I. DERIVATION OF THE KINETIC EQ.(4)

A. Primary asymptotics

The starting point is the NLS Eq.(1) (main text) written in the mode basis, i.e., Eq.(2) (main text). We consider the regime where linear propagation dominates over disorder, which in turn dominates over the nonlinearity. Accordingly, we introduce a small dimensionless parameter $\varepsilon$ and we consider the regime $\beta_1 \to \beta_j, C \to \varepsilon C, \gamma \to \varepsilon^2 \gamma$. For propagation distances of order $\varepsilon^{-2}$, the rescaled mode amplitudes $\tilde{a}_j^\varepsilon(z) = a_j(z/\varepsilon^2)$ satisfy

$$\partial_z \tilde{a}_j^\varepsilon = -i\beta_j \tilde{a}_j^\varepsilon + i\gamma \sum_{l,m=0}^{M-1} Q_{jlmn} \tilde{c}_l \overline{c}_m \overline{c}_n - i \varepsilon \sum_{l=0}^{M-1} C_{jl}(z) \tilde{a}_l^\varepsilon,$$

where the bar stands for complex conjugation. We set $c_j^\varepsilon(z) = a_j^\varepsilon(z) \exp \left(i \frac{\beta_j}{\varepsilon^2} z \right)$. The amplitudes $c_j^\varepsilon(z)$ satisfy:

$$\partial_z c_j^\varepsilon = i\gamma \sum_{l,m=0}^{M-1} Q_{jlmn} \tilde{c}_l \overline{c}_m \overline{c}_n \exp \left(i \frac{\beta_j - \beta_l - \beta_m + \beta_n}{\varepsilon^2} z \right) - i \varepsilon \sum_{l=0}^{M-1} C_{jl}(z) c_l^\varepsilon \exp \left(i \frac{\beta_j - \beta_l}{\varepsilon^2} z \right).$$

(S1)

This is the usual diffusion approximation framework [1]. We get the following result.

Proposition 1.1 The random process $(c_j^\varepsilon(z))_{j=0}^{M-1}$ converges in distribution in $C^0([0,\infty),\mathbb{C}^M)$, the space of continuous functions from $[0,\infty)$ to $\mathbb{C}^M$, to the Markov process $(\epsilon_j(z))_{j=0}^{M-1}$ with infinitesimal generator $L$:

$$L = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4 + \mathcal{L}_5,$$

with

$$\mathcal{L}_1 = \frac{1}{2} \sum_{j,l=0,\neq j}^{M-1} \Gamma_{jl}^D (c_j \partial_{\epsilon_l} + \overline{c}_l \partial_{\epsilon_j}) + \Gamma_{jl}^D (c_j \overline{c}_l \partial_{\epsilon_j}),$$

$$\mathcal{L}_2 = \frac{1}{2} \sum_{j=0}^{M-1} \Gamma_{j}^D (c_j \partial_{\epsilon_j} + \overline{c}_j \partial_{\epsilon_j}) + \Gamma_{j}^D (c_j \overline{c}_j \partial_{\epsilon_j}),$$

$$\mathcal{L}_3 = \frac{1}{2} \sum_{j=0}^{M-1} \Gamma_{jl}^D (c_j \partial_{\epsilon_l} + \overline{c}_l \partial_{\epsilon_j}) + \Gamma_{jl}^D (c_j \overline{c}_l \partial_{\epsilon_j}),$$

$$\mathcal{L}_4 = - \sum_{j=0}^{M-1} \Gamma_{j}^D (c_j \partial_{\epsilon_j} + \overline{c}_j \partial_{\epsilon_j}),$$

$$\mathcal{L}_5 = -i\gamma \sum_{l,m,n=0}^{M-1} \delta_{jlmn} Q_{jlmn} (c_l \partial_{\epsilon_l} - \overline{c}_l \partial_{\epsilon_l}),$$

where $\delta_{jlmn} = 1_{\beta_j - \beta_l - \beta_m + \beta_n = 0}$. In this definition we use the classical complex derivative: if $\zeta = \zeta_r + i\zeta_i$, then $\partial_\zeta = (1/2)(\partial_\zeta - i\partial_{\zeta_i})$ and $\partial_{\zeta_i} = (1/2)(\partial_\zeta + i\partial_{\zeta_i})$, and the coefficients of the operator $\mathcal{L}_k (k = 1,\ldots,5)$ are defined for $j,l = 0,\ldots, M - 1$, as follows:

- For all $j \neq l$, $\Gamma_{jl}$ and $\Gamma_{jl}^D$ are given by

$$\Gamma_{jl}^D = \frac{1}{2} \int_0^\infty \mathcal{R}_{jl}(z) \cos ((\beta_l - \beta_j)z) dz, \quad (S3)$$

$$\hat{\Gamma}_{jl}^D = \frac{1}{2} \int_0^\infty \mathcal{R}_{jl}(z) \sin ((\beta_l - \beta_j)z) dz, \quad (S4)$$

with $\mathcal{R}_{jl}(z)$ defined by

$$\mathcal{R}_{jl}(z) = E[C_{jl}(0)C_{jl}(z)]$$

$$= \int u_j(r)u_j(r')E[\delta V(0,r)\delta V(z,r')]|u_j(r)|^2dr dr'.$$

(S5)

- For all $j,l = 0,\ldots, M - 1$:

$$\Gamma_{jl}^D = \int_0^\infty E[C_{jl}(0)C_{jl}(z)] dz + \int_0^\infty E[C_{jl}(0)C_{jl}(z)] dz. \quad (S6)$$

- For all $j = 0,\ldots, M - 1$:

$$\Gamma_{jj}^D = - \sum_{l=0,\neq j}^{M-1} \Gamma_{jl}^D, \quad \hat{\Gamma}_{jj}^D = - \sum_{l=0,\neq j}^{M-1} \hat{\Gamma}_{jl}^D. \quad (S7)$$

B. Secondary asymptotics

We observe that $\Gamma_{OD}$ and $\hat{\Gamma}_{OD}$ depend on the power spectral density of the random index perturbation evaluated at the difference of distinct frequencies $\beta_j - \beta_l$, while $\Gamma_{D}$ depends on the power spectral density of the index perturbation evaluated at zero-frequency. Therefore, when $L_{lin} =$
$1/\beta_0 \ll \epsilon_c$, then $\Gamma^D$ is larger than $\Gamma^{OD}, \bar{\Gamma}^{OD}$. We consider this regime by introducing a small dimensionless parameter $\eta$ with $\Gamma^D \to \Gamma^D, \Gamma^{OD} \to \eta \bar{\Gamma}^{OD}, \bar{\Gamma}^{OD} \to \eta \bar{\Gamma}^{OD}, \gamma \to \eta \gamma$.

For propagation distances of order $\eta^{-2}$, we introduce the rescaled mode amplitudes $c_j^\eta(z) = c_j(z/\eta^2)$. By Proposition I.1 it is a Markov process with infinitesimal generator $L^\eta$:

$$L^\eta = \mathcal{L}_1 + \eta^{-2} \mathcal{L}_2 + \mathcal{L}_3 + \eta^{-2} \mathcal{L}_4 + \eta^{-1} \mathcal{L}_5,$$  \hspace{1cm} (S8)

where the operators $\mathcal{L}_k (k = 1, \ldots, 5)$ are given above. By (S8) the second-order moments for $j \neq j'$:

$$\partial_z E[c_j^\eta c_j^\eta] = -\frac{1}{2\eta^2} (\Gamma^D_j + \Gamma^{OD}_j - 2\Gamma^D_{jj'}) E[c_j^\eta c_j^\eta]$$
$$+ \frac{1}{2} (\Gamma^D_j + \Gamma^{OD}_j) E[c_j^\eta c_j^\eta] + \frac{i}{2} (\Gamma^D_j - \Gamma^{OD}_j) E[c_j^\eta c_j^\eta]$$
$$+ \frac{i}{\eta} \sum_{l,m,n=0}^{M-1} \delta^K_{lmn} Q_{jmn} \text{Re} E[c_j^\eta c_j^\eta c_j^\eta c_j^\eta],$$  \hspace{1cm} (S9)

up to negligible terms in $\eta$. Note that $\Gamma^D_j + \Gamma^{OD}_j - 2\Gamma^D_{jj'} = \int_{-\infty}^{\infty} E[(C_{jj'}(0) - C_{j'}(0))(C_{jj}(z) - C_{j'}(z))\text{dz}$ is positive (it is the power spectral density evaluated at frequency $\omega_j$ of the stationary process $C_{jj'}(z)$ for $j \neq j'$) by Bochner’s theorem. Therefore $E[c_j^\eta c_j^\eta]$ is exponentially damped and

$$E[c_j^\eta c_j^\eta] = O(\eta).$$  \hspace{1cm} (S9)

If $j = j'$, then the mean square amplitudes $w_j(\eta) = E[c_j^\eta(\eta)]$ satisfy

$$\partial_z w_j = \sum_{l=0}^{M-1} \Gamma^D_j (w_j - w_j)$$
$$- 2\frac{i}{\eta} \sum_{l,m,n=0}^{M-1} \delta^K_{lmn} Q_{jmn} \text{Re} \{E[c_j^\eta c_j^\eta c_j^\eta c_j^\eta]\}.$$  \hspace{1cm} (S10)

By (S8) the fourth-order moments satisfy

$$\partial_z E[c_j^\eta c_j^\eta c_j^\eta c_j^\eta] = -\frac{1}{2\eta^2} \Gamma^D_{jmn} E[c_j^\eta c_j^\eta c_j^\eta c_j^\eta] + i \frac{2}{\eta} \eta \gamma \sum_{j,j',m',n'} M_{jmn,j'j'm'n'} E[c_j^\eta c_j^\eta c_j^\eta c_j^\eta],$$  \hspace{1cm} (S11)

up to negligible terms in $\eta$. The coefficients $\Gamma^D_{jmn}$ and the sixth-order moment $Y_{jmn}$ are given by

$$\Gamma^D_{jmn} = \Gamma^D_j + \Gamma^{OD}_{mn} + \Gamma^D_{nn} + \Gamma^D_{jj} + 2\Gamma^D_{jm} - 2\Gamma^D_{jn},$$

and

$$Y_{jmn} = \sum_{j',m',n'=0}^{M-1} \delta^K_{nmn'} S_{j'j'm'n'} E[c_j^\eta c_j^\eta c_j^\eta c_j^\eta c_j^\eta c_j^\eta],$$  \hspace{1cm} (S12)

and

$$\partial_z w_j^\eta = Tj(\beta_j - \mu),$$  \hspace{1cm} (S13)

where $1/T$ and $-\mu/T$ are the Lagrange multipliers associated to the conservation of $E$ and $N$. There is a one to one relation between the pair $(N, E)$ and $(T, \mu)$: The values of the conserved quantities $(N, E)$ determine uniquely $(T, \mu)$, and thus the RJJ equilibrium distribution that maximizes the entropy $S[w_j]$ under the constraints that $N$ and $E$ are conserved, reads

$$w_j^\eta = Tj(\beta_j - \mu),$$  \hspace{1cm} (S14)

where $\beta_j$ and $\mu$ are the Lagrange multipliers associated to the conservation of $E$ and $N$. There is a one to one relation between the pair $(N, E)$ and $(T, \mu)$: The values of the conserved quantities $(N, E)$ determine uniquely $(T, \mu)$, and thus the RJJ equilibrium distribution that maximizes the entropy $S[w_j]$ under the constraints that $N$ and $E$ are conserved, reads

$$w_j^\eta = Tj(\beta_j - \mu).$$  \hspace{1cm} (S15)

C. Degenerate modes

In this section we assume that the modes may be degenerate. The detailed derivation of the kinetic equation accounting for mode degeneracy is cumbersome and will be reported elsewhere. Here we report the main results.

There are $G$ distinct wavenumbers:

$$\{\beta^{(g)}, g = 1, \ldots, G\},$$

and the mode indices can be partitioned into $G$ groups $G^{(g)}$, $g = 1, \ldots, G$:

$$G^{(g)} = \{p = 1, \ldots, N, \beta_p = \beta^{(g)}\}.$$  

We obtain the kinetic equation

$$\partial_z w^{(g)} = 8\gamma^2 \sum_{g_1, g_2, g_3=1}^{G} \delta^{(g_1)\beta^{(g_2)}\eta^{(g_3)}} \omega^{(g_1)\omega^{(g_2)\omega^{(g_3)}}} \left(\omega^{(g_1)\omega^{(g_2)\omega^{(g_3)}}} - \omega^{(g_1)\omega^{(g_2)\omega^{(g_3)}}}\right),$$  \hspace{1cm} (S16)
where
\[ q^{(g_9 9g_9 9g_9)} = \frac{1}{|G|} \sum_{jeG(s),lG(g_9),nmG(g_9),nG(g_9)} Q_{jlmn} Q^{(g_9 9g_9 9g_9)}_{jlmn} \]
where
\[ Q^{(g_9 9g_9 9g_9)}_{jlmn} = \left( Q^{(g_9 9g_9 9g_9)}_{jlmn} \right)_{jeG(s),lG(g_9),nmG(g_9),nG(g_9)} = (M^{(g_9 9g_9 9g_9)})^{-1} \left( (Q_{jlmn})_{jeG(s),lG(g_9),nmG(g_9),nG(g_9)} \right). \]

The tensor \( M^{(g_9 9g_9 9g_9)} \) (seen as a \( q \times q \) matrix with \( q = |G(s)| |G(g_9)| |G(g_9)| |G(g_9)| \)) is given by
\[
\sum_{j'G(s),l'G(g_9),m'G(g_9),n'G(g_9)} M^{(g_9 9g_9 9g_9)}_{j'lmn} j'lmn' = 2n^{(j'lmn)j} j'mn' + \sum_{n'G(g_9),j'G(s)} 2\eta_{n} j' j'mn' + \sum_{l'G(g_9),nG(g_9)} 2\eta_{l} n' j'mn' + \sum_{m'G(g_9),nG(g_9)} 2\eta_{m} n' j'mn' + \sum_{n',n'G(g_9)} 2\eta_{n} n' j'mn' + \sum_{j',j'G(g_9)} 2\eta_{j} n' j'mn'.
\]

\[ \gamma_{pq} = 2 \int_{-\infty}^{\infty} \mathbb{E}[C_{pq}(z) C_{pq}(0)] e^{(\beta_p - \beta_q)z} \, dz. \]

D. Numerical simulations

Implementation of disorder: To implement the disorder in the simulations of the NLS Eq. (2) (main text), we considered an exact discretization of the Ornstein-Uhlenbeck process. The propagation axis is divided in intervals with deterministic length \( \Delta z \), with \( \Delta z < \eta_{z} \). The random function \( \mu(z) \) is stepwise constant over each elementary interval \( z \in [k \Delta z, (k + 1) \Delta z] \), where \( \mu_0 \sim N(0, \sigma^2/2) \) denotes the Gaussian distribution, \( \mu_k = \sqrt{1 - 2\Delta z/\ell_{z} n_{\eta} - 1 + \sqrt{2\Delta z/\ell_{z} N(0, \sigma^2/2)}, \) with \( N(0, \sigma^2/2) \) \( \alpha \) \( \beta \) \( \gamma \) \( \delta \) \( \epsilon \) \( \zeta \) \( \eta \) \( \theta \) \( \iota \) \( \kappa \) \( \lambda \) \( \mu \) \( \nu \) \( \xi \) \( \omicron \) \( \pi \) \( \rho \) \( \sigma \) \( \tau \) \( \upsilon \) \( \phi \) \( \chi \) \( \psi \) \( \omega \) \( \Gamma \) \( \Delta \) \( \Theta \) \( \Lambda \) \( \Xi \) \( \Pi \) \( \Sigma \) \( \Upsilon \) \( \Phi \) \( \Psi \) \( \Omega \)
The energy $E$ provides the potential energy difference less than 1%. The conservation of energy at $N$ of injection of the speckle beam into the MMF: We measured the power has been verified by keeping fixed the conditions of propagation without strong disorder: The conservation of the energy $\frac{E}{N}$ for a large strength of random mode coupling corresponding to an increase of the energy due to disorder of $\Delta \frac{E}{N} \approx 19\%$. The black solid line reports the condensate fraction from the RJ theory, $w_0^{RJ}/N$ vs $E/N$. In the absence of strong disorder (squares): $w_0/N$ increases as the power increases, and reaches the value predicted by the RJ theory (solid line) — each color refers to a different value of the energy $E/N$. In the presence of strong disorder (big circles): the energy $E/N$ increases (the squares are shifted to the big circles of the same color). The big circles report the average over 10 different realizations of disorder (10 small circles for each color). At variance with Fig. 4, here the strength of random mode coupling is so large that RJ thermalization and condensation are inhibited by strong disorder.

Without altering the fiber launch conditions, the fiber is cut to 20cm to get $E_{in}$. The procedure is repeated for different speckle beams (i.e., for different values of the energy $E$), by moving the diffuser before injection into the MMF. We always obtained $|E_{out} - E_{in}|/E_{noy} < 1\%$ for values of the energy that span the range of the condensation curve, i.e. $w_0^{RJ}/N$ varying from 0 to 0.7.

3) Experimental observation of RJ thermalization: In the absence of strong disorder (i.e., absence of applied stress induced on the fiber), we observe the process of thermalization to the RJ equilibrium distribution, $w_j^{RJ} = T/(\beta_j - \mu)$. In the experiments, the modal populations ($w_j$) are computed by using the Gerchberg-Saxton algorithm, which allows us to retrieve the transverse phase profile of the field from the NF and the FF intensity distributions measured in the experiments [5]. By projecting the complex field over the modal populations of the MMF (Gauss-Hermite basis) we get the RJ thermalization curve $w_j^{RJ}/N$, $j = 0, 1, ..., M-1$. A typical example is reported in Fig. S1 showing the modal distribution $w_j/N$ recorded experimentally at low-power (linear regime) and high-power (nonlinear regime), and its comparison to the RJ equilibrium distribution. Fig. S1(a) reports a single realization of the speckle beam, Fig. S1(b) reports an average over 60 realizations of speckle beams. The quantitative agreement between the experimental results and the theoretical RJ distribution is obtained without using adjustable parameters.
We verify that $E/N$ regime), and compute FF intensity patterns at high power ($N$)

We return to the previous higher power ($N = 7kW$, non-linear regime) and we verify that we recover the same NF speckle beam as in step i).

Then we apply stress to a specific location of the MMF. The stress is applied by using clamps mounted on a linear translation manual stage whose position is controlled at the micrometer scale. We adjust the amount of stress by measuring the power losses (10% in Fig. 4(a), corresponding to $\Delta E/N \sim 6\%$). Once the stress is adjusted, the power is increased up to the same average power of step i). We then measure the NF and FF intensity patterns and compute $E/N$ and $w_0/N$ (small circles in Fig. 4(a)).

In a next step we decrease the power ($N = 0.23kW$, linear regime), we measure the NF and FF intensity patterns and compute $E/N$ and $w_0/N$ (small circles in Fig. 4(a)).

We return to the previous higher power ($N = 7kW$, non-linear regime) and we remove the applied stress. We verify that we recover the same initial NF speckle beam as in step i).

We repeat the steps iv)-v)-vi) 10 times to get 10 different realizations of strong disorder (small circles). Each disorder realization is achieved by applying stress to a different position of the MMF by rotating the drum on which it is wound.

The procedure i)-vii) is repeated for a larger amount of applied stress (disorder), corresponding to an increase of energy due to disorder of $\Delta E/N \approx 11\%$ in Fig. 4(b) (20% of power losses), and $\Delta E/N \approx 19\%$ in Fig. S2 (40% of power losses). In Fig. S2 the strength of random mode coupling is so large that RJ thermalization and condensation are inhibited by strong disorder.

Note that losses induced by strong disorder only weakly affect the condensate fraction through the propagation in the MMF, as illustrated in the simulation reported in Fig. S3. We have considered 10% of losses (over the propagation length $\varepsilon \beta_0 = 11 \times 10^5$), in the case where losses are distributed homogeneously in mode space, and non-homogeneously in mode space (only the higher-order modes experience losses). We have considered the parameters of the simulation reported in Fig. 2 (main text), which refers to the most interesting regime where linear disorder effects and nonlinear effects are of the same order, $E_{kin}^L \lesssim E_{kin}^N$. The condensate peak relevant to the experiments is only weakly affected by the losses, see the inset in Fig. S3(a). Note that, for larger propagation lengths, the losses concentrated on the higher-order modes reduce the effective number of modes and thus limit the increase of energy $E/(N\beta_0)$ due to disorder (Fig. S3(b)), which in turn leads to an increase of the condensate fraction (Fig. S3(a)).

---

1. J.-P. Fouque, J. Garnier, G. Papanicolaou, and K. Sølna, Wave Propagation and Time Reversal in Randomly Layered Media (Springer, 2007).
2. V.E. Zakharov, V.S. L’vov, G. Falkovich, Kolmogorov Spectra of Turbulence I (Springer, Berlin, 1992).
3. I.S. Gradshtein and I.M. Ryzhik, Tables of Integrals, Sums, Series, and Products (Academic Press, New York, 1980).
4. K. Baudin et al., Phys. Rev. Lett. 125, 244101 (2020).
5. J.R. Fienup, Applied Optics 21, 2758 (1982).