Nonlinear internal waves are found in many parts of the world ocean. Their widespread distribution is a result of their origin in the barotropic tide and in the variety of ways they can be generated, including by lee waves, tidal beams, resonance, plumes, and the transformation of the internal tide. The differing generation mechanisms and diversity of generation locations and conditions all combine to produce waves that range in scale from a few tens of meters to kilometers, but with all properly described by solitary wave theory. The ability of oceanic nonlinear internal waves to persist for days after generation and the key role internal waves play in connecting large-scale tides to smaller-scale turbulence make them important for understanding the ocean environment.

The photo shows the surface manifestations of a train of nonlinear internal waves (visible as alternating bands of smooth and rough water) off the northern tip of Cape Cod as viewed from a whale watcher boat in Massachusetts Bay (near position labeled B in Figure 6). Photograph taken 30 August 2006 (15:30 local time) by J.C.B. da Silva
INTRODUCTION

Nonlinear internal waves are just one type of wave present in the ocean’s interior. Other wave types include small-amplitude linear internal waves, baroclinic planetary Rossby waves, and the linear baroclinic tide. Each of these waves has a distinctive generation process, propagation manner, and propagation timescale. Nonlinear internal waves propagate along a pycnocline in the ocean, the portion of the water column with a sharp change in density, typically the result of differences in temperature or salinity. Generation can be achieved in a variety of ways, some of which involve the direct displacement of the pycnocline and others that involve a more complex conversion of tidal energy into pycnocline motion. These different mechanisms will be described in detail in the following sections.

Nonlinear internal waves are best described in terms of a special waveform called a soliton, or solitary wave (see Zabusky and Porter, 2010). The soliton has, as one of its defining features, the ability to retain its form over an extended time period as it propagates away from its generation site. This persistent coherence of form is explained by the balance between nonlinear steepening effects and nonhydrostatic dispersion, represented by different terms in the governing equations. The Korteweg-de Vries (KdV) equation, or a variant thereof, is typically used to represent these key physical processes (see Ostrovsky and Stepanyants, 1989; Helfrich and Melville, 2006). The nonlinearity term of the KdV equation contains the product of the wave amplitude and its spatial derivative.

Solitary waves are fascinating phenomena of considerable interest to a variety of fields, including fluid dynamics, plasma physics, condensed matter physics, optics, and biology (Scott, 2007). As nonlinear internal waves in the ocean, they represent a potential hazard to offshore drilling operations (Hyder et al., 2005) and have important consequences for underwater sound propagation (e.g., Zhou and Zhang, 1991). They represent an important energy transfer mechanism between the large-scale tides and vertical mixing, and they often play a key role in biological primary production (e.g., Sandstrom and Elliott, 1984) by affecting the exchange of heat, nutrients, and other properties between the shelf and the open ocean (e.g., Huthnance, 1995). In the atmosphere, solitary waves produce the Morning Glory cloud formations over northwestern Australia (e.g., Smith et al., 1982, Reeder et al., 1995) and similar formations over the Mozambique Channel and the Red Sea (da Silva and Magalhães, 2009; Magalhães et al., 2011).

The hydrodynamic soliton was first described by J. Scott Russell who, in 1834, reported the formation of a singular, unchanging hump on the surface of a Scottish canal that was generated when a towed barge was brought to a sharp halt. Russell followed the wave for several miles on horseback, remarking on how it retained a constant shape for a surprisingly long distance. Subsequent laboratory experiments by Russell produced the first quantitative description of the solitary wave (Russell, 1844). In 1895, Diederik Korteweg and Gustav de Vries produced a generic equation accounting (to lowest order) for the wave properties of dispersion and nonlinearity and its soliton solution. It was the work of Osborne and Burch (1980) in their study of the internal waves in the Andaman Sea during the 1970s that firmly tied the properties of solitary waves to the phenomena of oceanic nonlinear internal waves.

Although they propagate along a pycnocline in the interior of the ocean, nonlinear internal waves alter the sea surface roughness, making them readily detectable in satellite imagery. Currents within the nonlinear internal wave produce convergent and divergent zones on the surface that move in phase with the internal wave’s subsurface crests and troughs. These zones cause variations in sea surface roughness that appear as distinctive features in both synthetic aperture radar (SAR) and optical sunglint imagery. With an individual nonlinear internal wave remaining coherent for up to several days, waves generated on several successive tidal cycles are often visible in the same image. In many cases, the waves generated in each cycle tend to be grouped together into a “rank ordered” packet, with the largest wave in the lead, in accordance with KdV type theory. The distance between

Christopher R. Jackson (goa@internalwaveatlas.com) is Chief Scientist, Global Ocean Associates, Alexandria, VA, USA. José C.B. da Silva is Senior Scientist, Department of Geosciences, Environment and Spatial Planning and Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), University of Porto, Portugal, and Guest Investigator, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. Gus Jeans is Director, Oceanalysis, Wallingford, UK.
leading waves in an individual packet can vary from a few tens of meters to tens of kilometers, while the separation between packets (generated on different tidal cycles) can range from a few kilometers to more than 100 km, with both depending on the phase speed of the waves and the time since generation. In situ observations show that the phase speeds of oceanic nonlinear internal waves can vary from less than 0.3 m s\(^{-1}\) to more than 3 m s\(^{-1}\), with amplitudes (as measured as a displacement of the pycnocline) ranging from just a few meters to in excess of 100 m.

Beginning with early photographs taken by astronauts and continuing through to today’s dedicated Earth observing satellites, more than 40 years of orbital imagery shows that nonlinear internal waves commonly occur in many locations around the world (Jackson, 2004). Figure 1 presents a map of nonlinear internal wave occurrences detected in the sunglint imagery of the Moderate-Resolution Imaging Spectroradiometer (MODIS) sensor onboard the National Aeronautics and Space Administration’s (NASA’s) Earth Observing System satellites Terra and Aqua (Jackson, 2007). This survey, using 250 m resolution true-color imagery and covering 21 months between August 2002 and May 2004, identified more than 3,500 nonlinear internal wave occurrences. Limitations in the survey’s sampling means that this count represents a significant underestimate of the number of internal wave occurrences that took place during that period.

The widespread distribution of nonlinear internal waves is the result of their origin in barotropic tidal forcing and in the different number of ways in which they can be generated. The following sections will examine the various generation mechanisms, which include lee waves, tidal beams, resonance, plumes, and the conversion of the internal tide. The focus is on first mode (mode-1) waves, which are the most common and displace the pycnocline downward in deep water, but also includes a discussion of second mode (mode-2) waves.

**LEE WAVE GENERATION**

Nonlinear internal wave generation at an underwater sill or a bank has been traditionally explained by the lee wave mechanism put forth by Maxworthy (1979). The mechanism is based on the theory of supercritical flows, which uses a parameter called the internal Froude number to characterize the hydraulic state of a stratified shear flow. The Froude number \( F = \frac{u_f}{c} \) is a dimensionless quantity that expresses the ratio of the velocity of a fluid \( (u_f) \) to the velocity \( (c) \) (relative to the fluid velocity) of a small-amplitude internal wave on the interface. The velocity \( (c) \) is the linear phase speed of the first vertical internal mode. When the Froude number is less than 1 (subcritical), information can move upstream (opposite the direction of flow) because the wave velocity is greater than the flow velocity. When the Froude number is greater than 1 (supercritical), information can only move downstream (in the direction of the flow) because the fluid velocity is greater than the wave velocity. When the Froude number is critical

---

Figure 1. The location of nonlinear internal waves observed in 250 m resolution MODIS (Moderate-Resolution Imaging Spectroradiometer) satellite sunglint imagery acquired from August 2002 through May 2004. More than 3,500 nonlinear internal wave occurrences were identified by the survey. Adapted from Jackson (2007)
(F = 1), a disturbance on the interface will remain stationary and accumulate energy through resonance. This condition can occur when a moving fluid encounters a change in depth, with the Froude number transitioning from subcritical (F < 1) to supercritical (F > 1), or vice versa.

In the context of nonlinear internal wave generation at a sill or bank, the sloping seabed imparts a vertical component to the stratified flow, displacing the pycnocline. The seabed feature also provides the variation in depth necessary to set up the transition zone between subcritical and supercritical flow. The result is a stationary lee wave of depression in the pycnocline downstream of the sill, which attains a large phase speed relative to the moving water in order to remain stationary relative to the bank. As the flow slackens and turns subcritical, the lee wave can advance over the sill and propagate away. This initial disturbance can then evolve into a nonlinear internal wave train according to KdV theory.

Figure 2 shows this process schematically. In the figure, the arrows represent the direction of the overall flow. The patterns above the waveform represent the surface roughness associated with the convergence zone of the nonlinear internal wave. Hibiya (1986) modeled this generation as a transient process in response to a time-varying tidal flow. In his model, infinitesimal internal waves are continuously generated during the tidal cycle and constructively superimposed during the supercritical phase of the ebb tide.

Sills and banks are among the principal generation sites of large nonlinear internal waves in the ocean. The Camarinal and Spartel sills are responsible for the generation of the nonlinear internal waves found at the Strait of Gibraltar (Armi and Farmer, 1988; Farmer and Armi, 1988; La Violette and Arnone, 1988). Pearl Bank in the Sulu Archipelago is one of the principal places of origin for the nonlinear internal wave packets that propagate into the Sulu Sea (Apel et al., 1985). The Indonesian seas are home to a wide range of nonlinear internal wave activity due to the large number of shallow interisland sills separating deeper-water basins in the presence of strong currents. Along the western side of the Andaman Sea among the Nicobar Islands, lee waves are generated on both the east and west sides of sills connecting several of the islands. Figure 3, a true-color MODIS sunglint acquired on April 6, 2004, at 4:25 UTC (approximately the time when the east/west component of the barotropic tidal current was close to zero) shows this bidirectional generation situation. Wave packets 1 and 2 were released from the eastern side of the sill as the tide slackened. These packets will combine and evolve to resemble Packet 3, which was generated roughly 12 hours earlier.

Figure 2. A schematic showing nonlinear internal wave generation via the lee wave mechanism of Maxworthy (1979). (a) As the stratified flow moves over the sill, the initial lee wave develops on the downstream side. (b) As the flow slackens, the lee wave steepens nonlinearly and propagates upstream over the sill. (c) The lee wave evolves into a nonlinear internal wave packet in a manner consistent with the predictions of Korteweg-de Vries theory. Adapted from Haury et al. (1979)
Packet 4 was released from the western side of the sill during the prior slack tide, approximately six hours before the time of the image.

In a similar manner, a continental shelf break can cause a downward displacement of the pycnocline as the current flows off-shelf. Some nonlinear internal wave packets observed on the continental shelf appear to possess properties that are consistent with generation by the lee wave mechanism (Pingree and Mardell, 1985). Huthnance (1981) suggested that the lee wave mechanism may also be responsible for the generation of off-shelf propagating nonlinear internal waves, where a lee wave formed off the shelf during on-shelf tidal flow is released as the tide relaxes. However, Pingree et al. (1986) traced both on- and off-shelf propagating waves back to a generation time at the shelf break during the ebb tide. This simultaneous symmetric generation is inconsistent with the lee wave mechanism, but it is consistent with the mechanism of internal tide generation (Pingree et al., 1983). So while the lee wave mechanism can be useful for gaining insight into the origin of on-shelf nonlinear internal waves, it cannot fully explain nonlinear internal wave generation at the shelf break, so other generation mechanisms must be considered.

**INTERNAL TIDE EVOLUTION**

The lee wave is a specific manifestation of the general phenomenon of tide-topography interaction (see Baines, 1982) that is also responsible for the generation of the internal (or baroclinic) tide. In its simplest two-layer form, the baroclinic tide comprises a linear internal wave at tidal period on the density interface. Baroclinic tidal generation occurs in a manner similar to the lee wave mechanism, where vertical displacements of the pycnocline are caused by seabed features that induce vertical components into a tidal flow. The internal tide propagates away from the generation zone and may initially have a sinusoidal shape, but, if sufficiently energetic, it will steepen, break, and evolve into nonlinear internal waves.

A number of authors explain the origin of nonlinear internal waves in terms of the evolution of the internal

![Figure 3. True-color MODIS image acquired on April 6, 2004, at 4:25 UTC showing the bidirectional generation of nonlinear internal waves from the sills among the Nicobar Island chain. Adapted from Jackson (2007) ](image)
tide, including Holloway (1987), Smyth and Holloway (1988), Sandstrom et al. (1989), Brickman and Loder (1993), and Farmer et al. (2011). The relationship between nonlinear internal waves and the internal tide appears more subtle than merely a common tide-topography interaction. In some locations, nonlinear internal waves can maintain a fixed phase relationship with the baroclinic tide as both propagate together, with the nonlinear internal waves apparently “trapped” in the troughs of the longer tidal waves. This relationship has been observed for nonlinear internal waves generated at both shelf breaks (Pingree and Mardell, 1985; Pingree et al., 1986; New and Pingree 1992) and sills (Apel et al., 1985). Locations also include places where the nonlinear internal waves and the internal tide remain in phase after three to four tidal cycles (da Silva et al., 2011).

It is necessary to delve into a little more theory to appreciate the subtleties of the relationship between a nonlinear internal wave and the internal tide. The effect of Earth’s rotation enhances the phase speed of a two-layer internal tide (Defant, 1960), but does not affect internal solitons, which are typically much shorter in length than the internal Rossby radius. However, internal solitons also have an enhanced phase speed, this time due to the effect of nonlinearity (e.g., Osborne and Burch, 1980). It would be remarkable if the very distinct mechanisms of Earth’s rotation and nonlinearity produced the same enhancement to the wave phase speed. It is therefore likely that the match between these enhancements is only approximate, in which case nonlinear waves would no longer appear trapped after a long propagation distance. However, Henyey and Hoering (1997) describe a physical mechanism that could lead to a tendency for trapping in a scenario where slowly propagating nonlinear internal waves gain energy from a larger-scale internal bore, while fast-moving nonlinear internal waves lose energy. Energetic equilibrium occurs close to the linear internal wave phase speed, but it should be noted that this theory was formulated without considering the effects of rotation.

Rotation, as well as nonlinearity, should be accounted for in models of nonlinear internal wave generation and internal tide disintegration. They were brought together elegantly in the unified nonlinear internal tide model of Gerkema (1996), which includes generation of an initial disturbance and subsequent disintegration into nonlinear internal waves via KdV type physics. The effect of rotation is shown to have an important dispersive influence that could prevent internal tide disintegration if it is strong enough to overcome nonlinearity. The effects of rotation on the internal tide increase at higher latitudes as the Coriolis dispersion increases.

The Gerkema (1996) unified nonlinear internal tide generation model successfully reproduced the nonlinear internal wave observations made by Pingree and Mardell (1985) in the Celtic Sea. Coriolis dispersion prevented disintegration of the internal tide at neap tides but not at spring tides, when it was overcome by increased nonlinearity. The Gerkema (1996) model also reproduced the nonlinear internal waves observed by Halpern (1971) in Massachusetts Bay, where Coriolis dispersion was not strong enough to prevent disintegration of the internal tide, but did reduce the number of nonlinear internal waves produced. A similar result demonstrating the impact of rotation in slowing the development and evolution of nonlinear internal waves was also found by Farmer et al. (2009), who successfully used the Gerkema (1996) generation model to reproduce the distinctive features found in mooring observations of large nonlinear internal waves in the northern South China Sea.

Following the convention of the day, the Gerkema (1996) model represented tidal forcing in a two-dimensional slice perpendicular to the continental shelf break. It was common practice to include only the cross-slope component of tidal flow, because the along-slope component was thought to not contribute towards inducing vertical motions. However, on many continental margins around the world, the strongest tidal currents tend to be aligned along the bathymetry contours, with relatively weak cross-slope components. Strong internal tide generation occurs in regions where strong tidal flow perpendicular to the slope of the seabed. Internal waves can be generated where depth contours curve to become perpendicular to the tidal flow, creating a generation “hotspot.”

Sherwin et al. (2002) identified a classic example of this three-dimensional generation effect off the coast of Portugal, illustrated in Figure 4. The image shows two packets of nonlinear internal waves on the shelf (packets A and B), with one still over the slope (packet B). A third packet (packet C) out in the deep ocean appears to be propagating onshore, based on the orientation of the SAR wave signature. The generation site is shown to the south (marked X in Figure 4) where the continental margin bulges out into the deep ocean. Internal waves propagated...
offshore into the deep ocean toward the north before being refracted onshore to impact the shelf near 41°N.

**TIDAL BEAM GENERATION**

New and Pingree (1992) discuss an interesting connection between nonlinear internal waves and the internal tide, proposing a novel generation mechanism for nonlinear internal waves observed in the central Bay of Biscay. This mechanism incorporates the effects of continuous stratification, which produce a waveguide that allows internal tidal energy to propagate through the whole vertical extent of the ocean.

Over a flat bottom, linear wave theory predicts the combination of various vertical wave modes at tidal frequency that, superposed on each other, create beams along which the energy propagates and the particle velocities intensify (e.g., Gerkema and Zimmerman, 2008). Energy propagates upward and

![ERS-1 SAR (European Remote Sensing-1 synthetic aperture radar) image acquired August 8, 1994, at 11:20 UTC (orbit 16020, frame 2781). Light regions indicate high backscatter. The “X” denotes the generation site and the orange curves show the refraction of internal waves that propagate onshore to impact the shelf break. In situ measurements of nonlinear internal waves were made at the stations marked 1 to 3. The distance between the leading waves in the packets is approximately 500 m.](image-url)
downward along these internal tidal beams, reflecting from the ocean floor and ocean surface, in a manner similar to a laser beam propagating diagonally between two parallel mirrors (or within an optical fiber). The tidal beams follow a curve whose slope changes with depth due to varying stratification, similar to light propagating through a medium with a gradually increasing (or decreasing) index of refraction. The curvature is greatest where the stratification is strongest (usually near the ocean surface), and least where the stratification is weak (usually near the ocean floor). In homogeneous stratification, the beams would be straight lines reflecting between the ocean floor and the ocean surface.

In this generation method, the continental shelf edges excite the tidal beams where the slope of the bottom topography matches the angle of the beam to the horizontal. These locations are called “critical” slopes, where a downward-propagating ray can form, propagate to the seafloor, reflect upward, and impinge upon the pycnocline a long distance away from the continental shelf. In the early 1990s, Adrian New and Robin Pingree noted that the location of large solitons in the central Bay of Biscay coincided with the intersection of the thermocline with internal tidal beams generated 150 km away at the edge of the continental shelf. They suggested that the thermocline in the central bay is excited by tidal energy originating from the critical region of the continental slope and reflecting from the seabed along these beams. The model of New (1988) correctly predicted the location where a reflected beam would surface and produce large thermocline oscillations; observational evidence for the reflected beam was provided by Pingree and New (1991) and New and Pingree (1992). Later, New and da Silva (2002) reported the coherent two-dimensional structure of these nonlinear internal waves using SAR.

New and Pingree coined the term “local generation” to describe this new internal wave generation mechanism. Subsequent observations and modeling have confirmed soliton generation in this manner, provided certain requirements of the near-surface stratification are met. According to the studies of Gerkema (2001) and Akylas et al. (2007), in a given ocean basin (represented by an average depth $H$, an upper mixed layer of thickness $d$ and density $\rho$, an interfacial pycnocline of density step $\Delta \rho$, beneath which lies a layer of constant buoyancy frequency, $N_c$), the stratification should be of intermediate strength (quantified by the parameter $g' = g\Delta \rho / \rho$) for the internal tidal beam to produce large displacements of the pycnocline that consequently evolve into nonlinear internal waves. If the pycnocline strength is either very small or very large, there is no significant transfer of energy from beams to interfacial waves.

In an attempt to understand and quantify the way the pycnocline responds to an impinging internal wave beam, Gerkema (2001) used linear wave theory and identified a parameter (gamma) to “measure” this interaction. Gerkema defined gamma by

$$\gamma = \sqrt{\frac{g' d}{N_c H}},$$

where $H$ is the total water depth. The parameter $\gamma$ is a ratio of two phase speeds: the numerator is the phase speed of long interfacial waves in a two-layer system whose upper layer thickness is $d$; the denominator is a measure of the phase speed of vertical modes in a constantly stratified layer of depth $H$ (the mode-1 phase speed is $N_c H / \pi$). The parameter $\gamma$ therefore reflects the ratio of the phase speeds of waves on the pycnocline and the uniformly stratified lower layer. Gerkema (2001) put forward a theoretical criterion of $\gamma = 0.1$ for optimal excitations of solitons in the pycnocline. The value of $\gamma = 0.2$ appears to be the upper limit for local generation based on lab experiments and numerical modeling (see, e.g., Grisouard et al., 2011, and Mercier et al., in press).

Nonlinear internal wave generation via tidal beams occurs outside the tropics (latitudes > 20°) where the stratification is relatively moderate. In areas where the beam generation mechanism has been identified, values of gamma have been found to fall within this optimum range ($0.1 \leq \gamma \leq 0.2$), including in the Bay of Biscay ($\gamma = 0.14$; Pingree and New, 1991; New and Pingree, 1992), off the Iberian continental shelf in summer ($\gamma = 0.15$; Azevedo et al., 2006; da Silva et al., 2007), and in the Mozambique Channel in winter ($\gamma = 0.17$; da Silva et al., 2009). Local generation of nonlinear internal waves is very likely to occur in other regions, but identification is more difficult if other generation mechanisms are also at work. Da Silva et al. (2007) describe the role of SAR images in identifying beam-generated solitons.

Figure 5 shows an example of local generation of nonlinear internal waves off the west Iberian shelf. The critical slopes are situated below 1,000 m depth, some 50 km to the west of the shelf break. Tidal beams emanate from the critical slopes, one downward and offshore, and another upward and inshore. The offshore-propagating beam reflects from
the seafloor and impinges the thermocline from below some 60 km away from its origin (Figure 5b) where it is observed in the SAR image propagating offshore (Figure 5c). The thermocline at the point of impingement is sufficiently mild to allow the generation of a nonlinear internal wave train of considerable dimensions. Crest lengths extend for approximately 30 km and wavelengths (crest-to-crest length) are up to 1 km long (Figure 5c and d). For additional details, see da Silva et al. (2007).

**RESONANT GENERATION**

The section above on nonlinear internal wave generation via the lee wave mechanism discusses how stratified flow was perturbed in the vertical by the presence of a sill or bank and how the Froude number controls the release of the nonlinear internal wave. A stratified flow can also be perturbed by a lateral (horizontal) topographic contraction or small-scale undulations in bottom topography, with resulting critical flows and creation of nonlinear internal waves.

---

**Figure 5.** (a) Map of area southwest of the Iberian Peninsula with internal solitary wave crests marked in red based on one Envisat Advanced Synthetic Aperture Radar (ASAR) image in Wide-Swath Mode dated August 4, 2004 (22:27 UTC). (b) Ray-tracing diagram showing internal tide ray paths (in green; emanating from critical topography, in blue) along the black line in part (a). The small black squares in part (a) show where the ray path crosses the near-surface thermocline (taken at a depth of 50 m), and are also marked in part (b). (c) Full-resolution detail of a nonlinear internal wave train observed to propagate toward the west-northwest and believed to be generated by the tidal beam in part (b). (d) Radar backscatter profile showing the wavelength cross section of a nonlinear internal wave train generated by the tidal beam. The narrow rectangle in part (c) represents the cross section from which part (d) was obtained. The square in part (c) is a background backscatter reference used to normalize the radar profile in part (d). Zero in part (d) represents the average unperturbed backscatter of the SAR image in part (c).
Lateral contractions are not as common in the ocean as sills or banks, and it is much less intuitive to understand why a lateral contraction would result in vertical displacements of the interface and produce internal waves. But mathematical models using weakly nonlinear KdV type theory have successfully predicted “resonantly generated” upstream-propagating nonlinear solitary waves in a flat channel of slowly varying lateral dimension when the flow is nearly critical (e.g., Clarke and Grimshaw, 1994; Redekopp and You, 1995). Generation occurs especially in the presence of a vertical shear current. Laboratory experiments (e.g., Melville and Helfrich, 1987; Melville and Macomb, 1987) confirm these findings, but the situation has been less-comprehensively studied in the real ocean. When the bottom topography has undulating features whose crests are nearly perpendicular to the flow field, nonlinear internal waves can also appear as a result of near-critical flow (see, e.g., Pietrzak et al., 1990; Kranenburg et al., 1991; Bogucki et al., 1997; Groeskamp et al., 2011; Stastna, 2011). This kind of resonant generation of solitary waves, which theoretically includes mode-2 waves (as described below), is probably more frequently found in nature than nonlinear waves produced as a result of a lateral contraction.

In principle, shallow continental shelves where strong tidal currents can reach supercritical speeds are good candidates for the resonant generation mechanism of nonlinear internal waves. This mechanism has been measured and reported by Boguscki et al. (1997) on the California continental shelf, with implications for sediment resuspension and transport, and by da Silva and Helfrich (2008) at Race Point Channel in Massachusetts Bay. There, a flat channel between Race Point and Stellwagen Bank, with sill depth of just 50 m, contains tidal currents that reach supercritical speeds during the ebb phase of the tide. Da Silva and Helfrich (2008) show that the shear flow in Race Point Channel is supercritical in summer, and find numerous examples in SAR images of large-amplitude nonlinear internal waves generated within the channel, near the minimum breadth between Stellwagen Bank and the northern tip of Cape Cod. These solitons begin with short crest lengths (1–3 km) but very distinct amplitudes, detectable in the SAR images (see inset in Figure 6). As the ebb flow slackens, the nonlinear internal waves are observed to propagate upstream and develop into long-crested solitons whose crest lengths exceed 15 km. These long-lived waves propagate into shallow water and dissipate in Massachusetts Bay (see da Silva and Helfrich, 2008, for more details). Their horizontal structure is curved, in the form of a semicircle, which suggests a point source at Race Point (see Figure 6).

A remarkable feature of the nonlinear internal waves generated at the channel is the existence of two distinct wave trains that appear to be generated during the same tidal cycle at effectively the same location (wave trains A and B in Figure 6). This generation mechanism is different from the typical lee wave mechanism, which only produces a single train of internal waves each tidal cycle (Chereskin, 1983). Cummins et al. (2003) show evidence of nonlinear internal waves generated in the supercritical regime that propagate upstream in Knight Inlet, British Columbia, but only a single train of nonlinear internal waves are observed each tidal cycle. At Cape Cod’s Race Point Channel, a pair of internal wave trains is systematically observed to form during the ebb tide. Da Silva and Helfrich (2008) assume the resonant flow during the ebb phase of the tide generates both of these wave trains and speculate that the first of these pairs of wave trains is generated during the acceleration phase of the flow passing through the supercritical speeds, while the second is generated when the flow decelerates from supercritical to subcritical speeds (in a similar manner to the river plume mechanism described below, except here there is no density front). The timing of SAR images seems to be right for observing the generation that occurs during accelerating and decelerating flows through the supercritical regime (see Figure 11 of da Silva and Helfrich, 2008). Both of these nonlinear internal wave trains propagate upstream into Massachusetts and Cape Cod Bays, but with slightly different propagation directions, as Figure 6 shows (wave trains A and B). Although these observations are consistent with the theories of resonant generation, additional modeling and field measurements are required before a final conclusion can be made.

**Plume Generation**

Until this point, this paper has considered pycnocline displacements that evolve into nonlinear internal waves resulting from various forms of tide topography interaction. However, the pycnocline can also be displaced by the interaction between different water masses. Plumes, from either a river outflow or a throughflow current, have been found to produce nonlinear internal...
waves (Nash and Moum, 2005; Matthews et al., 2011). The plume generation can be thought of as an example of resonant generation, with the plume taking on the role previously played by the topography (i.e., as the method for imparting vertical displacements of the pycnocline). Rivers discharge fresh water into the ocean in tidally modulated pulses (Figure 7) that impinge upon the existing coastal waters. The initial momentum of the impinging fresh water plume, and the wind and current conditions at the time of discharge, determine the plume’s character when it enters the ocean. The impinging plume causes a buildup (or convergence) at the plume’s leading edge, due to the plume’s higher horizontal velocity relative to the ocean water. The plume represents a supercritical flow ($F > 1$), where an internal wave produced by the displacement would move more slowly than the plume, so the wave remains trapped at the plume’s leading edge. As the propagation speed of the plume decreases (i.e., the flow transitions from supercritical [$F > 1$] to subcritical [$F < 1$]), the internal wave separates from the plume front and radiates away as a freely propagating wave.

Nash and Moum (2005) documented nonlinear internal wave generation by the Columbia River outflow in Oregon. In subsequent observations of this plume, Kilcher and Nash (2010) noted how plume remnants from previous tidal cycles strengthened stratification and enhanced the generation of nonlinear internal waves. Matthews et al. (2011) showed how a plume associated with the Indonesian throughflow could produce “irregular internal waves” at Lombok Strait during certain times of the year (in addition to the “regular” internal waves generated by the tide topography interaction).

Two SAR images acquired during the summer of 2007 (Figure 7) show the evolution of nonlinear internal waves generated by the Columbia River plume. Figure 7A shows a strong river plume

Figure 6. A TerraSAR-X image of the northern tip of Cape Cod acquired April 7, 2008, at 22:25 UTC, showing a semicircle-shaped train of internal waves and “nascent solitons” in Race Point Channel. A pair of packets is generated every tidal cycle. A backscatter profile of the “nascent” solitons is shown at lower right.
signature along with the signature of an early-stage plume-generated internal wave packet leading the plume’s intrusion into the Pacific. Figure 7B shows the more mature signature of a classic rank ordered packet (after approximately five hours of evolution). Both images also show shoreward-propagating nonlinear internal waves, generated at the shelf break.

MODE-2 WAVE GENERATION

Although first-mode internal waves are most commonly observed in the ocean, higher-mode waveforms also occur under the proper conditions. A mode-2 wave (also known as varicose wave) is best thought of in terms of a three-layer stratified ocean. In this configuration, the wave travels as a bulge in the middle layer, displacing isopycnals upward into the upper layer and downward into the bottom layer (e.g., Shroyer, 2008).

Much of the work on mode-2 nonlinear internal waves has been theoretical or numerical, or involved laboratory experiments (see Yang et al., 2009, for an extensive reference list). This body of work has put forth no fewer than five possible generation mechanisms. Three of the mechanisms involve a mode-1 nonlinear internal wave: a mode-1 wave propagating onshore and entering the breaking instability stage (Helfrich and Melville, 1986), a mode-1 wave flowing over a sill (Vlasenko and Hutter, 2001), and reflection of a mode-1 wave (Chao et al., 2006). In addition, more recent results by Stastna (2011) show that mode-2 resonant generation is possible when appropriately small bottom length scale undulations are chosen as bottom topographic features (smaller than for mode-1 efficient resonant generation).

Because, for KdV type theory, mode-2 solitary waves generally have shorter length scales compared to mode-1 solitary waves, this result was expected.

There have been only a few reports of mode-2 wave observations in the ocean and none have included direct observation of the generation process. The waves have been found to occur around the Mascarene Ridge (Konyaev et al., 1995; da Silva et al., 2011), on the New Jersey shelf (Shroyer et al., 2010), in the South China Sea on the shelf north of the Dongsha coral reef (Yang et al., 2009), and at Knight Inlet in British Columbia (Farmer and Smith, 1980).

Yang et al. (2009) report an extensive set of in situ observations of mode-2 waves from the upper continental slope of the northern South China Sea. They describe the characteristics of 78 nonlinear internal mode-2 waves.

---

Figure 7. Radarsat-1 SAR images from July 28 and August 7, 2007, showing the evolution of nonlinear internal wave packet generated by the Columbia River plume.
(20 occurring during the summer and 58 during the winter) observed at a mooring deployed at 350 m depth from April 29 to July 28, 2005, and again from November 2, 2005, to February 24, 2006. Although the mooring observations of mode-2 waves observed by Yang et al. (2009) occurred in a region of the continental shelf, with extensive mode-1 nonlinear internal wave activity, the relationship between the mode-1 and mode-2 occurrences was not apparent. During the winter, 72% of the mode-2 nonlinear internal waves were observed without a prior mode-1 event. That number dropped to only 10% during the summer (at a time of extensive mode-1 activity). Yang et al. (2009) believe that the mode-2 nonlinear internal waves in winter could be related to the main thermocline being located near mid-depth. This scheme is more in line with the proposed generation mechanism of (Stastna and Peltier, 2005), but additional study is required to fully explain the occurrences.

Shroyer et al. (2010) found the mode-2 nonlinear internal waves on the New Jersey shelf to be much more ephemeral, observing just eight events during the summer of 2006. Not only did the waves appear sporadically in mooring records, they also appeared to be very short-lived; mode-2 packets were only tracked between two moorings separated by a kilometer, and they were not present at other moorings roughly 10 km distant (along the propagation direction). One reason for this observation is that mode-2 waves are subject to a resonance with mode-1 waves that may rapidly drain energy from the mode-2 waves, contributing to the sparsity of observations. Again, mode-2 wave generation was not directly observed in this study, but was believed to be the result of tidal forcing near the shelf break or frontal intrusions (Shroyer et al., 2010).

Although mode-2 waves are short-lived on the New Jersey shelf, they have significant lifetimes over the Mascarene Plateau in the Indian Ocean where mode-2 solitons were identified by means of in situ measurements and SAR observations (Konyaev et al., 1995; da Silva et al., 2011). Figure 8 shows an Envisat ASAR (Advanced SAR) image with clear evidence of a mode-2 wave.
structure, followed by a wave tail of short mode-1 waves. In Figure 8a, the SAR image intensity of mode-1 waves is characterized by bright, enhanced backscatter preceding dark reduced backscatter along the nonlinear internal wave propagation direction (in agreement with Alpers, 1985). For mode-2 solitary wave structures, the polarity of the SAR signature is reversed (because the location of the convergent and divergent zones are reversed), and thus a dark reduced backscatter crest precedes a bright, enhanced backscatter feature in the propagation direction of the wave. Moreover, Figure 8b shows a radar profile of a mode-2 large wave followed by a tail with several mode-1 shorter internal waves. It should be noted, however, that these mode-2 waves are not truly solitary, as mode-2 solitary waves develop oscillatory tails (of mode-1). These tails can drain energy from the main wave, as mentioned above, and grow in amplitude, as demonstrated in Akylas and Grimshaw (1992). Vlasenko et al. (2010) proposed a mechanism through which these mode-2 waves with mode-1 tails can be generated. For stable mode-2 structures that survive long enough, there is a chance that mode-2 solitary waves are “overtaken” by mode-1 solitons generated in previous tidal cycles. Zabusky and Kruskal (1965) showed that for weakly nonlinear theory, after a collision between mode-1 solitons, both waves preserve their shapes (but experience only a phase shift). For sufficiently nonlinear solitons, this collision may no longer be shape preserving, as in the case of a faster (mode-1) soliton overtaking a slower (mode-2) solitary wave (see, e.g., Akylas and Grimshaw, 1992, and Vlasenko et al., 2010).

SUMMARY
The understanding of nonlinear internal wave generation has continued to evolve since Maxworthy (1979) proposed the lee wave generation mechanism. This evolution has been driven by in situ studies, the ever-increasing amount of satellite imagery, and, in more recent years, the ability to numerically simulate the workings of the ocean over larger areas at high resolution. Important new results in nonlinear internal wave research have been published in the last few years, including remote-sensing observations of resonant generation (da Silva and Helfrich, 2008), significant first time in situ measurements of mode-2 waves (Yang et al., 2009; Shroyer et al., 2010), and advances in prediction ability (Simmons et al., 2011). But the understanding of nonlinear internal waves is still incomplete. The exact generation mechanism of some nonlinear internal wave occurrences is still unclear (see Jackson, 2007; Alford et al., 2012, in this issue) and mode-2 generation has not yet been observed in the ocean. Given their widespread distribution and the important role they play in connecting the large-scale tides to smaller-scale turbulence, nonlinear internal waves are expected to remain an important area of study for the future.

ACKNOWLEDGEMENTS
Christopher Jackson gratefully acknowledges the support of the Office of Naval Research through contract N0001409C0224. José C.B. da Silva thanks J.M. Magalhães for help with the preparation of Figures 4–6 and 8. The authors thank ESA for providing the ERS and Envisat imagery under project AOPT-2423 and DLR for providing the TerraSAR-X imagery under project OCE0056.

REFERENCES


}