CORRECTIONS TO THE KdV APPROXIMATION FOR WATER WAVES

J. DOUGLAS WRIGHT†

This article is dedicated to the memory of Dr. Joseph Hammack

Abstract. In order to investigate corrections to the common KdV approximation for surface water waves in a canal, we derive modulation equations for the evolution of long wavelength initial data. We work in Lagrangian coordinates. The equations which govern corrections to the KdV approximation consist of linearized and inhomogeneous KdV equations plus an inhomogeneous wave equation. These equations are explicitly solvable and we prove estimates showing that they do indeed give a significantly better approximation than the KdV equation alone.

Key words. water wave equation, KdV equation, linearized KdV equation, modulation equations, solitary waves, solitons, collisions of solitary waves

AMS subject classifications. 76B15, 35Q51, 35Q53

DOI. 10.1137/S0036141004444202

1. Introduction. It is often easier to write down a partial differential equation which models a physical phenomena than it is to study solutions of such an equation. Equations which model the evolution of the surface of a fluid in a canal have been known since at least the 19th century; however, it has only been in recent years that questions of existence and uniqueness for general initial data have been answered (see [35], [36], and [22]). Moreover, numerical simulations of water waves and similarly complex phenomena are frequently time consuming and challenging to implement. Consequently, it can be quite difficult to say much about the behavior of a general solution. Therefore scientists often restrict their attention to limiting cases—for instance, one may assume that solutions are of long wavelength and small amplitude (see Figure 1). Under such a supposition, a modulation equation may be (formally) derived. In particular, one hopes that the modulation equation

- is well-posed,
- is either explicitly solvable or easy to solve numerically, and
- captures the essential behavior of the original system.

Remarkably, many seemingly disparate physical phenomena possess modulation equations of the same form. For solutions of long wavelength, Korteweg–de Vries (KdV) equations are often used as modulation equations for a wide variety of nonlinear dispersive systems, including the water wave equation, the Euler–Poisson equations for plasma dynamics, and the Fermi–Pasta–Ulam equation for the interaction of particles in an infinite lattice.

Despite the fact that modulation equations have been in use for over a hundred years—the KdV equation was first proposed as a model for water waves by Boussinesq in 1872 and also by Korteweg and de Vries in 1895—only recently have attempts been made to rigorously connect the behavior of the modulation equations to the original
physical problem. In particular, through the work of Craig [10], Kano and Nishida [18], [19], Kalyakin [17], Schneider [27], Ben Youssef and Colin [2], and Schneider and Wayne [28], [29], the validity of KdV equations as a leading order approximation to the evolution of long wavelength water waves and to a number of other dispersive partial differential equations has been established.

In many respects, the KdV equation is an ideal modulation equation; it is simple in form and explicitly solvable via the inverse scattering transform. Nevertheless, both experimentally and numerically one observes deviations from the predictions of the KdV approximation. In this paper we derive a hierarchy of modulation equations which govern corrections to the KdV model and also prove rigorously that these higher order equations do indeed improve the accuracy of the approximation. While the correction is valid in general long wavelength/small amplitude settings, heuristically the model is set up to better approximate interactions between solitary waves—both counterpropagating collisions and unidirectional interactions.

Note that in [33], as a case study, Wayne and Wright examined higher order corrections to the KdV approximation to a Boussinesq equation. As the KdV equation is in some sense a universal approximation for long waves, we expect that the equations for corrections to this approximation will also be universal. Indeed, our results show that the higher order corrections for the water wave equation are nearly identical to those for the Boussinesq equation. Also, since a significant part of the work for the Boussinesq problem consists of showing that the modulation equations have well-behaved solutions over the time scales of interest, this is of use in tackling the water wave problem.

We now describe our results in some detail. The equations of motion for a water wave in an infinitely long canal (commonly called the water wave equation) are

\[
\begin{align*}
\frac{\partial^2 x}{\partial t^2} (1 + x_\alpha) + y_\alpha (1 + y_{tt}) &= 0, \\
y_t &= K(x, y)x_t, \\
\alpha &\in \mathbb{R}, \quad t \geq 0, \quad (x(\alpha, t), y(\alpha, t)) \in \mathbb{R}^2,
\end{align*}
\]

where \(K(x, y)\) is a complicated operator (see section 3 for the definition of \(K\)) and \((\alpha + x(\alpha, t), y(\alpha, t))\) parameterizes the free surface. According to the KdV approximation results of [28], to the order of the approximation long wavelength solutions of (WW) split up into two pieces, one a right-moving wave train and one a left-moving wave train. Each of these wave trains evolves according to a KdV equation, and there is no interaction between the left- and right-moving pieces. That is, for \(0 < \epsilon \ll 1\), if we scale amplitudes to be \(O(\epsilon^2)\) (i.e., small) and wavelengths to be \(O(\epsilon^{-1})\) (i.e., long), then for times of \(O(\epsilon^{-3})\), solutions to (WW) satisfy

\[
-x_\alpha(\alpha, t) = \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t) + O(\epsilon^4),
\]

\[\text{Fig. 1. The long wave, small amplitude scaling.}\]
where $U$ and $V$ satisfy the KdV equations

$$-2\partial_T U = \frac{1}{3} \partial^3_{\beta^-} U + \frac{3}{2} \partial_{\beta^-} (U^2),$$

(KdV)

$$2\partial_T V = \frac{1}{3} \partial^3_{\beta^+} V + \frac{3}{2} \partial_{\beta^+} (V^2).$$

See Figure 2. Here, $\beta_{\pm} = \epsilon (\beta \pm t)$ represent long wavelength moving reference frames and $T = \epsilon^3 t$ is the very long time scale coordinate. For technical reasons (which we discuss later), $-x_\alpha$ is the natural variable to estimate for the water wave equation. To lowest order, $-x_\alpha$ is proportional to the height of the wave. At higher order, this ceases to be true, though for purposes of intuition one can think of $-x_\alpha$ as representing the wave amplitude.

The KdV equation was initially derived from the water wave equation in an attempt to prove the existence of a solitary wave solution for waves in a canal. Famously, the KdV equation admits solitary wave solutions and also multisoliton solutions. (See Figure 3.) Note that the existence of a true solitary wave solution to (WW) was established first by Friedrichs and Hyers in [14] and subsequently by Beale in [1]. Also note that there exist standing wave solutions (also known as Stokes’ waves) to (WW) which in many ways are similar to solitary waves, though our methods do not carry over to studying such behavior—see Toland [32] and Constantin and Strauss [8] for an overview of Stokes’ waves in their irrotational and rotational problems, respectively.

We will frequently refer to multisoliton solutions as “overtaking wave” collisions. We remind the reader that the only notable first order effect after such a collision is that the waves are phase shifted after the collision.

Given the results in [28], one expects to see similar behavior in solutions systems modeled by KdV equations. Though it is unknown if these soliton-like solutions persist globally, analogous behavior is indeed observed for very long times (see [15]). The most notable deviation between true solutions and the KdV approximation is the size of the phase shift after a collision. In addition, soliton-like solutions to the type of systems we study frequently develop a very small amplitude dispersive wave train behind each soliton, which moves in the same direction; see Figure 4. The KdV approximation does not predict the existence of these dispersive wave trains. As these sorts of discrepancies are observed even in the case where there is only one wave train moving unidirectionally, we believe that they are, loosely, independent of interactions between the left- and right-moving wave trains. They reflect intrinsic differences between the approximation and the original system.

On the other hand, there is evidence that a noticeable interaction takes place between the left- and right-moving waves. One can see from the form of the approximation in (1) that during a head-on collision of waves moving in opposite directions, the KdV approximation predicts that the heights of the waves add linearly. In true head-on collisions in solutions to the water wave equation, however, the height of the
waves is slightly different from the sum of the heights of the waves taken separately—it is slightly larger. (See the works of Maxworthy [23], Byatt-Smith [6], [7], Cooker, Weidman, and Bale [9], and Su and Mirie [30], [31].) We sketch this in Figure 5.

Thus we might expect two types of corrections to the KdV approximation:

- corrections due to the fact that, even in the case of a purely right- (or left-) moving wave train, solutions to the water wave equation are not exactly described by solutions to the KdV equation—we will refer to this source of error as unidirectional error;
- corrections due the fact that the left- and right-moving wave trains will interact at higher order—we call such errors counterpropagation error.

Both of these types of corrections are apparent in our results, and to incorporate these two types of corrections, we add an additional three functions to the KdV wave trains. The first two, $F$ and $G$, will correct for unidirectional errors. The third, $P$, will correct for counterpropagation errors. We scale the amplitudes of these three functions so they are $O(\epsilon^4)$, which is the same as the order of the error in using only the KdV equations. $F$ and $G$ will take the same functional form as $U$ and $V$, as they correct for differences between the approximate and actual wave trains. Therefore we
add
\[ \epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t) \]
to the first order approximation (1).

We do not expect \( P \) to be moving strictly left or right, as it corrects for the interaction between waves moving in opposite directions. Thus its spatial dependence will be on \( \beta = \epsilon \alpha \). Suppose that the functions \( U \) and \( V \) are solitary wave solutions. In the long wavelength variables we are considering, this means that the right- and left-moving wave packets are large only over a length of \( O(\epsilon^{-1}) \). In addition, the reference frame moves with unit velocity. Thus we expect any interaction of the two waves to last a time of \( O(\epsilon^{-1}) \). Accordingly, we let \( P \) depend on the time variable \( \tau = \epsilon t \). That is, we add a correction term of the form
\[ \epsilon^4 P(\epsilon \alpha, \epsilon t) \]
to the KdV model.

Through formal means we find that \( P \) satisfies an inhomogeneous wave equation
\[ \partial_x^2 P - \partial_\beta^2 P = 3 \partial^2_{\beta} \left(U(\beta - \tau, \epsilon^2 \tau)V(\beta + \tau, \epsilon^2 \tau)\right). \]

Similarly, \( F \) and \( G \) satisfy a pair of driven, linearized KdV equations
\[ -2\partial_T F = \frac{1}{3} \partial^3_{\beta_+} F + 3 \partial_{\beta_-} (UF) + J^- , \]
\[ 2\partial_T G = \frac{1}{3} \partial^3_{\beta_+} G + 3 \partial_{\beta_-} (VG) + J^+ . \]

Notice that these equations are linearized about the KdV solutions \( U \) and \( V \). The inhomogeneous terms \( J^- \) and \( J^+ \) are made up of a combination of sums and products of \( U, V, \) and \( P \). For explicit forms of these driving terms see equations (22). Linearized KdV equations are explicitly solvable, though this is a complicated matter (see [25] and [16]). However, solutions are simple to compute numerically.

One can solve inhomogeneous wave equations explicitly and easily via the method of characteristics. Moreover, we can reduce such systems to a pair of transport equations by the following fact.
FACT 1. If $\partial_t f - \partial_x f = 1/2\partial_x h$ and $\partial_t g + \partial_x g = -1/2\partial_x h$, then $q = f + g$ satisfies $\partial^2_q - \partial^2_q = \partial^2_q h$.

Thus, we have

$$P^-_\tau + P^-_\beta = \frac{3}{2} \partial_\beta (U(\beta - \tau, e^2\tau)V(\beta + \tau, e^2\tau)),$$

$$P^+_\tau - P^+_\beta = \frac{3}{2} \partial_\beta (U(\beta - \tau, e^2\tau)V(\beta + \tau, e^2\tau)),$$

where

$$P(\beta, \tau) = P^+(\beta, \tau) + P^-(\beta, \tau).$$

We remark that the initial data for the modulation equations is determined from initial conditions for the original system in ways described in section 6. Also, this hierarchy of higher order modulation equations is nearly identical to that derived in Wayne and Wright [33] for the Boussinesq equation—the chief difference lying in the specific forms of the inhomogeneous terms $J^\pm$.

To enforce the notion of spatial localization, we will be considering initial data which is of rapid decay, that is, initial data in

$$H^s(m) = \left\{ f(\alpha) | (1 + \alpha^2)^{m/2} f(\alpha) \in H^s \right\}. $$

The inner product on $H^s(m)$ is given by

$$(f(\cdot), g(\cdot))_{H^s(m)} = \left( (1 + \cdot^2)^{m/2} f(\cdot), (1 + \cdot^2)^{m/2} g(\cdot) \right)_{H^s},$$

where we use the standard inner product in $H^s$. In particular, the known soliton solutions of the KdV equations are in such spaces.

That the KdV equations have solutions for all times with this sort of initial data is well known. In particular, we have the following theorem from [28].

**Theorem 1.** Let $\sigma \geq 4$. Then for all $C_1, T_0 > 0$ there exists $C_1 > 0$ such that if $U, V$ satisfy (KdV) with initial conditions $U_0, V_0$ and

$$\max \left\{ \|U_0\|_{H^s(4) \cap H^{s+3}(2) \cap H^{s+9}}, \|V_0\|_{H^s(4) \cap H^{s+4}(2) \cap H^{s+9}} \right\} < C_1, $$

then

$$\sup_{T \in [0, T_0]} \left\{ \|U(\cdot, T)\|_{H^s(4) \cap H^{s+3}(2) \cap H^{s+9}}, \|V(\cdot, T)\|_{H^s(4) \cap H^{s+4}(2) \cap H^{s+9}} \right\} < C_1. $$

On the other hand, it is less clear that solutions of (IW) and (LK) will remain bounded over the very long time scales necessary for the KdV approximation. In [33] we proved the following result which guarantees that the solutions of the modulation equations remain bounded for sufficiently long times.

**Proposition 1.** Fix $T_0 > 0$ and $\sigma > 11/2$. Suppose that $U_0, V_0$ satisfy (2) and $U, V, P^\pm, F, G$ satisfy (KdV), (T), and (LK). Then there exists a constant $C_2$, independent of $\epsilon$, such that the solutions of (T) and (LK) satisfy the estimates

$$\sup_{\tau \in [0, T_0 \epsilon^{-1}]} \|P^\pm(\cdot, \tau)\|_{H^{s+3}} \leq C_2,$$

$$\sup_{T \in [0, T_0]} \left\{ \|F(\cdot, T)\|_{H^s \cap H^{s-4}(2)}, \|G(\cdot, T)\|_{H^s \cap H^{s-4}(2)} \right\} \leq C_2.$$
Moreover, \( P^\pm(\beta, \tau) = \varphi^\pm(\beta_\pm, T) \) with
\[
\sup_{T \in [0, T_0]} \| \varphi^\pm(\cdot, T) \|_{H^{s+3} \cap H^{s-1}} \leq C_2.
\]

Finally we note that since \( \partial_\tau P = \partial_\beta P^+ - \partial_\beta P^- \) we have \( \| \partial_\tau P \|_s \leq \| \partial_\beta P \|_s \).

With this preliminary result in hand we can state our principal results. Denote the sum of the modulation functions, properly scaled, as
\[
-\psi^d(\alpha, t) = \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t)
+ \epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t) + \epsilon^4 P(\alpha, ct)
\]
\[
(4)
\]
and
\[
\psi^u(\alpha, t) = \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) - \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t)
+ \epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) - \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t)
+ \epsilon^4 \varphi^-(\epsilon(\alpha - t), \epsilon^3 t) - \epsilon^4 \varphi^+(\epsilon(\alpha + t), \epsilon^3 t)
\]
\[
(6)
\]
As mentioned earlier, \((x, y)\) are not the natural coordinates to study solutions to (WW). The coordinates we use are \((x_\alpha, y_\alpha, x_\tau)^T\). \(x_\alpha\) is approximated by \(\psi^d\), and the functions \(y\) and \(x_\tau\) are approximated by functions we denote \(\psi^y\) and \(\psi^u\), respectively. They are given by
\[
\psi^y(\alpha, t) = \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t)
+ \epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t) + \epsilon^4 P(\alpha, ct)
\]
\[
(5)
\]
We discuss the origin of these equations in section 4.

The approximation will be valid in the space
\[
S_f = H^s \times H^s \times H^{s-1/2}.
\]

Our main result is the following.

**Theorem.** Fix \( T_0, C_f > 0, s > 4, \sigma \geq s + 7 \). Suppose that \( U, V, P, F, \) and \( G \) satisfy (KdV), (IW), and (LK) and that \( \psi^d, \psi^y, \) and \( \psi^u \) are the combinations of these functions given in (4), (5), and (6). Then there exist \( \epsilon_0 > 0 \) and \( C_f > 0 \) such that the following is true. If the initial conditions for (WW) are of the form
\[
((x_\alpha(0, 0), y(0, 0), x_\tau(0, 0))^T = (0, \epsilon^2 \Theta_y(e\alpha), \epsilon^2 \Theta_u(e\alpha))^T \) with
\[
\max_{1 \leq i < j \leq 2} \{ ||\Theta_i(\cdot)||_{H^{s+1}} \cap H^{s+1} \cap H^{s+1} \} \leq C_I,
\]
then for \( \epsilon \in (0, \epsilon_0) \) there is a reparameterization of the free surface such that the unique solution to (WW) satisfies
\[
\left\| \begin{pmatrix} x_\alpha(\cdot, t) \\ y(\cdot, t) \\ x_\tau(\cdot, t) \end{pmatrix} - \begin{pmatrix} \psi^d(\cdot, t) \\ \psi^y(\cdot, t) \\ \psi^u(\cdot, t) \end{pmatrix} \right\|_{S_f} \leq C_f \epsilon^{1/2}
\]
for \( t \in [0,T_0\epsilon^{-3}] \). The constant \( C_\epsilon \) does not depend on \( \epsilon \).

**Remark 1.** The loss of the one-half power of \( \epsilon \) in Theorem 2 is caused by the long wave scaling and not a lack of sharpness in the estimates.

**Remark 2.** It is clear that the form of the initial conditions specified in the hypotheses of this theorem does not agree with that found by setting \( t = 0 \) in the approximation inequality (unless, of course, \( \psi^d \equiv 0 \)). This is precisely why we mention the need to reparameterize the free surface. We discuss this at length in section 6.

**Remark 3.** In [28], Schneider and Wayne prove the analogous result for the KdV (i.e., \( O(\epsilon^4) \)) approximation alone. The methods used here follow largely from their work.

**Remark 4.** The long time scale on which this theorem holds is natural given the form of the approximation—\( t \sim O(\epsilon^{-3}) \) corresponds to times \( T \) of \( O(1) \) in the KdV equation. The methods of the proof rely on a Gronwall-type estimate which imposes this time scale. Moreover, while Proposition 1 shows that the functions \( F \) and \( G \) remain \( O(1) \) for \( t \sim O(\epsilon^{-3}) \), estimates show that beyond this they are likely to grow very rapidly, ruining any chance for the higher order correction to be valid for longer times. Nonetheless, it is unclear whether or not the approximation of water waves by the KdV solutions alone is valid for even longer times. Numerical experiments done for the KdV approximation to the Euler–Poisson equations by Haragus, Nicholls, and Sattinger in [15] indicate that the approximation is valid for very long times, but little is known analytically.

Less technically, this theorem states that solutions to (WW), in the long wavelength limit, satisfy

\[
x_\alpha(\alpha, t) = \psi^d(\alpha, t) + O(\epsilon^6)
\]

for times of \( O(\epsilon^{-3}) \). This is, as expected, a marked improvement over the use of KdV alone.

We note that this is not the first time that linearized KdV equations have been put forward as a means to improve the accuracy of the KdV approximation. Other instances where linearized KdV equations appear include Sachs [26], Haragus, Nicholls, and Sattinger [15], Kodama and Taniuti [21], and Drazin and Johnson [12]. Moreover, there have been numerous models put forward over the years which model water waves in the same scaling regime we are considering. We refer the reader to Kodama [20], Olver [24], Bona, Pritchard, and Scott [5], Craig and Groves [11], Dullin, Gottwald, and Holm [13], and Bona and Chen [4]. Much of the work done in the above papers pertains to analyzing the behavior of the model equations and not to their connection to the original system. A notable exception is the recent work by Bona, Colin, and Lannes [3], wherein they prove the rigorous validity of a large number of Boussinesq-style models. There, they prove that such models are at least as accurate as the KdV approximation alone (though not as accurate as the model described here) and for the same time scales. Their models are designed, in part, to be easy to approximate numerically and are not, in general, explicitly solvable. Our particular combination of linearized KdV equations with an inhomogeneous wave equation appears to be unique and is asymptotically the most accurate model for long wavelength solutions to the water wave equation which has currently been justified rigorously.

The remainder of this paper is organized as follows. First, in section 2, we conduct a preliminary discussion of the water wave equation. Sections 3 and 5 contain a thorough discussion of the operator \( K(x,y) \). Then, in section 4 we derive the higher order modulation equations and prove an important estimate. In section 6 we prove
CORRECTIONS TO THE KdV APPROXIMATION

1169

\[ \Omega(t) = \{ (\alpha + x(\alpha, t), y(\alpha, t)) | \alpha \in \mathbb{R} \} \]

\[ \Gamma(t) = \{ (\alpha, y(\alpha, t)) | \alpha \in \mathbb{R} \} \]

**Fig. 6.** The water wave in Lagrangian coordinates.

the validity of the approximation, i.e., Theorem 2. Finally, section 7 contains the details for a number of proofs.

2. Preliminaries. We begin by discussing the water wave problem in greater detail. Consider an infinitely long canal of unit mean depth in two dimensions (see Figure 6). We denote the region occupied by the fluid at time \( t \) as \( \Omega(t) \), and the upper surface as \( \Gamma(t) \). We parameterize \( \Gamma(t) \) by \((\tilde{x}(\alpha, t), y(\alpha, t))\), where \( \alpha \in \mathbb{R} \) is the parameter and \( \tilde{x} \) and \( y \) are the real-valued coordinate functions. It is useful to break \( \tilde{x} \) up as follows:

\[ \tilde{x}(\alpha, t) = \alpha + x(\alpha, t). \]

We consider fluids which are inviscid and incompressible and flows which are irrotational. Also, we assume that the pressure on the top surface is constant and that the acceleration due to gravity is 1. With these assumptions, the evolution of \( x \) and \( y \) are given by the equations

\[
\begin{align*}
    x_{tt}(1 + x_{\alpha}) + y_{\alpha}(1 + y_{tt}) &= 0, \\
    y_t &= K(x, y)x_t.
\end{align*}
\]

(See [10].)

The first of these two equations is found from Euler’s equations for fluid motion. The operator \( K \) in the second line is a transformation which is linear in \( x_t \), but depends nonlinearly on \( (x, y) \). That such an operator exists and gives a relationship between \( x_t \) and \( y_t \) is discussed in sections 3 and 5, along with an analysis of \( K \). Much of the difficulty in answering questions about the water wave equation is related to this operator. If the surface of the water is perfectly flat, i.e., \( x = y = 0 \), then we have \( K(0, 0) = K_0 \), where \( K_0 \) is a linear operator defined by \( \hat{K}_0 f(k) = \hat{K}_0(k) \hat{f}(k) \), with \( \hat{K}_0(k) = -it \tanh(k) \). Notice that since \( \tanh \) is a bounded function, \( K_0 \) is bounded from \( H^s \) to \( H^s \). \( K_0 \) will be appearing frequently.

This formulation of the water wave problem is said to be in Lagrangian, or material, coordinates. In this point of view we are not in a fixed “lab” frame, but instead we are tracking the position of each “particle” of water separately. That is, \((\alpha + x(\alpha, t), y(\alpha, t))\) gives the location of the particle which was initially at \((\alpha + x(\alpha, 0), y(\alpha, 0))\). The laboratory, or Eulerian, point of view is to fix a system of coordinates on the fluid domain and to measure the velocity of the fluid at each point...
of this fixed reference frame. For our purposes, it is far more convenient to use Lagrangian coordinates; however, experimentalists work with Eulerian coordinates. The coordinate change from one frame to the other is nontrivial and adapting our approximation results to the Eulerian point of view requires some involved computations. We will carry the change of variables out and give formulae for the approximation in terms of Eulerian coordinates in a future publication. The interested reader may also see the author’s thesis [34] for this information. We remark that the results of Bona, Colin, and Lannes on the approximation of waterwaves by Boussinesq models in [3] are proven in Eulerian coordinates, and as such do not contend with this issue.

Since the waterwave equation is second order in time for both \( x \) and \( y \), one might suppose that four functions are necessary to specify the initial state of the system—\( x(\alpha, 0), y(\alpha, 0), x_t(\alpha, 0), \) and \( y_t(\alpha, 0) \). In fact, in general only three are needed, and for initial data with small amplitudes, only two are needed. The relationship \( y_t = K(x, y)x_t \) specifies the value of \( y_t(\alpha, 0) \) given the other three functions. We can also “do away” with the initial condition for \( x \), provided we are in the small amplitude, long wave limit. If \( x(\alpha, 0) \) and its first derivative are sufficiently small, then \( \alpha + x(\alpha, 0) \) will be invertible. This implies that \( \Gamma(0) \) will be a graph over the horizontal coordinate. And so, without loss of generality, we can reparameterize the initial conditions so that

\[
\Gamma(0) = \{ (\alpha, y(\alpha, 0)) | \alpha \in \mathbb{R} \}.
\]

Thus we need only to choose \( y(\alpha, 0) \) and \( x_t(\alpha, 0) \). As it turns out, we need to reparameterize the system one more time to prove the approximation theorem, but we will leave this technicality until section 6. The essential point here is that due to the freedom in choosing the initial parameterization, we can eliminate two of our initial conditions.

Even though we can assume that \( x(\alpha, 0) = 0 \) (or, alternately, is small), this coordinate grows linearly in time. (See the linear estimates in Chapter 2 of [28].) As we are concerned with very long time scales, this is a problem. As was shown in [28], one can replace \( x \) with the new coordinate \( z = K_0 x \), which is well behaved over long times. Rewriting (WW) with this new variable, we have

\[
\begin{align*}
\partial_t z &= K_0 u, \\
\partial_t y &= K(z, y)u, \\
\partial_t u &= -\partial_\alpha y(1 + \partial^2_t y) / (1 + K_0^{-1} \partial_\alpha z).
\end{align*}
\]

We will see that the operator \( K(x, y) \) in truth depends not on \( x \) but on \( z \), so the abuse of the notation \( K(z, y) \) above is in some sense legitimate (see section 3 for further discussion). Furthermore, even though \( K_0^{-1} \) is not well defined, as it blows up at frequency \( k = 0 \), the composition

\[
L = K_0^{-1} \partial_\alpha
\]

is well defined, as its symbol, \(-k / \tanh(k)\), has no singularities; \( L \) is also invertible. Finally, we notice that the Maclaurin expansion of \( K_0(k) = -ik + O(k^3) \). Thus to lowest order \( K_0 \sim -\partial_\alpha \), and so \( z \sim -x_\alpha \). This is precisely the reason why, in the introduction, we stated that \(-x_\alpha \) is a natural coordinate for the water wave equation.

Though we will primarily be working with the three-dimensional system (WW3), we will need to embed this system into a four-dimensional system to prove certain
aspects of the validity of the approximation. We introduce the new coordinate \( a = u_t \), and (WW3) becomes

\[
\begin{align*}
\partial_t z &= K_0 u, \\
\partial_t y &= K(z, y)u, \\
\partial_t u &= a, \\
\partial_t a &= -\frac{a\partial_a u + \partial_a \partial_t (1 + \partial_t^2 y) + \partial_t y \partial_t^3 y}{1 + K_0^{-1} \partial_a z}.
\end{align*}
\]

(WW4)

Though things appear to be getting out of hand, we remark that this is as large a system as we will need. Results in [28] prove that solutions to (WW), (WW3), and (WW4) do indeed exist for long times. We will be considering solutions to the four-dimensional system which are in

\[ \mathcal{S}_0^s = H^s \times H^s \times H^{s-1/2} \times H^{s-1}. \]

The main goal of this paper is to prove Theorem 2. To do this, we first prove a similar theorem for solutions to (WW3), from which Theorem 2 will follow. Let

\[ \Psi^d(\alpha, t) = \psi^d(\alpha, t) + \epsilon^6 W_3(\epsilon \alpha, \epsilon t). \]

The additional function \( W_3 \) will solve an equation we specify later. While we assure the reader that we will be following Chekhov’s rule and that this gun which appears in the first act will be fired in the third, interested parties may look ahead to equation (23) in section 4 for more information about \( W_3 \). Define the functions

\[
\Psi^\pi = L^{-1} \Psi^d, \\
\Psi^y = \Psi^\pi + \epsilon^4 \Delta_1 + \epsilon^6 \Delta_2, \\
\Psi^a = \partial_a^{-1} \partial_t \Psi^d.
\]

\( \Delta_1 \) and \( \Delta_2 \) are combinations of solutions to the modulation equations and are given in (17) and (19) in section 4. We justify the presence of the inverse derivative in \( \Psi^a \) by means of example. \( \Psi^d \) contains the term \( U \) which solves a KdV equation. Thus \( \partial_t \Psi^d \) contains the terms from the right-hand side of the KdV equation, all of which are perfect space derivatives. These then are cancelled by the inverse derivative. It is simple to check that this method applies to all terms in \( \Psi^d \).

With these function, we have the following results.

**Theorem 3.** Fix \( T_0, C_1 > 0, s > 4, \sigma \geq s + 7 \). Let \( \Psi^d, \Psi^z, \Psi^y, \) and \( \Psi^a \) be as above. Moreover assume that the initial conditions for (KdV) satisfy

\[
\max \{ \| U_0 \|_{H^s(4)} \cap H^{s+4}(2) \cap H^{s+9}, \| V_0 \|_{H^s(4)} \cap H^{s+4}(2) \cap H^{s+9} \} < C_1.
\]

Then there exist \( \epsilon_0 > 0 \) and \( C_F = C_F(T_0, C_1, s) > 0 \) such that if the initial conditions for (WW3) are of the form

\[
\begin{pmatrix}
z(\alpha, 0) \\
y(\alpha, 0) \\
u(\alpha, 0)
\end{pmatrix}
= \begin{pmatrix}
\Psi^z(\alpha, 0) \\
\Psi^y(\alpha, 0) \\
\Psi^a(\alpha, 0)
\end{pmatrix} + \epsilon^{11/2} \tilde{R}_0(\alpha)
\]

with \( \| \tilde{R}_0 \|_{\mathcal{S}_0^s} \leq C_1 \), then the unique solution of (WW3) satisfies the estimate

\[
\begin{pmatrix}
z(\cdot, t) \\
y(\cdot, t) \\
u(\cdot, t)
\end{pmatrix}
- \begin{pmatrix}
\Psi^z(\cdot, t) \\
\Psi^y(\cdot, t) \\
\Psi^a(\cdot, t)
\end{pmatrix}
\leq C_F \epsilon^{11/2}
\]

for \( t \in [0, T_0 \epsilon^{-3}] \). The constant \( C_F \) does not depend on \( \epsilon \).
3. The operator $K(x, y)$. Part I: Basics and basic expansions. This is the first of two sections where we discuss the operator $K(x, y)$ which gives the relation between $y_t$ and $x_t$ in the water wave equation. Here we briefly discuss the origin of this operator and report expansions of $K$ found in previous work. We also state some very basic facts about these expansions. In section 5 we quote more complicated results and prove some new technical extensions needed for our purposes.

$K(x, y)$ (or, more precisely $-K(x, y)$) is sometimes called the Hilbert transform for the region $\Omega(t)$. Loosely, given a region $\Omega$ in the complex plane, and any function $F$ which is analytic in $\Omega$, the Hilbert transform for $\Omega$, $H(\Omega)$, is a linear operator which relates the real and imaginary parts of $F$ on the boundary of $\Omega$. That is,

$$\text{Im}(F)|_{\partial\Omega} = H(\Omega)\text{Re}(F)|_{\partial\Omega}.$$  

For example, if $\Omega$ were the lower half-plane, then the Hilbert transform would be the operator $H$, given by

$$\hat{H}f = i \text{sgn}(k)\hat{f}.$$  

(This particular operator $H$ is also frequently called the Hilbert transform.) The nature of the operator depends greatly on the region being studied. Unsurprisingly, the proof that such an operator exists is connected to the Riemann mapping theorem and to techniques for solving boundary value problems for Laplace’s equation in the plane. In this problem, since the region $\Omega(t)$ is completely specified by the coordinate functions $x$ and $y$, we denote the Hilbert transform by $H(\Omega(t)) = -K(x, y)$.

As we are considering a fluid which is incompressible and a flow which is irrotational, $x_t(\alpha, t) - iy_t(\alpha, t)$ is the value, on the upper boundary $\Gamma(t)$, of an analytic function on $\Omega(t)$, $\omega(\alpha + i\beta) = v^x(\alpha, \beta) - iv^y(\alpha, \beta)$. Here $(v^x, v^y)$ is the velocity field for the fluid in the whole region. Thus, given that $K(x, y)$ exists, we have

$$y_t = K(x, y)x_t.$$  

Of course, the boundary of $\Omega(t)$ is not just $\Gamma(t)$, but also includes the bottom of the canal (i.e., where $\beta = -1$). As we do not have fluid flow through the bottom, we have $v^y(\alpha, -1) = 0$.

Under these conditions, $K(x, y)$ has been analyzed extensively by Craig in [10] and Schneider and Wayne in [28]. In particular Craig shows that $K(x, y)$ has the following expansion:

$$K(x, y)u = K_0u + K_1(x, y)u + S_2(x, y)u,$$

where

$$K_0u(k) = -i \tanh(k)\hat{u}(k),$$

$$K_1(x, y)u = [x, K_0]\partial_\alpha u - (y + K_0(yK_0))\partial_\alpha u,$$

and $S_2$ is quadratic in $(x, y)$.

First of all we note that $K_0$ is a bounded operator from $H^s$ to $H^s$ since $\tanh(k)$ is a bounded function. That is,

$$\|K_0u\|_s \leq \|u\|_s.$$  

The operator $L = K_0^{-1}\partial_\alpha$, which is well defined as we discussed in section 2, will also be used frequently. $L$ is not a bounded operator on $H^s$. It effectively takes one derivative. That is,

$$\|Lu\|_s \leq \|u\|_{s+1}.$$
On the other hand, $L^{-1}$ replaces one derivative. That is, since
\[ |\tanh(k)/k| \leq (1+k^2)^{-1/2}, \]
we know
\[ \|L^{-1}u\|_s \leq \|u\|_{s-1}. \]

We will be considering functions which are of long wavelength, that is, functions of the form $f(\alpha) = F(\beta)$ where $\beta = \epsilon \alpha$. We define operators $K_{0,\epsilon}$ and $L_\epsilon$ via
\[ K_{0,\epsilon}F(\beta) = K_0f(\alpha), \]
\[ L_\epsilon F(\beta) = Lf(\alpha). \]

Taking the Maclaurin series expansion for $\tanh(k)$ shows that formally
\[ K_{0,\epsilon} = -\epsilon \partial_\beta - \frac{1}{3} \epsilon^3 \partial_\beta^3 - \frac{2}{15} \epsilon^5 \partial_\beta^5 + O(\epsilon^7). \]

Similarly $L_\epsilon$ and $L_\epsilon^{-1}$ have expansions in terms of derivatives
\[ L_\epsilon = -1 + \frac{1}{3} \epsilon^2 \partial_\beta^2 + \frac{1}{45} \epsilon^4 \partial_\beta^4 + O(\epsilon^6), \]
\[ L_\epsilon^{-1} = -1 - \frac{1}{3} \epsilon^2 \partial_\beta^2 - \frac{2}{15} \epsilon^4 \partial_\beta^4 + O(\epsilon^6). \]

We call such expansions of Fourier multiplier operators long wave approximations. The rigorous connection between a long wave approximation and the original operator is given in the following lemma, whose simple proof is contained in section 7.

**Lemma 1.** Suppose $A$ and $A_n$ are linear operators defined by $\hat{A}f(k) = \hat{A}(k)\hat{f}(k)$ and $\hat{A_n}f(k) = \hat{A}_n(k)\hat{f}(k)$, where $\hat{A}(k)$ and $\hat{A}_n(k)$ are complex-valued functions. Also suppose that $|\hat{A}(k) - \hat{A}_n(k)| \leq C|k|^n$ (e.g., $\hat{A}_n$ is a Taylor polynomial for $\hat{A}$). Then for $f \in H^{s+n}$ we have
\[ \|Af(\cdot) - A_nf(\cdot)\|_s \leq C\|\partial_\alpha^n f(\cdot)\|_s. \]

Moreover, if $f(\alpha)$ is of long wavelength form—that is, if $f(\alpha) = F(\epsilon \alpha)$, with $F \in H^{s+n}$—then for $0 < \epsilon < 1$ there exists $C$ independent of $\epsilon$ such that
\[ \|Af(\cdot) - A_nf(\cdot)\|_s \leq C\epsilon^{n-1/2}\|F(\cdot)\|_{s+n}. \]

In [28], Schneider and Wayne show that the operator $K(x,y)$ does not depend on $x$ per se, but rather on $z = K_0x$. We confuse the notation for the operators intentionally. They showed that $K(z,y)$ has the following expansion:
\[ K(z,y)u = K_0u + K_1(z,y)u + S_2(z,y)u, \]
where
\[ K_1(z,y)u = M_1(z)\partial_\alpha u - (y + K_0yK_0)\partial_\alpha u \]
with
\[ \hat{\mathcal{F}}[M_1(z)v](k) = -\int \frac{\hat{K}_0(k) - \hat{K}_0(l)}{\hat{K}_0(k-l)}\hat{v}(k-l)\hat{v}(l)dl. \]
$S_2$ is an operator which depends quadratically on $z$ and $y$. Section 5 contains an analysis of these operators.

By using the hyperbolic trigonometric identity

$$\frac{\tanh(l) - \tanh(k)}{\tanh(l - k)} = 1 - \tanh(k) \tanh(l),$$

we can simplify the expression for $K_1$ to

$$K_1(z, y)u = M_1(z + y) \partial_\alpha u$$
$$= -(z + y + K_0(z + y)K_0) \partial_\alpha u.$$  

Since we know $K_0$ is a bounded operator, it is clear that

$$\|M_1(z)v\|_s \leq \|zv\|_s.$$  

### 4. The derivation.

In this section we will derive the higher order correction to the KdV approximation. For technical reasons, it is most convenient to work with the waterwave equation written in the form (WW3). Suppose that one is given the function $Ψ(α, t) = (Ψ^z(α, t), Ψ^y(α, t), Ψ^u(α, t))^{tr}$. The amount that this function fails to satisfy (WW3) is called the residual and is given by $Res[Ψ] = (Res^z, Res^y, Res^u)^{tr}$ with

$$Res^z = \partial_t Ψ^z - K_0Ψ^u,$$
$$Res^y = \partial_t Ψ^y - K(Ψ^z, Ψ^y)Ψ^u,$$
$$Res^u = \partial_t Ψ^u + Ψ^u \partial_\alpha Ψ^u + \frac{1 + \partial_t^2 Ψ^y}{1 + ΛΨ^z}.$$  

For a true solution, notice that $Res[Ψ]$ is identically zero.

**Remark 5.** We will also consider the four-dimensional system (WW4). If we let $Ψ^u = \partial_t Ψ^u$, then we have the additional Res function

$$Res^u = \partial_t Ψ^u + \frac{Ψ^u \partial_\alpha Ψ^u + \partial_\alpha \partial_t (1 + ΛΨ^y) + Ψ^u \partial_\alpha \partial_t Ψ^y}{1 + ΛΨ^z}$$
$$= \partial_t Res^u + \frac{Ψ^u \partial_\alpha Ψ^u}{1 + ΛΨ^z} Res^u.$$  

The main goal when deriving modulation equations is to choose a system of equations such that solutions to this system yield a very small residual. This is different from (but connected to) showing that solutions to the modulation equations are close to true solutions for the original problem. This latter issue is precisely that answered by the main results, Theorem 2 and Proposition 3, and is discussed in section 6. Here, we will perform a series of calculations on the residual and derive equations (KdV), (IW), and (LK). In this process we guarantee the smallness of the residual. While several of the steps will initially seem to have little mathematical justification (i.e., they are formal), once the calculation is completed it will be obvious that all steps are valid. For example, we will take

$$Ψ^u = K_0^{-1} \partial_t Ψ^z.$$  

With this choice

$$Res^z = 0.$$  

which is small indeed! However, \( K_0^{-1} \) is not in general a well-defined operator. Nonetheless, when we eventually select \( \Psi^z \), \( K_0^{-1} \partial \Psi^z \) will make perfect sense.

We are looking for solutions which are small in amplitude and long in wavelength. Thus we let

\[
\Psi^z(\alpha, t) = \epsilon^2 Z(\beta, \tau) = \epsilon^2 Z_1(\beta, \tau) + \epsilon^4 Z_2(\beta, \tau) + \epsilon^6 Z_3(\beta, \tau).
\]

Recall \( \beta = \epsilon \alpha \) and \( \tau = ct \). We require \( \text{Res}[\Psi] \) to be \( O(\epsilon^{17/2}) \). Loosely, we need three powers of \( \epsilon \) more than the expected error of \( O(\epsilon^{11/2}) \) to account for the long times \( (O(\epsilon^{-3})) \) over which our approximation will be valid. See Schneider and Wayne [28] and Wayne and Wright [33].

**Remark 6.** More specifically, if we wish to prove Theorem 2 in the space \( \mathcal{Y}^s \) we will need

\[
\| \text{Res}_z \|_s \leq C \epsilon^{17/2},
\]

\[
\| \text{Res}_y \|_s \leq C \epsilon^{17/2},
\]

\[
\| \text{Res}_u \|_{s-1} \leq C \epsilon^{17/2},
\]

\[
\| \text{Res}_a \|_{s-1} \leq C \epsilon^{19/2}
\]

for \( 0 \leq t \leq T_0 \epsilon^{-3} \). Given the definition of \( \text{Res}_a \), the final estimate will follow automatically from the estimate on \( \text{Res}_u \).

We have already chosen \( \Psi^u \) in terms of \( \Psi^z \). We will first use the expression for \( \text{Res}_y \) to similarly determine \( \Psi^y \) in terms of \( \Psi^z \). This is not as simple a matter because while \( K_0 \) commutes with \( \partial_t \), the full operator \( K(z, y) \) does not.

We have

\[
\text{Res}_y = -\partial_t \Psi^y + K_0 \Psi^u + M_1(\Psi^z + \Psi^y)\partial_\alpha \Psi^u + S_2(\Psi^y, \Psi^z)\Psi^u
\]

\[
= -\partial_t \Psi^y + \partial_t \Psi^z + M_1(\Psi^z + \Psi^y)K_0^{-1}\partial_\alpha \partial_t \Psi^z + S_2(\Psi^y, \Psi^z)\Psi^u.
\]

Notice that in the above expression we can cancel the linear terms by taking \( \Psi^y \sim \Psi^z \). More precisely, we set

\[
\Psi^y(\alpha, t) = \epsilon^2 Z(\beta, \tau) + \epsilon^4 \Delta_1(\beta, \tau) + \epsilon^6 \Delta_2(\beta, \tau)
\]

for as yet undetermined functions \( \Delta_i \). Thus

\[
\text{Res}_y = -\epsilon^2 \partial_t \Delta_1 - \epsilon^6 \partial_t \Delta_2 + M_1(2 \epsilon^2 Z)L_\epsilon \epsilon^3 \partial_\tau Z + M_1(\epsilon^4 \Delta_1 + \epsilon^6 \Delta_2)L_\epsilon \epsilon^3 \partial_\tau Z + S_2(\Psi^y, \Psi^z)\Psi^u
\]

\[
= -\epsilon^2 \partial_t \Delta_1 - \epsilon^6 \partial_t \Delta_2 - 2 \epsilon^6 ZL_\epsilon \partial_\tau Z - \epsilon^6 K_{0,\epsilon}(2 Z \partial_\beta \partial_\tau Z)
\]

\[
- \epsilon^7 \Delta_1 L_\epsilon \partial_\tau Z - \epsilon^8 K_{0,\epsilon}(\Delta_1 \partial_\beta \partial_\tau Z) + \epsilon^9 M_1(\Delta_2)L_\epsilon \partial_\tau Z
\]

\[
+ S_2(\Psi^y, \Psi^z)\Psi^u.
\]

A number of the terms in \( \text{Res}_y \) are already \( O(\epsilon^{17/2}) \). By Lemma 1 we have the following estimate on \( K_{0,\epsilon} \):

\[
\| K_{0,\epsilon} F \|_s \leq \epsilon^{1/2} \| F \|_{s+1}.
\]

Thus terms containing \( K_{0,\epsilon} \) can be considered to be a power of \( \epsilon \) smaller than they appear (though this costs a derivative). On the other hand, \( K_0 \) is bounded so we have

\[
\| K_{0,\epsilon} F \|_s \leq \epsilon^{-1/2} \| F \|_s.
\]
Thus we can use $K_{0,\epsilon}$ either as a bounded functional or to gain powers of $\epsilon$, but not both. Notice that $L_\epsilon$ does not contribute any additional powers of $\epsilon$ in any case. We separate out all the terms that are already sufficiently small into error terms. That is,

$$
\text{Res}_y = -\epsilon^5 \partial_\tau \Delta_1 - \epsilon^7 \partial_\tau \Delta_2 - 2\epsilon^5 Z_1 L_\epsilon \partial_\tau Z_1 \\
- 2\epsilon^7 Z_1 L_\epsilon \partial_\tau Z_2 - 2\epsilon^7 Z_3 L_\epsilon \partial_\tau Z_1 \\
- \epsilon^9 K_{0,\epsilon} (2Z_1 \partial_\beta \partial_\tau Z_1) - \epsilon^7 \Delta_1 L_\epsilon \partial_\tau Z_1 \\
+ E^y_{\text{small}} + E^y_{S_2}
$$

with

$$
E^y_{\text{small}} = 2\epsilon^9 Z_3 L_\epsilon \partial_\tau Z + 2\epsilon^9 (Z_1 + \epsilon^2 Z_2) L_\epsilon \partial_\tau Z_3 + 2\epsilon^9 Z_2 L_\epsilon \partial_\tau Z_2 \\
- \epsilon^8 K_{0,\epsilon} \left( (2Z_2 + \epsilon^2 Z_3) \partial_\beta \partial_\tau Z \right) - \epsilon^8 K_{0,\epsilon} \left( 2Z_1 \partial_\beta \partial_\tau (Z_2 + \epsilon^2 Z_3) \right) \\
- \epsilon^9 \Delta_1 L_\epsilon \partial_\tau (Z_2 + \epsilon^2 Z_3) - \epsilon^8 K_{0,\epsilon} \left( \Delta_1 \partial_\beta \partial_\tau Z \right) + \epsilon^9 M_1 (\Delta_2) L_\epsilon \partial_\tau Z
$$

$$
E^y_{S_2} = S_2 (\Psi^y, \Psi^z) \Psi^s.
$$

It is clear that $E^y_{\text{small}}$ is $O(\epsilon^{17/2})$. That is,

$$
\| E^y_{\text{small}} \|_s \leq C \epsilon^{17/2}.
$$

The constant $C$ depends on various norms of the functions $Z_i$, $\partial_\tau Z_i$, and $\Delta_i$. Specifically, chasing through the various terms in $E^y_{\text{small}}$ and applying the estimates in (9), (13), (15), and (16), one can show that $C$ depends on $\| Z_1 \|_{s+1}$, $\| Z_2 \|_{s+1}$, $\| Z_3 \|_s$, $\| \partial_\tau Z_1 \|_{s+2}$, $\| \partial_\tau Z_2 \|_{s+2}$, $\| \partial_\tau Z_3 \|_{s+1}$, $\| \Delta_1 \|_{s+1}$, and $\| \Delta_2 \|_s$.

The term $E^y_{S_2}$ is also $O(\epsilon^{17/2})$, though this is not as obvious. We prove this in Proposition 3 in section 5. The proof of this relies strongly on the fact that we have taken $\Psi^y$ and $\Psi^z$ such that $\Psi^y - \Psi^z$ is $O(\epsilon^4)$. This causes a cancellation in $S_2$, which in turn makes this term small.

We now expand $L_\epsilon$ and $K_{0,\epsilon}$ in the remaining low order terms in $\text{Res}_y$ to find

$$
\text{Res}_y = -\epsilon^5 \partial_\tau \Delta_1 - \epsilon^7 \partial_\tau \Delta_2 + 2\epsilon^5 Z_1 \partial_\tau Z_1 + \frac{4}{3} \epsilon^7 Z_1 \partial_\tau \partial_\beta^2 Z_1 \\
+ 2\epsilon^7 Z_1 \partial_\tau Z_2 + 2\epsilon^7 Z_2 \partial_\tau Z_1 \\
+ 2\epsilon^7 \partial_\beta Z_1 \partial_\beta \partial_\tau Z_1 + \epsilon^7 \Delta_1 \partial_\tau Z_1 \\
+ E^y_{\text{small}} + E^y_{S_2} + E^y_{\text{iwa}}
$$

with

$$
E^y_{\text{iwa}} = -2\epsilon^5 Z_1 \left( L_\epsilon + 1 - \frac{1}{3} \epsilon^2 \partial_\beta^2 \right) \partial_\tau Z_1 - 2\epsilon^7 Z_1 (L_\epsilon + 1) \partial_\tau Z_2 \\
- 2\epsilon^7 Z_2 (L_\epsilon + 1) \partial_\tau Z_1 - 2\epsilon^5 (K_{0,\epsilon} + \epsilon \partial_\beta)(Z_1 \partial_\beta \partial_\tau Z_1) \\
- \epsilon^7 \Delta_1 (L_\epsilon + 1) \partial_\tau Z_1.
$$

Each term in $E^y_{\text{iwa}}$ is $O(\epsilon^{17/2})$ by Lemma 1. That is,

$$
\| E^y_{\text{iwa}} \| \leq C \epsilon^{17/2},
$$

where $C$ depends on $\| Z_1 \|_{s+3}, \| Z_2 \|_s$, $\| \partial_\tau Z_1 \|_{s+4}, \| \partial_\tau Z_2 \|_{s+2}$, and $\| \Delta_1 \|_s$. (The subscript “iwa” stands for “long wave approximation.”)
The only \( O(\epsilon^5) \) terms remaining in \( \text{Res}_y \) are

\[
-\epsilon^5 \partial_\tau \Delta_1 + 2\epsilon^5 Z_1 \partial_\tau Z_1
\]

which we remove by selecting

\[
(17) \quad \Delta_1 = Z_1^2.
\]

Therefore

\[
\text{Res}_y = -\epsilon^7 \partial_\tau \Delta_2 + \frac{4}{3} \epsilon^7 Z_1 \partial_\tau \partial_\beta^2 Z_1 + 2\epsilon^7 Z_1 \partial_\tau Z_2 + 2\epsilon^7 Z_2 \partial_\tau Z_1 + 2\epsilon^7 \partial_\beta Z_1 \partial_\beta \partial_\tau Z_1 + \epsilon^7 Z_1^2 \partial_\tau Z_1 + E^y_{\text{small}} + E^y_{S_2} + E^y_{\text{iwa}}.
\]

The remaining \( O(\epsilon^7) \) terms in \( \text{Res}_y \) are all perfect time derivatives with the exception of

\[
\frac{4}{3} Z_1 \partial_\tau \partial_\beta^2 Z_1.
\]

Notice, however, that

\[
\partial_\tau \left( \frac{4}{3} Z_1 \partial_\beta^2 Z_1 - \frac{2}{3} (\partial_\tau Z_1)^2 \right) = \frac{4}{3} Z_1 \partial_\tau \partial_\beta Z_1 + \frac{4}{3} \partial_\tau Z_2 \left( \partial_\beta^2 Z_1 - \partial_\tau^2 Z_1 \right).
\]

Given the form of the approximation in (1), it is not unreasonable to suspect that

\[
(18) \quad \partial_\beta^2 Z_1 - \partial_\tau^2 Z_1 \sim O(\epsilon^2).
\]

We are now in a position to select \( \Delta_2 \). Taking

\[
(19) \quad \Delta_2 = (\partial_\beta Z_1)^2 + 2Z_1 Z_2 + \frac{1}{3} Z_1^3 + \frac{4}{3} Z_1 \partial_\beta^3 Z_1 - \frac{2}{3} (\partial_\tau Z_1)^2
\]

gives

\[
\text{Res}_y = E^y_{\text{small}} + E^y_{S_2} + E^y_{\text{iwa}} + E^y_{\text{switch}},
\]

where

\[
E^y_{\text{switch}} = \frac{4}{3} \epsilon^7 \partial_\tau Z_1 \left( \partial_\beta^2 Z_1 - \partial_\tau^2 Z_1 \right).
\]

Given that our assumption (18) is valid, we have shown, with our choices for \( \Psi^y \) and \( \Psi^u \) in terms of \( \Psi^z \), that \( \text{Res}_y = O(\epsilon^{17/2}) \). More specifically, we have shown that if

\[
\| \partial_\beta^2 Z_1 - \partial_\tau^2 Z_1 \|_s \leq C \epsilon^{3/2},
\]

then

\[
\text{Res}_y \leq C \epsilon^{17/2},
\]

where the constant depends only on \( \| Z_1 \|_{s+3}, \| Z_2 \|_{s+4}, \| Z_3 \|_s, \| \partial_\tau Z_1 \|_{s+4}, \| \partial_\tau Z_2 \|_{s+2}, \) and \( \| \partial_\tau Z_3 \|_{s+1} \).
Now that we have computed $\Psi^y$ and $\Psi^u$ in terms of $\Psi^z$, we now turn our attention to determining $\Psi^z$ by examining $\text{Res}_u$:

$$\text{Res}_u = K_0^{-1} \partial_t^2 \Psi^z + \partial_\alpha^2 \Psi^y \frac{1 + \partial_t^2 \Psi^y}{1 + L \Psi^z}.$$  

We expand $(1 + L \Psi^z)^{-1}$ by the geometric series to find

$$\text{Res}_u = K_0^{-1} \partial_t^2 \Psi^z + \partial_\alpha \Psi^y (1 + \partial_t^2 \Psi^y)(1 - L \Psi^z + (L \Psi^z)^2) + E_u^0,$$

where

$$E_{\text{geo}}^u = \partial_\alpha \Psi^y \frac{1 + \partial_t^2 \Psi^y}{1 + L \Psi^z} - \partial_\alpha \Psi^y (1 + \partial_t^2 \Psi^y)(1 - L \Psi^z + (L \Psi^z)^2).$$

Since $\Psi^z$ is “small” this error term can be shown to be $O(\epsilon^{17/2})$. We have the following lemma.

**Lemma 2.** Let $f \in H^{s+1}$, $s > 0$. Take $\epsilon_0$ such that $\epsilon_0^2 \|f\|_{L_\infty} \leq 1/2$. Then for $0 < \epsilon < \epsilon_0$ we have that the function

$$g(\epsilon x) = \frac{1}{1 + \epsilon^2 f(\epsilon x)} - 1 + \epsilon^2 f(\epsilon x)$$

satisfies $\|g(\epsilon \cdot)\|_s \leq C \epsilon^{7/2}$ for $C$ independent of $\epsilon$.

**Remark 7.** Under the same hypotheses as in Lemma 2, arguments similar to the proof of that lemma show that

- $(1 + \epsilon^2 f(\epsilon x))^{-1} \in C^s$ and is bounded there independent of $\epsilon$,
- $(1 + \epsilon^2 f(\epsilon x))^{-1} - 1 \in H^s$ and has norm there bounded by $C \epsilon^{3/2}$ for $C$ independent of $\epsilon$, and
- $(1 + \epsilon^2 f(\epsilon x))^{-1} - 1 + \epsilon^2 f(\epsilon x) - \epsilon^4 f^2(\epsilon x) \in H^s$ and has norm there bounded by $C \epsilon^{11/2}$ for $C$ independent of $\epsilon$.

Now, after substituting from the definitions of $\Psi^y$ and $\Psi^z$, we collect all the terms which are smaller than $O(\epsilon^{17/2})$ and find

$$\text{Res}_u = \epsilon^4 K_0^{-1} \partial_t^2 (Z_1 + \epsilon^2 Z_2 + \epsilon^4 Z_3) + \epsilon^3 \partial_\beta Z_1 - \epsilon^3 \partial_\beta Z_1 L \epsilon Z_1$$

$$+ \epsilon^5 \partial_\beta (Z_2 + Z_4^2) + \epsilon^7 \partial_\beta Z_1 \left( \partial_\beta^2 Z_1 - L \epsilon Z_2 + (L \epsilon Z_1)^2 \right) - \epsilon^7 \partial_\beta (Z_2 + Z_4^2) L \epsilon Z_1 + \epsilon^7 \partial_\beta Z_3$$

$$+ \epsilon^7 \partial_\beta \left( (\partial_\beta Z_1)^2 + 2Z_1 Z_2 + \frac{1}{3} Z_1^3 \right) + \epsilon^7 \partial_\beta \left( \frac{4}{3} Z_1 \partial_\beta^2 Z_1 - \frac{2}{3}(\partial_\beta Z_1)^2 \right) + E_{\text{geo}}^u + E_{\text{small}}^u + E_{\text{switch}}^u.$$  

We omit the exact expression for $E_{\text{small}}^u$ because it is both lengthy and uninteresting. We have

$$\|E_{\text{small}}^u\|_s \leq C \epsilon^{17/2},$$

where the constant $C$ depends on $H^{s+1}$ norms of the functions $Z_i$ and the $H^s$ norms of $\partial_\beta^2 Z_i$. We have also replaced one instance of $\partial_\beta^2 Z_1$ with $\partial_\beta^2 Z_1$ (much as we did earlier); thus

$$E_{\text{switch}}^u = \epsilon^7 \partial_\beta Z_1 \left( \partial_\beta^2 Z_1 - \partial_\beta^2 Z_1 \right).$$
Corrections to the KdV Approximation

We now define new functions $W_i$ by $Z_i = L_i^{-1}W_i$. This seemingly mysterious (and sudden!) change of variables will seem less so if we remind the reader that at the end of the day we wish to model not $z$ but rather the function $x_\alpha$. Accordingly, if we approximate $x_\alpha$ by a function

$$\Psi^d = \epsilon^2 W_1 + \epsilon^4 W_2 + \epsilon^6 W_3,$$

then it is logical to take

$$\Psi^z = L^{-1}\Psi^d,$$

and in the long wavelength limit we arrive at these functions $W_i$. Therefore we have

$$\text{Res}_u = \epsilon^3 \partial_\beta^{-1} \partial_\chi^2 (W_1 + \epsilon^2 W_2 + \epsilon^4 W_3) + \epsilon^2 K_{0,e} W_1 - \epsilon^4 W_1 K_{0,e} W_1$$

$$+ \epsilon^4 K_{0,e} W_2 + \epsilon^5 \partial_\beta (L^{-1}_e W_1)^2$$

$$+ \epsilon^2 K_{0,e} W_1 \left( \epsilon^3 K_{0,e} \partial_\beta W_1 - \epsilon^4 W_2 + \epsilon^2 W_1^2 \right)$$

$$- \left( \epsilon^6 K_{0,e} W_2 + \epsilon^7 \partial_\beta (L^{-1}_e W_1)^2 \right) W_1 + \epsilon^6 K_{0,e} W_3$$

$$+ \epsilon^7 \partial_\beta \left( (K_{0,e} W_1)^2 + 2\epsilon^2 L^{-1}_e W_1 L^{-1}_e W_2 + \frac{1}{3} \epsilon^2 (L^{-1}_e W_1)^3 \right)$$

$$+ \epsilon^7 \partial_\beta \left( \frac{4}{3} L^{-1}_e W_1 L^{-1}_e \partial_\beta^2 W_1 - \frac{2}{3} (L^{-1}_e \partial_\chi W_1)^2 \right)$$

$$+ E_{\text{geo}}^u + E_{\text{small}}^u + E_{\text{switch}}^u.$$

At this time, the presence of inverse $\beta$ derivatives may seem problematic. Notice that each such inverse derivative precedes a time derivative. Once we select the functions $W_i$ we will see that there can be an exchange between time and space derivatives, which will justify the instances of $\partial_\beta^{-1}$.

Now we replace $K_{0,e}$ and $L^{-1}_e$ with their long wave approximates and find

$$\text{Res}_u = \epsilon^3 \partial_\beta^{-1} \partial_\chi^2 (W_1 + \epsilon^2 W_2 + \epsilon^4 W_3)$$

$$- \epsilon^2 \left( \epsilon \partial_\beta + \frac{1}{3} \epsilon^3 \partial^3_\beta + \frac{2}{15} \epsilon^5 \partial^5_\beta \right) W_1$$

$$+ \epsilon^4 W_1 \left( \epsilon \partial_\beta + \frac{1}{3} \epsilon^3 \partial^3_\beta \right) W_1$$

$$- \epsilon^4 \left( \epsilon \partial_\beta + \frac{1}{3} \epsilon^3 \partial^3_\beta \right) W_2 + \epsilon^5 \partial_\beta (W_1)^2 + \frac{2}{3} \epsilon^7 \partial_\beta (W_1 \partial_\beta^2 W_1)$$

$$- \epsilon^7 \partial_\beta W_1 \left( - \partial_\beta^2 W_1 - W_2 + W_1^2 \right)$$

$$- \epsilon^7 W_1 \left( - \partial_\beta W_2 + \partial_\beta (W_1)^2 \right) - \epsilon^7 \partial_\beta W_3$$

$$+ \epsilon^7 \partial_\beta \left( (\partial_\beta W_1)^2 + 2W_1 W_2 - \frac{1}{3} W_1^3 \right)$$

$$+ \epsilon^7 \partial_\beta \left( \frac{4}{3} W_1 \partial_\beta^2 W_1 - \frac{2}{3} (\partial_\chi W_1)^2 \right)$$

$$+ E_{\text{geo}}^u + E_{\text{small}}^u + E_{\text{switch}}^u + E_{\text{lwa}}^u.$$

The error made by the long wave approximations is denoted $E_{\text{lwa}}^u$. By Lemma 1 we have

$$\|E_{\text{lwa}}^u\| s \leq C\epsilon^{17/2},$$
where $C$ depends on $\|W_1\|_{s+\tau}$, $\|W_2\|_{s+5}$, $\|W_3\|_{s+3}$, and $\|\partial_x W_1\|_{s+2}$.

Now we organize the above as

$$\text{Res}_y = \epsilon^3 \partial^3_\beta \partial_\tau^2 W_1 - \epsilon^3 \partial_\beta W_1$$

$$- \epsilon^3 \frac{1}{3} \partial^3_\beta W_1 + \epsilon^3 \frac{3}{2} \partial_\beta (W_1)^2$$

$$+ \epsilon^5 \partial^{-1}_\beta \partial^2_\tau W_2 - \epsilon^5 \partial_\beta W_2$$

$$- \epsilon^7 \frac{1}{3} \partial^3_\beta W_2 + \epsilon^7 3 \partial_\beta (W_1 W_2)$$

$$- \frac{2}{15} \epsilon^7 \partial^3_\beta W_1 + \epsilon^7 \frac{1}{3} W_1 \partial^3_\beta W_1$$

$$+ \epsilon^7 2 \partial_\beta (W_1 \partial^3_\beta W_1) + \epsilon^7 \frac{3}{2} \partial_\beta (\partial_\beta W_1)^2$$

$$- \epsilon^7 \frac{4}{3} \partial_\beta (W_1)^3 - \epsilon^7 \frac{2}{3} \partial_\beta (\partial_\tau W_1)^2$$

$$+ \epsilon^7 \partial^{-1}_\beta \partial^2_\tau W_3 - \epsilon^7 \partial_\beta W_3$$

$$+ E_{\text{geo}} + E_{\text{small}} + E_{\text{switch}} + E_{\text{lwa}}.$$  

The term on the first line of the right-hand side looks formally like an inverse derivative of a wave equation:

$$\partial^{-1}_\beta \partial^2_\beta W_1 - \partial_\beta W_1 = \partial^{-1}_\beta \left( \partial^2_\beta W_1 - \partial^3_\beta W_1 \right).$$

We cancel this term (to lowest order) by taking $W_1$ of the form

$$W_1(\beta, \tau) = -U(\beta - \tau, \epsilon^2 \tau) - V(\beta + \tau, \epsilon^2 \tau).$$

Recall $\beta_{\pm} = \beta \pm \tau$ and $T = \epsilon^2 \tau$. The “minus” signs may seem arbitrary, but are included at this stage so that they agree with previous work in the area. Noting that the third line looks very much like the first, we also set

$$W_2(\beta, \tau) = -F(\beta - \tau, \epsilon^2 \tau) - G(\beta + \tau, \epsilon^2 \tau) - P(\beta, \tau).$$

These choices for $W_1$ and $W_2$ are precisely those described heuristically in the introduction.

The first three lines in (21) become

$$\epsilon^5 \left( 2 \partial_\tau U + \frac{1}{3} \partial^3_\beta U + \frac{3}{2} \partial_\beta U^2 \right)$$

$$+ \epsilon^5 \left( -2 \partial_\tau V + \frac{1}{3} \partial^3_\beta V + \frac{3}{2} \partial_\beta V^2 \right)$$

$$+ \epsilon^5 \left( \partial^{-1}_\beta (\partial^2_\beta P - \partial^2_\beta P) + 3 \partial_\beta (UV) \right)$$

$$+ \epsilon^7 \left( 2 \partial_\tau F - 2 \partial_\tau G - \partial^{-1}_\beta (\partial^2_\beta U + \partial^2_\beta V) \right)$$

$$- \epsilon^9 \partial^{-1}_\beta \left( \partial^2_\beta F + \partial^2_\beta G \right).$$

We cancel everything multiplied by $\epsilon^5$ by taking

$$-2 \partial_\tau U = \frac{1}{3} \partial^3_\beta U + \frac{3}{2} \partial_\beta U^2,$$

$$2 \partial_\tau V = \frac{1}{3} \partial^3_\beta V + \frac{3}{2} \partial_\beta V^2,$$

$$\partial^2_\beta P - \partial^2_\beta P = 3 \partial^2_\beta (UV),$$
which are precisely equations (KdV) and (IW). By Proposition 1 we know that the solutions to these equations are well behaved over the long time scales.

Given that the functions \( U \) and \( V \) have been chosen to solve (KdV), one computes that

\[
\begin{align*}
\partial^2_T U &= \partial_{\beta_-} \left( \frac{1}{36} \partial^5_{\beta_-} U + \frac{9}{4} U^2 \partial_{\beta_-} U + \frac{1}{2} U \partial^3_{\beta_-} U + \frac{3}{4} \partial_{\beta_-} U \partial^2_{\beta_-} U \right), \\
\partial^2_T V &= \partial_{\beta_+} \left( \frac{1}{36} \partial^5_{\beta_+} V + \frac{9}{4} V^2 \partial_{\beta_+} V + \frac{1}{2} V \partial^3_{\beta_+} V + \frac{3}{4} \partial_{\beta_+} V \partial^2_{\beta_+} V \right).
\end{align*}
\]

Thus the term \( \partial^{-1}_{\beta} (\partial^2_T U + \partial^2_T V) \) is perfectly well defined. For brevity, we will continue to write these terms with the inverse derivatives instead of in the longer form above.

Moreover, now we can put more precise estimates on \( E_{\text{geo}}^u \) and \( E_{\text{switch}}^u \). In particular, since each time derivative for solutions to KdV equations counts for three space derivatives, we have

\[
\| \partial^2_T W_1 - \partial^2_T W_1 \|_s \leq C \epsilon^{3/2},
\]

where \( C \) depends on \( \| W_1 \|_{s+6} \).

Recall from Fact 1 and Proposition 1 that solutions to (IW) can be rewritten as

\[
\begin{align*}
P(\beta, \tau) &= P^+(\beta, \tau) + P^-(\beta, \tau) \\
&= \varphi^+(\beta_+, T) + \varphi^-(\beta_-, T).
\end{align*}
\]

The functions \( \varphi^\pm \) are rapidly decaying. We make this decomposition so that every remaining term in (21)

- will be a unidirectional term which is rapidly decaying;
- will be a product of two such terms which are moving in opposite directions;
- or will include a derivative of \( W_3 \).

That is,

\[
\begin{align*}
\text{Res}_y &= \epsilon^7 \left( 2 \partial_T F + \frac{1}{3} \partial^3_{\beta_-} F + 3 \partial_{\beta_-} (UF) + J^- \right) \\
&\quad + \epsilon^7 \left( -2 \partial_T G + \frac{1}{3} \partial^3_{\beta_+} G + 3 \partial_{\beta_+} (VG) + J^+ \right) \\
&\quad + \epsilon^7 \left( \partial^{-1}_{\beta} (\partial^2_T W_3 - \partial^2_T W_3) + J^p \right) \\
&\quad - \epsilon^9 \partial^{-1}_{\beta} (\partial^2_T F + \partial^2_T G) \\
&\quad + E_{\text{geo}}^u + E_{\text{small}}^u + E_{\text{switch}}^u + E_{\text{lwa}}^u + E_{\text{time}}^u.
\end{align*}
\]
where

\[ J^- = 3\partial_{\beta_-}(U\varphi^-) + 4U^2\partial_{\beta_-}U + \frac{7}{3}U\partial^3_{\beta_-}U \]
\[ + \frac{11}{3}\partial_{\beta_-}U\partial^2_{\beta_-}U + \frac{2}{15}\partial^5_{\beta_-}U + \frac{1}{3}\partial^3_{\beta_-}\varphi^- \]
\[ - \partial^1_{\beta_-}\partial^2_{T}U, \]

\[ J^+ = 3\partial_{\beta_+}(V\varphi^+) + 4V^2\partial_{\beta_+}V + \frac{7}{3}V\partial^3_{\beta_+}V \]
\[ + \frac{11}{3}\partial_{\beta_+}V\partial^2_{\beta_+}V + \frac{2}{15}\partial^5_{\beta_+}V + \frac{1}{3}\partial^3_{\beta_+}\varphi^+ \]
\[ - \partial^1_{\beta_+}\partial^2_{T}V, \]

(22)

\[ J^s = \partial_{\beta} \left( U \left( 3G + 3\varphi^+ + 4V^2 + \frac{7}{3}\partial^3_{\beta_+}V \right) \right) \]
\[ + \partial_{\beta} \left( V \left( 3F + 3\varphi^- + 4U^2 + \frac{7}{3}\partial^3_{\beta_-}U \right) \right) \]
\[ + 4\partial_{\beta} \left( \partial_{\beta_-}U\partial_{\beta_+}V \right), \]

and

\[ E^u_{time} = \epsilon^4 \frac{4}{3} \partial_{\beta} \left( (\partial_{\beta_-}U - \partial_{\beta_+}V) (\partial_T U + \partial_T V) \right) \]
\[ - \epsilon^{11} \frac{2}{3} \partial_{\beta} \left( (\partial_T U + \partial_T V)^2 \right). \]

Notice that \( J^\pm = J^\pm(\beta^\pm, T). \) \( E^u_{time} \) (so called because each term in it contains some sort of time derivative) is clearly \( O(\epsilon^{17/2}). \) That is,

\[ ||E^u_{time}||_s \leq C\epsilon^{17/2}. \]

The constant above depends on \( ||U||_{s+4} \) and \( ||V||_{s+4}. \)

The term \( \epsilon^2\partial_{\beta}^{-1} (\partial^2_T F + \partial^2_T G) \) is not included in \( E^u_{time} \) for the following reason. In a moment, when we select the equations \( F \) and \( G \) solve, a consequence will be that there will be terms in \( \partial^2_T F \) and \( \partial^2_T G \) which are \( O(\epsilon^{-2}). \)

By taking

\[ -2\partial_T F = \frac{1}{3}\partial_{\beta_-}(UF) + \frac{3}{2}\partial^3_{\beta_-}F + J^-, \]
\[ 2\partial_T G = \frac{1}{3}\partial_{\beta_+}(VG) + \frac{3}{2}\partial^3_{\beta_+}G + J^+, \]

we cancel nearly all the terms which are not in the various \( E^u \) terms. These are the linearized KdV equations (LK) discussed in the introduction. Proposition 1 guarantees that the solutions are well behaved. We are left with

\[ \text{Res}_y = + \epsilon^7 \left( \partial_{\beta}^{-1} \left( \partial^2_T W_3 - \partial^2_T W_3 \right) + J^s \right) \]
\[ - \epsilon^9 \partial_{\beta}^{-1} \left( \partial^2_T F + \partial^2_T G \right) \]
\[ + E^u_{geo} + E^u_{small} + E^u_{switch} + E^u_{lwa} + E^u_{time}. \]
Now we consider the terms in $\partial^{-1}_\beta (\partial^2 T + \partial^2 G)$. Notice that
\[
-\partial^{-1}_\beta \partial^2 T = \partial^{-1}_\beta \partial T \left( \frac{1}{6} \partial^3_\beta F + \frac{3}{2} \partial_\beta (UF) + \frac{1}{2} J^- \right)
= \frac{1}{6} \partial^3_\beta \partial T F + \frac{3}{2} \partial T (UF) + \frac{1}{2} \partial^{-1}_\beta \partial T J^-.
\]

$J^-$ contains the term $3 \partial_\beta (U \varphi^-) + \frac{1}{3} \partial^3_\beta \varphi^-$. From the definition of $\varphi^-$ we know that
\[
\partial T \varphi^- = -\varepsilon^2 \frac{3}{2} \partial_\beta (UV).
\]
Thus we have
\[
\frac{1}{2} \partial^{-1}_\beta \partial T \left( 3 \partial_\beta (U \varphi^-) + \frac{1}{3} \partial^3_\beta \varphi^- \right) = \frac{3}{2} \varphi^- \partial T U + \frac{3}{2} U \partial T \varphi^- + \frac{1}{6} \partial^3_\beta \partial T \varphi^-
= \frac{3}{2} \varphi^- \partial T U - \frac{9}{4} \varepsilon^2 U \partial_\beta (UV) + \frac{3}{2} \partial^{-1}_\beta (UV).
\]

We treat $\partial^2 G$ in the same fashion. Therefore we can write
\[
-\varepsilon^2 \partial^{-1}_\beta (\partial^2 F + \partial^2 G) = E^u_{F,G} - \varepsilon^7 \left( \frac{9}{4} (U + V) \partial_\beta (UV) + \frac{1}{2} \partial^3_\beta (UV) \right).
\]

By construction $E^u_{F,G}$ satisfies the estimate
\[
\|E^u_{F,G}\|_s \leq C \varepsilon^{7/2},
\]
where $C$ depends on $\|U\|_{s+7}$, $\|V\|_{s+7}$, $\|F\|_{s+5}$, and $\|G\|_{s+5}$.

We have
\[
\text{Res}_y = \varepsilon^7 \left( \partial^{-1}_\beta (\partial^2 W_3 - \partial^2_\beta W_3) + J^s \right)
- \varepsilon^7 \left( \frac{9}{4} (U + V) \partial_\beta (UV) + \frac{1}{2} \partial^3_\beta (UV) \right)
+ E^u_{geo} + E^u_{small} + E^u_{switch} + E^u_{lw} + E^u_{time} + E^u_{F,G}.
\]

By selecting
\[
\partial^2 W_3 - \partial^2_\beta W_3 = \partial_\beta \left( \frac{9}{4} (U + V) \partial_\beta (UV) + \frac{1}{2} \partial^3_\beta (UV) - J^s \right)
\]
the gun goes off and we cancel all remaining $O(\varepsilon^7)$ terms. Thus
\[
\text{Res}_y = E^u_{geo} + E^u_{small} + E^u_{switch} + E^u_{lw} + E^u_{time} + E^u_{F,G}.
\]
Each of the $E^u$ is $O(\varepsilon^{17/2})$.

Unlike the previous equations (KdV), (IW), and (LK), Proposition 1 does not tell us that the solutions to (23) are controllable. Nonetheless, (23) is an inhomogeneous wave equation where the inhomogeneity consists entirely of terms which are products of left- and right-moving rapidly decaying functions. From Wayne and Wright [33] we have the following lemma.
Lemma 3. Suppose
\[ \partial_{\tau} u \pm \partial_{\beta} u = l(\beta + \tau, \epsilon^2 \tau) r(\beta - \tau, \epsilon^2 \tau), \quad u(X, 0) = 0, \]
with \( \| l(\cdot, T) \| _{H^s(\mathbb{T})} \leq C \) and \( \| r(\cdot, T) \| _{H^s(\mathbb{T})} \leq C \) for \( T \in [0, T_0] \); then
\[ \| u(\beta, \tau) \| _s \leq C \]
for all \( \tau \in [0, T_0 \epsilon^{-2}] \). The constant C is uniform in \( \epsilon \).

Thus \( W_3 \) will remain \( O(1) \).

Remark 8. If we are in the situation in which Proposition 1 applies, we see that least regular parts in the driving term are \( \partial_{\beta}^2 (UG) \) and \( \partial_{\beta}^2 (VF) \), which are in \( H^s(\mathbb{T}) \). Thus, by this lemma we have that \( W_3 \in H^{s-5} \) for all times of interest.

At this time we have derived the modulation equations and shown that the residual is small. The only remaining order of business in this section is to determine how smooth the solutions to our modulation equations need to be in order for \( \text{Res}[\Psi] \) to be appropriately regular. This may seem to be a fairly tiresome task, but fortunately the least regular terms in all of the sundry \( E \) functions come from only one term—\( E_{\text{small}}^u \). This is because \( E_{\text{small}}^u \) contains many time derivatives.

We need to control \( \text{Res}_u \) in \( H^{s+1} \). For this we need \( \partial_t \text{Res}_u \in H^{s-1} \), which in turn implies that we must have \( \partial_t E_{\text{small}}^u \in H^{s-1} \). Recalling (20), we see that this will require \( \partial_{\beta}^2 Z_2 \in H^{s+1} \), or rather (since \( L_{\epsilon^{-1}} \) saves a derivative) \( \partial_{\beta}^3 W_2 \in H^{s+2} \). For this, we need \( \partial_{\beta} W \) and \( \partial_{\beta} G \) in \( H^{s-2} \). Given that \( F \) solves (LK) where \( J^+ \) contains the terms \( \partial_{\beta}^2 U, \partial_{\beta} \varphi^- \), one sees that \( \partial_{\beta} F \) will include the terms \( \partial_{\beta}^3 F, \partial_{\beta}^2 U, \) and \( \partial_{\beta}^3 \varphi^- \). Thus \( \| \partial_{\beta}^3 F \| _{s+2} \) is controlled by the \( H^{s+9} \) norms of \( U \) and \( V \), and the \( H^{s+7} \) norms of \( \varphi^- \), \( F \), and \( G \). The analogous result is true for \( \partial_{\beta} G \). We also need \( \partial_{\beta} W_3 \in H^{s-2} \). Since \( W_3 \) solves (23), we require \( W_3 \in H^{s+1} \).

In summary we have the following proposition.

Proposition 2. Take \( \Psi^d \) as in (7), with \( U, V, F, G, P, \) and \( W_3 \) solving their respective equations. Let
\[ \Psi^z = L^{-1} \Psi^d, \]
\[ \Psi^y = \Psi^z + \epsilon^4 \Delta_1 + \epsilon^6 \Delta_2, \]
\[ \Psi^u = \partial_{\alpha}^{-1} \partial_t \Psi^d, \]
\[ \Psi^a = \partial_{\alpha}^{-1} \partial_{\beta}^2 \Psi^d, \]
with \( \Delta_1 \) and \( \Delta_2 \) as in (17) and (19), and form \( \text{Res}[\Psi] \) as in (4). Then
\[ \| \text{Res}_z \| _s \leq C \epsilon^{17/2}, \]
\[ \| \text{Res}_y \| _s \leq C \epsilon^{17/2}, \]
\[ \| \text{Res}_u \| _{s-1} \leq C \epsilon^{17/2}, \]
\[ \| \text{Res}_a \| _{s-1} \leq C \epsilon^{19/2}, \]
where \( C \) is a constant which depends on \( \| U \| _{s+9}, \| V \| _{s+9}, \| P \| _{s+7}, \| F \| _{s+7}, \| G \| _{s+7}, \) and \( \| W_3 \| _{s+1} \). The estimate (6) holds as long as these quantities remain bounded. The constant \( C \) does not depend on \( \epsilon \).

In light of Proposition 1 and Remark 8, we see, if we take the initial conditions for \( U \) and \( V \) to satisfy (2) with \( s \geq s + 7 \), that \( \| U \| _{s+9}, \| V \| _{s+9}, \| P \| _{s+7}, \| F \| _{s+7}, \| G \| _{s+7}, \) and \( \| W_3 \| _{s+1} \) are all \( O(1) \) for \( t \in [0, T_0 \epsilon^{-3}] \). Thus we move on.
5. The operator $K(x,y)$. Part II: Estimates and extensions. In this section we will describe a few more estimates related to $K(x,y)$. All such estimates are either smoothing estimates or ones which show that certain terms are small in the long wavelength setting.

First, since $1 + \hat{K}_0^2(k)$ goes to zero exponentially fast as $|k| \to \infty$, the operator $1 + K_0^2$ is smoothing. That is, for all $s \geq 0$

$$\| (1 + K_0^2)u \|_s \leq C \| u \|_{L^2}.$$  

Also, commutators involving $K_0$ are smoothing. We quote the following lemma from [28].

**Lemma 4.** Let $r \geq 0$, $q > 1/2$, and $0 \leq p \leq q$. Then there exists a $C > 0$ such that

$$\| [f, K_0]g \|_r \leq C \| f \|_{r+p} \| g \|_{q-p}.$$  

*Proof.* See Lemma 3.12 on page 1498 of [28].

Schneider and Wayne show that $K_1(z,y)$ is a smoothing operator.

**Lemma 5.** For $r \geq 0$, $q \geq 1/2$ and $0 \leq p \leq q$, there is $C$ such that

$$\| K_1(z,y)u \|_r \leq C (\| z \|_{r+p} + \| y \|_{r+p}) \| u \|_{q-p}.$$  

*Proof.* See Corollary 3.13 on page 1499 of [28].

If we let $S_1(z,y) = K(z,y) - K_0$, we also have the following estimates from [28].

**Lemma 6.** Fix $s \geq 4$. If the free surface is sufficiently smooth, then for $j = 1, 2$ we have

(a) $\| S_j(z,y)u \|_s \leq C (\| z \|_s + \| y \|_s) \| u \|_3$,

that is, $S_j$ is a smoothing operator;

(b) $\| \partial_\alpha (S_j(z,y)u) \|_s \leq C (\| z \|_s^s + \| y \|_s^s) \| u \|_3$,

that is, $\partial_\alpha S_j$ is a smoothing operator;

(c) $\| [\partial_t, S_j]u \| \leq C (\| z \|_s + \| y \|_s) \| u \|_3$,

that is, $[\partial_t, S_j]$ is a smoothing operator and this operator can be bounded independently of $\partial_t u$; and

(d) $\| [\partial_t^2, S_j]u \| \leq C (\| z \|_s + \| y \|_s) (\| u \|_4 + \| \partial_t u \|_4),$

that is, $[\partial_t^2, S_j]$ is a smoothing operator and this operator can be bounded independently of $\partial_t^2 u$.

*Proof.* The proof is in [28]; see Lemmas 3.14 and 3.15 and Corollary 3.16 on pages 1500, 1506, and 1507, respectively.

We will also need the following propositions concerning the behavior of the remainder terms $S_1$ and $S_2$. The first of these says that more or less the remainder $S_2$ is negligible for the sort of scalings we are considering. That is to say, the term $E_{S_2}^\epsilon$ in section 4 is very small.

**Proposition 3.** Fix $s > 5/2$. Suppose $z = \epsilon^2 Z(\epsilon\alpha)$, $y = \epsilon^2 Y(\epsilon\alpha)$, and $f = \epsilon^2 F(\epsilon\alpha)$, with $Z, Y, F \in H^{s+1}(2)$. Moreover, assume $z - y = \epsilon^4 \Delta(\epsilon\alpha)$ with
\( \Delta \in H^{s+1}(2) \). Then there exists \( \epsilon_0 \) such that for \( \epsilon \in [0, \epsilon_0] \) there is a constant \( C \) independent of \( \epsilon \) such that

\[
\|S_2(z, y)f\|_s \leq C\epsilon^{17/2}.
\]

The second proposition is a technical version of the mean value theorem as applied to the operator \( S_1 \).

**Proposition 4.** Suppose \( z(\alpha, t) = \epsilon^2 Z(\epsilon(\alpha \pm t), \epsilon^3 t), y(\alpha, t) = \epsilon^2 Y(\epsilon(\alpha \pm t), \epsilon^3 t), u(\alpha, t) = \epsilon^2 U(\epsilon(\alpha \pm t), \epsilon^3 t) \), and \( f(\alpha, t) = \epsilon^2 F(\epsilon(\alpha \pm t), \epsilon^3 t) \) with \( Z, Y, U, F \in H^s(2) \) for \( t \in [0, T_0\epsilon^{-3}] \). Also suppose \( R^2(\alpha, t), R^0(\alpha, t), \) and \( R^s(\alpha, t) \in H^s \) for the same time interval. Then

\[
\|S_1(z(\cdot) + \epsilon^{1/2} R^z(\cdot), y(\cdot) + \epsilon^{1/2} R^y(\cdot))f(\cdot) - S_1(z(\cdot), y(\cdot))f(\cdot)\|_s \leq C\epsilon^{17/2}
\]

for \( t \in [0, T_0\epsilon^{-3}] \).

**Proof of Proposition 3.** First, notice that \( x(\alpha) = \int_0^\alpha Lz(\alpha)da = \epsilon X(\epsilon\alpha) \). We know that \( X \) is in \( L^\infty \) by the following lemma.

**Lemma 7.** Suppose \( f(\alpha) = \epsilon^2 F(\epsilon\alpha) \) with \( F \in H^s(2) \). Then for all \( \alpha \)

\[
\left| \int_0^\alpha f(\alpha)da \right| \leq C\epsilon\|F\|_{H^s(2)}.
\]

**Proof.** See section 7. \( \square \)

Let \( \Phi(\tilde{x}, y) = (\Phi_1, \Phi_2) \) be the analytic map which takes \( \Omega(t) \) to

\[ P^- = \{ (\xi, \gamma) | \gamma \in [-1, 0] \}. \]

That such a map exists and is analytic is guaranteed by the Riemann mapping theorem. Let

\[ h(\alpha) = \Phi_1(\tilde{x}(\alpha), y(\alpha)) \]

and \( Qf = f \circ h \). From [28], we know that

(25) \( K(x, y)f(\alpha) = Q \circ K_0 \circ Q^{-1}f(\alpha) \).

We can derive a very useful implicit formula for \( h^{-1} \) as follows. The function \( \Phi^{-1}(\xi, \gamma) \) is analytic on \( P^- \); thus it satisfies the Cauchy–Riemann equations. If we set

\[ \Phi^{-1}(\xi, \gamma) = (\xi + u_1(\xi, \gamma)) + i(\gamma - v_1(\xi, \gamma)) \]

and notice that \( \Phi^{-1} \) sends the bottom and top of \( P^- \) to the bottom and top of \( \Omega(t) \), respectively, we see that we have the following system:

\[
\begin{align*}
\partial_\xi u_1 + \partial_\gamma v_1 &= 0, \\
\partial_\gamma u_1 - \partial_\xi v_1 &= 0, \\
v_1(\xi, -1) &= 0, \quad v_1(\xi, 0) = \eta(\xi),
\end{align*}
\]

where \( \eta(\xi) = y(h^{-1}(\xi)) \). One can solve this system with the use of Fourier transforms relatively simply. If one does so, one finds that

\[
u_1(\xi, 0) = -\int_0^\xi \eta(\xi_1)d\xi_1 - M\eta(\xi), \]

\( M \) being some constant independent of \( \epsilon \).
where $M$ is the pseudodifferential operator given by

$$\hat{M} \eta(k) = \frac{k \cosh(k) - \sinh(k)}{ik \sinh(k)} \hat{\eta}(k).$$

Notice that to lowest order, $M$ is $C \partial_\alpha$.

Now, notice that $h^{-1}(\xi) = \tilde{x}^{-1}(u(\xi, 0))$ and so we have an implicit equation for $h^{-1}$:

$$h^{-1}(\xi) = \tilde{x}^{-1} \left( \xi - \int_0^\xi \eta(\xi_1) d\xi_1 - M \eta(\xi) \right)$$

$$= \tilde{x}^{-1} \left( \xi - \int_0^\xi y(h^{-1}(\xi_1)) d\xi_1 - M(y \circ h^{-1})(\xi) \right).$$

(26)

Remark 9. In [28, p. 1494], Schneider and Wayne make a minor error in calculating this same function. As a result, they claim that the above representation gives an explicit formula for $h^{-1}$. Our correction here changes nothing about subsequent steps in their proofs.

Since $\tilde{x} = \alpha + \epsilon X(\epsilon \alpha)$, where $X$ is well behaved, we can expect a similar form for $\tilde{x}^{-1}$.

Lemma 8. Suppose $f(\alpha) = \alpha + g(\alpha)$ with $\|g\|_{C^2} \leq 1/2$. Then

$$f^{-1}(\xi) = \xi - g(\xi) + g(\xi)g'(\xi) + E,$$

where $E = O(\|g\|_{W^{2,\infty}}^3)$. More specifically

$$E \leq C \left( \|g'\|_{L^\infty} \|g\|_{L^\infty} + \|g\|_{L^2}^2 \|g''\|_{L^\infty} \right).$$

In particular, notice that if $g(\alpha) = \epsilon G(\epsilon \alpha)$ this means that $E = O(\epsilon^3)$.

Proof. See section 7.

We apply this lemma to $\tilde{x}$ and find that

$$\tilde{x}^{-1}(\xi) = \xi - \epsilon X(\epsilon \xi) + \epsilon^3 X(\epsilon \xi) \partial_\xi X(\epsilon \xi) + O(\epsilon^3).$$

Combining this with (26) we can determine $h^{-1}(\xi)$ (and therefore $h$) in terms of $x$ and $y$ to any order we wish. To lowest order we see that

$$h^{-1}(\xi) = \xi + O(\epsilon).$$

(27)

Therefore now we have

$$h^{-1}(\xi) = \xi - \int_0^\xi y(h^{-1}(\xi_1)) d\xi_1 - M(y \circ h^{-1})(\xi)$$

$$- \epsilon X \left( \epsilon \left( \xi - \int_0^\xi y(h^{-1}(\xi_1)) d\xi_1 - M(y \circ h^{-1})(\xi) \right) \right)$$

$$+ O(\epsilon^3)$$

$$= \xi - \int_0^\xi \epsilon^2 Y(\epsilon h^{-1}(\xi_1)) d\xi_1 - M(\epsilon^2 Y \circ \epsilon h^{-1})(\xi)$$

$$- \epsilon X \left( \epsilon \left( \xi - \int_0^\xi \epsilon^2 Y(\epsilon h^{-1}(\xi_1)) d\xi_1 - M(\epsilon^2 Y \circ \epsilon h^{-1})(\xi) \right) \right)$$

$$+ O(\epsilon^3).$$
If we insert (27) into the above and expand we have

\[ h^{-1}(\xi) = \xi - \epsilon X(\epsilon \xi) - \int_0^\xi \epsilon^2 Y(\epsilon \xi) d\xi - \epsilon^2 M(\epsilon Y(\epsilon))(\xi) + O(\epsilon^3). \]

One can continue in this manner and determine the next order terms in the expansion of \( h^{-1} \). If we let

\[ \epsilon G_1(\epsilon \alpha) = -\epsilon X(\epsilon \alpha) - \int_0^\alpha \epsilon^2 Y(\epsilon \alpha) d\alpha - \epsilon^2 M(\epsilon Y(\epsilon))(\alpha), \]

the expansion is

\[ h^{-1}(\xi) = \xi + \epsilon G_1(\epsilon \xi) + \epsilon^3 B_1(\epsilon \xi) + O(\epsilon^5), \]

where

\[ B_1(\xi) = \int_0^\xi \epsilon G_1(\epsilon \alpha) \partial_\beta \epsilon^3 Y(\epsilon \alpha) d\alpha + M(\epsilon^3 G_1(\epsilon \cdot) \partial_\beta Y(\epsilon \cdot))(\xi) + \epsilon^3 G_1(\epsilon \xi) \partial_\beta X(\epsilon \xi). \]

Notice that since \( M \) is \( C\partial_\alpha \) to lowest order, \( -\epsilon^2 M(\epsilon Y(\epsilon))(\alpha) \) is \( O(\epsilon^3) \). Moreover, by hypothesis, we have \( \epsilon^2 Z(\epsilon \alpha) - \epsilon^2 Y(\epsilon \alpha) = \epsilon^4 \Delta(\epsilon \alpha) \). Thus

\[
-\epsilon X(\epsilon \alpha) - \int_0^\alpha \epsilon^2 Y(\epsilon \alpha) d\alpha = - \int_0^\alpha \left( \epsilon^2 \partial_\beta X(\epsilon \alpha) + \epsilon^2 Y(\epsilon \alpha) \right) d\alpha
= - \int_0^\alpha \left( \epsilon^2 L(Z(\epsilon \cdot))(\alpha) + \epsilon^2 Y(\epsilon \alpha) \right) d\alpha
= \int_0^\alpha \left( \epsilon^2 Z(\epsilon \alpha) - \epsilon^2 Y(\epsilon \alpha) \right) d\alpha + O(\epsilon^3)
= \int_0^\alpha \left( \epsilon^4 \Delta(\epsilon \alpha) \right) d\alpha + O(\epsilon^3)
= O(\epsilon^3).
\]

That is, \( \epsilon G_1 \) is really \( O(\epsilon^3)! \) This cancellation is the crucial step in this proof. Since \( \epsilon G_1 \) appears in each term in \( B_1 \), we have shown

\[ h^{-1}(\xi) = \xi + \epsilon^3 G(\epsilon \xi) + O(\epsilon^5) \]

with \( \epsilon^3 G = \epsilon G_1 \). We appeal to Lemma 8 again, and we have

\[ h(\alpha) = \alpha - \epsilon^3 G(\epsilon \alpha) + O(\epsilon^5). \]

Now that we have particularly good estimates on \( h \) and \( h^{-1} \), we can begin our discussion of \( K \) in earnest. For notational simplicity, we will let

\[ h(\alpha) = \alpha + g_1(\alpha), \]

\[ h^{-1}(\xi) = \xi + g_2(\xi). \]

If we let

\[ \tilde{f} = Q^{-1} f, \]
we can make the following formal approximation using Taylor’s theorem:

\[ K(x, y)f(\alpha) = Q \circ K_0 \tilde{f}(\alpha) \]
\[ = K_0 \tilde{f}(h(\alpha)) \]
\[ = K_0 \tilde{f}(\alpha + g_1(\alpha)) \]
\[ = K_0 \tilde{f}(\alpha) + g_1(\alpha)K_0 \partial_\alpha \tilde{f}(\alpha) + \text{h.o.t.} \]

Also by Taylor’s theorem,

\[ \tilde{f}(\alpha) = f(\alpha) + g_2(\alpha)\partial_\alpha f(\alpha) + \text{h.o.t.} \]

Putting these together we have

\[ K(x, y)f(\alpha) = K_0 f(\alpha) + g_1(\alpha)K_0 f'(\alpha) + K_0 (g_2 f')(\alpha) + \text{h.o.t.} \]

Thus let

\[ E_1 f = K_0 f(\alpha) + g_1(\alpha)K_0 f'(\alpha) + K_0 (g_2 f')(\alpha), \]
\[ E_2 f = K(x, y)f - E_1 f. \]

We prove Proposition 3 if we can prove

- \( \| E_1 f - K_0 f - K_1 (x, y)f\|_s \leq C\epsilon^{17/2} \)
- \( \| E_2 f\|_s \leq C\epsilon^{17/2} \).

Let us deal with \( E_2 f \) first. We can rewrite \( E_2 f \) as

\[ E_2 f = E_2^1 f + E_2^2 f + E_2^3 f \]

with

\[ E_2^1 f = K(x, y)f - K_0 \tilde{f} - g_1 K_0 \partial_\alpha \tilde{f}, \]
\[ E_2^2 f = K_0 \tilde{f} - K_0 f - K_0 (g_2 \partial_\alpha f), \]
\[ E_2^3 f = g_1 K_0 \partial_\alpha \tilde{f} - g_1 K_0 \partial_\alpha f. \]

As our approximation for \( K \) was determined by an application of Taylor’s theorem, we need to prove a lemma which shows that this formal step can be made rigorous, at least for functions in the weighted Sobolev spaces.

**Lemma 9.** Suppose \( F \in H^s(n), s > 1/2, n > 1/2. \) Then for all \( C_0 > 0 \) there exists \( \epsilon_0 \) such that for \( \epsilon \in [0, \epsilon_0] \) there is a constant \( C \) independent of \( \epsilon \) such that

\[ \left( \int_{|\alpha| > C_0 \epsilon^{-3}} |F(\epsilon \alpha)|^2 d\alpha \right)^{1/2} \leq C \epsilon^{2n - 3/2}. \]

Moreover, for \( 1 \leq j \leq s \) we have

\[ \left( \int_{|\alpha| > C_0 \epsilon^{-3}} |\partial^j_\alpha F(\epsilon \alpha)|^2 d\alpha \right)^{1/2} \leq C \epsilon^{2n - 3/2 + j}. \]

**Proof.** See section 7.  \[ \square \]
Remark 10. If instead we are considering

\[ \left( \int_{|\alpha| > C_0 \epsilon^{-3}} |F(\epsilon(\alpha + t), \epsilon^3 t)|^2 d\alpha \right)^{1/2} \]

with \( F(\cdot, T) \in H^s(n) \) for \( T \in [0, T_0] \), we can maintain the same bound as above by taking \( C_0 \geq 2T_0 \).

We can use the above lemma to prove a version of Taylor’s theorem.

**Lemma 10.** Suppose \( F \in H^s(2), s > 5/2, \) and \( g \in L^\infty \). Then

\[ \|F(\epsilon \cdot + \epsilon^2 g(\epsilon)) - F(\epsilon)\|_{L^2} \leq C \epsilon^{3/2}. \]

**Proof.** By Lemma 9 we have

\[ \|F(\epsilon \cdot + \epsilon^2 g(\epsilon)) - F(\epsilon)\|^2 = \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha)|^2 d\alpha \]

\[ + \int_{|\alpha| \geq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha)|^2 d\alpha \]

\[ \leq \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha)|^2 d\alpha + C \epsilon^5. \]

Now, we add and subtract \( \epsilon^2 g(\epsilon \alpha) F'(\epsilon \alpha) \) in the remaining integral,

\[ \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha)|^2 d\alpha \]

\[ \leq \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha) - \epsilon^2 g(\epsilon \alpha) F'(\epsilon \alpha)|^2 d\alpha \]

\[ + \int_{|\alpha| \leq \epsilon^{-3}} |\epsilon^2 g(\epsilon \alpha) F'(\epsilon \alpha)|^2 d\alpha \]

\[ \leq \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha) - \epsilon^2 g(\epsilon \alpha) F'(\epsilon \alpha)|^2 d\alpha \]

\[ + \epsilon^3 \|g(\cdot)\|_{L^\infty}^2 \|F'(\cdot)\|^2_{L^2}. \]

We naively bound the above integral and apply the mean value theorem. That is,

\[ \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha) - \epsilon^2 g(\epsilon \alpha) F'(\epsilon \alpha)|^2 d\alpha \]

\[ \leq C \epsilon^{-3} \sup_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon \alpha + \epsilon^2 g(\epsilon \alpha)) - F(\epsilon \alpha) - \epsilon^2 g(\epsilon \alpha) F'(\epsilon \alpha)|^2 \]

\[ \leq C \epsilon^{-3} \sup_{|\alpha| \leq \epsilon^{-3}} |\epsilon^4 g^2(\epsilon \alpha) F''(\epsilon \alpha)|^2 \]

\[ \leq C \epsilon^5 \|g\|^4_{L^\infty} \|F''(\cdot)\|^2_{L^\infty}. \]

With this, we have proven the lemma. \( \qed \)

**Remark 11.** With this general technique we are also able to show that

\[ \|\partial_{\alpha}^j (F(\epsilon \cdot + \epsilon^2 g(\epsilon)) - F(\epsilon \cdot))\|_{L^2} \leq C \epsilon^{3/2 + j}, \]
long wave approximation of can use Lemma 10. In what follows, that the first line is identically zero. Let 

Thus \( C \epsilon \) and so on.

Taking the Fourier transform of the second term we have

Now notice that \( \hat{f}(\alpha) = L_y \) so the above becomes

\[
\| \hat{f}(\alpha) \|_{L^2} \leq C \epsilon^{11/2 + j},
\]

Now we will be able to control \( E_2^{1/2} \). We can control the other functions in precisely the same fashion. Since \( f \) is of long wavelength and rapid decay, so is \( \hat{f} \). Thus we can use Lemma 10. In what follows, \( \epsilon^3 \hat{F}(\epsilon) = K_0 \hat{f}(\alpha) \). (The extra \( \epsilon \) comes from the long wave approximation of \( K_0 \).

\[
\| E_2^{1/2}(\cdot) \|_{L^1} = \| K_0 \hat{f}(\cdot + g_1(\cdot)) - K_0 \hat{f}(\cdot) - g_1(\cdot) K_0 \alpha \hat{f}(\cdot) \|_{L^1}
\]

\[
\leq C \epsilon^{3/2} \| \hat{F}(\cdot + \epsilon^4 G(\epsilon)) - \hat{F}(\cdot) - \epsilon^4 G(\epsilon) \hat{F}(\cdot) \|_{L^1}
\]

\[
\leq C \epsilon^{17/2}.
\]

Now we turn our attention to \( E_1 f - K_0 f - K_1(x, y) f \). A routine calculation shows that this is equal to

\[
- \epsilon^3 [G(\epsilon \alpha), K_0] \epsilon^3 \partial_\alpha F(\epsilon \alpha) - K_1(\epsilon X, \epsilon^2 Y) \epsilon^2 F(\epsilon \alpha)
\]

\[
+ (g_1 + \epsilon^3 G) K_0 \epsilon^3 \partial_\alpha F(\epsilon \alpha) + K_0 ((g_2 - \epsilon^3 G) \epsilon^3 \partial_\alpha F(\epsilon \alpha))(\alpha).
\]

Now, \( g_1 + \epsilon^3 G \) and \( g_2 - \epsilon^3 G \) are \( O(\epsilon^5) \), so the second line above can be bounded by \( C \epsilon^{17/2} \), if we make use of \( K_0 \)'s long wavelength approximation. Moreover, we claim that the first line is identically zero. Let

\[
b(\alpha) = - \int_0^\alpha \epsilon^2 Y(\epsilon \alpha_1) d\alpha_1 - \epsilon^2 M(Y(\epsilon))(\alpha)
\]

\[
= - \int_0^\alpha y(\alpha_1) d\alpha_1 - My(\alpha).
\]

Thus \( \epsilon^3 G(\epsilon \alpha) = b(\alpha) - x(\alpha) \). So

\[
\epsilon^3 [G(\epsilon \alpha), K_0] \epsilon^3 \partial_\alpha F(\epsilon \alpha) = - [x(\alpha), K_0] \partial_\alpha f(\alpha) + [b(\alpha), K_0] \partial_\alpha f(\alpha).
\]

Taking the Fourier transform of the second term we have

\[
\mathfrak{F} ([b(\alpha), K_0] \partial_\alpha f(\alpha))(k)
\]

\[
= \int \left( \hat{K}_0(l) - \hat{K}_0(k) \right) b(l - k) i l \hat{f}(l) dl
\]

\[
= \int \hat{K}_0(l) - \hat{K}_0(k) \frac{i(l - k)}{i(l - k)} b(l - k) i l \hat{f}(l) dl
\]

\[
= \int \hat{K}_0(l) - \hat{K}_0(k) \frac{\partial_\alpha b(l - k) i l \hat{f}(l) dl}{i(l - k)}.
\]

Now notice that \( \partial_\alpha b = Ly \), so the above becomes

\[
\int \hat{K}_0(l) - \hat{K}_0(k) \frac{L_y(l - k) i l \hat{f}(l) dl}{i(l - k)} = \int \hat{K}_0(l) - \hat{K}_0(k) \frac{\gamma'(l - k) i l \hat{f}(l) dl}{K_0(l - k)}
\]

\[
= \mathfrak{F} ((y + K_0 y K_0) \partial_\alpha f) (k),
\]
where we have used the trigonometric identity (11) from section 4.
That is,
\[ e^3[G(\alpha), K_0]e^{\beta}F(\alpha) = -[x(\alpha), K_0]\partial_\alpha f(\alpha) + ((y + K_0y)\partial_\alpha f)(k) \]
\[ = -K_1(x, y)f \]
and so we are done with the proof of Proposition 3. □

Proof of Proposition 4. Let \( h \) be as in Proposition 3 and \( h_2 \) be the analogous function for the configuration \((z_2, y_2) = (z + \epsilon^{11/2}R^z, y + \epsilon^{11/2}R^y)\). We also define the function \( \tilde{x}_2 = \alpha + x_2 \) by \( \partial_\alpha x_2 = L\tilde{z}_2 \). Unlike in the previous lemma, here the time dependence of the functions is important. Thus we determine \( x \) and \( x_2 \) by integrating in both space and time. That is,
\[ \hat{x}(\alpha, t) = \alpha + \epsilon\chi(\alpha, \epsilon t), \]
\[ \hat{x}_2(\alpha, t) = \alpha + \epsilon\chi(\alpha, \epsilon t) + \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho(\alpha, t) \]
with
\[ \epsilon\chi(\alpha, \epsilon t) = \left( \int_0^t u(0, s) \, ds + \int_0^\alpha L\tilde{z}(w, t) \, dw \right), \]
\[ \epsilon^{5/2}E(t) = \epsilon^{11/2} \int_0^t R^u(0, s) \, ds, \]
\[ \epsilon^{11/2}\rho = \int_0^\alpha LR^z(w, t) \, dw. \]

The functions satisfy the following estimates for all \( t \in [0, T_0\epsilon^{-3}] \):
\[ |\epsilon\chi(\alpha, \epsilon t)| \leq C\epsilon, \]
\[ |\epsilon^{5/2}E(t)| \leq C\epsilon^{5/2}, \]
\[ |\epsilon^{11/2}\rho(\alpha, t)| \leq C\sqrt{\alpha}\|R^z\|_{\mathcal{H}^{1/2}}. \]

The first estimate follows from similar estimates in the previous lemma, the second is the naive bound, and the final follows from the following simple fact.

**Fact 2.** If \( p(\alpha) = \int_0^\alpha r(\alpha) \, d\alpha \), where \( r \in L^2 \), then
\[ |p(\alpha)| \leq \sqrt{\alpha}\|r\|_{L^2}. \]

In what follows we will make strong use of the fact that \( E(t) \) does not depend on \( \alpha \).

Using the same techniques as were used in proving Lemma 8, one can show that
\[ \hat{x}_2^{-1}(\xi, t) = \hat{x}^{-1}(\xi, t) - \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_2(\xi), \]
where \( |\rho_2(\alpha)| \leq \sqrt{\alpha}\|R^z\|_{L^2} \). This sort of estimate carries over to the functions \( h \).
That is,
\[ h_2^{-1}(\xi, t) = h^{-1}(\xi, t) - \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_3(\xi, t), \]
\[ h_2(\alpha, t) = h(\alpha, t) + \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_4(\alpha, t), \]
where \( |\rho_3(\alpha, t)|, |\rho_4(\alpha, t)| \leq \sqrt{\alpha}(\|R^z\|_{L^2} + \|R^y\|_{L^2}) \) over the long time scale.
Now define $Q_2 f = f \circ h_2$. Therefore
\[
\begin{align*}
S_1(z + \epsilon^{11/2} R^z, y + \epsilon^{11/2})f - S_1(z, y)f =& \quad Q_2 \circ K_0 \circ Q_2^{-1} f - Q \circ K_0 \circ Q^{-1} f \\
= & \quad Q \circ (Q^{-1} \circ Q_2 \circ K_0 \circ Q_2^{-1} \circ Q - K_0) \circ Q^{-1} f.
\end{align*}
\]
Since $Q$ and its inverse are bounded operators from $H^s$ to $H^s$, we need only prove the estimate for the operator
\[
\hat{Q} \circ K_0 \circ \hat{Q}^{-1} - K_0,
\]
where $\hat{Q} = Q^{-1} \circ Q_2$. Notice that $\hat{Q} \circ K_0 \circ \hat{Q}^{-1}$ is the Hilbert operator $K$ for a domain with the “$h$” function given by $\hat{h}(\alpha) = h_2(h^{-1}(\alpha))$. Moreover, from the above calculations for $h$ and $h_2$ we have
\[
h_2(h^{-1}(\alpha)) = \alpha + \epsilon^{5/2} E(t) + \epsilon^{11/2} \rho_5(\alpha, t),
\]
with $\rho_5$ satisfying the same type of estimates as $\rho_4$.

At this point we can make an appeal to Lemma 3.14 on page 1500 of [28]. In this lemma they prove that $\|S_1(z, y)f\|_s \leq C (\|z\|_s + \|y\|_s)\|f\|_3$. In the course of their proof, they show that if $h(\alpha) = \alpha + g(\alpha)$, then
\[
\|Q \circ K_0 \circ Q^{-1} f(\cdot) - K_0 f(\cdot)\|_s \leq C \|\partial_\alpha g\|_{s-1} \|\partial_\alpha f\|_2.
\]
(See the inequalities in Cases I–IV on pages 1501–1506.) Therefore, if we set $\tilde{g} = \epsilon^{5/2} E(t) + \epsilon^{11/2} \rho_5(\alpha, t)$, we see that taking a spatial derivative leaves us with
\[
\partial_\alpha \tilde{g} = O(\epsilon^{11/2}).
\]
Thus, if we keep in mind that $f(\alpha, t) = \epsilon^{2} F(\epsilon(\alpha \pm t), \epsilon^3 t)$,
\[
\|\tilde{Q} \circ K_0 \circ \tilde{Q}^{-1} f(\cdot) - K_0 f(\cdot)\|_s \leq C \|\partial_\alpha \tilde{g}\|_{s-1} \|\partial_\alpha f\|_2 \leq C \epsilon^{17/2}.
\]
This completes the proof of Proposition 4. \qed

6. The error estimates. In this section we prove that the approximation is rigorous. That is we will prove Theorem 2. We will be working with the three- and four-dimensional formulations of the water wave problem (equations (WW3) and (WW4)). From [28], we know that for initial data of the type we are considering, solutions to these equations exist over the long times we are considering. If $(z, y, u)$ is a solution to (WW3), let
\[
\begin{align*}
z(\alpha, t) &= \Psi^z(\alpha, t) + \epsilon^{11/2} R^z(\alpha, t), \\
y(\alpha, t) &= \Psi^y(\alpha, t) + \epsilon^{11/2} R^y(\alpha, t), \\
u(\alpha, t) &= \Psi^u(\alpha, t) + \epsilon^{11/2} R^u(\alpha, t),
\end{align*}
\]
with the functions $\Psi$ defined as above. We call $R^z$, $R^y$, and $R^u$ “error” functions and we denote $\tilde{R} = (R^z, R^y, R^u)$. Our goal will be to show that $\tilde{R}$ remain $O(1)$ in $\tilde{g}^s = H^s \times H^s \times H^{s-1/2}$ over the long time scale. If we can do this, then we will have proven the main theorem. The first step will be to determine the equations which
these functions satisfy. Loosely, we want to be able to write for each of the error functions an evolution equation of the form
\[ \partial_t R = \text{quasi-linear + small and smooth}. \]

We will at times go to great lengths to achieve this!

Clearly,
\[ \partial_t R^z = K_0 R^u. \]

Finding the equations for \( R^y \) and \( R^u \) is a bit more complex. First we focus on \( R^y \).

Substituting from (28) into \( \partial_t y = K(z,y)u \), we have
\[
\partial_t R^y = \epsilon^{-11/2} \left( K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y)(\Psi^u + \epsilon^{11/2} R^u) - \partial_t \Psi^y \right)
= K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y)R^u
+ \epsilon^{-11/2} \left( K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y)\Psi^u - \partial_t \Psi^y \right)
= K_0 R^u + M_1(\Psi^z)\partial_u R^u - (\Psi^y + K_0(\Psi^y K_0))\partial_a R^u + N^y,
\]

where
\[
N^y = \epsilon^{-11/2} R_{\Psi^y} + \epsilon^{-11/2} \left( \left( S_1(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - S_1(\Psi^z, \Psi^y) \right) \Psi^u \right)
+ \left( K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - K_0 - K_1(\Psi^z, \Psi^y) \right) R^u.
\]

We claim that \( N^y \) is “small.” That is, we have the following lemma.

**Lemma 11.** For all \( C_R > 0 \), there exists \( \epsilon_0 \) such that for all \( \epsilon \in (0, \epsilon_0) \) and \( t \) such that \( \sup_{0 \leq t' \leq t} ||\tilde{R}(\cdot,t')||_{\mathcal{B}^*} \leq C_R \) we have
\[
||N^y||_s \leq C \left( \epsilon^3 + \epsilon^3 ||\tilde{R}||_{\mathcal{B}^*} + \epsilon^{11/2} ||\tilde{R}||_{\mathcal{B}^*}^2 \right).
\]

**Proof.** First we remark that the approximating functions \( \Psi \) and their derivatives are all bounded over the long time scales. Thus, we will not be keeping track of the dependence of the norm of \( N^y \) on the norms of these functions. By Proposition 2, we know that \( ||\epsilon^{-11/2} R_{\Psi^y}||_s \leq C \epsilon^3 \).

We can bound
\[
\epsilon^{-11/2} \left( S_1(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - S_1(\Psi^z, \Psi^y) \right) \Psi^u
\]
by Lemma 4.

Finally,
\[
\|(K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - K_0 - K_1(\Psi^z, \Psi^y))R^u\|_s
\leq C \left( \epsilon^3 ||\tilde{R}||_{\mathcal{B}^*} + \epsilon^{11/2} ||\tilde{R}||_{\mathcal{B}^*}^2 \right)
\]
by the estimates on \( K \) and its expansions which we saw in sections 3 and 5 (in particular Lemmas 5 and 6). \( \Box \)

Now we discuss \( R^u \). We know that
\[
\partial_t u(1 + Lz) + \partial_a y(1 + \partial_t^2 y) = 0.
\]
We would like an evolution-type equation for \( R^u \). Notice that since \( \partial_y = K(z,y)u \), there is a “hidden” \( \partial_y \) term in the term \( \partial^2_y \). Recall that the commutator \([\partial_t, S_1(z,y)]u\) can be bounded independently of \( \partial_t u \) (see Lemma 6 in section 5). Therefore, we can rewrite the above as

\[
(1 + Lz + \partial_y K(z,y)) \partial_t u + \partial_y (1 + [\partial_t, S_1(z,y)]u) = 0.
\]

Replacing \( u \) with its definition in (28) the above becomes

\[
0 = (1 + Lz + \partial_y K(z,y)) \partial_t \epsilon^{11/2} R^u
+ (1 + Lz + \partial_y K(z,y)) \partial_t \epsilon \Psi^u
+ \partial_y (1 + \partial_t (K(z,y) \Psi^u))
+ \partial_y \left( [\partial_t, S_1(z,y)] \epsilon^{11/2} R^u \right).
\]

or rather

\[
0 = (1 + Lz + \partial_y K(z,y)) \partial_t \epsilon^{11/2} R^u
+ (1 + Lz) \partial_t \epsilon \Psi^u
+ \partial_y (1 + \partial_t (K(z,y) \Psi^u))
+ \partial_y \left( [\partial_t, S_1(z,y)] \epsilon^{11/2} R^u \right).
\]

We rearrange this a bit and break up \( y \) and \( z \):

\[
0 = (1 + Lz + \partial_y K(z,y)) \partial_t R^u + \partial_y R^y
+ LR^z \partial_t \Psi^u + \partial_y R^y \partial_t (K(z,y) \Psi^u) + \partial_y \partial_t [\partial_t, S_1(z,y)] R^u
+ \epsilon^{-1/2} \left( (1 + L \Psi^u) \partial_t \epsilon \Psi^u + \partial_y \Psi^u (1 + \partial_t (K(z,y) \Psi^u)) \right).
\]

The operator

\[
A(z,y) = (1 + Lz + \partial_y K(z,y))
\]

is invertible since \( K(z,y) \) is a bounded operator on \( H^s \), provided \( z \) and \( y \) are small (which they are). Moreover, we can approximate \( A^{-1} \) via the Neumann series. Thus the above equation can be rewritten as

\[
\partial_t R^u = -(1 - \epsilon^2 W_1) \partial_y R^y + N^u,
\]

where \( N^u = N_1^u + N_2^u \) and

\[
N_1^u = - \epsilon^{-11/2} A^{-1} \left( (1 + L \Psi^u) \partial_t \epsilon \Psi^u + \partial_y \Psi^u (1 + \partial_t (K(z,y) \Psi^u)) \right),
\]

\[
N_2^u = - A^{-1} \left( LR^z \partial_t \Psi^u + \partial_y R^y \partial_t (K(z,y) \Psi^u) + \partial_y \partial_t [\partial_t, S_1(z,y)] R^u \right)
+ (-A^{-1} + (1 - \epsilon^2 W_1)) \partial_y R^y.
\]

**Lemma 12.** For all \( C_R > 0 \), there exists \( \epsilon_0 \) such that for all \( \epsilon \in (0, \epsilon_0) \) and \( t \) such that \( \sup_{0 \leq t' \leq t} \| \tilde{R}(\cdot, t') \|_{H^s} \leq C_R \) we have

\[
\| N^u \|_{s-1} \leq C \left( \epsilon^3 + \epsilon^3 \| \tilde{R} \|_{H^s} + \epsilon^{11/2} \| \tilde{R} \|_{H^s}^2 \right).
\]
Proof. First we point out that this estimate is in $H^{s-1}$. The loss of regularity here is easily seen. Both $LR^s$ and $\partial_t R^R$ explicitly appear in $N_2^u$, and are not smoothed by any operators. Thus, losing this derivative is unavoidable. In fact, it is easy to see that the above estimates hold for $N_2^u$ by noting that $A^{-1}$, $K$, and $[\partial_t, S_1]$ are bounded operators.

Bounding $N_3^u$ is also easily done once we recognize that this term is almost exactly $\epsilon^{-1/2}\text{Res}_u$. We have

$$
\epsilon^{-1/2}A^{-1} \left( (1 + L^2)^{-1} \partial_t \epsilon^2 \Psi + \partial_y \Psi (K(z, y) \Psi) \right) + N_3^u
$$

$$
= \epsilon^{-1/2}A^{-1} \left( (1 + L^2)^{-1} \partial_t \epsilon^2 \Psi + \partial_y \Psi (K(z, y) \Psi) \right) + N_3^u
$$

$$
= \epsilon^{-1/2} \left( \partial_t \epsilon^2 \Psi + \partial_y \Psi \frac{1 + \partial_t (K(z, y) \Psi)}{1 + L^2} \right) + N_3^u
$$

$$
= \epsilon^{-1/2} \text{Res}_u + N_3^u + N_4^u + N_5^u,
$$

where

$$
N_3^u = \epsilon^{-1/2}A^{-1} \partial_y \Psi \partial_z (K(z, y) \Psi - K(z, y) \Psi),
$$

$$
N_4^u = \epsilon^{-1/2}A^{-1} \left( (1 + L^2)^{-1} \partial_t \epsilon^2 \Psi + \partial_y \Psi \right),
$$

$$
N_5^u = \epsilon^{-1/2} \partial_y \Psi \frac{\partial_t \text{Res}_y}{1 + L^2}.
$$

We bound $N_3^u$ using mean value theorem arguments entirely analogous to those used when bounding $N_4^u$. To bound $N_4^u$, one observes that

$$
(1 + L^2)^{-1} \partial_t \epsilon^2 \Psi + \partial_y \Psi (K(z, y) \Psi)
$$

is very nearly $\text{Res}_u$ and is thus $O(\epsilon^{1/2})$. $N_5^u$ is clearly small, as it contains $\partial_t \text{Res}_y$. This completes the proof.

We need to make analogous calculations for the four-dimensional system. Let

$$
a(\alpha, t) = \epsilon^3 \Psi \left( c \alpha, ct \right) + \epsilon^{11/2} R^a(\alpha, t)
$$

and $\bar{R}_c = \left( R^2, R^R, R^R, R^R \right)$. This extended set of error functions lives in $S_3^s = H^s \times H^s \times H^{s-1/2} \times H^{s-1}$.

It is easy to see that

$$
\partial_t R^a = R^a
$$

but more difficult to determine the evolution of $R^a$. We begin by taking a time derivative of (29):

(30) \[ \partial_t^2 u + \partial_t u \partial_x u + \partial_x \partial_t y (1 + \partial_t^2 y) + \partial_y y \partial_x^2 y = 0. \]

Letting

$$
I = \partial_t^2 u + \partial_x y \partial_x^2 y,
$$

$$
II = \partial_x \partial_t y (1 + \partial_t^2 y) + \partial_t u \partial_x u,
$$

$$
\partial_t R^a = R^a
$$
(30) is $I + II = 0$.

Manipulations very similar to those carried out in determining $\partial_t R^u$ show that

$$I = A(z,y)\partial_t \varepsilon^{11/2} R^u + \partial_{\alpha} y[\partial_t^2, S_1(z,y)]\varepsilon^{11/2} R^u + (1 + Lz)\partial_t \varepsilon^{3} \Psi^u + \partial_{\alpha} y \partial_t^2 (K(z,y) \Psi^u).$$

For $II$, we have

$$II = (1 + \partial_t^2 y)\partial_{\alpha} \partial_t \varepsilon^{11/2} R^u + \partial_t u \partial_{\alpha} \varepsilon^{11/2} R^u + (1 + \partial_t^2 y)\partial_{\alpha} \partial_t \Psi^u + \partial_t u \partial_{\alpha} \Psi^u + \varepsilon^{11/2} (1 + \partial_t^2 y - \partial_t u K_0) K_0 \partial_{\alpha} R^u + \partial_{\alpha} K_1 (\Psi^z, \Psi^y) R^u + (1 + \partial_t^2 y)\partial_{\alpha} \partial_t \Psi^u + \partial_t u \partial_{\alpha} \Psi^u + \varepsilon^{11/2} B_{II},$$

where

$$B_{II} = \partial_t u (1 + K_0^2) \partial_{\alpha} R^u + \partial_t^2 y \partial_{\alpha} (K_1 (\Psi^z, \Psi^y) R^u + N^y) + \partial_{\alpha} N^y.$$ 

Noting that $\partial_t^2 y = K_0 a + [\partial_t, S_1] u + S_1 a$, we see that $B_{II}$ is smooth in the error functions, and is $O(\varepsilon^3)$.

Adding $I$ and $II$ gives

$$0 = A(z,y)\partial_t R^u + (1 + \partial_t^2 y - \partial_t u K_0) K_0 \partial_{\alpha} R^u + \partial_{\alpha} K_1 (\Psi^z, \Psi^y) R^u + B,$$

where

$$B = B_{II} + \partial_{\alpha} y[\partial_t^2, S_1(z,y)] R^u + B_{Res},$$

$$\varepsilon^{11/2} B_{Res} = (1 + Lz)\partial_t \varepsilon^{3} \Psi^u + \partial_{\alpha} y \partial_t^2 (K(z,y) \Psi^u) + (1 + \partial_t^2 y)\partial_{\alpha} \partial_t \Psi^u + \partial_t u \partial_{\alpha} \Psi^u.$$

The terms $B_{II}$ and $\partial_{\alpha} y[\partial_t^2, S_1(z,y)]$ are small and smooth, and we can bound $B_{Res}$ via the residual estimates, much as we did for $N^u$ above. That is, we have

$$\|B\|_{s-1} \leq C \left( \varepsilon^3 + \varepsilon^3 \|\tilde{R}\|_{\tilde{\alpha}^2} + \varepsilon^{11/2} \|\tilde{R}\|_{\tilde{\alpha}^2} \right)$$

under the same hypotheses as in the above lemmas.

At this time, it is tempting to simply invert $A(z,y)$. Though we could do this, the inverse of this operator is not smoothing. In particular the presence of the term $\partial_{\alpha} y K_0$ in $A$ will cause problems. We can eliminate $K_0$ to highest order by letting $H_0(z,y) = (1 + Lz - \partial_{\alpha} y K_0)$ act on (31). We have for the first term

$$H_0(z,y) A(z,y) \partial_t R^u$$

$$= (1 + Lz)^2 \partial_t R^u + (1 + Lz)\partial_{\alpha} y K_0 \partial_t R^u - \partial_{\alpha} y K_0 \left( (1 + Lz) \partial_t R^u \right)$$

$$- \partial_{\alpha} y K_0 \left( \partial_{\alpha} y K_0 \partial_t R^u \right) + H_0(z,y)(\partial_{\alpha} y S_1(z,y) \partial_t R^u)$$

$$= \left( (1 + Lz)^2 + (\partial_{\alpha} y)^2 + H_1(z,y) \right) \partial_t R^u,$$
where

\[ H_1(z, y) = \partial_\alpha y \left( [Lz, K_0] \cdot -K_0 \partial_\alpha y, K_0 \right) \cdot - (1 + K_0^2) \partial_\alpha y \cdot + H_0(z, y) \left( \partial_\alpha y S_1(z, y) \right) \cdot . \]

Notice that \( H_1 \) is made up of smoothing operators and is thus a smoothing operator.

Now, for the second term in (31) we have

\[ H_0(z, y)(1 + \partial_t^2 y - \partial_t u) K_0 \partial_\alpha R^a = (1 + Lz)(1 + \partial_t^2 y) K_0 \partial_\alpha R^a - (1 + Lz) \partial_t u K_0^2 \partial_\alpha R^a \]
\[ - \partial_\alpha y K_0 \left( (1 + \partial_t^2 y) K_0 \partial_\alpha R^a \right) + \partial_\alpha y K_0 \left( \partial_t u K_0^2 \partial_\alpha R^a \right) \]
\[ = (1 + Lz)(1 + \partial_t^2 y) K_0 \partial_\alpha R^a \]
\[ - \partial_\alpha y [K_0, \partial_t^2 y] K_0 \partial_\alpha R^a + \partial_\alpha y K_0 \left( \partial_t u K_0^2 \partial_\alpha R^a \right) \]
\[ + (\partial_t u (1 + Lz) + \partial_t y (1 + \partial_t^2 y) ) K_0^2 \partial_\alpha R^a. \]

Notice that by comparing the last line of the above with (29), we see that it is identically zero! One more rearrangement of this yields

\[ (1 + Lz)(1 + \partial_t^2 y - \partial_\alpha y \partial_t u) K_0 \partial_\alpha R^a + B_2, \]

where

\[ B_2 = \partial_\alpha y \left( -[K_0, \partial_t^2 y] + [K_0, \partial_t u] K_0 + \partial_t u (1 + K_0^2) \right) K_0 \partial_\alpha u \]

is a smooth and small function by Lemma 4 in section 5.

If we let

\[ f = ((1 + Lz)^2 + (\partial_\alpha y)^2)^{-1}, \]
\[ g = (1 + Lz)(1 + \partial_t^2 y) - \partial_\alpha y \partial_t u, \]

then we have transformed (31) into

\[ 0 = (1 + f H_1(z, y)) \partial_t R^a + f g K_0 \partial_\alpha R^a \]
\[ + f H_0(z, y) \left( \partial_\alpha (K_1(\Psi^x, \Psi^y) R^a) \right) \]
\[ + f (B_2 + H_0(z, y) B). \]

By the Neumann series,

\[ (1 + f H_1(z, y))^{-1} = 1 + \sum_{n=0}^\infty (-1)^n f^n H_1^n(z, y) \cdot . \]

Since \( H_1 \) is smoothing, this is the identity plus a smoothing piece. Let

\[ H_2(z, y) = \sum_{n=0}^\infty (-1)^n f^n H_1^n(z, y) \cdot . \]

Thus,

\[ 0 = \partial_t R^a + f g K_0 \partial_\alpha R^a \]
\[ + f H_0(z, y) \left( \partial_\alpha (K_1(\Psi^x, \Psi^y) R^a) \right) - N_1^a - N_2^a, \]
where
\[-N^a_1 = H_2(z, y) (fgK_0 \partial_\alpha R^u + f H_0(z, y) (\partial_\alpha (K_1(\Psi^z, \Psi^y) R^u))),\]
\[-N^a_2 = (1 + H_2(z, y)) f (B_2 + H_0(z, y) B).\]

Finally we rewrite the above as
\[\partial_t R^a = -(1 + K_0 a - Lz + N^a_s) \partial_\alpha K_0 R^u - \partial_\alpha (K_1(\Psi^z, \Psi^y) R^u) + N^a\]
with
\[N^a_s = fg - (1 + K_0 a - Lz),\]
\[N^a = N^a_s + N^a_2 + N^a_3,\]
\[N^a_4 = -(f H_0(z, y) - 1) (\partial_\alpha (K_1(\Psi^z, \Psi^y) R^u)).\]

Notice that \((1 + K_0 a - Lz)\) is the first order approximation to \(fg\). Thus, using techniques exactly like those we used in proving the bounds on \(N^u\) and \(N^a\), we have the following.

**Lemma 13.** For all \(C_R > 0\), there exists \(\epsilon_0\) such that for all \(\epsilon \in (0, \epsilon_0)\) and \(t\) such that \(\sup_{0 \leq t' \leq t} \|\bar{R}^e(\cdot, t')\|_{\mathcal{B}^2} \leq C_R\) we have
\[
\max \{\|N^a_s\|_{s-1}, \|N^a\|_{s-1}\} \leq C \left(\epsilon^3 + \epsilon^3 \|\bar{R}^e\|_{\mathcal{B}^2} + \epsilon^{11/2} \|\bar{R}^e\|_{\mathcal{B}^2}^2\right).
\]

Recapping, we have shown that the three-dimensional system may be rewritten as
\[
\begin{align*}
\partial_t R^z &= K_0 R^u, \\
\partial_t R^y &= K_0 R^u + M_1(\Psi^z) \partial_\alpha R^u - (\Psi^y + K_0(\epsilon^2 \Psi^y K_0)) \partial_\alpha R^u + N^y, \\
\partial_t R^u &= -(1 - \epsilon^2 W_1) \partial_\alpha R^u + N^u,
\end{align*}
\tag{32}
\]
and the four-dimensional system as
\[
\begin{align*}
\partial_t R^z &= K_0 R^u, \\
\partial_t R^y &= K_0 R^u + M_1(\Psi^z) \partial_\alpha R^u - (\Psi^y + K_0(\epsilon^2 \Psi^y K_0)) \partial_\alpha R^u + N^y, \\
\partial_t R^u &= R^u, \\
\partial_t R^a &= -(1 + K_0 a - Lz + N^a_s) \partial_\alpha K_0 R^u - \partial_\alpha (K_1(\Psi^z, \Psi^y) R^u) + N^a.
\end{align*}
\tag{33}
\]

We remark now that these are only cosmetically different from the equations which determine the evolution of the error for the KdV approximation alone in [28]. See page 1524 for the equations in three dimensions and page 1526 in four dimensions. Their variables
\[(Z_1, X_2, U_1, V_1)\]
correspond to our
\[(z, y, u, a),\]
and their functions

\[(N^2, N^3, N^4, N^5, N^8)\]

are our

\[(N^z, N^u, N^a, N^s, N^a).\]

The only difference of note is that their estimates contain a term they call \(q(t)\) while ours do not. This term, which is related to the interaction of the left- and right-moving wavetrains, has been removed in this paper by the inclusion of the function \(W_3\) in the approximating functions \(\Psi\). This simplification does not adversely affect the means which they employ to prove that the error functions remain \(O(1)\) over the long time scale. Therefore we appeal to their results on pages 1524–1533.

**Proposition 5.** For all \(T_0 > 0, s > 4\), and \(C_I > 0\), there exists \(\epsilon_0\) such that for all \(0 \leq \epsilon \leq \epsilon_0\), the unique solution \(\bar{R}_e\) of (33), with initial conditions such that

\[
\|\bar{R}_e(\cdot, 0)\|_{L^s_x} \leq C_I,
\]
satisfies

\[
\sup_{t \in [0, T_0 \epsilon^{-3}]} \|\bar{R}_e(\cdot, t)\|_{L^s_x} \leq C,
\]

where \(C\) is independent of \(\epsilon\).

Implicit in the above proposition is the assumption that the initial conditions for the water wave problem have the form

\[
\begin{pmatrix}
  z(\alpha, 0) \\
y(\alpha, 0) \\
u(\alpha, 0)
\end{pmatrix} = \begin{pmatrix}
  \Psi^z(\alpha, 0) \\
  \Psi^y(\alpha, 0) \\
  \Psi^u(\alpha, 0)
\end{pmatrix} + \epsilon^{11/2} \bar{R}_0(\alpha).
\]

Thus we see that this proposition immediately proves Theorem 3.

Now that we have this result, there are a few small steps and one big step needed to prove Theorem 2. The first simple step is to note that the \(z\) is not a very physical coordinate and that we would prefer estimates for \(x\). Since \(L\) is a bounded operator and gives the relationship between both \(z\) and \(x\) and \(\Psi^z\) and \(\Psi^d\), we automatically have

\[
\sup_{t \in [0, T_0 \epsilon^{-3}]} \|x(\cdot, t) - \Psi^d(\cdot, t)\|_{H^s} \leq C\epsilon^{11/2}.
\]

Second, the expressions for \(\Psi^d\), \(\Psi^y\), and \(\Psi^u\) contain terms of \(O(\epsilon^6)\). These terms were needed to make the residual sufficiently small, but they are unnecessary now. Moreover, the appearance of the operator \(L^{-1}\) and inverse derivatives in the definitions of \(\Psi^y\) and \(\Psi^u\) is not very intuitive. Thus, it is a simple consequence of Lemma 1 and the triangle inequality that

\[
\|\Psi^d - \psi^d\|_s \leq C\epsilon^{11/2},
\]

\[
\|\Psi^y - \psi^y\|_s \leq C\epsilon^{11/2},
\]

\[
\|\Psi^u - \psi^u\|_s \leq C\epsilon^{11/2},
\]
where $\psi^d$, $\psi^y$, and $\psi^u$ were given in the introduction in (4), (5), and (6). Therefore we have the following corollary.

**Corollary 1.** If the initial conditions for (WW) are of the form

$$(34) \quad \begin{pmatrix} x_{\alpha}(\alpha, 0) \\ y(\alpha, 0) \\ u(\alpha, 0) \end{pmatrix} = \begin{pmatrix} \psi^d(\alpha, 0) \\ \psi^y(\alpha, 0) \\ \psi^u(\alpha, 0) \end{pmatrix} + \epsilon^{11/2} \tilde{R}_1(\alpha)$$

with $\|\tilde{R}_1\|_{\mathcal{B}^s} \leq C_1$, then the solution of (WW) satisfies the estimate

$$\left\| \begin{pmatrix} x_{\alpha}(\cdot, t) \\ y(\cdot, t) \\ u(\cdot, t) \end{pmatrix} - \begin{pmatrix} \psi^d(\cdot, t) \\ \psi^y(\cdot, t) \\ \psi^u(\cdot, t) \end{pmatrix} \right\|_{\mathcal{B}^s} \leq C_F \epsilon^{11/2}$$

for $t \in [0, T_0 \epsilon^{-3}]$. The constant $C_F$ does not depend on $\epsilon$.

Finally, we must deal with initial conditions. Recall from the discussion in section 2 that it is typical to specify the initial data for the water wave problem in the long wavelength, small amplitude limit by

$$(35) \quad (\tilde{x}_{\alpha}(\tilde{\alpha}, 0), \tilde{y}(\tilde{\alpha}, 0), \tilde{u}(\tilde{\alpha}, 0)) = (0, \epsilon^2 \Theta_y(\epsilon \tilde{\alpha}), \epsilon^2 \Theta_u(\epsilon \tilde{\alpha})).$$

However, the above results are applicable if the initial data is of the form seen in (34). We eliminate this discrepancy by altering the initial parameterization of the free surface. What should this change be? Clearly,

$$(36) \quad \tilde{\alpha} = \alpha + x(\alpha, 0).$$

Now set $U(\beta, 0) = U_0(\beta)$, $V(\beta, 0) = V_0(\beta)$, $F(\beta, 0) = F_0(\beta)$, $G(\beta, 0) = G_0(\beta)$, and $P(\beta, 0) = 0$, and let

$$(37) \quad \tilde{\alpha} = \alpha + \int_0^\alpha \psi^d(a, 0) da = \alpha + \epsilon X_1(\epsilon \alpha) + \epsilon^3 X_2(\epsilon \alpha),$$

where

$$\epsilon X_1(\alpha) = -\epsilon^2 \int_0^\alpha (U_0(\epsilon a) + V_0(\epsilon a)) da,$$

$$\epsilon^3 X_2(\alpha) = -\epsilon^4 \int_0^\alpha (F_0(\epsilon a) + G_0(\epsilon a)) da.$$

With this definition, we clearly have satisfied the first condition in (34). We also want

$$\Theta_y(\epsilon \tilde{\alpha}) = \epsilon^2 \psi^y(\alpha) + O(\epsilon^{11/2}),$$

$$\Theta_u(\epsilon \tilde{\alpha}) = \epsilon^2 \psi^u(\alpha) + O(\epsilon^{11/2})$$

or rather

$$\epsilon^2 \Theta_y(\epsilon \alpha + \epsilon X_1(\epsilon \alpha) + \epsilon^3 X_2(\epsilon \alpha)) = \epsilon^2 \psi^y(\alpha) + O(\epsilon^{11/2}),$$

$$\epsilon^2 \Theta_u(\epsilon \alpha + \epsilon X_1(\epsilon \alpha) + \epsilon^3 X_2(\epsilon \alpha)) = \epsilon^2 \psi^u(\alpha) + O(\epsilon^{11/2}).$$
Applying Taylor’s theorem we have

\begin{align*}
\Theta_y + \epsilon^2 X_1 \Theta'_y &= (U_0 + V_0) + \epsilon^2 \left( \frac{1}{3} \partial_{\beta_+}^2 U_0 + \frac{1}{3} \partial_{\beta_-}^2 V_0 \right) \\
&\quad + \epsilon^2 (F_0 + G_0) + \epsilon^2 (U_0 + V_0)^2, \\
\Theta_u + \epsilon^2 X_1 \Theta'_u &= (U_0 - V_0) + \epsilon^2 \left( \frac{1}{6} \partial_{\beta_+}^2 U_0 - \frac{1}{6} \partial_{\beta_-}^2 V_0 \right) \\
&\quad + \epsilon^2 (F_0 - G_0) + \epsilon^2 \left( \frac{3}{4} U_0^2 - \frac{3}{4} V_0^2 \right). 
\end{align*}

We can solve the above by taking

\begin{align}
U_0 &= \frac{1}{2}(\Theta_y + \Theta_u), \\
V_0 &= \frac{1}{2}(\Theta_y - \Theta_u)
\end{align}

and

\begin{align}
F_0 &= \frac{1}{2}(h_y + h_u), \\
G_0 &= \frac{1}{2}(h_y - h_u),
\end{align}

where

\begin{align*}
h_y &= X_1 \Theta'_y - \frac{1}{3} \partial_{\beta_+}^2 U_0 - \frac{1}{3} \partial_{\beta_-}^2 V_0 - (U_0 + V_0)^2, \\
h_u &= X_1 \Theta'_u - \frac{1}{6} \partial_{\beta_+}^2 U_0 + \frac{1}{6} \partial_{\beta_-}^2 V_0 - \frac{3}{4} U_0^2 + \frac{3}{4} V_0^2.
\end{align*}

The functions \( U_0, V_0, F_0, \) and \( G_0 \) are all in \( H^s(4) \), and so the use of Taylor’s theorem is justified by Lemma 10. Thus we have proven the following lemma.

**Lemma 14.** Given initial conditions for the water wave equation in the form (35), define \( U_0, V_0, F_0, G_0, X_1, \) and \( X_2 \) as in (38), (39), and (37). Then the reparameterization of the initial profile given by

\[ \bar{\alpha} = \alpha + \epsilon X_1(\epsilon \alpha) + \epsilon X_2(\epsilon \alpha) \]

results in initial conditions given by (34).

**Remark 12.** Let \( \varphi^\pm(\beta_\pm, 0) = \varphi_0^\pm(\beta_\pm) \). Then this lemma will still be true if we replace \( F_0 \) with \( \varphi^- \) and \( G_0 \) with \( \varphi^+ \) and set \( F_0 \) and \( G_0 \) to be identically zero. That is, we have some choice in the way we select the initial conditions for the higher order equations.

Combining this lemma with Corollary 1 we prove Theorem 2. Thus we are done.

7. Assorted proofs.

**Proof of Lemma 8.** Let \( f^{-1}(\xi) = \xi - g_2(\xi) \). Since \( f^{-1}(f(\alpha)) = \alpha \) we have

\[ \alpha = f(\alpha) - g_2(f(\alpha)), \]

or rather

\[ g_2(f(\alpha)) = g(\alpha). \]

Notice that this relation implies \( \|g_2\|_{L^\infty} = \|g\|_{L^\infty} \). Taking a derivative, we have

\[ g'_2(f(\alpha)) = \frac{g'(\alpha)}{1 + g'(\alpha)}. \]
which implies that \( \|g_2\|_{L^\infty} \leq C\|g'\|_{L^\infty} \). If we expand the left-hand side of (40) by the mean value theorem, we see
\[
\begin{align*}
g_2(\alpha + g(\alpha)) &= g(\alpha), \\
g_2(\alpha) + g_2'(\alpha^*)g(\alpha) &= g(\alpha).
\end{align*}
\]
This implies \( g_2(\alpha) = g(\alpha) + O(\|g'\|_{L^\infty}\|g\|_{L^\infty}) \). Now, (40) can be rewritten and expanded using Taylor’s theorem,
\[
\begin{align*}
g_2(\alpha) &= g(\alpha), \\
g_2(\alpha) + g_2''(\alpha^*)(-g(\alpha) + O(\|g'\|_{L^\infty}\|g\|_{L^\infty})) &= g(\alpha).
\end{align*}
\]
which completes the proof.  

Proof of Lemma 9. Since \( (1 + \beta^2)^{n/2}F(\beta) \in H^s \), by the Sobolev embedding theorem there is a \( C \) such that
\[
F(\beta) \leq C(1 + \beta^2)^{-n/2}.
\]
So
\[
\int_{|\alpha| > C_0\epsilon^{-3}} |F(\epsilon \alpha)|^2 d\alpha \leq C \int_{|\alpha| > C_0\epsilon^{-3}} |1 + (\epsilon \alpha)^2|^{-n} d\alpha
\]
\[
\leq C \epsilon^{-2n} \int_{|\alpha| > C_0\epsilon^{-3}} |\alpha|^{-2n} d\alpha
\]
\[
\leq C \epsilon^{-2n} (\epsilon^{-3})^{-2n+1}
\]
\[
\leq C \epsilon^{4n-3}.
\]
The higher derivatives are bounded in exactly the same fashion. The extra powers of \( \epsilon \) come from the long wavelength scaling.

Proof of Lemma 1. The proof is a straightforward calculation:
\[
\begin{align*}
\|Af(\cdot) - A_n f(\cdot)\|_2^2 &= \int (1 + k^2)^n |\hat{A}(k) - \hat{A}_n(k)|^2 |\hat{f}(k)|^2 dk \\
&\leq C \int (1 + k^2)^n |\hat{f}(k)|^2 dk
\end{align*}
\]
\[
= C \int (1 + k^2)^n |\hat{f}(k)|^2 dk
\]
\[
= C \|\partial_x^n f(\cdot)\|_s^2.
\]
The proof for long wavelength data follows immediately from this.

Proof of Lemma 2. The fact that \( g(\epsilon x) \) is bounded as such in \( L^2 \) follows automatically from the geometric series approximation. That is, since \( |\epsilon^2 f(\epsilon x)| \leq 1/2 \), we know that
\[
|g(\epsilon x)| \leq C|\epsilon^2 f(\epsilon x)|^2.
\]
Thus we have
\[ \|g(\cdot)\|_{L^2}^2 \leq C \int e^4 |f(ex)|^4 \, dx \]
\[ \leq C e^4 \|f(\cdot)\|_{L^2}^2 \int f^2(\epsilon x) \, dx \]
\[ \leq C e^{7/2} \|f(\cdot)\|_s. \]

Now consider the $L^2$ norm of $g'(ex)$. A direct calculation shows that
\[ \frac{d}{dx}g(ex) = -\epsilon^3 f'(ex) \left( \frac{1}{(1 + \epsilon^2 f(ex))^2} - 1 \right). \]
Taylor’s theorem shows that
\[ \left| \frac{d}{dx}g(ex) \right| \leq C|\epsilon^3 f'(ex)||\epsilon^2 f(ex)|. \]

Therefore, just as before we have that this is bounded by $C \epsilon^{9/2}$ (which is of course bounded by $C \epsilon^{7/2}$).

We could keep on going in this fashion, showing each derivative of $g$ is bounded. This is, however, difficult as finding higher and higher derivatives is a notationally taxing job—see the expression of Faa di Bruno for proof of that!

Instead we take the following approach. Let
\[ h(y) = \frac{1}{1 + y} - 1 + y. \]
For $y \in [-1/2, 1/2]$, this function is real analytic and there exists another function $	ilde{h}(y)$ (real and analytic on the same interval) such that $h(y) = y \tilde{h}(y)$. Now, define $\tilde{h}_\epsilon(Y) = \tilde{h}(\epsilon^2 y)$. We have that
\[ \|\tilde{h}_\epsilon(\cdot)\|_{C^*[{-1/2\epsilon^{-2}, 1/2\epsilon^{-2}}]} \leq \|\tilde{h}(\cdot)\|_{C^*[{-1/2, 1/2}]} . \]

The point here is that the $C^*$ norm of $\tilde{h}_\epsilon$ can be bounded independently of $\epsilon$.

Now notice that $g(X) = \epsilon^2 f(X) \tilde{h}_\epsilon(f(X))$. Since $f \in H^{s+1}$, we know that $f \in C^s$. This implies that $\tilde{h}_\epsilon(f(X)) \in C^s$, with the $C^s$ norm bounded independent of $\epsilon$. Thus we have $f(X) \tilde{h}_\epsilon(f(X)) \in H^s$, with a norm bounded independent of $\epsilon$. With this in hand, we have that $\|g(\cdot)\|_s \leq C \epsilon^2$, with $C \neq C(\epsilon)$. Now the derivatives of $g$ can be bounded as follows:
\[ \left\| \frac{d^n}{dx^n} g(\cdot) \right\|_{L^2} = \epsilon^n \|g^{(n)}(\cdot)\|_{L^2} \]
\[ \leq C \epsilon^{n-1/2} \|g(\cdot)\|_s \]
\[ \leq C \epsilon^{n+3/2} . \]

Provided $n \geq 2$, this term is small enough. Therefore we have shown that the first $s$ derivatives are sufficiently small in $L^2$, and we have proved the lemma. \qed
Acknowledgments. As this work was done principally as the author’s dissertation, special thanks should go to Gene Wayne, who was the author’s thesis advisor, and to the Department of Mathematics and Statistics at Boston University, where he attended graduate school.

REFERENCES


