters would be located (this follows from Sandström's theorem, which applies to an ocean that is heated and cooled at the free surface). Under such conditions, the vertical fluid stratification would be very close to that of a two-layer fluid with a very thin upper layer. The ecological consequences of such a situation would be catastrophic. Understanding where and how oceanic waters are mixed may therefore be crucial for realistic modeling of global long-term processes such as climate change.

All the above-mentioned facts point to the important role played by internal gravity waves in ocean dynamics, but they do not form a complete list. Internal waves arise within a wide range of frequencies – from short-periodic waves with periods $T \sim 10\ \mathrm{s}^{-1}$ to inertial waves with periods of days. Those waves with astronomical tidal periods – the so-called baroclinic tides – occupy a central place because they are more intensive and more pronounced in the ocean than others. Calculations of power spectra for various sites of the World Ocean have shown that at practically all locations a peak can be found in the spectral density which is in the vicinity of the tidal frequency: the level of the spectrum in this frequency band usually exceeds the values of other frequencies by one to two orders of magnitude. Gregg and Briscoe [81] estimated that one-third of all vertical displacements related to internal waves can be assigned to baroclinic tides. An analysis by Müller and Briscoe [168] has confirmed the observation of spectral peaks at the inertial and semidiurnal tidal frequencies. Being more energetic, the baroclinic tides spread out their energy through the spectrum from large to small scales due to wave–wave interaction and thus develop universal spectral distribution, as was first recognized by Garrett and Munk [69], and specified more recently by Hibiya *et al.* [96].

Since baroclinic tides are one of the principal sources of oceanic internal waves, they assume a very important role in deep ocean mixing and ocean dynamics. In fact, there are indications that internal tides may be a dominant source of deep ocean mixing and a significant component in the thermohaline maintenance. On a global scale, the energy of the Earth–Moon and the Earth–Sun orbital systems is dissipated by lunar and solar tides, slowly increasing the length of the day (about

2 ms per century), while the orbital radius of the Moon increases by about 3.7 cm every year [166].

The total amount of tidal energy being dissipated in the Earth–Moon–Sun system comprises 3.7 TW, with 2.5 TW for the principal lunar tide M_2 [29], [113]. This value has been well determined using methods of space geodesy (e.g. altimetric measurements, satellite laser ranging, lunar laser ranging).

Within the oceans, the principal sink of energy occurs in shallow marginal seas due to the friction in the bottom boundary layer (2.6 TW); the traditional explanation is due to Jeffreys [110]. There is much evidence in favor of this statement [165]. Another possible energy sink occurs due to the conversion of energy into internal tides and other baroclinic waves. Based on satellite altimetry, Egbert and Ray [55] showed that as much as one-third of the global tidal dissipation (0.6–0.8 TW) occurs in the deep ocean. Models due to Kantha and Tierney [116] and Sjöberg and Stigebrandt [219] predict similar losses. As is summarized in ref. [167], up to one-quarter of all tidal energy in the open sea is transferred into baroclinic tides.

The main problem in Munk and Wunsch's analysis of the global oceanic circulation [167] was the definition of the basic sources of background turbulence, ensuring vertical mixing and formation of the observed vertical structure of the thermohaline fields. Theoretical estimations [164] show that, for the maintenance of the existing oceanic stratification, the coefficient of vertical turbulent exchange should be at the level of 10^{-4} m^2 s^{-1}, while the majority of direct *in situ* measurements performed far from the large-scale bottom roughness reveal a magnitude that is one order smaller.

The assumption that can be made from such considerations is that the basic sink of the tidal energy to turbulence forming the necessary degree of water mixing does not occur uniformly in all areas of the World Ocean but is concentrated in "hot spots" or "storm areas," where the main transfer of energy from the barotropic tidal motion to the internal waves and to turbulence takes place.

Dissipation in the open ocean is significantly enhanced around major bathymetric features. Although some of the shelf breaks such as the Bay of Biscay [192], [193] or the Queen Charlotte Islands [47] are sites of intense internal tide generation, surface tides tend to propagate along rather than across shelves, inducing only weak across-shelf currents to interact with shelf break topography. As was concluded in refs. [11] and [104], the generation of internal tides at the continental slopes is an insignificant sink, approximately 12 GW (1 GW = 10^9 W), for M_2 over the entire globe. But scattering of internal tides by deep-ocean topography may be more important than their generation at continental slopes [219].

Thus, an accurate identification of the regions of intense tidal dissipation in the oceans and its quantification is a very important, but still quite elusive, problem. The satellite altimeter data accumulated during the ten years of TOPEX/Poseidon indicated a possible way of detecting the areas of tidal dissipation in the World Ocean. Modern satellite altimetry can detect the sea-surface level to an accuracy of 1 cm. This is due to the fact that the tidal signal is time-coherent, which allows repeated measurements, so reducing noise. Taking into account that sea-surface elevations caused by baroclinic tidal waves with amplitudes of 5 m and more are of the same order makes it possible to map the "hot spots" in the World Ocean.

In doing this, Ray and Mitchum [204] report a coherent baroclinic tide propagating to the northeast from the Hawaiian Ridge. They found that both first and second baroclinic modes of the semidiurnal internal tides were present in the data, and that the signal of internal tide propagation could be tracked up to 1000 km from the Hawaiian Ridge. In a further study [205], they estimate that 15 GW of the semidiurnal internal tidal energy radiates away from the Hawaiian Ridge in the first baroclinic mode.

A further step was taken by Kantha and Tierney [116], who quantified the energy and energy dissipation rate in global baroclinic tides using a combination of precision altimetry and a numerical tidal model. They discovered more than 15 of the highest ridges of the World Ocean and found that approximately 15% of the barotropic tidal energy is transformed into the baroclinic component. A similar investigation was performed by Niwa and Hibiya [177] for the dissipation of the M_2 tide in the Pacific Ocean. It was found that the conversion rate from the semidiurnal tide to internal waves integrated over the whole Pacific Ocean amounts to 338 GW, 84% of which is located at the prominent topographic features. Thus, the global surveys of Kantha and Tiereny [116] and Niwa and Hibiya [177] show that the baroclinic tidal signals occur through the World Ocean and originate principally from large submerged ridge topographies. Recent *in situ* measurements have confirmed the existence of many of these "hot spots" (see refs. [3], [38], [48], [54], [64], [94], [126], and [127]). Through such "windows" the barotropic tidal energy is supplied to the ocean (it is transformed from the internal waves to turbulence), and is further distributed to all areas of the World Ocean.

Thus, a significantly difficult problem to be solved by oceanographers is to define "where," "how much," and "how" the tides dissipate their energy. Many works (e.g. the references listed above) address the questions of "where" and "how much." We try to answer the question "how" barotropic tidal energy transforms into the baroclinic component. For this purpose we shall highlight the most interesting aspects of the dynamics of baroclinic tides in a horizontally nonuniform ocean: generation and propagation of tidal internal waves over variable bottom topography and horizontal gradients of density (which are usually accompanied by bottom elevations).

A number of books published since the 1970s have considered different aspects of internal gravity waves. Some fragmentary data on internal waves and qualitative pictures of some particular problems are summarized in the corresponding chapters of the books by Monin *et al.* [159], Phillips [190], and Turner [236]. Despite the appearance of the fundamental work by Whitham [264], internal waves – due to their anisotropy and nonuniformity – still present a nontrivial object for investigations and for applying the methods outlined in these books. Specialists in wave dynamics must rely on what has been written in the more general textbooks by Le Blond and Mysak [132] and Tolstoy [232]. Other treatises are the books by Lighthill [139], which focuses on the general theory of the generation and propagation of internal waves, Gill [76], which discusses internal waves in atmospheric and oceanic processes, and Craik [46], which discusses wave interaction phenomena in fluids at rest and shear flows.

Only a few fundamental publications are known which deal entirely with internal waves: *Interne Wellen* by Krauss [123]; *Surface and Internal Waves* [37] and *Hydrodynamics of Surface and Internal Waves* [36] by Cherkesov; *Waves Inside the Ocean* by Konyaev and Sabinin [119]; Munk's *Internal Waves* [163] (an extensive literature review); and a more recent issue of Miropol'sky's *Dynamics of Internal Gravity Waves in the Ocean* [157], which is the English translation of the original version published in Russian in 1981. Another book, written by Baines, *Topographic Effects in Stratified Flows* [13], considers internal waves and hydraulic jumps that are generated by stationary flows interacting with bottom topography.

In contrast with the books given above, where the general wave theory is applied to study a wide range of baroclinic wave motions, this text focuses on the investigation of tidally generated internal waves. We will pay attention to linear as well as weakly and strongly nonlinear problems. The book presents results that have been obtained by the authors since the early 1980s.

1

General background

1.1 Introduction

Internal waves are present in the ocean because of the existence of water stratification. Properties of internal waves are different from the characteristics of waves which are observed at the ocean surface. Indeed, whereas ocean surface waves possess amplitudes of at most a few meters, internal wave amplitudes can be as large as 100 m or more. The following simple explanation can clarify the difference between surface and internal waves.

We consider wave motions in the Cartesian system of coordinates $Oxyz$, with Oxy within the undisturbed free surface of the fluid and the Oz-axis directed vertically upward. We assume that the water density consists of two parts: a stationary density $\rho_0(x, y, z)$ and density disturbances $\tilde{\rho}(x, y, z, t)$ introduced by the wave motions. In the real ocean, usually $\rho_{0z} \gg \{\rho_{0x}, \rho_{0y}\}$, which is why with good accuracy one can take $\rho_0 = \rho_0(z)$.[1] Some examples of such density distributions are presented in Figure 1.1(a). Imagine a unit volume of water is displaced vertically upward from its equilibrium position z_0 to $z_0 + \xi$ (Figure 1.1(b)). Newton's second law for this liquid particle reads

$$\rho_0(z_0)\,\frac{d^2\xi}{dt^2} = g\rho_0(z_0 + \xi) - g\rho_0(z_0), \tag{1.1}$$

where g is the acceleration due to gravity. The right-hand side of this equation represents the restoring force, which, in fact, is the difference between the Archimedean buoyancy force and the gravitational force.

Because the difference of the oceanic water density in (1.1) between $z = z_0$ and $z = z_0 + \xi$ is small (≤ 10 kg m^{-3}) in comparison with the density difference between water ($\rho_0 \sim 10^3$ kg m^{-3}) and air ($\rho_a = 1.210$ kg m^{-3}), the restoring force for internal waves is more than 100 times smaller than for surface waves. This

[1] Subscripts denote partial derivatives with respect to the subscripted variable.

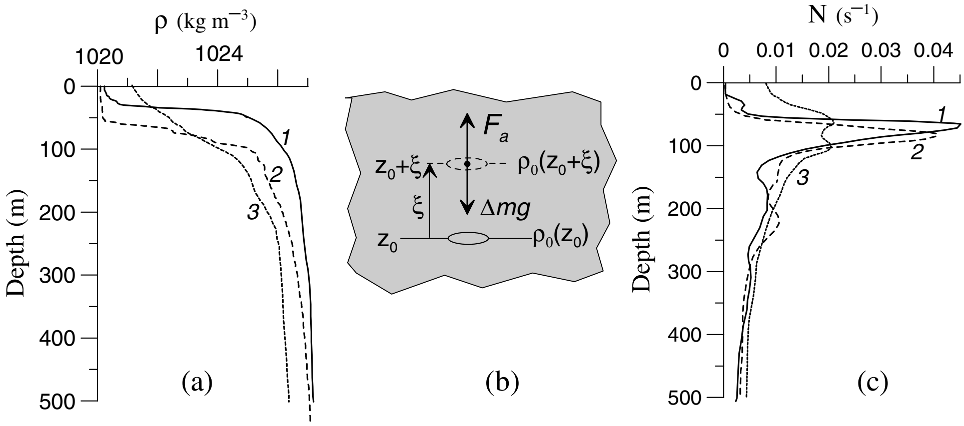

Figure 1.1. (a) Typical distributions of density and (c) buoyancy frequency measured in the Andaman Sea (1,2) and the Sulu Sea (3). (b) Schematic diagram of forces acting on a fluid particle as its position deviates from its state of equilibrium level, z_0. F_a represents the buoyancy force, the resultant of the pressure exerted by the ambient fluid on the particle's surface; Δmg is the weight of the fluid fragment.

is why the amplitudes of internal waves can be substantially larger than those of surface waves. Using Taylor series expansion one can express this as

$$\rho_0(z_0 + \xi) = \rho_0(z_0) + \xi \frac{d\rho_0}{dz} + O(\xi^2). \tag{1.2}$$

Thus, (1.1) can be rewritten as

$$\frac{d^2\xi}{dt^2} - \frac{g}{\rho_0(z)} \frac{d\rho_0}{dz}\xi = 0. \tag{1.3}$$

Depending upon the sign of $d\rho_0/dz$, solutions of this equation are bounded ($d\rho_0/dz < 0$) or unbounded ($d\rho_0/dz > 0$) as time approaches infinity. In fact, for stable stratification ($d\rho_0/dz < 0$), this equation describes vertical oscillations of fluid particles relative to their equilibrium positions with angular frequency $N(z) = [(-g/\rho_0(z_0))d\rho_0/dz]^{1/2}$, where N is known as the *buoyancy frequency*. This concept was introduced by Rayleigh [201] in his investigations of free convection processes. Alternatively, it is called the Brunt–Väisälä frequency because it was equally introduced by the Norwegian meteorologist Brunt [27] and the Finnish oceanographer Väisälä [240].

If the compressibility of the water is taken into account, the buoyancy frequency takes the form given by

$$N^2(z) = -\frac{g}{\rho_0(z_0)} \frac{d\rho_0}{dz} - \frac{g^2}{c_\mathrm{s}^2}, \tag{1.4}$$

Table 1.1. *Important tides and their characteristics.*

Darwinian symbol	Period (hours)
K_1	23.9345
O_1	25.8193
M_2	12.4206
S_2	12.0000

where c_s is the speed of sound in water. Vertical profiles of $N(z)$ for the Andaman and Sulu Seas for the density profiles of Figure 1.1(a) are shown in Figure 1.1(c). They are typical for most oceanic situations. Of course, the above simplified example does not reflect the wide variety of internal gravity waves in the real ocean. It shows only the physical mechanism which is responsible for their existence. In fact, real oceanic internal waves consist of a continuous spectrum of oscillations, from the buoyancy frequency up to the local inertial frequency.[2]

It is interesting that scrutiny of a vast number of field observations at various sites of the World Ocean points to the existence of internal waves with tidal periods.[3] The energy of such waves very often exceeds the wave energy within other frequency bands [15], [167], [168].

An analogous situation occurs for surface waves: the barotropic[4] tide is one of the most remarkable and most pronounced wave phenomena. This similarity in manifestation does not mean, however, that barotropic and baroclinic tides are excited by identical mechanisms. Barotropic oceanic tides are mostly the result of the perpetual variation of the gravitational attraction between the Earth and the Moon. The latter is the astronomical body closest to the Earth, and the Sun is the most massive heavenly body in our planetary system. The periods of the most pronounced tidal harmonics which are mentioned in the present book are shown in Table 1.1 (a detailed tidal theory explaining these is presented in ref. [115]).

The direct generation of internal waves by tidal forcing in an ocean of constant depth is insignificant [123]. Therefore, the question of how internal waves of tidal periods are generated is very important for an adequate understanding of ocean dynamics. In oceans of variable depth, the generation of baroclinic tides is possible because of the interaction of the barotropic tidal waves with the bottom topography.

[2] This statement ignores all acoustic waves, which arise due to the compressibility of the water and propagate much faster than barotropic or baroclinic waves. We will not discuss acoustic waves in this book.

[3] See refs. [15], [35], [51], [77], [81], [89], [97]–[100], [119], [133], [167], [192], [193], [194], [212], [217], and [234].

[4] Surface and internal waves are oscillating phenomena belonging to two classes of water motion, defined as barotropic and baroclinic processes, respectively.

The possibility of this mechanism of baroclinic wave generation was first indicated almost 100 years ago for an ocean with a vertical density jump (two-layer model) by Zeilon [271]. The efficiency of this mechanism (for a continuously stratified fluid) was further theoretically corroborated by Cox and Sandström [43]. Numerous laboratory and field experiments subsequently confirmed this conclusion. Laboratory investigations, performed in wave tanks ([12], [16], [28], [149], [269]), as well as theoretical studies ([7]–[11], [147], [202], [210]) led to the isolation of the most important factors influencing the topographic generation of the wave fields. It was found that two basic parameters that control the dynamics of internal wave fields near bottom features are the inclination angle α of the characteristic lines of the hyperbolic wave equation

$$w_{zz} - \alpha^{-2} w_{xx} = 0 \tag{1.5}$$

(which will be derived in Section 1.4) and the inclination of the bottom relief $\gamma = dH/dx$. Here, w is the vertical velocity component, and the bottom topography is given by $z = -H(x)$. The relation between α and γ is very important from the point of view of the generation of internal waves. Baines [12] found by performing laboratory experiments on the generation of internal waves by barotropic fluxes over an underwater canyon that, for *subcritical* inclinations of the bottom, when $\alpha \gg \gamma$, the influence of the stratification on the movement of a fluid was very weak, and in the vicinity of a canyon the motion was basically barotropic. If the inclination of the bottom was close to *critical*, $\alpha \sim \gamma$, the movement was essentially baroclinic through the entire water column. With $\alpha < \gamma$ (*supercritical* case) the baroclinic energy propagated downwards along the canyon as radiated waves, thereby creating a wave "beam."

Similar features were observed during *in situ* measurements in the ocean, for example near the coast of Africa [77], [81], over the continental slope off Oregon [234], in the coastal zones of Australia [44], [45], [97]–[100], on a shelf off Vancouver Island [15], near the coast of California [228], in the Bays of Massachusetts [35], [87] and Biscay [192], [193], and in many other regions of the World Ocean. In these observations, subcritical as well as supercritical situations of internal wave generation were encountered. Using the theory of empirical orthogonal functions, analysis of the observational data allowed the conclusion [77] that over the subcritical continental slopes a significant part of the energy of the generated internal tides was concentrated in the first baroclinic mode,[5] which is propagating shoreward from the shelf break. The density of the wave energy in that direction increased, and in the shallow part of the shelf it reached saturation.

[5] For a given distribution of the buoyancy frequency, a countably infinite number of internal wave modes exists, each of which has its own speed of propagation. The mode with the fastest wave speed is called the *first* mode. This fact will shortly be proven.

It was also found that many internal waves completely dissipated prior to reaching a coastal line where they would reflect. A similar conclusion on the predominance of the lowest baroclinic mode in the fields of the K_1 and O_1 internal tides was also obtained for the near-critical bottom features off the eastern part of Florida [133].

In those areas, where the inclination of the bottom is supercritical, the characteristic feature of internal tides is the formation of a wave beam, i.e. the confinement of the wave field to a long and rather narrow region, where the basic part of the baroclinic tidal energy is concentrated. Usually, this wave beam arises at the shelf break and extends into the deep part of the ocean downwards along the continental slope (to be more precise, along the characteristic lines of the wave equation (1.5)). This structure of the internal tide was found in the Pacific Ocean near the coast of Oregon [234], in the Atlantic Ocean in the Bay of Biscay [192], [193], in the eastern North Pacific near the Mendocino Escarpment [3], and also in the Indian Ocean in the area of the JASIN experiment [51].

Nonlinear wave effects were also of specific interest during observational investigations of the baroclinic tides. Probably the most conspicuous manifestation of the nonlinear behavior that internal tides take is the formation of an internal hydraulic jump (or a baroclinic bore) and its subsequent evolution into a packet of short nonlinear waves. This phenomenon usually takes place under conditions of sufficient intensity of the tidal forcing, as will be discussed below. Such baroclinic bores were often observed at different sites of the World Ocean, for instance in the Bays of Massachusetts [35], [87] and New York [261], close to the Mascarene Ridge in the Indian Ocean [120], [206], in the Strait of Gibraltar [6], [263], on a shelf off Australia [98], and in many other regions. A vast amount of observational material was collected on the structure and spatial–temporal characteristics of solitary internal waves (SIWs) and wave packets generated during the nonlinear stage of the evolution of the baroclinic tides (see, for instance, the review [100]). They were sufficiently reliably registered by contact[6] as well as remote sensing[7] methods.

The many forms in which tidal waves appear increase the urgency of providing a theoretical description of their dynamics. Various linear and nonlinear theoretical models (both analytical and numerical) have been constructed since the 1960s in order to study the generation mechanisms of baroclinic tides. We will give an overview of many published theoretical models in the introductions to all chapters where an appropriate topic of baroclinic tidal dynamics is considered. The list of studied phenomena includes all kinds of wave motions that are generated by the oscillating tidal flow over topographic features, such as underwater ridges and

[6] See refs. [1], [78], [88], [102], [107], [119], [120], [128], [140]–[142], [176], [180], [191], [195], [206]–[208], [212].
[7] See refs. [2], [23], [66], [70], [71].