# RANDOM TENSORS, PROPAGATION OF RANDOMNESS, AND NONLINEAR DISPERSIVE EQUATIONS

YU DENG[1], ANDREA R. NAHMOD[2], AND HAITIAN YUE[3]

Abstract. We introduce the *theory of random tensors*, which naturally extends the method of *random averaging operators* in our earlier work [36], to study the propagation of randomness under nonlinear *dispersive* equations. By applying this theory we establish almost-sure local well-posedness for semilinear Schrödinger equations in the *full subcritical* range relative to the *probabilistic scaling* (Theorem 1.1). The solution we construct has an explicit expansion in terms of multilinear Gaussians with adapted random tensor coefficients. As a byproduct we also obtain new results concerning regular data and long-time solutions, in particular Theorem 1.6, which provides long-time control for random homogeneous data, demonstrating the highly nontrivial fact that the first energy cascade happens at a much later time than in the deterministic setting.

In the random setting, the probabilistic scaling is the natural scaling for dispersive equations, and is *different* from the natural scaling for parabolic equations. Our theory of random tensors can be viewed as the dispersive counterpart of the existing parabolic theories (regularity structures, para-controlled calculus and renormalization group techniques).

Contents



[1]Y. D. is funded in part by NSF DMS-1900251 and Sloan Fellowship.

[2]A.N. is funded in part by NSF DMS-1800852, DMS-2101381 and the Simons Foundation Collaboration Grant on Wave Turbulence (Nahmod's Award ID 651469).

## 1. Introduction

The study of partial differential equations with randomness has become an important and influential subject in the last few decades. In this work we will be concerned with a major topic of this subject, namely the local in time Cauchy problems with either random initial data or additive stochastic forcing.

It is well known that in many situations, randomization or noise improves the behavior of solutions to PDEs. Usually this can be interpreted as *generic* solutions being genuinely better than *pathological* ones. This phenomenon, which has its roots in the various cancellation properties of independent random variables (e.g. Central Limit Theorem or Khintchine's inequality), has been extensively studied since the 70–80's. The key difficulty here is to analyze how the explicit randomness (given for example by a Wiener measure or Gaussian noise) *propagates* under the flow of nonlinear PDEs.

In the past few years, there has been significant progress in the setting of singular parabolic stochastic equations (SPDEs): the development of the theory of *regularity structures* of Hairer and the *para-controlled calculus* of Gubinelli-Imkeller-Perkowski has led to tremendous success in local well-posedness theory, essentially completing the full picture in what is known as the *subcritical* range. Unfortunately, both theories rely crucially on the *parabolic* nature of the equation, and have not achieved the same success in the other important class of PDEs, namely the *dispersive* equations.

The purpose of this work is to develop a new theory, which we will call *random tensors*, to fill this gap in the dispersive setting. This is a natural extension of the method of *random averaging operators* in our earlier work [36], but is much more powerful. In fact, in the random setting dispersive equations have a natural scaling, which we call the probabilistic scaling (see Section 1.2), that is *different* from the parabolic one[1], and our method—just like the theory of regularity structures and the para-controlled calculus in the parabolic setting—provides the complete picture in the full subcritical range with respect to this natural scaling.

In this work we will focus on the random data problem[2] for the semilinear Schrödinger equation, which is the most common (and most studied) nonlinear dispersive equation. Our method is general and can be applied to other dispersion relations (see Section 9.2).

The rest of this introduction is organized as follows. In Section 1.1 we describe the setup and state the main theorems. In Section 1.2 we present a heuristic scaling argument from [36] to justify the notion of criticality in this work. In Section 1.3 we briefly review the ideas in earlier works, and in Sections 1.4 we discuss the method of random averaging operators in [36]. Finally in Section 1.5 we provide the motivation behind our theory of random tensors; the detailed explanation of this theory is left to Section 2.

1.1. **Setup and main results.** Fix $d \geq 1$ and $p \geq 3$ odd, and assume $(d,p) \neq (1,3)$; in particular $d(p-1) \geq 4$. Consider the nonlinear Schrödinger (NLS) equation on $\mathbb{R}_t \times \mathbb{T}^d_x$, where $\mathbb{T}^d = (\mathbb{R}/2\pi\mathbb{Z})^d$:

$$\begin{cases} (i\partial_t + \Delta)u = W^p(u), \\ \qquad u(0) = f(\omega). \end{cases} \tag{1.1}$$

Here $f(\omega)$ is some choice of random initial data defined on an ambient probability space $(\Theta, \mathcal{B}, \mathbb{P})$, $\omega \in \Theta$, and $W^p(u)$ is either $|u|^{p-1}u$ or its *Wick ordering*, which will be defined precisely below. The Hamiltonian of (1.1) is linked to the $\Phi^{p+1}_d$ model in constructive quantum field theory.

1.1.1. *Almost-sure local well-posedness.* In the context of almost-sure local well-posedness, the random initial data will be given by

$$f(\omega) = \sum_{k\in\mathbb{Z}^d} \frac{g_k(\omega)}{\langle k\rangle^\alpha} e^{ik\cdot x}, \tag{1.2}$$

where $\{g_k(\omega)\}_{k\in\mathbb{Z}^d}$ are i.i.d. centered normalized (complex) Gaussian random variables. Such initial data was first considered by Bourgain [11, 12] and later by Burq-Tzvetkov [19]. In (1.2) we will fix

$$\alpha = s + \frac{d}{2}, \qquad s > s_{pr} := -\frac{1}{p-1}. \tag{1.3}$$

This value $s_{pr}$ is the critical exponent for the *probabilistic scaling*, which will be discussed in detail in Section 1.2. It is always lower than $s_{cr} := (d/2) - 2/(p-1)$, which is the critical exponent for the usual (deterministic) scaling. The random data $f(\omega)$ defined by (1.2) almost surely belongs to $H^{s-}(\mathbb{T}^d) := \cap_{s'<s} H^{s'}(\mathbb{T}^d)$, but not to $H^s(\mathbb{T}^d)$.

[1]See Remark 1.12, and the explanation in Section 9.1.

[2]Random data is a natural setting for dispersive equations (parallel to additive noise for parabolic ones) in view of the invariant measures. Of course one may also consider stochastic versions of (1.1), which are similar but correspond to different randomizations, see Section 9.2.

Our first main theorem, originally stated as Conjecture 1.7 in [36], proves that (1.1) is almost surely locally well-posed with random initial data (1.2), in the sense that canonical smooth approximations converge to a unique limit; here and throughout the paper, canonical smooth approximations always mean the ones described in Theorem 1.1 and Remark 1.4 below. This result can be interpreted as almost sure local well-posedness in $H^{s-}(\mathbb{T}^d)$ with respect to the canonical Gaussian measure—the law of $f(\omega)$ defined by (1.2)—for any $s > s_{pr}$, i.e. *in the full probabilistically subcritical range.*

To state the theorem, we need to define the canonical truncations and the associated Wick orderings. Given a dyadic number $N \geq 1$[1], define the truncation operators $\Pi_N$ and $\Delta_N$ by

$$(\Pi_N u)_k = \Pi_N u_k := \mathbf{1}_{\langle k\rangle\leq N}\cdot u_k, \quad \Delta_N = \Pi_N - \Pi_{\frac{N}{2}}, \tag{1.4}$$

where $u_k$ represents the Fourier coefficient. For notational simplicity, we will identify $\Pi_N$ with the multiplier $\Pi_N(k) = \mathbf{1}_{\langle k\rangle\leq N}$, and similarly for $\Delta_N$. Define the expectation of truncated mass

$$\sigma_N := \mathbb{E}\fint_{\mathbb{T}^d} |\Pi_N f(\omega)|^2 = \sum_{\langle k\rangle\leq N}\frac{1}{\langle k\rangle^{2\alpha}}, \tag{1.5}$$

and, for integers $r \geq 0$, the Wick-ordered monomials

$$W_N^{2r}(u) = \sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}\frac{\sigma_N^{r-j}r!}{j!}|u|^{2j}, \quad W_N^{2r+1}(u) = \sum_{j=0}^{r}(-1)^{r-j}\binom{r+1}{j+1}\frac{\sigma_N^{r-j}r!}{j!}|u|^{2j}u, \tag{1.6}$$

where $\sigma_N$ is as in (1.5). The first main theorem is then stated as follows.

**Theorem 1.1.** *Fix $s > s_{pr}$ and $\alpha$ as in (1.3), and let $f(\omega)$ be as in (1.2). Let $u_N$ be the solution to the canonically truncated system*

$$\begin{cases} (i\partial_t + \Delta)u_N = \Pi_N W_N^p(u_N), \\ \qquad u_N(0) = \Pi_N f(\omega). \end{cases} \tag{1.7}$$

*Then, for $0 < \tau \ll 1$, there exists a set $Z \subset \Theta$ with $\mathbb{P}(Z) \leq C_\theta e^{-\tau^{-\theta}}$, where $\theta$ is a small constant (ultimately determined by $(d, p, s)$, and independent of $\tau$) and $C_\theta$ is a constant determined by $\theta$, such that when $\omega \notin Z$, the sequence $\{u_N\}$ converges, as $N \to \infty$, to a unique limit $u$ in $C_t^0 H_x^{s-}[-\tau, \tau]$.*

*Moreover, for this $u$, the nonlinearity $W^p(u)$ in (1.1), which is the Wick ordering of $|u|^{p-1}u$, is well-defined as*

$$W^p(u) = \lim_{N\to\infty} W_N^p(\Pi_N u) = \lim_{N\to\infty}\Pi_N W_N^p(\Pi_N u) \tag{1.8}$$

*in the sense of spacetime distributions (where both limits exist and are equal), and $u$ solves the equation (1.1) in the distributional sense. Finally this solution $u$ has an explicit expansion in terms of multilinear Gaussians with adapted random tensor coefficients; see (8.5) for the precise form.*

*Remark* 1.2. Theorem 1.1 (and Theorem 1.6 below) can be shown for any rectangular torus $\mathbb{T}_\beta^d := (\mathbb{R}/2\pi\beta_1\mathbb{Z}) \times \cdots \times (\mathbb{R}/2\pi\beta_d\mathbb{Z})$, and for focusing nonlinearity, with almost the same proof.

When $(d, p) = (1, 3)$, instead of (1.6) one should look at the completely non-resonant nonlinearity $(|u|^2 - 2\fint |u|^2)u$, since in this case $\|\Pi_N u\|_{L^2}^2 - \mathbb{E}\|\Pi_N u\|_{L^2}^2$ does not converge as $N \to \infty$, see (5.5). With this change, it is known, see [45], that (1.1) is deterministically locally well-posed in a Fourier-Lebesgue space which the data (1.2) almost surely belongs to (as in this case $s_{pr} = s_{cr}$), so Theorem 1.1 remains true.

[1] We will assume $N \geq 1$ throughout, and only "formally" need to replace $N$ by $1/2$ in a few places.

*Remark* 1.3. The Wick ordering (1.6) is crucial in Theorem 1.1, as our solution has infinite mass when $s_{pr} < s < 0$; in this case the Wick-ordered NLS, rather than the original one, is the right equation to study. In fact, even in the simplest case $(d, p) = (1, 3)$, the NLS without the renormalization as in Remark 1.2 will have *no solution* for *any* infinite mass initial data; see [51]. As another example, for the dynamical $\Phi_2^4$ model (2D cubic heat equation with white noise forcing), the canonical smooth approximations will converge to a nontrivial limit only with Wick ordering, otherwise the limit would be identically zero for any initial condition (see [23, 56]).

*Remark* 1.4. Despite having low regularity $C_t^0 H_x^{s-}$, the local solution constructed in Theorem 1.1 is a *strong* solution, as it is the unique limit of canonical smooth approximations. In fact, by slightly modifying our proof (which will not be done here for simplicity of presentation), we can obtain the following general convergence result (see [67] for some discussions regarding different choices of approximations):

Let $\varphi$ be a function on $\mathbb{R}^d$, $\varphi(0) = 1$, and $\varphi$ is either Schwartz or equals the characteristic function $\mathbf{1}_{\mathbb{B}}$ of the unit ball $\mathbb{B}$ (the latter is the setting of Theorem 1.1). Let $\widetilde{\varphi}$ be either 1 or $\mathbf{1}_{\mathbb{B}}$, in the latter case we assume $\varphi \equiv 0$ outside $\mathbb{B}$. Define $\mathtt{P}_\lambda(k) = \varphi(\lambda^{-1}k)$ and $\widetilde{\mathtt{P}}_\lambda(k) = \widetilde{\varphi}(\lambda^{-1}k)$. Let $W_\lambda^p$ be defined as in (1.6) with $\sigma_N$ replaced by $\sigma_\lambda$, which is in turn defined as in (1.5) with $\Pi_N$ replaced by $\mathtt{P}_\lambda$. Consider the solution $u_\lambda$ to the system

$$\begin{cases} (i\partial_t + \Delta)u_\lambda = \widetilde{\mathtt{P}}_\lambda W_\lambda^p(u_\lambda), \\ \qquad\qquad u_\lambda(0) = \mathtt{P}_\lambda f(\omega). \end{cases} \tag{1.9}$$

Then Theorem 1.1 remains true with conclusion being the convergence of $\{u_\lambda\}$ as $\lambda \to \infty$. Here the exceptional set $Z$ and the limit $u$ *do not depend on* the choice of $(\varphi, \widetilde{\varphi})$.

More precisely, there exists a random time $\mathtt{T} = \mathtt{T}(\omega)$ satisfying $\mathbb{P}(\mathtt{T} < \tau) \leq C_\theta e^{-\tau^{-\theta}}$ for any $\tau > 0$, and a random function $u = u(t, x, \omega)$ defined for $|t| \leq \mathtt{T}(\omega)$, such that almost surely in $\omega$, we have $u_\lambda \to u$ in $C_t^0 H_x^{s-}[-\mathtt{T}, \mathtt{T}]$, as $\lambda \to \infty$, for *any* choice of $(\varphi, \widetilde{\varphi})$.

*Remark* 1.5. Although Theorem 1.1 concerns singular (i.e. low regularity) data and short-time solutions, the fundamental issue here is to understand how the randomness structure[1] propagates under the nonlinear Schrödinger flow. With this understanding, we can easily obtain new results for regular data and long-time solutions, such as Theorem 1.6 below.

1.1.2. *Long-time control for random homogeneous data.* Consider the random homogeneous data, which is the random initial data given by

$$f_{\mathrm{ho}}(\omega) = N^{-\alpha} \sum_{k \in \mathbb{Z}^d} \phi\Big(\frac{k}{N}\Big) g_k(\omega) e^{ik\cdot x}; \qquad \alpha = s + \frac{d}{2},\ s > s_{pr}. \tag{1.10}$$

Here $\phi$ is a fixed Schwartz function and $N$ is a *fixed* large parameter. Compared to (1.2), which is a superposition of multiple scales, in (1.10) we have only one scale $N$ and the Fourier modes are uniformly distributed in the ball $\langle k \rangle \lesssim N$. Such random data in fact are, up to rescaling, the ones appearing in the derivation of *wave kinetic equation* in weak turbulence problems [18, 28, 33]. Note that with high probability, $\|f_{\mathrm{ho}}(\omega)\|_{H^s} \sim 1$. The second main theorem is then stated as follows.

[1]Such structure lives on high frequencies and fine scales. It becomes more explicit when considering low regularity solutions, and may be obscured by the dominant coarse scale profile in high regularity solutions.

**Theorem 1.6.** *Fix $(s, \alpha)$ and $f_{\text{ho}}(\omega)$ as in (1.10), and $0 < \nu < (p-1)(s - s_{pr})$. Let $u_{\text{ho}}$ be the solution to the system*

$$\begin{cases} (i\partial_T + \Delta)u_{\text{ho}} = |u_{\text{ho}}|^{p-1}u_{\text{ho}}, \\ \qquad u_{\text{ho}}(0) = f_{\text{ho}}(\omega). \end{cases} \tag{1.11}$$

*Then there exists a set $Z \subset \Theta$ with $\mathbb{P}(Z) \leq C_\theta e^{-N^\theta}$, where $\theta$ is a small constant (ultimately determined by $(d, p, s)$, and independent of $N$) and $C_\theta$ is a constant determined by $\theta$, such that when $\omega \notin Z$, the solution $u_{\text{ho}}$ exists up to time $T = N^\nu$. Moreover, for some real valued gauge function $B(T)$, we have*

$$\sup_{0 \leq T \leq N^\nu} \|u_{\text{ho}}(T) - e^{-iB(T)}e^{iT\Delta}u_{\text{ho}}(0)\|_{H^s} \leq N^{-\theta}. \tag{1.12}$$

*Remark* 1.7. When $s_{cr} > 1$, i.e. (1.1) is $H^1$ supercritical in the usual (deterministic) sense, it is not a priori known whether $u_{\text{ho}}$ exists up to time $N^\nu$. Even in the $H^1$ subcritical case where $u_{\text{ho}}$ exists for all time, Theorem 1.6 demonstrates the highly nontrivial fact that there is no energy cascade in $u_{\text{ho}}$, (almost) until this very long time $T = N^{(p-1)(s-s_{pr})}$.

In comparison, for deterministic data, say when $g_k(\omega)$ in (1.10) are replaced by 1, one can describe the asymptotic behavior of $u_{\text{ho}}$ only up to time[1] $O(N^{(p-1)(s-s_{cr})})$ provided $s > s_{cr}$; see [17, 38] for the $p = 3$ case. Therefore, the randomization (1.10) effectively *extends the time of perturbative regime* for the given homogeneous initial data. In essence, this is the same as Theorem 1.1, where we keep the time of the perturbative regime constant (namely 1), and randomization allows us to increase the size of the initial data at a given frequency (equivalent to reducing regularity).

*Remark* 1.8. Since we are on the square torus $\mathbb{T}^d$, the behavior of $u_{\text{ho}}$ at long time is dominated by exact resonances. If $\mathbb{T}^d$ is replaced by a *generic irrational* torus, then we may expect Theorem 1.6 to hold on even longer time intervals, conjecturally up to $N^{2(p-1)(s-s_{pr})}$, at least for some range of $s$. This, after rescaling, would correspond to justifying the wave kinetic equation up to the *kinetic timescale* in the context of weak turbulence, which is still open at this point[2] (despite the recent success in [33] of the first author with Z. Hani; see also [18, 28]).

*Remark* 1.9. Unlike Theorem 1.1, in Theorem 1.6 we do not need the Wick ordering (1.6). Indeed, the worst contributions in the context of Theorem 1.1 (as well as the non-existence result of [51]), which can only be rescued by Wick ordering, are the high-low interactions where the high frequencies form a pairing and produce the mass term; for random homogeneous data there is no distinction between high and low frequencies, so such terms will not be a concern.

*Remark* 1.10. Our Theorem 1.1 provides the short time theory, and Theorem 1.6 yields also long time control for random homogeneous data. To pass from local to global (in time) results one needs to combine the random tensor theory with other methods, as is also the case in all previous works on parabolic and dispersive equations. See Section 9.2.4 for a discussion.

[1] In fact there is energy cascade in $u_{\text{ho}}$ at time $N^{(p-1)(s-s_{cr})}$, dictated by the *continuous resonance* (CR) equation; see [17, 38]. Note that, the approximation leading to the CR equation does not work in our case as randomization destroys the differentiability in rescaled Fourier space; in fact the solution $u_{\text{ho}}$ in Theorem 1.6 has no energy cascade at this time. It is currently unknown whether the solution $u_{\text{ho}}$ in Theorem 1.6, or the corresponding ensemble average, satisfies some effective equation at time $N^{(p-1)(s-s_{pr})}$.

[2] We remark that, after the submission of this manuscript, the justification of the wave kinetic equation has been done by the first author and Z. Hani in [34] in certain regimes, including the case $s = s_{pr}$ and $\nu = 0$ (the case $s > s_{pr}$ still remains open). This is a probabilistically critical result and is not covered by Theorem 1.6.

1.2. **The probabilistic scaling.** The probabilistic criticality threshold $s_{pr}$ plays a key role in the local theory of (1.1) with random initial data. In this section we recall a heuristic justification of this fact, which originally appeared in [36].

Start with (1.1) on $\mathbb{R}\times\mathbb{T}^d$, for simplicity we will replace $W^p(u)$ by the artificial nonlinearity $\mathcal{N}_{\mathrm{np}}$, defined by

$$(\mathcal{N}_{\mathrm{np}}(u))_k := \sum_{k_1-\cdots+k_p=k} u_{k_1}\overline{u_{k_2}}\cdots u_{k_p}, \tag{1.13}$$

assuming there is no pairing (i.e. $k_j\not\in\{k_{j'},k\}$ for any odd $j$ and even $j'$). Let the initial data $u(0)\in H^s$, to prove local well-posedness in $H^s$, one would like to control

$$u^{(1)}(t) := \int_0^t e^{i(t-t')\Delta}\mathcal{N}_{\mathrm{np}}(e^{it'\Delta}u(0))\,\mathrm{d}t', \tag{1.14}$$

namely the first nonlinear iteration, in $H^s$, at time $|t|\sim 1$. In the deterministic setting, if

$$u(0)=N^{-\alpha}\sum_{|k|\sim N}e^{ik\cdot x},\qquad \alpha=s+\frac{d}{2}, \tag{1.15}$$

then $\|u(0)\|_{H^s}\sim 1$. By (1.14) we can calculate the Fourier coefficients $u_k^{(1)}(t)$ of $u^{(1)}(t)$, where

$$u_k^{(1)}(t)\sim N^{-p\alpha}\sum_{\substack{k_j\in\mathbb{Z}^d,|k_j|\sim N\\ k_1-\cdots+k_p=k}}\frac{1}{\langle\Omega\rangle},\quad \Omega:=|k|^2-|k_1|^2+\cdots-|k_p|^2 \tag{1.16}$$

for $|k|\sim N$ and $|t|\sim 1$. We may restrict to (say) $\Omega=0$ in (1.16), and a dimension counting argument shows that the inner sum has size $N^{pd-d-2}$, hence

$$\|u^{(1)}(t)\|_{H^s}\sim N^{(p-1)(\frac{d}{2}-s)-2};\qquad \|u^{(1)}(t)\|_{H^s}\lesssim 1\Leftrightarrow s\geq\frac{d}{2}-\frac{2}{p-1}:=s_{cr}. \tag{1.17}$$

Indeed in the deterministic setting, (1.1) is locally well-posed for $s>s_{cr}$, and ill-posed for $s<s_{cr}$.

Now we switch to the random setting, where instead of (1.15) we have

$$u(0)=N^{-\alpha}\sum_{|k|\sim N}g_k(\omega)e^{ik\cdot x}, \tag{1.18}$$

where $g_k(\omega)$ are i.i.d. centered normalized Gaussians, and instead of (1.16) we have

$$u_k^{(1)}(t)\sim N^{-p\alpha}\sum_{\substack{k_j\in\mathbb{Z}^d,|k_j|\sim N\\ k_1-\cdots+k_p=k}}\frac{1}{\langle\Omega\rangle}g_{k_1}\overline{g_{k_2}}\cdots g_{k_p}. \tag{1.19}$$

Again we may restrict to $\Omega=0$; due to the square root cancellation in (1.19), now with high probability the inner sum only has size $N^{(pd-d-2)/2}$, hence instead of (1.17) we have

$$\|u^{(1)}(t)\|_{H^s}\sim N^{-(p-1)s-1};\qquad \|u^{(1)}(t)\|_{H^s}\lesssim 1\Leftrightarrow s\geq-\frac{1}{p-1}:=s_{pr}. \tag{1.20}$$

This justifies the role of $s_{pr}$ in Theorem 1.1, and explains why almost-sure local well-posedness is plausible, in the *probabilistically subcritical* range $s>s_{pr}$.

*Remark* 1.11. Theorem 1.1 establishes almost-sure local well-posedness when $s>s_{pr}$. In the probabilistically supercritical range $s<s_{pr}$, we believe (1.1) is almost surely ill-posed, in the sense that almost surely, the approximations $u_N$ defined in (1.7) *do not converge* in $C_t^0H_x^{s-}[-\tau,\tau]$ for *any*

$\tau > 0$. This seems out of reach with the current methods, but some weaker results (for example failure of convergence[1] in $C_t^\iota H_x^{s-}$ for any $\iota > 0$) might be possible.

*Remark* 1.12. The notion of *parabolic scaling* (also known as the *super-renormalizability* scaling in quantum field theory, see [22]) is central in the works on local theory of singular parabolic SPDEs, see for example [52, 46]. For the analog of (1.1), namely

$$(\partial_t - \Delta)u = \widetilde{W}^p(u) + \zeta, \qquad \zeta = \text{spacetime white (or colored) noise} \tag{1.21}$$

(where $\widetilde{W}^p$ is some renormalization beyond Wick ordering, see (1.22) below), this scaling has critical exponent $s_{pa} := -2/(p-1)$. Note that $s_{pa}$ is strictly lower than $s_{pr}$, which reflects a fundamental difference between Schrödinger and heat equations. See Section 9.1 for more detailed discussions.

1.3. **Propagating randomness: Earlier works.** The heuristics of Section 1.2 rely on the assumption that the Fourier coefficients of $u(0)$ are *independent*; this is no longer satisfied by $u(t)$, as soon as $t > 0$. Therefore, the key to the proof of Theorem 1.1 is to propagate the randomness of the initial data, for the anticipated amount of time, in such a way that the square cancellation in Section 1.2 remains valid.

The idea of propagating randomness, interpreted in one way or another, has been central in all previous works concerning local well-posedness in the random setting. In this section we briefly review the existing approaches, especially those developed in the past few years.

1.3.1. *The method of Bourgain and Da Prato-Debussche.* The important early results in this direction are proved by Bourgain [11] (for random initial data) and later by Da Prato-Debussche [29] (for additive noise). The idea is to propagate the random initial data (or the noise term) linearly, which preserves all the independence properties, and treat the nonlinearity as a perturbation.

For example, the equation studied in [11] is (1.1) with $(d, p) = (2, 3)$ and Gibbs measure initial data (i.e. $\alpha = 1$ in (1.2)), which is barely supercritical in the deterministic sense and subcritical in the probabilistic sense. In [11] Bourgain constructed the solution as $u = u_{\text{lin}} + w$, where $u_{\text{lin}} = e^{it\Delta}u(0)$ is the linear evolution which enjoys the same randomness properties as the initial data $u(0) = f(\omega)$, and the remainder $w$ has improved regularity, say $C_t^0 H_x^\sigma$ with $\sigma > 0$, thus becoming subcritical in the deterministic sense. Then the classical fixed point analysis together with large deviation estimates apply to control the hybrid nonlinearity—of the difference equation that $w$ satisfies—containing interactions of $u_{\text{lin}}$ with $w$. The situation in [29] is similar, except that Schrödinger is replaced by Navier-Stokes, and $u_{\text{lin}}$ is replaced by the linear evolution of the noise term.

Until recent years, the methods of Bourgain and of Da Prato-Debussche have been the dominant strategy in the study of local well-posedness theory for random PDEs. The weakness of this approach is that the improved regularity of the nonlinear contribution $w$ may not be enough for the deterministic theory to be applicable, especially when one gets close to probabilistic criticality. One may try to enhance this by moving to higher-order variants and bringing in self-interactions of $u_{\text{lin}}$, see [9, 64], but in many situations (like in [9]), there is an upper bound for all the regularity improvements obtained in this way, which may still fall short of the deterministic threshold.

[1] In comparison, when $s > s_{pr}$, the proof of Theorem 1.1 easily implies the convergence of $u_N$ (including the $u_\lambda$ in Remark 1.4) in $C_t^\iota H_x^{s-}$ for some $\iota > 0$ ultimately determined by $(d, p, s)$.

1.3.2. *The theory of regularity structures.* The theory of regularity structures was developed by Hairer [52, 53] in the context of singular parabolic SPDEs, to provide a natural and mathematically rigorous notion of solutions to such equations and prove their local well-posedness. The theory is based on the *local-in-space* properties of solutions at fine scales, hence it is well adapted to parabolic equations. It builds a general theory of distributions by means of an abstract generalization of local Taylor expansions of problem-dependent profiles (e.g. the spacetime white or colored noise, and self-interactions and Duhamel iterations thereof), in order to make sense of the equation—in particular the products of rough distributions emerging from the singular parabolic SPDE. Furthermore, the solutions obtained can be approximated locally to arbitrarily high degree by linear combinations of this fixed family of problem-dependent profiles. These expansions in the context of singular parabolic SPDEs should be compared and contrasted to the ones that we obtain in Theorem 1.1 (more precisely (8.5)).

The theory of regularity structures establishes local well-posedness results in the sense of convergence of canonical smooth approximations. When taking limits of such approximations, a suitable renormalization in the form of divergent counterterms is usually needed. Sometimes (for example in the dynamical $\Phi_2^4$ model, see Remark 1.3) this is just the Wick ordering, but for more sophisticated equations further renormalizations become necessary. A nice feature of the regularity structures theory is that these renormalization constants can always be calculated using the profiles defined for the specific equation.

We illustrate this renormalization process following [52, 54, 55] where Hairer studies the dynamical $\Phi_3^4$ model. Here the canonical smooth approximations $u_\varepsilon$ satisfy the renormalized equations with the spacetime white noise $\zeta$ replaced by its regularization $\zeta_\varepsilon$, namely

$$(\partial_t - \Delta)u_\varepsilon = 3(C_1 - 3C_2)u_\varepsilon - u_\varepsilon^3 + \zeta_\varepsilon, \tag{1.22}$$

where $C_1 \sim \varepsilon^{-1}$ corresponds to Wick ordering, and $C_2 \sim \log \varepsilon$ is an additional renormalization constant. The problem-dependent profiles are in the following space

$$T = \langle \bullet, \Psi^3, \Psi^2, \mathcal{I}(\Psi^3)\Psi^2, \Psi, \mathcal{I}(\Psi^3)\Psi, \mathcal{I}(\Psi^2)\Psi^2, x_i\Psi^2, \mathbf{1}, \mathcal{I}(\Psi^3), \mathcal{I}(\Psi^2)\Psi, \cdots \rangle, \tag{1.23}$$

with symbols ordered in increasing order of regularity and $\langle \cdot \rangle$ used to denote the linear span. The symbol $\bullet$ represents the regularized noise $\zeta_\varepsilon$, $x_i$ represent the coordinates of $x$, and $\Psi = (\partial_t - \Delta)^{-1}\bullet$, $\Psi^3$ and $\Psi^2$ represent the Wick renormalized[1] powers $(\Psi)^3$ and $(\Psi)^2$, etc. The renormalization constants are then calculated from these profiles, such as $C_1 = \mathbb{E}(\Psi)^2 \sim \varepsilon^{-1}$, $C_2 = \mathbb{E}(\mathcal{I}(\Psi^2) \cdot \Psi^2) \sim \log \varepsilon$, and convergence of $u_\varepsilon$ as $\varepsilon \to 0$ is proved for initial data in $C^\alpha$, $-\frac{2}{3} < \alpha < -\frac{1}{2}$.

In a series of papers [15, 16, 21], the regularity structures theory has been extended to general parabolic equations and now covers the whole subcritical range $s > s_{pa}$ relative to the parabolic scaling (as in Remark 1.12 above); for an example see the $\Phi_{4-\delta}^4$ model in [15].

*Remark* 1.13. We remark that, though in principle the regularity structures theory can cover the whole subcritical range for singular parabolic SPDEs, in reality there is an additional obstacle that sometimes requires slightly higher regularity when the noise is rougher than white. This obstacle can be traced back to the rough path theory [40], and is linked to the high-high-to-low frequency interactions. In the Schrödinger case, however, such obstacle is absent as these interactions can be treated in the same way as the main term, hence our theorem covers the full subcritical range

[1] In particular, $\Psi^2 = (\Psi)^2 - \mathbb{E}(\Psi)^2$ and $\Psi^3 = (\Psi)^3 - 3(\mathbb{E}(\Psi)^2) \cdot \Psi$.

$s > s_{pr}$, which is in parallel with the full subcritical range $s > s_{pa}$ in parabolic case, for all dimensions $d$ and powers $p$.

1.3.3. *The para-controlled calculus and the renormalization group technique.* The theory of para-controlled calculus put forward by Gubinelli, Imkeller and Perkowski [46] (see also Catellier and Chouk [20]) and the work of Kupiainen [58] based on renormalization group techniques provide alternative approaches to proving local well-posedness for singular parabolic SPDEs. The theory of para-controlled calculus, based on paradifferential calculus, leverages the key observation that the lack of regularity for $w$ in the approaches of Bourgain and of Da Prato-Debussche, as described at the end of Section 1.3.1, is only due to the high-low interactions where the high frequencies come from $u_{\mathrm{lin}}$ (or self-interactions and Duhamel evolutions thereof) and the low frequencies come from $w$. In this theory, such high-low interaction terms $X$ are *para-controlled* by the high-frequency inputs (for example $u_{\mathrm{lin}}$), in the sense that[1] $X = \pi_{>}(u_{\mathrm{lin}}, Y) + Z$ with $Z$ being smoother than $X$. Such terms, despite having insufficient regularity, are shown to enjoy similar randomness structures as $u_{\mathrm{lin}}$ itself, which allows for a fixed point argument, where $X$ is constructed in some low regularity space, and the remainder $Z$ is constructed in a higher regularity space. We refer the reader to [20, 48, 49, 50, 62, 6, 5, 47, 14] and references therein for nice expositions of these ideas and some other recent developments as well as a higher order variant of this method.

1.4. **Random averaging operators.** In view of these breakthrough works described in Sections 1.3.2 and 1.3.3 that deal with *parabolic* equations, it would be natural to think that something similar can be done in the context of *dispersive* equations. However, there are fundamental differences between dispersive (say Schrödinger) and heat equations, preventing these methods from being applicable (for more comparisons see Section 9.1):

(a) The heat equation is compatible with local-in-space analysis, as is clear from the maximum principle or the off-diagonal exponential decay for the heat kernel. The Schrödinger equation does not have these properties, which makes the application of the theory of regularity structures impossible, as the latter is based on pointwise Taylor expansions in physical space.

(b) Likewise, the heat equation is compatible with $C^\alpha$ (Hölder) spaces; in fact both the regularity structures theory and the para-controlled calculus rely heavily on such norms. On the other hand the Schrödinger flow is unbounded on $C^\alpha$, and is bounded only on $H^s$ type spaces, which require a lot more derivatives to reach the same scaling as $C^\alpha$.

(c) The heat equation gains two spatial derivatives in terms of the fundamental solution $(\partial_t - \Delta)^{-1}$ (wave gains one), while the Schrödinger equation gains none. The smoothing is seen only in terms of twisted temporal regularity by using $X^{s,b}$ type norms, which is clearly not compatible with either the regularity structures theory or the para-controlled calculus.

In our previous work [36], which studies (1.1) with $d = 2$, arbitrary $p \geq 5$ and Gibbs measure initial data ($\alpha = 1$), we introduced the method of *random averaging operators*. The idea is to take the high-low interaction $X$ described in Section 1.3.3, but instead of putting it in a low regularity space (as is done in the para-controlled calculus), we write it as an *operator* applied to the high frequency linear evolution $u_{\mathrm{lin}}$:

$$X = \sum_{N} \sum_{L \ll N} \mathcal{P}_{NL}(\Delta_N u_{\mathrm{lin}}), \tag{1.24}$$

[1] Here $\pi_{>}$ is the standard Bony paraproduct.

where $N$ and $L$ are the frequencies of the high and low inputs. The operator $\mathcal{P}_{NL}$, whose coefficients are independent with the modes of $\Delta_N u_{\text{lin}}$, contains all the randomness information of the low frequency components of $u$, which is carried by two operator norm estimates

$$\|\mathcal{P}_{NL}\|_{\text{OP}} \leq L^{-\delta_0}, \quad \|\mathcal{P}_{NL}\|_{\text{HS}} \leq N^{\frac{1}{2}+\delta_1} L^{-\frac{1}{2}}, \tag{1.25}$$

where $\delta_1 \ll \delta_0 \ll 1$. Here $\mathcal{P}_{NL}$ is viewed as a linear operator between two Hilbert spaces that can be $L^2$ or $X^{s,b}$ depending on the context, $\|\cdot\|_{\text{OP}}$ and $\|\cdot\|_{\text{HS}}$ are respectively operator and Hilbert-Schmidt norms.

The method of random averaging operators, compared to the regularity structures theory and the para-controlled calculus, has three advantages relative to the three difficulties listed above, which makes it suitable for the study of Schrödinger equations:

(a) The operator $\mathcal{P}_{NL}$ is a global-in-space object (in fact it is defined on the Fourier side), which is consistent with the non-local setting of Schrödinger equations;

(b) The role of $C^\alpha$ norms is replaced by (essentially) the $L^2 \to L^2$ operator norms, which is compatible with the well-established $L^2$ theory;

(c) The analysis for $\mathcal{P}_{NL}$ is performed in the category of $X^{s,b}$ spaces, which allows one to exploit the smoothing of the Schrödinger fundamental solution.

By applying this method, we have been able to propagate the randomness of $\mathcal{P}_{NL}$ in terms of the above operator norm bounds, as well as control the remainder in a deterministically subcritical space, leading to the full resolution of the Gibbs measure problem in 2D. See [36].

1.5. **Random tensors.** The core of this work is a broad extension of the method of random averaging operators, which we call the theory of *random tensors*. A detailed introduction to this theory will be given in Section 2, here we will restrict our attention to only the motivation.

Start by considering the random averaging operators; roughly speaking, the frequency $N$ piece of the high-low interaction $X$ in (1.24) is given by $\Delta_N X = \mathcal{P}_N(\Delta_N u_{\text{lin}})$ where $\mathcal{P}_N = \sum_L \mathcal{P}_{NL}$. In Fourier space it can be written as

$$(\Delta_N X)_k(t) = e^{-i|k|^2 t} \sum_{\langle k'\rangle \sim N} h_{kk'}(t) \frac{g_{k'}(\omega)}{\langle k'\rangle},$$

where subscripts denote Fourier coefficients, and $h_{kk'}(t)$ is essentially the kernel of the operator $\mathcal{P}_N$. For fixed $t$ this is a random matrix, or $(1,1)$ tensor, that depends on the low frequency components of the solution.

Now, to prove Theorem 1.1 we will need higher order expansions. This naturally leads to the multilinear expressions (here we denote $(u^+, u^-) = (u, \overline{u})$ as usual)

$$\Psi_k = \sum_{k_1,\cdots,k_r} h_{kk_1\cdots k_r} \prod_{j=1}^{r} \langle k_j\rangle^{-\alpha} g_{k_j}^{\pm}(\omega), \tag{1.26}$$

as well as the associated random $(r,1)$ tensors $h = h_{kk_1\cdots k_r}$, which depend on the low frequency components of the solution. For simplicity, in (1.26) we have omitted the dependence on $t$. We always assume there is no *pairing*, i.e. $k_{j'} \neq k_j$ if the corresponding $\pm$ signs are opposite.

Notice that, the product of $\Psi$ with another expression

$$\Psi'_{k'} = \sum_{k'_1,\cdots,k'_s} h'_{k'k'_1\cdots k'_s} \prod_{j=1}^{s} \langle k'_j\rangle^{-\alpha} g_{k'_j}^{\pm}(\omega)$$

can be written as a linear combination of similar multilinear expressions, depending on the possible set of pairings between $\{k_1, \cdots, k_r\}$ and $\{k_1', \cdots, k_s'\}$. For example, if

$$\Psi_k = \sum_{a,b,c,d} h_{kabcd} \cdot \langle a\rangle^{-\alpha} g_a \cdot \langle b\rangle^{-\alpha} \overline{g_b} \cdot \langle c\rangle^{-\alpha} g_c \cdot \langle d\rangle^{-\alpha} \overline{g_d},$$
$$\Psi_{k'}' = \sum_{u,v,w} h_{k'uvw}' \cdot \langle u\rangle^{-\alpha} \overline{g_u} \cdot \langle v\rangle^{-\alpha} g_v \cdot \langle w\rangle^{-\alpha} g_w,$$

and in the summation we assume $a = u$ and $b = w$, then we have one representative component of $\Psi \cdot \Psi'$ being

$$\begin{aligned}\Phi_m &= \sum_{c,d,v} H_{mcdv} \cdot \langle c\rangle^{-\alpha} g_c \cdot \langle d\rangle^{-\alpha} \overline{g_d} \cdot \langle v\rangle^{-\alpha} g_v, \\ H_{mcdv} &\sim \sum_{k+k'=m} \sum_{a,b} \langle a\rangle^{-\alpha} \langle b\rangle^{-\alpha} h_{kabcd} \cdot \langle a\rangle^{-\alpha} \langle b\rangle^{-\alpha} h_{k'avb}',\end{aligned} \tag{1.27}$$

where we have replaced $|g_a|^2$ and $|g_b|^2$ by 1. The process of going from $(h, h')$ to $H$ as above will be called *merging*, which gives the main algebraic structure of the tensors studied in this work.

For purposes related to independence of Fourier coefficients (which will be explained in Section 2.2.3), we also need another process called *trimming*, which means contracting against free Gaussians, namely going from $h = h_{kk_1\cdots k_r}$ to

$$h_{kk_1\cdots k_s}' = \sum_{k_{s+1},\cdots k_r} h_{kk_1\cdots k_r} \prod_{j=s+1}^{r} \langle k_j\rangle^{-\alpha} g_{k_j}^{\pm},$$

where $1 \leq s \leq r$. Note that $\Psi_k$ defined by (1.26) is formally invariant under trimming.

Now, as in [36], the central objects in our work will be the tensors $h$ (instead of the multilinear expressions $\Psi$), as well as suitable $L^2 \to L^2$ operator norms for these tensors. The theory of random tensors then provides a natural framework for studying such tensors, in particular proving estimates for such norms that are consistent with merging and trimming. In Section 2 below we provide a more detailed discussion of this theory, and application to the proof of Theorem 1.1.

*Remark* 1.14. The reason for introducing the random tensor theory is that, as the problem gets closer to probabilistic criticality, what is deemed a remainder term in the random averaging operator method will no longer have enough regularity, thus a higher order expansion is needed. In fact, by the arguments in Section 1.2, each iteration of the equation (1.1) gains regularity $\Delta s = (p-1)(s - s_{pr})$, so the order of the expansion needed would be $\sim 1/\Delta s$. This could be used as a measure for the difficulty of the problem, which goes to infinity when $p$ is fixed and $s \to s_{pr}$.

For example, for the problem studied in [36] one has $d = 2$ and $(p, s, s_{pr}) = (2r+1, 0, -1/(2r))$, hence $\Delta s = 1$ for any $r$ (note that in order for $s - s_{pr} \to 0$ we need $r \to \infty$) so the random averaging operator is always enough. However if $d = 3$, $p = 3$ and $s \to s_{pr} = -1/2$, then $\Delta s \to 0$ and the use of random tensors becomes necessary.

1.6. **Acknowledgement.** The authors would like to thank Hendrik Weber for helpful comments regarding the regularity structures theory and the reference [22]. They would also like to thank the referees for their useful comments and suggestions.

## 2. Overview of the theory

This section contains an overview of the theory of random tensors in the context of the proof of Theorem 1.1. We start with the definition and norms in Section 2.1, then develop the algebraic structure and main tools in Section 2.2. In Section 2.3 we introduce a simplified model, which is analyzed in Section 2.4 using tools from our theory. In Section 2.5 we explain the changes needed when moving to full generality, and finally in Section 2.6 we list an outline for the rest of the paper.

### 2.1. **Tensors and tensor norms.**

As discussed in Section 1.5, the central objects in this work are tensors and their $L^2 \to L^2$ operator norms. We therefore start with the following definition.

**Definition 2.1.** Let $A$ be a finite index set, we will denote $k_A = (k_j : j \in A)$. A *tensor* $h = h_{k_A}$ is a function $(\mathbb{Z}^d)^A \to \mathbb{C}$, with $k_A$ being the input variables. The *support* of $h$ is the set of $k_A$ such that $h_{k_A} \neq 0$. These tensors are usually random (i.e. depend on $\omega$ which belongs to the ambient probability space $\Theta$, though we may omit this dependence), hence the name *random tensors*.

A *partition* of $A$ is a pair of sets $(B, C)$ such that $B \cup C = A$ and $B \cap C = \varnothing$. For such $(B, C)$ define the norm $\|\cdot\|_{k_B \to k_C}$ by

$$\|h\|_{k_B \to k_C}^2 = \sup\bigg\{\sum_{k_C}\bigg|\sum_{k_B} h_{k_A} \cdot z_{k_B}\bigg|^2 : \sum_{k_B} |z_{k_B}|^2 = 1\bigg\}.$$

By duality we have that

$$\|h\|_{k_B \to k_C} = \sup\bigg\{\bigg|\sum_{k_B, k_C} h_{k_A} \cdot z_{k_B} \cdot y_{k_C}\bigg| : \sum_{k_B} |z_{k_B}|^2 = \sum_{k_C} |y_{k_C}|^2 = 1\bigg\}, \tag{2.1}$$

hence $\|h\|_{k_B \to k_C} = \|h\|_{k_C \to k_B} = \|\overline{h}\|_{k_B \to k_C}$. If $B = \varnothing$ or $C = \varnothing$ we get the norm $\|\cdot\|_{k_A}$ defined by

$$\|h\|_{k_A}^2 = \sum_{k_A} |h_{k_A}|^2.$$

Note that trivially $\|h\|_{k_B \to k_C} \leq \|h\|_{k_A}$.

Finally, we define a *subpartition* to be a pair $(B, C)$ such that $B \cup C \subset A$ and $B \cap C = \varnothing$. In such case let $E = A \backslash (B \cup C)$, then $(B, C)$ is a partition of $A \backslash E$ so we can define

$$\|h\|_{k_B \to k_C} = \sup_{k_E} \|h_{(k_E, \cdot)}\|_{k_B \to k_C}. \tag{2.2}$$

*Remark* 2.2. In the main proof the tensors may depend on other parameters, such as $t$ or $(k_F, \lambda_F)$, where $\lambda_F = (\lambda_j : j \in F)$, for some set $F$; in such cases we will write respectively $h_{k_A} = h_{k_A}(t)$ or $h_{k_A} = h_{k_A}(k_F, \lambda_F)$. Moreover, the norm (2.2) is designed only to treat some degenerate cases, so it will not appear in the simplified model of this section.

### 2.2. **Tensor algebra and basic tools.**

In this section we develop the algebra of random tensors given by merging and trimming as described in Section 1.5, and some important estimates which are the basic tools of our theory. The precise versions will be given in Sections 3.2 and 4.4 below. First we record the definition of *pairing*.

**Definition 2.3.** Define $u^\zeta$ for $\zeta \in \{\pm\}$ by $(u^+, u^-) = (u, \overline{u})$ (in doing algebra we may view such $\zeta$ as $\pm 1$). Given $k_i, k_j \in \mathbb{Z}^d$ with associated signs $\zeta_i, \zeta_j \in \{\pm\}$, we say $(k_i, k_j)$ is a *pairing* if $\zeta_i + \zeta_j = 0$ and $k_i = k_j$. We say it is *over-paired* (or an *over-pairing*) if $k_i = k_j = k_l$ for some $l \notin \{i, j\}$.

2.2.1. *Semi-products and merging.* As described in Section 1.5, our theory will focus on the analysis of the tensors $h_{kk_A}$ and associated multilinear expressions[1]

$$\Psi_k := \sum_{k_A} h_{kk_A} \prod_{j\in A} \langle k_j\rangle^{-\alpha} g_{k_j}^{\pm}, \tag{2.3}$$

where $k_A = (k_j : j \in A)$ as in Definition 2.1 and $k_{j'} \neq k_j$ when the signs are the opposite, as well as the merging and trimming operations loosely described in that section.

In fact the merging operation can be viewed as a special case of taking what we may call *semi-products* for two or more tensors, which means taking tensor products and contracting over the given set of repeated indices—note that, the repeated indices can be indices appearing in both tensors, or the result of specific pairings between the tensors. For example, if $h_{abcde}$ and $h'_{uvawx}$ are two tensors, then their semi-product, under the assumptions $b = u$ and $c = w$, will be[2]

$$H_{devx} = \sum_{a,b,c} h_{abcde} h'_{bvacx}.$$

In general, suppose $h_{k_A}$ are $h'_{k_B}$ are two tensors, and we have a particular set of repeated indices (coming from $A \cap B$ or pairings between $k_A$ and $k_B$). We may then assume these repeated indices belong to both $A$ and $B$, and define the corresponding semi-product as

$$H_{k_{A\Delta B}} = \sum_{k_{A\cap B}} h_{k_A} h'_{k_B}.$$

This can easily be generalized to semi-products of more than two tensors, for example the semi-product of the three tensors $h_{abcd}$, $h'_{aefg}$ and $h''_{cuv}$ under the assumptions $b = f$ and $g = v$ is

$$H_{deu} = \sum_{a,b,c,g} h_{abcd} h'_{aebg} h''_{cug}.$$

Note that for simplicity we are not considering over-pairings where an index is repeated two or more times, but such situations do appear in the actual definition of merging and need special treatment (though they are only associated with degenerate situations which are always much easier). See Definition 3.6 for details.

Now, with the notion of semi-products, we can define (in this simple case without over-pairing) the merging of finitely many tensors $h^{(1)}, \cdots, h^{(r)}$ via a *base tensor* $h$ as follows:

**Definition 2.4** (Merging: simple version)**.** Let $h^{(j)} = h^{(j)}_{k_j k_{A_j}}$ be tensors, where $1 \leq j \leq r$, $A_j$ are pairwise disjoint, and let $h = h_{kk_1\cdots k_r}$ be the base tensor. Also fix a set of pairings among the sets $A_1, \cdots, A_r$ (which creates paired indices that will be viewed as repeated indices; as before, assume there is no pairing within each $A_j$ and no pairing involving more than two indices), then the merging of $h^{(1)}, \cdots h^{(r)}$ via $h$, assuming the given choice of pairings, is defined to be the semi-product of $\widetilde{h}^{(1)}, \cdots, \widetilde{h}^{(r)}$ and $h$, where (i) each $\widetilde{h}^{(j)}$ is $h^{(j)}$ multiplied by the product of $\langle k_{\mathfrak{l}}\rangle^{-\alpha}$ over all $\mathfrak{l} \in A_j$ that appear in some pairing, and (ii) in addition to the paired indices, each $k_j$ $(1 \leq j \leq r)$ is also viewed as a repeated index and is summed over, as it appears in both $h^{(j)}$ and $h$.

[1] In practice we will use $\eta_k = |g_k|^{-1} g_k$, which is uniformly distributed on the unit circle, instead of $g_k$.

[2] In practice we will also have a $\pm$ sign for each index of each tensor, for example a $+$ sign for the index $a$ in $h_{abcde}$ or a $-$ sign for the index $x$ in $h'_{uvawx}$. When precisely defining the merging operations, see Definition 3.6, we will restrict to the cases where for each repeated or paired index, its signs in the two tensors that it appears are the opposite (for example if the sign of $b$ in $h_{abcde}$ is $+$ then the sign of $u$ in $h'_{uvawx}$ must be $-$, assuming $b = u$). In this section (including the simple case Definition 2.4) we will ignore this issue for simplicity.

For example, if $h^{(1)} = h^{(1)}_{k_1abcd}$, $h^{(2)} = h^{(2)}_{k_2efg}$, $h^{(3)} = h^{(3)}_{k_3uvw}$ and $h = h_{kk_1k_2k_3}$, then the merging of $h^{(1)}$, $h^{(2)}$ and $h^{(3)}$ via $h$, under the assumptions $a = w$, $b = f$ and $g = v$, will be

$$H_{kcdeu} = \sum_{k_1,k_2,k_3,a,b,g} h_{kk_1k_2k_3} \cdot \langle a\rangle^{-\alpha}\langle b\rangle^{-\alpha} h^{(1)}_{k_1abcd} \cdot \langle b\rangle^{-\alpha}\langle g\rangle^{-\alpha} h^{(2)}_{k_2ebg} \cdot \langle g\rangle^{-\alpha}\langle a\rangle^{-\alpha} h^{(3)}_{k_3uga}.$$

Similarly, the example (1.27) which gives a component of the product $\Psi \cdot \Psi'$, is the merging of $h_{kabcd}$ and $h'_{k'uvw}$ via $\widetilde{h}_{mkk'} = \mathbf{1}_{m=k+k'}$, under the assumption $a = u$ and $b = w$.

The general version of Definition 2.4, which includes the signs of indices, dependence on other parameters and additional structures, as well as over-pairings, will be given in Definition 3.6.

2.2.2. *Key bilinear and multilinear bounds.* Our first basic tool is the following lemma (together with the multilinear version thereof), where the $\|\cdot\|_{k_B\to k_C}$ norms for semi-products of tensors, as defined in Definition 2.1, are estimated by the $\|\cdot\|_{k_B\to k_C}$ norms of the individual tensors.

**Proposition 2.5** (A bilinear estimate)**.** *Let $h_{k_A}$ and $h'_{k_B}$ be two tensors, assume that all repeated indices are already in $A\cap B$. Then for any partition $(X, Y)$ of $A\Delta B$, the semi-product $H$ of $h$ and $h'$ satisfies that*

$$\|H\|_{k_X\to k_Y} \leq \|h\|_{k_{(X\cup B)\cap A}\to k_{Y\cap A}} \cdot \|h'\|_{k_{X\cap B}\to k_{(Y\cup A)\cap B}}.$$

For example, we have

$$\|H\|_{dv\to ex} \leq \|h\|_{abcd\to e} \cdot \|h'\|_{v\to xabc}, \quad \text{where} \quad H_{devx} = \sum_{a,b,c} h_{abcde} h'_{bvacx}.$$

Note that in the setting $b = u$ and $c = w$ as in Section 2.2.1, we can identify $h'_{bvacx}$ with $h'_{uvawx}$ and the norm $\|\cdot\|_{v\to xabc}$ with $\|\cdot\|_{v\to xauw}$; the same comment applies to Lemma 2.6 below.

An equivalent form of Proposition 2.5 will be stated and proved in Proposition 4.11.

**Proposition 2.6** (A multilinear estimate)**.** *Let $h^{(j)}_{k_{A_j}}$ $(1 \leq j \leq m)$ be tensors, assume all repeated indices coming from pairings between any $A_i$ and $A_j$ are already in $A_i \cap A_j$. Let $H = H_A$ be the semi-product of the $h^{(j)}$'s, where $A = A_1\Delta\cdots\Delta A_m$, then for any partition $(X, Y)$ of $A$ we have*

$$\|H\|_{k_X\to k_Y} \leq \prod_{j=1}^{m} \|h^{(j)}\|_{k_{(X\cap A_j)\cup B_j}\to k_{(Y\cap A_j)\cup C_j}}, \tag{2.4}$$

*where*

$$B_j = \bigcup_{\ell>j}(A_j\cap A_\ell), \quad C_j = \bigcup_{\ell<j}(A_j \cap A_\ell).$$

For example, we have

$$\|H\|_{e\to ud} \leq \|h\|_{abc\to d}\|h'\|_{eg\to ab}\|h''\|_{cug}, \quad \text{where} \quad H_{deu} = \sum_{a,b,c,g} h_{abcd} h'_{aebg} h''_{cug}.$$

An equivalent form of Proposition 2.6 will be stated and proved in Proposition 4.12.

2.2.3. *Trimming.* In the course the main proof, when considering the tensors $h_{kk_A}$ and associated multilinear expressions $\Psi_k$ as in (2.3), we will always assume that the tensor $h$ is *independent* with the Gaussians $g_{k_j}$, in order for the large deviation estimates (such as Lemma 4.4) to be applicable. In practice, this is guaranteed by requiring that $\langle k_j\rangle \geq R$ in the support of $h_{k_A}$ for some $R$, and that $h_{k_A}$ is a Borel function of $\{g_k : \langle k\rangle < R\}$. A problem then occurs, say when merging two tensors $h = h_{k_A}$ and $h' = h'_{k_B}$ with cutoffs $R_1$ and $R_2$ as above, because the merged tensor $H$ is a Borel function of $\{g_k : \langle k\rangle < \max(R_1, R_2)\}$ and may not be independent with $g_{k_j}$ for $j \in A\Delta B$. Because of this, we will introduce the operation of *trimming* as follows:

**Definition 2.7** (Trimming: simple version)**.** Let $h = h_{kk_A}$ be a tensor, assume for each $j \in A$ there is a dyadic $N_j$ such that $h$ is supported in $N_j/2 < \langle k_j\rangle \leq N_j$. Then, for any $R$, the trimming of $h$ at frequency $R$ is defined to be the contraction against free Gaussians, namely

$$h'_{kk_{A'}} = \sum_{k_{A\backslash A'}} h_{kk_A} \prod_{j\in A\backslash A'} \langle k_j\rangle^{-\alpha} g^{\pm}_{k_j},$$

where $A' = \{j \in A : N_j \geq R\}$. Note that those $g_{k_j}$ where $j \in A'$ are independent with those $g_{k_j}$ where $j \in A\backslash A'$. In particular we recover the expression $\Psi_k$ in (2.3) if $A' = \varnothing$.

For example, if $h = h_{kabcd}$, where $N_1/2 < \langle a\rangle \leq N_1$ etc., and assume $N_1 \leq N_3 < R \leq N_2 \leq N_4$, then the trimming of $h$ at frequency $R$ will be

$$h'_{kbd} = \sum_{a,c} h_{kabcd} \cdot \langle a\rangle^{-\alpha} g^{\pm}_a \cdot \langle c\rangle^{-\alpha} g^{\pm}_c.$$

The general version of Definition 2.7, which includes the signs of indices, as well as dependence on other parameters and additional structures, will be given in Definition 3.5.

2.2.4. *Method of descent.* Our second basic tool is the following lemma, where the $\|\cdot\|_{k_B\to k_C}$ norms for the contraction of a tensor against independent free Gaussians are estimated by the $\|\cdot\|_{k_B\to k_C}$ norms of this tensor. This inequality has an elegant form, and we believe it is of independent interest in the study of random matrices.

**Proposition 2.8.** *Let $h_{k_A}$ be a tensor, $A'$ be a subset of $A$ such that $\{g_{k_j} : j \in A\backslash A'\}$ is independent with $h_{k_A}$. Let $h' = h'_{k_{A'}}$ be the contraction of $h$ against the free Gaussians $\{g^{\pm}_{k_j} : j \in A\backslash A'\}$, namely*

$$h'_{k_{A'}} = \sum_{k_{A\backslash A'}} h_{k_A} \prod_{j\in A\backslash A'} g^{\pm}_{k_j},$$

*then for any partition $(X', Y')$ of $A'$, with high probability we have*[1]

$$\|h'\|_{k_{X'}\to k_{Y'}} \lesssim \sup_{(X,Y)} \|h\|_{h_X\to h_Y},$$

*where $(X, Y)$ runs over all partitions of $A$ such that $X' \subset X$ and $Y' \subset Y$.*

For example, under the independence assumption, with high probability we have

$$\|h'\|_{b\to d} \lesssim \max(\|h\|_{abc\to d}, \|h\|_{ab\to cd}, \|h\|_{bc\to ad}, \|h\|_{b\to acd}), \quad \text{where} \quad h'_{bd} = \sum_{a,c} h_{abcd} g^{\pm}_a g^{\pm}_c.$$

A more precise version of Proposition 2.8 will be stated and proved in Proposition 4.14. A slightly different version due to technical reasons will be stated and proved in Proposition 4.15.

[1] In practice this will have a small power loss $M^\theta$ where $\theta$ is an arbitrary small number and $M$ is the size of $k_A \in (\mathbb{Z}^d)^A$; see Propositions 4.14–4.15.

2.3. **A simple model.** We now turn to the proof of Theorem 1.1. In this section we introduce a much simplified model for (1.1) and (1.7) that still preserves the main difficulties.

First, we will replace the nonlinearities by the $\mathcal{N}_{\mathrm{np}}$ defined in (1.13). This $\mathcal{N}_{\mathrm{np}}$ is essentially the result of Wick ordering and a suitable gauge transform, but avoids the complications linked to over-pairings and deviation of mass around its expected value.

Second, we will remove the time variable. Indeed, if we believe our solution is close to a linear solution, and thus restrict to functions $u$ whose spacetime Fourier transform looks like $\widehat{u_k}(\xi) \sim u_k \cdot \psi(\xi + |k|^2)$ with some Schwartz function $\psi$ and some function $u_k$ of $k$ only, then by a formal calculation using Duhamel's formula, this $u_k$ will satisfy a fixed-point equation that essentially looks like

$$u_k = \frac{g_k}{\langle k\rangle^\alpha} - i \sum_{k_1 - \cdots + k_p = k;\, \Omega = 0} u_{k_1} \overline{u_{k_2}} \cdots u_{k_p}, \tag{2.5}$$

where $\Omega := |k|^2 - |k_1|^2 + \cdots - |k_p|^2$, and in (2.5) we also assume no-pairing as in (1.13).

Third, consistent with the setting of Section 2.2, in analyzing (2.5) we will disregard any over-pairings, and assume, when merging tensors, that no index is repeated more than once.

2.4. **The core ansatz.** We now start the analysis of (2.5). For simplicity we denote the terms on the right hand side of (2.5) by

$$\langle k\rangle^{-\alpha} g_k =: f_k, \qquad -i \sum_{k_1 - \cdots + k_p = k;\, \Omega = 0} u_{k_1} \overline{u_{k_2}} \cdots u_{k_p} =: \mathcal{M}_{\mathrm{np}}(u, \cdots, u)_k, \tag{2.6}$$

where $\mathcal{M}_{\mathrm{np}}$ is an $\mathbb{R}$-multilinear operator of degree $p$, so that (2.5) reads

$$u_k = f_k + \mathcal{M}_{\mathrm{np}}(u, \cdots, u)_k, \tag{2.7}$$

We also introduce the canonical truncations of (2.7), namely

$$(u_N)_k = \Pi_N f_k + \mathcal{M}_{\mathrm{np}}(u_N, \cdots, u_N)_k, \tag{2.8}$$

and define $y_N$ by

$$y_N = u_N - u_{N/2}; \quad u_N = \sum_{N' \leq N} y_{N'}. \tag{2.9}$$

Note that we do not put $\Pi_N$ before the nonlinearity in (2.8); however in this model we still assume $y_N$ and $u_N$ are supported in $\langle k\rangle \leq N$. Under these assumptions, $y_N$ satisfies the equation

$$(y_N)_k = \Delta_N f_k + \sum_{\max(N_1, \cdots, N_p) = N} \mathcal{M}_{\mathrm{np}}(y_{N_1}, \cdots, y_{N_p})_k. \tag{2.10}$$

The core ansatz for $y_N$ will be constructed by induction (assuming $u_{N'}$ and $y_{N'}$ are already defined for $N' < N$). Recall that in [36] the analogue ansatz for $y_N$ contains three parts: $\Delta_N f$ which corresponds to the linear evolution, the terms corresponding to the random averaging operators $\mathcal{P}_{NL}$, and a remainder $z_N$ of higher regularity. We start by a description of this simple case in Section 2.4.1. In order to prove Theorem 1.1 which covers the full subcritical range $s > s_{pr}$, in Section 2.4.2, we will further unravel the propagation of randomness from the remainder and make higher order expansions using the random tensors introduced in Sections 2.1–2.2.

2.4.1. *Random averaging operators.* As stated in Section 1.4, the random averaging operator $\mathcal{P}_{NL}$ describes the high-low interaction, which in this simple model is defined by

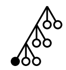
$$\mathcal{P}_{NL}(y) = \mathcal{M}_{\mathrm{np}}(y, u_L, \cdots, u_L), \tag{2.11}$$

where[1] $L = N^\delta$ for some small $\delta$. Note that when we discuss the ansatz for $y_N$, $u_L$ has already been defined by the induction hypothesis. The ansatz for $y_N$ then includes the following term:

$$(1 + \mathcal{P}_{NL} + \mathcal{P}_{NL}^2 + \mathcal{P}_{NL}^3 + \cdots)\, \Delta_N f = (1 - \mathcal{P}_{NL})^{-1} \Delta_N f, \tag{2.12}$$

where convergence is guaranteed by the operator bound (1.25) for $\mathcal{P}_{NL}$. In the case $p = 3$ for example, we may represent the terms in (2.12) by means of the following *iteration trees*

$\bullet : \Delta_N f, \quad \circ : u_L,$  $: \mathcal{M}_{\mathrm{np}}(\Delta_N f, u_L, u_L) = \mathcal{P}_{NL}(\Delta_N f),$

 $: \mathcal{P}_{NL}^2(\Delta_N f),$  $: \mathcal{P}_{NL}^3(\Delta_N f), \quad$ etc. (2.13)

For each term in (2.13) we can define the associated random $(1, 1)$ tensor (or equivalently random matrix). For example the random $(1, 1)$ tensor associated to $\bullet = \Delta_N f$ is just the identity matrix, while for the iteration tree , the associated random $(1, 1)$ tensor $h$ is such that

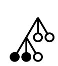
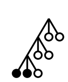
(  $)_k = \sum_{k_1} h$$_{kk_1} \cdot \Delta_N f_{k_1},$ (2.14)

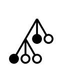
where $f_{k_1}$ is as in (2.6). By (2.6) and (2.11), we have the formula (where $\Omega = |k|^2 - |k_1|^2 + |k_2|^2 - |k_3|^2$)

$h$$_{kk_1} = -i \sum_{\substack{k_1 - k_2 + k_3 = k;\, \Omega = 0 \\ \langle k_2 \rangle, \langle k_3 \rangle \leq L,\, \langle k \rangle \leq N}} \overline{(u_L)_{k_2}} \cdot (u_L)_{k_3}.$ (2.15)

Similarly we can define the random $(1, 1)$ tensors associated to other iteration trees in (2.13) such as . Since $u_L$ (represented by $\circ$) has less importance in the estimates (in fact they will be trimmed out, see the arguments below), all these random $(1, 1)$ tensors can be treated in a similar way in our proof. Hence for simplicity we will denote them by the single notation $h$.

2.4.2. *Random tensors.* With the random $(1, 1)$ tensors $h$ defined as above, we proceed to construct the random $(r, 1)$ tensors in the ansatz for $y_N$ by induction, with $h$ being the base case. These tensors arise from high order iterations of the equation (2.10). We start with a simple case, namely the random $(2, 1)$ tensor terms in the ansatz for $y_N$, by using the following iteration trees (assume $p = 3$):

 $: \mathcal{M}_{\mathrm{np}}(\Delta_{N_{\mathfrak{l}_1}} f, \Delta_{N_{\mathfrak{l}_2}} f, u_L),$  $: \mathcal{P}_{NL}($$),$  $: \mathcal{P}_{NL}^2($$),$

 $: \mathcal{M}_{\mathrm{np}}(\mathcal{M}_{\mathrm{np}}(\Delta_{N_{\mathfrak{l}_1}} f, u_L, u_L), \Delta_{N_{\mathfrak{l}_2}} f, u_L), \quad$ etc. (2.16)

where $(\mathfrak{l}_1, \mathfrak{l}_2)$ is not a pairing (i.e. $k_{\mathfrak{l}_1} \neq k_{\mathfrak{l}_2}$ in (2.17) below, if the corresponding signs are the opposite), $N_{\mathfrak{l}_1}, N_{\mathfrak{l}_2} > L$ and $\max(N_{\mathfrak{l}_1}, N_{\mathfrak{l}_2}) = N$. These terms are similar to the high-low interactions in (2.12) and hence will also be added to the ansatz for $y_N$. We can define the random $(2, 1)$ tensors associated to terms in (2.16), proceeding similarly as in (2.14) and (2.15). Once again all these

[1]In [36] we used $L = N^{1-\delta}$; here we need a smaller value of $L$ which works better in the general setting.

random $(2,1)$ tensors can be treated in a similar way in our proof, hence for simplicity we also denote them by the single notation $h$, such that the $k$-th mode of terms in (2.16) are given by

$$\sum_{N_{\mathfrak{l}_i}/2<\langle k_{\mathfrak{l}_i}\rangle\leq N_{\mathfrak{l}_i},\ i=1,2} h_{kk_{\mathfrak{l}_1}k_{\mathfrak{l}_2}}(\Delta_{N_{\mathfrak{l}_1}} f_{k_{\mathfrak{l}_1}})^{\pm}(\Delta_{N_{\mathfrak{l}_2}} f_{k_{\mathfrak{l}_2}})^{\pm}. \tag{2.17}$$

Synthesizing the structures of these random $(1,1)$ and $(2,1)$ tensors, we describe the associated $(r,1)$ tensors in general. To that effect we introduce the *skeleton tree* $\mathcal{L}$ containing all solid leaves[1] $\bullet$ in the iteration trees associated to the $(r,1)$ tensors. For each leaf $\mathfrak{l}\in\mathcal{L}$ we also attach a frequency $N_{\mathfrak{l}}$ (such as $N_{\mathfrak{l}_1}$ and $N_{\mathfrak{l}_2}$ in (2.16) above) and a sign $\zeta_{\mathfrak{l}}\in\{\pm\}$ (for simplicity we will not explicitly write $\zeta_{\mathfrak{l}}$ below). Define also the frequency of the skeleton tree to be $N$, which always equals the maximum of $N_{\mathfrak{l}}$; in particular if $\mathfrak{l}$ is the only leaf then $N=N_{\mathfrak{l}}$. For example, the term in (2.16) that has the iteration tree will correspond to the skeleton tree, or $\mathcal{L}=\{\mathfrak{l}_1,\mathfrak{l}_2\}$, with two leaves, no pairing, and $\max(N_{\mathfrak{l}_1},N_{\mathfrak{l}_2})=N$. In such terms we always assume $N_{\mathfrak{l}}>L$ so that the tensor $h^{\mathcal{L}}$, which is a Borel function of $u_L$, is *independent* with the Gaussians $\Delta_{N_{\mathfrak{l}}}f$.

Let us now show with an example how the inductive definition of the $(r,1)$ tensors associated to the ansatz for $y_N$ proceeds using high order iterations, first in the no-pairing case. For $p=3$, consider the high order iteration term such as

$$= \mathcal{M}_{\mathrm{np}}(\ ,\bullet,\ );$$

$$\Big(\ \Big)_k = \sum_{N_{\mathfrak{l}_i}/2<\langle k_{\mathfrak{l}_i}\rangle\leq N_{\mathfrak{l}_i};\,1\leq i\leq 4}\left(\sum_{k_1,k_2,k_3} h_{kk_1k_2k_3}h^{\mathcal{L}_1}_{k_1k_{\mathfrak{l}_1}k_{\mathfrak{l}_2}}h^{\mathcal{L}_2}_{k_2k_{\mathfrak{l}_3}}h^{\mathcal{L}_3}_{k_3k_{\mathfrak{l}_4}}\right)\prod_{i=1}^{4}(\Delta_{N_{\mathfrak{l}_i}}f_{k_{\mathfrak{l}_i}})^{\pm}, \tag{2.18}$$

where $\mathcal{L}_1=\{\mathfrak{l}_1,\mathfrak{l}_2\}$ corresponds to the iteration tree and has frequency $N_1$, $\mathcal{L}_2=\{\mathfrak{l}_3\}$ corresponds to $\bullet$ and has frequency $N_2$, and $\mathcal{L}_3=\{\mathfrak{l}_4\}$ corresponds to the iteration tree and has frequency $N_3$. Also note that $N=\max(N_1,N_2,N_3)$ by (2.10). By the definition of $\mathcal{M}_{\mathrm{np}}$ in (2.6), the tensor $h_{kk_1k_2k_3}$ in (2.18) is (with the no-pairing restrictions which we omit)

$$h_{kk_1k_2k_3}=\mathbf{1}_{k=k_1-k_2+k_3}\cdot\mathbf{1}_{|k|^2=|k_1|^2-|k_2|^2+|k_3|^2}. \tag{2.19}$$

Consider the case when there is no pairing among $\{\mathfrak{l}_1,\mathfrak{l}_2,\mathfrak{l}_3,\mathfrak{l}_4\}$ (note that $(\mathfrak{l}_1,\mathfrak{l}_2)$ is already not a pairing in (2.16)). By (2.18) and Definition 2.4, the random tensor $h^{\mathcal{L}}$ associated to the iteration tree is the merging of $h^{\mathcal{L}_1}$, $h^{\mathcal{L}_2}$, $h^{\mathcal{L}_3}$ via $h$, assuming there is no pairing; it has skeleton tree , or $\mathcal{L}=\{\mathfrak{l}_1,\mathfrak{l}_2,\mathfrak{l}_3,\mathfrak{l}_4\}$.

It becomes unnecessarily complex to keep track of the iteration or skeleton trees such as and , as we iterate further and increase the depth. It turns out, see Section 2.4.3, that the desired estimates for the random $(r,1)$ tensors depend *only* on the set of solid leaves and their corresponding frequencies, not on the tree structure, except for some minor corrections. Therefore it will suffice to consider structures that we will refer to as *flattened trees* below, provided we keep the necessary information of the trees in a *memory set* $\mathcal{Y}$, which will yield the minor corrections alluded above. The pair $(\mathcal{L},\mathcal{Y})$, where $\mathcal{L}$ is viewed as a *set*, then plays the role of the trees (such as and ). The process of viewing $\mathcal{L}$ as a set—forgetting its tree structure—and finding the set $\mathcal{Y}$ associated to the trees, is then called the *flattening* of trees. More precisely, every time we merge the tensors, the flattening of trees proceeds by putting an element $\mathfrak{p}$ into $\mathcal{Y}$ and set $N_{\mathfrak{p}}$ to be

[1] In addition, we remove edges connecting a node to its *only* child; they correspond to the random averaging operator in (2.11) and can be dealt with as in Section 2.4.1.

the *second maximum* among all frequencies of the trees of the merged tensors, so $\mathcal{Y}$ can be viewed as a subset of selected nodes of the unflattened tree. For example, in the situation of (2.18), the skeleton tree $\mathcal{L} = \{\mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3, \mathfrak{l}_4\}$, which corresponds to the iteration tree , comes from merging $h^{\mathcal{L}_1}$, $h^{\mathcal{L}_2}$ and $h^{\mathcal{L}_3}$, so we have $\mathfrak{p}_1 \in \mathcal{Y}$ and $N_{\mathfrak{p}_1}$ equals the second maximum among $\{N_1, N_2, N_3\}$. Furthermore $\mathcal{L}_1$, which corresponds to the iteration tree , is constructed by merging the tensors $h^{\bullet}$ (from $\Delta_{N_{\mathfrak{l}_1}} f$), $h^{\bullet}$ (from $\Delta_{N_{\mathfrak{l}_2}} f$) and[1] $h^{\circ}$ (from $u_L$), hence we have one more element $\mathfrak{p}_2$ in $\mathcal{Y}$, and $N_{\mathfrak{p}_2}$ equals the second maximum among $\{N_{\mathfrak{l}_1}, N_{\mathfrak{l}_2}\}$. Since $h^{\mathcal{L}_2}$ and $h^{\mathcal{L}_3}$ are $(1,1)$ tensors which are defined directly without merging, we obtain the memory set $\mathcal{Y}$ associated to the skeleton tree , namely $\{\mathfrak{p}_1, \mathfrak{p}_2\}$. With this flattening process, we can forget the tree structures  and , and replace it by the flattened tree  together with $\mathcal{Y}$, so we may also denote $h^{\mathcal{L}} = h^{(\;,\mathcal{Y})}$, where $\mathcal{Y} = \{\mathfrak{p}_1, \mathfrak{p}_2\}$.

We now need to trim[2] the tensor $h^{(\;,\mathcal{Y})}$ as above at frequency $2L$, in order to maintain the property $N_{\mathfrak{l}} > L$ and the independence between the Gaussians $\Delta_{N_{\mathfrak{l}}} f$ and the tensor which contains the low frequency components. When $N_{\mathfrak{l}} > L$ for all $\mathfrak{l} \in \mathcal{L}$, no trimming is needed and we get the same random $(4,1)$ tensor $h^{(\;,\mathcal{Y})}$, with the $k$-th mode of the corresponding term in the ansatz being

$$\sum_{N_{\mathfrak{l}_i}/2 < \langle k_{\mathfrak{l}_i} \rangle \le N_{\mathfrak{l}_i};\, 1 \le i \le 4} h^{(\;,\mathcal{Y})}_{k k_{\mathfrak{l}_1} k_{\mathfrak{l}_2} k_{\mathfrak{l}_3} k_{\mathfrak{l}_4}} \cdot \prod_{i=1}^{4} (\Delta_{N_{\mathfrak{l}_i}} f_{k_{\mathfrak{l}_i}})^{\pm}. \tag{2.20}$$

Instead, if $N_{\mathfrak{l}_4} \le L$, then by Definition 2.7 the trimmed tensor is a $(3,1)$ tensor:

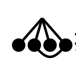
$$h^{(\;,\mathcal{Y})}_{k k_{\mathfrak{l}_1} k_{\mathfrak{l}_2} k_{\mathfrak{l}_3}} = \sum_{k_{\mathfrak{l}_4}} h^{(\;,\mathcal{Y})}_{k k_{\mathfrak{l}_1} k_{\mathfrak{l}_2} k_{\mathfrak{l}_3} k_{\mathfrak{l}_4}} \cdot (\Delta_{N_{\mathfrak{l}_4}} f_{k_{\mathfrak{l}_4}})^{\pm} \tag{2.21}$$

and the $k$-th mode of the corresponding term in the ansatz should be

$$\sum_{N_{\mathfrak{l}_i}/2 < \langle k_{\mathfrak{l}_i} \rangle \le N_{\mathfrak{l}_i};\, 1 \le i \le 3} h^{(\;,\mathcal{Y})}_{k k_{\mathfrak{l}_1} k_{\mathfrak{l}_2} k_{\mathfrak{l}_3}} \cdot \prod_{i=1}^{3} (\Delta_{N_{\mathfrak{l}_i}} f_{k_{\mathfrak{l}_i}})^{\pm} \tag{2.22}$$

which in fact is the same as (2.20). This means that the trimming process only changes the point of view by which we regard the terms in the ansatz, but not the terms themselves.

Finally, we consider the case when pairings (see Definition 2.3) occur in the merging process. In the above example, if we have a pairing $(\mathfrak{l}_2, \mathfrak{l}_4)$ in the merging process (2.18), then instead of the flattened tree  we will have a flattened tree *with pairing*, namely

;

for simplicity we also assume $N_{\mathfrak{l}_i} > L$ for $1 \le i \le 4$, i.e. no trimming is needed. The set of paired leaves is denoted by $\mathcal{P} = \{\mathfrak{l}_2, \mathfrak{l}_4\}$, and the set of unpaired leaves is denoted by $\mathcal{U} = \{\mathfrak{l}_1, \mathfrak{l}_3\}$. In this case we still merge $h^{\mathcal{L}_1}$, $h^{\mathcal{L}_2}$ and $h^{\mathcal{L}_3}$ via $h$ using Definition 2.4 as above, but assume now $(\mathfrak{l}_2, \mathfrak{l}_4)$ is a pairing, i.e. restricting $k_{\mathfrak{l}_2} = k_{\mathfrak{l}_4}$ in the sum (2.18). The merged tensor, denoted by $h^{(\;,\mathcal{Y})}$,

[1] Here $h^{\circ}_k = (u_L)_k$ can be understood as a $(0,1)$ tensor which has no input variable.

[2] This corresponds to removing $\circ$'s and low-frequency $\bullet$'s from the iteration trees, or removing low-frequency leaves from the skeleton and flattened trees. In the main proof, in addition to this trimming after merging, we also need to trim the tensors *before* merging; see (5.23) and (5.24).

is in fact a random $(2,1)$ tensor as only $k_{\mathfrak{l}}$ for *unpaired* leaves $\mathfrak{l} \in \mathcal{U}$ are input variables. The $k$-th mode of the corresponding term in the ansatz is then

$$\sum_{N_{\mathfrak{l}_i}/2<|k_{\mathfrak{l}_i}|\leq N_{\mathfrak{l}_i},\, i\in\{1,3\}} h^{(\;\cdot\;,\mathcal{Y})}_{kk_{\mathfrak{l}_1}k_{\mathfrak{l}_3}} \cdot (\Delta_{N_{\mathfrak{l}_1}} f_{k_{\mathfrak{l}_1}})^{\pm}(\Delta_{N_{\mathfrak{l}_3}} f_{k_{\mathfrak{l}_3}})^{\pm}. \tag{2.23}$$

In summary, in order to construct a random $(r,1)$ tensor, we start with a high order iteration which can be understood as the process of merging several lower order tensors as in (2.18), and then trim the merged tensor at a given frequency $2L$. This trimmed tensor is the $(r,1)$ tensor that we seek for the ansatz for $y_N$.

2.4.3. *The core ansatz.* Given a large parameter $D$, based on the above random $(r,1)$ tensors, we construct the ansatz for $y_N$ as follows:

$$(y_N)_k = \sum_{(\mathcal{L},\mathcal{Y})} h^{(\mathcal{L},\mathcal{Y})}_{kk_{\mathcal{U}}} \cdot \prod_{\mathfrak{l}\in\mathcal{U}} (\Delta_{N_{\mathfrak{l}}} f_{k_{\mathfrak{l}}})^{\pm} \;+\; (z_N)_k, \tag{2.24}$$

where $z_N$ is a smooth remainder, and the sum is taken over all flattened trees $\mathcal{L}$ with frequency $N$ and cardinality $|\mathcal{L}| \leq D$, and all possibilities of $\mathcal{Y}$ that arise from the above inductive process. In (2.24), $\mathcal{U}$ is the set of unpaired leaves in $\mathcal{L}$, and denote by $\mathcal{P} = \mathcal{L}\backslash\mathcal{U}$ the set of paired leaves in $\mathcal{L}$.

The main a priori estimates contain the bounds for various operator norms for the tensors $h^{\mathcal{L}}$ (as well as a high-regularity bound for the remainder $z_N$, which we omit). Here we look at a simplified example[1]: for any partition $(B,C)$ of $\mathcal{U}$, we would like to show

$$\left\| h^{(\mathcal{L},\mathcal{Y})}_{kk_{\mathcal{U}}} \right\|_{kk_B \to k_C} \leq \prod_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}^{\beta} \cdot \big(\max_{\mathfrak{l}\in C} N_{\mathfrak{l}}\big)^{-\beta} \cdot \prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-\varepsilon_1} \prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta_1}, \tag{2.25}$$

where $\beta$ is a constant which is a little bit smaller than $\alpha$, and $\varepsilon_1 \ll 1$ and $\delta_1$ is small compared to $\varepsilon_1$. In particular, the factor $\prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta_1}$ shows the decay we gain from the tree structures (e.g. ), hence we only need to keep the flattened trees (e.g. ) and the memory set $\mathcal{Y}$ abstracted from the full tree structures.

We will prove (2.25) by induction. The key step here is to show that if (2.25) is true for some tensors $h^{(\mathcal{L}_j,\mathcal{Y}_j)} = h^{(\mathcal{L}_j,\mathcal{Y}_j)}_{k_jk_{\mathcal{U}_j}}$ where $1 \leq j \leq p$, then it also holds for the tensor $h^{(\mathcal{L},\mathcal{Y})} = h^{(\mathcal{L},\mathcal{Y})}_{kk_{\mathcal{U}}}$ which is obtained by merging and trimming those tensors as in Section 2.4.2. This argument, which is the center of the whole paper, contains **three main ingredients**:

(1) The **inequalities** associated with the algebra of tensors, namely Propositions 2.6 and 2.8 (and their precise versions in Section 4). Note that these are problem non-specific and are not limited to Schrödinger equations.

(2) The operator norm bounds for the base tensor $h$ that appears in the merging process. The form of $h$ is similar to (2.19), and operator norm bounds for $h$ follow from various **counting estimates** and Schur's Lemma. This is proved in Proposition 4.9.

(3) A particular **selection algorithm**. This is crucial when we apply Proposition 2.6, since even though $H$ on the left hand side of (2.4) does not depend on the *order* of the tensors $h^{(j)}$, the right hand side does. Therefore we have to follow a particular algorithm in order to go from bounds of $h^{(\mathcal{L}_j,\mathcal{Y}_j)}$ to bounds of $h^{(\mathcal{L},\mathcal{Y})}$. This algorithm is described in the proof of Proposition 6.2.

[1]See Proposition 5.1 for the full detailed version. In particular there are distinctions between $h^{(*,0)}$ and $h^{(*,1)}$ tensors, which we omit here.

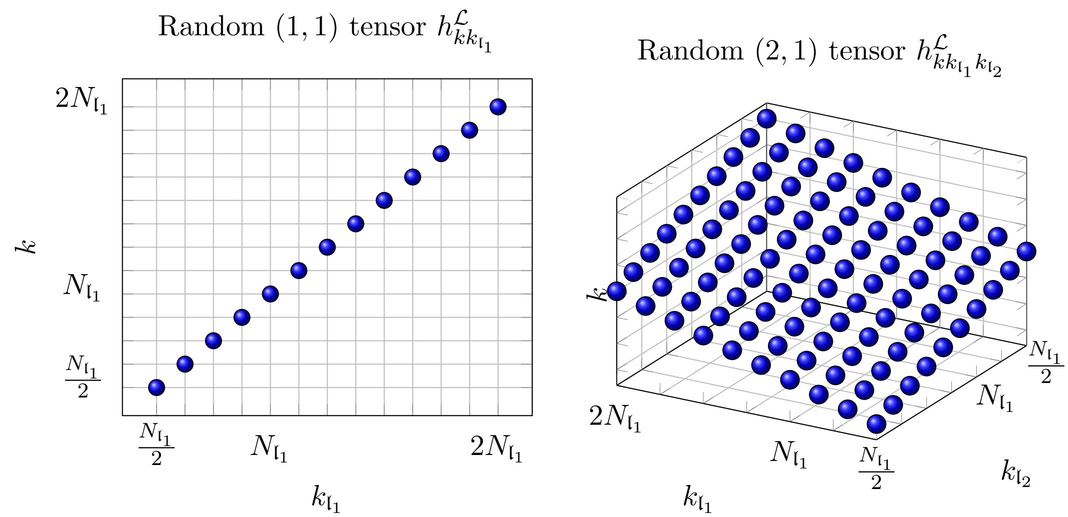



FIGURE 1. Localization hyperplanes of random $(1,1)$ and $(2,1)$ tensors

Finally, in addition to the operator norms, we need to control one more *localization* norm for the tensor $h^{(\mathcal{L},\mathcal{Y})}$, which localizes it around the hyperplane $k = \sum_{\mathfrak{l}\in\mathcal{U}} \zeta_{\mathfrak{l}} k_{\mathfrak{l}}$, where $\zeta_{\mathfrak{l}} \in \{\pm\}$ is the sign of $\mathfrak{l}$. This norm is essentially a weighted $L^2$ or Hilbert-Schmidt norm, and the corresponding estimates roughly look like

$$\left\| \left(1 + \frac{1}{L^2}\big|k - \sum_{\mathfrak{l}\in\mathcal{U}} \zeta_{\mathfrak{l}} k_{\mathfrak{l}}\big|\right)^{\kappa} h^{(\mathcal{L},\mathcal{Y})}_{kk_{\mathcal{U}}} \right\|_{kk_{\mathcal{U}}} \leq \prod_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-\varepsilon_1} \prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta_1}, \tag{2.26}$$

where $\beta$, $\varepsilon_1$ and $\delta_1$ are the same as in (2.25) and $\kappa$ is a large enough constant. Such localizations can be understood as our tensor $h^{\mathcal{L}}$ being close to a *multilinear Fourier multiplier*. See Figure 1 for an illustration of the regions around which the $(1,1)$ and $(2,1)$ tensors are localized.

2.5. **The extended ansatz, and general case.** The ansatz (2.24) in Section 2.4 exhibits the main idea, however the full ansatz has extra layers of complexity. Some of them come from passing from the model (2.5) to the full equation, such as the possibility of over-pairing (leading to the norms in (2.2) and the full Definition 3.6 of merging) and the role of time Fourier or modulation variables (leading to the spacetime norms defined in Section 3.3 and allowing $\Omega \neq 0$ in (2.5)). The main one, however, arises already in the model (2.5).

To demonstrate, suppose we plug the ansatz (2.24) into (2.10). Consider the nonlinear term where one (or more) of the inputs is the remainder term $z_{N'}$ with $N' \leq N/2$, which is a part of $y_{N'}$ in (2.24). If $N^{1/2} \gg N' \gg N^{\delta}$, then this $N'$ is not high enough for the resulting term to have enough regularity to be put in the remainder $z_N$, and not low enough for the tensor arising from the resulting term to be independent with the Gaussians.

To remedy this, we go back to the iteration trees in Section 2.4.2 and introduce more random tensor terms by considering all possible configurations of iteration trees where we replace at least one $\circ$ (meaning $u_L$) by a diamond $\diamond$ (meaning $z_{N'}$ with $N' \leq N/2$). For example consider the term

$$: \mathcal{M}_{\mathrm{np}}(\Delta_N f, z_{N'}, u_L), \tag{2.27}$$

where $N' \leq N/2$, whose $k$-th mode is given by

$$(\ )_k = \sum_{k_{\mathfrak{l}}, k_{\mathfrak{f}}} h_{kk_{\mathfrak{l}}}(k_{\mathfrak{f}}) \cdot (\Delta_N f_{k_{\mathfrak{l}}}) \cdot \overline{(z_{N'})_{k_{\mathfrak{f}}}}, \tag{2.28}$$

where

$$h_{kk_{\mathfrak{l}}}(k_{\mathfrak{f}}) = -i \sum_{\substack{k_{\mathfrak{l}} - k_{\mathfrak{f}} + k_3 = k;\, \Omega = 0 \\ \langle k_{\mathfrak{f}} \rangle \leq N', \langle k_3 \rangle \leq L, \langle k \rangle \leq N}} (u_L)_{k_3} \tag{2.29}$$

with $\Omega = |k|^2 - |k_{\mathfrak{l}}|^2 + |k_{\mathfrak{f}}|^2 - |k_3|^2$ and $u_{k_3}$ is the Fourier mode of $u_L$ in (2.27). The iteration term in (2.27) can be viewed, via (2.28)–(2.29), as a *linear combination* of the tensor terms in Section 2.4.2 with coefficients being the Fourier coefficients of $z_{N'}$, which are summable since the norm of $z_{N'}$ will be a large negative power of $N'$ (see part (4) of Proposition 5.1), Moreover the tensors $h_{kk_{\mathfrak{l}}}(k_{\mathfrak{f}})$ in (2.29) do not involve $z_{N'}$ and thus retain independence.

Then we flatten these new iteration trees and repeat what we did in Sections 2.4.2–2.4.3, except that the pair $(\mathcal{L}, \mathcal{Y})$ alone is not sufficient to represent the new random tensor terms, and we need to introduce one more set $\mathcal{V}$ which contains all the $\diamond$'s in the new iteration trees. Hence in the full ansatz, the sum in (2.24) should be taken over all triples $(\mathcal{L}, \mathcal{V}, \mathcal{Y})$, which will be defined as *plants*, see Definition 3.2.

2.6. **Outline of the paper.** Sections 3–4 are mainly preparations, with definitions listed in Section 3 and lemmas proved in Section 4. In Section 5 we introduce the main random tensor ansatz, thereby reducing Theorem 1.1 to Proposition 5.1, which is then proved in Sections 6–7. In Section 8 we finish the proof of Theorem 1.1, as well as the proof of Theorem 1.6, which is a simplified version of the former. Finally in Section 9 we make a few comments, including a comparison with parabolic equations. The structure of the proof is presented in Figure 2.

## 3. Preliminaries I: Definitions

In this section we list the main definitions used in this work. In Section 3.1 we fix and collect the various notations and parameters, in Section 3.2 we define the plant structure, associated tensors, and their operations. In Section 3.3 we define the norms used in the main proofs.

3.1. **Choice of parameters and notations.** We use $C$ to denote a generic large constant depending on $(d, p)$. Let $\alpha$ be fixed as in (1.3). Define

$$\alpha_0 := \frac{d}{2} - \frac{1}{p-1}; \quad \varepsilon := (10^3 dp)^{-1}(\alpha - \alpha_0) > 0, \quad \beta := \frac{\alpha + \alpha_0}{2}, \quad \beta_1 := \frac{\beta + \alpha_0}{2}. \tag{3.1}$$

We assume $0 < \varepsilon \ll_C 1$ and fix it throughout. We will use $C_\varepsilon$ to denote a generic large constant depending on $\varepsilon$; similarly $C_\delta$ etc. will depend on the small parameters $\delta$ etc. defined below. These constants, including $C$ and $(\theta, C_\theta)$ defined below, may be varying from line to line.

Fix $(\delta, D, \kappa)$ such that[1]

$$0 < \delta \ll_{C_\varepsilon} 1; \quad D \gg_{C_\delta} 1, \quad \kappa \gg_D 1. \tag{3.2}$$

Define also

$$D_1 := \delta^5 D, \quad b := \frac{1}{2} + 8\kappa^{-1}, \quad b^+ := \frac{1}{2} + 16\kappa^{-1}, \quad b_0 = \frac{1}{2} + 4\kappa^{-1}. \tag{3.3}$$

Let $\theta$ denote a generic positive constant that is sufficiently small depending on $\kappa$, and (as above) $C_\theta$ a generic large constant depending on $\theta$. We also fix $\theta_0$ to be a specific positive constant that is

[1]Roughly speaking $\delta = \varepsilon^{50}$, $D = \delta^{-50}$, $\kappa = D^{50}$ and $\theta \sim \kappa^{-50}$ should suffice.

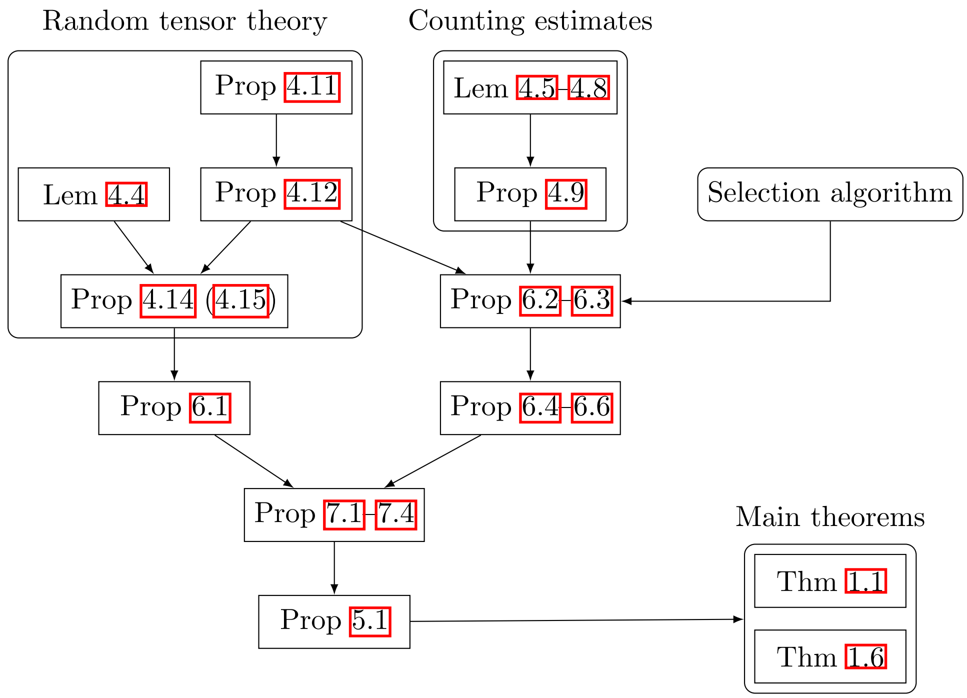



FIGURE 2. Structure of the proof

sufficiently small depending on $\kappa$. Unless otherwise stated, the implicit constants in the $\lesssim$ symbols will depend on $C_\theta$. We fix $0<\tau\ll_{C_\theta}1$, and let $J=[-\tau,\tau]$. If an event on the ambient probability space $(\Theta,\mathcal{B},\mathbb{P})$ happens with probability $\geq 1-C_\theta e^{-A^\theta}$ for some quantity $A>0$, we say this event is $A$-certain. In the proof below many quantities will depend on $\omega\in\Theta$; we may include $\omega$ in the expressions for emphasis, or omit it for simplicity. We use $\mathbf{1}_E$ to denote the characteristic function of a statement $E$.

In the proof, the capital letters $N,M,L$ etc. will denote dyadic numbers $\geq 1$; when they (formally) take the value of $1/2$, we will understand that the corresponding quantities are 0 (or the trivial case depending on the context). Define $N^{[\delta]}:=\max\{L:L<N^\delta\}$. The lower case letters $k,m$ etc. will denote integer vectors in $\mathbb{Z}^d$ or Cartesian powers of $\mathbb{Z}^d$. The letters $t,t'$ etc. will denote the time variable, and the letters $\lambda,\lambda',\lambda_j$ etc. will denote the Fourier dual of time (we call these *modulation variables*). For $k\in\mathbb{Z}^d$, let $\rho_k:=|g_k|$ and $\eta_k:=\rho_k^{-1}g_k$, which are independent random variables, such that each $\eta_k$ is uniformly distributed on the unit circle. We also define $\gamma_k=\langle k\rangle^{-\alpha}\rho_k$. For dyadic $N$, let $\mathcal{B}_N\subset\mathcal{B}$ be the $\sigma$-algebra generated by the random variables $\{\eta_k:\langle k\rangle\leq N\}$. The cardinality of a finite set $S$ is denoted by $|S|$ or $\#S$. Recall the notion of partition and subpartition in Definition 2.1, as well as the abbreviation $k_A=(k_j:j\in A)$; similarly we use $\lambda_A$ to denote $(\lambda_j:j\in A)$ and $\mathrm{d}\lambda_A$ to denote $\prod_{j\in A}\mathrm{d}\lambda_j$. Also recall the notion of $u^\zeta$, pairing and over-pairing in Definition 2.3.

We will use $u_k$ to denote the spatial Fourier coefficients of $u$, and the notation $\widehat{u}$ represents the time Fourier transform *only* (maybe in multiple time dimensions, see (3.4)–(3.5) below). We will be loose about powers of $2\pi$, and may write formulas like

$$\widehat{v}(\lambda)=\int_{\mathbb{R}}e^{-i\lambda t}v(t)\,\mathrm{d}t,\quad v(t)=\int_{\mathbb{R}}e^{it\lambda}\widehat{v}(\lambda)\,\mathrm{d}\lambda.$$

If $\mathscr{L}$ is an $\mathbb{R}$-linear operator acting on spacetime functions, we can uniquely decompose it into the sum of a $\mathbb{C}$-linear operator, $\mathscr{L}^+$, and a $\mathbb{C}$-conjugate linear operator, $\mathscr{L}^-$. We will denote the kernel of $\mathscr{L}^\zeta$, where $\zeta \in \{\pm\}$, by $(\mathscr{L}^\zeta)_{kk'}(t,t')$, so that

$$(\mathscr{L}^\zeta w)_k(t) = \sum_{k'} \int \mathrm{d}t' \cdot (\mathscr{L}^\zeta)_{kk'}(t,t') w_{k'}^\zeta(t'), \tag{3.4}$$

then on the time Fourier side we have

$$(\widehat{\mathscr{L}^\zeta w})_k(\lambda) = \sum_{k'} \int_{\mathbb{R}} (\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda, -\zeta\lambda')(\widehat{w})_{k'}^\zeta(\lambda')\,\mathrm{d}\lambda'. \tag{3.5}$$

Fix a smooth cutoff function $\chi(t)$ which equals 1 for $|t| \leq 1$ and equals 0 for $|t| \geq 2$. For $0 < \tau \lesssim 1$ define $\chi_\tau(t) := \chi(\tau^{-1}t)$. Define the standard and truncated Duhamel operators

$$\mathcal{I}v(t) = \int_0^t v(t')\,\mathrm{d}t', \quad \mathcal{I}_\chi v(t) = \chi(t) \int_0^t \chi(t')v(t')\,\mathrm{d}t'. \tag{3.6}$$

Note that these are not coming from the original Schrödinger equation (1.1), but a variant of it after conjugating by the linear Schrödinger flow (namely $v = e^{-it\Delta}u$). Finally we introduce the notion of *simplicity* for real polynomials and $\mathbb{R}$-multilinear operators; in practice the Wick-ordered and suitably gauged power nonlinearities will be simple.

**Definition 3.1.** Consider a real polynomial (or $\mathbb{R}$-multilinear operator) $\mathcal{N}$ of degree $r$, given by

$$\mathcal{N}(u)_k = \sum_{\zeta_1 k_1 + \cdots + \zeta_r k_r = k} c_{kk_1\cdots k_r} u_{k_1}^{\zeta_1} \cdots u_{k_r}^{\zeta_r}. \tag{3.7}$$

We say it is *simple* if the coefficients $c_{kk_1\cdots k_r}$ depend only on the *set of pairings*[1] in $(k, k_1, \cdots, k_r)$, and $c_{kk_1\cdots k_r} = 0$ unless each such pairing is over-paired.

3.2. **Plants and plant tensors.** In this section we introduce the main structure—namely *plants*—and the associated tensors, as well as two basic operations (`Trim` and `Merge`) of these objects.

**Definition 3.2** (Plants)**.** A *plant* $\mathcal{S}$ consists of the following objects:

(1) Three disjoint finite sets $\mathcal{L}$ (called the *tree*), $\mathcal{V}$ (called the *blossom set*), and $\mathcal{Y}$ (called the *memory set*); elements of $\mathcal{L}$, $\mathcal{V}$ and $\mathcal{Y}$ are called *leaves*, *blossoms* and *pasts*, and are denoted by $\mathfrak{l}$, $\mathfrak{f}$ and $\mathfrak{p}$. An arbitrary element of $\mathcal{L} \cup \mathcal{V} \cup \mathcal{Y}$ is denoted by $\mathfrak{n}$.
(2) A collection of pairwise disjoint 2-element subsets of $\mathcal{L}$, which we refer to as *pairings*; the set of paired leaves is denoted by $\mathcal{P}$, and the set of unpaired leaves is denoted by $\mathcal{U} := \mathcal{L} \backslash \mathcal{P}$.
(3) A dyadic number $N = N(\mathcal{S})$ (called the *frequency* of $\mathcal{S}$), a *sign* $\zeta_\mathfrak{n} \in \{\pm\}$ for each $\mathfrak{n} \in \mathcal{L} \cup \mathcal{V}$ (note that signs are not defined for pasts), and a dyadic *frequency* $N_\mathfrak{n}$ for each $\mathfrak{n} \in \mathcal{L} \cup \mathcal{V} \cup \mathcal{Y}$. We require that $N_{\mathfrak{l}'} = N_\mathfrak{l}$ and $\zeta_{\mathfrak{l}'} = -\zeta_\mathfrak{l}$ for any pairing $(\mathfrak{l}, \mathfrak{l}')$ in $\mathcal{L}$; that $N_\mathfrak{n} \leq N$ for $\mathfrak{n} \in \mathcal{L} \cup \mathcal{Y}$; and that $N_\mathfrak{f} \leq N/2$ for $\mathfrak{f} \in \mathcal{V}$.

We will denote a plant by $\mathcal{S} = (\mathcal{L}, \mathcal{V}, \mathcal{Y})$, and define $|\mathcal{S}| = |\mathcal{L}| + |\mathcal{V}| + |\mathcal{Y}|$ to be the *size* of the plant. Two plants will be identified if there is a bijection between them that preserves all these objects. We say a plant $\mathcal{S}$ is *regular* if $N_\mathfrak{n} \geq N^\delta$ for any $\mathfrak{n} \in \mathcal{L} \cup \mathcal{V} \cup \mathcal{Y}$, and *plain* if $\mathcal{V} = \varnothing$ and $\sum_{\mathfrak{l} \in \mathcal{L}} \zeta_\mathfrak{l} = 1$ (so in particular $|\mathcal{L}|$ is odd). We also define the *mini plant* $\mathcal{S}_N^\zeta$, where $\zeta \in \{\pm\}$, to be the plant where $\mathcal{V} = \mathcal{Y} = \varnothing$, $\mathcal{L}$ has only one element $\mathfrak{l}$ with sign $\zeta$ and frequency $N$, and $N(\mathcal{S}) = N$.

[1] Here a pairing $(k, k_j)$ means $k = k_j$ and $\zeta_j = +$.

This is regular, and is plain if $\zeta=+$. Finally we define the *conjugate* of a plant $\mathcal{S}$ to be $\overline{\mathcal{S}}$, which is the same as $\mathcal{S}$ except that the signs of all elements in $\overline{\mathcal{S}}$ are the opposite to the signs in $\mathcal{S}$.

*Remark* 3.3. Note that the sets $\mathcal{U}$, $\mathcal{L}$, $\mathcal{V}$ etc. are associated with a plant $\mathcal{S}$. We will keep this correspondence throughout; for example whenever there is a plant $\mathcal{S}_j$ in some context, the set $\mathcal{U}_j$ will always be the one coming from $\mathcal{S}_j$. We may encounter sets $\mathcal{U}_j$ that do not come from any plant; in such case there will simply be *no* plant called $\mathcal{S}_j$ that appears in the same context.

**Definition 3.4** (Plant tensors)**.** Given a plant $\mathcal{S}=(\mathcal{L},\mathcal{V},\mathcal{Y})$, let $\mathcal{U}$ be as in Definition 3.2. We say a tensor[1] $h=h_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ is *an $\mathcal{S}$-tensor*, if $k$ and each $k_{\mathfrak{n}}$ ($\mathfrak{n}\in\mathcal{U}\cup\mathcal{V}$) are vectors in $\mathbb{Z}^d$, and in the support of $h$ we have that

(1) $\langle k\rangle\leq N$, $\langle k_{\mathfrak{f}}\rangle\leq N_{\mathfrak{f}}$ and $|\lambda_{\mathfrak{f}}|\leq 2N^{\kappa^2}$ for each $\mathfrak{f}\in\mathcal{V}$, and $N_{\mathfrak{l}}/2<\langle k_{\mathfrak{l}}\rangle\leq N_{\mathfrak{l}}$ for each $\mathfrak{l}\in\mathcal{U}$;
(2) there is no pairing in $k_{\mathcal{U}}$, i.e. if $\mathfrak{l},\mathfrak{l}'\in\mathcal{U}$ and $\zeta_{\mathfrak{l}'}=-\zeta_{\mathfrak{l}}$ then $k_{\mathfrak{l}'}\neq k_{\mathfrak{l}}$.

Here $h$ may depend on other parameters like $t$, in which case we may write $h=h_{kk_{\mathcal{U}}}(t,k_{\mathcal{V}},\lambda_{\mathcal{V}})$.

Suppose we have defined functions $f_{N'}=(f_{N'})_{k'}$ for any $N'$, and $z_{N'}=(z_{N'})_k(t)$ for any $N'<N$. Define $\Psi_k=\Psi_k[\mathcal{S},h]$ by

$$\Psi_k=\sum_{k_{\mathcal{U}},k_{\mathcal{V}}}\int \mathrm{d}\lambda_{\mathcal{V}}\cdot h_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})\cdot\prod_{\mathfrak{l}\in\mathcal{U}}(f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}}\prod_{\mathfrak{f}\in\mathcal{V}}(\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}}(\lambda_{\mathfrak{f}}), \tag{3.8}$$

which is an expression determined by the tensor $h$. Note also that an $\mathcal{S}$ tensor $h$ is also an $\overline{\mathcal{S}}$ tensor, and $\overline{\Psi_k[\mathcal{S},h]}=\Psi_k[\overline{\mathcal{S}},\overline{h}]$.

**Definition 3.5** (Trimming)**.** Given a plant $\mathcal{S}=(\mathcal{L},\mathcal{V},\mathcal{Y})$ and $R\geq 1$, we can *trim $\mathcal{S}$ at frequency* $R$ to get $\mathcal{S}'=(\mathcal{L}',\mathcal{V}',\mathcal{Y}')$, such that $\mathcal{L}'=\{\mathfrak{l}\in\mathcal{L}:N_{\mathfrak{l}}\geq R\}$ and $(\mathcal{V}',\mathcal{Y}')$ are defined in the same way. The other objects (i.e. the frequency of $\mathcal{S}'$, the signs and frequencies of elements, and the pairings in $\mathcal{L}'$) are inherited from $\mathcal{S}$. Obviously, $\mathcal{S}'$ is regular if either $\mathcal{S}$ is regular or $R\geq N^{\delta}$.

Now suppose we have defined functions $f_{N'}=(f_{N'})_{k'}$ for any $N'$, and $z_{N'}=(z_{N'})_k(t)$ for any $N'<R$. Then, given an $\mathcal{S}$-tensor $h=h_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})$, we can *trim it at frequency* $R$ to get an $\mathcal{S}'$-tensor $h'=(h')_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'},\lambda_{\mathcal{V}'})$, which is defined by

$$(h')_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'},\lambda_{\mathcal{V}'})=\sum_{k_{\mathcal{U}\backslash\mathcal{U}'}}\sum_{k_{\mathcal{V}\backslash\mathcal{V}'}}\int \mathrm{d}\lambda_{\mathcal{V}\backslash\mathcal{V}'}\cdot h_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})\cdot\prod_{\mathfrak{l}\in\mathcal{U}\backslash\mathcal{U}'}(f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}}\prod_{\mathfrak{f}\in\mathcal{V}\backslash\mathcal{V}'}(\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}}(\lambda_{\mathfrak{f}}). \tag{3.9}$$

We shall write the above definitions as $\mathcal{S}'=\mathtt{Trim}(\mathcal{S},R)$ and $h'=\mathtt{Trim}(h,R)$. Note that the definition of $h'$ actually depends on the choices of $(f_{N'})$ and $(z_{N'})_{N'<R}$, but in practice these will be uniquely fixed whenever we apply `Trim` functions, so we will omit them from the list of parameters. If $N_{\mathfrak{n}}<R$ for all $\mathfrak{n}\in\mathcal{U}\cup\mathcal{V}$ then $\mathcal{U}'=\mathcal{V}'=\varnothing$, and $h'=(h')_k$ is just the $\Psi_k$ defined in (3.8).

**Definition 3.6** (Merging)**.** First, for any finite set $\mathcal{A}$ with a sign for each element, we will fix a maximal collection of pairwise disjoint two-element subsets of $\mathcal{A}$, such that each such subset contains two elements of opposite sign (i.e. pairings). Let $\mathscr{P}(\mathcal{A})$ be this collection, and $\mathcal{Q}(\mathcal{A})$ be the union of the two-element subsets in $\mathscr{P}(\mathcal{A})$.

Now let $3\leq q\leq p$ be odd and $0\leq r\leq q$. Given dyadic numbers $N$ and $N_j$ and signs $\zeta_j\in\{\pm\}$ for $1\leq j\leq q$, so that $N_j\leq N$ for $1\leq j\leq r$ and $N_j\leq N/2$ for $r+1\leq j\leq q$, and $\sum_{j=1}^q\zeta_j=1$, denote the collection of these parameters by $\mathscr{B}$. Given pairwise disjoint plants $\mathcal{S}_j=(\mathcal{L}_j,\mathcal{V}_j,\mathcal{Y}_j)$

[1] In this tensor $k_{\mathcal{V}}$ and $\lambda_{\mathcal{V}}$ appear as parameters. Note also that the definition does *not* involve $\mathcal{P}$ or $\mathcal{Y}$.

with frequency $N_j$ for $1 \leq j \leq r$, let $\mathcal{U}_j$ etc. be as in Definition 3.2. Let $\mathcal{L} := \mathcal{L}_1 \cup \cdots \cup \mathcal{L}_r$ and[1] $\mathcal{V} := \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_r \cup \{r+1, \cdots, q\}$, and for each $\mathfrak{n} \in \mathcal{L}_j \cup \mathcal{V}_j$ let the (new) sign of $\mathfrak{n}$ be $\zeta_\mathfrak{n}^* = \zeta_j \zeta_\mathfrak{n}$, where $\zeta_\mathfrak{n}$ is the sign of $\mathfrak{n}$ in $\mathcal{S}_j$. Let $\mathscr{O} = \{\mathcal{A}_1, \cdots, \mathcal{A}_m\}$ be an arbitrary collection of disjoint subsets of $\mathcal{W} := \mathcal{U}_1 \cup \cdots \cup \mathcal{U}_r$, such that:

(1) each $\mathcal{A}_i$ contains two elements of opposite $\zeta_\mathfrak{l}^*$ sign, but does not contain two elements of opposite $\zeta_\mathfrak{l}^*$ sign that belongs to *the same* $\mathcal{U}_j$;
(2) the frequencies of $\mathfrak{l} \in \mathcal{A}_i$ are the same for each $1 \leq i \leq m$.

For each possible $\mathscr{O}$, let

$$\mathscr{P} := \mathscr{P}(\mathcal{A}_1) \cup \cdots \cup \mathscr{P}(\mathcal{A}_m) \quad \text{and} \quad \mathcal{Q} := \mathcal{Q}(\mathcal{A}_1) \cup \cdots \cup \mathcal{Q}(\mathcal{A}_m) \tag{3.10}$$

with $\mathscr{P}(\cdot)$ and $\mathcal{Q}(\cdot)$ defined as above. We then *merge* $\mathcal{S}_j$ $(1 \leq j \leq r)$ *via* $\mathscr{B}$ *and* $\mathscr{O}$, to get a plant $\mathcal{S} = (\mathcal{L}, \mathcal{V}, \mathcal{Y})$ as follows. First $N(\mathcal{S}) = N$, $\mathcal{L}$ and $\mathcal{V}$ are as above, and the set of pairings in $\mathcal{L}$ is the union of the sets of pairings in each $\mathcal{L}_j$, together with $\mathscr{P}$ (the new pairings; in particular we have $\mathcal{U} = \mathcal{W} \backslash \mathcal{Q}$). Second, the sign and frequency of $j$ are given by $\zeta_j$ and $N_j$ for $r+1 \leq j \leq q$, and the sign and frequency of any $\mathfrak{n} \in \mathcal{L}_j \cup \mathcal{V}_j$ $(1 \leq j \leq r)$ is given by $\zeta_\mathfrak{n}^*$ and the frequency of $\mathfrak{n}$ in $\mathcal{S}_j$. Finally, $\mathcal{Y} = \mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_r \cup \{0\}$, with $N_0$ given by the *second maximum* of all the $N_j$ $(1 \leq j \leq q)$; for any $\mathfrak{p} \in \mathcal{Y}_j$ $(1 \leq j \leq r)$ the frequency of $\mathfrak{p}$ in $\mathcal{S}$ equals the frequency of $\mathfrak{p}$ in $\mathcal{S}_j$.

Now suppose we have defined a tensor $h = h_{kk_1 \cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q)$, where $k, k_1, \cdots, k_q$ are input variables and $\lambda_{r+1}, \cdots, \lambda_q$ are parameters; assume in the support of $h$ that $\langle k \rangle \leq N$ and $\langle k_j \rangle \leq N_j$ for $1 \leq j \leq q$, and that $|\lambda_j| \leq 2N^{\kappa^2}$ for $r+1 \leq j \leq q$. Then, given $\mathcal{S}_j$-tensors $h^{(j)} = h^{(j)}_{k_j k_{\mathcal{U}_j}}(k_{\mathcal{V}_j}, \lambda_{\mathcal{V}_j})$, where $1 \leq j \leq r$, we shall *merge these* $h^{(j)}$ *via* $h$, $\mathscr{B}$ *and* $\mathscr{O}$, to form a new tensor $H = H_{kk_\mathcal{U}}(k_\mathcal{V}, \lambda_\mathcal{V})$, namely (recall also $\gamma_k = \langle k \rangle^{-\alpha} \rho_k$)

$$\begin{aligned} H_{kk_\mathcal{U}}(k_\mathcal{V}, \lambda_\mathcal{V}) = \prod_{\mathfrak{l},\mathfrak{l}'}^{(1)} \mathbf{1}_{k_\mathfrak{l} = k_{\mathfrak{l}'}} \prod_{\mathfrak{l},\mathfrak{l}'}^{(2)} \mathbf{1}_{k_\mathfrak{l} \neq k_{\mathfrak{l}'}} \cdot \sum_{(k_1, \cdots, k_r)} h_{kk_1 \cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q) \\ \times \sum_{k_\mathcal{Q}}^{(3)} \prod_{\mathfrak{l} \in \mathcal{Q}} \Delta_{N_\mathfrak{l}} \gamma_{k_\mathfrak{l}} \prod_{j=1}^r \big[ h^{(j)}_{k_j k_{\mathcal{U}_j}}(k_{\mathcal{V}_j}, \lambda_{\mathcal{V}_j}) \big]^{\zeta_j}. \end{aligned} \tag{3.11}$$

In the above expression, the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1)}$ is taken over all leaves $\mathfrak{l}, \mathfrak{l}' \in \mathcal{U}$ such that they belong to the same $\mathcal{A}_i$ (in particular $\zeta_{\mathfrak{l}'}^* = \zeta_\mathfrak{l}^*$), the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}$ is taken over all leaves $\mathfrak{l}, \mathfrak{l}' \in \mathcal{U}$ such that[2] $\zeta_{\mathfrak{l}'}^* = -\zeta_\mathfrak{l}^*$ (so they do not belong to the same $\mathcal{A}_i$), and the summation $\sum_{k_\mathcal{Q}}^{(3)}$ is taken over all possible $k_\mathcal{Q}$ (with $\mathcal{Q}$ defined above) such that for each $i$, all the $k_\mathfrak{l}$ for $\mathfrak{l} \in \mathcal{Q} \cap \mathcal{A}_i$ are equal, and they equal $k_{\mathfrak{l}'}$ for $\mathfrak{l}' \in \mathcal{U} \cap \mathcal{A}_i$ (if such $\mathfrak{l}'$ exists). We can verify that $H$ is an $\mathcal{S}$-tensor.

We shall write the above definitions as

$$\mathcal{S} = \mathtt{Merge}(\mathcal{S}_1, \cdots, \mathcal{S}_r, \mathscr{B}, \mathscr{O}), \quad H = \mathtt{Merge}(h^{(1)}, \cdots, h^{(r)}, h, \mathscr{B}, \mathscr{O}). \tag{3.12}$$

**Proposition 3.7.** *Assume we have fixed the choices of $f_{N'}$ and $z_{N'}$ as in Definitions 3.4 and 3.5, and $h = h_{kk_1 \cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q)$, $\mathscr{B}$, $\mathcal{S}_j$ and $\mathcal{S}_j$-tensors $h^{(j)}$ for $1 \leq j \leq r$ as in Definition 3.6. In applying the* `Merge` *function below we will omit the parameters $h$ and $\mathscr{B}$. Then the following statements hold:*

[1] If necessary we may replace the unions $\cup$ by the disjoint unions $\sqcup$ to avoid repetition of elements.

[2] Here we may also require $N_\mathfrak{l} = N_{\mathfrak{l}'}$; whether we do so will not affect the result of this product.

*(1) Recall the definition of* $\Psi_k = \Psi_k[\cdot,\cdot]$ *as in (3.8). Then for any* $R$ *we have* $\Psi_k[\mathcal{S}_j, h^{(j)}] = \Psi_k[\mathtt{Trim}(\mathcal{S}_j, R), \mathtt{Trim}(h^{(j)}, R)]$. *Similarly, trimming at frequency* $R_1$ *and then* $R_2$ *is equivalent to trimming once at* $\max(R_1, R_2)$.

*(2) Let* $\Psi_{k_j}^{(j)} = \Psi_{k_j}[\mathcal{S}_j, h^{(j)}]$ *be defined as in (3.8) from the* $\mathcal{S}_j$*-tensor* $h^{(j)}$ *for* $1 \leq j \leq r$. *Then the quantity*

$$\Phi_k := \sum_{(k_1,\cdots,k_q)} \int \mathrm{d}\lambda_{r+1}\cdots\mathrm{d}\lambda_q \cdot h_{kk_1\cdots k_q}(\lambda_{r+1},\cdots,\lambda_q)\cdot\prod_{j=1}^{r}(\Psi_{k_j}^{(j)})^{\zeta_j}\prod_{j=r+1}^{q}(\widehat{z_{N_j}})_{k_j}^{\zeta_j}(\lambda_j) \tag{3.13}$$

*can be written as a linear combination*[1] *of* $\Psi_k = \Psi_k[\mathcal{S}, H]$ *(for different choices of* $\mathscr{O}$ *as in Definition 3.6), where*

$$\mathcal{S} = \mathtt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_r,\mathscr{O}), \quad H = \mathtt{Merge}(h^{(1)},\cdots,h^{(r)},\mathscr{O}).$$

*(3) Let* $\mathcal{S}_j' = (\mathcal{L}_j', \mathcal{V}_j', \mathcal{Y}_j') = \mathtt{Trim}(\mathcal{S}_j, N^\delta)$ *and* $(h^{(j)})' = \mathtt{Trim}(h^{(j)}, N^\delta)$ *for* $1 \leq j \leq r$. *For any* $\mathscr{O}$ *as in Definition 3.6, let* $\mathscr{O}'$ *be the sub-collection of* $\mathscr{O}$ *consisting of subsets that are contained in the union of* $\mathcal{U}_j'$ *for* $1 \leq j \leq r$. *Let*

$$\mathcal{S} = (\mathcal{L},\mathcal{V},\mathcal{Y}) = \mathtt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_r,\mathscr{O}), \qquad \mathcal{S}' = (\mathcal{L}',\mathcal{V}',\mathcal{Y}') = \mathtt{Trim}(\mathtt{Merge}(\mathcal{S}_1',\cdots,\mathcal{S}_r',\mathscr{O}'),N^\delta),$$
$$H = \mathtt{Merge}(h^{(1)},\cdots,h^{(r)},\mathscr{O}), \quad H' = \mathtt{Trim}(\mathtt{Merge}((h')^{(1)},\cdots,(h')^{(r)},\mathscr{O}'),N^\delta),$$

*then we have* $\mathcal{S}' = \mathtt{Trim}(\mathcal{S}, N^\delta))$. *Moreover, given any such* $\mathscr{O}'$, *the tensor* $H'$ *can be written as a linear combination of tensors* $\mathtt{Trim}(H, N^\delta)$ *(for different choices of* $\mathscr{O}$ *that are related to* $\mathscr{O}'$ *as above).*

*(4) Assume that some* $N_j = N$, $\mathcal{S}_j$ *is regular, and* $N_{j'} \geq N^\delta$ *for some* $j' \neq j$. *Then for the plant* $\mathcal{S} = \mathtt{Trim}(\mathtt{Merge}(\mathtt{Trim}(\mathcal{S}_1, N^\delta),\cdots,\mathtt{Trim}(\mathcal{S}_r, N^\delta),\mathscr{O}),N^\delta)$ *where* $\mathscr{O}$ *is as in Definition 3.6, we have* $|\mathcal{S}| > |\mathcal{S}_j|$.

*(5) Assume that each* $\mathcal{S}_j$ *is plain, and* $r = q$, *then the plant* $\mathcal{S} = \mathtt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_r,\mathscr{O})$, *where* $\mathscr{O}$ *is as in Definition 3.6, is also plain.*

*Proof.* First, (1) is obvious once we expand the $\Psi_k$ expressions using (3.8) and (3.9); also (5) directly follows from definition, noticing also that

$$\sum_{j=1}^{q}\sum_{\mathfrak{l}\in\mathcal{L}_j}\zeta_{\mathfrak{l}}^* = \sum_{j=1}^{q}\sum_{\mathfrak{l}\in\mathcal{L}_j}\zeta_j\zeta_{\mathfrak{l}} = \sum_{j=1}^{q}\zeta_j = 1.$$

Next, (4) is true because if $N_j = N$ and $\mathcal{S}_j$ is regular, then $\mathtt{Trim}(\mathcal{S}_j, N^\delta) = \mathcal{S}_j$, so by definition, if $\mathcal{S} = (\mathcal{L},\mathcal{V},\mathcal{Y})$ we have $\mathcal{L} \supset \mathcal{L}_j$ etc.; moreover as the second maximum of $N_1,\cdots,N_q$ is $\geq N^\delta$, by definition $\mathcal{Y}$ will have at least one more element than $\mathcal{Y}_j$, so $|\mathcal{S}| \geq |\mathcal{S}_j| + 1$.

Next consider (2). Recall $\mathcal{W} = \mathcal{U}_1\cup\cdots\cup\mathcal{U}_r$, using (3.8) and (3.13), we can expand the expression $\Phi_k$ as a sum over the variables $k_{\mathcal{W}}$ and $(k_1,\cdots,k_q)$ and $k_{\mathcal{V}_j}$ for $1 \leq j \leq r$, and an integration over the variables $(\lambda_{r+1},\cdots,\lambda_q)$ and $\lambda_{\mathcal{V}_j}$ for $1 \leq j \leq r$, of the quantity

$$h_{kk_1\cdots k_q}(\lambda_{r+1},\cdots,\lambda_q)\cdot\prod_{j=1}^{r}\big[h_{k_jk_{\mathcal{U}_j}}^{(j)}(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})\big]^{\zeta_j}\cdot\prod_{j=1}^{r}\prod_{\mathfrak{l}\in\mathcal{U}_j}(f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}^*}\cdot\prod_{j=1}^{r}\prod_{\mathfrak{f}\in\mathcal{V}_j}(\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}^*}(\lambda_{\mathfrak{f}})\prod_{j=r+1}^{q}(\widehat{z_{N_j}})_{k_j}^{\zeta_j}(\lambda_j), \tag{3.14}$$

[1]Here and below the phrase "linear combination" will refer to a linear combination with a fixed number of terms and fixed constant coefficients.

where in the summation we do not impose any pairing or no-pairing condition (with respect to the $\zeta_{\mathfrak{l}}^*$ signs, same below) for the variables $k_{\mathcal{W}}$ (of course if there is a pairing within $k_{\mathcal{U}_j}$ then $h^{(j)}=0$ so the quantity (3.14) is zero).

On the other hand, using (3.8) and (3.11), and noticing that $(f_N)_k\overline{(f_N)_k}=\Delta_N(\gamma_k)^2$, we can write $\Psi_k[\mathcal{S},H]$, corresponding to a certain choice of $\mathscr{O}=\{\mathcal{A}_1,\cdots,\mathcal{A}_m\}$ as in Definition 3.6, as a sum and integration of the same quantity (3.14) over the same set of variables as in $\Phi_k$, but with a set of additional pairing and no-pairing conditions on $k_{\mathcal{W}}$. Precisely, the extra conditions are (i) the $k_{\mathfrak{l}}$ are the same for $\mathfrak{l}$ in each $\mathcal{A}_i$; (ii) there is no pairing in $k_{\mathcal{W}\backslash\mathcal{Q}}$ where $\mathcal{Q}$ is defined in Definition 3.6. We denote this set of extra conditions by $(\mathscr{O},1)$. With these observations, it suffices to show that the sum $\sum_{k_{\mathcal{W}}}$ can be written as a linear combination of sums $\sum_{(\mathscr{O},1)}$ for different $\mathscr{O}$. Now, by identifying the exact set of pairings among $k_{\mathcal{W}}$, we can write $\sum_{k_{\mathcal{W}}}$ as a linear combination of sums $\sum_{(\mathscr{O},2)}$ for different $\mathscr{O}$'s , where $(\mathscr{O},2)$ represents a different set of extra conditions, namely (i) the $k_{\mathfrak{l}}$ are the same for $\mathfrak{l}$ in each $\mathcal{A}_i$; (ii) the $k_{\mathfrak{l}}$ for different $\mathcal{A}_i$ are different, and are different from any $k_{\mathfrak{l}}$ for $\mathfrak{l}$ not in any $\mathcal{A}_i$; (iii) there is no pairing among the $k_{\mathfrak{l}}$ where $\mathfrak{l}$ is not in any $\mathcal{A}_i$. Note that we may assume these $\mathscr{O}$'s are as in Definition 3.6, i.e. each $\mathcal{A}_i$ contains two elements of opposite sign but does not contain two elements of opposite sign that belong to the same $\mathcal{U}_j$, and $N_{\mathfrak{l}}$ for $\mathfrak{l}$ in each $\mathcal{A}_i$ are all the same, since otherwise the summand (3.14) would be zero by Definition 3.4.

Clearly the condition $(\mathscr{O},2)$ is stronger than $(\mathscr{O},1)$, and the difference $\sum_{(\mathscr{O},1)}-\sum_{(\mathscr{O},2)}$ can be written as a linear combination of sums $\sum_{(\mathscr{W},2)}$ for different $\mathscr{W}$'s, where $\mathscr{W}$ has the same form as $\mathscr{O}$, such that the condition $(\mathscr{W},2)$ gives *strictly* more pairings than $(\mathscr{O},2)$. Thus, we can inductively write $\sum_{k_{\mathcal{W}}}$ as a linear combination of sums $\sum_{(\mathscr{O},1)}$ for different $\mathscr{O}$. In this way we have written $\Phi_k$ as a linear combination of $\Psi_k=\Psi_k[\mathcal{S},H]$ for different choices of $\mathscr{O}$, which proves (2).

Finally look at (3). As $\mathscr{O}'$ consists of those subsets in $\mathscr{O}$ that are contained in the union of $\mathcal{U}_j'$ (note that any $\mathcal{A}_i$ is either contained in the union of $\mathcal{U}_j'$, or contained in the union of $\mathcal{U}_j\backslash\mathcal{U}_j'$), we know that $\mathscr{P}'$ (defined from $\mathscr{O}'$ as in Definition 3.6) also consists of those subsets in $\mathscr{P}$ that are contained in the union of $\mathcal{U}_j'$. Then the equality $\mathcal{S}'=\mathtt{Trim}(\mathcal{S},N^\delta)$ follows from Definitions 3.5 and 3.6, and straightforward verification. Note that if $N_j\geq N^\delta$ for $r+1\leq j\leq s$, and $N_j<N^\delta$ for $s+1\leq j\leq q$ (which we may assume), then $\{r+1,\cdots,s\}\subset\mathcal{V}'$ and $\{s+1,\cdots,q\}\subset\mathcal{V}\backslash\mathcal{V}'$.

Now look at the tensor $H'$. Let $\mathcal{R}$ be the union of all the $\mathcal{U}_j\backslash\mathcal{U}_j'$ and $\mathcal{Z}$ be the union of all the $\mathcal{V}_j\backslash\mathcal{V}_j'$, and let $\mathcal{Q}'$ be defined as in Definition 3.6, which occurs in the process of merging $\mathcal{S}_1',\cdots,\mathcal{S}_r'$ via $\mathscr{O}'$, then using (3.9) and (3.11), we can expand $(H')_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'},\lambda_{\mathcal{V}'})$ as a sum over the variables $(k_{s+1},\cdots,k_q)$ and $k_{\mathcal{Q}'}$ and $k_{\mathcal{R}}$ and $k_{\mathcal{Z}}$, and integration over the variables $(\lambda_{s+1},\cdots,\lambda_q)$ and $\lambda_{\mathcal{Z}}$, of the quantity

$$\begin{aligned}&\prod_{\mathfrak{l},\mathfrak{l}'}^{(1)}\mathbf{1}_{k_{\mathfrak{l}}=k_{\mathfrak{l}'}}\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}\cdot\sum_{(k_1,\cdots,k_r)}h_{kk_1\cdots k_q}(\lambda_{r+1},\cdots,\lambda_q)\\&\qquad\times\sum_{k_{\mathcal{Q}'}}^{(3)}\prod_{\mathfrak{l}\in\mathcal{Q}'}\Delta_{N_{\mathfrak{l}}}\gamma_{k_{\mathfrak{l}}}\prod_{j=1}^{r}\big[h_{k_jk_{\mathcal{U}_j}}^{(j)}(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})\big]^{\zeta_j}\prod_{\mathfrak{l}\in\mathcal{R}}(f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}^*}\prod_{\mathfrak{f}\in\mathcal{Z}}(\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}^*}(\lambda_{\mathfrak{f}})\prod_{j=s+1}^{q}(\widehat{z_{N_j}})_{k_j}^{\zeta_j}(\lambda_j),\end{aligned}\tag{3.15}$$

where the sums and products $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1)}$, $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}$ and $\sum_{k_{\mathcal{Q}'}}^{(3)}$ are defined as in Definition 3.6 in the process of merging $\mathcal{S}_1',\cdots,\mathcal{S}_r'$ via $\mathscr{O}'$, and in this summation we do not impose any pairing or no-pairing condition for the variables $k_{\mathcal{R}}$. The signs $\zeta_{\mathfrak{n}}^*$ are also defined as in Definition 3.6.

On the other hand, if $\mathscr{O} = \mathscr{O}' \cup \mathscr{W}$, then using (3.9) and (3.11) again, we can expand the tensor $(\texttt{Trim}(H, N^\delta))_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'}, \lambda_{\mathcal{V}'})$ as a sum and integration of the same quantity (3.15) over the same set of variables as in $(H')_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'}, \lambda_{\mathcal{V}'})$, but with the $k_{\mathcal{R}}$ variables satisfying a set of additional pairing and no-pairing conditions, given exactly by $(\mathscr{W}, 1)$. Therefore, the same arguments as in the proof of part (2) above also imply that $H'$ can be written as a linear combination of $\texttt{Trim}(H, N^\delta)$ for different choices of $\mathscr{O}$. This completes the proof. $\square$

3.3. **Working norms.** Based on the tensor norms of Definition 2.1, we can define the norms involving the modulation variables $\lambda, \lambda'$, etc., as well as some other parameters; these will be the norms used in the main proof.

Suppose $b_1, b_2 \in [0, 1]$, $h = h_{k_A}(t)$ depends on $t$, and let $\widehat{h}$ be the Fourier transform of $h$ in $t$. Let $(B, C)$ be a subpartition of $A$, we define

$$\|h\|^2_{X^{b_1}[k_B \to k_C]} = \int_{\mathbb{R}} \langle\lambda\rangle^{2b_1} \|\widehat{h}_{k_A}(\lambda)\|^2_{k_B \to k_C}\,\mathrm{d}\lambda. \tag{3.16}$$

If $h = h_{k_A}(k_F, \lambda_F)$ depends on some parameters $(k_F, \lambda_F)$, we define

$$\|h\|^2_{X_F^{-b_2}[k_B \to k_C]} = \sum_{k_F} \int \mathrm{d}\lambda_F \cdot \prod_{j \in F} \langle\lambda_j\rangle^{-2b_2} \|h_{k_A}(k_F, \lambda_F)\|^2_{k_B \to k_C}. \tag{3.17}$$

If $h = h_{k_A}(t, k_F, \lambda_F)$ depends on both $t$ and $(k_F, \lambda_F)$, we define

$$\|h\|^2_{X_F^{b_1, -b_2}[k_B \to k_C]} = \int_{\mathbb{R}} \langle\lambda\rangle^{2b_1} \sum_{k_F} \int \mathrm{d}\lambda_F \cdot \prod_{j \in F} \langle\lambda_j\rangle^{-2b_2} \|\widehat{h}_{k_A}(\lambda, k_F, \lambda_F)\|^2_{k_B \to k_C}\,\mathrm{d}\lambda, \tag{3.18}$$

where $\widehat{h}$ is the Fourier transform of $h$ in $t$ only; note that

$$\begin{aligned}\|h\|^2_{X_F^{b_1, -b_2}[k_B \to k_C]} &= \int_{\mathbb{R}} \langle\lambda\rangle^{2b_1} \|\widehat{h}_{k_A}(\lambda, \cdot, \cdot)\|^2_{X_F^{-b_2}[k_B \to k_C]}\,\mathrm{d}\lambda \\ &= \sum_{k_F} \int \mathrm{d}\lambda_F \cdot \prod_{j \in F} \langle\lambda_j\rangle^{-2b_2} \|h_{k_A}(\cdot, k_F, \lambda_F)\|^2_{X^{b_1}[k_B \to k_C]}.\end{aligned} \tag{3.19}$$

Now, given a tensor $h = h_{k_A}(t, t')$, a subpartition $(B, C)$, and $b_1, b_2 \in [0, 1]$, we can similarly define

$$\|h\|^2_{X^{b_1, -b_2}[k_B \to k_C]} = \int_{\mathbb{R}^2} \langle\lambda\rangle^{2b_1} \langle\lambda'\rangle^{-2b_2} \|\widehat{h}_{k_A}(\lambda, \lambda')\|^2_{k_B \to k_C}\,\mathrm{d}\lambda\mathrm{d}\lambda', \tag{3.20}$$

where $\widehat{h}$ is the Fourier transform of $h$ in $(t, t')$. Finally, if $s_1 \in \mathbb{R}$, the $X^{s_1, b_1}$ norm for a function $f = f_k(t)$ is defined by

$$\|f\|^2_{X^{s_1, b_1}} = \int_{\mathbb{R}} \langle\lambda\rangle^{2b_1} \|\langle k\rangle^{s_1} \widehat{f}_k(\lambda)\|^2_{\ell^2_k}\,\mathrm{d}\lambda. \tag{3.21}$$

When $s_1 = 0$ we simplify write $X^{b_1}$.

## 4. Preliminaries II: Estimates

In this section we collect the important core estimates. Sections 4.1–4.2 contain the basic linear and large deviation estimates. Section 4.3 contains counting estimates ultimately leading to Proposition 4.9, and Section 4.4 contains the main tensor norm estimates, Propositions 4.11–4.12 and 4.14–4.15.

4.1. **Linear estimates.** We record two estimates for Duhamel and time localization operators (recall from (3.6) the definition of $\mathcal{I}_\chi$), and another weighted estimate. For the proofs see [35, 36].

**Lemma 4.1.** *We have the formula*

$$\widehat{\mathcal{I}_\chi v}(\lambda) = \int_{\mathbb{R}} \mathcal{I}(\lambda,\lambda')\widehat{v}(\lambda')\,\mathrm{d}\lambda', \tag{4.1}$$

*where the kernel $\mathcal{I}$ satisfies that*

$$|\mathcal{I}| + |\partial_{\lambda,\lambda'}\mathcal{I}| \lesssim \left(\frac{1}{\langle\lambda\rangle^3} + \frac{1}{\langle\lambda-\lambda'\rangle^3}\right)\frac{1}{\langle\lambda'\rangle} \lesssim \frac{1}{\langle\lambda\rangle\langle\lambda-\lambda'\rangle}. \tag{4.2}$$

*Proof.* See [35], Lemma 3.1 whence by a similar proof, one can see that (4.2) also holds for $|\partial_{\lambda,\lambda'}\mathcal{I}|$. $\square$

**Lemma 4.2.** *Let $v = v(t)$ be a function on $\mathbb{R}$ valued in some Banach function space. For $b_1 \in [0,1]$ define the $Y^{b_1}$ norm by*

$$\|v\|_{Y^{b_1}}^2 = \int_{\mathbb{R}} \langle\lambda\rangle^{2b_1}\|\widehat{v}(\lambda)\|^2\,\mathrm{d}\lambda, \tag{4.3}$$

*where $\widehat{v}$ is the (vector-valued) Fourier transform of $u$. For $\tau \lesssim 1$ let $\chi_\tau(t) = \chi(\tau^{-1}t)$ be as in Section 3.1, then for any $0 < b_1 \le b_2 < 1/2$ and for any $v$, or for any $1/2 < b_1 \le b_2 < 1$ and for any $v$ satisfying*[1] *$v(0) = 0$, we have*

$$\|\chi_\tau \cdot v\|_{Y^{b_1}} \lesssim \tau^{b_2-b_1}\|v\|_{Y^{b_2}}. \tag{4.4}$$

*Proof.* See [36], Proposition 2.7 (which proves the scalar case, but the proof directly carries over to vector valued cases). $\square$

**Lemma 4.3.** *Fix $\kappa_1 > 0$. Let $\mathscr{M} = \mathscr{M}_{kk'}(\lambda,\lambda')$ be the kernel of an operator $\mathscr{M}$, namely*

$$(\mathscr{M}w)_k(\lambda) = \sum_{k'}\int_{\mathbb{R}} \mathscr{M}_{kk'}(\lambda,\lambda')w_{k'}(\lambda')\,\mathrm{d}\lambda',$$

*and assume that $\mathscr{M}$ is supported in $|k-k'| \le R$ for some dyadic $R$. Then uniformly in any $R$ and any $k^0 \in \mathbb{Z}^d$, we have*

$$\|(1+R^{-1}|k-k^0|)^{\kappa_1}(\mathscr{M}w)_k(\lambda)\|_{\ell_k^2 L_\lambda^2} \lesssim \|\mathscr{M}\|_{\ell_{k'}^2 L_{\lambda'}^2 \to \ell_k^2 L_\lambda^2} \cdot \|(1+R^{-1}|k'-k^0|)^{\kappa_1}w_{k'}(\lambda')\|_{\ell_{k'}^2 L_{\lambda'}^2}. \tag{4.5}$$

*Proof.* See [36], Proposition 2.5. $\square$

4.2. **Large deviation inequalities.** We state a large deviation estimate that works for uniform distributions on the unit circle, see [36].

**Lemma 4.4.** *Let $E \subset \mathbb{Z}^d$ be a finite set, $a = a_{k_1\cdots k_r}(\omega)$ be a random tensor such that the collection $\{a_{k_1\cdots k_r}\}$ is independent with the collection $\{\eta_k(\omega) : k \in E\}$. Let $\zeta_j \in \{\pm\}$ and assume that in the support of $a_{k_1\cdots k_r}$ there is no pairing in $(k_1,\cdots,k_r)$ associated with the signs $\zeta_j$. Let the random variable*

$$X(\omega) := \sum_{k_1,\cdots,k_r} a_{k_1\cdots k_r}(\omega)\prod_{j=1}^r \eta_j(\omega)^{\zeta_j}, \tag{4.6}$$

*then for any $A > 0$, we have $A$-certainly that*

$$|X(\omega)|^2 \le A^\theta \cdot \sum_{k_1,\cdots,k_r} |a_{k_1\cdots k_r}(\omega)|^2. \tag{4.7}$$

[1] In practice, the factor $\chi_\tau$ will always come with a $v$ which has the form $\mathcal{I}_\chi(\cdots)$, so we always have $v(0) = 0$.

*Proof.* This is a special case (i.e. no pairing) of [36], Lemma 4.1. □

4.3. **Lattice point counting bounds.** Here we state and prove the various counting bounds that eventually lead to Proposition 4.9.

**Lemma 4.5.** *Consider the set*

$$S^{(3)} = \big\{(k_1,k_2,k_3) \in (\mathbb{Z}^d)^3 : \zeta_1 k_1 + \zeta_2 k_2 + \zeta_3 k_3 = m,\ \zeta_1|k_1|^2 + \zeta_2|k_2|^2 + \zeta_3|k_3|^2 = \Gamma,$$
$$|k_1 - m_1| \leq M_1,\ |k_2 - m_2| \leq M_2,\ \text{and there is no pairing in } (k_1,k_2,k_3)\big\}, \quad (4.8)$$

*where $\zeta_j \in \{\pm\}$, $(m,\Gamma) \in \mathbb{Z}^d \times \mathbb{Z}$, $m_j \in \mathbb{Z}^d$ and $M_j$ dyadic are given. Then we have, uniformly in all parameters, that*

$$\#S^{(3)} \lesssim (M_1M_2)^{d-1+\theta}. \quad (4.9)$$

*Proof.* We may assume $M_1 \leq M_2$. If $d = 1$, by simple algebra, we can reduce to a divisor counting problem in $\mathbb{Z}[e^{2\pi i/3}]$ (if $\zeta_1 = \zeta_2 = \zeta_3$) or $\mathbb{Z}$ (otherwise). Since each $k_j$ belongs to an interval of length $O(M_2)$, the estimate (4.9) follows from Lemma 3.4 of [35].

Consider $d \geq 2$. Without loss of generality, we may assume (due to no-pairing) that either *the first coordinates* of $(k_1,k_2,k_3)$ do not contain a pairing, or the $j$-th coordinates of $(k_1,k_2,k_3)$ contain a pairing for each $j$, and this pairing is *not* from $(k_2,k_3)$ for $j = 1$. In the former case the $j$-th coordinates of $(k_1,k_2,k_3)$ have at most $M_1M_2$ choices for each $2 \leq j \leq d$, and then at most $(M_1M_2)^\theta$ choices for $j = 1$ thanks to the $d = 1$ case. In the latter case the $j$-th coordinates of $(k_1,k_2,k_3)$ have at most $M_2$ choices for $2 \leq j \leq d$, and at most $M_1$ choices for $j = 1$. In either case we get

$$\#S^{(3)} \lesssim \max((M_1M_2)^{d-1}(M_1M_2)^\theta, M_2^{d-1}M_1) \lesssim (M_1M_2)^{d-1+\theta}. \qquad \square$$

**Lemma 4.6.** *Recall that $p \geq 3$ is odd. For $1 \leq p_1 \leq p$, consider a partition of a set $A \subset \{1,\cdots,p\}$, $|A| = p_1$, into pairwise disjoint nonempty subsets $B_1,\cdots,B_t$, say $B_u = \{i_u(1),\cdots,i_u(b_u)\}$ $(1 \leq u \leq t)$. Given $m_j \in \mathbb{Z}^d$, $M_j$ dyadic, $\zeta_j \in \{\pm\}$ $(j \in A)$ and $\Gamma \in \mathbb{Z}$, consider the set $S$ consisting of vectors $k_A$ (where each $k_j \in \mathbb{Z}^d$) that satisfy*

$$\sum_{j\in A} \zeta_j |k_j|^2 = \Gamma; \quad \Big|\sum_{z=1}^{y} \zeta_{i_u(z)} k_{i_u(z)} - m_{i_u(y)}\Big| \leq M_{i_u(y)},\ \forall 1 \leq u \leq t, 1 \leq y \leq b_u = |B_u|. \quad (4.10)$$

*Assume $M_{i_u(b_u)} = 1$ for $1 \leq u \leq t$, and that there is no pairing in $k_A$. Then we have, uniformly in all parameters, that*

$$\#S \lesssim \prod_{j\in A} (M_j)^{2\alpha_0+\theta}, \quad (4.11)$$

*where $\alpha_0$ is as in (3.1).*

*Proof.* We may assume $p_1 = p$, since if $p_1 < p$ we can add some elements to the sets $B_u$ and reduce to the $p_1 = p$ case. We will prove (4.11) by induction. Suppose (4.11) is true for $p - 2$, we will prove it for $p$. For simplicity we define

$$w_{i_u(y)} := \sum_{z=1}^{y} \zeta_{i_u(z)} k_{i_u(z)}, \quad 1 \leq u \leq t,\ 1 \leq y \leq b_u.$$

Since $p$ is odd, at least one of $|B_u|$ (say $|B_1|$) must be odd. If $|B_1| = 1$ then the value of $k_{i_1(1)}$ is fixed, so we only need to count the vector $k_{A\setminus B_1}$. If $|B_2| = 1$ also then we may reduce to counting $k_{A\setminus(B_1\cup B_2)}$ and apply the induction hypothesis; otherwise $|B_2| \geq 2$ and we may add an element to

$B_2$ at no cost and reduce to the case $|B_2| \geq 3$. Therefore in any case we may assume some $|B_j| \geq 3$, say $|B_1| = b_1 \geq 3$.

Let $(i_1(b_1), i_1(b_1-1), i_1(b_1-2)) = (n, n', n'')$. Since by (4.10), each $w_{i_u(y)}$ belongs to a ball of radius $M_{i_u(y)}$, and $k_{i_u(y)} = \pm w_{i_u(y)} \pm w_{i_u(y-1)}$ for some independent choices of $\pm$ (same below), the number of choices for the vector $k_{B_1\backslash\{n,n',n''\}}$ is at most

$$M_{i_1(1)}^d \cdots M_{i_1(b_1-3)}^d = (M_{n'}M_{n''})^{-d} \prod_{j\in B_1} M_j^d,$$

noticing that $M_n = 1$. Similarly for $2 \leq u \leq t$, the number of choices for the vector $k_{B_u}$ is at most $\prod_{j\in B_u} M_j^d$. Once $k_{A\backslash\{n,n',n''\}}$ is fixed, $k_{n''} = \pm w_{n''} \pm w_{i_1(b_1-3)}$ belongs to a ball of radius $M_{n''}$, and $k_n = \pm w_n \pm w_{n'}$ belongs to a ball of radius $M_{n'}$. Then the number of choices for $(k_n, k_{n'}, k_{n''})$ can be bounded by $(M_{n'}M_{n''})^{d-1+\theta}$ by Lemma 4.5, thus

$$\#S \lesssim (M_{n'}M_{n''})^{d-1+\theta} \prod_{j\in A\backslash\{n,n',n''\}} M_j^d. \tag{4.12}$$

Note that this also settles the base case $p = 3$.

On the other hand, for $p \geq 5$, since $M_n = 1$, by (4.10) we know that $k_n = \pm w_n \pm w_{n'}$ belongs to a ball of radius $M_{n'}$, and when $k_n$ is fixed, $k_{n'} = \pm w_n \pm k_n \pm w_{n''}$ belongs to a ball of radius $M_{n''}$. Thus the number of choices for $(k_n, k_{n'})$ is at most $(M_{n'}M_{n''})^d$. When $(k_n, k_{n'})$ is fixed, we only need to count the vectors $k_{A\backslash\{n,n'\}}$. Now $w_{i_1(y)}$ belongs to a ball of radius $M_{i_1(y)}$ if $y \leq b_1 - 3$, and to a ball of radius 1 if $y = b_1 - 2$, so by the induction hypothesis we conclude that

$$\#S \lesssim (M_{n'}M_{n''})^d \prod_{j\in A\backslash\{n,n',n''\}} M_j^{d-\frac{2}{p-3}+\theta}. \tag{4.13}$$

Interpolating (4.12) and (4.13) we get

$$\#S \lesssim \prod_{j\in A\backslash\{n\}} M_j^{d-\frac{2}{p-1}+\theta},$$

which is just (4.11). ☐

**Lemma 4.7.** *For $1 \leq p_1 \leq p$, consider a partition of a set $A \subset \{1, \cdots, p\}$, $|A| = p_1$, into pairwise disjoint nonempty subsets $A_1, \cdots, A_s$ and $B_1, \cdots, B_t$, say $A_v = \{\ell_v(1), \cdots, \ell_v(a_v)\}\,(1 \leq v \leq s)$ and $B_u = \{i_u(1), \cdots, i_u(b_u)\}\,(1 \leq u \leq t)$. Given $m_j \in \mathbb{Z}^d$, $M_j$ dyadic, $\zeta_j \in \{\pm\}\,(j \in A)$ and $\Gamma, \Gamma_v \in \mathbb{Z}\,(1 \leq v \leq s)$, consider the set $S$ consisting of vectors $k_A$ (where each $k_j \in \mathbb{Z}^d$) that satisfy*

$$\sum_{j\in A} \zeta_j |k_j|^2 = \Gamma, \quad \left|\sum_{z=1}^{y} \zeta_{i_u(z)} k_{i_u(z)} - m_{i_u(y)}\right| \leq M_{i_u(y)}, \forall 1 \leq u \leq t, 1 \leq y \leq b_u = |B_u|, \tag{4.14}$$

$$\sum_{j\in A_v} \zeta_j |k_j|^2 = \Gamma_v, \quad \left|\sum_{z=1}^{y} \zeta_{\ell_v(z)} k_{\ell_v(z)} - m_{\ell_v(y)}\right| \leq M_{\ell_u(y)}, \forall 1 \leq v \leq s, 1 \leq y \leq a_v = |A_v|. \tag{4.15}$$

*Assume that $M_{\ell_v(a_v)} = 1$ and $|A_v| \leq p - 2$ for each $1 \leq v \leq s$, that $M_{i_u(b_u)} = 1$ for each $1 \leq u \leq t$, and that there is no pairing in $k_A$. Then we have, uniformly in all parameters, that*

$$\#S \lesssim \prod_{j\in A} (M_j)^{2\alpha_0+\theta} \prod_{v=1}^{s} (\min_{1\leq y<a_v} M_{\ell_v(y)})^{2\alpha_0-d}. \tag{4.16}$$

*Proof.* This essentially follows from Lemma 4.6. Let $B = B_1 \cup \cdots \cup B_t$, by (4.14) and (4.15) we may count $k_B$ and each $k_{A_v}$ $(1 \leq v \leq s)$ separately. By Lemma 4.6, the number of choices for $k_B$ is at most

$$\prod_{j\in B}(M_j)^{2\alpha_0+\theta};$$

thus it suffices to prove that the number of choices for $k_{A_v}$ is at most

$$\prod_{j=1}^{a_v-1}(M_{\ell_v(y)})^{2\alpha_0+\theta}\cdot(\min_{1\leq y<a_v}M_{\ell_v(y)})^{2\alpha_0-d}.$$

Let $|A_v| = a_v = n$, clearly we may assume $n \geq 2$. Lemma 4.6 then implies that the number of choices for $k_{A_v}$ is at most

$$\prod_{y=1}^{n-1}(M_{\ell_v(y)})^{d-\frac{2}{n'}+\theta},$$

where $n' = n$ if $n$ is even, and $n' = n-1$ if $n$ is odd. Now the desired estimate follows, since

$$\prod_{y=1}^{n-1}(M_{\ell_v(y)})^{d-\frac{2}{n'}}\leq\prod_{y=1}^{n-1}(M_{\ell_v(y)})^{2\alpha_0}\cdot(\min_{1\leq y<a_v}M_{\ell_v(y)})^{2\alpha_0-d},$$

due to the elementary inequality

$$(n-1)\big(2\alpha_0-d+\frac{2}{n'}\big)\geq d-2\alpha_0$$

which can be verified for $2 \leq n \leq p-2$. ☐

**Lemma 4.8.** *Consider the same setting and set $S$ as in Lemma 4.7, but instead of no pairing in $k_A$ we assume that (1) any pairing in $k_A$ must be over-paired, and (2) $d(p-1) \geq 8$. Then the bound (4.16) remains true.*

*Proof.* As in the proof of Lemma 4.6, we define

$$w_{i_u(y)}=\sum_{z=1}^{y}\zeta_{i_u(z)}k_{i_u(z)},\quad 1\leq u\leq t,\ 1\leq y\leq b_u;$$
$$w_{\ell_v(y)}=\sum_{z=1}^{y}\zeta_{\ell_v(z)}k_{\ell_v(z)},\quad 1\leq v\leq s,\ 1\leq y\leq a_v.$$

We also understand $M_{i_u(y)} = M_{\ell_v(y)} = 1$ when $y = 0$. It suffices to bound $\#S$ by

$$\mathfrak{N}=\prod_{v=1}^{s}\mathfrak{N}_v\prod_{u=1}^{t}\mathfrak{N}_u^*,\tag{4.17}$$

where for $1 \leq v \leq s$ and $1 \leq u \leq t$,

$$\mathfrak{N}_v=\prod_{1\leq y<a_v}(M_{\ell_v(y)})^{2\alpha_0+\theta}\cdot(\min_{1\leq y<a_v}M_{\ell_v(y)})^{2\alpha_0-d},\quad\text{and}\quad\mathfrak{N}_u^*=\prod_{1\leq y<b_u}(M_{i_u(y)})^{2\alpha_0+\theta}.\tag{4.18}$$

We proceed by induction. The no-pairing case is already known by Lemma 4.7. Now suppose there is some over-pairing in $k_A$; we list all the different $j$'s in $A$ (there are at least three of them) such that $k_j$ are the same, and let this common value be $k$. We will fix $k$, count the remaining variables, and then sum in $k$. With $k$ fixed, by using induction hypothesis, the number of choices for the remaining variables will be bounded by some product

$$\mathfrak{N}(k)=\prod_{v=1}^{s}\mathfrak{N}_v(k)\prod_{u=1}^{t}\mathfrak{N}_u^*(k),$$

and we need to bound the quotients $\mathfrak{N}_v(k)/\mathfrak{N}_v$ and $\mathfrak{N}_u^*(k)/\mathfrak{N}_u^*$. We only need to consider those $v$ and $u$ such that at least one $k_j$ ($j\in A_v$ or $B_u$) equals $k$ (otherwise the quotient is 1). There will be several cases depending on how many $k_j$ equal $k$, and their positions. First, if $|A_v|=1$ (or $|B_u|=1$), then the value of $k$ will be uniquely determined, so there is no summation in $k$, and the result follows from the induction hypothesis. From now on we will assume $|A_v|\geq 2$ and $|B_u|\geq 2$. Similarly, if $|A_v|=3$ or $|B_u|=3$ then the elements in $\{k_j:j\in A_v\}$ (or $\{k_j:j\in B_u\}$) cannot all equal $k$, since otherwise $k$ would also be uniquely determined.

(1) If all (or all but one) elements in $\{k_j:j\in A_v\}$ (or in $\{k_j:j\in B_u\}$) equal $k$, then $\mathfrak{N}_v(k)$ (or $\mathfrak{N}_u^*(k)$) will be equal to 1. Considering $B_u$, since $k_{i_u(y)}=\pm w_{i_u(y)}\pm w_{i_u(y-1)}$, by (4.14) we see that if $k_{i_u(y)}=k$, then there exists a vector $m^*$ not depending on $k$, such that $\max(M_{i_u(y)},M_{i_u(y-1)})\gtrsim |k-m^*|$. This implies that

$$\frac{\mathfrak{N}_u^*(k)}{\mathfrak{N}_u^*}\lesssim\begin{cases}|k-m^*|^{-2\alpha_0-\theta}, & \text{if } |B_u|\leq 3;\\ |k-m^*|^{-2\alpha_0-\theta}|k-m^{**}|^{-2\alpha_0-\theta}, & \text{if } |B_u|\geq 4,\end{cases}\tag{4.19}$$

where $m^{**}$ is another vector not depending on $k$.

Similarly considering $A_v$, in view of the extra factor in (4.18), the estimates will be

$$\frac{\mathfrak{N}_v(k)}{\mathfrak{N}_v}\lesssim\begin{cases}|k-m^*|^{d-4\alpha_0-\theta}, & \text{if } |A_v|=2;\\ |k-m^*|^{-2\alpha_0-\theta}, & \text{if } |A_v|\geq 3;\\ |k-m^*|^{-2\alpha_0-\theta}|k-m^{**}|^{d-4\alpha_0-\theta}, & \text{if } |A_v|\geq 4.\end{cases}\tag{4.20}$$

(2) If at least two elements in $\{k_j:j\in A_v\}$ (or $\{k_j:j\in B_u\}$) do not equal $k$, then in particular $|A_v|\geq 3$ (or $|B_u|\geq 3$). In this case we only need to consider $B_u$, and the bounds for $A_v$ will be the same (if not better) due to the negative power of $\min_y M_{\ell_v(y)}$ in (4.18). If we fix $k_{i_u(y)}=k$, then $w_{i_u(y-1)}$ belongs to a ball of radius $\min(M_{i_u(y)},M_{i_u(y-1)})$, so we get

$$\frac{\mathfrak{N}_u^*(k)}{\mathfrak{N}_u^*}\lesssim\max(M_{i_u(y)},M_{i_u(y-1)})^{-2\alpha_0-\theta},\tag{4.21}$$

which is bounded by $|k-m^*|^{-2\alpha_0-\theta}$, in the same way as in (1).

(3) By similar arguments as in (2), we know that if two non-adjacent elements in $B_u$ equal $k$ (similarly for $A_v$), say $k_{i_u(y)}=k_{i_u(y')}=k$, then $w_{i_u(y-1)}$ belongs to a ball of radius $\min(M_{i_u(y)},M_{i_u(y-1)})$ and $w_{i_u(y'-1)}$ belongs to a ball of radius $\min(M_{i_u(y')},M_{i_u(y'-1)})$, so we have

$$\frac{\mathfrak{N}_u^*(k)}{\mathfrak{N}_u^*}\lesssim\max(M_{i_u(y)},M_{i_u(y-1)})^{-2\alpha_0-\theta}\max(M_{i_u(y')},M_{i_u(y'-1)})^{-2\alpha_0-\theta}\lesssim|k-m^*|^{-2\alpha_0-\theta}|k-m^{**}|^{-2\alpha_0-\theta}.$$

In summary, since at least three elements in all the $A_v$'s and $B_u$'s equal $k$, we conclude that

$$\frac{\mathfrak{N}(k)}{\mathfrak{N}}=\prod_{v=1}^{s}\frac{\mathfrak{N}_v(k)}{\mathfrak{N}_v}\prod_{u=1}^{t}\frac{\mathfrak{N}_u^*(k)}{\mathfrak{N}_u^*}\lesssim|k-m^*|^{d-4\alpha_0-\theta}|k-m^{**}|^{d-4\alpha_0-\theta}.$$

Since $2(4\alpha_0 - d + \theta) > d$ because $d(p-1) \geq 8$, we conclude that

$$\sum_{k\in\mathbb{Z}^d} |k-m^*|^{d-4\alpha_0-\theta}|k-m^{**}|^{d-4\alpha_0-\theta} \leq O(1) \Rightarrow \sum_{k\in\mathbb{Z}^d} \mathfrak{N}(k) \lesssim \mathfrak{N},$$

which completes the proof. □

**Proposition 4.9.** *Partition $\{1,\cdots,p\}$ into disjoint nonempty subsets $A_1,\cdots,A_s$, $B_1,\cdots,B_t$ and $C$. Assume $A_v = \{\ell_v(1),\cdots,\ell_v(a_v)\}$, $B_u = \{i_u(1),\cdots,i_u(b_u)\}$, and $C = \{n_1,\cdots,n_c = 1\}$. Given $m_j \in \mathbb{Z}^d$, $M_j$ dyadic, $\zeta_j \in \{\pm\}\,(1 \leq j \leq p)$ and $\Gamma, \Gamma_v \in \mathbb{Z}\,(1 \leq v \leq s)$, consider a tensor $h = h_{kk_1\cdots k_p}$, where each $k, k_j \in \mathbb{Z}^d$, which satisfies $|h_{kk_1\cdots k_p}| \lesssim 1$, and in the support of $h$ we have*

$$\sum_{j\in A_v} \zeta_j|k_j|^2 = \Gamma_v, \quad \left|\sum_{z=1}^{y} \zeta_{\ell_v(z)}k_{\ell_v(z)} - m_{\ell_v(y)}\right| \leq M_{\ell_u(y)}, \forall 1 \leq v \leq s, 1 \leq y \leq a_v = |A_v|. \tag{4.22}$$

$$\sum_{j=1}^{p} \zeta_j|k_j|^2 - |k|^2 = \Gamma, \quad \left|\sum_{z=1}^{y} \zeta_{i_u(z)}k_{i_u(z)} - m_{i_u(y)}\right| \leq M_{i_u(y)}, \forall 1 \leq u \leq t, 1 \leq y \leq b_u = |B_u|, \tag{4.23}$$

$$\left|\sum_{z=1}^{y} \zeta_{n_z}k_{n_z} - m_{n_y}\right| \leq M_{n_y}, \forall 1 \leq y \leq c-1 = |C|-1. \tag{4.24}$$

*We assume that $M_{\ell_v(a_v)} = M_{i_u(b_u)} = M_{n_c} = 1$, and that*

$$\sum_{j\in A_v} \zeta_j = 0\,(\forall 1 \leq v \leq s), \ \sum_{j=1}^{p} \zeta_j = 1; \quad \sum_{j=1}^{p} \zeta_j k_j - k = \sum_{j\in A_v} \zeta_j k_j = 0. \tag{4.25}$$

*We also assume that any pairing in $(k, k_1,\cdots,k_p)$ is over-paired. Then, for any subset $P_0$ satisfying $P_0 \subset C\backslash\{1\}$, let $\{1,\cdots,p\}\backslash P_0 = Q_0$, then we have*

$$\|h\|_{kk_{P_0}\to k_{Q_0}} \lesssim \prod_{j=2}^{p}(M_j)^{\alpha_0+\theta}\prod_{v=1}^{s}(\min_{1\leq y<a_v} M_{\ell_v(y)})^{\alpha_0-\frac{d}{2}}, \tag{4.26}$$

unless $(d,p) = (1,7)$, *and (up to permutation) that* $|A_1| = |A_2| = 2$, $k_{\ell_1(1)} = k_{\ell_2(1)}$.

*Furthermore, if we do not assume that $h$ is supported in the set $\sum_{j=1}^{p}\zeta_j|k_j|^2 - |k|^2 = \Gamma$ as in (4.23), but instead assume*

$$|h_{k_1\cdots k_p}| \lesssim \frac{1}{\langle\Omega+\Gamma\rangle}, \quad \Omega := |k|^2 - \sum_{j=1}^{p}\zeta_j|k_j|^2, \tag{4.27}$$

*then the same result holds. Finally, all the above results remain true if we replace $p$ by any odd $3 \leq q \leq p$ (without changing the value of $\alpha_0$).*

*Proof.* First assume $d(p-1) \geq 8$. If there is some pairing between $(k, k_{P_0})$ and $k_{Q_0}$, then using the simple fact that

$$\|h_{k_1k_Ak_2k_B}\cdot \mathbf{1}_{k_1=k_2}\|_{k_1k_A\to k_2k_B} \leq \sup_k \|h_{kk_Akk_B}\|_{k_A\to k_B}$$

(this is proved in the same way as Lemma 4.10 below), we may fix the value $k$ of these paired variables with no summation (hence no cost of powers) and reduce to a problem involving a smaller set[1]. Therefore we may assume there is no pairing between $(k, k_{P_0})$ and $k_{Q_0}$. We may also assume

[1]Strictly speaking this reduction may not preserve (4.25), but (4.25) is only used to guarantee $|A_v| \leq p-2$ in order to apply Lemma 4.8; after this (4.25) can be replaced by the more general versions where the linear combinations of $k_j$ and $k$ are fixed $\mathbb{Z}^d$ vectors instead of 0, which is preserved under the reduction.

$|A_v| \leq p-2$ for each $v$, since otherwise $(k, k_1)$ will be a pairing due to (4.25), which has to be over-paired with an element in $A_v$, and after removing these over-paired variables[1], the remaining set will satisfy $|A_v| \leq p-2$.

At this point we are ready to apply Lemma 4.8. By Schur's Lemma, we have

$$\|h\|_{kk_{P_0}\to k_{Q_0}} \lesssim \Big(\sup_{k_{Q_0}} \sum_{k,k_{P_0}} 1\Big)^{1/2}\Big(\sup_{k,k_{P_0}} \sum_{k_{Q_0}} 1\Big)^{1/2},$$

so we just need to count $(k, k_{P_0})$ with $k_{Q_0}$ given, and also count $k_{Q_0}$ with $(k, k_{P_0})$ given. Now Lemma 4.8 implies that

$$\sup_{k_{Q_0}} \sum_{k,k_{P_0}} 1 \lesssim \prod_{j\in P_0} (M_j)^{2\alpha_0+\theta}, \tag{4.28}$$

$$\sup_{k,k_{P_0}} \sum_{k_{Q_0}} 1 \lesssim \prod_{j\in Q_0} (M_j)^{2\alpha_0+\theta} \cdot \prod_{v=1}^{s} (\min_{1\leq y<a_v} M_{\ell_v(y)})^{2\alpha_0-d}. \tag{4.29}$$

This is because, for example, once $(k, k_{P_0})$ is fixed, for any $n_y \in C \cap Q_0$ we have

$$\sum_{1\leq z\leq y;\, n_z\in Q_0} \zeta_{n_z} k_{n_z} = \sum_{z=1}^{y} \zeta_{n_z} k_{n_z} + (\text{constant vector}),$$

so the left hand side sum belongs to a ball of radius $M_{n_y}$. Using also (4.22) and (4.23), and noticing that any pairing in $k_{Q_0}$ must be over-paired, we can deduce (4.29) from Lemma 4.8, and similarly (4.28). Combining (4.28) and (4.29) then gives (4.26).

Now consider the exceptional cases $d(p-1) \leq 6$. If $p = 3$, then either there is no pairing at all and (4.26) follows from (4.28) and (4.29), which in turns follow from Lemma 4.7, or there is an over-pairing and each $k_j$ $(1 \leq j \leq 3)$ is uniquely fixed, in which case (4.26) is immediate.

In the remaining cases we must have $d = 1$, and $p \in \{5, 7\}$. Again we may assume there is an over-pairing (otherwise (4.26) follows from Lemma 4.7, which does require any condition on $(d, p)$); if $p = 5$, then an over-pairing takes at least 3 variables while there are 6 in total ($k$ and each $k_j$), so there are only 3 variables remaining. By using Schur's Lemma and the one-dimensional version of Lemma 4.5, one can show that in such cases we always have $\|h\|_{kk_{Q_0}\to k_{P_0}} \lesssim (M_2 \cdots M_5)^\theta$, which also implies (4.26). Finally if $p = 7$, then we have

$$2(2\alpha_0 + \theta) > (2\alpha_0 + \theta) + (4\alpha_0 - d + \theta) > d,$$

so we can apply the same arguments in the proof of Lemma 4.8 and get the same result, unless we are in the worst case, namely the first line of (4.20), which implies that (up to permutation) $|A_1| = |A_2| = 2$, and $k_{\ell_1(1)} = k_{\ell_2(1)}$.

Finally, we look at the case where (4.27) is assumed instead of the support condition (the result for $3 \leq q \leq p$ follows from the same arguments, which we will not repeat). Here again we can reduce to the case where Lemma 4.8 is applicable, and by Schur's Lemma and (4.27) we have

$$\|h\|_{kk_{P_0}\to k_{Q_0}} \lesssim \Big(\sup_{k_{Q_0}} \sum_{k,k_{P_0}} \frac{1}{\langle\Omega+\Gamma\rangle}\Big)^{1/2}\Big(\sup_{k,k_{P_0}} \sum_{k_{Q_0}} \frac{1}{\langle\Omega+\Gamma\rangle}\Big)^{1/2},$$

[1] These over-paired variables include a pairing between $(k, k_{P_0})$ and $k_{Q_0}$ as $1 \in Q_0$, and thus can be treated using the argument in the beginning of the proof.

so we just need to bound the sums on the right hand side. The idea is that, when $k_{Q_0}$ (or $(k, k_{P_0})$) is fixed, the number of choices for $(k, k_{P_0})$ (or $k_{Q_0}$) is at most $(M_2 \cdots M_p)^d$ due to (4.22)–(4.25) (without using the equality for $\Omega$), so by Hölder

$$\sup_{k_{Q_0}} \sum_{k,k_{P_0}} \frac{1}{\langle \Omega + \Gamma \rangle} \lesssim (M_2 \cdots M_p)^\theta \bigg( \sup_{k_{Q_0}} \sum_{k,k_{P_0}} \frac{1}{\langle \Omega + \Gamma \rangle^{1+\theta}} \bigg)^{1/(1+\theta)}.$$

Upon fixing the value of $\Omega + \Gamma$, the latter sum can be bounded by (4.28); similarly the sum in $k_{Q_0}$ can be bounded by (4.29) with a loss of $(M_2 \cdots M_p)^\theta$, which can always be incorporated into (4.26). This completes the proof. □

4.4. **Tensor norm estimates.** Here we prove the main estimates for tensor norms. Start with the following simple lemma.

**Lemma 4.10.** *Let $(B, C)$ be a subpartition of $A$, and let $E = A \backslash (B \cup C)$. Then the norm $\|h\|_{k_B \to k_C}$ increases by at most a constant multiple, under multiplication by:*

*(1) Any function $f(k_B, k_E)$, or any function $g(k_C, k_E)$, that is bounded;*

*(2) Any function of form $\varphi(L^{-1}[f(k_B, k_E) - g(k_C, k_E)])$, where $L > 0$ is a real number, $\varphi$ is defined on some $\mathbb{R}^m$ such that $\widehat{\varphi} \in L^1$, and $f, g$ are arbitrary $\mathbb{R}^m$-valued functions;*

*(3) Any function of form $\mathbf{1}_{k_i = k_j}$ or $\mathbf{1}_{k_i \neq k_j}$, regardless whether $i$ or $j$ belong to $B$, $C$ or $E$.*

*Proof.* (1) is obvious by definition, and (2) follows from (1) by writing

$$\varphi(L^{-1}[f(k_B, k_E) - g(k_C, k_E)]) = L^m \int_{\mathbb{R}^m} \widehat{\varphi}(L\xi) e^{i\xi \cdot f(k_B, k_E)} e^{-i\xi \cdot f(k_C, k_E)} \, \mathrm{d}\xi.$$

To prove (3), we may assume $i \in B$ and $j \in C$ (otherwise $i, j \in B$ or $i, j \in C$ or one of them belongs to $E$, and the proof will be easier), and also assume $E = \varnothing$. Let $k_{B \backslash \{i\}} = m$ and $k_{C \backslash \{j\}} = n$, and let $k_i = k_j = k$ after multiplying by $\mathbf{1}_{k_i = k_j}$, then it suffices to prove that

$$\sum_{k,n} \bigg| \sum_m h_{kmkn} \cdot z_{km} \bigg|^2 \leq \|h_{k_i m k_j n}\|^2_{k_i m \to k_j n} \cdot \sum_{k,m} |z_{km}|^2.$$

Clearly we may fix $k$, and consider the tensor $h_{kmkn}$, so the desired bound follows from the inequality

$$\sup_k \|h_{kmkn}\|_{m \to n} \leq \|h_{k_i m k_j n}\|_{k_i m \to k_j n},$$

which is obvious by definition. The result for $\mathbf{1}_{k_i \neq k_j}$ then follows, since $\mathbf{1}_{k_i \neq k_j} = 1 - \mathbf{1}_{k_i = k_j}$. □

Next we state and prove the bilinear semi-product estimate, equivalent to Proposition 2.5.

**Proposition 4.11** (Restatement of Proposition 2.5)**.** *Consider two tensors $h^{(1)}_{k_{A_1}}$ and $h^{(2)}_{k_{A_2}}$, where $A_1 \cap A_2 = C$. Let $A_1 \Delta A_2 = A$, define the semi-product*

$$H_{k_A} = \sum_{k_C} h^{(1)}_{k_{A_1}} h^{(2)}_{k_{A_2}}. \tag{4.30}$$

*Then, for any partition $(X, Y)$ of $A$, let $X \cap A_1 = X_1$, $Y \cap A_1 = Y_1$ etc., we have*

$$\|H\|_{k_X \to k_Y} \leq \|h^{(1)}\|_{k_{X_1 \cup C} \to k_{Y_1}} \cdot \|h^{(2)}\|_{k_{X_2} \to k_{C \cup Y_2}}. \tag{4.31}$$

*Proof.* Note that $X_1$, $X_2$, $Y_1$, $Y_2$ and $C$ are five pairwise disjoint sets; let the vectors $x := k_{X_1}$, $y := k_{Y_1}$, $z = k_C$, $u := k_{X_2}$ and $v := k_{Y_2}$, then we can write

$$h^{(1)} = h^{(1)}_{xyz}, \quad h^{(2)} = h^{(2)}_{uvz}; \quad H = H_{xyuv} = \sum_z h^{(1)}_{xyz} h^{(2)}_{uvz}.$$

Now for any $\alpha = \alpha_{xu}$, we can bound

$$\begin{aligned}\sum_{y,v}\Big|\sum_{x,u} H_{xyuv}\alpha_{xu}\Big|^2 &= \sum_{y,v}\Big|\sum_{x,u,z} h^{(1)}_{xyz}h^{(2)}_{uvz}\alpha_{xu}\Big|^2 = \sum_v\sum_y\Big|\sum_{x,z} h^{(1)}_{xyz}\Big(\sum_u h^{(2)}_{uvz}\alpha_{xu}\Big)\Big|^2\\ &\leq \|h^{(1)}\|^2_{xz\to y}\cdot\sum_{x,z,v}\Big|\sum_u h^{(2)}_{uvz}\alpha_{xu}\Big|^2 = \|h^{(1)}\|^2_{xz\to y}\cdot\sum_x\sum_{v,z}\Big|\sum_u h^{(2)}_{uvz}\alpha_{xu}\Big|^2\\ &\leq \|h^{(1)}\|^2_{xz\to y}\cdot\|h^{(2)}\|^2_{u\to vz}\cdot\sum_{x,u}|\alpha_{xu}|^2,\end{aligned}$$

so by definition, $\|H\|_{xu\to yv} \leq \|h^{(1)}\|_{xz\to y}\cdot\|h^{(2)}\|_{u\to vz}$, as desired. $\square$

A corollary is the following multilinear semi-product estimate, equivalent to Proposition 2.6.

**Proposition 4.12** (Restatement of Proposition 2.6)**.** *Let $A_j$ $(1 \leq j \leq m)$ be index sets, such that any index appears in at most two $A_j$'s, and let $h^{(j)} = h^{(j)}_{k_{A_j}}$ be tensors. Let $A = A_1\Delta\cdots\Delta A_m$ be the set of indices that belong to only one $A_j$, and $C = (A_1\cup\cdots\cup A_m)\backslash A$ be the set of indices that belong to two different $A_j$'s. Define the semi-product*

$$H_{k_A} = \sum_{k_C}\prod_{j=1}^m h^{(j)}_{k_{A_j}}. \tag{4.32}$$

*Let $(X, Y)$ be a partition of $A$. For $1 \leq j \leq m$ let $X_j = X\cap A_j$ and $Y_j = Y\cap A_j$, and define*

$$B_j := \bigcup_{\ell>j}(A_j\cap A_\ell), \quad C_j = \bigcup_{\ell<j}(A_j\cap A_\ell), \tag{4.33}$$

*then we have*

$$\|H\|_{k_X\to k_Y} \leq \prod_{j=1}^m \|h^{(j)}\|_{k_{X_j\cup B_j}\to k_{Y_j\cup C_j}}. \tag{4.34}$$

*Proof.* We induct in $m$. When $m = 2$, (4.34) is just (4.31); suppose (4.34) holds for $m-1$. Then, define $F = A_2\Delta\cdots\Delta A_m$, $E = (A_2\cup\cdots\cup A_m)\backslash F$ and

$$Y_{k_F} := \sum_{k_E}\prod_{j=2}^m h^{(j)}_{k_{A_j}},$$

then we have $A = A_1\Delta F$, and

$$H_{k_A} = \sum_{k_G} h^{(1)}_{k_{A_1}} Y_{k_F}, \quad G := A_1\cap F = \bigcup_{\ell>1}(A_1\cap A_\ell) = B_1.$$

Applying (4.31) we get

$$\|H\|_{k_X\to k_Y} \leq \|h^{(1)}\|_{k_{X_1\cup B_1}\to k_{Y_1}}\cdot\|Y\|_{k_{X\cap F}\to k_{(Y\cap F)\cup B_1}}.$$

Note that $X' := X \cap F$ and $Y' := (Y \cap F) \cup B_1$ form a partition of $F$, by induction hypothesis we have

$$\|Y\|_{k_{X'} \to k_{Y'}} \leq \prod_{j=2}^{m} \|h^{(j)}\|_{k_{(X' \cap A_j) \cup B_j} \to k_{(Y' \cap A_j) \cup (C_j \backslash B_1)}}.$$

Finally, note that $(X' \cap A_j) \cup B_j = X_j \cup B_j$ and $(Y' \cap A_j) \cup (C_j \backslash B_1) = Y_j \cup C_j$, this completes the inductive proof. □

*Remark* 4.13. Note that, if we fix $(X, Y)$ and rearrange the tensors $h^{(j)}$, then the expression (4.32) will not change, but the norms appearing on the right hand side of (4.34) will. We may take advantage of this and arrange these tensors in some order using a particular algorithm so that (4.34) gives the desired bound. This will be the key to the proof of Proposition 6.2 below.

Finally we state and prove the precise form of Proposition 2.8, and a similar variant. The proofs rely on higher order versions of Bourgain's $\mathscr{G}\mathscr{G}^*$ argument in [11].

**Proposition 4.14** (Precise form of Proposition 2.8)**.** *Let $A$ be a finite set and $h_{bck_A} = h_{bck_A}(\omega)$ be a tensor, where each $k_j \in \mathbb{Z}^d$ and $(b, c) \in (\mathbb{Z}^d)^q$ for some integer $q \geq 2$. Given signs $\zeta_j \in \{\pm\}$, we also assume that $\langle b \rangle, \langle c \rangle \lesssim M$ and $\langle k_j \rangle \lesssim M$ for all $j \in A$, where $M$ is a dyadic number, and that in the support of $h_{bck_A}$ there is no pairing in $k_A$. Define the tensor*

$$H_{bc} = \sum_{k_A} h_{bck_A} \prod_{j \in A} \eta_{k_j}^{\zeta_j}, \tag{4.35}$$

*where we restrict $k_j \in E$ in (4.35), $E$ being a finite set such that $\{h_{bck_A}\}$ is independent with $\{\eta_k : k \in E\}$. Then $\tau^{-1}M$-certainly, we have*

$$\|H_{bc}\|_{b \to c} \lesssim \tau^{-\theta} M^{\theta} \cdot \max_{(B,C)} \|h\|_{bk_B \to ck_C}, \tag{4.36}$$

*where $(B, C)$ runs over all partitions of $A$.*

*Proof.* By conditioning on $\{h_{bck_A}(\omega)\}$, we may assume $h_{bck_A}$ are constant tensors. View $H$ as a linear operator that maps functions of $c$ to functions of $b$, and consider the kernel of $(HH^*)^m$ for a large positive integer $m$.

Define $R_n = (HH^*)^m$ if $n = 2m$, and $R_n = (HH^*)^m H$ if $n = 2m + 1$. By induction in $n$, we will prove that the kernel of $R_n$ can be written as a linear combination of terms $\mathcal{R}_n$ which has the form

$$\begin{cases} (\mathcal{R}_n)_{bb'} = \displaystyle\sum_{k_Z} y_{bb'k_Z} \prod_{j \in Z} \eta_{k_j}^{\zeta_j}, & n \text{ even;} \\ (\mathcal{R}_n)_{bc} = \displaystyle\sum_{k_Z} y_{bck_Z} \prod_{j \in Z} \eta_{k_j}^{\zeta_j}, & n \text{ odd,} \end{cases} \tag{4.37}$$

where $Z$ is a finite set, $\zeta_j \in \{\pm\}$, $y_{bb'k_Z}$ (or $y_{bck_Z}$) is a tensor such that in its support, there is no pairing in $k_Z$, and satisfies the bound

$$\|y\|_{bb'k_Z} \text{ (or } \|y\|_{bck_Z}) \lesssim \big( \sup_{(B,C)} \|h\|_{bk_B \to ck_C} \big)^{n-1} \|h\|_{bck_A}. \tag{4.38}$$

In fact, when $n = 1$ this is obvious (with $Z = A$). Suppose (4.37) and (4.38) are true for $n - 1$, where $n$ is odd, then since $R_n = R_{n-1}H$ it suffices to consider the kernel (note that by relabeling we may assume $Z \cap A = \varnothing$)

$$(\mathcal{R}_n)_{bc} = \sum_{b'} (\mathcal{R}_{n-1})_{bb'} H_{b'c} = \sum_{b'} \sum_{k_Z, k_A} y_{bb'k_Z} h_{b'ck_A} \prod_{j \in Z} \eta_{k_j}^{\zeta_j} \prod_{j \in A} \eta_{k_j}^{\zeta_j}. \tag{4.39}$$

Now, by repeating the arguments in the proof of Proposition 3.7, we can write (4.39) as a linear combination of sums (for different choices of $\mathscr{O}$), which have the same summand as (4.39) and are taken over the same set of variables ($b'$ and $k_{Z\cup A}$), but with a set of additional pairing and no-pairing conditions for the variables $k_{Z\cup A}$ given by $(\mathscr{O},1)$. More precisely, here $\mathscr{O}=\{\mathcal{A}_1,\cdots,\mathcal{A}_m\}$ where $\mathcal{A}_i$ are pairwise disjoint subsets of $Z\cup A$ such that each subset contains two elements of $Z\cup A$ with opposite $\zeta_j$ sign, but does not contain two elements of $Z$ or two elements of $A$ with opposite $\zeta_j$ sign, and the set of conditions $(\mathscr{O},1)$ is defined by (i) the $k_j$ are the same for $j$ in each $\mathcal{A}_i$, and (ii) there is no pairing in $k_{(Z\cup A)\backslash\mathcal{Q}}$ where $\mathcal{Q}=\mathcal{Q}(\mathcal{A}_1)\cup\cdots\mathcal{Q}(\mathcal{A}_m)$ (see (3.10)), as in the proof of Proposition 3.7.

Since $|\eta_j|^2\equiv 1$, we may recast the sum corresponding to $\mathscr{P}$ defined above as

$$\mathcal{R}_{bc}=\sum_{k_Y}w_{bck_Y}\prod_{j\in Y}\eta_{k_j}^{\zeta_j},\quad w_{bck_Y}=\prod_{(j,j')}^{(1)}\mathbf{1}_{k_j\neq k_{j'}}\cdot\sum_{b'}\sum_{k_{\mathcal{Q}}}^{(2)}\widetilde{y}_{bb'k_Z}\widetilde{h}_{b'ck_A}. \tag{4.40}$$

Here $Y=(Z\cup A)\backslash\mathcal{Q}$, the product $\prod_{(j,j')}^{(1)}$ is taken over all $j,j'\in Y$ such that $\zeta_{j'}=-\zeta_j$, the sum $\sum_{k_{\mathcal{Q}}}^{(2)}$ is taken over the variables $k_{\mathcal{Q}}$ such that $k_j=k_{j'}$ whenever $\{j,j'\}$ is one of the opposite-sign 2-element subsets (pairings) selected when obtaining $\mathcal{Q}$ as in Definition 3.6, and

$$\widetilde{y}_{bb'k_Z}:=y_{bb'k_Z}\cdot\prod_{(j,j')}^{(3)}\mathbf{1}_{k_j=k_{j'}},\quad \widetilde{h}_{b'ck_A}:=h_{b'ck_A}\cdot\prod_{(j,j')}^{(3)}\mathbf{1}_{k_j=k_{j'}}, \tag{4.41}$$

where the products $\prod_{(j,j')}^{(3)}$ are taken over all $j,j'\in Z$ (for $y$, or $j,j'\in A$ for $h$) such that they belong to the same $\mathcal{A}_i$. We shall apply Proposition 4.11 to estimate $\|w_{bck_Y}\|_{bck_Y}$; in order to do so we need to make an adjustment in notations. Namely, for any pairing $\{j,j'\}$, as we always require $k_j=k_{j'}$ in the sum $\sum_{k_{\mathcal{Q}}}^{(2)}$, we may combine them into a single element and include this element in both $Z$ and $A$. In this way we are changing pairings between $Z$ and $A$ to intersections of $Z$ and $A$, which is the setting of Proposition 4.11.

With (4.40), (4.41) and these adjustments, by Lemma 4.10 and Proposition 4.11, we conclude that

$$\|w_{bck_Y}\|_{bck_Y}\lesssim\|y_{bb'k_Z}\|_{bb'k_Z}\cdot\|h_{b'ck_A}\|_{ck_C\to b'k_B},$$

where $B=\mathcal{Q}\cap A$, and $C=A\backslash\mathcal{Q}$. This completes the inductive proof of (4.37) and (4.38) when $n$ is odd. When $n$ is even noticing that $R_n=R_{n-1}H^*$, we have

$$(\mathcal{R}_n)_{bb'}=\sum_c(\mathcal{R}_{n-1})_{bc}\overline{H_{b'c}}=\sum_c\sum_{k_Z,k_A}y_{bck_Z}\overline{h_{b'ck_A}}\prod_{j\in Z}\eta_{k_j}^{\zeta_j}\prod_{j\in A}\eta_{k_j}^{-\zeta_j} \tag{4.42}$$

instead of (4.39), and the rest of proof goes analogously.

Now consider the product $(HH^*)^m$ with $n=2m$. Using (4.38), Lemma 4.4 and noticing that the number of choices for $(b,b')$ is at most $M^{O(1)}$, we conclude that $\tau^{-1}M$-certainly, we have

$$\begin{aligned}\|H_{bc}\|_{b\to c}^{4m}=\|(HH^*)^m\|_{\mathrm{OP}}^2\lesssim\sum_{b,b'}|(\mathcal{R}_n)_{bb'}|^2\lesssim(\tau^{-1}M)^\theta\|y\|_{bb'k_Z}^2\\ \lesssim(\tau^{-1}M)^\theta\big(\sup_{(B,C)}\|h\|_{bk_B\to ck_C}\big)^{4m-2}\|h\|_{bck_A}^2,\end{aligned}$$

and hence

$$\|H_{bc}\|_{b\to c}\lesssim(\tau^{-1}M)^\theta\big(\sup_{(B,C)}\|h\|_{bk_B\to ck_C}\big)^{1-\frac{1}{2m}}\|h\|_{bck_A}^{\frac{1}{2m}}.$$

Fixing $m$ large enough, and noticing that by the support condition,

$$\|h\|_{bck_A} \lesssim M^{\frac{q+|A|}{2}} \sup_{b,c,k_A} |h_{bck_A}| \lesssim M^{\frac{q+|A|}{2}} \sup_{(B,C)} \|h\|_{bk_B \to ck_C},$$

we deduce (4.36). □

**Proposition 4.15** (A variant of Proposition 4.14)**.** *Consider the same setting as in Proposition 4.14, with the following differences: (1) we only restrict $\langle k_j \rangle \lesssim M$ but do not impose any condition on $\langle b \rangle$ or $\langle c \rangle$; (2) we assume $b, c \in \mathbb{Z}^d$ also, and assume that in the support of $h_{bck_A}$ we have $|b - \zeta c| \lesssim M$ where $\zeta \in \{\pm\}$; (3) the tensor $h_{bck_A}$ only depends on $b - \zeta c$, $|b|^2 - \zeta|c|^2$ and $k_A$, and is supported in the set where $||b|^2 - \zeta|c|^2| \leq M^{\kappa^3}$. The other assumptions are the same. Then $\tau^{-1}M$-certainly we have*

$$\|H_{bc}\|_{b\to c} \lesssim \tau^{-\theta} M^{\theta} \cdot \sup_{(B,C)} \|h\|_{bk_B \to ck_C}.$$

*Proof.* We may assume $\zeta = +$, since the other case is much easier. Since $h_{bck_A}$ is supported in the set where $|b - c| \lesssim M$, by an orthogonality argument we may modify $h$ by restricting it to the set where $|b - f| \lesssim M$ and $|c - f| \lesssim M$, and to bound $\|H_{bc}\|_{b\to c}$ it suffices to bound these restricted operators *uniformly* in $f \in \mathbb{Z}^d$.

For any $f$, let $x = b - f$ and $y = c - f$, then $x$ and $y$ are both assumed to have size $\lesssim M$, and it suffices to estimate the norm

$$\|\widetilde{H}_{f;xy}\|_{x\to y}, \quad \text{where} \quad \widetilde{H}_{f;xy} = \sum_{k_A} \widetilde{h}_{f;xyk_A} \prod_{j\in A} \eta_{k_j}^{\zeta_j}, \quad \text{and} \quad \widetilde{h}_{f;xyk_A} := h_{x+f,y+f,k_A} \cdot \mathbf{1}_{|x|,|y|\lesssim M}.$$

For any *fixed* value of $f$, we may apply Proposition 4.14 to conclude that $\tau^{-1}M$-certainly we have

$$\|\widetilde{H}_{f;xy}\|_{x\to y} \lesssim \tau^{-\theta} M^{\theta} \cdot \sup_{(B,C)} \|\widetilde{h}_{f;xyk_A}\|_{xk_B \to yk_C} \leq \tau^{-\theta} M^{\theta} \cdot \sup_{(B,C)} \|h\|_{bk_B \to ck_C}, \tag{4.43}$$

so it suffices to establish (4.43) *uniformly* in $f$. Note that by assumption, $\widetilde{h}_{f;xyk_A}$ is in fact a function of $(x, y, k_A)$ and $|x+f|^2 - |y+f|^2 = 2f\cdot(x-y) + (|x|^2 - |y|^2)$. The desired uniform bound (4.43), and hence the proof of Proposition 4.15, will follow from the following statement:

**Claim 4.16.** *There exist finitely many integer-valued functions $g_j(z)$ (defined on a subset $E_j \subset \{z : \langle z \rangle \lesssim M\}$), where $1 \leq j \leq K \leq M^{\kappa^4}$, such that for* any *integer vector $f \in \mathbb{Z}^d$, there exists $1 \leq j \leq A$, such that for any $z$ satisfying $\langle z \rangle \lesssim M$, we have $|f \cdot z| \leq M^{2\kappa^3}$ if and only if $z \in E_j$, and in such case we have $f \cdot z = g_j(z)$.*

*Proof of claim 4.16.* The proof is a slight modification of (a special case of) the proof of [33], Claim 3.7, so we refer the reader to that paper. □

## 5. The random tensor ansatz

In this section we begin the main proof. We make several reductions to the equation (1.7) in Section 5.1, write down the central random tensor ansatz in Section 5.2, and state the key a priori estimates, Proposition 5.1, in Section 5.3.

5.1. **First reductions.** We start by analyzing (1.7). The first step is to reduce it to a more suitable form. This is done by using a gauge transform, conditioning on the norms of Fourier modes of (1.2), and conjugating by the linear Schrödinger flow.

5.1.1. *The gauge transform.* Define the *gauge transform*

$$\widetilde{u_N}(t) = u_N(t) \cdot \exp\left(\frac{(p+1)i}{2}\int_0^t \fint_{\mathbb{T}^d} W_N^{p-1}(u_N)\,\mathrm{d}t'\right), \tag{5.1}$$

which has inverse

$$u_N(t) = \widetilde{u_N}(t) \cdot \exp\left(\frac{-(p+1)i}{2}\int_0^t \fint_{\mathbb{T}^d} W_N^{p-1}(\widetilde{u_N})\,\mathrm{d}t'\right), \tag{5.2}$$

then $u_N$ satisfies (1.7) if and only if $\widetilde{u_N}$ satisfies the gauged equation

$$\begin{cases} (i\partial_t+\Delta)\widetilde{u_N} = \Pi_N\left(W_N^p(\widetilde{u_N}) - \dfrac{p+1}{2}\fint_{\mathbb{T}^d} W_N^{p-1}(\widetilde{u_N})\cdot\widetilde{u_N}\right), \\ \widetilde{u_N}(0) = \Pi_N f(\omega). \end{cases} \tag{5.3}$$

The nonlinearity in the big parenthesis of (5.3) can be recast in the following form:

$$W_N^p(\widetilde{u_N}) - \frac{p+1}{2}\fint_{\mathbb{T}^d} W_N^{p-1}(\widetilde{u_N})\cdot\widetilde{u_N} = \sum_{3\leq q\leq p} a_{pq}(m_N-\sigma_N)^{(p-q)/2}\mathcal{N}_q(\widetilde{u_N}), \tag{5.4}$$

where $q$ runs over odd integers, $a_{pq}$ are constants, $m_N$ denotes the conserved mass of $\widetilde{u_N}$ (and $u_N$),

$$m_N = \fint_{\mathbb{T}^d} |\widetilde{u_N}|^2 = \sum_{\langle k\rangle\leq N}\frac{|g_k|^2}{\langle k\rangle^{2\alpha}},$$

$\sigma_N$ is as in (1.5), and $\mathcal{N}_q$ is a degree $q$ real polynomial (regarded also as a $\mathbb{R}$-multilinear expression of degree $q$) that is simple in the sense of Definition 3.1. For the derivation of (5.4), see [36], Proposition 2.2.

5.1.2. *Conditioning and conjugating.* Note that each $m_N$ is a function of the norms $\rho_k = |g_k|$, moreover let $m_N^* := m_N - \sigma_N$, then by standard large deviation estimates,

$$\left|m_N^* - m_{\frac{N}{2}}^*\right|^2 = \left|\sum_{N/2<\langle k\rangle\leq N}\frac{\rho_k^2-1}{\langle k\rangle^{2\alpha}}\right|^2 \leq \tau^{-\theta}N^{\theta}\sum_{N/2<\langle k\rangle\leq N}\frac{1}{\langle k\rangle^{4\alpha}} \leq \tau^{-\theta}N^{d-4\alpha+\theta} \leq \tau^{-\theta}N^{-40\varepsilon} \tag{5.5}$$

holds $\tau^{-1}N$-certainly, as $4\alpha - d > 80\varepsilon$ by our assumptions.

Now, by excluding a set of probability $\leq C_\theta e^{-\tau^{-\theta}}$ and conditioning on $\{\rho_k\}$, we may fix the values of $\rho_k$ and hence $m_N^*$, so that $\widetilde{u_N}$ solves the equation

$$\begin{cases} (i\partial_t+\Delta)\widetilde{u_N} = \displaystyle\sum_{3\leq q\leq p} a_{pq}(m_N^*)^{(p-q)/2}\cdot\Pi_N\mathcal{N}_q(\widetilde{u_N}), \\ \widetilde{u_N}(0) = \displaystyle\sum_{k\in\mathbb{Z}^d}\Pi_N\gamma_k\cdot\eta_k(\omega)e^{ik\cdot x}, \end{cases} \tag{5.6}$$

where recall $\gamma_k = \langle k\rangle^{-\alpha}\rho_k$; they and the $m_N^*$ are constants that satisfy

$$|\gamma_k| \leq \tau^{-\theta}\langle k\rangle^{-\alpha+\theta}, \quad |m_N^*| \leq \tau^{-\theta}, \quad \left|m_N^* - m_{\frac{N}{2}}^*\right| \leq \tau^{-\theta}N^{-40\varepsilon}. \tag{5.7}$$

We may also assume, due to (5.5), that

$$\sum_{N/2<\langle k\rangle\leq N}\gamma_k^2 \leq \tau^{-\theta}N^{d-2\alpha}. \tag{5.8}$$

Finally, define $v_N$ by $(v_N)_k(t) := e^{it|k|^2}(\widetilde{u_N})_k(t)$, then $v_N$ satisfies the following equation

$$(v_N)_k(t) = (F_N)_k - i \sum_{3\leq q\leq p} a_{pq}(m_N^*)^{(p-q)/2} \int_0^t \Pi_N \mathcal{M}_q(v_N, \cdots, v_N)_k(t')\, \mathrm{d}t', \tag{5.9}$$

where the initial data $F_N$ is defined by

$$(F_N)_k := \Pi_N \gamma_k \cdot \eta_k(\omega) = \sum_{N'\leq N} (f_{N'})_k, \quad \text{where} \quad (f_N)_k := \Delta_N \gamma_k \cdot \eta_k(\omega) \tag{5.10}$$

(this $f_N$ will be the one appearing in Definitions 3.4 and 3.5 and in particular the `Trim` functions), $\mathcal{M}_q$ is a conjugated version of $\mathcal{N}_q$, and is given by

$$\mathcal{M}_q(v^{(1)}, \cdots, v^{(q)})_k(t') = \sum_{\zeta_1 k_1+\cdots+\zeta_q k_q = k} c_{kk_1\cdots k_q} \cdot e^{it'\Omega} \prod_{j=1}^{q} (v^{(j)})_{k_j}^{\zeta_j}(t'). \tag{5.11}$$

In (5.11), the signs $(\zeta_1, \cdots, \zeta_q) = (+, -, \cdots, +)$, and the coefficients $c_{kk_1\cdots k_q}$ satisfy the simplicity condition in the sense of Definition 3.1. Finally $\Omega$ is defined by

$$\Omega = |k|^2 - |k_1|^2 + \cdots - |k_q|^2 = |k|^2 - \sum_{j=1}^{q} \zeta_j |k_j|^2. \tag{5.12}$$

Below we will focus on the system (5.9)–(5.11) on $J = [-\tau, \tau]$, with the parameters satisfying (5.7)–(5.8). Using (3.6), (5.9) can be rewritten as

$$(v_N)_k(t) = (F_N)_k(t) - i \sum_{3\leq q\leq p} a_{pq}(m_N^*)^{(p-q)/2} \cdot \mathcal{I}\Pi_N \mathcal{M}_q(v_N, \cdots, v_N)_k(t), \tag{5.13}$$

where $\mathcal{I}$ is as in (3.6). In order to use the global-in-time norms defined in Section 3.3, we need to construct functions $v_N^\dagger$ that are well-controlled for all time, and agree with $v_N$ on $J$. The strategy is to construct $v_N^\dagger$ by the time truncated system

$$(v_N^\dagger)_k(t) = \chi(t) \cdot (F_N)_k - i \sum_{3\leq q\leq p} a_{pq}(m_N^*)^{(p-q)/2} \chi_\tau(t) \cdot \mathcal{I}_\chi \Pi_N \mathcal{M}_q(v_N, \cdots, v_N)_k(t), \tag{5.14}$$

where $\mathcal{I}_\chi$ is as in (3.6). Clearly if $v_N^\dagger$ solves (5.14) then they must agree with the solution $v_N$ to (5.13) on $J$. Unlike $v_N$, which always solve (5.13), the $v_N^\dagger$ we construct are solutions to (5.14) only $\tau^{-1}$-certainly, i.e. apart from a set of probability $\leq C_\theta e^{-\tau^{-\theta}}$.

5.2. **Construction of tensors.** In this section we present the random tensor ansatz. The reader may recall that the core idea of this ansatz was presented for a simpler model in Section 2.4. Suppose $v_N^\dagger$ solves (5.14), and let $y_N$ be defined by

$$y_N = v_N^\dagger - v_{\frac{N}{2}}^\dagger; \quad v_N^\dagger = \sum_{N'\leq N} y_{N'}, \tag{5.15}$$

then $y_N$ solves the system

$$\begin{aligned}(y_N)_k(t) = \chi(t) \cdot (f_N)_k - i &\sum_{3\leq q\leq p} a_{pq}(m_N^*)^{(p-q)/2} \chi_\tau(t) \cdot \mathcal{I}_\chi \Delta_N \mathcal{M}_q(v_{\frac{N}{2}}^\dagger, \cdots v_{\frac{N}{2}}^\dagger)_k(t) \\ - i &\sum_{3\leq q\leq p} a_{pq}(m_N^*)^{(p-q)/2} \chi_\tau(t) \cdot \mathcal{I}_\chi \Pi_N \big[\mathcal{M}_q(v_N^\dagger, \cdots, v_N^\dagger) - \mathcal{M}_q(v_{\frac{N}{2}}^\dagger, \cdots v_{\frac{N}{2}}^\dagger)\big]_k(t) \\ - i &\sum_{3\leq q\leq p} a_{pq}\big[(m_N^*)^{(p-q)/2} - (m_{\frac{N}{2}}^*)^{(p-q)/2}\big] \chi_\tau(t) \cdot \mathcal{I}_\chi \Pi_{\frac{N}{2}} \mathcal{M}_q(v_{\frac{N}{2}^\dagger}, \cdots v_{\frac{N}{2}}^\dagger)_k(t),\end{aligned} \tag{5.16}$$

where $\mathcal{M}_q$ is as in (5.11). Conversely if $y_N$ solves (5.16), then $v_N^\dagger$ solves (5.14) where we understand that $v_{1/2}^\dagger=0$. So it suffices to construct $y_N$.

We shall construct $y_N$ by an ansatz which involves $\mathcal{S}$-tensors $h^{(\mathcal{S},n)}=h_{kk_{\mathcal{U}}}^{(\mathcal{S},n)}(t,k_{\mathcal{V}},\lambda_{\mathcal{V}})$ for $n\in\{0,1\}$ and *regular* plants $\mathcal{S}$ having frequency $N(\mathcal{S})=N$ and size $|\mathcal{S}|\leq D$, as well as a remainder term $z_N$. Here $D$ is as in (3.2). This construction will be inductive, first in $N$ and then in $|\mathcal{S}|$ with fixed $N$. As the base case we understand that all these quantities are 0 when (formally) $N$ is $1/2$.

*Step 1: the induction hypothesis.* Now, given dyadic $M\geq 1$, assume that we have already defined the $\mathcal{S}$-tensors $h^{(\mathcal{S},n)}$ for all $n\in\{0,1\}$, all regular plants $\mathcal{S}$ with $N(\mathcal{S})<M$ and $|\mathcal{S}|\leq D$, as well as $z_{N'}=(z_{N'})_k(t)$ for $N'<M$. For $N<M$, define

$$\begin{aligned}(y_N)_k(t)&=\sum_{n\in\{0,1\}}\sum_{\substack{\mathcal{S}:N(\mathcal{S})=N\\|\mathcal{S}|\leq D}}\sum_{k_{\mathcal{U}},k_{\mathcal{V}}}\int \mathrm{d}\lambda_{\mathcal{V}}h_{kk_{\mathcal{U}}}^{(\mathcal{S},n)}(t,k_{\mathcal{V}},\lambda_{\mathcal{V}})\\&\qquad\times\prod_{\mathfrak{l}\in\mathcal{U}}(f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}}\prod_{\mathfrak{f}\in\mathcal{V}}(\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}}(\lambda_{\mathfrak{f}})+(z_N)_k(t),\\(v_N^\dagger)_k(t)&=\sum_{N'\leq N}(y_{N'})_k(t).\end{aligned}\tag{5.17}$$

Note that the first equation in (5.17) can also be written as

$$(y_N)_k(t)=\sum_{n\in\{0,1\}}\sum_{\substack{\mathcal{S}:N(\mathcal{S})=N\\|\mathcal{S}|\leq D}}\Psi_k[\mathcal{S},h^{(\mathcal{S},n)}(t)]+(z_N)_k(t)\tag{5.18}$$

in view of (3.8). Here and throughout the proof, the $f_{N'}$ in (3.8) will be fixed as in (5.10), and $(z_{N'})_{N'<M}$ will be fixed as above. Moreover, define the $\mathbb{R}$-linear operator (which plays the role of $\mathcal{P}_{NL}$ in Section 2.4.1)

$$(\mathscr{L}^Mw)_k(t)=-i\sum_{3\leq q\leq p}a_{pq}(m_M^*)^{(p-q)/2}\chi_\tau(t)\cdot\mathcal{I}_\chi\Pi_M\sum_{\mathrm{sym}}\mathcal{M}_q(w,v_{M^{[\delta]}}^\dagger,\cdots,v_{M^{[\delta]}}^\dagger)_k(t),\tag{5.19}$$

where $\sum_{\mathrm{sym}}$ represents the sum over all possible permutations, for example

$$\sum_{\mathrm{sym}}\mathcal{M}(w,v,v):=\mathcal{M}(w,v,v)+\mathcal{M}(v,w,v)+\mathcal{M}(v,v,w).$$

Let the components $\mathscr{L}^{M,\zeta}$, as well as the kernels $(\mathscr{L}^{M,\zeta})_{kk'}(t,t')$, be defined as in Section 3.1. Let the $\mathbb{R}$-linear operator $\mathscr{R}^M=(1-\mathscr{L}^M)^{-1}$, which is bounded from $X^{b_0}$ to itself, if $\|\mathscr{L}^M\|_{X^{b_0}\to X^{b_0}}<1/2$ (with $b_0$ in (3.3)); otherwise let $\mathscr{R}^M=\mathrm{Id}$. Define also $\mathscr{V}^M=\mathscr{R}^M-1$. The goal is to define the $\mathcal{S}$-tensors $h^{(\mathcal{S},n)}$ for $n\in\{0,1\}$ and regular plants $\mathcal{S}$ having $N(\mathcal{S})=M$ and $|\mathcal{S}|\leq D$, and the remainder term $z_M$, such that $y_M$ defined by (5.17) with $N$ replaced by $M$ solves (5.16) with high probability.

*Step 2: paralinearization.* If we assume $\mathscr{R}^M=(1-\mathscr{L}^M)^{-1}$, then using the operator $\mathscr{L}^M$, we can paralinearize (5.16) and rewrite it as

$$(y_M)_k(t)=\chi(t)\cdot(f_M)_k+(\mathscr{L}^My_M)_k(t)+\sum_{3\leq q\leq p}\sum_{N_1,\cdots,N_q\leq M}\Upsilon\cdot\chi_\tau(t)\big[\mathcal{I}_\chi\Pi\mathcal{M}_q(y_{N_1},\cdots y_{N_q})\big]_k(t).\tag{5.20}$$

In the above summation $q$ is odd, $\Pi$ is one of the projections $\Pi_M$, $\Delta_M$ or $\Pi_{\frac{M}{2}}$, and we require that if $N_j=M$ for some $j$, then there must be another $j'\neq j$ such that $N_{j'}\geq M^\delta$ (otherwise the term is contained in the second term $(\mathscr{L}^M y_M)_k(t))$. The coefficient $\Upsilon$ depends only on $q$ and the $N_j$'s, and satisfies $|\Upsilon| \leq \tau^{-\theta}$; moreover if $N_j \leq (50dp)^{-1}M$ for each $1 \leq j \leq q$, we have the stronger bound $|\Upsilon| \leq \tau^{-\theta} M^{-40dp\varepsilon}$.

Using the operators $\mathscr{R}^M$, $\mathscr{V}^M$ and their kernels defined by (3.4)–(3.5), we can solve (5.20) by

$$(y_M)_k(t) = \chi(t) \cdot (f_M)_k + \sum_{\zeta \in \{\pm\}} \sum_{k'} \int \mathrm{d}t' \cdot (\mathscr{V}^{M,\zeta})_{kk'}(t,t')\chi(t') \cdot (f_M)_{k'}^{\zeta}$$
$$+ \sum_{\zeta \in \{\pm\}} \sum_{k'} \int \mathrm{d}t' \cdot (\mathscr{R}^{M,\zeta})_{kk'}(t,t') \sum_{3 \leq q \leq p} \sum_{N_1,\cdots,N_q \leq M} \Upsilon \cdot \chi_\tau(t') \big[\mathcal{I}_\chi \Pi \mathcal{M}_q(y_{N_1}, \cdots y_{N_q})\big]_{k'}^{\zeta}(t'). \tag{5.21}$$

The strategy is to construct the tensors $h^{(\mathcal{S},n)}$ with $N(\mathcal{S}) = M$ inductively in $|\mathcal{S}|$, such that when we plug (5.17) into (5.21) allowing $N = M$, the terms on the left and right sides cancel to sufficiently high order so that the remainders can be put in $z_M$.

*Step 3: definition of $h^{(\mathcal{S},n)}$.* Expanding the right hand side of (5.21) using (5.17) and allowing $N = M$, we obtain a sum of terms of the form (omitting $\mathscr{R}^{M,\zeta}$ and other factors for the moment)

$$\sum_{\mathrm{sym}} \mathcal{M}_q(\Psi_{k_1}^{(\mathcal{S}_1,n_1)}, \cdots, \Psi_{k_r}^{(\mathcal{S}_r,n_r)}, z_{N_{r+1}}, \cdots, z_{N_q})_{k'}(t'), \tag{5.22}$$

where $\Psi_{k_j}^{(\mathcal{S}_j,n_j)} = \Psi_{k_j}^{(\mathcal{S}_j,n_j)}(t') = \Psi_{k_j}[\mathcal{S}_j, h^{(\mathcal{S}_j,n_j)}(t')]$.

Let $\mathscr{B} = (M, q, r, \zeta_1, \cdots, \zeta_q, N_1, \cdots, N_q)$, note that $\sum_{j=1}^q \zeta_j = 1$. By Proposition 3.7 (1) and (2), if $N_j \leq M/2$ for each $r+1 \leq j \leq q$, and each $z_{N_j}$ in (5.22) is replaced by its low-modulation cutoff[1] $z_{N_j}^{\mathrm{lo}}$ defined by

$$(\widehat{z_{N_j}^{\mathrm{lo}}})_{k_j}(\lambda_j) = (\widehat{z_{N_j}})_{k_j}(\lambda_j) \cdot \chi(M^{-\kappa^2}\lambda_j)$$

(see Section 3.1 for the definition of $\chi$), then (5.22) can be recast as a linear combination of $\widetilde{\Psi}_k^{(\mathcal{S})} = \widetilde{\Psi}_k^{(\mathcal{S})}(t') = \Psi_k[\mathcal{S}, H]$ (for different choices of $\mathscr{O}$ as in Definition 3.6), where

$$\mathcal{S} = \mathtt{Trim}(\mathtt{Merge}(\mathtt{Trim}(\mathcal{S}_1, M^\delta), \cdots, \mathtt{Trim}(\mathcal{S}_r, M^\delta), \mathscr{B}, \mathscr{O}), M^\delta) \tag{5.23}$$
$$H = \mathtt{Trim}(\mathtt{Merge}(\mathtt{Trim}(h^{(\mathcal{S}_1,n_1)}, M^\delta), \cdots, \mathtt{Trim}(h^{(\mathcal{S}_r,n_r)}, M^\delta), h, \mathscr{B}, \mathscr{O}), M^\delta). \tag{5.24}$$

In (5.24) the tensor $h = h_{kk_1\cdots k_q}(t', \lambda_{r+1}, \cdots, \lambda_q)$ is given by

$$h_{kk_1\cdots k_q}(t', \lambda_{r+1}, \cdots, \lambda_q) = \mathbf{1}_{k=\zeta_1 k_1 + \cdots + \zeta_q k_q} \cdot \mathbf{1}_{\langle k \rangle \leq M} \prod_{j=1}^q \mathbf{1}_{\langle k_j \rangle \leq N_j}$$
$$\times \prod_{j=r+1}^q \chi(M^{-\kappa^2}\lambda_j) c_{kk_1\cdots k_q} e^{it'(\Omega + \zeta_{r+1}\lambda_{r+1} + \cdots + \zeta_q \lambda_q)} \tag{5.25}$$

with $c_{kk_1\cdots k_q}$ as in (5.11) and $\Omega$ as in (5.12). Here and throughout the proof, when applying $\mathtt{Trim}$ functions, we always fix $f_{N'}$ as in (5.10), and $(z_{N'})_{N'<M}$ as in the beginning of *Step 1*.

We will define two types of tensors, $h^{(\mathcal{S},0)}$ and $h^{(\mathcal{S},1)}$. The former are constant tensors without randomness, while the latter may depend on low-frequency random variables, which are however independent with the inputs of the tensor. Define $h^{(\mathcal{S},n)}$ inductively in $|\mathcal{S}|$, by the equations

$$h_{kk_{\mathcal{U}}}^{(\mathcal{S},0)}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}}) = \chi(t)\mathbf{1}_{\mathcal{S}=\mathcal{S}_M^+} \cdot \mathbf{1}_{k=k_{\mathfrak{l}}} \mathbf{1}_{M/2<\langle k \rangle \leq M} + \sum_{\mathrm{sym}} \sum_{(a)} \Upsilon \cdot \chi_\tau(t) \big[\mathcal{I}_\chi \Pi H_{kk_{\mathcal{U}}}\big](t, k_{\mathcal{V}}, \lambda_{\mathcal{V}}), \tag{5.26}$$

[1]This matches Definition 3.4.

$$
\begin{aligned}
h_{kk_{\mathcal{U}}}^{(\mathcal{S},1)}(t,k_{\mathcal{V}},\lambda_{\mathcal{V}}) = &\sum_{\zeta\in\{\pm\}} \mathbf{1}_{\mathcal{S}=\mathcal{S}_M^{\zeta}} \int \mathrm{d}t' \cdot \mathbf{1}_{M/2<\langle k_{\mathfrak{l}}\rangle\leq M} \cdot \mathscr{V}_{kk_{\mathfrak{l}}}^{M,\zeta}(t,t')\chi(t') \\
&+ \sum_{\mathrm{sym}}\sum_{(b)} \Upsilon \cdot \chi_\tau(t) \big[\mathcal{I}_\chi \Pi H_{kk_{\mathcal{U}}}\big](t,k_{\mathcal{V}},\lambda_{\mathcal{V}}) \\
&+ \sum_{\zeta\in\{\pm\}}\sum_{k'} \int \mathrm{d}t' \cdot (\mathscr{V}^{M,\zeta})_{kk'}(t,t') \sum_{\mathrm{sym}}\sum_{(c[\zeta])} \Upsilon \cdot \chi_\tau(t') \big[\mathcal{I}_\chi \Pi H_{k'k_{\mathcal{U}}}\big](t',k_{\mathcal{V}},\lambda_{\mathcal{V}})^{\zeta}.
\end{aligned}
\tag{5.27}
$$

Here $\mathcal{S}_M^{\zeta}$ are the mini plants defined in Definition 3.2. The summation $\sum_{(c[\zeta])}$ is taken over $\mathscr{B}$, $n_j \in \{0,1\}$ and regular plants $\mathcal{S}_j$ with frequency $N_j \leq M$ and size $|\mathcal{S}_j| \leq D$ for $1 \leq j \leq r$, and[1] $\mathscr{O}$, such that

(i) if $N_j = M$ for some $1 \leq j \leq q$ then there is $q \geq j' \neq j$ with $N_{j'} \geq M^{\delta}$;
(ii) $N_j \leq M/2$ for $r+1 \leq j \leq q$;
(iii) if $\zeta = +$ then (5.23) is true with the given $\mathcal{S}$, and if $\zeta = -$ then (5.23) is true with the left hand side replaced by $\overline{\mathcal{S}}$.

The term $H_{k'k_{\mathcal{U}}}(t',k_{\mathcal{V}},\lambda_{\mathcal{V}})$ that appears in the summand is defined in (5.24) with $h^{(\mathcal{S}_j,n_j)}$ given by the induction hypothesis. The summation $\sum_{(a)}$ is taken over the same set of variables as $\sum_{(c[+])}$ but with the restrictions (in addition to those in $\sum_{(c[\zeta])}$) that $q = r$, $n_j = 0$ and $N_{\mathfrak{l}} \geq M^{\delta}$ for each $j$ and each $\mathfrak{l} \in \mathcal{L}_j$ (where $\mathcal{L}_j$ is the set of leaves of $\mathcal{S}_j$, see Remark 3.3), and $\sum_{(b)} = \sum_{(c[+])} - \sum_{(a)}$.

The above is a valid inductive definition, i.e. the tensors $h^{(\mathcal{S}_j,n_j)}$ in (5.24) are already defined when we use them to define $h^{(\mathcal{S},n)}$ via (5.26)–(5.27), thanks to Proposition 3.7 (4). Note that the first term on the right hand side of (5.26) and the first line in (5.27) are precisely the random $(1,1)$ tensors described in Section 2.4.1; the rest come from higher order iterations in Section 2.4.2.

*Step 4: definition of $z_M$.* Now we have finished the inductive definition of $\mathcal{S}$-tensors $h^{(\mathcal{S},n)}$ for $n \in \{0,1\}$ and regular plants $\mathcal{S}$ of frequency $N(\mathcal{S}) \leq M$ and $|\mathcal{S}| \leq D$. Using (3.8) this gives definition to terms $\Psi_k^{(\mathcal{S},n)} = \Psi_k^{(\mathcal{S},n)}(t') = \Psi_k[\mathcal{S}, h^{(\mathcal{S},n)}]$ for such $n$ and $\mathcal{S}$. Finally we shall construct $z_M$ to complete the inductive definition; this is simply defined to be the solution to the equation

$$
\begin{aligned}
(z_M)_k(t) = \sum_{\zeta\in\{\pm\}}\sum_{k'} \int \mathrm{d}t' \cdot (\mathscr{R}^{M,\zeta})_{kk'}(t,t') \sum_{\mathrm{sym}}\sum_{(d)} \Upsilon \cdot \chi_\tau(t') \\
\times \big[\mathcal{I}_\chi \Pi \mathcal{M}_q(\Psi_{k_1}^{(\mathcal{S}_1,n_1)},\cdots,\Psi_{k_r}^{(\mathcal{S}_r,n_r)}, z_{N_{r+1}}^*,\cdots,z_{N_q}^*)\big]_{k'}^{\zeta}(t'),
\end{aligned}
\tag{5.28}
$$

where $z_{N_j}^*$ $(r+1 \leq j \leq q)$ is either $z_{N_j}$ or $z_{N_j}^{\mathrm{lo}}$ or the high-modulation cutoff $z_{N_j}^{\mathrm{hi}} := z_{N_j} - z_{N_j}^{\mathrm{lo}}$. Here in (5.28), the sum $\sum_{(d)}$ is taken over $\mathscr{B}$, $n_j \in \{0,1\}$, regular plants $\mathcal{S}_j$ with frequency $N_j$ and size $|\mathcal{S}_j| \leq D$ for $1 \leq j \leq r$, and choices of $z_{N_j}^*$, under the restrictions that (i) if $N_j = M$ for some $1 \leq j \leq q$ then there is $q \geq j' \neq j$ with $N_{j'} \geq M^{\delta}$, (ii) either $N_j = M$ for at least one $r+1 \leq j \leq q$ and $z_{N_j}^* = z_{N_j}$ for all $r+1 \leq j \leq q$, or $N_j \leq M/2$ for all $r+1 \leq j \leq q$ and $z_{N_j}^* = z_{N_j}^{\mathrm{hi}}$ for at least one $r+1 \leq j \leq q$, or $(N_j \leq M/2) \wedge (z_{N_j}^* = z_{N_j}^{\mathrm{lo}})$ for all $r+1 \leq j \leq q$ and the plant

$$
\mathcal{S} = \mathtt{Trim}(\mathtt{Merge}(\mathtt{Trim}(\mathcal{S}_1, M^{\delta}),\cdots,\mathtt{Trim}(\mathcal{S}_r, M^{\delta}),\mathscr{B},\mathscr{O}), M^{\delta})
\tag{5.29}
$$

[1]Strictly speaking the sum over $\mathscr{O}$ should carry the coefficients in the linear combination of $\widetilde{\Psi}_k^{(\mathcal{S})}$ that gives (5.22) as above; these are constants, and for simplicity we will treat them as 1.

has size[1] $|\mathcal{S}| > D$.

Note that in (5.28), all terms $\Psi_{k_j}^{(\mathcal{S}_j,n_j)} = \Psi_{k_j}[\mathcal{S}_j, h^{(\mathcal{S}_j,n_j)}]$ are already defined for $n_j \in \{0,1\}$ and regular plants $\mathcal{S}_j$ with $N(\mathcal{S}_j) \leq M$ and $|\mathcal{S}_j| \leq D$, so (5.28) can be viewed as an equation for the function $z_M = (z_M)_k(t)$. If the mapping defined by the right hand side of (5.28) is a contraction mapping from the set $\{z_M : \|z_M\|_{X^{b_0}} \leq M^{-D_1}\}$ to itself, we define $z_M$ to be the unique fixed point of this mapping; otherwise define $z_M = 0$.

This finishes the inductive definition of the tensors $h^{(\mathcal{S},n)}$ and the remainder $z_N$, which give the definition of $y_N$ by the ansatz (5.17).

5.3. **The a priori estimates.** With the complete definition of $h^{(\mathcal{S},n)}$ tensors and $z_N$, we can now state the main a priori estimates for these terms. First notice that they satisfy the following simple properties, which are easily verified (using Definitions 3.5 and 3.6, and Proposition 3.7) during the construction process:

- The tensors $h^{(\mathcal{S},0)}$ are constant (i.e. do not depend on $\omega$) and are nonzero only when $\mathcal{S}$ is plain, the tensors $h^{(\mathcal{S},1)}$ are $\mathcal{B}_{N^{[\delta]}}$ measurable (recall Section 3.1 for definition) for any $\mathcal{S}$ with $N(\mathcal{S}) = N$ and $|\mathcal{S}| \leq D$, and the remainder $z_N$ is $\mathcal{B}_N$ measurable;
- All these terms are supported in $|t| \leq 1$, and $(z_N)_k(t)$ is supported in $\langle k\rangle \leq N$. In the support of $h_{kk_{\mathcal{U}}}^{(\mathcal{S},n)}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ we have $\langle k\rangle \leq N$, that $N_{\mathfrak{l}}/2 < \langle k_{\mathfrak{l}}\rangle \leq N_{\mathfrak{l}}$ for each $\mathfrak{l} \in \mathcal{U}$, $\langle k_{\mathfrak{f}}\rangle \leq N_{\mathfrak{f}}$ and $|\lambda_{\mathfrak{f}}| \leq 2N^{\kappa^2}$ for each $\mathfrak{f} \in \mathcal{V}$, and that there is no pairing in $k_{\mathcal{U}}$.

The main a priori estimates are listed in the following proposition.

**Proposition 5.1.** *Given a dyadic $M$, consider the following set of statements (viewed as an event for $\omega$), which we shall refer to as* $\mathtt{Local}(M)$ *below:*

*(1) For any plain regular plant $\mathcal{S} = (\mathcal{L}, \varnothing, \mathcal{Y})$ with $N(\mathcal{S}) = N < M$ and $|\mathcal{S}| \leq D$, we have that*

$$h^{(\mathcal{S},0)} = h_{kk_{\mathcal{U}}}^{(\mathcal{S},0)}(t) \text{ is supported in the set where } k = \sum_{\mathfrak{l}\in\mathcal{U}} \zeta_{\mathfrak{l}} k_{\mathfrak{l}}. \tag{5.30}$$

*For any $\Gamma \in \mathbb{Z}$, let $h^{(\mathcal{S},0,\Gamma)}$ be the restriction of $h^{(\mathcal{S},0)}$ to the set where*

$$|k|^2 - \sum_{\mathfrak{l}\in\mathcal{U}} \zeta_{\mathfrak{l}} |k_{\mathfrak{l}}|^2 = \Gamma, \tag{5.31}$$

*obtained by multiplying by the indicator function of this set. Let $(B, C)$ be a subpartition of $\mathcal{U}$ and let $E = \mathcal{U}\backslash(B \cup C)$. Then we have*

$$\int_{\mathbb{R}} \langle\lambda\rangle^{2b} \Big(\sum_{\Gamma\in\mathbb{Z}} \|\widehat{h_{kk_{\mathcal{U}}}^{(\mathcal{S},0,\Gamma)}}(\lambda)\|_{kk_B\to k_C}\Big)^2 \,\mathrm{d}\lambda \leq \Big(\prod_{\mathfrak{l}\in B\cup C} N_{\mathfrak{l}}^{\beta_1} \prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-8\varepsilon} \prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{4\varepsilon} \prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \cdot \mathcal{X}_0 \mathcal{X}_1\Big)^2, \tag{5.32}$$

*with the quantities*

$$\mathcal{X}_0 = \begin{cases} \big(\max_{\mathfrak{l}\in C} N_{\mathfrak{l}}\big)^{-\beta_1}, & \text{if } C \neq \varnothing; \\ \big(\min_{\mathfrak{l}\in\mathcal{L}} N_{\mathfrak{l}}\big)^{\frac{d}{2}-\beta_1}, & \text{if } C = E = \varnothing; \\ N^{-\varepsilon\delta}, & \text{if } B = C = \varnothing; \\ 1, & \text{otherwise}, \end{cases} \qquad \mathcal{X}_1 = \begin{cases} 1, & \text{if } \max_{\mathfrak{l}\in\mathcal{L}} N_{\mathfrak{l}} \geq (10^3 dp)^{-|\mathcal{L}|} N; \\ N^{-4\varepsilon}, & \text{if } \max_{\mathfrak{l}\in\mathcal{L}} N_{\mathfrak{l}} < (10^3 dp)^{-|\mathcal{L}|} N. \end{cases} \tag{5.33}$$

[1]Note that $\mathscr{O}$ actually does not appear in the summation (5.28). But this is fine, since it can easily be checked that the *size* of $\mathcal{S}$ defined by (5.29) does not depend on $\mathscr{O}$.

*(2) For any regular plant $\mathcal{S} = (\mathcal{L}, \mathcal{V}, \mathcal{Y})$ with $N(\mathcal{S}) = N < M$ and $|\mathcal{S}| \leq D$, consider the tensor $h^{(\mathcal{S},1)} = h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})$. Let $(B, C)$ be a subpartition of $\mathcal{U}$ and let $E = \mathcal{U}\backslash(B \cup C)$. Then, recall the norms defined in (3.17)–(3.18), for $C \neq \varnothing$ we have*

$$\|h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X^{1-b,-b_0}_{\mathcal{V}}[kk_B \to k_C]} \leq \prod_{\mathfrak{l}\in B\cup C} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l}\in \mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{8\varepsilon} \prod_{\mathfrak{p}\in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \prod_{\mathfrak{f}\in \mathcal{V}} N_{\mathfrak{f}}^{d} \cdot \big(\max_{\mathfrak{l}\in C} N_{\mathfrak{l}}\big)^{-\beta}, \tag{5.34}$$

*while for $C = \varnothing$ we have*

$$\|h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X^{\widetilde{b},-b_0}_{\mathcal{V}}[kk_B]} \leq \prod_{\mathfrak{l}\in B} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l}\in \mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{8\varepsilon} \prod_{\mathfrak{p}\in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \prod_{\mathfrak{f}\in \mathcal{V}} N_{\mathfrak{f}}^{d} \cdot N^{-\varepsilon}, \tag{5.35}$$

*where $\widetilde{b}$ equals $1-b$ if $E \neq \varnothing$ and equals $b$ if $E = \varnothing$. We also have a localization bound (for $C = \varnothing$)*

$$\left\|\left(1+\frac{1}{N^{2\delta}}\Big|k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell\Big|\right)^{\kappa} h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})\right\|_{X^{\widetilde{b},-b_0}_{\mathcal{V}}[kk_B]} \leq \prod_{\mathfrak{l}\in B} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l}\in \mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{8\varepsilon} \prod_{\mathfrak{p}\in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \prod_{\mathfrak{f}\in \mathcal{V}} N_{\mathfrak{f}}^{d}, \tag{5.36}$$

*where $\ell = \sum_{\mathfrak{f}\in\mathcal{V}} \zeta_{\mathfrak{f}}k_{\mathfrak{f}}$, and $\widetilde{b}$ is the one in (5.35).*

*Finally, we have an auxiliary bound for the $\lambda_{\mathcal{V}}$-derivative[1] of $h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ (for $C = E = \varnothing$),*

$$\|\partial_{\lambda_{\mathcal{V}}} h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X^{b,-b_0}_{\mathcal{V}}[kk_{\mathcal{U}}]} \leq \exp[(\log N)^5 + |\mathcal{S}|(\log N)^3]. \tag{5.37}$$

*(3) For $n \in \{0, 1\}$ and regular plant $\mathcal{S}$ with $N(\mathcal{S}) = N < M$ and $|\mathcal{S}| \leq D$, let the expression $\Psi^{(\mathcal{S},n)}_k = \Psi_k[\mathcal{S}, h^{(\mathcal{S},n)}]$ be defined as in (3.8), then we have*

$$\|\Psi^{(\mathcal{S},n)}\|_{X^{s',b_0}} \leq \tau^{-\theta_0} N^{s'-s} \prod_{\mathfrak{n}\in\mathcal{L}\cup\mathcal{V}\cup\mathcal{Y}} N_{\mathfrak{n}}^{-\delta^3} \tag{5.38}$$

*for any $s - \delta^2 < s' < s$, where $(s, b_0, \theta_0)$ are defined in (1.3) and Section 3.1.*

*(4) For all $N < M$, the mapping that defines $z_N$ (namely the right hand side of (5.28) but with $M$ replaced by $N$ and then $z_N$ replaced by an independent variable $z$) is a contraction mapping from $\{z : \|z\|_{X^{b_0}} \leq N^{-D_1}\}$ to itself with $D_1$ as in (3.3). In particular, we have $\|z_N\|_{X^{b_0}} \leq N^{-D_1}$ for all $N < M$.*

*(5) Let $y_N$ and $v^{\dagger}_N$ be defined as in (5.17) for $N < M$, then they solve the system (5.16) for $N < M$. Moreover, for any $N_2, \cdots, N_q < M$ and any $N$ (which may be $\geq M$), consider the operator $\mathscr{L}^{\zeta}$ (which is complex linear if $\zeta = +$, and conjugate complex linear if $\zeta = -$) defined by*

$$(\mathscr{L}^{\zeta}w)_k(t) = \chi_{\tau}(t) \cdot \mathcal{I}_{\chi}\Pi_N\mathcal{M}_q(y^*_{N_2}, \cdots, w, \cdots, y^*_{N_q})_k(t), \tag{5.39}$$

*and the corresponding kernel $(\mathscr{L}^{\zeta})_{kk'}(t, t')$, where each $y^*_{N_j}$ is either $y_{N_j}$ or one of its components, namely $z_{N_j}$ (possibly with Fourier truncations similar to the ones in $z^{\mathrm{hi}}_N$ and $z^{\mathrm{lo}}_N$ defined in Section 5.2) or $\Psi[\mathcal{S}_j, h^{(\mathcal{S}_j,n_j)}]$ for some $n_j \in \{0, 1\}$ and regular plant $\mathcal{S}_j$ with $N(\mathcal{S}_j) = N_j$ and $|\mathcal{S}_j| \leq D$ (see (3.8) and (5.17)), then they satisfy*

$$\|\mathscr{L}^{\zeta}\|_{X^{1-b,-b}[k\to k']} \leq \tau^{(5\kappa)^{-1}}\big(\max_{2\leq j\leq q} N_j\big)^{-4\varepsilon\delta}. \tag{5.40}$$

*Now, with the above definition of* `Local`$(M)$, *we have that* the probability that `Local`$(M)$ holds but `Local`$(2M)$ does not hold *is $\leq C_{\theta}e^{-(\tau^{-1}M)^{\theta}}$. In particular, $\tau^{-1}$-certainly,* `Local`$(M)$ *holds for all $M$.*

[1] Note that, once we have (5.37), we automatically have the same bound for the norm with the weight in (5.36), with the right hand side multiplied by $N^{\kappa}$, which is negligible as the right hand side is super-polynomial.

## 6. Trimming and merging estimates

In this and in the next section we prove Proposition 5.1. This section is devoted to the proof of some important estimates on trimming and merging of tensors, which will be crucial to the proof we will give in Section 7. These are: trimming bounds in Section 6.1, no-over-pairing merging bounds in Section 6.2 and general merging bounds in Section 6.3. Throughout this and the next section we will fix a dyadic $M$ (we may always assume $M \gg_{C_\theta} 1$, since otherwise the relevant bounds become trivial as $\tau \ll_{C_\theta} 1$), and assume that the statement $\mathtt{Local}(M)$, defined in Proposition 5.1, is already true. If results in this and the next section rely on more assumptions (such as parts of $\mathtt{Local}(2M)$ that have been established in preceding proofs), we will explicitly point this out.

All the quantities (functions, tensors, etc.) in this and the next section that depend on $t$ (or $t'$ etc.) will be supported in $|t| \leq 1$ (or $|t'| \leq 1$ etc.). This implies that their derivatives in the time-Fourier variables ($\lambda$, $\lambda_j$ etc.) automatically satisfy the same bounds as they do; these derivative bounds will be useful in applications of *meshing arguments* (see the proof of Proposition 6.1 for details) below. Moreover due to $\mathtt{Local}(M)$ part (4), the functions $z_{N'} = (z_{N'})_k(t)$ for $N' < M$ are already defined and satisfy that $\|z_{N'}\|_{X^{b_0}} \leq (N')^{-D_1}$. When applying $\mathtt{Trim}$ functions below we always fix these $z_{N'}$, and the $f_{N'}$ defined in (5.10).

### 6.1. Trimming estimates.

We first prove the trimming estimates.

**Proposition 6.1** (Trimmed tensor bounds)**.** *Let $\mathcal{S}$ be a regular plant, $N(\mathcal{S}) = N \leq M$ and $|\mathcal{S}| \leq D$. Let $h = h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ be an $\mathcal{S}$-tensor which is $\mathcal{B}_{N^{[\delta]}}$ measurable. For $N^\delta \leq R \leq M^\delta$, consider the trimmed plant $\mathcal{S}' = (\mathcal{L}', \mathcal{V}', \mathcal{Y}') = \mathrm{Trim}(\mathcal{S}, R)$ and the trimmed tensor $h' = (h')_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'}, \lambda_{\mathcal{V}'}) =$ $\mathtt{Trim}(h, R)$. Let $(B, C)$ be a subpartition of $\mathcal{U}$ with $E = \mathcal{U} \backslash (B \cup C)$ and $(B', C')$ be a subpartition of $\mathcal{U}'$ with $E' = \mathcal{U}' \backslash (B' \cup C')$.*

*(1) Assume $h$ satisfies that, for any $(B, C)$ with $C \neq \varnothing$,*

$$\|h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X_{\mathcal{V}}^{-b_0}[kk_B \to k_C]} \lesssim \mathfrak{X} \cdot \prod_{\mathfrak{l} \in B \cup C} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l} \in \mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{l} \in E} N_{\mathfrak{l}}^{8\varepsilon} \prod_{\mathfrak{p} \in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \prod_{\mathfrak{f} \in \mathcal{V}} N_{\mathfrak{f}}^{d} \cdot \big(\max_{\mathfrak{l} \in C} N_{\mathfrak{l}}\big)^{-\beta}, \tag{6.1}$$

*and for $C = \varnothing$,*

$$\|h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X_{\mathcal{V}}^{-b_0}[kk_B]} \lesssim \mathfrak{X} \cdot \prod_{\mathfrak{l} \in B} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l} \in \mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{l} \in E} N_{\mathfrak{l}}^{8\varepsilon} \prod_{\mathfrak{p} \in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \prod_{\mathfrak{f} \in \mathcal{V}} N_{\mathfrak{f}}^{d} \cdot N^{-\varepsilon}, \tag{6.2}$$

$$\left\|\left(1 + \frac{1}{N^{2\delta}}\Big|k - \sum_{\mathfrak{l} \in \mathcal{U}} \zeta_{\mathfrak{l}} k_{\mathfrak{l}} - \ell\Big|\right)^{\kappa} h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})\right\|_{X_{\mathcal{V}}^{-b_0}[kk_B]} \lesssim \mathfrak{X} \cdot \prod_{\mathfrak{l} \in B} N_{\mathfrak{l}}^{\beta} \prod_{\mathfrak{l} \in \mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{l} \in E} N_{\mathfrak{l}}^{8\varepsilon} \prod_{\mathfrak{p} \in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \prod_{\mathfrak{f} \in \mathcal{V}} N_{\mathfrak{f}}^{d}, \tag{6.3}$$

*where $\ell = \sum_{\mathfrak{f} \in \mathcal{V}} \zeta_{\mathfrak{f}} k_{\mathfrak{f}}$, and also the auxiliary bound for $C = E = \varnothing$,*

$$\|\partial_{\lambda_{\mathcal{V}}} h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X_{\mathcal{V}}^{-b_0}[kk_{\mathcal{U}}]} \leq \mathfrak{X} \cdot \exp[(\log N)^5 + |\mathcal{S}|(\log N)^3]. \tag{6.4}$$

*Then, $\tau^{-1}M$-certainly, the estimates (6.1)–(6.3) hold for $h = h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ replaced by $h' =$ $h'_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'}, \lambda_{\mathcal{V}'})$, the sets $B, C, E, \mathcal{P}, \mathcal{V}$ etc. replaced by $B', C', E', \mathcal{P}', \mathcal{V}'$ etc., the fraction $1/N^{2\delta}$ in (6.3) replaced by $1/\max(N^{2\delta}, R)$, and the factor $\mathfrak{X}$ replaced by $\mathfrak{X} \cdot \tau^{-\theta} M^{\theta}$.*

*(2) Assume $\mathcal{V} = \varnothing$, $h$ is supported in the set $k = \sum_{\mathfrak{l} \in \mathcal{U}} \zeta_{\mathfrak{l}} k_{\mathfrak{l}}$ and satisfies that, for any $(B, C)$,*

$$\|h_{kk_{\mathcal{U}}}\|_{kk_B \to k_C} \leq \mathfrak{X} \cdot \prod_{\mathfrak{l} \in B \cup C} N_{\mathfrak{l}}^{\beta_1} \prod_{\mathfrak{l} \in \mathcal{P}} N_{\mathfrak{l}}^{-8\varepsilon} \prod_{\mathfrak{l} \in E} N_{\mathfrak{l}}^{4\varepsilon} \prod_{\mathfrak{p} \in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \cdot \mathcal{X}_0 \mathcal{X}_1, \tag{6.5}$$

*where $\mathcal{X}_0$ and $\mathcal{X}_1$ are defined as in (5.33). Assume also that either $C' \cup E' \neq \varnothing$, or $\mathcal{L} \neq \mathcal{L}'$ (i.e. $N_{\mathfrak{l}} < R$ for at least one $\mathfrak{l} \in \mathcal{L}$), then $\tau^{-1}M$-certainly, the estimates (6.1)–(6.3) hold for $h' = h'_{kk_{\mathcal{U}'}}$, with the sets $B, C, E, \mathcal{P}$ etc. replaced by $B', C', E', \mathcal{P}'$ etc., the fraction $1/N^{2\delta}$ in (6.3) replaced by $1/R$, and the factor $\mathfrak{X}$ replaced by $\mathfrak{X} \cdot \tau^{-\theta} M^{\theta}$ for (6.1), and by $\mathfrak{X} \cdot \tau^{-\theta} M^{\theta}(1 + N^{-3\varepsilon} R^{d/2-\beta_1})$ for (6.2)–(6.3).*

*Proof.* (1) By definition we have

$$\begin{aligned} h' = h'_{kk_{\mathcal{U}'}}(k_{\mathcal{V}'}, \lambda_{\mathcal{V}'}) &= \sum_{k_{\mathcal{V}\setminus\mathcal{V}'}} \int \mathrm{d}\lambda_{\mathcal{V}\setminus\mathcal{V}'} \cdot \widetilde{h}_{kk_{\mathcal{U}'}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) \prod_{\mathfrak{f} \in \mathcal{V}\setminus\mathcal{V}'} (\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}}(\lambda_{\mathfrak{f}}), \\ \widetilde{h}_{kk_{\mathcal{U}'}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) &:= \sum_{k_{\mathcal{U}\setminus\mathcal{U}'}} h_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) \prod_{\mathfrak{l} \in \mathcal{U}\setminus\mathcal{U}'} (f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}}. \end{aligned} \tag{6.6}$$

By Cauchy-Schwartz we have

$$\|h'\|_{X_{\mathcal{V}'}^{-b_0}[\cdots]} \lesssim \|\widetilde{h}\|_{X_{\mathcal{V}}^{-b_0}[\cdots]} \cdot \prod_{\mathfrak{f} \in \mathcal{V}\setminus\mathcal{V}'} \|z_{N_{\mathfrak{f}}}\|_{X^{b_0}} \lesssim \|\widetilde{h}\|_{X_{\mathcal{V}}^{-b_0}[\cdots]} \cdot \prod_{\mathfrak{f} \in \mathcal{V}\setminus\mathcal{V}'} N_{\mathfrak{f}}^{-D_1}, \tag{6.7}$$

where $[\cdots]$ represents any $kk_{B'} \to k_{C'}$ or $kk_{B'}$ or weighted norm as in (6.1)–(6.3).

It thus suffices to bound the corresponding norms for $\widetilde{h}$. Note that if we *fix the values of* $k_{\mathcal{V}}$ and $\lambda_{\mathcal{V}}$, the tensor $\widetilde{h}_{kk_{\mathcal{U}'}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ can be estimated $\tau^{-1}M$-certainly, using Proposition 4.14; in order to transform this into a bound for the $X_{\mathcal{V}}^{-b_0}[\cdots]$ norms, we need to make this bound *uniform in all choices of* $k_{\mathcal{V}}$ *and* $\lambda_{\mathcal{V}}$. There is no problem in doing so for $k_{\mathcal{V}}$ since[1] the number of choices for $k_{\mathcal{V}}$ is at most $M^{\kappa}$. To deal with $\lambda_{\mathcal{V}}$, we will employ the following argument, which will be referred to as the *meshing argument* (see the proof of Lemma 4.2 in [36]), and will be used frequently in the proofs below. First note that $|\lambda_{\mathfrak{f}}| \lesssim N^{\kappa^2}$ for each $\mathfrak{f} \in \mathcal{V}$, then we divide the big box $\{\lambda_{\mathcal{V}} : |\lambda_{\mathfrak{f}}| \lesssim N^{\kappa^2}, \forall \mathfrak{f} \in \mathcal{V}\}$ into small boxes $\mathfrak{B}$ of size $\nu := \exp(-(\log N)^6)$. Now by taking averages on these small boxes and using Poincaré inequality, there exists a tensor $h_{\mathrm{avg}} = (h_{\mathrm{avg}})_{kk_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ such that $h_{\mathrm{avg}}$ satisfies the same measurability condition as $h$, that $h_{\mathrm{avg}}$ is supported in the big box and constant when $\lambda_{\mathcal{V}}$ moves within each small box (and other parameters are fixed), and that

$$\|h - h_{\mathrm{avg}}\|_{X_{\mathcal{V}}^{-b_0}[kk_{\mathcal{U}}]} \lesssim \nu \cdot (\|h\|_{X_{\mathcal{V}}^{-b_0}[kk_{\mathcal{U}}]} + \|\partial_{\lambda_{\mathcal{V}}} h\|_{X_{\mathcal{V}}^{-b_0}[kk_{\mathcal{U}}]}). \tag{6.8}$$

Let $\widetilde{h}_{\mathrm{avg}}$ be defined from $h_{\mathrm{avg}}$, in the same way as $\widetilde{h}$ is defined from $h$, then for *fixed values of* $\lambda_{\mathcal{V}}$, the tensor $(\widetilde{h}_{\mathrm{avg}})_{kk'_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ can be estimated $\tau^{-1}M$-certainly, using Proposition 4.14, in the same way that $\widetilde{h}$ is bounded in terms of $h$. Since $\lambda_{\mathcal{V}}$ has at most $\exp(\kappa(\log N)^6)$ different choices in studying $h_{\mathrm{avg}}$, and $N \leq M$, we know that the estimate for $\widetilde{h}_{\mathrm{avg}}$ is *uniform in all choices of* $\lambda_{\mathcal{V}}$, after removing an exceptional set whose probability is still $\leq C_{\theta} e^{-(\tau^{-1}M)^{\theta}}$. This gives the $X_{\mathcal{V}}^{-b_0}[\cdots]$ norm bounds for $\widetilde{h}_{\mathrm{avg}}$; but by (6.4), (6.8) and our choice for $\nu$, the $X_{\mathcal{V}}^{-b_0}[\cdots]$ norm of the difference $\widetilde{h} - \widetilde{h}_{\mathrm{avg}}$ is negligible, so we get the desired $X_{\mathcal{V}}^{-b_0}[\cdots]$ norm bounds for $\widetilde{h}$.

Armed with the meshing argument, we can now apply Proposition 4.14 to control the $X_{\mathcal{V}}^{-b_0}[\cdots]$ norms of $\widetilde{h}$. Given any subpartition $(B', C')$ of $\mathcal{U}'$, $C' \neq \varnothing$, let $E = E' = \mathcal{U}' \backslash (B' \cup C')$; since $h$ is $\mathcal{B}_{N^{[\delta]}}$ measurable and $N_{\mathfrak{l}} > N^{[\delta]}$ for all $\mathfrak{l} \in \mathcal{U}$, we can apply Proposition 4.14 (note that $(f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}} = \Delta_{N_{\mathfrak{l}}} \gamma_{k_{\mathfrak{l}}} \cdot \eta_{k_{\mathfrak{l}}}(\omega)$, where $\Delta_{N_{\mathfrak{l}}} \gamma_{k_{\mathfrak{l}}}$ can be replaced by $\tau^{-\theta} N_{\mathfrak{l}}^{-\alpha+\theta}$ due to (5.7) and Lemma

[1]The same applies to the $k_E$ variables when measuring $kk_B \to k_C$ norms, where $E = \mathcal{U} \backslash (B \cup C)$.

4.10) to get

$$\|\widetilde{h}_{kk_{\mathcal{U}'}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})\|_{X_{\mathcal{V}}^{-b_0}[kk_{B'}\to k_{C'}]}\lesssim \tau^{-\theta}M^{\theta}\prod_{\mathfrak{l}\in\mathcal{U}\backslash\mathcal{U}'}N_{\mathfrak{l}}^{-\alpha+\theta}\cdot\sup_{(F,G)}\|h_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})\|_{X_{\mathcal{V}}^{-b_0}[kk_B\to k_C]},$$

where $(F,G)$ is any partition of $\mathcal{U}\backslash\mathcal{U}'$, and $B=B'\cup F$, $C=C'\cup G$. The conclusion about the $X_{\mathcal{V}'}^{-b_0}[kk_{B'}\to k_{C'}]$ norms then follows by combining (6.1) and (6.7), and noticing that $\mathcal{P}'\subset\mathcal{P}$, $\mathcal{Y}'\subset\mathcal{Y}$, and $\max_{\mathfrak{l}\in C'}N_{\mathfrak{l}}=\max_{\mathfrak{l}\in C}N_{\mathfrak{l}}$. When $C'=\varnothing$, the proofs for the $X_{\mathcal{V}'}^{b}[kk_{B'}]$ norms are completely analogous (simply choose $G=\varnothing$ so $C=\varnothing$), and so is the weighted norm bound, where for the latter we notice that

$$1+\frac{1}{\max(N^{2\delta},R)}\bigg|k-\sum_{\mathfrak{l}\in\mathcal{U}'}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell'\bigg|\lesssim 1+\frac{1}{N^{2\delta}}\bigg|k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell\bigg|$$

with $\ell'=\sum_{\mathfrak{f}\in\mathcal{V}'}\zeta_{\mathfrak{f}}k_{\mathfrak{f}}$ and $\ell=\sum_{\mathfrak{f}\in\mathcal{V}}\zeta_{\mathfrak{f}}k_{\mathfrak{f}}$, since $\langle k_{\mathfrak{f}}\rangle\leq R$ for any $\mathfrak{f}\in\mathcal{V}\backslash\mathcal{V}'$ (hence $|\ell-\ell'|\lesssim R$) and $\langle k_{\mathfrak{l}}\rangle\leq R$ for any $\mathfrak{l}\in\mathcal{U}\backslash\mathcal{U}'$.

(2) The proof is similar to (1) but much easier since there is no blossom $\mathfrak{f}\in\mathcal{V}$ (hence no meshing argument) involved. Since $|k-\sum_{\mathfrak{l}\in\mathcal{U}'}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}|\lesssim R$ because $k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}=0$ and $\langle k_{\mathfrak{l}}\rangle\leq R$ for $\mathfrak{l}\in\mathcal{U}\backslash\mathcal{U}'$, we know that the weighted norm bound (6.3) follows from the unweighted norm bound (6.2). For the $kk_{B'}\to k_{C'}$ norms in (6.1)–(6.2), we simply apply Proposition 4.14; when $C'\neq\varnothing$ we readily get (6.1) with the indicated changes. When $C'=\varnothing$ (and $C=\varnothing$) we make two observations. First, when $E=E'=\varnothing$ there is an extra factor $\mathcal{X}_0$ in (6.5), but since $\min_{\mathfrak{l}\in\mathcal{L}}N_{\mathfrak{l}}\leq R$ due to the assumption $\mathcal{L}\neq\mathcal{L}'$, we have $\mathcal{X}_0\lesssim R^{d/2-\beta_1}$ by definition (5.33), which gives rise to the factor $R^{d/2-\beta_1}$ in the desired estimate. Second, thanks to the different powers between (6.1)–(6.3) and (6.5), in the process of using (6.5) for $h$ to deduce (6.1)–(6.2) for $h'$, we will be gaining a factor

$$\prod_{\mathfrak{l}\in\mathcal{U}\backslash E'}N_{\mathfrak{l}}^{\beta_1-\beta}\prod_{\mathfrak{l}\in E'\cup\mathcal{P}}N_{\mathfrak{l}}^{-4\varepsilon}\leq\big(\max_{\mathfrak{l}\in\mathcal{L}}N_{\mathfrak{l}}\big)^{-4\varepsilon},$$

which is at most $N^{-4\varepsilon}$ if $\max_{\mathfrak{l}\in\mathcal{L}}N_{\mathfrak{l}}\geq(10^3dp)^{-|\mathcal{L}|}N$. If $\max_{\mathfrak{l}\in\mathcal{L}}N_{\mathfrak{l}}<(10^3dp)^{-|\mathcal{L}|}N$ we gain exactly the same $N^{-4\varepsilon}$ from the factor $\mathcal{X}_1$ in (5.33). In any case, this gain will contribute the $N^{-3\varepsilon}$ in the desired estimate, after providing the $N^{-\varepsilon}$ factor for (6.2). □

6.2. **No-over-pairing merging estimates.** Next we will prove two merging estimates in the no-over-pairing case by introducing a *selection algorithm*. Note that in Propositions 6.2 and 6.3 below, the sets $\mathcal{U}_j$ are just sets by themselves and are *not* coming from any plant; nevertheless in applications, they do occur as suitable *subsets* of the sets coming from some plants, see the proof of Propositions 6.4 and 6.5.

**Proposition 6.2** (Selection algorithm: Case I)**.** *Let $\mathcal{U}_2,\cdots,\mathcal{U}_p$ be pairwise disjoint finite index sets (could be empty), $|\mathcal{U}_j|\leq D$. Given $\zeta_j\in\{\pm\}$ for $1\leq j\leq p$ and $\zeta_{\mathfrak{l}}\in\{\pm\}$ for any $\mathfrak{l}\in\mathcal{U}_j$, and $N_j$ for $2\leq j\leq p$ and $N_{\mathfrak{l}}$ for any $\mathfrak{l}\in\mathcal{U}_j$, let $N_*=\max(N_2,\cdots,N_p)$, and define $\zeta_{\mathfrak{l}}^*=\zeta_{\mathfrak{l}}\zeta_j$ for $\mathfrak{l}\in\mathcal{U}_j$. Assume that*

$$\sum_{j=1}^{p}\zeta_j=1,\quad N_j^{\delta}\leq N_{\mathfrak{l}}\leq N_j\,(\forall\mathfrak{l}\in\mathcal{U}_j,\,2\leq j\leq p).$$

*Assume there are some pairings in $\mathcal{W}:=\mathcal{U}_2\cup\cdots\cup\mathcal{U}_p$ (i.e. a collection of pairwise disjoint two-element subsets of $\mathcal{W}$, each containing two elements from two different $\mathcal{U}_j$), such that for any pair $(\mathfrak{l},\mathfrak{l}')$ we have $\zeta_{\mathfrak{l}'}^*=-\zeta_{\mathfrak{l}}^*$ and $N_{\mathfrak{l}'}=N_{\mathfrak{l}}$. Let $h^{(j)}=h^{(j)}_{k_jk_{\mathcal{U}_j}}$, where $2\leq j\leq p$, be tensors, and*

*$h = h_{kk_1\cdots k_p}$ be a tensor supported in $\langle k_j\rangle \le N_j$ for $2\le j\le p$. Let the set of paired elements in $\mathcal{W}$ be $\mathcal{Q}$ and the set of unpaired elements be $\mathcal{U}$, define the semi-product*

$$H = H_{kk_1k_{\mathcal{U}}} = \sum_{k_2,\cdots,k_p}\sum_{k_{\mathcal{Q}}} h_{kk_1\cdots k_p}\prod_{j=2}^{p}\big(h^{(j)}_{k_jk_{\mathcal{U}_j}}\big)^{\zeta_j}, \tag{6.9}$$

*where the sum is taken over all $(k_2,\cdots,k_p)$ and $k_{\mathcal{Q}}$ that satisfy $k_{\mathfrak{l}'} = k_{\mathfrak{l}}$ for any pairing $(\mathfrak{l},\mathfrak{l}')$.*

*For each $2\le j\le p$, in the support of $h^{(j)}$ we assume that $k_{\mathfrak{l}}\in\mathbb{Z}^d$, $\langle k_j\rangle\le N_j$ and $N_{\mathfrak{l}}/2<\langle k_{\mathfrak{l}}\rangle\le N_{\mathfrak{l}}$ for each $\mathfrak{l}\in\mathcal{U}_j$. Moreover, $h^{(j)}$ has one of the following three types:*

(1) *Type R0: where we assume, in the support of $h^{(j)}$, that*

$$\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}} = 1,\quad k_j = \sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}},\quad |k_j|^2 - \sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}|k_{\mathfrak{l}}|^2 = \Gamma_j. \tag{6.10}$$

*Moreover, for any partition $(P_j,Q_j)$ of $\mathcal{U}_j$, we assume*

$$\|h^{(j)}\|_{k_jk_{P_j}\to k_{Q_j}} \le \mathfrak{X}_j\cdot\prod_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}^{\beta_1}\cdot\mathcal{Z}_{0,j}\mathcal{Z}_{1,j}, \tag{6.11}$$

*where $\mathcal{Z}_{0,j}$ and $\mathcal{Z}_{1,j}$ are defined similarly as in (5.33) but with the following modifications,*

$$\mathcal{Z}_{0,j} = \begin{cases}\big(\max_{\mathfrak{l}\in Q_j}N_{\mathfrak{l}}\big)^{-\beta_1}, & \text{if } Q_j\neq\varnothing;\\ \big(\min_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}\big)^{\frac{d}{2}-\beta_1}, & \text{if } Q_j=\varnothing,\end{cases}\qquad \mathcal{Z}_{1,j} = \begin{cases}1, & \text{if } \max_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}\ge(10^3dp)^{-D}N_j;\\ N_j^{-4\varepsilon}, & \text{if } \max_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}<(10^3dp)^{-D}N_j.\end{cases} \tag{6.12}$$

(2) *Type R0$^+$: similar to type R0, but instead of (6.10), in the support of $h^{(j)}$ we only assume that*

$$k_j - \sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}} = m_j \tag{6.13}$$

*for some given $m_j\in\mathbb{Z}^d$. Moreover the bounds (6.11)–(6.12) also hold (in particular $\mathcal{Z}_{1,j} = N_j^{-4\varepsilon}$ if $\mathcal{U}_j=\varnothing$), except that in (6.12), when $Q_j=\varnothing$, we have $\mathcal{Z}_{0,j}=1$ instead of $(\min_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}})^{d/2-\beta_1}$.*

(3) *Type R1: where we assume, in the support of $h^{(j)}$, that*

$$\Big|k_j - \sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}} - m_j\Big| \le (N_*)^{3\delta} \tag{6.14}$$

*for some given $m_j\in\mathbb{Z}^d$. Moreover, for any partition $(P_j,Q_j)$ of $\mathcal{U}_j$ we have*

$$\|h^{(j)}\|_{k_jk_{P_j}\to k_{Q_j}} \le \begin{cases}\mathfrak{X}_j\cdot\prod_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}^{\beta}\cdot\big(\max_{\mathfrak{l}\in Q_j}N_{\mathfrak{l}}\big)^{-\beta}, & Q_j\neq\varnothing;\\ \mathfrak{X}_j\cdot\prod_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}^{\beta}\cdot N_j^{-\varepsilon}, & Q_j=\varnothing.\end{cases} \tag{6.15}$$

*Regarding the tensor $h$, we assume that $|h_{kk_1\cdots k_p}|\lesssim 1$, and that $h=h_{kk_1\cdots k_p}$ is supported in the set*

$$k = \sum_{j=1}^{p}\zeta_jk_j,\quad |k|^2 - \sum_{j=1}^{p}\zeta_j|k_j|^2 = \Gamma, \tag{6.16}$$

*and that any pairing in $(k,k_1,\cdots,k_p)$ must be over-paired.*

FIGURE 3. Classification of $\mathcal{U}_j$, $j \in \{2, \cdots, p\}$

*Then, for any partition $(P, Q)$ of $\mathcal{U}$, we have the bound*

$$\prod_{\mathfrak{l} \in \mathcal{Q}} N_{\mathfrak{l}}^{-\alpha+4\varepsilon+\theta} \cdot \|H\|_{kk_P \to k_1 k_Q} \lesssim \prod_{j=2}^{p} \mathfrak{X}_j \cdot \prod_{\mathfrak{l} \in \mathcal{U}} N_{\mathfrak{l}}^{\beta} \cdot \prod_{j}^{(0,0^+)} N_j^{-2\varepsilon} \cdot (N_*)^{-\varepsilon^3}, \tag{6.17}$$

*where the product* $\prod_j^{(0,0^+)}$ *is taken over all* $2 \leq j \leq p$ *such that* $h^{(j)}$ *is of type* $R0$ *or type* $R0^+$. *The result is uniform in all parameters* $\Gamma, \Gamma_j, m_j$ *etc., and remains true if we replace* $p$ *by an odd* $3 \leq q \leq p$. *It also remains true if instead of the second equation in the support condition in (6.16), we assume that* $h$ *satisfies (4.27).*

*Proof.* The two key ingredients in the proof are Proposition 4.12 and Proposition 4.9.

*Step 1: first reductions.* We start by making an adjustment in notation, just like the one in the proof of Proposition 4.14, which will allow us to apply Proposition 4.12 below. For each pairing $(\mathfrak{l}, \mathfrak{l}')$, where $\mathfrak{l} \in \mathcal{U}_j$ and $\mathfrak{l}' \in \mathcal{U}_{j'}$, since in the sum (6.9) we are always assuming $k_{\mathfrak{l}} = k_{\mathfrak{l}'}$, we may combine them into a single element and include this element in both $\mathcal{U}_j$ and $\mathcal{U}_{j'}$. In this way we are changing pairings between different $\mathcal{U}_j$'s to intersections of different $\mathcal{U}_j$'s, which is the setting of Proposition 4.12. Then $\mathcal{U}$ will be the set of elements that occurs once in all the $\mathcal{U}_j$'s, and $\mathcal{Q}$ is the set of elements that occur twice.

Next, we will identify all subsets $A \subset \{2, \cdots, p\}$ such that each and every element in the union of $\mathcal{U}_j \, (j \in A)$ occurs twice in these sets $\mathcal{U}_j \, (j \in A)$. We only need to consider *minimal* subsets $A$ that satisfy this, which will be pairwise disjoint. Let them be $A_v \, (1 \leq v \leq s)$ and $B_u \, (1 \leq u \leq t)$, where for each $v$, the tensor $h^{(j)}$ is of type $R0$ for each $j \in A_v$, while for each $u$, there is at least one $j \in B_u$ such that the tensor $h^{(j)}$ is of type $R0^+$ or $R1$. Let $\mathcal{E}$ be the union of all these sets $A_v$ and $B_u$ and $\mathcal{D} = \{2, \cdots, p\} \backslash \mathcal{E}$; See Figure 3 for an illustration of the above subsets of $\{2, \cdots, p\}$. We know that $\mathcal{U}_j \subset \mathcal{Q}$ for $j \in \mathcal{E}$. Let $\mathcal{R} \subset \mathcal{Q}$ be the set of elements that occur twice in the sets

$\mathcal{U}_j (j \in \mathcal{D})$. Then we have

$$H = H_{kk_1k_{\mathcal{U}}} = \sum_{(k_{A_1},\cdots k_{A_s},k_{B_1},\cdots k_{B_t})} \prod_{v=1}^{s} H^{[v]}_{k_{A_v}} \prod_{u=1}^{t} R^{[u]}_{k_{B_u}} \cdot H^{\mathrm{sg}}_{kk_1k_{\mathcal{E}}k_{\mathcal{U}}}, \tag{6.18}$$

where

$$H^{[v]}_{k_{A_v}} := \sum_{(k_{\mathcal{U}_j}:j\in A_v)} \prod_{j\in A_v} h^{(j)}_{k_jk_{\mathcal{U}_j}}, \qquad R^{[u]}_{k_{B_u}} := \sum_{(k_{\mathcal{U}_j}:j\in B_u)} \prod_{j\in B_u} h^{(j)}_{k_jk_{\mathcal{U}_j}}, \tag{6.19}$$

$$H^{\mathrm{sg}}_{kk_1k_{\mathcal{E}}k_{\mathcal{U}}} := \sum_{(k_{\mathcal{D}},k_{\mathcal{R}})} h_{kk_1\cdots k_p} \cdot \prod_{j\in\mathcal{D}} h^{(j)}_{k_jk_{\mathcal{U}_j}}. \tag{6.20}$$

By (6.18) and Proposition 4.12 we have

$$\|H\|_{kk_P\to k_1k_Q} \le \prod_{q=1}^{s} \|H^{[v]}\|_{k_{A_v}} \prod_{u=1}^{t} \|R^{[u]}\|_{k_{B_u}} \cdot \|H^{\mathrm{sg}}\|_{kk_P\to k_1k_{\mathcal{E}}k_Q}, \tag{6.21}$$

so it suffices to bound the norm $\|H^{\mathrm{sg}}\|_{kk_P\to k_1k_{\mathcal{E}}k_Q}$ and the other norms $\|H^{[v]}\|_{k_{A_v}}$ and $\|R^{[u]}\|_{k_{B_u}}$.

By (6.20), the tensor $H^{\mathrm{sg}}$ is a semi-product the tensors $h$ and $h^{(j)}$ $(j \in \mathcal{D})$ in the sense of (4.32) in Proposition 4.12, so our strategy is to select these tensors in some particular order and apply Proposition 4.12. The selection algorithm is described as follows.

*Step 2: the selection algorithm for* $H^{\mathrm{sg}}$. First, by our choice of the set $\mathcal{D}$, there will be at least one element in the union of the sets $\mathcal{U}_j$ $(j \in \mathcal{D})$ that appears only once in these sets. Consider such an element $\mathfrak{l}$ with $N_{\mathfrak{l}}$ being the biggest. Denote this $\mathfrak{l}$ by $\mathfrak{l}_{c-1}$ and the $j \in \mathcal{D}$ such that $\mathfrak{l}_{c-1} \in \mathcal{U}_j$ by $n_{c-1}$, where $c-1 = |\mathcal{C}|-1 = |\mathcal{D}|$ and $\mathcal{C} = \mathcal{D}\cup\{1\}$. Next, there will be at least one element in the union of the sets $\mathcal{U}_j$ $(j \in \mathcal{D}\backslash\{n_{c-1}\})$ that appears only once in these sets. Consider such an element $\mathfrak{l} = \mathfrak{l}_{c-2}$ such that $N_{\mathfrak{l}}$ is the biggest, and suppose such $\mathfrak{l}_{c-2} \in \mathcal{U}_{j'}$, where $j' \in \mathcal{D}\backslash\{n_{c-1}\}$; we shall denote $n_{c-2} = j'$. Next there will be at least one element in the union of the sets $\mathcal{U}_j$ $(j \in \mathcal{D}\backslash\{n_{c-1}, n_{c-2}\})$ that appears only once in these sets, and so on. Repeating this process, we can label the elements of $\mathcal{D}$ as $n_1, n_2, \cdots, n_{c-1}$. Notice that by (6.10) and (6.14) and our selection algorithm, for each $1 \le y \le c-1$ we must have

$$\left| \sum_{z=1}^{y} \zeta_{n_z}k_{n_z} - m'_{n_y} \right| \lesssim \max((N_*)^{3\delta}, N_{\mathfrak{l}_y}) := M_{n_y}, \tag{6.22}$$

where $m'_{n_y} \in \mathbb{Z}^d$ is some fixed vector, otherwise the product in (6.20) will be zero. Let us provide some details to explain (6.22). When $y = 1$, $\mathfrak{l}_1 \in \mathcal{U}_{n_1}$ and $N_{\mathfrak{l}_1}$ is the biggest among all $\mathfrak{l}$ that appear only once in $\mathcal{U}_{n_1}$ (here these $\mathfrak{l}$'s are just all elements $\mathfrak{l} \in \mathcal{U}_{n_1}$). By (6.10), (6.13) and (6.14), setting $m'_{n_1} = \zeta_{n_1}m_{n_1}$, we then have

$$|\zeta_{n_1}k_{n_1} - m'_{n_1}| \le |\mathcal{U}_{n_1}|N_{\mathfrak{l}_1} + (N_*)^{3\delta}, \tag{6.23}$$

where we understand that $m_{n_1} = 0$ if $h^{(n_1)}$ has type $R0$ (same below), which implies (6.22) since $|\mathcal{U}_{n_1}| \le D$. When $y = 2$, $\mathfrak{l}_2 \in \mathcal{U}_{n_2}$ and $N_{\mathfrak{l}_2}$ is the biggest among all $\mathfrak{l}$ that appear only once in $\mathcal{U}_{n_1}\cup\mathcal{U}_{n_2}$ (i.e. all $\mathfrak{l} \in \mathcal{U}_{n_1}\Delta\mathcal{U}_{n_2}$), then by (6.10), (6.13) and (6.14), we have $|\zeta_{n_1}k_{n_1}+\zeta_{n_2}k_{n_2}-m'_{n_2}| \lesssim \max((N_*)^{3\delta}, N_{\mathfrak{l}_2})$ with $m'_{n_2} = \zeta_{n_1}m_{n_1} + \zeta_{n_2}m_{n_2}$, since the $k_{\mathfrak{l}}$ terms for $\mathfrak{l} \in \mathcal{U}_{n_1}\cap\mathcal{U}_{n_2}$ always cancel themselves thanks to our assumption about signs of paired elements. For the other $y$'s, (6.22) is obtained similarly. The above selection for $n_{c-1}, \cdots, n_1$ is designed to fit the hypothesis (4.24) in Proposition 4.9 via (6.22). Proposition 4.9 will be applied to $h$ later in *Step 4*.

Next, we divide these $n_y$ into two classes: first $n_{c-1}$ will be class $P$ (or $Q$) if $\mathfrak{l}_{c-1}$ which belongs to $\mathcal{U}$, is in $P$ (or $Q$). Next, if $\mathfrak{l}_{c-2}\in\mathcal{U}$, then $n_{c-2}$ will be class $P$ (or $Q$) if $\mathfrak{l}_{c-2}$ is in $P$ (or $Q$); if $\mathfrak{l}_{c-2}\in\mathcal{Q}$, then $\mathfrak{l}_{c-2}$ also belongs to $\mathcal{U}_j$ for some previously selected $j$ (here $j=n_{c-1}$), then $n_{c-2}$ will be same class as $j$. Then consider $\mathfrak{l}_{c-3}$, and so on. Repeating this process we can assign a class $P$ or $Q$ to each $n_y\,(1\leq y\leq c-1)$. Now we can apply Proposition 4.12 by arranging the tensors $h$ and $h^{(j)}\,(j\in\mathcal{D})$ in a particular order (which will correspond to the superscripts in the tensors in Proposition 4.12): first list all the $h^{(n_y)}$, where $n_y$ has class $Q$, in the *decreasing* order for $y$, then list $h$, then list all the $h^{(n_y)}$, where $n_y$ has class $P$, in the *increasing* order for $y$. By applying Proposition 4.12 to (6.20) with $(k_A,k_C,k_X,k_Y)=(kk_1k_{\mathcal{E}}k_{\mathcal{U}},k_{\mathcal{D}}k_{\mathcal{R}},kk_P,k_1k_{\mathcal{E}}k_Q)$ in the above order we then have

$$\|H^{\mathrm{sg}}\|_{kk_P\to k_1k_{\mathcal{E}}k_Q}\lesssim\|h\|_{kk_{P_0}\to k_{Q_0}}\prod_{j\text{ of class }Q}\|h^{(j)}\|_{k_jk_{P_j}\to k_{Q_j}}\prod_{j\text{ of class }P}\|h^{(j)}\|_{k_{Q_j}\to k_jk_{P_j}},\tag{6.24}$$

where $(P_0,Q_0)$ and $(P_j,Q_j)\,(j\in\mathcal{D})$ are sets defined according to Proposition 4.12, which are explained next. Next we analyze the individual factors on the right hand side of (6.24).

(1) For $j=n_y$ of class $Q$, by Proposition 4.12 and our algorithm, the set $Q_j$ will consist of all $\mathfrak{l}\in\mathcal{U}_{n_y}$ such that either $\mathfrak{l}\in Q$, or $\mathfrak{l}$ belongs to $\mathcal{U}_{n_{y'}}$ for some $y'>y$ with $n_{y'}$ of class $Q$. Furthermore $P_j=\mathcal{U}_j\backslash Q_j$. By the definition of classes, this implies that $\mathfrak{l}_y\in Q_j$. By (6.11) and (6.15) we then have

$$\|h^{(j)}\|_{k_jk_{P_j}\to k_{Q_j}}\lesssim\mathfrak{X}_j\cdot\prod_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}^{\beta}\cdot N_{\mathfrak{l}_y}^{-\beta}\lesssim\mathfrak{X}_j\cdot\prod_{\mathfrak{l}\in\mathcal{U}_j}N_{\mathfrak{l}}^{\beta}\cdot(N_*)^{C\delta}M_{n_y}^{-\beta};\tag{6.25}$$

moreover if $h^{(j)}$ has type $R0$ or $R0^+$, we gain an extra factor namely $\mathcal{Z}_{1,j}$ from (6.12).

(2) Similarly, for $j=n_y$ of class $P$, by Proposition 4.12 and our algorithm, the set $Q_j$ will consist of all $\mathfrak{l}\in\mathcal{U}_{n_y}$ such that either $\mathfrak{l}\in P$, or $\mathfrak{l}$ belongs to $\mathcal{U}_{n_{y'}}$ for some $y'>y$ with $n_{y'}$ of class $P$. Furthermore $P_j=\mathcal{U}_j\backslash Q_j$. By the definition of class, we also have $\mathfrak{l}_y\in Q_j$. Now by (6.11), (6.15), and using the duality of the operator norm $\|h^{(j)}\|_{k_{Q_j}\to k_jk_{P_j}}=\|h^{(j)}\|_{k_jk_{P_j}\to k_{Q_j}}$, we know that (6.25) is true (with the gain $\mathcal{Z}_{1,j}$ for types $R0$ and $R0^+$) also in this case.

(3) For the tensor $h$, by our algorithm we have that $P_0$ consists of all $j\in\mathcal{D}$ of class $P$, and $Q_0$ consists of all $j\in\mathcal{D}$ of class $Q$, as well as 1 and all $j\in\mathcal{E}$.

We illustrate the above algorithm with an explicit example. Suppose $p=7$, $\mathcal{D}=\{2,3,4,5\}$, $\mathcal{E}=\{6,7\}$ and $\mathcal{U}_2=\{\mathfrak{a},\mathfrak{b}\}$, $\mathcal{U}_3=\{\mathfrak{c},\mathfrak{d},\mathfrak{e}\}$, $\mathcal{U}_4=\{\mathfrak{e},\mathfrak{f},\mathfrak{g}\}$, $\mathcal{U}_5=\{\mathfrak{a},\mathfrak{f},\mathfrak{h}\}$ with $N_{\mathfrak{a}}\geq N_{\mathfrak{b}}\geq N_{\mathfrak{c}}\geq N_{\mathfrak{d}}\geq N_{\mathfrak{e}}\geq N_{\mathfrak{f}}\geq N_{\mathfrak{g}}\geq N_{\mathfrak{h}}$, then $\mathcal{U}=\{\mathfrak{b},\mathfrak{c},\mathfrak{d},\mathfrak{g},\mathfrak{h}\}$, $\mathcal{R}=\{\mathfrak{a},\mathfrak{e},\mathfrak{f}\}$ and

$$H^{\mathrm{sg}}_{kk_1k_6k_7k_{\mathcal{U}}}=\sum_{(k_2,\cdots,k_5)}\sum_{(k_{\mathfrak{a}},k_{\mathfrak{e}},k_{\mathfrak{f}})}h_{kk_1\cdots k_7}\cdot h^{(2)}_{k_2k_{\mathfrak{a}}k_{\mathfrak{b}}}\cdot h^{(3)}_{k_3k_{\mathfrak{c}}k_{\mathfrak{d}}k_{\mathfrak{e}}}\cdot h^{(4)}_{k_4k_{\mathfrak{e}}k_{\mathfrak{f}}k_{\mathfrak{g}}}\cdot h^{(5)}_{k_5k_{\mathfrak{a}}k_{\mathfrak{f}}k_{\mathfrak{h}}}.$$

Suppose $P=\{\mathfrak{b},\mathfrak{d}\}$ and $Q=\{\mathfrak{c},\mathfrak{g},\mathfrak{h}\}$ is a partition of all unpaired leaves. Then, by our algorithm we have $n_4=2$ since $N_{\mathfrak{b}}=\max_{\mathfrak{l}\in\{\mathfrak{b},\mathfrak{c},\mathfrak{d},\mathfrak{g},\mathfrak{h}\}}(N_{\mathfrak{l}})$ and $\mathfrak{b}\in\mathcal{U}_2$. Then $n_3=5$ since $N_{\mathfrak{a}}=\max_{\mathfrak{l}\in\{\mathfrak{a},\mathfrak{c},\mathfrak{d},\mathfrak{g},\mathfrak{h}\}}N_{\mathfrak{l}}$ and $\mathfrak{a}\in\mathcal{U}_5$. Similarly we have $n_2=3$ and $n_1=4$. Next, by definition $n_4=2$ and $n_3=5$ will have class $P$, while $n_2=3$ and $n_1=4$ will have class $Q$. Then we can apply Proposition 4.12 to (6.20) in the following order: $h^{(3)}\to h^{(4)}\to h\to h^{(5)}\to h^{(2)}$ and obtain that

$$\|H^{\mathrm{sg}}\|_{kk_{\mathfrak{b}}k_{\mathfrak{d}}\to k_1k_6k_7k_{\mathfrak{c}}k_{\mathfrak{g}}k_{\mathfrak{h}}}\lesssim\|h^{(3)}\|_{k_3k_{\mathfrak{d}}k_{\mathfrak{e}}\to k_{\mathfrak{c}}}\cdot\|h^{(4)}\|_{k_4k_{\mathfrak{f}}\to k_{\mathfrak{e}}k_{\mathfrak{g}}}\\ \times\|h\|_{kk_2k_5\to k_1k_3k_4k_6k_7}\cdot\|h^{(5)}\|_{k_{\mathfrak{a}}\to k_5k_{\mathfrak{f}}k_{\mathfrak{h}}}\cdot\|h^{(2)}\|_{k_{\mathfrak{b}}\to k_2k_{\mathfrak{a}}}.$$

Moreover we have the following inequalities (assume $m_j = 0$, and up to error $(N_*)^{3\delta}$):

$$|k_2 \pm k_3 \pm k_4 \pm k_5| \lesssim N_{\mathfrak{b}}, \quad |k_3 \pm k_4 \pm k_5| \lesssim N_{\mathfrak{a}}, \quad |k_3 \pm k_4| \lesssim N_{\mathfrak{c}}, \quad |k_4| \lesssim N_{\mathfrak{e}}.$$

*Step 3: the selection algorithms for $H^{[v]}$ and $R^{[u]}$.* Now we discuss the estimates for the norms $\|H^{[v]}\|_{k_{A_v}}$ and $\|R^{[u]}\|_{k_{B_u}}$. The basic idea is the same as before, but as each element in the sets $\mathcal{U}_j$ ($j \in A_v$ or $B_u$) occurs twice in these sets, we have to make some small adjustments. Let us first look at $A_v$. Let $a_v = |A_v|$, and first choose $\mathfrak{l} = \mathfrak{l}_{a_v}$ in the sets $\mathcal{U}_j$ ($j \in A_v$) such that $N_{\mathfrak{l}}$ is the smallest over all $\mathfrak{l}$ in the sets $\mathcal{U}_j$ ($j \in A_v$). We denote the $j$ such that $\mathfrak{l}_{a_v} \in \mathcal{U}_j$ by $\ell_v(a_v)$. Next, as in *Step 2* there will be at least one element in the union of the sets $\mathcal{U}_j$ ($j \in A_v \backslash \{\ell_v(a_v)\}$) that appears only once in these sets. Consider such an element $\mathfrak{l}$ with $N_{\mathfrak{l}}$ being the biggest. We shall denote this $\mathfrak{l}$ by $\mathfrak{l}_{a_v-1}$ and the $j \in A_v \backslash \{\ell_v(a_v)\}$ such that $\mathfrak{l}_{a_v-1} \in \mathcal{U}_j$ by $\ell_v(a_v-1)$. Then we can repeat the process in *Step 2* and label the elements of $A_v$ as $\ell_v(1), \cdots, \ell_v(a_v)$. Recall that $h^{(j)}$ has type $R0$ for all $j \in A_v$. By (6.10), in the same way as in *Step 2*, for $1 \leq y \leq a_v - 1$ we have

$$\left|\sum_{z=1}^{y} \zeta_{\ell_v(z)} k_{\ell_v(z)}\right| \lesssim N_{\mathfrak{l}_y} := M_{\ell_v(y)}, \tag{6.26}$$

as well as

$$\sum_{z=1}^{a_v} \zeta_{\ell_v(z)} k_{\ell_v(z)} = 0, \quad \sum_{z=1}^{a_v} \zeta_{\ell_v(z)} |k_{\ell_v(z)}|^2 = \widetilde{\Gamma}, \tag{6.27}$$

where $\widetilde{\Gamma} = \sum_{z=1}^{a_v} \zeta_{\ell_v(z)} \Gamma_{\ell_v(z)}$. Now we apply Proposition 4.12 to $H^{[v]}$ in (6.19) with $(A, C, X, Y) = (A_v, \bigcup\{\mathcal{U}_j : j \in A_v\}, A_v, \varnothing)$ by arranging the tensors $h^{(\ell_v(y))}$ in the decreasing order for $y$, hence

$$\|H^{[v]}\|_{k_{A_v}} \lesssim \prod_{j \in A_v} \|h^{(j)}\|_{k_j k_{P_j} \to k_{Q_j}}, \tag{6.28}$$

where $Q_j = \varnothing$ for $j = \ell_v(a_v)$, and for $j = \ell_v(y)$ with $y < a_v$, $Q_j$ consists of all $\mathfrak{l} \in \mathcal{U}_{\ell_v(y)}$ that belongs to $\mathcal{U}_{\ell_v(y')}$ for some $y' > y$, so in particular $\mathfrak{l}_y \in Q_j$. Furthermore $P_j = \mathcal{U}_j \backslash Q_j$ for all $j \in A_v$. By (6.11) and (6.12) we then get

$$\|h^{(j)}\|_{k_j k_{P_j} \to k_{Q_j}} \lesssim \mathfrak{X}_j \cdot \prod_{\mathfrak{l} \in \mathcal{U}_j} N_{\mathfrak{l}}^{\beta_1} \cdot M_{\ell_v(y)}^{-\beta_1} \cdot \mathcal{Z}_{1,j}, \quad \text{if } j = \ell_v(y) \quad (y < a_v), \tag{6.29}$$

$$\|h^{(j)}\|_{k_j k_{P_j} \to k_{Q_j}} \lesssim \mathfrak{X}_j \cdot \prod_{\mathfrak{l} \in \mathcal{U}_j} N_{\mathfrak{l}}^{\beta_1} \cdot (N_{\mathfrak{l}_{a_v}})^{\frac{d}{2} - \beta_1} \cdot \mathcal{Z}_{1,j}, \quad \text{if } j = \ell_v(a_v), \tag{6.30}$$

noticing also that $N_{\mathfrak{l}_{a_v}} \leq M_{\ell_v(y)}$ for any $1 \leq y \leq a_v - 1$ by our choice.

As for $B_u$, the argument is essentially the same as above. We choose $i_u(b_u)$ such that $h^{(j)}$ has type $R0^+$ or $R1$ for $j = i_u(b_u)$; then $i_u(y)$ for $y < b_u$ are chosen in the same manner as $\ell_v(y)$ above. In this case, (6.26) holds after translating by some fixed vector $m'_{i_u(y)}$, and with a loss of $(N_*)^{C\delta}$ due to the weaker bound (6.14). For (6.27) in this case we don't have the equation for $\sum_{z \leq b_u} \zeta_{i_u(z)} |k_{i_u(z)}|^2$, and the sum $\sum_{z \leq b_u} \zeta_{i_u(z)} k_{i_u(z)}$ only belongs to a ball of radius $(N_*)^{3\delta}$. However by losing a factor $(N_*)^{C\delta}$ in the operator bound for the tensor $h$ we may assume $\sum_{z \leq b_u} \zeta_{i_u(z)} k_{i_u(z)}$ is constant. Therefore, for the norms appearing in (6.28), instead of (6.29) and (6.30), we now have

$$\|h^{(j)}\|_{k_j k_{P_j} \to k_{Q_j}} \lesssim \mathfrak{X}_j \cdot (N_*)^{C\delta} \prod_{\mathfrak{l} \in \mathcal{U}_j} N_{\mathfrak{l}}^{\beta} \cdot M_{i_u(y)}^{-\beta}, \quad \text{if } j = i_u(y) \quad (y < b_u), \tag{6.31}$$

$$\|h^{(j)}\|_{k_j k_{P_j} \to k_{Q_j}} \lesssim \mathfrak{X}_j \cdot (N_*)^{C\delta} \prod_{\mathfrak{l} \in \mathcal{U}_j} N_{\mathfrak{l}}^{\beta}, \quad \text{if } j = i_u(b_u); \tag{6.32}$$

moreover we gain an extra factor $\mathcal{Z}_{1,j}$ in (6.31) and (6.32) if $h^{(j)}$ has type $R0^+$.

*Step 4: putting together.* We now come back to the estimate for $\|h\|_{kk_{P_0}\to k_{Q_0}}$ that appears in (6.24). By all the previous discussions, we may assume that in the support of $h = h_{kk_1\cdots k_p}$ we have the equalities and inequalities (6.16), (6.22), (6.26), (6.27), as well as the variants of (6.26) and (6.27) for $B_u$ (see *Step 3*), and that any pairing in $(k, k_1, \cdots, k_p)$ must be over-paired. All these allow us to apply Proposition 4.9 (unless we are in the exceptional case, namely $(d,p) = (1,7)$, and up to permutation $|A_v| = 2$ for $v = 1, 2$, $k_{\ell_1(1)} = k_{\ell_2(1)}$) to obtain the bound

$$\|h\|_{kk_{P_0}\to k_{Q_0}} \lesssim (N_*)^{\theta} \prod_{j=2}^{p} M_j^{\alpha_0} \prod_{v=1}^{s} (\min_{1\leq y<a_v} M_{\ell_v(y)})^{\alpha_0-\frac{d}{2}}. \tag{6.33}$$

Now combining (6.21), (6.24), (6.25) and (6.28)–(6.33), we conclude that

$$\prod_{\mathfrak{l}\in\mathcal{Q}} N_{\mathfrak{l}}^{-\alpha+4\varepsilon+\theta} \cdot \|H\|_{kk_P\to k_1k_Q} \lesssim \prod_{j=2}^{p} \mathfrak{X}_j \prod_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}^{\beta} \cdot (N_*)^{C\delta} (\max_{\mathfrak{l}\in\mathcal{Q}} N_{\mathfrak{l}})^{-4p\varepsilon} (\max_{2\leq j\leq p} M_j)^{-4p\varepsilon} \prod_{j}^{(0,0^+)} \mathcal{Z}_{1,j}, \tag{6.34}$$

where $\prod_j^{(0,0^+)}$ is defined as in (6.17). By definition of $M_j$ we have

$$\max_{2\leq j\leq p} M_j \geq \max_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}};$$

using also (6.34) and the fact that $\mathcal{U} \cup \mathcal{Q} = \mathcal{U}_2 \cup \cdots \cup \mathcal{U}_p$, we then have

$$\prod_{\mathfrak{l}\in\mathcal{Q}} N_{\mathfrak{l}}^{-\alpha+4\varepsilon+\theta} \cdot \|H\|_{kk_P\to k_1k_Q} \lesssim \prod_{j=2}^{p} \mathfrak{X}_j \prod_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}^{\beta} \cdot (N_*)^{C\delta} (\max_{\mathfrak{l}\in\mathcal{U}_j, 2\leq j\leq p} N_{\mathfrak{l}})^{-4p\varepsilon} \prod_{j}^{(0,0^+)} \mathcal{Z}_{1,j}. \tag{6.35}$$

The factor $\prod_j^{(0,0^+)} N_j^{-2\varepsilon}$ in (6.17) will be provided by separating a square root of the last two factors in the right hand side of (6.35). To gain the other factor $(N_*)^{-\varepsilon^3}$ in (6.17), we consider two cases. If $\max\{N_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{U}_j, 2 \leq j \leq p\} \geq (N_*)^{\varepsilon^2}$, then the factor $(\max_{\mathfrak{l}\in\mathcal{U}_j,2\leq j\leq p} N_{\mathfrak{l}})^{-2p\varepsilon}$ in the other square root is bounded by $(N_*)^{-C\delta-\varepsilon^3}$, hence (6.17) is proved. Otherwise, let $2 \leq j \leq p$ be such that $N_j = N_*$, then $N_{\mathfrak{l}} \ll N_j^{\varepsilon^2}$ for all $\mathfrak{l} \in \mathcal{U}_j$. If $h^{(j)}$ has type $R0$ or $R0^+$, the factor $\mathcal{Z}_{1,j}^{1/2} = N_j^{-2\varepsilon}$ in the other square root is bounded by $(N_*)^{-C\delta-\varepsilon^3}$. If $h^{(j)}$ has type $R1$, by using the bound of $\|h^{(j)}\|_{k_jk_{\mathcal{U}_j}}$ (the second bound in (6.15)) and that $N_j^{-\varepsilon/2} \leq (\max_{\mathfrak{l}\in Q_j} N_{\mathfrak{l}})^{-\beta}$, we have

$$\|h^{(j)}\|_{k_jk_{P_j}\to k_{Q_j}} \leq \mathfrak{X}_j \cdot N_j^{-\varepsilon/2} \prod_{\mathfrak{l}\in\mathcal{U}_j} N_{\mathfrak{l}}^{\beta} \cdot \big(\max_{\mathfrak{l}\in Q_j} N_{\mathfrak{l}}\big)^{-\beta} \tag{6.36}$$

for any partition $(P_j, Q_j)$ of $\mathcal{U}_j$. Then we gain an extra factor $N_j^{-\varepsilon/2}$ (which is less than $(N_*)^{-C\delta-\varepsilon^3}$) by using (6.36) instead of (6.15) in (6.25), (6.31) or (6.32), and hence (6.17) is proved.

Finally, in the exceptional case mentioned above, we may assume (up to permutation) that $A_1 = \{2,3\}$ and $A_2 = \{4,5\}$, so $(d,p) = (1,7)$ and $k_2 = k_3 = k_4 = k_5 := k_*$ by the setting of $A_v$ in *Step 1*. Here we may fix and sum in $k_*$, while for fixed $k_*$ the corresponding part of the tensor $H$ can be bounded as above, with $h^{(j)}$ $(2 \leq j \leq 5)$ measured in the norm

$$\sup_{k_j} \|h^{(j)}\|_{k_{\mathcal{U}_j}} \leq \|h^{(j)}\|_{k_j\to k_{\mathcal{U}_j}}.$$

One can check that the power gain coming from using these norms is enough to cancel the summation in $k$, and the rest of the proof goes just like above. The cases when $p$ is replaced by odd

$3 \leq q \leq p$, or when $h$ satisfies (4.27) can be proved in the same way, since Proposition 4.9 works equally well. $\square$

**Proposition 6.3** (Selection algorithm: Case II)**.** *Consider the same setting as in Proposition 6.2. Here we assume that each $h^{(j)}$ has type $R0$ or $R0^+$ (in the sense of Proposition 6.2), but in both cases the factor $\mathcal{Z}_{1,j}$ in (6.11) is replaced by 1 (the factor $\mathcal{Z}_{0,j}$ remains the same). Then we have*

$$\prod_{\mathfrak{l}\in\mathcal{Q}} N_{\mathfrak{l}}^{-\alpha+8\varepsilon+\theta} \cdot \|H\|_{k k_P \to k_1 k_Q} \lesssim \prod_{j=2}^{p} \mathfrak{X}_j \cdot \prod_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}^{\beta_1} \cdot \big(\max_{2\leq j\leq p} \max_{\mathfrak{l}\in\mathcal{U}_j} N_{\mathfrak{l}}\big)^{-4p\varepsilon}. \tag{6.37}$$

*The same holds if $p$ is replaced by odd $3 \leq q \leq p$, without changing the power $4p\varepsilon$ in (6.37).*

*Proof.* The proof is the same as Proposition 6.2, with the following adjustments. First due to the absence of type $R1$ tensors, we will not lose any $(N_*)^{C\delta}$ factor in the proof process; second, we do not gain any extra factor as in the proof of Proposition 6.2, since $\mathcal{Z}_{1,j}$ has been replaced by 1. With these, we are still able to gain a factor

$$\big(\max_{\mathfrak{l}\in\mathcal{Q}} N_{\mathfrak{l}}\big)^{-4p\varepsilon}\big(\max_{2\leq j\leq p} M_j\big)^{-4p\varepsilon}$$

as in (6.34), which implies (6.37) in the same way as in the proof of Proposition 6.2. $\square$

6.3. **Merging estimates.** Finally we prove the general merging bounds, Propositions 6.4–6.6.

**Proposition 6.4** (Merged tensor bounds: Case I)**.** *Let $3 \leq q \leq p$ be odd, $0 \leq r \leq q$, and let $\mathcal{S}_j\,(1 \leq j \leq r)$ be regular plants with frequency $N(\mathcal{S}_j) = N_j \leq M$ and $|\mathcal{S}_j| \leq D$. Fix $N_j \leq M/2$ for $r+1 \leq j \leq q$ and $\zeta_j \in \{\pm\}$ for $1 \leq j \leq q$, and assume that $\sum_{j=1}^q \zeta_j = 1$. Denote $\mathscr{B} = (M, q, r, \zeta_1, \cdots, \zeta_q, N_1, \cdots, N_q)$. Let $h = h_{kk_1\cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q)$ be a constant tensor[1] supported in the set*

$$\begin{aligned} &\langle k\rangle \leq M, \qquad \langle k_j\rangle \leq N_j\,(1\leq j\leq q), \quad \langle \lambda_j\rangle \leq 2M^{\kappa^2}\,(r+1\leq j\leq q),\\ &k=\sum_{j=1}^q \zeta_j k_j, \quad \Big||k|^2-\sum_{j=1}^q \zeta_j|k_j|^2+\sum_{j=r+1}^q \zeta_j\lambda_j+\widetilde{\Gamma}\Big|\lesssim 1, \end{aligned} \tag{6.38}$$

*where $\widetilde{\Gamma} \in \mathbb{Z}$ is fixed. Assume that*

$$|h| + |\partial_{\lambda_j} h| \lesssim \tau^{-\theta}, \quad r+1 \leq j \leq q, \tag{6.39}$$

*and that any pairing in $(k, k_1, \cdots, k_q)$ must be over-paired. Now let $h^{(j)} = h^{(j)}_{k_j k_{\mathcal{U}_j}}(k_{\mathcal{V}_j}, \lambda_{\mathcal{V}_j})$ be an $\mathcal{S}_j$-tensor for $1 \leq j \leq r$, $\mathscr{O}$ be as in Definition 3.6, and let*

$$\begin{aligned} \mathcal{S} = (\mathcal{L}, \mathcal{V}, \mathcal{Y}) &= \mathtt{Trim}(\mathtt{Merge}(\mathcal{S}_1, \cdots, \mathcal{S}_r, \mathscr{B}, \mathscr{O}), M^\delta),\\ H &= \mathtt{Trim}(\mathtt{Merge}(h^{(1)}, \cdots, h^{(r)}, h, \mathscr{B}, \mathscr{O}), M^\delta). \end{aligned} \tag{6.40}$$

*Let $N_* := \max(N_2, \cdots, N_q)$, assume $N_* \geq M^\delta$, and $\Upsilon$ be a factor such that*

$$\Upsilon \leq \tau^{-\theta}; \qquad \Upsilon \leq \tau^{-\theta} M^{-40dp\varepsilon} \text{ if } \max_{1\leq j\leq q} N_j \leq (50dp)^{-1} M. \tag{6.41}$$

*We assume that the tensor $h^{(1)}$ is $\mathcal{B}_{M^{[\delta]}}$ measurable, and $N_{\mathfrak{l}} \geq M^\delta$ for $\mathfrak{l} \in \mathcal{L}_1$; for $2 \leq j \leq r$, the tensor $h^{(j)}$ is $\mathcal{B}_{(N_*)^{[\delta]}}$ measurable, and $N_{\mathfrak{l}} \geq (N_*)^\delta$ for $\mathfrak{l} \in \mathcal{L}_j$. Furthermore we assume that for $1 \leq j \leq r$, $\mathcal{S}_j$ and $h^{(j)}$ have one of the following two types:*

[1] i.e. which does not depend on $\omega$.

*(1) Type 0: where $\mathcal{V}_j=\varnothing$, and in the support of $h^{(j)}$ we have*

$$\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}=1,\quad k_j=\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}},\quad |k_j|^2-\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}|k_{\mathfrak{l}}|^2=\Gamma_j,\tag{6.42}$$

*where $\Gamma_j\in\mathbb{Z}$ is fixed, and $h^{(j)}$ satisfies the bound*

$$\|h^{(j)}\|_{k_jk_{B_j}\to k_{C_j}}\lesssim\mathfrak{X}_j\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B_j\cup C_j}N_{\mathfrak{l}}^{\beta_1}\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-8\varepsilon}\prod_{\mathfrak{l}\in E_j}N_{\mathfrak{l}}^{4\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}\cdot\mathcal{X}_{0,j}\mathcal{X}_{1,j},\tag{6.43}$$

*for any subpartition $(B_j,C_j)$ of $\mathcal{U}_j$, where $E_j=\mathcal{U}_j\backslash(B_j\cup C_j)$, and $\mathcal{X}_{0,j}$ and $\mathcal{X}_{1,j}$ are defined as in (5.33) but are associated with $\mathcal{S}_j$ and $(B_j,C_j)$ instead.*

*(2) Type 1: where $h^{(j)}$ satisfies the bounds (with $B_j,C_j,E_j$ same as (1))*

$$\|h^{(j)}\|_{X^{-b_0}_{\mathcal{V}_j}[k_jk_{B_j}\to k_{C_j}]}\lesssim\mathfrak{X}_j\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B_j\cup C_j}N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E_j}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_j}N_{\mathfrak{f}}^{d}\cdot\big(\max_{\mathfrak{l}\in C_j}N_{\mathfrak{l}}\big)^{-\beta},\quad\text{if } C_j\neq\varnothing,\tag{6.44}$$

$$\|h^{(j)}\|_{X^{-b_0}_{\mathcal{V}_j}[k_jk_{B_j}]}\lesssim\mathfrak{X}_j\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B_j}N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E_j}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_j}N_{\mathfrak{f}}^{d}\cdot N_j^{-\varepsilon},\quad\text{if } C_j=\varnothing;\tag{6.45}$$

*and we also assume for $C_j=\varnothing$ the localization bound*

$$\bigg\|\bigg(1+\frac{1}{R_j}\Big|k_j-\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell_j\Big|\bigg)^{\kappa}h^{(j)}\bigg\|_{X^{-b_0}_{\mathcal{V}_j}[k_jk_{B_j}]}\lesssim\mathfrak{X}_j\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B_j}N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E_j}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_j}N_{\mathfrak{f}}^{d},\tag{6.46}$$

*where $R_1=M^{2\delta}$, $R_j=(N_*)^{2\delta}$ for $j\geq 2$ and $\ell_j=\sum_{\mathfrak{f}\in\mathcal{V}_j}\zeta_{\mathfrak{f}}k_{\mathfrak{f}}$, and the $\lambda_{\mathcal{V}_j}$-derivative bound (for $C_j=E_j=\varnothing$),*

$$\|\partial_{\lambda_{\mathcal{V}_j}}h^{(j)}\|_{X^{-b_0}_{\mathcal{V}_j}[k_jk_{\mathcal{U}_j}]}\lesssim\mathfrak{X}_j\cdot\tau^{-\theta}\exp[(\log N_j)^5+|\mathcal{S}_j|(\log N_j)^3].\tag{6.47}$$

*Finally fix a subpartition $(B,C)$ of $\mathcal{U}$, and let $E=\mathcal{U}\backslash(B\cup C)$. Then, under all of the above assumptions, we have the following results, where we denote*

$$\mathfrak{Y}=\prod_{j=1}^{r}\mathfrak{X}_j\cdot\tau^{-\theta}M^{\theta}(N_*)^{-2\varepsilon^4}:\tag{6.48}$$

*(i) If $C\neq\varnothing$, and assume that $\max\{N_{\mathfrak{l}}:\mathfrak{l}\in C\cap\mathcal{U}_1\}\sim\max\{N_{\mathfrak{l}}:\mathfrak{l}\in C\}$, then $\tau^{-1}M$-certainly, the tensor $H=H_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ satisfies the bounds*

$$\sqrt{\Upsilon}\cdot\|H\|_{X^{-b_0}_{\mathcal{V}}[kk_B\to k_C]}\lesssim\mathfrak{Y}\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B\cup C}N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}}N_{\mathfrak{f}}^{d}\cdot\big(\max_{\mathfrak{l}\in C}N_{\mathfrak{l}}\big)^{-\beta}.\tag{6.49}$$

*(ii) If $C=\varnothing$, and assume that $h^{(1)}$ has type 1, and $N_1\gtrsim N_j$ for all $1\leq j\leq r$ such that $h^{(j)}$ has type 1, then $\tau^{-1}M$-certainly we have*

$$\sqrt{\Upsilon}\cdot\|H\|_{X^{-b_0}_{\mathcal{V}}[kk_B]}\lesssim\mathfrak{Y}\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B}N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}}N_{\mathfrak{f}}^{d}\cdot M^{-\varepsilon}.\tag{6.50}$$

*(iii) If $C=\varnothing$, and assume that $h^{(1)}$ has type 1. Moreover, assume we restrict the tensors $H_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ and $h^{(1)}_{k_1k_{\mathcal{U}_1}}(k_{\mathcal{V}_1},\lambda_{\mathcal{V}_1})$ to the sets*

$$1+\frac{1}{M^{2\delta}}\Big|k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}^*k_{\mathfrak{l}}-\ell\Big|\sim K,\quad 1+\frac{1}{M^{2\delta}}\Big|k_1-\sum_{\mathfrak{l}\in\mathcal{U}_1}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell_1\Big|\sim K_1,\tag{6.51}$$

*where $K \lesssim K_1$ are two dyadic numbers, $\zeta_{\mathfrak{n}}^*$ represents the sign of $\mathfrak{n}$ in $\mathcal{S}$, $\ell = \sum_{\mathfrak{f}\in\mathcal{V}} \zeta_{\mathfrak{f}}^* k_{\mathfrak{f}}$, and $\ell_1 = \sum_{\mathfrak{f}\in\mathcal{V}_1} \zeta_{\mathfrak{f}} k_{\mathfrak{f}}$ as above. Then $\tau^{-1}M$-certainly we have*

$$\sqrt{\Upsilon}\cdot\Big\|\Big(1+\frac{1}{M^{2\delta}}\big|k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}^* k_{\mathfrak{l}}-\ell\big|\Big)^{\kappa} H\Big\|_{X_{\mathcal{V}}^{-b_0}[kk_B]} \lesssim \mathfrak{Y}\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B} N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}} N_{\mathfrak{f}}^{d}. \tag{6.52}$$

*Proof.* We would like to apply Proposition 6.2. The technical difficulty before doing so is two-folded. On the one hand, we must separate the tensor $h^{(1)}$ from the tensor $H$ defined in (6.40). On the other hand, since there are no over-pairings in the statement of Proposition 6.2 we must remove these from $H$. Once we have dealt with this technical difficulty, the heart of the matter lies in implementing Proposition 6.2 to obtain the desired bounds above.

In the proof below we will mainly focus on (6.49). The proof of (6.50) will be analogous, and we will only point out the necessary changes in the course of the proof. Moreover, in proving (6.50) we will only use the bound (6.45) for the tensor $h^{(1)}$, so (6.52) will follow from the same arguments as (6.50) once we use (6.46) instead of (6.45), in view of (6.51) and $K \lesssim K_1$, $N_1 \leq M$.

The proof will proceed in four steps. In *Step 1*, we reduce the desired estimates for the tensor $H$ to those for the tensor $H^\circ$ defined in (6.53) below. In *Step 2*, we remove and estimate the over-pairings, and reduce the desired estimates for $H^\circ$ to those for $(H^\circ)^\dagger$ defined in (6.63). In *Step 3*, we first single out $h^{(1)}$ in $(H^\circ)^\dagger$ and in turn apply Propositions 4.11 and 4.14, then remove and estimate the over-pairings as in *Step 2*, to reduce the desired estimates for $(H^\circ)^\dagger$ to those for $\mathcal{H}^\dagger$ defined in (6.72). Finally in *Step 4*, we implement Proposition 6.2 and conclude the proof.

*Step 1: pre-processing.* Define

$$\widetilde{\mathcal{S}} = (\widetilde{\mathcal{L}},\widetilde{\mathcal{V}},\widetilde{\mathcal{Y}}) = \mathtt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_r,\mathscr{B},\mathscr{O}),$$

$$\widetilde{H} = \widetilde{H}_{kk_{\widetilde{\mathcal{U}}}}(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}}) = \mathtt{Merge}(h^{(1)},\cdots,h^{(r)},h,\mathscr{B},\mathscr{O}),$$

noticing that $\widetilde{\mathcal{V}} = \mathcal{V}_1\cup\cdots\cup\mathcal{V}_r\cup\{r+1,\cdots,q\}$, and define

$$H^\circ = (H^\circ)_{kk_{\mathcal{U}}}(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}}) = \sum_{k_{\widetilde{\mathcal{U}}\setminus\mathcal{U}}} \widetilde{H}_{kk_{\widetilde{\mathcal{U}}}}(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})\prod_{\mathfrak{l}\in\widetilde{\mathcal{U}}\setminus\mathcal{U}} (f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}^*}, \tag{6.53}$$

then we have

$$H_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}}) = \sum_{k_{\widetilde{\mathcal{V}}\setminus\mathcal{V}}}\int \mathrm{d}\lambda_{\widetilde{\mathcal{V}}\setminus\mathcal{V}}\cdot(H^\circ)_{kk_{\mathcal{U}}}(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})\prod_{\mathfrak{f}\in\widetilde{\mathcal{V}}\setminus\mathcal{V}} (\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}^*}(\lambda_{\mathfrak{f}}).$$

By the same proof as part (1) of Proposition 6.1, for any $X_{\mathcal{V}}^{-b_0}[\cdots]$ norm we have

$$\|H\|_{X_{\mathcal{V}}^{-b_0}[\cdots]} \lesssim \|H^\circ\|_{X_{\widetilde{\mathcal{V}}}^{-b_0}[\cdots]}\cdot\prod_{\mathfrak{f}\in\widetilde{\mathcal{V}}\setminus\mathcal{V}} N_{\mathfrak{f}}^{-D_1}, \tag{6.54}$$

as well as for the weighted norm in (6.52). Therefore, it suffices to estimate the $X_{\widetilde{\mathcal{V}}}^{-b_0}[\cdots]$ norms (as well as the weighted ones) for $H^\circ$.

For each $1\leq j\leq r$, if $h^{(j)}$ has type 1, we can define $\mathfrak{X}_j^* = \mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$ to be the smallest positive number such that the bounds (6.44)–(6.47) are true for this choice of $(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$ with $X_{\mathcal{V}_j}^{-b_0}$ in the norms removed (for example $X_{\mathcal{V}_j}^{-b_0}[k_jk_{B_j}\to k_{C_j}]$ replaced by $k_jk_{B_j}\to k_{C_j}$), and with $\mathfrak{X}_j$

replaced by $\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$; for example one of the inequalities satisfied by $\mathfrak{X}_j^*=\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$, which corresponds to (6.44), would be

$$\|h^{(j)}\|_{k_jk_{B_j}\to k_{C_j}}\lesssim\mathfrak{X}_j^*\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B_j\cup C_j}N_\mathfrak{l}^\beta\prod_{\mathfrak{l}\in\mathcal{P}_j}N_\mathfrak{l}^{-4\varepsilon}\prod_{\mathfrak{l}\in E_j}N_\mathfrak{l}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_\mathfrak{p}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_j}N_\mathfrak{f}^d\cdot\big(\max_{\mathfrak{l}\in C_j}N_\mathfrak{l}\big)^{-\beta}\tag{6.55}$$

with $C_j\neq\varnothing$, for fixed $(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$. If $h^{(j)}$ has type 0, we simply define $\mathfrak{X}_j^*=\mathfrak{X}_j$. Then we have

$$\sum_{k_{\mathcal{V}_j}}\int\mathrm{d}\lambda_{\mathcal{V}_j}\cdot\prod_{\mathfrak{f}\in\mathcal{V}_j}\langle\lambda_\mathfrak{f}\rangle^{-2b_0}\cdot\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})^2\lesssim\mathfrak{X}_j^2\tag{6.56}$$

for type 1 tensors. When $(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})$ is fixed, which means $(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$ are fixed for $1\leq j\leq r$ and $(k_j,\lambda_j)$ are fixed for $r+1\leq j\leq q$, we shall view $h=h_{kk_1\cdots k_r}$ as a tensor depending on $(k,k_1,\cdots,k_r)$; for $1\leq j\leq r$ we shall view $h^{(j)}=h^{(j)}_{k_jk_{\mathcal{U}_j}}$ as a tensor depending on $(k_j,k_{\mathcal{U}_j})$.

With these reductions, we can view $\widetilde{H}=\widetilde{H}_{kk_{\widetilde{\mathcal{U}}}}$ as a tensor depending on $(k,k_{\widetilde{\mathcal{U}}})$ and $H^\circ=H^\circ_{kk_\mathcal{U}}$ depending on $(k,k_\mathcal{U})$, namely

$$\widetilde{H}_{kk_{\widetilde{\mathcal{U}}}}=\prod_{\mathfrak{l},\mathfrak{l}'}^{(1)}\mathbf{1}_{k_\mathfrak{l}=k_{\mathfrak{l}'}}\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}\mathbf{1}_{k_\mathfrak{l}\neq k_{\mathfrak{l}'}}\cdot\sum_{(k_1,\cdots,k_r)}^{(3)}h_{kk_1\cdots k_r}\sum_{k_\mathcal{Q}}\prod_{\mathfrak{l}\in\mathcal{Q}}\Delta_{N_\mathfrak{l}}\gamma_{k_\mathfrak{l}}\prod_{j=1}^r[h^{(j)}_{k_jk_{\mathcal{U}_j}}]^{\zeta_j},\tag{6.57}$$

$$(H^\circ)_{kk_\mathcal{U}}=\sum_{k_{\widetilde{\mathcal{U}}\setminus\mathcal{U}}}\widetilde{H}_{kk_{\widetilde{\mathcal{U}}}}\prod_{\mathfrak{l}\in\widetilde{\mathcal{U}}\setminus\mathcal{U}}(f_{N_\mathfrak{l}})_{k_\mathfrak{l}}^{\zeta_\mathfrak{l}^*},\tag{6.58}$$

where in (6.57), the sum and products are taken in the same way as (3.11) when merging the tensors $(h^{(1)},\cdots,h^{(r)})$ via $(h,\mathscr{B},\mathscr{O})$. Similarly $\mathcal{Q}$ is defined as in Definition 3.6. Also note that $\mathcal{U}=\{\mathfrak{l}\in\widetilde{\mathcal{U}}:N_\mathfrak{l}\geq M^\delta\}$.

Our goal is to prove that, for fixed values of $(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})$ defined in (6.53), the tensor $H^\circ=H^\circ_{kk_\mathcal{U}}$ satisfies (6.49), (6.50) and (6.52) $\tau^{-1}M$-certainly, but with the following three adjustments: (a) we remove the $X_\mathcal{V}^{-b_0}$ parts in the norms (for example $X_\mathcal{V}^{-b_0}[kk_B\to k_C]$ is replaced by $kk_B\to k_C$), and multiply the left hand sides by the extra factor $\prod_{j=r+1}^q N_j^{-d/2}$; (b) the set $\mathcal{V}$ in the factors $\prod_{\mathfrak{f}\in\mathcal{V}}N_\mathfrak{f}^d$ on the right hand sides is replaced by $\mathcal{V}_1\cup\cdots\cup\mathcal{V}_r$ and $\mathcal{Y}$ is replaced by $\widetilde{\mathcal{Y}}$; and (c) the $\mathfrak{X}_j$ in the definition (6.48) of $\mathfrak{Y}$ is replaced by $\mathfrak{X}_j^*$. For example, the analogue of (6.49) with the above adjustments (a)–(c) amounts to showing:

$$\begin{aligned}\sqrt{\Upsilon}\cdot\prod_{j=r+1}^q N_j^{-d/2}\,\|H^\circ\|_{kk_B\to k_C}\lesssim&\,\tau^{-\theta}\prod_{\mathfrak{l}\in B\cup C}N_\mathfrak{l}^\beta\prod_{\mathfrak{l}\in\mathcal{P}}N_\mathfrak{l}^{-4\varepsilon}\prod_{\mathfrak{l}\in E}N_\mathfrak{l}^{8\varepsilon}\big(\max_{\mathfrak{l}\in C}N_\mathfrak{l}\big)^{-\beta}\\&\times\prod_{\mathfrak{p}\in\widetilde{\mathcal{Y}}}N_\mathfrak{p}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_1\cup\cdots\cup\mathcal{V}_r}N_\mathfrak{f}^d\cdot\left(\prod_{j=1}^r\mathfrak{X}_j^*\cdot\tau^{-\theta}M^\theta(N_*)^{-2\varepsilon^4}\right),\end{aligned}\tag{6.59}$$

where $\prod_{j=1}^r\mathfrak{X}_j^*\cdot\tau^{-\theta}M^\theta(N_*)^{-2\varepsilon^4}$ is $\mathfrak{Y}$ in (6.48) with replacements of $\mathfrak{X}_j$ by $\mathfrak{X}_j^*$. The corresponding analogues of (6.50) and (6.52) with the adjustments (a)–(c) are similar.

If we can prove (6.59) for a fixed choice of $(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})$ in $H^\circ$, then by applying the *meshing argument* in the same way as we did in the proof of Proposition 6.1, using (6.39) and (6.47), we can reduce to the case of at most $\exp(\kappa(\log M)^6)$ choices for $(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})$, hence $\tau^{-1}M$-certainly we may assume that (6.59) holds for $H^\circ_{kk_\mathcal{U}}$ for all choices of $(k_{\widetilde{\mathcal{V}}},\lambda_{\widetilde{\mathcal{V}}})$. Since $\widetilde{\mathcal{V}}=\mathcal{V}_1\cup\cdots\cup\mathcal{V}_r\cup\{r+1,\cdots,q\}$ and the definition of $X_{\widetilde{\mathcal{V}}}^{-b_0}[\cdots]$ involves summing and integrating over $(k_{\widetilde{\mathcal{V}}}, \lambda_{\widetilde{\mathcal{V}}})$, then $\tau^{-1}M$-certainly we have the following estimate

$$\sqrt{\Upsilon}\cdot\|H^\circ\|_{X_{\widetilde{\mathcal{V}}}^{-b_0}[kk_B\to k_C]} \lesssim \mathfrak{Y}\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B\cup C} N_{\mathfrak{l}}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\widetilde{\mathcal{Y}}} N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\widetilde{\mathcal{V}}} N_{\mathfrak{f}}^{d}\cdot\big(\max_{\mathfrak{l}\in C} N_{\mathfrak{l}}\big)^{-\beta}, \quad (6.60)$$

which follows from (6.59) and

$$\Big(\sum_{k_{\widetilde{\mathcal{V}}}}\int \mathrm{d}\lambda_{\widetilde{\mathcal{V}}}\cdot\prod_{\mathfrak{f}\in\widetilde{\mathcal{V}}}\langle\lambda_{\mathfrak{f}}\rangle^{-2b_0}\prod_{j=1}^{r}\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})^2\Big)^{1/2} \lesssim \prod_{j=1}^{r}\mathfrak{X}_j\prod_{j=r+1}^{q} N_j^{d/2}. \quad (6.61)$$

Note that (6.61) follows from taking the tensor product of (6.56) for $1\le j\le r$ and summing and integrating over $(k_j,\lambda_j)$ for $r+1\le j\le q$. Finally, the desired bound (6.49) for $H_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ follows from (6.60) and (6.54) in view of the $N_{\mathfrak{f}}^{-D_1}$ powers on the right hand side of (6.54). The desired bounds (6.50) and (6.52) for $H_{kk_{\mathcal{U}}}(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ can be obtained in a similar way.

*Step 2: removing over-pairings.* From now on we will fix the value of $(k_{\widetilde{\mathcal{V}}}, \lambda_{\widetilde{\mathcal{V}}})$ and reduce to the setting of the tensor $\widetilde{H}$ in (6.57) and the tensor $H^\circ$ in (6.58).

We will first focus on (6.49), fix a subpartition $(B,C)$ of $\mathcal{U}$ with $C\neq\varnothing$, and denote $E=\mathcal{U}\backslash(B\cup C)$. Recall that $\mathscr{O}=\{\mathcal{A}_1,\cdots,\mathcal{A}_m\}$ is the collection of all pairings and over-pairings (see Definition 3.6), and notice that the frequencies $N_{\mathfrak{l}}$ are the same for $\mathfrak{l}\in\mathcal{A}_i$. Without loss of generality, we may assume the frequency $N_{\mathfrak{l}}$ for $\mathfrak{l}\in\mathcal{A}_i$ is $\ge M^\delta$ for all $1\le i\le v$ (where $1\le v\le m$), and $\le M^{[\delta]}$ for $v+1\le i\le m$. In particular, we have that $\mathcal{A}_i\cap\mathcal{U}=\varnothing$ and that $\mathcal{A}_i\cap\mathcal{U}_1=\varnothing$ for $i\ge v+1$ (since $N_{\mathfrak{l}}\ge M^\delta$ for $\mathfrak{l}\in\mathcal{U}$ and $\mathfrak{l}\in\mathcal{U}_1$).

Next, in (6.57), the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}$ has two parts, namely $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,\ge)}$ containing $(\mathfrak{l},\mathfrak{l}')$ such that $N_{\mathfrak{l}}=N_{\mathfrak{l}'}\ge M^\delta$, and $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}$ containing $(\mathfrak{l},\mathfrak{l}')$ such that $N_{\mathfrak{l}}=N_{\mathfrak{l}'}<M^\delta$. In the proof below we will slightly modify (6.57) by changing $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$ into $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$. This will be necessary in order to separate the tensor $h^{(1)}$ from the rest later in the proof, and will not cause a problem, since the original $\widetilde{H}$ equals the modified $\widetilde{H}$ multiplied by $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,\ge)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$, and the original $H^\circ$ equals the modified $H^\circ$ multiplied by $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,\ge)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$, which is a bounded operation due to Lemma 4.10. The reason we need to keep the factors in the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$ is to guarantee the no-pairing assumption required to apply Proposition 4.14 later in *Step 3*.

Recall that when $C\neq\varnothing$ we have $\max\{N_{\mathfrak{l}}:\mathfrak{l}\in C\cap\mathcal{U}_1\}\sim\max\{N_{\mathfrak{l}}:\mathfrak{l}\in C\}$ (see part (i) in the assumption). Denote the particular $\mathfrak{l}\in C\cap\mathcal{U}_1$ where the maximum is attained by $\mathfrak{l}_{\mathrm{top}}$. In this step we shall fix the values of $k_{\mathfrak{l}}$ for $\mathfrak{l}\in E$ and for $\mathfrak{l}\in\mathcal{A}_i$, where $1\le i\le v$ and $|\mathcal{A}_i|\ge 3$. These $\mathfrak{l}$'s are divided into groups according to the pairing and over-pairing relations, and there are four possible cases:

*Case 1*: For each $\mathfrak{l}\in E$ that does not belong to any $\mathcal{A}_i$, we form a group with only one element $\mathfrak{l}$. Each of these $\mathfrak{l}$ belongs to a unique $\mathcal{U}_j$ for some $1\le j\le r$.

*Case 2*: For each $i$ such that $\mathcal{A}_i\cap E\neq\varnothing$, we form a group containing all elements of this $\mathcal{A}_i$. Define[1] $y_i$ and $z_i$ such that $|\mathcal{A}_i\cap\mathcal{Q}|=2y_i$ and $|\mathcal{A}_i\cap\widetilde{\mathcal{U}}|=z_i$. We then have $|\mathcal{A}_i|=2y_i+z_i$.

*Case 3*: For each $i$ such that $\mathcal{A}_i\cap E=\varnothing$ and $\mathcal{A}_i\cap\widetilde{\mathcal{U}}\neq\varnothing$, we form a group containing all elements of this $\mathcal{A}_i$. Define $y_i$ and $z_i$ such that $|\mathcal{A}_i\cap\mathcal{Q}|=2y_i$ and $|\mathcal{A}_i\cap\widetilde{\mathcal{U}}|=z_i$, then $|\mathcal{A}_i|=2y_i+z_i$.

*Case 4*: For each $i$ such that $\mathcal{A}_i\cap\widetilde{\mathcal{U}}=\varnothing$, we form a group containing all elements of this $\mathcal{A}_i$. Note that in this case $|\mathcal{A}_i|\ge 4$. Define $y_i$ such that $|\mathcal{A}_i|=2y_i$.

[1]Recall that $\mathcal{Q}$ is defined in (3.10). Since $\mathcal{Q}$ contains the two-element pairings in $\mathcal{A}_i$, $|\mathcal{A}_i\cap\mathcal{Q}|$ is even.

Note that in *Cases 1–3*, some $\mathfrak{l}$ in the group belong to $\widetilde{\mathcal{U}}$ (and hence $\mathcal{U}$) and thus appear as a variable of $\widetilde{H}$ and $H^\circ$, while some $\mathfrak{l}$ in the group may not. In *Case 4*, no $\mathfrak{l}$ in the group appear as a variable of $\widetilde{H}$ or $H^\circ$, and they only appear in the summation $\sum_{k_{\mathcal{Q}}}^{(3)}$ in (6.57).

Now let $G$ be the union of all groups in *Cases 1–4*. Define $\widetilde{\mathcal{U}}^\dagger=\widetilde{\mathcal{U}}\backslash G$ and similarly $\mathcal{U}^\dagger=\mathcal{U}\backslash G$, and similarly $\mathcal{Q}^\dagger$, $B^\dagger$ and $C^\dagger$. Note that $\widetilde{\mathcal{U}}^\dagger\backslash\mathcal{U}^\dagger=\widetilde{\mathcal{U}}\backslash\mathcal{U}$. Let $\mathcal{U}_j^\dagger=\mathcal{U}_j\backslash G$ for $1\leq j\leq r$, and define[1] $G_1=\mathcal{U}_1\backslash\mathcal{U}_1^\dagger=\mathcal{U}_1\cap G$. Let $\mathscr{O}^\dagger$ be $\mathscr{O}$ after removing the $\mathcal{A}_i$'s involved in *Cases 2–4* above. Thus, for any $\mathcal{A}_i\in\mathscr{O}^\dagger$, if $1\leq i\leq v$ (in particular if $\mathcal{A}_i\cap\mathcal{U}_1^\dagger\neq\varnothing$), then $|\mathcal{A}_i|=2$.

Once we fix all the variables $k_\mathfrak{l}$ for $\mathfrak{l}\in G$ described above, we can view $\widetilde{H}$ as a tensor depending on $(k,k_{\widetilde{\mathcal{U}}^\dagger})$, and $H^\circ$ as a tensor depending on $(k,k_{\mathcal{U}^\dagger})$. In the same way $h^{(j)}$ can be viewed as a tensor depending on $(k_j,k_{\mathcal{U}_j^\dagger})$. More precisely, we define $h^{(j,\dagger)}=h^{(j,\dagger)}_{k_jk_{\mathcal{U}_j^\dagger}}$ to be $h^{(j)}_{k_jk_{\mathcal{U}_j}}$ with the values of $k_\mathfrak{l}$ for $\mathfrak{l}\in\mathcal{U}_j\backslash\mathcal{U}_j^\dagger=G\cap\mathcal{U}_j$ fixed.

If we use the triangle inequality, as well as the simple fact that

$$\|h_{k_Xk_Yk_Zk_W}\|_{k_Xk_Y\to k_Zk_W}\lesssim\prod_{\mathfrak{l}\in X\cup Z}N_\mathfrak{l}^{d/2}\cdot\sup_{k_X,k_Z}\|h_{k_Xk_Yk_Zk_W}\|_{k_Y\to k_W}$$

(where $X,Y,Z,W$ are arbitrary sets) under the assumption that $\langle k_\mathfrak{l}\rangle\lesssim N_\mathfrak{l}$ for $\mathfrak{l}\in X\cup Z$, then we can deduce that

$$\|H^\circ\|_{kk_B\to k_C}\lesssim\prod_i^{(\mathscr{O},2,\geq)}N_{\mathfrak{l}_i}^{-2y_i(\alpha-\theta)}\prod_i^{(\mathscr{O},3,\geq)}N_{\mathfrak{l}_i}^{(d/2)-2y_i(\alpha-\theta)}\prod_i^{(\mathscr{O},4,\geq)}N_{\mathfrak{l}_i}^{d-2y_i(\alpha-\theta)}\cdot\sup_{(k_\mathfrak{l})}\|(H^\circ)^\dagger\|_{kk_{B^\dagger}\to k_{C^\dagger}},\tag{6.62}$$

where $\prod_i^{(\mathscr{O},n,\geq)}$ is taken over all groups $\mathcal{A}_i$ of *Case* $n$ for $2\leq n\leq 4$, $\mathfrak{l}_i$ is any element of $\mathcal{A}_i$, and $\sup_{(k_\mathfrak{l})}$ is taken over all choices of the $k_\mathfrak{l}$'s with $\mathfrak{l}\in G$. The tensor $(H^\circ)^\dagger$ is defined by

$$(H^\circ)^\dagger=(H^\circ)^\dagger_{kk_{\mathcal{U}^\dagger}}=\sum_{k_{\widetilde{\mathcal{U}}^\dagger\backslash\mathcal{U}^\dagger}}(\widetilde{H})^\dagger_{kk_{\widetilde{\mathcal{U}}^\dagger}}\prod_{\mathfrak{l}\in\widetilde{\mathcal{U}}^\dagger\backslash\mathcal{U}^\dagger}(f_{N_\mathfrak{l}})_{k_\mathfrak{l}}^{\zeta_\mathfrak{l}^*},\tag{6.63}$$

$$(\widetilde{H})^\dagger_{kk_{\widetilde{\mathcal{U}}^\dagger}}=\prod_{\mathfrak{l},\mathfrak{l}'}^{(1,\dagger)}\mathbf{1}_{k_\mathfrak{l}=k_{\mathfrak{l}'}}\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}\mathbf{1}_{k_\mathfrak{l}\neq k_{\mathfrak{l}'}}\cdot\sum_{(k_1,\cdots,k_r)}h_{kk_1\cdots k_r}\sum_{k_{\mathcal{Q}^\dagger}}^{(3,\dagger)}\prod_{\mathfrak{l}\in\mathcal{Q}^\dagger}\Delta_{N_\mathfrak{l}}\gamma_{k_\mathfrak{l}}\prod_{j=1}^r[h^{(j,\dagger)}_{k_jk_{\mathcal{U}_j^\dagger}}]^{\zeta_j}.\tag{6.64}$$

Here the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}$ is the same as the one defined above in the third paragraph of *Step 2*, and the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1,\dagger)}$ is the same as $\prod_{(\mathfrak{l},\mathfrak{l}')}^{(1)}$ defined in (6.57) (which is taken from (3.11)), except that the product here is only taken over $\mathfrak{l},\mathfrak{l}'\in\mathcal{A}_i$ for $\mathcal{A}_i\in\mathscr{O}^\dagger$. Similarly the sum $\sum_{k_{\mathcal{Q}^\dagger}}^{(3,\dagger)}$ is the same as $\sum_{k_{\mathcal{Q}}}^{(3)}$ defined in (6.57) (which is taken from (3.11)), except that the sum here does not involve the variables $k_{\mathcal{Q}\backslash\mathcal{Q}^\dagger}$. Also note that, the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1,\dagger)}$ and sum $\sum_{k_{\mathcal{Q}^\dagger}}^{(3,\dagger)}$ are exactly the same as the ones defined in (3.11) when merging the tensors $(h^{(1,\dagger)},\cdots,h^{(r,\dagger)})$ via $(h,\mathscr{B},\mathscr{O}^\dagger)$.

Let us illustrate the above *Step 2* with an explicit example, which is an extension of the example in *Step 2* of the proof of Proposition 6.2; for simplicity, assume there is no blossoms (i.e. $\mathcal{V}_j=\varnothing$) and all frequencies are $\geq M^\delta$, so no trimming is needed throughout the process. Suppose $q=r=7$,

$$(\mathcal{U}_1,\mathcal{U}_2,\cdots,\mathcal{U}_7)=(\{\mathfrak{l}_{\mathrm{top}},\mathfrak{c},\mathfrak{j},\mathfrak{k},\mathfrak{m}\},\ \{\mathfrak{a},\mathfrak{b},\mathfrak{k},\mathfrak{o}\},\ \{\mathfrak{c},\mathfrak{d},\mathfrak{e},\mathfrak{k}\},\ \{\mathfrak{e},\mathfrak{f},\mathfrak{g}\},\ \{\mathfrak{a},\mathfrak{f},\mathfrak{h},\mathfrak{j},\mathfrak{m}\},\ \{\mathfrak{i},\mathfrak{j},\mathfrak{k},\mathfrak{m}\},\ \{\mathfrak{k},\mathfrak{m}\}),$$

[1] Note this definition is only for $j=1$ and not for $2\leq j\leq r$, which we will define later.

and the corresponding frequencies satisfy

$$N_{\mathfrak{l}_{\mathrm{top}}} \geq N_{\mathfrak{a}} \geq N_{\mathfrak{b}} \geq N_{\mathfrak{c}} \geq N_{\mathfrak{d}} \geq N_{\mathfrak{e}} \geq N_{\mathfrak{f}} \geq N_{\mathfrak{g}} \geq N_{\mathfrak{h}} \geq N_{\mathfrak{i}} \geq N_{\mathfrak{j}} \geq N_{\mathfrak{k}} \geq N_{\mathfrak{m}} \geq N_{\mathfrak{o}} \geq M^{\delta}.$$

Like in the proof of Proposition 6.2, a leaf occurring in at least two sets represents a pairing or over-pairing; for these leaves $\mathfrak{l}$, we also use $\mathfrak{l}_j$ to indicate the copy of $\mathfrak{l}$ in $\mathcal{U}_j$. Then $\mathcal{U} = \widetilde{\mathcal{U}} = \{\mathfrak{l}_{\mathrm{top}}, \mathfrak{b}, \mathfrak{d}, \mathfrak{g}, \mathfrak{h}, \mathfrak{i}, \mathfrak{j}, \mathfrak{k}, \mathfrak{o}\}$ and $\widetilde{H} = H^{\circ}$, and

$$(H^{\circ})_{kk_{\mathcal{U}}} = \sum_{(k_1,\cdots,k_7)} \sum_{(k_{\mathfrak{a}},k_{\mathfrak{c}},k_{\mathfrak{e}},k_{\mathfrak{f}},k_{\mathfrak{m}})} h_{kk_1\cdots k_7} \cdot h^{(1)}_{k_1k_{\mathfrak{l}_{\mathrm{top}}}k_{\mathfrak{c}}k_{\mathfrak{j}}k_{\mathfrak{k}}k_{\mathfrak{m}}} \cdot h^{(2)}_{k_2k_{\mathfrak{a}}k_{\mathfrak{b}}k_{\mathfrak{k}}k_{\mathfrak{o}}}$$
$$\times h^{(3)}_{k_3k_{\mathfrak{c}}k_{\mathfrak{d}}k_{\mathfrak{e}}k_{\mathfrak{k}}} \cdot h^{(4)}_{k_4k_{\mathfrak{e}}k_{\mathfrak{f}}k_{\mathfrak{g}}} \cdot h^{(5)}_{k_5k_{\mathfrak{a}}k_{\mathfrak{f}}k_{\mathfrak{h}}k_{\mathfrak{j}}k_{\mathfrak{m}}} \cdot h^{(6)}_{k_6k_{\mathfrak{i}}k_{\mathfrak{j}}k_{\mathfrak{k}}k_{\mathfrak{m}}} \cdot h^{(7)}_{k_7k_{\mathfrak{k}}k_{\mathfrak{m}}},$$

where for simplicity, we have omitted the various $\Delta_{N_{\mathfrak{l}}}\gamma_{k_{\mathfrak{l}}}$ factors.

Suppose $B = \{\mathfrak{b}, \mathfrak{d}\}$ and $C = \{\mathfrak{l}_{\mathrm{top}}, \mathfrak{g}, \mathfrak{h}, \mathfrak{k}\}$ is a subpartition of $\mathcal{U}$ and $E = \{\mathfrak{i}, \mathfrak{j}, \mathfrak{o}\}$. Our goal is to estimate $\|H^{\circ}\|_{kk_B \to k_C}$. The groups in *Cases 1–4* in the above process are then:

- *Case 1*: $\{\mathfrak{i}\}$ and $\{\mathfrak{o}\}$;
- *Case 2*: $\mathcal{A}_1 = \{\mathfrak{j}_1, \mathfrak{j}_5, \mathfrak{j}_6\}$;
- *Case 3*: $\mathcal{A}_2 = \{\mathfrak{k}_1, \mathfrak{k}_2, \mathfrak{k}_3, \mathfrak{k}_6, \mathfrak{k}_7\}$;
- *Case 4*: $\mathcal{A}_3 = \{\mathfrak{m}_1, \mathfrak{m}_5, \mathfrak{m}_6, \mathfrak{m}_7\}$.

After fixing the values of $(k_{\mathfrak{i}}, k_{\mathfrak{o}}, k_{\mathfrak{j}}, k_{\mathfrak{k}}, k_{\mathfrak{m}})$, which are considered in *Cases 1–4* above, we can reduce $\|H^{\circ}\|_{kk_B \to k_C}$ to $\|(H^{\circ})^{\dagger}\|_{kk_{B^{\dagger}} \to k_{C^{\dagger}}}$ by the above argument, where the relevant sets

$$B^{\dagger} = \{\mathfrak{b}, \mathfrak{d}\},\ C^{\dagger} = \{\mathfrak{l}_{\mathrm{top}}, \mathfrak{g}, \mathfrak{h}\},\ \mathcal{U}^{\dagger} = \{\mathfrak{l}_{\mathrm{top}}, \mathfrak{b}, \mathfrak{d}, \mathfrak{g}, \mathfrak{h}\},\ \mathcal{Q}^{\dagger} = \{\mathfrak{a}_2, \mathfrak{a}_5, \mathfrak{c}_1, \mathfrak{c}_3, \mathfrak{e}_3, \mathfrak{e}_4, \mathfrak{f}_4, \mathfrak{f}_5\}, \tag{6.65}$$
$$(\mathcal{U}_1^{\dagger}, \cdots, \mathcal{U}_7^{\dagger}) = (\{\mathfrak{l}_{\mathrm{top}}, \mathfrak{c}\},\ \{\mathfrak{a}, \mathfrak{b}\},\ \{\mathfrak{c}, \mathfrak{d}, \mathfrak{e}\},\ \{\mathfrak{e}, \mathfrak{f}, \mathfrak{g}\},\ \{\mathfrak{a}, \mathfrak{f}, \mathfrak{h}\},\ \varnothing,\ \varnothing),$$

and the tensors

$$(H^{\circ})^{\dagger}_{kk_{\mathcal{U}^{\dagger}}} = \sum_{(k_1,\cdots,k_7)} \sum_{(k_{\mathfrak{a}},k_{\mathfrak{c}},k_{\mathfrak{e}},k_{\mathfrak{f}})} h_{kk_1\cdots k_7} \cdot h^{(1,\dagger)}_{k_1k_{\mathfrak{l}_{\mathrm{top}}}k_{\mathfrak{c}}} \cdot h^{(2,\dagger)}_{k_2k_{\mathfrak{a}}k_{\mathfrak{b}}} \cdot h^{(3,\dagger)}_{k_3k_{\mathfrak{c}}k_{\mathfrak{d}}k_{\mathfrak{e}}} \cdot h^{(4,\dagger)}_{k_4k_{\mathfrak{e}}k_{\mathfrak{f}}k_{\mathfrak{g}}} \cdot h^{(5,\dagger)}_{k_5k_{\mathfrak{a}}k_{\mathfrak{f}}k_{\mathfrak{h}}} \cdot h^{(6,\dagger)}_{k_6} \cdot h^{(7,\dagger)}_{k_7},$$

where $h^{(j,\dagger)}_{k_jk_{\mathcal{U}_j^{\dagger}}}$ is $h^{(j)}_{k_jk_{\mathcal{U}_j}}$ after fixing the $k_{\mathcal{U}_j \backslash \mathcal{U}_j^{\dagger}}$ for $1 \leq j \leq 7$.

*Step 3: the method of descent.* Now we need to estimate $(H^{\circ})^{\dagger}$. A key step is to implement Proposition 6.2 by singling out $h^{1,\dagger}$. To that effect, consider those $\mathcal{A}_i \in \mathscr{O}^{\dagger}$ such that $\mathcal{A}_i \cap \mathcal{U}_1^{\dagger} \neq \varnothing$; we know that each such $\mathcal{A}_i$ contains exactly one pair. Let $(\mathscr{O}^{\dagger})'$ be $\mathscr{O}^{\dagger}$ after removing these $\mathcal{A}_i$'s, and $\mathcal{D}$ be the union of these $\mathcal{A}_i$'s. Define $\mathcal{F} = (\mathcal{U}^{\dagger} \cup \mathcal{D}) \backslash \mathcal{U}_1^{\dagger}$ and similarly define $\widetilde{\mathcal{F}} = (\widetilde{\mathcal{U}}^{\dagger} \cup \mathcal{D}) \backslash \mathcal{U}_1^{\dagger}$, so that $\widetilde{\mathcal{U}}^{\dagger} \backslash \mathcal{U}^{\dagger} = \widetilde{\mathcal{F}} \backslash \mathcal{F}$. Then by (6.63)–(6.64) we have

$$(H^{\circ})^{\dagger}_{kk_{\mathcal{U}^{\dagger}}} = \sum_{k_1} \sum_{k_{\mathcal{D}}} \prod_{\mathfrak{l} \in \mathcal{D}} \Delta_{N_{\mathfrak{l}}}\gamma_{k_{\mathfrak{l}}} \cdot \big[h^{(1,\dagger)}_{k_1k_{\mathcal{U}_1^{\dagger}}}\big]^{\zeta_1} \cdot (H^*)_{kk_1k_{\mathcal{F}}}, \tag{6.66}$$

where $\sum_{k_{\mathcal{D}}}$ is taken over all $k_{\mathcal{D}}$ such that $k_{\mathfrak{l}} = k_{\mathfrak{l}'}$ for any pairing $\mathcal{A}_i = \{\mathfrak{l}, \mathfrak{l}'\} \subset \mathcal{D}$, and the tensor $(H^*)_{kk_1k_{\mathcal{F}}}$ is defined by

$$(H^*)_{kk_1k_{\mathcal{F}}} = \sum_{k_{\widetilde{\mathcal{F}} \backslash \mathcal{F}}} \mathcal{H}_{kk_1k_{\widetilde{\mathcal{F}}}} \prod_{\mathfrak{l} \in \widetilde{\mathcal{F}} \backslash \mathcal{F}} (f_{N_{\mathfrak{l}}})^{\zeta_{\mathfrak{l}}^*}_{k_{\mathfrak{l}}}, \tag{6.67}$$

$$\mathcal{H}_{kk_1k_{\widetilde{\mathcal{F}}}} = \prod_{\mathfrak{l},\mathfrak{l}'}^{(1',\dagger)} \mathbf{1}_{k_{\mathfrak{l}}=k_{\mathfrak{l}'}} \prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)} \mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}} \cdot \sum_{(k_2,\cdots,k_r)} h_{kk_1\cdots k_r} \sum_{k_{\mathcal{R}}} \prod_{\mathfrak{l} \in \mathcal{R}}^{(3',\dagger)} \Delta_{N_{\mathfrak{l}}}\gamma_{k_{\mathfrak{l}}} \prod_{j=2}^{r} \big[h^{(j,\dagger)}_{k_jk_{\mathcal{U}_j^{\dagger}}}\big]^{\zeta_j}. \tag{6.68}$$

Here $\mathcal{R}=\mathcal{Q}^\dagger\backslash\mathcal{D}$; in (6.68) the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}$ is the same as the one defined above in *Step 2*, and the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1',\dagger)}$ is the same as $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1,\dagger)}$ defined in (6.64). The sum $\sum_{k_{\mathcal{R}}}^{(3',\dagger)}$ is the same as $\sum_{k_{\mathcal{Q}^\dagger}}^{(3,\dagger)}$ defined in (6.64), except that the sum here does not involve the variables $k_{\mathcal{D}}$. The tensor $H^*$ can be understood as a "partial trimming" of $\mathcal{H}$ at frequency $M^\delta$, i.e. only the tree part $\widetilde{\mathcal{F}}$ is trimmed.

Note that $N_{\mathfrak{l}}^{\alpha-\theta}\cdot\Delta_{N_{\mathfrak{l}}}\gamma_{k_{\mathfrak{l}}}$ is a bounded function of $k_{\mathfrak{l}}$ and can be absorbed into one of the tensors $h^{(1,\dagger)}$ or $H^*$. Applying Proposition 4.11 to (6.66) with $(k_X,k_Y,k_{A_1},k_{A_2})=(kk_{B^\dagger},k_{C^\dagger},k_1k_{\mathcal{U}_1^\dagger},kk_1k_{\mathcal{F}})$, where we combine any two elements in $\mathcal{D}$ that form a pairing into a single element as we did in the proof of Proposition 4.14, we obtain

$$\|(H^\circ)^\dagger\|_{kk_{B^\dagger}\to k_{C^\dagger}}\lesssim\prod_{\mathfrak{l}\in\mathcal{D}}N_{\mathfrak{l}}^{-\alpha+\theta}\cdot\|h^{(1,\dagger)}_{k_1k_{\mathcal{U}_1^\dagger}}\|_{k_1k_{(B^\dagger\cup\mathcal{D})\cap\mathcal{U}_1^\dagger}\to k_{C^\dagger\cap\mathcal{U}_1^\dagger}}\cdot\|(H^*)_{kk_1k_{\mathcal{F}}}\|_{kk_{B^\dagger\cap\mathcal{F}}\to k_1k_{(C^\dagger\cup\mathcal{D})\cap\mathcal{F}}}.\tag{6.69}$$

To estimate $H^*$ in (6.69) we shall apply Proposition 4.14. First recall $N_*=\max(N_2,\cdots,N_q)$, and that since the expression (6.68) for $\mathcal{H}$ involves only $h$ and $h^{(j,\dagger)}$ for $2\leq j\leq r$, by assumption this $\mathcal{H}$ *is* $\mathcal{B}_{(N_*)^{[\delta]}}$ *measurable.* Next note that by assumption $N_{\mathfrak{l}}\geq(N_*)^\delta$ for each $\mathfrak{l}\in\widetilde{\mathcal{F}}\backslash\mathcal{F}$ and that there is no pairing among $k_{\widetilde{\mathcal{F}}\backslash\mathcal{F}}$ in (6.67) in view of the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$ in the definition (6.68) of $\mathcal{H}$. Now we apply Proposition 4.14 to (6.67), setting $(b,c)=(kk_{B^\dagger\cap\mathcal{F}},\,k_1k_{(C^\dagger\cup\mathcal{D})\cap\mathcal{F}})$, then for some partition $(B_0,C_0)$ of $\widetilde{\mathcal{F}}\backslash\mathcal{F}$ we have $\tau^{-1}M$-certainly that

$$\|(H^*)_{kk_1k_{\mathcal{F}}}\|_{kk_{B^\dagger\cap\mathcal{F}}\to k_1k_{(C^\dagger\cup\mathcal{D})\cap\mathcal{F}}}\lesssim(\tau^{-1}M)^\theta\prod_{\mathfrak{l}\in\widetilde{\mathcal{F}}\backslash\mathcal{F}}N_{\mathfrak{l}}^{-\alpha+\theta}\cdot\|\mathcal{H}\|_{kk_S\to k_1k_T},\tag{6.70}$$

where $S=B_0\cup(B^\dagger\cap\mathcal{F})$ and $T=C_0\cup((C^\dagger\cup\mathcal{D})\cap\mathcal{F})$. Note that $(S,T)$ form a partition of $\widetilde{\mathcal{F}}$ such that $S\supset B^\dagger\cap\mathcal{F}$ and $T\supset(C^\dagger\cup\mathcal{D})\cap\mathcal{F}$.

It now remains to estimate $\mathcal{H}$. By applying Lemma 4.10 we may remove the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2,<)}\mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}}$ from the definition (6.68). Then, we can repeat the arguments in *Step 2* above, and fix the values of $k_{\mathfrak{l}}$ for $\mathfrak{l}\in\mathcal{A}_i$ with $\mathcal{A}_i\in(\mathscr{O}^\dagger)'$ and $|\mathcal{A}_i|\geq 3$ (where necessarily $i\geq v+1$). Here we only have two cases:

*Case 3*: For each $i$ such that $\mathcal{A}_i\cap\widetilde{\mathcal{F}}\neq\varnothing$, we form a group containing all elements of this $\mathcal{A}_i$. Define $y_i$ and $z_i$ such that $|\mathcal{A}_i\cap\mathcal{R}|=2y_i$ and $|\mathcal{A}_i\cap\widetilde{\mathcal{F}}|=z_i$, then $|\mathcal{A}_i|=2y_i+z_i$.

*Case 4*: For each $i$ such that $\mathcal{A}_i\cap\widetilde{\mathcal{F}}=\varnothing$, we form a group containing all elements of this $\mathcal{A}_i$. Note that in this case $|\mathcal{A}_i|\geq 4$. Define $y_i$ such that $|\mathcal{A}_i|=2y_i$.

As in *Step 2*, we define $G^\dagger$ to be the union of all groups in *Cases 3–4* defined above. Define $(\mathscr{O}^{\dagger\dagger})'$ to be $(\mathscr{O}^\dagger)'$ after removing the $\mathcal{A}_i$'s involved in *Cases 3–4* above, $\widetilde{\mathcal{F}}^\dagger=\widetilde{\mathcal{F}}\backslash G^\dagger$, and similarly define $S^\dagger$, $T^\dagger$ and $\mathcal{R}^\dagger$. We also define $\mathcal{U}_j^{\dagger\dagger}=\mathcal{U}_j^\dagger\backslash G^\dagger$. With these variables $k_{\mathfrak{l}}$ for $\mathfrak{l}\in G^\dagger$ fixed, we can view $\mathcal{H}$ as a tensor depending on $(k,k_1,k_{\widetilde{\mathcal{F}}^\dagger})$, and view $h^{(j,\dagger)}$ as a tensor depending on $(k_j,k_{\mathcal{U}_j^{\dagger\dagger}})$. We define $h^{(j,\dagger\dagger)}=h^{(j,\dagger\dagger)}_{kk_{\mathcal{U}_j^{\dagger\dagger}}}$ to be $h^{(j,\dagger)}_{kk_{\mathcal{U}_j^\dagger}}$ with the values of $k_{\mathfrak{l}}$ for $\mathfrak{l}\in\mathcal{U}_j^\dagger\backslash\mathcal{U}_j^{\dagger\dagger}=G^\dagger\cap\mathcal{U}_j^\dagger$ fixed. By the same arguments as in *Step 2*, we deduce that

$$\|\mathcal{H}\|_{kk_S\to k_1k_T}\lesssim\prod_i^{(\mathscr{O},3,<)}N_{\mathfrak{l}_i}^{(d/2)-2y_i(\alpha-\theta)}\prod_i^{(\mathscr{O},4,<)}N_{\mathfrak{l}_i}^{d-2y_i(\alpha-\theta)}\cdot\sup_{(k_{\mathfrak{l}})}\|(\mathcal{H}^\dagger)_{kk_1k_{\widetilde{\mathcal{F}}^\dagger}}\|_{kk_{S^\dagger}\to k_1k_{T^\dagger}},\tag{6.71}$$

where the products $\prod_i^{(\mathscr{O},n,<)}$ and the supremum $\sup_{(k_\mathfrak{l})}$ are defined in the same way as in *Step 2* above, and

$$(\mathcal{H}^\dagger)_{kk_1k_{\widetilde{\mathcal{F}}^\dagger}}=\prod_{\mathfrak{l},\mathfrak{l}'}^{(1',\dagger\dagger)}\mathbf{1}_{k_\mathfrak{l}=k_{\mathfrak{l}'}}\sum_{(k_2,\cdots,k_r)}h_{kk_1\cdots k_r}\sum_{k_{\mathcal{R}^\dagger}}^{(3',\dagger\dagger)}\prod_{\mathfrak{l}\in\mathcal{R}^\dagger}\Delta_{N_\mathfrak{l}}\gamma_{k_\mathfrak{l}}\prod_{j=2}^{r}[h^{(j,\dagger\dagger)}_{k_jk_{\mathcal{U}_j^{\dagger\dagger}}}]^{\zeta_j}. \tag{6.72}$$

In (6.72) the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1',\dagger\dagger)}$ is the same as $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1',\dagger)}$ defined in (6.68), except that the product here is only taken over $\mathfrak{l},\mathfrak{l}'\in\mathcal{A}_i$ for some $\mathcal{A}_i\in(\mathscr{O}^{\dagger\dagger})'$; similarly, the sum $\sum_{k_{\mathcal{R}^\dagger}}^{(3',\dagger\dagger)}$ is the same as $\sum_{k_\mathcal{R}}^{(3',\dagger)}$ defined in (6.68), except that the sum here does not involve the variables $k_{\mathcal{R}\backslash\mathcal{R}^\dagger}$. Note that $|\mathcal{A}_i|=2$ and $\mathcal{A}_i\cap\widetilde{\mathcal{F}}^\dagger=\varnothing$ for any $\mathcal{A}_i\in(\mathscr{O}^{\dagger\dagger})'$, so the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1',\dagger\dagger)}$ vacuously equals 1, and the condition of the sum $\sum_{k_{\mathcal{R}^\dagger}}^{(3',\dagger\dagger)}$ is just $k_\mathfrak{l}=k_{\mathfrak{l}'}$ for each pairing $\{\mathfrak{l},\mathfrak{l}'\}\in(\mathscr{O}^{\dagger\dagger})'$ (recall that $\mathcal{R}^\dagger=\mathcal{R}\backslash G^\dagger=\mathcal{Q}\backslash(\mathcal{D}\cup G\cup G^\dagger)$).

Now recall the example in *Step 2*. To estimate $\|(H^\circ)^\dagger\|_{kk_{B^\dagger}\to k_{C^\dagger}}$, we separate $h^{(1)}$ from the others as in (6.66)–(6.68), where $\mathcal{D}=\{\mathfrak{c}_1,\mathfrak{c}_3\}$, $\mathcal{U}_1^\dagger=\{\mathfrak{l}_{\mathrm{top}},\mathfrak{c}\}$, $\mathcal{F}=\widetilde{\mathcal{F}}=\{\mathfrak{b},\mathfrak{c},\mathfrak{d},\mathfrak{g},\mathfrak{h}\}$ and $\mathcal{R}=\mathcal{Q}^\dagger\backslash\mathcal{D}=\{\mathfrak{a}_2,\mathfrak{a}_5,\mathfrak{e}_3,\mathfrak{e}_4,\mathfrak{f}_4,\mathfrak{f}_5\}$, namely

$$(H^\circ)^\dagger_{kk_{\mathcal{U}^\dagger}}=\sum_{k_1,k_\mathfrak{c}}h^{(1,\dagger)}_{k_1k_{\mathfrak{l}_{\mathrm{top}}}k_\mathfrak{c}}(H^*)_{kk_1k_\mathcal{F}},$$

$$(H^*)_{kk_1k_\mathcal{F}}=\sum_{(k_2,\cdots k_7)}\sum_{(k_\mathfrak{a},k_\mathfrak{e},k_\mathfrak{f})}h_{kk_1\cdots k_7}\cdot h^{(2,\dagger)}_{k_2k_\mathfrak{a}k_\mathfrak{b}}\cdot h^{(3,\dagger)}_{k_3k_\mathfrak{c}k_\mathfrak{d}k_\mathfrak{e}}\cdot h^{(4,\dagger)}_{k_4k_\mathfrak{e}k_\mathfrak{f}k_\mathfrak{g}}\cdot h^{(5,\dagger)}_{k_5k_\mathfrak{a}k_\mathfrak{f}k_\mathfrak{h}}\cdot h^{(6,\dagger)}_{k_6}\cdot h^{(7,\dagger)}_{k_7},$$

where for simplicity, we have again omitted the various $\Delta_{N_\mathfrak{l}}\gamma_{k_\mathfrak{l}}$ factors (and also the power factors below). Note that $H^*=\mathcal{H}=\mathcal{H}^\dagger$ as all frequencies are $\geq M^\delta$, in particular no partial trimming or *Cases 3–4* in *Step 3* above is involved. By (6.69) we have

$$\|(H^\circ)^\dagger_{kk_{\mathcal{U}^\dagger}}\|_{kk_\mathfrak{b}k_\mathfrak{d}\to k_{\mathfrak{l}_{\mathrm{top}}}k_\mathfrak{g}k_\mathfrak{h}}\leq\|h^{(1,\dagger)}_{k_1k_{\mathfrak{l}_{\mathrm{top}}}k_\mathfrak{c}}\|_{k_1k_\mathfrak{c}\to k_{\mathfrak{l}_{\mathrm{top}}}}\cdot\|(H^*)_{kk_1k_\mathcal{F}}\|_{kk_\mathfrak{b}k_\mathfrak{d}\to k_1k_\mathfrak{c}k_\mathfrak{g}k_\mathfrak{h}}.$$

The norm $\|h^{(1,\dagger)}\|_{k_1k_\mathfrak{c}\to k_{\mathfrak{l}_{\mathrm{top}}}}$ is then controlled using (6.43)–(6.45), and $\|H^*\|_{kk_\mathfrak{b}k_\mathfrak{d}\to k_1k_\mathfrak{c}k_\mathfrak{g}k_\mathfrak{h}}$ is controlled using Proposition 6.2; note that here $\mathcal{E}=\{6,7\}$ as in the example in *Step 2* of the proof of Proposition 6.2, and the corresponding $H^{\mathrm{sg}}$ is the same as the one in that example, but with $h^{(j)}$ replaced by $h^{(j,\dagger)}$. After putting these two bounds together and calculating the various powers involved (see *Step 4*), we can obtain the desired estimate (6.49) for this example.

*Step 4: putting together.* We now need to estimate $\mathcal{H}^\dagger$. As in *Step 3*, we may replace $\Delta_{N_\mathfrak{l}}\gamma_{k_\mathfrak{l}}$ by $N_\mathfrak{l}^{-\alpha+\theta}$ and absorb the resulting factor into one of the tensors $h^{(j,\dagger\dagger)}$ using Lemma 4.10, so instead of $\mathcal{H}^\dagger$ we only need to consider

$$\mathcal{M}_{kk_1k_{\widetilde{\mathcal{F}}^\dagger}}:=\prod_{\mathfrak{l}\in\mathcal{R}^\dagger}N_\mathfrak{l}^{-\alpha+\theta}\sum_{(k_2,\cdots,k_r)}h_{kk_1\cdots k_r}\sum_{k_{\mathcal{R}^\dagger}}\prod_{j=2}^{r}[h^{(j,\dagger\dagger)}_{k_jk_{\mathcal{U}_j^{\dagger\dagger}}}]^{\zeta_j}, \tag{6.73}$$

where $\sum_{k_{\mathcal{R}^\dagger}}$ is the sum such that $k_\mathfrak{l}=k_{\mathfrak{l}'}$ for each pairing $\{\mathfrak{l},\mathfrak{l}'\}\in(\mathscr{O}^{\dagger\dagger})'$ which is just $\sum_{k_{\mathcal{R}^\dagger}}^{(3',\dagger\dagger)}$ in (6.72). We shall apply Proposition 6.2 to estimate (6.73), but we first need to make a few adjustments to fit the framework of (6.9)–(6.17).

First, for $r+1\leq j\leq q$, we may define $\mathcal{U}_j^{\dagger\dagger}=\varnothing$ and $h^{(j,\dagger\dagger)}=h^{(j,\dagger\dagger)}_{k_j}$ to be supported at a single point $k_j$ (the one fixed in *Step 1*). Moreover, in view of the extra factor $\prod_{j=r+1}^qN_j^{-d/2}$ described in *Step 1* when stating the norm bounds we want to prove for $H^\circ_{kk_\mathcal{U}}$, we may assume $|h^{(j,\dagger\dagger)}_{k_j}|\lesssim N_j^{-d/2}$

for $r+1\leq j\leq q$, so it satisfies the type $R0^+$ conditions[1] of Proposition 6.2 with $\mathfrak{X}_j$ replaced by 1. Second, we also view $h_{kk_1\cdots k_r}$ as a tensor depending on $(k,k_1,\cdots,k_q)$, in the support of which $k_j\,(r+1\leq j\leq q)$ takes a single value, and denote it by $h_{kk_1\cdots k_q}$. Moreover, the tensor $h=h_{kk_1\cdots k_q}$ satisfies the support condition (6.16) for some choice of $\Gamma$, as well as other conditions of $h$ listed in Proposition 6.2.

Next we check that $h^{(j,\dagger\dagger)}_{k_jk_{\mathcal{U}_j^{\dagger\dagger}}}$ for $2\leq j\leq r$ satisfy the conditions in Proposition 6.2. Note that $\mathcal{U}_j^{\dagger\dagger}$ is a subset of $\mathcal{U}_j$, and $h^{(j,\dagger\dagger)}$ is formed from $h^{(j)}$ by fixing the values of $k_{\mathcal{U}_j\backslash\mathcal{U}_j^{\dagger\dagger}}$ (for simplicity we denote it by $k_{G_j}$, where $G_j=\mathcal{U}_j\backslash\mathcal{U}_j^{\dagger\dagger}=(G\cap\mathcal{U}_j)\cup(G^\dagger\cap\mathcal{U}_j))$. We consider the following scenarios.

*Scenario 1*: when $h^{(j)}$ has type 1. Then $h^{(j,\dagger\dagger)}$ will have type $R1$ in the sense of Proposition 6.2. Indeed by (6.46) (actually the modified version of it[2] in *Step 1*), we may assume

$$\Big|k_j-\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell_j\Big|\leq(N_*)^{3\delta},\quad \ell_j=\sum_{\mathfrak{f}\in\mathcal{V}_j}\zeta_{\mathfrak{f}}k_{\mathfrak{f}},$$

otherwise we gain a huge power of $M$ from (6.46) that would suffice (note that $\kappa\gg_{C_\delta}1$ and $N_*\geq M^\delta$). When $k_{G_j}$ is fixed, the above implies that $h^{(j,\dagger\dagger)}$ satisfies (6.14) with $\mathcal{U}_j$ replaced by $\mathcal{U}_j^{\dagger\dagger}$ and some fixed $m_j=\ell_j+\sum_{\mathfrak{l}\in G_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}$. Moreover, by (6.44)–(6.45) we deduce that $h^{(j,\dagger\dagger)}$ satisfies (6.15) with $\mathcal{U}_j$ replaced by $\mathcal{U}_j^{\dagger\dagger}$, and $\mathfrak{X}_j$ replaced by

$$\mathfrak{X}_j^*\cdot\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in G_j}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_j}N_{\mathfrak{f}}^{d}.$$

*Scenario 2*: when $h^{(j)}$ has type 0, and $G_j=\varnothing$. Then in this case, $h^{(j,\dagger\dagger)}=h^{(j)}$ and it will have type $R0$ in the sense of Proposition 6.2. It satisfies (6.11)–(6.12) with $\mathfrak{X}_j$ replaced by

$$\mathfrak{X}_j^*\cdot\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}.\tag{6.74}$$

*Scenario 3*: when $h^{(j)}$ has type 0, and $G_j\neq\varnothing$. Then $h^{(j,\dagger\dagger)}$ will have type $R0^+$ in the sense of Proposition 6.2. It satisfies the modified versions of (6.11)–(6.12) with $\mathcal{U}_j$ replaced by $\mathcal{U}_j^{\dagger\dagger}$, and $\mathfrak{X}_j$ replaced by

$$\mathfrak{X}_j^*\cdot\prod_{\mathfrak{l}\in\mathcal{P}_j}N_{\mathfrak{l}}^{-4\varepsilon}\prod_{\mathfrak{l}\in G_j}N_{\mathfrak{l}}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_j}N_{\mathfrak{p}}^{-\delta^3}.\tag{6.75}$$

In *Scenarios 2–3*, note the different powers of $N_{\mathfrak{l}}$ for $\mathfrak{l}\in\mathcal{P}_j\cup G_j$ between (6.74)–(6.75) and (6.43) for type 0 tensors, which allow us to bridge from $\mathcal{X}_{1,j}$ in (6.43) and (5.33) to $\mathcal{Z}_{1,j}$ in (6.11)–(6.12).

Therefore, by applying Proposition 6.2 with $P$, $Q$, $\mathcal{U}$ and $\mathcal{Q}$ replaced by $S^\dagger$, $T^\dagger$, $\widetilde{\mathcal{F}}^\dagger$ and $\mathcal{R}^\dagger$ respectively, we obtain that

$$\begin{aligned}\prod_{j=r+1}^{q}N_j^{-d/2}\cdot\|\mathcal{M}\|_{kk_{S^\dagger}\to k_1k_{T^\dagger}}\lesssim\prod_{j=2}^{r}\mathfrak{X}_j^*\cdot\prod_{\mathfrak{l}\in\mathcal{P}_2\cup\cdots\cup\mathcal{P}_r\cup\mathcal{R}^\dagger}N_{\mathfrak{l}}^{-4\varepsilon}\cdot\prod_{\mathfrak{l}\in G_2\cup\cdots\cup G_r}N_{\mathfrak{l}}^{8\varepsilon}\cdot\prod_{\mathfrak{p}\in\mathcal{Y}_2\cup\cdots\cup\mathcal{Y}_r}N_{\mathfrak{l}}^{-\delta^3}\\ \times\prod_{\mathfrak{l}\in\mathcal{V}_2\cup\cdots\cup\mathcal{V}_r}N_{\mathfrak{l}}^{d}\cdot\prod_{\mathfrak{l}\in\widetilde{\mathcal{F}}^\dagger}N_{\mathfrak{l}}^{\beta}\cdot\prod_{j}^{(0)}N_j^{-2\varepsilon}\cdot(N_*)^{-\varepsilon^3},\end{aligned}\tag{6.76}$$

where $\prod_j^{(0)}$ is taken over all $j$ such that $r+1\leq j\leq q$, or $1\leq j\leq r$ and $h^{(j)}$ has type 0.

[1]The support condition (6.13) for type $R0^+$ can be immediately verified, since $\mathcal{U}_j^{\dagger\dagger}=\varnothing$.

[2]In the analogous sense that (6.55) is the modified version of (6.44).

This implies the same bound for $\mathcal{H}^\dagger$. Next we can control the norm

$$\|h^{(1,\dagger)}_{k_1 k_{\mathcal{U}_1^\dagger}}\|_{k_1 k_{(B^\dagger\cup\mathcal{D})\cap\mathcal{U}_1^\dagger}\to k_{C^\dagger\cap\mathcal{U}_1^\dagger}} \tag{6.77}$$

using the modified versions of (6.43)–(6.45) in *Step 1*. In fact, if $h^{(1)}$ has type 1, the norm (6.77) is bounded as follows

$$(6.77)\lesssim \mathfrak{X}_1^*\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in\mathcal{U}_1^\dagger}N_\mathfrak{l}^{\beta}\prod_{\mathfrak{l}\in\mathcal{P}_1}N_\mathfrak{l}^{-4\varepsilon}\prod_{\mathfrak{l}\in G_1}N_\mathfrak{l}^{8\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_1}N_\mathfrak{p}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_1}N_\mathfrak{f}^{d}\cdot\mathcal{X}_1^{\mathrm{type1}}, \tag{6.78}$$

by the modified versions of (6.44) and (6.45), where $\mathcal{X}_1^{\mathrm{type1}}$ equals $(\max_{\mathfrak{l}\in C^\dagger\cap\mathcal{U}_1^\dagger}N_\mathfrak{l})^{-\beta}$ if $C^\dagger\cap\mathcal{U}_1^\dagger\neq\varnothing$, and equals $N_1^{-\varepsilon}$ if $C^\dagger\cap\mathcal{U}_1^\dagger=\varnothing$. If $h^{(1)}$ has type 0, the norm (6.77) is bounded as follows

$$(6.77)\lesssim \mathfrak{X}_1^*\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in\mathcal{U}_1^\dagger}N_\mathfrak{l}^{\beta_1}\prod_{\mathfrak{l}\in\mathcal{P}_1}N_\mathfrak{l}^{-8\varepsilon}\prod_{\mathfrak{l}\in G_1}N_\mathfrak{l}^{4\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}_1}N_\mathfrak{p}^{-\delta^3}\cdot\mathcal{X}_{0,1}\mathcal{X}_{1,1} \tag{6.79}$$

by (6.43), where $\mathcal{X}_{0,1}$ and $\mathcal{X}_{1,1}$ are as in (6.43) (which is taken from (5.33)), corresponding to $B_1=(B^\dagger\cup\mathcal{D})\cap\mathcal{U}_1^\dagger$ and $C_1=C^\dagger\cap\mathcal{U}_1^\dagger$ (and $E_1=G_1$).

Now, by plugging the bound for $\mathcal{H}^\dagger$ (which follows from (6.76)) back into (6.71), (6.70), (6.69) and (6.62), using the bounds for (6.77) described above, and exploiting the relations between the various sets defined before, we can eventually obtain the bound for $H^\circ$ that implies (6.59), and hence (6.49). In fact, by separating the two cases according to whether $\mathfrak{l}_{\mathrm{top}}\in\mathcal{U}_1^\dagger$ or not (recall that $\mathfrak{l}_{\mathrm{top}}\in C\cap\mathcal{U}_1$ is such that $N_{\mathfrak{l}_{\mathrm{top}}}\sim\max\{N_\mathfrak{l}:\mathfrak{l}\in C\}$), and by using (6.78) if $h^{(1)}$ has type 1 or (6.79) if[1] $h^{(1)}$ has type 0, we can obtain in both two cases that

$$\begin{aligned}\prod_{j=r+1}^{q}N_j^{-d/2}\cdot\|H^\circ\|_{kk_B\to k_C}\lesssim(\tau^{-1}M)^{\theta}(N_*)^{-\varepsilon^3/2}\prod_{j=1}^{r}\mathfrak{X}_j^*\cdot\prod_{(\mathfrak{l}\in B\cup C)\setminus\{\mathfrak{l}_{\mathrm{top}}\}}N_\mathfrak{l}^{\beta}\cdot\prod_{\mathfrak{l}\in\mathcal{P}}N_\mathfrak{l}^{-4\varepsilon}\\ \times\prod_{\mathfrak{l}\in E}N_\mathfrak{l}^{8\varepsilon}\cdot\prod_{\mathfrak{p}\in\widetilde{\mathcal{Y}}}N_\mathfrak{p}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_1\cup\cdots\cup\mathcal{V}_r}N_\mathfrak{f}^{d}\cdot\prod_{j}^{(0)}N_j^{-2\varepsilon}\cdot\mathfrak{Z},\end{aligned} \tag{6.80}$$

where $\prod_j^{(0)}$ is defined in (6.76) and $\mathfrak{Z}$ is a quantity such that

$$\mathfrak{Z}\leq\prod_i^{(\mathscr{O},2)}N_{\mathfrak{l}_i}^{-2y_i(\alpha-\theta)}\cdot\prod_i^{(\mathscr{O},3)}N_{\mathfrak{l}_i}^{(d/2)-2y_i(\alpha-\theta)}\cdot\prod_i^{(\mathscr{O},4)}N_{\mathfrak{l}_i}^{d-2y_i(\alpha-\theta)}\cdot\prod_{\mathfrak{l}}N_\mathfrak{l}^{50\varepsilon}\cdot\big(\max_{\mathfrak{l}}N_\mathfrak{l}\big)^{50\varepsilon}. \tag{6.81}$$

In (6.81), the product $\prod_i^{(\mathscr{O},n)}$, for $2\leq n\leq 4$, is the product of $\prod_i^{(\mathscr{O},n,\geq)}$ defined in *Step 2* and $\prod_i^{(\mathscr{O},n,<)}$ defined in *Step 3* (there is no $\prod_i^{(\mathscr{O},2,<)}$, which is replaced by 1). In the last two factors in (6.81), the product is taken over all $\mathfrak{l}\in(G\cup G^\dagger)\cap\mathcal{Q}$, and the maximum is taken over all $\mathfrak{l}\in G\cup G^\dagger$ involved in *Cases 2–4* in *Step 2* and *Cases 3–4* in *Step 3*.

We make a few remarks regarding the calculations leading to (6.80)–(6.81):

(I) The factor $\prod_{\mathfrak{p}\in\widetilde{\mathcal{Y}}}N_\mathfrak{p}^{-\delta^3}$ in (6.80) is obtained from $N_*^{-\varepsilon^3/2}$ and $\prod_{\mathfrak{l}\in\mathcal{Y}_1\cup\cdots\cup\mathcal{Y}_r}N_\mathfrak{l}^{-\delta^3}$ coming from (6.78) and (6.76), given that $\widetilde{\mathcal{Y}}$ arises from the merging process as in Definition 3.6.

[1]In fact in the cases we consider the bound in (6.79) is always better than that in (6.78) due to the $\mathcal{X}_{1,1}$ factor in (6.79), so below we will always assume that we are using (6.78).

(II) By combining all terms involving $N_{\mathfrak{l}}^{\beta}$ or $N_{\mathfrak{l}}^{-\alpha+\theta}$ in (6.78), (6.76), (6.70) and (6.69), we obtain $\prod_{(\mathfrak{l}\in B\cup C)\backslash\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{\beta}$ in (6.80) with extra decays $\prod_{\mathfrak{l}\in((B\cup C)\cap G)\backslash\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{-\beta}$ and $\prod_{\mathfrak{l}\in\widetilde{\mathcal{F}}\cap G^{\dagger}} N_{\mathfrak{l}}^{-\beta}$ and $\prod_{\mathfrak{l}\in\mathcal{D}} N_{\mathfrak{l}}^{-(\alpha-\beta)+\theta}$.

(III) By comparing $\prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{8\varepsilon}$ in (6.80) with $\prod_{\mathfrak{l}\in G_1\cup\cdots\cup G_r} N_{\mathfrak{l}}^{8\varepsilon}$ coming from (6.78) and (6.76), we observe that the quotient between these two products arises from the $\mathfrak{l}$'s in $(G\cup G^{\dagger})\backslash E$. This quotient, multiplied by the extra decays $\prod_{\mathfrak{l}\in((B\cup C)\cap G)\backslash\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{-\beta}$ and $\prod_{\mathfrak{l}\in\widetilde{\mathcal{F}}\cap G^{\dagger}} N_{\mathfrak{l}}^{-\beta}$ obtained in (II), accounts for one square root of the factor $\prod_{\mathfrak{l}} N_{\mathfrak{l}}^{50\varepsilon}\cdot\big(\max_{\mathfrak{l}} N_{\mathfrak{l}}\big)^{50\varepsilon}$ in (6.81) (the other square root will be used in (IV)).

(IV) Recall from Definition 3.6 that $\mathcal{P}\subset\widetilde{\mathcal{P}}=\mathcal{P}_1\cup\cdots\cup\mathcal{P}_r\cup\mathcal{Q}$, also $\mathcal{R}^{\dagger}=\mathcal{Q}\backslash(\mathcal{D}\cup G\cup G^{\dagger})$. Considering $\prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon}$ in (6.80), we know that $\prod_{\mathfrak{l}\in\mathcal{P}_1\cup\cdots\cup\mathcal{P}_r\cup\mathcal{R}^{\dagger}} N_{\mathfrak{l}}^{-4\varepsilon}$ coming from (6.78) and (6.76), multiplied by the extra decay $\prod_{\mathfrak{l}\in\mathcal{D}} N_{\mathfrak{l}}^{-(\alpha-\beta)+\theta}$ obtained in (II), can be bounded by $\prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-4\varepsilon}$ multiplied by the other square root of $\prod_{\mathfrak{l}} N_{\mathfrak{l}}^{50\varepsilon}\cdot\big(\max_{\mathfrak{l}} N_{\mathfrak{l}}\big)^{50\varepsilon}$ in (6.81).

Now we have verified (6.80)–(6.81). Since $y_i\geq 1$ in *Cases 2–3* and $y_i\geq 2$ in *Case 4* (this holds in both *Step 2* and *Step 3*), it is easy to verify that $\mathfrak{Z}$ is a product of negative powers of $N_{\mathfrak{l}_i}$ in (6.81), thus $\mathfrak{Z}\leq 1$, hence (6.80) implies the desired estimate (6.59) for $H^{\circ}_{kk_{\mathcal{U}}}$. This finishes the proof of (6.49).

Finally we prove (6.50) ((6.52) follows from the same arguments as explained in *Step 1*). When $C=\varnothing$, we know that $h^{(1)}$ has type 1 and $C^{\dagger}\cap\mathcal{U}_1^{\dagger}=\varnothing$, thus we only need to use the modified version of the bound (6.45) for $h^{(1)}$. The same arguments as above yield, instead of (6.80), that

$$\prod_{j=r+1}^{q} N_j^{-d/2}\cdot\|H^{\circ}\|_{kk_B}\lesssim(\tau^{-1}M)^{\theta}(N_*)^{-\varepsilon^3/2}\prod_{j=1}^{r}\mathfrak{X}_j^{*}\cdot\prod_{\mathfrak{l}\in B}N_{\mathfrak{l}}^{\beta}\cdot\prod_{\mathfrak{l}\in\mathcal{P}}N_{\mathfrak{l}}^{-4\varepsilon}$$
$$\times\prod_{\mathfrak{l}\in E}N_{\mathfrak{l}}^{8\varepsilon}\cdot\prod_{\mathfrak{p}\in\widetilde{\mathcal{Y}}}N_{\mathfrak{p}}^{-\delta^3}\prod_{\mathfrak{f}\in\mathcal{V}_1\cup\cdots\cup\mathcal{V}_r}N_{\mathfrak{f}}^{d}\cdot\Big(N_1^{-\varepsilon}\prod_{j}^{(0)}N_j^{-2\varepsilon}\Big)\cdot\mathfrak{Z}. \tag{6.82}$$

Therefore, it suffices to prove that

$$\sqrt{\Upsilon}\cdot N_1^{-\varepsilon}\cdot\prod_{j}^{(0)}N_j^{-2\varepsilon}\lesssim M^{-\varepsilon},$$

which easily follows from the definition of $\Upsilon$, the product $\prod_j^{(0)}$, and the assumption that $N_1\gtrsim N_j$ for all $2\leq j\leq r$ such that $h^{(j)}$ has type 1. □

**Proposition 6.5** (Merged tensor bounds: Case II)**.** *Consider the same setting as Proposition 6.4, but assume $q=r$ and each $\mathcal{S}_j$ and $h^{(j)}$ have type 0. Moreover, assume that $N_{\mathfrak{l}}\geq M^{\delta}$ for each $\mathfrak{l}\in\mathcal{L}_j$ and $1\leq j\leq q$. Then, $H$ satisfies the bound*

$$\sqrt{\Upsilon}\cdot\|H\|_{kk_B\to k_C}\lesssim\mathfrak{Y}\cdot\tau^{-\theta}\prod_{\mathfrak{l}\in B\cup C}N_{\mathfrak{l}}^{\beta_1}\prod_{\mathfrak{l}\in\mathcal{P}}N_{\mathfrak{l}}^{-8\varepsilon}\prod_{\mathfrak{l}\in E}N_{\mathfrak{l}}^{4\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}}N_{\mathfrak{p}}^{-\delta^3}\cdot\mathcal{X}_0\mathcal{X}_1, \tag{6.83}$$

*where $\Upsilon$ is defined in (6.41), $\mathcal{X}_0,\mathcal{X}_1$ are defined as in (5.33) but with $N$ replaced by $M$, and*

$$\mathfrak{Y}=\prod_{j=1}^{q}\mathfrak{X}_j\cdot\tau^{-\theta}M^{-\delta^3}. \tag{6.84}$$

*Note that unlike Proposition 6.4, this bound (6.83) is deterministic, i.e. we do not need to remove any exceptional set.*

*Proof.* The proof will be similar to the proof of Proposition 6.4, so we mainly focus on the parts where the two proofs are different.

Since $q = r$ and each $\mathcal{S}_j$ and $h^{(j)}$ have type 0, we know that $\mathcal{V} = \varnothing$. Since also $N_{\mathfrak{l}} \geq M^{\delta}$ for each $\mathfrak{l}$, the possible trimming process in this proof will only affect the $\mathcal{Y}$ set and we may omit it; in particular we will not need Propositions 4.14, and can obtain (6.83) deterministically. Next, it can be easily verified that

$$\Upsilon^{1/4} \cdot \prod_{j=1}^{q} \mathcal{X}_{1,j} \lesssim \mathcal{X}_1$$

where $\mathcal{X}_1$ and $\mathcal{X}_{1,j}$ are defined as in (5.33) relative to $\mathcal{S}$ and $\mathcal{S}_j$ respectively, so in the proof below we will replace $\sqrt{\Upsilon}$ by $\Upsilon^{1/4}$, and replace $\mathcal{X}_1$ and $\mathcal{X}_{1,j}$ for all $j$ by 1.

By rearranging the tensors, we may assume that (i) if $C \neq \varnothing$, then $\max\{N_{\mathfrak{l}} : \mathfrak{l} \in C \cap \mathcal{U}_1\} \sim \max\{N_{\mathfrak{l}} : \mathfrak{l} \in C\}$, denote the particular $\mathfrak{l} \in C \cap \mathcal{U}_1$ where the maximum is attained by $\mathfrak{l}_{\mathrm{top}}$; (ii) if $C = \varnothing$ and $E \neq \varnothing$ then $E \cap \mathcal{U}_1 \neq \varnothing$; and (iii) if $C = E = \varnothing$ then $\min\{N_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{L}_1\} \sim \min\{N_{\mathfrak{l}} : \mathfrak{l} \in \mathcal{L}\}$, denote the particular $\mathfrak{l} \in C \cap \mathcal{L}_1$ where the minimum is attained by $\mathfrak{l}_{\mathrm{bot}}$. We shall repeat the same arguments in *Steps 2–3* of the proof of Proposition 6.4. Namely, we first remove the over-pairings as in *Step 2* (in particular, the *Cases 1–4* are the same as in *Step 2*, except that $\widetilde{\mathcal{U}}$ there is replaced by $\mathcal{U}$, and we do not require $i \leq v$ since there is no $v$ in the current case), then separate $h^{(1)}$ from the others as in *Step 3* up to (6.69). After these, we shall apply Proposition 6.3 (instead of Proposition 6.2 as in *Step 4* of the proof of Proposition 6.4) to get that

$$\begin{aligned}\|H\|_{kk_B \to k_C} \lesssim \tau^{-\theta} \prod_{j=1}^{q} \mathfrak{X}_j \cdot \prod_{i}^{(\mathscr{O},2)} N_{\mathfrak{l}_i}^{-2y_i(\alpha-\theta)} \prod_{i}^{(\mathscr{O},3)} N_{\mathfrak{l}_i}^{(d/2)-2y_i(\alpha-\theta)} \prod_{i}^{(\mathscr{O},4)} N_{\mathfrak{l}_i}^{d-2y_i(\alpha-\theta)} \\ \times \prod_{\mathfrak{l} \in \mathcal{D}} N_{\mathfrak{l}}^{-\alpha+\theta} \prod_{\mathfrak{l} \in \mathcal{U}_1^{\dagger}} N_{\mathfrak{l}}^{\beta_1} \prod_{\mathfrak{l} \in \mathcal{F}} N_{\mathfrak{l}}^{\beta_1} \cdot \prod_{\mathfrak{l} \in \mathcal{P}_1 \cup \cdots \mathcal{P}_q \cup \mathcal{R}} N_{\mathfrak{l}}^{-8\varepsilon} \prod_{\mathfrak{l} \in G_1 \cup \cdots \cup G_q} N_{\mathfrak{l}}^{4\varepsilon} \prod_{\mathfrak{p} \in \mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_q} N_{\mathfrak{p}}^{-\delta^3} \cdot \mathcal{X}\mathcal{X}'. \quad (6.85)\end{aligned}$$

Here in the above:

- The product $\prod_{i}^{(\mathscr{O},n)}$ for $2 \leq n \leq 4$, as well as the parameters $y_i$ and $z_i$, are defined in the same way as in *Step 2* of the proof of Proposition 6.4.
- The set $G_j = G \cap \mathcal{U}_j$, where $G$ is the union of all groups in *Cases 1–4*, and $\mathcal{U}_j^{\dagger} = \mathcal{U}_j \backslash G_j$.
- The set $\mathcal{R} = \mathcal{Q}^{\dagger} \backslash \mathcal{D}$, where $\mathcal{Q}^{\dagger}$ is the union of all $\mathcal{A}_i$'s with $|\mathcal{A}_i| = 2$ (equivalently $\mathcal{Q}^{\dagger} = \mathcal{Q} \backslash G$), and $\mathcal{D}$ is the union of all $\mathcal{A}_i$'s with $|\mathcal{A}_i| = 2$ and $\mathcal{A}_i \cap \mathcal{U}_1^{\dagger} \neq \varnothing$. The set $\mathcal{F} = (\mathcal{U}^{\dagger} \cup \mathcal{D}) \backslash \mathcal{U}_1^{\dagger}$ where $\mathcal{U}^{\dagger} = \mathcal{U} \backslash G$.
- The factor $\mathcal{X}$ is such that (i) if $C \neq \varnothing$ and $\mathfrak{l}_{\mathrm{top}} \in \mathcal{U}_1^{\dagger}$ then $\mathcal{X} = N_{\mathfrak{l}_{\mathrm{top}}}^{-\beta_1}$, (ii) if $C = G_1 = \varnothing$ (which implies $E = \varnothing$) then $\mathcal{X} = (N_{\mathfrak{l}_{\mathrm{bot}}})^{(d/2)-\beta_1}$, (iii) if $\mathcal{U}_1^{\dagger} = \varnothing$ then $\mathcal{X} = N_1^{-\varepsilon\delta}$, (iv) in other cases $\mathcal{X} = 1$.
- The factor $\mathcal{X}'$ is such that (i) if $\mathcal{U}_j^{\dagger} \neq \varnothing$ for some $2 \leq j \leq q$, then $\mathcal{X}' = (\max N_{\mathfrak{l}})^{-4p\varepsilon}$ where the maximum is taken over all $\mathfrak{l} \in \mathcal{U}_j^{\dagger}$ for all $2 \leq j \leq q$, (ii) if $\mathcal{U}_j^{\dagger} = \varnothing$ for all $2 \leq j \leq q$ then $\mathcal{X}' = (N_2 \cdots N_q)^{-\varepsilon\delta}$.

In the last two points above regarding $\mathcal{X}$ and $\mathcal{X}'$, the powers $N_{\mathfrak{l}_{\mathrm{top}}}^{-\beta_1}$ and $(N_{\mathfrak{l}_{\mathrm{bot}}})^{(d/2)-\beta_1}$ in cases (i) and (ii) for $\mathcal{X}$ are obtained from the first and second lines in the definition (5.33) of $\mathcal{X}_{0,j}$ (which is the factor appearing in (6.43)) with $j = 1$. The powers $N_1^{-\varepsilon\delta}$ in case (iii) for $\mathcal{X}$, and $(N_2 \cdots N_q)^{-\varepsilon\delta}$ in case (ii) for $\mathcal{X}'$ are obtained from the third line in the definition (5.33) of $\mathcal{X}_{0,j}$, which is only used here and not needed in other parts of the proof. Finally, the power $(\max N_{\mathfrak{l}})^{-4p\varepsilon}$ in case (i) for $\mathcal{X}'$ is obtained from the last term of (6.37).

Now, by the same arguments as in the proof of Proposition 6.4 (in particular using the quantity $\mathfrak{Z}$ in (6.81)), we can simplify the right hand side of (6.85) as follows. When $C \neq \varnothing$ we obtain that

$$\Upsilon^{1/4} \cdot \|H\|_{kk_B \to k_C} \lesssim \tau^{-\theta} \prod_{j=1}^{q} \mathfrak{X}_j \cdot \prod_{\mathfrak{l} \in (B\cup C)\setminus\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{\beta_1} \prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-8\varepsilon} \prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{4\varepsilon} \prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \cdot M^{\delta^3}\mathcal{X}', \tag{6.86}$$

where we notice by definition that $\mathcal{X}' \leq M^{-\varepsilon\delta^2} \leq M^{-4\delta^3}$. This easily implies (6.83). When $C = \varnothing$ we obtain that

$$\Upsilon^{1/4} \cdot \|H\|_{kk_B} \lesssim \tau^{-\theta} \prod_{j=1}^{q} \mathfrak{X}_j \cdot \prod_{\mathfrak{l}\in B} N_{\mathfrak{l}}^{\beta_1} \prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-8\varepsilon} \prod_{\mathfrak{l}\in E} N_{\mathfrak{l}}^{4\varepsilon} \prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \cdot M^{\delta^3}\mathcal{X}\mathcal{X}'; \tag{6.87}$$

in this case, if $E = \varnothing$, then we have $\mathcal{X} \leq (N_{\mathfrak{l}_{\mathrm{bot}}})^{(d/2)-\beta_1}$, which is acceptable as $\mathcal{X}_0 = (N_{\mathfrak{l}_{\mathrm{bot}}})^{(d/2)-\beta_1}$ in (6.83), and the bound $\mathcal{X}' \leq M^{-\varepsilon\delta^2}$ provides the needed gain as above; if $E \neq \varnothing$ and $B \neq \varnothing$, then $\mathcal{X} \leq 1$ and $\mathcal{X}_0 = 1$ in (6.83), so once again we can use the bound of $\mathcal{X}'$. Finally, if $B = \varnothing$, then in particular $\mathcal{U}_j^{\dagger} = \varnothing$ for each $1 \leq j \leq q$. In this case $\mathcal{X} = N_1^{-\varepsilon\delta}$, $\mathcal{X}' = (N_2 \cdots N_q)^{-\varepsilon\delta}$ and $\mathcal{X}_0 = M^{-\varepsilon\delta}$ in (6.83), so it suffices to prove that

$$M^{\delta^3}\Upsilon^{1/4} \prod_{j=1}^{q} N_j^{-\varepsilon\delta} \leq M^{-\delta^3} M^{-\varepsilon\delta}.$$

This is obviously true if $N_j \ll M$ for each $j$, due to the definition (6.41) of $\Upsilon$; if $N_j \sim M$ for some $j$, this is also true since the other $N_{j'}$ still satisfy $N_{j'}^{-\varepsilon\delta} \leq M^{-\varepsilon\delta^2}$. This completes the proof. □

**Proposition 6.6** (A special case: Operator bounds). *Let $3 \leq q \leq p$ be odd and $1 \leq r \leq q$. For $2 \leq j \leq r$, assume that $\Psi_{k_j}^{(j)} = \Psi_{k_j}[\mathcal{S}_j, h^{(j)}]$ as in (3.8), where $\mathcal{S}_j$ is a regular plant with $|\mathcal{S}_j| \leq D$ and $N(\mathcal{S}_j) = N_j \leq M$ such that $N_{\mathfrak{l}} \geq M^{\delta}$ for each $\mathfrak{l} \in \mathcal{L}_j$, $h^{(j)}$ is an $\mathcal{S}_j$-tensor that is $\mathcal{B}_{M^{[\delta]}}$ measurable. Moreover we assume that $h^{(j)}$ either has type 0 and satisfies the assumptions of Proposition 6.4 (1), or has type 1 and satisfies the assumptions of Proposition 6.4 (2).*

*We also fix $\zeta_j$ $(1 \leq j \leq q)$ and $N_j \leq M/2$ $(r+1 \leq j \leq q)$, and assume that $\max(N_2, \cdots, N_q) = M$. Let $h = h_{kk_1\cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q)$ be a constant tensor (which does not depend on $\omega$) supported in the set*

$$\langle k_j\rangle \leq N_j\,(2 \leq j \leq q), \quad \langle \lambda_j\rangle \leq 2M^{\kappa^2}\,(r+1 \leq j \leq q), \quad k = \sum_{j=1}^{q} \zeta_j k_j; \tag{6.88}$$

*also assume that $h$ can be written as a function of*[1] *$k - \zeta_1 k_1$, $|k|^2 - \zeta_1|k_1|^2$, and $(k_2, \cdots, k_q, \lambda_{r+1}, \cdots, \lambda_q)$, and satisfies that*

$$|h| + |\partial_{\lambda_j} h| \lesssim \frac{\tau^{-\theta}}{\langle \Omega + \zeta_{r+1}\lambda_{r+1} + \cdots + \zeta_q\lambda_q + \widetilde{\Xi}\rangle}, \quad r+1 \leq j \leq q, \tag{6.89}$$

[1]We may also need to multiply this $h$ by functions $\mathbf{1}_{\langle k\rangle \geq M^2}$ or $\mathbf{1}_{\langle k_1\rangle \geq M^2}$, but they do not affect Proposition 4.15 (which can be easily checked), so the proof below will proceed in the same way.

*where* $\Omega = |k|^2 - \sum_{j=1}^{q} \zeta_j |k_j|^2$, $\widetilde{\Xi} \in \mathbb{R}$ *is fixed with* $|\widetilde{\Xi}| \lesssim M^{\kappa^2}$, *and that any pairing in* $(k, k_1, \cdots, k_q)$ *must be over-paired. Now define*

$$\mathscr{M}_{kk_1} = \sum_{k_2,\cdots,k_q} \int \mathrm{d}\lambda_{r+1} \cdots \mathrm{d}\lambda_q \cdot h_{kk_1\cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q) \prod_{j=2}^{r} (\Psi_{k_j}^{(j)})^{\zeta_j} \prod_{j=r+1}^{q} (\widehat{z_{N_j}})_{k_j}^{\zeta_j}(\lambda_j), \tag{6.90}$$

*possibly with Fourier truncations on* $z_{N_j}$ *as in part (5) of Proposition 5.1, then* $\tau^{-1}M$*-certainly we have*

$$\|\mathscr{M}_{kk_1}\|_{k\to k_1} \le \prod_{j=2}^{r} \mathfrak{X}_j \cdot \tau^{-\theta} M^{-\varepsilon^4}. \tag{6.91}$$

*Proof.* The proof will be similar to the proof of Proposition 6.4, so we mainly focus on the parts where the two proofs are different.

Let $\mathcal{S}_1$ be the empty plant (each component being empty). For any $\mathscr{O}$ as in Definition 3.6, define

$$\mathcal{S} = (\mathcal{L}, \mathcal{V}, \mathcal{Y}) = \mathtt{Merge}(\mathcal{S}_1, \cdots, \mathcal{S}_r, \mathscr{B}, \mathscr{O}),$$

and define

$$\mathcal{H}_{kk_1k_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) = \prod_{\mathfrak{l},\mathfrak{l}'}^{(1)} \mathbf{1}_{k_{\mathfrak{l}}=k_{\mathfrak{l}'}} \prod_{\mathfrak{l},\mathfrak{l}'}^{(2)} \mathbf{1}_{k_{\mathfrak{l}}\neq k_{\mathfrak{l}'}} \cdot \sum_{(k_2,\cdots,k_r)} h_{kk_1\cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q)$$
$$\times \sum_{k_{\mathcal{Q}}}^{(3)} \prod_{\mathfrak{l}\in\mathcal{Q}} \Delta_{N_{\mathfrak{l}}} \gamma_{k_{\mathfrak{l}}} \prod_{j=2}^{r} \left[h_{k_jk_{\mathcal{U}_j}}^{(j)}(k_{\mathcal{V}_j}, \lambda_{\mathcal{V}_j})\right]^{\zeta_j}. \tag{6.92}$$

Here in (6.92), the set $\mathcal{Q}$ (as well as $\mathcal{U}$ etc. below) is defined from $\mathscr{O}$ in the same way as in Definition 3.6; the products $\prod_{\mathfrak{l},\mathfrak{l}'}^{(1)}$ and $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}$, and the sum $\sum_{k_{\mathcal{Q}}}^{(3)}$, are defined as in (3.11). By the same proof as Proposition 3.7 (2), we can write $\mathscr{M}_{kk_1}$ as a linear combination (for different choices of $\mathscr{O}$) of

$$\mathscr{N}_{kk_1} = \sum_{k_{\mathcal{U}},k_{\mathcal{V}}} \int \mathrm{d}\lambda_{\mathcal{V}} \cdot \mathcal{H}_{kk_1k_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) \cdot \prod_{\mathfrak{l}\in\mathcal{U}} (f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}^*} \prod_{\mathfrak{f}\in\mathcal{V}} (\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}^*}(\lambda_{\mathfrak{f}}), \tag{6.93}$$

where $\zeta_{\mathfrak{l}}^*$ and $\zeta_{\mathfrak{f}}^*$ are defined as in Definition 3.6 when merging $(\mathcal{S}_1, \cdots, \mathcal{S}_r)$ via $(\mathscr{B}, \mathscr{O})$. If we assume $\big||k|^2 - \zeta_1|k_1|^2\big| > M^{\kappa^3}$, then by (6.89) and (6.88), we have

$$|h| + |\partial_{\lambda_j} h| \lesssim \tau^{-\theta} M^{-\kappa^3}, \quad r+1 \le j \le q,$$

which easily implies (6.91) thanks to the dominant decay $M^{-\kappa^3}$. Therefore, below we will focus on the estimate for $\mathscr{N}$ with a fixed $\mathscr{O}$ as in Definition 3.6, and assume that $\big||k|^2 - \zeta_1|k_1|^2\big| \le M^{\kappa^3}$ (which will allow us to apply Proposition 4.15).

First, notice that $\mathcal{V} = \mathcal{V}_2 \cup \cdots \cup \mathcal{V}_r \cup \{r+1, \cdots, q\}$; define $H^* = (H^*)_{kk_1}(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ so that we have

$$(H^*)_{kk_1}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) = \sum_{k_{\mathcal{U}}} \mathcal{H}_{kk_1k_{\mathcal{U}}}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) \cdot \prod_{\mathfrak{l}\in\mathcal{U}} (f_{N_{\mathfrak{l}}})_{k_{\mathfrak{l}}}^{\zeta_{\mathfrak{l}}^*}, \tag{6.94}$$

$$\mathscr{N}_{kk_1} = \sum_{k_{\mathcal{V}}} \int \mathrm{d}\lambda_{\mathcal{V}} \cdot (H^*)_{kk_1}(k_{\mathcal{V}}, \lambda_{\mathcal{V}}) \cdot \prod_{\mathfrak{f}\in\mathcal{V}} (\widehat{z_{N_{\mathfrak{f}}}})_{k_{\mathfrak{f}}}^{\zeta_{\mathfrak{f}}^*}(\lambda_{\mathfrak{f}}). \tag{6.95}$$

By the same argument as in *Step 1* of the proof of Proposition 6.4, using Cauchy-Schwartz, we obtain that

$$\|\mathscr{N}_{kk_1}\|_{k\to k_1} \lesssim \|H^*\|_{X_{\mathcal{V}}^{-b_0}[k\to k_1]} \cdot \prod_{\mathfrak{f}\in\mathcal{V}} N_{\mathfrak{f}}^{-D_1}, \tag{6.96}$$

so it suffices to control $\|H^*\|_{X_{\mathcal{V}}^{-b_0}[k\to k_1]}$. For $2\leq j\leq r$, define $\mathfrak{X}_j^*=\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})$ as in the proof of Proposition 6.4. Using the inequality

$$\Big(\sum_{k_{\mathcal{V}}}\int \mathrm{d}\lambda_{\mathcal{V}}\cdot\prod_{\mathfrak{f}\in\mathcal{V}}\langle\lambda_{\mathfrak{f}}\rangle^{-2b_0}\prod_{j=2}^r\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})^2\Big)^{1/2}\lesssim\prod_{j=2}^r\mathfrak{X}_j\prod_{j=r+1}^q N_j^{d/2},$$

which is proved in the same way as (6.61), it suffices to prove $\tau^{-1}M$-certainly that

$$\prod_{j=r+1}^q N_j^{-d/2}\cdot\|(H^*)_{kk_1}(k_{\mathcal{V}},\lambda_{\mathcal{V}})\|_{k\to k_1}\lesssim\prod_{j=2}^r\mathfrak{X}_j^*(k_{\mathcal{V}_j},\lambda_{\mathcal{V}_j})\cdot\tau^{-\theta}M^{-\varepsilon^4}\prod_{\mathfrak{f}\in\mathcal{V}_2\cup\cdots\cup\mathcal{V}_r}N_{\mathfrak{f}}^d \tag{6.97}$$

for any choice of $(k_{\mathcal{V}},\lambda_{\mathcal{V}})$. By a meshing argument as in the proof of Proposition 6.4, we may fix a single choice of $(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ and view $h^{(j)}=h^{(j)}_{k_jk_{\mathcal{U}_j}}$ as depending only on $(k_j,k_{\mathcal{U}_j})$, $h=h_{kk_1\cdots k_r}$ as depending only on $(k,k_1,\cdots,k_r)$, $\mathcal{H}=\mathcal{H}_{kk_1k_{\mathcal{U}}}$ as depending only on $(k,k_1,k_{\mathcal{U}})$, and $(H^*)_{kk_1}$ as depending only on $(k,k_1)$. Also note the extra factor $\prod_{j=r+1}^q N_j^{-d/2}$ in the norm bound (6.97) that we want to prove for $(H^*)_{kk_1}$, which is analogous to *Step 1* of the proof of Proposition 6.4, and can be exploited in exactly the same way.

Now, using (6.94), noticing that $\mathcal{H}$ is $\mathcal{B}_{M^{[\delta]}}$ measurable and $N_{\mathfrak{l}}\geq M^\delta$ for $\mathfrak{l}\in\mathcal{U}$, that $h$ (and hence $\mathcal{H}$) depends on $(k,k_1)$ only via the quantities $k-\zeta_1k_1$ and $|k|^2-\zeta_1|k_1|^2$ and is supported in $||k|^2-\zeta_1|k_1|^2|\leq M^{\kappa^3}$, and that no pairing occurs in $k_{\mathcal{U}}$ in view of the product $\prod_{\mathfrak{l},\mathfrak{l}'}^{(2)}$ in the definition (6.92) of $\mathcal{H}$, we can apply Proposition 4.15 to (6.94) with $(b,c,A)=(k,k_1,\mathcal{U})$ and obtain that $\tau^{-1}M$-certainly,

$$\|H^*\|_{k\to k_1}\lesssim(\tau^{-1}M)^\theta\prod_{\mathfrak{l}\in\mathcal{U}}N_{\mathfrak{l}}^{-\alpha+\theta}\cdot\max_{(S,T)}\|\mathcal{H}\|_{kk_S\to k_1k_T}, \tag{6.98}$$

where $(S,T)$ is any partition of $\mathcal{U}$. Then, repeating the same arguments in *Steps 3–4* of the proof of Proposition 6.4 (namely, first removing the over-pairings as in *Step 3* after (6.70), then applying Proposition 6.2 as in *Step 4*—note that here we do not have the set $E$), we obtain (after omitting factors that are $\leq 1$) that

$$\prod_{j=r+1}^q N_j^{-d/2}\cdot\|\mathcal{H}\|_{kk_S\to k_1k_T}\lesssim\prod_{j=2}^r\mathfrak{X}_j^*\cdot\prod_{\mathfrak{f}\in\mathcal{V}_2\cup\cdots\cup\mathcal{V}_r}N_{\mathfrak{f}}^d\cdot\prod_{\mathfrak{l}\in\mathcal{U}}N_{\mathfrak{l}}^\beta\cdot M^{-2\varepsilon^4}. \tag{6.99}$$

Plugging (6.99) into (6.98) we get (6.97), as desired. □

## 7. Proof of Proposition 5.1

In this section we apply Propositions 6.2–6.6 to complete the inductive proof of Proposition 5.1. Namely, assuming $\mathtt{Local}(M)$, we shall prove that $\mathtt{Local}(2M)$ holds $\tau^{-1}M$-certainly. Recall the choice of $M$ and the basic assumptions and facts listed in the beginning of Section 6.

### 7.1. The operator $\mathscr{V}^M$.

We start by obtaining suitable bounds for the operator $\mathscr{V}^M$ (as well as $\mathscr{R}^M=\mathscr{V}^M+1$), which will follow from the corresponding bounds for $\mathscr{L}^M$, which in turn follow from the bounds for $\mathscr{L}^\zeta$ in part (5) of $\mathtt{Local}(M)$ in Proposition 5.1.

**Proposition 7.1.** *Assume* $\mathtt{Local}(M)$ *is true. Then* $\|\mathscr{L}^{M,\zeta}\|_{X^{1-b}\to X^b}\leq\tau^{(6\kappa)^{-1}}$ *for* $\zeta\in\{\pm\}$, *so in particular* $\mathscr{R}^M=(1-\mathscr{L}^M)^{-1}$ *and* $\|\mathscr{V}^{M,\zeta}\|_{X^{1-b}\to X^b}\leq\tau^{(7\kappa)^{-1}}$. *For the kernel* $\mathscr{L}^{M,\zeta}$ *we also have*

$$\int_{\mathbb{R}}\langle\lambda\rangle^{2(1-b)}\|\langle\lambda'\rangle^{-(1-b)}(\widehat{\mathscr{L}^{M,\zeta}})_{kk'}(\lambda,\lambda')\|^2_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}\,\mathrm{d}\lambda\leq\tau^{(3\kappa)^{-1}}, \tag{7.1}$$

*and the same bound (7.1) holds also for the kernel $\mathscr{V}^{M,\zeta}$, with the power $(3\kappa)^{-1}$ on the right hand side replaced by $(7\kappa/2)^{-1}$. Moreover for any $\widetilde{N}$ we have (recall that $\alpha_0$ and $\beta_1$ are defined in (3.1))*

$$\|\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}(\mathscr{V}^{M,\zeta})_{kk'}\|_{X^{b,-(1-b)}[kk']}\leq\tau^{(8\kappa)^{-1}}(\widetilde{N})^{\alpha_0}\cdot M^{C\delta},\tag{7.2}$$

$$\|(1+M^{-\delta}|k-\zeta k'|)^{\kappa^2}(\mathscr{V}^{M,\zeta})_{kk'}\|_{X^{b,-(1-b)}[kk']}\leq\tau^{(8\kappa)^{-1}}M^{\beta_1-\varepsilon}.\tag{7.3}$$

*Proof. Step 1: bounds for $\mathscr{L}^{M,\zeta}$.* For $L<M^\delta$, define the operators

$$(\mathscr{L}^{ML}w)_k(t)=-i\sum_{3\leq q\leq p}a_{pq}(m_M^*)^{(p-q)/2}\chi_\tau(t)\cdot\mathcal{I}_\chi\Pi_M\sum_{\mathrm{sym}}\mathcal{M}_q(w,v_L^\dagger,\cdots,v_L^\dagger)_k(t)\tag{7.4}$$

and $\widetilde{\mathscr{L}}^{ML}=\mathscr{L}^{ML}-\mathscr{L}^{M,L/2}$, then by definitions (7.4) and (5.19) we have

$$\mathscr{L}^{M,\zeta}=\mathscr{L}^{M,M^{[\delta]},\zeta}=\sum_{L<M^\delta}\widetilde{\mathscr{L}}^{ML,\zeta},$$

where the corresponding operators with $\zeta$ are defined as in Section 3.1. Moreover, each $\widetilde{\mathscr{L}}^{ML,\zeta}$ can be written as a superposition of at most $(\log L)^{C_\theta}$ operators of the form $\mathscr{L}^\zeta$, defined by (5.39) with this fixed $\zeta$, where $N$ is replaced by $M$ and $\max(N_2,\cdots,N_q)=L$. Therefore, to bound $\|\mathscr{L}^{M,\zeta}\|_{X^{1-b}\to X^b}$, it suffices to control the same norm for $\mathscr{L}^\zeta$ with a gain of a power of $L$.

Let the kernel of $\mathscr{L}^\zeta$ be $(\mathscr{L}^\zeta)_{kk'}(t,t')$, with Fourier transform $(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,\lambda')$, then by (3.5),

$$(\widehat{\mathscr{L}^\zeta w})_k(\lambda)=\sum_{k'}\int\mathrm{d}\lambda'\cdot(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,-\zeta\lambda')(\widehat{w})_{k'}(\lambda').$$

For any $w$ with $\|w\|_{X^b}=1$, we can estimate

$$\begin{aligned}\|\mathscr{L}^\zeta w\|_{X^{1-b}}^2&\leq\int_{\mathbb{R}}\langle\lambda\rangle^{2(1-b)}\,\mathrm{d}\lambda\cdot\bigg(\int_{\mathbb{R}}\|(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,-\zeta\lambda')\|_{k\to k'}\cdot\|(\widehat{w})_{k'}(\lambda')\|_{\ell^2_{k'}}\,\mathrm{d}\lambda'\bigg)^2\\&\leq\int_{\mathbb{R}^2}\langle\lambda\rangle^{2(1-b)}\langle\lambda'\rangle^{-2b}\|(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,-\zeta\lambda')\|_{k\to k'}^2\,\mathrm{d}\lambda\mathrm{d}\lambda'\cdot\|w\|_{X^b}^2,\end{aligned}$$

so we have by (5.40) that

$$\|\mathscr{L}^\zeta\|_{X^b\to X^{1-b}}\leq\|\mathscr{L}^\zeta\|_{X^{1-b,-b}[k\to k']}\leq\tau^{(5\kappa)^{-1}}L^{-4\varepsilon\delta}.\tag{7.5}$$

On the other hand, noticing that the $X^0$ and $X^1$ norms can be viewed as Sobolev $L^2$ and $H^1$ norms in the $t$ variable (for $\ell^2_k$ valued functions), and using the elementary inequalities leading to

$$\|\chi_\tau(t)\cdot\mathcal{I}_\chi v(t)\|_{H^1_t}\lesssim\|v(t)\|_{L^2_t}\tag{7.6}$$

for both scalar and vector valued functions, we can deduce that

$$\begin{aligned}\|\mathscr{L}^\zeta\|_{X^0\to X^1}&\lesssim\sup\big\{\|\mathcal{M}_q(y_{N_2}^*,\cdots,w,\cdots,y_{N_q}^*)\|_{X^0}:\|w\|_{X^0}=1\big\}\\&\lesssim\prod_{j=2}^q\|(\widehat{y_{N_j}^*})_{k_j}(\lambda_j)\|_{\ell^1_{k_j}L^1_{\lambda_j}}\lesssim\tau^{-\theta}L^{dp},\end{aligned}\tag{7.7}$$

where we have used $\texttt{Local}(M)$ parts (3) and (5) to control the norms of $y_{N_j}^*$. Interpolating (7.5) and (7.7) gives $\|\mathscr{L}^\zeta\|_{X^{1-b}\to X^b}\lesssim\tau^{(6\kappa)^{-1}}L^{-3\varepsilon\delta}$, which implies the desired bound for $\mathscr{L}^{M,\zeta}$. In particular we also get that $\mathscr{R}^M=(1-\mathscr{L}^M)^{-1}$ and the corresponding bound for $\mathscr{V}^{M,\zeta}$. Note that the estimates for $\mathscr{L}^\zeta$ in this step *do not require* that $L<M^\delta$.

*Step 2: more bounds for $\mathscr{L}^{\zeta}$.* We need to prove (7.1) for $\mathscr{L}^{M,\zeta}$. Clearly we may replace $\mathscr{L}^{M,\zeta}$ by $\mathscr{L}^{\zeta}$, provided we can prove (7.1) with right hand side replaced by $\tau^{(3\kappa)^{-1}}L^{-6\varepsilon\delta}$. For the purpose of (7.2)–(7.3) we will also prove an additional bound, which holds *assuming* $L < M^{\delta}$, namely:

$$\|\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}(\mathscr{L}^{\zeta})_{kk'}\|_{X^{b,-(1-b)}[kk']}\leq\tau^{(6\kappa)^{-1}}\min(\widetilde{N},M)^{\alpha_0}M^{C\delta}. \tag{7.8}$$

The proof of (7.8) is straightforward. In fact, since $N_j\leq L<M^{\delta}$ for $2\leq j\leq q$, we may expand the functions $y_{N_j}^*$ using their Fourier transforms which satisfy the $\ell_{k_j}^1L_{\lambda_j}^1$ bounds as in *Step 1*, and then reduce to fixed values of $k_j$ and $\lambda_j$ at a loss of $M^{C\delta}$. Using also Lemma 4.2 (note that $(\mathscr{L}^{\zeta})_{kk'}(0,t')\equiv 0$), we may get rid of the $\chi_\tau$ factor at the price of replacing the $X^{b,-(1-b)}[kk']$ norm by the slightly larger $X^{b^+,-(1-b)}[kk']$ norm. With these reductions, let the resulting operator be $\mathscr{L}^{*,\zeta}$, then by Lemma 4.1 we have

$$|(\widehat{\mathscr{L}^{*,\zeta}})_{kk'}(\lambda,\lambda')|\lesssim\mathbf{1}_{\langle k\rangle\leq M}\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}\mathbf{1}_{k-\zeta k'=k^*}\left(\frac{1}{\langle\lambda\rangle^3}+\frac{1}{\langle\lambda\pm(\lambda'\pm\Omega\pm\lambda^*)\rangle^3}\right)\frac{1}{\langle\lambda'\pm\Omega\pm\lambda^*\rangle}, \tag{7.9}$$

where $k^*$ and $\lambda^*$ are fixed, and $\Omega=|k|^2-\zeta|k'|^2$. The bound (7.8) for $\mathscr{L}^{*,\zeta}$ then follows from (7.9), elementary integral bounds and the fact that the number of choices for $(k,k')$ with $\langle k\rangle\leq M$ and $\langle k'\rangle\leq\widetilde{N}$, and the values of $k-\zeta k'=k^*$ and $|k|^2-\zeta|k'|^2=\Omega$ fixed, under the simplicity assumption[1], is $\lesssim\min(\widetilde{N},M)^{d-1}\leq\min(\widetilde{N},M)^{2\alpha_0}$.

Now we prove (7.1) for $\mathscr{L}^{\zeta}$ with right hand side replaced by $\tau^{(3\kappa)^{-1}}L^{-6\varepsilon\delta}$. Since

$$\|\langle\lambda'\rangle^{-b}(\widehat{\mathscr{L}^{\zeta}})_{kk'}(\lambda,\lambda')\|^2_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}\leq\int_{\mathbb{R}}\langle\lambda'\rangle^{-2b}\|(\widehat{\mathscr{L}^{\zeta}})_{kk'}(\lambda,\lambda')\|^2_{\ell^2_{k'}\to\ell^2_k}\,\mathrm{d}\lambda'$$

and thanks to (5.40), we know that the desired bound is true with $\langle\lambda'\rangle^{-(1-b)}$ replaced by $\langle\lambda'\rangle^{-b}$, and the right hand side replaced by $\tau^{(5\kappa/2)^{-1}}L^{-8\varepsilon\delta}$, in (7.1). By interpolation, it then suffices to verify that

$$\int_{\mathbb{R}}\langle\lambda\rangle^{2(1-b)}\|(\widehat{\mathscr{L}^{\zeta}})_{kk'}(\lambda,\lambda')\|^2_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}\,\mathrm{d}\lambda\leq\tau^{-\theta}L^{2dp}. \tag{7.10}$$

Here we first use Lemma 4.2 (again since $(\mathscr{L}^{\zeta})_{kk'}(0,t')\equiv 0$) to get rid of the $\chi_\tau$ factor in $\mathscr{L}^{\zeta}$, then apply the same arguments as in (7.7) in *Step 1* to control the spacetime $L^2$ norm of $\mathcal{M}_q(y_{N_2}^*,\cdots,w,\cdots,y_{N_q}^*)$, and reduce (7.10) to the bound (viewing $w$ as an $\ell_k^2$ valued function)

$$|\mathcal{F}_t\mathcal{I}_\chi w(\lambda)|\lesssim\langle\lambda\rangle^{-1}\|w(\lambda')\|_{L^2_{\lambda'}},$$

which easily follows from Lemma 4.1.

*Step 3: bounds for $\mathscr{V}^M$.* Now we can prove (7.1)–(7.3) for $\mathscr{V}^{M,\zeta}$. First look at (7.2)–(7.3); note that $\mathscr{V}^M=\mathscr{L}^M+\mathscr{L}^M\mathscr{V}^M$, so in terms of kernels we have

$$(\widehat{\mathscr{V}^{M,\zeta}})_{kk'}(\lambda,\lambda')=(\widehat{\mathscr{L}^{M,\zeta}})_{kk'}(\lambda,\lambda')+\sum_{\substack{\iota_1,\iota_2\in\{\pm\}\\ \iota_1\iota_2=\zeta}}\sum_m\int\mathrm{d}\mu\cdot(\widehat{\mathscr{L}^{M,\iota_1}})_{km}(\lambda,\mu)(\widehat{\mathscr{V}^{M,\iota_2}})^{\iota_2}_{mk'}(-\mu,\lambda').$$

[1]See Definition 3.1. Here simplicity implies that, if $\zeta=+$ and $k=k'$, then $k$ must also equal some other $k_j$, which has already been fixed.

We may multiply by the truncation $\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}$ on both sides; then, by fixing $(k',\lambda')$ and applying the kernel $(\widehat{\mathscr{L}^{M,\iota_1}})_{km}(\lambda,\mu)$ to $(\widehat{\mathscr{V}^{M,\iota_2}})^{\iota_2}_{mk'}(-\mu,\lambda')$ as a function of $(m,\mu)$, we obtain that

$$\begin{aligned}\sum_{\zeta\in\{\pm\}}\|\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}(\mathscr{V}^{M,\zeta})_{kk'}\|_{X^{b,-(1-b)}[kk']}&\leq\sum_{\zeta\in\{\pm\}}\|\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}(\mathscr{L}^{M,\zeta})_{kk'}\|_{X^{b,-(1-b)}[kk']}\\&+\sum_{\iota_1\in\{\pm\}}\|\mathscr{L}^{M,\iota_1}\|_{X^b\to X^b}\cdot\sum_{\iota_2\in\{\pm\}}\|\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}\mathscr{V}^{M,\iota_2}_{kk'}\|_{X^{b,-(1-b)}[kk']}.\end{aligned}\tag{7.11}$$

Using (7.8) for $\mathscr{L}^{M,\zeta}$, and the estimates for $\mathscr{L}^{M,\zeta}$ obtained in *Step 1*, we get (7.2). The proof for (7.3) is similar, where we use Lemma 4.3 (with $\kappa_1=\kappa^2$) to control the weighted norm of $\mathscr{V}^M$, noticing that $(\widehat{\mathscr{L}^{M,\iota_1}})_{km}(\lambda,\mu)$ is supported in $|k-\iota_1 m|\lesssim M^\delta$.

Finally we prove (7.1) for $\mathscr{V}^{M,\zeta}$. Since $\mathscr{V}^M=\mathscr{L}^M+\mathscr{V}^M\mathscr{L}^M$, similar to the above argument we can write

$$(\widehat{\mathscr{V}^{M,\zeta}})_{kk'}(\lambda,\lambda')=(\widehat{\mathscr{L}^{M,\zeta}})_{kk'}(\lambda,\lambda')+\sum_{\substack{\iota_1,\iota_2\in\{\pm\}\\ \iota_1\iota_2=\zeta}}\sum_m\int\mathrm{d}\mu\cdot(\widehat{\mathscr{V}^{M,\iota_1}})_{km}(\lambda,\mu)(\widehat{\mathscr{L}^{M,\iota_2}})^{\iota_2}_{mk'}(-\mu,\lambda').$$

This implies, for any fixed $\lambda$, that

$$\begin{aligned}\sum_{\zeta\in\{\pm\}}\|\langle\lambda'\rangle^{-(1-b)}(\widehat{\mathscr{V}^{M,\zeta}})_{kk'}(\lambda,\lambda')\|_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}&\leq\sum_{\zeta\in\{\pm\}}\|\langle\lambda'\rangle^{-(1-b)}(\widehat{\mathscr{L}^{M,\zeta}})_{kk'}(\lambda,\lambda')\|_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}\\&+\sum_{\iota_2\in\{\pm\}}\|\mathscr{L}^{M,\iota_2}\|_{X^{1-b}\to X^{1-b}}\cdot\sum_{\iota_1\in\{\pm\}}\|\langle\mu\rangle^{-(1-b)}(\widehat{\mathscr{V}^{M,\iota_1}})_{km}(\lambda,\mu)\|_{\ell^2_m L^2_\mu\to\ell^2_k}.\end{aligned}\tag{7.12}$$

Using (7.1) for $\mathscr{L}^{M,\zeta}$, and the estimates for $\mathscr{L}^{M,\zeta}$ obtained in *Step 1*, we get (7.1) for $\mathscr{V}^{M,\zeta}$. □

7.2. **The $h^{(\mathcal{S},0)}$ tensors.** In this section we prove part (1) of `Local`$(2M)$.

**Proposition 7.2.** *Assume* `Local`$(M)$ *is true. Then part (1) of* `Local`$(2M)$ *is true. More precisely,* $h^{(\mathcal{S},0)}$ *satisfies (5.30) and (5.32), for each plain regular plant* $\mathcal{S}$ *with* $N(\mathcal{S})=M$ *and* $|\mathcal{S}|\leq D$.

*Proof.* We induct in $|\mathcal{S}|$, using the inductive definition (5.26). For the first term on the right hand of (5.26), which corresponds to the mini-plant $\mathcal{S}=\mathcal{S}_M^+$, the desired bounds are obvious, so we just need to consider the second term, which is a multilinear expression of the input tensors $h^{(\mathcal{S}_j,0)}$. By induction hypothesis, each input tensor satisfies (5.30) and (5.32) associated with $\mathcal{S}_j$. Recall also that $\mathcal{V}_j=\mathcal{V}=\varnothing$ for $1\leq j\leq q$ when considering $h^{(\mathcal{S},0)}$ tensors.

First, to prove (5.30) for $h^{(\mathcal{S},0)}$, we notice that the sign of $\mathfrak{l}\in\mathcal{U}$ in $\mathcal{S}$ is given by $\zeta^*_{\mathfrak{l}}=\zeta_j\zeta_{\mathfrak{l}}$ where $\mathfrak{l}\in\mathcal{U}_j$ and $\zeta_{\mathfrak{l}}$ is the sign of $\mathfrak{l}$ in $\mathcal{S}_j$ (see Definition 3.6). In the support of $\mathcal{S}$ we have

$$\sum_{\mathfrak{l}\in\mathcal{U}}\zeta^*_{\mathfrak{l}}k_{\mathfrak{l}}=\sum_{\mathfrak{l}\in\mathcal{W}}\zeta^*_{\mathfrak{l}}k_{\mathfrak{l}}=\sum_{j=1}^q\zeta_j\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}=\sum_{j=1}^q\zeta_jk_j=k,\tag{7.13}$$

where $\mathcal{W}=\mathcal{U}_1\cup\cdots\cup\mathcal{U}_q$, using the induction hypothesis, and the definition (5.25) of the tensor $h$ used in the merging process. Now let us prove (5.32) for $h^{(\mathcal{S},0)}$.

*Step 1: first reductions.* By definition of $\sum_{(a)}$ in (5.26), we know $N_{\mathfrak{l}}\geq M^\delta$ for each $\mathfrak{l}\in\mathcal{L}_j$ and $1\leq j\leq q$, so we can omit the trimmings in (5.24), as they involve only the $\mathcal{Y}$ sets which do not appear in the tensors. Applying Lemma 4.2, we may remove the localization factor $\chi_\tau(t)$ on the right hand side of (5.26) and gain a power $\tau^{8\kappa^{-1}}$ (which would overwhelm any possible $\tau^{-\theta}$ loss), provided we estimate this expression without $\chi_\tau$ in the stronger norm with the power $\langle\lambda\rangle^{2b}$ in (5.32) replaced by $\langle\lambda\rangle^{2b^+}$. By abusing notation we will still use $h^{(\mathcal{S},0)}$ to denote the expression after these reductions; moreover, since $\sum_{\mathrm{sym}}$ and $\sum_{(a)}$ involve at most $(\log M)^\kappa$ terms, we may focus on one single term in the discussion below.

With the above reductions, and applying also Lemma 4.1 and the definition (5.25) of $h$, we can then take the Fourier transform in time and obtain that

$$\widehat{h^{(\mathcal{S},0)}}(\lambda)=\int \mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_q\cdot H(\lambda,\lambda_1,\cdots,\lambda_q), \tag{7.14}$$

where $H(\lambda,\lambda_1,\cdots,\lambda_q)=[H(\lambda,\lambda_1,\cdots,\lambda_q)]_{kk_{\mathcal{U}}}$ is a tensor (with $(\lambda,\lambda_1,\cdots,\lambda_q)$ being parameters) defined by

$$H(\lambda,\lambda_1,\cdots,\lambda_q)=\Upsilon\cdot\mathtt{Merge}(\widehat{h^{(\mathcal{S}_1,0)}}(\lambda_1),\cdots,\widehat{h^{(\mathcal{S}_q,0)}}(\lambda_q),\widetilde{h},\mathscr{B},\mathscr{O}). \tag{7.15}$$

In the above formula $(\Upsilon,\mathscr{B},\mathscr{O})$ are as in Section 5.2, $\widetilde{h}=[\widetilde{h}(\lambda,\lambda_1,\cdots,\lambda_q)]_{kk_1\cdots k_q}$ is a function of $(k,k_1,\cdots,k_q)$ with parameters $(\lambda,\lambda_1,\cdots,\lambda_q)$, that is supported in the set $k=\sum_{j=1}^q\zeta_jk_j$, and satisfies the bound

$$|\widetilde{h}|\lesssim\frac{\tau^{-\theta}}{\langle\lambda\rangle\langle\lambda-\Omega-\zeta_1\lambda_1-\cdots-\zeta_q\lambda_q\rangle};\quad \Omega:=|k|^2-\sum_{j=1}^q\zeta_j|k_j|^2. \tag{7.16}$$

We shall separate two cases: the *high modulation case* where $|\lambda|\geq M^{\sqrt{\kappa}}$, and the *low modulation case* where $|\lambda|\leq M^{\sqrt{\kappa}}$.

*Step 2: the high modulation case.* Assume $|\lambda|\geq M^{\sqrt{\kappa}}$. If we can estimate the norm in (5.32) for $h^{(\mathcal{S},0)}$, but with $\langle\lambda\rangle^{2b^+}$ replaced by $\langle\lambda\rangle^2$, then this gain of power in $\lambda$ will overwhelm any possible loss coming from any summation of any $k_j$ and $k_{\mathfrak{l}}$ variables (the latter summation loses at most a power $M^{C\cdot D}$, while $\sqrt{\kappa}\gg D$). Because of this we can fix the values of $k$, $k_j$ and all $k_{\mathfrak{l}}$, and view $\widehat{h^{(\mathcal{S},0)}}(\lambda)$ as a function of $\lambda$ only, and $\widehat{h^{(\mathcal{S}_j,0)}}(\lambda_j)$ as a function of $\lambda_j$ only. Moreover by induction hypothesis (5.32) and Hölder, this function of $\lambda_j$ can be controlled in $L^1_{\lambda_j}$, so upon integrating in $\lambda_j$, we can also fix the value of $\lambda_j$, in which case $\widehat{h^{(\mathcal{S},0)}}(\lambda)$ satisfies the bound

$$|\widehat{h^{(\mathcal{S},0)}}(\lambda)|\lesssim\frac{\tau^{-\theta}}{\langle\lambda\rangle\langle\lambda-\Xi\rangle}$$

due to (7.16), where $\Xi$ is a fixed real number depending on the choices of $k$, $k_j$, $k_{\mathfrak{l}}$ and $\lambda_j$. Clearly this implies $\int_{\mathbb{R}}\langle\lambda\rangle^2|\widehat{h^{(\mathcal{S},0)}}(\lambda)|^2\mathrm{d}\lambda\lesssim\tau^{-\theta}$ uniformly in all choices of the fixed parameters, so the desired estimate (5.32) follows.

*Step 3: the low modulation case.* Assume $|\lambda|\leq M^{\sqrt{\kappa}}$. Recall the formula (7.14); we shall further decompose $h^{(\mathcal{S},0)}$ into $h^{(\mathcal{S},0,\Gamma)}$ and $h^{(\mathcal{S}_j,0)}$ into $h^{(\mathcal{S}_j,0,\Gamma_j)}$, where $\Gamma$ and $\Gamma_j$ are integers, as in part (1) of Proposition 5.1. By induction hypothesis (5.32), if we define $\mathfrak{X}_j=\mathfrak{X}_j(\lambda_j,\Gamma_j)$ to be the smallest value such that $\widehat{h^{(\mathcal{S}_j,0,\Gamma_j)}}(\lambda_j)$ satisfies the type 0 bounds (6.42)–(6.43) in Proposition 6.4, then

$$\bigg(\sum_{\Gamma_j}\int_{\mathbb{R}}\mathfrak{X}_j(\lambda_j,\Gamma_j)\,\mathrm{d}\lambda_j\bigg)^2\lesssim\int_{\mathbb{R}}\langle\lambda_j\rangle^{2b}\bigg(\sum_{\Gamma_j}\mathfrak{X}(\lambda_j,\Gamma_j)\bigg)^2\mathrm{d}\lambda_j\lesssim 1. \tag{7.17}$$

For fixed values of $(\lambda,\lambda_j)$ and $(\Gamma,\Gamma_j)$, if we replace $h^{(\mathcal{S}_j,0)}$ by $h^{(\mathcal{S}_j,0,\Gamma_j)}$ in (7.15), restrict to the set $|k|^2-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}^*|k_{\mathfrak{l}}|^2=\Gamma$, and denote the resulting tensor by $H^{(\Gamma,\Gamma_1,\cdots,\Gamma_q)}(\lambda,\lambda_1,\cdots,\lambda_q)$, then by similar arguments as in (7.13) (but with $k_{\mathfrak{l}}$ replaced by $|k_{\mathfrak{l}}|^2$), we know that in this situation, we may further restrict $\widetilde{h}$ to the set $|k|^2 - \sum_{j=1}^{q} \zeta_j |k_j|^2 = \Gamma - \widetilde{\Gamma}$ in (7.15), where $\widetilde{\Gamma}$ depends only on the fixed parameters $\Gamma_j$. Therefore we can apply Proposition 6.5, also using (7.16), to deduce that

$$\begin{aligned}&\|H^{(\Gamma,\Gamma_1,\cdots,\Gamma_q)}(\lambda,\lambda_1,\cdots,\lambda_q)\|_{kk_B\to k_C}\\&\lesssim \tau^{-\theta}M^{-\delta^3}\langle\lambda\rangle^{-1}\langle\lambda-\Gamma+\Xi\rangle^{-1}\prod_{j=1}^{q}\mathfrak{X}_j(\lambda_j,\Gamma_j)\cdot\prod_{\mathfrak{l}\in B\cup C}N_{\mathfrak{l}}^{\beta_1}\prod_{\mathfrak{l}\in\mathcal{P}}N_{\mathfrak{l}}^{-8\varepsilon}\prod_{\mathfrak{l}\in E}N_{\mathfrak{l}}^{4\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}}N_{\mathfrak{p}}^{-\delta^3}\cdot\mathcal{X}_0\mathcal{X}_1,\end{aligned}\tag{7.18}$$

where $\mathcal{X}_0$ and $\mathcal{X}_1$ are as in (5.33), and $\Xi\in\mathbb{R}$ is a quantity depending only on the fixed parameters $\lambda_j$ and $\Gamma_j$. Note that here no meshing argument is required since Proposition 6.5 holds deterministically. Then, after summing in $\Gamma$, then taking the weighted $L^2$ norm in $\lambda$ within the set $|\lambda|\leq M^{\sqrt{\kappa}}$, then integrating in $\lambda_j$ and summing in $\Gamma_j$ using (7.17), we deduce that

$$\int_{\mathbb{R}}\langle\lambda\rangle^{2b^+}\bigg(\sum_{\Gamma\in\mathbb{Z}}\|\widehat{h^{(\mathcal{S},0,\Gamma)}}(\lambda)\|_{kk_B\to k_C}\bigg)^2\mathrm{d}\lambda\lesssim\bigg(\tau^{-\theta}M^{-\delta^4}\prod_{\mathfrak{l}\in B\cup C}N_{\mathfrak{l}}^{\beta_1}\prod_{\mathfrak{l}\in\mathcal{P}}N_{\mathfrak{l}}^{-8\varepsilon}\prod_{\mathfrak{l}\in E}N_{\mathfrak{l}}^{4\varepsilon}\prod_{\mathfrak{p}\in\mathcal{Y}}N_{\mathfrak{p}}^{-\delta^3}\cdot\mathcal{X}_0\mathcal{X}_1\bigg)^2,$$

which implies (5.32) in view of Lemma 4.2. □

### 7.3. The $h^{(\mathcal{S},1)}$ tensors.
In this section we prove part (2) of $\texttt{Local}(2M)$.

**Proposition 7.3.** *Assume* $\texttt{Local}(M)$ *and part (1) of* $\texttt{Local}(2M)$ *are true. Then $\tau^{-1}M$-certainly, part (2) of* $\texttt{Local}(2M)$ *is true. More precisely, $h^{(\mathcal{S},1)}$ satisfies (5.34)–(5.37), for each regular plant $\mathcal{S}$ with $N(\mathcal{S})=M$ and $|\mathcal{S}|\leq D$.*

*Proof.* Again we proceed by induction on $|\mathcal{S}|$, using the inductive definition (5.27). We first focus on the *main case*, namely the estimates (5.34)–(5.36) for the second line of (5.27), assuming the second maximum of $N_j$ $(1\leq j\leq q)$ is $\geq M^{\delta}$; then we will treat the remaining estimates. By induction hypothesis, if $n_j=0$ then $h^{(\mathcal{S}_j,n_j)}$ satisfies (5.30) and (5.32) associated with $\mathcal{S}_j$; if $n_j=1$ it satisfies the bounds (5.34)–(5.37) associated with $\mathcal{S}_j$. In various steps below, we will abuse notation and refer to some components of $h^{(\mathcal{S},1)}$ in (5.27) still as $h^{(\mathcal{S},1)}$ for simplicity.

*Step 1: the main case.* We start with the second line of (5.27), assuming the second maximum of $N_j$ $(1\leq j\leq q)$ is $\geq M^{\delta}$. Here we will prove (5.34)–(5.36) with all norms replaced by the stronger ones $X_{\mathcal{V}}^{b,-b_0}[\cdots]$. By Lemma 4.2, we can get rid of the $\chi_\tau(t)$ localization with a gain of $\tau^{8\kappa^{-1}}$, as long as we estimate the expression without $\chi_\tau$ in the $X_{\mathcal{V}}^{b^+,-b_0}[\cdots]$ norms. Repeating the proof of Proposition 7.2, we can reduce to

$$\widehat{h^{(\mathcal{S},1)}}(\lambda)=\int\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_r\cdot H(\lambda,\lambda_1,\cdots,\lambda_r)\tag{7.19}$$

in the same way as (7.14), but instead of (7.15) we have

$$H(\lambda,\lambda_1,\cdots,\lambda_r)=\Upsilon\cdot\texttt{Trim}(\texttt{Merge}(\texttt{Trim}(\widehat{h^{(\mathcal{S}_1,n_1)}}(\lambda_1),M^{\delta}),\cdots,\texttt{Trim}(\widehat{h^{(\mathcal{S}_r,n_r)}}(\lambda_r),M^{\delta}),\widetilde{h},\mathscr{B},\mathscr{O}),M^{\delta}),\tag{7.20}$$

where $\widetilde{h}=[\widetilde{h}(\lambda,\lambda_1,\cdots,\lambda_r)]_{kk_1\cdots k_q}(\lambda_{r+1},\cdots,\lambda_q)$ satisfies the same bound (7.16), as does any $\lambda_j$ derivative of $\widetilde{h}$ for $r+1\leq j\leq q$.

Similar to the proof of Proposition 7.2, we shall consider two cases, the high modulation case where $\max(|\lambda|,|\lambda_1|,\cdots,|\lambda_r|)\geq M^{\sqrt{\kappa}}$, and the low modulation case where $\max(|\lambda|,|\lambda_1|,\cdots,|\lambda_r|)\leq M^{\sqrt{\kappa}}$. In the high-modulation case, we may again fix the values of $k$, $k_j$ and all $k_{\mathfrak{l}}$ and $k_{\mathfrak{f}}$; then $\widehat{h^{(\mathcal{S}_j,n_j)}}(\lambda_j)$ can be viewed as a function of $\lambda_j$ and $\lambda_{\mathcal{V}_j}$ only, and $\widehat{h^{(\mathcal{S},1)}}(\lambda)$ can be viewed as a function of $\lambda$ and $\lambda_{\mathcal{V}}$ only, where recall $\mathcal{V}$ is associated to the plant $\mathcal{S}$ in (5.23). The trimming steps follow easily from Cauchy-Schwartz as in the proof of Proposition 6.1, so we may omit them and consider only the merging step. We may take the $X_{\mathcal{V}_j}^{-b_0}$ norm for $\widehat{h^{(\mathcal{S}_j,n_j)}}(\lambda_j)$ and denote the result by $\mathfrak{h}_j(\lambda_j)$; similarly we may take the $X_{\mathcal{V}}^{-b_0}$ norm of $\widehat{h^{(\mathcal{S},1)}}(\lambda)$ and denote it by $\mathfrak{h}(\lambda)$. Then each $\mathfrak{h}_j$ is bounded in a weighted $L^2$ space embedded in $L^1_{\lambda_j}$, so we may fix the value of $\lambda_j$ as in the proof of Proposition 7.2 whenever wanted. Now if $|\lambda|=\max(|\lambda|,|\lambda_1|,\cdots,|\lambda_r|)$, then the same argument as in *Step 2* of the proof of Proposition 7.2 works and implies the desired bound (with significant decay) for $\int_{\mathbb{R}}\langle\lambda\rangle^{2b^+}|\mathfrak{h}(\lambda)|^2\mathrm{d}\lambda$ uniformly in all choices of the fixed parameters; instead, if (say) $|\lambda_1|$ is the maximum, then we can fix $\lambda_j$ ($j\geq 2$), also using the definition of $X_{\mathcal{V}}^{-b_0}$ norm, to get that

$$|\mathfrak{h}(\lambda)|^2\lesssim\tau^{-\theta}\int\bigg(\frac{1}{\langle\lambda\rangle}\int_{\mathbb{R}}\frac{1}{\langle\lambda\pm\lambda_1\pm\Xi\rangle}|\mathfrak{h}_1(\lambda_1)|\,\mathrm{d}\lambda_1\bigg)^2\cdot\prod_{j=r+1}^{q}\langle\lambda_j\rangle^{-2b_0}\,\mathrm{d}\lambda_{r+1}\cdots\mathrm{d}\lambda_q,$$

where $\Xi$ is a real number depending on $(\lambda_{r+1},\cdots,\lambda_q)$, and the choices of the fixed variables. We can then fix $(\lambda_{r+1},\cdots,\lambda_q)$, estimate the $\lambda_1$ integral using Cauchy-Schwartz and the $L^2$ norm of $\mathfrak{h}_1(\lambda_1)$, save the $\langle\lambda_1\rangle^b$ weight to gain an $M^{\sqrt{\kappa}}$ power, and bound $\int_{\mathbb{R}}\langle\lambda\rangle^{2b^+}|\mathfrak{h}^*(\lambda)|^2\mathrm{d}\lambda$ (with significant decay) uniformly in $(\lambda_{r+1},\cdots,\lambda_q)$ and all choices of the fixed variables, where $\mathfrak{h}^*(\lambda)$ is the above integral in $\lambda_1$. This implies (5.34)–(5.35).

As for (5.36), just notice that $\widetilde{h}$ is supported in $k=\sum_{j=1}^q\zeta_jk_j$, which implies that

$$1+\frac{1}{M^{2\delta}}\bigg|k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}^*k_{\mathfrak{l}}-\ell\bigg|\lesssim\max_{1\leq j\leq r}\bigg(1+\frac{1}{M^{2\delta}}\bigg|k_j-\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell_j\bigg|\bigg),\tag{7.21}$$

where $\ell_j=\sum_{\mathfrak{f}\in\mathcal{V}_j}\zeta_{\mathfrak{f}}k_{\mathfrak{f}}$ and $\ell=\sum_{\mathfrak{f}\in\mathcal{V}}\zeta_{\mathfrak{f}}^*k_{\mathfrak{f}}$ (note that trimming at frequency $M^\delta$ or lower will not affect this inequality). This allows to control the weight in (5.36) for $h^{(\mathcal{S},1)}$ by the weights in (5.36) for $h^{(\mathcal{S}_j,n_j)}$, so the $M^{\sqrt{\kappa}}$ power gain above also implies (5.36).

From now on we can restrict to the low modulation case. We first look at (5.34)–(5.35); the proof of (5.36) requires slightly different arguments and is left to the end of this step. Recall the bounds (5.34)–(5.37) for the norms $X_{\mathcal{V}_j}^{\widetilde{b},-b_0}[\cdots]$ where $1\leq j\leq r$ and $\widetilde{b}\in\{b,1-b\}$; since $|\lambda_j|\leq M^{\sqrt{\kappa}}$, we may replace $\widetilde{b}$ by $b$ in all these bounds, at a price of $M^{C/\sqrt{\kappa}}$ which in the end will be negligible as $\kappa\gg_{C_\delta}1$. Suppose we want to estimate the $X_{\mathcal{V}}^{b^+,-b_0}[kk_B\to k_C]$ norm of $h^{(\mathcal{S},1)}$. If $C\neq\varnothing$, we shall rearrange the tensors such that $\max\{N_{\mathfrak{l}}:\mathfrak{l}\in C\cap\mathcal{U}_1\}\sim\max\{N_{\mathfrak{l}}:\mathfrak{l}\in C\}$; if $C=\varnothing$, we shall select all $1\leq j\leq q$ such that either $j>r$, or $n_j=1$ or $\min_{\mathfrak{l}\in\mathcal{L}_j}N_{\mathfrak{l}}<M^\delta$ (such $j$ exists by definition of $\sum_{(b)}$ in (5.27)), and by rearranging the tensors we may assume that the maximum of $N_j$ for such $j$ corresponds to[1] $j=1$.

Next, for $1\leq j\leq r$, define $h^{(j)}=\mathtt{Trim}(h^{(\mathcal{S}_j,n_j)},R_j)$ where $R_1=M^\delta$ and $R_j=(N_*)^\delta$ for $j\geq 2$, with $N_*=\max(N_2,\cdots,N_q)$. Then, by the rearrangements we made above, it can be verified that the relevant parts (measurability, etc.) of the assumptions of Proposition 6.4 are satisfied. Moreover, since $\partial_{\lambda_j}\widehat{h^{(\mathcal{S}_j,n_j)}}$ can be controlled using the fact that $h^{(\mathcal{S}_j,n_j)}$ is compactly supported in $t$, we can apply a meshing argument in $\lambda_j$ in the same way as in the proof of Proposition 6.1. By combining the bounds (5.34)–(5.37) for $h^{(\mathcal{S}_j,n_j)}$, Proposition 6.1, and the above meshing argument in $\lambda_j$, we obtain that the following holds $\tau^{-1}M$-certainly:

[1]Here we have assumed $1\leq j\leq r$. If $r+1\leq j\leq q$, then this $j$ would correspond to the input function $z_{N_j}$ in (5.22), which gains a big power $N_j^{-D_1}$. Thus the case where the maximum $N_j$ occurs at this $j$ will be strictly easier than the cases we actually treat in the proof.

(a) If $j$ is such that either $n_j = 1$, or $n_j = 0$ and $\min_{\mathfrak{l}\in\mathcal{L}_j} N_{\mathfrak{l}} < R_j$, define $\mathfrak{X}_j = \mathfrak{X}_j(\lambda_j)$ to be the smallest value such that the type 1 bounds (6.44)–(6.47) in Proposition 6.4 hold for $\widehat{h^{(j)}}(\lambda_j)$, then we have

$$\left(\int_{\mathbb{R}} \mathfrak{X}_j(\lambda_j)\,\mathrm{d}\lambda_j\right)^2 \lesssim \int_{\mathbb{R}} \langle\lambda_j\rangle^{2b}\mathfrak{X}_j(\lambda_j)^2\,\mathrm{d}\lambda_j \lesssim \left[\tau^{-\theta}M^{C/\sqrt{\kappa}}(1+R_j^C N_j^{-3\varepsilon})\right]^2. \tag{7.22}$$

(b) If $n_j = 0$ and $\min_{\mathfrak{l}\in\mathcal{L}_j} N_{\mathfrak{l}} \geq R_j$ (in particular $h^{(j)} = h^{(\mathcal{S}_j,0)}$), we may decompose $h^{(j)}$ into $h^{(j,\Gamma_j)} = h^{(\mathcal{S}_j,0,\Gamma_j)}$ as in part (1) of Proposition 5.1; define $\mathfrak{X}_j = \mathfrak{X}_j(\lambda_j,\Gamma_j)$ to be the smallest value such that the type 0 bounds (6.42)–(6.43) in Proposition 6.4 hold for $\widehat{h^{(j,\Gamma_j)}}(\lambda_j)$, then we have (7.17).

Moreover, since $\widetilde{h}$ satisfies (7.16), we can apply a further meshing argument to replace it by some function (which we still denote by $\widetilde{h}$ for simplicity) that is supported in the big box $|\lambda|, |\lambda_j| \leq M^{\sqrt{\kappa}}$ and is constant on each small box of size (say) $\exp(-(\log M)^6)$. Let $\Xi := \lambda - \Omega - \zeta_1\lambda_1 - \cdots - \zeta_q\lambda_q$ (see (7.16)), we may also decompose $\widetilde{h}$ into $\widetilde{h}_{\Xi^*}$, which are restrictions of $\widetilde{h}$ to the set $\lfloor\Xi\rfloor = \Xi^*$ for $\Xi^* \in \mathbb{Z}, |\Xi^*| \lesssim M^{2\sqrt{\kappa}}$. Then $\widetilde{h}_{\Xi^*}$ satisfies the assumptions (6.38) and (6.39) in Proposition 6.4, with $\widetilde{\Gamma}$ in (6.38) depending on $(\lambda, \lambda_1, \cdots, \lambda_r, \Xi^*)$, and the right hand side of (6.39) multiplied by $\langle\lambda\rangle^{-1}\langle\Xi^*\rangle^{-1}$.

Let the tensor $H(\lambda, \lambda_1, \cdots, \lambda_r, \Gamma_j, \Xi^*)$ be defined as in (7.20) but with $\widetilde{h}$ replaced by $\widetilde{h}_{\Xi^*}$, and $\mathtt{Trim}(\widehat{h^{(\mathcal{S}_j,n_j)}}(\lambda_j), M^\delta)$ replaced by $\widehat{h^{(j)}}(\lambda_j)$ or $\widehat{h^{(j,\Gamma_j)}}(\lambda_j)$ in case (a) or (b) above, and $\mathscr{O}$ replaced by some $\mathscr{O}^+$ containing $\mathscr{O}$. By Proposition 3.7 (3), $\widehat{h^{(\mathcal{S},1)}}(\lambda)$ can be written as a linear combination of

$$\int \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_r \cdot \sum_{\Xi^*}\sum_{(\Gamma_j)} H(\lambda, \lambda_1, \cdots, \lambda_r, \Gamma_j, \Xi^*),$$

where $\sum_{(\Gamma_j)}$ are present only for those $j$ in case (b) above. We now apply Proposition 6.4 to conclude that the tensor $H(\lambda, \lambda_1, \cdots, \lambda_r, \Gamma_j, \Xi^*)$ satisfies (6.49)–(6.50) with

$$\mathfrak{Y} = \sqrt{\Upsilon}\frac{1}{\langle\lambda\rangle\langle\Xi_*\rangle}\prod_{j=1}^r \mathfrak{X}_j \cdot \tau^{-\theta}M^{\theta}(N_*)^{-\varepsilon^4}, \tag{7.23}$$

where $\Upsilon$ is as in Proposition 6.4, and $\mathfrak{X}_j = \mathfrak{X}_j(\lambda_j)$ or $\mathfrak{X}_j(\lambda_j, \Gamma_j)$ in case (a) or (b) above; the meshing argument guarantees that the above holds for all values of $(\lambda, \lambda_j)$ after removing a single exceptional set of probability $\leq C_\theta e^{-(\tau^{-1}M)^\theta}$. Therefore, by taking the weighted $L^2$ norm in $\lambda$ within the set $|\lambda| \leq M^{\sqrt{\kappa}}$, then summing in $\Xi^*$ and $\Gamma_j$ and integrating in $\lambda_j$, we will obtain the bounds (5.34)–(5.35) for this component of $h^{(\mathcal{S},1)}$ under consideration, once we show that

$$\sqrt{\Upsilon}\cdot\tau^{-\theta}M^{C/\sqrt{\kappa}}\prod_{j=1}^r(1+R_j^C N_j^{-3\varepsilon}) \leq (N_*)^{\varepsilon^5}. \tag{7.24}$$

But this is true since $R_j = (N_*)^\delta$ for $j \geq 2$, so any power $R_j^C$ for $j \geq 2$ will be negligible. Moreover $R_1 = M^\delta$, so either $N_* \gtrsim M$ and $R_1^C$ is also negligible, or $N_1 \sim M$ and the $R_1^C$ loss is covered by the $N_1^{-3\varepsilon}$ gain, or $\max(N_1, N_*) \ll M$ and the $R_1^C$ loss is covered by the $\sqrt{\Upsilon}$ gain. Finally since $N_* \geq M^\delta$, the gain $(N_*)^{\varepsilon^4}$ in (7.23) will overwhelm the loss $M^{C/\sqrt{\kappa}}$ by our choice of $\kappa$. This proves (5.34)–(5.35) in the main case.

Now we turn to the proof of (5.36). Starting with $H(\lambda,\lambda_1,\cdots,\lambda_r)$, we shall further decompose it by attaching smooth truncations supported in sets

$$1+\frac{1}{M^{2\delta}}\bigg|k-\sum_{\mathfrak{l}\in\mathcal{U}}\zeta_{\mathfrak{l}}^*k_{\mathfrak{l}}-\ell\bigg|\sim K,\quad 1+\frac{1}{M^{2\delta}}\bigg|k_j-\sum_{\mathfrak{l}\in\mathcal{U}_j}\zeta_{\mathfrak{l}}k_{\mathfrak{l}}-\ell_j\bigg|\sim K_j\,(1\leq j\leq r), \tag{7.25}$$

where $1\leq K,K_j\leq M$ are dyadic numbers, $\ell$ and $\ell_j$ as in (7.21). The number of terms in this decomposition is $\leq(\log M)^C$, so we only need to consider a single term. By (7.21) we know $K\lesssim\max(K_1,\cdots,K_r)$. By rearrangement we may assume $K\lesssim K_1$; in particular we may assume $n_1=1$ (otherwise $K\lesssim K_1\sim 1$, so (5.36) follows directly from (5.35)). At this point we can repeat the arguments in the above proof of (5.34)–(5.35) (namely trimming $h^{(\mathcal{S}_j,n_j)}$ at frequency $R_j$ with $R_1=M^\delta$ and $R_j=(N_*)^\delta\,(j\geq 2)$, decomposing $H(\lambda,\lambda_1,\cdots,\lambda_r)$ into $H(\lambda,\lambda_1,\cdots,\lambda_r,\Gamma_j,\Xi^*)$, defining $\mathfrak{X}_j$ as above, etc.) and then apply Proposition 6.4 to conclude that $H(\lambda,\lambda_1,\cdots,\lambda_r,\Gamma_j,\Xi^*)$ satisfies (6.52) with $\mathfrak{Y}$ defined as in (7.23). Here the assumptions of Proposition 6.4 are satisfied, since $K\lesssim K_1$, and multiplying by any of the smooth truncations we introduced does not increase any of the $X^{\widetilde{b},-b_0}_{\mathcal{V}_j}[k_jk_{B_j}\to k_{C_j}]$ norms due to Lemma 4.10. After obtaining (6.52), we can again repeat the arguments in the above proof (namely taking the weighted $L^2$ norm in $\lambda$ within the set $|\lambda|\leq M^{\sqrt{\kappa}}$, then summing in $\Xi^*$ and $\Gamma_j$ and integrating in $\lambda_j$) and deduce (5.36) for this component of $h^{(\mathcal{S},1)}$ under consideration, using (7.24). This proves (5.36) and completes the main case.

*Step 2: adding the $\mathbb{R}$-linear operators.* Now we prove the estimates (5.34)–(5.36) for the second line of (5.27) assuming the second maximum of $N_j\,(1\leq j\leq q)$ is $<M^\delta$ (and in particular the maximum of $N_j$ is $\leq N/2$), and the same estimates for the third line of (5.27). First look at the second line of (5.27); we may assume the maximum of $N_j$ is $N_1\geq M^\delta$ (and $L:=\max(N_2,\cdots,N_q)<M^\delta$), since the cases when the maximum of $N_j$ occurs at $r+1\leq j\leq q$, or when $N_1<M^\delta$ also, are much easier. With such assumptions we must have $\mathscr{O}=\varnothing$, and the current term can be written as an $\mathbb{R}$-linear operator[1] $\mathscr{L}^\zeta$ (for some $\zeta\in\{\pm\}$, see (5.39) for definition, where $N$ should be replaced by $M$) applied to the tensor $h':=\mathtt{Trim}(h^{(\mathcal{S}_1,n_1)},M^\delta)$. More precisely, $\mathcal{S}':=\mathtt{Trim}(\mathcal{S}_1,M^\delta)$ has the same sets of leaves, pairings, blossoms and pasts as $\mathcal{S}$ (in particular the sets $\mathcal{L},\mathcal{U},\mathcal{V}$ etc. are common to both tensors), the only differences being that $N(\mathcal{S}')=N_1$ while $N(\mathcal{S})=M$, and the sign $\zeta_{\mathfrak{n}}^*$ of $\mathfrak{n}\in\mathcal{L}\cup\mathcal{V}$ in $\mathcal{S}$ equals the sign $\zeta_{\mathfrak{n}}$ of $\mathfrak{n}$ in $\mathcal{S}'$ multiplied by $\zeta$; for the tensors we have

$$(\widehat{h^{(\mathcal{S},1)}})_{kk_{\mathcal{U}}}(\lambda,k_{\mathcal{V}},\lambda_{\mathcal{V}})=\Upsilon\cdot\sum_{k'}\int_{\mathbb{R}}\mathrm{d}\lambda'\cdot(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,-\zeta\lambda')(\widehat{h'})^\zeta_{k'k_{\mathcal{U}}}(\lambda',k_{\mathcal{V}},\lambda_{\mathcal{V}}). \tag{7.26}$$

To estimate the $X^{\widetilde{b},-b_0}[kk_B\to k_C]$ norms of $h^{(\mathcal{S},1)}$ (including the weighted ones in (5.36)), where $(B,C)$ is a subpartition of $\mathcal{U}$ and we denote $E:=\mathcal{U}\backslash(B\cup C)$, we will consider four cases.

(a) Assume $C\neq\varnothing$, then for any fixed $\lambda$, $(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ and $k_E$, by (7.26) and a variant of Proposition 4.11 we have

$$\|(\widehat{h^{(\mathcal{S},1)}})_{kk_{\mathcal{U}}}\|_{kk_B\to k_C}\leq\Upsilon\cdot\|\langle\lambda'\rangle^{1-b}(\widehat{h'})_{k'k_{\mathcal{U}}}(\lambda')\|_{\ell^2_{k'k_B}L^2_{\lambda'}\to\ell^2_{k_C}}\cdot\|\langle\lambda'\rangle^{-(1-b)}(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,-\zeta\lambda')\|_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}.$$

[1]When $\mathcal{S}$ is fixed, the sum in $(N_j,\mathcal{S}_j)$ etc. for $j\geq 2$ involves at most $(\log L)^\kappa$ terms, which is negligible in view of the $L^{\varepsilon\delta}$ gain we will obtain below. The sum in $(N_1,\mathcal{S}_1)$ involves at most $\kappa$ terms if $N_1\sim M$ and $\mathcal{S}_1=\mathcal{S}'$, and at most $(\log M)^\kappa$ terms otherwise; either way this is negligible in view of the gain from $\Upsilon$, and the gain of at least $M^{\delta^5}$ coming from trimming assuming $\mathcal{S}_1\neq\mathcal{S}'$, which is evident from the proof of Proposition 6.1.

The third factor above is a function of $\lambda$ only, and we shall temporarily denote it by $G(\lambda)$; the second factor is bounded by

$$\big\|\langle\lambda'\rangle^{1-b}\cdot\|(\widehat{h'})_{k'k_{\mathcal{U}}}(\lambda')\|_{k'k_B\to k_C}\big\|_{L^2_{\lambda'}},$$

so upon taking supremum in $k_E$, and then taking the weighted $L^2$ norm in $\lambda$ and the weighted $\ell^2L^2$ norm in $(k_{\mathcal{V}},\lambda_{\mathcal{V}})$, we obtain that

$$\|h^{(\mathcal{S},1)}\|_{X_{\mathcal{V}}^{1-b,-b_0}[kk_B\to k_C]}\leq\Upsilon\cdot\|h'\|_{X_{\mathcal{V}}^{1-b,-b_0}[k'k_B\to k_C]}\cdot\|\langle\lambda\rangle^{1-b}G(\lambda)\|_{L^2_\lambda}.$$

By the proof of Proposition 7.1, the weighted norm of $G$ above is bounded by $\tau^{(6\kappa)^{-1}}L^{-3\varepsilon\delta}$, hence

$$\|h^{(\mathcal{S},1)}\|_{X_{\mathcal{V}}^{1-b,-b_0}[kk_B\to k_C]}\lesssim\tau^{(6\kappa)^{-1}}L^{-3\varepsilon\delta}\cdot\Upsilon\cdot\|h'\|_{X_{\mathcal{V}}^{1-b,-b_0}[k'k_B\to k_C]}.$$

By using the induction hypothesis for $h^{(\mathcal{S}_1,n_1)}$, Proposition 6.1, and controlling the potential loss factor $1+M^{C\delta}N_1^{-3\varepsilon}$ occurring in Proposition 6.1 by the $\Upsilon$ factor, we deduce (5.34) for $h^{(\mathcal{S},1)}$. Note that, if $\mathcal{S}_1=\mathcal{S}'$ then the above estimate has no loss. If $\mathcal{S}_1\neq\mathcal{S}'$, then applying Proposition 6.1 loses a factor $M^\theta$, but by examining the proof of Proposition 6.1 we see that we can also gain a small power of $N_1$ (which is $\leq N_1^{-\delta^4}$). This, in view of the $\Upsilon$ factor, is enough to cover this loss together with the potential log loss coming from summing over all plants; moreover the continuous variable $\lambda'$ can be handled by restricting to $|\lambda'|\leq M^{\kappa^2}$ and performing another meshing argument exploiting the above power gain. The same comment also applies in the other cases below.

(b) Assume $C=\varnothing$ and $E\neq\varnothing$, then the same argument as in case (a) is enough to control the norm $\|h^{(\mathcal{S},1)}\|_{X_{\mathcal{V}}^{1-b,-b_0}[kk_B]}$; note that to prove (5.35) for $h^{(\mathcal{S},1)}$ we need to gain a power $M^{-\varepsilon}$, which is provided by the corresponding power $N_1^{-\varepsilon}$ (or the better powers and the $\mathcal{X}_{1,1}$ factor in (5.32) corresponding to $\mathcal{S}_1$) from the induction hypothesis if $N_1\sim M$, and by the $\Upsilon$ factor if $N_1\ll M$. The weighted norm in (5.36) is bounded in the same way using the induction hypothesis, Proposition 6.1, (the proof of) Proposition 7.1, and (a variant of) Lemma 4.3, using the fact that $(\mathscr{L}^\zeta)_{kk'}$ is supported in $|k-\zeta k'|\lesssim M^\delta$.

(c) Assume $C=E=\varnothing$, and either $n_1=1$, or $\min_{\mathfrak{l}\in\mathcal{L}_1}N_{\mathfrak{l}}<M^\delta$, then the norm in question, namely $X_{\mathcal{V}}^{b,-b_0}[kk_{\mathcal{U}}]$, is a weighted $\ell^2L^2$ norm in $(k_{\mathcal{V}},\lambda_{\mathcal{V}})$, while for fixed $(k_{\mathcal{V}},\lambda_{\mathcal{V}})$ it is an $\ell^2L^2$ norm in $(k,k_{\mathcal{U}},\lambda)$ weighted by $\langle\lambda\rangle^b$. Therefore, by (7.26) we have

$$\|h^{(\mathcal{S},1)}\|_{X_{\mathcal{V}}^{b,-b_0}[kk_{\mathcal{U}}]}\leq\Upsilon\cdot\|h'\|_{X_{\mathcal{V}}^{1-b,-b_0}[k'k_{\mathcal{U}}]}\cdot\|\mathscr{L}^\zeta\|_{X^{1-b}\to X^b}.$$

Using the induction hypothesis, Proposition 6.1 and the bound $\|\mathscr{L}^\zeta\|_{X^{1-b}\to X^b}\leq\tau^{(6\kappa)^{-1}}L^{-3\varepsilon\delta}$, which also follows from the proof of Proposition 7.1, we can prove the desired estimates in the same way as in parts (a) and (b).

(d) Assume $C=E=\varnothing$, $n_1=0$, and that $\min_{\mathfrak{l}\in\mathcal{L}_1}N_{\mathfrak{l}}\geq M^\delta$ (in particular $\mathcal{V}=\varnothing$ and $\mathcal{L}=\mathcal{L}_1$). In this case we will use a different estimate. Still starting with (7.26), we have

$$\|h^{(\mathcal{S},1)}\|_{X^b[kk_{\mathcal{U}}]}\leq\Upsilon\cdot\|(\mathscr{L}^\zeta)_{kk'}\|_{X^{b,-(1-b)}[kk']}\cdot\|h^{(\mathcal{S}_1,0)}\|_{X^{1-b}[k'\to k_{\mathcal{U}}]}.$$

Now let $\widetilde{N}=\max_{\mathfrak{l}\in\mathcal{U}}N_{\mathfrak{l}}$, then in the above formula we may assume $\langle k'\rangle\lesssim\widetilde{N}$. By the proof of Proposition 7.1 we have

$$\|\mathbf{1}_{\langle k'\rangle\lesssim\widetilde{N}}\cdot(\mathscr{L}^\zeta)_{kk'}\|_{X^{b,-(1-b)}[kk']}\lesssim\tau^{(6\kappa)^{-1}}(\widetilde{N})^{\alpha_0}M^{C\delta};$$

combining with the induction hypothesis (in particular summing over the $\Gamma$ variable in (5.32)) we obtain that

$$\|h^{(\mathcal{S},1)}\|_{X^b[kk_{\mathcal{U}}]} \lesssim \Upsilon \cdot \tau^{(6\kappa)^{-1}} (\widetilde{N})^{\alpha_0} M^{C\delta} \cdot \prod_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}^{\beta_1} \prod_{\mathfrak{l}\in\mathcal{P}} N_{\mathfrak{l}}^{-8\varepsilon} \prod_{\mathfrak{p}\in\mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \cdot (\widetilde{N})^{-\beta_1} \cdot \mathcal{X}_{1,1}, \tag{7.27}$$

where $\mathcal{X}_{1,1}$ is the one in (5.33) corresponding to $\mathcal{S}_1$. To prove (5.35) we just need to gain an extra $M^{-2\varepsilon}$ power, which is provided either by the difference in the powers of $N_{\mathfrak{l}}$ for $\mathfrak{l} \in \mathcal{L}$ between (5.35) and (7.27), or by the $\mathcal{X}_{1,1}$ and $\Upsilon$ factors. The proof of (5.36) is the same, as the weight in (5.36) is in fact bounded by 1 using the induction hypothesis and the support condition for $(\mathscr{L}^\zeta)_{kk'}$.

Next we look at the third line of (5.27). For simplicity, we will consider the case $\zeta = +$; the case $\zeta = -$ is analogous, with $\mathcal{S}$ replaced by $\overline{\mathcal{S}}$ in a few places. This term can be written as

$$\widehat{(h^{(\mathcal{S},1)})}_{kk_{\mathcal{U}}}(\lambda, k_{\mathcal{V}}, \lambda_{\mathcal{V}}) = \sum_{k'} \int_{\mathbb{R}} \mathrm{d}\lambda' \cdot (\widehat{\mathscr{V}^{M,+}})_{kk'}(\lambda, -\lambda')(\widehat{\mathcal{H}})_{k'k_{\mathcal{U}}}(\lambda', k_{\mathcal{V}}, \lambda_{\mathcal{V}}). \tag{7.28}$$

The term (7.28) is similar to (7.26), except that $\mathscr{L}^+$ is replaced by $\mathscr{V}^{M,+}$, and $h'$ is replaced by $\mathcal{H}$, which is either the second term on the right hand side of (5.26) (if we take $\sum_{(a)}$ instead of $\sum_{(c)}$ in the third line of (5.27)), or the second line of (5.27) (if we take $\sum_{(b)}$ instead of $\sum_{(c)}$). By what we proved in Proposition 7.2 and the above arguments, we know that $\mathcal{H}$ is an $\mathcal{S}$-tensor, which either satisfies (5.30) and (5.32), or satisfies the bounds (5.34)–(5.36). Moreover for the $\mathbb{R}$-linear operator $\mathscr{V}^{M,+}$ by Proposition 7.1 we have

$$\begin{aligned} \|\mathscr{V}^{M,+}\|_{X^{1-b}\to X^b} &\leq \tau^{(7\kappa)^{-1}}, \\ \big\|\langle\lambda\rangle^{1-b}\|\langle\lambda'\rangle^{-(1-b)}(\widehat{\mathscr{V}^{M,+}})_{kk'}(\lambda,\lambda')\|_{\ell^2_{k'}L^2_{\lambda'}\to\ell^2_k}\big\|_{L^2_\lambda} &\leq \tau^{(7\kappa)^{-1}}, \\ \|\mathbf{1}_{\langle k'\rangle\leq\widetilde{N}}\mathscr{V}^{M,+}\|_{X^{b,-(1-b)}[kk']} &\leq \tau^{(8\kappa)^{-1}}(\widetilde{N})^{\alpha_0}\cdot M^{C\delta}, \\ \|(1+M^{-\delta}|k-k'|)^{\kappa^2}\mathscr{V}^{M,+}\|_{X^{b,-(1-b)}[kk']} &\leq \tau^{(8\kappa)^{-1}}M^{\beta_1-\varepsilon}, \end{aligned} \tag{7.29}$$

which are similar to the bounds for $\mathscr{L}^+$ we have used above. The estimate for the third line of (5.27) can then be deduced by considering cases (a)–(d), in the same way as above. We only mention a few important points: (i) going from $\mathcal{H}$ to $h^{(\mathcal{S},1)}$ does not involve trimming, so in these proofs no meshing argument is needed, hence they do not require the derivative bound (5.37) for $\mathcal{H}$ (which has not been proved yet); (ii) the $\Upsilon$ factor is not needed, because $\mathcal{H}$ is an $\mathcal{S}$-tensor, therefore the $N_1$ in the above proof will be replaced by $M$; (iii) the fact that $\mathscr{L}^+$ is supported in $|k-k'| \leq M^\delta$ is replaced by the last bound in (7.29), which leads to the restriction $|k-k'| \leq M^{2\delta}$, since otherwise we gain a sufficiently high power of $M$. This allows us to apply Lemma 4.3 in estimating the weighted norms in (5.36) in cases (b) and (c), and also bound the weight in (5.36) in case (d).

*Step 3: remaining estimates.* Next we shall prove (5.34)–(5.36) for the first line of (5.27). In fact, this term can be treated in the same way as the third line of (5.27), see *Step 2* above, where the only difference is that $\mathcal{H}$ is replaced by the $\mathcal{S}_M^+$ tensor occurring as the first term of the right hand side of (5.26). Since this tensor also satisfies (5.30) and (5.32) by Proposition 7.2, the same arguments as in *Step 2* above suffice to estimate this term.

Finally we prove the derivative bound (5.37) for all terms in (5.27). This is a very loose bound, so it can be proved by very loose estimates. Just notice that:

- The $\mathbb{R}$-linear operator $\mathscr{V}^{M,\zeta}$ commutes with $\partial_{\lambda_{\mathfrak{f}}}$ and increases the norm in consideration by at most a constant multiple;
- If $h' = \mathtt{Trim}(h, R)$, where $h$ is an $\mathcal{S}$-tensor and $h'$ is an $\mathcal{S}'$-tensor, then $\partial_{\lambda_{\mathfrak{f}}} h' = \mathtt{Trim}(\partial_{\lambda_{\mathfrak{f}}} h, R)$ for any $\mathfrak{f} \in \mathcal{V}'$;
- If $H = \mathtt{Merge}(h^{(1)}, \cdots, h^{(r)}, h, \mathscr{B}, \mathscr{O})$, where $h^{(j)}$ is an $\mathcal{S}_j$-tensor and $H$ is an $\mathcal{S}$-tensor, then $\partial_{\lambda_{\mathfrak{f}}} H = \mathtt{Merge}(h^{(1)}, \cdots, \partial_{\lambda_{\mathfrak{f}}} h^{(j)}, \cdots, h^{(r)}, h, \mathscr{B}, \mathscr{O})$ for any $\mathfrak{f} \in \mathcal{V}_j$; in the same way we also have $\partial_{\lambda_j} H = \mathtt{Merge}(h^{(1)}, \cdots, h^{(r)}, \partial_{\lambda_j} h, \mathscr{B}, \mathscr{O})$ for any $r+1 \leq j \leq q$.

Therefore, in order to estimate $\partial_{\lambda_{\mathfrak{f}}} h^{(\mathcal{S},1)}$ for $\mathfrak{f} \in \mathcal{V}$, we only need to consider the same `Trim-Merge` combination, where *one* of the inputs $h^{(\mathcal{S}_j,n_j)}$ is replaced by $\partial_{\mathcal{V}_j} h^{(\mathcal{S}_j,n_j)}$. This input is then bounded by the induction hypothesis, noting that either $N_j \leq M/2$ or $|\mathcal{S}_j| < |\mathcal{S}|$, and the other inputs are bounded trivially (say using part (3) of $\mathtt{Local}(M)$) by a power $M^{C\cdot\kappa}$. Therefore we get, without removing any exceptional set, that

$$\|\partial_{\lambda_{\mathcal{V}}} h^{(\mathcal{S},1)}_{kk_{\mathcal{U}}}(t, k_{\mathcal{V}}, \lambda_{\mathcal{V}})\|_{X^{b,-b_0}_{\mathcal{V}}[kk_{\mathcal{U}}]} \lesssim M^{C\cdot\kappa} + M^{C\cdot\kappa} \exp[(\log \widetilde{N})^5 + |\widetilde{\mathcal{S}}|(\log \widetilde{N})^3],$$

where $\widetilde{N} \leq M$, $|\widetilde{\mathcal{S}}| \leq D$, and either $\widetilde{N} \leq M/2$ or $|\widetilde{\mathcal{S}}| < |\mathcal{S}|$. Therefore (5.37) follows from the bound

$$\exp[(\log M)^5 + |\mathcal{S}|(\log M)^3] \geq \exp((\log M)^2) \cdot \exp[(\log \widetilde{N})^5 + |\widetilde{\mathcal{S}}|(\log \widetilde{N})^3]$$

under these assumptions. This completes the proof. □

7.4. **The remaining parts.** In this section we prove parts (3)–(5) of $\mathtt{Local}(2M)$.

**Proposition 7.4.** *Assume* $\mathtt{Local}(M)$ *and parts (1) and (2) of* $\mathtt{Local}(2M)$ *are true. Then* $\tau^{-1}M$*-certainly, parts (3)–(5) of* $\mathtt{Local}(2M)$ *are true. More precisely, (5.38) is true for each* $n \in \{0, 1\}$ *and regular plant* $\mathcal{S}$ *with* $N(\mathcal{S}) = M$ *and* $|\mathcal{S}| \leq D$*, and the mapping that defines* $z_M$ *(see the right hand side of (5.28)) is a contraction mapping from the ball* $\{z : \|z\|_{X^{b_0}} \leq M^{-D_1}\}$ *to itself, and (5.40) is true for the kernel* $\mathscr{L}^\zeta$ *defined by (5.39), if* $\max(N_2, \cdots, N_q) = M$.

*Proof.* First we shall prove (5.38) for $\Psi^{(\mathcal{S},n)}$, assuming either (5.30) and (5.32), or the estimates (5.34)–(5.37) for $h^{(\mathcal{S},n)}$. If $n = 1$, by applying Cauchy-Schwartz (in the $(k_{\mathcal{V}}, \lambda_{\mathcal{V}})$ variables), Lemma 4.4, and a meshing argument in $\lambda$ and $\lambda_{\mathcal{V}}$ variables, we can get from (5.34)–(5.37) (where we choose $C = E = \varnothing$) that $\tau^{-1}M$-certainly,

$$\|\Psi^{(\mathcal{S},1)}\|_{X^{b_0}} \leq (\tau^{-1}M)^{\theta_0} M^{-\varepsilon} \prod_{\mathfrak{n} \in \mathcal{L}\cup\mathcal{V}\cup\mathcal{Y}} N_{\mathfrak{n}}^{-\delta^3},$$

which clearly implies (5.38). Here for the meshing argument, just notice that the $\lambda$ and $\lambda_{\mathcal{V}}$ derivatives of $\widehat{h^{(\mathcal{S},1)}}$ are bounded, and that the choice of $b_0$ in (5.38) (compared to $b$) allows us to restrict $\lambda$ to the big box $|\lambda| \leq M^{\kappa^2}$, so we can apply the same arguments as in the proof of Proposition 6.1.

Next, if $n = 0$ and $\mathcal{S} \neq \mathcal{S}_M^+$ (which is the mini-tensor defined in Definition 3.2), then $|\mathcal{L}| \geq 3$. For each fixed $k \in \mathbb{Z}^d$ with $\langle k \rangle \leq M$, using Lemma 4.4, a meshing argument in $\lambda$ as above, and the simple inequality

$$\sup_k \|h_{kk_{\mathcal{U}}}\|_{k_{\mathcal{U}}} \leq \|h_{kk_{\mathcal{U}}}\|_{k \to k_{\mathcal{U}}}$$

for any tensor $h = h_{kk_{\mathcal{U}}}$, we deduce from (5.32) (where we choose $B = E = \varnothing$ *instead of* $C = E = \varnothing$) that $\tau^{-1}M$-certainly,

$$\left(\int_{\mathbb{R}} \langle\lambda\rangle^{2b_0} |\widehat{\Psi_k^{(\mathcal{S},0)}}(\lambda)|^2\right)^{1/2} \leq (\tau^{-1}M)^{\theta_0} \mathcal{X}_1 \cdot \prod_{\mathfrak{l} \in \mathcal{L}\setminus\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{-4\varepsilon} \prod_{\mathfrak{p} \in \mathcal{Y}} N_{\mathfrak{p}}^{-\delta^3} \cdot N_{\mathfrak{l}_{\mathrm{top}}}^{-\alpha+\theta},$$

where $N_{\mathfrak{l}_{\mathrm{top}}} := \max_{\mathfrak{l}\in\mathcal{U}} N_{\mathfrak{l}}$, and $\mathcal{X}_1$ is as in (5.33) but with $N$ replaced by $M$. We may assume that $\tau^{-1}M$-certainly the above holds for every $k$; since $\langle k\rangle \lesssim N_{\mathfrak{l}_{\mathrm{top}}}$ by (5.30), we conclude that

$$\|\Psi^{(\mathcal{S},0)}\|_{X^{s',b_0}} \leq \tau^{-\theta_0} N_{\mathfrak{l}_{\mathrm{top}}}^{s'-s} M^{2\delta^3}\mathcal{X}_1 \cdot \prod_{\mathfrak{l}\in\mathcal{L}\setminus\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{-2\varepsilon} \cdot \prod_{\mathfrak{n}\in\mathcal{L}\cup\mathcal{Y}} N_{\mathfrak{l}}^{-\delta^3},$$

which implies (5.38), noticing that

$$N_{\mathfrak{l}_{\mathrm{top}}}^{s'-s} M^{2\delta^3}\mathcal{X}_1 \cdot \prod_{\mathfrak{l}\in\mathcal{L}\setminus\{\mathfrak{l}_{\mathrm{top}}\}} N_{\mathfrak{l}}^{-2\varepsilon} \leq M^{s'-s},$$

which easily follows from the assumption $0 < s - s' < \delta^2$ and the definition (5.33) of $\mathcal{X}_1$. Finally, if $n = 0$ and $\mathcal{S} = \mathcal{S}_M^+$, then we simply have $\Psi_k^{(\mathcal{S},0)}(t) = \chi(t)\cdot \mathbf{1}_{M/2<\langle k\rangle\leq M}(f_M)_k$, so (5.38) follows from (5.8).

Next we prove the contraction mapping part in the statement. We will only prove that the right hand side of (5.28) maps the given ball to itself, since the contraction part follows in the same way. Suppose $\|z_M\|_{X^{b_0}} \leq M^{-D_1}$. The right hand side of (5.28), which we shall denote by $z_{\mathrm{out}}$, contains three types of terms, which we shall analyze below. Like in the proof of Proposition 7.3, we will abuse notation and refer to some components of $z_{\mathrm{out}}$ on the right hand side of (5.28) still as $z_{\mathrm{out}}$.

(1) Consider the terms on the right hand side of (5.28) that contain no factor $z_M$ (that is, $N_j \leq M/2$ for all $r+1 \leq j \leq q$). If some $z_{N_j}$ is replaced by $z_{N_j}^{\mathrm{hi}}$ whose Fourier transform is supported in $|\lambda_j| \geq M^{\kappa^2}$, then we can gain a power $M^{\kappa}$ by using the bound $\|z_{N_j}\|_{X^{b_0}} \leq 1$, which will overwhelm all loss and easily imply the desired estimate. Below we will assume each $z_{N_j}$ is replaced by $z_{N_j}^{\mathrm{lo}}$, so by definition of $\sum_{(d)}$ in (5.28), the corresponding component (up to linear combination over different $\mathscr{O}$) can be written in Fourier space as

$$(\widehat{z_{\mathrm{out}}})_k(\lambda) = \Upsilon \cdot \sum_{\zeta\in\{\pm\}}\sum_{k'}\int \mathrm{d}\lambda' \cdot (\widehat{\mathscr{R}^{M,\zeta}})_{kk'}(\lambda,-\zeta\lambda')(\widehat{\Psi^{(\mathcal{S},n)}})_{k'}^{\zeta}(\lambda'), \quad \Psi_{k'}^{(\mathcal{S},n)} = \Psi_{k'}[\mathcal{S},h^{(\mathcal{S},n)}],$$

where

$$\mathcal{S} = \mathtt{Trim}(\mathtt{Merge}(\mathtt{Trim}(\mathcal{S}_1, M^{\delta}),\cdots,\mathtt{Trim}(\mathcal{S}_r, M^{\delta}),\mathscr{B},\mathscr{O}), M^{\delta})$$

satisfies $|\mathcal{S}| > D$, and $h^{(\mathcal{S},n)}$ is defined in the same way as the second term on the right hand side of (5.26) (if $n = 0$), or as the second line of (5.27) (if $n = 1$).

Note that we now have $D < |\mathcal{S}| \leq C\cdot D$; however in the proof of Propositions 7.2 and 7.3 we have not used the assumption $|\mathcal{S}| \leq D$, so the same proof also works in the current case and gives ($\tau^{-1}M$-certainly, i.e. after removing an exceptional set of probability $\leq C_\theta e^{-(\tau^{-1}M)^\theta}$) the bound (5.32) or the bounds (5.34)–(5.37) for $h^{(\mathcal{S},n)}$. Then, applying once more Lemma 4.4 and the meshing argument as before, we deduce that

$$\|\Psi^{(\mathcal{S},n)}\|_{X^{b_0}} \leq M^{d/2}\prod_{\mathfrak{n}\in\mathcal{L}\cup\mathcal{V}\cup\mathcal{Y}} N_{\mathfrak{n}}^{-\delta^3} \leq M^{d/2}M^{-\delta^4 D} \leq M^{-2D_1},$$

using the fact that $|\mathcal{L}|+|\mathcal{V}|+|\mathcal{Y}| = |\mathcal{S}| > D$ and $N_{\mathfrak{n}} \geq M^{\delta}$ for each $\mathfrak{n}\in\mathcal{L}\cup\mathcal{V}\cup\mathcal{Y}$. Using then the $X^{b_0}\to X^{b_0}$ norm bound for $\mathscr{R}^{M,\zeta}$ (which follows from the corresponding bound for $\mathscr{V}^{M,\zeta}$ proved in Proposition 7.1), we deduce the same bound for $z_{\mathrm{out}}$.

(2) Consider the terms on the right hand side of (5.28) that contain at least two factors $z_M$ (that is, $N_j = M$ for at least two $r+1\leq j\leq q$). Then, these $z_M$ factors can be estimated in $X^{b_0}$ and lead to a gain of at least $M^{-2D_1}$. The other factors that are not $z_M$ can be bounded trivially using either the induction hypothesis or (5.38) which we just proved and contributes at most an $M^C$

power. Since $D_1 = \delta^5 D$ and $D \gg_{C_\delta} 1$, the power $M^{-2D_1}$ will be more than enough to close the estimate.

(3) It remains to consider the terms on the right hand side of (5.28) that contains exactly one factor $z_M$ (that is, $N_j = M$ for exactly one $r+1 \le j \le q$). By rearranging, we can write this term as $z_{\mathrm{out}} = \mathscr{R}^{M,\iota}\mathscr{L}^\zeta z_M$, where $\iota, \zeta \in \{\pm\}$, and $\mathscr{L}^\zeta$ is the $\mathbb{R}$-linear operator defined in (5.39). Here in (5.39), each $N_j$ $(2 \le j \le q)$ is assumed to be $\le M$ instead of $< M$, but if $N_j = M$ then $y_{N_j}^*$ can only be one of the $\Psi[\mathcal{S}_j, h^{(\mathcal{S}_j,n_j)}]$ terms with $\mathcal{S}_j$ a regular plant of frequency $M$ and size at most $D$ (i.e. $y_{N_j}^*$ is not allowed to be $z_M$ or any Fourier truncation thereof). Our goal here is to prove that (5.40) holds for such $\mathscr{L}^\zeta$. In fact, if (5.40) holds, then by repeating *Step 1* of the proof of Proposition 7.1 we obtain that

$$\|\mathscr{L}^\zeta\|_{X^{b_0}\to X^{b_0}} \le \tau^{(6\kappa)^{-1}}\big(\max_{2\le j\le q} N_j\big)^{-3\varepsilon\delta};$$

summing over all possible choices of $\mathscr{L}^\zeta$ and using also the $X^{b_0} \to X^{b_0}$ bound of $\mathscr{R}^{M,\iota}$, we obtain

$$\|z_{\mathrm{out}}\|_{X^{b_0}} \le \tau^{(8\kappa)^{-1}}\|z_M\|_{X^{b_0}} \le \tau^{(8\kappa)^{-1}} M^{-D_1},$$

which is acceptable. This means that, if we can prove (5.40) for the $\mathscr{L}^\zeta$ as above, then part (4), i.e. the contraction mapping part of $\mathtt{Local}(2M)$ is true, and thus $z_M$, being the unique fixed point of a contraction mapping, does satisfy $\|z_M\|_{X^{b_0}} \le M^{-D_1}$. Combining with Proposition 7.1 and the construction in Section 5.2, we also obtain that $y_M$ defined by (5.17) solves (5.16) with $N$ replaced by $M$. Finally, the bound (5.40) for the $\mathscr{L}^\zeta$ as above also implies part (5) of $\mathtt{Local}(2M)$, since if any $y_{N_j}^*$ in (5.39) is replaced by $z_M$ or its Fourier truncation, then the $M^{-D_1}$ decay will overwhelm any possible loss and immediately imply (5.40).

In summary, we now only need to prove (5.40) for $\mathscr{L}^\zeta$ as in (5.39), where either $N_j < M$ or $N_j = M$ and $y_{N_j}^* = \Psi[\mathcal{S}_j, h^{(\mathcal{S}_j,n_j)}]$. We may assume $\max(N_2, \cdots, N_q) = M$ (otherwise use the induction hypothesis) and replace $z_{N_j}$ by $z_{N_j}^{\mathrm{lo}}$. Applying Lemma 4.2, we can remove the $\chi_\tau$ factor in (5.39) and gain a power $\tau^{\kappa^{-1}}$, which will overwhelm all possible $\tau^{-\theta}$ losses, provided we estimate the expression without $\chi_\tau$ in the stronger $X^{1-b_0,-b}[k \to k']$ norm.

By Proposition 3.7 (1) we can write

$$(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda, \lambda') = \int [\mathscr{M}^\zeta(\lambda, \lambda', \lambda_2, \cdots, \lambda_r)]_{kk'}\,\mathrm{d}\lambda_2\cdots\mathrm{d}\lambda_r, \tag{7.30}$$

where for fixed values of $(\lambda, \lambda', \lambda_2, \cdots, \lambda_r)$, $\mathscr{M}^\zeta(\lambda, \lambda', \lambda_2, \cdots, \lambda_r) = [\mathscr{M}^\zeta(\lambda, \lambda', \lambda_2, \cdots, \lambda_r)]_{kk'}$ is the tensor $\mathscr{M} = \mathscr{M}_{kk_1}$ defined in (6.90), Proposition 6.6 (where we rename $k_1$ as $k'$). Here in (6.90), we assume that $\Psi_{k_j}^{(j)} = \Psi_{k_j}[\mathcal{S}_j', \widehat{h^{(j)}}(\lambda_j)]$ where $\mathcal{S}_j' = \mathtt{Trim}(\mathcal{S}_j, M^\delta)$ and $h^{(j)} = \mathtt{Trim}(h^{(\mathcal{S}_j,n_j)}, M^\delta)$ for $2 \le j \le r$; moreover, for fixed values of $(\lambda, \lambda', \lambda_2, \cdots, \lambda_r)$, the tensor $h = h(\lambda, \lambda', \lambda_2, \cdots, \lambda_r) = [h(\lambda, \lambda', \lambda_2, \cdots, \lambda_r)]_{kk'\cdots k_q}(\lambda_{r+1}, \cdots, \lambda_q)$ satisfies (6.88) and

$$|h| + |\partial_{\lambda_j} h| \lesssim \frac{\tau^{-\theta}}{\langle\lambda\rangle\langle\Omega + \zeta_{r+1}\lambda_{r+1} + \cdots + \zeta_q\lambda_q + \widetilde{\Xi}\rangle},\ r+1 \le j \le q;\quad \widetilde{\Xi} := \zeta_2\lambda_2 + \cdots + \zeta_r\lambda_r + \zeta\lambda' - \lambda.$$

Moreover, $h$ is in fact a function of $(k - \zeta k', |k|^2 - \zeta|k'|^2)$ and $(k_2, \cdots, k_q, \lambda_{r+1}, \cdots, \lambda_q)$ only, in the same manner as in Proposition 6.6.

Like before, we will separate the high modulation case $\max(|\lambda_2|, \cdots, |\lambda_r|) \ge M^{\sqrt{\kappa}}$, and the low modulation case $\max(|\lambda_2|, \cdots, |\lambda_r|) \le M^{\sqrt{\kappa}}$. In the high modulation case we may assume (say) $|\lambda_2| = \max(|\lambda_2|, \cdots, |\lambda_r|) \ge M^{\sqrt{\kappa}}$, then as before, using the induction hypothesis and (5.38), we can fix the values of $k_j$ $(2 \leq j \leq q)$ and $\lambda_j$ $(3 \leq j \leq q)$, and view $\Psi^{(2)}$ as a function of $\lambda_2$ only. Then we obtain, up to a loss of $M^C$, that

$$|(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,\lambda')| \lesssim \frac{\tau^{-\theta}}{\langle\lambda\rangle}\mathbf{1}_{k-\zeta k'=k^*}\int_{\mathbb{R}}\frac{1}{\langle\lambda-\zeta\lambda'-|k|^2+\zeta|k'|^2-\zeta_2\lambda_2+\Xi\rangle}|\Psi^{(2)}(\lambda_2)|\,\mathrm{d}\lambda_2,$$

where $k_*$ is a fixed $\mathbb{Z}^d$ vector, and $\Xi$ is a fixed real number, depending on the choices of the fixed variables. Since $\Psi^{(2)}$ is bounded in $L^2_{\lambda_2}$ with the weight $\langle\lambda_2\rangle^{b_0} \geq M^{\sqrt{\kappa}/2}$, we can gain this $M^{\sqrt{\kappa}/2}$ power (which overwhelms all $M^C$ losses) and apply Cauchy-Schwartz, estimating $\Psi^{(2)}$ only in $L^2_{\lambda_2}$, to obtain that

$$\|(\widehat{\mathscr{L}^\zeta})_{kk'}(\lambda,\lambda')\|_{k\to k'} \lesssim \frac{\tau^{-\theta}}{\langle\lambda\rangle}M^{-\sqrt{\kappa}/4}$$

uniformly in $(\lambda,\lambda')$, which is sufficient to prove (5.40).

Now we can restrict to the low modulation case. With a loss of $M^{C/\sqrt{\kappa}}$, which will be negligible compare to the gain, we can replace the exponents $1-b$ and $\widetilde{b}$ in (5.34)–(5.37) all by $b$. Therefore, by the same arguments as in the proof of Proposition 7.3, we obtain the following:

(a) If $j$ is such that either $n_j = 1$, or $n_j = 0$ and $\min_{\mathfrak{l}\in\mathcal{L}_j} N_{\mathfrak{l}} < M^\delta$, define $\mathfrak{X}_j = \mathfrak{X}_j(\lambda_j)$ to be the smallest value such that the type 1 bounds (6.44)–(6.47) in Proposition 6.4 hold for $\widehat{h^{(j)}}(\lambda_j)$, then we have

$$\left(\int_{\mathbb{R}}\mathfrak{X}_j(\lambda_j)\,\mathrm{d}\lambda_j\right)^2 \lesssim \int_{\mathbb{R}}\langle\lambda_j\rangle^{2b}\mathfrak{X}_j(\lambda_j)^2\,\mathrm{d}\lambda_j \lesssim \tau^{-\theta}M^{C/\sqrt{\kappa}+C\delta}. \tag{7.31}$$

(b) If $n_j = 0$ and $\min_{\mathfrak{l}\in\mathcal{L}_j} N_{\mathfrak{l}} \geq M^\delta$ (in particular $h^{(j)} = h^{(\mathcal{S}_j,0)}$), we may decompose $h^{(j)}$ into $h^{(j,\Gamma_j)} = h^{(\mathcal{S}_j,0,\Gamma_j)}$ as in part (1) of Proposition 5.1; define $\mathfrak{X}_j = \mathfrak{X}_j(\lambda_j,\Gamma_j)$ to be the smallest value such that the type 0 bounds (6.42)–(6.43) in Proposition 6.4 hold for $\widehat{h^{(j,\Gamma_j)}}(\lambda_j)$, then we have (7.17).

Moreover we may also assume $\max(|\lambda|,|\lambda'|) \leq M^{\kappa^2}$, otherwise we exploit the room coming from the exponents $1-b_0 < 1/2$ and $b > 1/2$ to gain a power $M^\kappa$ that will overwhelm all possible losses. These assumptions allow us to apply Proposition 6.6 and get

$$\|[\mathscr{M}^\zeta(\lambda,\lambda',\lambda_2,\cdots,\lambda_r)]_{kk'}\|_{k\to k'} \leq \frac{1}{\langle\lambda\rangle}\prod_{j=2}^r \mathfrak{X}_j \cdot M^{-\varepsilon^5},$$

where $\mathfrak{X}_j = \mathfrak{X}_j(\lambda_j)$ or $\mathfrak{X}_j(\lambda_j,\Gamma_j)$ in case (a) or (b) above. Since $\max(|\lambda_2|,\cdots,|\lambda_r|) \leq M^{\sqrt{\kappa}}$ and $\max(|\lambda|,|\lambda'|) \leq M^{\kappa^2}$, by a meshing argument we can remove a single exceptional set of probability $\leq C_\theta e^{-(\tau^{-1}M)^\theta}$ such that the above holds for all values of $(\lambda,\lambda',\lambda_2,\cdots,\lambda_r)$. Therefore, by taking the $L^2$ norm weighted by $\langle\lambda\rangle^{1-b_0}\langle\lambda'\rangle^{-b}$ in $(\lambda,\lambda')$, then summing in $\Gamma_j$ and integrating in $\lambda_j$, we obtain the bound (5.40) for $\mathscr{L}^\zeta$. This completes the proof of parts (3)–(5) of `Local`$(2M)$ and finishes the inductive proof of Proposition 5.1. □

## 8. Proof of the main results

In this section we prove our main theorems, Theorems 1.1 and 1.6. Theorem 1.1 follows from Proposition 5.1 together with some arguments similar to those in Sections 6–7; Theorem 1.6 is easier and follows from simplified versions of these arguments.

8.1. **Proof of Theorem 1.1.** First, by Proposition 5.1, after removing an exceptional set of probability at most $C_\theta e^{-\tau^{-\theta}}$, we may assume that $\mathtt{Local}(M)$ holds for all $M$. In particular, by (5.18), (5.38) and part (4) of $\mathtt{Local}(M)$ in Proposition 5.1, and in view of the fact that the number of plants $\mathcal{S}$ with frequency $N(\mathcal{S})=N$ and size $|\mathcal{S}|\leq D$ is at most $(\log N)^\kappa$, we conclude that

$$\|y_N\|_{X^{s',b_0}}\leq \tau^{-\theta}N^{(s'-s)/2} \tag{8.1}$$

for each $s'<s$ and each $N$, thus

$$\lim_{N\to\infty}v_N^\dagger=\lim_{N\to\infty}\sum_{N'\leq N}y_{N'}\qquad \text{exists in } X^{s',b_0}. \tag{8.2}$$

Moreover, under all these $\mathtt{Local}(M)$ assumptions, $v_N^\dagger$ solves (5.14), thus it must equal $v_N$ on $J=[-\tau,\tau]$. By definition $e^{it\Delta}v_N^\dagger$ also equals the solution $\widetilde{u_N}$ of (5.6) on $J$.

Now, define

$$B_N(t)=\frac{p+1}{2}\chi(t)\int_0^t\chi(t')\fint_{\mathbb{T}^d}W_N^{p-1}(e^{it'\Delta}v_N^\dagger)\,\mathrm{d}t', \tag{8.3}$$

akin to the one in (5.2) but with the smooth cutoff $\chi$, and define $u_N^\dagger(t)=e^{it\Delta}v_N^\dagger(t)\cdot e^{-iB_N(t)}$, then $u_N^\dagger$ equals $u_N$, which is the solution to (1.7), on $J$. By analyzing the term $W_N^{p-1}$ as in [36], Proposition 2.2, and applying the same calculations as in Section 5.1.2, we can rewrite $B_N(t)$ as

$$B_N(t)=\sum_{3\leq q\leq p}a'_{pq}(m_N^*)^{(p-q)/2}\cdot\mathcal{I}_\chi\mathbb{A}\mathcal{M}_{q-1}(v_N^\dagger,\cdots,v_N^\dagger)(t). \tag{8.4}$$

In the above $q$ runs over odd integers, $a'_{pq}$ are constants, $m_N^*$ is defined as in Section 5.1.2, $\mathbb{A}$ is the projection onto frequency $k=0$, $\mathcal{M}_{q-1}$ is defined as in (5.11)–(5.12) but with $q$ replaced by $q-1$. Note that instead of the *simplicity* condition as in Definition 3.1, here the coefficients $c_{kk_1\cdots k_{q-1}}$ satisfy the slightly different *input-simplicity* condition, namely that $c_{kk_1\cdots k_{q-1}}$ depends only on the set of pairings in $(k_1,\cdots,k_{q-1})$, and $c_{kk_1\cdots k_{q-1}}=0$ unless any pairing in $(k_1,\cdots,k_{q-1})$ is over-paired. However, in view of the projection $\mathbb{A}$ to $k=0$, this input-simplicity condition will imply the same tensor norm estimates in Section 4 that are proved under the simplicity condition.

Therefore, by decomposing $v_N^\dagger$ using (5.17) and repeating the proofs[1] of Sections 6–7, after possibly removing another exceptional set of probability not exceeding $C_\theta e^{-\tau^{-\theta}}$, we conclude that $B_N(t)$ converges to some $B(t)$ in $H_t^{b_0}\hookrightarrow C_t^0$ as $N\to\infty$, where recall $b_0>1/2$ as in (3.3). Therefore $u_N$, which equals $u_N^\dagger=e^{it\Delta}v_N^\dagger\cdot e^{-iB_N}$ on $J$, converges in $C_t^0H_x^{s-}(J)$ as $N\to\infty$. This limit $u$ has the explicit expansion (which is valid on $J$)

$$u_k(t)=e^{-i(|k|^2t+B(t))}\bigg[\sum_{n\in\{0,1\}}\sum_{|\mathcal{S}|\leq D}\sum_{k_\mathcal{U}}\bigg\{\sum_{k_\mathcal{V}}\int\mathrm{d}\lambda_\mathcal{V}\cdot h_{kk_\mathcal{U}}^{(\mathcal{S},n)}(k_\mathcal{V},\lambda_\mathcal{V})\\ \times\prod_{\mathfrak{f}\in\mathcal{V}}(\widehat{z_{N_\mathfrak{f}}})_{k_\mathfrak{f}}^{\zeta_\mathfrak{f}}(\lambda_\mathfrak{f})\bigg\}\prod_{\mathfrak{l}\in\mathcal{U}}\frac{\Delta_{N_\mathfrak{l}}(k_\mathfrak{l})}{\langle k_\mathfrak{l}\rangle^\alpha}\cdot g_{k_\mathfrak{l}}^{\zeta_\mathfrak{l}}(\omega)+z_k(t)\bigg], \tag{8.5}$$

where the sum is taken over all regular plants $\mathcal{S}$ with $|\mathcal{S}|\leq D$, the random tensors $h^{(\mathcal{S},n)}$ and the functions $z_N$ are defined as in Section 5.2, and the remainder $z$ belongs to $C_t^0H_x^{D_1-1}(J)$.

[1] If needed, we can always view an $\mathbb{R}$-multilinear operator of degree $q-1$ as one of degree $q$ by adding a trivial input function.

It now remains to prove that the nonlinearity $W^p(u)$ defined by (1.8) exists as a spacetime distribution, and that $u$ solves (1.1) in the distributional sense. Define

$$v^\dagger = \lim_{N\to\infty} v_N^\dagger = \sum_{N'} y_{N'}, \quad \text{and} \quad u^\dagger(t) = e^{it\Delta}v^\dagger(t)\cdot e^{-iB(t)}, \tag{8.6}$$

then $u$ equals $u^\dagger$ on $J$, so it suffices to prove that

$$\lim_{N\to\infty} W_N^p(\Pi_N u^\dagger) = \lim_{N\to\infty} \Pi_N W_N^p(\Pi_N u^\dagger) = \lim_{N\to\infty} W_N^p(u_N^\dagger) = \lim_{N\to\infty} \Pi_N W_N^p(u_N^\dagger) \tag{8.7}$$

in the sense of distributions. As $B_N \to B$, we may replace $u_N^\dagger$ by $e^{it\Delta}v_N^\dagger$ and $u^\dagger$ by $e^{it\Delta}v^\dagger$; then arguing as in Section 5.1.2, we can reduce to analyzing the terms

$$\sum_{3\leq q\leq p} a''_{pq}(m_N^*)^{(p-q)/2}\,\Pi\mathcal{M}_q(w,\cdots,w)(t), \tag{8.8}$$

where $q$ is odd as before, $a''_{pq}$ are constants, $\Pi$ is either 1 or $\Pi_N$, $w$ is either $v_N^\dagger$ or $\Pi_N v^\dagger$, and $\mathcal{M}$ is as in (5.11)–(5.12) but is input-simple instead of simple. Decomposing $w$ using (5.17), it then suffices to show that

$$\Phi := \mathcal{M}_q(y_{N_1},\cdots,y_{N_q}) \to 0, \quad \text{as } N_{\max} := \max(N_1,\cdots,N_q)\to\infty, \tag{8.9}$$

in the sense of distributions. In fact we shall control the term $\mathcal{I}_\chi\Phi$ which, by Lemma 4.1, satisfies

$$\widehat{\mathcal{I}_\chi\Phi_k}(\lambda) = \int_{\mathbb{R}} \mathcal{I}(\lambda,\lambda')\widehat{\Phi_k}(\lambda')\,\mathrm{d}\lambda', \qquad |\mathcal{I}|+|\partial\mathcal{I}| \lesssim \left(\frac{1}{\langle\lambda\rangle^3}+\frac{1}{\langle\lambda-\lambda'\rangle^3}\right)\frac{1}{\langle\lambda'\rangle}. \tag{8.10}$$

Note that, apart from simple modifications, $\mathcal{I}_\chi\Phi$ essentially has the same structure as $\Psi^{(\mathcal{S},n)}$ in (5.38) with the associated tensor $h^{(\mathcal{S},n)}$ as in (5.26) and (5.27), so it can be estimated in the same way as in Sections 6–7. We only make two additional observations:

(a) The proofs of Sections 6–7 do not depend on any cancellation in (8.10), so the same arguments can be applied for the term $\mathcal{I}_\chi^{\mathrm{abs}}\Phi$ defined by

$$\widehat{\mathcal{I}_\chi^{\mathrm{abs}}\Phi_k}(\lambda) = \int_{\mathbb{R}} \langle\lambda'\rangle^{-1}(\langle\lambda\rangle^{-3}+\langle\lambda-\lambda'\rangle^{-3})|\widehat{\Phi_k}(\lambda')|\,\mathrm{d}\lambda', \tag{8.11}$$

leading to the control of $\mathcal{I}_\chi^{\mathrm{abs}}\Phi$ in the $X^{-d,b_0}$ norm (here the exponent $-d$ has to do with the potential power loss associated with fixing $k$, see observation (b) below), which in turn implies the control of $\Phi$ in the sense of distributions—for example, due to the trivial bound

$$\|\mathcal{I}_\chi^{\mathrm{abs}}\Phi\|_{X^{-d,b_0}} \gtrsim \sum_k \int_{\mathbb{R}} \langle k\rangle^{-2d}\langle\lambda'\rangle^{-4}|\widehat{\Phi_k}(\lambda')|\,\mathrm{d}\lambda'.$$

(b) The $\mathbb{R}$-multilinear operator $\mathcal{M}_q$ is input-simple instead of simple. However, in order to control the $X^{-d,b_0}$ norm of $\mathcal{I}_\chi^{\mathrm{abs}}\Phi$, we may fix the value of $k$ in $\widehat{\Phi_k}(\lambda)$ and it suffices to get a bound uniform in $k$ thanks to the exponent $-d$. Now once $k$ is fixed, the pairings between $k$ and any $k_j$ become unimportant (they no longer cause losses in any counting estimate as the paired and over-paired variables now have only one choice), so an input-simple $\mathbb{R}$-multilinear operator can be treated in the same way as a simple one, similar to the analysis of (8.4) above.

With the observations above, after possibly removing another exceptional set of probability not exceeding $C_\theta e^{-\tau^{-\theta}}$, the same proofs of Sections 6–7 can be carried out to obtain, for example, that

$$\|\mathcal{I}_\chi^{\mathrm{abs}}\Phi\|_{X^{-d,b_0}} \leq \tau^{-\theta}(N_{\max})^{-\delta^6}. \tag{8.12}$$

This proves (8.9) and thus finishes the proof of Theorem 1.1.

8.2. **Proof of Theorem 1.6.** Fix $\varepsilon > 0$ small enough depending on $(d,p)$, $s - s_{pr}$ and $(p-1)(s - s_{pr}) - \nu$ (note that this is different from (3.1)). Let $(\delta, D, \kappa, \theta)$ etc. be defined as in Section 3.1, we may assume $N \gg_{C_\theta} 1$. Let

$$B(T) = \frac{p+1}{2}\int_0^T \fint_{\mathbb{T}^d} |u_{\mathrm{ho}}|^{p-1}\,\mathrm{d}T',$$

and repeat the gauging, conditioning and conjugating arguments as in Section 5.1, except that $\sigma_N$ is replaced by 0 since we are dealing with the nonlinearity $|u_{\mathrm{ho}}|^{p-1}u_{\mathrm{ho}}$ instead of the Wick-ordered one. After also rescaling time, we can write $N$-certainly that

$$\widetilde{u}(T,x) = \sum_k w_k(N^{-\nu}T)e^{i(k\cdot x - |k|^2T - B(T))}, \tag{8.13}$$

where $w_k(t)$ is the solution to the system

$$w_k(t) = f_k - iN^\nu \sum_{3\leq q\leq p} a_{pq}(m_{\mathrm{ho}})^{(p-q)/2}\int_0^t \mathcal{M}_q(w,\cdots,w)_k(t')\,\mathrm{d}t' \tag{8.14}$$

similar to (5.9), where

$$m_{\mathrm{ho}} = \fint_{\mathbb{T}^d}|u_{\mathrm{ho}}|^2 = N^{-2\alpha}\sum_k \phi^2\Big(\frac{k}{N}\Big)|g_k|^2,$$

which is $N$-certainly bounded by $N^{d-2\alpha}$; like in (5.14) we also consider the solution $w^\dagger$ to

$$w_k^\dagger(t) = \chi(t)f_k - iN^\nu \sum_{3\leq q\leq p} a_{pq}(m_{\mathrm{ho}})^{(p-q)/2}\cdot \mathcal{I}_\chi\mathcal{M}_q(w^\dagger,\cdots,w^\dagger)_k(t). \tag{8.15}$$

Here in (8.14)–(8.15) we have $f_k = \gamma_k\cdot\eta_k(\omega)$ with constants $|\gamma_k| \leq N^{-\alpha+\theta}\langle k\rangle^\theta|\phi(k/N)|$, $a_{pq}$ are bounded constants and $|m_{\mathrm{ho}}| \leq N^{d-2\alpha}$ ($\alpha$ is defined as in (1.10)), and the $\mathbb{R}$-multilinear expression

$$\mathcal{M}_q(w^{(1)},\cdots,w^{(q)})_k(t') = \sum_{\zeta_1k_1+\cdots+\zeta_qk_q=k} c_{kk_1\cdots k_q}\cdot e^{iN^\nu t'\Omega}\prod_{j=1}^q (w^{(j)})_{k_j}^{\zeta_j}(t'), \tag{8.16}$$

with the signs $\zeta_j$ and coefficients $c_{kk_1\cdots k_q}$ as in (5.11).

Since the initial data $f_k$ is uniformly distributed in $k$, the analysis of (8.14) will be significantly simpler than the arguments in Sections 5–7. More precisely, we will only need the tensors[1] $h^{(\mathcal{S},n)}$ with $n=0$ (which will be constant tensors, i.e. do not depend on $\omega$), and instead of the full plant structure, we will only need its tree part (which is called $\mathcal{L}$ before, see Definition 3.2), leaf pairings and signs of leaves. As such, we will define (in this proof only) $\mathcal{S}$ to be a set of leaves $\mathfrak{l}$ with possible pairings, together with the sign $\zeta_\mathfrak{l} \in \{\pm\}$ for each $\mathfrak{l}\in\mathcal{S}$. Let $\mathcal{P}$ (resp. $\mathcal{U}$) be the set of paired (resp. unpaired) leaves, we require that $\zeta_{\mathfrak{l}'} = -\zeta_\mathfrak{l}$ for any pair $(\mathfrak{l},\mathfrak{l}')$, and that $\sum_{\mathfrak{l}\in\mathcal{L}}\zeta_\mathfrak{l} = 1$. The $\mathcal{S}$ tensors $h = h_{kk_\mathcal{U}}$ are defined as in Definition 3.4 but instead of condition (1) we only assume[2] $\langle k_\mathfrak{l}\rangle \leq N^{1+\theta}$ for each $\mathfrak{l}\in\mathcal{S}$, and $\Psi_k = \Psi_k[\mathcal{S},h]$ is defined as in (3.8). We do not need the `Trim` function, and

[1]Because we do not need to distinguish the low-frequency inputs as there is only one scale.

[2]The choice of $N^{1+\theta}$ is because $\phi$ is not compactly supported and may have a Schwartz tail.

$\texttt{Merge}$ is defined in the same way as in Definition 3.6 (with only the tree part, and without the frequency parameters such as $N_{\mathfrak{l}}$ for leaves $\mathfrak{l}$; also the factors $\Delta_{N_{\mathfrak{l}}}\gamma_{k_{\mathfrak{l}}}$ in (3.11) are replaced by $\gamma_{k_{\mathfrak{l}}}$).

Next, for any $\mathcal{S}$ with $|\mathcal{S}|\leq D$, define the $\mathcal{S}$ tensor $h^{(\mathcal{S})}=h^{(\mathcal{S})}_{kk_{\mathcal{U}}}(t)$ inductively by

$$h^{(\mathcal{S})}_{kk_{\mathcal{U}}}(t)=\mathbf{1}_{|\mathcal{S}|=1}\cdot\mathbf{1}_{k=k_{\mathfrak{l}}}\mathbf{1}_{\langle k\rangle\leq N^{1+\theta}}+N^{\nu}\sum_{3\leq q\leq p}a_{pq}(m_{\mathrm{ho}})^{(p-q)/2}\sum_{(*)}\mathcal{I}_{\chi}H_{kk_{\mathcal{U}}}(t),\tag{8.17}$$

similar to (5.26). Here the sum $\sum_{(*)}$ is taken over $\mathscr{B}=(\zeta_1,\cdots,\zeta_q)$, $\mathcal{S}_j$ as defined above, and $\mathscr{O}$, such that $\mathcal{S}=\texttt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_q,\mathscr{B},\mathscr{O})$; $H=H_{kk_{\mathcal{U}}}$ is defined by $H=\texttt{Merge}(h^{(\mathcal{S}_1)},\cdots,h^{(\mathcal{S}_q)},h,\mathscr{B},\mathscr{O})$ with

$$h=h_{k_1\cdots k_q}(t')=\mathbf{1}_{k=\zeta_1k_1+\cdots+\zeta_qk_q}\cdot\mathbf{1}_{\langle k\rangle\leq N^{1+2\theta}}\prod_{j=1}^{q}\mathbf{1}_{\langle k_j\rangle\leq N^{1+2\theta}}\cdot c_{kk_1\cdots k_q}e^{iN^{\nu}t'\Omega}\tag{8.18}$$

similar to (5.23)–(5.25). By arguing similarly as in Propositions 4.9, 6.3, 6.5 and 7.2, we can prove inductively that $h^{(\mathcal{S})}_{kk_{\mathcal{U}}}$ satisfies the support condition (5.30); moreover, for each subpartition $(B,C)$ of $\mathcal{U}$, we can prove that

$$\int_{\mathbb{R}}\langle\lambda\rangle^{2b}\bigg(\sum_{\Gamma\in\mathbb{Z}}\|\widehat{h^{(\mathcal{S},\Gamma)}_{kk_{\mathcal{U}}}}(\lambda)\|_{kk_B\to k_C}\bigg)^2\,\mathrm{d}\lambda\leq\big(N^{(\alpha-2\varepsilon)|B\cup C|}N^{-\varepsilon|\mathcal{P}|}N^{\varepsilon|E|}\cdot\mathcal{X}_0\big)^2,\tag{8.19}$$

similar to (5.32). Here $\mathcal{P}$ is the set of paired leaves as defined above, $E=\mathcal{U}\backslash(B\cup C)$, and $h^{(\mathcal{S},\Gamma)}$ is the restriction of $h^{(\mathcal{S})}$ to the set (5.31) as in part (1) of Proposition 5.1; moreover, $\mathcal{X}_0$ is defined to be $N^{-(\alpha-2\varepsilon)}$ if $C\neq\varnothing$, and $N^{\max(0,d/2-\alpha+2\varepsilon)}$ if $C=E=\varnothing$, and $N^{-\varepsilon\delta}$ if $B=C=\varnothing$, and 1 otherwise, as in (5.33).

We note that the proof of (8.19) is much easier than that of (5.32), as we do not need to apply the careful selection algorithm in Proposition 6.2. There are only two nontrivial differences. The first is due to the extra $N^{\nu}$ factor in $e^{iN^{\nu}t'\Omega}$ in (8.18), which actually helps us as $\nu>0$, since $N^{\nu}\Omega$ belonging to an interval of length $O(1)$ will force $\Omega$ to belong to an interval of length $O(1)$. The second is the extra factor $N^{\nu}$ on the right hand side of (8.17), which gets cancelled by the $N^{-\alpha}$ decay of $\gamma_k$ and the $(m_{\mathrm{ho}})^{(p-q)/2}$ factor in (8.17), in view of the inequality $\nu\leq(p-1)(\alpha-\alpha_0-10\varepsilon)$ ($\alpha_0$ is defined as in (3.1)). In particular we have

$$(m_{\mathrm{ho}})^{(p-q)/2}N^{\nu}\cdot\prod_{j=2}^{q}N^{\alpha_0'+\theta}\leq\prod_{j=2}^{q}N^{\alpha-8\varepsilon},\quad\alpha_0':=\frac{d}{2}-\frac{1}{q-1},$$

so the bound for the $h$ tensor we see when merging $h^{(\mathcal{S}_j)}$ tensors—which is the one on the right hand side of (4.26) with $p$ replaced by $q$—can be cancelled by the $N^{-\alpha}$ decay of $\gamma_k$ with extra gain of $N^{\varepsilon}$ powers, after being multiplied by $N^{\nu}$ and $(m_{\mathrm{ho}})^{(p-q)/2}$ factors.

Now, with (8.19) available, we can construct the solution $w^{\dagger}$ to (8.15) by the ansatz

$$w_k^{\dagger}(t)=\sum_{|\mathcal{S}|\leq D}\Psi_k[\mathcal{S},h^{(\mathcal{S})}(t)]+z_k^{\dagger}(t),\tag{8.20}$$

where $z^{\dagger}$ satisfies the equation

$$z_k^{\dagger}(t)=\chi(t)\mathbf{1}_{\langle k\rangle\geq N^{1+\theta}}\cdot f_k-iN^{\nu}\sum_{3\leq q\leq p}a_{pq}(m_*)^{(p-q)/2}\sum_{(v^{(1)},\cdots,v^{(q)})}\mathcal{I}_{\chi}\mathcal{M}_q(v^{(1)},\cdots,v^{(q)})_k(t),\tag{8.21}$$

with the sum taken over $(v^{(1)},\cdots,v^{(q)})$ such that each $v^{(j)}$ is either $z^{\dagger}$ or $\Psi_k[\mathcal{S}_j,h^{(\mathcal{S})j)}]$ for some $\mathcal{S}_j$ with $|\mathcal{S}_j|\leq D$, and that either (i) at least one $v^{(j)}=z$, or (ii) $v^{(j)}=\Psi_k[\mathcal{S}_j,h^{(\mathcal{S}_j)}]$ for each $j$, and $\mathcal{S}=\texttt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_r,\mathscr{B},\mathscr{O})$ satisfies $|\mathcal{S}|>D$ (for any $\mathscr{O}$).

Note that, due to the lack of projection $\Pi_N$ on the right hand side of (1.11), the remainder $z^\dagger$ is not guaranteed to have compact support in $k$. Thus, instead of the $X^{b_0}$ norm as in Proposition 5.1, we should control the $\widetilde{X}^{C_0,b_0}$ norm of $z^\dagger$, defined by

$$\|z^\dagger\|_{\widetilde{X}^{C_0,b_0}}^2 := \sum_k \int_{\mathbb{R}} \left(1+\frac{|k|}{N^{1+\theta}}\right)^{2C_0} \langle\lambda\rangle^{2b_0} |(\widehat{z^\dagger})_k(\lambda)|^2 \,\mathrm{d}\lambda,$$

where $C_0>0$ is a large absolute constant depending only on $(d,p)$ and $b_0$ is as in (3.3). Indeed we shall prove that $\|z^\dagger\|_{\widetilde{X}^{C_0,b_0}} \leq N^{-D_1}$ by a contraction mapping. Note that again we only need to prove that the right hand side of (8.21) satisfies this same inequality assuming that $z^\dagger$ does. This right hand side contains four types of terms, which are listed as follows[1]:

- The first term on the right hand side of (8.21), which is acceptable because $|f_k| \leq |\phi(k/N)|$ where $\phi$ is Schwartz and $\langle k\rangle \geq N^{1+\theta}$.
- The term where at least two $v^{(j)}$ equal $z^\dagger$, which is acceptable thanks to the decay $N^{-D_1}$ of $z^\dagger$ and the choice of large $C_0$.
- The term where $v^{(j)} = \Psi_k[\mathcal{S}_j, h^{(\mathcal{S}_j)}]$ for each $j$. This term is acceptable because it can be written as a linear combination of $\Psi_k[\mathcal{S}, h^{(\mathcal{S})}]$ for some $\mathcal{S} = \mathtt{Merge}(\mathcal{S}_1,\cdots,\mathcal{S}_q,\mathscr{B},\mathscr{O})$ with $|\mathcal{S}|>D$, and the corresponding tensor $h^{(\mathcal{S})} = \mathtt{Merge}(h^{(\mathcal{S}_1)},\cdots,h^{(\mathcal{S}_q)},h,\mathscr{B},\mathscr{O})$, which can be shown to satisfy (8.19) by repeating the proofs above. By applying Lemma 4.4 again, this $\Psi_k$ term can be bounded by $N^{-D_1}$ in $X^{b_0}$, and hence in $\widetilde{X}^{C_0,b_0}$ because it is supported in $|k| \lesssim N^{1+\theta}$.
- Finally, the term where exactly one $v^{(j)}$ equals $z^\dagger$. This term can be written as an $\mathbb{R}$-linear operator $\mathscr{L}^\zeta$ applied to $z^\dagger$, where this $\mathscr{L}^\zeta$ has similar form as the one in (5.39). Now by repeating the same arguments as in Propositions 4.9, 6.2, 6.6, 7.1 and 7.4, we can bound the $X^{b_0}\to X^{b_0}$ norm of this operator by a negative power of $N$. As the kernel $(\mathscr{L}^\zeta)_{kk'}$ is supported in $|k-\zeta k'| \lesssim N^{1+\theta}$, by Lemma 4.3, the $\widetilde{X}^{C_0,b_0}\to\widetilde{X}^{C_0,b_0}$ norm of $\mathscr{L}^\zeta$ is also bounded by a negative power of $N$, so this term is also acceptable.

As such, we have closed the estimates for $z^\dagger$ and obtained the solution $w^\dagger$ in the form of (8.20). This means that the equation (8.15) is well-posed at least up to time $t=1$, so the equation (1.11) is well-posed at least up to time $T=N^\nu$. Moreover (1.12) easily follows from the bound for $z^\dagger$, as well as the bounds for nonlinear components $\Psi_k[\mathcal{S}, h^{(\mathcal{S})}]$ with $|\mathcal{S}|>1$, which in turn follow from (8.19) and Lemma 4.4. This completes the proof of Theorem 1.6.

## 9. Final remarks

In this section we make some final remarks. These include a comparison with parabolic equations in Section 9.1 and some future directions in Section 9.2. We also list some open problems.

### 9.1. **Comparison with parabolic equations.**

The random data Schrödinger equation

$$(i\partial_t+\Delta)u = W^p(u), \quad u(0)=f(\omega), \tag{9.1}$$

is closely linked to, and fundamentally different from, the stochastic heat equation

$$(\partial_t-\Delta)u = \widetilde{W^p}(u)+\zeta, \tag{9.2}$$

if both are suitably renormalized. In this section we will explain their differences and connections.

[1] The reader may notice the similarity with the construction of $z_M$ in (5.28).

9.1.1. *Difference in scaling.* In Section 1.2 we explained the heuristics behind the *probabilistic scaling* critical index $s_{pr}$ for (9.1). In fact the same philosophy can be applied to (9.2), leading to the *parabolic scaling* critical index $s_{pa}$ in Remark 1.12, which is the one appearing in works such as [52, 46]. Note that $s_{pa}$ is strictly lower than $s_{pr}$.

Fix some value of $s$, and let $\alpha = s + d/2$. As in Section 1.2 we will make simplifications to (9.2) by replacing the nonlinearity by $\mathcal{N}_{\mathrm{np}}$ defined in (1.13) and neglecting the renormalization[1]; we also set initial data $u(0) = 0$. Similar to (1.18), assume the noise $\zeta$ (or its regularization) has the form

$$\zeta(t,x) = N^{-\alpha+1} \sum_{|k|\sim N} \partial_t \beta_k(t) \cdot e^{ik\cdot x},$$

where $\beta_k(t)$ are independent Brownian motions. Let $\psi = (\partial_t - \Delta)^{-1}\zeta$ be the linear evolution of noise (which plays the same role as $e^{it\Delta}u(0)$ in Section 1.2), then

$$\psi(t,x) = N^{-\alpha} \sum_{|k|\sim N} G_k(t) e^{ik\cdot x}, \quad G_k(t) := N \int_0^t e^{-(t-t')|k|^2} \,\mathrm{d}\beta_k(t').$$

For fixed $|t| \sim 1$ these $G_k(t)$ form a collection of independent Gaussian variables with $\mathbb{E}|G_k(t)|^2 \sim 1$, hence $\psi$ is bounded in $C_t^0 H_x^s$ (also in $C_t^0 C_x^s$ by Khintchine's inequality), just as in Section 1.2.

Now, plugging into (the simplified version of) (9.2), we need to control the first nonlinear iteration

$$u^{(1)}(t) = \int_0^t e^{(t-t')\Delta} \mathcal{N}_{\mathrm{np}}(\psi(t'))\,\mathrm{d}t', \tag{9.3}$$

where $|t| \sim 1$, in $H^s$ (or equivalently $C^s$). Similar to Section 1.2, on the Fourier side we have

$$u_k^{(1)}(t) \sim N^{-p\alpha} \sum_{\substack{k_j\in\mathbb{Z}^d, |k_j|\sim N\\ k_1-\cdots+k_p=k}} \int_0^t e^{-(t-t')|k|^2} G_{k_1}(t')\overline{G_{k_2}(t')}\cdots G_{k_p}(t')\,\mathrm{d}t'. \tag{9.4}$$

Suppose $|k| \sim N$, using the square root cancellation in the sum in (9.4) (from independence, as in Section 1.2) and the $N^{-2}$ gain from the $t'$ integral, we see that with high probability, the inner sum-integral has size $N^{(pd-d)/2-2}$, hence

$$\|u^{(1)}(t)\|_{H^s} \sim N^{-(p-1)s-2}; \qquad \|u^{(1)}(t)\|_{H^s} \lesssim 1 \Leftrightarrow s \geq -\frac{2}{p-1} := s_{pa}.$$

We make a few observations on the above heuristic calculation:

(a) It is no surprise that $s_{pa} = s_{cr} - d/2$. Indeed this makes $C^{s_{pa}}$ and $H^{s_{cr}}$ have equal scaling, and in the usual (deterministic) sense $H^{s_{cr}}$, thus also $C^{s_{pa}}$, is critical for the heat equation. The effect of randomness then comes through Khintchine's inequality, where a Gaussian random function which belongs to $H^{s_{pa}}$ must also belong to $C^{s_{pa}}$ which scaling-wise equals the critical space $H^{s_{cr}}$. This is essentially how $s_{pa}$ is calculated in [52].

(b) The above argument does not work for Schrödinger equations, because even though $H^{s_{cr}}$ and $C^{s_{pa}}$ are still scaling critical in the usual sense, the latter is not compatible with Schrödinger flows. However, this does not tell us what *is* the right notion of criticality for Schrödinger.

(c) To exactly see the difference between $s_{pa}$ and $s_{pr}$, we have to compare the calculations in here and in Section 1.2. Note that in (9.4) the $t'$ integral gains two derivatives $N^{-2}$; in comparison

[1] Even with the $\mathcal{N}_{\mathrm{np}}$ nonlinearity, some renormalization may still be needed for higher order iterations, but not for the first nonlinear iteration which is discussed here. Also whether $u$ is real or complex valued, and whether $\mathcal{N}_{\mathrm{np}}$ contains complex conjugates, does not affect the scaling heuristics.

in (1.16) and (1.19) there is no derivative gain—since the Schrödinger flow has no smoothing—only the denominator $\langle\Omega\rangle^{-1}$ which restricts to the submanifold $\Omega = 0$. This is the fundamental difference between heat and Schrödinger that eventually leads to different scalings in the random setting.

(d) More precisely, note that restricting to $\Omega = 0$ reduces the number of dimensions by two. In the deterministic setting, this gains two derivatives $N^{-2}$ in the summation in (1.16), which matches the two-derivative gain from heat, leading to the same criticality threshold $s_{cr}$; however in the random setting the summation in (1.19) gets square rooted due to randomness, which means the $N^{-2}$ gain *also* gets square rooted, leading to the different criticality thresholds $s_{pa}$ and $s_{pr}$.

9.1.2. *Necessary renormalizations.* Another difference between our theory and the parabolic theories is that, in the latter more and more renormalization terms are needed when one gets close to criticality (for example with the $\Phi^4_{4-\delta}$ model [15, 22]), while in the former we stay with Wick ordering in the full subcritical range.

The main reason for this is the difference in the notions of scaling. For example, in the $\Phi^4_3$ setting where $(d, p) = (3, 3)$, if solutions have regularity $C^{-1/2-}$ or equivalently $H^{-1/2-}$, then (9.2) is still subcritical though needs a log correction $3C_2$ beyond Wick ordering (see (1.22)), but (9.1) is already *critical* relative to the probabilistic scaling (comparable to the *four dimensional* $\Phi^4_4$ problem which is critical relative to the parabolic scaling, due to the reason explained in Section 9.1.1). Conversely, if (9.1) is subcritical, then a calculation shows that the $3C_2$ in (1.22) will not appear as $\mathbb{E}(\Upsilon \cdot \vee)$ is not divergent in the limit, so (9.2) only needs Wick ordering.

More precisely, for (9.2) there are two types of renormalization terms[1], namely those coming from the mass (which is just Wick ordering for $\Phi^4_3$ in (1.22)) and those not coming from the mass (such as the log term in (1.22) and the further corrections described in [15], Section 2.8.2 for $\Phi^4_{4-\delta}$). Now for (9.1) only the mass terms diverge (and need to be renormalized) in the probabilistically subcritical range; moreover since the Schrödinger equation *conserves* mass, we can always replace the mass by the mass of initial data, which just leads to Wick ordering and no further renormalization is needed.

Given this difference, one might ask whether for (9.1) we can go strictly below $s_{pr}$ and down to $s_{pa}$ by including additional counterterms such as the ones in [15]. We believe the answer is *no* due to the following reason. In all previous works, the counterterms in the renormalization process are needed because certain *specific terms* in the formal expansion of the solution with respect to the random initial data or noise become unbounded when $\varepsilon \to 0$ (or $N \to \infty$ in the setting of Theorem 1.1); however if one considers (9.1) in the supercritical range $s < s_{pr}$, then an extended version of the calculation in Section 1.2 shows that *every* term in the formal expansion—not just those with a renormalizable structure—becomes unbounded (say with respect to the regularity of initial data) as $N \to \infty$. This is similar to[2] what happens to (9.2) when $s < s_{pa}$. Therefore, it is at least highly improbable that the local-in-time problem for (9.1) can be solved perturbatively via a renormalization process similar to those in the theory of regularity structures [52, 54, 55, 15].

[1] This distinction may be artificial from the regularity structure perspective, but is convenient in comparison with the dispersive case here.

[2] Put in another way, imagine one has all the input and output frequencies being the same in the nonlinearity. Then (9.2) is locally well-posed, without any renormalization or special arguments, if and only if $s > s_{pa}$, while for (9.1) the same thing holds if and only if $s > s_{pr}$.

9.1.3. *Invariant measures and quantum field theory.* If the random initial data of (9.1) is given by (1.2) with $\alpha = 1$, or if $\zeta$ in (9.2) is the spacetime white noise, then both equations will possess the same formally invariant *Gibbs measure*, which is the $\Phi_d^{p+1}$ measure in quantum field theory (up to real/complex distinction), formally defined by

$$d\mu \sim \exp\left[\frac{-2}{p+1}\int_{\mathbb{T}^d} \widetilde{W}^{p+1}(u)\,dx\right]\cdot \exp\left[-\int_{\mathbb{T}^d}|\nabla u|^2\,dx\right]\prod_{x\in\mathbb{T}^d} dx \tag{9.5}$$

for some renormalization $\widetilde{W}^{p+1}$ of $|u|^{p+1}$. The justification of the formal definition (9.5) is a major problem in constructive quantum field theory, see [44, 66, 63] and recently [7]. It has been done in dimension $d \in \{1,2\}$ for any $p$, and in dimension $d = 3$ for $p = 3$. The other cases are not super-renormalizable in the sense of [22], and such constructions are either unknown or proved impossible [1, 2, 41].

The study of the dynamics of the measure (9.5) under the flow of (9.2), commonly known as *stochastic quantization*, starts with [65]. The invariance of the $\Phi_d^{p+1}$ measure with $d \in \{1,2\}$ and any $p$ is proved in [30]. Recent developments of parabolic theories has led to the resolution of the $\Phi_3^4$ case, with proof of invariance in [62, 4].

On the other hand, the Gibbs measure problem for (9.1) is harder, both conceptually and technically, due to lack of smoothing and (consequently) the different scalings as described above. The invariance of $\Phi_1^{p+1}$ (for any $p$) and $\Phi_2^4$ measures are proved by Bourgain [10, 11] (see also [59]). The $\Phi_2^{p+1}$ case for $p \geq 5$ is much more challenging and is resolved only in our recent work [36]. This matches the results of the Gibbs measure problem for (9.1) with those of the measure construction problem, and of the stochastic quantization problem for (9.2), *except* in the $\Phi_3^4$ case $(d,p) = (3,3)$.

The Gibbs measure problem for $(d,p) = (3,3)$ has two main difficulties. First it is probabilistically critical. This is not as bad as supercritical cases which we believe—as mentioned above—cannot be renormalized (at least through a process similar to [52]), but still log divergences seem unavoidable in all aspects, even for short time. Second, the $\Phi_3^4$ measure is mutually singular with the reference Gaussian measure, as proved in [8], thus $f(\omega)$ in (9.1) will not be given by the simple formula[1] (1.2). Therefore the hope is to somehow get rid of the log divergences by moving to the right measure, i.e. Gibbs instead of Gaussian, but then a local solution theory has to be developed without independence of Fourier coefficients.

**Open problem 1.** Prove invariance of the Gibbs measure for (9.1), possibly with suitable renormalizations, when $d = p = 3$.

9.2. **Future directions.** Though in this paper we have restricted to Schrödinger equations, our method can be applied to more general settings. In this last section we list some future directions.

9.2.1. *The stochastic setting.* Consider (9.1), but with additive noise $\zeta$ instead of random data, for example

$$(i\partial_t + \Delta)u = W^p(u) + \zeta, \quad u(0) = 0,$$

[1] This should be compared to the stochastic quantization problem for (9.2) where the solution theory relies on the *Gaussian noise* instead of the non-Gaussian measure, as observed in [8].

see [31, 39, 37, 25] for some previous works. Here the role of the linear evolution $e^{it\Delta}f(\omega)$ is played by $\psi := (i\partial_t - \Delta)^{-1}\zeta$. Note that if we formally periodize the time, then $\psi$ will have the form

$$\psi(t,x) = \sum_{k,\lambda} a_{k,\lambda} g_{k,\lambda}(\omega) e^{i(k\cdot x - |k|^2 t + \lambda t)},$$

where $\lambda$ is the modulation variable, $a_{k,\lambda}$ is some fixed function of $(k,\lambda)$ and $g_{k,\lambda}$ are i.i.d. Gaussian random variables. Therefore, in addition to the $k$ variables, we should include also the $\lambda$ variables as input variables for our random tensor, which will then look like $h_{kk_A,\lambda\lambda_A}$ for some sets $A$. The counting estimate should be adjusted, which may lead to changes in the selection algorithm.

9.2.2. *Other dispersion relations.* Similarly we may consider other dispersion relations, still in the semilinear setting. The main difference is again in the counting estimates: suppose the new dispersion relation is $\Lambda(k)$ for some function $\Lambda$, then we should look at the cardinality of sets

$$\{(k_1,\cdots,k_p) : k_1 - \cdots + k_p = k,\ \Lambda(k_1) - \cdots + \Lambda(k_p) = \Gamma + O(1)\}$$

with fixed $k$ and $\Gamma$, perhaps with some additional linear relations between $k_j$ like those in Section 4.3. Note that while *parabolic equations are all alike, each dispersive equation is dispersive in its own way.* As a result, the above counting bound will depend on the exact form of $\Lambda$ (not just its homogeneity), and a selection algorithm is then designed to match the counting bound. In particular we will not have a general black-box argument working for all $\Lambda$, and the proof has to be done in a dispersion-specific way.

9.2.3. *Quasilinear problems.* Recently there have been attempts to extend the existing parabolic theories to quasilinear equations [42, 43]. This is also of interest in the dispersive setting, especially in view of the recent results in low regularity deterministic local well-posedness [3, 57].

Of course, compared to parabolic equations, moving to the quasilinear (or even variable-coefficient semilinear) setting completely changes the methodology for dispersive equations. The $X^{s,b}$-based approaches become unavailable and dispersion has to be observed on the level of energy estimates or parametrices. In the deterministic setting, it is expected that the local well-posedness threshold is higher than $s_{cr}$, but the precise value is only known in some cases; in the random setting we also expect the threshold to be higher than $s_{pr}$, but are unable to decide or even guess the correct value. The method of random averaging operators can be applied to the quasilinear setting but may not achieve the same power as the semilinear version, and the quasilinear version of random tensor theory still needs to be explored.

**Open problem 2.** Build a random data theory for quasilinear dispersive (including wave) equations, and determine the threshold for almost-sure local well-posedness.

9.2.4. *Long-time propagation of randomness.* It is natural to ask whether the short-time solutions for (9.1) constructed in Theorem 1.1 can be extended to longer or infinite time; i.e. whether the randomness structure can be propagated beyond the perturbative regime. Such global-in-time extensions are immediate if an invariant Gibbs measure is available at the regularity we are considering, but as discussed in Section 9.1.3 this happens only in a few specific cases.

Note that the theory of random tensors, like the theory of regularity structures, is a short-time theory by nature; thus to get global results it has to be combined with separate global or large-scale techniques. In the context of (9.2), the work [62] combines the para-controlled calculus with energy estimates, and the more recent works [61, 22] combine (a reformulated version of) the regularity structures theory with the maximum principle. In the context of (9.1), the main global technique known is energy conservation, and the associated high-low and $I$-methods [13, 24, 26, 27]. Note that these require deterministic analysis at the $H^1$ (energy) level, so they need (deterministic) $H^1$ subcriticality, i.e. $s_{cr} < 1$, to work.

In the $H^1$ supercritical ($s_{cr} > 1$) case, another natural question is whether *classical* solutions with random initial data (such as (1.2) with $\alpha$ suitably large, as opposed to low regularity solutions of Theorem 1.1) are almost surely global. This is also important from the PDE point of view, as it would mean that blowup for defocusing $H^1$ supercritical nonlinear Schrödinger equations is *non-generic* and unstable[1]. Note that the blowup example in $\mathbb{R}^d$, recently constructed in [60], is indeed non-generic.

**Open problem 3.** In the energy subcritical case, do the singular solutions constructed in Theorem 1.1 extend to all time? In the energy supercritical case, does almost-sure *global* well-posedness hold for random initial data of high regularity?

## References


[1] M. Aizenman. Geometric analysis of $\Phi^4$ fields and Ising models. Part I and II. *Comm. Math. Phys.* 86 (1982), issue 1, 1–48.

[2] M. Aizenman and H. Copin. Marginal triviality of the scaling limits of critical 4D Ising and $\Phi_4^4$ models. *Ann. of Math.* (2) 194 (2021), no. 1, 163–235.

[3] T. Alazard, N. Burq, and C. Zuily. Strichartz estimates and the Cauchy problem for the gravity water waves equations. *Memoirs of the AMS* 256 (2014), no. 1229.

[4] S. Albeverio and S. Kusuoka. The invariant measure and the flow associated to the $\Phi_3^4$-quantum field model. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) Vol. XX (2020), 1359–1427.

[5] I. Bailleul and F. Bernicot. Heat semigroup and singular PDEs. *J. Funct. Anal.* 270 (2016), no. 9, 3344–3452.

[6] I. Bailleul and F. Bernicot. High order paracontrolled calculus. *Forum of Mathematics Sigma* 7, e44 (2019), 1–94.

[7] N. Barashkov and M. Gubinelli. A variational method for $\Phi_3^4$. *Duke Math. J.* 169 (2020), no. 17, 3339–3415.

[8] N. Barashkov and M. Gubinelli. The $\Phi_3^4$ measure via Girsanov's theorem. *Electron. J. Probab.* 26 (2021), 1–29.

[9] 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.* 6 (2019), 114–160.

[10] J. Bourgain. Periodic nonlinear Schrödinger equation and invariant measures. *Comm. Math. Phys.* 166 (1994), 1–26.

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

[12] J. Bourgain. Invariant measures for the Gross-Pitaevskii equation. *J. Math. Pures Appl.* 76 (1997), 649–702.

[13] J. Bourgain. Refinements of Strichartz inequality and applications to 2D–NLS with critical nonlinearity. *Int. Math. Res. Notices* (1998), no. 5, 253–283.

[14] B. Bringmann. Almost sure local well-posedness for a derivative nonlinear wave equation. *Int. Math. Res. Notices* (2021), no. 11, 8657–8697.

[15] Y. Bruned, A. Chandra, I. Chevyrev and M. Hairer. Renormalising SPDEs in regularity structures. *J. Eur. Math. Soc.* 23 (2020), no. 3, 869–947.

[16] Y. Bruned, M. Hairer and L. Zambotti. Algebraic renormalisation of regularity structures. *Invent. Math.* 215 (2019), no. 3, 1039–1156.

[17] 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), no. 7, 1407–1460.

[18] T. Buckmaster, P. Germain, Z. Hani and J. Shatah. Onset of the wave turbulence description of the longtime behavior of the nonlinear Schrödinger equation. *Invent. math.* 225 (2021), 787–855.


[1] For example, invariance of Gibbs measure and the associated almost-sure global well-posedness result would imply that there is no stable blowup mechanism at the regularity of the support of the Gibbs measure.

[19] N. Burq and N. Tzvetkov. Random data Cauchy theory for supercritical wave equations I: local theory. *Invent. Math.* 173 no. 3 (2008), 449–475.

[20] R. Catellier and K. Chouk. Paracontrolled distributions and the 3-dimensional stochastic quantization equation. *Ann. Prob.* 46 (2018), no. 5, 2621–2679.

[21] A. Chandra and M. Hairer. An analytic BPHZ theorem for regularity structures. arXiv:1612.08138.

[22] A. Chandra, A. Moinat and H. Weber. A priori bounds for the $\Phi^4$ equation in the full sub-critical regime. arXiv:1910.13854.

[23] A. Chandra and H. Weber. Stochastic PDEs, Regularity structures, and interacting particle systems.*Ann. Fac. Sci. Toulouse Math.* (6) 26 (2017), no. 4, 847–909.

[24] K. Cheung, G. Li and T. Oh. Almost conservation laws for stochastic nonlinear Schrödinger equations. *J. Evol. Equ.* 21 (2021), 1865–1894.

[25] K. Cheung and R. Mosincat. Stochastic nonlinear Schrödinger equations on tori. *Stoch. Partial Differ. Equ. Anal. Comput.* 7 (2019), no. 2, 169–208.

[26] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao. Almost Conservation Laws and Global Rough Solutions to a Nonlinear Schrödinger Equation. *Math. Res. Letters* 9 (2002), no. 5, 659-682.

[27] J. Colliander and T. Oh. Almost sure well-posedness of the cubic nonlinear Schrödinger equation below $L^2(\mathbb{T})$. *Duke Math. J.* 161 (2012), no. 3, 367–414.

[28] C. Collot and P. Germain. On the derivation of the homogeneous kinetic wave equation. arXiv:1912.10368.

[29] G. Da Prato and A. Debussche. Two-dimensional Navier-Stokes equations driven by a space-time white noise. *J. Funct. Anal.* 196 (2002), no. 1, 180–210.

[30] G. Da Prato and A. Debussche. Strong solutions to the stochastic quantization equations. *Ann. Probab.* 31 (2003), no. 4, 1900–1916.

[31] A. de Bouard and A. Debussche. The Stochastic Nonlinear Schrödinger Equation in $H^1$. *Stochastic Anal. Appl.* 21 (2003), no. 1, 97–126.

[32] Y. Deng. Two dimensional nonlinear Schrödinger equation with random radial data. *Anal. PDE* 5 (2012), no. 5, 913–960.

[33] Y. Deng and Z. Hani. On the derivation of the wave kinetic equation for NLS. *Forum Math. Pi*, 9 (2021), E6, 1–37.

[34] Y. Deng and Z. Hani. Full derivation of the wave kinetic equation. arXiv:2104.11204.

[35] Y. Deng, Andrea R. Nahmod and H. Yue. Optimal local well-posedness for the periodic derivative nonlinear Schrödinger equation. *Comm. Math. Phys.* 384 (2021), 1061–1107.

[36] Y. Deng, A. Nahmod and H. Yue. Invariant Gibbs measures and global strong solutions for nonlinear Schrödinger equations in dimension two. arXiv:1910.08492.

[37] C. Fan and W. Xu. Global well-posedness for the defocusing mass-critical stochastic nonlinear Schrödinger equation on $\mathbb{R}$ at $L^2$ regularity. arXiv:1810.07925.

[38] E. Faou, P. Germain, and Z. Hani. The weakly nonlinear large-box limit of the 2D cubic nonlinear Schröinger equation. *J. Amer. Math. Soc.* 29 (2016), no. 4, 915–982.

[39] J. Forlano, T. Oh, and Y. Wang. Stochastic nonlinear Schrödinger equation with almost space-time white noise. *J. Aust. Math. Soc.* (2018), 1–24.

[40] P, Friz and M. Hairer. A course on rough paths. With an introduction to regularity structures. Second ed., Universitext. Springer, Cham, 2020. xvi+346 pp.

[41] J. Fröhlich. On the triviality of $\lambda\Phi_d^4$ theories and the approach to the critical point in $d_{(-)} > 4$ dimensions. *Nuclear Physics B* 200 (1982), issue 2, 281–296.

[42] M. Furlan and M. Gubinelli. Paracontrolled quasilinear SPDEs. *Ann. Probab.* 47 (2019), no. 2, 1096–1135.

[43] M. Gerencsér and M. Hairer. A solution theory for quasilinear singular SPDEs. *Comm. Pure Appl. Math.* 72 (2019), no. 9, 1983–2005.

[44] J. Glimm and A. Jaffe. *Quantum physics, A functional integral point of view*, Second edition, Springer-Verlag, New York, 1987. xxii+535 pp.

[45] A. Grünrock and S. Herr. Low regularity local well-posedness of the derivative nonlinear Schrödinger equation with periodic initial data. *SIAM J. Math. Anal.* 39 (2008), no. 6, 1890–1920.

[46] M. Gubinelli, P. Imkeller and N. Perkowski. Paracontrolled distributions and singular PDEs. *Forum Math Pi* 3 (2015), e6, 75 pp.

[47] M. Gubinelli, H. Koch and T. Oh. Paracontrolled approach to the three-dimensional stochastic nonlinear wave equation with quadratic nonlinearity. arXiv:1811.07808.

[48] M. Gubinelli and N. Perkowski. Lectures on singular stochastic PDEs. *Ensaios Matemáticos, Mathematical Surveys*, 29. Sociedade Brasileira de Matemática, Rio de Janeiro. (2015), 89 pp.

[49] M. Gubinelli and N. Perkowski. KPZ reloaded. *Comm. Math. Phys.* 349 (2017), no. 1, 165–269.

[50] M. Gubinelli and N. Perkowski. An introduction to singular SPDEs. in *Stochastic partial differential equations and related fields*, 69–99, *Springer Proc. Math. Stat.*, 229, Springer, Cham, 2018.

[51] Z. Guo and T. Oh. Non-Existence of Solutions for the Periodic Cubic NLS below $L^2$. *Int. Math. Res. Notices* (2018), no. 6, 1656–1729.

[52] M. Hairer. A theory of regularity structures. *Invent. Math.* 198 (2014), no. 2, 269–504.

[53] M. Hairer. Singular Stochastic PDE. *Proceedings of the ICM-Seoul*, Vol. I, (2014), 685–709.

[54] M. Hairer. Introduction to regularity structures. *Braz. J. Probab. Stat.* 29 (2015), no. 2, 175–210.

[55] M. Hairer. Regularity structures and the dynamical $\Phi_3^4$ model. *Current Developments in Mathematics 2014*, Int. Press, Somerville, MA, (2016), 1–49.

[56] M. Hairer, M. D. Ryser and H. Weber. Triviality of the 2D stochastic Allen-Cahn equation. *Electron. J. Probab* 17 (2012), no. 39, p. 1–14.

[57] S. Klainerman, I. Rodnianski and J. Szeftel. The bounded $L^2$ curvature conjecture. *Invent. Math.* 202 (2015) no. 1, 91–216.

[58] A. Kupiainen. Renormalization group and stochastic PDEs. *Ann. Henri Poincaré* 17 (2016), no. 3, 497–535.

[59] J. Lebowitz, R. Rose and E. Speer. Statistical mechanics of the nonlinear Schrödinger equation. *J. Statist. Phys.* 50 (1988), 657–687.

[60] F. Merle, P. Raphael, I. Rodnianski and J. Szeftel. On blow up for the energy super critical defocusing non linear Schrödinger equations. *Invent. Math.* (2021). https://doi.org/10.1007/s00222-021-01067-9.

[61] A. Moinat and H. Weber. Space-time localisation for the dynamic $\Phi_3^4$ model. *Comm. Pure Appl. Math.* 73 (2020), no. 12, 2519–2555.

[62] J.C. Mourrat and H. Weber. The dynamic $\Phi_3^4$ model comes down from infinity. *Comm. Math. Phys.* 356 (2017), no. 3 673–753.

[63] E. Nelson. Construction of quantum fields from Markoff fields, *J. Functional Analysis* 12 (1973), 97–112.

[64] T. Oh, N. Tzvetkov and Y. Wang. Solving the 4NLS with white noise initial data. *Forum Math. Sigma*, 8 (2020), E48.

[65] G. Parisi and Y. S. Wu. Perturbation theory without gauge fixing. *Scientia Sinica. Zhongguo Kexue* 24 (1981), no.4, 483–496.

[66] B. Simon. *The $P(\varphi)_2$ Euclidean (quantum) field theory*, Princeton Series in Physics. Princeton University Press, Princeton, N.J., 1974. xx+392 pp.

[67] C. Sun and N. Tzvetkov. Gibbs measure dynamics for the fractional NLS. *SIAM J. Math. Anal.* 52 (2020), no. 5, 4638–4704.

[1] Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA
*E-mail address*: yudeng@usc.edu

[2] Department of Mathematics, University of Massachusetts, Amherst MA 01003
*E-mail address*: nahmod@math.umass.edu

[3] Institute of Mathematical Sciences, ShanghaiTech University, Shanghai, 201210, China
*E-mail address*: yuehaitian@shanghaitech.edu.cn