Forum of Mathematics, Pi (2021), **Vol. 9**:e6 1–37
doi:10.1017/fmp.2021.6

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

On the derivation of the wave kinetic equation for NLS

Yu Deng [1] and Zaher Hani [2]

[1]Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA; E-mail: yudeng@usc.edu.
[2]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA; E-mail: zhani@umich.edu.

Received: 22 February 2021; **Revised:** 19 May 2021; **Accepted:** 9 June 2021

2020 Mathematics Subject Classification: *Primary* – 35Q55; *Secondary* – 35Q82, 82C03

Abstract

A fundamental question in wave turbulence theory is to understand how the wave kinetic equation describes the long-time dynamics of its associated nonlinear dispersive equation. Formal derivations in the physics literature, dating back to the work of Peierls in 1928, suggest that such a kinetic description should hold (for well-prepared random data) at a large *kinetic time scale* $T_{\text{kin}} \gg 1$ and in a limiting regime where the size L of the domain goes to infinity and the strength α of the nonlinearity goes to 0 (weak nonlinearity). For the cubic nonlinear Schrödinger equation, $T_{\text{kin}} = O\left(\alpha^{-2}\right)$ and α is related to the conserved mass λ of the solution via $\alpha = \lambda^2 L^{-d}$.

In this paper, we study the rigorous justification of this monumental statement and show that the answer seems to depend on the particular *scaling law* in which the (α, L) limit is taken, in a spirit similar to how the Boltzmann–Grad scaling law is imposed in the derivation of Boltzmann's equation. In particular, there appear to be *two* favourable scaling laws: when α approaches 0 like $L^{-\varepsilon+}$ or like $L^{-1-\frac{\varepsilon}{2}+}$ (for arbitrary small ε), we exhibit the wave kinetic equation up to time scales $O(T_{\text{kin}} L^{-\varepsilon})$, by showing that the relevant Feynman-diagram expansions converge absolutely (as a sum over paired trees). For the other scaling laws, we justify the onset of the kinetic description at time scales $T_* \ll T_{\text{kin}}$ and identify specific interactions that become very large for times beyond T_*. In particular, the relevant tree expansion diverges absolutely there. In light of those interactions, extending the kinetic description beyond T_* toward T_{kin} for such scaling laws seems to require new methods and ideas.

Contents

1 Introduction **2**
1.1 Statement of the results . . . 4
1.2 Ideas of the proof . . . 6
1.3 Organisation of the paper . . . 11
1.4 Notation . . . 11
2 Tree expansions and long-time existence **12**
2.1 First reductions . . . 12
2.2 The tree expansion . . . 13
2.3 Statement of main estimates . . . 15
3 Proof of main estimates **16**
3.1 Large deviation and basic counting estimates . . . 17
3.2 Bounds for $\mathcal{J}_n$ . . . 20
3.3 Bounds for $\mathcal{P}_{\pm}$ . . . 25
3.4 The worst terms . . . 29
4 Proof of the main theorem **30**
5 Number-theoretic results **32**

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

1. Introduction

The *kinetic framework* is a general paradigm that aims to extend Boltzmann's kinetic theory for dilute gases to other types of microscopic interacting systems. This approach has been highly informative, and became a cornerstone of the theory of nonequilibrium statistical mechanics for a large body of systems [43, 44]. In the context of nonlinear dispersive waves, this framework was initiated in the first half of the past century [41] and developed into what is now called *wave turbulence theory* [51, 39]. There, waves of different frequencies interact nonlinearly at the microscopic level, and the goal is to extract an effective macroscopic picture of how the energy densities of the system evolve.

The description of such an effective evolution comes via the *wave kinetic equation* (WKE), which is the analogue of Boltzmann's equation for nonlinear wave systems [46]. Such kinetic equations have been derived at a formal level for many systems of physical interest (nonlinear Schrödinger (NLS) and nonlinear wave (NLW) equations, water waves, plasma models, lattice crystal dynamics, etc.; compare [39] for a textbook treatment) and are used extensively in applications (thermal conductivity in crystals [45], ocean forecasting [31, 49], and more). This kinetic description is conjectured to appear in the limit where the number of (locally interacting) waves goes to infinity and an appropriate measure of the interaction strength goes to zero (weak nonlinearity[1]). In such kinetic limits, the total energy of the whole system often diverges.

The fundamental mathematical question here, which also has direct consequences for the physical theory, is to provide a rigorous justification of such wave kinetic equations starting from the microscopic dynamics given by the nonlinear dispersive model at hand. The importance of such an endeavour stems from the fact that it allows an understanding of the exact regimes and the limitations of the kinetic theory, which has long been a matter of scientific interest (see [20, 1]). A few mathematical investigations have recently been devoted to studying problems in this spirit [23, 7, 35], yielding some partial results and useful insights.

This manuscript continues the investigation initiated in [7], aimed at providing a rigorous justification of the wave kinetic equation corresponding to the nonlinear Schrödinger equation,

$$i\partial_t v - \Delta v + |v|^2 v = 0.$$

As we shall explain later, the sign of the nonlinearity has no effect on the kinetic description, so we choose the defocussing sign for concreteness. The natural setup for the problem is to start with a spatial domain given by a torus $\mathbb{T}^d_L$ of size L, which approaches infinity in the thermodynamic limit we seek. This torus can be rational or irrational, which amounts to rescaling the Laplacian into

$$\Delta_\beta := \frac{1}{2\pi}\sum_{i=1}^{d}\beta_i\partial_i^2, \qquad \beta := (\beta_1,\dots,\beta_d) \in [1,2]^d,$$

and taking the spatial domain to be the standard torus of size L, namely $\mathbb{T}^d_L = [0,L]^d$ with periodic boundary conditions. With this normalisation, an irrational torus would correspond to taking the β_j to be rationally independent. Our results cover both cases, and in part of them β is assumed to be generic – that is, avoiding a set of Lebesgue measure 0.

The strength of the nonlinearity is related to the characteristic size λ of the initial data (say in the conserved L^2 space). Adopting the ansatz $v = \lambda u$, we arrive at the following equation:

$$\begin{cases} i\partial_t u - \Delta_\beta u + \lambda^2 |u|^2 u = 0, \quad x \in \mathbb{T}^d_L = [0,L]^d, \\ u(0,x) = u_{\text{in}}(x). \end{cases} \tag{NLS}$$

The kinetic description of the long-time behaviour is akin to a law of large numbers, and therefore one has to start with a random distribution of the initial data. Heuristically, a randomly distributed, L^2-normalised field would (with high probability) have a roughly uniform spatial distribution, and

[1] It is for this reason that the theory is sometimes called *weak turbulence theory*.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

consequently an L^∞_x norm $\sim L^{-d/2}$. This makes the strength of the nonlinearity in (NLS) comparable to $\lambda^2 L^{-d}$ (at least initially[2]), which motivates us to introduce the quantity

$$\alpha = \lambda^2 L^{-d}$$

and phrase the results in terms of α instead of λ. The *kinetic conjecture* states that at sufficiently long time scales, the effective dynamics of the Fourier-space mass density $\mathbb{E}\,|\widehat{u}(t,k)|^2$ $\left(k \in \mathbb{Z}^d_L = L^{-1}\mathbb{Z}^d\right)$ is well approximated – in the limit of large L and vanishing α – by an appropriately scaled solution $n(t,\xi)$ of the following WKE:

$$\partial_t n(t,\xi) = \mathcal{K}\left(n(t,\cdot)\right),$$
$$\mathcal{K}(\phi)(\xi) := \int_{\substack{(\xi_1,\xi_2,\xi_3)\in\mathbb{R}^{3d}\\ \xi_1-\xi_2+\xi_3=\xi}} \phi\phi_1\phi_2\phi_3\left(\frac{1}{\phi}-\frac{1}{\phi_1}+\frac{1}{\phi_2}-\frac{1}{\phi_3}\right)\delta_{\mathbb{R}}\left(|\xi_1|^2_\beta - |\xi_2|^2_\beta + |\xi_3|^2_\beta - |\xi|^2_\beta\right) d\xi_1 d\xi_2 d\xi_3, \tag{WKE}$$

where we used the shorthand notations $\phi_j := \phi\left(\xi_j\right)$ and $|\xi|^2_\beta = \sum_{j=1}^d \beta_j \left(\xi^{(j)}\right)^2$ for $\xi = \left(\xi^{(1)},\cdots,\xi^{(d)}\right)$. More precisely, one expects this approximation to hold at the kinetic timescale $T_{\text{kin}} \sim \alpha^{-2} = \frac{L^{2d}}{\lambda^4}$, in the sense that

$$\mathbb{E}\,|\widehat{u}(t,k)|^2 \approx n\left(\frac{t}{T_{\text{kin}}},k\right) \quad \text{as } L\to\infty, \alpha\to 0. \tag{1.1}$$

Of course, for such an approximation to hold at time $t = 0$, one has to start with a *well-prepared* initial distribution for $\widehat{u}_{\text{in}}(k)$ as follows: denoting by n_{in} the initial data for (WKE), we assume

$$\widehat{u}_{\text{in}}(k) = \sqrt{n_{\text{in}}(k)}\eta_k(\omega), \tag{1.2}$$

where $\eta_k(\omega)$ are mean-0 complex random variables satisfying $\mathbb{E}\,|\eta_k|^2 = 1$. In what follows, $\eta_k(\omega)$ will be independent, identically distributed complex random variables, such that the law of each η_k is either the normalised complex Gaussian or the uniform distribution on the unit circle $|z| = 1$.

Before stating our results, it is worth remarking on the regime of data and solutions covered by this kinetic picture in comparison to previously studied and well-understood regimes in the nonlinear dispersive literature. For this, let us look back at the (pre-ansatz) NLS solution v, whose conserved energy is given by

$$\mathcal{E}[v] := \int_{\mathbb{T}^d_L} \frac{1}{2}|\nabla v|^2 + \frac{1}{4}|v|^4\,\mathrm{d}x.$$

We are dealing with solutions having an L^∞-norm of $O\left(\sqrt{\alpha}\right)$ (with high probability) and whose total mass is $O\left(\alpha L^d\right)$, in a regime where α is vanishingly small and L is asymptotically large. These bounds on the solutions are true initially, as we have already explained, and will be propagated in our proof. In particular, the mass and energy are very large and will diverge in this kinetic limit, as is common in taking thermodynamic limits [42, 37]. Moreover, the potential part of the energy is dominated by the kinetic part – the former of size $O\left(\alpha^3 L^d\right)$ and the latter of size $O\left(\alpha L^d\right)$ – which explains why there is no distinction between the defocussing and focussing nonlinearities in the kinetic limit. It would be interesting to see how the kinetic framework can be extended to regimes of solutions which are sensitive to the sign of the nonlinearity; this has been investigated in the physics literature (e.g., [22, 25, 50]).

[2]Formal derivations of the wave kinetic equation often involve heuristic arguments (like a propagation of quasi-Gaussianity of the initial data through time) which effectively imply that the strength of the nonlinearity stays $\sim \lambda^2 L^{-d}$. Such heuristic arguments are hard to justify rigorously; however, this bound on the nonlinearity strength will be propagated and proved as a consequence of our estimates.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

1.1. Statement of the results

It is not *a priori* clear how the limits $L \to \infty$ and $\alpha \to 0$ need to be taken for formula (1.1) to hold or whether there is an additional scaling law (between α and L) that needs to be satisfied in the limit. In comparison, such scaling laws are imposed in the rigorous derivation of Boltzmann's equation [34, 10, 26], which is derived in the so-called Boltzmann–Grad limit [27]: namely, the number N of particles goes to ∞ while their radius r goes to 0 in such a way that $Nr^{d-1} \sim O(1)$. To the best of our knowledge, this central point has not been adequately addressed in the wave-turbulence literature.

Our results seem to suggest some key differences depending on the chosen scaling law. Roughly speaking, we identify *two special scaling laws* for which we are able to justify the approximation (1.1) up to time scales $L^{-\varepsilon}T_{\text{kin}}$ for any arbitrarily small $\varepsilon > 0$. For other scaling laws, we identify significant absolute divergences in the power-series expansion for $\mathbb{E}\,|\widehat{u}(t,k)|^2$ at much earlier times. We can therefore only justify this approximation at such shorter times (which are still better than those in [7]). In these cases, whether or not formula (1.1) holds up to time scales $L^{-\varepsilon}T_{\text{kin}}$ depends on whether such series converge conditionally instead of absolutely, and thus would require new methods and ideas, as we explain later.

We start by identifying the two favourable scaling laws. We use the notation $\sigma+$ for any numerical constant σ (e.g., $\sigma = -\varepsilon$ or $\sigma = -1 - \frac{\varepsilon}{2}$, where ε is as in Theorem 1.1) to denote a constant that is strictly larger than and sufficiently close to σ.

Theorem 1.1. *Set $d \geq 2$ and let $\beta \in [1,2]^d$ be arbitrary. Suppose that $n_{\text{in}} \in \mathcal{S}\big(\mathbb{R}^d \to [0,\infty)\big)$ is Schwartz[3] and $\eta_k(\omega)$ are independent, identically distributed complex random variables, such that the law of each η_k is either complex Gaussian with mean 0 and variance 1 or the uniform distribution on the unit circle $|z| = 1$. Assume well-prepared initial data u_{in} for (NLS) as in equation (1.2).*

Fix $0 < \varepsilon < 1$ (in most interesting cases ε will be small); recall that λ and L are the parameters in (NLS) and let $\alpha = \lambda^2 L^{-d}$ be the characteristic strength of the nonlinearity. If α has the scaling law $\alpha \sim L^{(-\varepsilon)+}$ or $\alpha \sim L^{\left(-1-\frac{\varepsilon}{2}\right)+}$, then we have

$$\mathbb{E}\,|\widehat{u}(t,k)|^2 = n_{\text{in}}(k) + \frac{t}{T_{\text{kin}}}\mathcal{K}(n_{\text{in}})(k) + o_{\ell^\infty_k}\left(\frac{t}{T_{\text{kin}}}\right)_{L\to\infty} \tag{1.3}$$

for all $L^{0+} \leq t \leq L^{-\varepsilon}T_{\text{kin}}$, where $T_{\text{kin}} = \alpha^{-2}/2$, $\mathcal{K}$ is defined in (WKE) and $o_{\ell^\infty_k}\left(\frac{t}{T_{\text{kin}}}\right)_{L\to\infty}$ is a quantity that is bounded in ℓ^∞_k by $L^{-\theta}\frac{t}{T_{\text{kin}}}$ for some $\theta > 0$.

We remark that in the time interval of the approximation we have been discussing, the right hand sides of formulas (1.1) and (1.3) are equivalent. Also note that any type of scaling law of the form $\alpha \sim L^{-s}$ gives an upper bound of $t \leq L^{-\varepsilon}T_{\text{kin}} \sim L^{2s-\varepsilon}$ for the times considered. Consequently, for the two scaling laws in Theorem 1.1, the time t always satisfies $t \ll L^2$, and it is for this reason that the rationality type of the torus is not relevant. As will be clear later, no similar results can hold for $t \gg L^2$ in the case of a rational torus,[4] as this would require rational quadratic forms to be equidistributed on scales $\ll 1$, which is impossible. However, if the aspect ratios β are assumed to be generically irrational, then one can access equidistribution scales that are as small as L^{-d+1} for the resulting irrational quadratic forms [4, 7]. This allows us to consider scaling laws for which T_{kin} can be as big as L^{d-} on generically irrational tori.

Remark 1.2. Strictly speaking, in evaluating equation (1.3) one has to first ensure the existence of the solution u. This is guaranteed if $d \in \{2,3,4\}$ (when (NLS) is H^1-critical or subcritical). When $d \geq 5$ we shall interpret equation (1.3) such that the expectation is taken only when the long-time smooth solution u exists. Moreover, from our proof it follows that the probability that this existence fails is at most $O\left(e^{-L^\theta}\right)$, which quickly becomes negligible when $L \to \infty$.

[3] In fact, only a finite amount of decay and smoothness is needed on n_{in}. We chose $n_{\text{in}} \in \mathcal{S}$ to simplify the exposition and avoid minor distracting technicalities.

[4] Even at the endpoint case where $t \sim L^2$, the number-theoretic components of the proof would yield different answers to those anticipated by the theory; see Section 5.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

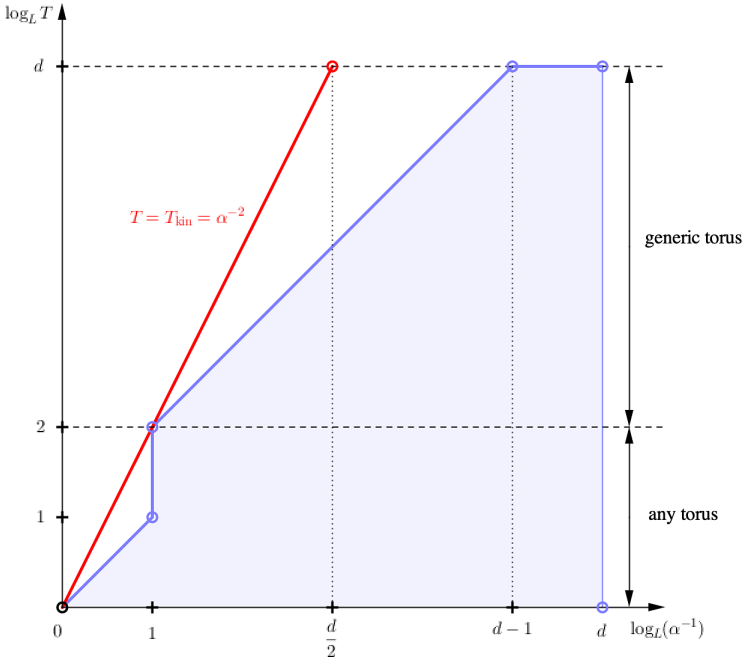

Figure 1. *Admissible range for (α, L, T) in the $\left(\log_L\left(\alpha^{-1}\right), \log_L T\right)$ plot when $d \geq 3$. The coloured region is the range of Theorem 1.3 (up to ε endpoint accuracy). The red line denotes the case when $T = T_{\text{kin}} = \alpha^{-2}$, which our coloured region touches at two points corresponding to $T \sim 1$ and $T \sim L^2$.*

The following theorem covers general scaling laws, including the ones that can only be accessed for the generically irrational torus. By a simple calculation of exponents, we can see that it implies Theorem 1.1.

Theorem 1.3. *With the same assumptions as in Theorem 1.1, we impose the following conditions on (α, L, T) for some $\delta > 0$:*

$$T \leq \begin{cases} L^{2-\delta} & \text{if } \beta_i \text{ is arbitrary,} \\ L^{d-\delta} & \text{if } \beta_i \text{ is generic,} \end{cases} \qquad \alpha \leq \begin{cases} L^{-\delta} T^{-1} & \text{if } T \leq L, \\ L^{-1-\delta} & \text{if } L \leq T \leq L^2, \\ L^{1-\delta} T^{-1} & \text{if } T \geq L^2. \end{cases} \tag{1.4}$$

Then formula (1.3) holds for all $L^\delta \leq t \leq T$.

It is best to read this theorem in terms of the $\left(\log_L\left(\alpha^{-1}\right), \log_L T\right)$ plot in Figure 1. The kinetic conjecture corresponds to justifying the approximation in formula (1.1) up to time scales $T \lesssim T_{\text{kin}} = \alpha^{-2}$. As we shall explain later, the time scale $T \sim T_{\text{kin}}$ represents a *critical scale* for the problem from a probabilistic point of view. This is depicted by the red line in the figure, and the region below this line corresponds to a *probabilistically subcritical* regime (see Section 1.2.1). The shaded blue region corresponds to the (α, T) region in Theorem 1.3, neglecting δ losses. This region touches the line $T = \alpha^{-2}$ at the two points corresponding to $\left(\alpha^{-1}, T\right) = (1, 1)$ and $\left(L, L^2\right)$, whereas the two scaling laws of Theorem 1.1, where $\left(\alpha^{-1}, T\right) \sim \left(L^{\varepsilon-}, L^{\varepsilon-}\right)$ and $\left(\alpha^{-1}, T\right) \sim \left(L^{1+\frac{\varepsilon}{2}-}, L^{2-}\right)$, approach these two points when ε is small.

These results rely on a diagrammatic expansion of the NLS solution in Feynman diagrams akin to a Taylor expansion. The shaded blue region depicting the result of Theorem 1.3 corresponds to the cases when such a diagrammatic expansion is absolutely convergent for very large L. In the complementary

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

region between the blue region and the line $T = T_{\text{kin}}$, we show that some (arbitrarily high-degree) terms of this expansion do not converge to 0 as their degree goes to ∞, which means that the diagrammatic expansion cannot converge absolutely in this region. Therefore, the only way for the kinetic conjecture to be true in the scaling regimes not included in Theorem 1.1 is for those terms to exhibit a highly nontrivial cancellation, which would make the series converge conditionally but not absolutely.

Finally, we remark on the restriction in formula (1.4). The upper bounds on T on the left are necessary from number-theoretic considerations: indeed, if $T \gg L^2$ for a rational torus, or if $T \gg L^d$ for an irrational one, the exact resonances of the NLS equation dominate the quasi-resonant interactions that lead to the kinetic wave equation. One should therefore not expect the kinetic description to hold in those ranges of T (see Lemma 3.2 and Section 4). The second set of restrictions in formula (1.4) correspond exactly to the requirement that the size of the Feynman diagrams of degree n can be bounded by ρ^n with some $\rho \ll 1$. In fact, if one aims only at proving existence with high probability (not caring about the asymptotics of $\mathbb{E}\,|\widehat{u}(t,k)|^2$), then the restrictions on the left of formula (1.4) will not be necessary, and one obtains control for longer times. See also the following remark:

Remark 1.4 (Admissible scaling laws). The foregoing restrictions on T impose the limits of the admissible scaling laws, in which $\alpha \to 0$ and $L \to \infty$, for which the kinetic description of the long-time dynamics can appear. Indeed, since $T_{\text{kin}} = \alpha^{-2}$, then the *necessary* (up to L^{δ} factors) restrictions $T \ll L^{2-\delta}$ (resp., $T \ll L^{d-\delta}$) on the rational (resp., irrational) torus already mentioned imply that one should only expect the previous kinetic description in the regime where $\alpha \gtrsim L^{-1}$ (resp., $\gtrsim L^{-d/2}$). In other words, the kinetic description requires the nonlinearity to be weak, but not too weak! In the complementary regime of very weak nonlinearity, the exact resonances of the equation dominate the quasi-resonances – a regime referred to as *discrete wave turbulence* (see [36, 32, 39]), in which different effective equations, like the (CR) equation in [24, 6], can arise.

1.2. Ideas of the proof

As Theorem 1.1 is a consequence of Theorem 1.3, we will focus on Theorem 1.3. The proof of Theorem 1.3 contains three components: (1) a long-time well-posedness result, where we expand the solution to (NLS) into Feynman diagrams for sufficiently long time, up to a well-controlled error term; (2) computation of $\mathbb{E}\,|\widehat{u}_k(t)|^2$ $\left(k \in \mathbb{Z}^d_L\right)$ using this expansion, where we identify the leading terms and control the remainders; and (3) a number-theoretic result that justifies the large box approximation, where we pass from the sums appearing in the expansion in the previous component to the integral appearing on the right-hand side of (WKE).

The main novelty of this work is in the first component, which is the hardest. The second component follows similar lines to those in [7]. Regarding the third component, the main novelty of this work is to complement the number-theoretic results in [7] (which dealt only with the generically irrational torus) by the cases of general tori (in the admissible range of time $T \ll L^2$). This provides an essentially full (up to L^{ε} losses) understanding of the number-theoretic issues arising in wave-turbulence derivations for (NLS). Therefore, we will limit this introductory discussion to the first component.

1.2.1. The scheme and probabilistic criticality

Though technically involved, the basic idea of the long-time well-posedness argument is in fact quite simple. Starting from (NLS) with initial data of equation (1.2), we write the solution as

$$u = u^{(0)} + \cdots + u^{(N)} + \mathcal{R}_{N+1}, \tag{1.5}$$

where $u^{(0)} = e^{-it\Delta_\beta} u_{\text{in}}$ is the linear evolution, $u^{(n)}$ are iterated self-interactions of the linear solution $u^{(0)}$ that appear in a formal expansion of u and $\mathcal{R}_{N+1}$ is a sufficiently regular remainder term.

Since $u^{(0)}$ is a linear combination of independent random variables, and each $u^{(n)}$ is a multilinear combination, each of them will behave strictly better (both linearly and nonlinearly) than its deterministic analogue (i.e., with all $\eta_k = 1$). This is due to the well-known large deviation estimates, which yield a

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

'square root' gain coming from randomness, akin to the central limit theorem (for instance, $\|u_{\text{in}}\|_{L^\infty}$ is bounded by $L^{-d/2} \cdot \|u_{\text{in}}\|_{L^2}$ in the probabilistic setting, as opposed to $1 \cdot \|u_{\text{in}}\|_{L^2}$ deterministically by Sobolev embedding, assuming compact Fourier support). This gain leads to a new notion of criticality for the problem, which can be defined[5] as the edge of the regime of (α, T) for which the iterate $u^{(1)}$ is better bounded than the iterate $u^{(0)}$. It is not hard to see that $u^{(1)}$ can have size up to $O(\alpha\sqrt{T})$ (in appropriate norms), compared to the $O(1)$ size of $u^{(0)}$ (see, e.g., formula (2.25) for $n = 1$). This justifies the notion that $T \sim T_{\text{kin}} = \alpha^{-2}$ corresponds to *probabilistically critical scaling*, whereas the time scales $T \ll T_{\text{kin}}$ are *subcritical*.[6]

As it happens, a certain notion of criticality might not capture all the subtleties of the problem. As we shall see, some higher-order iterates $u^{(n)}$ will not be better bounded than $u^{(n-1)}$ in the full subcritical range $T \ll \alpha^{-2}$ we have postulated, but instead only in a subregion thereof. This is what defines our admissible blue region in Figure 1.

We should mention that the idea of using the gain from randomness goes back to Bourgain [3] (in the random-data setting) and to Da Prato and Debussche [14] (later, in the stochastic PDE setting). They first noticed that the ansatz $u = u^{(0)} + \mathcal{R}$ allows one to put the remainder $\mathcal{R}$ in a higher regularity space than the linear term $u^{(0)}$. This idea has since been applied to many different situations (see, e.g., [5, 8, 11, 15, 21, 33, 38]), though most of these works either involve only the first-order expansion (i.e., $N = 0$) or involve higher-order expansions with only suboptimal bounds (e.g., [2]). To the best of our knowledge, the present paper is the first work where the sharp bounds for these $u^{(j)}$ terms are obtained to arbitrarily high order (at least in the dispersive setting).

Remark 1.5. There are two main reasons why the high-order expansion (1.5) gives the sharp time of control, in contrast to previous works. The first is that we are able to obtain sharp estimates for the terms $u^{(j)}$ with arbitrarily high order, which were not known previously due to the combinatorial complexity associated with trees (see Section 1.2.2).

The second reason is more intrinsic. In higher-order versions of the original Bourgain–Da Prato–Debussche approach, it usually stops improving in regularity beyond a certain point, due to the presence of the *high-low* interactions (heuristically, the gain of powers of low frequency does not transform to the gain in regularity). This is a major difficulty in random-data theory, and in recent years a few methods have been developed to address it, including *regularity structure* [29], *para-controlled calculus* [28] and *random averaging operators* [18]. Fortunately, in the current problem this issue is absent, since the well-prepared initial data (1.2) bound the high-frequency components (where $|k| \sim 1$) and low-frequency components (where $|k| \sim L^{-1}$) uniformly, so the high-low interaction is simply controlled in the same way as the high-high interaction, allowing one to gain regularity indefinitely as the order increases.

1.2.2. Sharp estimates of Feynman trees

We start with the estimate for $u^{(n)}$. As is standard with the cubic nonlinear Schrödinger equation, we first perform the Wick ordering by defining

$$w := e^{-2i\lambda^2 M_0 t} \cdot u, \qquad M_0 := \fint_{\mathbb{T}^d_L} |u|^2 .$$

Note that M_0 is essentially the mass which is conserved. Now w satisfies the renormalised equation

$$i\partial_t w - \Delta_\beta w + \lambda^2 \left(|w|^2 - 2 \fint_{\mathbb{T}^d_L} |w|^2 \right) w = 0, \tag{1.6}$$

[5] One can interpret the usual scaling criticality for (NLS) in the same way: it corresponds to the minimum regularity s for which the first iterate of an H^s-normalised rescaled bump function like $N^{-s+\frac{d}{2}} \phi(Nx)$ is better bounded than the linear solution $\left(\text{comparing } |u|^p u \text{ to } \Delta u \text{ for such data gives the critical regularity } s_{\text{critical}} = \frac{d}{2} - \frac{2}{p}\right)$.

[6] It may be supercritical under *deterministic scaling*. See [18] for a discussion of these notions in the more customary context of Sobolev regularity of local well-posedness in deterministic versus probabilistic settings.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

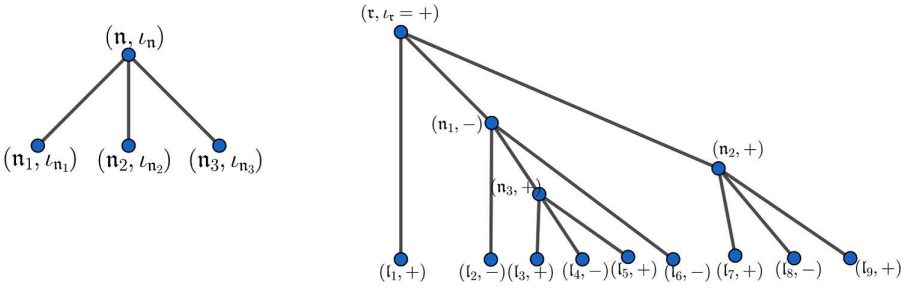

Figure 2. *On the left, a node* $\mathfrak{n}$ *with its three children* $\mathfrak{n}_1, \mathfrak{n}_2, \mathfrak{n}_3$*, with signs* $\iota_1 = \iota_3 = \iota = -\iota_2$*. On the right, a tree of scale* 4 ($\mathfrak{s}(\mathcal{T}) = 4$) *with root* $\mathfrak{r}$*, four branching nodes* ($\mathfrak{r}, \mathfrak{n}_1, \mathfrak{n}_2, \mathfrak{n}_3$) *and* $l = 9$ *leaves, along with their signatures.*

and $|\widehat{w_k}(t)|^2 = |\widehat{u_k}(t)|^2$. This gets rid of the worst resonant term, which would otherwise lead to a suboptimal time scale.

Let $w^{(n)}$ be the nth-order iteration of the nonlinearity in equation (1.6), corresponding to the $u^{(n)}$ in equation (1.5). Since this nonlinearity is cubic, by induction it is easy to see that $w^{(n)}$ can be written (say, in Fourier space) as a linear combination of terms[7] $\mathcal{J}_{\mathcal{T}}$, where $\mathcal{T}$ runs over all ternary trees with exactly n branching nodes (we will say it has *scale* $\mathfrak{s}(\mathcal{T}) = n$). After some further reductions, the estimate for $\mathcal{J}_{\mathcal{T}}$ can be reduced to the estimate for terms of the form

$$\Sigma_k := \sum_{(k_1,\dots,k_{2n+1})\in S} \eta_{k_1}^{\pm} \cdots \eta_{k_{2n+1}}^{\pm}, \qquad \left(\eta_k^+, \eta_k^-\right) := \left(\eta_k(\omega), \overline{\eta_k}(\omega)\right), \tag{1.7}$$

where $\eta_k(\omega)$ is as in equation (1.2), $(k_1, \dots, k_{2n+1}) \in \left(\mathbb{Z}_L^d\right)^{2n+1}$, S is a suitable finite subset of $\left(\mathbb{Z}_L^d\right)^{2n+1}$ and the $(2n+1)$ subscripts correspond to the $(2n+1)$ leaves of $\mathcal{T}$ (see Definition 2.2 and Figure 2).[8]

To estimate Σ_k defined in formaul (1.7) we invoke the standard large deviation estimate (see Lemma 3.1), which essentially asserts that $|\Sigma_k| \lesssim (\#S)^{1/2}$ with overwhelming probability, provided that there is no *pairing* in $(k_1, \dots, k_{2n+1})$, where a pairing $\left(k_i, k_j\right)$ means $k_i = k_j$ and the signs of η_{k_i} and η_{k_j} in formula (1.7) are opposites. Moreover, in the case of a pairing $\left(k_i, k_j\right)$ we can essentially replace $\eta_{k_i}^{\pm}\eta_{k_j}^{\pm} = \left|\eta_{k_i}\right|^2 \approx 1$, so in general we can bound, with overwhelming probability,

$$|\Sigma_k|^2 \lesssim \sum_{(\text{unpaired } k_i)} \left(\sum_{\substack{(\text{paired } k_i):\\ (k_1,\dots,k_{2n+1})\in S}} 1 \right)^2 \lesssim \sum_{(k_1,\dots,k_{2n+1})\in S} 1 \cdot \sup_{(\text{unpaired } k_i)} \sum_{\substack{(\text{paired } k_i):\\ (k_1,\dots,k_{2n+1})\in S}} 1.$$

It thus suffices to bound the number of choices for $(k_1, \dots, k_{2n+1})$ given the pairings, as well as the number of choices for the paired k_js given the unpaired k_js.

In the no-pairing case, such counting bounds are easy to prove, since the set S is well adapted to the tree structure of $\mathcal{T}$; what makes the counting nontrivial is the pairings, especially those between leaves that are far away or from different levels (see Figure 3, where a pairing is depicted by an extra link between the two leaves). Nevertheless, we have developed a counting algorithm that specifically deals with the given pairing structure of $\mathcal{T}$ and ultimately leads to sharp counting bounds and consequently sharp bounds for Σ_k (see Proposition 3.5).

[7] We will first perform rescaling and conjugation by the linear Schrödinger flow (see Section 2.1); for simplicity we still use $\mathcal{J}_{\mathcal{T}}$ to denote these terms.

[8] In reality one may have coefficients $m = m(k, k_1, \dots, k_{2n+1})$ in the expression of M_k in formula (1.7), but one can always reduce to the form of formula (1.7) by restricting to the level sets of m.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

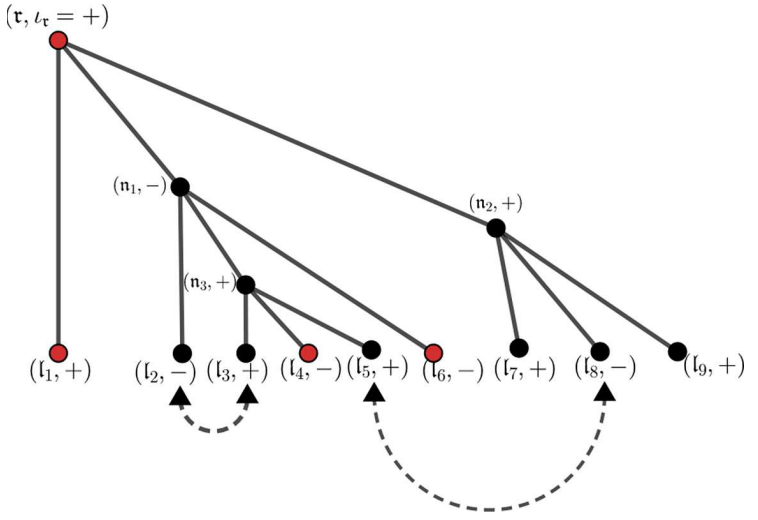

Figure 3. *A paired tree with two pairings* ($p = 2$). *The set* $\mathcal{S}$ *of single leaves is* $\{\mathfrak{l}_1, \mathfrak{l}_4, \mathfrak{l}_6, \mathfrak{l}_7, \mathfrak{l}_9\}$. *The subset* $\mathcal{R} \subset \mathcal{S} \cup \{\mathfrak{r}\}$ *of red-coloured vertices is* $\{\mathfrak{r}, \mathfrak{l}_1, \mathfrak{l}_4, \mathfrak{l}_6\}$. *Here* $(l, p, r) = (9, 2, 4)$. *A strongly admissible assignment with respect to this pairing, colouring and a certain fixed choice of the red modes* $(k_{\mathfrak{r}}, k_{\mathfrak{l}_4}, k_{\mathfrak{l}_6})$ *corresponds to having the modes* $k_{\mathfrak{l}_2} = k_{\mathfrak{l}_3}$, $k_{\mathfrak{l}_5} = k_{\mathfrak{l}_8}$ *and* $|k_{\mathfrak{l}}| \leq L^{\theta}$ *for all the uncoloured leaves. The rest of the modes are determined according to Definition 2.2.*

1.2.3. An ℓ^2 operator norm bound

In contrast to the tree terms $\mathcal{J}_{\mathcal{T}}$, the remainder term $\mathcal{R}_{N+1}$ has no explicit random structure. Indeed, the only way it feels the 'chaos' of the initial data is through the equation it satisfies, which in integral form and spatial Fourier variables looks like

$$\mathcal{R}_{N+1} = \mathcal{J}_{\sim N} + \mathcal{L}(\mathcal{R}_{N+1}) + \mathcal{Q}(\mathcal{R}_{N+1}) + \mathcal{C}(\mathcal{R}_{N+1}),$$

where $\mathcal{J}_{\sim N}$ is a sum of Feynman trees $\mathcal{J}_{\mathcal{T}}$ (already described) of scale $\mathfrak{s}(\mathcal{T}) \sim N$, and $\mathcal{L}$, $\mathcal{Q}$ and $\mathcal{C}$ are, respectively, linear, bilinear and trilinear operator in $\mathcal{R}_{N+1}$. The main point here is that one would like to propagate the estimates on $\mathcal{J}_{\sim N}$ to $\mathcal{R}_{N+1}$ itself; this is how we make rigorous the so-called 'propagation of chaos or quasi-Gaussianity' claims that are often adopted in formal derivations. In another aspect, qualitative results on propagation of quasi-Gaussianity, in the form of absolute continuity of measures, have been obtained in some cases (with different settings) by exploiting almost-conservation laws (e.g., [48]).

Since we are bootstrapping a smallness estimate on $\mathcal{R}_{N+1}$, any quadratic and cubic form of $\mathcal{R}_{N+1}$ will be easily bounded. It therefore suffices to propagate the bound for the term $\mathcal{L}(\mathcal{R}_{N+1})$, which reduces to bounding the $\ell^2 \to \ell^2$ operator norm for the linear operator $\mathcal{L}$. By definition, the operator $\mathcal{L}$ will have the form $v \mapsto \mathcal{I}\mathcal{W}(\mathcal{J}_{\mathcal{T}_1}, \mathcal{J}_{\mathcal{T}_2}, v)$, where $\mathcal{I}$ is the Duhamel operator, $\mathcal{W}$ is the trilinear form coming from the cubic nonlinearity and $\mathcal{J}_{\mathcal{T}_1}, \mathcal{J}_{\mathcal{T}_2}$ are trees of scale $\leq N$; thus in Fourier space it can be viewed as a matrix with *random coefficients*. The key to obtaining the sharp estimate for $\mathcal{L}$ is then to exploit the cancellation coming from this randomness, and the most efficient way to do this is via the TT^* method.

In fact, the idea of applying the TT^* method to random matrices has already been used by Bourgain [3]. In that paper one is still far above (probabilistic) criticality, so applying the TT^* method once already gives adequate control. In the present case, however, we are aiming at obtaining sharp estimates, so applying TT^* once will not be sufficient.

The solution is thus to apply TT^* sufficiently many times (say, $D \gg 1$), which leads to the analysis of the kernel of the operator $(\mathcal{L}\mathcal{L}^*)^D$. At first sight this kernel seems to be a complicated multilinear expression which is difficult to handle; nevertheless, we make one key observation, namely that this

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

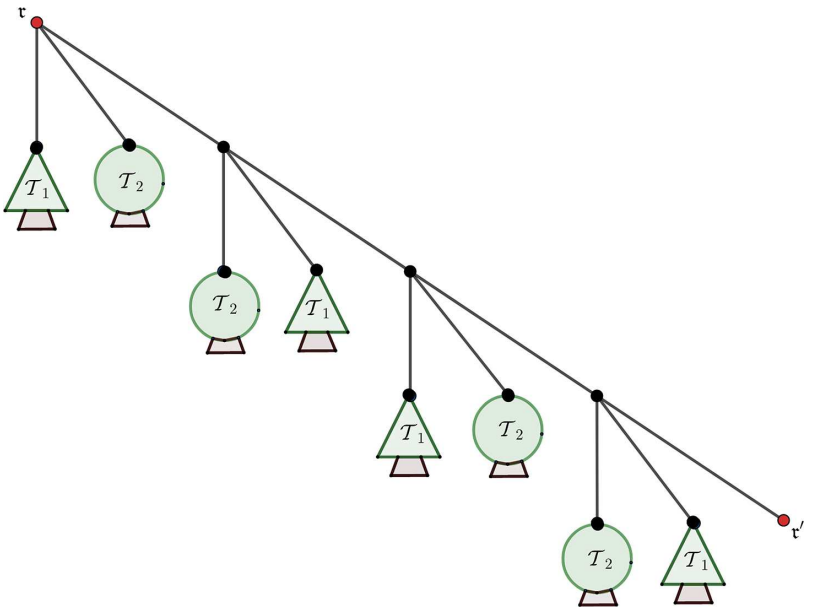

Figure 4. *Construction of the tree $\mathcal{T}^D$ by successive plantings of trees $\mathcal{T}_1$ and $\mathcal{T}_2$ onto the first two nodes of a ternary tree, starting with a root $\mathfrak{r}$ and stopping after $2D$ steps, leaving a leaf node $\mathfrak{r}'$. In the figure, $D = 2$.*

kernel can essentially be recast in the form of formula (1.7) for some large auxiliary tree $\mathcal{T} = \mathcal{T}^D$, which is obtained from a single root node by attaching copies of the trees $\mathcal{T}_1$ and $\mathcal{T}_2$ successively a total of $2D$ times (see Figure 4). With this observation, the arguments in the previous section then lead to sharp bounds of the kernel of $(\mathcal{L}\mathcal{L}^*)^D$, up to some loss that is a power of L independent of D; taking the $1/(2D)$ power and choosing D sufficiently large makes this power negligible and implies the sharp bound for the operator norm of $\mathcal{L}$ (see Section 3.3).

1.2.4. Sharpness of estimates

We remark that the estimates we prove for $\mathcal{J}_{\mathcal{T}}$ are sharp up to some finite power of L (independent of $\mathcal{T}$). More precisely, from Proposition 2.5 we know that for any ternary tree $\mathcal{T}$ of scale n and possible pairing structure (see Definition 3.3), with overwhelming probability,

$$\sup_k \|(\mathcal{J}_{\mathcal{T}})_k\|_{h^b} \leq L^{0+}\rho^n, \tag{1.8}$$

where ρ is some quantity depending on α, L and T (see formula (2.24)), k is the spatial Fourier variable and h^b is a time-Sobolev norm defined in equation (2.22); on the other hand, we will show that that for some particular choice of trees $\mathcal{T}$ of scale n and some particular choice of pairings, with high probability,

$$\sup_k \|(\mathcal{J}_{\mathcal{T}})_k\|_{h^b} \geq L^{-d}\rho^n. \tag{1.9}$$

The timescale T of Theorem 1.3 is the largest that makes $\rho \ll 1$; thus if one wants to go beyond T in cases other than Theorem 1.1, it would be necessary to address the divergence of formula (1.9) with $\rho \gg 1$ by exploiting the cancellation between different tree terms or different pairing choices (see Section 3.4).

1.2.5. Discussions

Shortly after the completion of this paper, work of Collot and Germain [12] was announced that studies the same problem, but only in the rational-torus setting. In the language of this paper, their result

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

corresponds to the validity of equation (1.3) for $L \le t \le L^{2-\delta}$, under the assumption $\alpha \le L^{-1-\delta}$. This is a special case of Theorem 1.3, essentially corresponding to the rectangle below the horizontal line $\log_L T = 2$ and to the right of the vertical line $\log_L(\alpha^{-1}) = 1$ in Figure 1. We also mention later work by the same authors [13], where they consider a *generic nonrectangular* torus (as opposed to the rectangular tori here and in [12]) and prove the existence of solutions (but without justifying equation (1.3)) up to time $t \le L^{-\delta} T_{\mathrm{kin}}$ for a wider range of power laws between α and L.

While the present paper was being peer-reviewed, we submitted new work to arXiv [16], in which we provide the first *full derivation* of (WKE) from (NLS). Those results reach the kinetic time scale $t = \tau \cdot T_{\mathrm{kin}}$, where τ is *independent* of L (compared to Theorem 1.1 here, where $\tau \le L^{-\varepsilon}$), for the scaling law $\alpha \sim L^{-1}$ on generic (irrational) rectangular tori and the scaling laws $\alpha \sim L^{-\gamma}$ (where $\gamma < 1$ and is close to 1) on arbitrary rectangular tori.

Shortly after completing [16], we received a preprint of a forthcoming deep work by Staffilani and Tran [47]. It concerns a high-dimensional (on $\mathbb{T}^d$ for $d \ge 14$) KdV equation under a time-dependent Stratonovich stochastic forcing, which effectively randomises the phases without injecting energy into the system. The authors derive the corresponding wave kinetic equation up to the kinetic time scale, for the scaling law $\alpha \sim L^{-0}$ (i.e., first taking $L \to \infty$ and then taking $\alpha \to 0$). They also prove a conditional result without such forcing, where the condition is verified for some particular initial densities converging to the equilibrium state (stationary solution to the wave kinetic equation) in the limit.

1.3. Organisation of the paper

In Section 2 we explain the diagrammatic expansion of the solution into Feynman trees, and state the *a priori* estimates on such trees and remainder terms, which yield the long-time existence of such expansions. Section 3 is devoted to the proof of those *a priori* estimates. In Section 4 we prove the main theorems already mentioned, and in Section 5 we prove the necessary number-theoretic results that allow us to replace the highly oscillating Riemann sums by integrals.

1.4. Notation

Most notation will be standard. Let $z^+ = z$ and $z^- = \bar{z}$. Define $|k|_\beta$ by $|k|_\beta^2 = \beta_1 k_1^2 + \cdots + \beta_d k_d^2$ for $k = (k_1, \ldots, k_d)$. The spatial Fourier series of a function $u : \mathbb{T}_L^d \to \mathbb{C}$ is defined on $\mathbb{Z}_L^d := L^{-1}\mathbb{Z}^d$ by

$$\widehat{u}_k = \int_{\mathbb{T}_L^d} u(x) e^{-2\pi i k \cdot x}, \quad \text{so that} \quad u(x) = \frac{1}{L^d} \sum_{k \in \mathbb{Z}_L^d} \widehat{u}_k e^{2\pi i k \cdot x}. \tag{1.10}$$

The temporal Fourier transform is defined by

$$\widetilde{f}(\tau) = \int_{\mathbb{R}} e^{-2\pi i t \tau} f(t) \mathrm{d}t.$$

Let $\delta > 0$ be fixed throughout the paper. Let N, s and $b > \frac{1}{2}$ be fixed, such that N and s are large enough and $b - \frac{1}{2}$ is small enough, depending on d and δ. The quantity C will denote any large absolute constant, not dependent on $(N, s, b - \frac{1}{2})$, and θ will denote any small positive constant, which is dependent on $(N, s, b - \frac{1}{2})$; these may change from line to line. The symbols $O(\cdot)$, $\lesssim$ and so on will have their usual meanings, with implicit constants depending on θ. Let L be large enough depending on all these implicit constants. If some statement S involving ω is true with probability $\ge 1 - Ke^{-L^\theta}$ for some constant K (depending on θ), then we say this statement S is L-certain.

When a function depends on many variables, we may use notations like

$$f = f\left(x_i : i \in A,\, y_j : 1 \le j \le m\right)$$

to denote a function f of variables $(x_i : i \in A)$ and $y_1, \ldots, y_m$.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

2. Tree expansions and long-time existence

2.1. First reductions

Let $\widehat{u}_k(t)$ be the Fourier coefficients of $u(t)$, as in equation (1.10). Then with $c_k(t) := e^{2\pi i|k|_\beta^2 t}\widehat{u}_k(t) = \left(\mathcal{F}_{\mathbb{T}^d_L} e^{-it\Delta_\beta} u\right)(k)$, we arrive at the following equation for the Fourier modes:

$$\begin{cases} ic_k^{\cdot} = \left(\frac{\lambda}{L^d}\right)^2 \sum\limits_{\substack{(k_1,k_2,k_3)\in\left(\mathbb{Z}^d_L\right)^3 \\ k-k_1+k_2-k_3=0}} c_{k_1}\overline{c_{k_2}}c_{k_3}e^{2\pi i\Omega(k_1,k_2,k_3,k)t} \\ c_k(0) = (c_k)_{\text{in}} = \widehat{u}_k(0), \end{cases} \tag{2.1}$$

where $\Omega(k_1,k_2,k_3,k) = |k_1|_\beta^2 - |k_2|_\beta^2 + |k_3|_\beta^2 - |k|_\beta^2$. Note that the sum can be written as

$$\sum_{\substack{(k_1,k_2,k_3)\in\left(\mathbb{Z}^d_L\right)^3 \\ k-k_1+k_2-k_3=0}} = 2\sum_{k_1=k} - \sum_{k_1=k_2=k_3} + \sum_{k_1,k_3\neq k},$$

which, defining $M = \sum_{k_3}\left|c_{k_3}\right|^2$ (which is conserved), allows us to write

$$ic_k^{\cdot} = \left(\frac{\lambda}{L^d}\right)^2\left(2Mc_k - |c_k|^2 c_k + \sum_{(k_1,k_2,k_3)}^{\times} c_{k_1}\overline{c_{k_2}}c_{k_3}e^{2\pi i\Omega(k_1,k_2,k_3,k)t}\right).$$

Here and later, $\sum^{\times}$ represents summation under the conditions $k_j\in\mathbb{Z}^d_L$, $k_1-k_2+k_3 = k$ and $k\notin\{k_1,k_3\}$. Introducing $b_k(t) = c_k(t)e^{-2i\left(L^{-d}\lambda\right)^2 Mt}$, we arrive at the following equation for $b_k(t)$:

$$\begin{cases} ib_k^{\cdot} = \left(\frac{\lambda}{L^d}\right)^2\left(-|b_k|^2 b_k + \sum\limits_{(k_1,k_2,k_3)}^{\times} b_{k_1}\overline{b_{k_2}}b_{k_3}e^{2\pi i\Omega(k_1,k_2,k_3,k)t}\right) \\ b_k(0) = (b_k)_{\text{in}} = \widehat{u}_k(0). \end{cases} \tag{2.2}$$

In Theorem 1.3 we will be studying the solution $u(t)$, or equivalently the sequence $(b_k(t))_{k\in\mathbb{Z}^d_L}$, on a time interval $[0,T]$. It will be convenient, to simplify some notation later, to work on the unit time interval $[0,1]$. For this we introduce the final ansatz

$$a_k(t) = b_k(Tt),$$

which satisfies the equation

$$\begin{cases} ia_k^{\cdot} = \left(\frac{\alpha T}{L^d}\right)\left(-|a_k|^2 a_k + \sum\limits_{(k_1,k_2,k_3)}^{\times} a_{k_1}\overline{a_{k_2}}a_{k_3}e^{2\pi iT\Omega(k_1,k_2,k_3,k)t}\right) \\ a_k(0) = (a_k)_{\text{in}} = \widehat{u}_k(0). \end{cases} \tag{2.3}$$

Here we have also used the relation $\alpha = \lambda^2 L^{-d}$. Recall the well-prepared initial data (1.2), which transform into the initial data for a_k:

$$(a_k)_{\text{in}} = \sqrt{n_{\text{in}}}\cdot\eta_k(\omega), \tag{2.4}$$

where $\eta_k(\omega)$ are the same as in equation (1.2).

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

2.2. The tree expansion

Let $\boldsymbol{a}(t) = (a_k(t))_{k\in\mathbb{Z}^d_L}$ and $\boldsymbol{a}_{\text{in}} = \boldsymbol{a}(0)$. Let $J = [0,1]$; we will fix a smooth compactly supported cutoff function χ such that $\chi \equiv 1$ on J. Then by equation (2.3), we know that for $t \in J$ we have

$$\boldsymbol{a}(t) = \chi(t)\boldsymbol{a}_{\text{in}} + \mathcal{I}\mathcal{W}(\boldsymbol{a},\boldsymbol{a},\boldsymbol{a})(t), \tag{2.5}$$

where the Duhamel term is defined by

$$\mathcal{I}F(t) = \chi(t)\int_0^t \chi(t')F(t')\mathrm{d}t', \tag{2.6}$$

$$\mathcal{W}(\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})_k(t) := -\frac{i\alpha T}{L^d}\left(-\left(b_k\overline{c_k}d_k\right)(t) + \sum_{(k_1,k_2,k_3)}^{\times}\left(b_{k_1}\overline{c_{k_2}}d_{k_3}\right)(t)e^{2\pi iT\Omega(k_1,k_2,k_3,k)t}\right). \tag{2.7}$$

Since we will only be studying $\boldsymbol{a}$ for $t \in J$, from now on *we will replace* $\boldsymbol{a}$ *by the solution to equation (2.5) for* $t \in \mathbb{R}$ (the existence and uniqueness of the latter will be clear from a proof to follow). We will be analysing the temporal Fourier transform of this (extended) $\boldsymbol{a}$, so let us first record a formula for $\mathcal{I}$ on the Fourier side:

Lemma 2.1. *Let* $\mathcal{I}$ *be defined as in equation (2.6), and recall that* $\widetilde{G}$ *means the temporal Fourier transform of* G*; then we have*

$$\widetilde{\mathcal{I}F}(\tau) = \int_{\mathbb{R}}(I_0+I_1)(\tau,\sigma)\widetilde{F}(\sigma)\mathrm{d}\sigma, \qquad \left|\partial^a_{\tau,\sigma}I_d(\tau,\sigma)\right| \lesssim_{a,A} \frac{1}{\langle\tau-d\sigma\rangle^A}\frac{1}{\langle\sigma\rangle}. \tag{2.8}$$

Proof. See [17]. □

Now define $\mathcal{J}_n$ recursively by

$$\begin{aligned}\mathcal{J}_0(t) &= \chi(t)\cdot\boldsymbol{a}_{\text{in}},\\ \mathcal{J}_n(t) &= \sum_{n_1+n_2+n_3=n-1}\mathcal{I}\mathcal{W}\left(\mathcal{J}_{n_1},\mathcal{J}_{n_2},\mathcal{J}_{n_3}\right)(t),\end{aligned} \tag{2.9}$$

and define

$$\mathcal{J}_{\le N} = \sum_{n\le N}\mathcal{J}_n, \qquad \mathcal{R}_{N+1} = \boldsymbol{a} - \mathcal{J}_{\le N}. \tag{2.10}$$

By plugging in equation (2.5), we get that $\mathcal{R}_{N+1}$ satisfies the equation

$$\mathcal{R}_{N+1} = \mathcal{J}_{\sim N} + \mathcal{L}(\mathcal{R}_{N+1}) + \mathcal{Q}(\mathcal{R}_{N+1}) + \mathcal{C}(\mathcal{R}_{N+1}), \tag{2.11}$$

where the relevant terms are defined as

$$\mathcal{J}_{\sim N} := \sum_{\substack{n_1,n_2,n_3\le N\\ n_1+n_2+n_3\ge N}}\mathcal{I}\mathcal{W}\left(\mathcal{J}_{n_1},\mathcal{J}_{n_2},\mathcal{J}_{n_3}\right), \tag{2.12}$$

$$\mathcal{L}(v) := \sum_{n_1,n_2\le N}\left(2\mathcal{I}\mathcal{W}\left(\mathcal{J}_{n_1},\mathcal{J}_{n_2},v\right) + \mathcal{I}\mathcal{W}\left(\mathcal{J}_{n_1},v,\mathcal{J}_{n_2}\right)\right), \tag{2.13}$$

$$\mathcal{Q}(v) := \sum_{n_1\le N}\left(2\mathcal{I}\mathcal{W}\left(v,v,\mathcal{J}_{n_1}\right) + \mathcal{I}\mathcal{W}\left(v,\mathcal{J}_{n_1},v\right)\right), \tag{2.14}$$

$$\mathcal{C}(v) := \mathcal{I}\mathcal{W}(v,v,v). \tag{2.15}$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Next we will derive a formula for the time Fourier transform of $\mathcal{J}_n$; for this we need some preparation regarding multilinear forms associated with ternary trees.

Definition 2.2.

1. Let $\mathcal{T}$ be a ternary tree. We use $\mathcal{L}$ to denote the set of leaves and l their number, $\mathcal{N} = \mathcal{T}\backslash\mathcal{L}$ the set of branching nodes and n their number, and $\mathfrak{r} \in \mathcal{N}$ the root node. The *scale* of a ternary tree $\mathcal{T}$ is defined as $\mathfrak{s}(\mathcal{T}) = n$ (the number of branching nodes).[9] A tree of scale n has $l = 2n + 1$ leaves and a total of $3n + 1$ vertices.
2. (Signs on a tree) For each node $\mathfrak{n} \in \mathcal{N}$, let its children from left to right be $\mathfrak{n}_1, \mathfrak{n}_2, \mathfrak{n}_3$. We fix the sign $\iota_{\mathfrak{n}} \in \{\pm\}$ as follows: first $\iota_{\mathfrak{r}} = +$, then for any node $\mathfrak{n} \in \mathcal{N}$, define $\iota_{\mathfrak{n}_1} = \iota_{\mathfrak{n}_3} = \iota_{\mathfrak{n}}$ and $\iota_{\mathfrak{n}_2} = -\iota_{\mathfrak{n}}$.
3. (Admissible assignments) Suppose we assign to each $\mathfrak{n} \in \mathcal{T}$ an element $k_{\mathfrak{n}} \in \mathbb{Z}_L^d$. We say such an assignment $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ is *admissible* if for any $\mathfrak{n} \in \mathcal{N}$ we have $k_{\mathfrak{n}} = k_{\mathfrak{n}_1} - k_{\mathfrak{n}_2} + k_{\mathfrak{n}_3}$ and either $k_{\mathfrak{n}} \notin \{k_{\mathfrak{n}_1}, k_{\mathfrak{n}_3}\}$ or $k_{\mathfrak{n}} = k_{\mathfrak{n}_1} = k_{\mathfrak{n}_2} = k_{\mathfrak{n}_3}$. Clearly an admissible assignment is completely determined by the values of $k_{\mathfrak{l}}$ for $\mathfrak{l} \in \mathcal{L}$. For any assignment, we denote $\Omega_{\mathfrak{n}} := \Omega\left(k_{\mathfrak{n}_1}, k_{\mathfrak{n}_2}, k_{\mathfrak{n}_3}, k_{\mathfrak{n}}\right)$. Suppose we also fix[10] $d_{\mathfrak{n}} \in \{0, 1\}$ for each $\mathfrak{n} \in \mathcal{N}$; then we can define $q_{\mathfrak{n}}$ for each $\mathfrak{n} \in \mathcal{T}$ inductively by

$$q_{\mathfrak{n}} = 0 \text{ if } \mathfrak{n} \in \mathcal{L} \quad \text{or} \quad q_{\mathfrak{n}} = d_{\mathfrak{n}_1} q_{\mathfrak{n}_1} - d_{\mathfrak{n}_2} q_{\mathfrak{n}_2} + d_{\mathfrak{n}_3} q_{\mathfrak{n}_3} + \Omega_{\mathfrak{n}} \text{ if } \mathfrak{n} \in \mathcal{N}. \tag{2.16}$$

Proposition 2.3. *For each ternary tree $\mathcal{T}$, define $\mathcal{J}_{\mathcal{T}}$ inductively by*

$$\mathcal{J}_{\bullet}(t) = \chi(t) \cdot \boldsymbol{a}_{\text{in}}, \qquad \mathcal{J}_{\mathcal{T}}(t) = \mathcal{DW}\left(\mathcal{J}_{\mathcal{T}_1}, \mathcal{J}_{\mathcal{T}_2}, \mathcal{J}_{\mathcal{T}_3}\right)(t), \tag{2.17}$$

where $\bullet$ represents the tree with a single node and $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$ are the subtrees rooted at the three children of the root node of $\mathcal{T}$. Then we have

$$\mathcal{J}_n = \sum_{\mathfrak{s}(\mathcal{T})=n} \mathcal{J}_{\mathcal{T}}. \tag{2.18}$$

Moreover, for any $\mathcal{T}$ of scale $\mathfrak{s}(\mathcal{T}) = n$ we have the formula

$$\left(\widetilde{\mathcal{J}_{\mathcal{T}}}\right)_k(\tau) = \left(\frac{\alpha T}{L^d}\right)^n \sum_{(k_{\mathfrak{n}}:\mathfrak{n}\in\mathcal{T})} \mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}) \prod_{\mathfrak{l}\in\mathcal{L}} \sqrt{n_{\text{in}}(k_{\mathfrak{l}})} \cdot \prod_{\mathfrak{l}\in\mathcal{L}} \left[\eta_{k_{\mathfrak{l}}}(\omega)\right]^{\iota_{\mathfrak{l}}}, \tag{2.19}$$

where the sum is taken over all admissible assignments $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ such that $k_{\mathfrak{r}} = k$, and the function $\mathcal{K} = \mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ satisfies

$$\left|\partial_\tau^a \mathcal{K}\right| \lesssim_{a,A} \sum_{(d_{\mathfrak{n}}:\mathfrak{n}\in\mathcal{N})} \langle \tau - T d_{\mathfrak{r}} q_{\mathfrak{r}} \rangle^{-A} \cdot \prod_{\mathfrak{n}\in\mathcal{N}} \langle T q_{\mathfrak{n}} \rangle^{-1}, \tag{2.20}$$

where $q_{\mathfrak{n}}$ is defined in equation (2.16).

Proof. First, equation (2.18) follows from the definitions in equations (2.9) and (2.17) and an easy induction. We now prove formulas (2.19) and (2.20) inductively, noting also that $(a_k)_{\text{in}} = \sqrt{n_{\text{in}}(k)} \cdot \eta_k(\omega)$. For $\mathcal{T} = \bullet$, equation (2.19) follows from equation (2.17) with $\mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{r}}) = \widetilde{\chi}(\tau)$ that satisfies formula (2.20). Now suppose formulas (2.19) and (2.20) are true for smaller trees; then by formulas (2.7) and (2.17) and Lemma 2.1, up to unimportant coefficients, we can write

$$\left(\widetilde{\mathcal{J}_{\mathcal{T}}}\right)_k(\tau) = \frac{i\alpha T}{L^d} \sum_{d\in\{0,1\}} \sum_{(k_1,k_2,k_3)}^{*} \int_{\mathbb{R}^3} I_d(\tau, \sigma) \prod_{j=1}^{3} \left[\left(\widetilde{\mathcal{J}_{\mathcal{T}_j}}\right)_{k_j}(\tau_j)\right]^{\iota_j} \mathrm{d}\tau_j,$$

where $\sum^*$ represents summation under the conditions $k_j \in \mathbb{Z}_L^d$, $k_1 - k_2 + k_3 = k$ and either $k \notin \{k_1, k_3\}$ or $k = k_1 = k_2 = k_3$, the signs $(\iota_1, \iota_2, \iota_3) = (+, -, +)$, and $\sigma = \tau_1 - \tau_2 + \tau_3 + T\Omega(k_1, k_2, k_3, k)$. Now

[9] By convention, the scale of a single node is 0.
[10] This assignment is arbitrary but will usually be omitted, since there are finitely many choices.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

applying the induction hypothesis, we can write $\left(\widetilde{\mathcal{J}_{\mathcal{T}}}\right)_k(\tau)$ in the form of equation (2.19) with the function

$$\mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}) = \sum_{d\in\{0,1\}} \int_{\mathbb{R}^3} I_d(\tau, \tau_1 - \tau_2 + \tau_3 + T\Omega_{\mathfrak{r}}) \prod_{j=1}^{3} \left[\mathcal{K}_{\mathcal{T}_j}\left(\tau_j, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}_j\right)\right]^{\iota_j} \mathrm{d}\tau_j, \quad (2.21)$$

where $\mathfrak{r}$ is the root of $\mathcal{T}$ with children $\mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{r}_3$ and $\mathcal{T}_j$ is the subtree rooted at $\mathfrak{r}_j$.

It then suffices to prove that $\mathcal{K}_{\mathcal{T}}$ defined by equation (2.21) satisfies formula (2.20). By the induction hypothesis, we may fix a choice $d_{\mathfrak{n}}$ for each nonleaf node $\mathfrak{n}$ of each $\mathcal{T}_j$, and let $d_{\mathfrak{r}} = d$. Then plugging formula (2.20) into equation (2.21), we get

$$\left|\partial_\tau^a \mathcal{K}_{\mathcal{T}}\right| \lesssim_{a,A} \prod_{\mathfrak{r}\neq\mathfrak{n}\in\mathcal{N}} \frac{1}{\langle Tq_{\mathfrak{n}}\rangle} \int_{\mathbb{R}^3} \frac{1}{\langle \tau - d(\tau_1-\tau_2+\tau_3+T\Omega_{\mathfrak{r}})\rangle^A} \frac{1}{\langle \tau_1-\tau_2+\tau_3+T\Omega_{\mathfrak{r}}\rangle} \prod_{j=1}^{3} \frac{\mathrm{d}\tau_j}{\left\langle \tau_j - Td_{\mathfrak{r}_j} q_{\mathfrak{r}_j}\right\rangle^A},$$

which upon integration in τ_j gives equation (2.20). This completes the proof. □

2.3. Statement of main estimates

Define the h^b space by

$$\|a(t)\|_{h^b} = \left(\int_{\mathbb{R}} \langle\tau\rangle^{2b} \left|\widetilde{a}(\tau)\right|^2 d\tau\right)^{\frac{1}{2}}, \quad (2.22)$$

and similarly the $h^{s,b}$ space for $\boldsymbol{a}(t) = (a_k(t))_{k\in\mathbb{Z}^d_L}$ by

$$\|\boldsymbol{a}\|_{h^{s,b}} = \left(L^{-d} \sum_{k\in\mathbb{Z}^d_L} \int_{\mathbb{R}} \langle\tau\rangle^{2b} \langle k\rangle^{2s} \left|\widetilde{a}_k(\tau)\right|^2 d\tau\right)^{\frac{1}{2}}. \quad (2.23)$$

We shall estimate the solution u in an appropriately rescaled $X^{s,b}$ space, which is equivalent to estimating the sequence $\boldsymbol{a}(t) = (a_k(t))_{k\in\mathbb{Z}^d_L}$ in the space $h^{s,b}$. Define the quantity

$$\rho := \begin{cases} \alpha T & \text{if } 1 \le T \le L, \\ \alpha L & \text{if } L \le T \le L^2, \\ \alpha T L^{-1} & \text{if } T \ge L^2 \text{ and } \beta_i \text{ is generic.} \end{cases} \quad (2.24)$$

By the definition of $\delta > 0$ in formula (1.4), we can verify that $\alpha T^{1/2} \le \rho \le L^{-\delta}$.

Proposition 2.4 (Well-posedness bounds). *Let ρ be defined as in formula (2.24); then L-certainly, for all $1 \le n \le 3N$, we have*

$$\sup_k \langle k\rangle^{2s} \|(\mathcal{J}_n)_k\|_{h^b} \le L^{\theta+C\left(b-\frac{1}{2}\right)} \rho^{n-1} \left(\alpha T^{\frac{1}{2}}\right) \le L^{\theta+C\left(b-\frac{1}{2}\right)} \rho^n, \quad (2.25)$$

$$\|\mathcal{R}_{N+1}\|_{h^{s,b}} \le \rho^N. \quad (2.26)$$

Proposition 2.4 follows from the following two bounds, which will be proved in Section 3:

Proposition 2.5 (Bounds of tree terms). *We have, L-certainly, that*

$$\sup_k \langle k\rangle^{2s} \|(\mathcal{J}_{\mathcal{T}})_k\|_{h^b} \le L^{\theta+C\left(b-\frac{1}{2}\right)} \rho^{n-1} \left(\alpha T^{\frac{1}{2}}\right) \le L^{\theta+C\left(b-\frac{1}{2}\right)} \rho^n \quad (2.27)$$

for any ternary tree of depth n, where $0 \le n \le 3N$.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Proposition 2.6 (An operator norm bound). *We have, L-certainly, that for any trees $\mathcal{T}_1, \mathcal{T}_2$ with $|\mathcal{T}_j| = 3n_j + 1$ and $0 \le n_1, n_2 \le N$, the operators*

$$\mathcal{P}_+ : v \mapsto \mathcal{DW}\left(\mathcal{J}_{\mathcal{T}_1}, \mathcal{J}_{\mathcal{T}_2}, v\right) \quad \text{and} \quad \mathcal{P}_- : v \mapsto \mathcal{DW}\left(\mathcal{J}_{\mathcal{T}_1}, v, \mathcal{J}_{\mathcal{T}_2}\right) \tag{2.28}$$

satisfy the bounds

$$\|\mathcal{P}_\pm\|_{h^{s,b}\to h^{s,b}} \le L^\theta \rho^{n_1+n_2+\frac{1}{2}}. \tag{2.29}$$

Remark 2.7. The bound (2.29) is a result of the probabilistic subcriticality of the problem. Similar bounds are also used in recent work by the first author, Nahmod and Yue [19] to get sharp probabilistic local well-posedness of nonlinear Schrödinger equations. The proof in both cases relies on high-order TT^* arguments, although in [19] one needs to use the more sophisticated tensor norms due to the different ansatz caused by the inhomogeneity of initial data.

Proof of Proposition 2.4 (assuming Propositions 2.5 and 2.6). Assume we have already excluded an exceptional set of probability $\lesssim e^{-L^\theta}$. The bound (2.25) follows directly from formulas (2.18) and (2.27); it remains to bound $\mathcal{R}_{N+1}$. Recall that $\mathcal{R}_{N+1}$ satisfies equation (2.11), so it suffices to prove that the mapping

$$v \mapsto \mathcal{J}_{\sim N} + \mathcal{L}(v) + \mathcal{Q}(v) + \mathcal{C}(v)$$

is a contraction mapping from the set $\mathcal{Z} = \left\{v : \|v\|_{h^{s,b}} \le \rho^N\right\}$ to itself. We will prove only that it maps $\mathcal{Z}$ into $\mathcal{Z}$, as the contraction part follows in a similar way. Now suppose $\|v\|_{h^{s,b}} \le \rho^N$; then by formulas (2.18) and (2.27), we have

$$\|\mathcal{J}_{\sim N}\|_{h^{s,b}}^2 \sim L^{-d} \sum_{k\in\mathbb{Z}_L^d} \langle k\rangle^{2s} \|(\mathcal{J}_{\sim N})_k\|_{h^b}^2 \lesssim \left(L^{\theta+C\left(b-\frac{1}{2}\right)}\rho^{N+1}\right)^2 \cdot L^{-d} \sum_{k\in\mathbb{Z}_L^d} \langle k\rangle^{-2s} \ll \rho^{2N},$$

so $\|\mathcal{J}_{\sim N}\|_{h^{s,b}} \ll \rho^N$. Next we may use formula (2.29) to bound

$$\|\mathcal{L}(v)\|_{h^{s,b}} \le L^\theta \rho^{\frac{1}{2}} \cdot \|v\|_{h^{s,b}} \le L^\theta \rho^{\frac{1}{2}} \cdot \rho^N \ll \rho^N.$$

As for the terms $\mathcal{Q}(v)$ and $\mathcal{C}(v)$, we apply the simple bound

$$\begin{aligned} \|\mathcal{DW}(u,v,w)\|_{h^{s,b}} &\lesssim \|\mathcal{DW}(u,v,w)\|_{h^{s,1}} \lesssim \|\mathcal{DW}(u,v,w)\|_{h^{s,0}} + \|\partial_t \mathcal{DW}(u,v,w)\|_{h^{s,0}} \\ &\lesssim \frac{\alpha T}{L^d} \sum_{\text{cyc}} \|u\|_{h^{s,0}} \|v_k(t)\|_{\ell_k^1 L_t^\infty} \|w_k(t)\|_{\ell_k^1 L_t^\infty} \lesssim \alpha T L^d \|u\|_{h^{s,b}} \|v\|_{h^{s,b}} \|w\|_{h^{s,b}} \end{aligned} \tag{2.30}$$

(which easily follows from formula (2.7)), where $\sum_{\text{cyc}}$ means summing in permutations of (u, v, w). As $\alpha T \le L^d$, we conclude (also using Proposition 2.5) that

$$\|\mathcal{Q}(v)\|_{h^{s,b}} + \|\mathcal{C}(v)\|_{h^{s,b}} \lesssim \alpha T L^{\theta+d+C\left(b-\frac{1}{2}\right)} \rho^{2N} \ll \rho^N,$$

since $\rho \le L^{-\delta}$ and $N \gg \delta^{-1}$. This completes the proof. □

3. Proof of main estimates

In this section we prove Propositions 2.5 and 2.6.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

3.1. Large deviation and basic counting estimates

We start by making some preparations, namely the large deviation and counting estimates that will be used repeatedly in the proof later.

Lemma 3.1. *Let $\{\eta_k(\omega)\}$ be independent, identically distributed complex random variables, such that the law of each η_k is either Gaussian with mean 0 and variance 1 or the uniform distribution on the unit circle. Let $F = F(\omega)$ be defined by*

$$F(\omega) = \sum_{k_1,\dots,k_n} a_{k_1\cdots k_n} \prod_{j=1}^{n} \eta_{k_j}^{\iota_j}, \tag{3.1}$$

where $a_{k_1\cdots k_n}$ are constants; then F can be divided into finitely many terms, and for each term there is a choice of $X = \{i_1,\dots,i_p\}$ and $Y = \{j_1,\dots,j_p\}$, which are two disjoint subsets of $\{1,2,\dots,n\}$, such that

$$\mathbb{P}\left(|F(\omega)| \geq A \cdot M^{\frac{1}{2}}\right) \leq Ce^{-cA^{\frac{2}{n}}} \tag{3.2}$$

holds with

$$M = \sum_{(k_\ell):\ell\notin X\cup Y} \left(\sum_{\text{pairing } (k_{i_s},k_{j_s}):1\leq s\leq p} \left|a_{k_1\cdots k_n}\right| \right)^2, \tag{3.3}$$

where a pairing (k_i,k_j) means $(\iota_i+\iota_j, \iota_i k_i + \iota_j k_j) = 0$.

Proof. First assume η_k is Gaussian. Then by the standard hypercontractivity estimate for an Ornstein–Uhlenbeck semigroup (see, e.g., [40]), we know that formula (3.2) holds with M replaced by $\mathbb{E}\,|F(\omega)|^2$. Now to estimate $\mathbb{E}\,|F(\omega)|^2$, by dividing the sum (3.1) into finitely many terms and rearranging the subscripts, we may assume in a monomial of equation (3.1) that

$$k_1 = \cdots = k_{j_1}, k_{j_1+1} = \cdots = k_{j_2}, \cdots, k_{j_{r-1}+1} = \cdots = k_{j_r}, \quad 1 \leq j_1 < \cdots < j_r = n, \tag{3.4}$$

and the k_{j_s} are different for $1 \leq s \leq r$. Such a monomial has the form

$$\prod_{s=1}^{r} \eta_{k_{j_s}}^{b_s} \left(\overline{\eta_{k_{j_s}}}\right)^{c_s}, \qquad b_s + c_s = j_s - j_{s-1}\ (j_0 = 0),$$

where the factors for different s are independent. We may also assume $b_s = c_s$ for $1 \leq s \leq q$ and $b_s \neq c_s$ for $q+1 \leq s \leq r$, and for $1 \leq j \leq j_q$ we may assume ι_j has the same sign as $(-1)^j$. Then we can further rewrite this monomial as a linear combination of

$$\prod_{s=1}^{p} b_s! \prod_{s=p+1}^{q} \left(\left|\eta_{k_{j_s}}\right|^{2b_s} - b_s!\right) \prod_{s=q+1}^{r} \eta_{k_{j_s}}^{b_s} \left(\overline{\eta_{k_{j_s}}}\right)^{c_s}$$

for $1 \leq p \leq q$. Therefore, $F(\omega)$ is a finite linear combination of expressions of the form

$$\sum_{k_{j_1},\dots,k_{j_r}} a_{k_{j_1},\dots,k_{j_1},\dots,k_{j_r},\dots k_{j_r}} \prod_{s=1}^{p} b_s! \prod_{s=p+1}^{q} \left(\left|\eta_{k_{j_s}}\right|^{2b_s} - b_s!\right) \prod_{s=q+1}^{r} \eta_{k_{j_s}}^{b_s} \left(\overline{\eta_{k_{j_s}}}\right)^{c_s}.$$

Due to independence and the fact that $\mathbb{E}\left(|\eta|^{2b} - b!\right) = \mathbb{E}\left(\eta^b\,(\overline{\eta})^c\right) = 0$ for a normalised Gaussian η and $b \neq c$, we conclude that

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

$$\mathbb{E}\,|F(\omega)|^2 \lesssim \sum_{k_{j_{p+1}},\dots,k_{j_r}} \left(\sum_{k_{j_1},\dots,k_{j_p}} \left| a_{k_{j_1},\dots,k_{j_1},\dots,k_{j_r},\dots k_{j_r}} \right| \right)^2, \tag{3.5}$$

which is bounded by the right-hand side of equation (3.3), by choosing $X = \{1, 3, \dots, j_p - 1\}$ and $Y = \{2, 4, \dots, j_p\}$, as under our assumptions (k_{2i-1}, k_{2i}) is a pairing for $2i \le j_p$.

Now assume η_k is uniformly distributed on the unit circle. Let $\{g_k(\omega)\}$ be independent, identically distributed normalised Gaussians as in the first part, and consider the random variable

$$H(\omega) = \sum_{k_1,\dots,k_n} \left| a_{k_1 \cdots k_n} \right| \prod_{j=1}^{n} g_{k_j}^{\iota_j}.$$

We can calculate

$$\mathbb{E}\left(|F(\omega)|^{2q} \right) = \sum_{\left(k_j^i, \ell_j^i\right)} \prod_{i=1}^{q} a_{k_1^i \cdots k_n^i} \overline{a_{\ell_1^i \cdots \ell_n^i}} \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{n} \eta_{k_j^i}^{\iota_j} \overline{\eta_{\ell_j^i}^{\iota_j}} \right), \tag{3.6}$$

where $1 \le i \le q$ and $1 \le j \le n$, and similarly for H,

$$\mathbb{E}\left(|H(\omega)|^{2q} \right) = \sum_{\left(k_j^i, \ell_j^i\right)} \prod_{i=1}^{q} \left| a_{k_1^i \cdots k_n^i} \right| \left| a_{\ell_1^i \cdots \ell_n^i} \right| \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{n} g_{k_j^i}^{\iota_j} \overline{g_{\ell_j^i}^{\iota_j}} \right). \tag{3.7}$$

The point is that we always have

$$\left| \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{r} \eta_{k_j^i}^{\iota_j} \overline{\eta_{\ell_j^i}^{\iota_j}} \right) \right| \le \mathrm{Re}\mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{r} g_{k_j^i}^{\iota_j} \overline{g_{\ell_j^i}^{\iota_j}} \right).$$

In fact, in order for either side to be nonzero, for any particular k we must have

$$\begin{aligned} &\#\left\{ (i,j) : k_j^i = k, \iota_j = + \right\} + \#\left\{ (i,j) : \ell_j^i = k, \iota_j = - \right\} \\ &= \#\left\{ (i,j) : k_j^i = k, \iota_j = - \right\} + \#\left\{ (i,j) : \ell_j^i = k, \iota_j = + \right\}. \end{aligned}$$

Let both be equal to m; then by independence, the factor that the $\eta_k^{\pm}$s contribute to the expectation on the left-hand side will be $\mathbb{E}\,|\eta_k|^{2m} = 1$, while for the right-hand side it will be $\mathbb{E}\,|g_k|^{2m} = m! \ge 1$.

This implies that $\mathbb{E}\left(|F|^{2q} \right) \le \mathbb{E}\left(|H|^{2q} \right)$ for any positive integer q; since formula (3.2) holds for H, we have

$$\mathbb{E}\left(|H|^{2q} \right) \le (Cq)^{nq} M^q$$

with an absolute constant C. This gives an upper bound for $\mathbb{E}\left(|F|^{2q} \right)$, and by Chebyshev inequality, we deduce formula (3.2) for F. □

Lemma 3.2. *Let $\beta = (\beta_1, \dots, \beta_d) \in [1,2]^d$ and $0 < T \le L^d$. Assume that β is generic for $T \ge L^2$. Then, uniformly in $(k, a, b, c) \in \left(\mathbb{Z}_L^d\right)^4$ and $m \in \mathbb{R}$, the sets*

$$\begin{aligned} S_3 = \Big\{ (x,y,z) \in \left(\mathbb{Z}_L^d\right)^3 : x - y + z = k,\ \left| |x|_\beta^2 - |y|_\beta^2 + |z|_\beta^2 - |k|_\beta^2 - m \right| \le T^{-1}, \\ |x - a| \le L^\theta,\ |y - b| \le L^\theta,\ |z - c| \le L^\theta \text{ and } k \notin \{x, z\} \Big\}, \end{aligned} \tag{3.8}$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

$$S_2 = \left\{ (x,y) \in \left(\mathbb{Z}_L^d\right)^3 : x \pm y = k,\ \left| |x|_\beta^2 \pm |y|_\beta^2 - |k|_\beta^2 - m \right| \le T^{-1},\right.$$
$$\left. |x-a| \le L^\theta,\ |y-b| \le L^\theta \text{ and } x \ne y \text{ if the sign } \pm \text{ is } - \right\} \tag{3.9}$$

satisfy the bounds

$$\#S_3 \lesssim L^{2d+\theta}T^{-1}, \qquad \#S_2 \lesssim \begin{cases} L^{d+\theta} & \text{if } T \le L, \\ L^{d+1+\theta}T^{-1} & \text{if } T \in [L, L^2], \\ L^{d-1+\theta} & \text{if } T \ge L^2 \text{ and } \beta_i \text{ is generic}, \end{cases} \tag{3.10}$$

where in the first inequality of formula (3.10) we also assume $|k|, |a|, |b| \le L^\theta$.

Moreover, with ρ *defined as in formula (2.24), we have*

$$\max\left((\#S_3)^{\frac{1}{2}}, \#S_2 \right) \le L^\theta Q, \qquad Q := \frac{L^d \rho}{\alpha T}, \tag{3.11}$$

without any assumption on (k, a, b).

Proof. We first consider S_3. Let $k - x = p$ and $k - z = q$; then we may write $p = \left(L^{-1}u_1, \dots, L^{-1}u_d\right)$ and similarly for q, where each u_i and v_i is an integer and belongs to a fixed interval of length $O\left(L^{1+\theta}\right)$. Moreover, from $(x, y, z) \in S_3$ we deduce that

$$\left| \sum_{i=1}^d \beta_i u_i v_i + \frac{L^2 m}{2} \right| \le \frac{L^2 T^{-1}}{2}.$$

We may assume $u_i v_i = 0$ for $1 \le i \le r$, and $\sigma_i := u_i v_i \ne 0$ for $r+1 \le i \le d$; then the number of choices for $(u_i, v_i : 1 \le i \le r)$ is $O\left(L^{r+\theta}\right)$. It is known (see [17, 18]) that given $\sigma \ne 0$, the number of integer pairs (u, v) such that u and v each belongs to an interval of length $O\left(L^{1+\theta}\right)$ and $uv = \sigma$ is $O\left(L^\theta\right)$. Therefore, if $|k|, |a|, |b| \le L^\theta$, then $\#S_3$ is bounded by $O\left(L^{r+\theta}\right)$ times the number of choices for $(\sigma_{r+1}, \dots, \sigma_d)$ that satisfy

$$|\sigma_j| \le L^{2+\theta}\ (r+1 \le j \le d), \qquad \sum_{j=r+1}^d \beta_i \sigma_i = -\frac{L^2 m}{2} + O\left(L^2 T^{-1}\right). \tag{3.12}$$

Using the assumption $T \le L^d$, it suffices to show that the number of choices for $(\sigma_{r+1}, \dots, \sigma_d)$ satisfying formula (3.12) is at most $O\left(1 + L^{2(d-r)+\theta}T^{-1}\right)$. This latter bound is trivial if $d - r = 1$ or $L^2T^{-1} \ge 1$, so we may assume $d - r \ge 2$, $T \ge L^2$ and β_i is generic. It is well known in Diophantine approximation (see, e.g., [9]) that for generic β_i we have

$$\left| \sum_{i=r+1}^d \beta_i \eta_i \right| \gtrsim \left(\max_{r+1 \le i \le d} \langle \eta_i \rangle \right)^{-(d-r-1)-\theta} \qquad \text{if } \eta_i \text{ are not all 0},$$

so the distance between any two points $(\sigma_i : r+1 \le i \le d)$ and $(\sigma_i' : r+1 \le i \le d)$ satisfying formula (3.12) is at least $\left(L^2T^{-1}\right)^{-\frac{1}{d-r-1}-\theta}$. Since all these points belong to a box which has size $O(1)$ in one direction and size $O\left(L^{2+\theta}\right)$ in other orthogonal directions, we deduce that the number of solutions to formula (3.12) is at most $1 + L^\theta L^{2(d-r-1)} L^2 T^{-1}$, as desired.

Next, without any assumption on (k, a, b), we need to prove formula (3.11). By definition (2.24) we can check that $Q^2 \ge L^{2d}\left(\min\left(T, L^2\right)\right)^{-1}$, so it suffices to prove the first inequality of formula (3.10),

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

assuming $T \leq L^2$. But this again follows from formula (3.12), noting that now $|\sigma_j| \leq L^{2+\theta}$ is no longer true, but each σ_j still has at most $L^{2+\theta}$ possible values.

Finally we consider S_2, which is much easier. In fact, formula (3.11) follows from formula (3.10), so we only need to prove the latter. Now if $T \leq L$, we trivially have $\#S_2 \leq L^{d+\theta}$, as y will be fixed once x is; if $T \geq L$, then we may assume $x_d - y_d \neq 0$ if the sign $\pm$ is $-$, and then fix the first coordinates $x_j (1 \leq j \leq d-1)$ and hence $y_j (1 \leq j \leq d-1)$. Then we have that $x_d \pm y_d$ is fixed, and $x_d^2 \pm y_d^2$ belongs to a fixed interval of length $O\left(T^{-1}\right)$. Since $x_d, y_d \in L^{-1}\mathbb{Z}$, we know that x_d has at most $1 + L^2 T^{-1}$ choices, which implies what we want to prove. □

3.2. Bounds for $\mathcal{J}_n$

In this section we prove Proposition 2.5. We will need to extend the notion of ternary trees to *paired, coloured* ternary trees:

Definition 3.3 (Tree pairings and colourings). Let $\mathcal{T}$ be a ternary tree as in Definition 2.2. We will pair some of the leaves of $\mathcal{T}$ such that each leaf belongs to at most one pair. The two leaves in a pair are called *partners* of each other, and the unpaired leaves are called *single*. We assume $\iota_{\mathfrak{l}} + \iota_{\mathfrak{l}'} = 0$ for any pair $(\mathfrak{l}, \mathfrak{l}')$. The set of single leaves is denoted $\mathcal{S}$. The number of pairs is denoted by p, so that $|\mathcal{S}| = l - 2p$. Moreover, we assume that some nodes in $\mathcal{S} \cup \{\mathfrak{r}\}$ are coloured red, and let $\mathcal{R}$ be the set of red nodes. We shall denote $r = |\mathcal{R}|$.

We shall use red colouring to denote that the frequency assignments to the corresponding red vertex are fixed in the counting process. We also introduce the following definition:

Definition 3.4 (Strong admissibility). Suppose we fix $n_{\mathfrak{m}} \in \mathbb{Z}_L^d$ for each $\mathfrak{m} \in \mathcal{R}$. An assignment $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ is called *strongly admissible* with respect to the given pairing, colouring and $(n_{\mathfrak{m}} : \mathfrak{m} \in \mathcal{R})$ if it is admissible in the sense of Definition 2.2, and

$$k_{\mathfrak{m}} = n_{\mathfrak{m}} \ \forall \mathfrak{m} \in \mathcal{R}, \qquad |k_{\mathfrak{l}}| \leq L^{\theta} \ \forall \mathfrak{l} \in \mathcal{L}, \qquad k_{\mathfrak{l}} = k_{\mathfrak{l}'} \ \forall \text{ pairs of leaves } (\mathfrak{l}, \mathfrak{l}'). \tag{3.13}$$

The key to the proof of Proposition 2.5 is the following combinatorial counting bound:

Proposition 3.5. *Let $\mathcal{T}$ be a paired and coloured ternary tree such that $\mathcal{R} \neq \emptyset$, and let $(n_{\mathfrak{m}} : \mathfrak{m} \in \mathcal{R})$ be fixed. We also fix $\sigma_{\mathfrak{n}} \in \mathbb{R}$ for each $\mathfrak{n} \in \mathcal{N}$. Let $l = |\mathcal{L}|$ be the total number of leaves, p be the number of pairs and $r = |\mathcal{R}|$ be the number of red nodes. Then the number of strongly admissible assignments $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ which also satisfy*

$$|\Omega_{\mathfrak{n}} - \sigma_{\mathfrak{n}}| \leq T^{-1} \ \forall \mathfrak{n} \in \mathcal{N} \tag{3.14}$$

is – recalling Q defined in formula (3.11) – bounded by

$$M \leq \begin{cases} L^{\theta} Q^{l-p-r} & \text{if } \mathcal{R} \neq \mathcal{S} \cup \{\mathfrak{r}\}, \\ L^{\theta} Q^{l-p-r+1} & \text{if } \mathcal{R} = \mathcal{S} \cup \{\mathfrak{r}\}. \end{cases} \tag{3.15}$$

Proof. We proceed by induction. The base cases directly follow from formula (3.11). Now suppose the desired bound holds for all smaller trees, and consider $\mathcal{T}$. Let $\mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{r}_3$ be the children of the root node $\mathfrak{r}$ and $\mathcal{T}_j$ be the subtree rooted at $\mathfrak{r}_j$. Let l_j be the number of leaves in $\mathcal{T}_j$, p_j the number of pairs within $\mathcal{T}_j$ and p_{ij} the number of pairings between $\mathcal{T}_i$ and $\mathcal{T}_j$, and let $r_j = \left|\mathcal{R} \cap \mathcal{T}_j\right|$; then we have

$$l = l_1 + l_2 + l_3, \qquad p = p_1 + p_2 + p_3 + p_{12} + p_{13} + p_{23}, \qquad r = r_1 + r_2 + r_3 + \mathbf{1}_{\mathfrak{r} \in \mathcal{R}}.$$

Also note that $|k_{\mathfrak{n}}| \lesssim L^{\theta}$ for all $\mathfrak{n} \in \mathcal{T}$.

The proof will be completely algorithmic, with the discussion of a lot of cases. The general strategy is to perform the following four operations, which we refer to as $\mathcal{O}_j (0 \leq j \leq 3)$, in a suitable order.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Here in operation $\mathcal{O}_0$ we apply formula (3.11) to count the number of choices for the values among $\{k_{\mathfrak{r}}, k_{\mathfrak{r}_1}, k_{\mathfrak{r}_2}, k_{\mathfrak{r}_2}\}$ that are not already fixed (this step may be trivial if three of these four vectors are already fixed –i.e., coloured – or if one of them is already fixed and $k_{\mathfrak{r}} = k_{\mathfrak{r}_1} = k_{\mathfrak{r}_2} = k_{\mathfrak{r}_3}$). In operations $\mathcal{O}_j (1 \le j \le 3)$, we apply the induction hypothesis to one of the subtrees $\mathcal{T}_j$ and count the number of choices for $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}_j)$. Let the number of choices associated with $\mathcal{O}_j (0 \le j \le 3)$ be M_j, with superscripts indicating different cases. In the whole process we may colour new nodes $\mathfrak{n}$ red if $k_{\mathfrak{n}}$ is already fixed during the previous operations, namely when $\mathfrak{n} = \mathfrak{r}$ and we have performed $\mathcal{O}_0$ before, when $\mathfrak{n} = \mathfrak{r}_j$ and we have performed $\mathcal{O}_0$ or $\mathcal{O}_j$ before or when $\mathfrak{n}$ is a leaf that has a partner in $\mathcal{T}_j$ and we have performed $\mathcal{O}_j$ before.

(1) Suppose $\mathfrak{r} \notin \mathcal{R}$; then we may assume that there is a red leaf from $\mathcal{T}_1$.[11] We first perform $\mathcal{O}_1$ and get a factor

$$M_1^{(1)} := L^{\theta} Q^{l_1 - p_1 - r_1}.$$

Now $\mathfrak{r}_1$ is coloured red, as is any leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$ which has a partner in $\mathcal{T}_1$. There are then two cases.

(1.1) Suppose now there is a leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$, say from $\mathcal{T}_2$, that is red. Then we perform $\mathcal{O}_2$ and get a factor

$$M_2^{(1.1)} := L^{\theta} Q^{l_2 - p_2 - r_2 - p_{12}}.$$

Now $\mathfrak{r}_2$ is coloured red, as is any leaf of $\mathcal{T}_3$ which has a partner in $\mathcal{T}_2$. There are again two cases.

(1.1.1) Suppose now there is a red leaf in $\mathcal{T}_3$; then we perform $\mathcal{O}_3$ and get a factor

$$M_3^{(1.1.1)} := L^{\theta} Q^{l_3 - p_3 - r_3 - p_{13} - p_{23}},$$

then colour $\mathfrak{r}_3$ red and apply $\mathcal{O}_0$ to get a factor $M_0^{(1.1.1)} := 1$. Thus

$$M \le M_1^{(1)} M_2^{(1.1)} M_3^{(1.1.1)} M_0^{(1.1.1)} = L^{l - p - r + \theta},$$

which is what we need.

(1.1.2) Suppose after step (1.1) there is no red leaf in $\mathcal{T}_3$; then $r_3 = p_{13} = p_{23} = 0$. We perform $\mathcal{O}_0$ and get a factor $M_0^{(1.1.2)} := L^{\theta} Q^1$ (perhaps with slightly enlarged θ; the same applies later). Now we may colour $\mathfrak{r}_3$ red and perform $\mathcal{O}_3$ to get a factor

$$M_3^{(1.1.2)} := L^{\theta} Q^{l_3 - p_3 - 1}.$$

Thus

$$M \le M_1^{(1)} M_2^{(1.1)} M_0^{(1.1.2)} M_3^{(1.1.2)} = L^{\theta} Q^{l - p - r},$$

which is what we need.

(1.2) Now suppose that after step (1) there is no red leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$; then $r_2 = r_3 = p_{12} = p_{13} = 0$. There are two cases.

(1.2.1) Suppose there is a single leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$, say from $\mathcal{T}_2$. Then we will perform $\mathcal{O}_0$ and get a factor $M_0^{(1.2.1)} := L^{\theta} Q^2$. Now we may colour $\mathfrak{r}_2$ and $\mathfrak{r}_3$ red and perform $\mathcal{O}_3$ to get a factor

$$M_3^{(1.2.1)} := L^{\theta} Q^{l_3 - p_3 - 1}.$$

Now any leaf of $\mathcal{T}_2$ which has a partner in $\mathcal{T}_3$ is coloured red, so we may perform $\mathcal{O}_2$ and get a factor

$$M_2^{(1.2.1)} := L^{\theta} Q^{l_2 - p_2 - p_{23} - 1}.$$

[11] Strictly speaking, the roles of $\mathcal{T}_1$ and $\mathcal{T}_2$ are not exactly symmetric, due to the sign difference, but this will not affect the proof, because formula (3.11) includes all choices of signs.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Thus

$$M \leq M_1^{(1)} M_0^{(1.2.1)} M_3^{(1.2.1)} M_2^{(1.2.1)} = L^\theta Q^{l-p-r},$$

which is what we need.

(1.2.2) Suppose there is no single leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$; then all leaves in $\mathcal{T}_2 \cup \mathcal{T}_3$ are paired to one another, which implies that $k_{\mathfrak{r}_2} = k_{\mathfrak{r}_3}$ and that $\mathfrak{r}_2$ and $\mathfrak{r}_3$ have opposite signs, and hence by the admissibility condition we must have $k_{\mathfrak{r}} = k_{\mathfrak{r}_1} = k_{\mathfrak{r}_2} = k_{\mathfrak{r}_3}$. This allows us to perform $\mathcal{O}_0$ and colour $\mathfrak{r}_2$ and $\mathfrak{r}_3$ red with $M_0^{(1.2.2)} := 1$, then perform $\mathcal{O}_3$ and colour red any leaf of $\mathcal{T}_2$ which has a partner in $\mathcal{T}_3$, then perform $\mathcal{O}_2$ (for which we use the second bound in formula (3.15)). This leads to the factors

$$M_3^{(1.2.2)} := L^\theta Q^{l_3-p_3-1}, \qquad M_2^{(1.2.2)} \leq L^\theta Q^{l_2-p_2-p_{23}-1+1},$$

and thus

$$M \leq M_1^{(1)} M_0^{(1,2,2)} M_3^{(1.2.2)} M_2^{(1.2.2)} = L^\theta Q^{l-p-r-1},$$

which is better than what we need.

(2) Now suppose $\mathfrak{r} \in \mathcal{R}$; then $r = r_1 + r_2 + r_3 + 1$. There are two cases.

(2.1) Suppose there is one single leaf that is not red, say from $\mathcal{T}_1$. There are again two cases.

(2.1.1) Suppose there is a red leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$, say $\mathcal{T}_2$. Then we perform $\mathcal{O}_2$ and get a factor

$$M_2^{(2.1.1)} := L^\theta Q^{l_2-p_2-r_2}.$$

We now colour red $\mathfrak{r}_2$ and any leaf in $\mathcal{T}_1 \cup \mathcal{T}_3$ which has a partner in $\mathcal{T}_2$. There are a further two cases.

(2.1.1.1) Suppose now there is a red leaf in $\mathcal{T}_3$; then we perform $\mathcal{O}_3$ and get a factor

$$M_3^{(2.1.1.1)} := L^\theta Q^{l_3-p_3-r_3-p_{23}}.$$

Now we perform $\mathcal{O}_0$ and get a factor $M_0^{(2.1.1.1)} := 1$, then colour red $\mathfrak{r}_1$ as well as any leaf of $\mathcal{T}_1$ which has a partner in $\mathcal{T}_3$, and perform $\mathcal{O}_1$ to get a factor

$$M_1^{(2.1.1.1)} := L^\theta Q^{l_1-p_1-r_1-p_{12}-p_{13}-1}.$$

Thus

$$M \leq M_2^{(2.1.1)} M_3^{(2.1.1.1)} M_0^{(2.1.1.1)} M_1^{(2.1.1.1)} = L^\theta Q^{l-p-r},$$

which is what we need.

(2.1.1.2) Suppose after step (2.1.1) there is no red leaf in $\mathcal{T}_3$; then $r_3 = p_{23} = 0$. We perform $\mathcal{O}_0$ and get a factor $M_0^{(2.1.1.2)} := L^\theta Q^1$. Then we colour $\mathfrak{r}_1$ and $\mathfrak{r}_3$ red and perform $\mathcal{O}_3$ to get a factor

$$M_3^{(2.1.1.2)} := L^\theta Q^{l_3-p_3-1}.$$

Finally we colour red any leaf of $\mathcal{T}_1$ which has a partner in $\mathcal{T}_3$, and perform $\mathcal{O}_1$ to get a factor

$$M_1^{(2.1.1.2)} := L^\theta Q^{l_1-p_1-r_1-p_{12}-p_{13}-1}.$$

Thus

$$M \leq M_2^{(2.1.1)} M_0^{(2.1.1.2)} M_3^{(2.1.1.2)} M_1^{(2.1.1.2)} = L^\theta Q^{l-p-r},$$

which is what we need.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

(2.1.2) Suppose in the beginning there is no red leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$; then $r_2 = r_3 = 0$. There are again two cases.

(2.1.2.1) Suppose there is a leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$, say from $\mathcal{T}_2$, that is either single or paired with a leaf in $\mathcal{T}_1$. Then we perform $\mathcal{O}_0$ and get a factor $M_0^{(2.1.2.1)} := L^\theta Q^2$. After this we colour $\mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{r}_3$ red and perform $\mathcal{O}_3$ to get a factor

$$M_3^{(2.1.2.1)} := L^\theta Q^{l_3 - p_3 - 1}.$$

We then colour red any leaf of $\mathcal{T}_1$ and $\mathcal{T}_2$ which has a partner in $\mathcal{T}_3$, and perform $\mathcal{O}_2$ to get a factor

$$M_2^{(2.1.2.1)} := L^\theta Q^{l_2 - p_2 - p_{23} - 1}.$$

Finally we colour red any leaf of $\mathcal{T}_1$ which has a partner in $\mathcal{T}_2$, and perform $\mathcal{O}_1$ to get a factor

$$M_1^{(2.1.2.1)} := L^\theta Q^{l_1 - p_1 - r_1 - p_{12} - p_{13} - 1}.$$

Thus

$$M \le M_0^{(2.1.2.1)} M_3^{(2.1.2.1)} M_2^{(2.1.2.1)} M_1^{(2.1.2.1)} = L^\theta Q^{l-p-r},$$

which is what we need.

(2.1.2.2) Suppose there is no leaf in $\mathcal{T}_2 \cup \mathcal{T}_3$ that is either single or paired with a leaf in $\mathcal{T}_1$; then in the same way as in case (1.2.2), we must have $k_{\mathfrak{r}} = k_{\mathfrak{r}_1} = k_{\mathfrak{r}_2} = k_{\mathfrak{r}_3}$. Moreover, we have $p_{12} = p_{13} = 0$. Then we perform $\mathcal{O}_0$ and get a factor $M_0^{(2.1.2.2)} := 1$. After this we colour $\mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{r}_3$ red and perform $\mathcal{O}_3$ to get a factor

$$M_3^{(2.1.2.2)} := L^\theta Q^{l_3 - p_3 - 1}.$$

We then colour red any leaf of $\mathcal{T}_2$ which has a partner in $\mathcal{T}_3$ and perform $\mathcal{O}_2$ to get a factor

$$M_2^{(2.1.2.2)} \le L^\theta Q^{l_2 - p_2 - p_{23} - 1 + 1}.$$

Finally, we perform $\mathcal{O}_1$, again using the second part of estimate (3.15), to get a factor

$$M_1^{(2.1.2.2)} := L^\theta Q^{l_1 - p_1 - r_1 - 1}.$$

Thus

$$M \le M_0^{(2.1.2.2)} M_3^{(2.1.2.2)} M_2^{(2.1.2.2)} M_1^{(2.1.2.2)} = L^\theta Q^{l-p-r-1},$$

which is better than what we need.

(2.2) Now suppose that in the beginning all single leaves are red – that is, $\mathcal{R} = \mathcal{S} \cup \{\mathfrak{r}\}$. Then we can argue in exactly the same way as in case (2.1), except that in the last step where we perform $\mathcal{O}_1$, it may happen that the root $\mathfrak{r}_1$ as well as all leaves of $\mathcal{T}_1$ are red at that time, so we lose one power of Q in view of the weaker bound from the induction hypothesis. However, since $\mathcal{R} = \mathcal{S} \cup \{\mathfrak{r}\}$, we are in fact allowed to lose this power, so we can still close the inductive step, in the same way as in case (2.1). This completes the proof. □

Corollary 3.6. *In Proposition 3.5, suppose $\mathcal{R} = \{\mathfrak{r}\}$. Then formula (3.15) can be improved to*

$$M \le L^\theta Q^{l-p-3} L^{2d} T^{-1}. \tag{3.16}$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Proof. In the proof of Proposition 3.5, we are now in case (2.1.2). In each subcase, either (2.1.2.1) or (2.1.2.2), we perform the operation $\mathcal{O}_0$ first. In case (2.1.2.1), by formula (3.10) – noting that the extra conditions are satisfied – we can replace the bound $M_0^{(2.1.2.1)}$ by $M_0' := L^\theta L^{2d} T^{-1}$, so we get

$$M \le M_0' M_3^{(2.1.2.1)} M_2^{(2.1.2.1)} M_1^{(2.1.2.1)} = L^\theta Q^{l-p-3} L^{2d} T^{-1}.$$

In case (2.1.2.2) we get an improvement: we have $M \le L^\theta Q^{l-p-2}$, which also implies formula (3.16), since we can check $Q \le L^{2d} T^{-1} \le Q^2$ by definition. □

Now we are ready to prove Proposition 2.5.

Proof of Proposition 2.5. We start with equation (2.19). Let $|\mathcal{T}| = 3n + 1$. Due to the rapid decay of $\sqrt{n_{\text{in}}}$, we may assume in the summation that $|k_{\mathfrak{I}}| \le L^\theta$ for any $\mathfrak{I} \in \mathcal{L}$, and so $|k| \le L^\theta$ also. For any *fixed* value of τ, we may apply Lemma 3.1 to the L-certain estimate $\left(\widetilde{\mathcal{J}_{\mathcal{T}}}\right)_k(\tau)$. Namely, L-certainly, we have, for some choice of pairing and with colouring $\mathcal{R} = \{\mathfrak{r}\}$ and $n_{\mathfrak{r}} = k$,

$$\langle k\rangle^{4s} \left|\left(\widetilde{\mathcal{J}_{\mathcal{T}}}\right)_k(\tau)\right|^2 \le L^\theta \left(\frac{\alpha T}{L^d}\right)^{2n} \sum_{(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{S})} \left[\sum_{(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\setminus\mathcal{S})}^{**} |\mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})|\right]^2, \tag{3.17}$$

where $\sum^{**}$ represents summation under the condition that the unique admissible assignment determined by $(k_{\mathfrak{I}} : \mathfrak{I} \in \mathcal{S})$ and $(k_{\mathfrak{I}} : \mathfrak{I} \in \mathcal{L}\setminus\mathcal{S})$ is strongly admissible. Next we would like to assume formula (3.17) for *all* τ, which can be done by the following trick. First, due to the decay factor in formula (2.20) and the assumption $|k_{\mathfrak{I}}| \le L^\theta$, we may assume $|\tau| \le L^{d+\theta}$; moreover, choosing a large power D, we may divide the interval $\left[-L^\theta, L^\theta\right]$ into subintervals of length L^{-D} and pick one point τ_j from each interval. Due to the differentiability of $\mathcal{K}_{\mathcal{T}}$ (see formula (2.20)), we can bound the difference

$$\left|\mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}) - \mathcal{K}_{\mathcal{T}}(\tau_j, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})\right|$$

by a large negative power of L, provided τ is in the same interval as τ_j. Therefore, as long as formula (3.17) is true for each τ_j, we can assume it is true for each τ up to negligible errors. Since the number of τ_js is at most $O\left(L^{2D}\right)$ and formula (3.17) holds L-certainly for each fixed τ_j, we conclude that, L-certainly, formula (3.17) holds for all τ.

Now, by expanding the square in formula (3.17), it suffices to bound the quantity

$$\int_{\mathbb{R}} \langle\tau\rangle^{2b} \sum_{(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{S})} \sum_{(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\setminus\mathcal{S})}^{**} \sum_{\left(k'_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\setminus\mathcal{S}\right)}^{**'} |\mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})| \cdot \left|\mathcal{K}_{\mathcal{T}}\left(\tau, k'_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}\right)\right| \mathrm{d}\tau,$$

where $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ is the unique admissible assignment determined by $(k_{\mathfrak{I}} : \mathfrak{I} \in \mathcal{L})$ and $(k_{\mathfrak{I}} : \mathfrak{I} \in \mathcal{L}\setminus\mathcal{S})$, and $\left(k'_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}\right)$ is the one determined by $(k_{\mathfrak{I}} : \mathfrak{I} \in \mathcal{L})$ and $\left(k'_{\mathfrak{I}} : \mathfrak{I} \in \mathcal{L}\setminus\mathcal{S}\right)$. The conditions in the summations $\sum^{**}$ and $\sum^{**'}$ correspond to these two assignments being strongly admissible. By formula (2.20), we have (for some choice of $d_{\mathfrak{n}}$)

$$\begin{aligned}
&\langle\tau\rangle^{2b} |\mathcal{K}_{\mathcal{T}}(\tau, k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})| \cdot \left|\mathcal{K}_{\mathcal{T}}\left(\tau, k'_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}\right)\right| \\
&\lesssim \langle\tau\rangle^{2b} \langle\tau - T d_{\mathfrak{r}} q_{\mathfrak{r}}\rangle^{-10} \left\langle\tau - T d_{\mathfrak{r}} q'_{\mathfrak{r}}\right\rangle^{-10} \prod_{\mathfrak{n}\in\mathcal{N}} \langle T q_{\mathfrak{n}}\rangle^{-1} \left\langle T q'_{\mathfrak{n}}\right\rangle^{-1},
\end{aligned}$$

where $q_{\mathfrak{n}}$ and $q'_{\mathfrak{n}}$ are defined from the assignments $(k_{\mathfrak{n}})$ and $\left(k'_{\mathfrak{n}}\right)$, respectively, via equation (2.16). Thus the integral in τ gives

$$\max\left(\langle T q_{\mathfrak{r}}\rangle, \left\langle T q'_{\mathfrak{r}}\right\rangle\right)^{-2+2b} \left\langle T\left(q_{\mathfrak{r}} - q'_{\mathfrak{r}}\right)\right\rangle^{-5} \prod_{\mathfrak{r}\neq\mathfrak{n}\in\mathcal{N}} \langle T q_{\mathfrak{n}}\rangle^{-1} \left\langle T q'_{\mathfrak{n}}\right\rangle^{-1},$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

and it suffices to bound

$$\sum_{(k_{\mathfrak{l}}:\mathfrak{l}\in\mathcal{S})}\ \sum_{(k_{\mathfrak{l}}:\mathfrak{l}\in\mathcal{L}\backslash\mathcal{S})}^{**}\ \sum_{\left(k'_{\mathfrak{l}}:\mathfrak{l}\in\mathcal{L}\backslash\mathcal{S}\right)}^{**'} \max\left(\langle Tq_{\mathfrak{r}}\rangle,\left\langle Tq'_{\mathfrak{r}}\right\rangle\right)^{-2+2b}\left\langle T\left(q_{\mathfrak{r}}-q'_{\mathfrak{r}}\right)\right\rangle^{-5}\prod_{\mathfrak{r}\neq\mathfrak{n}\in\mathcal{N}}\langle Tq_{\mathfrak{n}}\rangle^{-1}\left\langle Tq'_{\mathfrak{n}}\right\rangle^{-1}.$$

Since all the qs are bounded by L^{θ}, and $T\leq L^{d}$, we may fix the integer parts of each $Tq_{\mathfrak{n}}$ and $Tq'_{\mathfrak{n}}$ for each $\mathfrak{n}\in\mathcal{N}$, and reduce the foregoing sum to a counting bound, at the price of losing a power $L^{C\left(b-\frac{1}{2}\right)}$. Now by definition (2.16), each $q_{\mathfrak{n}}$ is a linear combination of $\Omega_{\mathfrak{n}}$s, and conversely, each $\Omega_{\mathfrak{n}}$ is a linear combination of $q_{\mathfrak{n}}$s. So once the integer parts of each $Tq_{\mathfrak{n}}$ and $Tq'_{\mathfrak{n}}$ are fixed, we have also fixed $\sigma_{\mathfrak{n}}\in\mathbb{R}$ and $\sigma'_{\mathfrak{n}}\in\mathbb{R}$, such that

$$\left|\Omega_{\mathfrak{n}}-\sigma_{\mathfrak{n}}\right|\leq T^{-1},\qquad\left|\Omega'_{\mathfrak{n}}-\sigma'_{\mathfrak{n}}\right|\leq T^{-1}. \tag{3.18}$$

Therefore we are reduced to counting the number of $(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{S})$, $(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\backslash\mathcal{S})$ and $\left(k'_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\backslash\mathcal{S}\right)$ such that the assignments $(k_{\mathfrak{n}})$ and $\left(k'_{\mathfrak{n}}\right)$ are both strongly admissible and satisfy formula (3.18). Now let $|\mathcal{L}|=l=2n+1$ and p be the number of pairs; then $|\mathcal{S}|=2n+1-2p$. First we count the number of choices for $(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{S})$ and $(k_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\backslash\mathcal{S})$, where we apply Corollary 3.6 with $\mathcal{R}=\{\mathfrak{r}\}$ and get the factor $M:=L^{\theta}Q^{2n-p-2}L^{2d}T^{-1}$; then, with $k_{\mathfrak{I}}$ fixed for all $\mathfrak{I}\in\mathcal{S}$, we count the number of choices for $\left(k'_{\mathfrak{I}}:\mathfrak{I}\in\mathcal{L}\backslash\mathcal{S}\right)$ by applying Proposition 3.5 with $\mathcal{R}=\mathcal{S}\cup\{\mathfrak{r}\}$ and get the factor $M':=L^{\theta}Q^{p}$. In the end we have, L-certainly,

$$\begin{aligned}\sup_{k}\langle k\rangle^{4s}\left|\left(\widetilde{\mathcal{J}_{\mathcal{T}}}\right)_{k}(\tau)\right|^{2}&\leq L^{\theta+C\left(b-\frac{1}{2}\right)}\left(\frac{\alpha T}{L^{d}}\right)^{2n}MM'\\&\leq L^{\theta+C\left(b-\frac{1}{2}\right)}\left(\frac{\alpha T}{L^{d}}\right)^{2n}Q^{2n-2}L^{2d}T^{-1}=L^{\theta+C\left(b-\frac{1}{2}\right)}\rho^{2n-2}\left(\alpha^{2}T\right)\end{aligned}$$

by the definition of Q in formula (3.11), as desired. □

3.3. *Bounds for* $\mathcal{P}_{\pm}$

In this section we prove Proposition 2.6. The proofs for both $\mathcal{P}_{\pm}$ are similar, so we consider only $\mathcal{P}_{+}$.

Proof of Proposition 2.6. There are three steps.

Step 1: First reductions. We start with some simple observations. The operator $\mathcal{P}_{+}(v)=\mathcal{I}\mathcal{W}\left(\mathcal{J}_{\mathcal{T}_1},\mathcal{J}_{\mathcal{T}_2},v\right)$, where $\mathcal{I}$ and $\mathcal{W}$ are defined in formulas (2.6) and (2.7). Now in formula (2.7) we may assume $|k_1|,|k_2|\leq L^{\theta}$, for the same reason as in the proof of Proposition 2.5. We thus have

$$L^{-\theta}\leq\frac{\langle k\rangle^{s}}{\langle k_3\rangle^{s}}\leq L^{\theta},$$

so instead of $h^{s,b}$ bounds we only need to consider $h^{0,b}$ bounds. Next notice that if $\mathcal{I}$ is defined by equation (2.6) and $\mathcal{I}_1$ is defined by $\mathcal{I}_1F=\chi\cdot(\operatorname{sgn}*(\chi\cdot F))$, then we have the identity $2\mathcal{I}F(t)=\mathcal{I}_1F(t)-\chi(t)\mathcal{I}_1F(0)$, so for $b>\frac{1}{2}$ we have $\|\mathcal{I}F\|_{h^{s,b}}\lesssim\|\mathcal{I}_1F\|_{h^{s,b}}$. Therefore, in estimating $\mathcal{P}_{+}$ we may replace the operator G that appears in the formula for $\mathcal{I}$ by $\mathcal{I}_1$. The advantage is that $\mathcal{I}_1$ has a formula

$$\widetilde{\mathcal{I}_1F}(\tau)=\int_{\mathbb{R}}I_1(\tau,\sigma)\widetilde{F}(\sigma)\mathrm{d}\sigma,$$

where I_1 is as in Lemma 2.1, so we may get rid of the I_0 term. From now on we will stick to the renewed definition of $\mathcal{I}$. Next, by Proposition 2.5 we have the trivial bound

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

$$\|\mathcal{P}_+ v\|_{h^{0,1}} \sim \|\mathcal{P}_+ v\|_{\ell_k^2 L_t^2} + \|\partial_t \mathcal{P}_+ v\|_{\ell_k^2 L_t^2}$$
$$\lesssim \|v\|_{\ell_k^2 L_t^2} \cdot \frac{\alpha T}{L^d} \sum_{k_1,k_2} \left\| \left(\mathcal{J}_{\mathcal{T}_1}\right)_{k_1} \right\|_{L_t^\infty} \left\| \left(\mathcal{J}_{\mathcal{T}_2}\right)_{k_2} \right\|_{L_t^\infty} \lesssim \alpha T L^{d+\theta+C\left(b-\frac{1}{2}\right)} \rho^{n_1+n_2} \cdot \|v\|_{h^{0,0}} .$$

Note also that $\alpha T \le L^d$ and $\rho \le L^{-\varepsilon}$, so by interpolation it suffices to L-certainly bound the $h^{0,b} \to h^{0,1-b}$ norm of (the renewed version of) $\mathcal{P}_+$ by $L^\theta \rho^{n_1+n_2+1}$.

Now, using Lemma 2.1 and noticing that the bound (2.8) is symmetric in σ and τ, we have the formula

$$\left(\widetilde{\mathcal{P}_+ v}\right)_k (\tau) = \frac{i\alpha T}{L^d} \langle\tau\rangle^{-1} \sum_{(m_1,m_2,k')}^{*} \int_{\mathbb{R}^3} J\left(\tau, \sigma_1 - \sigma_2 + \tau' + T\Omega(m_1,m_2,k',k)\right)$$
$$\times \left(\widetilde{\mathcal{J}_{\mathcal{T}_1}}\right)_{m_1} (\sigma_1) \overline{\left(\widetilde{\mathcal{J}_{\mathcal{T}_2}}\right)_{m_2} (\sigma_2)} \cdot \widetilde{v}_{k'}(\tau') \mathrm{d}\sigma_1 \mathrm{d}\sigma_2 \mathrm{d}\tau', \tag{3.19}$$

where $J = J(\tau,\eta)$ and all its derivatives are bounded by $\langle\tau-\eta\rangle^{-10}$. By elementary estimates we have

$$\|\widetilde{w}_k(\tau)\|_{L_\tau^1 \ell_k^2} \lesssim \left\|\langle\tau\rangle^b \widetilde{w}_k(\tau)\right\|_{\ell_k^2 L_\tau^2}, \qquad \left\|\langle\tau\rangle^{1-b} \langle\tau\rangle^{-1} w_k(\tau)\right\|_{\ell_k^2 L_\tau^2} \lesssim \|\widetilde{w}_k(\tau)\|_{L_\tau^\infty \ell_k^2}, \tag{3.20}$$

and thus it suffices to L-certainly bound the $\ell^2 \to \ell^2$ norm of the operator

$$\mathcal{X} : (\mathcal{X}v)_k = \frac{\alpha T}{L^d} \sum_{(m_1,m_2,k')}^{*} v_{k'} \cdot \int_{\mathbb{R}^2} J\left(\tau, \sigma_1 - \sigma_2 + \tau' + T\Omega(m_1,m_2,k',k)\right)$$
$$\times \left(\widetilde{\mathcal{J}_{\mathcal{T}_1}}\right)_{m_1} (\sigma_1) \overline{\left(\widetilde{\mathcal{J}_{\mathcal{T}_2}}\right)_{m_2} (\sigma_2)} \mathrm{d}\sigma_1 \mathrm{d}\sigma_2 \tag{3.21}$$

uniformly in τ and τ'.

Step 2: Second reductions. At this point we apply similar arguments as in the proof of Proposition 2.5. Namely, we first restrict $|\tau|, |\tau'| \le L^{\theta^{-1}}$ (otherwise we can gain a power of either $|\tau|^{\frac{1}{2}\left(b-\frac{1}{2}\right)}$ or $|\tau'|^{\frac{1}{2}\left(b-\frac{1}{2}\right)}$ from the extra room when applying formula (3.20), which turns into a large power of L and closes the whole estimate), and then divide this interval into subintervals of length $L^{-\theta^{-1}}$ and apply differentiability to reduce to $O\left(L^{C\theta^{-1}}\right)$ choices of (τ,τ'). Therefore, it suffices to *fix* τ and τ' and L-certainly bound $\|\mathcal{X}\|_{\ell^2\to\ell^2}$. Let $\tau - \tau' = \zeta$ be fixed.

Now use equation (2.19) for the $\mathcal{J}_{\mathcal{T}_j}$ factors, assuming also $|k_{\mathfrak{l}}| \le L^\theta$ in each tree, and integrate in (σ_1,σ_2). This leads to further reduced expression for $\mathcal{X}$, which can be described as follows. First let the tree $\mathcal{T}$ be defined such that its root is $\mathfrak{r}$ and three subtrees from left to right are $\mathcal{T}_1$, $\mathcal{T}_2$ and a single node $\mathfrak{r}'$. Then we have

$$(\mathcal{X}v)_k = \sum_{k'} \mathcal{X}_{kk'} v_{k'},$$

where the matrix coefficients are given by

$$\mathcal{X}_{kk'} = \left(\frac{\alpha T}{L^d}\right)^{n_1+n_2+1} \sum_{(k_{\mathfrak{n}}:\mathfrak{n}\in\mathcal{T})} \mathcal{K}\left(|k|_\beta^2 - |k'|_\beta^2,\ k_{\mathfrak{l}} : \mathfrak{r}' \ne \mathfrak{l} \in \mathcal{L}\right) \cdot \frac{1}{\langle Tq_{\mathfrak{r}} - \zeta\rangle^5} \prod_{\mathfrak{n}\in\mathcal{N}\setminus\{\mathfrak{r}\}} \frac{1}{\langle Tq_{\mathfrak{n}}\rangle} \prod_{\mathfrak{l}\in\mathcal{L}\setminus\{\mathfrak{r}'\}} \eta_{k_{\mathfrak{l}}}^{\iota_{\mathfrak{l}}},$$

where the sum is taken over all admissible assignments $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ which satisfy $k_{\mathfrak{r}} = k$, $k_{\mathfrak{r}'} = k'$ and $|k_{\mathfrak{n}}| \le L^\theta$ for $\mathfrak{n} \notin \{\mathfrak{r},\mathfrak{r}'\}$, and the coefficient satisfies $|\mathcal{K}| \le L^\theta$ and $|\partial\mathcal{K}| \le L^\theta T$. Moreover, we observe that $\mathcal{K}$ and $q_{\mathfrak{r}}$ depend on the variables $k_{\mathfrak{r}} = k$ and $k_{\mathfrak{r}'} = k'$ *only through the quantity* $|k|_\beta^2 - |k'|_\beta^2$.

Next we argue in the same way as in the proof of Proposition 2.5 and fix the integer parts of $Tq_{\mathfrak{n}}$ for $\mathfrak{n} \in \mathcal{N}\setminus\{\mathfrak{r}\}$, as well as the integer part of $Tq_{\mathfrak{r}} - \zeta$, at a cost of $(\log L)^{O(1)}$. All these can be assumed to

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

be $\le L^{\theta^{-1}}$ due to the decay $\langle Tq_{\mathfrak{r}} - \zeta\rangle^{-5}$ and the bounds on τ and τ'. This is equivalent to fixing some real numbers $\sigma_{\mathfrak{n}} = O\left(L^{\theta^{-1}}\right)$ and requiring the assignment $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T})$ to satisfy $|\Omega_{\mathfrak{n}} - \sigma_{\mathfrak{n}}| \le T^{-1}$ for each $\mathfrak{n} \in \mathcal{N}$. Let this final operator, obtained by all the previous reductions, be $\mathcal{G}$. Schematically, the operator $\mathcal{G}$ can be viewed as 'attaching' two trees $\mathcal{T}_1$ and $\mathcal{T}_2$ to a single node $\mathfrak{r}'$.

Step 3: The high-order $\mathcal{G}\mathcal{G}^*$ argument. For this, we consider the adjoint operator $\mathcal{G}^*$. A similar argument gives a formula for $\mathcal{G}^*$, which is associated with a tree $\mathcal{T}^*$ formed by attaching the two trees $\mathcal{T}_2$ and $\mathcal{T}_1$ (with $\mathcal{T}_2$ on the left of $\mathcal{T}_1$) to a single node $\mathfrak{r}'$, in the same way that $\mathcal{G}$ is associated with $\mathcal{T}$. Given a large positive integer D, we will consider $(\mathcal{G}\mathcal{G}^*)^D$, which is associated with a tree $\mathcal{T}^D$. The precise description is as follows.

First, $\mathcal{T}^D$ is a tree with root node $\mathfrak{r}_0 = \mathfrak{r}$, and its first two subtrees (from the left) are $\mathcal{T}_1$ and $\mathcal{T}_2$. The third subtree has root $\mathfrak{r}_1$, and its first two subtrees (from the left) are $\mathcal{T}_2$ and $\mathcal{T}_1$. The third subtree has root $\mathfrak{r}_2$, and its first two subtrees (from the left) are $\mathcal{T}_1$ and $\mathcal{T}_2$, and so on. This process repeats and eventually stops at $\mathfrak{r}_{2D} = \mathfrak{r}'$, which is a single node, finishing the construction of $\mathcal{T}^D$. As usual, denote by $\mathcal{L}^D$ and $\mathcal{N}^D$ the set of leaves and branching nodes, respectively. Then the kernel of $(\mathcal{G}\mathcal{G}^*)^D$ is given by

$$\left((\mathcal{G}\mathcal{G}^*)^D\right)_{kk'} = \sum_{(k_{\mathfrak{n}}:\mathfrak{n}\in\mathcal{T}^D)} \mathcal{K}^{(D)}\left(\left|k_{\mathfrak{r}_j}\right|_\beta^2 - \left|k_{\mathfrak{r}_{j+1}}\right|_\beta^2 : 0 \le j \le 2D-1,\ k_{\mathfrak{l}} : \mathfrak{r}' \ne \mathfrak{l} \in \mathcal{L}^D\right) \times \left(\frac{\alpha T}{L^d}\right)^{2D(n_1+n_2+1)} \prod_{\mathfrak{r}'\ne\mathfrak{l}\in\mathcal{L}^D} \eta_{k_{\mathfrak{l}}}^{\iota_{\mathfrak{l}}}, \tag{3.22}$$

where $\left|\mathcal{K}^{(D)}\right| \le L^\theta$ and $\left|\partial\mathcal{K}^{(D)}\right| \le L^\theta T$, and the sum is taken over all admissible assignments $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}^D)$ that satisfy $(k_{\mathfrak{r}}, k_{\mathfrak{r}'}) = (k, k')$, $|k_{\mathfrak{l}}| \le L^\theta$ for $\mathfrak{r}' \ne \mathfrak{l} \in \mathcal{L}$ and $|\Omega_{\mathfrak{n}} - \sigma_{\mathfrak{n}}| \le T^{-1}$ for $\mathfrak{n} \in \mathcal{N}^D$, where $\sigma_{\mathfrak{n}} = O\left(L^{\theta^{-1}}\right)$ are fixed. Moreover, $\mathcal{K}^{(D)}$ depends on the variables $k_{\mathfrak{r}} = k$ and $k_{\mathfrak{r}'} = k'$ *only through the quantities* $\left|k_{\mathfrak{r}_j}\right|_\beta^2 - \left|k_{\mathfrak{r}_{j+1}}\right|_\beta^2$ for $0 \le j \le 2D-1$.

Now note that each $\left((\mathcal{G}\mathcal{G}^*)^D\right)_{kk'}$ is an explicit multilinear Gaussian expression. Since for fixed k (or k') the number of choices for k' (or k) is $O\left(L^{d+\theta}\right)$, by Schur's estimate we know

$$\left\|(\mathcal{G}\mathcal{G}^*)^D\right\|_{\ell^2\to\ell^2} \lesssim L^{d+\theta} \sup_{k,k'} \left|\left((\mathcal{G}\mathcal{G}^*)^D\right)_{kk'}\right|.$$

So it suffices to L-certainly bound $\left|\left((\mathcal{G}\mathcal{G}^*)^D\right)_{kk'}\right|$ *uniformly* in k and k'. We first consider this estimate with fixed (k, k'). Applying Lemma 3.1, we can fix some pairings of $\mathcal{T}^D$ and the set $\mathcal{S}^D$ of single leaves, and argue as in the proof of Proposition 2.5 to conclude L-certainly that

$$\left|\left((\mathcal{G}\mathcal{G}^*)^D\right)_{kk'}\right|^2 \lesssim L^\theta \left(\frac{\alpha T}{L^d}\right)^{4D(n_1+n_2+1)} \sum_{(k_{\mathfrak{l}}:\mathfrak{l}\in\mathcal{S}^D)} \sideset{}{^{**}}\sum_{(k_{\mathfrak{l}}:\mathfrak{l}\in\mathcal{L}^D\setminus\mathcal{S}^D)} \sideset{}{^{**'}}\sum_{(k_{\mathfrak{l}'}:\mathfrak{l}\in\mathcal{L}^D\setminus\mathcal{S}^D)} 1,$$

where the condition for summation, as in the proof of Proposition 2.5, is that the unique admissible assignment $(k_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}^D)$ determined by $(k_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{S}^D)$ and $(k_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{L}^D\setminus\mathcal{S}^D)$ satisfies all the conditions already listed, and the same happens for $(k'_{\mathfrak{n}} : \mathfrak{n} \in \mathcal{T}^D)$ corresponding to $(k_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{S}^D)$ and $\left(k'_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{L}^D\setminus\mathcal{S}^D\right)$. We know that $\mathcal{T}^D$ is a tree of scale $2D(n_1+n_2+1)$,, and so $\left|\mathcal{L}^D\right| = 4D(n_1+n_2+1)+1$; let the number of pairings be p, and then $\left|\mathcal{S}^D\right| = 4D(n_1+n_2+1) - 2p$. By Proposition 3.5 we can bound the number of choices[12] for $(k_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{S}^D)$ and $\left(k'_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{L}^D\setminus\mathcal{S}^D\right)$ by $M = L^\theta Q^{4D(n_1+n_2+1)-p}$, and bound the number of choices for $\left(k'_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{L}^D\setminus\mathcal{S}^D\right)$ given $(k_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{S}^D)$ by $M' = L^\theta Q^p$. In the end, for any

[12]Strictly speaking, we need to modify Proposition 3.5 a little, as we do not assume $|k'| \le L^\theta$. But this will not affect the proof, which relies on the translation-invariant inequality (3.11).

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

fixed (k, k'), we have that L-certainly,

$$L^{d+\theta} \sup_{k,k'} \left| \left((\mathcal{G}\mathcal{G}^*)^D \right)_{kk'} \right| \leq L^{d+\theta} \left(\frac{\alpha T}{L^d} \right)^{2D(n_1+n_2+1)} (MM')^{1/2} \leq L^{d+\theta} \rho^{2D(n_1+n_2+1)}.$$

Finally we need to L-certainly make this bound uniform in all choices of (k, k'). This is not obvious, since we impose no upper bound on $|k|$ and $|k'|$, so the number of exceptional sets we remove in the L-certain condition could presumably be infinite. However, note that the coefficient $\mathcal{K}$ depends on k and k' only through the quantities $|k|_\beta^2 - |k_{\mathfrak{r}_j}|_\beta^2$. Let $\mathcal{D} = \mathcal{L} \backslash \{\mathfrak{r}'\}$; then $|k_{\mathfrak{l}}| \leq L^\theta$ for $\mathfrak{l} \in \mathcal{D}$, and the condition for summation creates the restriction that $\left| |k|_\beta^2 - |k_{\mathfrak{r}_j}|_\beta^2 \right| \leq L^{\theta^{-1}}$. The reduction from infinitely many possibilities for k (and hence k') to finitely many is done by invoking the following result, whose proof will be left to the end:

Claim 3.7. *Let* $k \in \mathbb{Z}_L^d$, *and consider the function*

$$f_{(k)} : m \mapsto |k|_\beta^2 - |k+m|_\beta^2, \qquad \mathrm{Dom}\left(f_{(k)}\right) = \left\{ m \in \mathbb{Z}_L^d : |m| \leq L^\theta, \left| |k|_\beta^2 - |k+m|_\beta^2 \right| \leq L^{\theta^{-1}} \right\}.$$

Then there exist finitely many functions $f_1, \ldots, f_A$, *where* $A \leq L^{C\theta^{-1}}$, *such that for any* $k \in \mathbb{Z}_L^d$ *there exists* $1 \leq j \leq A$ *such that* $\left| f_{(k)} - f_j \right| \leq L^{-\theta^{-1}}$ *on* $\mathrm{Dom}\left(f_{(k)}\right)$.

Remark 3.8. We may view Claim 3.7 as a 'finiteness' or 'compactness' lemma. Similar results are also used in [18] and [19] for similar purposes.

Now it is not hard to see that Claim 3.7 allows us to obtain a bound of the form proved that is uniform in (k, k'), after removing at most $O\left(L^{C\theta^{-1}}\right)$ exceptional sets, each with probability $\lesssim e^{-L^\theta}$. This then implies

$$\left\| (\mathcal{G}\mathcal{G}^*)^D \right\|_{\ell^2 \to \ell^2} \lesssim L^{d+\theta} \rho^{2D(n_1+n_2+1)},$$

hence

$$\|\mathcal{G}\|_{\ell^2 \to \ell^2} \lesssim L^{\frac{d+\theta}{2D}} \rho^{n_1+n_2+1}.$$

By fixing D to be a sufficiently large positive integer, we deduce the correct operator bound for $\mathcal{G}$, and hence for $\mathcal{X}$ and $\mathcal{P}_+$. This completes the proof of Proposition 2.6. □

Proof of Claim 3.7. We will prove the result for any linear function $g(m) = x \cdot m + X$, where $x \in \mathbb{R}^d$ and $X \in \mathbb{R}$ are arbitrary. We may also assume $m \in \mathbb{Z}^d$ instead of $\mathbb{Z}_L^d$; the domain $\mathrm{Dom}(g)$ will then be the set E of m such that $|m| \leq L^{1+\theta}$ and $|g(m)| \leq L^{2+\theta^{-1}}$.

Let the affine dimension $\dim(E) = r \leq d$; then E contains a maximal affine independent set $\{q_j : 0 \leq j \leq r\}$. The number of choices for these q_j is at most L^{d+1}, so we may fix them. Let $\mathcal{L}$ be the primitive lattice generated by $\{q_j - q_0 : 1 \leq j \leq r\}$, and fix a reduced basis $\{\ell_j : 1 \leq j \leq r\}$ of $\mathcal{L}$. For any $m \in E$ there is a unique integer vector $k = (k_1, \ldots, k_r) \in \mathbb{Z}^r$ such that $|k| \lesssim L^{1+\theta}$, $m - q_0 = k_1\ell_1 + \cdots + k_r\ell_r$, and as a linear function we can write $g(m) = y \cdot k + Y$, where $y \in \mathbb{R}^r$ and $Y = g(q_0) \in \mathbb{R}$.

Now let the $k \in \mathbb{Z}^r$ corresponding to $m = q_j$ be $k^{(j)}$, where $1 \leq j \leq r$; then since $q_0 \in E$ and $q_j \in E$, we conclude that $\left| y \cdot k^{(j)} \right| \leq L^{3+\theta^{-1}}$. As the $k^{(j)}$ are linear independent integer vectors in $\mathbb{Z}^r$ with norm bounded by $L^{1+\theta}$, we conclude that $|y| \leq L^{C+\theta^{-1}}$, and consequently $|Y| \leq L^{C+\theta^{-1}}$. We may then approximate $g(m)$ for $m \in E$ by $y_j \cdot k + Y_j$, where y_j and Y_j are one of the $L^{C\theta^{-1}}$ choices that approximate y and Y up to error $L^{-\theta^{-1}}$, and choose $g_{(j)} = y_j \cdot k + Y_j$. □

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

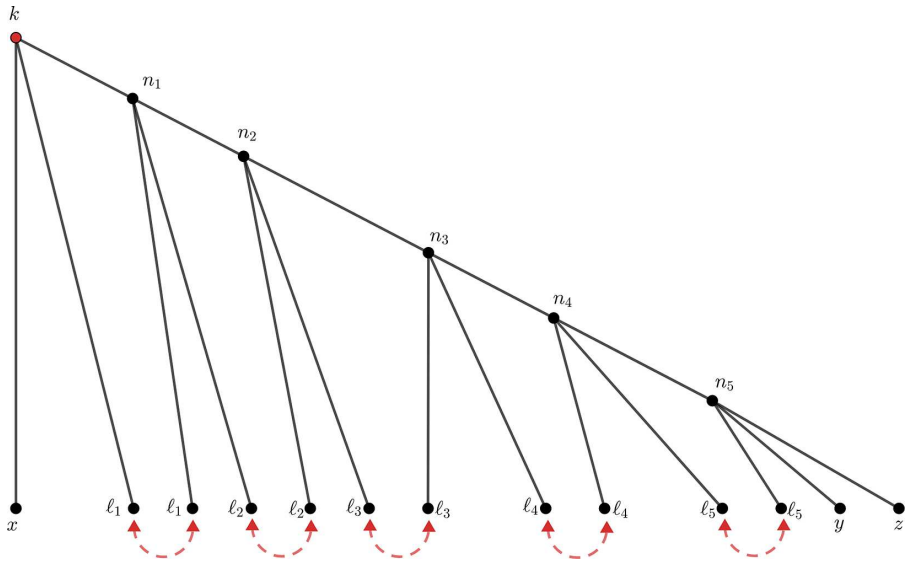

Figure 5. *A tree of scale* $\mathfrak{s}(\mathcal{T}) = 6$ *and* $p = 6 - 1 = 5$ *pairings. The pairings force* $|y - z| = |n_5 - \ell_5| = |n_4 - \ell_4| = \cdots = |k - x|$.

3.4. The worst terms

In this section we exhibit terms $\mathcal{J}_{\mathcal{T}}$ that satisfy the lower bound (1.9). These are the terms corresponding to trees $\mathcal{T}$ and pairings (see Remark 3.9) as shown in Figure 5, where $\mathcal{T}$ is formed from a single node by successively attaching two leaf nodes, and the 'left' node attached at each step is paired with the 'right' node attached in the next step. Let the scale $\mathfrak{s}(\mathcal{T}) = r$; then $\mathcal{T}$ has exactly $r - 1$ pairings. For simplicity we will consider the rational case $\beta_j = 1$ and $T \le L^{2-\delta}$; the irrational case is similar.

Here it is more convenient to work with the time variable t (instead of its Fourier dual τ). To show formula (1.9), since $b > 1/2$, we just need to bound $(\mathcal{J}_{\mathcal{T}})_k(t)$ from below for some k and some $t \in [0, 1]$; moreover, since $\chi \equiv 1$ on $[0, 1]$, and using the recursive definition (2.17), we can write

$$(\mathcal{J}_{\mathcal{T}})_k(t) = \left(\frac{\alpha T}{L^d}\right)^r \sum_{x-y+z=k} \left(\sum_{\ell_1,\dots,\ell_{r-1}} \mathcal{B} \cdot \prod_{j=1}^{r-1} n_{\mathrm{in}}(\ell_j) \left|\eta_{\ell_j}\right|^2 \right) \cdot \eta_x \overline{\eta_y} \eta_z \cdot \sqrt{n_{\mathrm{in}}(x) n_{\mathrm{in}}(y) n_{\mathrm{in}}(z)}, \tag{3.23}$$

where (due to admissibility) the variables in the summation satisfy

$$k - x = n_1 - \ell_1 = n_2 - \ell_2 = \cdots = n_{r-1} - \ell_{r-1} = y - z := q$$

and the coefficient $\mathcal{B}$ is given by

$$\mathcal{B} = \int_{t > t_1 > \cdots > t_r > 0} e^{2\pi i T (t_1 \Omega_1 + \cdots + t_r \Omega_r)} \mathrm{d}t_1 \cdots \mathrm{d}t_r, \tag{3.24}$$

with Ω_j being the resonance factors, namely

$$\Omega_1 = 2q \cdot (k - n_1), \Omega_2 = 2q \cdot (n_1 - n_2), \dots, \Omega_{r-1} = 2q \cdot (n_{r-2} - n_{r-1}), \Omega_r = 2q \cdot (n_{r-1} - z).$$

In equation (3.23) we may replace $\left|\eta_{\ell_j}\right|^2$ by 1, so the factor in the big parentheses, which we denote by $\mathcal{A}_{kxyz}$, involves no randomness. Therefore, with high probability,

$$|(\mathcal{J}_{\mathcal{T}})_k(t)|^2 \sim \left(\frac{\alpha T}{L^d}\right)^{2r} \sum_{x-y+z=k} \left|\mathcal{A}_{kxyz}\right|^2.$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

In the sum, we may fix $q \in \mathbb{Z}_L^d$ with $0 < |q| \lesssim L^{-1}$, which has $O(1)$ choices, and write

$$\mathcal{A}_{kxyz} = \int_{t>t_1>\cdots>t_r>0} e^{4\pi i T q\cdot[t_1(k-q)+t_r(q-z)]} \times \left[\sum_{\ell_1,\ldots,\ell_{r-1}} e^{4\pi i T[(t_2-t_1)q\cdot\ell_1+\cdots+(t_r-t_{r-1})q\cdot\ell_{r-1}]}\prod_{j=1}^{r-1} n_{\text{in}}(\ell_j)\right] \mathrm{d}t_1\cdots\mathrm{d}t_r.$$

By Poisson summation, and noticing that $|Tt_jq| \lesssim L^{1-\delta}$, we conclude that up to constants,

$$\sum_{\ell_1,\ldots,\ell_{r-1}} e^{4\pi i T[(t_2-t_1)q\cdot\ell_1+\cdots+(t_r-t_{r-1})q\cdot\ell_{r-1}]}\prod_{j=1}^{r-1} n_{\text{in}}(\ell_j) = L^{(r-1)d}\prod_{j=1}^{r-1}\widehat{n_{\text{in}}}\left(Tq\left(t_{j+1}-t_j\right)\right) + O\left(L^{-\infty}\right).$$

By making change of variables $s_j = t_j - t_{j+1} (1 \le j \le r-1)$ and $s_0 = t - t_1$, $s_r = t_r$, we can reduce to

$$\mathcal{A}_{kxyz} \approx L^{(r-1)d}\int_{s_0+\cdots+s_r=t} e^{4\pi i T[(t-s_0)q\cdot(k-q)+s_rq\cdot(q-z)]}\prod_{j=1}^{r-1}\widehat{n_{\text{in}}}\left(Tqs_j\right)\mathrm{d}s_1\cdots\mathrm{d}s_r.$$

By choosing some particular (k,q,z), we may assume $q\cdot(k-q) = q\cdot(q-z) = 0$, and if we also choose n_{in} such that $\widehat{n_{\text{in}}}$ is positive, say $n_{\text{in}}(k) = e^{-|k|^2}$, and $t = \min\left(1, LT^{-1}\right)$, then we have

$$\left|\mathcal{A}_{kxyz}\right| \sim L^{(r-1)d}\min\left(1, LT^{-1}\right)^r,$$

and hence, with high probability,

$$\sup_{k,t}\left|(\mathcal{J}_{\mathcal{T}})_k(t)\right| \gtrsim L^{-d}\min(\alpha T, \alpha L)^r = L^{-d}\rho^r$$

for any fixed r – thus formula (1.9).

Remark 3.9. Here, strictly speaking, we are further decomposing $\mathcal{J}_{\mathcal{T}}$ into the sum of terms $\mathcal{J}_{\mathcal{T},\mathcal{P}}$, where $\mathcal{P}$ represents the pairing structure of $\mathcal{T}$. In the proof of Proposition 2.5, we are actually making the same decomposition (by identifying the set of pairings) and proving the same bound for each $\mathcal{J}_{\mathcal{T},\mathcal{P}}$. On the other hand, the example here shows that individual terms $\mathcal{J}_{\mathcal{T},\mathcal{P}}$ can be very large in absolute value. Thus to get any improvement on the results of this paper, one would need to explore the subtle cancellations between the $\mathcal{J}_{\mathcal{T},\mathcal{P}}$ terms with different $\mathcal{T}$ or different $\mathcal{P}$.

4. Proof of the main theorem

In this section we prove Theorem 1.3 (which also implies Theorem 1.1). Since we may alter the value of T, in proving Theorem 1.3 we may restrict to the case $T/2 \le t \le T$.

First note that $\mathbb{E}\left|\widehat{u}(k,t)\right|^2 = \mathbb{E}\left|a_k(s)\right|^2$, where $s := \frac{t}{T} \in [1/2, 1]$. By mass conservation, we have $L^{-d/2}\sum_{k\in\mathbb{Z}_L^d}|a_k|^2 = O(1)$ and hence $\|a_k\|_{\ell^\infty} \lesssim L^{d/2}$. Therefore, if we denote by Γ the intersection of all the L-certain events in Propositions 2.4 and 2.5, we have, for $0 \le s \le 1$ (denoting by $\mathbb{E}_\Gamma G = \mathbb{E}\mathbf{1}_\Gamma G$),

$$\begin{aligned}\mathbb{E}\left|\widehat{u}(k,Ts)\right|^2 = \mathbb{E}_\Gamma\Big[&|(\mathcal{J}_0)_k(s)|^2 + |(\mathcal{J}_1)_k(s)|^2 + 2\mathrm{Re}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_1)_k(s) + 2\mathrm{Re}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_2)_k(s)\Big]\\ &+ \sum_{3\le n\le N} 2\mathbb{E}_\Gamma\mathrm{Re}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_n)_k(s) + \sum_{1\le n_1,n_2\le N;n_1+n_2\ge 3}\mathbb{E}_\Gamma\overline{\left(\mathcal{J}_{n_1}\right)_k(s)}\left(\mathcal{J}_{n_2}\right)_k(s)\\ &+ \sum_{n\le N} 2\mathbb{E}_\Gamma\mathrm{Re}\overline{(\mathcal{R}_{N+1})_k(s)}(\mathcal{J}_n)_k(s) + \mathbb{E}_\Gamma\left|(\mathcal{R}_{N+1})_k(s)\right|^2 + O\left(e^{-L^\theta}\right). \qquad (4.1)\end{aligned}$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

By using Proposition 2.4, we can bound the last three terms by

$$\left|\mathbb{E}_\Gamma \overline{\left(\mathcal{J}_{n_1}\right)_k(s)} \left(\mathcal{J}_{n_2}\right)_k(s)\right| \le L^{\theta+c(b-1/2)}\rho^{n_1+n_2-2}\left(\alpha^2 T\right) \le L^{-\delta}/10\frac{T}{T_{\mathrm{kin}}},$$
$$\left|\mathbb{E}_\Gamma \mathrm{Re}\overline{(\mathcal{R}_{N+1})_k(s)}(\mathcal{J}_n)_k(s)\right| + \mathbb{E}_\Gamma\left|(\mathcal{R}_{N+1})_k(s)\right|^2 \le L^{\theta+C(b-1/2)}\rho^N \le L^{-10d}.$$

As with the first term on the second line of equation (4.1), since $(\mathcal{J}_0)_k(s) = \chi(t)\sqrt{n_{\mathrm{in}}}\cdot\eta_k(\omega)$, by direct calculations and similar arguments as in the proof of Proposition 2.5 we can bound, for any tree $\mathcal{T}$ with $\mathfrak{s}(\mathcal{T}) = n$,

$$\left|\mathbb{E}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_\mathcal{T})_k(s)\right| \le L^\theta\left(\frac{\alpha T}{L^d}\right)^n M,$$

where M is the quantity estimated in Proposition 3.5 (i.e., the number of strongly admissible assignments satisfying formula (3.14)), with all but one leaf of $\mathcal{T}$ being paired, and $\mathcal{R} = \{\mathfrak{r}\}$. By Corollary 3.6 we have

$$\left|\mathbb{E}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_\mathcal{T})_k(s)\right| \le L^\theta\left(\frac{\alpha T}{L^d}\right)^n Q^{n-2}L^{2d}T^{-1} \le L^\theta\rho^{n-2}\left(\alpha^2 T\right) \le L^{-\delta}/10\frac{T}{T_{\mathrm{kin}}}.$$

It then suffices to calculate the main term, which is the first line of equation (4.1). Up to an error of size $O\left(e^{-L^\theta}\right)$, we can replace $\mathbb{E}_\Gamma$ by $\mathbb{E}$; also, we can easily show that $\mathrm{Re}\,\mathbb{E}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_1)_k(s) = 0$. For $|s| \le 1$, clearly $\mathbb{E}\left|(\mathcal{J}_0)_k(s)\right|^2 = n_{\mathrm{in}}$; as for the other two terms, namely $\mathbb{E}\left|(\mathcal{J}_1)_k(s)\right|^2$ and $2\mathbb{E}\mathrm{Re}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_2)_k(s)$, we compute as follows: Recall that $(a_{\mathrm{in}})_k = \sqrt{n_{\mathrm{in}}(k)}\eta_k(\omega)$ and

$$(\mathcal{J}_1)_k(s) = -\frac{\alpha T}{L^d}\left[\sum_{(k_1,k_2,k_3);\Omega\neq 0}^{\times}(a_{\mathrm{in}})_{k_1}\overline{(a_{\mathrm{in}})_{k_2}}(a_{\mathrm{in}})_{k_3}\frac{e^{2\pi i T\Omega s}-1}{2\pi T\Omega}\right.$$
$$\left. + is\sum_{(k_1,k_2,k_3);\Omega=0}^{\times}(a_{\mathrm{in}})_{k_1}\overline{(a_{\mathrm{in}})_{k_2}}(a_{\mathrm{in}})_{k_3} - is\left|(a_{\mathrm{in}})_k\right|^2(a_{\mathrm{in}})_k\right],$$

and therefore we have

$$\mathbb{E}\left|(\mathcal{J}_1)_k(t)\right|^2 = \frac{\alpha^2 s^2 T^2}{L^{2d}}\left[\sum_{(k_1,k_2,k_3);\Omega\neq 0}^{\times} 2(n_{\mathrm{in}})_{k_1}\overline{(n_{\mathrm{in}})_{k_2}}(n_{\mathrm{in}})_{k_3}\left|\frac{\sin\pi\Omega T s}{\pi\Omega T s}\right|^2\right.$$
$$\left. + \sum_{(k_1,k_2,k_3);\Omega=0}^{\times}(n_{\mathrm{in}})_{k_1}\overline{(n_{\mathrm{in}})_{k_2}}(n_{\mathrm{in}})_{k_3} + \left|(n_{\mathrm{in}})_k\right|^2(n_{\mathrm{in}})_k\right]$$
$$= 2\frac{\alpha^2 t^2}{L^{2d}}\sum_{(k_1,k_2,k_3);\Omega\neq 0}^{\times}(n_{\mathrm{in}})_{k_1}\overline{(n_{\mathrm{in}})_{k_2}}(n_{\mathrm{in}})_{k_3}\left|\frac{\sin\pi\Omega t}{\pi\Omega t}\right|^2 + O\left(\frac{T}{T_{\mathrm{kin}}}L^{-\delta}\right),$$

where we used $T < L^{2d-\delta}$ for the third term and estimated the second term by $L^{2d-2+\theta}$ for general β_j and by $L^{d+\theta}$ if β_j are irrational (e.g., using Lemma 3.2 with $m = 0$ and $T = L^2$ and L^d, respectively).

A similar computation for $2\mathbb{E}\mathrm{Re}\overline{(\mathcal{J}_0)_k(s)}(\mathcal{J}_2)_k(s)$ (see [7]) gives

$$\mathbb{E}\left[\left|(\mathcal{J}_1)_k(t)\right|^2 + 2\mathrm{Re}\overline{(\mathcal{J}_0)_k(t)}(\mathcal{J}_2)_k(t)\right] = \frac{2\alpha^2 t^2}{L^{2d}}\cdot\mathcal{S}_t(n_{\mathrm{in}}) + O\left(\frac{T}{T_{\mathrm{kin}}}L^{-\delta}\right),$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

where

$$\mathcal{S}_t(\phi) := \sum_{\substack{k_i \in \mathbb{Z}_L^d \\ k-k_1+k_2-k_3=0}} \phi_k \phi_{k_1} \phi_{k_2} \phi_{k_3} \left[\frac{1}{\phi_k} - \frac{1}{\phi_{k_1}} + \frac{1}{\phi_{k_2}} - \frac{1}{\phi_{k_3}} \right] \left| \frac{\sin\left(\pi t \Omega\left(\vec{k}\right)\right)}{\pi t \Omega\left(\vec{k}\right)} \right|^2, \tag{4.2}$$

with $\Omega\left(\vec{k}\right) = \Omega(k, k_1, k_2, k_3) = |k_1|_\beta^2 - |k_2|_\beta^2 + |k_3|_\beta^2 - |k|_\beta^2$. Therefore, we conclude that

$$\mathbb{E}\,|\widehat{u}(k,t)|^2 = n_{\text{in}} + \frac{2\alpha^2 t^2}{L^{2d}} \mathcal{S}_t(n_{\text{in}}) + O\left(\frac{T}{T_{\text{kin}}} L^{-\delta/10}\right).$$

In the following section, we derive the asymptotic formula for the sum $\mathcal{S}_t$ – namely, we show that $\mathcal{S}_t(\phi) = \mathscr{K}_t(\phi) + O\left(t^{-1} L^{2d-\theta}\right)$ for some $\theta > 0$, where $\mathscr{K}_t$ is given by

$$\begin{aligned} \mathscr{K}_t(\phi) := L^{2d} \int_{\xi_1 - \xi_2 + \xi_3 = \xi} & \phi(\xi)\phi(\xi_1)\phi(\xi_2)\phi(\xi_3) \left[\frac{1}{\phi(\xi)} - \frac{1}{\phi(\xi_1)} + \frac{1}{\phi(\xi_2)} - \frac{1}{\phi(\xi_3)} \right] \\ & \times \left| \frac{\sin\left(\pi t \Omega\left(\vec{\xi}\right)\right)}{\pi t \Omega\left(\vec{\xi}\right)} \right|^2 d\xi_1 d\xi_2 d\xi_3. \end{aligned} \tag{4.3}$$

Finally, the proof is complete by using the fact that for a smooth function f,

$$t \int \left| \frac{\sin(\pi t x)}{\pi t x} \right|^2 f(x) \mathrm{d}x = f(0) + O\left(t^{-1}\right). \qquad \square$$

5. Number-theoretic results

The purpose of this section is to prove the asymptotic formula for $\mathcal{S}_t$ defined in formula (4.2). The sum $\mathcal{S}_t$ should be regarded as a Riemann sum that approximates the integral $\mathscr{K}_t$ in formula (4.3). However, this approximation is far from trivial, because of the highly oscillating factor $\left| \frac{\sin\left(\pi t \Omega\left(\vec{\xi}\right)\right)}{\pi t \Omega\left(\vec{\xi}\right)} \right|^2$, which makes the problem intimately related to the equidistribution properties of the values of the quadratic form Ω.

Theorem 5.1. *Set $\phi \in \mathcal{S}\left(\mathbb{R}^d\right)$ with $d \geq 3$. For any $\delta > 0$, there exists $\theta > 0$ such that the asymptotic holds:*

1. *(General tori) For any $\beta_i \in [1,2]^d$ and any $t < L^{2-\delta}$,*

$$\mathcal{S}_t = \mathscr{K}_t + O\left(L^{2d-\theta} t^{-1}\right).$$

2. *(Generic tori) For generic $\beta_i \in [1,2]^d$ and any $t < L^{d-\delta}$,*

$$\mathcal{S}_t = \mathscr{K}_t + O\left(L^{2d-\theta} t^{-1}\right).$$

It is not hard to see that $\mathscr{K}_t = O\left(L^{2d}\right) t^{-1}$, which justifies the sufficiency of the error-term bound.

Remark 5.2. It is interesting that in the case of the rational torus for which $\beta_j = 1$, this asymptotic ceases to be true at the end point $t = L^2$. This corresponds to $\mu = 1$ in formula (5.1), whose asymptotic was studied in [24, 6] and yields a logarithmic divergence when $d = 2$ and a different multiplicative constant for $d \geq 3$ compared to the asymptotic in our theorem.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Proof of Theorem 5.1. The proof of part (2) is contained in [7], we will focus only on the first part, which is less sophisticated. To simplify the notation, we will drop the subscript t from $\mathcal{S}_t$ and $\mathscr{K}_t$. We use a refinement of [7] which basically covers the case $t < L^{1-\delta}$. First, observe that $\Omega\left(\vec{k}\right) = -2\mathcal{Q}(k_1-k, k_3-k)$, where $\mathcal{Q}(x,y) := \sum_{j=1}^{d} \beta_j x_j y_j$. Therefore, changing variables $N_1 = L(k_1-k) \in \mathbb{Z}^d$ and $N_2 = L(k_3-k) \in \mathbb{Z}^d$, we write the sum $\mathcal{S}$ in the form

$$\mathcal{S} = \sum_{N=(N_1,N_2)\in\mathbb{Z}^{2d}} W\left(\frac{N}{L}\right) g(4\mu\mathcal{Q}(N)), \quad g(x) := \left|\frac{\sin \pi x}{\pi x}\right|^2, \quad \mu := \frac{1}{2} t L^{-2} < L^{-\delta}, \tag{5.1}$$

where $W \in \mathcal{S}\left(\mathbb{R}^{2d}\right)$. Thus we have

$$\mathscr{K} = L^{2d} \int_{(z_1,z_2)\in\mathbb{R}^{2d}} W(z_1,z_2) g(4\mu z_1 \cdot z_2) dz_1 dz_2. \tag{5.2}$$

Step 1: Truncating in N. We first notice that the main contribution of the sum $\mathcal{S}$ (resp., the integral $\mathscr{K}$) comes from the region $|N| \lesssim L^{1+\delta_1}$ (resp., $|(\xi_1,\xi_2)| \lesssim L^{\delta_1}$), where $\delta_1 = \frac{\delta}{100}$. This uses the fact that W is a Schwartz function with sufficient decay. We can therefore without loss of generality include in the sum $\mathcal{S}$ (resp., the integral $\mathcal{K}$) a factor $\chi\left(\frac{N}{L^{1+\delta_1}}\right)$ $\left(\text{resp., } \chi\left(\frac{z}{L^{1+\delta_1}}\right)\right)$, where $\chi \in C_c^\infty\left(\mathbb{R}^d\right)$ is 1 on the unit ball $B\left(0, \frac{1}{10}\right)$ and vanishes outside $B\left(0, \frac{2}{10}\right)$.

Step 2: Isolating the main term. We now use the fact that the Fourier transform of g is given by the tent function $\widehat{g}(x) = 1 - |x|$ on the interval $[-1,1]$ and vanishes otherwise to write (using the notation $e(x) := e^{2\pi i x}$)

$$\begin{aligned}
\mathcal{S} =& \sum_{N=(N_1,N_2)\in\mathbb{Z}^{2d}} W\left(\frac{N}{L}\right) \chi\left(\frac{N}{L^{1+\delta_1}}\right) \int_{-1}^{1} \widehat{g}(\tau) e(4\mu\tau\mathcal{Q}(N)) d\tau \\
=& \mu^{-1} \sum_{N=(N_1,N_2)\in\mathbb{Z}^{2d}} W\left(\frac{N}{L}\right) \chi\left(\frac{N}{L^{1+\delta_1}}\right) \int_{-\mu}^{\mu} \widehat{g}\left(\frac{\tau}{\mu}\right) e(4\tau\mathcal{Q}(N)) d\tau \\
=& \mathcal{S}_A + \mathcal{S}_B,
\end{aligned}$$

where A is the contribution of $|\tau| \le L^{-1-\delta_1}$ and B is the contribution of the complementary region, which could be empty if $\mu < L^{-1-\delta_1}$, in which case we assume $B = 0$. This decomposition can be understood as the analogue of the classical minor versus major arc splitting in the circle method. For the major arc $\mathcal{S}_A$, we use Poisson summation to replace the sum in N by an integral which will give the needed asymptotic up to acceptable errors:

$$\begin{aligned}
\mathcal{S}_A &= \mu^{-1} \int_{|\tau|\le \min\left(\mu, L^{-1-\delta_1}\right)} \widehat{g}\left(\frac{\tau}{\mu}\right) \sum_{c\in\mathbb{Z}^{2d}} \int_{z\in\mathbb{R}^{2d}} W\left(\frac{z}{L}\right) \chi\left(\frac{z}{L^{1+\delta_1}}\right) e(4\tau\mathcal{Q}(z) - c\cdot z) dz d\tau \\
&= \mu^{-1} L^{2d} \int_{|\tau|\le \min\left(\mu, L^{-1-\delta_1}\right)} \widehat{g}\left(\frac{\tau}{\mu}\right) \int_{z\in\mathbb{R}^{2d}} W(z) \chi\left(\frac{z}{L^{\delta_1}}\right) e\left(4\tau L^2 \mathcal{Q}(z)\right) dz d\tau \\
&\quad + \mu^{-1} L^{2d} \int_{|\tau|\le \min\left(\mu, L^{-1-\delta_1}\right)} \widehat{g}\left(\frac{\tau}{\mu}\right) \sum_{\substack{c\in\mathbb{Z}^{2d} \\ c\neq 0}} \int_{z\in\mathbb{R}^{2d}} W(z) \chi\left(\frac{z}{L^{\delta_1}}\right) e\left(4\tau L^2 \mathcal{Q}(z) - Lc\cdot z\right) dz d\tau \\
&= \mu^{-1} L^{2d} \int_{|\tau|\le \mu} \widehat{g}\left(\frac{\tau}{\mu}\right) \int_{z\in\mathbb{R}^{2d}} W(z) \chi\left(\frac{z}{L^{\delta_1}}\right) e\left(4\tau L^2 \mathcal{Q}(z)\right) dz d\tau \\
&\quad + \mu^{-1} L^{2d} \int_{\min\left(\mu, L^{-1-\delta_1}\right)<|\tau|\le\mu} \widehat{g}\left(\frac{\tau}{\mu}\right) \int_{z\in\mathbb{R}^{2d}} W(z) \chi\left(\frac{z}{L^{\delta_1}}\right) e\left(4\tau L^2 \mathcal{Q}(z)\right) dz d\tau
\end{aligned}$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

$$+\mu^{-1}L^{2d}\int_{|\tau|\leq\min\left(\mu,L^{-1-\delta_1}\right)}\widehat{g}\left(\frac{\tau}{\mu}\right)\sum_{\substack{c\in\mathbb{Z}^{2d}\\ c\neq 0}}\int_{z\in\mathbb{R}^{2d}}W(z)\chi\left(\frac{z}{L^{\delta_1}}\right)e\left(4\tau L^2\mathcal{Q}(z)-Lc\cdot z\right)dzd\tau$$

$$=\mathscr{K}+\mathcal{S}_{A1}+\mathcal{S}_{A2},$$

where $\mathcal{S}_{A1}$ and $\mathcal{S}_{A2}$ are respectively the second and third terms in the second-to-last equality.

The remainder of the proof is to show that $\mathcal{S}_{A1}, \mathcal{S}_{A2}$ and $\mathcal{S}_B$ are error terms.

Step 3: Showing that $\mathcal{S}_{A1}$ and $\mathcal{S}_{A2}$ are error terms. To estimate $\mathcal{S}_{A1}$, we use the stationary phase estimate

$$\left|\int_{z\in\mathbb{R}^{2d}}W(z)\chi\left(\frac{z}{L^{\delta_1}}\right)e\left(4\tau L^2\mathcal{Q}(z)\right)dz\right|\lesssim\left(\tau L^2\right)^{-d}$$

and the fact that the term is only nonzero if $\mu > L^{-1-\delta_1}$ to bound

$$|\mathcal{S}_{A1}|\lesssim\mu^{-1}\int_{L^{-1-\delta_1}<|\tau|\leq\mu}\left|\widehat{g}\left(\frac{\tau}{\mu}\right)\right||\tau|^{-d}\,d\tau\lesssim\mu^{-1}L^{(1+\delta_1)(d-1)}=t^{-1}L^{d+1+\delta_1(d-1)}\ll L^{2d-\delta}t^{-1}.$$

For $\mathcal{S}_{A2}$, we use nonstationary phase techniques relying on the fact that the phase function $\Phi(z) = 4\tau L^2\mathcal{Q}(z) - Lc\cdot z$ satisfies $|\nabla_z\Phi(z)| = |L(4\tau L(z_2, z_1) - c)| \gtrsim L|c|$ for $c \neq 0$, since $|z| \leq \frac{L^{\delta_1}}{5}$. Therefore, one can integrate by parts in z sufficiently many times and show that $|\mathcal{S}_{A1}| \ll L^{2d-\delta}t^{-1}$ as well.

Step 4: Showing that $\mathcal{S}_B$ is an error term. Here we assume without loss of generality that $L^{-1-\delta_1} < \mu \leq L^{-\delta}$ (otherwise $\mathcal{S}_B = 0$). Therefore,

$$\mathcal{S}_B=\mu^{-1}\int_{L^{-1-\delta_1}<|\tau|\leq\mu}\widehat{g}\left(\frac{\tau}{\mu}\right)F(\tau)d\tau,\quad F(\tau)=\sum_{N=(N_1,N_2)\in\mathbb{Z}^{2d}}W\left(\frac{N}{L}\right)\chi\left(\frac{N}{L^{1+\delta_1}}\right)e\left(4\tau\mathcal{Q}(N)\right).$$

Recall that $\mathcal{Q}(N) = \sum_{j=1}^{d}\beta_j(N_1)_j(N_2)_j$, so we perform the following change of variables:

$$(N_1)_j=\frac{p_j+q_j}{2},\qquad (N_2)_j=\frac{p_j-q_j}{2},\qquad p_j=q_j\pmod 2.$$

Therefore, the sum in $(N_1)_j, (N_2)_j \in \mathbb{Z}^2$ becomes a sum

$$\sum_{(p_j,q_j)\in\mathbb{Z}^2}-\sum_{p_j\in2\mathbb{Z},q_j\in\mathbb{Z}}-\sum_{p_j\in\mathbb{Z},q_j\in2\mathbb{Z}}+2\sum_{(p_j,q_j)\in2\mathbb{Z}^2}.$$

We will estimate the contribution of the first sum, and it will be obvious from the proof that the other sums are estimated similarly. Also, by symmetry, we only need to consider the sums for which $p_j, q_j \geq 0$, which reduces us to

$$F(\tau)=\sum_{\substack{p_j,q_j\geq0\\ j=1,\dots,d}}\widetilde{W}\left(\frac{(p,q)}{L}\right)\widetilde{\chi}\left(\frac{(p,q)}{L^{1+\delta_1}}\right)\prod_{j=1}^{d}e\left(\tau\beta_jp_j^2\right)e\left(-\tau\beta_jq_j^2\right).$$

Let $G(s,n) = \sum_{p=0}^{n}e\left(sp^2\right)$ be the Gauss sum, and abusing notation, also denote by $G(s,x) = G(s,[x])$ for $x \in \mathbb{R}$, where $[x]$ is the floor function. Then

$$F(\tau)=\int_{\substack{x_j,y_j\geq0\\ j=1,\dots,d}}\widetilde{W}\left(\frac{(x,y)}{L}\right)\widetilde{\chi}\left(\frac{(x,y)}{L^{1+\delta_1}}\right)\prod_{j=1}^{d}G'\left(\tau\beta_j,x_j\right)G'\left(\tau\beta_j,y_j\right).$$

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

Integrating by parts in all the variables (or equivalently, performing an Abel summation), one obtains

$$F(\tau) = \int_{\substack{x_j, y_j \geq 0 \\ j=1,\dots,d}} \partial_{x_1} \cdots \partial_{y_d} \widetilde{W}\left(\frac{(x,y)}{L}\right) \widetilde{\chi}\left(\frac{(x,y)}{L^{1+\delta_1}}\right) \prod_{j=1}^{d} G\left(\tau\beta_j, x_j\right) G\left(\tau\beta_j, y_j\right) + \text{l.o.t.},$$

where 'l.o.t.' is lower-order terms that can be bounded is a similar or simpler way than the main term. Here and in what follows, $\partial_{x_1} \cdots \partial_{y_d} \widetilde{W}\left(\frac{(x,y)}{L}\right)$ is understood as $\partial_{x_1} \cdots \partial_{y_d} \left(\widetilde{W}\left(\frac{(x,y)}{L}\right)\right)$.

We now recall the Gauss sum estimate for $G(s,n)$: let $0 \leq a < q \leq n$ be integers such that $(a,q) = 1$ and $\left|s - \frac{a}{q}\right| < \frac{1}{qn}$ (for any s and n, such a pair exists by Dirichlet's approximation theorem); then

$$|G(s,n)| \leq \frac{n}{\sqrt{q}\left(1 + n\left\|s - \frac{a}{q}\right\|^{1/2}\right)} \leq \frac{n}{\sqrt{q}}.$$

Here $s = \tau\beta_j$, with $\tau \in \left[L^{-1-\delta_1}, L^{-\delta}\right]$, $\beta_j \in [1,2]$. This means that either $|n| < L^{2\delta}$ or $\frac{a}{q} \lesssim L^{-\delta}$ ($\Rightarrow q \gtrsim L^{\delta}$), and in either case we get $|G(s,n)| \lesssim L^{\left(1+\delta_1-\frac{\delta}{2}\right)}$, since $|n| \lesssim L^{1+\delta_1}$ (note that this argument works when $a > 0$; if $a = 0$ we have the better bound $|G(s,n)| \lesssim |s|^{-1/2} \leq L^{2/3}$).

As a result, we have

$$|F(\tau)| \lesssim L^{\left(1+\delta_1-\frac{\delta}{2}\right)(2d-4)} \int_{\substack{x_j, y_j \geq 0 \\ j=1,\dots,d}} \left|\partial_{x_1} \cdots \partial_{y_d} \widetilde{W}\left(\frac{(x,y)}{L}\right) \widetilde{\chi}\left(\frac{(x,y)}{L^{1+\delta_1}}\right)\right| \times \prod_{j=1}^{2} \left|G\left(\tau\beta_j, [x_j]\right)\right| \left|G\left(\tau\beta_j, [y_j]\right)\right|.$$

This gives

$$|\mathcal{S}_B| \lesssim L^{\left(1+\delta_1-\frac{\delta}{2}\right)(2d-4)} \mu^{-1} \int_{\substack{x_j, y_j \geq 0 \\ j=1,\dots,d}} \left|\partial_{x_1} \cdots \partial_{y_d} \widetilde{W}\left(\frac{(x,y)}{L}\right) \widetilde{\chi}\left(\frac{(x,y)}{L^{1+\delta_1}}\right)\right| \times \int_{|\tau| \leq \mu} \prod_{j=1}^{2} \left|G\left(\tau\beta_j, [x_j]\right)\right| \left|G\left(\tau\beta_j, [y_j]\right)\right| d\tau.$$

Now using Hua's lemma (compare [30]), we have $\left\|G\left(\tau, n_j\right)\right\|_{L^4[0,1]} \lesssim n_j^{1/2+\delta_1} \lesssim L^{(1+\delta_1)\left(\frac{1}{2}+\delta_1\right)}$, which gives

$$|\mathcal{S}_B| \lesssim L^{\left(1+\delta_1-\frac{\delta}{2}\right)(2d-4)} \mu^{-1} L^{(1+\delta_1)(2+4\delta_1)} = L^{2d} t^{-1} L^{-\delta(d-2)+\delta_1(2d-2+4\delta_1)} \ll L^{2d-\theta} t^{-1},$$

provided that $\theta < \min\left(1, \frac{(d-2)\delta}{2d}\right)$ and recalling that $\delta_1 = \frac{\delta}{100}$. □

Acknowledgements. The authors would like to thank Andrea Nahmod, Sergey Nazarenko and Jalal Shatah for illuminating discussions. They also thank the anonymous referee for valuable suggestions that helped improve the exposition.

Conflict of Interest. None.

Financial support. The first author was supported by NSF grant DMS 1900251. The second author was supported by NSF grants DMS-1852749 and DMS-1654692, a Sloan Fellowship and the Simons Collaboration Grant on Wave Turbulence. The results of this work were announced on 1 November, 2019, at a Simons Collaboration Grant meeting.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

References

[1] Q. Aubourg, A. Campagne, C. Peureux, F. Ardhuin, J. Sommeria, S. Viboud and N. Mordant, 'Three-wave and four-wave interactions in gravity wave turbulence', *Phys. Rev. Fluids* **2** (2017), 114802.

[2] A. Bényi, T. Oh and O. Pocovnicu, 'Higher order expansions for the probabilistic local Cauchy theory of the cubic nonlinear Schrödinger equation on $\mathbb{R}^3$', *Trans. Amer. Math. Soc. Ser. B* **6**(4) (2019), 114–160.

[3] J. Bourgain, 'Invariant measures for the 2D-defocusing nonlinear Schrödinger equation', *Comm. Math. Phys.* **176** (1996), 421–445.

[4] J. Bourgain, 'On pair correlation for generic diagonal forms', Preprint, 2016, arXiv:1606.06173.

[5] J. Bourgain and A. Bulut, 'Almost sure global well-posedness for the radial nonlinear Schrödinger equation on the unit ball II: The 3d case', *J. Eur. Math. Soc.* **16**(6) (2014), 1289–1325.

[6] T. Buckmaster, P. Germain, Z. Hani and J. Shatah, 'Effective dynamics of the nonlinear Schrödinger equation on large domains', *Comm. Pure Appl. Math.* **71** (2018), 1407–1460.

[7] T. Buckmaster, P. Germain, Z. Hani and J. Shatah, 'Onset of the wave turbulence description of the longtime behavior of the nonlinear Schrödinger equation', Preprint, 2019, arXiv:1907.03667.

[8] N. Burq and N. Tzvetkov, 'Random data Cauchy theory for supercritical wave equations I: Local theory', *Invent. Math.* **173**(3) (2008), 449–475.

[9] J. W. S. Cassels, *An Introduction to Diophantine Approximation*, *Cambridge Tracts in Mathematics and Mathematical Physics* vol. 45 (Hafner Publishing Co., New York, 1972). Facsimile reprint of the 1957 edition.

[10] C. Cercignani, R. Illner and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, (Springer, Berlin, 1994).

[11] J. Colliander and T. Oh, 'Almost sure well-posedness of the cubic nonlinear Schrödinger equation below $L^2(T)$', *Duke Math. J* **161**(3) (2012), 367–414.

[12] C. Collot and P. Germain, 'On the derivation of the homogeneous kinetic wave equation', Preprint, 2019, arXiv:1912.10368.

[13] C. Collot and P. Germain, 'Derivation of the homogeneous kinetic wave equation: Longer time scales', Preprint, 2020, arXiv:2007.03508.

[14] G. Da Prato and A. Debussche, 'Two-dimensional Navier-Stokes equations driven by a space-time white noise', *J. Funct. Anal.* **196**(1) (2002), 180–210.

[15] Y. Deng, 'Two dimensional nonlinear Schrödinger equation with random radial data', *Anal. PDE* **5**(5) (2012), 913–960.

[16] Y. Deng and Z. Hani, 'Full derivation of the wave kinetic equation', Preprint, 2021, arXiv:2104.11204.

[17] Y. Deng, A. R. Nahmod and H. Yue, 'Optimal local well-posedness for the periodic derivative nonlinear Schrödinger equation', Preprint, 2019, arXiv:1905.04352.

[18] Y. Deng, A. R. Nahmod and H. Yue, 'Invariant Gibbs measures and global strong solutions for nonlinear Schrödinger equations in dimension two', Preprint, 2019, arXiv1910.08492.

[19] Y. Deng, A. R. Nahmod and H. Yue, 'Random tensors, propagation of randomness, and nonlinear dispersive equations', Preprint, 2020, arXiv:2006.09285.

[20] P. Denissenko, S. Lukaschuk and S. Nazarenko, 'Gravity wave turbulence in a laboratory flume, *Phys. Rev. Lett.* **99** (2007), 014501.

[21] B. Dodson, J. Lührmann and D. Mendelson, 'Almost sure local well-posedness and scattering for the 4D cubic nonlinear Schrödinger equation', *Adv. Math.* **347** (2019), 619–676.

[22] S. Dyachenko, A. C. Newell, A. Pushkarev and V. E. Zakharov, 'Optical turbulence: Weak turbulence, condensates and collapsing filaments in the nonlinear Schrödinger equation', *Physica D: Nonlinear Phenomena*, **57**(1– 2) (1992), 96–160.

[23] E. Faou, 'Linearized wave turbulence convergence results for three-wave systems', Preprint, 2018, arXiv:1805.11269.

[24] E. Faou, P. Germain and Z. Hani, 'The weakly nonlinear large-box limit of the 2D cubic nonlinear Schrödinger equation', *J. Amer. Math. Soc.* **NN**(NN) (2015), 915–982.

[25] N. Fitzmaurice, D. Gurarie, F. McCaughan and W. A. Woyczynski (eds), *Nonlinear Waves and Weak Turbulence, with applications in Oceanography and Condensed Matter Physics* (Springer Science+Business Media, New York (1993).

[26] I. Gallagher, L. Saint-Raymond and B. Texier, *From Newton to Boltzmann: The Case of Hard-Spheres and Short-Range Potentials* vol. **18**, (European Mathematical Society, Zürich, 2014).

[27] H. Grad, 'Principles of the kinetic theory of gases, in *Handbuch der Physik, Vol.* **12** (Springer, Berlin, 1958), 205–294.

[28] M. Gubinelli, P. Imkeller and N. Perkowski, 'Paracontrolled distributions and singular PDEs', *Forum Math Pi* **3** (2015), e6.

[29] M. Hairer, 'A theory of regularity structures', *Invent. Math.* **198**(2) (2014), 269–504.

[30] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS Colloquium Publications, *American Math. Society*, **53**(2004).

[31] P. A. Janssen, 'Progress in ocean wave forecasting', *J. Comput. Phys.* **227**(7) (2008), 3572–3594.

[32] E. Kartashova 'Exact and quasiresonances in discrete water wave turbulence', *Phys. Rev. Lett.* **98**(21) (2007), 214502.

[33] C. Kenig and D. Mendelson, 'The focusing energy-critical nonlinear wave equation with random initial data', *International Mathematics Research Notices*, 2019.

[34] O. E. Lanford, *Time Evolution of Large Classical Systems*, *Lecture Notes in Physics vol.* **38** (Springer, Heidelberg, 1975).

[35] J. Lukkarinen and H. Spohn, 'Weakly nonlinear Schrödinger equation with random initial data', *Invent. Math.* **183** (2011), 79–188.

[36] V. S. L'vov and S. Nazarenko, 'Discrete and mesoscopic regimes of finite-size wave turbulence', *Phys. Rev. E* **82** (2010), 056322.

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press

[37] R. A. Minlos, Introduction to Mathematical Statistical Physics, *University Lecture Series vol.*: **19** (American Math Society, Providence, 2000).
[38] N. Nahmod and G. Staffilani, 'Almost sure well-posedness for the periodic 3D quintic nonlinear Schrödinger equation below the energy space', *J. Eur. Math. Soc. (JEMS)* **17**(7) (2015), 1687–1759.
[39] S. Nazarenko, *Wave Turbulence*, Lecture Notes in Physics vol. **825** (Springer, Heidelberg, 2011).
[40] T. Oh and L. Thomann, 'A pedestrian approach to the invariant Gibbs measures for the 2-d defocusing nonlinear Schrödinger equations', *Stoch. Partial Differ. Equ. Anal. Comput.* **6**(3) (2018), 397–445.
[41] R.E. Peierls, 'Zur kinetischen Theorie der Wärmeleitung in Kristallen', *Ann. Phys.* **3** (1929), 1055–1101.
[42] D. Ruelle, *Statistical Mechanics: Rigorous Results*, 2nd edn (World Scientific Publishing, Singapore,1999).
[43] H. Spohn, 'Kinetic equations from Hamiltonian dynamics: Markovian limits', *Rev. Modern Phys.* **63**(3) (1980), 569–615.
[44] H. Spohn, *Large Scale Dynamics of Interacting Particles*, Texts and Monographs in Physics (Springer Verlag, Heidelberg, 1991).
[45] H. Spohn, 'The phonon Boltzmann equation, properties and link to weakly anharmonic lattice dynamics', *J. Stat. Phys.* **124** (2006), 1041–1104.
[46] H. Spohn, 'On the Boltzmann equation for weakly nonlinear wave equations', in *Boltzmann's Legacy, ESI Lectures in Mathematics and Physics* (Zürich, Switzerland: European Mathematical Society 2008) 145–159.
[47] G. Staffilani and M.-B. Tran, 'On the wave turbulence theory for stochastic and random multidimensional KdV type equations', Unpublished manuscript.
[48] N. Tzvetkov, 'Quasi-invariant Gaussian measures for one-dimensional Hamiltonian partial differential equations', *Forum Math. Sigma* **3** (2015), **e28**.
[49] T. Burns, et al. *Guide to Wave Analysis and Forecasting* (Secretariat of the World Meteorological Organization, Geneva, 1998).
[50] V. E. Zakharov, A. O. Korotkevich, A. Pushkarev and D. Resio, 'Coexistence of weak and strong wave turbulence in a swell propagation', *Phys. Rev. Lett.* **99** (2007), 164501.
[51] V. E. Zakharov, V. S. L'vov and G. Falkovich, *Kolmogorov Spectra of Turbulence: I WaveTurbulence* (Springer, Berlin, 1992).

https://doi.org/10.1017/fmp.2021.6 Published online by Cambridge University Press