

# *Two-bubble dynamics for threshold solutions to the wave maps equation*

**Jacek Jendrej & Andrew Lawrie**

**Inventiones mathematicae**

ISSN 0020-9910

Volume 213

Number 3

Invent. math. (2018) 213:1249–1325

DOI 10.1007/s00222-018-0804-2

## *Inventiones mathematicae*

Vol. 213 Fasc. 3 2018

J. De Simoi, C. Liverani:	<b>Limit theorems for fast–slow partially hyperbolic systems</b>	811
D. Maulik, Y. Toda:	<b>Gopakumar–Vafa invariants via vanishing cycles</b>	1017
H. Bao, W. Wang:	<b>Canonical bases arising from quantum symmetric pairs</b>	1099
J. Huang:	<b>Commensurability of groups quasi-isometric to RAAGs</b>	1179
J. Jendrej, A. Lawrie:	<b>Two-bubble dynamics for threshold solutions to the wave maps equation</b>	1249
F. Loray, J.V. Pereira, F. Touzet:	<b>Singular foliations with trivial canonical class</b>	1327

Comprehensively covered by Zentralblatt MATH,  
Mathematical Reviews, and Current Contents

 Springer

 Springer

**Your article is protected by copyright and all rights are held exclusively by Springer-Verlag GmbH Germany, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](https://link.springer.com)".**

# Two-bubble dynamics for threshold solutions to the wave maps equation

Jacek Jendrej<sup>1</sup> · Andrew Lawrie<sup>2</sup>

Received: 9 June 2017 / Accepted: 29 April 2018 / Published online: 10 May 2018  
 © Springer-Verlag GmbH Germany, part of Springer Nature 2018

**Abstract** We consider the energy-critical wave maps equation  $\mathbb{R}^{1+2} \rightarrow \mathbb{S}^2$  in the equivariant case, with equivariance degree  $k \geq 2$ . It is known that initial data of energy  $< 8\pi k$  and topological degree zero leads to global solutions that scatter in both time directions. We consider the threshold case of energy  $8\pi k$ . We prove that the solution is defined for all time and either scatters in both time directions, or converges to a superposition of two harmonic maps in one time direction and scatters in the other time direction. In the latter case, we describe the asymptotic behavior of the scales of the two harmonic maps. The proof combines the classical concentration-compactness techniques of Kenig–Merle with a modulation analysis of interactions of two harmonic maps in the absence of excess radiation.

## 1 Introduction

This paper concerns energy critical wave maps  $\Psi : (\mathbb{R}_{t,x}^{1+2}, \mathbf{m}) \rightarrow (\mathcal{M}, \mathbf{g})$ , where  $\mathbf{m}$  is the Minkowski metric and  $\mathcal{M}$  is a Riemannian manifold with a

---

✉ Andrew Lawrie  
 alawrie@mit.edu

Jacek Jendrej  
 jendrej@math.univ-paris13.fr

<sup>1</sup> LAGA, UMR 7539, CNRS and Université Paris 13, 99 av J.-B. Clément, 93430 Villetaneuse, France

<sup>2</sup> Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Ave, 2-267, Cambridge, MA 02139, USA

metric  $\mathbf{g}$ . Wave maps arise in the physics literature as examples of nonlinear  $\sigma$ -models. A particularly interesting case is when the target manifold admits nontrivial finite energy stationary wave maps, or harmonic maps, as these give simple examples of topological (albeit unstable) solitons. Mathematically, wave maps simultaneously generalize the classical harmonic maps equation to Lorentzian domains as well as the free wave equation to manifold-valued maps.

Viewing  $(\mathcal{M}, \mathbf{g})$  as an isometrically embedded sub-manifold of Euclidean space  $(\mathbb{R}^N, \langle \cdot, \cdot \rangle_{\mathbb{R}^N})$ , a *wave map* is defined as a formal critical point of the Lagrangian action

$$\mathcal{L}(\Psi) = \frac{1}{2} \int_{\mathbb{R}^{1+2}} \mathbf{m}^{\alpha\beta} \langle \partial_\alpha \Psi, \partial_\beta \Psi \rangle_{\mathbb{R}^N} \, dx \, dt.$$

The Euler–Lagrange equations are given by

$$\square \Psi \perp T_\Psi \mathcal{M},$$

which can be rewritten as

$$\square \Psi = \mathcal{S}(\Psi)(\partial \Psi, \partial \Psi), \quad (1.1)$$

where  $\mathcal{S}$  denotes the second fundamental form of the embedding  $(\mathcal{M}, \mathbf{g}) \hookrightarrow (\mathbb{R}^N, \langle \cdot, \cdot \rangle)$ . The conserved energy is given by

$$\mathcal{E}(\Psi, \partial_t \Psi)(t) = \frac{1}{2} \int_{\mathbb{R}^2} |\partial_t \Psi(t)|^2 + |\nabla \Psi(t)|^2 \, dx = \text{constant}. \quad (1.2)$$

Smooth finite energy initial data for (1.1) consist of a pair  $\vec{\Psi}(0) = (\Psi_0, \Psi_1)$ , where

$$\Psi_0(x) \in \mathcal{M} \subset \mathbb{R}^N, \quad \Psi_1(x) \in T_{\Psi_0(x)} \mathcal{M}, \quad \forall x \in \mathbb{R}^2.$$

We assume here that we can find a fixed vector  $\Psi_\infty \in \mathcal{M}$  so that

$$\Psi_0(x) \rightarrow \Psi_\infty \text{ as } |x| \rightarrow \infty. \quad (1.3)$$

Wave maps on  $\mathbb{R}_{t,x}^{1+2}$  are called *energy critical* because the conserved energy and the equation are invariant under the same scaling: If  $\vec{\Psi}(t)$  solves (1.1) then so does

$$\vec{\Psi}_\lambda(t, x) := (\Psi_\lambda(t, x), \partial_t \Psi_\lambda(t, x)) := \left( \Psi(t/\lambda, x/\lambda), \frac{1}{\lambda} \partial_t \Psi(t/\lambda, x/\lambda) \right) \quad (1.4)$$

and it also holds that  $\mathcal{E}(\vec{\Psi}_\lambda) = \mathcal{E}(\vec{\Psi})$ .

The geometry of the target manifold, and in particular the existence of non-constant finite energy harmonic maps  $\Psi : \mathbb{R}^2 \rightarrow \mathcal{M}$ , plays a crucial role in determining the possible dynamics of solutions to the wave maps equation. Here we'll focus on a special case when the target manifold is the 2-sphere,  $\mathcal{M} = \mathbb{S}^2 \subset \mathbb{R}^3$  with the round metric  $\mathbf{g}$ . One advantage is that here the harmonic maps are explicit: by a classical theorem of Eells and Wood [20] they are either holomorphic or anti-holomorphic with respect to the complex structure on  $\mathbb{S}^2$ , and by (1.3) with the removable singularity theorem [47] they can thus be identified with the rational functions  $\rho : \mathbb{C}_\infty \rightarrow \mathbb{C}_\infty$ . It follows that each harmonic map  $\mathbb{R}^2 \rightarrow \mathbb{S}^2$  has a topological degree given by the degree of the corresponding rational map.

In fact, the condition (1.3) allows us to assign a topological degree to each smooth finite energy data. Given data  $(\Psi_0, \Psi_1)$ , we can identify  $\Psi_0$  with a map  $\tilde{\Psi}_0 : \mathbb{S}^2 \rightarrow \mathbb{S}^2$  by assigning the point at  $\infty$  to the vector  $\Psi_\infty := \lim_{|x| \rightarrow \infty} \Psi(x)$ . Abusing notation slightly by writing  $\tilde{\Psi}_0 = \Psi_0$ , the degree of the map  $\Psi_0$  is defined by

$$\deg(\Psi_0) := \frac{1}{\text{Area}(\mathbb{S}^2)} \int_{\mathbb{S}^2} \Psi_0^*(\omega) = \frac{1}{4\pi} \int_{\mathbb{R}^2} \Psi_0^*(\omega) \in \mathbb{Z}$$

where  $\omega$  is the area element of  $\mathbb{S}^2 \subset \mathbb{R}^3$ . The degree  $\deg(\Psi_0)$  is preserved by the smooth wave map flow on its maximal interval of existence  $I_{\max}$ , that is,

$$\deg(\Psi_0) = \deg(\Psi(t)) \quad \forall t \in I_{\max}.$$

Importantly, any harmonic map  $\mathcal{Q}_k$  of degree  $k$  minimizes the energy amongst all degree  $k$  wave maps, and in fact

$$\mathcal{E}(\mathcal{Q}_k) = 4\pi |\deg(\mathcal{Q}_k)| = 4\pi |k|.$$

## 1.1 $k$ -equivariant wave maps

To simplify the analysis we'll take advantage of a symmetry reduction and study a restricted class of maps  $\Psi$  satisfying the equivariance relation  $\Psi \circ \rho^k = \rho^k \circ \Psi$  for all rotations  $\rho \in SO(2)$ . We consider a subclass of such maps known as  $k$ -equivariant, or  $k$ -corotational, which correspond to equivariant maps that in local coordinates take the form

$$\Psi(t, r, \theta) = (\psi(t, r), k\theta) \hookrightarrow (\sin \psi \cos k\theta, \sin \psi \sin k\theta, \cos \psi) \in \mathbb{S}^2 \subset \mathbb{R}^3,$$

where  $\psi$  is the colatitude measured from the north pole of the sphere and the metric on  $\mathbb{S}^2$  is given by  $ds^2 = d\psi^2 + \sin^2 \psi d\omega^2$ . The Euler–Lagrange

equations (1.1) reduce to an equation for  $\psi$  and we are led to the Cauchy problem:

$$\begin{cases} \psi_{tt} - \psi_{rr} - \frac{1}{r}\psi_r + k^2 \frac{\sin 2\psi}{2r^2} = 0, \\ (\psi(0), \partial_t \psi(0)) = (\psi_0, \psi_1). \end{cases} \quad (1.5)$$

We'll often use the notation  $\vec{\psi}(t)$  to denote the pair

$$\vec{\psi}(t, r) := (\psi(t, r), \psi_t(t, r))$$

and we remark that the scaling (1.4) can be expressed as follows: If  $\vec{\psi}(t, r)$  is a solution to (1.5) then so is

$$\vec{\psi}_\lambda(t, r) = \left( \psi(t/\lambda, r/\lambda), \frac{1}{\lambda} \psi_t(t/\lambda, r/\lambda) \right)$$

for each fixed  $\lambda > 0$ .

The conserved energy from (1.2) takes the form

$$\mathcal{E}(\vec{\psi}(t)) = 2\pi \frac{1}{2} \int_0^\infty \left( (\partial_t \psi(t, r))^2 + (\partial_r \psi(t, r))^2 + k^2 \frac{\sin^2 \psi(t, r)}{r^2} \right) r \, dr.$$

From the above it's clear that any  $k$ -equivariant data  $\vec{\psi}(0, r)$  of finite energy must satisfy  $\lim_{r \rightarrow 0} \psi(0, r) = m\pi$  and  $\lim_{r \rightarrow \infty} \psi(0, \infty) = n\pi$  for some  $m, n \in \mathbb{Z}$ . Since the smooth wave map flow depends continuously on the initial data these integers are fixed over any time interval  $t \in I$  on which the solution is defined. This splits the energy space into disjoint classes according to this topological condition and it is natural to consider the Cauchy problem (1.5) within a fixed class

$$\mathcal{H}_{m\pi, n\pi}$$

$$:= \{(\psi_0, \psi_1) \mid \mathcal{E}(\psi_0, \psi_1) < \infty \text{ and } \lim_{r \rightarrow 0} \psi_0(r) = m\pi, \lim_{r \rightarrow \infty} \psi_0(r) = n\pi\}.$$

We can restrict to  $\mathcal{H}_{0, n\pi}$  and we'll denote these by  $\mathcal{H}_{n\pi} := \mathcal{H}_{0, n\pi}$ . We also define  $\mathcal{H} = \bigcup_{n \in \mathbb{Z}} \mathcal{H}_{n\pi}$  to be the full energy space.

The equivariant reduction introduces a good deal of rigidity into the problem, but still allows us access to the family of harmonic maps. Indeed the degree  $k$  harmonic map  $Q_k$  corresponding to  $z \mapsto z^k$  as a holomorphic map  $\mathbb{C}_\infty \rightarrow \mathbb{C}_\infty$ , can be expressed uniquely (up to scaling) as the  $|k|$ -equivariant map

$$Q_k(r, \theta) = (Q_k(r), k\theta) \hookrightarrow (\sin Q_k \cos k\theta, \sin Q_k \sin k\theta, \cos Q_k) \in \mathbb{S}^2 \subset \mathbb{R}^3,$$

where  $Q_k$  is the explicit finite energy stationary solution to (1.5) given by

$$Q_k(r) := 2 \arctan r^k.$$

Note that  $Q_k(r)$  satisfies

$$Q_k(0) = 0, \quad \lim_{r \rightarrow \infty} Q_k(r) = \pi.$$

We often write  $\vec{Q}_k := (Q_k, 0)$ . We see that  $\mathcal{E}(\vec{Q}_k) = 4\pi k$ , which is minimal amongst all  $k$ -equivariant maps in the energy class  $\mathcal{H}_\pi$ ; see Sect. 2.4 for a direct argument.

Here we consider  $k$ -equivariant maps  $\vec{\psi} = (\psi_0, \psi_1)$  in the class  $\mathcal{H}_0$ , i.e., that satisfy

$$\lim_{r \rightarrow 0} \psi_0(r) = 0 \quad \text{and} \quad \lim_{r \rightarrow \infty} \psi_0(r) = 0,$$

so that  $\psi_0$  is the polar angle of a finite energy map  $\Psi_0$  into  $\mathbb{S}^2$  with  $\deg(\Psi_0) = 0$ . Before stating our main results let us first motivate this restriction with a brief summary of recent developments.

## 1.2 Threshold theorems and bubbling

The energy critical wave maps Eq. (1.1) has been extensively studied over the past several decades; [5, 6, 28–34, 48, 49, 53, 54, 56, 57]. In recent years the focus has centered on understanding the nonlinear dynamics of solutions with large energy. At the end of the last decade, the following remarkable *sub-threshold conjecture* was established [38, 50, 51, 55]: Every wave map with energy less than that of the first nontrivial harmonic map is globally regular on  $\mathbb{R}^{1+2}$  and scatters to a constant map. The role of the least energy harmonic map in the statement of the sub-threshold conjecture is based on fundamental work of Struwe [52], who showed that the smooth equivariant wave map flow can only develop a singularity by concentrating energy at the tip of a light cone by bubbling off at least one non-trivial finite energy harmonic map. In breakthrough works, Krieger et al. [39], Rodnianski and Sterbenz [46], Raphaël and Rodnianski [44], constructed examples solutions of such blow-up by bubbling, with the latter two works yielding a stable blow-up regime; see also the recent stability analysis of Krieger [35] for type-II blow ups solutions to the energy critical NLW, which suggests that the solutions from [39] should also exhibit stability properties.

The starting point for the present work is the following natural question: Can one give a satisfactory description of the possible dynamics for arbitrary initial data? In dispersive models such as (1.1) this is typically referred to as the *soliton*

*resolution conjecture*, which states roughly that any smooth solution asymptotically decouples into weakly interacting (possibly concentrating) solitons plus free radiation. The wave maps Eq. (1.1) with  $\mathcal{N} = \mathbb{S}^2$  is an intriguing model in which to study this question: all stationary solutions (the harmonic maps) are known explicitly, the conserved topological degree of the solution introduces additional rigidity, and the equivariant reduction (1.5) greatly simplifies certain aspects of the analysis without destroying the essential mechanisms of truly nonlinear behavior, e.g., solitons, blow-up. There has been exciting recent progress in this direction for the general Eq. (1.1), see [12, 21]. Here we focus on the equivariant model (1.5) where more is known.

Our analysis is motivated by several results proved in the last few years; we will in fact use some of them explicitly. First, note that, by continuity,  $\lim_{r \rightarrow 0} \psi(t, r)$  and  $\lim_{r \rightarrow \infty} \psi(t, r)$  are independent of  $t$ . Hence scattering to a constant map is only possible if  $\lim_{r \rightarrow 0} \psi_0(r) = \lim_{r \rightarrow \infty} \psi_0(r)$ . We can assume without loss of generality that both these limits equal 0, i.e. the initial data  $(\psi_0, \psi_1)$  is in  $\mathcal{H}_0$ . For such maps, the following refined threshold theorem was proved in [9].

**Theorem 1.1** ( $2\mathcal{E}(\vec{Q})$  Threshold Theorem). [9, Theorem 1.1] *For any smooth initial data  $\vec{\psi}(0) \in \mathcal{H}_0$  with*

$$\mathcal{E}(\vec{\psi}(0)) < 2\mathcal{E}(\vec{Q}_k) = 8\pi k,$$

*there exists a unique global evolution  $\vec{\psi} \in C^0(\mathbb{R}; \mathcal{H}_0)$ . Moreover,  $\vec{\psi}(t)$  scatters to zero in both time directions, i.e., there exist solutions  $\vec{\varphi}_L^\pm$  to the linearized Eq. (2.2) such that*

$$\vec{\psi}(t) = \vec{\varphi}_L^\pm(t) + o_{\mathcal{H}_0}(1) \text{ as } t \rightarrow \pm\infty.$$

The analogous result for the full model without symmetries was obtained by the second author and Oh in [40], as a consequence of the bubbling analysis in [51]. The heuristic reasoning behind the threshold  $2\mathcal{E}(\vec{Q})$  is as follows. The topological degree counts (with orientation) the number of times a map ‘wraps around’  $\mathbb{S}^2$ . If a harmonic map of degree  $k$  bubbles off from a wave map  $\vec{\psi}(t)$ , then, in order for  $\vec{\psi}(t)$  to satisfy  $\deg(\psi) = 0$ , it must ‘unwrap’ precisely  $k$  times away from the bubble. The minimum energy cost for wrapping and unwrapping is  $4\pi k$ , which is also the energy of  $Q_k$ . The total energy cost is at least  $8\pi k = 2\mathcal{E}(\vec{Q}_k)$ .

Similar intuition motivated the works [9, 10], which established soliton resolution for 1-equivariant maps with energies that only allow for one concentrating bubble, namely for data in  $\mathcal{H}_\pi$  with energy below  $3\mathcal{E}(Q)$ . These works showed that for any such solution there exists a regular map  $\vec{\varphi} \in \mathcal{H}_0$  (free radiation if the solution is global) and a continuous dynamical scale  $\lambda(t) \in [0, \infty)$

such that

$$\vec{\psi}(t) = \vec{Q}_{\lambda(t)} + \vec{\varphi}(t) + o_{\mathcal{H}_0}(1) \text{ as } t \rightarrow T_+. \quad (1.6)$$

Cote [8] and later Jia and Kenig [25] extended the theory to handle arbitrary energies, the latter work in all equivariance classes. It was shown that in this case the decomposition (1.6) holds with possibly many concentrating harmonic maps, but only along at least one *sequence of times*  $t_n \rightarrow T_+$ . The proofs of [8–10, 25] rely heavily on concentration compactness techniques and were all inspired by the remarkable series of papers by Duyckaerts et al. [14–17] on the focusing quintic nonlinear wave equation in 3 space dimensions. We'll discuss these latter works more below; see Remark 1.13.

**Theorem 1.2** (Sequential Decomposition). [8, 25] *Let  $\vec{\psi}(t) \in \mathcal{H}_{\ell\pi}$  be a smooth solution to (1.5) on  $[0, T_+)$ . Then there exists a sequence of times  $t_n \rightarrow T_+$ , an integer  $J \in \mathbb{N}$ , a regular map  $\vec{\varphi} \in \mathcal{H}_0$ , sequences of scales  $\lambda_{n,j}$  and signs  $\iota_j \in \{-1, 1\}$  for  $j \in \{1, \dots, J\}$ , so that*

$$\vec{\psi}(t_n) = \sum_{j=1}^J \iota_j \vec{Q}_{\lambda_{n,j}} + \vec{\varphi}(t_n) + o_{\mathcal{H}_0}(1) \text{ as } n \rightarrow \infty \quad (1.7)$$

*In the case of finite time blow-up at least one scale  $\lambda_{n,1} \rightarrow 0$  as  $n \rightarrow \infty$  and  $\vec{\varphi}(t) \rightarrow \vec{\varphi}(1)$  is a finite energy map with  $\mathcal{E}(\vec{\varphi}(1)) = \mathcal{E}(\vec{\psi}) - J\mathcal{E}(\vec{Q})$ . In the case of a global solution,  $\vec{\varphi}(t)$  can be taken to be a solution to the linear wave Eq. (2.2) and signs  $\iota_j$  are required to match up so that*

$$\lim_{r \rightarrow \infty} \vec{\psi}(0, r) = \ell\pi = \lim_{r \rightarrow \infty} \sum_{j=1}^J \iota_j \vec{Q}_{\lambda_{n,j}}(r).$$

**Remark 1.3** A decomposition into bubbles for a sequence of times for the full non-equivariant model was obtained by Grinis [21] up to an error that vanishes in a weaker Besov-type norm. Duyckaerts et al. [12] proved that for energies slightly above  $\mathcal{E}(\vec{Q})$  a one-bubble decomposition holds for continuous time. The same authors obtained in [13] a sequential decomposition into bubbles in the case of the focusing energy critical power-type nonlinear wave equation (NLW).

Theorem 1.2 raises two natural questions:

- Are there any solutions to (1.5) with  $J \geq 2$  in (1.7), i.e., are there any solutions that form more than one bubble?
- And, if so, does the decomposition (1.7) hold continuously in time, i.e., does soliton resolution hold for (1.5)?

In view of Theorems 1.1 and 1.2 it is natural to ask both questions at the minimal possible energy level where multiple bubble dynamics can occur, namely for solutions  $\vec{\psi}(t) \in \mathcal{H}_0$  having *threshold energy*, that is such that

$$\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(\vec{Q}).$$

In [22] the first author obtained an affirmative answer to the first question, proving the following result.

**Theorem 1.4** [22, Theorem 2] *Let  $k > 2$ . There exists a solution  $\vec{\psi} : (-\infty, T_0] \rightarrow \mathcal{H}_0$  of (1.5) such that*

$$\lim_{t \rightarrow -\infty} \|\vec{\psi}(t) - (-\vec{Q} + \vec{Q}_{\gamma_k |t|^{-\frac{2}{k-2}}})\|_{\mathcal{H}_0} = 0,$$

where  $\gamma_k > 0$  is an explicit constant depending on  $k$ . □

*Remark 1.5* Similar solutions could be obtained for  $k = 2$  by the same method.

### 1.3 Main result

In this paper, we address the problem of *classification* of solutions at threshold energy level, in the spirit of the works of Duyckaerts and Merle [18, 19]. The major difficulty in the analysis is that in our case the threshold solutions contain two bubbles, which leads to significantly more complicated dynamics.

Let  $\vec{\psi}(t) : (T_-, T_+) \rightarrow \mathcal{H}_0$  be a solution to (1.5) with  $\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(\vec{Q})$ . We will say that  $\vec{\psi}(t)$  is a *two-bubble in the forward time direction* if there exist  $\iota \in \{1, -1\}$  and continuous functions  $\lambda(t), \mu(t) > 0$  such that

$$\lim_{t \rightarrow T_+} \|(\psi(t) - \iota(Q_{\lambda(t)} - Q_{\mu(t)}), \psi_t(t))\|_{\mathcal{H}_0} = 0, \quad \lambda(t) \ll \mu(t) \text{ as } t \rightarrow T_+.$$

The notion of a *two-bubble in the backward time direction* is defined similarly. We prove the following result.

**Theorem 1.6** (Main Theorem). *Fix any equivariance class  $k \geq 2$ . Let  $\vec{\psi}(t) : (T_-, T_+) \rightarrow \mathcal{H}_0$  be a solution to (1.5) such that*

$$\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(\vec{Q}) = 8\pi k.$$

*Then  $T_- = -\infty$ ,  $T_+ = +\infty$  and one the following alternatives holds:*

- $\vec{\psi}(t)$  scatters in both time directions,
- $\vec{\psi}(t)$  scatters in one time direction and is a two-bubble in the other time direction with the scales of the bubbles  $\lambda(t), \mu(t)$  satisfying

$$\mu(t) \rightarrow \mu_0 \in (0, +\infty), \quad \lambda(t) \rightarrow 0.$$

**Remark 1.7** As a by-product of the proof, we will determine the rate of decay of  $\lambda(t)$  in the two-bubble case. Suppose a two-bubble solution forms as  $t \rightarrow \infty$ : If  $k \geq 3$  there exists a constant  $C_k > 0$  such that  $\frac{1}{C_k} \mu_0 t^{-\frac{2}{k-2}} \leq \lambda(t) \leq C_k \mu_0 t^{-\frac{2}{k-2}}$  for  $t$  large enough, see (4.34). In the case  $k = 2$  there exists a constant  $C > 0$  such that we have  $\exp(-Ct) \leq \lambda(t) \leq \exp(-t/C)$  for  $t$  large enough, see (4.33).

**Remark 1.8** In particular, the two-bubble solutions from Theorem 1.4 scatter in forward time, which provides an example of an orbit connecting different types of dynamical behavior for positive and negative times.

**Remark 1.9** Non-existence of solutions which form a pure two-bubble in both time directions is reminiscent of the work of Martel and Merle [41, 42]. This seems to be a typical feature of models which are not completely integrable.

One of the main points of our paper is an analysis of what we could call a *collision of bubbles* in the simplest possible case of threshold energy.

**Remark 1.10** Recall that in [18] a *complete classification* at the threshold energy was obtained. It is tempting to believe that the solutions from Theorem 1.4 should play a similar role as the solution  $W^-$  from [18], in which case they would be unique non-dispersive solutions up to rescaling. This remains an open question.

**Remark 1.11** We conjecture that for  $k = 1$  a similar result holds, but in the two-bubble case  $\lambda(t) \rightarrow 0$  in finite time. The slower decay of  $Q$  would be a source of additional technical difficulties in Section 3, but the general scheme could be applied without major changes.

**Remark 1.12** Our method establishes the exact analog of Theorem 1.6 in the case of the equivariant Yang-Mills equation, by making the usual analogy between equivariant Yang-Mills and  $k = 2$ -equivariant wave maps, see for example [9, Appendix] for the analog of the Threshold Theorem 1.1 and [22] for the analog of Theorem 1.4. There the harmonic map  $Q$  is replaced by the first instanton, the notion of topological degree is replaced by the second Chern number, and the threshold energy is exactly twice the energy of the first instanton.

**Remark 1.13** The full soliton resolution conjecture was established for the radial solutions of focusing energy critical NLW by Duyckaerts, Kenig, and Merle in the landmark work [17]. This result is the only known case of a complete *continuous-in-time* classification for a model that is not completely integrable. The proof relies on a particularly strong form of the “channels of energy” method introduced by the same authors. However, proving channel of energy estimates in other settings is a delicate issue, and the strong form of

these estimates used in [17] is known to fail for the linear wave equation in even dimensions, see [11].

Aside from [17], Theorem 1.6 is the only other classification result for a dispersive equation that holds for continuous times in the presence of more than one non-trivial elliptic profile. Upgrading sequential decompositions such as Theorem 1.2 or the one in [13] to hold for continuous times is regarded as a major open problem.

## 1.4 Structure of the proof

Inspired by the work of Duyckaerts and Merle [18], we merge the concentration-compactness techniques with a careful analysis of the *modulation equations* governing the evolution of the scales  $\lambda(t)$  and  $\mu(t)$ . As mentioned above, the main difference with respect to [18] consists in the fact that our threshold solutions contain two bubbles, one of which is concentrating, whereas in [18] the modulation happens essentially around one stationary bubble. Thus our analysis requires substantially new technique. Our proof can be summarized as follows.

*Step 1* If the solution does not scatter, then, by a special case of Theorem 1.2, it approaches a two-bubble configuration for a sequence of times.

*Step 2* We divide the time axis into regions where the solution is close to a two-bubble configuration, which we can call the *bad intervals*  $[a_m, b_m]$ , and regions where it is not, which are the *good intervals*  $[b_m, a_{m+1}]$ .

*Step 3* On a bad interval  $[a_m, b_m]$ , we decompose the solution as follows:

$$\vec{\psi}(t) = (-Q_{\mu(t)} + Q_{\lambda(t)} + g(t), \partial_t \psi(t)).$$

In order to specify the values of  $\lambda(t)$  and  $\mu(t)$ , we use suitable orthogonality conditions, see Lemma 3.1. For technical reasons, we introduce a parameter  $\zeta(t)$  such that  $|\zeta(t) - \lambda(t)| \ll \lambda(t)$ . We consider  $c_m \in (a_m, b_m)$  where the quantity  $\zeta(t)/\mu(t)$  attains its global minimum on  $[a_m, b_m]$  (we make sure that the minimum is not attained at one of the endpoints).

The orthogonality conditions yield modulation equations for the evolution of  $\zeta(t)$  and  $\mu(t)$ . From these equations we can deduce crucial information about the behavior of  $\mu(t)$  and  $\zeta(t)$  for  $t \geq c_m$  and  $t \leq c_m$ . We refer to the beginning of Sect. 3 for a short description of the method. The main conclusion can be intuitively phrased as follows:  $\mu(t)$  does not change much on a bad interval, whereas  $\zeta(t)$  grows in a controlled way both for  $t \geq c_m$  and  $t \leq c_m$ . The decisive point is that the bad interval  $[a_m, b_m]$  can be long if  $\zeta(c_m)/\mu(c_m)$  is small, but

$$\int_{a_m}^{b_m} \left( \frac{\zeta(t)}{\mu(t)} \right)^{\frac{k}{2}} dt \leq C_k, \quad C_k \text{ depending only on } k. \quad (1.8)$$

Note that the only information about the solution which is used in this process is the fact that  $\mathcal{E}(\vec{\psi}(t)) = 2\mathcal{E}(\vec{Q})$ ,  $\vec{\psi}(c_m)$  is close to a two-bubble configuration and

$$\left. \frac{d}{dt} \right|_{t=c_m} (\zeta(t)/\mu(t)) = 0.$$

*Step 4* Using concentration-compactness arguments and Theorem 1.1 we obtain that the solution has the *compactness property* on the union of the good intervals. Now the idea is to run a convexity argument based on a monotonicity formula between two times where  $\vec{\psi}(t)$  is close to a two-bubble. It is at this stage that we reach a contradiction – if the solution has exited a neighborhood of two-bubble configurations during the interim, the total cost in terms of time derivative is too great to allow it to return. This is a type of *no-return* result and one can draw parallels here to the ignition and ejection lemmas from the work of Krieger et al. [36, 37] concerning near ground-state dynamics for the energy critical NLW.

There can potentially be many good and bad intervals between the two times where  $\vec{\psi}(t)$  is close to a two-bubble. It is well-known that one needs to use a cut-off in the monotonicity formula, which introduces an error in the estimates. On the good intervals, this error is controlled thanks to the compactness property. On the bad intervals, the bound (1.8) comes into play. More precisely, we obtain that the error on a bad interval is absorbed by positive terms obtained on intervals preceding and following the bad interval.

*Step 5* Once the convergence to a two-bubble for continuous time is proved, we deduce easily from the modulation equations that the solution is global and  $\mu(t) \rightarrow \mu_0 \in (0, +\infty)$ . Scattering on at least one side follows easily from the previous analysis. Namely, if the solution is non-scattering in both time directions, then the time axis is divided into two bad regions near  $\pm\infty$  and one good interval in between. We reach a contradiction by a similar (but simpler) argument as in Step 4.

## 2 Preliminaries and technical lemmas

In this section we establish a few preliminary facts about solutions to (1.5) that will be required in our analysis. We first aggregate here some notation.

### 2.1 Notation

Given a radial function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  we'll abuse notation and simply write  $f = f(r)$ , where  $r = |x|$ . We'll also drop the factor  $2\pi$  in our notation for the

$L^2$  pairing of radial functions on  $\mathbb{R}^2$

$$\langle f | g \rangle := \frac{1}{2\pi} \langle f | g \rangle_{L^2(\mathbb{R}^2)} = \int_0^\infty f(r)g(r) r dr$$

Recall the definition of the space  $\mathcal{H}_0$ :

$$\mathcal{H}_0 := \left\{ (\psi_0, \psi_1) \mid \mathcal{E}(\psi_0, \psi_1) < \infty, \quad \lim_{r \rightarrow 0} \psi_0(r) = \lim_{r \rightarrow \infty} \psi_0(r) = 0 \right\}$$

We define a norm  $H$  by

$$\|\psi_0\|_H^2 := \int_0^\infty \left( (\partial_r \psi_0(r))^2 + k^2 \frac{(\psi_0(r))^2}{r^2} \right) r dr$$

and for pairs  $\vec{\psi} = (\psi_0, \psi_1) \in \mathcal{H}_0$  we write

$$\|\vec{\psi}\|_{\mathcal{H}_0} := \|(\psi_0, \psi_1)\|_{H \times L^2}.$$

The change of variables  $r \mapsto e^x$  gives us an identification between the radial functions  $H(\mathbb{R}^2)$  and  $H^1(\mathbb{R})$ , i.e.,  $\psi_0(r) \in H \Leftrightarrow \psi_0(e^x) \in H^1(\mathbb{R})$ . In particular this means that

$$\|\psi_0\|_{L^\infty} \leq C \|\psi_0\|_H$$

Scaling invariance plays a key role in our analysis. Given a radial function  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  we denote the  $\dot{H}^1$  and  $L^2$  re-scalings as follows

$$\phi_\lambda(r) = \phi(r/\lambda), \quad \phi_{\underline{\lambda}}(r) = \frac{1}{\lambda} \phi(r/\lambda)$$

The corresponding infinitesimal generators are given by

$$\begin{aligned} \Lambda \phi &:= -\frac{\partial}{\partial \lambda} \Big|_{\lambda=1} \phi_\lambda = r \partial_r \phi \quad (\dot{H}_{\text{rad}}^1(\mathbb{R}^2) \text{ scaling}) \\ \Lambda_0 \phi &:= -\frac{\partial}{\partial \lambda} \Big|_{\lambda=1} \phi_{\underline{\lambda}} = (1 + r \partial_r) \phi \quad (L_{\text{rad}}^2(\mathbb{R}^2) \text{ scaling}) \end{aligned} \quad (2.1)$$

## 2.2 Review of the Cauchy theory

For initial data  $(\varphi_0, \varphi_1)$  in the class  $\mathcal{H}_0$  the formulation of the Cauchy problem (1.5) can be modified to take into account the strong repulsive potential term in the nonlinearity:

$$\frac{k^2 \sin(2\phi)}{2r^2} = \frac{k^2}{r^2} \phi + \frac{k^2}{2r^2} (\sin(2\phi) - 2\phi) = \frac{k^2}{r^2} \phi + \frac{O(\phi^3)}{r^2}$$

The presence of the potential  $\frac{k^2}{r^2}$  indicates that the linear wave equation,

$$\left( \partial_t^2 - \Delta_{\mathbb{R}^2} + \frac{k^2}{r^2} \right) \psi = 0, \quad (2.2)$$

of (1.5) has more dispersion than the  $2d$  wave equation. In fact, it has the same dispersion as the free wave equation in dimension  $d = 2k + 2$  as can be seen from the following change of variables: given a radial function  $\phi \in H$ , define  $v(r)$  by  $\phi(r) = r^k v(r)$ . Then

$$\frac{1}{r^k} \left( -\Delta_{\mathbb{R}^2} + \frac{k^2}{r^2} \right) \phi = -\Delta_{\mathbb{R}^{2k+2}} v, \quad \|\phi\|_H = \|v\|_{\dot{H}^1(\mathbb{R}^{2k+2})}. \quad (2.3)$$

Thus one way of studying solutions  $\vec{\psi}(t) \in \mathcal{H}_0$  of Cauchy problem (1.5) is to define  $\vec{v}(t) = (r^{-k}\psi(t), r^{-k}\psi_t(t)) \in \dot{H}^1 \times L^2(\mathbb{R}^{2k+2})$  and analyze the equivalent Cauchy problem for the radial nonlinear wave equation in  $\mathbb{R}_{t,x}^{1+(2k+2)}$  satisfied by  $\vec{v}(t)$ . Unfortunately, this route leads to unpleasant technicalities when  $k > 2$  (spatial dimension  $= 2k + 2 > 6$ ) due to the high dimension and the particular structure of the nonlinearity.

There is a simpler approach that allows us to treat the scattering theory for the Cauchy problem (1.5) for all equivariance classes  $k \geq 1$  in a unified fashion. The idea is to make use of some, but not all, of the extra dispersion in  $-\Delta_{\mathbb{R}^2} + k^2/r^2$ . Indeed, given a solution  $\vec{\psi}(t)$  to (1.5) we define  $u$  by  $ru = \psi$  and obtain the following Cauchy problem for  $u$ .

$$u_{tt} - u_{rr} - \frac{3}{r}u_r + \frac{k^2 - 1}{r^2}u = k^2 \frac{2ru - \sin(2ru)}{2r^3} =: Z(ru)u^3$$

$$\vec{u}(0) = (u_0, u_1). \quad (2.4)$$

where the function  $Z$  defined above is a clearly smooth, bounded, even function. The linear part of (2.4) is the radial wave equation in  $\mathbb{R}^{1+4}$  with a *repulsive* inverse square potential, namely

$$v_{tt} - v_{rr} - \frac{3}{r}v_r + \frac{k^2 - 1}{r^2}v = 0. \quad (2.5)$$

For each  $k \geq 1$ , define the norm  $H_k$  for radially symmetric functions  $v$  on  $\mathbb{R}^4$  by

$$\|v\|_{H_k(\mathbb{R}^4)}^2 := \int_0^\infty \left[ (\partial_r v)^2 + \frac{(k^2 - 1)}{r^2} v^2 \right] r^3 dr$$

Solutions to (2.5) conserve the  $H_k$  norms. By Hardy's inequality we have

$$\|v\|_{H_k(\mathbb{R}^4)} \simeq \|v\|_{\dot{H}^1(\mathbb{R}^4)}$$

The mapping,

$$H_k \times L^2(\mathbb{R}^4) \ni (u_0, u_1) \mapsto (\psi_0, \psi_1) := (ru_0, ru_1) \in H \times L^2(\mathbb{R}^2)$$

satisfies

$$\|(u_0, u_1)\|_{\dot{H}^1 \times L^2(\mathbb{R}^4)} \simeq \|(u_0, u_1)\|_{H_k \times L^2(\mathbb{R}^4)} = \|(\psi_0, \psi_1)\|_{H \times L^2(\mathbb{R}^2)} \quad (2.6)$$

Thus we can conclude that the Cauchy problem for (2.4) with initial data in  $\dot{H}^1 \times L^2(\mathbb{R}^4)$  is equivalent to the Cauchy problem for (1.5) for initial data  $(\psi_0, \psi_1) \in \mathcal{H}_0$ , allowing us to give a scattering criterion for solutions  $\vec{\psi}(t) \in \mathcal{H}_0$  to (1.5).

**Lemma 2.1** *Let  $\vec{\psi}(0) = (\psi_0, \psi_1) \in \mathcal{H}_0$ . Then there exists a unique solution  $\vec{\psi}(t) \in \mathcal{H}_0$  to (1.5) defined on a maximal interval of existence  $I_{\max}(\vec{\psi}) := (-T_-(\vec{\psi}), T_+(\vec{\psi}))$  with the following properties: Define*

$$\vec{u}(t, r) = (r^{-1}\psi(t, r), r^{-1}\psi_t(t, r)) \in \dot{H}^1 \times L^2(\mathbb{R}^4)$$

*Then for any compact time interval  $J \Subset I_{\max}$  we have*

$$\|u\|_{L_t^3(J; L_x^6(\mathbb{R}^4))} \leq C(J) < \infty$$

*In addition, if*

$$\|u\|_{L_t^3([0, T_+(\vec{\psi})); L_x^6(\mathbb{R}^4))} < \infty$$

*then  $T_+ = \infty$  and  $\vec{\psi}(t)$  scatters  $t \rightarrow \infty$ , i.e., there exists a solution  $\vec{\phi}_L(t) \in \mathcal{H}_0$  to (2.2) so that*

$$\|\vec{\psi}(t) - \vec{\phi}_L(t)\|_{\mathcal{H}_0} \rightarrow 0 \text{ as } t \rightarrow \infty.$$

*Conversely, any solution  $\vec{\psi}(t)$  that scatters as  $t \rightarrow \infty$  satisfies*

$$\|\psi/r\|_{L_t^3 L_x^6([0, \infty) \times \mathbb{R}^4)} < \infty.$$

The proof of Lemma 2.1 is standard consequence of Strichartz estimates for (2.5) and the equivalence of the Cauchy problems (1.5) and (2.4). In this case, we need Strichartz estimates for the radial wave equation in  $\mathbb{R}^{1+4}$  with a repulsive inverse square potential. For these we can cite the more general results of Planchon et al. [43]; see also [2, 3] which cover the non-radial case.

**Lemma 2.2** (Strichartz estimates). [43, Corollary 3.9] *Fix  $k \geq 1$  and let  $\vec{v}(t)$  be a radial solution to the linear equation*

$$v_{tt} - v_{rr} - \frac{3}{r}v_r + \frac{k^2 - 1}{r^2}v = F(t, r), \quad \vec{v}(0) = (v_0, v_1) \in \dot{H}^1 \times L^2(\mathbb{R}^4)$$

Then, for any time interval  $I \subset \mathbb{R}$  we have

$$\|v\|_{L_t^3 L_x^6(I \times \mathbb{R}^4)} + \sup_{t \in I} \|\vec{v}(t)\|_{\dot{H}^1 \times L^2(\mathbb{R}^4)} \lesssim \|\vec{v}(0)\|_{\dot{H}^1 \times L^2(\mathbb{R}^4)} + \|F\|_{L_t^1, L_x^2(I \times \mathbb{R}^4)}$$

where the implicit constant above is independent of  $I$ .

We'll also explicitly require the following nonlinear perturbation lemma from [27]; see also [9, Lemma 2.18].

**Lemma 2.3** (Perturbation Lemma). [27, Theorem 2.20] [9, Lemma 2.18] *There are continuous functions  $\varepsilon_0, C_0 : (0, \infty) \rightarrow (0, \infty)$  such that the following holds: Let  $I \subset \mathbb{R}$  be an open interval, (possibly unbounded),  $\psi, \varphi \in C^0(I; H) \cap C^1(I; L^2)$  radial functions satisfying for some  $A > 0$*

$$\begin{aligned} \|\vec{\psi}\|_{L^\infty(I; H \times L^2(\mathbb{R}^2))} + \|\vec{\varphi}\|_{L^\infty(I; H \times L^2(\mathbb{R}^2))} + \|\varphi/r\|_{L_t^3(I; L_x^6(\mathbb{R}^4))} &\leq A \\ \|eq(\psi/r)\|_{L_t^1(I; L_x^2(\mathbb{R}^4))} + \|eq(\varphi/r)\|_{L_t^1(I; L_x^2(\mathbb{R}^4))} + \|w_0\|_{L_t^3(I; L_x^6(\mathbb{R}^4))} &\leq \varepsilon \leq \varepsilon_0(A) \end{aligned}$$

where  $eq(\psi/r) := (\square_{\mathbb{R}^4} + \frac{k^2-1}{r^2})(\psi/r) + (\psi/r)^3 Z(\psi)$  in the sense of distributions, and  $\vec{w}_0(t) := S(t-t_0)(\vec{\psi} - \vec{\varphi})(t_0)$  with  $t_0 \in I$  arbitrary, but fixed and  $S$  denoting the linear wave evolution operator in  $\mathbb{R}^{1+4}$  (i.e., the propagator for (2.2)). Then,

$$\|\vec{\psi} - \vec{\varphi} - \vec{w}_0\|_{L_t^\infty(I; H \times L^2(\mathbb{R}^2))} + \left\| \frac{1}{r}(\psi - \varphi) \right\|_{L_t^3(I; L_x^6(\mathbb{R}^4))} \leq C_0(A)\varepsilon$$

In particular,  $\|\psi/r\|_{L_t^3(I; L_x^6(\mathbb{R}^4))} < \infty$ .

## 2.3 Concentration compactness

Another consequence of (2.6) and Lemma 2.2 is that we can translate the concentration compactness theory of Bahouri and Gérard to solutions to (2.2) and (1.5). We begin by stating the linear profile decompositions in the 4d setting for solutions to (2.5).

**Lemma 2.4** (Linear 4d profile decomposition). [1, Main Theorem] *Let  $k \geq 1$  be fixed. Consider a sequence  $\vec{u}_n = (u_{n,0}, u_{n,1}) \in H_k \times L^2(\mathbb{R}^4)$  which is bounded in the sense that  $\|\vec{u}_n\|_{H_k \times L^2(\mathbb{R}^4)} \lesssim 1$ . Then, up to passing to a subsequence, there exists a sequence of solutions to (2.5),  $\vec{V}_L^j \in H_k \times L^2(\mathbb{R}^4)$ , sequences of times  $\{t_{n,j}\} \subset \mathbb{R}$ , and sequences of scales  $\{\lambda_{n,j}\} \subset (0, \infty)$ , and  $\vec{w}_n^k$  defined by*

$$\begin{aligned} \vec{u}_n(r) = & \sum_{j=1}^k \left( \frac{1}{\lambda_{n,j}} V_L^j \left( \frac{-t_{n,j}}{\lambda_{n,j}}, \frac{r}{\lambda_{n,j}} \right), \frac{1}{(\lambda_{n,j})^2} \partial_t V_L^j \left( \frac{-t_{n,j}}{\lambda_{n,j}}, \frac{r}{\lambda_{n,j}} \right) \right) \\ & + (w_{n,0}^k, w_{n,1}^k)(r) \end{aligned}$$

so that the following statements hold: Let  $w_{n,L}^k(t)$  denote the linear evolution of the data  $\vec{w}_n^k$ , i.e., solutions to (2.5). Then, for any  $j \leq k$ ,

$$(\lambda_n^j w_{n,L}^k(t_{n,j}, \lambda_{n,j} \cdot), \lambda_{n,j}^2 w_n^k(t_{n,j}, \lambda_{n,j} \cdot)) \rightarrow 0 \text{ weakly in } H_k \times L^2(\mathbb{R}^4).$$

In addition, for any  $j \neq k$  we have

$$\frac{\lambda_{n,j}}{\lambda_{n,k}} + \frac{\lambda_{n,k}}{\lambda_{n,j}} + \frac{|t_{n,j} - t_{n,k}|}{\lambda_{n,j}} + \frac{|t_{n,j} - t_{n,k}|}{\lambda_{n,k}} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Moreover, the errors  $\vec{w}_n^k$  vanish asymptotically in the sense that

$$\limsup_{n \rightarrow \infty} \|w_{n,L}^k\|_{L_t^\infty L_x^4 \cap L_t^3 L_x^6(\mathbb{R} \times \mathbb{R}^4)} \rightarrow 0 \text{ as } k \rightarrow \infty. \quad (2.7)$$

Finally, we have almost-orthogonality of the  $H_k \times L^2$  norms of the decomposition:

$$\|\vec{u}_n\|_{H_k \times L^2}^2 = \sum_{1 \leq j \leq k} \|\vec{V}_L^j(-t_{n,j}/\lambda_{n,j})\|_{H_k \times L^2}^2 + \|\vec{w}_n^k\|_{H_k \times L^2}^2 + o_n(1) \text{ as } n \rightarrow \infty$$

**Remark 2.5** The difference between Lemma 2.4 and the main theorem in [1] is that here we have phrased matters in terms of solutions to the 4d linear wave equation with a repulsive inverse square potential (2.5) (which conserve the  $H_k \times L^2$  norm), as opposed to the free wave equation in 4d with data in  $\dot{H}^1 \times L^2$ . However, a proof identical to the one in [1] can be used to establish Lemma 2.4. Alternatively, one can establish Lemma 2.4 by conjugating (2.5) to the free wave equation in dimension  $d = 2k + 2$  via the map  $v(r) \mapsto r^{-k+1}v(r) = u$ . This map induces an isometry  $H_k(\mathbb{R}^4) \rightarrow \dot{H}^1(\mathbb{R}^{2k+2})$ ; see (2.3). Then the usual Bahouri-Gérard profile decomposition in  $d = 2k + 2$  induces a profile decomposition in  $H_k$ . One must check that the errors  $w_{n,L}^j$  can be made to vanish as in (2.7), but this follows by combining the vanishing of  $r^{-k+1}w_{n,L}^j$  in appropriate  $\dim = 2k + 2$  Strichartz norms with the Strauss estimate,

$$\sup_{t \in \mathbb{R}, r > 0} |r w_{n,L}^j(t, r)| \lesssim \|w_{n,L}^j\|_{L_t^\infty H_k(\mathbb{R}^4)},$$

and interpolation.

A direct consequence of Lemma 2.4 and (2.6) with the identifications

$$\begin{aligned}\psi_n(r) &:= ru_n(r), \quad \gamma_n^J(r) := rw_n^J, \\ \varphi_L^j(-t_{n,j}/\lambda_{n,j}, r/\lambda_{n,j}) &:= \frac{r}{\lambda_{n,j}} V_L^j(-t_{n,j}/\lambda_{n,j}, r/\lambda_{n,j}),\end{aligned}$$

is the following profile decomposition for bounded sequences  $\vec{\psi}_n \in \mathcal{H}_0$ .

**Corollary 2.6** (Linear profile decomposition). *Consider a sequence  $\vec{\psi}_n \in \mathcal{H}_0$  that is uniformly bounded in  $\mathcal{H}_0$ . Then, up to passing to a subsequence, there exists a sequence of solutions  $\vec{\varphi}_L^j \in \mathcal{H}_0$  to (2.2), sequences of times  $\{t_{n,j}\} \subset \mathbb{R}$ , sequences of scales  $\{\lambda_{n,j}\} \subset (0, \infty)$ , and errors  $\vec{\gamma}_n^J$  defined by*

$$\vec{\psi}_n = \sum_{j=1}^J (\varphi_L^j(-t_{n,j}/\lambda_{n,j}, \cdot/\lambda_{n,j}), \frac{1}{\lambda_{n,j}} \partial_t \varphi_L^j(-t_{n,j}/\lambda_{n,j}, \cdot/\lambda_{n,j})) + (\gamma_{n,0}^J, \gamma_{n,1}^J)$$

so that the following statements hold: Let  $\gamma_{n,L}^J(t) \in \mathcal{H}_0$  denote the linear evolution, [i.e., solution to (2.2)] of the data  $\vec{\gamma}_n^J \in \mathcal{H}_0$ . Then, for any  $j \leq \ell$ ,

$$(\gamma_n^\ell(t_{n,j}, \lambda_{n,j} \cdot), \lambda_{n,j} \gamma_n^\ell(t_{n,j}, \lambda_{n,j} \cdot)) \rightharpoonup 0 \text{ weakly in } \mathcal{H}_0.$$

In addition, for any  $j \neq \ell$  we have

$$\frac{\lambda_{n,j}}{\lambda_{n,\ell}} + \frac{\lambda_{n,\ell}}{\lambda_{n,j}} + \frac{|t_{n,j} - t_{n,\ell}|}{\lambda_{n,j}} + \frac{|t_{n,j} - t_{n,\ell}|}{\lambda_{n,\ell}} \rightarrow \infty \text{ as } n \rightarrow \infty. \quad (2.8)$$

Moreover, the errors  $\vec{\gamma}_n^J$  vanish asymptotically in the sense that

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{r} \gamma_{n,L}^J \right\|_{L_t^\infty L_x^4 \cap L_t^3 L_x^6(\mathbb{R} \times \mathbb{R}^4)} \rightarrow 0 \text{ as } J \rightarrow \infty.$$

Finally, we have almost-orthogonality of the  $\mathcal{H}_0$  norms of the decomposition:

$$\|\vec{\psi}_n\|_{\mathcal{H}_0}^2 = \sum_{1 \leq j \leq J} \|\vec{\varphi}_L^j(-t_{n,j}/\lambda_{n,j})\|_{\mathcal{H}_0}^2 + \|\vec{\gamma}_n^J\|_{\mathcal{H}_0}^2 + o_n(1) \text{ as } n \rightarrow \infty$$

Our applications of the concentration-compactness techniques developed by Kenig and Merle [26, 27] requires a “Pythagorean decomposition” of the nonlinear energy proved in [9].

**Lemma 2.7** [9, Lemma 2.16] *Let  $\vec{\psi}_n \in \mathcal{H}_0$  be a bounded sequence with a linear profile decomposition as in Corollary 2.6. Then the following Pythagorean decomposition holds for the nonlinear energy of the sequence:*

$$\mathcal{E}(\vec{\psi}_n) = \sum_{j=1}^J \mathcal{E}(\vec{\varphi}_L^j(-t_{n,j}/\lambda_{n,j})) + \mathcal{E}(\vec{\gamma}_n^J) + o_n(1) \text{ as } n \rightarrow \infty. \quad (2.9)$$

We will also require the following nonlinear profile decomposition analogous to [9, Proposition 2.17], or [14, Proposition 2.8]. We'll use the following notation: Given a linear profile decomposition as in Corollary 2.6 with profiles  $\{\varphi_L^j\}$  and parameters  $\{t_{n,j}\}, \{\lambda_{n,j}\}$  we denote by  $\{\varphi^j\}$  the nonlinear profile associated to  $\{\varphi_L^j(-t_n^j/\lambda_n^j), \vec{\varphi}_L^j(-t_n^j/\lambda_n^j)\}$ , i.e., the unique solution to (1.5) so that for all  $-t_n^j/\lambda_n^j \in I_{\max}(\vec{\varphi}^j)$  we have

$$\lim_{n \rightarrow \infty} \|\vec{\varphi}^j(-t_{n,j}/\lambda_{n,j}) - \vec{\varphi}_L^j(-t_{n,j}/\lambda_{n,j})\|_{\mathcal{H}_0} = 0.$$

The existence of a non-linear profile is immediate from the local well-posedness theory for (1.5) in the case that  $-t_{n,j}/\lambda_{n,j} \rightarrow \tau_{\infty,j} \in \mathbb{R}$ . If  $-t_{n,j}/\lambda_{n,j} \rightarrow \pm\infty$  then the existence of the nonlinear profile follows from the existence of wave operators for (1.5) and it follows that the maximal forward/backward time of existence  $T_{\pm}(\vec{\varphi}) = \infty$ . Each of these facts are now standard consequences of the Strichartz estimates in Lemma 2.2.

**Lemma 2.8** (Nonlinear Profile Decomposition). [9, Proposition 2.17] [14, Proposition 2.8] [1] *Let  $\vec{\psi}_n(0) \in \mathcal{H}_0$  be a uniformly bounded sequence with a profile decomposition as in Corollary 2.6. Assume that the nonlinear profile  $\varphi^j$  associated to the linear profile  $\varphi_L^j$  has maximal forward time of existence  $T_+(\vec{\varphi}^j)$ . Let  $s_n \in (0, \infty)$  be any sequence such that for all  $j$  and for all  $n$ ,*

$$\frac{s_n - t_{n,j}}{\lambda_{n,j}} < T_+(\vec{\varphi}^j), \quad \limsup_{n \rightarrow \infty} \|\varphi^j/r\|_{L_t^3\left(\left[-\frac{t_{n,j}}{\lambda_{n,j}}, \frac{s_n - t_{n,j}}{\lambda_{n,j}}\right]; L_x^6(\mathbb{R}^4)\right)} < \infty.$$

*Let  $\vec{\psi}_n(t)$  denote the solution of (1.5) with initial data  $\vec{\psi}_n(0)$ . Then for  $n$  large enough  $\vec{\psi}_n(t)$  exists on the interval  $s \in (0, s_n)$  and satisfies,*

$$\limsup_{n \rightarrow \infty} \|\psi_n/r\|_{L_t^3([0, s_n]; L_x^6(\mathbb{R}^4))} < \infty.$$

*Moreover, the following non-linear profile decomposition holds for all  $s \in [0, s_n)$ ,*

$$\begin{aligned} \vec{\psi}_n(s, r) &= \sum_{j=1}^J \left( \varphi^j \left( \frac{s - t_{n,j}}{\lambda_{n,j}}, \frac{r}{\lambda_{n,j}} \right), \frac{1}{\lambda_{n,j}^j} \partial_t \varphi^j \left( \frac{s - t_{n,j}}{\lambda_{n,j}}, \frac{r}{\lambda_{n,j}} \right) \right) \\ &\quad + \vec{\gamma}_{n,L}^J(s, r) + \vec{\theta}_n^J(s, r) \end{aligned}$$

with  $\gamma_{n,L}^J(t)$  as in (2.7) and

$$\lim_{J \rightarrow \infty} \limsup_{n \rightarrow \infty} \left( \|\theta_n^J / r\|_{L_t^3([0, s_n]; L_x^6(\mathbb{R}^4))} + \|\bar{\theta}_n^J\|_{L_t^\infty([0, s_n]; \mathcal{H}_0)} \right) = 0.$$

The analogous statement holds for sequences  $s_n \in (-\infty, 0)$ .

Our main application of these ideas can be summarized in the following compactness lemma.

**Lemma 2.9** *Let  $\vec{\psi}(t) \in \mathcal{H}_0$  be a solution to (1.5) defined on its forward maximal interval of existence  $[0, T_+(\vec{\psi}))$ . Suppose that  $\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(Q_k)$  and that  $\vec{\psi}(t)$  does not scatter as  $t \rightarrow T_+(\vec{\psi})$ . Then the following holds: Suppose that  $t_n \rightarrow T_+$  is any sequence of times such that*

$$\sup_n \|\vec{\psi}(t_n)\|_{\mathcal{H}_0} \leq C < \infty \quad (2.10)$$

*Then, up to passing to a subsequence of the  $t_n$ , there exists scales  $v_n > 0$  and a nonzero  $\vec{\varphi} \in \mathcal{H}_0$  such that*

$$\vec{\psi}(t_n) \underset{v_n}{\perp} \rightarrow \vec{\varphi} \in \mathcal{H}_0$$

*strongly in  $\mathcal{H}_0$  and  $\mathcal{E}(\vec{\varphi}) = 2\mathcal{E}(Q_k)$ . Moreover, the nonlinear evolution  $\vec{\varphi}(s)$  of the data  $\vec{\varphi}(0) = \vec{\varphi}$  is non-scattering in both forwards and backwards time.*

**Remark 2.10** One consequence the main result, Theorem 1.6, is that the hypothesis of Lemma 2.9 are not satisfied by any solution! However, we'll use Lemma 2.9 in the context of a contradiction argument in the proof of Proposition 4.1 in Sect. 4. Since the proof of the lemma uses only standard facts about profile decompositions, the local Cauchy theory, and the Threshold Theorem 1.1 we include it here in Sect. 2.3.

**Proof of Lemma 2.9** By (2.10) we can perform a linear profile decomposition as in Corollary 2.6 on  $\vec{\psi}(t_n)$ .

First we observe that there can only be one non-zero profile  $\vec{\varphi} = \vec{\varphi}^1$  and that the errors  $\vec{\gamma}_{n,L}^J$  must vanish strongly  $\mathcal{H}_0$  as  $n \rightarrow \infty$ . Indeed, if there were two non-trivial profiles, or if the errors did not vanish strongly in  $\mathcal{H}_0$ , then (2.9) along with our hypothesis that  $\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(Q)$  imply that every nonzero profile must have energy  $< 2\mathcal{E}(\vec{Q})$ . Thus each non-zero nonlinear profile scatters in both directions by the Threshold Theorem 1.1. A now standard argument based on the nonlinear Perturbation Lemma 2.3, and the orthogonality of the parameters in (2.8) implies that  $\vec{\psi}(t)$  must also scatter in forward time, a contradiction.

Thus, there exists times  $t_{n,1}$  and scales  $\nu_{n,1}$ , and a single limiting profile  $\vec{\varphi} = (\varphi_0, \varphi_1)$  so that

$$(\psi(t_n + \nu_n t_{n,1}, \nu_{n,1} \cdot), \nu_{n,1} \psi_t(t_n + \nu_n t_{n,1}, \nu_{n,1} \cdot)) \rightarrow \vec{\varphi} \in \mathcal{H}_0 \text{ as } n \rightarrow \infty$$

Next we claim both  $-\frac{t_{n,1}}{\nu_{n,1}} \rightarrow \pm\infty$  are impossible and we can therefore assume without loss of generality that  $t_{n,1} = 0$  for all  $n$ . To see this, first assume first  $-\frac{t_{n,1}}{\nu_{n,1}} \rightarrow +\infty$ . Then  $\vec{\varphi}$  scatters in forward time and we can deduce that

$$\|\varphi/r\|_{L_t^3 L_x^6([-\frac{t_{n,1}}{\nu_{n,1}}, \infty) \times \mathbb{R}^4)} \rightarrow 0 \text{ as } n \rightarrow \infty$$

by the definition of the nonlinear profile. But then the Nonlinear Perturbation Lemma 2.3 implies that  $\vec{\psi}(t)$  must also scatter in forward time, which contradicts our initial assumptions on  $\vec{\psi}(t)$ .

Now assume that  $-t_{n,1}/\nu_{n,1} \rightarrow -\infty$ . Then the nonlinear profile  $\vec{\varphi}(s)$  scatters in backwards time, and the Nonlinear Perturbation Lemma 2.3 implies that

$$\|\psi/r\|_{L_t^3 L_x^6([0, t_n] \times \mathbb{R}^4)} = \|\varphi/r\|_{L_t^3 L_x^6([\frac{-t_n - t_{n,1}}{\nu_{n,1}}, \frac{t_{n,1}}{\nu_{n,1}}] \times \mathbb{R}^4)} + o_n(1) \rightarrow 0,$$

a contradiction. Thus we can assume that  $t_{n,1} \equiv 0$  and we simply write  $\nu_{n,1} = \nu_n$ . At this point we've shown that up to passing to a subsequence in  $t_n$  we have

$$\vec{\psi}(t_n) \xrightarrow{\frac{1}{\nu_n}} \varphi \in \mathcal{H}_0, \quad \mathcal{E}(\vec{\varphi}) = 2\mathcal{E}(Q)$$

We can now run a nearly identical argument to show that nonlinear evolution  $\vec{\varphi}(s) \in \mathcal{H}_0$  can not scatter in either time direction. To see this, first suppose that  $\vec{\varphi}$  scatters in forward time. Then, the Nonlinear Perturbation Lemma 2.3 implies that  $\vec{\psi}(t)$  must also scatter as  $t \rightarrow \infty$ . If  $\vec{\varphi}(s)$  were to scatter as  $s \rightarrow -\infty$ , then,

$$\|\psi/r\|_{L_t^3 L_x^6([0, t_n] \times \mathbb{R}^4)} = \|\vec{\varphi}/r\|_{L_t^3 L_x^6([(-t_n)/\nu_{n,1}, 0] \times \mathbb{R}^4)} + o_n(1) \leq C < \infty.$$

Letting limit  $n \rightarrow \infty$ , we see that  $\|\psi/r\|_{L_t^3 L_x^6([0, T_+(\vec{\psi})) \times \mathbb{R}^4)} \leq C$ , which again means that  $\vec{\psi}(t)$  scatters in forward time, a contradiction. Hence  $\vec{\varphi}(s)$  does not scatter in either direction.  $\square$

## 2.4 The harmonic maps $Q = Q_k$

We record a few properties about the unique (up to scaling)  $k$ -equivariant harmonic map  $Q = Q_k(r) = 2 \arctan r^k$  and some consequences of the fact that each  $Q_k$  minimizes the energy functional amongst all  $k$ -equivariant maps.

First observe that  $Q$  satisfies

$$r \partial_r Q(r) = k \sin Q(r), \quad Q(0) = 0, \quad Q(\infty) = \pi.$$

Recall that  $\mathcal{H}_\pi$  the set of all finite energy  $k$ -equivariant maps, with  $\phi_0(0) = 0$  and  $\phi_0(\infty) = \pi$ ,

$$\mathcal{H}_\pi := \{(\phi_0, \phi_1) \mid \mathcal{E}(\vec{\phi}) < \infty, \quad \phi_0(0) = 0, \quad \lim_{r \rightarrow \infty} \phi_0(r) = \pi\}$$

The fact that  $Q$  minimizes the energy in  $\mathcal{H}_\pi$  can be easily seen from the following Bogomol'nyi factorization of the energy:

$$\begin{aligned} \mathcal{E}(\phi_0, \phi_1) &= \pi \|\phi_1\|_{L^2}^2 + \pi \int_0^\infty \left( \partial_r \phi_0 - k \frac{\sin(\phi_0)}{r} \right)^2 r \, dr \\ &\quad + 2\pi k \int_0^\infty \sin(\phi_0) \partial_r \phi_0 \, dr \\ &= \pi \|\phi_1\|_{L^2}^2 + \pi \int_0^\infty \left( \partial_r \phi_0 - k \frac{\sin(\phi_0)}{r} \right)^2 r \, dr \\ &\quad + 2\pi k \int_{\phi_0(0)}^{\phi_0(\infty)} \sin(\rho) \, d\rho \\ &= \pi \|\phi_1\|_{L^2}^2 + \pi \int_0^\infty \left( \partial_r \phi_0 - k \frac{\sin(\phi_0)}{r} \right)^2 r \, dr + 4\pi k \quad (2.11) \end{aligned}$$

Hence,

$$\mathcal{E}(\phi_0, \phi_1) \geq \pi \|\phi_1\|_{L^2}^2 + 4\pi k = \pi \|\phi_1\|_{L^2}^2 + \mathcal{E}(Q_k, 0)$$

where the inequality in the last line above is in fact strict if  $\phi_0 \neq Q_k$ .

We define a functional on maps  $\Phi : \mathbb{R}^2 \rightarrow \mathbb{S}^2$  of finite energy. Let  $\omega_{\mathbb{S}^2}$  denote the volume form on  $\mathbb{S}^2$ . Given  $\Omega \subset \mathbb{R}^2$  set

$$G(\Phi, \Omega) := \int_{\Phi(\Omega)} \omega_{\mathbb{S}^2} = \int_\Omega \Phi^*(\omega_{\mathbb{S}^2})$$

where  $\Phi^*(\omega_{\mathbb{S}^2})$  denotes the pull-back. Given  $k$ -equivariant  $\Phi$  with polar angle  $\phi$ , this reduces to

$$G(\phi_0(r)) := 2\pi \int_{\phi_0(0)}^{\phi_0(r)} k |\sin \rho| \, d\rho$$

Observe that for any  $(\phi, 0)$  with  $\mathcal{E}(\vec{\phi}) < \infty$  and for any  $R \in [0, \infty)$  we have

$$\begin{aligned} |G(\phi_0(R))| &= \left| 2\pi \int_{\phi_0(0)}^{\phi_0(R)} k |\sin \rho| \, d\rho \right| = \left| 2\pi \int_0^R |k \sin(\phi_0(r))| \, \partial_r \phi_0(r) \, dr \right| \\ &\leq \mathcal{E}_0^R(\phi_0, 0) \end{aligned}$$

where for any  $0 \leq a < b$  we define the localized energy  $\mathcal{E}_b^a$  by

$$\mathcal{E}_a^b(\phi_0, \phi_1) := 2\pi \int_a^b \frac{1}{2} \left( \phi_1^2 + (\partial_r \phi_0)^2 + k^2 \frac{\sin^2 \phi_0}{r^2} \right) r \, dr.$$

The same argument shows that

$$|G(\phi_0(R))| \leq \mathcal{E}_R^\infty(\phi_0, 0)$$

On the other hand, since  $Q$  satisfies  $r \partial_r Q(r) = k \sin(Q)$ , for any  $0 \leq a \leq b < \infty$  we see that

$$G(Q(b)) - G(Q(a)) = 2\pi \int_a^b |\sin(Q(r))| \, Q_r(r) \, dr = \mathcal{E}_a^b(Q, 0) \quad (2.12)$$

Letting  $a \rightarrow 0$  and  $b \rightarrow \infty$  we recover the fact that  $\mathcal{E}(Q, 0) = G(\pi) = 4\pi k$ .

We recall the following variational characterization of  $Q$  in  $\mathcal{H}_\pi$  from [7], which amounts to the coercivity of the energy functional near  $Q$ .

**Lemma 2.11** [7, Proposition 2.3] *There exists a function  $c : [0, \infty) \rightarrow [0, \infty)$  such that  $c(\alpha) \rightarrow 0$  as  $\alpha \rightarrow 0$  and such that the following holds: Let  $(\phi_0, 0) \in \mathcal{H}_\pi$ . Suppose*

$$\alpha := \mathcal{E}(\phi_0, 0) - \mathcal{E}(Q, 0) \geq 0$$

*Then for  $\lambda > 0$  defined so that  $\mathcal{E}_0^\lambda(\phi_0, 0) = \mathcal{E}_0^1(Q) = \mathcal{E}(Q)/2$ , we have*

$$\|\phi_0 - Q_\lambda\|_H \leq c(\alpha)$$

*Moreover,  $\alpha = 0$  if and only if  $\phi_0(r) = Q(r/\lambda)$  for some  $\lambda > 0$ .*

## 2.5 Threshold solutions near a 2-bubble configuration

The goal of this section is to relate the proximity of a map  $\phi \in \mathcal{H}_0$  to a 2-bubble configuration to the size of the  $\mathcal{H}_0$ -norm of  $\vec{\phi}$ . With this in mind we make the following definition.

**Definition 2.12** (*Proximity to a 2-bubble*). Given a map  $\vec{\phi} = (\phi_0, \phi_1) \in \mathcal{H}_0$  we define its proximity  $\mathbf{d}(\vec{\phi})$  to a pure 2-bubble by

$$\mathbf{d}(\vec{\phi}) := \inf_{\lambda, \mu > 0, \iota \in \{+1, -1\}} \left( \|(\phi_0 - \iota(Q_\lambda - Q_\mu), \phi_1)\|_{\mathcal{H}_0}^2 + (\lambda/\mu)^k \right) \quad (2.13)$$

The proof of Theorem 1.6 will require a few technical lemmas concerning  $\mathbf{d}$ . We'll state the lemmas first and postpone the proofs until the end of this section.

**Lemma 2.13** Suppose that  $\vec{\phi} = (\phi_0, \phi_1) \in \mathcal{H}_0$  is  $k$ -equivariant and satisfies,

$$\mathcal{E}(\vec{\phi}) = 2\mathcal{E}(\vec{Q}_k).$$

Then for each  $\beta > 0$  there exists a constant  $C(\beta) > 0$  such that

$$\mathbf{d}(\vec{\phi}) \geq \beta \implies \|(\phi_0, \phi_1)\|_{\mathcal{H}_0} \leq C(\beta) \quad (2.14)$$

Conversely, for each  $A > 0$  we can find  $\alpha = \alpha(A)$  such that

$$\mathbf{d}(\vec{\phi}) \leq \alpha(A) \implies \|(\phi_0, \phi_1)\|_{\mathcal{H}_0} \geq A \quad (2.15)$$

Note that  $\mathbf{d}$  is small when  $\vec{\phi}$  is close to either a bubble/anti-bubble ( $\iota = +$  in the definition of  $\mathbf{d}$ ) or anti-bubble/bubble configuration ( $\iota = -$ ). The next lemma makes precise the intuitive notion that a map  $\vec{\phi}$  cannot be simultaneously close to both configurations. With this in mind we define

$$\mathbf{d}_\pm(\vec{\phi}) := \inf_{\lambda, \mu > 0} \left( \|(\phi_0 \mp (Q_\lambda - Q_\mu), \phi_1)\|_{\mathcal{H}_0}^2 + (\lambda/\mu)^k \right) \quad (2.16)$$

**Lemma 2.14** There exists  $\alpha_0 > 0$  with the following property: Let  $\vec{\phi} \in \mathcal{H}_0$ . Then,

$$\mathbf{d}_\pm(\vec{\phi}) \leq \alpha_0 \implies \mathbf{d}_\mp(\vec{\phi}) \geq \alpha_0.$$

We begin by proving Lemma 2.13.

*Proof of Lemma 2.13* It suffices to consider  $\vec{\phi}$  of the form  $\vec{\phi} = (\phi, 0)$ . First we prove (2.14). To see this we'll first show that for each  $\beta > 0$  there exists a constant  $\delta = \delta(\beta)$  so that for any  $\vec{\phi} \in \mathcal{H}_0$  with  $\mathcal{E}(\vec{\phi}) = 2\mathcal{E}(Q)$  we have

$$\mathbf{d}(\vec{\phi}) \geq \beta \implies \|\phi\|_{L^\infty} \leq \pi - \delta(\beta), \quad (2.17)$$

Suppose (2.17) fails. Then we can find  $\beta > 0$ , a sequence  $\vec{\phi}_n = (\phi_n, 0) \in \mathcal{H}_0$  with  $\mathcal{E}(\vec{\phi}_n) = 2\mathcal{E}(Q)$ , and numbers  $r_n > 0$  so that

$$\mathbf{d}(\phi_n, 0) \geq \beta \quad \text{and} \quad |\phi_n(r_n) - \pi| = o_n(1) \quad \text{as } n \rightarrow \infty \quad (2.18)$$

Define scales  $\lambda_n$  and  $\mu_n$  by

$$\mathcal{E}_0^{\lambda_n}(\vec{\phi}_n) = \mathcal{E}(Q)/2, \quad \mathcal{E}_{\mu_n}^\infty(\vec{\phi}_n) = \mathcal{E}(Q)/2 \quad (2.19)$$

Then, by (2.12) we see that for  $n$  large enough  $\lambda_n < r_n$  and  $\mu_n > r_n$ . Now define  $\phi_{n,1}$  and  $\phi_{n,2}$  as follows

$$\phi_{n,1}(r) = \begin{cases} \phi_n(r) & \text{if } 0 \leq r \leq r_n \\ \pi + \frac{\pi - \phi(r_n)}{r_n}(r - 2r_n) & \text{if } r \in [r_n, 2r_n] \\ \pi & \text{if } r \geq 2r_n \end{cases}$$

$$\phi_{n,2}(r) = \begin{cases} \pi + \frac{\phi(r_n) - \pi}{r_n}r & \text{if } r \leq r_n \\ \phi_n(r) & \text{if } r \geq r_n \end{cases}$$

And define  $\eta_n(r)$  by

$$\eta_n(r) := \phi_n(r) - \phi_{n,1}(r) - \phi_{n,2}(r) + \pi$$

We claim that

$$\mathcal{E}(\phi_{n,1}, 0) = \mathcal{E}(Q, 0) + o_n(1) \quad \text{as } n \rightarrow \infty \quad (2.20)$$

$$\mathcal{E}(\phi_{n,2}, 0) = \mathcal{E}(Q, 0) + o_n(1) \quad \text{as } n \rightarrow \infty$$

$$\|\eta_n\|_H \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad (2.21)$$

First we prove (2.20) (2.21). Since  $\phi_n(r_n) \rightarrow \pi$  we have

$$\mathcal{E}_0^{r_n}(\phi_{n,1}, 0) = \mathcal{E}_0^{r_n}(\phi_n, 0) \geq G(\phi_n(r_n)) \rightarrow G(\pi) = \mathcal{E}(Q, 0) \quad \text{as } n \rightarrow \infty$$

$$\mathcal{E}_{r_n}^\infty(\phi_n, 0) = \mathcal{E}_{r_n}^\infty(\phi_n, 0) \geq G(\phi_n(r_n)) \rightarrow G(\pi) = \mathcal{E}(Q, 0) \quad \text{as } n \rightarrow \infty$$

From the above and the fact that  $\mathcal{E}(\vec{\phi}) = 2\mathcal{E}(Q)$  we see that in fact

$$\begin{aligned}\mathcal{E}_0^{r_n}(\phi_{n,1}, 0) &= \mathcal{E}(Q, 0) + o_n(1) \text{ as } t \rightarrow \infty \\ \mathcal{E}_{r_n}^\infty(\phi_{n,2}, 0) &= \mathcal{E}(Q, 0) + o_n(1) \text{ as } t \rightarrow \infty\end{aligned}\quad (2.22)$$

Direct computations using the definitions of  $\phi_{n,1}, \phi_{n,2}$  then show that

$$\begin{aligned}\mathcal{E}_{r_n}^\infty(\phi_{n,1}, 0) &\lesssim (\pi - \phi_n(r_n))^2 \rightarrow 0 \text{ as } n \rightarrow \infty \\ \mathcal{E}_0^{r_n}(\phi_{n,2}, 0) &\lesssim (\pi - \phi_n(r_n))^2 \rightarrow 0 \text{ as } n \rightarrow \infty\end{aligned}\quad (2.23)$$

Combining (2.22) and (2.23) gives (2.20) and (2.21). By construction

$$\eta_n(r) = \pi - \phi_{n,2}(r) \text{ if } r \leq r_n, \quad \eta_n(r) = \pi - \phi_{n,1}(r) \text{ if } r \geq r_n$$

A direct computation using the above and the definitions of  $\phi_{n,1}, \phi_{n,2}$  on the relevant intervals then yields

$$\|\eta_n\|_H^2 \lesssim (\pi - \phi_n(r_n))^2 \rightarrow 0 \text{ as } n \rightarrow \infty$$

By (2.20), and (2.21), and  $\lambda_n, \mu_n$  defined in (2.19) we use Lemma 2.11 to find  $\eta_{n,1}, \eta_{n,2} \in H$  so that

$$\begin{aligned}\phi_{n,1}(r) &= Q_{\lambda_n} + \eta_{n,1}(r), \quad \phi_{n,2}(r) = \pi - Q_{\mu_n} - \eta_{n,2}(r) \\ \|\eta_{n,j}\|_H &\rightarrow 0 \text{ as } n \rightarrow \infty\end{aligned}$$

for  $j = 1, 2$ . Thus,

$$\|\phi_n - Q_{\lambda_n} + Q_{\mu_n}\|_H = \|\eta_n + \eta_{n,1} - \eta_{n,2}\|_H \rightarrow 0 \text{ as } n \rightarrow \infty \quad (2.24)$$

Moreover, we must have  $\lambda_n/\mu_n \rightarrow 0$  as  $n \rightarrow \infty$ . To see this, simply note that if  $\lambda_n/\mu_n \simeq 1$  then  $Q_{\lambda_n} - Q_{\mu_n}$  stays bounded away from  $\pi$ . But this contradicts (2.24) and the assumption that  $\phi_n(r_n) \rightarrow \pi$  as  $n \rightarrow \infty$ . Hence,

$$\|\phi_n - Q_{\lambda_n} + Q_{\mu_n}\|_H^2 + \left(\frac{\lambda_n}{\mu_n}\right)^k \rightarrow 0$$

and thus  $\mathbf{d}(\phi_n, 0) \rightarrow 0$ , which contradicts (2.18). To finish the proof, note that (2.17) implies the estimate

$$\phi^2(r) \leq C(\beta) \sin^2 \phi(r)$$

which means we can control the  $H$  norm of  $\phi$  by a constant (which depends only on  $\beta$ ) times  $\mathcal{E}(\phi) = 2\mathcal{E}(Q)$ .

Lastly, we prove (2.15). Suppose that  $\mathbf{d}(\vec{\phi}) \leq \alpha$ . Then we can find, say,  $\lambda_0, \mu_0$  such that

$$\alpha \leq \|(\phi_0 - Q_{\lambda_0} + Q_{\mu_0}, \phi_1)\|_{\mathcal{H}_0} + (\lambda_0/\mu_0)^k \leq 2\alpha$$

A direct computation then shows,

$$\begin{aligned} \|\phi_0\|_H &\geq \|Q_{\lambda_0} - Q_{\mu_0}\|_H - \|\phi_0 - Q_{\lambda_0} + Q_{\mu_0}\|_H \\ &\gtrsim |\log(\lambda_0/\mu_0)| - 2\alpha \rightarrow \infty \text{ as } \alpha \rightarrow 0 \end{aligned}$$

which completes the proof.  $\square$

We next prove Lemma 2.14.

*Proof of Lemma 2.14* If the conclusion fails we could find a sequence  $\phi_n \in H$ , and two sequences of scales  $\lambda_n^+, \mu_n^+, \lambda_n^-, \mu_n^-$  so that

$$\begin{aligned} \|\phi_n - Q_{\lambda_n^+} + Q_{\mu_n^+}\|_H + \frac{\lambda_n^+}{\mu_n^+} &\rightarrow 0 \text{ as } n \rightarrow \infty \\ \|\phi_n + Q_{\lambda_n^-} - Q_{\mu_n^-}\|_H + \frac{\lambda_n^-}{\mu_n^-} &\rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

It follows that

$$\begin{aligned} 0 &= \|(\phi_n - Q_{\lambda_n^+} + Q_{\mu_n^+}) - (\phi_n + Q_{\lambda_n^-} - Q_{\mu_n^-}) + (Q_{\lambda_n^+} - Q_{\mu_n^+} \\ &\quad + Q_{\lambda_n^-} - Q_{\mu_n^-})\|_H \\ &\geq \|Q_{\lambda_n^+} - Q_{\mu_n^+} + Q_{\lambda_n^-} - Q_{\mu_n^-}\|_H - o_n(1) \text{ as } n \rightarrow \infty \end{aligned} \quad (2.25)$$

Passing to subsequences if necessary, relabeling  $\pm$ , or rescaling, we can assume that  $\lambda_n^+ \leq \lambda_n^-$  for all  $n$  and that one of the following three possibilities holds

$$\frac{\lambda_n^-}{\mu_n^+} \rightarrow 0, \text{ or } \frac{\lambda_n^-}{\mu_n^+} \rightarrow \infty, \text{ or } \frac{\lambda_n^-}{\mu_n^+} \rightarrow 1 > 0, \text{ as } n \rightarrow \infty$$

Assume we are in the first situation. Then, we can choose  $n$  large enough so that

$$\begin{aligned} Q_{\lambda_n^+}(r) + Q_{\lambda_n^-}(r) &\geq \pi \quad \forall r \in [\lambda_n^-, 2\lambda_n^-] \\ Q_{\mu_n^+}(r) + Q_{\mu_n^-}(r) &\leq \frac{\pi}{2} \quad \forall r \in [\lambda_n^-, 2\lambda_n^-] \end{aligned}$$

and thus

$$\|Q_{\lambda_n^+} - Q_{\mu_n^+} + Q_{\lambda_n^-} - Q_{\mu_n^-}\|_H^2 \geq \frac{\pi^2}{4} \int_{\lambda_n^-}^{2\lambda_n^-} \frac{dr}{r} \geq \frac{\pi^2}{4} \log 2$$

for all  $n$  large enough, which is impossible by (2.25). Now suppose we are in the second case  $\frac{\lambda_n^-}{\mu_n^+} \rightarrow \infty$ . This means that

$$\lambda_n^+ \ll \mu_n^+ \ll \lambda_n^- \ll \mu_n^-$$

and so for large enough  $n$  we have

$$(Q_{\lambda_n^+} - Q_{\mu_n^+} + Q_{\lambda_n^-} - Q_{\mu_n^-})(r) \geq \frac{\pi}{4}, \quad \forall r \in [\lambda_n^+, 2\lambda_n^+]$$

which similarly leads (2.25) into a contradiction. Finally, if  $\frac{\lambda_n^-}{\mu_n^+} \rightarrow 1$  we have

$$\|Q_{\lambda_n^+} - Q_{\mu_n^+} + Q_{\lambda_n^-} - Q_{\mu_n^-}\|_H \geq \|Q_{\lambda_n^+} - Q_{\mu_n^+}\|_H - o_n(1)$$

Then setting  $\varphi_n := Q_{\lambda_n^+} - Q_{\mu_n^+}$  we see that  $\mathbf{d}((\varphi_n, 0)) \rightarrow 0$  and hence the right-hand-side above is bounded below by a fixed constant by (2.15) in Lemma 2.13. This again leads to a contradiction in (2.25), which completes the proof.  $\square$

## 2.6 Virial identity

In this section we record a *nonlinear* estimate related to a virial-type identity that will be used in the proof of Theorem 1.6.

We begin with a virial-type identity for solutions to (1.5). In what follows we fix a smooth radial cut-off function  $\chi \in C_{\text{rad}}^\infty(\mathbb{R}^2)$ , so that, writing  $\chi = \chi(r)$  we have

$$\chi(r) = 1 \text{ if } r \leq 1 \text{ and } \chi(r) = 0 \text{ if } r \geq 3 \text{ and } |\chi'(r)| \leq 1 \quad \forall r \geq 0$$

For each  $R > 0$  we then define

$$\chi_R(r) := \chi(r/R)$$

**Lemma 2.15** *Let  $\vec{\psi}(t)$  be a solution to (1.5) on a time interval  $I$ . Then for any time  $t \in I$  and  $R > 0$  fixed we have*

$$\frac{d}{dt} \langle \psi_t \mid \chi_R r \partial_r \psi \rangle_{L^2}(t) = - \int_0^\infty \psi_t^2(t, r) r \, dr + \Omega_R(\vec{\psi}(t))$$

where

$$\begin{aligned}\Omega_R(\vec{\psi}(t)) &:= \int_0^\infty \psi_t^2(t)(1 - \chi_R) r \, dr \\ &\quad - \frac{1}{2} \int_0^\infty \left( \psi_t^2(t) + \psi_r^2(t) - k^2 \frac{\sin^2 \psi(t)}{r^2} \right) \frac{r}{R} \chi'(r/R) r \, dr\end{aligned}\quad (2.26)$$

satisfies

$$\begin{aligned}|\Omega_R(\vec{\psi}(t))| &\lesssim \int_R^\infty \psi_t^2(t, r) r \, dr \, dt + \int_R^\infty \left| \psi_r^2 - k^2 \frac{\sin^2 \psi}{r^2} \right| r \, dr \, dt \\ &\lesssim \mathcal{E}_R^\infty(\vec{\psi}(t))\end{aligned}$$

*Proof* By direct calculation, using (1.5) we have

$$\begin{aligned}&\frac{d}{dt} \langle \psi_t \mid \chi_R r \partial_r \psi \rangle_{L^2}(t) \\ &= - \int_0^\infty \psi_t^2(t) r \, dr + \int_0^\infty \psi_t^2(t)(1 - \chi_R) r \, dr \\ &\quad - \frac{1}{2} \int_0^\infty \left( \psi_t^2(t) + \psi_r^2(t) - k^2 \frac{\sin^2 \psi(t)}{r^2} \right) \frac{r}{R} \chi'(r/R) r \, dr\end{aligned}$$

□

We show below how the quantities appearing on the right hand side of the virial identity can be estimated in terms of  $\mathbf{d}(\vec{\psi})$  in the vicinity of a two-bubble.

**Lemma 2.16** *There exists a number  $C_0 > 0$  depending only on  $k$  such that for all  $\vec{\phi} = (\phi_0, \phi_1) \in \mathcal{H}_0$  with  $\mathcal{E}(\vec{\phi}) = 2\mathcal{E}(Q)$  and all  $R > 0$  there holds*

$$|\langle \phi_1, \chi_R r \partial_r \phi_0 \rangle| \leq C_0 R \sqrt{\mathbf{d}(\vec{\phi})}, \quad (2.27)$$

$$\Omega_R(\vec{\phi}) \leq C_0 \sqrt{\mathbf{d}(\vec{\phi})}. \quad (2.28)$$

*Proof* By Cauchy–Schwarz, we get

$$|\langle \phi_1, \chi_R r \partial_r \phi_0 \rangle| \lesssim R \|\phi_1\|_{L^2} \|\partial_r \phi_0\|_{L^2}. \quad (2.29)$$

We have  $\|\partial_r(Q_\lambda - Q_\mu)\|_{L^2} \lesssim 1$  for all  $\lambda$  and  $\mu$ , hence by the triangle inequality  $\|\partial_r \phi_0\|_{L^2} \lesssim 1 + \sqrt{\mathbf{d}(\vec{\phi})}$ . If  $\mathbf{d}(\vec{\phi}) \leq 1$ , then we obtain  $\|\partial_r \phi_0\|_{L^2} \lesssim 1$ . If  $\mathbf{d}(\vec{\phi}) \geq 1$ , then Lemma 2.13 gives  $\|\vec{\phi}\|_{\mathcal{H}_0} \lesssim 1$ , in particular again  $\|\partial_r \phi_0\|_{L^2} \lesssim 1$ . Thus (2.29) yields (2.27).

To prove (2.28), we write

$$|\Omega_R(\vec{\phi})| \lesssim \int_0^{+\infty} \phi_1^2 r dr + \left| (\partial_r \phi_0)^2 - k^2 \frac{\sin^2 \phi_0}{r^2} \right| r dr.$$

Again, the conclusion is clear if  $\mathbf{d}(\vec{\phi}) \geq 1$ , we can assume  $\mathbf{d}(\vec{\phi}) \leq 1$ . Find  $\lambda, \mu > 0$  such that, say,

$$\left(\frac{\lambda}{\mu}\right)^k \leq 2\mathbf{d}(\vec{\phi}) \text{ and } \|(\phi_0 - Q_\lambda + Q_\mu), \phi_1\|_{\mathcal{H}_0}^2 \leq 2\mathbf{d}(\vec{\phi})$$

By the above it suffices to show that for  $g := \phi_0 - Q_\lambda + Q_\mu$  we have

$$\int_0^{+\infty} \left| (\partial_r \phi_0)^2 - k^2 \frac{\sin^2 \phi_0}{r^2} \right| r dr \lesssim \left( \left(\frac{\lambda}{\mu}\right)^{k/2} + \|g\|_H \right)$$

Using trigonometric identities we expand

$$\begin{aligned} \sin^2(Q_\lambda - Q_\mu + g) &= \sin^2 Q_\lambda + \sin^2 Q_\mu - \frac{1}{2} \sin 2Q_\lambda \sin 2Q_\mu \\ &\quad - 2 \sin^2 Q_\lambda \sin^2 Q_\mu + \frac{1}{2} \sin 2g \sin 2(Q_\lambda - Q_\mu) \\ &\quad + \sin^2 g \cos 2(Q_\lambda - Q_\mu) \end{aligned}$$

Then, since  $\Lambda Q_\lambda := r \partial_r Q_\lambda = k \sin Q_\lambda$  we have

$$\begin{aligned} \int_0^\infty \left| (r \partial_r \phi_0)^2 - k^2 \sin^2 \phi_0 \right| \frac{dr}{r} &\lesssim \int_0^\infty |\Lambda Q_\lambda \Lambda Q_\mu| + |\Lambda Q_\lambda r \partial_r g| \\ &\quad + |\Lambda Q_\mu r \partial_r g| \frac{dr}{r} \\ &\quad + \int_0^\infty |g \Lambda Q_\lambda| + |g \Lambda Q_\mu r| \\ &\quad + |r \partial_r g|^2 + |g|^2 \frac{dr}{r} \end{aligned}$$

To estimate the first term above we see that setting  $\sigma = \lambda/\mu$  we have

$$\begin{aligned} \int_0^\infty |\Lambda Q_\lambda \Lambda Q_\mu| \frac{dr}{r} &\lesssim \int_0^\infty \frac{(r/\lambda)^k (r/\mu)^k}{(1 + (r/\lambda)^{2k})(1 + (r/\mu)^{2k})} \frac{dr}{r} \\ &= \sigma^k \int_0^\infty \frac{r^{2k-1}}{(\sigma^{2k} + r^{2k})(1 + r^{2k})} dr \lesssim \sigma^k |\log \sigma| \lesssim \left(\frac{\lambda}{\mu}\right)^{\frac{k}{2}} \end{aligned}$$

The remaining terms can be controlled by  $\|g\|_H$  by Cauchy–Schwarz.  $\square$

### 3 The modulation method: analysis of 2-bubble collisions

In this section we give a careful analysis of the modulation equations that govern the evolution of 2-bubble configurations. The intuition is that the less concentrated bubble does not change its scale and influences the dynamics of the more concentrated bubble. We will quantify this influence.

#### 3.1 Modulation equations

We consider solutions  $\vec{\psi}(t)$  to (1.5) that are close to a 2-bubble configuration on a time interval  $J$  in the sense that  $\mathbf{d}(\vec{\psi}(t))$ , defined in (2.13), is small for all  $t \in J$ . Recall that  $\mathbf{d}(\vec{\psi}(t))$  is the smaller of the numbers  $\mathbf{d}_+(\vec{\psi}(t))$  and  $\mathbf{d}_-(\vec{\psi}(t))$  defined in (2.16).

Linearizing (1.5) about  $Q_\lambda$  leads to the Schrödinger operator

$$\mathcal{L}_\lambda := -\partial_r^2 - \frac{1}{r}\partial_r + k^2 \frac{\cos 2Q_\lambda}{r^2}$$

We write  $\mathcal{L} := \mathcal{L}_1$ . Recall from (2.1) that  $\Lambda = r\partial_r$  is the infinitesimal generator of dilations in  $\dot{H}^1(\mathbb{R}^2)$ . One can check that  $\Lambda Q$  is a zero energy eigenfunction for  $\mathcal{L}$ , i.e.,

$$\mathcal{L}\Lambda Q = 0, \text{ and } \Lambda Q \in L_{\text{rad}}^2(\mathbb{R}^2).$$

When  $k = 1$ ,  $\mathcal{L}\Lambda Q = 0$  still holds but in this case  $\Lambda Q \notin L^2$  due to slow decay as  $r \rightarrow \infty$  and is 0 is referred to as a threshold resonance.

In fact,  $\Lambda Q$  spans the kernel of  $\mathcal{L}$ . This can be seen using the following well known factorization of  $\mathcal{L}$ ,

$$\mathcal{L} = A^*A \text{ where } A^* = \partial_r + \frac{1 + k \cos(Q)}{r}, \quad A = -\partial_r + \frac{k \cos(Q)}{r} \quad (3.1)$$

together with the observation that  $A(\Lambda Q) = 0$ ; we note that (3.1) is a consequence of the Bogomol'nyi factorization (2.11); see [44, 46] for more.

The fact that  $\mathcal{L}_\lambda \Lambda Q_\lambda = 0$  will play an important role in the modulation estimates.

We fix a radial function  $\mathcal{Z} \in C_0^\infty(\mathbb{R}^2)$  so that

$$\int_0^\infty \mathcal{Z}(r) \cdot \Lambda Q(r) r \, dr > 0, \quad \left| \frac{\mathcal{Z}(r)}{r^k} \right| \lesssim 1 \quad \forall r \leq 1. \quad (3.2)$$

**Lemma 3.1** (Modulation Lemma). *There exist  $\eta_0 > 0$  and  $C > 0$  with the following property: Let  $J \subset \mathbb{R}$  be a time interval,  $\vec{\psi}(t)$  a solution to (1.5) defined on  $J$ , and assume that*

$$\mathbf{d}_+(\vec{\psi}(t)) \leq \eta_0 \quad \forall t \in J.$$

Then, there exist unique  $C^1(J)$  functions  $\lambda(t), \mu(t)$  so that, defining  $g(t) \in H$  by

$$g(t) := \psi(t) - Q_{\lambda(t)} + Q_{\mu(t)}, \quad (3.3)$$

we have, for each  $t \in J$ ,

$$\langle \underline{Z_{\lambda(t)}} \mid g(t) \rangle = 0, \quad (3.4)$$

$$\langle \underline{Z_{\mu(t)}} \mid g(t) \rangle = 0, \quad (3.5)$$

$$\mathbf{d}_+(\vec{\psi}(t)) \leq \|(g(t), \psi_t(t))\|_{\mathcal{H}_0}^2 + (\lambda(t)/\mu(t))^k \leq C \mathbf{d}_+(\vec{\psi}(t)). \quad (3.6)$$

Moreover,

$$\|(g(t), \psi_t(t))\|_{\mathcal{H}_0} \leq C \left( \frac{\lambda(t)}{\mu(t)} \right)^{\frac{k}{2}}, \quad (3.7)$$

and hence

$$\mathbf{d}_+(\vec{\psi}(t)) \simeq \left( \frac{\lambda(t)}{\mu(t)} \right)^k. \quad (3.8)$$

**Remark 3.2** The following version of the implicit function theorem will be used in the proof.

Let  $X, Y, Z$  be Banach spaces. Let  $(x_0, y_0) \in X \times Y$ , let  $\delta_1, \delta_2 > 0$  and consider a mapping  $G : B(x_0, \delta_1) \times B(y_0, \delta_2) \rightarrow Z$  that is continuous in  $x$  and  $C^1$  in  $y$ . Suppose that  $G(x_0, y_0) = 0$  and  $(D_y G)(x_0, y_0)$  has bounded inverse  $L_0$ . Moreover, suppose that

$$\begin{aligned} \|L_0 - D_y G(x, y)\|_{\mathcal{L}(Y, Z)} &\leq \frac{1}{3\|L_0^{-1}\|_{\mathcal{L}(Z, Y)}} \\ \|G(x, y_0)\|_Z &\leq \frac{\delta_2}{3\|L_0^{-1}\|_{\mathcal{L}(Z, Y)}} \end{aligned} \quad (3.9)$$

for all  $\|x - x_0\|_X \leq \delta_1$  and  $\|y - y_0\|_Y \leq \delta_2$ . Then, there exists a continuous function  $\varsigma : B(x_0, \delta_1) \rightarrow B(y_0, \delta_2)$  such that for all  $x \in B(x_0, \delta_1)$ ,  $y = \varsigma(x)$  is the unique solution of  $G(x, \varsigma(x)) = 0$  in  $B(y_0, \delta_2)$ .

The above is proved in the same fashion as the usual implicit function theorem, see, e.g., [4, Section 2.2]. The key point is that the bounds (3.9) give uniform control on the size of the open set on which the Banach contraction mapping theorem can be applied.

*Proof* The proof follows by standard techniques that we outline below; we refer the reader to [23, Lemma 3.3] for a detailed proof of a similar statement.

We begin by showing that for each  $t \in J$  there exist unique  $\lambda(t)$ ,  $\mu(t)$ , and  $g(t)$  that satisfy (3.3), (3.6) and the orthogonality conditions (3.4) (3.5) using an argument based on the implicit function theorem stated in Remark 3.2. That  $\lambda(t)$  and  $\mu(t)$  are actually  $C^1(J)$  is then proved via a standard ODE argument, which we postpone until Remark 3.13 in Sect. 3.2.

To establish the former statement let  $\vec{\phi} \in \mathcal{H}_0$  be such that  $\mathbf{d}_+(\vec{\phi}) \leq \eta_0$ . This means we can find  $\lambda_0, \mu_0 > 0$  such that for  $g_0 \in H$  defined by

$$g_0 := \phi_0 - (Q_{\lambda_0} - Q_{\mu_0}) \quad (3.10)$$

we have

$$\|(g_0, \phi_1)\|_{H \times L^2}^2 + \left(\frac{\lambda_0}{\mu_0}\right)^k \leq 2\eta_0$$

Define the mapping  $F : H \times (0, \infty) \times (0, \infty) \rightarrow H$  by

$$F(g, \lambda, \mu) := g - (Q_\lambda - Q_\mu) + (Q_{\lambda_0} - Q_{\mu_0})$$

Note that  $F(0, \lambda_0, \mu_0) = 0$  and moreover that

$$\|F(g, \mu, \lambda)\|_H \lesssim \|g\|_H + |(\lambda/\lambda_0) - 1|^{\frac{1}{2}} + |(\mu/\mu_0) - 1|^{\frac{1}{2}}$$

Next define a mapping  $G : H \times (0, \infty) \times (0, \infty) \rightarrow \mathbb{R}^2$  by

$$G(g, \lambda, \mu) := \left( \left\langle \frac{1}{\lambda} \mathcal{Z}_\lambda \mid F(g, \lambda, \mu) \right\rangle, \left\langle \frac{1}{\mu} \mathcal{Z}_\mu \mid F(g, \lambda, \mu) \right\rangle \right)$$

For  $g \in H$  we have

$$\frac{1}{\lambda} \langle \mathcal{Z}_\lambda \mid g \rangle \leq \|r\lambda^{-1} \mathcal{Z}_\lambda\|_{L^2} \|r^{-1}g\|_{L^2} \lesssim \|g\|_H, \quad \frac{1}{\mu} \langle \mathcal{Z}_\mu \mid g \rangle \lesssim \|g\|_H$$

which ensures that the mapping  $G$  is well-defined and continuous. Taking the  $\lambda, \mu$  derivatives of  $G$ , we have

$$\begin{aligned} \frac{d}{d\lambda} \left\langle \frac{1}{\lambda} \mathcal{Z}_\lambda \mid F(g, \lambda, \mu) \right\rangle &= \frac{1}{\lambda} \langle \mathcal{Z}_\lambda \mid \Lambda Q_\lambda \rangle - \frac{1}{\lambda^2} \langle (\Lambda_0 + 1) \mathcal{Z}_\lambda \mid F(g, \lambda, \mu) \rangle \\ &=: \frac{1}{\lambda} A_{11}(g, \lambda, \mu) \\ \frac{d}{d\mu} \left\langle \frac{1}{\lambda} \mathcal{Z}_\lambda \mid F(g, \lambda, \mu) \right\rangle &= -\frac{1}{\lambda} \langle \mathcal{Z}_\lambda \mid \Lambda Q_\mu \rangle =: \frac{1}{\lambda} A_{12}(g, \lambda, \mu) \end{aligned} \quad (3.11)$$

and

$$\begin{aligned} \frac{d}{d\lambda} \left\langle \frac{1}{\mu} \mathcal{Z}_{\underline{\mu}} \mid F(g, \lambda, \mu) \right\rangle &= \frac{1}{\mu} \left\langle \mathcal{Z}_{\underline{\mu}} \mid \Lambda Q_{\underline{\lambda}} \right\rangle =: \frac{1}{\mu} A_{21}(g, \lambda, \mu) \\ \frac{d}{d\mu} \left\langle \frac{1}{\mu} \mathcal{Z}_{\underline{\mu}} \mid F(g, \lambda, \mu) \right\rangle &= -\frac{1}{\mu} \left\langle \mathcal{Z}_{\underline{\mu}} \mid \Lambda Q_{\underline{\mu}} \right\rangle - \frac{1}{\mu^2} \left\langle (\Lambda_0 + 1) \mathcal{Z}_{\underline{\mu}} \mid F(g, \lambda, \mu) \right\rangle \\ &=: \frac{1}{\mu} A_{22}(g, \lambda, \mu) \end{aligned} \quad (3.12)$$

For convenience in applying the implicit function theorem we change variables, setting  $\ell := \log \lambda$  and  $m := \log \mu$ . In the new variables we write

$$\tilde{G}(g, \ell, m) = G(g, \lambda, \mu), \quad \tilde{F}(g, \ell, m) = F(g, \lambda, \mu)$$

We now check that the conditions (3.9) are satisfied for  $x_0 = 0 \in H$ ,  $y_0 = (\ell_0, m_0) \in \mathbb{R}^2$  and  $\tilde{G} : B_H(0, 2\eta_0) \times B_{\mathbb{R}^2}((\ell_0, m_0), C_0\eta) \rightarrow \mathbb{R}^2$ , for  $\delta_1 = 2\eta_0 > 0$  small enough and  $C_0$  a uniform constant. Since  $\partial_\ell = \lambda \partial_\lambda$  and  $\partial_m = \mu \partial_\mu$  we deduce using (3.11), (3.12) that

$$D_{\ell, m} \tilde{G}(g, \ell, m) = \begin{pmatrix} A_{11}(g, \lambda, \mu) & A_{12}(g, \lambda, \mu) \\ A_{21}(g, \lambda, \mu) & A_{22}(g, \lambda, \mu) \end{pmatrix}$$

Restricting to  $(g, \lambda, \mu) = (0, \lambda_0, \mu_0)$ , this yields

$$\begin{aligned} L_0 &:= D_{\ell, m} \tilde{G}|_{(g=0, \ell=\ell_0, m=m_0)} = \begin{pmatrix} \left\langle \mathcal{Z}_{\underline{\lambda}_0} \mid \Lambda Q_{\underline{\lambda}_0} \right\rangle - \left\langle \mathcal{Z}_{\underline{\lambda}_0} \mid \Lambda Q_{\underline{\mu}_0} \right\rangle \\ \left\langle \mathcal{Z}_{\underline{\mu}_0} \mid \Lambda Q_{\underline{\lambda}_0} \right\rangle - \left\langle \mathcal{Z}_{\underline{\mu}_0} \mid \Lambda Q_{\underline{\mu}_0} \right\rangle \end{pmatrix} \\ &=: \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \end{aligned}$$

The diagonal terms in the matrix  $L_0$  above are size  $O(1)$  by (3.2). We can estimate the off-diagonal terms as follows:

**Claim 3.3** *For  $\lambda \ll \mu$  we have*

$$\left| \left\langle \mathcal{Z}_{\underline{\lambda}} \mid \Lambda Q_{\underline{\mu}} \right\rangle \right| \lesssim (\lambda/\mu)^{k+1}, \quad \left| \left\langle \mathcal{Z}_{\underline{\mu}} \mid \Lambda Q_{\underline{\lambda}} \right\rangle \right| \lesssim (\lambda/\mu)^{k-1} \quad (3.13)$$

To prove the Claim, without loss of generality we can assume  $\mu = 1$  and  $\lambda \ll 1$ . Let  $B > 0$  be such that  $\text{supp } \mathcal{Z} \subset \{r \leq B\}$ . Then, using (3.2) we have

$$\begin{aligned} \left| \left\langle \mathcal{Z}_{\underline{\lambda}} \mid \Lambda Q \right\rangle \right| &\lesssim \frac{1}{\lambda} \int_0^\infty \mathcal{Z}(r/\lambda) \Lambda Q(r) r \, dr \\ &\lesssim \frac{1}{\lambda} \int_0^\lambda (r/\lambda)^k \frac{r^{k+1}}{1+r^{2k}} \, dr + \frac{1}{\lambda} \int_\lambda^{B\lambda} \frac{r^{k+1}}{1+r^{2k}} \, dr \lesssim \lambda^{k+1} \end{aligned}$$

Similarly,

$$\begin{aligned} |\langle \mathcal{Z} \mid \Lambda Q_{\lambda} \rangle| &\lesssim \frac{1}{\lambda} \int_0^{\infty} \Lambda Q(r/\lambda) \mathcal{Z}(r) \, dr \\ &\lesssim \lambda^{k-1} \int_0^{\lambda} \frac{r^k}{\lambda^{2k} + r^{2k}} r^{k+1} \, dr + \lambda^{k-1} \int_{\lambda}^B \frac{\mathcal{Z}(r)}{r^k} \frac{r^{2k+1}}{\lambda^{2k} + r^{2k}} \, dr \\ &\lesssim \lambda^{-k-1} \int_0^{\lambda} r^{2k+1} \, dr + \lambda^{k-1} \lesssim \lambda^{k+1} + \lambda^{k-1} \lesssim \lambda^{k-1} \end{aligned}$$

which proves Claim 3.3.

This proves that the off diagonal terms in  $L_0$  are of size  $O((\lambda_0/\mu_0)^{k-1})$ . Hence for  $k \geq 2$  the matrix  $L_0$  is invertible as long as  $(\lambda_0/\mu_0)^{k-1}$  is small enough.

The second condition in (3.9) is clear since  $F(g, \lambda_0, \mu_0) = g$  and hence

$$|G(g, \lambda_0, \mu_0)| = \left| \left( \frac{1}{\lambda} \langle \mathcal{Z}_{\lambda} \mid g \rangle, \frac{1}{\mu} \langle \mathcal{Z}_{\mu} \mid g \rangle \right) \right| \lesssim \|g\|_H$$

The first condition in (3.9) follows from a direct computation checking that for  $1 \leq j, k \leq 2$ ,

$$|A_{ij}(g, \lambda, \mu) - A_{ij}| \lesssim (|\lambda/\lambda_0 - 1|^{\frac{1}{2}} + |\mu/\mu_0 - 1|^{\frac{1}{2}} + \|g\|_H) \ll 1$$

as long as  $\eta_0 > 0$  is chosen small enough. Here let us just remark that the factors involving  $\lambda/\lambda_0$  and  $\mu/\mu_0$  on the right-hand-side above appear from the estimates

$$\|\Lambda Q_{\sigma} - \Lambda Q_{\sigma_0}\|_{L^2} + \|\mathcal{Z}_{\sigma} - \mathcal{Z}_{\sigma_0}\|_{L^2} \lesssim |\sigma/\sigma_0 - 1|^{\frac{1}{2}}$$

An application of Remark 3.2 yields the following: There exists  $\eta_0 > 0$  small enough and a continuous mapping  $\varsigma : B_H(0, 2\eta_0) \rightarrow B_{\mathbb{R}^2}((\ell_0, m_0), C_0\eta_0)$  so that for all  $(g, \ell, m) \in B_H(0, \eta_0) \times B_{\mathbb{R}^2}((\ell_0, m_0), C_0\eta_0)$  we have

$$G(g, \lambda, \mu) \equiv 0 \iff (\ell, m) = \varsigma(g), \quad \lambda = e^{\ell}, \quad \mu = e^m$$

Finally, we observe that if we let  $g_0$  be as in (3.10), and define  $(\lambda, \mu) = (e^{\ell}, e^m)$  and  $g \in H$  by

$$(\ell, m) := \varsigma(g_0), \quad g := F(g_0, \lambda, \mu)$$

we see that  $\phi_0 = Q_{\lambda} - Q_{\mu} + g$  and moreover that

$$\langle \mathcal{Z}_{\lambda} \mid g \rangle = 0 \quad \text{and} \quad \langle \mathcal{Z}_{\mu} \mid g \rangle = 0.$$

Lastly, since  $|\ell - \ell_0| \leq C_0\eta_0$  and  $|m - m_0| \leq C_0\eta_0$ , we have

$$|\lambda/\lambda_0 - 1| + |\mu/\mu_0 - 1| \lesssim \eta_0 \ll 1, \quad \|g\|_H \lesssim \sqrt{\eta_0} \ll 1.$$

In particular, we have

$$(\lambda/\mu)^k \lesssim (\lambda_0/\mu_0)^k \leq \mathbf{d}_+(\vec{\phi}). \quad (3.14)$$

Now we establish the estimate (3.7). This follows by expanding the nonlinear energy. We'll make use of trigonometric identities here for simplicity but note that the following computation relies only on the fact that the nonlinearity in (1.5) is smooth, that  $Q$  is a solution, and the orthogonality conditions (3.4) (3.5).

$$\begin{aligned} \frac{2}{\pi} \mathcal{E}(Q) &= \frac{1}{\pi} \mathcal{E}(\vec{\phi}) = \frac{2}{\pi} \mathcal{E}(Q) + \int_0^\infty g_r^2 r \, dr + \int_0^\infty \psi_t^2 r \, dr \\ &\quad - 2 \int_0^\infty \partial_r Q_\lambda \partial_r Q_\mu r \, dr + 2 \int_0^\infty \partial_r Q_\lambda g_r r \, dr - 2 \int_0^\infty \partial_r Q_\mu g_r r \, dr \\ &\quad + k^2 \int_0^\infty \frac{\sin^2(Q_\lambda - Q_\mu + g)}{r^2} r \, dr - k^2 \int_0^\infty \frac{(\sin^2 Q_\lambda + \sin^2 Q_\mu)}{r^2} r \, dr \end{aligned} \quad (3.15)$$

We expand the nonlinear terms on the last line using trigonometric identities

$$\begin{aligned} \sin^2(Q_\lambda - Q_\mu + g) &= \sin^2(Q_\lambda - Q_\mu) + \frac{1}{2} \sin 2g \sin 2(Q_\lambda - Q_\mu) \\ &\quad + \sin^2 g \cos 2(Q_\lambda - Q_\mu) \\ &= \sin^2 Q_\lambda + \sin^2 Q_\mu - \frac{1}{2} \sin 2Q_\lambda \sin 2Q_\mu \\ &\quad - 2 \sin^2 Q_\lambda \sin^2 Q_\mu + g \sin 2(Q_\lambda - Q_\mu) \\ &\quad + g^2 \cos 2(Q_\lambda - Q_\mu) + O(|g|^3) \end{aligned}$$

which further reduces to

$$\begin{aligned} &= \sin^2 Q_\lambda + \sin^2 Q_\mu + g^2 \cos 2(Q_\lambda - Q_\mu) - 2 \sin^2 Q_\lambda \sin^2 Q_\mu \\ &\quad + g \sin 2Q_\lambda - g \sin 2Q_\mu - \sin 2Q_\lambda Q_\mu + \frac{1}{2} \sin 2Q_\lambda [2Q_\mu - \sin 2Q_\mu] \\ &\quad - g(2 \sin 2Q_\lambda \sin^2 Q_\mu - 2 \sin 2Q_\mu \sin^2 Q_\lambda) + O(|g|^3) \end{aligned} \quad (3.16)$$

Next we observe that the first three terms in the second line of (3.16) will give exact cancelations with the terms in the second line of (3.15). Indeed, using the identity

$$\frac{1}{r} \partial_r (r \partial_r Q_\lambda) = k^2 \frac{\sin 2Q_\lambda}{2r^2}$$

we integrate by parts to obtain

$$\begin{aligned} k^2 \int_0^\infty \frac{\sin 2Q_\lambda}{r^2} Q_\mu r \, dr &= 2 \int_0^\infty \frac{1}{r} \partial_r (r \partial_r Q_\lambda) Q_\mu r \, dr = -2 \int_0^\infty \partial_r Q_\lambda \partial_r Q_\mu r \, dr \\ k^2 \int_0^\infty \frac{\sin 2Q_\lambda}{r^2} g r \, dr &= 2 \int_0^\infty \frac{1}{r} \partial_r (r \partial_r Q_\lambda) g r \, dr = -2 \int_0^\infty \partial_r Q_\lambda g r \, dr \\ k^2 \int_0^\infty \frac{\sin 2Q_\mu}{r^2} g r \, dr &= 2 \int_0^\infty \frac{1}{r} \partial_r (r \partial_r Q_\mu) g r \, dr = -2 \int_0^\infty \partial_r Q_\mu g r \, dr \end{aligned}$$

We can use the same identity to integrate by parts the terms arising from the rest of (3.16).

$$\begin{aligned} -2k^2 \int_0^\infty \frac{\sin 2Q_\lambda}{r^2} \sin^2 Q_\mu g r \, dr &= -4 \int_0^\infty \frac{1}{r} \partial_r (r \partial_r Q_\lambda) \sin^2 Q_\mu g r \, dr \\ &= 4 \int_0^\infty \partial_r Q_\lambda \partial_r Q_\mu \sin 2Q_\mu g r \, dr + 4 \int_0^\infty \partial_r Q_\lambda \sin^2 Q_\mu g r \, dr \end{aligned}$$

and the same for the symmetric term in  $\lambda, \mu$ . Finally, we have

$$\begin{aligned} k^2 \int_0^\infty \frac{\sin 2Q_\lambda}{2r^2} [2Q_\mu - \sin 2Q_\mu] r \, dr &= \int_0^\infty \frac{1}{r} \partial_r (r \partial_r Q_\lambda) [2Q_\mu - \sin 2Q_\mu] r \, dr \\ &= - \int_0^\infty \partial_r Q_\lambda \partial_r [2Q_\mu - \sin 2Q_\mu] r \, dr = -4 \int_0^\infty \partial_r Q_\lambda \partial_r Q_\mu \sin^2 Q_\mu r \, dr \end{aligned}$$

Therefore, using the above, (3.16), the identity  $\Lambda Q = k \sin Q$ , and the fact that  $\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(Q)$ , we can deduce from (3.15) that

$$\begin{aligned} &\int_0^\infty \psi_t^2 r \, dr + \int_0^\infty g_r^2 r \, dr + k^2 \int_0^\infty \frac{\cos 2(Q_\lambda - Q_\mu)}{r^2} g^2 r \, dr \\ &= \frac{4}{k^2} \int_0^\infty \Lambda Q_\lambda (\Lambda Q_\mu)^3 \frac{dr}{r} - \frac{2}{k^2} \int_0^\infty (\Lambda Q_\lambda)^2 (\Lambda Q_\mu)^2 \frac{dr}{r} \\ &\quad - 4 \int_0^\infty \Lambda Q_\lambda \Lambda Q_\mu \sin 2Q_\mu g \frac{dr}{r} - \frac{4}{k^2} \int_0^\infty \Lambda Q_\lambda (\Lambda Q_\mu)^2 (r g_r) \frac{dr}{r} \\ &\quad + 4 \int_0^\infty \Lambda Q_\lambda \Lambda Q_\mu \sin 2Q_\lambda g \frac{dr}{r} + \frac{4}{k^2} \int_0^\infty \Lambda Q_\mu (\Lambda Q_\lambda)^2 (r g_r) \frac{dr}{r} \\ &\quad + O\left(\int_0^\infty |g|^3 \frac{dr}{r}\right) \end{aligned} \tag{3.17}$$

Next, we estimate each of the terms on the right-hand-side of (3.17). Denote  $\sigma := \lambda/\mu$ . We claim that first term on the right-hand-side of (3.17) gives the leading order, i.e., we claim that

$$\frac{4}{k^2} \int_0^\infty \Lambda Q_\lambda (\Lambda Q_\mu)^3 \frac{dr}{r} = 16k\sigma^k (1 + O(\sigma^{2k})) \quad (3.18)$$

We compute, using the identity  $k \sin Q_\lambda = \Lambda Q_\lambda$ , and setting  $\sigma = \lambda/\mu$ ,

$$\begin{aligned} \frac{4}{k^2} \int_0^\infty \Lambda Q_\lambda (\Lambda Q_\mu)^3 \frac{dr}{r} &= \frac{4}{k^2} \int_0^\infty \Lambda Q_\sigma (\Lambda Q)^3 \frac{dr}{r} \\ &= 64k^2 \sigma^k \int_0^\infty \frac{r^{4k-1}}{(\sigma^{2k} + r^{2k})(1 + r^{2k})^3} dr \end{aligned} \quad (3.19)$$

First we estimate the contribution of the integral on the interval  $[0, \sigma]$ . Since we can assume that  $\sigma \ll 1$ , we have

$$\int_0^\sigma \frac{r^{4k-1}}{(\sigma^{2k} + r^{2k})(1 + r^{2k})^3} dr \simeq \sigma^{-2k} \int_0^\sigma r^{4k-1} dr \simeq \sigma^{2k}$$

Next, we estimate the integral on  $[\sigma, \infty]$ . For  $\sigma < r$  we have

$$\begin{aligned} \frac{1}{\sigma^{2k} + r^{2k}} &= \frac{1}{r^{2k}} + \left( \frac{1}{\sigma^{2k} + r^{2k}} - \frac{1}{r^{2k}} \right) \\ &= \frac{1}{r^{2k}} + \frac{1}{r^{2k}} \left( \frac{1}{1 + (\sigma/r)^{2k}} - 1 \right) \\ &= \frac{1}{r^{2k}} + \frac{1}{r^{2k}} \left( -(\sigma/r)^{2k} + O((\sigma/r)^{4k}) \right) \end{aligned}$$

Hence,

$$\begin{aligned} \int_\sigma^\infty \frac{r^{4k-1}}{(\sigma^{2k} + r^{2k})(1 + r^{2k})^3} dr &= \int_0^\infty \frac{r^{2k-1}}{(1 + r^{2k})^3} dr \\ &\quad - \int_0^\sigma \frac{r^{2k-1}}{(1 + r^{2k})^3} dr + O(\sigma^{2k}) \\ &= \int_0^\infty \frac{r^{2k-1}}{(1 + r^{2k})^3} dr + O(\sigma^{2k}) \\ &= \frac{1}{4k} + O(\sigma^{2k}) \end{aligned}$$

where the integral on the second to last line can be computed explicitly by contour integration. Inserting the above into the last line of (3.19) yields (3.18).

Next we observe that all of the remaining terms on the right-hand-side of (3.17) are  $o(\sigma^k)$ . Indeed, a similar computation to the one performed above yields,

$$\int_0^\infty (\Lambda Q_\lambda)^2 (\Lambda Q_\mu)^2 \frac{dr}{r} \lesssim \sigma^{2k} |\log \sigma|$$

Moreover since  $\|g\|_{L^\infty} \lesssim \|g\|_H$  we have

$$\int_0^\infty |g|^3 \frac{dr}{r} \lesssim \|g\|_H^3 \lesssim \sqrt{\eta_0} \|g\|_H^2$$

And the remaining terms in (3.17) can be controlled by a combination of these last two estimates together with Cauchy–Schwarz. Therefore as long as  $\eta_0$  is small enough and since  $\sigma = \lambda/\mu \lesssim \eta_0$  we have

$$\begin{aligned} & \int_0^\infty \psi_t^2 r \, dr + \int_0^\infty g_r^2 r \, dr + k^2 \int_0^\infty \frac{\cos 2(Q_\lambda - Q_\mu)}{r^2} g^2 r \, dr \\ &= 16k\sigma^k - O(\sigma^{\frac{3k}{2}} |\log \sigma|) \end{aligned}$$

To complete the proof of (3.7) we claim the following coercivity statement: there exists a uniform constant  $c > 0$  so that

$$\int_0^\infty g_r^2 r \, dr + k^2 \int_0^\infty \frac{\cos 2(Q_\lambda - Q_\mu)}{r^2} g^2 r \, dr \geq c \|g\|_H^2$$

for all  $g \in H$  such that (3.4) (3.5) hold and such that  $\|g\|_H$  is small enough. This is a standard consequence of the orthogonality conditions (3.4), (3.5) and the smallness of  $\lambda/\mu$ ,  $\|g\|_H$  and we refer the reader to [22, Lemma 5.4] for a detailed proof.

The left inequality in (3.6) is trivial and the right inequality follows from (3.7) and (3.14).  $\square$

## 3.2 Dynamical control of the modulation parameters

In this section we obtain precise control of the evolution of the modulation parameters  $\lambda(t)$ ,  $\mu(t)$  on any time interval  $J$  on which  $\mathbf{d}_+(\vec{\psi}(t))$  is small. We'll show that any solution  $\vec{\psi}(t)$  that lies within a small enough  $\epsilon$ -neighborhood of a 2-bubble at some time  $t_0$ , must be ejected from this  $\epsilon$ -neighborhood in at least one time direction.

This ejection happens by a defocalisation of the more concentrated bubble  $Q_\lambda$  until its scale becomes comparable with the less concentrated bubble  $Q_\mu$  (which does not change in the process). The influence of the bubble  $Q_\mu$  on the evolution of  $Q_\lambda$  is reflected in the time derivative of the function  $b(t)$  defined in (3.32) below. Indeed, the main term of  $b'(t)$  is given precisely by the interaction between the two bubbles, see (3.39). The main term of  $b(t)$  is

related to  $\lambda'(t)$ . Hence  $b'(t)$  is related to  $\lambda''(t)$  so that the interaction influences the acceleration, as it should be expected.

In this subsection we define a truncated virial functional and state some estimates related to it. The same functional was used crucially in the two-bubble construction by the first author in [22]. For the proofs of the following statements we refer the reader to [22, Lemma 4.6] and [22, Lemma 5.5].

**Lemma 3.4** [22, Lemma 4.6] *For each  $c, R > 0$  there exists a function  $q(r) = q_{c,R}(r) \in C^{3,1}((0, +\infty))$  with the following properties:*

- (P1)  $q(r) = \frac{1}{2}r^2$  for  $r \leq R$ ,
- (P2) there exists  $\tilde{R} = \tilde{R}(R, c) > R$  such that  $q(r) \equiv \text{const}$  for  $r \geq \tilde{R}$ ,
- (P3)  $|q'(r)| \lesssim r$  and  $|q''(r)| \lesssim 1$  for all  $r > 0$ , with constants independent of  $c, R$ ,
- (P4)  $q''(r) \geq -c$  and  $\frac{1}{r}q'(r) \geq -c$ , for all  $r > 0$ ,
- (P5)  $\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr}\right)^2 q(r) \leq c \cdot r^{-2}$ , for all  $r > 0$ ,
- (P6)  $|r(\frac{q'(r)}{r})'| \leq c$ , for all  $r > 0$ .

For each  $\lambda > 0$  we define the operators  $\mathcal{A}(\lambda)$  and  $\mathcal{A}_0(\lambda)$  as follows:

$$[\mathcal{A}(\lambda)g](r) := q' \left( \frac{r}{\lambda} \right) \cdot \partial_r g(r), \quad (3.20)$$

$$[\mathcal{A}_0(\lambda)g](r) := \left( \frac{1}{2\lambda} q'' \left( \frac{r}{\lambda} \right) + \frac{1}{2r} q' \left( \frac{r}{\lambda} \right) \right) g(r) + q' \left( \frac{r}{\lambda} \right) \cdot \partial_r g(r). \quad (3.21)$$

Note the similarity between  $\mathcal{A}$  and  $\frac{1}{\lambda}\Lambda$  and between  $\mathcal{A}_0$  and  $\frac{1}{\lambda}\Lambda_0$ . For technical reasons we introduce the space

$$X := \left\{ g \in H \mid \frac{g}{r}, \partial_r g \in H \right\}$$

Let  $f(\rho) := \frac{k^2}{2} \sin 2\rho$  denote the nonlinearity in (1.5).

**Lemma 3.5** [22, Lemma 5.5] *Let  $c_0 > 0$  be arbitrary. There exists  $c > 0$  small enough and  $R, \tilde{R} > 0$  large enough in Lemma 3.4 so that the operators  $\mathcal{A}(\lambda)$  and  $\mathcal{A}_0(\lambda)$  defined in (3.20) and (3.21) have the following properties:*

- the families  $\{\mathcal{A}(\lambda) : \lambda > 0\}$ ,  $\{\mathcal{A}_0(\lambda) : \lambda > 0\}$ ,  $\{\lambda \partial_\lambda \mathcal{A}(\lambda) : \lambda > 0\}$  and  $\{\lambda \partial_\lambda \mathcal{A}_0(\lambda) : \lambda > 0\}$  are bounded in  $\mathcal{L}(H; L^2)$ , with the bound depending only on the choice of the function  $q(r)$ ,
- For all  $\lambda > 0$  and  $g_1, g_2 \in X$  there holds

$$\left| \left\langle \mathcal{A}(\lambda)g_1 \mid \frac{1}{r^2} (f(g_1 + g_2) - f(g_1) - f'(g_1)g_2) \right\rangle \right. \\ \left. + \left\langle \mathcal{A}(\lambda)g_2 \mid \frac{1}{r^2} (f(g_1 + g_2) - f(g_1) - k^2 g_2) \right\rangle \right| \leq \frac{c_0}{\lambda} \|g_2\|_H^2, \quad (3.22)$$

- For all  $g \in X$  we have

$$\left\langle \mathcal{A}_0(\lambda)g \mid \left( \partial_r^2 + \frac{1}{r} \partial_r - \frac{k^2}{r^2} \right) g \right\rangle \\ \leq \frac{c_0}{\lambda} \|g\|_H^2 - \frac{1}{\lambda} \int_0^{R\lambda} \left( (\partial_r g)^2 + \frac{k^2}{r^2} g^2 \right) r dr, \quad (3.23)$$

- Moreover, for  $\lambda, \mu > 0$  with  $\lambda/\mu \ll 1$ ,

$$\|\Lambda_0 \Lambda Q_{\underline{\lambda}} - \mathcal{A}_0(\lambda) \Lambda Q_{\lambda}\|_{L^2} \leq c_0, \quad (3.24)$$

$$\|\Lambda Q_{\underline{\lambda}} - \mathcal{A}(\lambda) Q_{\lambda}\|_{L^\infty} \leq \frac{c_0}{\lambda}, \quad (3.25)$$

$$\|\mathcal{A}(\lambda) Q_{\mu}\|_{L^\infty} + \|\mathcal{A}_0(\lambda) Q_{\mu}\|_{L^\infty} \lesssim \frac{1}{\lambda} (\lambda/\mu)^k, \quad (3.26)$$

and, for any  $g \in H$ ,

$$\left| \int_0^{+\infty} \frac{1}{2} \left( q''\left(\frac{r}{\lambda}\right) + \frac{\lambda}{r} q'\left(\frac{r}{\lambda}\right) \right) \frac{1}{r^2} (f(-Q_{\mu} + Q_{\lambda} + g) \right. \right. \\ \left. - f(-Q_{\mu} + Q_{\lambda}) - k^2 g) g r dr \right. \\ \left. - \int_0^{+\infty} \frac{1}{r^2} (f'(Q_{\lambda}) - k^2) g^2 r dr \right| \leq c_0 (\|g\|_H^2 + (\lambda/\mu)^k). \quad (3.27)$$

**Remark 3.6** The conditions  $g, g_1, g_2 \in X$  is required only to ensure that the left-hand-side of (3.22) and (3.23) are well defined, but do not appear on the right-hand-side of the estimates. Note also that in (3.22), (3.23) and (3.27) we have extracted the linear part of  $f$ . Lastly, the estimate (3.26) is not stated in [22] but follows immediately from  $P_2, P_3$  in Lemma 3.4 and the explicit formula for  $Q$ .

We are now ready to state the main modulation estimates. Proofs are given in Sect. 3.3. Our first estimate is a consequence of the orthogonality conditions (3.4) and (3.5).

**Proposition 3.7** (Modulation Control Part 1). *Let  $\eta_0 > 0$  be as in Lemma 3.1, let  $J \subset \mathbb{R}$  be a time interval, and let  $\vec{\psi}(t)$  be a solution to (1.5) on  $J$  such that*

$$\mathbf{d}(\vec{\psi}(t)) \leq \eta_0 \quad \forall t \in J.$$

*Let  $\lambda(t), \mu(t)$  be given by Lemma 3.1. Then the following estimates hold for  $t \in J$ :*

$$|\lambda'(t)| \lesssim \lambda(t)^{\frac{k}{2}} / \mu(t)^{\frac{k}{2}}, \quad (3.28)$$

$$|\mu'(t)| \lesssim \lambda(t)^{\frac{k}{2}} / \mu(t)^{\frac{k}{2}}, \quad (3.29)$$

The control we obtain on  $\lambda(t), \mu(t)$  above is not sufficient for our purposes. In particular, we'd like to show that the ratio  $\lambda(t)/\mu(t)$  grows in a controlled fashion away from any small enough local minimum value. For this purpose we introduce a virial-type correction  $b(t)$  to  $\lambda'(t)$ . The idea of modifying a modulation parameter by a virial term was used in [24] and, in a different context of minimal mass blow-up for non-homogeneous  $L^2$ -critical NLS, in an earlier work of Raphaël and Szeftel [45].

Given scaling parameters  $\lambda(t), \mu(t)$  we write

$$\begin{aligned} g(t) &:= \psi(t) - Q_{\lambda(t)} + Q_{\mu(t)} \\ \dot{g}(t) &:= \psi_t(t) \end{aligned}$$

so that the vector  $\vec{g} := (g, \dot{g})$  satisfies the system of equations

$$\partial_t g = \dot{g} + \lambda' \Lambda Q_{\underline{\lambda}} - \mu' \Lambda Q_{\underline{\mu}} \quad (3.30)$$

$$\partial_t \dot{g} = \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_{\lambda} - Q_{\mu} + g) - f(Q_{\lambda}) + f(Q_{\mu})) \quad (3.31)$$

We then define the auxiliary function  $b(t)$  by

$$b(t) := -\left\langle \Lambda Q_{\underline{\lambda(t)}} \mid \dot{g}(t) \right\rangle - \langle \dot{g}(t) \mid \mathcal{A}_0(\lambda(t))g(t) \rangle \quad (3.32)$$

We'll show below that we can think of  $b(t)$  as a subtle monotonic correction to the derivative  $\lambda'(t)$ .

Before stating the estimates satisfied by  $b(t)$  we record the following numbers, which can be computed using contour integration:

$$\|\Lambda Q\|_2^2 = \frac{2\pi i \operatorname{Res}[(\Lambda Q(z))^2 z; \omega_k]}{1 - \omega_k^4}, \quad \omega_k := \exp(2\pi i / 4k)$$

which means that

$$\|\Lambda Q\|_2^2 = \frac{2\pi}{\sin(\pi/k)} =: \kappa = \kappa(k) > 0$$

We will also use fact that

$$\int_0^\infty (\Lambda Q(r))^3 r^{k-1} dr = 2k^2 \quad (3.33)$$

Lastly, the modulation parameter  $\lambda(t)$  itself is an imprecise proxy for the true dynamics because it was defined with respect to a somewhat arbitrary function  $\mathcal{Z}$  as in (3.2). To account for this imprecision we introduce a correction to  $\lambda(t)$  as follows. Fix a radial cutoff  $\chi \in C_0^\infty(\mathbb{R}^2)$  such that  $\chi(r) = 1$  if  $r \leq 1$ ,  $\text{supp}(\chi) \in B(0, 2)$ . Define

$$\zeta(t) := \lambda(t) - \frac{1}{\kappa} \langle \chi_{\mu(t)} \Lambda Q_{\lambda(t)} \mid g(t) \rangle \quad (3.34)$$

Note that  $\zeta(t)$  is  $C^1$  (because  $\partial_t g(t)$  is continuous in  $L^2$  with respect to  $t$ ).

**Proposition 3.8** (Modulation Control Part 2). *Fix  $k \geq 2$ . Assume the same hypothesis as in Proposition 3.7. Let  $0 < \delta < 1 - 2^{-\frac{1}{k-1}}$  be arbitrary and let  $\eta_0$  be as in Lemma 3.1. Let  $b(t)$  be as in (3.32) and let  $\zeta(t)$  be as in (3.34). Then, there exists  $\eta_1 = \eta_1(\delta) < \eta_0$  such that if  $\mathbf{d}_+(\vec{\psi}(t)) \leq \eta_1$  for all  $t \in J$  we have*

$$|\zeta(t)/\lambda(t) - 1| \leq \delta, \quad (3.35)$$

$$|\zeta'(t) - b(t)| \leq \delta \lambda(t)^{\frac{k}{2}} / \mu(t)^{\frac{k}{2}} \leq 2\delta \zeta(t)^{\frac{k}{2}} / \mu(t)^{\frac{k}{2}}, \quad (3.36)$$

$$|b(t)| \leq 4\sqrt{\kappa k} (\lambda(t)/\mu(t))^{\frac{k}{2}} + \delta \lambda(t)^{\frac{k}{2}} / \mu(t)^{\frac{k}{2}} \leq 10\sqrt{\kappa k} \zeta(t)^{\frac{k}{2}} / \mu(t)^{\frac{k}{2}}, \quad (3.37)$$

In addition,  $b(t)$  is locally Lipschitz and the derivative  $b'(t)$  satisfies

$$|b'(t)| \leq C_0 \lambda(t)^{k-1} / \mu(t)^k \leq 2C_0 \zeta(t)^{k-1} / \mu(t)^k \quad (3.38)$$

$$b'(t) \geq 8k^2 \lambda^{k-1}(t) / \mu^k(t) - \delta \lambda^{k-1}(t) / \mu(t)^k \geq 2k^2 \zeta(t)^{k-1} / \mu(t)^k \quad (3.39)$$

where  $C_0 > 0$  depends only on  $k$ .

**Remark 3.9** If  $k \geq 3$ , then we can take  $\mathcal{Z} = \Lambda Q$  in Lemma 3.1 and use no cut-off function in the definition of  $\zeta$ . Then  $\zeta(t) \equiv \lambda(t)$ . This fails for  $k = 2$ , which was pointed out to us by one of the referees.

We'll deduce the following consequence of Propositions 3.7 and 3.8.

**Proposition 3.10** *Let  $C > 0$ . For any  $\epsilon_0 > 0$  small enough, and for all  $\epsilon > 0$  sufficiently small relative to  $\epsilon_0$  the following conclusions hold true. Let  $\vec{\psi}(t) : [T_0, T_+) \rightarrow \mathcal{H}_0$  be a solution of (1.5). Assume that  $t_0 \in [T_0, T_+)$  is such that  $\mathbf{d}(\vec{\psi}(t_0)) \leq \epsilon$  and  $\frac{d}{dt}(\zeta(t)/\mu(t))|_{t=t_0} \geq 0$ . Then there exist  $t_1$  and  $t_2$ ,  $T_0 \leq t_0 \leq t_1 \leq t_2 < T_+$ , such that*

$$\mathbf{d}(\vec{\psi}(t)) \geq 2\epsilon, \quad \text{for } t \in [t_1, t_2], \quad (3.40)$$

$$\mathbf{d}(\vec{\psi}(t)) \leq \frac{1}{4}\epsilon_0, \quad \text{for } t \in [t_0, t_1], \quad (3.41)$$

$$\mathbf{d}(\vec{\psi}(t_2)) \geq 2\epsilon_0, \quad (3.42)$$

$$\int_{t_1}^{t_2} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq C \int_{t_0}^{t_1} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt \quad (3.43)$$

Analogous statements hold with times  $t_2 \leq t_1 \leq t_0$  if  $\frac{d}{dt}(\zeta(t)/\mu(t))|_{t=t_0} \leq 0$ .

*Remark 3.11* We will take  $\epsilon < \eta_0$ , so that  $\mathbf{d}(\vec{\psi}(t_0)) \leq \epsilon$  implies that the modulation parameters  $\lambda(t)$ ,  $\mu(t)$  and also  $\zeta(t)$  are well-defined  $C^1$  functions in a neighborhood of  $t = t_0$ .

*Remark 3.12* We will actually deduce (3.43) from the following stronger statement. There exist  $\epsilon_0, C_k$  depending only on  $k$  such that for any  $\epsilon > 0$  small enough, Proposition 3.10 holds and additionally

$$\begin{aligned} \int_{t_0}^{t_1} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt &\leq C_k \epsilon^{\frac{1}{k}}, \\ \int_{t_1}^{t_2} \|\partial_t \psi(t)\|_{L^2}^2 dt &\geq \frac{1}{C_k}. \end{aligned} \quad (3.44)$$

### 3.3 Proofs of the modulation estimates

We first assume the conclusions of Propositions 3.7 and 3.8 and prove Proposition 3.10. We record here a few useful formulae:

$$\Lambda Q := r \partial_r Q = k \sin Q = \frac{2kr^k}{1+r^{2k}} \quad (3.45)$$

$$\Lambda^2 Q = \frac{k^2}{2} \sin 2Q = 2k^2 r^k \left( \frac{1-r^{2k}}{(1+r^{2k})^2} \right) \quad (3.46)$$

$$\Lambda^3 Q = 2k^3 r^k \left( \frac{1+r^{2k}-5r^{4k}-r^{6k}}{(1+r^{2k})^4} \right)$$

$$\Lambda_0 \Lambda Q = (r \partial_r + 1)(r \partial_r Q) = 2\Lambda Q + r^2 \partial_r^2 Q \quad (3.47)$$

*Proof of Proposition 3.10* From (3.7) and (3.35) it follows that there exists  $\epsilon_1 > 0$  such that if  $\zeta(t)/\mu(t) \leq \epsilon_1$ , the modulation estimates hold in a neighborhood of time  $t$ . If needed, we will assume that  $\epsilon_1$  is sufficiently small, but depending only on  $k$ . Let  $t_2$  be the first time  $t_2 \geq t_0$  such that  $\zeta(t_2)/\mu(t_2) = \epsilon_1$  (if there is no such time, we set  $t_2 = T_+$ ). By the estimate (3.8) in Lemma 3.1 along with (3.35), which accounts for the difference between  $\lambda(t)$  and  $\zeta(t)$ , there exists  $\epsilon_0$  sufficiently small such that  $\zeta(t_2)/\mu(t_2) = \epsilon_1$  implies (3.42). The number  $\epsilon > 0$  will be chosen later in the proof and should be thought of as being much smaller than  $(\epsilon_1)^{\frac{k}{2}}$ , whereas we can think of  $\epsilon_0$  as comparable to  $(\epsilon_1)^{\frac{k}{2}}$ .

Without loss of generality we can assume that  $\mu(t_0) = 1$ . Let  $t_3 \leq t_2$  be the last time such that  $\mu(t) \in [\frac{1}{2}, 2]$  for all  $t \in [t_0, t_3]$ . If there is no such final time we set  $t_3 = t_2$ . (Later we will see that we can always take  $t_3 = t_2$  as long as  $\epsilon_1 > 0$  is small enough.)

For  $t \in [t_0, t_3]$ , from (3.39) we obtain

$$b'(t) \geq \frac{k^2}{2^{k-1}} \zeta(t)^{k-1}. \quad (3.48)$$

We also obtain from (3.36)

$$\zeta'(t) \geq \frac{1}{\kappa} b(t) - \sqrt{\frac{k}{2^{k-1}\kappa}} \zeta(t)^{\frac{k}{2}}.$$

Let  $\kappa_1 := \sqrt{\frac{k\kappa}{2^{k-1}}}$  and consider  $\xi(t) := b(t) + \kappa_1 \zeta(t)^{\frac{k}{2}}$ . Using the two inequalities above we obtain

$$\begin{aligned} \xi'(t) &\geq \frac{k^2}{2^{k-1}} \zeta(t)^{k-1} + \kappa_1 \frac{k}{2} \zeta(t)^{\frac{k}{2}-1} \left( \frac{1}{\kappa} b(t) - \sqrt{\frac{k}{2^{k-1}\kappa}} \zeta(t)^{\frac{k}{2}} \right) \\ &= \frac{\kappa_1 k}{2\kappa} \zeta(t)^{\frac{k}{2}-1} b(t) + \left( \frac{k^2}{2^{k-1}} - \frac{k\kappa_1}{2} \sqrt{\frac{k}{2^{k-1}\kappa}} \right) \zeta(t)^{k-1} \\ &= k \sqrt{\frac{k}{\kappa 2^{k+1}}} \zeta(t)^{\frac{k}{2}-1} b(t) + \frac{k^2}{2^k} \zeta(t)^{k-1} \\ &= k \sqrt{\frac{k}{\kappa 2^{k+1}}} \zeta(t)^{\frac{k}{2}-1} \xi(t). \end{aligned} \quad (3.49)$$

It is easy to compute that (3.37) yields  $|b(t)| \leq \kappa_1 2^{k+4} \zeta(t)^{\frac{k}{2}}$ , so we have

$$\xi(t) \leq \kappa_1 2^{k+5} \zeta(t)^{\frac{k}{2}} \quad (3.50)$$

and (3.49) leads to

$$\xi'(t) \geq \kappa_2 \xi(t)^{\frac{2k-2}{k}}, \quad (3.51)$$

for some constant  $\kappa_2 > 0$  depending only on  $k$ .

Let  $\xi_1(t) := b(t) + \frac{\kappa_1}{2} \zeta(t)^{\frac{k}{2}} = \frac{1}{2} b(t) + \frac{1}{2} \xi(t)$ . Since  $b'(t) \geq 0$ , we have

$$\xi_1'(t) \geq \frac{1}{2} \xi'(t) \geq \frac{k}{2} \sqrt{\frac{k}{\kappa 2^{k+1}}} \zeta(t)^{\frac{k}{2}-1} \xi(t) \geq \frac{k}{2} \sqrt{\frac{k}{\kappa 2^{k+1}}} \zeta(t)^{\frac{k}{2}-1} \xi_1(t). \quad (3.52)$$

Since  $\mu(t_0) = 1$ , we have  $0 \leq \frac{d}{dt}(\lambda(t)/\mu(t))|_{t=t_0} = \zeta'(t_0) - \zeta(t_0)\mu'(t_0)$ , so (3.29) and (3.35) imply that  $\kappa \zeta'(t_0) \geq -\frac{\kappa_1}{4} \zeta(t_0)^{\frac{k}{2}}$  as long as  $\epsilon$  is taken small enough. Now (3.36) gives  $b(t_0) \geq -\frac{\kappa_1}{3} \zeta(t_0)^{\frac{k}{2}}$ , so  $\xi_1(t_0) > 0$  and (3.52) yields  $\xi_1(t) > 0$  for all  $t \in [t_0, t_3]$ . Thus

$$\xi(t) \geq \frac{\kappa_1}{2} \zeta(t)^{\frac{k}{2}}, \quad \text{for } t \in [t_0, t_3]. \quad (3.53)$$

In particular, (3.51) implies that  $\xi(t)$  is strictly increasing on  $[t_0, t_3]$  and by (3.50) we see that  $\zeta(t)$  and thus  $\lambda(t)$  is far from 0 on  $[t_0, t_3]$ .

Bounds (3.50) and (3.35) imply that there exists a constant  $\kappa_3$  depending only on  $k$  such that  $\xi(t) \geq \kappa_3 \sqrt{\epsilon}$  forces  $\mathbf{d}(\vec{\psi}(t)) \geq 2\epsilon$ . Let  $t_1 \in [t_0, t_3]$  be the last time such that  $\xi(t_1) = \kappa_3 \sqrt{\epsilon}$  (set  $t_1 = t_3$  if no such time exists). Then by (3.53) and (3.35) we have

$$\frac{1}{2} \lambda(t)^{\frac{k}{2}} \leq \zeta(t)^{\frac{k}{2}} \leq \frac{2\kappa_3}{\kappa_1} \sqrt{\epsilon} \quad \text{for } t \in [t_0, t_1],$$

which yields (3.41) if  $\epsilon$  is small enough.

**Case  $k = 2$ .** In this case (3.51) reads

$$\xi'(t) \geq \kappa_2 \xi(t). \quad (3.54)$$

Integrating between  $t$  and  $t_3$  we get  $\xi(t) \leq e^{\kappa_2(t-t_3)} \xi(t_3)$ . Thus (3.53) and (3.50) yield

$$\zeta(t) \leq \kappa_4 e^{\kappa_2(t-t_3)} \zeta(t_3) \leq 2\kappa_4 e^{\kappa_2(t-t_3)} \epsilon_1,$$

with a universal constant  $\kappa_4 > 0$ . Thus integrating (3.29) and using  $\mu(t_0) = 1$  we get  $\mu(t_3) \in [2/3, 3/2]$  if  $\epsilon_1$  is small enough, which implies that  $t_3 = t_2$ . Also, suppose that there is no  $t_2 \geq t_0$  such that  $\zeta(t_2)/\mu(t_2) = \epsilon_1$ . Then, since  $\zeta(t)$  (and hence  $\lambda(t)$ ) is far from 0, by known arguments, see for instance [22, Corollary A.4], the solution is global and (3.51) implies that  $\xi(t)$  is unbounded. Thus  $\lambda(t)$  is also unbounded, which is a contradiction. We infer that there must be  $t_2 < T_+$  such that  $\zeta(t_2)/\mu(t_2) = \epsilon_1$ , which implies (3.42) by choosing  $\epsilon_0$  comparable to  $(\epsilon_1)^{\frac{k}{2}}$ .

We have  $|\zeta'(t)| \lesssim |\zeta(t)|$ , see (3.36) and (3.37), hence there exists a constant  $\kappa_5$  such that  $\zeta(t) \geq \frac{1}{4}\epsilon_1$  for  $t \in [t_2 - \kappa_5, t_2]$ . Thus (3.48) yields

$$b(t) - b(t_0) \geq \kappa_6(t - (t_2 - \kappa_5))\epsilon_1, \quad \text{for } t \in [t_2 - \kappa_5, t_2].$$

Thus, if  $\epsilon$  is small enough, we get  $b(t) \geq \kappa_7\epsilon_1$  for  $t \in [t_2 - \frac{1}{2}\kappa_5, t_2]$ . Note that  $\kappa_7$  is independent of  $\epsilon_1$ . From the definition of  $b(t)$  and the Cauchy–Schwarz inequality we can deduce, if  $\epsilon_1$  is small enough, that  $\|\dot{g}(t)\|_{L^2} \geq \kappa_8\epsilon_1$  for  $t \in [t_2 - \frac{1}{2}\kappa_5, t_2]$ , which leads to

$$\int_{t_2 - \frac{1}{2}\kappa_5}^{t_2} \|\dot{g}(t)\|_{L^2}^2 dt \geq \kappa_9\epsilon_1^2. \quad (3.55)$$

Integrating (3.54) between  $t$  and  $t_1$  and using (3.53), (3.50) and the definition of  $t_1$  we obtain

$$\frac{1}{2}\lambda(t) \leq \zeta(t) \leq \kappa_{10}e^{\kappa_2(t-t_1)}\sqrt{\epsilon}, \quad \text{for } t \in [t_0, t_1].$$

Thus

$$\int_{t_0}^{t_1} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt \leq \kappa_{11}\sqrt{\epsilon}.$$

Comparing this bound with (3.55) and choosing  $\epsilon$  small enough, we get (3.43).

**Case  $k \geq 3$ .** Most of the argument can be repeated without major changes. We can rewrite (3.51) as

$$\left(\xi(t)^{\frac{2-k}{k}}\right)' \leq -\frac{(k-2)\kappa_2}{k}.$$

Integrating and using (3.53), (3.50) we obtain

$$\zeta(t)^{\frac{k}{2}} \leq \kappa_4\left(\zeta(t_3)^{\frac{2-k}{2}} + (t_3 - t)\right)^{\frac{k}{2-k}}, \quad (3.56)$$

with  $\kappa_4$  only depending on  $k$ . Thus

$$\begin{aligned} \frac{1}{2} \int_{t_0}^{t_3} \lambda(t)^{\frac{k}{2}} dt &\leq \int_{t_0}^{t_3} \zeta(t)^{\frac{k}{2}} dt \leq \kappa_4 \int_{\zeta(t_3)^{\frac{2-k}{2}}}^{+\infty} \tau^{\frac{k}{2-k}} d\tau \\ &\leq \frac{k-2}{2} \kappa_4 \lambda(t_3) \leq (k-2)\kappa_4\epsilon_1. \end{aligned}$$

As in the case  $k = 2$ , we can deduce that  $t_3 = t_2$  and that  $\zeta(t_2)/\mu(t_2) = \epsilon_1$ , which implies (3.42).

The proof of (3.55) applies without significant changes and yields

$$\int_{t_2 - \frac{1}{2}\kappa_5}^{t_2} \|\dot{g}(t)\|_{L^2}^2 dt \geq \kappa_9 \epsilon_1^{2k-2}. \quad (3.57)$$

The proof of (3.56) yields

$$\frac{1}{2} \lambda(t)^{\frac{k}{2}} \leq \zeta(t)^{\frac{k}{2}} \leq \kappa_{10} (\epsilon^{\frac{2-k}{2k}} + (t_1 - t))^{\frac{k}{2-k}}, \quad \text{for } t \in [t_0, t_1].$$

After integrating, this implies

$$\int_{t_0}^{t_1} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt \leq \kappa_{11} \epsilon^{\frac{1}{k}},$$

With  $\kappa_{11}$  depending only on  $k$ . Comparing this bound with (3.57) and choosing  $\epsilon$  small enough, we get (3.43).  $\square$

*Proof of Proposition 3.7* Let  $t_0 \in J$  be any point in  $J$ . By rescaling  $\vec{\psi}(t) \mapsto \vec{\psi}(t)_{\mu(t_0)^{-1}}$  we can assume without loss of generality that  $\mu(t_0) = 1$ . We can also assume that

$$\frac{1}{2} \leq \mu(t) \leq 2$$

for all  $t \in J$  (we work in a small neighborhood of  $t_0$ ).

We begin by differentiating the orthogonality conditions (3.4) and (3.5) to derive a linear system for  $(\lambda', \mu')$ .

Differentiating (3.4) yields

$$0 = \frac{d}{dt} \langle \mathcal{Z}_{\underline{\lambda}} | g \rangle = -\lambda' \left\langle \frac{1}{\lambda} [\Lambda_0 \mathcal{Z}]_{\underline{\lambda}} | g \right\rangle + \langle \mathcal{Z}_{\underline{\lambda}} | \partial_t g \rangle$$

Plugging in (3.30) above and rearranging we have

$$-\langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle = \lambda' \left( \langle \mathcal{Z}_{\underline{\lambda}} | \Lambda Q_{\underline{\lambda}} \rangle - \left\langle \frac{1}{\lambda} [\Lambda_0 \mathcal{Z}]_{\underline{\lambda}} | g \right\rangle \right) - \mu' \langle \mathcal{Z}_{\underline{\lambda}} | \Lambda Q_{\underline{\mu}} \rangle$$

Differentiating (3.5) yields

$$0 = \frac{d}{dt} \langle \mathcal{Z}_{\underline{\mu}} | g \rangle = -\mu' \left\langle \frac{1}{\mu} [\Lambda_0 \mathcal{Z}]_{\underline{\mu}} | g \right\rangle + \langle \mathcal{Z}_{\underline{\mu}} | \partial_t g \rangle$$

Plugging in (3.30) above and rearranging we have

$$-\langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle = \lambda' \langle \mathcal{Z}_{\underline{\mu}} | \Lambda Q_{\underline{\lambda}} \rangle + \mu' \left( -\langle \mathcal{Z}_{\underline{\mu}} | \Lambda Q_{\underline{\mu}} \rangle - \left\langle \frac{1}{\mu} [\Lambda_0 \mathcal{Z}]_{\underline{\mu}} | g \right\rangle \right)$$

We then arrive at the following linear system for  $(\lambda', \mu')$ ,

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} \lambda' \\ \mu' \end{pmatrix} = \begin{pmatrix} -\langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle \\ -\langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle \end{pmatrix}$$

where

$$\begin{aligned} M_{11} &:= \langle \mathcal{Z}_{\underline{\lambda}} | \Lambda Q_{\underline{\lambda}} \rangle - \left\langle \frac{1}{\lambda} [\Lambda_0 \mathcal{Z}]_{\underline{\lambda}} | g \right\rangle \\ M_{12} &:= -\langle \mathcal{Z}_{\underline{\lambda}} | \Lambda Q_{\underline{\mu}} \rangle \\ M_{21} &:= \langle \mathcal{Z}_{\underline{\mu}} | \Lambda Q_{\underline{\lambda}} \rangle \\ M_{22} &:= -\langle \mathcal{Z}_{\underline{\mu}} | \Lambda Q_{\underline{\mu}} \rangle - \left\langle \frac{1}{\mu} [\Lambda_0 \mathcal{Z}]_{\underline{\mu}} | g \right\rangle \end{aligned} \quad (3.58)$$

We first claim that  $M = (M_{ij})$  is diagonally dominant with coefficients of size  $\simeq 1$  on the diagonal. This will allow us to invert  $M$  and estimate  $\lambda', \mu'$ . Indeed, in Claim 3.3 we showed that the off-diagonal terms  $M_{12}$  and  $M_{21}$  satisfy  $|M_{12}| \lesssim \lambda^{k+1}$  and  $|M_{21}| \lesssim \lambda^{k-1}$ . To estimate the diagonal terms define  $\beta := \langle \mathcal{Z} | \Lambda Q \rangle$  and note that  $\beta > 0$  is a fixed positive number by (3.2). Then

$$|M_{11} - \beta| + |M_{22} + \beta| \lesssim \lambda^{\frac{k}{2}}$$

To see this, note that by the definitions of  $M_{11}, M_{22}$  and the fact that  $\mathcal{Z} \in C_0^\infty$  we have

$$\begin{aligned} |M_{11} - \beta| + |M_{22} + \beta| &\leq \left| \left\langle \frac{r}{\lambda} [\Lambda_0 \mathcal{Z}]_{\underline{\lambda}} | r^{-1} g \right\rangle \right| + \left| \left\langle \frac{r}{\mu} [\Lambda_0 \mathcal{Z}]_{\underline{\mu}} | r^{-1} g \right\rangle \right| \\ &\lesssim \|g\|_H \lesssim \lambda^{\frac{k}{2}} \end{aligned}$$

where the last inequality above follows from (3.7).

We solve for  $(\lambda', \mu')$ , by inverting  $M$ ,

$$\begin{pmatrix} \lambda' \\ \mu' \end{pmatrix} = \frac{1}{\det M} \begin{pmatrix} -M_{22} \langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle + M_{12} \langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle \\ M_{21} \langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle - M_{11} \langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle \end{pmatrix}$$

Now note that by (3.7) we have

$$|\langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle| + \left| \langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle \right| \lesssim \|\dot{g}\|_{L^2} \lesssim \lambda^{\frac{k}{2}} \quad (3.59)$$

Our estimates for the coefficients of  $M$  imply that

$$\det M = M_{11}M_{22} + O(\lambda^{2k}), \quad \frac{1}{\det M} = \frac{1}{M_{11}M_{22}} + O(\lambda^{2k})$$

Using the above we now write  $\lambda'$  and  $\mu'$  as follows,

$$\lambda' = \left( \frac{1}{M_{11}M_{22}} + O(\lambda^{2k}) \right) \left( -M_{22} \langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle + M_{12} \langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle \right)$$

and thus using (3.13) and (3.59) we conclude that

$$|\lambda'| \lesssim \lambda^{\frac{k}{2}}$$

Similarly, for  $\mu'$  we have

$$\mu' = \left( \frac{1}{M_{11}M_{22}} + O(\lambda^{2k}) \right) \left( M_{21} \langle \mathcal{Z}_{\underline{\lambda}} | \dot{g} \rangle - M_{11} \langle \mathcal{Z}_{\underline{\mu}} | \dot{g} \rangle \right)$$

and hence,

$$|\mu'| \lesssim \lambda^{\frac{k}{2}}$$

which proves (3.29) and completes the proof of Proposition 3.7.  $\square$

*Remark 3.13* We remark here that  $\lambda(t)$ ,  $\mu(t)$  obtained in the proof of Lemma 3.1 can be easily seen to be  $C^1$  functions. Recall the ODE

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} \lambda' \\ \mu' \end{pmatrix} = \begin{pmatrix} -\langle \mathcal{Z}_{\underline{\lambda}} | \psi_t(t) \rangle \\ -\langle \mathcal{Z}_{\underline{\mu}} | \psi_t(t) \rangle \end{pmatrix}, \quad (3.60)$$

obtained by formally differentiating the orthogonality conditions (3.4) and (3.5); the coefficients  $M_{ij}$  are given explicitly in (3.58). For any  $t_0 \in J$  the smallness of  $\lambda(t_0)/\mu(t_0)$  guarantees the existence of a unique  $C^1$  solution  $(\tilde{\lambda}(t), \tilde{\mu}(t))$  with initial data  $(\tilde{\lambda}, \tilde{\mu})(t_0) = (\lambda(t_0), \mu(t_0))$  in a neighborhood of  $t_0$ . Because of how the system (3.60) was derived,  $\tilde{\lambda}(t)$ ,  $\tilde{\mu}(t)$  and  $g(t) := \psi(t) - (Q_{\tilde{\lambda}(t)} + Q_{\tilde{\mu}(t)})$  satisfy (3.4) (3.5) and (3.6) in a small enough neighborhood of  $t_0$ . Since the  $\lambda(t)$ ,  $\mu(t)$  obtained by the implicit value theorem are unique with these properties we have  $\lambda(t) = \tilde{\lambda}(t)$  and  $\mu(t) = \tilde{\mu}(t)$  proving that  $\lambda(t)$ ,  $\mu(t)$  are indeed  $C^1$ .

*Proof of Proposition 3.8* As in the proof of Proposition 3.7 we can assume that  $\frac{1}{2} \leq \mu(t) \leq 2$  below.

We first prove (3.35). From (3.7) we have  $\|g\|_{L^\infty} \lesssim \lambda^{\frac{k}{2}}$ , so it suffices to check that

$$\|\chi_\mu \Lambda Q_{\underline{\lambda}}\|_{L^1(rdr)} \ll \lambda^{1-\frac{k}{2}} \text{ as } \lambda \rightarrow 0 \quad (3.61)$$

which follows from

$$\|\chi_\mu \Lambda Q_{\underline{\lambda}}\|_{L^1(rdr)} \leq \lambda \int_0^{4/\lambda} \Lambda Q(r) r dr \lesssim \lambda \int_0^{4/\lambda} (1+r)^{-k+1} dr \lesssim \lambda |\log(4/\lambda)|$$

Now we prove (3.36). From (3.30) we have

$$\begin{aligned} \frac{d}{dr} \langle \chi_\mu \Lambda Q_{\underline{\lambda}} | g \rangle &= \langle \chi_\mu \Lambda Q_{\underline{\lambda}} | \dot{g} \rangle + \lambda' \langle \chi_\mu \Lambda Q_{\underline{\lambda}} | \Lambda Q_{\underline{\lambda}} \rangle - \mu' \langle \chi_\mu \Lambda Q_{\underline{\lambda}} | \Lambda Q_{\underline{\mu}} \rangle \\ &\quad - \frac{\lambda'}{\lambda} \langle \chi_\mu \Lambda_0 \Lambda Q_{\underline{\lambda}}, g \rangle - \mu' \langle \Lambda \chi_\mu \Lambda Q_{\underline{\lambda}} | g \rangle, \end{aligned} \quad (3.62)$$

Most terms on the right hand side are negligible (that is,  $\ll \lambda^{\frac{k}{2}}$ ). Since  $\lambda \ll 1$ , we have

$$\begin{aligned} \|(1 - \chi_\mu) \Lambda Q_{\underline{\lambda}}\|_{L^2}^2 &\lesssim \int_{1/2}^\infty (\Lambda Q_{\underline{\lambda}})^2 r dr = \int_{\frac{1}{2}\lambda^{-1}}^\infty (\Lambda Q)^2 r dr \\ &\lesssim \int_{\frac{1}{2}\lambda^{-1}}^\infty r^{-2k+1} dr \lesssim \lambda^{2k-2}. \end{aligned} \quad (3.63)$$

Together with (3.7) this yields

$$|\langle (1 - \chi_\mu) \Lambda Q_{\underline{\lambda}} | \dot{g} \rangle| \lesssim \lambda^{k-1} \lambda^{\frac{k}{2}} \ll \lambda^{\frac{k}{2}},$$

so in the first term we can erase  $\chi_\mu$ . Similarly, from (3.63) and (3.28) we have  $|\lambda' \langle (1 - \chi_\mu) \Lambda Q_{\underline{\lambda}} | \Lambda Q_{\underline{\lambda}} \rangle| \ll \lambda^{\frac{k}{2}}$ , so  $\chi_\mu$  can be erased also in the second term.

Regarding the third term, since we are assuming  $\frac{1}{2} \leq \mu \leq 2$ , we have

$$\begin{aligned} \left| \langle \chi_\mu \Lambda Q_{\underline{\lambda}} | \Lambda Q_{\underline{\mu}} \rangle \right| &\lesssim \int_0^4 \frac{1}{\lambda} \frac{r}{\lambda} Q_r(r/\lambda) \frac{1}{\mu} \frac{r}{\mu} Q_r(r/\mu) r dr \\ &\lesssim \frac{1}{\lambda} \int_0^4 \frac{(r/\lambda)^k}{1 + (r/\lambda)^{2k}} \frac{r^{k+1}}{1 + r^{2k}} dr \\ &\lesssim \lambda^{k-1} \int_0^\lambda \frac{r^{2k+1}}{\lambda^{2k} + r^{2k}} \frac{1}{1 + r^{2k}} dr \\ &\quad + \lambda^{k-1} \int_\lambda^4 \frac{r^{2k+1}}{\lambda^{2k} + r^{2k}} \frac{1}{1 + r^{2k}} dr \end{aligned}$$

To estimate the first integral on the right above on the interval  $[0, \lambda]$  we note that since  $\lambda \ll 1$  we have

$$\lambda^{k-1} \int_0^\lambda \frac{r^{2k+1}}{\lambda^{2k} + r^{2k}} \frac{1}{1 + r^{2k}} dr \lesssim \lambda^{k-1} \lambda^{-2k} \int_0^\lambda r^{2k+1} dr \lesssim \lambda^{k+1}$$

On the interval  $[\lambda, 4]$  we write

$$\begin{aligned} \frac{1}{\lambda^{2k} + r^{2k}} &= \frac{1}{r^{2k}} + \left( \frac{1}{\lambda^{2k} + r^{2k}} - \frac{1}{r^{2k}} \right) \\ &= \frac{1}{r^{2k}} + \frac{1}{r^{2k}} \left( \frac{1}{1 + (\lambda/r)^{2k}} - 1 \right) \\ &= \frac{1}{r^{2k}} + \frac{1}{r^{2k}} \left( -\lambda^{2k} r^{-2k} + O(\lambda^{4k} r^{-4k}) \right) \end{aligned}$$

This yields,

$$\begin{aligned} \lambda^{k-1} \int_\lambda^4 \frac{r^{2k+1}}{\lambda^{2k} + r^{2k}} \frac{1}{1 + r^{2k}} dr &= \lambda^{k-1} \int_0^\infty \frac{r}{1 + r^{2k}} dr \\ &\quad - \lambda^{k-1} \int_0^\lambda \frac{r}{1 + r^{2k}} dr + O(\lambda^{k+1}) \\ &= C\lambda^{k-1} + O(\lambda^{k+1}) \end{aligned}$$

This and (3.29) imply that the third term of the right hand side in (3.62) is negligible.

As for the fourth term, we have

$$\left| \frac{\lambda'}{\lambda} \langle \chi_\mu \Lambda_0 \Lambda Q_\lambda, g \rangle \right| \lesssim |\lambda'| \|g\|_{L^\infty} \|\chi_{\mu/\lambda} \Lambda_0 \Lambda Q\|_{L^1} \lesssim \lambda^k \|\chi_{2/\lambda} \Lambda_0 \Lambda Q\|_{L^1},$$

which is  $\ll \lambda^{\frac{k}{2}}$ , see the proof of (3.61). The fifth term is even smaller (we gain an additional factor  $\lambda$ ).

Summarizing, from the definition of  $\zeta(t)$  and (3.64), we obtain

$$|\kappa \zeta' - \langle \Lambda Q_\lambda | \dot{g} \rangle| \ll \lambda^{\frac{k}{2}}. \quad (3.64)$$

Recall that

$$b(t) := -\langle \Lambda Q_\lambda | \dot{g} \rangle - \langle \dot{g} | \mathcal{A}_0(\lambda)g \rangle$$

By (3.7) and the fact that  $\mathcal{A}_0 : H \rightarrow L^2$  is bounded independently of  $\lambda$ —see Lemma 3.5—we have

$$\langle \dot{g} | \mathcal{A}_0(\lambda)g \rangle \lesssim \|\dot{g}\|_{L^2} \|\mathcal{A}_0(\lambda)g\|_{L^2} \lesssim \|(g, \dot{g})\|_{\mathcal{H}_0}^2 \lesssim \lambda^k \ll \lambda^{\frac{k}{2}}.$$

Together with (3.64) this means that

$$|\kappa \zeta' - b| \ll \lambda^{\frac{k}{2}},$$

which is (3.36).

Arguing as above we have

$$|b(t)| \leq \|\Lambda Q_{\underline{\lambda}}\|_2 \|\psi_t\|_2 - O(\lambda^k) = \sqrt{\kappa} \|\psi_t\|_2 - O(\lambda^k)$$

From the expansion of the nonlinear energy in the proof of (3.7) and our assumption that  $\mu(t) \simeq 1$  we see that

$$\|\psi_t(t)\|_2^2 \leq 16k(\lambda/\mu)^k + o(\lambda^k)$$

Plugging this in above yields (3.37).

Finally, we begin the delicate proof of (3.39); we note that (3.38) will also be a consequence of this analysis. It is sufficient to prove the result for smooth solutions. Indeed, we can then use a standard approximation procedure. We approximate a solution  $\tilde{\psi} : J \rightarrow \mathcal{H}_0$  by a sequence of smooth solutions  $\tilde{\psi}_n$ . Then  $b_n(t)$  converges to  $b(t)$  uniformly for  $t \in J$ , and we can pass to a limit in (3.38) and (3.39). Differentiating  $b(t)$  and recalling the formulae (3.30), (3.31) we have

$$\begin{aligned} b'(t) &= \frac{\lambda'}{\lambda} \langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} | \dot{g} \rangle - \langle \Lambda Q_{\underline{\lambda}} | \partial_t \dot{g} \rangle - \langle \partial_t \dot{g} | \mathcal{A}_0(\lambda) g \rangle \\ &\quad - \frac{\lambda'}{\lambda} \langle \dot{g} | (\lambda \partial_{\lambda} \mathcal{A}_0(\lambda)) g \rangle - \langle \dot{g} | \mathcal{A}_0(\lambda) \partial_t g \rangle \\ &= \frac{\lambda'}{\lambda} \langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} | \dot{g} \rangle \\ &\quad - \left\langle \Lambda Q_{\underline{\lambda}} | \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_{\lambda} - Q_{\mu} + g) - f(Q_{\lambda}) + f(Q_{\mu})) \right\rangle \quad (3.65) \\ &\quad - \left\langle \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_{\lambda} - Q_{\mu} + g) - f(Q_{\lambda}) + f(Q_{\mu})) | \mathcal{A}_0(\lambda) g \right\rangle \\ &\quad - \frac{\lambda'}{\lambda} \langle \dot{g} | (\lambda \partial_{\lambda} \mathcal{A}_0(\lambda)) g \rangle - \langle \dot{g} | \mathcal{A}_0(\lambda) \dot{g} \rangle - \lambda' \langle \dot{g} | \mathcal{A}_0(\lambda) \Lambda Q_{\underline{\lambda}} \rangle \\ &\quad + \mu' \langle \dot{g} | \mathcal{A}_0(\lambda) \Lambda Q_{\underline{\mu}} \rangle \end{aligned}$$

Let us first identify terms above that we've already established to be  $\ll \lambda^{k-1}$  and discard them. First note that since  $(\lambda \partial_{\lambda} \mathcal{A}_0(\lambda)) : H \rightarrow L^2$  is bounded, and since we've already shown  $|\lambda'| \lesssim \lambda^{\frac{k}{2}}$  we have

$$\frac{\lambda'}{\lambda} \langle \dot{g} | (\lambda \partial_{\lambda} \mathcal{A}_0(\lambda)) g \rangle \lesssim \|(g, \dot{g})\|_{\mathcal{H}_0}^2 \lesssim \lambda^k \ll \lambda^{k-1}$$

Then we note that

$$\langle \dot{g} \mid \mathcal{A}_0(\lambda) \dot{g} \rangle = 0,$$

which can be shown directly by integration by parts. Next, using the fact that  $\mu \simeq 1$ , along with the boundedness of  $\mathcal{A}_0(\lambda) : H \rightarrow L^2$  we have

$$\begin{aligned} \mu' \langle \dot{g} \mid \mathcal{A}_0(\lambda) \Lambda Q_\mu \rangle &= \frac{\mu'}{\mu} \langle \dot{g} \mid \mathcal{A}_0(\lambda) \Lambda Q_\mu \rangle \\ &\lesssim |\mu'| \|\dot{g}\|_{L^2} \|\Lambda Q_\mu\|_H \lesssim |\mu'| \|\dot{g}\|_{L^2} \lesssim \lambda^k \ll \lambda^{k-1} \end{aligned}$$

Next, the combination of the first and sixth terms on the right-hand-side of (3.65) can be estimated using (3.24),

$$\begin{aligned} \left| \frac{\lambda'}{\lambda} \langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid \dot{g} \rangle - \lambda' \langle \dot{g} \mid \mathcal{A}_0(\lambda) \Lambda Q_{\underline{\lambda}} \rangle \right| &= \left| \frac{\lambda'}{\lambda} \langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} - \mathcal{A}_0(\lambda) \Lambda Q_{\underline{\lambda}} \mid \dot{g} \rangle \right| \\ &\lesssim \lambda^{\frac{k}{2}-1} \|[\Lambda_0 \Lambda Q]_{\underline{\lambda}} - \mathcal{A}_0(\lambda) \Lambda Q_{\underline{\lambda}}\|_{L^2} \|\dot{g}\|_{L^2} \\ &\lesssim c_0 \lambda^{\frac{k}{2}-1} \lambda^{\frac{k}{2}} \ll \lambda^{k-1} \end{aligned}$$

where in the last line above we rely on our ability to take  $c_0$  as small as we like in the estimate (3.24) from Lemma 3.5.

Thus we've show that up to terms of order  $\ll \lambda^{k-1}$ , which can be absorbed into the error, we have

$$\begin{aligned} b'(t) &\simeq - \left\langle \Lambda Q_{\underline{\lambda}} \mid \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda) + f(Q_\mu)) \right\rangle \\ &\quad - \left\langle \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda) + f(Q_\mu)) \mid \mathcal{A}_0(\lambda) g \right\rangle \end{aligned} \quad (3.66)$$

Next, rescaling the equation  $\mathcal{L} \Lambda Q = 0$ , we see that

$$\mathcal{L}_\lambda \Lambda Q_{\underline{\lambda}} := \left( -\partial_{rr} - \frac{1}{r} \partial_r + \frac{f'(Q_\lambda)}{r^2} \right) \Lambda Q_{\underline{\lambda}} = 0$$

And since  $\mathcal{L}_\lambda$  is symmetric we have

$$\left\langle \Lambda Q_{\underline{\lambda}} \mid \partial_r^2 g + \frac{1}{r} \partial_r g \right\rangle = \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{f'(Q_\lambda)}{r^2} g \right\rangle$$

We thus rewrite (3.66) as

$$\begin{aligned} b'(t) &\simeq \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda) + f(Q_\mu) - f'(Q_\lambda) g) \right\rangle \\ &\quad - \left\langle \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda) + f(Q_\mu)) \mid \mathcal{A}_0(\lambda) g \right\rangle \end{aligned}$$

where the symbol  $\simeq$  above means “up to terms of order  $\ll \lambda^{k-1}$ ”. Adding and subtracting we have

$$\begin{aligned} b'(t) &\simeq \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} \left( f(Q_{\lambda} - Q_{\mu}) - f(Q_{\lambda}) + f(Q_{\mu}) \right) \right\rangle \\ &\quad + \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} \left( f'(Q_{\lambda} - Q_{\mu}) - f'(Q_{\lambda}) \right) g \right\rangle \end{aligned} \quad (3.67)$$

$$\begin{aligned} &\quad + \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} \left( f(Q_{\lambda} - Q_{\mu} + g) - f(Q_{\lambda} - Q_{\mu}) - f'(Q_{\lambda} - Q_{\mu})g \right) \right\rangle \\ &\quad (3.68) \end{aligned}$$

$$\begin{aligned} &\quad - \left\langle \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} \left( f(Q_{\lambda} - Q_{\mu} + g) - f(Q_{\lambda}) + f(Q_{\mu}) \right) \mid \mathcal{A}_0(\lambda)g \right\rangle \\ &\quad (3.69) \end{aligned}$$

Let’s begin by estimating the first term on the right-hand-side above, which we’ll show contributes the leading order:

### Claim 3.14

$$\left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} \left( f(Q_{\lambda} - Q_{\mu}) - f(Q_{\lambda}) + f(Q_{\mu}) \right) \right\rangle \simeq 8k^2 \frac{\lambda^{k-1}}{\mu^k}$$

where again  $\simeq$  means “up to terms of order  $\ll \lambda^{k-1}$ .”

*Proof* Let’s prove the claim. Recall that the nonlinearity  $f(\rho)$  is given by

$$f(\rho) := \frac{k^2}{2} \sin(2\rho), \quad f'(\rho) = k^2 \cos(2\rho)$$

Using trigonometric identities we can write

$$\begin{aligned} f(Q_{\lambda} - Q_{\mu}) - f(Q_{\lambda}) + f(Q_{\mu}) &= \frac{k^2}{2} (\sin 2Q_{\lambda} (\cos 2Q_{\mu} - 1) + \sin 2Q_{\mu} (1 - \cos 2Q_{\lambda})) \\ &= -k^2 \sin 2Q_{\lambda} \sin^2 Q_{\mu} + k^2 \sin 2Q_{\mu} \sin^2 Q_{\lambda} \\ &= -\sin 2Q_{\lambda} (\Lambda Q_{\mu})^2 + \sin 2Q_{\mu} (\Lambda Q_{\lambda})^2. \end{aligned} \quad (3.70)$$

We show that the leading order contribution comes from the second term above. Indeed, writing  $\sigma = \lambda/\mu$  and changing variables we have

$$\begin{aligned} \left\langle \Lambda Q_{\lambda} \mid \frac{1}{r^2} (\Lambda Q_{\lambda})^2 \sin 2Q_{\mu} \right\rangle &= \frac{1}{\lambda} \int_0^{\infty} (\Lambda Q_{\sigma}(r))^3 \sin 2Q(r) \frac{dr}{r} \\ &= \frac{1}{\lambda} \int_0^{\sqrt{\sigma}} (\Lambda Q_{\sigma}(r))^3 \sin 2Q(r) \frac{dr}{r} + \frac{1}{\lambda} \int_{\sqrt{\sigma}}^{\infty} (\Lambda Q_{\sigma}(r))^3 \sin 2Q(r) \frac{dr}{r} \end{aligned} \quad (3.71)$$

Since  $\sigma = \lambda/\mu \ll 1$ , on the interval  $[0, \sqrt{\sigma}]$  we write

$$\sin 2Q = \frac{2}{k^2} 2k^2 r^k \frac{1 - r^{2k}}{(1 + r^{2k})^2} = 4r^k + O(r^{3k}) \quad (3.72)$$

Changing variables again we have

$$\begin{aligned} \frac{1}{\lambda} \int_0^{\sqrt{\sigma}} (\Lambda Q_{\sigma})^3 4r^k \frac{dr}{r} &= 4 \frac{\sigma^k}{\lambda} \int_0^{\frac{1}{\sqrt{\sigma}}} (\Lambda Q)^3(r) r^{k-1} dr \\ &= 4 \frac{\sigma^k}{\lambda} \int_0^{\infty} (\Lambda Q)^3(r) r^{k-1} dr \\ &\quad - 4 \frac{\sigma^k}{\lambda} \int_{\frac{1}{\sqrt{\sigma}}}^{\infty} (\Lambda Q)^3(r) r^{k-1} dr \\ &= 8k^2 \frac{\lambda^{k-1}}{\mu^k} + O(\lambda^{-1} \sigma^{2k}) \end{aligned}$$

where we used (3.33) in the last line above. Moreover, using (3.72) we estimate

$$\begin{aligned} \frac{1}{\lambda} \int_0^{\sqrt{\sigma}} (\Lambda Q_{\sigma})^3 (\sin 2Q - 4r^k) \frac{dr}{r} &\lesssim \frac{1}{\lambda} \int_0^{\sqrt{\sigma}} (\Lambda Q_{\sigma})^3 r^{3k-1} dr \\ &= \frac{\sigma^{3k}}{\lambda} \int_0^{\frac{1}{\sqrt{\sigma}}} (\Lambda Q)^3 r^{3k-1} dr \lesssim \frac{\sigma^{3k} |\log \sigma|}{\lambda} \end{aligned}$$

Finally, we estimate the second term on the last line (3.71) by

$$\begin{aligned} \left| \frac{1}{\lambda} \int_{\sqrt{\sigma}}^{\infty} (\Lambda Q_{\sigma}(r))^3 \sin 2Q(r) \frac{dr}{r} \right| &\lesssim \frac{1}{\lambda} \int_{\sqrt{\sigma}}^{\infty} (\Lambda Q_{\sigma}(r))^3 \frac{dr}{r} \\ &= \frac{1}{\lambda} \int_{\frac{1}{\sqrt{\sigma}}}^{\infty} (\Lambda Q(r))^3 \frac{dr}{r} \\ &\lesssim \frac{1}{\lambda} \int_{\frac{1}{\sqrt{\sigma}}}^{\infty} r^{-3k-1} dr \lesssim \frac{\sigma^{\frac{3}{2}k}}{\lambda} \end{aligned}$$

Next we estimate the contribution of the first term in (3.70). Recalling that  $\sin 2Q_\lambda = \frac{2}{k^2} \Lambda^2 Q_\lambda = \frac{2}{k^2} r \partial_r (\Lambda Q_\lambda)$  we integrate by parts to obtain

$$\begin{aligned} - \left\langle \Lambda Q_\lambda \mid \frac{1}{r^2} \sin 2Q_\lambda (\Lambda Q_\mu)^2 \right\rangle &= \frac{2}{k^2 \lambda} \left\langle (\Lambda Q_\lambda)^2 \mid \frac{1}{r^2} \Lambda Q_\mu \Lambda^2 Q_\mu \right\rangle \\ &= \frac{2}{k^2 \lambda} \left\langle (\Lambda Q_\sigma)^2 \mid \frac{1}{r^2} \Lambda Q \Lambda^2 Q \right\rangle \end{aligned}$$

where  $\sigma = \lambda/\mu$  as before. We first estimate the last line above on the interval  $[0, \sigma]$  and  $[\sigma, 1]$  using (3.45) (3.46) to obtain the bound

$$|\Lambda Q(r) \Lambda^2 Q(r)| \lesssim r^{2k}$$

which gives

$$\begin{aligned} \frac{2}{\lambda k^2} \int_0^\sigma (\Lambda Q_\sigma)^2 \Lambda Q \Lambda^2 Q \frac{dr}{r} &\lesssim \frac{1}{\lambda} \int_0^\sigma \Lambda Q_\sigma^2 r^{2k-1} dr \\ &= \frac{\sigma^{2k}}{\lambda} \int_0^1 \Lambda Q^2 r^{2k-1} dr \lesssim \frac{\sigma^{2k}}{\lambda} \end{aligned}$$

and

$$\frac{2}{\lambda k^2} \int_\sigma^1 (\Lambda Q_\sigma)^2 \Lambda Q \Lambda^2 Q \frac{dr}{r} \lesssim \frac{\sigma^{2k}}{\lambda} \int_\sigma^1 \frac{r^{4k-1}}{(\sigma^{2k} + r^{2k})^2} dr \lesssim \frac{\sigma^{2k} |\log \sigma|}{\lambda}$$

On the interval  $[1, \infty]$  we use the formulas (3.45) (3.46) to estimate

$$|\Lambda Q(r) \Lambda^2 Q(r)| \lesssim \frac{r^{4k}}{(1 + r^{2k})^3}$$

which means

$$\begin{aligned} \frac{2}{\lambda k^2} \int_\sigma^\infty (\Lambda Q_\sigma)^2 \Lambda Q(r) \Lambda^2 Q(r) \frac{dr}{r} &\lesssim \frac{\sigma^{2k}}{\lambda} \int_\sigma^\infty \frac{r^{6k-1} dr}{(\sigma^{2k} + r^{2k})^2 (1 + r^{2k})^3} \\ &\lesssim \frac{\sigma^{2k}}{\lambda} \int_0^\infty \frac{r^{2k-1}}{(1 + r^{2k})^3} dr \lesssim \frac{\sigma^{2k}}{\lambda} \end{aligned}$$

Putting this all together, we've shown that

$$\left\langle \Lambda Q_\lambda \mid \left( f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu) \right) \right\rangle = 8k^2 \frac{\lambda^{k-1}}{\mu^k} + o \left( \frac{\lambda^{k-1}}{\mu^k} \left( \frac{\lambda}{\mu} \right)^{\frac{k}{2}} \right)$$

which is precisely Claim 3.14.  $\square$

Next, we claim the second term (3.67) in our expansion of  $b'(t)$  satisfies

$$\left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} \left( f'(Q_{\lambda} - Q_{\mu}) - f'(Q_{\lambda}) \right) g \right\rangle = o(\lambda^{k-1}) \quad (3.73)$$

and can thus be absorbed into the error. First note that the we have

$$\begin{aligned} f'(Q_{\lambda} - Q_{\mu}) - f'(Q_{\lambda}) &= k^2 \sin 2Q_{\lambda} \sin 2Q_{\mu} - 2k^2 \cos 2Q_{\lambda} \sin^2 Q_{\mu} \\ &= \frac{4}{k^2} \Lambda^2 Q_{\lambda} \Lambda^2 Q_{\mu} - (\Lambda Q_{\mu})^2 \cos 2Q_{\lambda} \end{aligned}$$

For the contribution from the first term above we integrate by parts, change variables, and use the explicit formulae (3.45) (3.46) (3.47) to estimate

$$\begin{aligned} \left| \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{4}{k^2 r^2} (\Lambda^2 Q_{\lambda} \Lambda^2 Q_{\mu}) g \right\rangle \right| &\lesssim \frac{1}{\lambda} \left| \int_0^{\infty} (\Lambda Q_{\lambda})^2 \Lambda^3 Q_{\mu} g \frac{dr}{r} \right| \\ &\quad + \left| \int_0^{\infty} (\Lambda Q_{\lambda})^2 \Lambda^2 Q_{\mu} r \partial_r g \frac{dr}{r} \right| \\ &\lesssim \frac{1}{\lambda} \|g\|_H \left[ \left( \int_0^{\infty} (\Lambda Q_{\sigma})^4 (\Lambda^3 Q)^2 \frac{dr}{r} \right)^{\frac{1}{2}} + \left( \int_0^{\infty} (\Lambda Q_{\sigma})^4 (\Lambda^2 Q)^2 \frac{dr}{r} \right)^{\frac{1}{2}} \right] \\ &= o(\sigma^k / \lambda) \end{aligned}$$

where  $\sigma = \lambda/\mu$  as before. For the second term we write

$$\begin{aligned} \left| \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} ((\Lambda Q_{\mu})^2 \cos 2Q_{\lambda}) g \right\rangle \right| &\lesssim \|g\|_H \left( \int_0^{\infty} (\Lambda Q_{\sigma})^2 (\Lambda Q)^4 \frac{dr}{r} \right)^{\frac{1}{2}} \\ &\lesssim \frac{1}{\lambda} \sigma^{\frac{k}{2}} \sigma^k |\log \sigma|^{\frac{1}{2}} = o(\sigma^k / \lambda) \end{aligned}$$

which finishes the proof of (3.73).

Finally, we consider the last two terms (3.68) (3.69). We will reorganize these terms in anticipation of applications of Lemma 3.5. First we rewrite (3.68) as follows:

$$\begin{aligned} &\left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{r^2} (f(-Q_{\mu} + Q_{\lambda} + g) - f(-Q_{\mu} + Q_{\lambda}) - f'(-Q_{\mu} + Q_{\lambda})g) \right\rangle \\ &= - \left\langle \mathcal{A}(\lambda)g \mid \frac{1}{r^2} (f(-Q_{\mu} + Q_{\lambda} + g) - f(-Q_{\mu} + Q_{\lambda}) - k^2 g) \right\rangle \\ &\quad + \left\langle \mathcal{A}(\lambda)g \mid \frac{1}{r^2} (f(-Q_{\mu} + Q_{\lambda} + g) - f(-Q_{\mu} + Q_{\lambda}) - k^2 g) \right\rangle \\ &\quad + \left\langle \mathcal{A}(\lambda)(Q_{\lambda} - Q_{\mu}) \mid \frac{1}{r^2} (f(-Q_{\mu} + Q_{\lambda} + g) - f(-Q_{\mu} + Q_{\lambda}) - f'(-Q_{\mu} + Q_{\lambda})g) \right\rangle \end{aligned}$$

$$\begin{aligned}
 & + \left\langle \mathcal{A}(\lambda) \mathcal{Q}_\mu \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - f'(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda)g) \right\rangle \\
 & + \left\langle \Lambda \mathcal{Q}_\lambda - \mathcal{A}(\lambda) \mathcal{Q}_\lambda \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - f'(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda)g) \right\rangle
 \end{aligned}$$

The second two terms on the right-hand-side can be controlled by setting  $g_1 = \mathcal{Q}_\lambda - \mathcal{Q}_\mu$  and  $g_2 = g$  in (3.22):

$$\begin{aligned}
 & \left| \left\langle \mathcal{A}(\lambda)g \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - k^2g) \right\rangle \right. \\
 & \left. + \left\langle \mathcal{A}(\lambda)(\mathcal{Q}_\lambda - \mathcal{Q}_\mu) \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - f'(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda)g) \right\rangle \right| \\
 & \lesssim c_0 \lambda^{k-1}
 \end{aligned}$$

Using the pointwise bound

$$\begin{aligned}
 & |f(\mathcal{Q}_\lambda - \mathcal{Q}_\mu + g) - f(\mathcal{Q}_\lambda - \mathcal{Q}_\mu) - f'(\mathcal{Q}_\lambda - \mathcal{Q}_\mu)g| \\
 & = \frac{k^2}{2} |\sin(2\mathcal{Q}_\lambda - 2\mathcal{Q}_\mu)[\cos 2g - 1] + \cos(2\mathcal{Q}_\lambda - 2\mathcal{Q}_\mu)[\sin 2g - 2g]| \lesssim |g|^2
 \end{aligned}$$

along with (3.26) the second to last line of the above can be estimated by

$$\left\langle \mathcal{A}(\lambda) \mathcal{Q}_\mu \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - f'(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda)g) \right\rangle \lesssim c_0 \lambda^{k-1}.$$

Similarly, the last line of the expansion of (3.68) can be controlled as follows

$$\begin{aligned}
 & \left| \left\langle \Lambda \mathcal{Q}_\lambda - \mathcal{A}(\lambda) \mathcal{Q}_\lambda \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - f'(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda)g) \right\rangle \right| \\
 & \lesssim \|\Lambda \mathcal{Q}_\lambda - \mathcal{A}(\lambda) \mathcal{Q}_\lambda\|_{L^\infty} \|g\|_H^2 \leq C c_0 \lambda^{k-1} \ll \lambda^{k-1}
 \end{aligned}$$

In the last line we used (3.25) and the fact that  $c_0$  can be taken small independently of  $\lambda$  in Lemma 3.5.

Thus, up to terms of order  $\ll \lambda^{k-1}$  we have

$$\begin{aligned}
 & \left\langle \Lambda \mathcal{Q}_\lambda \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - f'(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda)g) \right\rangle \\
 & \simeq - \left\langle \mathcal{A}(\lambda)g \mid \frac{1}{r^2} (f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda + g) - f(-\mathcal{Q}_\mu + \mathcal{Q}_\lambda) - k^2g) \right\rangle
 \end{aligned} \tag{3.74}$$

We now transform the last line (3.69) adding and subtracting terms as before. Using (3.23) we have

$$\begin{aligned}
 & - \left\langle \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda) + f(Q_\mu)) \mid \mathcal{A}_0(\lambda)g \right\rangle \\
 & = - \left\langle \mathcal{A}_0(\lambda)g \mid \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{k^2}{r^2} g \right\rangle \\
 & \quad + \left\langle \mathcal{A}_0(\lambda)g \mid \frac{1}{r^2} (f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu)) \right\rangle \\
 & \quad + \left\langle \mathcal{A}_0(\lambda)g \mid \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda - Q_\mu) - k^2 g) \right\rangle \quad (3.75) \\
 & \geq - \frac{c_0}{\lambda} \|g\|_H^2 + \frac{1}{\lambda} \int_0^{R\lambda} \left( (\partial_r g)^2 + \frac{k^2}{r^2} g^2 \right) r dr \\
 & \quad + \left\langle \mathcal{A}_0(\lambda)g \mid \frac{1}{r^2} (f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu)) \right\rangle \\
 & \quad + \left\langle \mathcal{A}_0(\lambda)g \mid \frac{1}{r^2} (f(-Q_\mu + Q_\lambda + g) + f(-Q_\mu + Q_\lambda) - k^2 g) \right\rangle
 \end{aligned}$$

where  $R$  is as in Lemma 3.5. Note that from (3.70) we have the pointwise inequality

$$|f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu)| \lesssim (\Lambda Q_\lambda)^2 (\Lambda Q_\mu) + \Lambda Q_\lambda (\Lambda Q_\mu)^2$$

Since  $\|\mathcal{A}_0(\lambda)g\|_{L^2} \lesssim \|g\|_H$ , and since  $\mathcal{A}_0(\lambda)g$  is supported on a ball of radius  $R\lambda$ , the term on the second to last line above can be estimated as follows,

$$\begin{aligned}
 & \left| \left\langle \mathcal{A}_0(\lambda)g \mid \frac{1}{r^2} (f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu)) \right\rangle \right| \\
 & \lesssim \|g\|_H \left[ \left( \int_0^{R\sigma} r^{-2} (\Lambda Q_\sigma)^4 (\Lambda Q)^2 \frac{dr}{r} \right)^{\frac{1}{2}} + \left( \int_0^{R\sigma} r^{-2} (\Lambda Q)^4 (\Lambda Q_\sigma)^2 \frac{dr}{r} \right)^{\frac{1}{2}} \right] \\
 & \lesssim \sigma^{\frac{k}{2}} \sigma^{k-1} \ll \lambda^{k-1}
 \end{aligned}$$

where  $\sigma = \lambda/\mu$  as usual and  $\mu \simeq 1$ .

Therefore, up to terms of order  $\ll \lambda^{k-1}$ , we can put together (3.74) and (3.75) to estimate the combination of (3.68) and (3.69) from below by

$$\begin{aligned}
 & \left\langle \Lambda Q_\lambda \mid \frac{1}{r^2} (f(-Q_\mu + Q_\lambda + g) - f(-Q_\mu + Q_\lambda) - f'(-Q_\mu + Q_\lambda)g) \right\rangle \\
 & - \left\langle \partial_r^2 g + \frac{1}{r} \partial_r g - \frac{1}{r^2} (f(Q_\lambda - Q_\mu + g) - f(Q_\lambda) + f(Q_\mu)) \mid \mathcal{A}_0(\lambda)g \right\rangle
 \end{aligned}$$

$$\begin{aligned} &\geq \frac{1}{\lambda} \int_0^{R\lambda} \left( (\partial_r g)^2 + \frac{k^2}{r^2} g^2 \right) r dr \\ &\quad + \left\langle (\mathcal{A}_0(\lambda) - \mathcal{A}(\lambda))g \mid \frac{1}{r^2} (f(-Q_\mu + Q_\lambda + g) - f(-Q_\mu + Q_\lambda) - k^2 g) \right\rangle \end{aligned} \quad (3.76)$$

Since  $\mathcal{A}_0(\lambda) - \mathcal{A}(\lambda)$  is the operator of multiplication by  $\frac{1}{2\lambda} \left( q''\left(\frac{r}{\lambda}\right) + \frac{\lambda}{r} q'\left(\frac{r}{\lambda}\right) \right)$ , we can use (3.27) to estimate the last term above,

$$\begin{aligned} &\left\langle (\mathcal{A}_0(\lambda) - \mathcal{A}(\lambda))g \mid \frac{1}{r^2} (f(-Q_\mu + Q_\lambda + g) - f(-Q_\mu + Q_\lambda) - k^2 g) \right\rangle \\ &= \frac{1}{\lambda} \int_0^{+\infty} \frac{1}{r^2} (f'(Q_\lambda) - k^2) g^2 r dr + O(c_0 \lambda^k) \end{aligned} \quad (3.77)$$

where  $c_0 > 0$  is as in Lemma 3.5.

Putting together the estimates from Claims 3.14, (3.73), (3.76), and (3.77) we obtain the estimate

$$\begin{aligned} b'(t) &\geq 8k^2 \frac{\lambda^{k-1}}{\mu^k} + o(\lambda^{k-1}) + \frac{1}{\lambda} \int_0^{R\lambda} \left( (\partial_r g)^2 + \frac{k^2}{r^2} g^2 \right) r dr \\ &\quad + \frac{1}{\lambda} \int_0^{+\infty} \frac{1}{r^2} (f'(Q_\lambda) - k^2) g^2 r dr \end{aligned}$$

Finally we conclude by using the following localized coercivity estimate,

$$\frac{1}{\lambda} \int_0^{R\lambda} \left( (\partial_r g)^2 + \frac{k^2}{r^2} g^2 \right) r dr + \frac{1}{\lambda} \int_0^{+\infty} \frac{1}{r^2} (f'(Q_\lambda) - k^2) g^2 r dr \geq -\frac{c_1}{\lambda} \|g\|_H^2$$

where we use again crucially here that  $\langle \mathcal{Z}_\lambda \mid g \rangle = 0$ ; see [22, Lemma 5.4, eq. (5.28)] for the proof. Above the constant  $c_1 > 0$  can be made as small as we like by taking  $R > 0$  large enough, which we are free to do. This completes the proof.  $\square$

## 4 Dynamics of non-scattering threshold solutions

### 4.1 Overall scheme

In this section we prove the Main Theorem. We deduce it from the following proposition, whose proof will be split into several lemmas.

**Proposition 4.1** *Let  $\psi(t) : (T_-, T_+) \rightarrow \mathcal{H}_0$  be a solution to (1.5) with  $\mathcal{E}(\psi) = 2\mathcal{E}(\vec{Q})$  which does not scatter in forward time. Then*

$$\lim_{t \rightarrow T_+} \mathbf{d}(\vec{\psi}(t)) = 0. \quad (4.1)$$

To begin with, note the following special case of Theorem 1.2.

**Proposition 4.2** *Let  $\vec{\psi}(t) : (T_-, T_+) \rightarrow \mathcal{H}_0$  be a solution to (1.5) with  $\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(\vec{Q})$  which does not scatter in forward time. Then*

$$\liminf_{t \rightarrow T_+} \mathbf{d}(\vec{\psi}(t)) = 0.$$

An analogous statement holds if  $\vec{\psi}(t)$  does not scatter in backwards time.  $\square$

Let us summarize the main idea of the proof of Proposition 4.1. We know from Proposition 4.2 that (4.1) holds *for a sequence of times*. Thus, in order to obtain (4.1), we should prove that  $\vec{\psi}(t)$ , after exiting a small neighborhood of a two-bubble configuration, can never approach a two-bubble again. Such a result is similar in nature to the *no-return lemma* proved by Krieger et al. [37] in their study of the dynamics near the ground state stationary solution for the energy critical NLW. Such results are usually obtained by means of a convexity argument based on monotonicity formulas, which is also the scheme that we adopt here.

Until the end of this section,  $\vec{\psi}(t)$  always denotes a solution to (1.5),  $\vec{\psi}(t) : (T_-, T_+) \rightarrow \mathcal{H}_0$ , such that  $\mathcal{E}(\vec{\psi}) = 2\mathcal{E}(\vec{Q})$  and  $\vec{\psi}(t)$  does not scatter in forward time. Let  $T_- < \tau_1 \leq \tau_2 < T_+$ . Integrating the virial identity from Lemma 2.15 for  $t \in [\tau_1, \tau_2]$  yields

$$\begin{aligned} \int_{\tau_1}^{\tau_2} \|\partial_t \psi(t)\|_{L^2}^2 dt &\leq |\langle \partial_t \psi | \chi_{Rr} \partial_r \psi \rangle(\tau_1)| + |\langle \partial_t \psi | \chi_{Rr} \partial_r \psi \rangle(\tau_2)| \\ &\quad + \int_{\tau_1}^{\tau_2} \left| \Omega_R(\vec{\psi}(t)) \right| dt \end{aligned}$$

where  $\Omega_R(\vec{\psi}(t))$  is defined in (2.26). Note that for any  $R > 0$  we can use Lemma 2.16 to bound the first two terms on the right-hand-side above and obtain

$$\begin{aligned} \int_{\tau_1}^{\tau_2} \|\partial_t \psi(t)\|_{L^2}^2 dt &\leq C_0 \left( R \sqrt{\mathbf{d}(\vec{\psi}(\tau_1))} + R \sqrt{\mathbf{d}(\vec{\psi}(\tau_2))} \right) \\ &\quad + \int_{\tau_1}^{\tau_2} \left| \Omega_R(\vec{\psi}(t)) \right| dt. \end{aligned} \quad (4.2)$$

Our goal is to show that with a good choice of  $R$ ,  $\tau_1$  and  $\tau_2$  the right hand side can be absorbed into the left hand side. As mentioned in the Introduction, we use different arguments depending whether  $\mathbf{d}(\vec{\psi}(t))$  is small or not.

## 4.2 Splitting of the time axis

We would like to divide the time axis into *good intervals* where  $\mathbf{d}(\vec{\psi}(t))$  is large and *bad intervals* where it is small. We begin with a preliminary splitting, which will then need to be refined.

**Claim 4.3** *Suppose that (4.1) fails. Then for any  $\epsilon_0 > 0$  sufficiently small there exist sequences  $p_n, q_n$  such that*

$$T_- < p_0 < q_0 < p_1 < q_1 < \cdots < p_{n-1} < q_{n-1} < p_n < q_n < \dots$$

such that the following holds for all  $n \in \{0, 1, 2, 3, \dots\}$ :

$$\forall t \in [p_n, q_n] : \mathbf{d}(\vec{\psi}(t)) \leq \epsilon_0, \quad (4.3)$$

$$\forall t \in [q_n, p_{n+1}] : \mathbf{d}(\vec{\psi}(t)) \geq \frac{1}{2}\epsilon_0, \quad (4.4)$$

$$\lim_{n \rightarrow +\infty} p_n = \lim_{n \rightarrow +\infty} q_n = T_+. \quad (4.5)$$

*Proof* Suppose that (4.1) fails and let  $\epsilon_0$  be any number such that

$$0 < \epsilon_0 < \min(\limsup_{t \rightarrow T_+} \mathbf{d}(\vec{\psi}(t)), \eta_1) \quad (4.6)$$

(recall that  $\mathbf{d}(\vec{\psi}(t_0)) < \eta_1$  guarantees that the modulation estimates hold for  $t$  in some neighborhood of  $t_0$ ). Let  $T_0 \in (T_-, T_+)$  be such that  $\mathbf{d}(\vec{\psi}(T_0)) > \epsilon_0$ . We set

$$p_0 := \sup \left\{ t : \mathbf{d}(\vec{\psi}(\tau)) \geq \frac{1}{2}\epsilon_0, \forall \tau \in [T_0, t] \right\}.$$

Proposition 4.1 implies that  $p_0 < T_+$  and  $\mathbf{d}(\vec{\psi}(p_0)) = \frac{1}{2}\epsilon_0$ . Then we define inductively for  $n \geq 1$ :

$$\begin{aligned} q_{n-1} &:= \sup \left\{ t : \mathbf{d}(\vec{\psi}(\tau)) \leq \epsilon_0, \forall \tau \in [p_{n-1}, t] \right\}, \\ p_n &:= \sup \left\{ t : \mathbf{d}(\vec{\psi}(\tau)) \geq \frac{1}{2}\epsilon_0, \forall \tau \in [q_{n-1}, t] \right\}. \end{aligned}$$

By a simple inductive argument using (4.6) and Proposition 4.2 we can show that for  $n \in \{1, 2, \dots\}$  there holds

$$\begin{aligned} p_{n-1} &< q_{n-1} < T_+, \\ q_{n-1} &< p_n < T_+, \\ \mathbf{d}(\vec{\psi}(p_n)) &= \frac{1}{2}\epsilon_0, \end{aligned} \quad (4.7)$$

$$\mathbf{d}(\vec{\psi}(q_n)) = \epsilon_0. \quad (4.8)$$

Bounds (4.4) and (4.3) follow directly from the definitions of  $p_n$  and  $q_n$ . Suppose that (4.5) does not hold. Then, by monotonicity,

$$\lim_{n \rightarrow +\infty} p_n = \lim_{n \rightarrow +\infty} q_n = T_1 < T_+.$$

By the local well-posedness  $\mathbf{d}(\vec{\psi}(t))$  has a limit as  $t \rightarrow T_1$ , which is in contradiction with (4.7) and (4.8).  $\square$

**Claim 4.4** *Let  $\epsilon > 0$ . There exist  $\lambda_0, \epsilon' > 0$  having the following property. Assume that  $\mathbf{d}(\vec{\psi}(t)) < \eta_1$ , with  $\eta_1$  as in Proposition 3.8, and let  $\lambda(t), \mu(t)$  be the modulation parameters given by Lemma 3.1 and let  $\zeta(t)$  be the correction to  $\lambda(t)$  defined in (3.34). Then*

$$\frac{\zeta(t)}{\mu(t)} \geq \lambda_0 \Rightarrow \mathbf{d}(\vec{\psi}(t)) > \epsilon', \quad (4.9)$$

$$\frac{\zeta(t)}{\mu(t)} \leq \lambda_0 \Rightarrow \mathbf{d}(\vec{\psi}(t)) < \epsilon. \quad (4.10)$$

**Remark 4.5** Note that  $\epsilon' < \epsilon$ .

*Proof* Lemma 3.1 yields  $\mathbf{d}(\vec{\psi}(t)) \leq (C^2 + 1)(\lambda(t)/\mu(t))^k \leq 2(C^2 + 1)(\zeta(t)/\mu(t))^k$ , so we get (4.10) with any  $\lambda_0 < (\epsilon/2(C^2 + 1))^{\frac{1}{k}}$ .

In order to prove (4.9), we notice that from Lemma 3.1 and (3.35) we get  $\mathbf{d}(\vec{\psi}(t)) \geq \frac{1}{C}(\zeta(t)/\mu(t))^k$ , hence it suffices to take  $\epsilon' < \frac{1}{C}\lambda_0^k$ .  $\square$

**Lemma 4.6** *Suppose that (4.1) fails. Let  $\epsilon_0 > 0$  be small enough so that Claim 4.3 and Proposition 3.10 hold. Then there exist  $\epsilon, \epsilon' > 0$  with  $\epsilon' < \epsilon$  and  $\epsilon < \frac{1}{10}\epsilon_0$  as in Proposition 3.10, and a splitting of the time axis*

$$T_- < a_1 < c_1 < b_1 < \dots < a_m < c_m < b_m < a_{m+1} < \dots$$

*such that the following holds for all  $m \in \{2, 3, 4, \dots\}$ :*

$$\forall t \in [b_m, a_{m+1}] : \mathbf{d}(\vec{\psi}(t)) \geq \epsilon', \quad (4.11)$$

$$\begin{aligned} \exists t \in [b_m, a_{m+1}] : \mathbf{d}(\vec{\psi}(t)) &\geq 2\epsilon, \\ \mathbf{d}(\vec{\psi}(a_m)) &= \mathbf{d}(\vec{\psi}(b_m)) = \epsilon, \end{aligned} \quad (4.12)$$

$$C_0 \int_{a_{m+1}}^{c_{m+1}} \sqrt{\mathbf{d}(\vec{\psi}(t))} \, dt \leq \frac{1}{10} \int_{b_m}^{a_{m+1}} \|\partial_t \psi(t)\|_{L^2}^2 \, dt \quad (4.13)$$

$$C_0 \int_{c_m}^{b_m} \sqrt{\mathbf{d}(\vec{\psi}(t))} \, dt \leq \frac{1}{10} \int_{b_m}^{a_{m+1}} \|\partial_t \psi(t)\|_{L^2}^2 \, dt \quad (4.14)$$

and

$$\liminf_{m \rightarrow +\infty} \mathbf{d}(\vec{\psi}(c_m)) = 0. \quad (4.15)$$

*Proof* We choose  $\epsilon, \epsilon_0 > 0$  such that Claim 4.3 and Proposition 3.10 hold, where the constant  $C_0$  in Proposition 3.10 is given by Lemma 2.16. We can assume that  $\epsilon < \frac{1}{10}\epsilon_0$ . Let  $\lambda_0$  and  $\epsilon'$  be given by Claim 4.4.

We begin by defining the times  $c_m$ . Let  $0 \leq n_1 < n_2 < \dots$  be the sequence of these indices  $n_m$  for which

$$\inf_{t \in [p_{n_m}, q_{n_m}]} \mathbf{d}(\vec{\psi}(t)) \leq \epsilon'. \quad (4.16)$$

Recall that the modulation parameters  $\lambda(t)$ ,  $\mu(t)$ , and  $\zeta(t) \simeq \lambda(t)$  are well defined on  $[p_{n_m}, q_{n_m}]$ . Let  $c_m \in [p_{n_m}, q_{n_m}]$  be such that

$$\zeta(c_m)/\mu(c_m) = \inf_{t \in [p_{n_m}, q_{n_m}]} \zeta(t)/\mu(t).$$

Claim 4.4 and (4.16) imply that  $\zeta(c_m)/\mu(c_m) < \lambda_0$ , which implies again by Claim 4.4 that  $\mathbf{d}(\vec{\psi}(c_m)) < \epsilon < \frac{1}{10}\epsilon_0$ . Hence  $c_m \in (p_{n_m}, q_{n_m})$  and

$$\left. \frac{d}{dt} \right|_{t=c_m} \left( \frac{\zeta(t)}{\mu(t)} \right) = 0.$$

We will use Proposition 3.10 with various  $t_0$ , in forward and backward direction. Thus the meaning of  $t_0$ ,  $t_1$  and  $t_2$  will change depending on the context.

Using Proposition 3.10 with  $t_0 = c_m$  in the backward time direction we obtain times  $t_1 \leq c_m$  and  $t_2 \leq t_1$ . Note that (3.41) and (4.4) imply that  $t_1 \in (p_{n_m}, c_m]$ . We set

$$a_m := \sup\{t \geq t_1 : \mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \forall \tau \in [t_1, t]\}.$$

By (3.40) we have  $\mathbf{d}(\vec{\psi}(t_1)) > \epsilon$ . Since  $\mathbf{d}(\vec{\psi}(c_m)) < \epsilon$ , we have  $a_m \in (p_{n_m}, c_m)$  and  $\mathbf{d}(\vec{\psi}(a_m)) = \epsilon$ .

Denote  $\sigma_m := t_2$ . Then (3.40) yields  $\mathbf{d}(\vec{\psi}(t)) \geq \epsilon$  for  $t \in [\sigma_m, t_1]$ . By definition of  $a_m$  we also have  $\mathbf{d}(\vec{\psi}(t)) \geq \epsilon$  for  $t \in [t_1, a_m]$ , hence

$$\mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \quad \forall t \in [\sigma_m, a_m]. \quad (4.17)$$

Bound (3.42) together with (4.3) yields  $\sigma_m < p_{n_m}$ , so (4.17) implies that

$$\mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \quad \forall t \in [p_{n_m}, a_m]. \quad (4.18)$$

Finally, (3.43) yields

$$\int_{\sigma_m}^{a_m} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq C \int_{a_m}^{c_m} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt. \quad (4.19)$$

Now using Proposition 3.10 with  $t_0 = c_m$  in the forward time direction we obtain times  $t_1 \geq c_m$  and  $t_2 \geq t_1$ . Note that (3.41) and (4.4) imply that  $t_1 \in [c_m, q_{n_m}]$ . We set

$$b_m := \inf\{t \leq t_1 : \mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \quad \forall \tau \in [t, t_1]\}.$$

As in the case of  $a_m$ , we obtain  $b_m \in (c_m, q_{n_m})$  and  $\mathbf{d}(\vec{\psi}(b_m)) = \epsilon$ . Denote  $\tau_m := t_2$ . Adapting the proofs of (4.17) and (4.18) we get

$$\mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \quad \forall t \in [b_m, \tau_m], \quad (4.20)$$

$$\mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \quad \forall t \in [b_m, q_{n_m}], \quad (4.21)$$

$$\int_{b_m}^{\tau_m} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq C \int_{c_m}^{b_m} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt. \quad (4.22)$$

We will prove that  $\tau_m < a_{m+1}$ . Suppose not. Since  $\mathbf{d}(\vec{\psi}(\tau_m)) \geq 2\epsilon_0$ , see (3.42), the fact that  $a_{m+1} \in [p_{n_{m+1}}, q_{n_{m+1}}]$  would imply that  $\tau_m > q_{n_{m+1}}$ . Thus by (4.20) we would have  $\mathbf{d}(\vec{\psi}(t)) \geq \epsilon$ ,  $\forall t \in [b_m, q_{n_{m+1}}]$ . But  $b_m < q_{n_m} < p_{n_{m+1}}$ , so we obtain

$$\mathbf{d}(\vec{\psi}(t)) \geq \epsilon, \quad \forall t \in [p_{n_{m+1}}, q_{n_{m+1}}].$$

Since  $\epsilon' < \epsilon$ , this contradicts the definition of  $n_{m+1}$ . Thus  $\tau_m < a_{m+1}$ , so (4.22) implies (4.14). Also  $\mathbf{d}(\vec{\psi}(\tau_m)) \geq 2\epsilon_0$  yields (4.12), since  $\epsilon_0 \geq \epsilon$ .

Analogously, we have  $\sigma_{m+1} > b_m$ , so (4.19) yields (4.13).

It remains to prove (4.11). Take  $t$  such that  $\mathbf{d}(\vec{\psi}(t)) < \epsilon'$ . Then  $t \in [p_{n_m}, q_{n_m}]$  for some  $m$ , so (4.18) and (4.21) yield  $t \in [a_m, b_m]$ , which is exactly (4.11).

Finally, (4.15) follows from Proposition 4.2, (3.8), and (3.35).  $\square$

*Remark 4.7* It follows from the proof that  $\epsilon$  can be taken as small as we wish.

Until the end of the proof of Proposition 4.1, we fix  $\epsilon, \epsilon' > 0$  and a partition of the time axis given by the last lemma. In particular, all the constants are allowed to depend on  $\epsilon$  and  $\epsilon'$ . We denote

$$I_m := [b_{m-1}, a_m], \quad I := \bigcup_{m \geq 1} I_m.$$

Then (4.11) is equivalent to

$$\forall t \in I : \mathbf{d}(\vec{\psi}(t)) \geq \epsilon'. \quad (4.23)$$

We will see that  $\vec{\psi}(t)$  has a *compactness property* for  $t \in I$ , which allows to deal with the right hand side of (4.2) for  $t \in I$ . For  $t \notin I$  we will rely on (4.13) and (4.14).

### 4.3 Compactness on $I$

The objective of this step is to deduce a compactness statement on  $I$  that will allow us to obtain a lower bound for the left-hand-side of (4.2) restricted to  $I$  and to uniformly control the errors  $\Omega_R(\vec{\psi}(t))$  on  $I$ , by choosing  $R$  large enough.

**Lemma 4.8** *There exists a continuous function  $v : I \rightarrow (0, +\infty)$  such that the set*

$$\mathcal{K} := \{\vec{\psi}(t)_{1/v(t)} \mid t \in I\} \subset \mathcal{H}_0$$

*is pre-compact in  $\mathcal{H}_0$ .*

*Proof* We will first prove that for any sequence  $\{t_n\} \in I$  there exists a subsequence (still denoted by  $t_n$ ) and a sequence of scales  $v_n$ , so that

$$\vec{\psi}(t_n)_{1/v_n} \rightarrow \vec{\varphi} \in \mathcal{H}_0 \quad (4.24)$$

for some  $\vec{\varphi} \in \mathcal{H}_0$ .

We observe that by Lemmas 2.13 and (4.23) we have the uniform bound

$$\|\vec{\psi}(t)\|_{\mathcal{H}_0} \leq C(\epsilon') \quad \forall t \in I,$$

which means in particular that

$$\|\vec{\psi}(t_n)\|_{\mathcal{H}_0} \leq C(\epsilon') < \infty.$$

Thus (4.24) follows from Lemma 2.9.

We are now ready to construct the function  $v(t)$ . For each  $t \in I$  let  $v(t)$  be the unique number such that

$$\begin{aligned} & \int_0^\infty e^{-r} \left( (\partial_t \psi_{1/v(t)}(t, r))^2 + (\partial_r \psi_{1/v(t)}(t, r))^2 + k^2 \frac{(\psi_{1/v(t)}(t, r))^2}{r^2} \right) r \, dr \\ &= \frac{1}{2} \|\vec{\psi}(t)\|_{\mathcal{H}_0}^2 \end{aligned}$$

(the function  $e^{-r}$  could be replaced by any continuous strictly decreasing function whose value is 1 for  $r = 0$  and tending to 0 as  $r \rightarrow +\infty$ ). We see that  $v(t)$  is a continuous function.

Suppose that  $\vec{\psi}(t)_{1/v(t)}$  is not pre-compact in  $\mathcal{H}_0$ . Thus there exists a sequence  $\vec{\psi}(t_n)_{1/v(t_n)}$  which has no convergent subsequence. But we know (by assumption) that there exist a subsequence (still denoted  $t_n$ ) and numbers  $v_n$  such that  $\vec{\psi}(t_n)_{1/v_n}$  converges in  $\mathcal{H}_0$  to some  $\vec{\varphi} = (\varphi_0, \varphi_1)$ . This implies that

$$\int_0^\infty e^{-r} \left( (\partial_t \psi_{1/v_n}(t_n, r))^2 + (\partial_r \psi_{1/v_n}(t_n, r))^2 + k^2 \frac{(\psi_{1/v_n}(t_n, r))^2}{r^2} \right) r \, dr$$

converges to

$$\int_0^\infty e^{-r} \left( \varphi_1(r)^2 + (\partial_r \varphi_0(r))^2 + k^2 \frac{\varphi_0(r)^2}{r^2} \right) r \, dr \in (0, \|\vec{\varphi}\|_{\mathcal{H}_0}).$$

Since  $\|\vec{\psi}(t_n)\|_{\mathcal{H}_0}$  converges to  $\|\vec{\varphi}\|_{\mathcal{H}_0}$ , we deduce that  $v_n/v(t_n)$  is bounded. This implies that  $v_n/v(t_n)$  has a convergent subsequence, hence  $\vec{\psi}(t_n)_{1/v(t_n)}$  has a convergent subsequence, so we have a contradiction. This completes the proof.  $\square$

For  $m \in \{1, 2, 3, \dots\}$  we define

$$v_m := |I_m| = a_m - b_{m-1}.$$

**Lemma 4.9** *There exists  $C_1 > 0$  such that for all  $m \geq 1$  and all  $t \in I_m$  there holds*

$$\frac{1}{C_1} v(t) \leq v_m \leq C_1 v(t).$$

**Remark 4.10** The last lemma tells us that  $v(t)$  is comparable to  $v_m$  for  $t \in I_m$ . In particular, the set

$$\mathcal{K}_1 := \bigcup_{m \geq 1} \{\vec{\psi}(t)_{1/v_m} \mid t \in I_m\} \quad (4.25)$$

is pre-compact in  $\mathcal{H}_0$ .

*Proof of Lemma 4.9* Suppose that there exists a sequence  $m_\ell$  and times  $t_\ell \in I_{m_\ell}$  such that

$$\lim_{\ell \rightarrow +\infty} \frac{v_{m_\ell}}{v(t_\ell)} = 0. \quad (4.26)$$

Let  $\vec{\psi}_\ell(s)$  be the solution of (1.5) with initial data  $\vec{\psi}_\ell(0) = \vec{\psi}(t_\ell)_{1/v(t_\ell)}$ . After extracting a subsequence,  $\vec{\psi}_\ell(0)$  converges in  $\mathcal{H}_0$  to some  $\vec{\varphi}_0$ . Let  $\vec{\varphi}(s) : [-s_0, s_0] \rightarrow \mathcal{H}_0$  be the solution of (1.5) with initial data  $\vec{\varphi}(0) = \vec{\varphi}_0$  (where  $s_0 > 0$ ). By the standard Cauchy theory, for sufficiently large  $\ell$  the solution  $\vec{\psi}_\ell(s)$  is defined for  $s \in [-s_0, s_0]$  and  $\vec{\psi}_\ell(s) \rightarrow \vec{\varphi}(s)$  in  $\mathcal{H}_0$ , uniformly for  $s \in [-s_0, s_0]$ .

Let  $t'_\ell \in I_{m_\ell}$  be any sequence. Let  $s_\ell = \frac{t'_\ell - t_\ell}{v(t_\ell)}$ . Then (4.26) implies that  $\lim_{\ell \rightarrow +\infty} s_\ell = 0$ . Thus  $s_\ell \in [-s_0, s_0]$  for large  $\ell$  and we deduce that

$$\lim_{\ell \rightarrow +\infty} \|\vec{\psi}_\ell(s_\ell) - \vec{\varphi}(s_\ell)\|_{\mathcal{H}_0} = 0.$$

But of course  $\lim_{\ell \rightarrow +\infty} \|\vec{\varphi}(s_\ell) - \vec{\varphi}_0\|_{\mathcal{H}_0} = 0$ , so the triangle inequality yields

$$\lim_{\ell \rightarrow +\infty} \|\vec{\psi}_\ell(s_\ell) - \vec{\varphi}_0\|_{\mathcal{H}_0} = 0.$$

In particular,  $\lim_{\ell \rightarrow +\infty} \mathbf{d}(\vec{\psi}_\ell(s_\ell)) = \mathbf{d}(\vec{\varphi}_0)$ . We have  $\vec{\psi}_\ell(s_\ell) = \vec{\psi}(t'_\ell)_{1/v(t_\ell)}$ , thus  $\mathbf{d}(\vec{\psi}(t'_\ell)) = \mathbf{d}(\vec{\psi}_\ell(s_\ell))$  and we obtain

$$\lim_{\ell \rightarrow +\infty} \mathbf{d}(\vec{\psi}(t'_\ell)) = \mathbf{d}(\vec{\varphi}_0),$$

for any sequence  $t'_\ell \in I_{m_\ell}$ . This is impossible, because we know that  $\mathbf{d}(\vec{\psi}(a_{m_\ell})) = \epsilon$  and on the other hand for each  $\ell$  we have  $\sup_{t \in I_{m_\ell}} \mathbf{d}(\vec{\psi}(t)) \geq 2\epsilon$ .

Now suppose that there exist a sequence  $m_\ell$  and times  $t_\ell \in I_{m_\ell}$  such that

$$\lim_{\ell \rightarrow +\infty} \frac{v(t_\ell)}{v_{m_\ell}} = 0.$$

Without loss of generality we can assume that

$$\lim_{\ell \rightarrow +\infty} \frac{v(t_\ell)}{a_{m_\ell} - t_\ell} = 0$$

(the case  $v(t_\ell)/(t_\ell - b_{m_\ell-1}) \rightarrow 0$  can be treated similarly).

Again, let  $\vec{\psi}_\ell(s)$  be the solution of (1.5) with initial data  $\vec{\psi}_\ell(0) = \vec{\psi}(t_\ell)_{1/v(t_\ell)}$ , and let  $\vec{\psi}_\ell(0) \rightarrow \vec{\varphi}_0 \in \mathcal{H}_0$ . Let  $\vec{\varphi}(s) : (-T_-(\vec{\varphi}_0), T_+(\vec{\varphi}_0)) \rightarrow \mathcal{H}_0$  be the solution of (1.5) with initial data  $\vec{\varphi}(0) = \vec{\varphi}_0$ . By Lemma 2.9 we know that  $\vec{\varphi}(s)$  is non-scattering in both time directions and satisfies

$$\mathcal{E}(\vec{\varphi}) = \mathcal{E}(\vec{\psi}) = 2\mathcal{E}(\vec{Q}).$$

Thus Proposition 4.2 implies that there exists  $\sigma \in [0, T_+(\vec{\varphi}_0))$  such that  $\mathbf{d}(\vec{\varphi}(\sigma)) \leq \frac{1}{2}\epsilon'$ . By Cauchy theory, for  $\ell$  large enough  $\vec{\psi}_\ell(s)$  is defined for  $s \in [0, \sigma]$  and  $\vec{\psi}_\ell(\sigma) \rightarrow \vec{\varphi}(\sigma)$  in  $\mathcal{H}_0$ , in particular  $\mathbf{d}(\vec{\psi}_\ell(\sigma)) \rightarrow \mathbf{d}(\vec{\varphi}(\sigma)) \leq \frac{1}{2}\epsilon'$ .

Let  $t'_\ell := t_\ell + v(t_\ell)\sigma$ . Then  $\vec{\psi}(t'_\ell) = \vec{\psi}_\ell(\sigma)_{v(t_\ell)}$ , so we have

$$\lim_{\ell \rightarrow +\infty} \mathbf{d}(\vec{\psi}(t'_\ell)) = \lim_{\ell \rightarrow +\infty} \mathbf{d}(\vec{\psi}_\ell(\sigma)) \leq \frac{1}{2}\epsilon'.$$

However, (4.26) implies that for  $\ell$  large enough there holds  $t_\ell \leq t'_\ell \leq a_{m_\ell}$ , thus (4.11) yields  $\mathbf{d}(\vec{\psi}(t'_\ell)) \geq \epsilon'$ . The contradiction finishes the proof.  $\square$

**Lemma 4.11** *There exists  $\delta_1 > 0$  such that for all  $m$  there holds*

$$\int_{I_m} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq \delta_1^2 v_m.$$

*Proof* Let  $t_m := \frac{1}{2}(b_{m-1} + a_m)$  and recall that  $v_m := a_m - b_{m-1}$ . Then for any  $0 < s_1 \leq 1/2$ ,

$$b_{m-1} \leq t_m - v_m s_1 \leq t_m + v_m s_1 \leq a_m. \quad (4.27)$$

We consider the following sequence of solutions of (1.5):

$$\vec{\psi}_m(s) := \vec{\psi}(t_m + v_m s)_{1/v_m} \quad \text{for } s \in [-s_1, s_1].$$

Then (4.27) implies that

$$\int_{I_m} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq v_m \int_{-s_1}^{s_1} \|\partial_s \psi_m(s)\|_{L^2}^2 ds.$$

Suppose that the conclusion fails. Then there exists a sequence  $m_1, m_2, \dots$  such that

$$\lim_{l \rightarrow +\infty} \int_{-s_1}^{s_1} \|\partial_s \psi_{m_l}(s)\|_{L^2}^2 ds = 0. \quad (4.28)$$

After extraction of a subsequence,  $\vec{\psi}_{m_l}(0) \rightarrow \vec{\varphi}_0 \in \mathcal{H}_0$ . Let  $\vec{\varphi}(s)$  be the solution of (1.5) such that  $\vec{\varphi}(0) = \vec{\varphi}_0$ . Then by the standard Cauchy theory  $\vec{\psi}_{m_l}(s) \rightarrow \vec{\varphi}(s)$  in  $\mathcal{H}_0$ , uniformly for  $s \in [-s_1, s_1]$  for  $s_1 > 0$  small enough. In particular, (4.28) yields

$$\int_{-s_1}^{s_1} \|\partial_s \varphi(s)\|_{L^2}^2 ds = 0,$$

so the limiting wave map  $\vec{\varphi}(s)$  must be time-independent. Hence  $\varphi(s) \in H$  is a harmonic map. But then  $\varphi(s) \equiv 0$  since the constant map is the unique harmonic map with topological degree 0. However, we also have  $\mathcal{E}(\vec{\varphi}) = 2\mathcal{E}(\vec{Q}) > 0$ , which gives a contradiction.  $\square$

**Lemma 4.12** *There exists  $R_0 > 0$  such that if  $R_1 \geq R_0$ , then for all  $m \in \{2, 3, \dots\}$  there holds*

$$\int_{I_m} \Omega_{v_m R_1}(\vec{\psi}(t)) dt \leq \frac{\delta_1^2}{10} v_m.$$

*Proof* With a change of variables, it suffices to prove that for all  $t \in I$  there holds

$$\Omega_{R_1}(\vec{\psi}(t)_{1/v_m}) \leq \frac{\delta_1^2}{10}.$$

This is a standard consequence of the pre-compactness of the set  $\mathcal{K}_1$  defined in (4.25).  $\square$

## 4.4 Conclusions

*Proof of Proposition 4.1* Choose  $1 \leq m_1 < m_2$  such that

$$\sqrt{\mathbf{d}(\vec{\psi}(c_{m_1}))} + \sqrt{\mathbf{d}(\vec{\psi}(c_{m_2}))} \leq \frac{\delta_1^2}{10C_0 R_0}. \quad (4.29)$$

This is possible thanks to (4.15). Let

$$R := R_0 \max_{m_1 < m \leq m_2} v_m.$$

Inequalities (4.13) and (2.28) yield

$$\frac{1}{5} \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq \sum_{m=m_1+1}^{m_2} \frac{1}{5} \int_{b_{m-1}}^{a_m} \|\partial_t \psi(t)\|_{L^2}^2 dt$$

$$\begin{aligned} &\geq \sum_{m=m_1+1}^{m_2} 2C_0 \int_{a_m}^{c_m} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt \\ &\geq 2 \sum_{m=m_1+1}^{m_2} \int_{a_m}^{c_m} \Omega_R(\vec{\psi}(t)) dt. \end{aligned}$$

Similarly, (4.14) and (2.28) yield

$$\begin{aligned} \frac{1}{5} \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt &\geq \sum_{m=m_1}^{m_2-1} \frac{1}{5} \int_{b_m}^{a_{m+1}} \|\partial_t \psi(t)\|_{L^2}^2 dt \\ &\geq \sum_{m=m_1}^{m_2-1} 2C_0 \int_{c_m}^{b_m} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt \\ &\geq 2 \sum_{m=m_1}^{m_2-1} \int_{c_m}^{b_m} \Omega_R(\vec{\psi}(t)) dt. \end{aligned}$$

Next, from Lemma 4.11 we have

$$\frac{1}{5} \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq \sum_{m=m_1+1}^{m_2} \frac{1}{5} \int_{b_{m-1}}^{a_m} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq \frac{\delta_1^2}{5} \sum_{m=m_1+1}^{m_2} \nu_m. \quad (4.30)$$

By the definition of  $R$ , for each  $m \in \{m_1 + 1, m_1 + 2, \dots, m_2\}$  we have  $R = R_1 \nu_m$  with  $R_1 \geq R_0$ . Thus Lemma 4.12 gives

$$\frac{1}{5} \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq 2 \sum_{m=m_1+1}^{m_2} \int_{b_{m-1}}^{a_m} \Omega_R(\vec{\psi}(t)) dt.$$

Finally, (4.30) and (4.29) imply

$$\begin{aligned} \frac{1}{5} \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt &\geq \frac{\delta_1^2}{5} \max_{m_1 < m \leq m_2} \nu_m \\ &\geq 2C_0 R \left( \sqrt{\mathbf{d}(\vec{\psi}(c_{m_1}))} + \sqrt{\mathbf{d}(\vec{\psi}(c_{m_2}))} \right). \end{aligned}$$

Summing the four inequalities above and using (4.2) for  $(\tau_1, \tau_2) = (c_{m_1}, c_{m_2})$  we get

$$\frac{4}{5} \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt \geq 2 \int_{c_{m_1}}^{c_{m_2}} \|\partial_t \psi(t)\|_{L^2}^2 dt.$$

This contradiction finishes the proof.  $\square$

*Proof of the Main Theorem 1.6 Step 1* Let  $\epsilon > 0$  be such that Proposition 3.10 holds with some  $\epsilon_0 \geq 10\epsilon$ . Define

$$T_1 := \sup\{t : \exists t' \geq t \text{ such that } \mathbf{d}(\vec{\psi}(t')) \geq \epsilon\}.$$

By Proposition 4.1 we have  $T_1 < T_+$ , and we know that the modulation parameters  $\lambda(t)$  and  $\mu(t)$  are well-defined for  $t \in [T_1, T_+)$ . Define  $\zeta(t)$  as in (3.34). Assume without loss of generality that  $\mu(T_1) = 1$ .

There exists a sequence  $\tau_n \rightarrow T_+$  such that

$$\frac{d}{dt} \Big|_{t=\tau_n} \left( \frac{\zeta(t)}{\mu(t)} \right) \leq 0.$$

For any such  $t_0 = \tau_n$  we are in the setting of Proposition 3.10 in the backward time direction, so we obtain times  $t_1 \leq \tau_n$  and  $t_2 \leq t_1$ . By the definition of  $T_1$  and (3.40) we have  $t_1 \leq T_1$ , so the proof of Proposition 3.10 yields in particular  $\frac{1}{2} \leq \mu(t) \leq 2$  for  $t \in [t_1, \tau_n]$ , thus  $\frac{1}{2} \leq \mu(t) \leq 2$  for  $t \in [T_1, T_+)$ .

Furthermore, (3.44) implies that  $\int_{T_1}^{\tau_n} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt$  is bounded as  $\tau_n \rightarrow T_+$ . Thus (3.29) implies that  $\int_{T_1}^{T_+} |\mu'(t)| dt < +\infty$ , hence  $\mu(t)$  converges to some  $\mu_0 \in [\frac{1}{2}, 2]$ . Eventually rescaling again, we can assume that  $\mu_0 = 1$ .

As in the proof of Proposition 3.10, we consider  $\xi(t) := b(t) + \kappa_1 \zeta(t)^{\frac{k}{2}}$  and we find that it is strictly decreasing on  $[T_1, \tau_n]$  and satisfies

$$\xi'(t) \leq -\kappa_2 \xi(t)^{\frac{2k-2}{k}}. \quad (4.31)$$

Hence  $\xi(t)$  is strictly decreasing on  $[T_1, T_+)$  and satisfies (4.31) for  $t \in [T_1, T_+)$ . From the modulation equations we also obtain

$$\xi'(t) \geq -\kappa_3 \xi(t)^{\frac{2k-2}{k}}. \quad (4.32)$$

for some  $\kappa_3$  depending only on  $k$ . Indeed, (3.28) and (3.38) yield

$$|\xi'(t)| \lesssim |b'(t)| + |\zeta'(t)| \zeta(t)^{\frac{k-2}{2}} \lesssim \zeta(t)^{k-1} \lesssim \xi(t)^{\frac{2k-2}{k}},$$

Since  $\lim_{t \rightarrow T_+} \xi(t) = 0$  and  $\frac{2k-2}{k} \geq 1$ , (4.32) implies that  $T_+ = +\infty$ .

Showing that the sign  $\iota$  is constant is standard. Lemma 2.14 implies that  $\mathbf{d}_+(\vec{\psi}(t)) \leq \epsilon$  for  $t \in [T_1, +\infty)$  or  $\mathbf{d}_-(\vec{\psi}(t)) \leq \epsilon$  for  $t \in [T_1, +\infty)$ . Indeed,

suppose that  $t_1, t_2 \geq T_1, t_1 \leq t_2$  are such that  $\mathbf{d}_+(\vec{\psi}(t_1)) \leq \epsilon$  and  $\mathbf{d}_-(\vec{\psi}(t_2)) \leq \epsilon$ . Without loss of generality we can assume that  $\epsilon < \frac{1}{2}\alpha_0$ , where  $\alpha_0$  is the constant from Lemma 2.14. Then Lemma 2.14 yields  $\mathbf{d}_+(\vec{\psi}(t_2)) \geq \alpha_0$ , hence there exists  $t_0 \in [t_1, t_2]$  such that  $\mathbf{d}_+(\vec{\psi}(t_0)) = \frac{1}{2}\alpha_0 > \epsilon$ . But Lemma 2.14 gives that also  $\mathbf{d}_-(\vec{\psi}(t_0)) \geq \alpha_0 > \epsilon$ , which contradicts the choice of  $T_1$ .

*Step 2* We now deduce the rate of decay of  $\lambda(t)$  as  $t \rightarrow +\infty$ . Bounds (3.50), (3.53), and (3.35) imply that  $\xi(t)$  is comparable to  $\lambda(t)^{\frac{k}{2}}$ . Rewrite (4.31) and (4.32) as follows:

$$-\kappa_3 \leq \frac{\xi'(t)}{\xi(t)^{\frac{2k-2}{k}}} \leq -\kappa_2, \quad \kappa_2, \kappa_3 > 0,$$

In the case  $k = 2$ , after integrating and possibly changing the values of the constants, we obtain

$$e^{-\kappa_3 t} \leq \xi(t) \leq e^{-\kappa_2 t},$$

which implies that there exists a constant  $C$  such that

$$e^{-Ct} \leq \lambda(t) \leq e^{-\frac{1}{C}t} \quad \text{as } t \rightarrow +\infty. \quad (4.33)$$

(recall that we rescale the solution so that  $\mu_0 = \lim_{t \rightarrow +\infty} \mu(t) = 1$ ).

Similarly, for  $k > 2$  we obtain

$$\frac{1}{C}t^{-\frac{2}{k-2}} \leq \lambda(t) \leq Ct^{-\frac{2}{k-2}} \quad \text{as } t \rightarrow +\infty, \quad (4.34)$$

with a constant depending on  $k$ .

*Step 3* Suppose that  $\vec{\psi}$  does not scatter in either time direction. Take any  $\delta > 0$ . Bounds (4.33) and (4.34) imply that

$$\int_{-\infty}^{+\infty} \sqrt{\mathbf{d}(\vec{\psi}(t))} dt < +\infty. \quad (4.35)$$

Indeed, it suffices to consider the behavior as  $t \rightarrow \pm\infty$ . In this situation we have well-defined modulation parameters  $\lambda(t)$ ,  $\mu(t)$  and (3.7) together with the fact that  $\mu(t) \rightarrow \mu_0 > 0$  imply that  $\mathbf{d}(\vec{\psi}(t)) \lesssim \lambda(t)^k$ , so (4.35) follows from time integrability of  $\lambda(t)^{\frac{k}{2}}$ .

Thus (2.28) implies that there exist  $T_1, T_2$  such that for all  $R > 0$

$$\begin{aligned} \int_{-\infty}^{T_1} \Omega_R(\vec{\psi}(t)) dt &\leq \frac{1}{3}\delta, \\ \int_{T_2}^{+\infty} \Omega_R(\vec{\psi}(t)) dt &\leq \frac{1}{3}\delta. \end{aligned}$$

Since  $[T_1, T_2]$  is a finite time interval, there exists  $R > 0$  such that

$$\int_{T_1}^{T_2} \Omega_R(\vec{\psi}(t)) \, dt \leq \frac{1}{3}\delta.$$

But  $\mathbf{d}(\vec{\psi}(t)) \rightarrow 0$  as  $t \rightarrow \pm\infty$ , hence (4.2) yields

$$\int_{-\infty}^{+\infty} \|\partial_t \psi(t)\|_{L^2}^2 \, dt \leq \delta.$$

This would imply that  $\vec{\psi}(t)$  is a constant in time solution (because  $\delta$  was any strictly positive number), which is impossible.  $\square$

*Remark 4.13* Suppose that  $\vec{\psi}(t)$  does not scatter in the forward time direction. From the modulation equations we know that

$$b'(t) \leq -\kappa_6 \lambda(t)^{k-1}, \quad \kappa_6 > 0,$$

so integration yields

$$b(t) \geq \kappa_7 \int_t^\infty \lambda(s)^{k-1} \, ds.$$

But, as noticed in Step 1. above, we also have  $|(\lambda(t)^{\frac{k}{2}})'| \lesssim \lambda(t)^{k-1}$ , which yields  $\lambda(t)^{\frac{k}{2}} \lesssim \int_t^\infty \lambda(s)^{k-1} \, ds$ , so we obtain

$$b(t) \geq \kappa_8 \lambda(t)^{\frac{k}{2}} \quad \Rightarrow \quad b(t)^2 \geq \kappa_9 \mathbf{d}(\vec{\psi}(t)),$$

which in turn implies

$$\left| \left\langle \frac{1}{\lambda(t)} \Lambda \mathcal{Q}_{\lambda(t)}, \partial_t \psi \right\rangle \right|^2 \geq c \mathbf{d}(\vec{\psi}(t)), \quad \text{as } t \rightarrow +\infty,$$

where  $c > 0$  is a constant depending only on  $k$ . Thus the projection of the time derivative of the solution constitutes at least a fixed fraction of the total distance from a two-bubble. In fact, if we were more precise in our computations, we could probably obtain that this projection is the leading term of the error.

**Acknowledgements** J. Jendrej was supported by the ERC Grant 291214 BLOWDISOL and by the NSF Grant DMS-1463746. This work was completed during his postdoc at the University of Chicago. A. Lawrie was supported by NSF Grant DMS-1700127. We would like to thank Raphaël Côte for many helpful discussions. And lastly, we would like to thank the anonymous referees for their careful reading of an earlier version of the manuscript and for suggesting substantial improvements.

## References

1. Bahouri, H., Gérard, P.: High frequency approximation of solutions to critical nonlinear wave equations. *Am. J. Math.* **121**, 131–175 (1999)
2. Burq, N., Planchon, F., Stalker, J.G., Tahvildar-Zadeh, A.S.: Strichartz estimates for the wave and Schrödinger equations with the inverse-square potential. *J. Funct. Anal.* **203**(2), 519–549 (2003)
3. Burq, N., Planchon, F., Stalker, J.G., Tahvildar-Zadeh, A.S.: Strichartz estimates for the wave and Schrödinger equations with potentials of critical decay. *Indiana Univ. Math. J.* **53**(6), 1665–1680 (2004)
4. Chow, S.-N., Hale, J.K.: *Methods of Bifurcation Theory*, volume 251 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. Springer, New York (1982)
5. Christodoulou, D., Tahvildar-Zadeh, A.S.: On the asymptotic behavior of spherically symmetric wave maps. *Duke Math. J.* **71**(1), 31–69 (1993)
6. Christodoulou, D., Tahvildar-Zadeh, A.S.: On the regularity of spherically symmetric wave maps. *Commun. Pure Appl. Math.* **46**(7), 1041–1091 (1993)
7. Côte, R.: Instability of nonconstant harmonic maps for the  $(1+2)$ -dimensional equivariant wave map system. *Int. Math. Res. Not.* **57**, 3525–3549 (2005)
8. Côte, R.: On the soliton resolution for equivariant wave maps to the sphere. *Commun. Pure Appl. Math.* **68**(11), 1946–2004 (2015)
9. Côte, R., Kenig, C., Lawrie, A., Schlag, W.: Characterization of large energy solutions of the equivariant wave map problem: I. *Am. J. Math.* **137**(1), 139–207 (2015)
10. Côte, R., Kenig, C., Lawrie, A., Schlag, W.: Characterization of large energy solutions of the equivariant wave map problem: II. *Am. J. Math.* **137**(1), 209–250 (2015)
11. Côte, R., Kenig, C.E., Schlag, W.: Energy partition for the linear radial wave equation. *Math. Ann.* **358**(3–4), 573–607 (2014)
12. Duyckaerts, T., Jia, H., Kenig, C., Merle, F.: Universality of blow up profile for small blow up solutions to the energy critical wave map equation (2016). [arXiv:1612.04927](https://arxiv.org/abs/1612.04927)
13. Duyckaerts, T., Jia, H., Kenig, C.E., Merle, F.: Soliton resolution along a sequence of times for the focusing energy critical wave equation. *Geom. Funct. Anal.* **27**(4), 798–862 (2017)
14. Duyckaerts, T., Kenig, C., Merle, F.: Universality of the blow-up profile for small radial type II blow-up solutions of the energy critical wave equation. *J. Eur. Math. Soc. (JEMS)* **13**(3), 533–599 (2011)
15. Duyckaerts, T., Kenig, C., Merle, F.: Profiles of bounded radial solutions of the focusing, energy-critical wave equation. *Geom. Funct. Anal.* **22**(3), 639–698 (2012)
16. Duyckaerts, T., Kenig, C., Merle, F.: Universality of the blow-up profile for small type II blow-up solutions of the energy-critical wave equation: the nonradial case. *J. Eur. Math. Soc. (JEMS)* **14**(5), 1389–1454 (2012)
17. Duyckaerts, T., Kenig, C., Merle, F.: Classification of radial solutions of the focusing, energy critical wave equation. *Camb. J. Math.* **1**(1), 75–144 (2013)
18. Duyckaerts, T., Merle, F.: Dynamics of threshold solutions for energy-critical wave equation. *Int. Math. Res. Pap.* **2008**, rpn2002 (2008). <https://doi.org/10.1093/imrp/rpn002>
19. Duyckaerts, T., Merle, F.: Dynamic of threshold solutions for energy-critical NLS. *Geom. Funct. Anal.* **18**(6), 1787–1840 (2009)
20. Eells, J., Wood, J.C.: Restrictions on harmonic maps of surfaces. *Topology* **15**(3), 263–266 (1976)
21. Grinis, R.: Quantization of time-like energy for wave maps into spheres. *Commun. Math. Phys.* **352**(2), 641–702 (2017)
22. Jendrej, J.: Construction of two-bubble solutions for energy-critical wave equations. *Am. J. Math.* (to appear). [arXiv:1602.06524](https://arxiv.org/abs/1602.06524)

23. Jendrej, J.: Nonexistence of radial two-bubbles with opposite signs for the energy-critical wave equation. *Ann. Sci. Norm. Super. Pisa Cl. Sci. (to appear)*. [arXiv:1510.03965](https://arxiv.org/abs/1510.03965)
24. Jendrej, J.: Construction of two-bubble solutions for the energy-critical NLS. *Anal. PDE* **10**(8), 1923–1959 (2017)
25. Jia, H., Kenig, C.: Asymptotic decomposition for semilinear wave and equivariant wave map equations. *Am. J. Math.* **139**(6), 1521–1603 (2017)
26. Kenig, C., Merle, F.: Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case. *Invent. Math.* **166**(3), 645–675 (2006)
27. Kenig, C., Merle, F.: Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation. *Acta Math.* **201**(2), 147–212 (2008)
28. Klainerman, S., Machedon, M.: Space-time estimates for null forms and the local existence theorem. *Commun. Pure Appl. Math.* **46**(9), 1221–1268 (1993)
29. Klainerman, S., Machedon, M.: Smoothing estimates for null forms and applications. *Int. Math. Res. Not.* **1994**(9), 383–389 (1994). <https://doi.org/10.1155/S1073792894000425>
30. Klainerman, S., Machedon, M.: Smoothing estimates for null forms and applications. *Duke Math. J.* **81**(1), 99–133 (1995)
31. Klainerman, S., Machedon, M.: On the regularity properties of a model problem related to wave maps. *Duke Math. J.* **87**(3), 553–589 (1997)
32. Klainerman, S., Selberg, S.: Remark on the optimal regularity for equations of wave maps type. *Commun. Partial Differ. Equ.* **22**(5–6), 901–918 (1997)
33. Klainerman, S., Selberg, S.: Bilinear estimates and applications to nonlinear wave equations. *Commun. Contemp. Math.* **4**(2), 223–295 (2002)
34. Krieger, J.: Global regularity of wave maps from  $R^{2+1}$  to  $H^2$ . Small energy. *Commun. Math. Phys.* **250**(3), 507–580 (2004)
35. Krieger, J.: On stability of type II blow up for the critical NLW on  $\mathbb{R}^3$  (2017). [arXiv:1705.03907](https://arxiv.org/abs/1705.03907)
36. Krieger, J., Nakanishi, K., Schlag, W.: Global dynamics away from the ground state for the energy-critical nonlinear wave equation. *Am. J. Math.* **135**(4), 935–965 (2013)
37. Krieger, J., Nakanishi, K., Schlag, W.: Center-stable manifold of the ground state in the energy space for the critical wave equation. *Math. Ann.* **361**(1–2), 1–50 (2015)
38. Krieger, J., Schlag, W.: Concentration Compactness for Critical Wave Maps. EMS Monographs. European Mathematical Society, Zürich (2012)
39. Krieger, J., Schlag, W., Tataru, D.: Renormalization and blow up for charge one equivariant wave critical wave maps. *Invent. Math.* **171**(3), 543–615 (2008)
40. Lawrie, A., Oh, S.-J.: A refined threshold theorem for  $(1+2)$ -dimensional wave maps into surfaces. *Commun. Math. Phys.* **342**(3), 989–999 (2016)
41. Martel, Y., Merle, F.: Description of two soliton collision for the quartic gKdV equation. *Ann. Math. (2)* **174**(2), 757–857 (2011)
42. Martel, Y., Merle, F.: Inelastic interaction of nearly equal solitons for the quartic gKdV equation. *Invent. Math.* **183**(3), 563–648 (2011)
43. Planchon, F., Stalker, J.G., Tahvildar-Zadeh, A.S.:  $L^p$  estimates for the wave equation with the inverse-square potential. *Discrete Cont. Dyn. Syst.* **9**(2), 427–442 (2003)
44. Raphaël, P., Rodnianski, I.: Stable blow up dynamics for the critical co-rotational wave maps and equivariant Yang-Mills problems. *Publ. Math. Inst. Hautes Études Sci.* **115**, 1–122 (2012)
45. Raphaël, P., Szeftel, J.: Existence and uniqueness of minimal mass blow up solutions to an inhomogeneous  $L^2$ -critical NLS. *J. Am. Math. Soc.* **24**(2), 471–546 (2011)
46. Rodnianski, I., Sterbenz, J.: On the formation of singularities in the critical  $O(3)$   $\sigma$ -model. *Ann. Math.* **172**, 187–242 (2010)
47. Sacks, J., Uhlenbeck, K.: The existence of minimal immersions of 2-spheres. *Ann. Math. (2)* **113**(1), 1–24 (1981)

48. Shatah, J., Tahvildar-Zadeh, A.: Regularity of harmonic maps from the Minkowski space into rotationally symmetric manifolds. *Commun. Pure Appl. Math.* **45**(8), 947–971 (1992)
49. Shatah, J., Tahvildar-Zadeh, A.S.: On the Cauchy problem for equivariant wave maps. *Commun. Pure Appl. Math.* **47**(5), 719–754 (1994)
50. Sterbenz, J., Tataru, D.: Energy dispersed large data wave maps in  $2 + 1$  dimensions. *Commun. Math. Phys.* **1**, 139–230 (2010)
51. Sterbenz, J., Tataru, D.: Regularity of wave maps in  $2 + 1$  dimensions. *Commun. Math. Phys.* **1**, 231–264 (2010)
52. Struwe, M.: Equivariant wave maps in two space dimensions. *Commun. Pure Appl. Math.* **56**(7), 815–823 (2003)
53. Tao, T.: Global regularity of wave maps. I. Small critical Sobolev norm in high dimension. *Int. Math. Res. Not.* **6**, 299–328 (2001)
54. Tao, T.: Global regularity of wave maps II: small energy in two dimensions. *Commun. Math. Phys.* **224**(2), 443–544 (2001)
55. Tao, T.: Global regularity of wave maps III–VII (2008–2009). [arXiv:0805.4666](#), [arXiv:0806.3592](#), [arXiv:0808.0368](#), [arXiv:0906.2833](#), [arXiv:0908.0776](#)
56. Tataru, D.: Local and global results for wave maps. I. *Commun. Partial Differ. Equ.* **23**(9–10), 1781–1793 (1998)
57. Tataru, D.: On global existence and scattering for the wave maps equation. *Am. J. Math.* **123**(1), 37–77 (2001)