Capacity-achieving Spatially Coupled Sparse Superposition Codes with AMP Decoding

Cynthia Rush* Kuan Hsieh† Ramji Venkataramanan‡

May 11, 2021

Abstract

Sparse superposition codes, also referred to as sparse regression codes (SPARCs), are a class of codes for efficient communication over the AWGN channel at rates approaching the channel capacity. In a standard SPARC, codewords are sparse linear combinations of columns of an i.i.d. Gaussian design matrix, while in a spatially coupled SPARC the design matrix has a block-wise structure, where the variance of the Gaussian entries can be varied across blocks. A well-designed spatial coupling structure can significantly enhance the error performance of iterative decoding algorithms such as Approximate Message Passing (AMP).

In this paper, we obtain a non-asymptotic bound on the probability of error of spatially coupled SPARCs with AMP decoding. Applying this bound to a simple band-diagonal design matrix, we prove that spatially coupled SPARCs with AMP decoding achieve the capacity of the AWGN channel. The bound also highlights how the decay of error probability depends on each design parameter of the spatially coupled SPARC.

An attractive feature of AMP decoding is that its asymptotic mean squared error (MSE) can be predicted via a deterministic recursion called state evolution. Our result provides the first proof that the MSE concentrates on the state evolution prediction for spatially coupled designs. Combined with the state evolution prediction, this result implies that spatially coupled SPARCs with the proposed band-diagonal design are capacity-achieving. Using the proof technique used to establish the main result, we also obtain a concentration inequality for the MSE of AMP applied to compressed sensing with spatially coupled design matrices. Finally, we provide numerical simulation results that demonstrate the finite length error performance of spatially coupled SPARCs. The performance is compared with coded modulation schemes that use LDPC codes from the DVB-S2 standard.

Contents

*Department of Statistics, Columbia University, New York, NY 10027, USA. Email: cynthia.rush@columbia.edu

†Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK. Email: kh525@cam.ac.uk

‡Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK. Email: rv285@cam.ac.uk This work was supported in part by an EPSRC Doctoral Training Award and a Turing Fellowship from the Alan Turing Institute, the National Science Foundation (NSF CCF #1849883), the Simons Institute for the Theory of Computing, and NTT Research. This paper was presented in part at the 2018 IEEE International Symposium on Information Theory and at the 2018 IEEE Information Theory Workshop.

1 Introduction

We consider communication over the memoryless additive white Gaussian noise (AWGN) channel, where the output symbol y is generated from input symbol u as $y = u + w$. The noise w is Gaussian with zero mean and variance σ^2. The input has an average power constraint P: for a codeword $\boldsymbol{x} = x_1, x_2, \ldots, x_n$ transmitted over n uses of the channel,

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P. \tag{1}$$

The Shannon capacity of the channel is $\mathcal{C} = \frac{1}{2}\ln\left(1 + \frac{P}{\sigma^2}\right)$ nats/transmission.

Sparse superposition codes, or sparse regression codes (SPARCs), were introduced by Joseph and Barron [1,2] for efficient communication over the AWGN channel. These codes have been proven to be reliable at rates approaching $\mathcal{C}$ with various low complexity iterative decoders [2–4]. As shown in Fig. 1, a SPARC is defined by a design matrix $\boldsymbol{A}$ of dimension $n \times ML$, where n is the code length and M, L are integers such that $\boldsymbol{A}$ has L sections with M columns each. Codewords are generated as linear combinations of L columns of $\boldsymbol{A}$, with one column from each section. Thus a codeword can be represented as $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$, with $\boldsymbol{\beta}$ being an $ML \times 1$ *message vector* with exactly one non-zero entry in each of its L sections. The message is indexed by the locations of the non-zero entries in $\boldsymbol{\beta}$. The values of the non-zero entries are fixed a priori.

Since there are M choices for the location of the non-zero entry in each of the L sections, there are M^L codewords. To achieve a communication rate of R nats/transmission, we therefore require

$$M^L = e^{nR} \quad \text{or} \quad nR = L\ln M. \tag{2}$$

The decoding task is to recover the message vector $\boldsymbol{\beta}$ from the received sequence $\boldsymbol{y} \in \mathbb{R}^n$ given by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}. \tag{3}$$

In the standard SPARC construction introduced in [1,2], the design matrix $\boldsymbol{A}$ is constructed with i.i.d. standard Gaussian entries. The values of the non-zero coefficients in the message vector $\boldsymbol{\beta}$ then define a *power allocation* across sections. With an appropriately chosen power allocation (e.g., one that is exponentially decaying across sections), the feasible decoders proposed in [2–4] have been shown to be asymptotically capacity-achieving. The choice of power allocation has also been shown to be crucial for obtaining good finite length performance with the standard SPARC construction [5]. A detailed discussion of the error rates of optimal versus feasible decoders for standard SPARCs can be found in [6].

In a *spatially coupled SPARC*, the design matrix $\boldsymbol{A}$ is composed of blocks with different variances, i.e., the entries of $\boldsymbol{A}$ are still independent and Gaussian but not identically distributed. Spatially coupled SPARCs were introduced by Barbier and co-authors in [7–10]. In these works, an approximate message passing (AMP) algorithm was used for decoding, whose performance can be predicted via a deterministic recursion called *state evolution*. Empirical results indicate that spatially coupled SPARCs can have better error performance than power allocated SPARCs at finite code lengths (see, e.g., [9,11]). Moreover, both standard SPARCs and power allocated SPARCs can be viewed as special cases of spatially coupled SPARCs (see Section 2). It is therefore of interest to rigorously characterize the achievable rates and the decay of error probability for spatially coupled SPARCs.

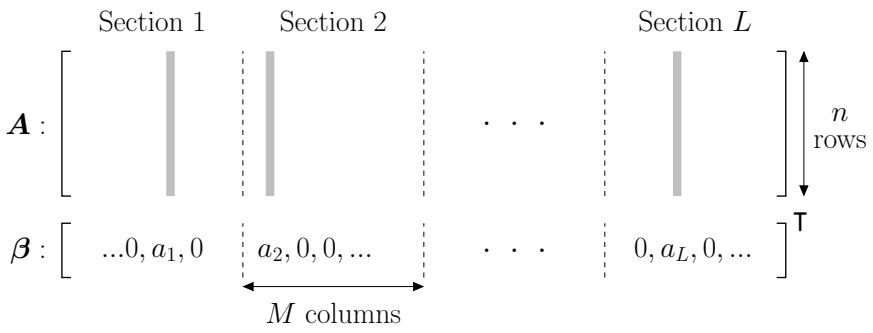

Figure 1: $\boldsymbol{A}$ is an $n \times ML$ design matrix and $\boldsymbol{\beta}$ is an $ML \times 1$ message vector with one non-zero entry in each of its L sections. Codewords are of the form $\boldsymbol{A\beta}$. The non-zero values $a_1, \ldots, a_L$ are fixed a priori.

Two key steps are required to prove that spatially coupled SPARCs achieve vanishingly small error probability with AMP decoding for rates $R < \mathcal{C}$:

1. Prove that in a suitable limit (corresponding to increasing code length), the fixed point of the state evolution recursion is one that corresponds to vanishing probability of decoding error;

2. Prove that the error rate of the AMP decoder is closely tracked by the state evolution parameters for sufficiently large code length.

The first step was proved by Barbier et al. [8], using the potential function method [12]. The result in [8] shows 'threshold saturation' for a class of spatially coupled SPARCs with AMP decoding. For any fixed rate $R < \mathcal{C}$, this implies that state evolution predicts vanishing probability of decoding error in the large system limit. (Throughout, the terminology 'large system limit' or 'asymptotic limit' refers to (L, M, n) all tending to infinity with $L \ln M = nR$.)

In this paper, we rigorously prove the second step. We also provide an alternative proof of the first step which gives insight into how the parameters defining the spatially coupled matrix influence the decoding progression. These two steps together yield the main result (Theorem 1), which is a non-asymptotic bound on the probability of decoding error of the AMP decoder. To our knowledge, this is the first complete proof that spatially coupled SPARCs are capacity-achieving with efficient decoding.

Related work: Approximate message passing (AMP) refers to a class of iterative algorithms for statistical estimation in models corresponding to dense factor graphs. AMP algorithms, which are obtained via Gaussian and quadratic approximations of standard message passing techniques like belief propagation, have been successfully applied to compressed sensing [13–16] and its applications in communications such as MIMO detection [17] and unsourced random access [18]. Other applications of AMP include estimation in generalized linear models [19,20], robust estimation [21,22], and several variants of low-rank matrix estimation [23–28].

The idea of spatial coupling was introduced in the context of LDPC codes [29–34], and first used for compressed sensing in [35]. AMP algorithms for compressed sensing with spatially coupled matrices were first proposed by Krzakala et al. in [15,36]. Takeuchi et al. [37,38] proposed a method for analyzing the state evolution recursion of spatially coupled systems using a potential function defined for the uncoupled system. This method characterizes the fixed points of the spatially coupled state evolution in terms of the stationary points of the potential function. The potential function method was made rigorous by Yedla et al. in [12]. In [39], Donoho et al. proved that

a class of spatially coupled Gaussian designs achieve the optimal information-theoretic limit for compressed sensing by analyzing the spatially coupled state evolution recursion in the continuum limit. In a complementary work [40], Javanmard and Montanari proved that the mean-squared error of AMP algorithm for spatially coupled compressed sensing converges (almost surely) to the state evolution prediction in each iteration.

Though the SPARC model is similar to compressed sensing, the result of [40] cannot be directly applied to AMP decoder since the SPARC message vector has a section-wise i.i.d. structure, with a growing section size M in the large system limit. Moreover, our result is non-asymptotic and generalizes the proof technique developed for power-allocated SPARCs [41] to the present setting of spatially coupled designs.

1.1 Structure of the paper and main contributions

After describing the construction of spatially coupled SPARCs (SC-SPARCs) in Section 2, we explain the AMP decoder and the associated state evolution recursion in Section 3. In Section 4, we obtain upper and lower bounds on the state evolution parameters (Lemma 4.1) which help explain the decoding progression of the AMP decoder for large M (Proposition 4.1). The main theoretical results of the paper are stated in Section 5:

- Theorem 1 gives a non-asymptotic bound on the probability of excess section error rate of the AMP decoder for any fixed rate $R < \mathcal{C}$. The bound shows how the error performance depends on each parameter of the SC-SPARC, and highlights the tradeoffs involved in choosing these parameters.
- Theorem 2 gives a concentration inequality for the mean-squared error (MSE) of the AMP decoder in each iteration. Theorem 1 is a straightforward consequence of Theorem 2.
- With the same technique used for proving Theorem 2, we can obtain a concentration inequality for the MSE of AMP for compressed sensing with spatially coupled measurement matrices. This result, stated in Theorem 3, refines the asymptotic result for this model obtained in [40], and makes explicit how the probability of deviation of the MSE (from the state evolution prediction) depends on the problem dimension as well as the parameters defining the spatial coupling.

Section 6 presents numerical simulation results showing the finite length error performance of SC-SPARCs over the complex AWGN channel. The error performance is compared with coded modulation schemes which use Quadrature Amplitude Modulation (QAM) with LDPC codes from the DVB-S2 standard. We observe that at moderately high rates (around 1.5 bits/dimension), SC-SPARCs have better error performance than the coded modulation schemes considered.

In Section 7, we prove Theorem 2. The proof has two key technical ingredients. The first is a conditional distribution lemma (Lemma 7.4) which provides a non-asymptotic characterization of the conditional distribution of the AMP iterates. This characterization is then used to prove that various scalar products involving the AMP iterates concentrate around deterministic values (Lemma 7.6). In Section 7, we give an overview of the key ideas in the proof of Theorem 2, then state the main technical lemmas and use them to prove Theorem 2. The full proofs of the lemmas are deferred to Section 8.

Though our approach to proving Theorems 1 and 2 is broadly similar to that used for power-allocated SPARCs in [41], the block-wise structure of the spatially coupled design matrix introduces several new technical challenges. For example, we need to define block-adjusted versions of the AMP iterates (vectors) to obtain the appropriate linear constraints for the conditional distribution lemma

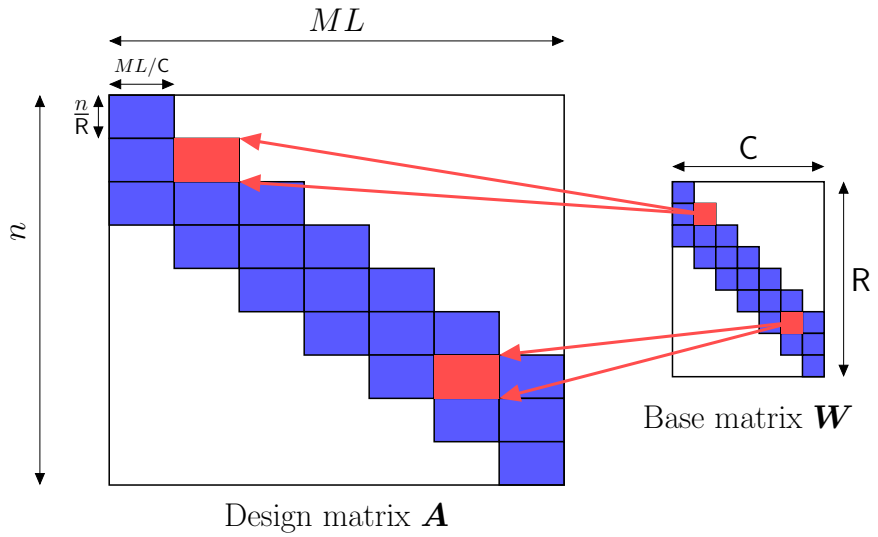

Figure 2: An $n \times ML$ spatially coupled design matrix $\boldsymbol{A}$ consists of $\mathsf{R} \times \mathsf{C}$ blocks, each of size $\frac{n}{\mathsf{R}} \times \frac{ML}{\mathsf{C}}$. The entries of $\boldsymbol{A}$ are independent and normally distributed with $A_{ij} \sim \mathcal{N}(0, \frac{1}{L} W_{\mathsf{r}(i)\mathsf{c}(j)})$, where $\boldsymbol{W}$ is the base matrix. The base matrix shown here is an (ω, Λ, ρ) base matrix with parameters $\omega = 3, \Lambda = 7$ and $\rho = 0$. The white parts of $\boldsymbol{A}$ and $\boldsymbol{W}$ correspond to zeros.

(see (56)-(61)). Similarly, in the main concentration lemma (Lemma 7.6) we establish concentration results for scalar products scaled block-wise by the entries of the base matrix.

Notation: For a positive integer m, we use $[m]$ to denote the set $\{1, \ldots, m\}$. For $x \in \mathbb{R}$, we let $x_+ = \max\{x, 0\}$. Throughout the paper, we use plain font for scalars, bold font for vectors and matrices, and subscripts to denote entries of a vector or matrix. For example, if $\boldsymbol{x}$ is a vector, we write x_i for its i^{th} component. Similarly, if $\boldsymbol{X}$ is a matrix, we write X_{ij} for its $(i,j)^{th}$ entry. The transpose of $\boldsymbol{X}$ is denoted by $\boldsymbol{X}^*$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$.

We write $\boldsymbol{I}_t$ for the $t \times t$ identity matrix; the subscript is dropped when the dimension is clear from context. The indicator function of an event $\mathcal{A}$ is denoted by $\mathbb{I}\{\mathcal{A}\}$. For deterministic sequences $(s_n)_{n \geq 0}, (x_n)_{n \geq 0}$, we write $s_n = \Theta(x_n)$ if s_n / x_n is bounded above and below by strictly positive constants for all sufficiently large n.

2 Spatially coupled SPARC construction

As in the standard construction, a spatially coupled (SC) SPARC is defined by a design matrix $\boldsymbol{A}$ of dimension $n \times ML$, where n is the code length. The codeword is $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ has one non-zero entry in each of the L sections. In an SC-SPARC, since the variances of the entries of different blocks of $\boldsymbol{A}$ can be varied, without loss of generality we will set the value of each non-zero entry of $\boldsymbol{\beta}$ to 1.

In an SC-SPARC, the matrix $\boldsymbol{A}$ consists of independent zero-mean normally distributed entries whose variances are specified by a *base matrix* $\boldsymbol{W}$ of dimension $\mathsf{R} \times \mathsf{C}$. The design matrix $\boldsymbol{A}$ is obtained from the base matrix $\boldsymbol{W}$ by replacing each entry W_{rc} by an $(n/\mathsf{R}) \times (ML/\mathsf{C})$ block with i.i.d. entries $\sim \mathcal{N}(0, W_{\mathsf{rc}}/L)$, for $\mathsf{r} \in [\mathsf{R}]$, $\mathsf{c} \in [\mathsf{C}]$. This is analogous to the "graph lifting" procedure for constructing spatially coupled LDPC codes from protographs [34]. See Fig. 2 for an example.

From the construction, the design matrix $\boldsymbol{A}$ has independent normal entries

$$A_{ij} \sim \mathcal{N}\Big(0, \frac{1}{L} W_{\mathsf{r}(i)\mathsf{c}(j)}\Big) \quad i \in [n],\ j \in [ML]. \tag{4}$$

The operators $\mathsf{r}(\cdot) : [n] \to [\mathsf{R}]$ and $\mathsf{c}(\cdot) : [ML] \to [\mathsf{C}]$ in (4) map a particular row or column index in $\boldsymbol{A}$ to its corresponding *row block* or *column block* index in $\boldsymbol{W}$. We require C to divide L, resulting in $\frac{L}{\mathsf{C}}$ sections per column block.

The non-zero coefficients of $\boldsymbol{\beta}$ are all set to 1. Then it can be verified that $\mathbb{E}[\|\boldsymbol{x}\|^2] = nP$ (and the power constraint is satisfied with high probability) if the entries of the base matrix $\boldsymbol{W}$ satisfy

$$\frac{1}{\mathsf{RC}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} = P \tag{5}$$

The trivial base matrix with $\mathsf{R} = \mathsf{C} = 1$ corresponds to a standard (non-coupled) SPARC with uniform power allocation, while a base matrix consisting of a single row $\mathsf{R} = 1$, $\mathsf{C} = L$ is equivalent to a standard SPARC with power allocation. In this paper, we will use the following base matrix inspired by the coupling structure of SC-LDPC codes constructed from protographs [34].

Definition 2.1. *An (ω, Λ, ρ) base matrix $\boldsymbol{W}$ for SC-SPARCs is described by three parameters: coupling width $\omega \geq 1$ coupling length $\Lambda \geq 2\omega - 1$, and $\rho \in [0, 1)$ which determines the fraction of power allocated to the coupled entries in each column. The matrix has $\mathsf{R} = \Lambda + \omega - 1$ rows, $\mathsf{C} = \Lambda$ columns, with each column having ω identical non-zero entries. For an average power constraint P, the (r, c)th entry of the base matrix, for $\mathsf{r} \in [\mathsf{R}], \mathsf{c} \in [\mathsf{C}]$, is given by*

$$W_{\mathsf{rc}} = \begin{cases} (1-\rho)P \cdot \frac{\Lambda+\omega-1}{\omega} & \text{if } \mathsf{c} \leq \mathsf{r} \leq \mathsf{c} + \omega - 1, \\ \rho P \cdot \frac{\Lambda+\omega-1}{\Lambda-1} & \text{otherwise.} \end{cases} \tag{6}$$

It is easy to verify that this definition satisfies the power allocation constraint in (5). For example, the base matrix in Fig. 2 has parameters $\omega = 3$ and $\Lambda = 7$. For our simulations in Section 6, we use $\rho = 0$, whereas for our main theoretical result (Theorem 1) we choose ρ to be a small positive value proportional to the rate gap from capacity. (Choosing $\rho = 0$ causes some technical difficulties in the proof, which can be addressed by picking a suitable $\rho > 0$.) The $(\omega, \Lambda, \rho = 0)$ base matrix construction was previously used for SC-SPARCs in [42]. Other base matrix constructions can be found in [8, 9, 15, 39].

Each non-zero entry in a base matrix $\boldsymbol{W}$ corresponds to an $(n/\mathsf{R}) \times (ML/\mathsf{C})$ block in the design matrix $\boldsymbol{A}$. Each block can be viewed as a standard (non-coupled) SPARC with $\frac{L}{\mathsf{C}}$ sections (with M columns in each section), code length n/R, and rate $R_{\text{inner}} = \frac{(L/\mathsf{C}) \ln M}{(n/\mathsf{R})}$ nats. Using (2), the overall rate of the SC-SPARC is related to R_{inner} according to

$$R = \frac{\mathsf{C}}{\mathsf{R}} R_{\text{inner}} = \frac{\Lambda}{\Lambda + \omega - 1} R_{\text{inner}}, \tag{7}$$

where the last equality holds for an (ω, Λ, ρ) base matrix.

With spatial coupling, ω is an integer greater than 1, so $R < R_{\text{inner}}$. The difference $(R_{\text{inner}} - R)$ is sometimes referred to as the rate loss due to spatial coupling. From (7), we see that rate loss depends on the ratio $(\omega - 1)/\Lambda$, which becomes negligible when Λ is large w.r.t. ω. For our theoretical results, we will be interested in the regime where $L \gg \mathsf{C} = \Lambda \gg \omega$. Without loss of generality, we will henceforth assume that $\omega < \sqrt{\Lambda}$.

Remark 2.1. SC-SPARC constructions usually have a 'seed' to jumpstart decoding. In [8], a small fraction of $\boldsymbol{\beta}$'s sections are fixed a priori — this pinning condition is used to analyze the state evolution equations via the potential function method. Analogously, the construction in [9] introduces additional rows in the design matrix for blocks corresponding to the first row of the base matrix. In an (ω, Λ, ρ) base matrix, the fact that the number of rows in the base matrix exceeds the number of columns by $(\omega - 1)$ helps decoding start from both ends. The asymptotic state evolution equations in Sec. 4 describe how AMP decoding progresses in an (ω, Λ, ρ) base matrix.

In the remainder of the paper, we use subscripts in sans-serif font (r or c) to denote row or column block indices. Thus, $\boldsymbol{\beta}_{\mathsf{c}} \in \mathbb{R}^{ML/\mathsf{C}}$ denotes the c-th column block of $\boldsymbol{\beta} \in \mathbb{R}^{ML}$, for $\mathsf{c} \in [\mathsf{C}]$.

3 The AMP decoder for spatially coupled SPARCs

Recall that the decoding task is to recover the message vector $\boldsymbol{\beta} \in \mathbb{R}^{ML}$ from the channel output sequence $\boldsymbol{y} \in \mathbb{R}^n$ produced according to (3). An Approximate Message Passing (AMP) decoding algorithm can be derived using an approach similar to the one for standard SPARCs [4, Appendix A], with modifications to account for the different variances for the blocks of $\boldsymbol{A}$ specified by the base matrix. The AMP decoder can also be derived from the Generalized AMP algorithm in [19].

The AMP decoder initializes $\boldsymbol{\beta}^0$ to the all-zero vector, and for $t \geq 0$, iteratively computes:

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}^t + \boldsymbol{v}^t \odot \boldsymbol{z}^{t-1}, \tag{8}$$

$$\boldsymbol{\beta}^{t+1} = \eta^t(\boldsymbol{\beta}^t + (\overline{\boldsymbol{S}}^t \odot \boldsymbol{A})^* \boldsymbol{z}^t). \tag{9}$$

Here $\odot$ denotes the Hadamard (entry-wise) product. The vector $\boldsymbol{v}^t \in \mathbb{R}^n$, the matrix $\overline{\boldsymbol{S}}^t \in \mathbb{R}^{n \times ML}$, and the denoising function η^t are defined below in terms of the state evolution parameters. Quantities with negative time indices are set to zero.

For any rate $R < \mathcal{C}$, the AMP decoder is run for a finite number of iterations T, where T is specified later in Section 4. After T iterations, the maximum value in each section $\ell \in [L]$ of $\boldsymbol{\beta}^T$ is set to 1 and remaining entries are set to 0 to obtain the decoded message $\widehat{\boldsymbol{\beta}}$.

3.1 State evolution

Given a base matrix $\boldsymbol{W}$, state evolution (SE) iteratively defines a sequence of scalars $(\phi_{\mathsf{r}}^t)_{\mathsf{r} \in \mathsf{R}}$ and $(\psi_{\mathsf{c}}^t)_{\mathsf{c} \in \mathsf{C}}$, for $t \geq 0$. Initialize $\psi_{\mathsf{c}}^0 = 1$ for $\mathsf{c} \in [\mathsf{C}]$, and for $t = 0, 1, \ldots$, compute

$$\sigma_{\mathsf{r}}^t = \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \psi_{\mathsf{c}}^t, \qquad \phi_{\mathsf{r}}^t = \sigma^2 + \sigma_{\mathsf{r}}^t, \qquad \mathsf{r} \in [\mathsf{R}], \tag{10}$$

$$\psi_{\mathsf{c}}^{t+1} = 1 - \mathcal{E}(\tau_{\mathsf{c}}^t), \qquad \mathsf{c} \in [\mathsf{C}], \tag{11}$$

where

$$\tau_{\mathsf{c}}^t = \frac{R}{\ln M} \left[\frac{1}{\mathsf{R}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \frac{W_{\mathsf{rc}}}{\phi_{\mathsf{r}}^t} \right]^{-1}, \tag{12}$$

and $\mathcal{E}(\tau_{\mathsf{c}}^t)$ is defined with $U_1, \ldots, U_M \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ as

$$\mathcal{E}(\tau_{\mathsf{c}}^t) = \mathbb{E}\left[\frac{e^{U_1/\sqrt{\tau_{\mathsf{c}}^t}}}{e^{U_1/\sqrt{\tau_{\mathsf{c}}^t}} + e^{-1/\tau_{\mathsf{c}}^t} \sum_{j=2}^{M} e^{U_j/\sqrt{\tau_{\mathsf{c}}^t}}} \right]. \tag{13}$$

For $t \geq 1$, the vector $\boldsymbol{v}^t \in \mathbb{R}^n$ in (8) has a block-wise structure, with the ith entry defined as

$$v_i^t = \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}, \qquad \text{if } \mathsf{r}(i) = \mathsf{r}, \tag{14}$$

where we recall that $\mathsf{r}(i)$ denotes the row block index of the ith entry. (The vector $\boldsymbol{v}^0$ is defined to be all-zeros.) Similarly, $\overline{\boldsymbol{S}}^t \in \mathbb{R}^{n \times ML}$ in (9) has a block-wise structure, with entries defined as follows. For $i \in [n], j \in [ML]$,

$$\overline{S}_{ij}^t = \frac{\tau_{\mathsf{c}}^t}{\phi_{\mathsf{r}}^t}, \qquad \text{if } \mathsf{r}(i) = \mathsf{r} \text{ and } \mathsf{c}(j) = \mathsf{c}. \tag{15}$$

The function $\eta^t = (\eta_1^t, \ldots, \eta_{ML}^t) : \mathbb{R}^{ML} \to \mathbb{R}^{ML}$ in (9) is defined as follows, for $j \in [ML]$. For j in section $\ell \in [L]$, with section ℓ in column block $\mathsf{c} \in [\mathsf{C}]$,

$$\eta_j^t(\boldsymbol{s}) = \frac{e^{s_j/\tau_{\mathsf{c}}^t}}{\sum_{j' \in \mathrm{sec}(\ell)} e^{s_{j'}/\tau_{\mathsf{c}}^t}}, \tag{16}$$

where $\mathrm{sec}(\ell) := \{(\ell-1)M + 1, \ldots, \ell M\}$ refers to the set of indices in section ℓ. We note that $\eta_j^t(\boldsymbol{s})$ depends on all the components of $\boldsymbol{s}$ in the section containing j.

3.2 Interpretation of the AMP decoder

The input to $\eta_j^t(\cdot)$ in (16) can be viewed as a noisy version of $\boldsymbol{\beta}$. In particular, consider an index j in section $\ell \in [L]$ which belongs to column block $\mathsf{c} \in [\mathsf{C}]$. Recall that $\boldsymbol{\beta}_\ell \in \mathbb{R}^M$ is section ℓ of the message vector, and let $\boldsymbol{s}_\ell^t = \boldsymbol{s}_\ell$ denote section ℓ of the input vector to the function $\eta_j^t(\cdot)$. Then, $\boldsymbol{s}_\ell^t$ is approximately distributed as $\boldsymbol{\beta}_\ell + \sqrt{\tau_{\mathsf{c}}^t}\boldsymbol{Z}_\ell$, where $\boldsymbol{Z}_\ell \in \mathbb{R}^M$ is a standard normal random vector independent of $\boldsymbol{\beta}_\ell$. Under the above distributional assumption, the denoising function η_j^t in (16) is the minimum mean squared error (MMSE) estimator for $\boldsymbol{\beta}_j$, i.e.,

$$\eta_j^t(\boldsymbol{s}) = \mathbb{E}\left[\boldsymbol{\beta}_j \mid \boldsymbol{\beta}_\ell + \sqrt{\tau_{\mathsf{c}}^t}\,\boldsymbol{Z}_\ell = \boldsymbol{s}_\ell\right], \qquad \text{for } j \in [ML], \tag{17}$$

where the expectation is calculated over and $\boldsymbol{Z}_\ell \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M)$ and $\boldsymbol{\beta}_\ell$, which is uniformly distributed over the M vectors with a single non-zero entry equal to 1.

The entries of the modified residual $\boldsymbol{z}^t$ in (8) are approximately Gaussian and independent, with the variance determined by the block index. (A precise characterization of the distribution is given in Lemmas 7.4 and 7.5.) For $\mathsf{r} \in [\mathsf{R}]$, the SE parameter ϕ_{r}^t approximates the variance of $\boldsymbol{z}_{\mathsf{r}}^t$, the rth block of the residual. The 'Onsager' term $\boldsymbol{v}^t \odot \boldsymbol{z}^{t-1}$ in (8) reflects the block-wise structure of $\boldsymbol{z}^t$. To summarize, the key difference from the state evolution parameters for standard SPARCs is that here the variances of the effective observation $\boldsymbol{s}^t$ and the modified residual $\boldsymbol{z}^t$ depend on their column- and row-block indices, respectively. These variances are captured by $\{\tau_{\mathsf{c}}^t\}_{\mathsf{c} \in [\mathsf{C}]}$ and $\{\phi_{\mathsf{r}}^t\}_{\mathsf{r} \in [\mathsf{R}]}$.

3.3 Measuring the performance of the AMP decoder

The performance of a SPARC decoder is measured by the *section error rate*, defined as

$$\mathcal{E}_{\mathrm{sec}} := \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}\{\widehat{\boldsymbol{\beta}}_{\mathrm{sec}(\ell)} \neq \boldsymbol{\beta}_{\mathrm{sec}(\ell)}\}. \tag{18}$$

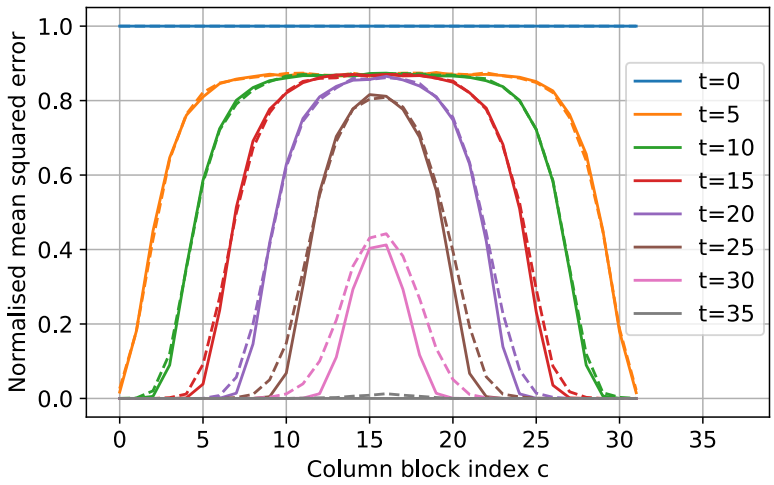

Figure 3: NMSE $\frac{\|\boldsymbol{\beta}_{\mathsf{c}}^t-\boldsymbol{\beta}_{\mathsf{c}}\|_2^2}{L/\mathsf{C}}$ vs. column block index $\mathsf{c} \in [\mathsf{C}]$ for several iteration numbers. The SC-SPARC with an $(\omega, \Lambda, \rho = 0)$ base matrix has the following parameters: $R = 1.5$ bits, $\mathcal{C} = 2$ bits, $\omega = 6$, $\Lambda = 32$, $M = 512$, $L = 2048$ and $n = 12284$. The solid lines are the state evolution predictions from (11), and the dotted lines are the average NMSE over 100 instances of AMP decoding.

If the AMP decoder is run for T iterations, the section error rate can be bounded in terms of the squared error $\|\boldsymbol{\beta}^T - \boldsymbol{\beta}\|^2$ as follows. Since the unique non-zero entry in any section $\ell \in [L]$ of $\boldsymbol{\beta}$ equals 1 and $\widehat{\boldsymbol{\beta}}_{\mathrm{sec}(\ell)} \neq \boldsymbol{\beta}_{\mathrm{sec}(\ell)}$ implies that the corresponding element of $\boldsymbol{\beta}^T_{\mathrm{sec}(\ell)}$ is less than or equal to $1/2$,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{sec}(\ell)} \neq \boldsymbol{\beta}_{\mathrm{sec}(\ell)} \quad \Rightarrow \quad \|\boldsymbol{\beta}^T_{\mathrm{sec}(\ell)} - \boldsymbol{\beta}_{\mathrm{sec}(\ell)}\|_2^2 \geq \frac{1}{4}. \tag{19}$$

We recall that $\boldsymbol{\beta}_{\mathsf{c}}$ is the part of the message vector corresponding to column block c of the design matrix. There are $\frac{L}{\mathsf{C}}$ sections in $\boldsymbol{\beta}_{\mathsf{c}}$, with the non-zero entry in each section being equal to 1; we denote by $\boldsymbol{\beta}_{\mathsf{c}_\ell}$ the ℓth of these sections, for $\ell \in [L/\mathsf{C}]$. Then, (19) implies

$$\begin{aligned}
\mathcal{E}_{\mathrm{sec}} &= \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}\{\widehat{\boldsymbol{\beta}}_{\mathrm{sec}(\ell)} \neq \boldsymbol{\beta}_{\mathrm{sec}(\ell)}\} = \frac{1}{L}\sum_{\mathsf{c}=1}^{\mathsf{C}}\sum_{\ell=1}^{L/\mathsf{C}} \mathbb{I}\left\{\widehat{\boldsymbol{\beta}}_{\mathsf{c}_\ell} \neq \boldsymbol{\beta}_{\mathsf{c}_\ell}\right\} \\
&\leq \frac{4}{L}\sum_{\mathsf{c}=1}^{\mathsf{C}}\sum_{\ell=1}^{L/\mathsf{C}} \|\boldsymbol{\beta}^T_{\mathsf{c}_\ell} - \boldsymbol{\beta}_{\mathsf{c}_\ell}\|_2^2 = 4\left[\frac{1}{\mathsf{C}}\sum_{\mathsf{c}=1}^{\mathsf{C}} \frac{\|\boldsymbol{\beta}^T_{\mathsf{c}} - \boldsymbol{\beta}_{\mathsf{c}}\|_2^2}{L/\mathsf{C}}\right] = 4\left[\frac{1}{L}\|\boldsymbol{\beta}^T - \boldsymbol{\beta}\|^2\right].
\end{aligned} \tag{20}$$

We can therefore focus on bounding the bracketed term on the RHS of (20) which is the overall *normalized mean square error* (NMSE). Fig. 3 shows that ψ^t ((11)) closely tracks the NMSE of each block of the message vector, i.e., $\psi^t_{\mathsf{c}} \approx \frac{\|\boldsymbol{\beta}^t_{\mathsf{c}}-\boldsymbol{\beta}_{\mathsf{c}}\|_2^2}{L/\mathsf{C}}$ for $\mathsf{c} \in [\mathsf{C}]$. We additionally observe from the figure that as AMP iterates, the NMSE reduction propagates from the ends towards the center blocks.

4 Decoding progression according to state evolution

In this section, we derive bounds for the state evolution parameters which help explain the decoding propagation illustrated in Fig. 3. These bounds lead to a succinct asymptotic characterization of

state evolution (as $M \to \infty$). The non-asymptotic version of these bounds (for large but finite M) will be used to establish the main result in Theorems 1 and 2.

Lemma 4.1. *Let $\boldsymbol{W} \in \mathbb{R}^{R \times C}$ be a base matrix having row and column averages that are bounded above and below by strictly positive constants. That is, there exist constants $\kappa_{\mathrm{L}}, \kappa_{\mathrm{U}} > 0$ such that*

$$\kappa_{\mathrm{L}} \leq \frac{1}{C}\sum_{c'} W_{rc'},\ \frac{1}{R}\sum_{r'} W_{r'c} \leq \kappa_{\mathrm{U}}, \qquad r \in [R],\ c \in [C].$$

Let

$$\nu_c^t := \frac{1}{\tau_c^t \ln M} = \frac{1}{R}\left[\frac{1}{R}\sum_{r=1}^{R}\frac{W_{rc}}{\phi_r^t}\right]. \tag{21}$$

For sufficiently large M and any $\delta \in (0, \frac{1}{2})$, $\tilde{\delta} \in (0,1)$,

$$\left(1 - M^{-k_1\tilde{\delta}^2}\right)\mathbb{I}\{\nu_c^t < 2 - \tilde{\delta}\} \leq \psi_c^{t+1} \leq 1 - \left(1 - \frac{M^{-k\delta^2}}{\delta\sqrt{\log M}}\right)\mathbb{I}\{\nu_c^t > 2 + \delta\}, \qquad c \in [C], \tag{22}$$

where k, k_1 are positive constants depending only on κ_{L} and κ_{U}.

The proof of the lemma is given in Appendix A.1.

Lemma 4.1 implies the following asymptotic state evolution recursion as $M \to \infty$. Initialise $\bar{\psi}_{\mathsf{c}}^0 = 1$, for $\mathsf{c} \in [\mathsf{C}]$, and for $t = 0, 1, 2, \ldots$,

$$\bar{\phi}_{\mathsf{r}}^t = \sigma^2 + \frac{1}{\mathsf{C}}\sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}}\bar{\psi}_{\mathsf{c}}^t, \qquad \mathsf{r} \in [\mathsf{R}], \tag{23}$$

$$\bar{\psi}_{\mathsf{c}}^{t+1} = 1 - \mathbb{I}\left\{\frac{1}{R\,\mathsf{R}}\sum_{\mathsf{r}=1}^{\mathsf{R}}\frac{W_{\mathsf{rc}}}{\bar{\phi}_{\mathsf{r}}^t} > 2\right\}, \qquad \mathsf{c} \in [\mathsf{C}], \tag{24}$$

where $\bar{\phi}, \bar{\psi}$ indicate asymptotic values as $M \to \infty$.

The asymptotic SE recursion (23)-(24) is given for a general base matrix $\boldsymbol{W}$. To get some insight into the decoding progression, we specialise the result to the $(\omega, \Lambda, \rho = 0)$ base matrix introduced in Definition 2.1. Recall that an $(\omega, \Lambda, \rho = 0)$ base matrix has $\mathsf{R} = \Lambda + \omega - 1$ rows and $\mathsf{C} = \Lambda$ columns, with each column having ω non-zero entries, all equal to $P \cdot \frac{\Lambda+\omega-1}{\omega}$.

Corollary 4.1. *The asymptotic state evolution recursion (23)-(24) for an $(\omega, \Lambda, \rho = 0)$ base matrix is as follows. Initialise $\bar{\psi}_c^0 = 1\ \forall\ c \in [\Lambda]$, and for $t = 0, 1, 2, \ldots$,*

$$\bar{\phi}_r^t = \sigma^2 + \frac{\vartheta\, P}{\omega}\sum_{c=\underline{c}_r}^{\bar{c}_r}\bar{\psi}_c^t, \quad r \in [\Lambda + \omega - 1], \tag{25}$$

$$\bar{\psi}_c^{t+1} = 1 - \mathbb{I}\left\{\frac{P}{R\,\omega}\sum_{r=c}^{c+\omega-1}\frac{1}{\bar{\phi}_r^t} > 2\right\}, \quad c \in [\Lambda], \tag{26}$$

where $\vartheta = \frac{\Lambda+\omega-1}{\Lambda}$, and

$$(\underline{c}_r,\, \bar{c}_r) = \begin{cases} (1,\, r) & \text{if } \ 1 \leq r \leq \omega \\ (r-\omega+1,\, r) & \text{if } \ \omega \leq r \leq \Lambda \\ (r-\omega+1,\, \Lambda) & \text{if } \ \Lambda \leq r \leq \Lambda+\omega-1. \end{cases} \tag{27}$$

Proof. Substitute the value of W_{rc} from (6), with $\rho = 0$ and $\mathsf{C} = \Lambda$, $\mathsf{R} = \Lambda + \omega - 1$ in (23)-(24). ■

Observe that the $\bar{\phi}_{\mathsf{r}}^t$'s and $\bar{\psi}_{\mathsf{c}}^t$'s are symmetric about the middle indices, i.e. $\bar{\phi}_{\mathsf{r}}^t = \bar{\phi}_{\mathsf{R}-\mathsf{r}+1}^t$ for $\mathsf{r} \leq \lfloor \frac{\mathsf{R}}{2} \rfloor$ and $\bar{\psi}_{\mathsf{c}}^t = \bar{\psi}_{\mathsf{C}-\mathsf{c}+1}^t$ for $\mathsf{c} \leq \lfloor \frac{\mathsf{C}}{2} \rfloor$.

Consider the initial step ($t = 0$): from (25) the value of $\bar{\phi}_{\mathsf{r}}^0$ for each r depends on the number of non-zero entries in row r of $\boldsymbol{W}$, which is equal to $\bar{\mathsf{c}}_{\mathsf{r}} - \underline{\mathsf{c}}_{\mathsf{r}} + 1$, with $\bar{\mathsf{c}}_{\mathsf{r}}, \underline{\mathsf{c}}_{\mathsf{r}}$ given by (27). Therefore, $\bar{\phi}_{\mathsf{r}}^0$ increases from $\mathsf{r} = 1$ until $\mathsf{r} = \omega$, is constant for $\omega \leq \mathsf{r} \leq \Lambda$, and then starts decreasing again for $\Lambda < \mathsf{r} \leq \Lambda + \omega - 1$. As a result, $\bar{\psi}_{\mathsf{c}}^1$ is smallest for c at either end of the base matrix ($\mathsf{c} \in \{1, \Lambda\}$) and increases as c moves towards the middle, since the $\sum_{\mathsf{r}=\mathsf{c}}^{\mathsf{c}+\omega-1} (\bar{\phi}_{\mathsf{r}}^0)^{-1}$ term in (26) is largest for $\mathsf{c} \in \{1, \Lambda\}$, followed by $\mathsf{c} \in \{2, \Lambda - 1\}$, and so on. Therefore, we expect the blocks of the message vector corresponding to column index $\mathsf{c} \in \{1, \Lambda\}$ to be decoded most easily, followed by $\mathsf{c} \in \{2, \Lambda - 1\}$, and so on. Fig. 3 shows that this is indeed the case.

The decoding propagation phenomenon seen in Fig. 3 can also be explained using Corollary 4.1 by tracking the evolution of the $\bar{\phi}_{\mathsf{r}}^t$'s and $\bar{\psi}_{\mathsf{c}}^t$'s. In particular, one finds that if column c^* decodes in iteration t, i.e. $\bar{\psi}_{\mathsf{c}^*}^t = 0$, then columns within a coupling width away, i.e. columns $\mathsf{c} \in \{\mathsf{c}^* - (\omega - 1), \ldots, \mathsf{c}^* + (\omega - 1)\}$, will become easier to decode in iteration $(t + 1)$. This wave-like decoding propagation also occurs in spatially coupled LDPC codes decoded with belief propagation. The propagation of the LDPC decoding wave (in the large system limit) was studied in [43].

4.1 Decoding progression

We make the above discussion precise by characterizing the decoding progression for an (ω, Λ, ρ) base matrix (with $\rho > 0$) using Lemma 4.1. Recalling that

$$\vartheta = 1 + \frac{\omega - 1}{\Lambda}, \tag{28}$$

we will consider rates R such that

$$R < \frac{1}{2\vartheta} \ln(1 + \vartheta\, \mathsf{snr}), \qquad \text{where} \qquad \mathsf{snr} = \frac{P}{\sigma^2}. \tag{29}$$

Note that the RHS of (29) can be made arbitrarily close to the channel capacity $\mathcal{C}$ by making $\frac{\omega - 1}{\Lambda}$ small enough. Indeed, since the expression in (29) is decreasing in ϑ for $\vartheta > 1$, we have

$$\mathcal{C} > \frac{1}{2\vartheta} \ln(1 + \vartheta\, \mathsf{snr}) > \frac{\mathcal{C}}{\vartheta}. \tag{30}$$

Proposition 4.1. *Consider a rate R SC-SPARC with an $n \times ML$ design matrix constructed using an (ω, Λ, ρ) base matrix and a constant $\delta \in (0, \min\{\frac{\Delta}{2R}, \frac{1}{2}\})$, where $0 \leq \rho \leq \min\{\frac{\Delta}{3\mathsf{snr}}, \frac{1}{2}\}$, and*

$$\Delta := \frac{1}{2\vartheta} \ln(1 + \vartheta \mathsf{snr}) - R. \tag{31}$$

If the rate satisfies $R < \frac{(1-\rho)\mathsf{snr}}{(2+\delta)(1+\vartheta\mathsf{snr})}$, then all the column blocks of the message vector simultaneously decode in one iteration, i.e., for all $\mathsf{c} \in [\Lambda]$,

$$\psi_{\mathsf{c}}^1 \leq f_{M,\delta} := \frac{M^{-k\delta^2}}{\delta\sqrt{\log M}} \tag{32}$$

for sufficiently large M, where $k > 0$ is a universal constant.

Otherwise, if the rate satisfies $\frac{(1-\rho)\mathsf{snr}}{(2+\delta)(1+\vartheta\mathsf{snr})} \leq R < \frac{1}{2\vartheta}\ln(1+\vartheta\,\mathsf{snr})$, the coupling width ω satisfies

$$\omega > \left(\frac{\vartheta\,\mathsf{snr}^2}{1+\vartheta\mathsf{snr}}\right)\frac{1}{\Delta}, \tag{33}$$

and

$$g = \frac{(1+\vartheta\,\mathsf{snr})\Delta}{\vartheta\,\mathsf{snr}^2}\,\omega, \tag{34}$$

then, for $t \geq 1$ and

$$\mathsf{c} \leq \max\left\{tg,\ \left\lceil\frac{\Lambda}{2}\right\rceil\right\}, \tag{35}$$

we have

$$\psi_{\mathsf{c}}^t = \psi_{\Lambda-\mathsf{c}+1}^t \leq f_{M,\delta} \tag{36}$$

for sufficiently large M.

The proof of the proposition is given in Appendix A.2.

Remark 4.1 (BP threshold)**.** For rates R smaller than $\frac{\mathsf{snr}}{2(1+\mathsf{snr})}$, one does not require spatial coupling (or power allocation) for reliable SPARC decoding. Indeed, consider a standard non-coupled SPARC where the 1-by-1 base matrix is a single entry equal to P. Using Lemma 4.1 in the state evolution recursion (10)-(13), we see that if $R < \frac{\mathsf{snr}}{(2+\delta)(1+\mathsf{snr})}$ for some $\delta \in (0,1)$, then the whole message vector decodes in one iteration, i.e., $\psi^1 \leq f_{M,\delta}$. The threshold $\frac{\mathsf{snr}}{2(1+\mathsf{snr})}$ can be interpreted as the BP threshold in the $M \to \infty$ limit.

Remark 4.2 (Choice of base matrix parameters)**.** For any fixed rate $\frac{\mathsf{snr}}{2(1+\mathsf{snr})} \leq R < \mathcal{C} = \frac{1}{2}\ln(1+\mathsf{snr})$, Proposition 4.1 requires the parameters (ω, Λ) to be chosen such that: i) the ratio (ω/Λ) is small enough that the rate gap Δ in (31) is positive, and ii) ω is large enough that (33) is satisfied. These two conditions can be satisfied for any fixed $R < \mathcal{C}$ by taking (for example) $\Lambda > \omega^2$, and ω sufficiently large.

For rates larger than the threshold $\frac{\mathsf{snr}}{2(1+\mathsf{snr})}$, the proposition says that if the coupling width ω is large enough (as specified by (33)), then in iteration t at least the first and last $\lfloor gt \rfloor$ column blocks from each end are expected to decode. Furthermore, the proof shows that if $g_t \leq \Lambda/2$ is the exact number of column blocks such that $\psi_{\mathsf{c}}^t = \psi_{\Lambda-\mathsf{c}+1}^t \leq f_{M,\delta}$ for $\mathsf{c} \leq g_t$, then $g_{t+1} \geq \lfloor g_t + g \rfloor$, i.e., in each iteration at least $\lfloor g \rfloor \geq 1$ additional column blocks of the message vector from each end are expected to decode.

This decoding progression continues until iteration T when all column blocks have been decoded, i.e., $\psi_{\mathsf{c}}^T \leq f_{M,\delta}$ for $\mathsf{c} \in [\mathsf{C}]$. More precisely, we run the AMP decoder for T iterations where

$$T := \min\{t : \psi_{\mathsf{c}}^t \leq f_{M,\delta} \text{ for } \mathsf{c} \in [\mathsf{C}]\}. \tag{37}$$

Proposition 4.1 implies that for rates larger than the threshold

$$T \leq \left\lceil\frac{\Lambda}{2g}\right\rceil. \tag{38}$$

Using the interpretation of the AMP decoder in Sections 3.2 and 3.3, after iteration T we expect the mean squared error $\frac{1}{L}\|\boldsymbol{\beta} - \boldsymbol{\beta}^T\|^2$ to be small. We note that g is proportional to Δ, which represents the rate gap from capacity (see (30), (31)). Therefore, from (38) the number of iterations T grows as the rate approaches the channel capacity. For a fixed R the quantity $f_{M,\delta}$ tends to 0 with growing M.

5 Main Result

The main result, stated in the theorem below is a bound on the probability of the section error rate of the AMP decoder exceeding a target level ϵ, for any $\epsilon \in (0,1)$.

Theorem 1. *Consider an (ω, Λ, ρ) base matrix $\boldsymbol{W} \in \mathbb{R}^{R\times C}$ with $(R, C) = (\Lambda + \omega - 1, \Lambda)$ and $\rho = \min\{\frac{\Delta}{3\mathsf{snr}}, \frac{1}{2}\}$, where Δ is the rate gap defined in (31). Fix rate $\frac{\mathsf{snr}}{2(1+\mathsf{snr})} \le R < \frac{1}{2\vartheta}\ln(1+\vartheta\,\mathsf{snr})$, where $\vartheta = 1 + \frac{\omega-1}{\Lambda}$, and let ω be large enough that the condition in (33) is satisfied. Let $\mathcal{S}_n$ be a SC-SPARC of rate R defined via an $n \times ML$ design matrix constructed using the base matrix $\boldsymbol{W}$. The parameters (n, M, L) satisfy (2).*

Fix $\epsilon \in (0,1)$, and for $f_{M,\frac{\Delta}{R}}$ defined in Proposition 4.1, let M be large enough that $f_{M,\delta} \le \frac{\epsilon}{8}$ for $\delta = \min\{\frac{\Delta}{3R}, \frac{1}{3}\}$. Then after the AMP decoder is run for T iterations, with T defined in (37), the section error rate (defined in (18)) satisfies

$$P\left(\mathcal{E}_{sec}(\mathcal{S}_n) > \epsilon\right) \le K_{T-1}(RC)^T \exp\Big\{\frac{-\kappa_{T-1} n\epsilon^2}{64(\log M)^{2T}(R/\omega)^{2T-1}}\Big\}. \tag{39}$$

For $t \ge 0$, the constants κ_t and K_t are given by $\kappa_t = [\xi^{2t}(t!)^{24}]^{-1}$ and $K_t = \Xi^{2t}(t!)^{14}$ where $\xi, \Xi > 0$ are universal constants (not depending on the AMP parameters (L, M, n, R, C) or ϵ), but not explicitly specified.

Remark 5.1. *The theorem is stated for rates $\frac{\mathsf{snr}}{2(1+\mathsf{snr})} \le R < \frac{1}{2\vartheta}\ln(1+\vartheta\,\mathsf{snr})$ as this is the region where spatial coupling is required. Indeed, for $R < \frac{\mathsf{snr}}{2(1+\mathsf{snr})}$, Remark 4.1 and the proof of Theorem 1 imply that the probability bound (39) holds with $R = C = \omega = 1$ and $T = 1$. This result also follows from the analysis in [41], applied with a uniform power allocation.*

The bound (39) on the probability of excess section error rate is obtained via a concentration inequality on the normalized MSE of the AMP decoder. Theorem 2 below gives a concentration inequality bounding the probability of deviation of the normalized MSE from the state evolution prediction in each iteration. Recall, by (20), the section error rate can be bounded in terms of the normalized MSE, $\|\boldsymbol{\beta}^T - \boldsymbol{\beta}\|^2/L$. This connection is used to prove Theorem 1 from Theorem 2.

Theorem 2. *With the same assumptions as Theorem 1, for $1 \le t \le T$ and $\epsilon > 0$ we have*

$$P\Big(\Big|\frac{\|\boldsymbol{\beta}^t - \boldsymbol{\beta}\|^2}{L} - \frac{1}{C}\sum_{c\in[C]}\psi_c^t\Big| \ge \epsilon\Big) \le K_{t-1}(RC)^t \exp\Big\{\frac{-\kappa_{t-1} n\epsilon^2}{(\log M)^{2t}(R/\omega)^{2t-1}}\Big\}. \tag{40}$$

Here $\{\psi_c^t\}_{c\in[C]}$ are the state evolution parameters defined in (11), and the constants K_{t-1} and κ_{t-1} are as defined in the statement of Theorem 1.

The dependence of the constants K_t and κ_t on $t!$ arises due to the induction-based proof of the concentration lemma (Lemma 7.6). These constants have not been optimized, but we believe that these constants will depend on $t!$ in any induction-based proof of the result. It is an open question whether the $t!$ factors are fundamental to the problem or if a different analysis of the AMP can yield a different t dependence in these constants.

Theorem 2 is proved in Section 7. We now show how Theorem 1 follows from Theorem 2.

Proof of Theorem 1. Without loss of generality, we can assume that rate gap Δ (defined in (31)) satisfies $\Delta < 2R$. Otherwise the arguments below hold with (Δ/R) replaced by 1.

Taking $t = T$, and noting that $\psi_{\mathsf{c}}^T \leq f_{M,\Delta/(3R)}$ for $\mathsf{c} \in [\mathsf{C}]$ (from Proposition 4.1), Theorem 2 implies that for any $\tilde{\epsilon} > 0$,

$$P\left(\frac{\|\boldsymbol{\beta}^T - \boldsymbol{\beta}\|^2}{L} \geq \tilde{\epsilon} + f_{M,\Delta/(3R)}\right) \leq K_{T-1}(\mathsf{RC})^T \exp\left\{\frac{-\kappa_{T-1} n \tilde{\epsilon}^2}{(\log M)^{2T}(\mathsf{R}/\omega)^{2T-1}}\right\}. \tag{41}$$

Furthermore, from (20) we have

$$P\left(\mathcal{E}_{\text{sec}}(\mathcal{S}_n) > \epsilon\right) \leq P\left(\frac{\|\boldsymbol{\beta}^T - \boldsymbol{\beta}\|^2}{L} \geq \frac{\epsilon}{4}\right). \tag{42}$$

Combining (41) and (42), and taking $\tilde{\epsilon} = \frac{\epsilon}{8}$ and M large enough so that $f_{M,\Delta/(3R)} \leq \frac{\epsilon}{8}$ (see (32)) yields Theorem 1. ∎

5.1 Choosing the SC-SPARC parameters

Theorems 1 and 2 give guidance on how to systematically choose parameters of the spatially coupled SPARC for any fixed rate $R < \mathcal{C} = \frac{1}{2}\ln(1+\mathsf{snr})$ and a target section error rate ϵ. First, choose (ω, Λ) so that the rate gap Δ in (31) is positive and (33) is satisfied. As described in Remark 4.2, this can be done by choosing $\Lambda > \omega^2$, and ω sufficiently large. This determines $\mathsf{R} = (\Lambda + \omega - 1)$ the number of iterations T, which from (38), is bounded by a value proportional to $\Lambda/(\omega\Delta)$.

Next choose M large enough for $f_{M,\delta} \leq \frac{\epsilon}{8}$ for $\delta = \min\{\frac{\Delta}{3R}, \frac{1}{3}\}$. For fixed values of (ω, Λ, M), (39) shows that the probability that the section error rate exceeds ϵ decays exponentially in the block length n. Once n is chosen to be a large multiple of $\mathsf{R} = (\Lambda + \omega - 1)$, the number of sections is $L = \frac{nR}{\ln M}$, which completes the specification of the SPARC.[1]

Remark 5.2. Theorem 1 implies that for any fixed $R < \mathcal{C}$ and $\epsilon \in (0,1)$, one can construct a sequence of rate R spatially coupled SPARCs $\{\mathcal{S}_n\}$ (indexed by code length n) for which

$$\lim_{n\to\infty} \mathcal{E}_{\text{sec}}(\mathcal{S}_n) = 0 \quad \text{almost surely.} \tag{43}$$

Indeed, once (Λ, ω, M) are chosen to satisfy the conditions in Theorem 1, the bound in (39) decreases exponentially in n. The Borel-Cantelli lemma then yields the asymptotic result in (43).

Remark 5.3. As described in [1], to obtain a small probability of codeword error $P(\widehat{\boldsymbol{\beta}} \neq \boldsymbol{\beta})$, one can use a concatenated code with the SPARC as the inner code and an outer Reed-Solomon code. A suitably chosen Reed-Solomon code of rate $(1-2\epsilon)$ ensures that $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ whenever the section error rate $\mathcal{E}_{sec} < \epsilon$, for any $\epsilon > 0$. The overall rate for such a concatenated code is $(1-2\epsilon)R$ and the $P(\widehat{\boldsymbol{\beta}} \neq \boldsymbol{\beta})$ is bounded by the RHS of (39). The reader is referred to [6, Sec. 2.2] for details of how to choose an appropriate Reed-Solomon code.

[1] The parameters L and $N_{\mathsf{C}} = ML/\Lambda$ determined in this way need to be integer-valued, which can be ensured by picking suitable values for ω, Λ, M.

5.2 Tradeoffs in choosing the parameters

We first consider the effect of the base matrix parameters ω, Λ. The condition in (28) and the bounds in (30) together imply that the minimum gap from capacity $\mathcal{C} - R$ is of order $\frac{\omega}{\Lambda}$, or equivalently, of order $\frac{\omega}{\Lambda+\omega-1} = \frac{\omega}{\mathsf{R}}$. Therefore, decreasing the ratio $\frac{\omega}{\mathsf{R}}$ allows rates closer to capacity, but weakens the probability bound in Theorem 1. Indeed, the bound in (39) is exponential in $(\omega/\mathsf{R})^{2T-1}$, with T also increasing with (R/ω).

Next consider the effect of increasing M. From Theorem 1 increasing M allows for a smaller target section error rate ϵ, since we require $\epsilon \geq 8 f_{M,\delta}$. Equivalently, from (36) and Theorem 2, increasing M allows for a smaller state evolution estimate of the normalized MSE, $\|\boldsymbol{\beta}^T - \boldsymbol{\beta}\|^2/L$. On the other hand, the probability bounds in Theorems 1 and 2 worsen with increasing M.

With Λ, ω and M fixed, increasing the code length n (or, equivalently L) exponentially improves the probability bound in Theorem 1.

The per-iteration computational complexity of the AMP decoder is determined by the complexity of matrix-vector multiplications with the design matrix $\boldsymbol{A}$. This complexity is $O(nML)$ for Gaussian design matrices. For our empirical results in Section 6, we use DFT-based design matrices which reduce the per-iteration complexity to $O(ML\log(ML))$.

5.3 Compressed sensing with spatially coupled design matrices

In this section, we establish a nonasymptotic result analogous to Theorem 2 for compressed sensing with a spatially coupled measurement matrix. In compressed sensing, the goal is to estimate a vector $\boldsymbol{\beta} \in \mathbb{R}^p$ from a linear measurement $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$. This model is similar to SPARC decoding, with the main difference being that the entries of the signal vector $\boldsymbol{\beta}$ are now assumed to be drawn from a generic prior P_β rather than the section-wise structure of a SPARC message vector.

We consider a spatially coupled measurement matrix $\boldsymbol{A} \in \mathbb{R}^{n\times p}$ defined via a base matrix $\boldsymbol{W} \in \mathbb{R}^{\mathsf{R}\times\mathsf{C}}$ by replacing each entry W_{rc} by an $(n/\mathsf{R}) \times (p/\mathsf{C})$ block with i.i.d. $\mathcal{N}(0, \frac{W_{\mathsf{rc}}}{n/\mathsf{R}})$ entries, for $\mathsf{r} \in [\mathsf{R}]$ and $\mathsf{c} \in [\mathsf{C}]$. The AMP algorithm for spatially coupled compressed sensing has the same form as the one in (8)-(9), with the main difference being the denoising function η^t and the corresponding changes in the state evolution parameters. Before describing these differences, we state the assumptions on the model:

(1) The components of the signal vector $\boldsymbol{\beta} \in \mathbb{R}^p$ are i.i.d. with a sub-Gaussian distribution P_β.

(2) The denoising function $\eta^t : \mathbb{R}^p \to \mathbb{R}^p$ used in the AMP algorithm is separable, and its components $\eta^t_j : \mathbb{R} \to \mathbb{R}$ are Lipschitz continuous, for $j \in [p]$.

(3) As the signal dimension p grows, the sampling ratio n/p is constant and denoted by δ.

(4) The entries of the noise vector $\boldsymbol{w} \in \mathbb{R}^n$ are i.i.d. Gaussian with zero mean and variance σ^2.

(5) The entries of the base matrix $\boldsymbol{W} \in \mathbb{R}^{\mathsf{R}\times\mathsf{C}}$ are bounded below by a strictly positive constant, and for $\mathsf{r} \in [\mathsf{R}]$, the row sums satisfy $\frac{1}{2} \leq \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \leq 2$. (The assumption on the row sums is made to ensure that the definitions of the state evolution parameters are consistent with [39].)

The state evolution recursion for the spatially coupled compressed sensing AMP is as follows.

Initialize $\psi_{\mathsf{c}}^0 = 1$ for $\mathsf{c} \in [\mathsf{C}]$, and for $t \geq 0$:

$$\phi_{\mathsf{r}}^t = \sigma^2 + \frac{1}{\delta} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \psi_{\mathsf{c}}^t, \qquad \mathsf{r} \in [\mathsf{R}], \tag{44}$$

$$\psi_{\mathsf{c}}^{t+1} = \mathbb{E}\left\{ \left[\beta - \eta^t \left((\tau_{\mathsf{c}}^t)^{-1/2} \beta + G \right) \right]^2 \right\}, \qquad \mathsf{c} \in [\mathsf{C}], \tag{45}$$

where the expectation in (45) is over the independent pair $\beta \sim P_\beta$ and $G \sim \mathcal{N}(0,1)$. Furthermore, $\tau_{\mathsf{c}}^t = \left(\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}} / \phi_{\mathsf{r}}^t \right)^{-1}$ for $\mathsf{c} \in [\mathsf{C}]$.

The entries of the vector $\boldsymbol{\upsilon}^t$ and the matrix $\boldsymbol{S}$ in the AMP algorithm are defined as in (14) and (15), respectively. If the prior P_β is known, the Bayes optimal choice for scalar denoising function η^t is the MMSE estimator. Indeed, ψ_{c}^t is minimized by taking $\eta_j(s) = \mathbb{E}[\beta \,|\, (\tau_{\mathsf{c}}^t)^{-1/2}\beta + G = s]$ for an index j in column block c.

Theorem 3. *Consider the spatially coupled compressed sensing model under the assumptions listed above. For $t \geq 1$, the mean-squared error of the AMP satisfies*

$$P\Big(\Big| \frac{\|\boldsymbol{\beta}^t - \boldsymbol{\beta}\|^2}{p} - \frac{1}{\mathsf{C}} \sum_{\mathsf{c} \in [\mathsf{C}]} \psi_{\mathsf{c}}^t \Big| \geq \epsilon \Big) \leq K_t (\mathsf{RC})^t \exp\Big\{ \frac{-k_t n \epsilon^2}{[\mathsf{R}(\max_{r,\mathsf{c}} W_{r,\mathsf{c}})]^{2t}} \Big\}. \tag{46}$$

Here K_t, k_t are positive constants that depend only on t. (These constants are not the same as the ones in Theorem 2).

The proof of Theorem 3 is similar to that of Theorem 2, with appropriate changes (along the lines of the result in [44]) to account for the fact that $\boldsymbol{\beta}$ now has i.i.d. entries rather than the block-wise structure of a SPARC message vector. A few remarks about Theorem 3 and the underlying assumptions:

1. Theorem 3 refines the asymptotic result established in [40] for spatially coupled design matrices, which showed that $\lim_{p\to\infty} \|\boldsymbol{\beta}^t - \boldsymbol{\beta}\|^2/p = \frac{1}{\mathsf{C}} \sum_{\mathsf{c}} \psi_{\mathsf{c}}^t$ almost surely.
2. The assumption that the entries of the base matrix are lower bounded by a positive constant is due to a technical detail in the proof. For the same reason, we need ρ to be strictly positive (rather than 0) in Theorems 1 and 2.
3. The scaling of the base matrix entries (implied by the column sums condition in Assumption 5) differs from the one used for SPARCs in (5) by a factor of order R. Therefore, the maximum entry in the equivalent of the (ω, Λ, ρ) base matrix for compressed sensing would be of order $1/\omega$, leading to an exponent of $\frac{-k_t n \epsilon^2}{(\mathsf{R}/\omega)^{2t}}$ in (46). Thus the probability bound in Theorem 3 is similar to that of Theorem 2.

6 Empirical performance of SC-SPARCs

In this section, we investigate the finite length error performance of SC-SPARCs with AMP decoding via numerical simulations. We use the $(\omega, \Lambda, \rho = 0)$ base matrix construction in all the simulations.

We would like to compare the performance of SC-SPARCs with that of standard coded modulation schemes such as LDPC codes with Quadrature Amplitude Modulation (QAM), which produce complex-valued symbols to be transmitted over the channel. Therefore, in these simulations we consider the communication over the *complex* AWGN channel, where the noise is circularly-symmetric

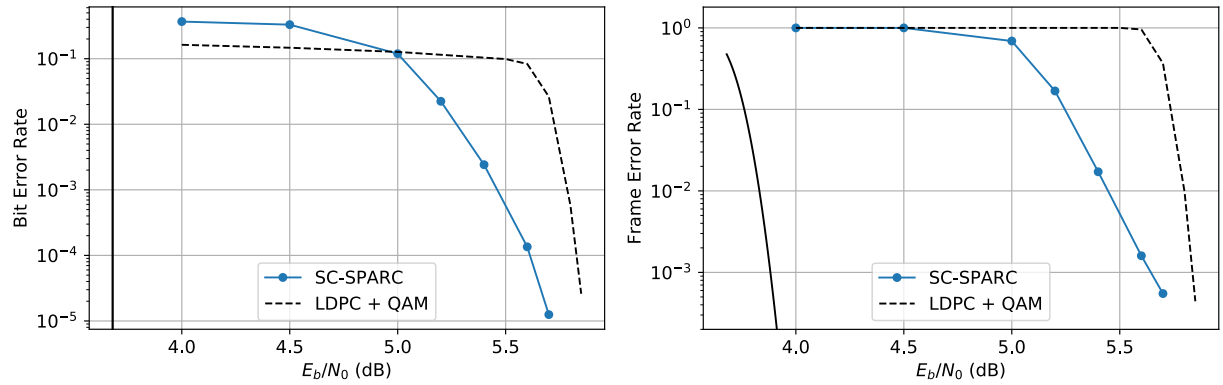

Figure 4: Error performance of complex SC-SPARCs defined via a $(\omega = 4, \Lambda = 32, \rho = 0)$ base matrix. $R = 1.5$ bits/dimension, $L = 2944$, $M = 2048$, code length $n = 10795$. The dashed lines show the performance of coded modulation: $(K = 32400, N = 64800)$ DVB-S2 LDPC + 64 QAM, frame length = 10800 symbols. The solid black line in the BER plot is the AWGN Shannon limit for $R = 1.5$ bits/dimension, and in the FER plot, it is the normal approximation to the AWGN finite length error probability bound in [46].

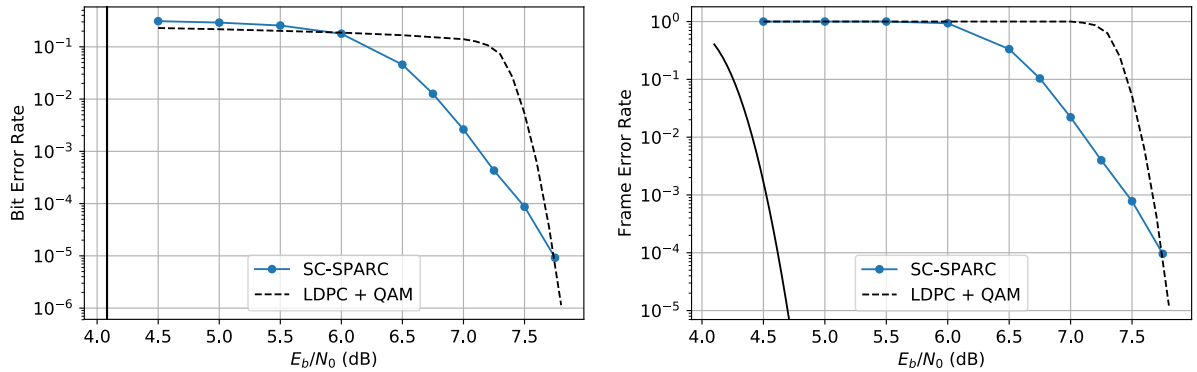

Figure 5: Error performance of complex SC-SPARCs defined via an $(\omega = 6, \Lambda = 32, \rho = 0)$ base matrix. $R = 1.6$ bits/dimension, $L = 960$, $M = 128$, code length $n = 2100$. The dashed lines show the performance of coded modulation: $(K = 6480, N = 16200)$ DVB-S2 LDPC + 256 QAM, frame length = 2025 symbols. The solid black line in the BER plot is the AWGN Shannon limit for $R = 1.6$ bits/dimension, and in the FER plot, it is the normal approximation to the AWGN finite length error probability bound in [46].

complex Gaussian. We use complex SC-SPARCs, which are defined as described in Section 2, except that the design matrix now has independent circularly-symmetric complex Gaussian entries instead of real-valued Gaussian entries. The AMP decoder for complex SC-SPARCs is similar to the one in (8)-(9): we take $\boldsymbol{A}^*$ to be the conjugate transpose of $\boldsymbol{A}$, and modify the definition of η_j^t in (16) according to (17). For additional details on complex SC-SPARCs and its AMP decoder, see [45].

In Figures 4, 5, and 6, we provide numerical simulation results demonstrating the finite length error performance of complex SC-SPARCs with AMP decoding at different code rates and code lengths. The error performance is evaluated using both the bit error rate (BER) and the frame error rate (FER). (The FER is the message/codeword error rate.) We also simulate and plot the error performance of coded modulation schemes (LDPC + QAM) for reference using the AFF3CT toolbox [47]. The LDPC codes are chosen from the DVB-S2 standard and a belief propagation

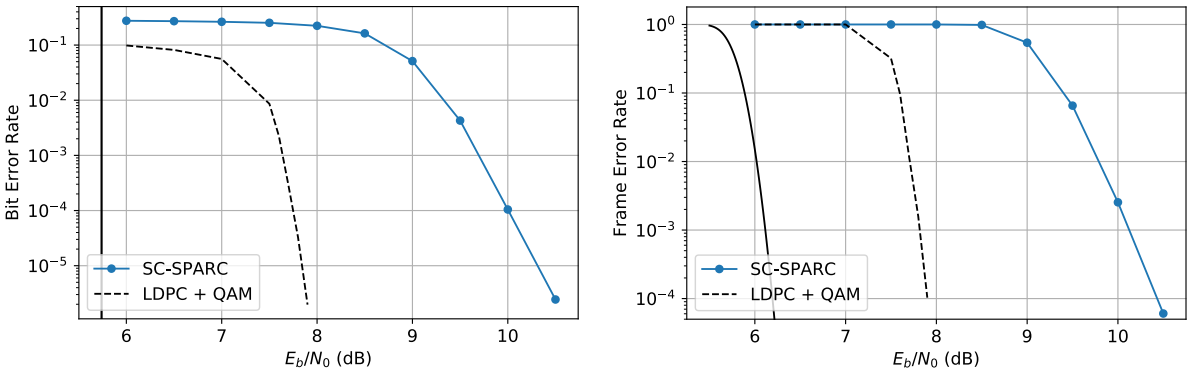

Figure 6: Error performance of complex SC-SPARCs defined via an $(\omega = 6, \Lambda = 32, \rho = 0)$ base matrix. $R = 2$ bits/dimension, $L = 2688$, $M = 16$, code length $n = 2688$. The dashed lines show the performance of coded modulation: $(K = 10800, N = 16200)$ DVB-S2 LDPC + 64 QAM, frame length = 2700 symbols. The solid black line in the BER plot is the AWGN Shannon limit for $R = 2$ bits/dimension, and in the FER plot, it is the normal approximation to the AWGN finite length error probability bound in [46].

(BP) decoder is used which runs for 50 iterations. For fair comparison, in each figure, the frame length of the coded modulation scheme is chosen to be close to the code length of the SC-SPARC.

Fig. 4 shows the performance of SC-SPARCs with rate 1.5 bits/dimension and code length $n = 10795$. The AMP decoder for the SC-SPARC is run for a maximum of 200 iterations (details in Sec. 6.1). The coded modulation scheme uses a rate $\frac{1}{2}$ $(32400, 64800)$ DVB-S2 LDPC code with 64-QAM modulation, for the same overall rate of 1.5 bits/dimension and a frame length of 10,800 symbols. We observe that the SC-SPARC requires a smaller E_b/N_0 to achieve BERs in the range 10^{-1} to 10^{-5}, and FERs down to 5×10^{-4} compared to the coded modulation scheme. However, for much lower FERs, we expect the coded modulation scheme to require a smaller E_b/N_0 because its frame error rate drops faster as E_b/N_0 increases.

Fig. 5 shows the performance of an SC-SPARC with a shorter code length $n = 2100$, and a rate of 1.6 bits/dimension. The AMP decoder for the SC-SPARC is run for a maximum of 100 iterations. The coded modulation scheme uses a rate $\frac{1}{2}$ (6480, 16200) DVB-S2 LDPC code with 256-QAM modulation, for the same overall rate of 1.6 bits/dimension and a frame length of 2025 symbols. We observe that the SC-SPARC requires a smaller E_b/N_0 to achieve BERs in the range 10^{-1} to 10^{-5} and FERs down to 10^{-4} compared to the coded modulation scheme. However, for BERs and FERs lower than 10^{-5} and 10^{-4}, respectively, we expect the coded modulation scheme to require a smaller E_b/N_0 because its error rate drops faster as E_b/N_0 increases.

In Fig. 6, the rate of the SC-SPARC is 2 bits/dimension and the code length is $n = 2688$. The AMP decoder for the SC-SPARC is run for a maximum of 100 iterations. The coded modulation scheme uses a rate $\frac{2}{3}$ (10800, 16200) DVB-S2 LDPC code with 64-QAM modulation, for the same overall rate of 2 bits/dimension and a frame length of 2700 symbols. We observe that the SC-SPARC has a higher BER and FER compared to the coded modulation scheme for all values of E_b/N_0, and its error rate also drops more slowly as E_b/N_0 increases.

In the above plots, the SC-SPARC parameters $(\omega, \Lambda, L, M, n)$ have not been carefully optimized. An interesting direction for future work is to develop good finite length design guidelines for choosing these parameters as a function of rate and snr. Another direction is to explore whether alternative base matrix designs could improve the finite length performance at higher rates like 2 bits/dimension.

6.1 Implementation details

The $(\omega, \Lambda, \rho = 0)$ base matrix was used for all the simulations. Furthermore, to reduce the decoding complexity and the memory required, a few modifications were made to the SC-SPARC construction and the AMP decoder:

1) DFT based design matrices: We replaced the complex Gaussian design matrix with a Discrete Fourier Transform (DFT) based design matrix. This enables the matrix-vector multiplications in the AMP decoder (8)–(9) to be computed via the Fast Fourier Transform (FFT), which significantly lowers the decoding complexity and memory requirement. Our approach is similar to that of [4, 9] where Hadamard based design matrices were used for real-valued SPARCs.

The computational complexity of the AMP decoder is dominated by the two matrix-vector multiplications associated with the design matrix $\boldsymbol{A}$. These operations have complexity $O(nLM)$ when $\boldsymbol{A}$ has independent Gaussian entries. The memory requirements of the encoder and decoder are also proportional to nLM since the Gaussian design matrix has to be stored. By constructing the design matrix using randomly sampled rows and columns of the (deterministic) DFT matrix, the complexity of the matrix-vector multiplications (replaced by FFTs) is reduced to $O(LM \log(LM))$, and the memory requirements of the encoder and decoder are proportional to ωLM. The error performance of DFT based design matrices was found to be similar to that of Gaussian matrices for large matrix sizes.

2) Online estimation of state evolution parameters: The AMP decoder in (8)–(9) contains parameters computed using the state evolution (SE) recursion (10)–(13). In particular, the vector $\boldsymbol{v}^t \in \mathbb{R}^n$, and the matrix $\overline{\boldsymbol{S}}^t \in \mathbb{R}^{n \times ML}$ are determined via SE parameters computed offline. Instead of computing the SE parameters offline, the SE parameters can be estimated online (at runtime) using the outputs of the AMP decoder in each iteration. The SE parameters $\{\sigma_{\mathsf{r}}^t\}_{\mathsf{r}\in[\mathsf{R}]}$, $\{\phi_{\mathsf{r}}^t\}_{\mathsf{r}\in[\mathsf{R}]}$ and $\{\tau_{\mathsf{c}}^t\}_{\mathsf{c}\in[\mathsf{C}]}$, which are needed to compute $\boldsymbol{v}^t$ and $\overline{\boldsymbol{S}}^t$ (see (14) and (15)) can be estimated online in the following way. For $\mathsf{r} \in [\mathsf{R}]$ and $\mathsf{c} \in [\mathsf{C}]$,

$$\widehat{\sigma}_{\mathsf{r}}^t = \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \left(1 - \frac{\|\boldsymbol{\beta}_{\mathsf{c}}^t\|^2}{L/\mathsf{C}} \right), \tag{47}$$

$$\widehat{\phi}_{\mathsf{r}}^t = \begin{cases} \sigma^2 + \widehat{\sigma}_{\mathsf{r}}^t & \text{if the decoder knows } \sigma^2, \\ \frac{\|\boldsymbol{z}_{\mathsf{r}}^t\|^2}{n/\mathsf{R}} & \text{otherwise,} \end{cases} \tag{48}$$

$$\widehat{\tau}_{\mathsf{c}}^t = \frac{L}{n} \left[\frac{1}{\mathsf{R}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \frac{W_{\mathsf{rc}}}{\widehat{\phi}_{\mathsf{r}}^t} \right]^{-1}. \tag{49}$$

The justification for these estimates comes from Lemma 7.6, which proves that the estimates $\widehat{\sigma}_{\mathsf{r}}^t, \widehat{\phi}_{\mathsf{r}}^t$ concentrate on $\sigma_{\mathsf{r}}^t, \phi_{\mathsf{r}}^t$, respectively (eqs. (94) and (103)), for large (n, L). We observe that using online estimates of the SE parameters results in a better error performance than using deterministic SE parameters. A similar improvement was observed in [5] for power allocated SPARCs.

3) Early stopping of AMP: Since the empirical estimates of SE parameters in (47)-(49) are estimates of certain noise variances related to the decoding error in each iteration of the AMP, we chose to stop the AMP decoder early if the change in $\widehat{\boldsymbol{\sigma}}^t, \widehat{\boldsymbol{\phi}}^t$ or $\widehat{\boldsymbol{\tau}}^t$ fell below a prescribed threshold over consecutive iterations. A similar stopping criterion was used in [5] to terminate the AMP decoder for power allocated SPARCs.

A Python implementation of SPARCs (both power allocated and spatially coupled) with AMP decoding is available at [48].

7 Proof of Theorem 2

The main ingredients in the proof of Theorem 2 are two technical lemmas (Lemmas 7.4 and 7.6). After laying down some definitions and notation, we give a brief overview of the proof in Section 7.2. We then state the key technical lemmas, and use them to prove Theorem 2. For consistency with earlier analyses of AMP, we use notation similar to [4,14,41], with modifications to account for the row- and column-block dependence induced due to spatial coupling.

7.1 Definitions and Preliminaries

Recall that $\boldsymbol{\beta}_0 \in \mathbb{R}^{ML}$ is the message vector chosen by the transmitter, $\boldsymbol{w} \in \mathbb{R}^n$ is the channel noise vector, and the AMP decoder is intialized with $\boldsymbol{\beta}^0 = \mathbf{0}$ and $\boldsymbol{z}^0 = \boldsymbol{y}$. Throughout the proof, we use the notation

$$N_{\mathsf{R}} := n/\mathsf{R}, \quad N_{\mathsf{C}} := ML/\mathsf{C}. \tag{50}$$

Define the column vectors $\boldsymbol{h}^{t+1}, \breve{\boldsymbol{q}}^{t+1} \in \mathbb{R}^{ML}$ and $\boldsymbol{b}^t, \breve{\boldsymbol{m}}^t \in \mathbb{R}^n$ for $t \geq 1$ recursively as follows. Starting with the initial conditions

$$\begin{aligned} \breve{\boldsymbol{q}}^0 &= -\boldsymbol{\beta}_0, \qquad & \boldsymbol{h}^1 &= \boldsymbol{\beta}_0 - (\overline{\boldsymbol{S}}^0 \odot \boldsymbol{A})^* \boldsymbol{z}^0, \\ \breve{\boldsymbol{m}}^0 &= -\boldsymbol{z}^0, & \boldsymbol{b}^0 &= \boldsymbol{w} - \boldsymbol{z}^0, \end{aligned} \tag{51}$$

for $t \geq 1$, the vectors $\breve{\boldsymbol{q}}^t, \boldsymbol{h}^{t+1}, \breve{\boldsymbol{m}}^t, \boldsymbol{b}^t$ are defined as

$$\begin{aligned} \breve{\boldsymbol{q}}^t &= \boldsymbol{\beta}^t - \boldsymbol{\beta}_0, \qquad & \boldsymbol{h}^{t+1} &:= \boldsymbol{\beta}_0 - \Big([(\overline{\boldsymbol{S}}^t \odot \boldsymbol{A})^* \boldsymbol{z}^t] + \boldsymbol{\beta}^t \Big), \\ \breve{\boldsymbol{m}}^t &= -\boldsymbol{z}^t, & \boldsymbol{b}^t &= \boldsymbol{w} - \boldsymbol{z}^t, \end{aligned} \tag{52}$$

where $\overline{\boldsymbol{S}}^t \in \mathbb{R}^{n \times ML}$ is the matrix with entries defined in (15). For notational convenience, we define the matrix $\boldsymbol{S}^t \in \mathbb{R}^{\mathsf{R} \times \mathsf{C}}$ with entries

$$S^t_{\mathsf{rc}} = \tau^t_{\mathsf{c}} / \phi^t_{\mathsf{r}}, \quad \mathsf{r} \in [\mathsf{R}],\ \mathsf{c} \in [\mathsf{C}]. \tag{53}$$

We define a modified design matrix $\mathbf{A} \in \mathbb{R}^{n \times ML}$ having entries given by

$$\mathsf{A}_{ij} = \frac{A_{ij}}{\sqrt{W_{\mathsf{r}(i)\mathsf{c}(j)}}}, \qquad i \in [n],\ j \in [ML]. \tag{54}$$

Since $A_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{L} W_{\mathsf{r}(i)\mathsf{c}(j)})$, the modified matrix has entries $\mathsf{A}_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{L})$. We note that $\boldsymbol{A} = \sqrt{\widetilde{\boldsymbol{W}}} \odot \mathbf{A}$ where $\widetilde{\boldsymbol{W}} \in \mathbb{R}^{n \times ML}$ is the matrix with entries $\widetilde{W}_{ij} = W_{\mathsf{r}(i)\mathsf{c}(j)}$.

Using the definitions (51)–(54) in the AMP update equations (8)-(9), we find that the following block-wise relationships are satisfied for $t \geq 0$:

$$\begin{aligned} \boldsymbol{b}^t_{\mathsf{r}} - \frac{\sigma^t_{\mathsf{r}}}{\phi^{t-1}_{\mathsf{r}}} \breve{\boldsymbol{m}}^{t-1}_{\mathsf{r}} &= [(\sqrt{\widetilde{\boldsymbol{W}}} \odot \mathbf{A})\, \breve{\boldsymbol{q}}^t]_{\mathsf{r}} = \sum_{\mathsf{c} \in [\mathsf{C}]} \sqrt{W_{\mathsf{rc}}}\, \mathbf{A}_{\mathsf{rc}}\, \breve{\boldsymbol{q}}^t_{\mathsf{c}}, \qquad \text{for } \mathsf{r} \in [\mathsf{R}], \\ \boldsymbol{h}^{t+1}_{\mathsf{c}} + \breve{\boldsymbol{q}}^t_{\mathsf{c}} &= [(\overline{\boldsymbol{S}}^t \odot \sqrt{\widetilde{\boldsymbol{W}}} \odot \mathbf{A})^* \breve{\boldsymbol{m}}^t]_{\mathsf{r}} = \sum_{\mathsf{r} \in [\mathsf{R}]} S^t_{\mathsf{rc}} \sqrt{W_{\mathsf{rc}}}\, (\mathbf{A}_{\mathsf{rc}})^* \breve{\boldsymbol{m}}^t_{\mathsf{r}}, \qquad \text{for } \mathsf{c} \in [\mathsf{C}]. \end{aligned} \tag{55}$$

We define complementary quantities for the $\breve{\boldsymbol{m}}^t$ and $\breve{\boldsymbol{q}}^t$ vectors that will be useful in the conditional distribution lemma that follows. For $t \geq 0$ and $\mathsf{r} \in [\mathsf{R}]$ and $\mathsf{c} \in [\mathsf{C}]$, let

$$\boldsymbol{m}^{t,\mathsf{c}} = \begin{bmatrix} S_{1\mathsf{c}}^t \sqrt{W_{1\mathsf{c}}}\, \breve{\boldsymbol{m}}_{\mathsf{r}=1}^t \\ S_{2\mathsf{c}}^t \sqrt{W_{2\mathsf{c}}}\, \breve{\boldsymbol{m}}_{\mathsf{r}=2}^t \\ \vdots \\ S_{\mathsf{R}\mathsf{c}}^t \sqrt{W_{\mathsf{R}\mathsf{c}}}\, \breve{\boldsymbol{m}}_{\mathsf{r}=\mathsf{R}}^t \end{bmatrix} \in \mathbb{R}^{n \times 1} \quad \text{and} \quad \boldsymbol{q}^{t,\mathsf{r}} = \begin{bmatrix} \sqrt{W_{\mathsf{r}1}}\, \breve{\boldsymbol{q}}_{\mathsf{c}=1}^t \\ \sqrt{W_{\mathsf{r}2}}\, \breve{\boldsymbol{q}}_{\mathsf{c}=2}^t \\ \vdots \\ \sqrt{W_{\mathsf{r}\mathsf{C}}}\, \breve{\boldsymbol{q}}_{\mathsf{c}=\mathsf{C}}^t \end{bmatrix} \in \mathbb{R}^{ML \times 1}. \tag{56}$$

(Here $\breve{\boldsymbol{m}}_{\mathsf{r}=1}^t \in \mathbb{R}^{N_\mathsf{R}}$ refers to the first row block of $\breve{\boldsymbol{m}}^t$, and $\breve{\boldsymbol{q}}_{\mathsf{c}=1}^t \in \mathbb{R}^{N_\mathsf{C}}$ to the first column block of $\breve{\boldsymbol{q}}^t$.) A word about the notation before we proceed: when a row- or column-block index (r or c) appears as a subscript of a vector or a matrix (e.g., as in $\mathbf{A}_{\mathsf{r}\mathsf{c}}$), it denotes the corresponding block of that vector/matrix, but when a row- or column-block index appears as a *superscript* of a vector/matrix (e.g., $\boldsymbol{m}^{t,\mathsf{c}}$ and $\boldsymbol{q}^{t,\mathsf{r}}$), it denotes the dependence of the vector/matrix on that index.

Using the vectors defined in (56), we can rewrite (55) as

$$\begin{aligned} \boldsymbol{b}_\mathsf{r}^t - \frac{\sigma_\mathsf{r}^t}{\phi_\mathsf{r}^{t-1}} \breve{\boldsymbol{m}}_\mathsf{r}^{t-1} &= [\mathbf{A}\, \boldsymbol{q}^{t,\mathsf{r}}]_\mathsf{r} = \sum_{\mathsf{c} \in [\mathsf{C}]} \mathbf{A}_{\mathsf{r}\mathsf{c}}\, \boldsymbol{q}_\mathsf{c}^{t,\mathsf{r}} \quad \text{for } \mathsf{r} \in [\mathsf{R}], \\ \boldsymbol{h}_\mathsf{c}^{t+1} + \breve{\boldsymbol{q}}_\mathsf{c}^t &= [\mathbf{A}^* \boldsymbol{m}^{t,\mathsf{c}}]_\mathsf{c} = \sum_{\mathsf{r} \in [\mathsf{R}]} [\mathbf{A}_{\mathsf{r}\mathsf{c}}]^* \boldsymbol{m}_\mathsf{r}^{t,\mathsf{c}} \quad \text{for } \mathsf{c} \in [\mathsf{C}]. \end{aligned} \tag{57}$$

It will be useful to write the equations in (57) in matrix form. For this, we define the following matrices for $t \geq 1$:

$$\begin{aligned} \breve{\boldsymbol{Q}}_t &:= [\breve{\boldsymbol{q}}^0 \mid \ldots \mid \breve{\boldsymbol{q}}^{t-1}] \in \mathbb{R}^{ML \times t}, \\ \boldsymbol{H}_t &:= [\boldsymbol{h}^1 | \ldots | \boldsymbol{h}^t] \in \mathbb{R}^{ML \times t}, \\ \boldsymbol{X}_t &:= [\boldsymbol{h}^1 + \breve{\boldsymbol{q}}^0 \mid \boldsymbol{h}^2 + \breve{\boldsymbol{q}}^1 \mid \ldots \mid \boldsymbol{h}^t + \breve{\boldsymbol{q}}^{t-1}] \in \mathbb{R}^{ML \times t}, \\ \breve{\boldsymbol{M}}_t &:= [\breve{\boldsymbol{m}}^0 \mid \ldots \mid \breve{\boldsymbol{m}}^{t-1}] \in \mathbb{R}^{n \times t}, \\ \boldsymbol{B}_t &:= [\boldsymbol{b}^0 | \ldots | \boldsymbol{b}^{t-1}] \in \mathbb{R}^{n \times t} \\ \boldsymbol{Y}_t &:= [\boldsymbol{b}^0 \mid \boldsymbol{b}^1 - \boldsymbol{v}^1 \odot \breve{\boldsymbol{m}}^0 \mid \ldots \mid \boldsymbol{b}^{t-1} - \boldsymbol{v}^{t-1} \odot \breve{\boldsymbol{m}}^{t-2}] \in \mathbb{R}^{n \times t}, \\ \boldsymbol{\Upsilon}_t^\mathsf{r} &:= \operatorname{diag}\left(0, \frac{\sigma_\mathsf{r}^1}{\phi_\mathsf{r}^0}, \ldots, \frac{\sigma_\mathsf{r}^{t-1}}{\phi_\mathsf{r}^{t-2}}\right) \in \mathbb{R}^{t \times t}. \end{aligned} \tag{58}$$

In the equations above, the notation $[\boldsymbol{a}_1 \mid \boldsymbol{a}_2 \mid \ldots \mid \boldsymbol{a}_k]$ is used to denote a matrix with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$. We also recall from (14) that $v_i^t = \sigma_\mathsf{r}^t / \phi_\mathsf{r}^{t-1}$ for $t \geq 1$ if $i \in [n]$ is in row block r. For $t = 0$, the matrices above are all defined as all-zeros. For $\mathsf{c} \in [\mathsf{C}]$, we define $\breve{\boldsymbol{Q}}_{t,\mathsf{c}}, \boldsymbol{X}_{t,\mathsf{c}}, \boldsymbol{H}_{t,\mathsf{c}} \in \mathbb{R}^{N_\mathsf{C} \times t}$ to correspond only to rows c of the corresponding matrix. We similarly define $\breve{\boldsymbol{M}}_{t,\mathsf{r}}, \boldsymbol{B}_{t,\mathsf{r}}, \boldsymbol{Y}_{t,\mathsf{r}} \in \mathbb{R}^{N_\mathsf{R} \times t}$ for $\mathsf{r} \in [\mathsf{R}]$. Using these definitions we have

$$\boldsymbol{Y}_{t,\mathsf{r}} = \boldsymbol{B}_{t,\mathsf{r}} - [\boldsymbol{0} | \breve{\boldsymbol{M}}_{t-1,\mathsf{r}}] \boldsymbol{\Upsilon}_t^\mathsf{r}, \qquad \boldsymbol{X}_{t,\mathsf{c}} = \boldsymbol{H}_{t,\mathsf{c}} + \breve{\boldsymbol{Q}}_{t,\mathsf{c}}. \tag{59}$$

Let $\boldsymbol{Q}_0^\mathsf{r}$ and $\boldsymbol{M}_0^\mathsf{c}$ be all-zero vectors. For $t \geq 1$ and $\mathsf{c} \in [\mathsf{C}]$, $\mathsf{r} \in [\mathsf{R}]$, let

$$\boldsymbol{M}_t^\mathsf{c} = [\boldsymbol{m}^{0,\mathsf{c}} | \ldots | \boldsymbol{m}^{t-1,\mathsf{c}}] \in \mathbb{R}^{n \times t} \quad \text{and} \quad \boldsymbol{Q}_t^\mathsf{r} = [\boldsymbol{q}^{0,\mathsf{r}} | \ldots | \boldsymbol{q}^{t-1,\mathsf{r}}] \in \mathbb{R}^{ML \times t}. \tag{60}$$

With this notation, we can compactly write (57) for $t \geq 1$ as

$$\left[\mathbf{A} \boldsymbol{Q}_t^\mathsf{r}\right]_{(\mathsf{r},\cdot)} = \boldsymbol{Y}_{t,\mathsf{r}}, \quad \mathsf{r} \in [\mathsf{R}], \qquad \text{and} \qquad \left[\mathbf{A}^* \boldsymbol{M}_t^\mathsf{c}\right]_{(\mathsf{c},\cdot)} = \boldsymbol{X}_{t,\mathsf{c}}, \quad \mathsf{c} \in [\mathsf{C}]. \tag{61}$$

Here the subscript $(\mathsf{r},\cdot)$ on a matrix denotes the rth block of rows of the matrix.

We use the notation $\boldsymbol{m}_{\|}^{t,\mathsf{c}}$ and $\boldsymbol{q}_{\|}^{t,\mathsf{r}}$ to denote the projection of $\boldsymbol{m}^{t,\mathsf{c}}$ and $\boldsymbol{q}^{t,\mathsf{r}}$ onto the column space of $\boldsymbol{M}_t^{\mathsf{c}}$ and $\boldsymbol{Q}_t^{\mathsf{r}}$, respectively. Let

$$\boldsymbol{\alpha}^{t,\mathsf{c}} := (\alpha_0^{t,\mathsf{c}},\dots,\alpha_{t-1}^{t,\mathsf{c}})^*, \qquad \boldsymbol{\gamma}^{t,\mathsf{r}} := (\gamma_0^{t,\mathsf{r}},\dots,\gamma_{t-1}^{t,\mathsf{r}})^* \tag{62}$$

be the coefficient vectors of these projections, i.e.,

$$\boldsymbol{m}_{\|}^{t,\mathsf{c}} := \sum_{i=0}^{t-1} \alpha_i^{t,\mathsf{c}} \boldsymbol{m}^{i,\mathsf{c}}, \qquad \boldsymbol{q}_{\|}^{t,\mathsf{r}} := \sum_{i=0}^{t-1} \gamma_i^{t,\mathsf{r}} \boldsymbol{q}^{i,\mathsf{r}}. \tag{63}$$

Writing $\mathbf{P}^{\|}_{\boldsymbol{M}_t^{\mathsf{c}}}$ for the orthogonal projection matrix onto the column space of $\boldsymbol{M}_t^{\mathsf{c}}$, we have $\boldsymbol{m}_{\|}^{t,\mathsf{c}} = \mathbf{P}^{\|}_{\boldsymbol{M}_t^{\mathsf{c}}} \boldsymbol{m}^{t,\mathsf{c}} = \boldsymbol{M}_t^{\mathsf{c}}((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}$ and so $\boldsymbol{\alpha}^{t,\mathsf{c}} = ((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}$. (If the columns of $\boldsymbol{M}_t^{\mathsf{c}}$ are linearly dependent, $((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{M}_t^{\mathsf{c}})^*$ is interpreted as the pseudoinverse of $\boldsymbol{M}_t^{\mathsf{c}}$.) We can similarly write $\boldsymbol{\gamma}^{t,\mathsf{r}} = ((\boldsymbol{Q}_t^{\mathsf{r}})^*\boldsymbol{Q}_t^{\mathsf{r}})^{-1}(\boldsymbol{Q}_t^{\mathsf{r}})^*\boldsymbol{q}^{t,\mathsf{r}}$. The projections of $\boldsymbol{m}^{t,\mathsf{c}}$ and $\boldsymbol{q}^{t,\mathsf{r}}$ onto the orthogonal complements of $\boldsymbol{M}_t^{\mathsf{c}}$ and $\boldsymbol{Q}_t^{\mathsf{r}}$, respectively, are denoted by

$$\boldsymbol{m}_{\perp}^{t,\mathsf{c}} := \boldsymbol{m}^{t,\mathsf{c}} - \boldsymbol{m}_{\|}^{t,\mathsf{c}}, \qquad \boldsymbol{q}_{\perp}^{t,\mathsf{r}} := \boldsymbol{q}^{t,\mathsf{r}} - \boldsymbol{q}_{\|}^{t,\mathsf{r}}. \tag{64}$$

In Lemma 7.6, we show that the entries of $\boldsymbol{\alpha}^{t,\mathsf{c}}$ and $\boldsymbol{\gamma}^{t,\mathsf{r}}$ concentrate around constants. We now specify these constants. For $\mathsf{c} \in [\mathsf{C}]$, $\mathsf{r} \in [\mathsf{R}]$, define matrices $\widetilde{\boldsymbol{C}}^{t,\mathsf{r}}, \breve{\boldsymbol{C}}^{t,\mathsf{c}} \in \mathbb{R}^{t\times t}$ for $t \geq 1$ such that

$$\widetilde{C}^{t,\mathsf{r}}_{i+1,j+1} = \sigma_{\mathsf{r}}^{\max(i,j)}, \quad \text{and} \quad \breve{C}^{t,\mathsf{c}}_{i+1,j+1} = \frac{n}{L}\tau_{\mathsf{c}}^{\max(i,j)}, \quad 0 \leq i,j \leq t-1. \tag{65}$$

The concentrating values for $\boldsymbol{\gamma}^{t,\mathsf{r}}$ and $\boldsymbol{\alpha}^{t,\mathsf{c}}$ are

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}^{t,\mathsf{r}} &:= \sigma_{\mathsf{r}}^t(\widetilde{\boldsymbol{C}}^{t,\mathsf{r}})^{-1}(1,\dots,1)^* \overset{(a)}{=} (0,\dots,0,\sigma_{\mathsf{r}}^t/\sigma_{\mathsf{r}}^{t-1})^* \in \mathbb{R}^t, \\ \widehat{\boldsymbol{\alpha}}^{t,\mathsf{c}} &:= \frac{n}{L}\tau_{\mathsf{c}}^t(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}(1,\dots,1)^* \overset{(b)}{=} (0,\dots,0,\tau_{\mathsf{c}}^t/\tau_{\mathsf{c}}^{t-1})^* \in \mathbb{R}^t. \end{aligned} \tag{66}$$

To see that (a) holds, we observe that $(\widetilde{\boldsymbol{C}}^{t,\mathsf{r}})^{-1}\widetilde{\boldsymbol{C}}^{t,\mathsf{r}} = \boldsymbol{I}_t$ implies that $(\widetilde{\boldsymbol{C}}^{t,\mathsf{r}})^{-1}(\sigma_{\mathsf{r}}^{t-1},\dots,\sigma_{\mathsf{r}}^{t-1})^* = (0,\dots,0,1)^* \in \mathbb{R}^t$. The equality (b) is obtained similarly.

Let $\sigma_{\perp,\mathsf{r}}^0 := \sigma_{\mathsf{r}}^0$ and $\tau_{\perp,\mathsf{c}}^0 := \tau_{\mathsf{c}}^0$, and for $t \geq 1$ define

$$\sigma_{\perp,\mathsf{r}}^t := \sigma_{\mathsf{r}}^t\Big(1 - \frac{\sigma_{\mathsf{r}}^t}{\sigma_{\mathsf{r}}^{t-1}}\Big), \quad \text{and} \quad \tau_{\perp,\mathsf{c}}^t := \tau_{\mathsf{c}}^t\Big(1 - \frac{\tau_{\mathsf{c}}^t}{\tau_{\mathsf{c}}^{t-1}}\Big). \tag{67}$$

Lemma 7.1. *Under the assumptions of Prop. 4.1, for sufficiently large M, the constants $\sigma_{\perp,\mathsf{r}}^t$ and $\frac{n}{L}\tau_{\perp,\mathsf{c}}^t$ are bounded below for $0 \leq k < T$:*

$$\sigma_{\perp,r}^t \geq \overline{C}_1\left(\frac{\omega}{\Lambda}\right)^2, \quad r \in [R], \qquad \tau_{\perp,\mathsf{c}}^t \geq \overline{C}_2\left(\frac{\omega}{\Lambda}\right), \quad \mathsf{c} \in [\mathsf{C}], \tag{68}$$

where

$$\overline{C}_1 = \left(1 + \frac{1}{\vartheta \mathsf{snr}}\right)^2 \frac{P\rho^2}{\vartheta \mathsf{snr}^2}\,\Delta^2, \qquad \overline{C}_2 = \frac{\rho(1+\vartheta \mathsf{snr})}{\vartheta R}\Delta, \tag{69}$$

where Δ is the rate gap defined in (31).

Proof. In Appendix A.3. ■

Lemma 7.2. *If the $\sigma^t_{\perp,r}$ and $\frac{n}{L}\tau^t_{\perp,\mathsf{c}}$ are bounded below by some positive constants for $0 \le k < T$, then the matrices $\widetilde{\boldsymbol{C}}^{k,r}$ and $\breve{\boldsymbol{C}}^{k,\mathsf{c}}$ defined in (65) are invertible for $1 \le k \le T$.*

Proof. The proof can be found in [44, Lemma 2]. ■

We will use the following notation. Given two random vectors $\boldsymbol{x}_1, \boldsymbol{x}_2$ and a sigma-algebra $\mathscr{S}$, $\boldsymbol{x}_1|_{\mathscr{S}} \overset{d}{=} \boldsymbol{x}_2$ denotes that the conditional distribution of $\boldsymbol{x}_1$ given $\mathscr{S}$ equals the distribution of $\boldsymbol{x}_2$.

7.2 Outline of the proof of Theorem 2

Theorem 2 gives a concentration inequality for the normalized MSE of the AMP decoder, which for iteration $(t+1)$ can be written as

$$\frac{1}{L}\|\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}\|^2 = \frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} \|\eta^t_{\mathsf{c}}(\boldsymbol{\beta}_{\mathsf{c}} - \boldsymbol{h}^{t+1}_{\mathsf{c}}) - \boldsymbol{\beta}_{\mathsf{c}}\|^2, \tag{70}$$

where we recall that $\boldsymbol{h}^{t+1}_{\mathsf{c}}$ is the cth column block of $\boldsymbol{h}^{t+1} \in \mathbb{R}^{ML}$. (We choose iteration $(t+1)$ rather than t for notational convenience.)

The proof of the theorem is based on showing that for $t \ge 0$, the vector $\boldsymbol{h}^{t+1}_{\mathsf{c}}$ is approximately Gaussian, for $\mathsf{c} \in [\mathsf{C}]$. In particular, we show that $\boldsymbol{h}^{t+1}_{\mathsf{c}}$ is approximately distributed as $\sqrt{\tau^t_{\mathsf{c}}}\widetilde{\boldsymbol{Z}}_{t,\mathsf{c}}$, where $\widetilde{\boldsymbol{Z}}_{t,\mathsf{c}}$ is standard Gaussian and independent across $\mathsf{c} \in [\mathsf{C}]$. If we assume that $\boldsymbol{h}^{t+1}_{\mathsf{c}}$ is *exactly* distributed as $\sqrt{\tau^t_{\mathsf{c}}}\widetilde{\boldsymbol{Z}}_{t,\mathsf{c}}$ for $\mathsf{c} \in [\mathsf{C}]$, then obtaining a concentration inequality for the MSE in (70) is straightforward. Indeed, for a fixed $\boldsymbol{\beta}$, the MSE $\frac{1}{L}\sum_{\mathsf{c}} \|\eta^t_{\mathsf{c}}(\boldsymbol{\beta}_{\mathsf{c}} - \boldsymbol{h}^{t+1}_{\mathsf{c}}) - \boldsymbol{\beta}_{\mathsf{c}}\|^2$ is a bounded and Lipschitz function of $\boldsymbol{h}^{t+1}$. Therefore, if $\boldsymbol{h}^{t+1}$ were Gaussian, one could obtain a concentration inequality for the MSE of the AMP decoder via standard Gaussian concentration results [49]. The bulk of the technical work is in precisely quantifying and controlling the deviation from Gaussianity of the vectors $\boldsymbol{h}^{t+1}$, for $t \ge 0$.

To study the distribution of $\boldsymbol{h}^{t+1}$, we use the recursion in (57), or equivalently, the matrix version in (61). Note that (57) is a restatement of the dynamics of the AMP algorithm, although AMP cannot be run this way in practice (since it is initialized with $\breve{\boldsymbol{q}}^0 = -\boldsymbol{\beta}_0$ which is unknown). The first ingredient in the proof is Lemma 7.4, which specifies the conditional distribution of $\boldsymbol{b}^t_{\mathsf{r}}$ and $\boldsymbol{h}^{t+1}_{\mathsf{c}}$ given the past iterates of the algorithm in (57), for $\mathsf{r} \in [\mathsf{R}], \mathsf{c} \in [\mathsf{C}]$. More precisely, for $t \ge 0$, the lemma specifies the conditional distribution of $\boldsymbol{b}^t_{\mathsf{r}}|_{\mathscr{S}_{t,t}}$ and $\boldsymbol{h}^{t+1}_{\mathsf{c}}|_{\mathscr{S}_{t+1,t}}$, where $\mathscr{S}_{t_a,t}$ is the sigma-algebra generated by the collection of vectors

$$\boldsymbol{b}^0, ..., \boldsymbol{b}^{t_a-1}, \breve{\boldsymbol{m}}^0, ..., \breve{\boldsymbol{m}}^{t_a-1}, \boldsymbol{h}^1, ..., \boldsymbol{h}^t, \breve{\boldsymbol{q}}^0, ..., \breve{\boldsymbol{q}}^t, \text{ and } \boldsymbol{\beta}_0, \boldsymbol{w}. \tag{71}$$

Lemmas 7.4 and 7.5 together show that the conditional distributions have the following form:

$$\boldsymbol{b}^t_{\mathsf{r}}|_{\mathscr{S}_{t,t}} \overset{d}{=} \sum_{i=0}^{t} \frac{\sigma^t_{\mathsf{r}}}{\sigma^i_{\mathsf{r}}} \left(\sqrt{\sigma^i_{\perp,\mathsf{r}}}\boldsymbol{Z}'_{i,\mathsf{r}} + \boldsymbol{\Delta}_{i,i,\mathsf{r}}\right), \quad \mathsf{r} \in [\mathsf{R}], \tag{72}$$

$$\boldsymbol{h}^{t+1}_{\mathsf{c}}|_{\mathscr{S}_{t+1,t}} \overset{d}{=} \sum_{i=0}^{t} \frac{\tau^t_{\mathsf{c}}}{\tau^i_{\mathsf{c}}} \left(\sqrt{\tau^i_{\perp,\mathsf{c}}}\boldsymbol{Z}_{i,\mathsf{c}} + \boldsymbol{\Delta}_{i+1,i,\mathsf{c}}\right), \quad \mathsf{c} \in [\mathsf{C}]. \tag{73}$$

Here, $\boldsymbol{Z}'_i = [\boldsymbol{Z}_{i,1}, \ldots, \boldsymbol{Z}_{i,\mathsf{R}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_n)$ is independent of the sigma algebra $\mathscr{S}_{i,i}$, and $\boldsymbol{Z}_i = [\boldsymbol{Z}_{i,1}, \ldots, \boldsymbol{Z}_{i,\mathsf{C}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_{ML})$ is independent of $\mathscr{S}_{i+1,i}$, for $0 \leq i \leq t$. The deviation vectors $\boldsymbol{\Delta}_{i,i} = [\boldsymbol{\Delta}_{i,i,1}, \ldots, \boldsymbol{\Delta}_{i,i,\mathsf{R}}]^*$ and $\boldsymbol{\Delta}_{i+1,i} = [\boldsymbol{\Delta}_{i+1,i,1}, \ldots, \boldsymbol{\Delta}_{i+1,i,\mathsf{C}}]^*$ are measurable with respect to the sigma algebras $\mathscr{S}_{i,i}$ and $\mathscr{S}_{i+1,i}$, respectively. Their precise definitions are given in Lemma 7.4.

If we ignore the deviation terms in (73), then $\boldsymbol{h}^{t+1}$ would be an i.i.d. Gaussian vector with the variance of the entries equal to $\tau_{\mathsf{c}}^t \sum_{i=1}^t \tau_{\perp,\mathsf{c}}^i / (\tau_{\mathsf{c}}^i)^2 = \tau_{\mathsf{c}}^t$. (The equality can be seen by using the definition of $\tau_{\perp,\mathsf{c}}^i$ in (67).) In this case, a concentration inequality for the MSE could be directly obtained using standard concentration results, as described above.

The deviation terms in (72) and (73) are controlled via results in Lemma 7.6, specifically the concentration results in (92), (101) and (102). The definitions of the terms $\boldsymbol{\Delta}_{t,t,\mathsf{r}}$ and $\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}$ (see (81) and (82)) involve a combination of vectors that are measurable with respect to $\mathscr{S}_{t,t}$ and $\mathscr{S}_{t+1,t}$, respectively. We need several concentration results for scalar products involving these vectors in order to show that the deviation terms are negligible. Lemma 7.6 lists all the required concentration results, which are proved using an induction based argument in Section 8.3.

7.3 Conditional distribution lemma

For $t \geq 1$ and $t_a \in \{t, t+1\}$, we recall that $\mathscr{S}_{t_a,t}$ be the sigma-algebra generated by the collection of vectors in (71). Furthermore, let $\mathscr{S}_{0,0}$ and $\mathscr{S}_{1,0}$ be the sigma-algebras generated by $\{\breve{\boldsymbol{q}}^0, \boldsymbol{\beta}_0, \boldsymbol{w}\}$ and $\{\boldsymbol{b}^0, \boldsymbol{m}^0, \breve{\boldsymbol{q}}^0, \boldsymbol{\beta}_0, \boldsymbol{w}\}$, respectively. Given the vectors in (71), the vectors $\boldsymbol{m}^{0,\mathsf{c}}, ..., \boldsymbol{m}^{t_a-1,\mathsf{c}}, \boldsymbol{q}^{0,\mathsf{r}}, ..., \boldsymbol{q}^{t,\mathsf{r}}$ are determined via (56) for $\mathsf{r} \in [\mathsf{R}]$, $\mathsf{c} \in [\mathsf{C}]$.

The conditional distribution on $\mathbf{A}$ given $\mathscr{S}_{t_a,t}$ is the same as the conditional distribution given the following linear constraints:

$$[\mathbf{A}\boldsymbol{Q}_{t_a}^{\mathsf{r}}]_{(\mathsf{r},\cdot)} = \boldsymbol{Y}_{t_a,\mathsf{r}}, \quad \mathsf{r} \in [\mathsf{R}], \qquad [\mathbf{A}^* \boldsymbol{M}_t^{\mathsf{c}}]_{(\mathsf{c},\cdot)} = \boldsymbol{X}_{t,\mathsf{c}}, \quad \mathsf{c} \in [\mathsf{C}]. \tag{74}$$

where $\boldsymbol{Y}_{t_a,\mathsf{r}}, \boldsymbol{X}_{t,\mathsf{c}}$ are defined in (59), and $\boldsymbol{Q}_{t_a}^{\mathsf{r}}$, $\boldsymbol{M}_t^{\mathsf{c}}$ in (60). When conditioning on the linear constraints in (74), we emphasize that only $\mathbf{A}$ is treated as random. In the following lemma, we characterize the conditional distributions of the vectors $[\mathbf{A}^* \boldsymbol{m}^{t,\mathsf{c}}]_{\mathsf{c}}|_{\mathscr{S}_{t+1,t}}$ and $[\mathbf{A}\, \boldsymbol{q}^{t,\mathsf{r}}]_{\mathsf{r}}|_{\mathscr{S}_{t,t}}$. This result is then used in Lemma 7.4 to compute the conditional distributions of $\boldsymbol{b}^t|_{\mathscr{S}_{t,t}}$ and $\boldsymbol{h}^{t+1}|_{\mathscr{S}_{t+1,t}}$.

We write $\mathbf{A}_{\mathsf{rc}}$ for the $(\mathsf{r},\mathsf{c})^{th}$ block of $\mathbf{A}$, $\mathbf{A}_{(\mathsf{r},\cdot)}$ for the r^{th} block of rows, and $\mathbf{A}_{(\cdot,\mathsf{c})}$ for the c^{th} block of columns of $\mathbf{A}$. We also recall that when r or c are used as a subscript, it refers to a row or column block of a larger vector, whereas when r or c are used as a superscript it represents a 'full' vector whose entires depend on row block r or column index c. For a projection matrix $\mathbf{P}^{\|} \in \mathbb{R}^{ML \times ML}$ we let $[\mathbf{P}^{\|}]_{\mathsf{cc}'} \in \mathbb{R}^{N_{\mathsf{C}} \times N_{\mathsf{C}}}$ be the sub-matrix of of $\mathbf{P}^{\|}$ composed of the c^{th} block of rows and the c'^{th} block of columns. The sub-matrix $[\mathbf{P}^{\|}]_{(\mathsf{c},\cdot)} \in \mathbb{R}^{N_{\mathsf{C}} \times ML}$ is composed of the c^{th} block of rows and all columns and a similar definition is given to $[\mathbf{P}^{\|}]_{(\cdot,\mathsf{c})} \in \mathbb{R}^{ML \times N_{\mathsf{C}}}$. For a projection matrix $\mathbf{P}^{\|} \in \mathbb{R}^{n \times n}$, the sub-matrices $[\mathbf{P}^{\|}]_{\mathsf{rr}'} \in \mathbb{R}^{N_{\mathsf{R}} \times N_{\mathsf{R}}}$ and $[\mathbf{P}^{\|}]_{(\mathsf{r},\cdot)} \in \mathbb{R}^{N_{\mathsf{R}} \times n}$ are similarly defined.

Lemma 7.3. *The conditional distributions of the vectors in (57) satisfy the following for $\mathsf{c} \in [C]$ and $\mathsf{r} \in [R]$, assuming $n > t$ and $\boldsymbol{Q}_{t+1}^{\mathsf{r}}$ and $\boldsymbol{M}_t^{\mathsf{c}}$ have full column rank.*

$$\left[\mathbf{A}\, \boldsymbol{q}^{0,r}\right]_r \big|_{\mathscr{S}_{0,0}} \overset{d}{=} \boldsymbol{b}_r^0 \quad \text{and} \quad \left[\mathbf{A}^* \boldsymbol{m}^{0,c}\right]_c \big|_{\mathscr{S}_{1,0}} \overset{d}{=} \sum_{r \in [R]} \left[\left[\mathbf{P}^{\perp}_{\boldsymbol{q}^{0,r}} \mathbf{A}'^*\right]_{(c,r)} + \boldsymbol{q}_c^{0,r} \left\|\boldsymbol{q}^{0,r}\right\|^{-2} (\boldsymbol{b}_r^0)^* \right] \boldsymbol{m}_r^{0,c}, \tag{75}$$

and for $t \geq 1$,

$$[\mathbf{A}\, \boldsymbol{q}^{t,r}]_r|_{\mathscr{S}_{t,t}} \stackrel{d}{=} \boldsymbol{Y}_{t,r}\boldsymbol{\gamma}^{t,r} + \sum_{c\in[C]} \left([\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{c'}} \widehat{\mathbf{A}}]_{(r,c)} + \boldsymbol{M}_{t,r}^{c}((\boldsymbol{M}_t^{c})^*\boldsymbol{M}_t^{c})^{-1}(\boldsymbol{X}_{t,c})^*\right)\boldsymbol{q}_{\perp,c}^{t,r}, \tag{76}$$

$$[\mathbf{A}^*\boldsymbol{m}^{t,c}]_c|_{\mathscr{S}_{t+1,t}} \stackrel{d}{=} \boldsymbol{X}_{t,c}\boldsymbol{\alpha}^{t,c} + \sum_{r\in[R]} \left([\mathbf{P}^{\perp}_{\boldsymbol{Q}_{t+1}^{r'}} \mathbf{A}'^*]_{(c,r)} + \boldsymbol{Q}_{t+1,c}^{r}((\boldsymbol{Q}_{t+1}^{r})^*\boldsymbol{Q}_{t+1}^{r})^{-1}(\boldsymbol{Y}_{t+1,r})^*\right)\boldsymbol{m}_{\perp,r}^{t,c}. \tag{77}$$

Here $\mathbf{A}' \stackrel{d}{=} \mathbf{A}$ *and* $\widehat{\mathbf{A}} \stackrel{d}{=} \mathbf{A}$ *are random matrices independent of* $\mathscr{S}_{t+1,t}$ *and* $\mathscr{S}_{t,t}$, *and* $\mathbf{P}^{\perp}_{\boldsymbol{Q}}$ *denotes the projection matrix onto the orthogonal complement of the space spanned by the columns of* $\boldsymbol{Q}$.

The proof of the lemma is given in Section 8.1.

Lemma 7.4 (Conditional Distribution Lemma)**.** *For the vectors* $\boldsymbol{h}^{t+1}$ *and* $\boldsymbol{b}^t$ *defined in* (52), *the following hold for* $t \geq 1$, *provided* $n > t$ *and* $\boldsymbol{M}_t^c, \boldsymbol{Q}_t^r$ *have full column rank.*

$$\boldsymbol{h}_c^1|_{\mathscr{S}_{1,0}} \stackrel{d}{=} \sqrt{\tau_{\perp,c}^0}\,\boldsymbol{Z}_{0,c} + \boldsymbol{\Delta}_{1,0,c}, \quad \text{and} \quad \boldsymbol{h}_c^{t+1}|_{\mathscr{S}_{t+1,t}} \stackrel{d}{=} \frac{\tau_c^t}{\tau_c^{t-1}}\boldsymbol{h}_c^t + \sqrt{\tau_{\perp,c}^t}\,\boldsymbol{Z}_{t,c} + \boldsymbol{\Delta}_{t+1,t,c}, \tag{78}$$

$$\boldsymbol{b}_r^0|_{\mathscr{S}_{0,0}} \stackrel{d}{=} \sqrt{\sigma_{\perp,r}^0}\,\boldsymbol{Z}'_{0,r}, \quad \text{and} \quad \boldsymbol{b}_r^t|_{\mathscr{S}_{t,t}} \stackrel{d}{=} \frac{\sigma_r^t}{\sigma_r^{t-1}}\boldsymbol{b}_r^{t-1} + \sqrt{\sigma_{\perp,r}^t}\,\boldsymbol{Z}'_{t,r} + \boldsymbol{\Delta}_{t,t,r}. \tag{79}$$

For each $t \geq 0$, *the vectors* $\boldsymbol{Z}_t = [\boldsymbol{Z}_{t,1}, \ldots, \boldsymbol{Z}_{t,C}]^* \sim \mathcal{N}(0, \boldsymbol{I}_{ML})$ *and* $\boldsymbol{Z}'_t = [\boldsymbol{Z}_{t,1}, \ldots, \boldsymbol{Z}_{t,R}]^* \sim \mathcal{N}(0, \boldsymbol{I}_n)$ *are independent of the corresponding conditioning sigma algebras. The terms* $\widehat{\gamma}_i^{t,r}$ *and* $\widehat{\alpha}_i^{t,c}$ *for* $i \in [t-1]$ *are defined in* (66) *and* $\sigma_{t,r}^{\perp}$ *and* $\tau_{t,c}^{\perp}$ *are defined in* (67). *The deviation terms are*

$$\boldsymbol{\Delta}_{1,0,c} = \left[\frac{1}{\sqrt{L}}\left\|\boldsymbol{m}^{0,c}\right\| - \sqrt{\tau_c^0}\right]\boldsymbol{Z}_{0,c} - \sum_{r\in[R]}\frac{1}{\sqrt{L}}\left\|\boldsymbol{m}_r^{0,c}\right\|[\mathbf{P}^{\|}_{\boldsymbol{Q}_1^r}\boldsymbol{Z}_0^r]_c + \breve{\boldsymbol{q}}_c^0\left(\sum_{r\in[R]}\frac{\sqrt{W_{rc}}(\boldsymbol{b}_r^0)^*\boldsymbol{m}_r^{0,c}}{L\sigma_r^0} - 1\right), \tag{80}$$

and for $t > 0$,

$$\begin{aligned}\boldsymbol{\Delta}_{t,t,r} = &\sum_{i=0}^{t-2}\boldsymbol{b}_r^i\gamma_i^{t,r} + \boldsymbol{b}_r^{t-1}\left[\gamma_{t-1}^{t,r} - \frac{\sigma_r^t}{\sigma_r^{t-1}}\right] + \left[\frac{1}{\sqrt{L}}\left\|\boldsymbol{q}_{\perp}^{t,r}\right\| - \sqrt{\sigma_{\perp,r}^t}\right]\boldsymbol{Z}'_{t,r} - \sum_{c\in[C]}\frac{1}{\sqrt{L}}\left\|\boldsymbol{q}_{\perp,c}^{t,r}\right\|[\mathbf{P}^{\|}_{\boldsymbol{M}_t^c}\boldsymbol{Z}_t'^c]_r \\ &+ \sum_{c\in[C]}\boldsymbol{M}_{t,r}^{c}((\boldsymbol{M}_t^{c})^*\boldsymbol{M}_t^{c})^{-1}(\boldsymbol{X}_{t,c})^*\,\boldsymbol{q}_{\perp,c}^{t,r} - \sum_{i=1}^{t-1}\gamma_i^{t,r}\upsilon_r^i\breve{\boldsymbol{m}}_r^{i-1} + \upsilon_r^t\breve{\boldsymbol{m}}_r^{t-1},\end{aligned} \tag{81}$$

$$\begin{aligned}\boldsymbol{\Delta}_{t+1,t,c} = &\sum_{i=0}^{t-2}\boldsymbol{h}_c^{i+1}\alpha_i^{t,c} + \boldsymbol{h}_c^t\left[\alpha_{t-1}^{t,c} - \frac{\tau_c^t}{\tau_c^{t-1}}\right] + \left[\frac{1}{\sqrt{L}}\left\|\boldsymbol{m}_{\perp}^{t,c}\right\| - \sqrt{\tau_{\perp,c}^t}\right]\boldsymbol{Z}_{t,c} - \sum_{r\in[R]}\frac{1}{\sqrt{L}}\left\|\boldsymbol{m}_{\perp,r}^{t,c}\right\|[\mathbf{P}^{\|}_{\boldsymbol{Q}_{t+1}^r}\boldsymbol{Z}_t^r]_c \\ &+ \sum_{r\in[R]}\boldsymbol{Q}_{t+1,c}^{r}((\boldsymbol{Q}_{t+1}^{r})^*\boldsymbol{Q}_{t+1}^{r})^{-1}(\boldsymbol{Y}_{t+1,r})^*\boldsymbol{m}_{\perp,r}^{t,c} + \sum_{i=0}^{t-1}\alpha_i^{t,c}\breve{\boldsymbol{q}}_c^i - \breve{\boldsymbol{q}}_c^t.\end{aligned} \tag{82}$$

In (81), *we recall that* $\upsilon_r^i = \sigma_r^i/\phi_r^{i-1}$ *for* $i \in [t]$. *The vectors* $\boldsymbol{Z}_t'^c \sim \mathcal{N}(0, \boldsymbol{I}_n)$ *are i.i.d. for* $c \in [C]$, *and independent of* $\mathscr{S}_{t,t}$. *Similarly, in* (80) *and* (82), $\boldsymbol{Z}_t^r \sim \mathcal{N}(0, \boldsymbol{I}_{ML})$ *are i.i.d. for* $r \in [R]$ *and independent of* $\mathscr{S}_{t+1,t}$. *Furthermore,* $\boldsymbol{Z}'_t = \frac{1}{\sqrt{C}}\sum_{c\in[C]}\boldsymbol{Z}_t'^c$ *and* $\boldsymbol{Z}_t = \frac{1}{\sqrt{R}}\sum_{r\in[R]}\boldsymbol{Z}_t^r$.

The proof of the lemma is given in Section 8.2.

The next lemma uses the representation in Lemma 7.4 to show that $\boldsymbol{h}_{\mathsf{c}}^{t+1}$ is the sum of a $\mathcal{N}(0, \tau_{\mathsf{c}}^t \boldsymbol{I}_{N_{\mathsf{C}}})$ random vector and a deviation term for each $t \geq 0$. Similarly $\boldsymbol{b}_{\mathsf{r}}^t$ is the sum of a $\mathcal{N}(0, \sigma_{\mathsf{r}}^t \boldsymbol{I}_{N_{\mathsf{R}}})$ random vector and a deviation term.

Lemma 7.5. *For $t \geq 0$, the conditional distributions in Lemma 7.4 can be expressed as*

$$\boldsymbol{h}_{\mathsf{c}}^{t+1}|_{\mathscr{S}_{t+1,t}} \stackrel{d}{=} \widetilde{\boldsymbol{h}}_{\mathsf{c}}^t + \widetilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}}, \qquad \boldsymbol{b}_{\mathsf{r}}^t|_{\mathscr{S}_{t,t}} \stackrel{d}{=} \breve{\boldsymbol{b}}_{\mathsf{r}}^t + \breve{\boldsymbol{\Delta}}_{t,r}, \tag{83}$$

where

$$\widetilde{\boldsymbol{h}}_{\mathsf{c}}^{t+1} := \tau_{\mathsf{c}}^t \sum_{i=0}^{t} \sqrt{\tau_{\perp,\mathsf{c}}^i} \left(\frac{1}{\tau_{\mathsf{c}}^i}\right) \boldsymbol{Z}_{i,\mathsf{c}}, \qquad \widetilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}} := \tau_{\mathsf{c}}^t \sum_{i=0}^{t} \left(\frac{1}{\tau_{\mathsf{c}}^i}\right) \boldsymbol{\Delta}_{i+1,i,\mathsf{c}}, \tag{84}$$

$$\breve{\boldsymbol{b}}_{\mathsf{r}}^t := \sigma_{\mathsf{r}}^t \sum_{i=0}^{t} \sqrt{\sigma_{\perp,\mathsf{r}}^i} \left(\frac{1}{\sigma_{\mathsf{r}}^i}\right) \boldsymbol{Z}'_{i,\mathsf{r}}, \qquad \breve{\boldsymbol{\Delta}}_{t,\mathsf{r}} := \sigma_{\mathsf{r}}^t \sum_{i=0}^{t} \left(\frac{1}{\sigma_{\mathsf{r}}^i}\right) \boldsymbol{\Delta}_{i,i,\mathsf{r}}. \tag{85}$$

Here, for each $0 \leq i \leq t$, the standard Gaussian vectors $\boldsymbol{Z}_i = [\boldsymbol{Z}_{i,1}, \ldots, \boldsymbol{Z}_{i,\mathsf{C}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_{ML})$ and $\boldsymbol{Z}'_i = [\boldsymbol{Z}_{i,1}, \ldots, \boldsymbol{Z}_{i,\mathsf{R}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_n)$ are the ones used in Lemma 7.4, Eqs. (78) and (79).

Consequently, $\widetilde{\boldsymbol{h}}_{\mathsf{c}}^{t+1} \stackrel{d}{=} \sqrt{\tau_{\mathsf{c}}^t} \widetilde{\boldsymbol{Z}}_{t,\mathsf{c}}$, and $\breve{\boldsymbol{b}}_{\mathsf{r}}^t \stackrel{d}{=} \sqrt{\sigma_{\mathsf{r}}^t} \breve{\boldsymbol{Z}}_{t,r}$, where $\widetilde{\boldsymbol{Z}}_t = [\widetilde{\boldsymbol{Z}}_{t,1}, \ldots, \widetilde{\boldsymbol{Z}}_{t,\mathsf{C}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_{ML})$ and $\breve{\boldsymbol{Z}}_t = [\breve{\boldsymbol{Z}}_{t,1}, \ldots, \breve{\boldsymbol{Z}}_{t,\mathsf{R}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_n)$ such that for any $j \in [N_{\mathsf{C}}]$ and $i \in [N_{\mathsf{R}}]$, the length-t vectors $([\widetilde{Z}_{0,\mathsf{c}}]_j, \ldots, [\widetilde{Z}_{t,\mathsf{c}}]_j)$ and $([\breve{Z}_{0,r}]_i, \ldots, [\breve{Z}_{t,r}]_i)$ are each jointly Gaussian with

$$\mathbb{E}\{[\widetilde{Z}_{\widetilde{s},\mathsf{c}}]_j [\widetilde{Z}_{s,\mathsf{c}}]_j\} = \sqrt{\tau_{\mathsf{c}}^{\widetilde{s}}/\tau_{\mathsf{c}}^s}, \qquad \mathbb{E}\{[\breve{Z}_{\widetilde{s},r}]_i [\breve{Z}_{s,r}]_i\} = \sqrt{\sigma_r^{\widetilde{s}}/\sigma_r^s} \qquad \text{for } 0 \leq s \leq \widetilde{s} \leq t. \tag{86}$$

Hence for any $s \leq t$ we can write

$$\widetilde{\boldsymbol{Z}}_{t,\mathsf{c}} \stackrel{d}{=} \widetilde{\boldsymbol{Z}}_{s,\mathsf{c}} \sqrt{\frac{\tau_{\mathsf{c}}^t}{\tau_{\mathsf{c}}^s}} + \widetilde{\mathbf{U}}_{t,\mathsf{c}} \sqrt{1 - \frac{\tau_{\mathsf{c}}^t}{\tau_{\mathsf{c}}^s}}, \quad \text{and} \quad \breve{\boldsymbol{Z}}_{t,r} \stackrel{d}{=} \breve{\boldsymbol{Z}}_{s,r} \sqrt{\frac{\sigma_r^t}{\sigma_r^s}} + \breve{\mathbf{U}}_{t,r} \sqrt{1 - \frac{\sigma_r^t}{\sigma_r^s}}, \tag{87}$$

where $\widetilde{\boldsymbol{Z}}_s$ is independent of $\widetilde{\mathbf{U}}_t = [\widetilde{U}_{t,1}, \ldots, \widetilde{U}_{t,\mathsf{C}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_{ML})$ and $\breve{\boldsymbol{Z}}_s$ is independent of $\breve{\mathbf{U}}_t = [\breve{U}_{t,1}, \ldots, \breve{U}_{t,\mathsf{R}}]^* \sim \mathcal{N}(0, \boldsymbol{I}_n)$.

Proof. The proof is similar to that of [41, Lemma 6] and is omitted. ∎

7.4 Order of SPARC parameters and state evolution constants

We recall a few facts about the SC-SPARCs construction from Section 2 that will be used throughout the proof. There are L/C sections per column block of $\boldsymbol{\beta}$, with the non-zero coefficient in each section equal to 1. Each block in the design matrix $\boldsymbol{A}$ can be viewed as a standard (non-SC) SPARC with L/C sections (with M columns in each section), code length $N_{\mathsf{R}} = n/\mathsf{R}$, and rate

$$R_{\text{inner}} = \frac{(L/\mathsf{C}) \ln M}{(n/\mathsf{R})} = R \cdot \frac{\mathsf{R}}{\mathsf{C}}.$$

For an (ω, Λ, ρ) base matrix, $\mathsf{C} = \Lambda$ and $\mathsf{R} = (\Lambda + \omega - 1)$. Since $L \gg \mathsf{C} = \Lambda \gg \omega$, we have $\frac{\mathsf{R}}{\mathsf{C}} = 1 + \frac{\omega - 1}{\Lambda} < 2$, hence $R < R_{\text{inner}} < 2R$.

From (6), we have $\max_{\mathsf{r},\mathsf{c}} W_{\mathsf{r},\mathsf{c}} \le P\frac{\mathsf{R}}{\omega}$. From (6), it can also be verified that for $\mathsf{r} \in [\mathsf{R}]$,

$$\frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}} = \kappa_1, \qquad \frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^2 = \kappa_2\Big(\frac{\mathsf{R}}{\omega}\Big), \qquad \frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^4 = \kappa_3\Big(\frac{\mathsf{R}}{\omega}\Big)^3, \tag{88}$$

where $\kappa_1, \kappa_2, \kappa_3$ are absolute positive constants. A similar statement holds when the summations in (88) are over $\mathsf{r} \in [\mathsf{R}]$, with c fixed.

Studying the state evolution equations (10)-(11), we see that $\psi_{\mathsf{c}}^t, \sigma_{\mathsf{r}}^t$, and ϕ_{r}^t are all $\Theta(1)$ for all $t \ge 0, \mathsf{r} \in [\mathsf{R}]$, and $\mathsf{c} \in [\mathsf{C}]$, while $\tau_{\mathsf{c}}^t = \Theta(1/\ln M) = \Theta(L/n)$ for all $t \ge 0$ and $\mathsf{c} \in [\mathsf{C}]$. This implies that $S_{\mathsf{rc}}^t = \tau_{\mathsf{c}}^t(\phi_{\mathsf{r}}^t)^{-1} = \Theta(1/\ln M) = \Theta(L/n)$. We will use the following facts implicitly. For $t \ge 0$, $\min_{\mathsf{r}\in[\mathsf{R}]} \phi_{\mathsf{r}}^t = \phi_1^t$ and $\min_{\mathsf{r}\in[\mathsf{R}]} \sigma_{\mathsf{r}}^t = \sigma_1^t$, additionally, $\max_{\mathsf{r}\in[\mathsf{R}]} \phi_{\mathsf{r}}^t = \phi_{\lfloor \mathsf{R}/2 \rfloor}^t$ and $\max_{\mathsf{r}\in[\mathsf{R}]} \sigma_{\mathsf{r}}^t = \sigma_{\lfloor \mathsf{R}/2 \rfloor}^t$. Furthermore, $\max_{\mathsf{r}} \sigma_{\mathsf{r}}^t \le \max_{\mathsf{r}} \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \le 2P$. It follows that for $\mathsf{c} \in [\mathsf{C}]$,

$$S_{\mathsf{rc}}^t = \tau_{\mathsf{c}}^t(\phi_{\mathsf{r}}^t)^{-1} \le \tau_{\mathsf{c}}^t(\phi_1^t)^{-1} = S_{1,\mathsf{c}}^t. \tag{89}$$

7.5 Concentration lemma

The next lemma uses the conditional distribution given by Lemma 7.4 to prove concentration results for various inner products and functions involving $\{\boldsymbol{b}^t, \boldsymbol{m}^{t,\mathsf{c}}, \boldsymbol{h}^{t+1}, \boldsymbol{q}^{t,\mathsf{r}}\}$. The concentration lemma is stated in two parts. The first part gives concentration inequalities for inner products involving the vectors $\{\boldsymbol{b}^s, \boldsymbol{m}^{s,\mathsf{c}}\}_{0\le s\le t}$ (Eqs. (93)–(100)). The second part gives concentration inequalities for inner products involving the vectors $\{\boldsymbol{h}^{s+1}, \boldsymbol{q}^{s,\mathsf{r}}\}_{0\le s\le t}$ (Eqs. (102)–(109)). These results are proved using an induction argument that includes two other concentration inequalities (Eqs. (92) and (101)) showing that that the deviation terms in Lemma 7.4 are small.

The proof of Theorem 2 requires only one of the results in the concentration lemma, namely, Eq. (103). However, the other results in the lemma are required for the induction argument.

To keep the notation compact, we use K, K', κ, and κ' to denote generic positive universal constants whose values may change throughout the lemma statement and proof.

Let $\xi, \Xi > 0$ be universal constants not depending on n, ϵ, or t. For $t \ge 0$, let

$$K_t = \Xi^{2t}(t!)^{14}, \quad \kappa_t = \frac{1}{\xi^{2t}(t!)^{24}}, \quad K_t' = \Xi(t+1)^7 K_t, \quad \kappa_t' = \frac{\kappa_t}{\xi(t+1)^{12}}. \tag{90}$$

We also define the following iteration-dependent quantities that summarize the problem parameters:

$$\Pi_t = (\mathsf{RC})^{t+1}, \quad \pi_t = \frac{(N_{\mathsf{R}}\omega)(\omega/\mathsf{R})^{2(t+1)}}{(\log M)^{2(t+1)}}, \quad \Pi_t' = \mathsf{C}\Pi_t, \quad \pi_t' = \pi_t. \tag{91}$$

Lemma 7.6. *The following results hold for $\epsilon \in (0,1)$ and $1 \le t < T$, where T is defined in (38).*

1. *Let u be an integer with $\mathsf{u} \in \{0,1,2\}$. Let $X_n \doteq$ const be shorthand for*

$$P(|X_n - \textit{const}| \ge \epsilon) \le t^4 K K_{t-1} \Pi_{t-1} \exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/R)^{2(\mathsf{u}-1)_+}\pi_{t-1}\epsilon^2}{t^8}\Big\}.$$

(a) *For $\mathsf{c} \in [C]$*

$$P\Big(\frac{1}{n}\sum_{r\in[R]} W_{rc}^{\mathsf{u}} \|\boldsymbol{\Delta}_{t,t,r}\|^2 \ge \epsilon\Big) \le t^3 K K_{t-1}\Pi_{t-1} \exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/R)^{(\mathsf{u}-1)_+}\pi_{t-1}\epsilon}{t^6}\Big\}, \tag{92}$$

(b) *For all* $\mathsf{c} \in [\mathsf{C}]$,

$$\frac{1}{n}\sum_{r\in[R]} W_{r\mathsf{c}}^{\mathsf{u}}(\boldsymbol{b}_r^t)^*\boldsymbol{w}_r \doteq 0, \tag{93}$$

(c) *For all* $0 \le s \le t$ *and* $\mathsf{c} \in [\mathsf{C}]$,

$$\frac{1}{n}\sum_{r\in[R]} W_{r\mathsf{c}}^{\mathsf{u}}(\boldsymbol{b}_r^s)^*\boldsymbol{b}_r^t \doteq \frac{1}{R}\sum_{r\in[R]} W_{r\mathsf{c}}^{\mathsf{u}}\sigma_r^t, \tag{94}$$

(d) *For all* $0 \le s, \widetilde{s} \le t$ *and* $\mathsf{c} \in [\mathsf{C}]$,

$$\frac{1}{L}\sum_{r\in[R]} W_{r\mathsf{c}}^{\mathsf{u}-1/2}(\boldsymbol{b}_r^{\widetilde{s}})^*\boldsymbol{m}_r^{s,\mathsf{c}} \doteq \frac{N_R}{L}\sum_{r\in[R]} S_{r\mathsf{c}}^s W_{r\mathsf{c}}^{\mathsf{u}}\,\sigma_r^{\max(\widetilde{s},s)}, \tag{95}$$

(e) *For all* $0 \le s \le t$ *and* $\mathsf{c} \in [\mathsf{C}]$,

$$\frac{n}{L^2}\sum_{r\in[R]} W_{r\mathsf{c}}^{\mathsf{u}-1}(\boldsymbol{m}_r^{s,\mathsf{c}})^*\boldsymbol{m}_r^{t,\mathsf{c}} \doteq \frac{n^2}{L^2R}\sum_{r\in[R]} S_{r\mathsf{c}}^s S_{r\mathsf{c}}^t W_{r\mathsf{c}}^{\mathsf{u}}\,\phi_r^t \quad (= \frac{n}{L}\tau_{\mathsf{c}}^t \text{ when } \mathsf{u}=1), \tag{96}$$

(f) *For* $\tau_{\perp,\mathsf{c}}^t$ *defined in* (67) *and shown to be positive in Lemma* 7.1 *and* $\widehat{\boldsymbol{\alpha}}^{t,\mathsf{c}}$ *defined in* (62)*, for all* $\mathsf{c} \in [\mathsf{C}]$ *and* $0 < i \le t$,

$$P\Big(|\alpha_{i-1}^{t,\mathsf{c}} - \widehat{\alpha}_{i-1}^{t,\mathsf{c}}| \ge \epsilon\Big) \le t^5 K K_{t-1}\Pi_{t-1}\exp\Big\{\frac{-\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2}{t^{10}}\Big\}, \quad t \ge 1, \tag{97}$$

$$P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{m}_\perp^{t,\mathsf{c}}\|^2 - \tau_{\perp,\mathsf{c}}^t\Big| \ge \epsilon\Big) \le t^6 K K_{t-1}\Pi_{t-1}\exp\Big\{\frac{-\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2}{t^{12}}\Big\}. \tag{98}$$

(g) *Let* $\mathbb{M}_{t+1}^{\mathsf{c}} := \frac{n}{L^2}(\boldsymbol{M}_{t+1}^{\mathsf{c}})^*\boldsymbol{M}_{t+1}^{\mathsf{c}}$. *Then for* $\mathsf{c} \in [\mathsf{C}]$,

$$P\Big(\mathbb{M}_{t+1}^{\mathsf{c}} \text{ is singular}\Big) \le t^7 K K_{t-1}\Pi_{t-1}\exp\Big\{\frac{-\kappa\kappa_{t-1}\pi_{t-1}}{t^{12}}\Big\}. \tag{99}$$

For matrix $\breve{\boldsymbol{C}}^{t+1,\mathsf{c}}$ *defined in* (65)*, when the inverses of* $\mathbb{M}_{t+1}^{\mathsf{c}}$ *exist,*

$$P\Big(\Big|[(\mathbb{M}_{t+1}^{\mathsf{c}})^{-1} - (\breve{\boldsymbol{C}}^{t+1,\mathsf{c}})^{-1}]_{ij}\Big| \ge \epsilon\Big) \le t^6 K K_{t-1}\Pi_{t-1}\exp\Big\{\frac{-\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2}{t^{12}}\Big\}, \tag{100}$$

2. *For* $\mathsf{v} \in \{0,1\}$, *we have the following:*
(a) *For all* $r \in [R]$,

$$P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{r\mathsf{c}}^{2\mathsf{v}}\sum_{\ell\in\mathsf{c}}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}]_j|^2 \ge \epsilon\Big) \le t^3 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^{2\mathsf{v}}\pi'_{t-1}\epsilon}{t^6}\Big\}. \tag{101}$$

(b) *For all* $r \in [R]$,

$$P\Big(\Big|\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{r\mathsf{c}}}(\boldsymbol{h}_{\mathsf{c}}^{t+1})^*\boldsymbol{q}_{\mathsf{c}}^{0,r}\Big| \ge \epsilon\Big) \le t^3 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^2\pi'_{t-1}\epsilon^2}{t^6}\Big\}. \tag{102}$$

(c) *For all* $0 \le s \le t+1$ *and* $r \in [R]$,

$$P\Big(\Big|\sum_{c\in[C]} W_{rc}^{\mathsf{v}}\Big[\frac{1}{L}(\breve{\boldsymbol{q}}_c^s)^*\breve{\boldsymbol{q}}_c^{t+1} - \frac{1}{C}\psi_c^{t+1}\Big]\Big| \ge \epsilon\Big) \le t^4 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^{2\mathsf{v}}\pi'_{t-1}\epsilon^2}{t^8(\log M)^2}\Big\}, \quad (103)$$

where for $\mathsf{v}=1$, *we note that* $\frac{1}{C}\sum_{c\in[C]} W_{rc}\psi_c^{t+1} = \sigma_r^{t+1}$.

(d) *For all* $0 \le s, \widetilde{s} \le t$ *and* $r \in [R]$,

$$P\Big(\Big|\sum_{c\in[C]}\sqrt{W_{rc}}\Big[\frac{(\boldsymbol{h}_c^{\widetilde{s}+1})^*\boldsymbol{q}_c^{s+1,r}}{L} + \frac{\sqrt{W_{rc}}\psi_c^{s+1}\tau_c^{\max(\widetilde{s},s)}}{C\tau_c^s}\Big]\Big| \ge \epsilon\Big) \le t^4 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^2\pi'_{t-1}\epsilon^2}{t^8(\log M)^2}\Big\}. \quad (104)$$

(e)

$$P\Big(\sum_{c\in[C]}\sum_{\ell\in c}\frac{W_{rc}^{2\mathsf{v}}}{L}\max_{j\in sec(\ell)}([\boldsymbol{h}_c^{t+1}]_j)^2 \ge 6(\max_c W_{rc}^{2\mathsf{v}})\tau_c^t\log M + \epsilon\Big)$$
$$\le t^4 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^{2\mathsf{v}}\pi'_{t-1}\epsilon}{t^8}\Big\}. \quad (105)$$

(f) *For* $\sigma_{\perp,r}^{t+1}$ *defined in* (67) *and shown to be positive in Lemma* 7.1, *and* $\widehat{\gamma}^{t+1,r}$ *is defined in* (62), *for all* $r \in [R]$ *and* $0 \le i \le (t+1)$,

$$P\Big(|\gamma_{i-1}^{t+1,r} - \widehat{\gamma}_{i-1}^{t+1,r}| \ge \epsilon\Big) \le t^5 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^2\pi'_{t-1}\epsilon^2}{t^{10}(\log M)^2}\Big\}. \quad (106)$$

$$P\Big(\Big|\frac{1}{L}\|\boldsymbol{q}_\perp^{t+1,r}\|^2 - \sigma_{\perp,r}^{t+1}\Big| \ge \epsilon\Big) \le t^6 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^2\pi'_{t-1}\epsilon^2}{t^{12}(\log M)^2}\Big\}. \quad (107)$$

(g) *Let* $\mathbb{Q}_{t+2}^r := \frac{1}{L}(\boldsymbol{Q}_{t+2}^r)^*\boldsymbol{Q}_{t+2}^r$. *Then for all* $r \in [R]$,

$$P(\mathbb{Q}_{t+2}^r \text{ is singular}) \le t^7 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^2\pi'_{t-1}\epsilon^2}{t^{12}(\log M)^2}\Big\}, \quad (108)$$

For matrix $\widetilde{\boldsymbol{C}}^{t+2,r}$ *defined in* (65), *when the inverses of* $\mathbb{Q}_{t+2}^r$ *exist,*

$$P\Big(\Big|[(\mathbb{Q}_{t+2}^r)^{-1} - (\widetilde{\boldsymbol{C}}^{t+2,r})^{-1}]_{ij}\Big| \ge \epsilon\Big) \le t^6 K K'_{t-1}\Pi'_{t-1}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/R)^2\pi'_{t-1}\epsilon^2}{t^{12}(\log M)^2}\Big\}. \quad (109)$$

7.6 Proof of Theorem 2

Recall from (52) that $\breve{\boldsymbol{q}}^s = \boldsymbol{\beta}^s - \boldsymbol{\beta}_0$. With $\mathsf{v}=0$, Eq. (103) implies that for $0 \le s \le (T-1)$,

$$P\Big(\frac{1}{L}\|\boldsymbol{\beta}^{s+1} - \boldsymbol{\beta}_0\|^2 - \frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}\psi_{\mathsf{c}}^{s+1}\Big| \ge \epsilon\Big) \le s^4 K K'_{s-1}(\mathsf{RC})^s\mathsf{C}\exp\Big\{\frac{-\kappa\kappa'_{s-1}(N_{\mathsf{R}}\omega)\,(\omega/\mathsf{R})^{2s}\,\epsilon^2}{t^8(\log M)^{2s+2}}\Big\}$$
$$\le K_s(\mathsf{RC})^{s+1}\exp\Big\{\frac{-\kappa_s(N_{\mathsf{R}}\omega)\epsilon^2}{(\log M)^{2s+2}(\mathsf{R}/\omega)^{2s}}\Big\}. \quad (110)$$

Setting $s = (t-1)$ and recalling that $N_{\mathsf{R}} = n/\mathsf{R}$ yields the statement of Theorem 2. ■

Remark 7.1. Although Theorem 2 is an immediate consequence of (103), the other concentration results in Lemma 7.6 are required for the induction based proof of (103). Indeed, we write

$$\frac{1}{L}\|\breve{\boldsymbol{q}}_{\mathsf{c}}^{s+1}\|^2 = \frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\|\eta_{\mathsf{c}}^t(\boldsymbol{\beta}_{\mathsf{c}} - \boldsymbol{h}_{\mathsf{c}}^{s+1}) - \boldsymbol{\beta}_{\mathsf{c}}\|^2, \tag{111}$$

and use the representation for $\boldsymbol{h}_{\mathsf{c}}^{s+1}$ from Lemma 7.4. The concentration results in the Lemma 7.6 are used in an induction argument to show that the deviation term is negligible in the sense of (101). Consequently, $\boldsymbol{h}_{\mathsf{c}}^{s+1}$ is approximately Gaussian, and hence $\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\|\eta_{\mathsf{c}}^t(\boldsymbol{\beta}_{\mathsf{c}} - \boldsymbol{h}_{\mathsf{c}}^{s+1}) - \boldsymbol{\beta}_{\mathsf{c}}\|^2$ concentrates on a deterministic value, as described in Section 7.2. The detailed proof of Lemma 7.6 is given in Section 8.2.

8 Proofs of conditional distribution and concentration lemmas

8.1 Proof of Lemma 7.3

For $t_a \geq 1$, let $\mathbf{P}^{\parallel}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}$ and $\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}$ denote the orthogonal projectors onto the column space of $\boldsymbol{Q}^{\mathsf{r}}_{t_a}$ and its orthogonal complement, respectively. Given the constraints in (74), for $\mathsf{r} \in [\mathsf{R}]$, we can write

$$\begin{aligned}\mathbf{A}_{(\mathsf{r},\cdot)} = \mathbf{A}_{(\mathsf{r},\cdot)}\left(\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}} + \mathbf{P}^{\parallel}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}\right) &= \mathbf{A}_{(\mathsf{r},\cdot)}\left(\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}} + \boldsymbol{Q}^{\mathsf{r}}_{t_a}((\boldsymbol{Q}^{\mathsf{r}}_{t_a})^*\boldsymbol{Q}^{\mathsf{r}}_{t_a})^{-1}(\boldsymbol{Q}^{\mathsf{r}}_{t_a})^*\right)\\ &= \mathbf{A}_{(\mathsf{r},\cdot)}\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}} + \boldsymbol{Y}_{t_a,\mathsf{r}}((\boldsymbol{Q}^{\mathsf{r}}_{t_a})^*\boldsymbol{Q}^{\mathsf{r}}_{t_a})^{-1}(\boldsymbol{Q}^{\mathsf{r}}_{t_a})^*.\end{aligned} \tag{112}$$

Similarly, for $\mathsf{c} \in [\mathsf{C}]$ and $t \geq 0$, we can write

$$\mathbf{A}_{(\cdot,\mathsf{c})} = \left(\mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}}_t} + \mathbf{P}^{\parallel}_{\boldsymbol{M}^{\mathsf{c}}_t}\right)\mathbf{A}_{(\cdot,\mathsf{c})} = \mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}}_t}\mathbf{A}_{(\cdot,\mathsf{c})} + \boldsymbol{M}^{\mathsf{c}}_t((\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{M}^{\mathsf{c}}_t)^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*, \tag{113}$$

where we interpret the $t = 0$ case as follows: $\mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}}_0} = \mathbf{P}^{\perp}_{\mathbf{0}} = \boldsymbol{I}$ and $\mathbf{P}^{\parallel}_{\boldsymbol{M}^{\mathsf{c}}_0} = \mathbf{P}^{\parallel}_{\mathbf{0}} = \mathbf{0}$. From (112) and (113), we have two equivalent representations for the submatrix $\mathbf{A}_{(\mathsf{r},\mathsf{c})} \in \mathbb{R}^{N_{\mathsf{R}}\times N_{\mathsf{C}}}$:

$$\mathbf{A}_{\mathsf{rc}} = \sum_{\mathsf{c}'\in[\mathsf{C}]}\mathbf{A}_{\mathsf{rc}'}[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}]_{\mathsf{c}'\mathsf{c}} + \boldsymbol{Y}_{t_a,\mathsf{r}}((\boldsymbol{Q}^{\mathsf{r}}_{t_a})^*\boldsymbol{Q}^{\mathsf{r}}_{t_a})^{-1}(\boldsymbol{Q}^{\mathsf{r}}_{t_a,\mathsf{c}})^* \tag{114}$$

$$= \sum_{\mathsf{r}'\in[\mathsf{R}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}}_t}]_{\mathsf{r}\mathsf{r}'}\mathbf{A}_{\mathsf{r}'\mathsf{c}} + \boldsymbol{M}^{\mathsf{c}}_{t,\mathsf{r}}((\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{M}^{\mathsf{c}}_t)^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*. \tag{115}$$

Using the representation (115) in (114), we obtain

$$\begin{aligned}\mathbf{A}_{\mathsf{rc}} = \sum_{\mathsf{c}'\in[\mathsf{C}]}\sum_{\mathsf{r}'\in[\mathsf{R}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}'}_t}]_{\mathsf{r}\mathsf{r}'}\mathbf{A}_{\mathsf{r}'\mathsf{c}'}[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}]_{\mathsf{c}'\mathsf{c}} + \sum_{\mathsf{c}'\in[\mathsf{C}]}\boldsymbol{M}^{\mathsf{c}'}_{t,\mathsf{r}}((\boldsymbol{M}^{\mathsf{c}'}_t)^*\boldsymbol{M}^{\mathsf{c}'}_t)^{-1}(\boldsymbol{X}_{t,\mathsf{c}'})^*[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}]_{\mathsf{c}'\mathsf{c}}\\ + \boldsymbol{Y}_{t_a,\mathsf{r}}((\boldsymbol{Q}^{\mathsf{r}}_{t_a})^*\boldsymbol{Q}^{\mathsf{r}}_{t_a})^{-1}(\boldsymbol{Q}^{\mathsf{r}}_{t_a,\mathsf{c}})^*,\end{aligned} \tag{116}$$

and using (114) in (115), we obtain

$$\begin{aligned}\mathbf{A}_{\mathsf{rc}} = \sum_{\mathsf{r}'\in[\mathsf{R}]}\sum_{\mathsf{c}'\in[\mathsf{C}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}}_t}]_{\mathsf{r}\mathsf{r}'}\mathbf{A}_{\mathsf{r}'\mathsf{c}'}[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}'}_{t_a}}]_{\mathsf{c}'\mathsf{c}} + \sum_{\mathsf{r}'\in[\mathsf{R}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}^{\mathsf{c}}_t}]_{\mathsf{r}\mathsf{r}'}\boldsymbol{Y}_{t_a,\mathsf{r}'}((\boldsymbol{Q}^{\mathsf{r}'}_{t_a})^*\boldsymbol{Q}^{\mathsf{r}'}_{t_a})^{-1}(\boldsymbol{Q}^{\mathsf{r}'}_{t_a,\mathsf{c}})^*\\ + \boldsymbol{M}^{\mathsf{c}}_{t,\mathsf{r}}((\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{M}^{\mathsf{c}}_t)^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*.\end{aligned} \tag{117}$$

The first term in the RHS of (116) can be written as $\sum_{\mathsf{c}'\in[\mathsf{C}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}'}}\mathbf{A}]_{(\mathsf{r},\mathsf{c}')}[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}]_{\mathsf{c}'\mathsf{c}}$. Recall that conditioning on the sigma-algebra $\mathscr{S}_{t,t}$ is equivalent to conditioning on the linear constraints in (74) with $t_a = t$. We note that the constraint corresponding to r in (74) involves only the rth row block of $\mathbf{A}$, and the constraint corresponding to c' in (74) involves only the c'th column block of $\mathbf{A}$, for each $\mathsf{r}\in[\mathsf{R}], \mathsf{c}'\in[\mathsf{C}]$. Therefore, since the entries of $\mathbf{A}$ are i.i.d. Gaussian, the conditional distribution of the above term given $\mathscr{S}_{t,t}$ satisfies [14, Lemmas 10, 12],

$$\sum_{\mathsf{c}'\in[\mathsf{C}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}'}}\mathbf{A}]_{(\mathsf{r},\mathsf{c}')}[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}]_{\mathsf{c}'\mathsf{c}}\Big|_{\mathscr{S}_{t,t}} \overset{d}{=} \sum_{\mathsf{c}'\in[\mathsf{C}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}'}}\hat{\mathbf{A}}]_{(\mathsf{r},\mathsf{c}')}[\mathbf{P}^{\perp}_{\boldsymbol{Q}^{\mathsf{r}}_{t_a}}]_{\mathsf{c}'\mathsf{c}}, \tag{118}$$

where $\hat{\mathbf{A}} \overset{d}{=} \mathbf{A}$ is independent of $\mathscr{S}_{t,t}$.

We first prove the two results in (75). We note that $\mathbf{A}$ is independent of $\mathscr{S}_{0,0}$ and $\boldsymbol{b}^0_{\mathsf{r}} = \mathbf{A}\boldsymbol{q}^{0,\mathsf{r}}$ by definition. When $t_a = 1$ and $t = 0$, that the result in (117) gives

$$\mathbf{A}_{\mathsf{rc}} = \sum_{\mathsf{r}'\in[\mathsf{R}]}\sum_{\mathsf{c}'\in[\mathsf{C}]}\boldsymbol{I}_{\mathsf{r}\mathsf{r}'}\mathbf{A}_{\mathsf{r}'\mathsf{c}'}[\mathbf{P}^{\perp}_{\boldsymbol{q}^{0,\mathsf{r}'}}]_{\mathsf{c}'\mathsf{c}} + \sum_{\mathsf{r}'\in[\mathsf{R}]}\boldsymbol{I}_{\mathsf{r}\mathsf{r}'}\boldsymbol{b}^0_{\mathsf{r}'}\left\|\boldsymbol{q}^{0,\mathsf{r}'}\right\|^{-2}(\boldsymbol{q}^{0,\mathsf{r}'}_{\mathsf{c}})^* = [\mathbf{A}\mathbf{P}^{\perp}_{\boldsymbol{q}^{0,\mathsf{r}}}]_{(\mathsf{r},\mathsf{c})} + \boldsymbol{b}^0_{\mathsf{r}}\left\|\boldsymbol{q}^{0,\mathsf{r}}\right\|^{-2}(\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}})^*.$$

Using the above and noting that $\mathbf{A}\mathbf{P}^{\perp}_{\boldsymbol{q}^{0,\mathsf{r}}}|_{\mathscr{S}_{1,0}} \overset{d}{=} \mathbf{A}'\mathbf{P}^{\perp}_{\boldsymbol{q}^{0,\mathsf{r}}}$ (by a similar argument to (118)), we have

$$\left[\mathbf{A}^*\boldsymbol{m}^{0,\mathsf{c}}\right]_{\mathsf{c}}\Big|_{\mathscr{S}_{1,0}} = \sum_{\mathsf{r}\in[\mathsf{R}]}(\mathbf{A}_{(\mathsf{r},\mathsf{c})})^*\boldsymbol{m}^{0,\mathsf{c}}_{\mathsf{r}}\Big|_{\mathscr{S}_{1,0}} \overset{d}{=} \sum_{\mathsf{r}\in[\mathsf{R}]}\left[\left[\mathbf{P}^{\perp}_{\boldsymbol{q}^{0,\mathsf{r}}}\mathbf{A}'^*\right]_{(\mathsf{c},\mathsf{r})} + \boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}\left\|\boldsymbol{q}^{0,\mathsf{r}}\right\|^{-2}(\boldsymbol{b}^0_{\mathsf{r}})^*\right]\boldsymbol{m}^{0,\mathsf{c}}_{\mathsf{r}}.$$

Next, using (116) and (118) and multiplying by $\boldsymbol{q}^{t,\mathsf{r}}$, we obtain

$$\begin{aligned}\left[\mathbf{A}\boldsymbol{q}^{t,\mathsf{r}}\right]_{\mathsf{r}}\Big|_{\mathscr{S}_{t,t}} &= \sum_{\mathsf{c}\in[\mathsf{C}]}\mathbf{A}_{(\mathsf{r},\mathsf{c})}\boldsymbol{q}^{t,\mathsf{r}}_{\mathsf{c}}\Big|_{\mathscr{S}_{t,t}}\\ &\overset{d}{=} \sum_{\mathsf{c}'\in[\mathsf{C}]}[\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}'}}\hat{\mathbf{A}}]_{(\mathsf{r},\mathsf{c}')}\,\boldsymbol{q}^{t,\mathsf{r}}_{\perp,\mathsf{c}'} + \sum_{\mathsf{c}'\in[\mathsf{C}]}\boldsymbol{M}^{\mathsf{c}'}_{t,\mathsf{r}}((\boldsymbol{M}^{\mathsf{c}'}_t)^*\boldsymbol{M}^{\mathsf{c}'}_t)^{-1}(\boldsymbol{X}_{t,\mathsf{c}'})^*\boldsymbol{q}^{t,\mathsf{r}}_{\perp,\mathsf{c}'} + \boldsymbol{Y}_{t,\mathsf{r}}((\boldsymbol{Q}^{\mathsf{r}}_t)^*\boldsymbol{Q}^{\mathsf{r}}_t)^{-1}(\boldsymbol{Q}^{\mathsf{r}}_{t,\mathsf{c}})^*\boldsymbol{q}^{t,\mathsf{r}}.\end{aligned}$$

Noting that $((\boldsymbol{Q}^{\mathsf{r}}_t)^*\boldsymbol{Q}^{\mathsf{r}}_t)^{-1}(\boldsymbol{Q}^{\mathsf{r}}_{t,\mathsf{c}})^*\boldsymbol{q}^{t,\mathsf{r}} = \boldsymbol{\gamma}^{t,\mathsf{r}}$ completes the proof of (76). The result (77) can be similarly obtained, by using the representation in (117) to express $\left[\mathbf{A}^*\boldsymbol{m}^{t,\mathsf{c}}\right]_{\mathsf{c}}$.

8.2 Proof of Lemma 7.4

Proof. We begin by demonstrating (79). By (52) it follows that

$$\boldsymbol{b}^0_{\mathsf{r}}|_{\mathscr{S}_{0,0}} = [\mathbf{A}\boldsymbol{q}^{0,\mathsf{r}}]_{\mathsf{r}} = \sum_{\mathsf{c}\in[\mathsf{C}]}\mathbf{A}_{\mathsf{rc}}\,\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}} \overset{d}{=} \boldsymbol{Z}'_{0,\mathsf{r}}\left(\frac{1}{\sqrt{L}}\left\|\boldsymbol{q}^{0,\mathsf{r}}\right\|\right),$$

where $\boldsymbol{Z}'_{0,\mathsf{r}} \sim \mathcal{N}(0, \boldsymbol{I}_{N_{\mathsf{R}}})$ independent of $\mathscr{S}_{0,0}$. Furthermore, we have

$$\frac{1}{L}\left\|\boldsymbol{q}^{0,\mathsf{r}}\right\|^2 = \frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}\left\|\breve{\boldsymbol{q}}^0_{\mathsf{c}}\right\|^2 = \frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}\left\|\boldsymbol{\beta}_{0,\mathsf{c}}\right\|^2 = \frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}\psi^0_{\mathsf{c}} = \sigma^0_{\mathsf{r}}. \tag{119}$$

For the case $t \geq 1$, we use (57) and (76) to write

$$\begin{aligned}\boldsymbol{b}_{\mathsf{r}}^t|_{\mathscr{S}_{t,t}} &= ([\mathbf{A}\boldsymbol{q}^{t,\mathsf{r}}]_{\mathsf{r}} + \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}\breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1})|_{\mathscr{S}_{t,t}} \\ &\stackrel{d}{=} \sum_{\mathsf{c}\in[\mathsf{C}]} [\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}}}]_{(\mathsf{r},\cdot)}\, \hat{\mathbf{A}}_{(\cdot,\mathsf{c})}\, \boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} + \boldsymbol{Y}_{t,\mathsf{r}}\boldsymbol{\gamma}^{t,\mathsf{r}} + \sum_{\mathsf{c}\in[\mathsf{C}]} \boldsymbol{M}_{t,\mathsf{r}}^{\mathsf{c}}((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*\, \boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} + \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}\breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1}.\end{aligned} \tag{120}$$

Since $\hat{\mathbf{A}}$ is independent of $\mathscr{S}_{t,t}$, with i.i.d. $\sim \mathcal{N}(0, \frac{1}{L})$ entries, the first term on the right side of (120) can be written as:

$$\sum_{\mathsf{c}\in[\mathsf{C}]} [\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}}}]_{(\mathsf{r},\cdot)}\, \hat{\mathbf{A}}_{(\cdot,\mathsf{c})}\, \boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} \stackrel{d}{=} \sum_{\mathsf{c}\in[\mathsf{C}]} \frac{\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|}{\sqrt{L}}\, [\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}}}\, \boldsymbol{Z}_t'^{\mathsf{c}}]_{\mathsf{r}} \stackrel{d}{=} \frac{\|\boldsymbol{q}_{\perp}^{t,\mathsf{r}}\|}{\sqrt{L}}\, \boldsymbol{Z}'_{t,\mathsf{r}} - \sum_{\mathsf{c}\in[\mathsf{C}]} \frac{\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|}{\sqrt{L}}\, [\mathbf{P}^{\|}_{\boldsymbol{M}_t^{\mathsf{c}}}\, \boldsymbol{Z}_t'^{\mathsf{c}}]_{\mathsf{r}}, \tag{121}$$

where in the first step, the vectors $\boldsymbol{Z}_t'^{\mathsf{c}} \sim \mathcal{N}(0, \boldsymbol{I}_n)$ are independent for $\mathsf{c} \in [\mathsf{C}]$. For the second equality we use $\mathbf{P}^{\perp}_{\boldsymbol{M}_t^{\mathsf{c}}} = \boldsymbol{I}_n - \mathbf{P}^{\|}_{\boldsymbol{M}_t^{\mathsf{c}}}$, and writing $\boldsymbol{Z}'_{t,\mathsf{r}} = \sum_{\mathsf{c}\in[\mathsf{C}]} \frac{\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|}{\|\boldsymbol{q}_{\perp}^{t,\mathsf{r}}\|}\boldsymbol{Z}_{t,\mathsf{r}}'^{\mathsf{c}}$, we note that $\boldsymbol{Z}'_{t,\mathsf{r}} \sim \mathcal{N}(0, \boldsymbol{I}_{N_{\mathsf{R}}})$. Using (121) and $\boldsymbol{Y}_{t,\mathsf{r}} = \boldsymbol{B}_{t,\mathsf{r}} - [0|\breve{\boldsymbol{M}}_{t-1,\mathsf{r}}]\boldsymbol{\Upsilon}_t^{\mathsf{r}}$, we simplify (120) to the following:

$$\begin{aligned}\boldsymbol{b}_{\mathsf{r}}^t|_{\mathscr{S}_{t,t}} \stackrel{d}{=}\; & \boldsymbol{B}_{t,\mathsf{r}}\boldsymbol{\gamma}^{t,\mathsf{r}} + \frac{1}{\sqrt{L}}\|\boldsymbol{q}_{\perp}^{t,\mathsf{r}}\|\, \boldsymbol{Z}'_{t,\mathsf{r}} - \frac{1}{\sqrt{L}}\sum_{\mathsf{c}\in[\mathsf{C}]} \|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|\, [\mathbf{P}^{\|}_{\boldsymbol{M}_t^{\mathsf{c}}}\, \boldsymbol{Z}_t'^{\mathsf{c}}]_{\mathsf{r}} \\ & + \sum_{\mathsf{c}\in[\mathsf{C}]} \boldsymbol{M}_{t,\mathsf{r}}^{\mathsf{c}}((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*\, \boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} - [\boldsymbol{0}\,|\,\breve{\boldsymbol{M}}_{t-1,\mathsf{r}}]\boldsymbol{\Upsilon}_t^{\mathsf{r}}\boldsymbol{\gamma}^{t,\mathsf{r}} + \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}\breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1}.\end{aligned} \tag{122}$$

All the quantities in the RHS of (122) except the vectors $\boldsymbol{Z}'_{t,\mathsf{r}}$ and $\boldsymbol{Z}_t'^{\mathsf{c}}$ are in the conditioning sigma-field. We can rewrite (120) with the following pair of values:

$$\begin{aligned}\boldsymbol{b}_{\mathsf{r}}^t|_{\mathscr{S}_{t,t}} &\stackrel{d}{=} \boldsymbol{b}_{\mathsf{r}}^{t-1}\frac{\sigma_{\mathsf{r}}^t}{\sigma_{\mathsf{r}}^{t-1}} + \sqrt{\sigma_{\perp,\mathsf{r}}^t}\boldsymbol{Z}'_{t,\mathsf{r}} + \boldsymbol{\Delta}_{t,t,\mathsf{r}}, \\ \boldsymbol{\Delta}_{t,t,\mathsf{r}} &= \sum_{i=0}^{t-2} \boldsymbol{b}_{\mathsf{r}}^i\gamma_i^{t,\mathsf{r}} + \boldsymbol{b}_{\mathsf{r}}^{t-1}\left[\gamma_{t-1}^{t,\mathsf{r}} - \frac{\sigma_{\mathsf{r}}^t}{\sigma_{\mathsf{r}}^{t-1}}\right] + \left[\frac{1}{\sqrt{L}}\|\boldsymbol{q}_{\perp}^{t,\mathsf{r}}\| - \sqrt{\sigma_{\perp,\mathsf{r}}^t}\right]\boldsymbol{Z}'_{t,\mathsf{r}} - \frac{1}{\sqrt{L}}\sum_{\mathsf{c}\in[\mathsf{C}]} \|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|\, [\mathbf{P}^{\|}_{\boldsymbol{M}_t^{\mathsf{c}}}\, \boldsymbol{Z}_t'^{\mathsf{c}}]_{\mathsf{r}} \\ &\quad + \sum_{\mathsf{c}\in[\mathsf{C}]} \boldsymbol{M}_{t,\mathsf{r}}^{\mathsf{c}}((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*\, \boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} - [\boldsymbol{0}\,|\,\breve{\boldsymbol{M}}_{t-1,\mathsf{r}}]\boldsymbol{\Upsilon}_t^{\mathsf{r}}\boldsymbol{\gamma}^{t,\mathsf{r}} + \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}\breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1}.\end{aligned}$$

To complete the proof of (79), we note that $[0|\breve{\boldsymbol{M}}_{t-1,\mathsf{r}}]\boldsymbol{\Upsilon}_t^{\mathsf{r}}\boldsymbol{\gamma}^{t,\mathsf{r}} = \sum_{i=1}^{t-1}\gamma_i^{t,\mathsf{r}}\frac{\sigma_{\mathsf{r}}^i}{\phi_{\mathsf{r}}^{i-1}}\breve{\boldsymbol{m}}_{\mathsf{r}}^{i-1}$.

For the result (78) we start by writing $\boldsymbol{h}_{\mathsf{c}}^{t+1} = [\mathbf{A}^*\boldsymbol{m}^{t,\mathsf{c}}]_{\mathsf{c}} - \breve{\boldsymbol{q}}_{\mathsf{c}}^t$, and using (77) for the conditional distribution of $[\mathbf{A}^*\boldsymbol{m}^{t,\mathsf{c}}]_{\mathsf{c}}$ given $\mathscr{S}_{t+1,t}$. We omit the proof as the argument is along similar lines as the above. ■

8.3 Proof of Lemma 7.6

The proof proceeds by induction on t. We label the results in (92) through (100) as step $\mathcal{B}_t$, and those in (101) through (109) as step $\mathcal{H}_{t+1}$. The proof consists of four steps, inductively showing that: (1) $\mathcal{B}_0$ holds, (2) $\mathcal{H}_1$ holds, (3) if $\mathcal{B}_{\tilde{s}}, \mathcal{H}_s$ hold for all $\tilde{s} < t$ and $s \leq t$, then $\mathcal{B}_t$ holds, and (4) if $\mathcal{B}_{\tilde{s}}, \mathcal{H}_s$ hold for all $\tilde{s} \leq t$ and $s \leq t$, then $\mathcal{H}_{t+1}$ holds. Appendix B lists a few basic concentration inequalities and other lemmas that are used in the proof.

8.3.1 Step 1: Showing $\mathcal{B}_0$ holds

(a) $\boldsymbol{\Delta}_{0,0} = \mathbf{0}$, so there is nothing to prove.
(b) First, $\boldsymbol{b}_{\mathsf{r}}^0 \overset{d}{=} \sqrt{\sigma_{\mathsf{r}}^0}\, \boldsymbol{Z}'_{0,\mathsf{r}}$ where $\boldsymbol{Z}'_{0,\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$ by Lemma 7.4. Further, $\boldsymbol{w}_{\mathsf{r}} \overset{d}{=} \sigma \mathbf{U}_{\mathsf{r}}$ where $\mathbf{U}_{\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$ independent of $\boldsymbol{Z}'_{0,\mathsf{r}}$. Hence,

$$\begin{aligned} &P\Big(\frac{1}{n}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{w}_{\mathsf{r}}\Big| \geq \epsilon\Big) \leq P\Big(\frac{1}{n}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} \sigma\sqrt{\sigma_{\mathsf{r}}^0} W_{\mathsf{rc}}^{\mathsf{u}}(\boldsymbol{Z}'_{0,\mathsf{r}})^*\mathbf{U}_{\mathsf{r}}\Big| \geq \epsilon\Big) \\ &\overset{(a)}{\leq} K\exp\Big\{\frac{-\kappa n\epsilon^2}{\max\{\frac{1}{\mathsf{R}}\sum_{\mathsf{r}} W_{\mathsf{rc}}^{2\mathsf{u}},\, \max_{\mathsf{r}} W_{\mathsf{rc}}^{\mathsf{u}},\, 1\}}\Big\} \overset{(b)}{\leq} K\exp\{-\kappa n(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}}\epsilon^2\}. \end{aligned} \tag{123}$$

In step (a), we use Lemma B.9 (254) and that $\sigma_{\mathsf{r}}^0 \leq 2P$ for $\mathsf{r} \in [\mathsf{R}]$. In step (b), we use $\max_{\mathsf{r},\mathsf{c}} W_{\mathsf{rc}}^{\mathsf{u}} \leq (P\mathsf{R}/\omega)^{\mathsf{u}}$, and $\frac{1}{\mathsf{R}}\sum_{\mathsf{r}} W_{\mathsf{rc}}^{2\mathsf{u}} \leq P(P\mathsf{R}/\omega)^{2\mathsf{u}-1}$ for $\mathsf{u} \in \{1,2\}$ (see (88)). Notice that this is consistent with the Π, π notation defined in (91) and used in the stated concentration results since $\Pi_{-1} = 1$ and $\pi_{-1} = N_{\mathsf{R}}\omega$, and in particular, $n(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}} = N_{\mathsf{R}}\omega(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}-1} \geq N_{\mathsf{R}}\omega(\omega/\mathsf{R})^{\max\{2(\mathsf{u}-1),0\}}$ (the only inequality is when $\mathsf{u} = 0$).
(c) Proving $\mathcal{B}_0(c)$ is similar to $\mathcal{B}_0(b)$ and results in the same bound: $\|\boldsymbol{b}_{\mathsf{r}}^0\|^2 \overset{d}{=} \sigma_{\mathsf{r}}^0\|\boldsymbol{Z}'_{0,\mathsf{r}}\|^2$ by Lemma 7.4, therefore the result follows from Lemma B.9 (253) and arguments as used in step (b) of (123).
(d) Recall, $\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}} = S_{\mathsf{rc}}^0\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^0 - \boldsymbol{w}_{\mathsf{r}})$, therefore $\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}} = S_{\mathsf{rc}}^0 W_{\mathsf{rc}}(\|\boldsymbol{b}_{\mathsf{r}}^0\|^2 - (\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{w}_{\mathsf{r}}$. Then,

$$\begin{aligned} &P\Big(\frac{N_{\mathsf{R}}}{L}\Big|\frac{1}{N_{\mathsf{R}}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}-\frac{1}{2}}(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}} - \sum_{\mathsf{r}\in[\mathsf{R}]} S_{\mathsf{rc}}^0 W_{\mathsf{rc}}^{\mathsf{u}}\,\sigma_{\mathsf{r}}^0\Big| \geq \epsilon\Big) \\ &= P\Big(\Big|\frac{n}{L\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} S_{\mathsf{rc}}^0 W_{\mathsf{rc}}^{\mathsf{u}}\Big(\frac{\|\boldsymbol{b}_{\mathsf{r}}^0\|^2}{N_{\mathsf{R}}} - \frac{(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{w}_{\mathsf{r}}}{N_{\mathsf{R}}} - \sigma_{\mathsf{r}}^0\Big)\Big| \geq \epsilon\Big) \\ &\overset{(a)}{\leq} P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}\Big(\frac{n}{L}S_{\mathsf{rc}}^0\Big)W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{\|\boldsymbol{b}_{\mathsf{r}}^0\|^2}{N_{\mathsf{R}}} - \sigma_{\mathsf{r}}^0\Big]\Big| \geq \frac{\epsilon}{2}\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}\Big(\frac{n}{L}S_{\mathsf{rc}}^0\Big)W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{w}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \geq \frac{\epsilon}{2}\Big) \\ &\overset{(b)}{\leq} K\exp\{-\kappa n\epsilon^2(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}}\}, \end{aligned} \tag{124}$$

where we have used Lemma B.2 in step (a), and step (b) follows by $\mathcal{B}_0(b)$ and $\mathcal{B}_0(c)$, using that fact that $\frac{n}{L}S_{\mathsf{rc}}^0$ is bounded above and below by positive constants for all (r,c).
(e) Proving $\mathcal{B}_0(e)$ is similar to $\mathcal{B}_0(d)$ and results in the same bound. We sketch the detais. First, $\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}} = S_{\mathsf{rc}}^0\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^0 - \boldsymbol{w}_{\mathsf{r}})$, therefore,

$$\sum_{\mathsf{r}\in[\mathsf{R}]}\|\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}}\|^2 = \sum_{\mathsf{r}\in[\mathsf{R}]}(S_{\mathsf{rc}}^0)^2 W_{\mathsf{rc}}\|\boldsymbol{b}_{\mathsf{r}}^0 - \boldsymbol{w}_{\mathsf{r}}\|^2 = \sum_{\mathsf{r}\in[\mathsf{R}]}(S_{\mathsf{rc}}^0)^2 W_{\mathsf{rc}}(\|\boldsymbol{b}_{\mathsf{r}}^0\|^2 + \|\boldsymbol{w}_{\mathsf{r}}\|^2 - 2(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{w}_{\mathsf{r}}),$$

Next, $\phi_{\mathsf{r}}^0 = (\sigma^2 + \sigma_{\mathsf{r}}^0)$, and letting $\boldsymbol{w}_{\mathsf{r}} \overset{d}{=} \sigma\mathbf{U}_{\mathsf{r}}$ where $\mathbf{U}_{\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$, the result follows by Lemma B.2, $\mathcal{B}_0(c)$, $\mathcal{B}_1(b)$, and Lemma B.9 as well as an argument as in step (b) of (123).
(f) There is nothing to prove for (97) since $\|\boldsymbol{m}_{\perp}^{0,\mathsf{c}}\|^2 = \|\boldsymbol{m}^{0,\mathsf{c}}\|^2$ and $\tau_{\perp,\mathsf{c}}^0 = \tau_{\mathsf{c}}^0$ and result (98) follows from $\mathcal{B}_0(e)$ with $\mathsf{u} = 1$.
(g) First, $\mathbb{M}_1^{\mathsf{c}} := \frac{n}{L^2}(\boldsymbol{M}_1^{\mathsf{c}})^*\boldsymbol{M}_1^{\mathsf{c}} = \frac{n}{L^2}\|\boldsymbol{m}^{0,\mathsf{c}}\|^2$ concentrates to $\frac{n}{L}\tau_{\mathsf{c}}^0$ by $\mathcal{B}_0(e)$ with $\mathsf{u} = 1$. Result (100) then follows from $\mathcal{B}_0(e)$ and Lemma B.7. By Fact C.1, if $\frac{n}{L^2}\|\boldsymbol{m}_{\perp}^{0,\mathsf{c}}\|^2 \geq c > 0$, then $\mathbb{M}_1^{\mathsf{c}}$ is invertible.

Note from $\mathcal{B}_0(f)$ that $\frac{n}{L^2}\|\boldsymbol{m}_\perp^{0,\mathsf{c}}\|^2$ concentrates on $\frac{n}{L}\tau_{\perp,\mathsf{c}}^0 = \frac{n}{L}\tau_\mathsf{c}^0$. From (12), since $\phi_\mathsf{r}^0 \geq \sigma^2$, we have $\frac{n}{L}\tau_\mathsf{c}^0 \geq \frac{\sigma^2}{P}$ for all c. Choosing $\kappa = \min\{1, \frac{n}{2L}\tau_{\perp,\mathsf{c}}^0\}$, we therefore have from $\mathcal{B}_0(f)$ that

$$P\Big(\mathbb{M}_1^\mathsf{c} \text{ singular}\Big) \leq P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{m}_\perp^{0,\mathsf{c}}\|^2 - \tau_{\perp,\mathsf{c}}^0\Big| \geq \kappa\Big) \leq K\exp\{-cN_\mathsf{R}\omega\}.$$

8.3.2 Step 2: Showing $\mathcal{H}_1$ holds

(a) We use the expression for $\boldsymbol{\Delta}_{1,0,\mathsf{c}}$ from Lemma 7.4, and write the second term in (80) as

$$\sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|}{\sqrt{L}}\,[\mathsf{P}_{\boldsymbol{Q}_1^\mathsf{r}}^\perp\,\boldsymbol{Z}_0^\mathsf{r}]_\mathsf{c} = \sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|}{\sqrt{L}}\Big[\frac{\boldsymbol{q}_\mathsf{c}^{0,\mathsf{r}}\,(\boldsymbol{q}^{0,\mathsf{r}})^*\,\boldsymbol{Z}_0^\mathsf{r}}{\|\boldsymbol{q}^{0,\mathsf{r}}\|^2}\Big] \stackrel{d}{=} \sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|}{\sqrt{L}}\Big[\frac{\boldsymbol{q}_\mathsf{c}^{0,\mathsf{r}}\,Z^\mathsf{r}}{\|\boldsymbol{q}^{0,\mathsf{r}}\|}\Big],$$

where $Z^\mathsf{r} \sim \mathcal{N}(0,1)$ are i.i.d. for $\mathsf{r}\in[\mathsf{R}]$. Recall, $\boldsymbol{q}_\mathsf{c}^{0,\mathsf{r}} = \sqrt{W_{\mathsf{rc}}}\breve{\boldsymbol{q}}_\mathsf{c}^0$ and $\big\|\boldsymbol{q}^{0,\mathsf{r}}\big\|^2 = L\sigma_\mathsf{r}^0$ from (119), so

$$\sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|}{\sqrt{L}}\Big[\frac{\boldsymbol{q}_\mathsf{c}^{0,\mathsf{r}}\,Z^\mathsf{r}}{\|\boldsymbol{q}^{0,\mathsf{r}}\|}\Big] = \breve{\boldsymbol{q}}_\mathsf{c}^0 \sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\sqrt{W_{\mathsf{rc}}}\,\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|Z^\mathsf{r}}{\sqrt{L}\|\boldsymbol{q}^{0,\mathsf{r}}\|} = Z\,\breve{\boldsymbol{q}}_\mathsf{c}^0\Big(\sum_{\mathsf{r}\in[\mathsf{R}]} \frac{W_{\mathsf{rc}}\,\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|^2}{L^2\,\sigma_\mathsf{r}^0}\Big)^{1/2},$$

where $Z\sim\mathcal{N}(0,1)$. Using $\max_{j\in sec(\ell)}\big|[\breve{\boldsymbol{q}}_\mathsf{c}^0]_j\big|^2 = \max_{j\in sec(\ell)}\big|[\boldsymbol{\beta}_{0,\mathsf{c}}]_j\big|^2 = 1$, for any section ℓ in c and the above, we have

$$\begin{aligned}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|^2 \leq \Big|\frac{\big\|\boldsymbol{m}^{0,\mathsf{c}}\big\|}{\sqrt{L}} - \sqrt{\tau_\mathsf{c}^0}\Big|^2 \max_{j\in sec(\ell)}|[\boldsymbol{Z}_{0,\mathsf{c}}]_j|^2 + |Z|^2\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{W_{\mathsf{rc}}\,\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|^2}{L^2\,\sigma_\mathsf{r}^0} \\ + \Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{\sqrt{W_{\mathsf{rc}}}}{L\sigma_\mathsf{r}^0}(\boldsymbol{b}_\mathsf{r}^0)^*\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}} - 1\Big|^2.\end{aligned} \tag{125}$$

We therefore have the following bound for $\mathsf{r}'\in[\mathsf{R}]$, and $\mathsf{v}=0$ or $\mathsf{v}=1$ using Lemma B.2.

$$\begin{aligned}&P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{r}'\mathsf{c}}^{2\mathsf{v}}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|^2 \geq \epsilon\Big) \\ &\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{LW_{\mathsf{r}'\mathsf{c}}^{2\mathsf{v}}}{n}\Big|\sqrt{\frac{n\|\boldsymbol{m}^{0,\mathsf{c}}\|^2}{L^2}} - \sqrt{\frac{n\tau_\mathsf{c}^0}{L}}\Big|^2\sum_{\ell\in\mathsf{c}}\max_{j\in sec(\ell)}|[\boldsymbol{Z}_{0,\mathsf{c}}]_j|^2 \geq \frac{\epsilon}{3}\Big) \\ &+ P\Big(\frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}}^{2\mathsf{v}}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{\sqrt{W_{\mathsf{rc}}}}{L\sigma_\mathsf{r}^0}(\boldsymbol{b}_\mathsf{r}^0)^*\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}} - 1\Big|^2 \geq \frac{\epsilon}{3}\Big) + P\Big(\frac{|Z|^2}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{r}'\mathsf{c}}^{2\mathsf{v}}\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{W_{\mathsf{rc}}\,\|\boldsymbol{m}_\mathsf{r}^{0,\mathsf{c}}\|^2}{L\,\sigma_\mathsf{r}^0} \geq \frac{\epsilon}{3}\Big).\end{aligned} \tag{126}$$

Label the terms on the RHS of (126) as T_1, T_2, T_3 and we bound each separately.

Consider term T_1 and let Π_0 be the event under consideration, so that $T_1 = P(\Pi_0)$, and define an event $\mathcal{F}$ as

$$\mathcal{F} = \Big\{\max_{\mathsf{c}\in[\mathsf{C}]}\frac{L}{n}\Big|\sqrt{\frac{n}{L}\,\|\boldsymbol{m}^{0,\mathsf{c}}\|^2} - \sqrt{\frac{n}{L}\tau_\mathsf{c}^0}\Big|^2 \geq \frac{\epsilon}{(\max_{\mathsf{c}'\in[\mathsf{C}]}W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}})9\log M}\Big\}. \tag{127}$$

With this definition, $T_1 = P(\Pi_0) \le P(\mathcal{F}) + P(\Pi_0|\mathcal{F}^c)$. First,

$$\begin{aligned} P(\mathcal{F}) &\le \sum_{\mathsf{c}\in[\mathsf{C}]} P\Big(\Big|\sqrt{\frac{n}{L}\|\boldsymbol{m}^{0,\mathsf{c}}\|^2} - \sqrt{\frac{n}{L}\tau_{\mathsf{c}}^0}\Big| \ge \sqrt{\frac{n\epsilon}{(\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}})9L\log M}}\Big) \\ &\overset{(a)}{\le} \sum_{\mathsf{c}\in[\mathsf{C}]} K\exp\Big\{\frac{-\kappa N_{\mathsf{R}}\omega(n\tau_{\mathsf{c}}^0/L)\epsilon}{\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}}}\Big\} \overset{(b)}{\le} K\mathsf{C}\exp\Big\{\frac{-\kappa N_{\mathsf{R}}\omega\epsilon}{\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}}}\Big\}, \end{aligned} \tag{128}$$

where step (a) uses $\mathcal{B}_0(f)$ and Lemma B.5, and step (b) the fact that $n\tau_{\mathsf{c}}^0/L \in \Theta(1)$ and $nR = L\log M$. Next, by Lemma B.9, using that $\boldsymbol{Z}_{0,\mathsf{c}}$ is independent of the event $\mathcal{F}$ for $\mathsf{c}\in[\mathsf{C}]$,

$$P(\Pi_0|\mathcal{F}^c) \le P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} \max_{j\in sec(\ell)} |[\boldsymbol{Z}_{0,\mathsf{c}}]_j|^2 \ge 3\log M\Big) \le K\exp\{-\kappa L\log M\}. \tag{129}$$

From (128) and (129), we find

$$T_1 \le K\mathsf{C}\exp\Big\{\frac{-\kappa N_{\mathsf{R}}\omega\epsilon}{\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}}}\Big\} \le K\mathsf{C}\exp\{-\kappa' n(\omega/\mathsf{R})^{1+2\mathsf{v}}\epsilon\}. \tag{130}$$

The final inequality uses $\max_{\mathsf{r},\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}} \le (P\mathsf{R}/\omega)^{2\mathsf{v}}$, thus $N_{\mathsf{R}}\omega/(\max_{\mathsf{r},\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}}) \ge \kappa N_{\mathsf{R}}\omega(\omega/\mathsf{R})^{2\mathsf{v}}$.

For term T_2, we use result $\mathcal{B}_0(d)$. To see this,

$$\begin{aligned} T_2 &\overset{(a)}{=} P\Big(\frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}}^{2\mathsf{v}}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\Big(\frac{\sqrt{W_{\mathsf{rc}}}}{L\sigma_{\mathsf{r}}^0}(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}} - \frac{N_{\mathsf{R}}}{L}S_{\mathsf{rc}}^0 W_{\mathsf{rc}}\Big)\Big|^2 \ge \frac{\epsilon}{3}\Big) \\ &\overset{(b)}{\le} \sum_{\mathsf{c}\in[\mathsf{C}]} P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\Big(\frac{\sqrt{W_{\mathsf{rc}}}}{L}(\boldsymbol{b}_{\mathsf{r}}^0)^*\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}} - \frac{N_{\mathsf{R}}}{L}S_{\mathsf{rc}}^0 W_{\mathsf{rc}}\sigma_{\mathsf{r}}^0\Big)\Big|^2 \ge \frac{\kappa\epsilon}{\sum_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}}/\mathsf{C}}\Big) \overset{(c)}{\le} K\mathsf{C}e^{-\kappa n(\omega/\mathsf{R})^{1+\mathsf{v}}\epsilon}. \end{aligned}$$

In the above, step (a) holds because $\frac{N_{\mathsf{R}}}{L}\sum_{\mathsf{r}\in[\mathsf{R}]} S_{\mathsf{rc}}^0 W_{\mathsf{rc}} = 1$, which can be verified using $S_{\mathsf{rc}}^0 = \tau_{\mathsf{c}}^0/\phi_{\mathsf{r}}^0$, and the state evolution equations in (10)–(12). Step (b) uses that $\sigma_{\mathsf{r}}^0 \in \Theta(1)$ and Lemma B.2. Step (c) follows from $\mathcal{B}_0(d)$ (Eq. (124)) with $\mathsf{u} = 1$, noting from (88) that $\sum_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}'}^{2\mathsf{v}}/\mathsf{C} = \Theta((\mathsf{R}/\omega)^{\mathsf{v}})$.

Finally we bound term T_3. Note that for any $\mathsf{r},\mathsf{r}'\in[\mathsf{R}]$, we have $\frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{r}'\mathsf{c}}^{2\mathsf{v}} = \Theta((\mathsf{R}/\omega)^{\mathsf{v}})$ since $\mathsf{v}\in\{0,1\}$ and $\sigma_{\mathsf{r}}^0 = \Theta(1)$. Furthermore, from $\mathcal{B}_0(e)$ with $\mathsf{u} = 2$, for $\mathsf{c}\in[\mathsf{C}]$ the term $\frac{n}{L^2}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}\|\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}}\|^2$ concentrates on $\frac{nN_{\mathsf{R}}}{L^2}\sum_{\mathsf{r}\in[\mathsf{R}]}(S_{\mathsf{rc}}^0)^2 W_{\mathsf{rc}}^2\phi_{\mathsf{r}}^0 = \Theta(\mathsf{R}/\omega)$, with the deviation probability bounded by $\mathcal{B}_0(e)$. We therefore have

$$\begin{aligned} T_3 &\le P\Big(\cup_{\mathsf{c}\in[\mathsf{C}]}\Big\{\frac{|Z|^2}{N_{\mathsf{R}}\omega}\kappa\Big(\frac{\mathsf{R}}{\omega}\Big)^{\mathsf{v}}\cdot\frac{n\omega}{L^2\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}\|\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}}\|^2 \ge \frac{\epsilon}{3}\Big\}\Big) \\ &\le \sum_{\mathsf{c}\in[\mathsf{C}]} P\Big(\frac{n\omega}{L^2\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}\|\boldsymbol{m}_{\mathsf{r}}^{0,\mathsf{c}}\|^2 \le 2\cdot\frac{n^2\omega}{L^2\mathsf{R}^2}\sum_{\mathsf{r}\in[\mathsf{R}]}(S_{\mathsf{rc}}^0)^2 W_{\mathsf{rc}}^2\phi_{\mathsf{r}}^0\Big) + P\Big(|Z|^2 \ge \frac{N_{\mathsf{R}}\omega\epsilon}{\kappa(\mathsf{R}/\omega)^{\mathsf{v}}}\Big) \\ &\overset{(a)}{\le} \mathsf{C}K\exp\{-\kappa n(\omega/\mathsf{R})\} + \exp\{-n(\omega/\mathsf{R})^{\mathsf{v}+1}\epsilon/\kappa\}, \end{aligned} \tag{131}$$

where step (a) follows from $\mathcal{B}_0(e)$ and Lemma B.8. The overall upper bound is $\exp\{-\kappa n(\omega/\mathsf{R})^{2\mathsf{v}+1}\epsilon\}$, which is again consistent with the Π,π notation defined in (91) for $t = 0$.

(b) From Lemma 7.4, we have $\boldsymbol{h}^1_{\mathsf{c}}|_{\mathscr{S}_{1,0}} \stackrel{d}{=} \sqrt{\tau^0_{\mathsf{c}}}\,\boldsymbol{Z}_{0,\mathsf{c}} + \boldsymbol{\Delta}_{1,0,\mathsf{c}}$. Using this, we obtain

$$P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\frac{(\boldsymbol{h}^1_{\mathsf{c}})^*\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}}{L}\Big|\geq\epsilon\Big)\leq P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}\tau^0_{\mathsf{c}}}\frac{\boldsymbol{Z}^*_{0,\mathsf{c}}\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}}{L}\Big|\geq\frac{\epsilon}{2}\Big)+P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big|\frac{\boldsymbol{\Delta}^*_{1,0,\mathsf{c}}\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}}{L}\Big|\geq\frac{\epsilon}{2}\Big). \tag{132}$$

Label the terms on the right side as T_1 and T_2.

Consider term T_1. Since $\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}$ is independent of $\boldsymbol{Z}_{0,\mathsf{c}}$, we have $\boldsymbol{Z}^*_{0,\mathsf{c}}\,\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}} \stackrel{d}{=} \|\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}\|Z^{\mathsf{c}}$, where $Z^{\mathsf{c}}\sim\mathcal{N}(0,1)$ and $\|\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}\|^2 = W_{\mathsf{rc}}\|\boldsymbol{\beta}_{0,\mathsf{c}}\|^2 = W_{\mathsf{rc}}(L/\mathsf{C})$. Therefore,

$$\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}\tau^0_{\mathsf{c}}}\,\boldsymbol{Z}^*_{0,\mathsf{c}}\,\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}\Big| \stackrel{d}{=} \Big|\sqrt{\frac{L}{\mathsf{C}}}\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{\tau^0_{\mathsf{c}}}\,W_{\mathsf{rc}}\,Z^{\mathsf{c}}\Big| \stackrel{d}{=} \Big|Z\Big(\frac{L}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}\tau^0_{\mathsf{c}}\,W^2_{\mathsf{rc}}\Big)^{\frac{1}{2}}\Big| \leq \kappa\,|Z|\,\sqrt{\frac{L^2\mathsf{R}}{n\omega}},$$

where $Z\sim\mathcal{N}(0,1)$. Recall, $n = N_{\mathsf{R}}\mathsf{R}$, therefore $T_1 \leq P(\kappa\,|Z|\,\sqrt{\mathsf{R}/(n\omega)}\geq\frac{\epsilon}{2}) \leq \exp\{-\kappa N_{\mathsf{R}}\omega\epsilon^2\}$.

For term T_2, observe that $\sum_{i\in\mathrm{sec}(\ell)}[\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}]_i = -\sqrt{W_{\mathsf{rc}}}$ for each $\mathsf{c}\in[\mathsf{C}]$ and section $\ell\in\mathsf{c}$,

$$\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\,|\boldsymbol{\Delta}^*_{1,0,\mathsf{c}}\,\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}| = \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\sqrt{W_{\mathsf{rc}}}\Big|([\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_\ell)^*\,[\boldsymbol{q}^{0,\mathsf{r}}_{\mathsf{c}}]_\ell\Big| \leq \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}\max_{i\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_i|\,.$$

Now $(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}\max_{i\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_i|)^2 \leq \frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W^2_{\mathsf{rc}}\max_{i\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_i|^2$ by Lemma C.2 and therefore by $\mathcal{H}_1(a)$ with $\mathsf{v}=1$,

$$T_2 \leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W^2_{\mathsf{rc}}\max_{i\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_i|^2\geq\kappa\epsilon^2\Big) \leq K\mathsf{C}\exp\{-\kappa n(\omega/\mathsf{R})^3\epsilon^2\}. \tag{133}$$

(c) We begin by showing the result (103) for $s=1$. Recall that $\eta^0_{\mathsf{c}}(\boldsymbol{\beta}_{0,\mathsf{c}}-\boldsymbol{h}^1_{\mathsf{c}})-\boldsymbol{\beta}_{0,\mathsf{c}}$ and $\boldsymbol{h}^1_{\mathsf{c}}|_{\mathscr{S}_{1,0}} \stackrel{d}{=} \sqrt{\tau^0_{\mathsf{c}}}\,\boldsymbol{Z}_{0,\mathsf{c}}+\boldsymbol{\Delta}_{1,0,\mathsf{c}}$ from Lemma 7.4. Therefore, by Lemma B.2,

$$\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}W^{\mathsf{v}}_{\mathsf{rc}}\Big(\frac{\|\breve{\boldsymbol{q}}^1_{\mathsf{c}}\|^2}{L}-\frac{\psi^1_{\mathsf{c}}}{\mathsf{C}}\Big)\Big|\geq\epsilon\Big)\\
&=P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}W^{\mathsf{v}}_{\mathsf{rc}}\Big(\frac{\mathsf{C}}{L}\Big\|\eta^0_{\mathsf{c}}(\boldsymbol{\beta}_{0,\mathsf{c}}-\sqrt{\tau^0_{\mathsf{c}}}\boldsymbol{Z}_{0,\mathsf{c}}-\boldsymbol{\Delta}_{1,0,\mathsf{c}})-\boldsymbol{\beta}_{0,\mathsf{c}}\Big\|^2-\psi^1_{\mathsf{c}}\Big)\Big|\geq\epsilon\Big)\\
&\leq P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}W^{\mathsf{v}}_{\mathsf{rc}}\Big(\frac{\mathsf{C}}{L}\Big\|\eta^0_{\mathsf{c}}(\boldsymbol{\beta}_{0,\mathsf{c}}-\sqrt{\tau^0_{\mathsf{c}}}\boldsymbol{Z}_{0,\mathsf{c}})-\boldsymbol{\beta}_{0,\mathsf{c}}\Big\|^2-\psi^1_{\mathsf{c}}\Big)\Big|\geq\frac{\epsilon}{2}\Big)\\
&+P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}W^{\mathsf{v}}_{\mathsf{rc}}\Big(\Big\|\eta^0_{\mathsf{c}}(\boldsymbol{\beta}_{0,\mathsf{c}}-\sqrt{\tau^0_{\mathsf{c}}}\boldsymbol{Z}_{0,\mathsf{c}}-\boldsymbol{\Delta}_{1,0,\mathsf{c}})-\boldsymbol{\beta}_{0,\mathsf{c}}\Big\|^2-\Big\|\eta^0_{\mathsf{c}}(\boldsymbol{\beta}_{0,\mathsf{c}}-\sqrt{\tau^0_{\mathsf{c}}}\boldsymbol{Z}_{0,\mathsf{c}})-\boldsymbol{\beta}_{0,\mathsf{c}}\Big\|^2\Big)\Big|\geq\frac{\epsilon}{2}\Big).
\end{aligned}$$

Label the two terms on the RHS as T_1 and T_2.

To bound T_1, we write

$$T_1 = P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}W^{\mathsf{v}}_{\mathsf{rc}}\sum_{\ell\in\mathsf{c}}\Big(\|\eta^0_\ell(\boldsymbol{\beta}_{0,\mathsf{c}}-\sqrt{\tau^0_{\mathsf{c}}}\boldsymbol{Z}_{0,\mathsf{c}})-[\boldsymbol{\beta}_{0,\mathsf{c}}]_\ell\|^2-\psi^1_{\mathsf{c}}\Big)\Big|\geq\frac{\epsilon}{2}\Big), \tag{134}$$

and apply Hoeffding's inequality (Lemma B.1). To do this, notice that $\big(\|\eta_\ell^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_{0,\mathsf{c}}) - [\boldsymbol{\beta}_{0,\mathsf{c}}]_\ell\|^2 - \psi_\mathsf{c}^1\big)$ is bounded in absolute value by 1, and has zero mean. Indeed,

$$\mathbb{E}_{\boldsymbol{Z}_0}\|\eta_\ell^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_{0,\mathsf{c}}) - [\boldsymbol{\beta}_{0,\mathsf{c}}]_\ell\|^2 = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{Z}_0}\|\eta_\ell^0(\boldsymbol{\beta}_\mathsf{c} - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_{0,\mathsf{c}}) - [\boldsymbol{\beta}_\mathsf{c}]_\ell\|^2 = \psi_\mathsf{c}^1, \tag{135}$$

where the first equality is true for each $\boldsymbol{\beta}_0 \in \mathcal{B}_{M,L}$ because of the uniform distribution of the non-zero entry in each section of $\boldsymbol{\beta}_0$ over the M possible locations and the i.i.d. distribution of $\boldsymbol{Z}_{0,\mathsf{c}}$. The second equality follows by Lemma C.4. Applying Hoeffding's inequality to (134), we obtain

$$T_1 \le 2\exp\Big\{\frac{-\kappa L^2\epsilon^2}{(\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}})}\Big\} = 2\exp\Big\{\frac{-\kappa L\epsilon^2}{(\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}/\mathsf{C})}\Big\} = 2\exp\{-\kappa L(\omega/\mathsf{R})^{\mathsf{v}}\epsilon^2\}, \tag{136}$$

where for the last equality we have used (88).

Next, we bound term T_2. To save space, we write $\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0})$ to mean $\eta_\mathsf{c}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_{0,\mathsf{c}} - \boldsymbol{\Delta}_{1,0,\mathsf{c}})$. First,

$$\begin{aligned}
&\|\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \boldsymbol{\beta}_{0,\mathsf{c}}\|^2 - \|\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0) - \boldsymbol{\beta}_{0,\mathsf{c}}\|^2 \\
&= [\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \boldsymbol{\beta}_{0,\mathsf{c}}]^*[\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0)] \\
&\quad + [\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0) - \boldsymbol{\beta}_{0,\mathsf{c}}]^*[\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0)].
\end{aligned} \tag{137}$$

Using this, we have

$$\begin{aligned}
T_2 &\le P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{\mathsf{v}}\Big|[\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0) - \boldsymbol{\beta}_{0,\mathsf{c}}]^*[\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0)]\Big| \ge \frac{\kappa\epsilon}{4}\Big) \\
&+ P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{\mathsf{v}}\Big|[\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \beta_{0,\mathsf{c}}]^*[\eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \eta_\mathsf{c}^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0)]\Big| \ge \frac{\kappa\epsilon}{4}\Big).
\end{aligned}$$

Label the terms on the right side above as $T_{2,a}$ and $T_{2,b}$. Then,

$$\begin{aligned}
T_{2,a} &\overset{(a)}{\le} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\sum_{j\in sec(\ell)} W_{\mathsf{rc}}^{\mathsf{v}}\Big|\eta_j^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \eta_j^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0)\Big| \ge \frac{\kappa\epsilon}{4}\Big) \\
&\overset{(b)}{\le} P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{W_{\mathsf{rc}}^{\mathsf{v}}}{L\tau_\mathsf{c}^0}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j| \ge \frac{\kappa\epsilon}{8}\Big) \\
&\overset{(c)}{\le} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|^2 \ge \frac{\kappa^2\epsilon^2}{(\log M)^2}\Big) \overset{(d)}{\le} K\mathsf{C}\exp\Big\{\frac{-\kappa n(\omega/\mathsf{R})^{2\mathsf{v}+1}\epsilon^2}{(\log M)^2}\Big\}.
\end{aligned} \tag{138}$$

Step (a) holds since $|\eta_j^0(\boldsymbol{\beta}_0 - \sqrt{\tau_\mathsf{c}^0}\boldsymbol{Z}_0) - \beta_{0,j}| \le 1$ for $j \in [ML]$, step (b) by Lemma C.5, and step (c) by Lemma C.2 and $\tau_\mathsf{c}^0 = \Theta(1/\log M)$. Finally, step (d) follows from $\mathcal{H}_1(a)$.

Using Lemma C.5, it can be shown that term $T_{2,b}$ also has the same upper bound. This proves the concentration result (103) for $s = 1$. Proving the result for $s = 0$ is similar: we use Lemma B.2 followed by Hoeffding's inequality and Lemma C.5.

(d) Recalling $\boldsymbol{q}_{\mathsf{c}}^{1,\mathsf{r}} = \sqrt{W_{\mathsf{rc}}}[\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_0 - \boldsymbol{h}^1) - \boldsymbol{\beta}_{0,\mathsf{c}}]$, we write $\sum_{\mathsf{c}\in[\mathsf{C}]}(\boldsymbol{h}_{\mathsf{c}}^1)^*\boldsymbol{q}_{\mathsf{c}}^{1,\mathsf{r}} = \sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^1)^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_0 - \boldsymbol{h}^1) + (\boldsymbol{h}_{\mathsf{c}}^1)^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}$, and therefore, by Lemma B.2,

$$
\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\Big(\frac{\sqrt{W_{\mathsf{rc}}}}{L}(\boldsymbol{h}_{\mathsf{c}}^1)^*\boldsymbol{q}_{\mathsf{c}}^{1,\mathsf{r}} + \frac{W_{\mathsf{rc}}}{\mathsf{C}}\,\psi_{\mathsf{c}}^1\Big)\Big| \geq \epsilon\Big)\\
&\leq P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big[\frac{\mathsf{C}}{L}(\boldsymbol{h}_{\mathsf{c}}^1)^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_0 - \boldsymbol{h}^1) + \psi_{\mathsf{c}}^1\Big]\Big| \geq \frac{\epsilon}{2}\Big) + P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^1)^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}\Big| \geq \frac{\epsilon}{2}\Big). \qquad (139)
\end{aligned}
$$

By $\mathcal{H}_1(b)$, the second term in (139) is bounded by $K\mathsf{C}\exp\{-\kappa n(\omega/\mathsf{R})^3\epsilon^2\}$. Using the conditional distribution of $\boldsymbol{h}^1$ stated in Lemma 7.4 and Lemma B.2, for the first term of (139) we write

$$
\begin{aligned}
&P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big(\frac{\mathsf{C}}{L}(\boldsymbol{h}_{\mathsf{c}}^1)^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_0 - \boldsymbol{h}^1) + \psi_{\mathsf{c}}^1\Big)\Big| \geq \frac{\epsilon}{2}\Big)\\
&\leq P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big(\frac{\mathsf{C}}{L}(\sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}} + \boldsymbol{\Delta}_{1,0,\mathsf{c}})^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}} - \boldsymbol{\Delta}_{1,0,\mathsf{c}}) + \psi_{\mathsf{c}}^1\Big)\Big| \geq \epsilon\Big)\\
&\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big|(\boldsymbol{\Delta}_{1,0,\mathsf{c}})^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}} - \boldsymbol{\Delta}_{1,0,\mathsf{c}})\Big| \geq \kappa\epsilon\Big)\\
&\quad + P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big(\frac{\mathsf{C}}{L}\sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}}^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}}) + \psi_{\mathsf{c}}^1\Big)\Big| \geq \kappa\epsilon\Big) \qquad (140)\\
&\quad + P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{\tau_{\mathsf{c}}^0}W_{\mathsf{rc}}\Big|\boldsymbol{Z}_{0,\mathsf{c}}^*\Big[\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}} - \boldsymbol{\Delta}_{1,0,\mathsf{c}}) - \eta_{\mathsf{c}}(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}})\Big]\Big| \geq \kappa\epsilon\Big).
\end{aligned}
$$

Label the terms of the above as T_1, T_2, T_3 and we bound each individually.

For term T_1 notice that $T_1 \leq P(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}\max_{i\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_i| \geq \kappa\epsilon)$ and therefore the term can be bounded as in (133) using $\mathcal{H}_1(a)$ with $\mathsf{v} = 1$.

Next consider term T_2 of (140). Because of the uniform distribution of the non-zero entry in each section of $\boldsymbol{\beta}$ over the M possible locations and the i.i.d. distribution of $\boldsymbol{Z}_{0,\mathsf{c}}$, for any $\boldsymbol{\beta}_0 \in \mathcal{B}_{M,L}$, we have $\mathbb{E}_{\boldsymbol{Z}_0}\{\boldsymbol{Z}_{0,\mathsf{c}}^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}})\} = \mathbb{E}_{\boldsymbol{Z}_0,\boldsymbol{\beta}}\{\boldsymbol{Z}_{0,\mathsf{c}}^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}})\}$. The expectation equals

$$
\sqrt{\tau_{\mathsf{c}}^0}\mathbb{E}\{\boldsymbol{Z}_{0,\mathsf{c}}^*\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}})\} \overset{(a)}{=} \mathbb{E}\|\eta_{\mathsf{c}}^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}})\|^2 - \frac{L}{\mathsf{C}} \overset{(b)}{=} -\psi_{\mathsf{c}}^1\frac{L}{\mathsf{C}}, \qquad (141)
$$

where equality (a) is obtained using Stein's lemma, Lemma C.3 (see [4, p.1491, Eqs. (102) – (104)] for details) and equality (b) from Lemma C.4. Now, using (141) and the fact that $\tau_{\mathsf{c}}^0 = \Theta(1/\log M)$, the concentration result from Lemma C.7 yields

$$
\begin{aligned}
T_2 &= P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{\tau_{\mathsf{c}}^0}W_{\mathsf{rc}}\sum_{\ell\in\mathsf{c}}\Big[([\boldsymbol{Z}_{0,\mathsf{c}}]_\ell)^*\eta_\ell^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}}) - \mathbb{E}\{([\boldsymbol{Z}_{0,\mathsf{c}}]_\ell)^*\eta_\ell^0(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_{0,\mathsf{c}})\}\Big]\Big| \geq \kappa\epsilon\Big)\\
&\leq \exp\{-\kappa L(\omega/\mathsf{R})\epsilon^2\}.
\end{aligned}
$$

Finally consider term T_3 in (140). First,

$$
\begin{aligned}
T_3 &\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{\tau_{\mathsf{c}}^0}W_{\mathsf{rc}}\sum_{\ell\in\mathsf{c}}\max_{k\in sec(\ell)}|[\boldsymbol{Z}_0]_k|\sum_{j\in sec(\ell)}|\eta_j^0(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_0 - \boldsymbol{\Delta}_{1,0}) - \eta_j^0(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^0}\boldsymbol{Z}_0)| \geq \kappa\epsilon\Big)\\
&\overset{(a)}{\leq} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{W_{\mathsf{rc}}}{\sqrt{\tau_{\mathsf{c}}^0}}\sum_{\ell\in\mathsf{c}}\max_{k\in sec(\ell)}|[\boldsymbol{Z}_0]_k|\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0}]_j| \geq \kappa\epsilon\Big) \overset{(b)}{\leq} K\mathsf{C}\exp\Big\{\frac{-\kappa n(\omega/\mathsf{R})^3\epsilon^2}{(\log M)^2}\Big\}. \qquad (142)
\end{aligned}
$$

Step (a) follows from Lemma C.5 and step (b) is obtained as follows.

$$
\begin{aligned}
&P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{W_{\mathsf{rc}}}{\sqrt{\tau_{\mathsf{c}}^0}}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|\max_{k\in sec(\ell)}|[\boldsymbol{Z}_{0,\mathsf{c}}]_k|\geq\kappa\epsilon\Big)\\
&\overset{(c)}{\leq} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|\max_{k\in sec(\ell)}|[\boldsymbol{Z}_{0,\mathsf{c}}]_k|\geq\kappa\epsilon\sqrt{L/n}\Big)\\
&\overset{(d)}{\leq} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\max_{k\in sec(\ell)}|[Z_{0,\mathsf{c}}]_k|^2\cdot\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^2\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|^2\geq\frac{\kappa\epsilon^2L}{n}\Big)\\
&\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\max_{k\in sec(\ell)}|[Z_{0,\mathsf{c}}]_k|^2\geq 3\log M\Big)+P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^2\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j|^2\geq\frac{\kappa\epsilon^2L}{3n\log M}\Big)\\
&\overset{(e)}{\leq} e^{-\kappa n}+K\mathsf{C}\exp\Big\{\frac{-\kappa n(\omega/\mathsf{R})^3\epsilon^2}{(\log M)^2}\Big\}.
\end{aligned}
\tag{143}
$$

Step (c) follows by using $\tau_{\mathsf{c}}^0=\Theta(L/n)$ and step (d) by Cauchy-Schwarz. Finally step (e) follows from Lemma B.9, $\mathcal{H}_1(a)$ with $\mathsf{v}=1$ along with $nR=L\log M$ and $n=\mathsf{R}N_{\mathsf{R}}$.

(e) From the conditional distribution of $\boldsymbol{h}_{\mathsf{c}}^1$ stated in Lemma 7.4 and Lemma C.2, it follows that $([\boldsymbol{h}_{\mathsf{c}}^1]_j)^2\leq 2\tau_{\mathsf{c}}^0([\boldsymbol{Z}_{0,\mathsf{c}}]_j)^2+2([\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j)^2$ for $j\in sec(\ell)$. We therefore have the following bound:

$$
\begin{aligned}
&P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\boldsymbol{h}_{\mathsf{c}}^1]_j)^2\geq 6(\max_{\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}})\tau_{\mathsf{c}}^0\log M+\epsilon\Big)\\
&\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}}\Big[\tau_{\mathsf{c}}^0\max_{j\in sec(\ell)}([\boldsymbol{Z}_{0,\mathsf{c}}]_j)^2+\max_{j\in sec(\ell)}([\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j)^2\Big]\geq 3(\max_{\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}})\tau_{\mathsf{c}}^0\log M+\epsilon/2\Big)\\
&\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\boldsymbol{Z}_{0,\mathsf{c}}]_j)^2\geq 3(\max_{\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}})\log M\Big)+P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\max_{j\in sec(\ell)}W_{\mathsf{rc}}^{2\mathsf{v}}([\boldsymbol{\Delta}_{1,0,\mathsf{c}}]_j)^2\geq\frac{\epsilon}{2}\Big)\\
&\overset{(a)}{\leq} 2\exp\{-\kappa L\log M\}+K\mathsf{C}\exp\{-\kappa n(\omega/\mathsf{R})^{2\mathsf{v}+1}\epsilon\}.
\end{aligned}
\tag{144}
$$

The inequality (a) follows from Lemma B.9 and $\mathcal{H}_1(a)$.

(f) We first prove (106), then (107). Recall, $\gamma_0^{1,\mathsf{r}}=\frac{(\boldsymbol{q}^{0,\mathsf{r}})^*\boldsymbol{q}^{1,\mathsf{r}}}{\|\boldsymbol{q}^{0,\mathsf{r}}\|^2}=\frac{(\boldsymbol{q}^{0,\mathsf{r}})^*\boldsymbol{q}^{1,\mathsf{r}}}{L\sigma_{\mathsf{r}}^0}$. Then since $\gamma_0^{1,\mathsf{r}}=\sigma_{\mathsf{r}}^1/\sigma_{\mathsf{r}}^0$, result (106) follows directly from $\mathcal{H}_1(c)$ with $\mathsf{v}=1$ as $\frac{1}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}\psi_{\mathsf{c}}^1=\sigma_{\mathsf{r}}^1$ and $\sigma_{\mathsf{r}}^0\in\Theta(1)$:

$$
P\Big(\Big|\gamma_0^{1,\mathsf{r}}-\frac{\sigma_{\mathsf{r}}^1}{\sigma_{\mathsf{r}}^0}\Big|\geq\epsilon\Big)=P\Big(\Big|\frac{1}{L}(\boldsymbol{q}^{0,\mathsf{r}})^*\boldsymbol{q}^{1,\mathsf{r}}-\sigma_{\mathsf{r}}^1\Big|\geq\epsilon\sigma_{\mathsf{r}}^0\Big)\leq K\mathsf{C}\exp\Big\{\frac{-\kappa n(\omega/\mathsf{R})^3\epsilon^2}{(\log M)^2}\Big\}.
\tag{145}
$$

For (107), by definition, $\|\boldsymbol{q}_\perp^{1,\mathsf{r}}\|^2=\|\boldsymbol{q}^{1,\mathsf{r}}\|^2-(\gamma_0^{1,\mathsf{r}})^2\|\boldsymbol{q}^{0,\mathsf{r}}\|^2=\|\boldsymbol{q}^{1,\mathsf{r}}\|^2-(\gamma_0^{1,\mathsf{r}})^2L\sigma_{\mathsf{r}}^0$. Using this and the fact that $\sigma_{\perp,\mathsf{r}}^1=\sigma_{\mathsf{r}}^1[1-(\sigma_{\mathsf{r}}^1/\sigma_{\mathsf{r}}^0)]$, we obtain the following upper bound:

$$
\begin{aligned}
&P\Big(\Big|\frac{\|\boldsymbol{q}_\perp^{1,\mathsf{r}}\|^2}{L}-\sigma_{\perp,\mathsf{r}}^1\Big|\geq\epsilon\Big)=P\Big(\Big|\frac{\|\boldsymbol{q}^{1,\mathsf{r}}\|^2}{L}-(\gamma_0^{1,\mathsf{r}})^2\sigma_{\mathsf{r}}^0-\sigma_{\mathsf{r}}^1\Big[1-\frac{\sigma_{\mathsf{r}}^1}{\sigma_{\mathsf{r}}^0}\Big]\Big|\geq\epsilon\Big)\\
&\leq P\Big(\Big|\frac{\|\boldsymbol{q}^{1,\mathsf{r}}\|^2}{L}-\sigma_{\mathsf{r}}^1\Big|\geq\frac{\epsilon}{2}\Big)+P\Big(\Big|(\gamma_0^{1,\mathsf{r}})^2-\frac{(\sigma_{\mathsf{r}}^1)^2}{(\sigma_{\mathsf{r}}^0)^2}\Big|\geq\frac{\epsilon}{2\sigma_{\mathsf{r}}^0}\Big)\leq K\mathsf{C}\exp\Big\{\frac{-\kappa n(\omega/\mathsf{R})^3\epsilon^2}{(\log M)^2}\Big\}.
\end{aligned}
$$

The second inequality follows from $\mathcal{H}_1(c)$ with $\mathsf{v}=1$ and (145) above, along with Lemma B.6.

(g) Note that $\left\|\boldsymbol{q}^{0,\mathsf{r}}\right\|^2 = L\sigma_{\mathsf{r}}^0$ by (119) and, therefore, $\boldsymbol{Q}_1^{\mathsf{r}} := \frac{1}{L}\left\|\boldsymbol{q}^{0,\mathsf{r}}\right\|^2 = \sigma_{\mathsf{r}}^0 = \tilde{C}^1$. So results (108) and (109) are trivially true for $\boldsymbol{Q}_1^{\mathsf{r}}$.

We now show (108) for $\boldsymbol{Q}_2^{\mathsf{r}}$. Recall, $[\boldsymbol{Q}_2^{\mathsf{r}}]_{\tilde{s}+1,s+1} = \frac{1}{L}(\boldsymbol{q}^{\tilde{s},\mathsf{r}})^*\boldsymbol{q}^{s,\mathsf{r}}$ for $0 \le \tilde{s}, s \le 1$ and, therefore, by $\mathcal{H}_1(c)$ with $\mathsf{v}=1$, $[\boldsymbol{Q}_2^{\mathsf{r}}]_{\tilde{s}+1,s+1}$ concentrates on $\sigma_{\mathsf{r}}^{\max\{\tilde{s},s\}}$. By Lemma C.1, if $\frac{1}{L}\|\boldsymbol{q}_\perp^{s,\mathsf{r}}\|^2 \ge \kappa > 0$ for all $0 \le s \le 1$, then $\boldsymbol{Q}_2^{\mathsf{r}}$ is invertible. By $\mathcal{H}_1(f)$, $\frac{1}{L}\left\|\boldsymbol{q}_\perp^{s,\mathsf{r}}\right\|^2$ concentrates on $\sigma_{\perp,\mathsf{r}}^s$, and $\sigma_{\perp,\mathsf{r}}^s > 0$ by Lemma 7.1. Choosing $\kappa = \frac{1}{2}\min\{2, \sigma_{\perp,\mathsf{r}}^0, \sigma_{\perp,\mathsf{r}}^1\}$, using $\mathcal{H}_1(f)$, we therefore have

$$P\Big(\boldsymbol{Q}_2^{\mathsf{r}} \text{ singular}\Big) \le P\Big(\Big|\frac{\|\boldsymbol{q}_\perp^{0,\mathsf{r}}\|^2}{L} - \sigma_{\perp,\mathsf{r}}^0\Big| \ge \kappa\Big) + P\Big(\Big|\frac{\|\boldsymbol{q}_\perp^{1,\mathsf{r}}\|^2}{L} - \sigma_{\perp,\mathsf{r}}^1\Big| \ge \kappa\Big) \le K\mathsf{C}\exp\Big\{\frac{-\kappa n(\omega/\mathsf{R})^3\epsilon^2}{(\log M)^2}\Big\}.$$

Now we show (109). Since

$$(\boldsymbol{Q}_2^{\mathsf{r}})^{-1} = \frac{L}{\|\boldsymbol{q}^{0,\mathsf{r}}\|^2\|\boldsymbol{q}^{1,\mathsf{r}}\|^2 - ((\boldsymbol{q}^{0,\mathsf{r}})^*\boldsymbol{q}^{1,\mathsf{r}})^2}\begin{bmatrix}\|\boldsymbol{q}^{1,\mathsf{r}}\|^2 & -(\boldsymbol{q}^{0,\mathsf{r}})^*\boldsymbol{q}^{1,\mathsf{r}} \\ -(\boldsymbol{q}^{1,\mathsf{r}})^*\boldsymbol{q}^{0,\mathsf{r}} & \|\boldsymbol{q}^{0,\mathsf{r}}\|^2\end{bmatrix},$$

and

$$(\widetilde{\boldsymbol{C}}^{2,\mathsf{r}})^{-1} = \frac{1}{\sigma_{\mathsf{r}}^0\sigma_{\mathsf{r}}^1 - (\sigma_{\mathsf{r}}^1)^2}\begin{bmatrix}\sigma_{\mathsf{r}}^1 & -\sigma_{\mathsf{r}}^1 \\ -\sigma_{\mathsf{r}}^1 & \sigma_{\mathsf{r}}^0,\end{bmatrix}.$$

element-wise concentration of $(\boldsymbol{Q}_2^{\mathsf{r}})^{-1}$ to $(\widetilde{\boldsymbol{C}}^{2,\mathsf{r}})^{-1}$ follows from $\mathcal{H}_1(c)$ with $\mathsf{v}=1$ using Lemma B.2, Lemma B.3, and Lemma B.7.

8.4 Step 3: Showing $\mathcal{B}_t$ holds

We prove the statements in $\mathcal{B}_t$ assuming that $\mathcal{B}_0, \ldots, \mathcal{B}_{t-1}$, and $\mathcal{H}_1, \ldots, \mathcal{H}_t$ hold due to the induction hypothesis. We begin with a lemma that is used to prove $\mathcal{B}_t(a)$. The lemma as well as other parts of $\mathcal{B}_t$ assume the invertibility of $\mathbb{M}_1^{\mathsf{c}}, \ldots, \mathbb{M}_t^{\mathsf{c}}$ for all $\mathsf{c} \in [\mathsf{C}]$, but for the sake of brevity, we do not explicitly specify the conditioning. The induction hypothesis implies that for $0 \le s \le (t-1)$, the large deviation bound in (92) gives

$$P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\|\boldsymbol{\Delta}_{s,s,\mathsf{r}}\|^2 \ge \epsilon\Big) \le KK_{t-1}\Pi_{t-2}\exp\{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^{(\mathsf{u}-1)_+}\pi_{t-2}\epsilon\}.$$

Similarly, for $0 \le s \le (t-1)$, we can use the values K_{t-1}, κ_{t-1} in prefactor and exponent, respectively, of the bounds given in (93)-(109).

Lemma 8.1. *Let* $\mathbb{M}_t^{\mathsf{c}} := \frac{n}{L^2}(\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}}$ *and* $\mathbf{v}^{\mathsf{r},\mathsf{c}} = \frac{n}{L^2}(\boldsymbol{X}_{t,\mathsf{c}})^*\,\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}$. *If* $\mathbb{M}_1^{\mathsf{c}}, \ldots, \mathbb{M}_t^{\mathsf{c}}$ *are invertible, we have for* $j \in [t-1]$ *and for all* $\mathsf{r} \in [\mathsf{R}]$,

$$P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j - \gamma_j^{t,\mathsf{r}}\frac{\sigma_{\mathsf{r}}^j}{\phi_{\mathsf{r}}^{j-1}}\Big| \ge \epsilon\Big) \le t^2KK_{t-1}\Pi_{t-2}'\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/R)^2\pi_{t-2}'\epsilon^2}{t^4(\log M)^2}\Big\}, \tag{146}$$

$$P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_t + \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}\Big| \ge \epsilon\Big) \le t^2KK_{t-1}\Pi_{t-2}'\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/R)^2\pi_{t-2}'\epsilon^2}{t^4(\log M)^2}\Big\}. \tag{147}$$

Proof. First note that $(\boldsymbol{M}_t^{\mathsf{c}})^{-1}$ concentrates element-wise to $(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}$ by $\mathcal{B}_{t-1}(g)$. So before proving (146) and (147), we state two results providing concentration and boundedness guarantees for the elements of $\mathbf{v}^{\mathsf{r},\mathsf{c}}$. For for $i \in [t]$, define

$$\mathbb{E}_{i,\mathsf{c}} = -\frac{1}{\mathsf{C}}\hat{\gamma}_{t-1}^{t,\mathsf{r}}\Psi_{\mathsf{c}}^{t-1}\sqrt{W_{\mathsf{rc}}}\Big(1 - \frac{\tau_{\mathsf{c}}^{\max\{i-1,\,t-2\}}}{\tau_{\mathsf{c}}^{t-2}}\Big). \tag{148}$$

Notice $\mathbb{E}_{1,\mathsf{c}} = \ldots = \mathbb{E}_{t-1,\mathsf{c}} = 0$. Then, we will prove that for $i \in [t]$ and $\mathsf{B} > 0$, a universal constant,

$$P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big(\frac{L}{n}\mathrm{v}_i^{\mathsf{r},\mathsf{c}} - \mathbb{E}_{i,\mathsf{c}}\Big)\Big| \geq \epsilon\Big) \leq tKK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}\epsilon^2}{t^2(\log M)^2}\Big\}, \tag{149}$$

$$P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big|\frac{L}{n}\mathrm{v}_i^{\mathsf{r},\mathsf{c}}\Big| \geq \frac{\mathsf{BR}}{\omega}\Big) \leq tKK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}}{t^2(\log M)^2}\Big\}. \tag{150}$$

We prove the main result using (149) and (150), and then prove (149) and (150). We first claim:

$$\frac{-n}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{k=1}^{t} S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}\,[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{t,k}\,\mathbb{E}_{k,\mathsf{c}} = \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}}, \tag{151}$$

$$\frac{n}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{k=1}^{t} S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{jk}\,\mathbb{E}_{k,\mathsf{c}} = \frac{\hat{\gamma}_j^{t,\mathsf{r}}\sigma_{\mathsf{r}}^j}{\phi_{\mathsf{r}}^{j-1}}, \qquad 1 \leq j \leq (t-1). \tag{152}$$

The result (151) is obtained as follows using (148).

$$\begin{aligned}\frac{-n}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{k=1}^{t} S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}\,[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\,\mathbb{E}_{k,\mathsf{c}} &= \frac{-n}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}\,[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tt}\,\mathbb{E}_{t,\mathsf{c}} \qquad (153)\\ = \frac{n\hat{\gamma}_{t-1}^{t,\mathsf{r}}}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} S_{\mathsf{rc}}^{t-1}W_{\mathsf{rc}}\,[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tt}\Psi_{\mathsf{c}}^{t-1}\Big(1 - \frac{\tau_{\mathsf{c}}^{t-1}}{\tau_{\mathsf{c}}^{t-2}}\Big) &\overset{(a)}{=} \frac{\hat{\gamma}_{t-1}^{t,\mathsf{r}}}{\mathsf{C}\phi_{\mathsf{r}}^{t-1}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Psi_{\mathsf{c}}^{t-1} = \frac{\hat{\gamma}_{t-1}^{t,\mathsf{r}}\sigma_{\mathsf{r}}^{t-1}}{\phi_{\mathsf{r}}^{t-1}} = \frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^{t-1}},\end{aligned}$$

where step (a) uses $S_{\mathsf{rc}}^{t-1} = \tau_{\mathsf{c}}^{t-1}/\phi_{\mathsf{r}}^{t-1}$ and $[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tt} = \frac{L}{n}(\tau_{\perp,\mathsf{c}}^{t-1})^{-1}$ which can be seen as follows. From the definition of $\breve{\boldsymbol{C}}^{t,\mathsf{c}}$ in (65), if $\breve{\boldsymbol{C}}^{t-1,\mathsf{c}}$ is invertible, using the block inversion formula,

$$(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1} = \begin{pmatrix} (\breve{\boldsymbol{C}}^{t-1,\mathsf{c}})^{-1} + \frac{L}{n}(\tau_{\perp,\mathsf{c}}^{t-1})^{-1}\hat{\alpha}^{t-1,\mathsf{c}}(\hat{\alpha}^{t-1,\mathsf{c}})^* & -\frac{L}{n}(\tau_{\perp,\mathsf{c}}^{t-1})^{-1}\hat{\alpha}^{t-1,\mathsf{c}} \\ -\frac{L}{n}(\tau_{\perp,\mathsf{c}}^{t-1})^{-1}(\hat{\alpha}^{t-1,\mathsf{c}})^* & \frac{L}{n}(\tau_{\perp,\mathsf{c}}^{t-1})^{-1}\end{pmatrix}, \tag{154}$$

where we have used $\hat{\alpha}^{t-1,\mathsf{c}} := \frac{n}{L}\tau_{\mathsf{c}}^{t-1}(\breve{\boldsymbol{C}}^{t-1,\mathsf{c}})^{-1}(1,\ldots,1)^*$ and $\tau_{\mathsf{c}}^{t-1} - \tau_{\mathsf{c}}^{t-1}(1,\ldots,1)\hat{\alpha}^{t-1,\mathsf{c}} = \tau_{\mathsf{c}}^{t-1} - (\tau_{\mathsf{c}}^{t-1})^2/\tau_{\mathsf{c}}^{t-2} = \tau_{\perp,\mathsf{c}}^{t-1}$. Result (152) is obtained using steps similar to (153) to show that the LHS of (152) equals

$$\frac{\hat{\gamma}_{t-1}^{t,\mathsf{r}}}{\phi_{\mathsf{r}}^{j-1}\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\,\Psi_{\mathsf{c}}^{t-1}\hat{\alpha}_{j-1}^{t-1,\mathsf{c}}\Big[\frac{\tau_{\mathsf{c}}^{j-1}}{\tau_{\mathsf{c}}^{t-1}}\Big] = \frac{\hat{\gamma}_j^{t,\mathsf{r}}\sigma_{\mathsf{r}}^j}{\phi_{\mathsf{r}}^{j-1}},$$

where the last equality follows from the fact that for $j \in [t-2]$ we have $\hat{\gamma}_j^{t,\mathsf{r}} = \hat{\alpha}_{j-1}^{t-1,\mathsf{c}} = 0$, and if $j = t-1$ then since $\hat{\alpha}_{t-1}^{t-1,\mathsf{c}} = \tau_{\mathsf{c}}^{t-1}/\tau_{\mathsf{c}}^{t-2}$.

We now prove (147). Using the result in (151), the LHS of (147) can be expressed as follows:

$$
\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t}\Big[[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{tk}\,\mathrm{v}_k^{\mathsf{r},\mathsf{c}}-\frac{n}{L}[(\breve{\boldsymbol{C}}^{t\mathsf{c}})^{-1}]_{tk}\,\mathbb{E}_{k,\mathsf{c}}\Big]\Big|\geq\epsilon\Big)\\
&\overset{(a)}{\leq}\sum_{k=1}^{t}P\Big(\frac{n}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\tau_{\mathsf{c}}^{t-1}\sqrt{W_{\mathsf{rc}}}\Big[[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{tk}\Big(\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}\Big)-[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\,\mathbb{E}_{k,\mathsf{c}}\Big]\Big|\geq\frac{\kappa\epsilon}{t}\Big)\\
&\overset{(b)}{\leq}\sum_{k=1}^{t}P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\Big(\frac{n}{L}\tau_{\mathsf{c}}^{t-1}\Big)\sqrt{W_{\mathsf{rc}}}\Big(\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}-\mathbb{E}_{k,\mathsf{c}}\Big)\Big|\geq\frac{\kappa\epsilon}{2t}\Big)\\
&\quad+\sum_{k=1}^{t}P\Big(\frac{n}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\tau_{\mathsf{c}}^{t-1}\sqrt{W_{\mathsf{rc}}}\Big[[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{tk}-[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\Big]\Big(\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}\Big)\Big|\geq\frac{\kappa\epsilon}{2t}\Big).
\end{aligned}
\tag{155}
$$

In the above, step (a) follows by Lemma B.2 and $S_{\mathsf{rc}}^{t-1}=\tau_{\mathsf{c}}^{t-1}/\phi_{\mathsf{r}}^{t-1}$ with $\phi_{\mathsf{r}}^{t-1}\in\Theta(1)$. Step (b) follows by Lemma B.2 again. Label the terms on the RHS of (155) as T_1 and T_2. Note that term T_1 has the desired upper bound using (149) and that the non-zero elements of $[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\,(\frac{n}{L}\tau_{\mathsf{c}}^{t-1})$ are $\Theta(1)$. Now consider term T_2. Using that $\frac{n}{L}\tau_{\mathsf{c}}^{t-1}$ is $\Theta(1)$ for all $\mathsf{c}\in[\mathsf{C}]$,

$$
\begin{aligned}
T_2&\leq\sum_{k=1}^{t}P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big|[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{tk}-[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\Big|\Big|\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}\Big|\geq\frac{\kappa\epsilon}{2t}\Big)\\
&\leq\sum_{k=1}^{t}\Big[P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big|\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}\Big|\geq\frac{\mathsf{BR}}{\omega}\Big)+P\Big(\cup_{\mathsf{c}\in[\mathsf{C}]}\Big\{\Big|[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{tk}-[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\Big|\geq\frac{\kappa\epsilon\omega}{2t\mathsf{BR}}\Big\}\Big)\Big]\\
&\leq\sum_{k=1}^{t}\Big[\sum_{\mathsf{c}\in[\mathsf{C}]}P\Big(\Big|[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{tk}-[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{tk}\Big|\geq\frac{\kappa\epsilon\omega}{2t\mathsf{BR}}\Big)+P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big|\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}\Big|\geq\frac{\mathsf{BR}}{\omega}\Big)\Big]\\
&\leq t\mathsf{C}KK_{t-1}\Pi_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi_{t-2}\epsilon^2}{t^2\mathsf{B}^2}\Big\}+t^2KK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}\epsilon^2}{t^2(\log M)^2}\Big\}.
\end{aligned}
\tag{156}
$$

The last inequality follows by induction hypothesis $\mathcal{B}_{t-1}(g)$ and (150). The final result follows since $\mathsf{C}\Pi_{t-2}=\Pi'_{t-2}$ and $\pi_{t-2}=\pi'_{t-2}$.

We next consider the bound (146). First, using Lemma B.2,

$$
\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}\,[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}\,\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j-\gamma_j^{t,\mathsf{r}}\frac{\sigma_{\mathsf{r}}^j}{\phi_{\mathsf{r}}^{j-1}}\Big|\geq\epsilon\Big)\\
&\leq P\Big(|\gamma_j^{t,\mathsf{r}}-\hat{\gamma}_j^{t,\mathsf{r}}|\frac{\sigma_{\mathsf{r}}^j}{\phi_{\mathsf{r}}^{j-1}}\geq\frac{\kappa\epsilon}{2}\Big)+P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t}[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{jk}\,\mathrm{v}_k^{\mathsf{r},\mathsf{c}}-\frac{\hat{\gamma}_j^{t,\mathsf{r}}\sigma_{\mathsf{r}}^j}{\phi_{\mathsf{r}}^{j-1}}\Big|\geq\frac{\kappa\epsilon}{2}\Big).
\end{aligned}
\tag{157}
$$

The first term on the RHS of (157) is upper bounded by $KK_{t-1}\Pi'_{t-1}\exp\{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-1}\epsilon^2/(\log M)^2\}$ using $\mathcal{H}_t(f)$. Using result (152), the second term on the RHS of (157) can be written as

$$
P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t}\Big[[(\boldsymbol{M}_t^{\mathsf{c}})^{-1}]_{jk}\,\mathrm{v}_k^{\mathsf{r},\mathsf{c}}-\frac{n}{L}[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{jk}\,\mathbb{E}_{k,\mathsf{c}}\Big]\Big|\geq\frac{\kappa\epsilon}{2}\Big),
$$

which can be bounded using steps similar to that in (155).

To complete the proof, it remains to prove the bounds in (149) and (150). Note that for $i \in [t]$,

$$\frac{L}{n}\sqrt{W_{\mathsf{rc}}}\mathrm{v}_i^{\mathsf{r},\mathsf{c}} = \frac{\sqrt{W_{\mathsf{rc}}}}{L}(\boldsymbol{h}_{\mathsf{c}}^i + \breve{\boldsymbol{q}}_{\mathsf{c}}^{i-1})^* \boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} = \frac{1}{L}\Big(\frac{\sqrt{W_{\mathsf{rc}}}}{L}\boldsymbol{h}_{\mathsf{c}}^i - \boldsymbol{q}_{\mathsf{c}}^{i-1,\mathsf{r}}\Big)^* (\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}} - \sum_{j=0}^{t-1}\gamma_j^{t,\mathsf{r}}\boldsymbol{q}^{j,\mathsf{r}}), \tag{158}$$

where we have used the fact that $\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} = \boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}} - \boldsymbol{q}_{\parallel,\mathsf{c}}^{t,\mathsf{r}} = \boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}} - \sum_{j=0}^{t-1}\gamma_j^{t,\mathsf{r}}\boldsymbol{q}^{j,\mathsf{r}}$. We first prove result (149). Using (148) and (158), we can bound for the probability in (149) as:

$$\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big(\frac{L}{n}\mathrm{v}_i^{\mathsf{r},\mathsf{c}} - \mathbb{E}_{i,\mathsf{c}}\Big)\Big| \ge \epsilon\Big) \\
\le& P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\Big[\frac{\sqrt{W_{\mathsf{rc}}}}{L}(\boldsymbol{h}_{\mathsf{c}}^i)^*(\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}} - \sum_{j=0}^{t-1}\gamma_j^{t,\mathsf{r}}\boldsymbol{q}^{j,\mathsf{r}}) + \frac{W_{\mathsf{rc}}}{\mathsf{C}}\Big(\Psi_{\mathsf{c}}^t - \hat{\gamma}_{t-1}^{t,\mathsf{r}}\Psi_{\mathsf{c}}^{t-1}\frac{\tau_{\mathsf{c}}^{\max\{i-1,t-2\}}}{\tau_{\mathsf{c}}^{t-2}}\Big)\Big]\Big| \ge \frac{\epsilon}{2}\Big) \\
&+ P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\Big[\frac{1}{L}(\boldsymbol{q}_{\mathsf{c}}^{i-1,\mathsf{r}})^*(\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}} - \sum_{j=0}^{t-1}\gamma_j^{t,\mathsf{r}}\boldsymbol{q}^{j,\mathsf{r}}) - \frac{W_{\mathsf{rc}}}{\mathsf{C}}\Big(\Psi_{\mathsf{c}}^t - \hat{\gamma}_{t-1}^{t,\mathsf{r}}\Psi_{\mathsf{c}}^{t-1}\Big)\Big]\Big| \ge \frac{\epsilon}{2}\Big).
\end{aligned} \tag{159}$$

Label the terms on the RHS of (159) as T_1 and T_2, and we bound both. First, for any $i \in [t]$,

$$\begin{aligned}
T_1 \overset{(a)}{\le}& P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big[\frac{(\boldsymbol{h}_{\mathsf{c}}^i)^*\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}}}{L} + \frac{\sqrt{W_{\mathsf{rc}}}}{\mathsf{C}}\Psi_{\mathsf{c}}^t\Big]\Big| \ge \frac{\epsilon}{6}\Big) + \sum_{j=0}^{t-2}P\Big(\Big|\gamma_j^{t,\mathsf{r}}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{\sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^i)^*\boldsymbol{q}_{\mathsf{c}}^{j,\mathsf{r}}}{L}\Big| \ge \frac{\epsilon}{6(t-1)}\Big) \\
&+ P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big[\gamma_{t-1}^{t,\mathsf{r}}\frac{(\boldsymbol{h}_{\mathsf{c}}^i)^*\boldsymbol{q}_{\mathsf{c}}^{t-1,\mathsf{r}}}{L} + \hat{\gamma}_{t-1}^{t,\mathsf{r}}\frac{\sqrt{W_{\mathsf{rc}}}\Psi_{\mathsf{c}}^{t-1}\tau_{\mathsf{c}}^{\max\{i-1,t-2\}}}{\mathsf{C}\tau_{\mathsf{c}}^{t-2}}\Big]\Big| \ge \frac{\epsilon}{6}\Big) \\
\overset{(b)}{\le}& tKK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}\epsilon^2}{t^2(\log M)^2}\Big\}.
\end{aligned} \tag{160}$$

Step (a) follows from Lemma B.2, and step (b) uses $\mathcal{H}_t(d)$ and $\mathcal{H}_t(f)$: the bound for the first term follows directly from $\mathcal{H}_t(d)$, while the bound for the other two terms uses $\mathcal{H}_t(d)$ and $\mathcal{H}_t(f)$, along with Lemma B.3 and Lemma B.4. Noting that $\hat{\gamma}_{t-1}^{t,\mathsf{r}}$, $(\tau_{\mathsf{c}}^{\max\{i-1,t-2\}}/\tau_{\mathsf{c}}^{t-2})$, and Ψ_{c}^{t-1} are all $\Theta(1)$ terms, we observe from (88) that $\frac{1}{\mathsf{C}}\sum_{\mathsf{c}} W_{\mathsf{rc}}\Psi_{\mathsf{c}}^{t-1}(\tau_{\mathsf{c}}^{\max\{i-1,t-2\}}/\tau_{\mathsf{c}}^{t-2})$ is also $\Theta(1)$.

The second term T_2 on the RHS of (159) can be bounded similarly using Lemma B.2, $\mathcal{H}_t(c)$ with $\mathsf{v} = 1$,$\mathcal{H}_t(f)$, Lemma B.3, and Lemma B.4.

Next we prove result (150). Using (158) we have for $k \in [t]$,

$$\begin{aligned}
&P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big|\frac{L}{n}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}\Big| \ge \frac{\mathsf{BR}}{\omega}\Big) = P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{1}{L}\Big|\Big(\sqrt{W_{\mathsf{rc}}}\boldsymbol{h}_{\mathsf{c}}^k - \boldsymbol{q}_{\mathsf{c}}^{k-1,\mathsf{r}}\Big)^*(\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}} - \sum_{j=0}^{t-1}\gamma_j^{t,\mathsf{r}}\boldsymbol{q}_{\mathsf{c}}^{j,\mathsf{r}})\Big| \ge \frac{\mathsf{BR}}{\omega}\Big) \\
\le& P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\Big|\sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^k)^*\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}}\Big| \ge \frac{\mathsf{BR}}{4\omega}\Big) + P\Big(\frac{1}{L}\sum_{j=0}^{t-1}\Big|\gamma_j^{t,\mathsf{r}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\Big|\sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^k)^*\boldsymbol{q}_{\mathsf{c}}^{j,\mathsf{r}}\Big| \ge \frac{\mathsf{BR}}{4\omega}\Big) \\
&+ P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\Big|(\boldsymbol{q}_{\mathsf{c}}^{k-1,\mathsf{r}})^*\boldsymbol{q}_{\mathsf{c}}^{t,\mathsf{r}}\Big| \ge \frac{\mathsf{BR}}{4\omega}\Big) + P\Big(\frac{1}{L}\sum_{j=0}^{t-1}\Big|\gamma_j^{t,\mathsf{r}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\Big|(\boldsymbol{q}_{\mathsf{c}}^{k-1,\mathsf{r}})^*\boldsymbol{q}_{\mathsf{c}}^{j,\mathsf{r}}\Big| \ge \frac{\mathsf{BR}}{4\omega}\Big).
\end{aligned} \tag{161}$$

We now provide upper bounds for each term of (161), labeled as $A_1 - A_4$. We note that these results don't follow directly from the induction hypothesis as they involve the sums of the *absolute values* of the inner products over the column blocks c. First notice that by (88), for any $0 \le s \le t$,

$$\frac{\|\boldsymbol{q}^{s,\mathsf{r}}\|^2}{L} = \sum_{\mathsf{c}\in[\mathsf{C}]} \frac{W_{\mathsf{rc}}\,\|\breve{\boldsymbol{q}}^s_{\mathsf{c}}\|^2}{L} \le \sum_{\mathsf{c}\in[\mathsf{C}]} \frac{4W_{\mathsf{rc}}}{\mathsf{C}} \le 4\kappa_1, \qquad \sum_{i\in\mathrm{sec}(\ell)} |[\boldsymbol{q}^{s,\mathsf{r}}_{\mathsf{c}}]_i| = \sqrt{W_{\mathsf{rc}}} \sum_{i\in\mathrm{sec}(\ell)} |[\breve{\boldsymbol{q}}^s_{\mathsf{c}}]_i| \le 2\sqrt{W_{\mathsf{rc}}}. \tag{162}$$

For reasons that will become clear in a few steps, we take

$$\mathsf{B} = 4\Big(1 + (\hat{\gamma}^{t,\mathsf{r}}_{t-1} + 1)\max\Big\{4\kappa_1,\ \sqrt{24\tau^0_{\mathsf{c}}\log M + 1}\Big\}\Big). \tag{163}$$

(Note that $\tau^0_{\mathsf{c}}\log M \in \Theta(1)$). First, we claim that third term, A_3, in (161) equals 0. Indeed, using the Cauchy-Schwarz inequality and (162),

$$\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} |(\boldsymbol{q}^{k-1,\mathsf{r}}_{\mathsf{c}})^*\,\boldsymbol{q}^{j,\mathsf{r}}_{\mathsf{c}}| \le \frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} \|\boldsymbol{q}^{k-1,\mathsf{r}}_{\mathsf{c}}\|\|\boldsymbol{q}^{j,\mathsf{r}}_{\mathsf{c}}\| \le \frac{1}{L}\|\boldsymbol{q}^{k-1,\mathsf{r}}\|\|\boldsymbol{q}^{j,\mathsf{r}}\| \le 4\kappa_1. \tag{164}$$

Since $\frac{\mathsf{BR}}{4\omega} > \frac{4\kappa_1\mathsf{R}}{\omega} \ge 4\kappa_1$, it follows that $A_3 = 0$. To bound A_4, from (163) we note that $\mathsf{B} \ge 4 + 16\kappa_1|\hat{\gamma}^{t,\mathsf{r}}_{t-1}| = 4 + 16\kappa_1\sum_{j=0}^{t-1}\hat{\gamma}^{t,\mathsf{r}}_j$ since $\hat{\gamma}^{t,\mathsf{r}}_0 = \ldots = \hat{\gamma}^{t,\mathsf{r}}_{t-2} = 0$. Then, using (164), we have

$$\begin{aligned} A_4 &\le P\Big(\sum_{j=0}^{t-1}\Big|\gamma^{t,\mathsf{r}}_j\Big| \ge \frac{\mathsf{BR}}{16\kappa_1\omega}\Big) \le P\Big(\sum_{j=0}^{t-1}\Big|\gamma^{t,\mathsf{r}}_j\Big| \ge \sum_{j=0}^{t-1}\Big|\hat{\gamma}^{t,\mathsf{r}}_j\Big| + \frac{1}{4\kappa_1}\Big) \\ &\le \sum_{j=0}^{t-1} P\Big(|\gamma^{t,\mathsf{r}}_j - \hat{\gamma}^{t,\mathsf{r}}_j| \ge \frac{1}{t(4\kappa_1)}\Big) \le tKK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}\epsilon^2}{t^2(\log M)^2}\Big\}. \end{aligned}$$

The last inequality uses inductive hypothesis $\mathcal{H}_t(f)$. Now consider A_1. Using (162), for $0 \le j,k \le t$,

$$\sum_{\mathsf{c}\in[\mathsf{C}]}\Big|\frac{\sqrt{W_{\mathsf{rc}}}}{L}(\boldsymbol{h}^k_{\mathsf{c}})^*\,\boldsymbol{q}^{j,\mathsf{r}}_{\mathsf{c}}\Big| \le \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{\sqrt{W_{\mathsf{rc}}}}{L}\max_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{h}^k_{\mathsf{c}}]_i\Big|\sum_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{q}^{j,\mathsf{r}}_{\mathsf{c}}]_i\Big| \le \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{2W_{\mathsf{rc}}}{L}\max_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{h}^k_{\mathsf{c}}]_i\Big|. \tag{165}$$

From (163), we have

$$\frac{\mathsf{BR}}{4\omega} \ge \frac{\mathsf{R}}{\omega}(\hat{\gamma}^{t,\mathsf{r}}_{t-1} + 1)\max\Big\{4\kappa_1,\ \sqrt{24\tau^0_{\mathsf{c}}\log M + 1}\Big\}\Big) > (\hat{\gamma}^{t,\mathsf{r}}_{t-1} + 1)\sqrt{24(\max_{\mathsf{c}} W^2_{\mathsf{rc}})\tau^0_{\mathsf{c}}\log M + 1}, \tag{166}$$

where the last inequality holds when the constant κ_1 is chosen to be large enough since $\max_{\mathsf{c}} W_{\mathsf{rc}} = \Theta(\mathsf{R}/\omega)$. Therefore, using Lemma C.2 and $\mathcal{H}_t(e)$ with $\mathsf{v} = 1$ since $k \in [t]$, we obtain

$$\begin{aligned} A_1 &\le P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{2W_{\mathsf{rc}}}{L}\max_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{h}^k_{\mathsf{c}}]_i\Big| \ge \frac{\mathsf{BR}}{4\omega}\Big) \\ &\le P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} 4W^2_{\mathsf{rc}}\max_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{h}^k_{\mathsf{c}}]_i\Big|^2 \ge 24(\max_{\mathsf{c}} W^2_{\mathsf{rc}})\tau^0_{\mathsf{c}}\log M + 1\Big) \le KK_{t-1}\Pi'_{t-2}e^{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}\epsilon^2}. \end{aligned}$$

Finally, using (165), Lemma C.2, and (166), we bound term A_2:

$$A_2 \le P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{2W_{\mathsf{rc}}}{L}\max_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{h}_{\mathsf{c}}^k]_i\Big|\sum_{j=0}^{t-1}\Big|\gamma_j^{t,\mathsf{r}}\Big| \ge \frac{\mathsf{BR}}{4\omega}\Big)$$

$$\le P\Big(\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\frac{4W_{\mathsf{rc}}^2}{L}\max_{i\in\mathrm{sec}(\ell)}\Big|[\boldsymbol{h}_{\mathsf{c}}^k]_i\Big|^2 \ge 24(\max_{\mathsf{c}} W_{\mathsf{rc}}^2)\tau_{\mathsf{c}}^0\log M+1\Big) + P\Big(\sum_{j=0}^{t-1}\Big|\gamma_j^{t,\mathsf{r}}\Big| \ge \sum_{j=0}^{t-1}\Big|\hat{\gamma}_j^{t,\mathsf{r}}\Big|+1\Big).$$

Then the upper bound follows along the same lines as that of A_1 and A_4. This completes the proof of (150), and the lemma. ■

(a) Recall the definition of $\boldsymbol{\Delta}_{t,t,\mathsf{r}}$ from Lemma 7.4 Eq. (81). Consider the term $\frac{1}{\sqrt{L}}\sum_{\mathsf{c}}\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|\,[\mathbf{P}^{\parallel}_{\boldsymbol{M}_t^{\mathsf{c}}}\boldsymbol{Z}_t'^{\mathsf{c}}]_{\mathsf{r}}$, where $\boldsymbol{Z}_t'^{\mathsf{c}}\sim\mathcal{N}(0,\mathbb{I}_n)$. Using Lemma C.6, $\mathbf{P}^{\parallel}_{\boldsymbol{M}_t^{\mathsf{c}}}\boldsymbol{Z}_t'^{\mathsf{c}} \overset{d}{=} \sum_{j=0}^{t-1}\frac{\boldsymbol{m}_{\perp,\mathsf{r}}^{j,\mathsf{c}}}{\|\boldsymbol{m}_{\perp}^{j,\mathsf{c}}\|}Z_j^{\mathsf{c}}$ for i.i.d. $Z_j^{\mathsf{c}}\sim\mathcal{N}(0,1)$. Then

$$\sum_{\mathsf{c}}\frac{\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|}{\sqrt{L}}\,[\mathbf{P}^{\parallel}_{\boldsymbol{M}_t^{\mathsf{c}}}\boldsymbol{Z}_t'^{\mathsf{c}}]_{\mathsf{r}} \overset{d}{=} \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t-1}\frac{\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|}{\sqrt{L}}\frac{\boldsymbol{m}_{\perp,\mathsf{r}}^{j,\mathsf{c}}}{\|\boldsymbol{m}_{\perp}^{j,\mathsf{c}}\|}Z_j^{\mathsf{c}}, \qquad \{Z_j^{\mathsf{c}}\}_{\mathsf{c}\in[\mathsf{C}],\,0\le j\le(t-1)} \overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,1). \tag{167}$$

Noe we simplify the final three terms of $\boldsymbol{\Delta}_{t,t,\mathsf{r}}$ in (81). Using the Lemma 8.1 notation $\mathbb{M}_t^{\mathsf{c}}, \mathbf{v}^{\mathsf{r},\mathsf{c}}$,

$$\sum_{\mathsf{c}\in[\mathsf{C}]}\boldsymbol{M}_{t,\mathsf{r}}^{\mathsf{c}}((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} = \sum_{\mathsf{c}\in[\mathsf{C}]}\boldsymbol{M}_{t,\mathsf{r}}^{\mathsf{c}}(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}} = \sum_{j=1}^{t}\sum_{\mathsf{c}\in[\mathsf{C}]}\boldsymbol{m}_{\mathsf{r}}^{j-1,\mathsf{c}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j.$$

Therefore,

$$\sum_{\mathsf{c}\in[\mathsf{C}]}\boldsymbol{M}_{t,\mathsf{r}}^{\mathsf{c}}((\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{X}_{t,\mathsf{c}})^*\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}} - \sum_{i=1}^{t-1}\gamma_i^{t,\mathsf{r}}\upsilon_{\mathsf{r}}^i\breve{\boldsymbol{m}}_{\mathsf{r}}^{i-1} + \upsilon_{\mathsf{r}}^t\breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1} \tag{168}$$

$$= \sum_{j=1}^{t-1}\breve{\boldsymbol{m}}_{\mathsf{r}}^{j-1}[\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j - \gamma_j^{t,\mathsf{r}}\upsilon_{\mathsf{r}}^j] + \breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1}[\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_t + \upsilon_{\mathsf{r}}^t].$$

Using the expressions in (167) and (168) in the definition of $\boldsymbol{\Delta}_{t,t,\mathsf{r}}$ in (81), by Lemma C.2,

$$\frac{\|\boldsymbol{\Delta}_{t,t,\mathsf{r}}\|^2}{2(t+1)} \le \sum_{i=0}^{t-1}(\gamma_i^{t,\mathsf{r}}-\hat{\gamma}_i^{t,\mathsf{r}})^2\,\|\boldsymbol{b}_{\mathsf{r}}^i\|^2 + \|\boldsymbol{Z}_{t,\mathsf{r}}'\|^2\Big[\frac{\|\boldsymbol{q}_{\perp}^{t,\mathsf{r}}\|}{\sqrt{L}} - \sqrt{\sigma_{\perp,\mathsf{r}}^t}\Big]^2 + Z^2\sum_{j=0}^{t-1}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{\|\boldsymbol{q}_{\perp,\mathsf{c}}^{t,\mathsf{r}}\|^2\|\boldsymbol{m}_{\perp,\mathsf{r}}^{j,\mathsf{c}}\|^2}{L\|\boldsymbol{m}_{\perp}^{j,\mathsf{c}}\|^2}$$

$$+\sum_{j=1}^{t-1}\|\breve{\boldsymbol{m}}_{\mathsf{r}}^{j-1}\|^2\Big[\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j - \gamma_j^{t,\mathsf{r}}\upsilon_{\mathsf{r}}^j\Big]^2 + \|\breve{\boldsymbol{m}}_{\mathsf{r}}^{t-1}\|^2\Big[\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{t-1}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_t + \upsilon_{\mathsf{r}}^t\Big]^2 \tag{169}$$

Applying Lemma B.2, with $\tilde{\epsilon}_t = \frac{\epsilon}{4(t+1)^2}$, we obtain

$$P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}\,\|\boldsymbol{\Delta}_{t,t,\mathsf{r}}\|^2 \geq \epsilon\Big) \leq \sum_{i=0}^{t-1} P\Big(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}|\gamma_i^{t,\mathsf{r}} - \hat{\gamma}_i^{t,\mathsf{r}}|^2 \frac{\|\boldsymbol{b}^i_{\mathsf{r}}\|^2}{N_{\mathsf{R}}} \geq \tilde{\epsilon}_t\Big)$$
$$+ P\Big(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}\Big|\frac{\|\boldsymbol{q}^{t,\mathsf{r}}_{\perp}\|}{\sqrt{L}} - \sqrt{\sigma^t_{\perp,\mathsf{r}}}\Big|^2 \frac{\|\boldsymbol{Z}'_{t,\mathsf{r}}\|^2}{N_{\mathsf{R}}} \geq \tilde{\epsilon}_t\Big) + P\Big(\frac{Z^2}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}\sum_{j=0}^{t-1}\sum_{\mathsf{c}\in[\mathsf{C}]} \frac{W^{\mathsf{u}}_{\mathsf{rc}'}\|\boldsymbol{q}^{t,\mathsf{r}}_{\perp,\mathsf{c}}\|^2\|\boldsymbol{m}^{j,\mathsf{c}}_{\perp,\mathsf{r}}\|^2}{L\|\boldsymbol{m}^{j,\mathsf{c}}_{\perp}\|^2} \geq \tilde{\epsilon}_t\Big)$$
$$+ \sum_{j=1}^{t-1} P\Big(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} \frac{W^{\mathsf{u}}_{\mathsf{rc}'}\|\breve{\boldsymbol{m}}^{j-1}_{\mathsf{r}}\|^2}{N_{\mathsf{R}}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} S^{j-1}_{\mathsf{rc}}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}^{\mathsf{c}}_t)^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j - \gamma^{t,\mathsf{r}}_j \upsilon^j_{\mathsf{r}}\Big|^2 \geq \tilde{\epsilon}_t\Big)$$
$$+ P\Big(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} \frac{W^{\mathsf{u}}_{\mathsf{rc}'}\|\breve{\boldsymbol{m}}^{t-1}_{\mathsf{r}}\|^2}{N_{\mathsf{R}}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} S^{t-1}_{\mathsf{rc}}\sqrt{W_{\mathsf{rc}}}[(\mathbb{M}^{\mathsf{c}}_t)^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_t + \upsilon^t_{\mathsf{r}}\Big|^2 \geq \tilde{\epsilon}_t\Big). \tag{170}$$

We label the terms in (170) $T_1 - T_5$, and show that each has the desired upper bound.

First, using Lemma B.4 and induction hypotheses $\mathcal{B}_0(c) - \mathcal{B}_{t-1}(c)$ and $\mathcal{H}_t(f)$:

$$T_1 \leq \sum_{i=0}^{t-1} P\Big(\max_{\mathsf{r}'\in[\mathsf{R}]}|\gamma_i^{t,\mathsf{r}'} - \hat{\gamma}_i^{t,\mathsf{r}'}|^2 \cdot \sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}\frac{\|\boldsymbol{b}^i_{\mathsf{r}}\|^2}{n} \geq \tilde{\epsilon}_t\Big)$$
$$\leq \sum_{i=0}^{t-1}\Big[P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}\Big[\frac{\|\boldsymbol{b}^i_{\mathsf{r}}\|^2}{n} - \frac{\sigma^i_{\mathsf{r}}}{\mathsf{R}}\Big]\Big| \geq \sqrt{\tilde{\epsilon}_t}\Big) + \sum_{\mathsf{r}\in[\mathsf{R}]} P\Big(|\gamma_i^{t,\mathsf{r}} - \hat{\gamma}_i^{t,\mathsf{r}}|^2 \geq \tilde{\epsilon}_t/\Big(2\max\Big\{1, \sum_{\mathsf{r}\in[\mathsf{R}]}\frac{W^{\mathsf{u}}_{\mathsf{rc}'}\sigma^i_{\mathsf{r}}}{\mathsf{R}}\Big\}\Big)\Big)\Big]$$
$$\leq tKK_{t-1}\Pi_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^{2(\mathsf{u}-1)_+}\pi_{t-2}\epsilon}{t^2}\Big\} + t\mathsf{R}KK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^{2+(\mathsf{u}-1)_+}\pi'_{t-2}\epsilon}{t^2(\log M)^2}\Big\}. \tag{171}$$

For the second term we have used that $\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}/\mathsf{R} = 1$ if $\mathsf{u} = 0$ and for $\mathsf{u} \in \{1,2\}$ we have $\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}/\mathsf{R} \leq \kappa(\mathsf{R}/\omega)^{\mathsf{u}-1}$ (see (88)). We have also used the fact that σ^i_{r} is bounded above and below by positive constants for $0 \leq i \leq t-1$ and $\mathsf{r} \in [\mathsf{R}]$.

The second term in (170) is bounded using Lemma B.4 along with induction hypothesis $\mathcal{H}_t(f)$, Lemma B.5, and Lemma B.9,

$$T_2 \leq P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}\|\boldsymbol{Z}'_{t,\mathsf{r}}\|^2 \max_{\mathsf{r}'\in[\mathsf{R}]}\Big|\frac{\|\boldsymbol{q}^{t,\mathsf{r}'}_{\perp}\|}{\sqrt{L}} - \sqrt{\sigma^t_{\perp,\mathsf{r}'}}\Big|^2 \geq \tilde{\epsilon}_t\Big)$$
$$\leq \sum_{\mathsf{r}'\in[\mathsf{R}]} P\Big(\Big|\frac{\|\boldsymbol{q}^{t,\mathsf{r}'}_{\perp}\|}{\sqrt{L}} - \sqrt{\sigma^t_{\perp,\mathsf{r}'}}\Big|^2 \geq \tilde{\epsilon}_t/\big(2\max\{1, \sum_{\mathsf{r}\in[\mathsf{R}]}\frac{W^{\mathsf{u}}_{\mathsf{rc}'}}{\mathsf{R}}\}\big)\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}'}\Big[\frac{\|\boldsymbol{Z}'_{t,\mathsf{r}}\|^2}{N_{\mathsf{R}}} - 1\Big]\Big| \geq \sqrt{\tilde{\epsilon}_t}\Big)$$
$$\leq \mathsf{R}KK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^{2+(\mathsf{u}-1)_+}\pi'_{t-2}\epsilon}{t^2(\log M)^2}\Big\} + 2\exp\Big\{\frac{-\kappa n\epsilon(\omega/\mathsf{R})^{2\mathsf{u}-1}}{t^2}\Big\}. \tag{172}$$

We note that $\sigma^t_{\perp,\mathsf{r}}$ is bounded below for all r and we have used that $\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}/\mathsf{R} = 1$ if $\mathsf{u} = 0$ and for $\mathsf{u} \in \{1,2\}$ we have $\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}/\mathsf{R} \leq \kappa(\mathsf{R}/\omega)^{\mathsf{u}-1}$ (see (88)).

The third term in (170) is bounded as follows. For all $\mathsf{r} \in [\mathsf{R}]$ and $\mathsf{c} \in [\mathsf{C}]$, using Lemma C.2,

$$\frac{\|\boldsymbol{q}^{t,\mathsf{r}}_{\perp,\mathsf{c}}\|^2}{L} \leq \frac{W_{\mathsf{rc}}}{L}\Big\|\breve{\boldsymbol{q}}^t_{\mathsf{c}} - \sum_{i=0}^{t-1}\gamma_i^{t,\mathsf{r}}\breve{\boldsymbol{q}}^i_{\mathsf{c}}\Big\|^2 \leq \frac{W_{\mathsf{rc}}(t+1)}{L}\Big[\big\|\breve{\boldsymbol{q}}^t_{\mathsf{c}}\big\|^2 + \sum_{i=0}^{t-1}(\gamma_i^{t,\mathsf{r}})^2\big\|\breve{\boldsymbol{q}}^i_{\mathsf{c}}\big\|^2\Big] \leq \frac{4W_{\mathsf{rc}}(t+1)}{\mathsf{C}}\Big[1 + \sum_{i=0}^{t-1}(\gamma_i^{t,\mathsf{r}})^2\Big],$$

where the final inequality follows as $\|\breve{\boldsymbol{q}}_{\mathsf{c}}^{i}\|^2 \le 4L/\mathsf{C}$ for $0 \le i \le t$. Therefore,

$$T_3 \le P\Big(\frac{Z^2}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}\sum_{j=0}^{t-1}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{4W^{\mathsf{u}}_{\mathsf{rc}'}W_{\mathsf{rc}}(t+1)}{\mathsf{C}}\Big[1+\sum_{i=0}^{t-1}(\gamma_i^{t,\mathsf{r}})^2\Big]\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}}^{j,\mathsf{c}}\|^2}{\|\boldsymbol{m}_{\perp}^{j,\mathsf{c}}\|^2} \ge \tilde{\epsilon}_t\Big). \tag{173}$$

Now considering the RHS of (173), note that if $\Big\{(\gamma_i^{t,\mathsf{r}})^2 \le (\hat{\gamma}_i^{t,\mathsf{r}})^2+1\Big\}$ for $\mathsf{r}\in[\mathsf{R}]$ and $0\le i\le (t-1)$, then since $(\hat{\gamma}_i^{t,\mathsf{r}})^2+1 \in \Theta(1)$ for $\mathsf{r}\in[\mathsf{R}]$ (meaning $1+\sum_{i=0}^{t-1}(\gamma_i^{t,\mathsf{r}})^2 \le \kappa t$ for a constant $\kappa>0$),

$$\begin{aligned}
&\frac{Z^2}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}\sum_{j=0}^{t-1}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{4W^{\mathsf{u}}_{\mathsf{rc}'}W_{\mathsf{rc}}(t+1)}{\mathsf{C}}\Big[1+\sum_{i=0}^{t-1}(\gamma_i^{t,\mathsf{r}})^2\Big]\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}}^{j,\mathsf{c}}\|^2}{\|\boldsymbol{m}_{\perp}^{j,\mathsf{c}}\|^2}\\
&\le \frac{\kappa Z^2t^2\max_{\mathsf{rc}}\{W^{\mathsf{u}}_{\mathsf{rc}'}W_{\mathsf{rc}}\}}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}\sum_{j=0}^{t-1}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{1}{\mathsf{C}}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}}^{j,\mathsf{c}}\|^2}{\|\boldsymbol{m}_{\perp}^{j,\mathsf{c}}\|^2} = \frac{\kappa Z^2t^3\max_{\mathsf{rc}}\{W^{\mathsf{u}}_{\mathsf{rc}'}W_{\mathsf{rc}}\}}{n}.
\end{aligned}$$

Using this argument in (173) and noting that $\max_{\mathsf{rc}}\{W^{\mathsf{u}}_{\mathsf{rc}'}W_{\mathsf{rc}}\} \le (P\mathsf{R}/\omega)^{\mathsf{u}+1}$ we obtain

$$\begin{aligned}
T_3 &\le \sum_{i=0}^{t-1}\sum_{\mathsf{r}\in[\mathsf{R}]}P\Big((\gamma_i^{t,\mathsf{r}})^2 \ge (\hat{\gamma}_i^{t,\mathsf{r}})^2+1\Big) + P\Big(Z^2 \ge \kappa n\tilde{\epsilon}_t(\omega/\mathsf{R})^{\mathsf{u}+1}/t^3\Big)\\
&\le tK\mathsf{R}K_{t-1}\Pi'_{t-2}\exp\Big\{-\frac{\kappa\kappa_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}}{(\log M)^2}\Big\} + 2\exp\Big\{-\frac{1}{t^5}\kappa N_{\mathsf{R}}\omega(\omega/\mathsf{R})^{\mathsf{u}}\epsilon\Big\}.
\end{aligned} \tag{174}$$

The final inequality in (174) is obtained using $\mathcal{H}_t(f)$ and Lemma B.6 for the first term, and Lemma B.9 for the second term.

Now consider the fourth term of (170):

$$\begin{aligned}
T_4 &\le \sum_{j=1}^{t-1}P\Big(\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{W^{\mathsf{u}}_{\mathsf{r}'\mathsf{c}'}\|\breve{\boldsymbol{m}}_{\mathsf{r}'}^{j-1}\|^2}{n}\max_{\mathsf{r}\in[\mathsf{R}]}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}W_{\mathsf{rc}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j - \gamma_j^{t,\mathsf{r}}\upsilon_{\mathsf{r}}^{j}\Big|^2 \ge \tilde{\epsilon}_t\Big)\\
&\overset{(a)}{\le} \sum_{j=1}^{t-1}P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}'}\Big(\frac{\|\breve{\boldsymbol{m}}_{\mathsf{r}}^{j-1}\|^2}{N_{\mathsf{R}}} - \phi_{\mathsf{r}}^{j-1}\Big)\Big| \ge \sqrt{\tilde{\epsilon}_t}\Big)\\
&\quad + \sum_{j=1}^{t-1}\sum_{\mathsf{r}\in[\mathsf{R}]}P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}S_{\mathsf{rc}}^{j-1}W_{\mathsf{rc}}[(\mathbb{M}_t^{\mathsf{c}})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_j - \gamma_j^{t,\mathsf{r}}\upsilon_{\mathsf{r}}^{j}\Big|^2 \ge \tilde{\epsilon}_t/\Big(2\max\{1,(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}'}\phi_{\mathsf{r}}^{j-1})\}\Big)\Big)\\
&\overset{(b)}{\le} tKK_{t-1}\Pi_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^{2(\mathsf{u}-1)_+}\pi_{t-2}\epsilon}{t^2}\Big\} + t^3\mathsf{R}KK_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa_{t-1}(\omega/\mathsf{R})^{2+(\mathsf{u}-1)_+}\pi'_{t-2}\epsilon}{t^6(\log M)^2}\Big\}.
\end{aligned} \tag{175}$$

Step (a) follows by Lemma B.4 and (b) by the induction hypothesis $\mathcal{B}_{t-1}(e)$ and Lemma 8.1, noting $\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}}/\mathsf{R}=1$ if $\mathsf{u}=0$ and $\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}}/\mathsf{R} \le P(\mathsf{R}/\omega)^{\mathsf{u}-1}$ for $\mathsf{u}\in\{1,2\}$. In particular, $\mathcal{B}_{t-1}(e)$ is used as follows. First, recall $\|\boldsymbol{m}_{\mathsf{r}}^{j-1,\mathsf{c}}\|^2 = (S_{\mathsf{rc}}^{j-1})^2W_{\mathsf{rc}}\|\breve{\boldsymbol{m}}_{\mathsf{r}}^{j-1}\|^2$ and $\frac{n}{L}S_{\mathsf{rc}}^{j-1}\in\Theta(1)$. Therefore,

$$\begin{aligned}
&P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}'}\Big[\frac{\|\breve{\boldsymbol{m}}_{\mathsf{r}}^{j-1}\|^2}{N_{\mathsf{R}}} - \phi_{\mathsf{r}}^{j-1}\Big]\Big| \ge \sqrt{\tilde{\epsilon}_t}\Big) \le P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}'}\Big(\frac{n}{L}S_{\mathsf{rc}}^{j-1}\Big)^2\Big[\frac{\|\breve{\boldsymbol{m}}_{\mathsf{r}}^{j-1}\|^2}{N_{\mathsf{R}}} - \phi_{\mathsf{r}}^{j-1}\Big]\Big| \ge \kappa\sqrt{\tilde{\epsilon}_t}\Big)\\
&= P\Big(\Big|\frac{n}{L^2}\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}-1}_{\mathsf{rc}'}\|\boldsymbol{m}_{\mathsf{r}}^{j-1,\mathsf{c}}\|^2 - \frac{n^2}{\mathsf{R}L^2}\sum_{\mathsf{r}\in[\mathsf{R}]}W^{\mathsf{u}}_{\mathsf{rc}'}(S_{\mathsf{rc}}^{j-1})^2\phi_{\mathsf{r}}^{j-1}\Big| \ge \kappa\sqrt{\tilde{\epsilon}_t}\Big).
\end{aligned}$$

Now the result from $\mathcal{B}_{t-1}(e)$ can be applied directly.

Finally, the last term in (170) can be bounded using the same arguments as for the fourth term.

To see the overall concentration result, notice that $\Pi_{t-2} = (\mathsf{RC})^{t-1} \leq \mathsf{R}\Pi'_{t-2} = \mathsf{RC}\Pi_{t-2} = \Pi_{t-1}$ and $(\omega/\mathsf{R})^{2(\mathsf{u}-1)_+}\pi_{t-2} \geq (\omega/\mathsf{R})^{2+(\mathsf{u}-1)_+}\pi'_{t-2}/(\log M)^2$ since $\pi_{t-2} = \pi'_{t-2}$ and $\mathsf{u} \in \{0,1,2\}$. Finally, $(\omega/\mathsf{R})^{2+(\mathsf{u}-1)_+}\pi'_{t-2}/(\log M)^2 = (\omega/\mathsf{R})^{(\mathsf{u}-1)_+}\pi_{t-1}$. Therefore we have the bound in (92).

(b) Using $\boldsymbol{b}^t_{\mathsf{r}}|_{\mathscr{S}_{t,t}} \overset{d}{=} \sqrt{\sigma^t_{\mathsf{r}}}\breve{\boldsymbol{Z}}_{t,\mathsf{r}} + \breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}$, by Lemma 7.5, where $\breve{\boldsymbol{Z}}_{t,\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$, and $\boldsymbol{w}_{\mathsf{r}} \overset{d}{=} \sigma \mathbf{U}_{\mathsf{r}}$ for $\mathbf{U}_{\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$ independent of $\breve{\boldsymbol{Z}}_{t,\mathsf{r}}$, we have by Lemma B.2,

$$P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\Big[\frac{(\boldsymbol{b}^t_{\mathsf{r}})^*\boldsymbol{w}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \geq \epsilon\Big) = P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\Big(\sqrt{\sigma^t_{\mathsf{r}}}\Big[\frac{(\breve{\boldsymbol{Z}}_{t,\mathsf{r}})^*\mathbf{U}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big] + \frac{\breve{\boldsymbol{\Delta}}^*_{t,\mathsf{r}}\mathbf{U}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big)\Big| \geq \frac{\epsilon}{\sigma}\Big)$$

$$\leq P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\sqrt{\sigma^t_{\mathsf{r}}}\Big[\frac{(\breve{\boldsymbol{Z}}_{t,\mathsf{r}})^*\mathbf{U}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \geq \frac{\epsilon}{2\sigma}\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\Big[\frac{\breve{\boldsymbol{\Delta}}^*_{t,\mathsf{r}}\mathbf{U}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \geq \frac{\epsilon}{2\sigma}\Big). \tag{176}$$

To bound the first term in the above, we recall that $\breve{\boldsymbol{Z}}_{t,\mathsf{r}}$ is independent of $\mathbf{U}_{\mathsf{r}}$. Hence, using Lemma B.9 and the same argument as in (123), this term is bounded by $2\exp\{-\kappa n\epsilon^2(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}}\}$.

For the second term in (176), we first obtain a concentration result on the norm of $\breve{\boldsymbol{\Delta}}_t = \sum_{i=0}^t(\sigma^t_{\mathsf{r}}/\sigma^i_{\mathsf{r}})\boldsymbol{\Delta}_{i,i,\mathsf{r}}$, as defined in Lemma 7.5. We have

$$P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|^2 \geq \epsilon^2\Big) \overset{(a)}{\leq} P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}\sum_{i=0}^{t}\Big(\frac{\sigma^t_{\mathsf{r}}}{\sigma^i_{\mathsf{r}}}\Big)^2 W^{\mathsf{u}}_{\mathsf{rc}}\|\boldsymbol{\Delta}_{i,i,\mathsf{r}}\|^2 \geq \frac{\epsilon^2}{t+1}\Big) \tag{177}$$

$$\overset{(b)}{\leq} \sum_{i=0}^{t} P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\|\boldsymbol{\Delta}_{i,i,\mathsf{r}}\|^2 \geq \frac{\kappa\epsilon^2}{(t+1)^2}\Big) \overset{(c)}{\leq} t^4 K K_{t-1}\Pi_{t-1}\exp\Big\{-\frac{1}{t^8}\kappa\kappa_{t-1}(\omega/\mathsf{R})^{(\mathsf{u}-1)_+}\pi_{t-1}\epsilon^2\Big\}.$$

Step (a) follows from Lemma C.2 and step (b) by Lemma B.2, using that $\sigma^i_{\mathsf{r}} \in \Theta(1)$ for all $0 \leq i \leq t$. Finally step (c) uses the result from $\mathcal{B}_t(a)$ above.

Next, using $\Big|\breve{\boldsymbol{\Delta}}^*_{t,\mathsf{r}}\mathbf{U}_{\mathsf{r}}\Big| \leq \|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|\|\mathbf{U}_{\mathsf{r}}\|$, the second term can be bounded as

$$P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\Big[\frac{\breve{\boldsymbol{\Delta}}^*_{t,\mathsf{r}}\mathbf{U}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \geq \frac{\epsilon}{2\sigma}\Big) \leq P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\frac{\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|\|\mathbf{U}_{\mathsf{r}}\|}{n} \geq \kappa\epsilon\Big) \tag{178}$$

$$\overset{(a)}{\leq} P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|^2 \cdot \frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\|\mathbf{U}_{\mathsf{r}}\|^2 \geq \kappa\epsilon^2\Big)$$

$$\overset{(b)}{\leq} P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|^2 \geq \kappa\epsilon^2/\Big(2\max\Big\{1, \frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\Big\}\Big)\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W^{\mathsf{u}}_{\mathsf{rc}}\Big[\frac{\|\mathbf{U}_{\mathsf{r}}\|^2}{N_{\mathsf{R}}} - 1\Big]\Big| \geq \epsilon\Big)$$

$$\overset{(c)}{\leq} t^4 K K_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^8}\kappa\kappa_{t-1}(\omega/\mathsf{R})^{2(\mathsf{u}-1)_+}\pi_{t-1}\epsilon^2\}. + 2\exp\{-\kappa n\epsilon^2(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}}\}.$$

Step (a) follows by Cauchy-Schwarz, (b) by Lemma B.4, and (c) by Lemma B.9 and (177).

(c) Let $0 \leq s \leq t$. By Lemma 7.5, $\boldsymbol{b}^t_{\mathsf{r}}|_{\mathscr{S}_{t,t}} \overset{d}{=} \sqrt{\sigma^t_{\mathsf{r}}}\breve{\boldsymbol{Z}}_{t,\mathsf{r}} + \breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}$ and $\boldsymbol{b}^s_{\mathsf{r}}|_{\mathscr{S}_{t,t}} \overset{d}{=} \sqrt{\sigma^s_{\mathsf{r}}}\breve{\boldsymbol{Z}}_{s,\mathsf{r}} + \breve{\boldsymbol{\Delta}}_{s,\mathsf{r}}$, where $\breve{\boldsymbol{Z}}_{t,\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$ and $\breve{\boldsymbol{Z}}_{s,\mathsf{r}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{R}}})$, such that for any $i \in [N_{\mathsf{R}}]$, the entries $[\breve{Z}_{s,\mathsf{r}}]_i$, $[\breve{Z}_{t,\mathsf{r}}]_i$ are

jointly Gaussian with covariance $\mathbb{E}\{[\breve{Z}_{s,\mathsf{r}}]_i[\breve{Z}_{t,\mathsf{r}}]_i\} = \sqrt{\sigma_{\mathsf{r}}^t/\sigma_{\mathsf{r}}^s}$. Now we use Lemma B.2 and the fact that $\sigma_{\mathsf{r}}^s \in \Theta(1)$ to write

$$
\begin{aligned}
&P\Big(\Big|\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}(\boldsymbol{b}_{\mathsf{r}}^s)^*\boldsymbol{b}_{\mathsf{r}}^t - \frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\sigma_{\mathsf{r}}^t\Big| \ge \epsilon\Big) \\
&= P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{(\sqrt{\sigma_{\mathsf{r}}^t}\breve{\boldsymbol{Z}}_{t,\mathsf{r}} + \breve{\boldsymbol{\Delta}}_{t,\mathsf{r}})^*(\sqrt{\sigma_{\mathsf{r}}^s}\breve{\boldsymbol{Z}}_{s,\mathsf{r}} + \breve{\boldsymbol{\Delta}}_{s,\mathsf{r}})}{N_{\mathsf{R}}} - \sigma_{\mathsf{r}}^t\Big]\Big| \ge \epsilon\Big) \\
&\le P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\sqrt{\sigma_{\mathsf{r}}^t\sigma_{\mathsf{r}}^s}\Big[\frac{\breve{\boldsymbol{Z}}_{t,\mathsf{r}}^*\breve{\boldsymbol{Z}}_{s,\mathsf{r}}}{N_{\mathsf{R}}} - \sqrt{\sigma_{\mathsf{r}}^t/\sigma_{\mathsf{r}}^s}\Big]\Big| \ge \frac{\epsilon}{4}\Big) + P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\frac{|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}^*\breve{\boldsymbol{\Delta}}_{s,\mathsf{r}}|}{n} \ge \frac{\epsilon}{4}\Big) \\
&\quad + P\Big(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\frac{|\breve{\boldsymbol{\Delta}}_{s,\mathsf{r}}^*\breve{\boldsymbol{Z}}_{t,\mathsf{r}}|}{N_{\mathsf{R}}} \ge \frac{\epsilon}{4}\Big) + P\Big(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\frac{|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}^*\breve{\boldsymbol{Z}}_{s,\mathsf{r}}|}{N_{\mathsf{R}}} \ge \frac{\epsilon}{4}\Big).
\end{aligned} \tag{179}
$$

Label the terms of the above as $T_1 - T_4$.

For T_1 recall from Lemma 7.5 that $\breve{\boldsymbol{Z}}_{t,\mathsf{c}} \stackrel{d}{=} \breve{\boldsymbol{Z}}_{s,\mathsf{r}}\sqrt{\sigma_{\mathsf{r}}^t/\sigma_{\mathsf{r}}^s} + \breve{\mathbf{U}}_{t,\mathsf{r}}\sqrt{1-\sigma_{\mathsf{r}}^t/\sigma_{\mathsf{r}}^s}$ where $\breve{\mathbf{U}}_{t,\mathsf{r}}$ and $\breve{\boldsymbol{Z}}_{s,\mathsf{r}}$ are independent. Therefore, using Lemma B.2,

$$
\begin{aligned}
T_1 &\le P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\sigma_{\mathsf{r}}^t\Big[\frac{\|\breve{\boldsymbol{Z}}_{s,\mathsf{r}}\|^2}{N_{\mathsf{R}}} - 1\Big]\Big| \ge \frac{\epsilon}{8}\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\sqrt{\sigma_{\mathsf{r}}^t(\sigma_{\mathsf{r}}^s-1)}\Big[\frac{\breve{\mathbf{U}}_{t,\mathsf{r}}^*\breve{\boldsymbol{Z}}_{s,\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \ge \frac{\epsilon}{8}\Big) \\
&\le K\exp\{-\kappa n\epsilon^2(\omega/\mathsf{R})^{\max\{2\mathsf{u}-1,0\}}\},
\end{aligned} \tag{180}
$$

where the last inequality is obtained by using Lemma B.9 to bound each of the two probabilities. We use (88) and the fact that $\sigma_{\mathsf{r}}^s \in \Theta(1)$.

Next for T_2, by two applications of Cauchy-Schwarz,

$$
\begin{aligned}
T_2 &\le P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|\|\breve{\boldsymbol{\Delta}}_{s,\mathsf{r}}\| \ge \frac{\kappa\epsilon}{4}\Big) \le P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|^2 \cdot \frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\|\breve{\boldsymbol{\Delta}}_{s,\mathsf{r}}\|^2 \ge \kappa\epsilon^2\Big) \\
&\le \Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\|\breve{\boldsymbol{\Delta}}_{t,\mathsf{r}}\|^2 \ge \kappa\epsilon\Big) + \Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\|\breve{\boldsymbol{\Delta}}_{s,\mathsf{r}}\|^2 \ge \kappa\epsilon\Big) \le t^4KK_{t-1}\Pi_{t-1}e^{-\frac{1}{t^8}\kappa\kappa_{t-1}(\omega/\mathsf{R})^{(\mathsf{u}-1)}+\pi_{t-1}\epsilon}.
\end{aligned} \tag{181}
$$

The final inequality follows by (177). Finally terms T_3, T_4 can be shown to have the desired upper bounded by work similar to that in (178).

(d) Let $0 \le \tilde{s}, s \le t$ where either $\tilde{s} = t$, $s = t$, or both $\tilde{s} = s = t$. Since $\boldsymbol{m}_{\mathsf{r}}^{s,\mathsf{c}} = S_{\mathsf{rc}}^s\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^s - \boldsymbol{w}_{\mathsf{r}})$,

$$
\begin{aligned}
&P\Big(\frac{N_{\mathsf{R}}}{L}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\Big(W_{\mathsf{rc}}^{\mathsf{u}-1/2}\frac{(\boldsymbol{b}_{\mathsf{r}}^{\tilde{s}})^*\boldsymbol{m}_{\mathsf{r}}^{s,\mathsf{c}}}{N_{\mathsf{R}}} - S_{\mathsf{rc}}^s W_{\mathsf{rc}}^{\mathsf{u}}\sigma_{\mathsf{r}}^{\max(\tilde{s},s)}\Big)\Big| \ge \epsilon\Big) \\
&\le P\Big(\frac{1}{\mathsf{R}}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\Big(\frac{n}{L}S_{\mathsf{rc}}^s\Big)\Big[\frac{(\boldsymbol{b}_{\mathsf{r}}^{\tilde{s}})^*\boldsymbol{b}_{\mathsf{r}}^s}{N_{\mathsf{R}}} - \sigma_{\mathsf{r}}^{\max(\tilde{s},s)}\Big]\Big| \ge \kappa\epsilon\Big) + P\Big(\frac{1}{\mathsf{R}}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}}\Big(\frac{n}{L}S_{\mathsf{rc}}^s\Big)\Big[\frac{(\boldsymbol{b}_{\mathsf{r}}^{\tilde{s}})^*\boldsymbol{w}_{\mathsf{r}}}{N_{\mathsf{R}}}\Big]\Big| \ge \kappa\epsilon\Big).
\end{aligned} \tag{182}
$$

Now the result follows from $\mathcal{B}_t(b)$ and $\mathcal{B}_t(c)$ using that $(\frac{n}{L}S_{\mathsf{rc}}^s) \in \Theta(1)$.

(e) Since $\boldsymbol{m}_{\mathsf{r}}^{s,\mathsf{c}} \stackrel{d}{=} S_{\mathsf{r},\mathsf{c}}^s\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^s - \boldsymbol{w}_{\mathsf{r}})$, we have $(\boldsymbol{m}_{\mathsf{r}}^{s,\mathsf{c}})^*\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}} = S_{\mathsf{r},\mathsf{c}}^s S_{\mathsf{r},\mathsf{c}}^t W_{\mathsf{rc}}[(\boldsymbol{b}_{\mathsf{r}}^s)^*\boldsymbol{b}_{\mathsf{r}}^t - \boldsymbol{w}_{\mathsf{r}}^*\boldsymbol{b}_{\mathsf{r}}^s - \boldsymbol{w}_{\mathsf{r}}^*\boldsymbol{b}_{\mathsf{r}}^t + \|\boldsymbol{w}_{\mathsf{r}}\|^2]$. Therefore, using Lemma B.2 and that $\phi_{\mathsf{r}}^t = \sigma^2 + \sigma_{\mathsf{r}}^t$, we obtain

$$
\begin{aligned}
&P\Big(\frac{n}{L^2}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} W_{\mathsf{rc}}^{\mathsf{u}-1}(\boldsymbol{m}_{\mathsf{r}}^{s,\mathsf{c}})^*\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}} - N_{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]} S_{\mathsf{r},\mathsf{c}}^s\, S_{\mathsf{r},\mathsf{c}}^t\, W_{\mathsf{rc}}^{\mathsf{u}}\,\phi_{\mathsf{r}}^t\Big| \ge \epsilon\Big)\\
&= P\Big(\frac{n^2}{\mathsf{R}L^2}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} S_{\mathsf{r},\mathsf{c}}^s S_{\mathsf{r},\mathsf{c}}^t W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{(\boldsymbol{b}_{\mathsf{r}}^s)^*\boldsymbol{b}_{\mathsf{r}}^t}{N_{\mathsf{R}}} - \frac{\boldsymbol{w}_{\mathsf{r}}^*\boldsymbol{b}_{\mathsf{r}}^t}{N_{\mathsf{R}}} - \frac{\boldsymbol{w}_{\mathsf{r}}^*\boldsymbol{b}_{\mathsf{r}}^s}{N_{\mathsf{R}}} + \frac{\|\boldsymbol{w}_{\mathsf{r}}\|^2}{N_{\mathsf{R}}} - (\sigma^2+\sigma_{\mathsf{r}}^t)\Big]\Big| \ge \epsilon\Big)\\
&\le P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}(\frac{n^2}{L^2}S_{\mathsf{r},\mathsf{c}}^s S_{\mathsf{r},\mathsf{c}}^t)W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{(\boldsymbol{b}_{\mathsf{r}}^s)^*\boldsymbol{b}_{\mathsf{r}}^t}{N_{\mathsf{R}}} - \sigma_{\mathsf{r}}^t\Big]\Big| \ge \frac{\epsilon}{4}\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}(\frac{n^2}{L^2}S_{\mathsf{r},\mathsf{c}}^s S_{\mathsf{r},\mathsf{c}}^t)W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{\boldsymbol{w}_{\mathsf{r}}^*\boldsymbol{b}_{\mathsf{r}}^t}{N_{\mathsf{R}}}\Big]\Big| \ge \frac{\epsilon}{4}\Big)\\
&+ P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}(\frac{n^2}{L^2}S_{\mathsf{r},\mathsf{c}}^s S_{\mathsf{r},\mathsf{c}}^t)W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{\boldsymbol{w}_{\mathsf{r}}^*\boldsymbol{b}_{\mathsf{r}}^s}{N_{\mathsf{R}}}\Big]\Big| \ge \frac{\epsilon}{4}\Big) + P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}(\frac{n^2}{L^2}S_{\mathsf{r},\mathsf{c}}^s S_{\mathsf{r},\mathsf{c}}^t)W_{\mathsf{rc}}^{\mathsf{u}}\Big[\frac{\|\boldsymbol{w}_{\mathsf{r}}\|^2}{N_{\mathsf{R}}} - \sigma^2\Big]\Big| \ge \frac{\epsilon}{4}\Big).
\end{aligned}
$$

Now the result follows from Lemma B.9, $\mathcal{B}_t(b)$, and $\mathcal{B}_t(c)$ using that $(\frac{n}{L}S_{\mathsf{rc}}^s) \in \Theta(1)$.

(f) We first prove (97). Recall, $\boldsymbol{\alpha}^{t,\mathsf{c}} = \frac{n}{L^2}(\mathbb{M}_t^{\mathsf{c}})^{-1}(\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}$ where $\mathbb{M}_t^{\mathsf{c}} := \frac{n}{L^2}(\boldsymbol{M}_t^{\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}}$, so for $1 \le k \le t$, we have $\alpha_{k-1}^{t,\mathsf{c}} = \frac{n}{L^2}\sum_{i=1}^t[(\mathbb{M}_t^{\mathsf{c}})^{-1}]_{ki}(\boldsymbol{m}^{i-1,\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}$. From the definition of $\hat{\alpha}^{t,\mathsf{c}}$ in (66),

$$
\begin{aligned}
&P\Big(|\alpha_{k-1}^{t,\mathsf{c}} - \hat{\alpha}_{k-1}^{t,\mathsf{c}}| \ge \epsilon\Big) = P\Big(\Big|\sum_{i=1}^t\Big[\frac{n(\boldsymbol{m}^{i-1,\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}}{L^2}[(\mathbb{M}_t^{\mathsf{c}})^{-1}]_{ki} - \frac{n\tau_{\mathsf{c}}^t}{L}[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{ki}\Big]\Big| \ge \epsilon\Big)\\
&\stackrel{(a)}{\le} \sum_{i=1}^t P\Big(\Big|\frac{n(\boldsymbol{m}^{i-1,\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}}{L^2}[(\mathbb{M}_t^{\mathsf{c}})^{-1}]_{ki} - \frac{n\tau_{\mathsf{c}}^t}{L}[(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{ki}\Big| \ge \frac{\epsilon}{t}\Big) \stackrel{(b)}{\le} t^5 K K_{t-1}\Pi_{t-1}e^{-\frac{1}{t^{10}}\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2}.
\end{aligned}
$$

Step (a) follows by Lemma B.2 and (b) by Lemmas B.3, B.4 using $\mathcal{B}_t(e)$ with $\mathsf{u}=1$ and $\mathcal{B}_{t-1}(g)$ result (100). Note that $n\tau_{\mathsf{c}}^t/L$ and the absolute values of the non-zero entries of $(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}$ are $\Theta(1)$.

Next we prove (98). First, note that $\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2 = \|\boldsymbol{m}^{t,\mathsf{c}}\|^2 - \|\boldsymbol{m}_{\parallel}^{t,\mathsf{c}}\|^2 = \|\boldsymbol{m}^{t,\mathsf{c}}\|^2 - \|\boldsymbol{M}_t^{\mathsf{c}}\alpha^{t,\mathsf{c}}\|^2$. Using the definition of $\tau_{\perp,\mathsf{c}}^t$ in (67) and Lemma B.2, we have

$$
\begin{aligned}
&P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2 - \tau_{\perp,\mathsf{c}}^t\Big| \ge \epsilon\Big) = P\Big(\frac{n}{L}\Big|\frac{1}{L}\left\|\boldsymbol{m}^{t,\mathsf{c}}\right\|^2 - \frac{1}{L}\|\boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}}\|^2 - \tau_{\mathsf{c}}^t + \frac{(\tau_{\mathsf{c}}^t)^2}{\tau_{\mathsf{c}}^{t-1}}\Big| \ge \epsilon\Big)\\
&\le P\Big(\frac{n}{L}\Big|\frac{1}{L}\left\|\boldsymbol{m}^{t,\mathsf{c}}\right\|^2 - \tau_{\mathsf{c}}^t\Big| \ge \frac{\epsilon}{2}\Big) + P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}}\|^2 - \frac{(\tau_{\mathsf{c}}^t)^2}{\tau_{\mathsf{c}}^{t-1}}\Big| \ge \frac{\epsilon}{2}\Big). \qquad (183)
\end{aligned}
$$

The first term has the desired bound by $\mathcal{B}_t(e)$ with $\mathsf{u}=1$. For the second term, using $\boldsymbol{m}_{\parallel}^{t,\mathsf{c}} = \boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}} = \sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\boldsymbol{m}^{i,\mathsf{c}}$, we have $\|\boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}}\|^2 = (\boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}})^*\boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}} = \sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}(\boldsymbol{m}_{\parallel}^{t,\mathsf{c}})^*\boldsymbol{m}^{i,\mathsf{c}} = \sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}(\boldsymbol{m}^{i,\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}}$. Hence, recalling the definition of $\hat{\alpha}^{t,\mathsf{c}}$ in (66),

$$
\begin{aligned}
&P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{M}_t^{\mathsf{c}}\boldsymbol{\alpha}^{t,\mathsf{c}}\|^2 - \frac{(\tau_{\mathsf{c}}^t)^2}{\tau_{\mathsf{c}}^{t-1}}\Big| \ge \frac{\epsilon}{2}\Big) = P\Big(\frac{n}{L}\Big|\sum_{i=0}^{t-1}\Big(\frac{1}{L}\alpha_i^{t,\mathsf{c}}(\boldsymbol{m}^{i,\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}} - \hat{\alpha}_i^{t,\mathsf{c}}\tau_{\mathsf{c}}^t\Big)\Big| \ge \frac{\epsilon}{2}\Big)\\
&\le \sum_{i=0}^{t-1} P\Big(\frac{n}{L}\Big|\frac{1}{L}\alpha_i^{t,\mathsf{c}}(\boldsymbol{m}^{i,\mathsf{c}})^*\boldsymbol{m}^{t,\mathsf{c}} - \hat{\alpha}_i^{t,\mathsf{c}}\tau_{\mathsf{c}}^t\Big| \ge \frac{\epsilon}{2t}\Big) \stackrel{(a)}{\le} t^6 K K_{t-1}\Pi_{t-1}\exp\Big\{-\frac{1}{t^{12}}\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2\Big\}.
\end{aligned}
$$

Step (a) is obtained using Lemma B.3 when $i = t-1$ and Lemma B.4 otherwise, along with the results $\mathcal{B}_t(e)$ with $\mathsf{u}=1$ and $\mathcal{B}_t(f)$ proved in (97) above.

(g) We first show (99). For $0 \le \tilde{s}, s \le t$, $[\mathbb{M}^{\mathsf{c}}_{t+1}]_{\tilde{s}+1,s+1} = \frac{n}{L^2}(\boldsymbol{m}^{\tilde{s},\mathsf{c}})^* \boldsymbol{m}^{s,\mathsf{c}}$ and therefore, $[\mathbf{M}^{\mathsf{c}}_{t+1}]_{\tilde{s}+1,s+1}$ concentrates on $\frac{n}{L}\tau_{\mathsf{c}}^{\max\{\tilde{s},s\}}$ by $\mathcal{B}_t(e)$ with $\mathsf{u}=1$. By Fact C.1, if $\frac{n}{L^2}\left\|\boldsymbol{m}_{\perp}^{s,\mathsf{c}}\right\|^2 \ge \kappa > 0$ for all $0 \le s \le t$, then $\mathbb{M}^{\mathsf{c}}_{t+1}$ is invertible. Note from $\mathcal{B}_0(f) - \mathcal{B}_t(f)$ that $\frac{n}{L^2}\left\|\boldsymbol{m}_{\perp}^{s,\mathsf{c}}\right\|^2$ concentrates on $\frac{n}{L}\tau^s_{\perp,\mathsf{c}}$, and $\frac{n}{L}\tau^s_{\perp,\mathsf{c}} > 0$ by Lemma 7.1. Choosing $\kappa = \frac{1}{2}\min\{2, \frac{n}{L}\tau^0_{\perp,\mathsf{c}}, \ldots, \frac{n}{L}\tau^t_{\perp,\mathsf{c}}\}$, we therefore have

$$P\Big(\mathbb{M}^{\mathsf{c}}_t \text{ singular}\Big) \le \sum_{s=0}^{t} P\Big(\frac{n}{L}\Big|\frac{1}{L}\left\|\boldsymbol{m}_{\perp}^{s,\mathsf{c}}\right\|^2 - \tau^s_{\perp,\mathsf{c}}\Big| \ge \kappa\Big) \le t^7 K K_{t-1}\Pi_{t-1} e^{-\frac{1}{t^{12}}\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2}, \tag{184}$$

where the second inequality follows from $\mathcal{B}_0(f) - \mathcal{B}_t(f)$.

Next, we show (100). We first note that each non-zero element of $(\breve{\boldsymbol{C}}^{t+1,\mathsf{c}})^{-1}$ is $\Theta(1)$. To see this, recall the definition of $\breve{\boldsymbol{C}}^{t+1,\mathsf{c}}$, from which it follows with work as in (154) that if $\breve{\boldsymbol{C}}^{t,\mathsf{c}}$ is invertible, by the block inversion formula we have

$$(\breve{\boldsymbol{C}}^{t+1,\mathsf{c}})^{-1} = \begin{pmatrix} (\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1} + \frac{L}{n}(\tau^t_{\perp,\mathsf{c}})^{-1}\hat{\boldsymbol{\alpha}}^{t,\mathsf{c}}(\hat{\boldsymbol{\alpha}}^{t,c})^* & -\frac{L}{n}(\tau^t_{\perp,\mathsf{c}})^{-1}\hat{\boldsymbol{\alpha}}^{t,\mathsf{c}} \\ -\frac{L}{n}(\tau^t_{\perp,\mathsf{c}})^{-1}(\hat{\boldsymbol{\alpha}}^{t,c})^* & \frac{L}{n}(\tau^t_{\perp,\mathsf{c}})^{-1} \end{pmatrix}. \tag{185}$$

Each non-zero element of (185) is $\in \Theta(1)$ since $\tau^s_{\mathsf{c}} \in \Theta(L/n)$, for $0 \le s \le T$, and $\tau^t_{\perp,\mathsf{c}}$ is bounded below (by Lemma 7.1). We can similarly represent $[\mathbb{M}^{\mathsf{c}}_{t+1}]^{-1}$ by block inversion. Noting that

$$\mathbb{M}^{\mathsf{c}}_{t+1} = \frac{n}{L^2}\begin{pmatrix} \frac{n}{L^2}\mathbb{M}^{\mathsf{c}}_t & (\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{m}^{t,\mathsf{c}} \\ ((\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{m}^{t,\mathsf{c}})^* & \left\|\boldsymbol{m}^{t,\mathsf{c}}\right\|^2 \end{pmatrix}, \tag{186}$$

if $\mathbb{M}^{\mathsf{c}}_t$ is invertible, by the block inversion formula we have

$$(\mathbb{M}^{\mathsf{c}}_{t+1})^{-1} = \begin{pmatrix} (\mathbb{M}^{\mathsf{c}}_t)^{-1} + \frac{n}{L^2}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2}\boldsymbol{\alpha}^{t,\mathsf{c}}(\boldsymbol{\alpha}^{t,\mathsf{c}})^* & -\frac{n}{L^2}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2}\boldsymbol{\alpha}^{t,\mathsf{c}} \\ -\frac{n}{L^2}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2}(\boldsymbol{\alpha}^{t,\mathsf{c}})^* & \frac{n}{L^2}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2} \end{pmatrix}, \tag{187}$$

where we have used $\boldsymbol{\alpha}^{t,\mathsf{c}} = \frac{n}{L^2}(\mathbb{M}^{\mathsf{c}}_t)^{-1}(\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{m}^{t,\mathsf{c}}$ and $((\boldsymbol{M}^{\mathsf{c}}_t)^*\boldsymbol{m}^{t,\mathsf{c}})^*\boldsymbol{\alpha}^{t,\mathsf{c}} = (\boldsymbol{m}^{t,\mathsf{c}})^*\boldsymbol{m}_{\|}^{t,\mathsf{c}}$.

In what follows we prove concentration for each of the elements in (187) to the corresponding element of (185). First, by $\mathcal{B}_t(f)$ and Lemma B.7,

$$P\Big(\Big|\frac{L^2}{n\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2} - \frac{L}{n}(\tau^t_{\perp,\mathsf{c}})^{-1}\Big| \ge \epsilon\Big) \le t^6 K K_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^{12}}\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2\}. \tag{188}$$

Next, consider the i^{th} element of $-\frac{L^2}{n}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2}\boldsymbol{\alpha}^{t,\mathsf{c}}$. For $i \in [t]$,

$$P\Big(\Big|\frac{L^2\alpha^{t,\mathsf{c}}_{i-1}}{n\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2} - \frac{L\hat{\alpha}^{t,\mathsf{c}}_{i-1}}{n\tau^t_{\perp,\mathsf{c}}}\Big| \ge \epsilon\Big) \le t^6 K K_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^{12}}\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2\}. \tag{189}$$

The above follows from (188), Lemma B.3 for $i = t$ or Lemma B.4 when $i \in [t-1]$, and $\mathcal{B}_t(f)$. Finally, consider element (i,j) of $(\mathbb{M}^{\mathsf{c}}_t)^{-1} + (L^2/n)\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2}\boldsymbol{\alpha}^{t,\mathsf{c}}(\boldsymbol{\alpha}^{t,\mathsf{c}})^*$ for $i,j \in [t]$.

$$\begin{aligned}
&P\Big(\Big|[(\mathbb{M}^{\mathsf{c}}_t)^{-1}]_{ij} + \frac{L^2}{n}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^{-2}\alpha^{t,\mathsf{c}}_{i-1}\alpha^{t,\mathsf{c}}_{j-1} - [(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{ij} - \frac{L}{n}(\tau^t_{\perp,\mathsf{c}})^{-2}\hat{\alpha}^{t,\mathsf{c}}_{i-1}\hat{\alpha}^{t,\mathsf{c}}_{j-1}\Big| \ge \epsilon\Big) \\
&\overset{(a)}{\le} P\Big(\Big|[(\mathbb{M}^{\mathsf{c}}_t)^{-1}]_{ij} - [(\breve{\boldsymbol{C}}^{t,\mathsf{c}})^{-1}]_{ij}\Big| \ge \frac{\epsilon}{2}\Big) + P\Big(\Big|\alpha^{t,\mathsf{c}}_{j-1} - \hat{\alpha}^{t,\mathsf{c}}_{j-1}\Big| \ge \frac{\epsilon'}{2}\Big) + P\Big(\Big|\frac{L^2\alpha^{t,\mathsf{c}}_{i-1}}{n\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2} - \frac{L\hat{\alpha}^{t,\mathsf{c}}_{i-1}}{n\tau^t_{\perp,\mathsf{c}}}\Big| \ge \frac{\epsilon'}{2}\Big) \\
&\overset{(b)}{\le} t^6 K K_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^{12}}\kappa\kappa_{t-1}\pi_{t-1}\epsilon^2\}.
\end{aligned} \tag{190}$$

Step (a) follows from Lemma B.2 and Lemma B.3/B.4 with $\epsilon' = \min(\sqrt{\epsilon/3}, \frac{n\tau_{\perp,\mathsf{c}}^t}{3L\hat{\alpha}_{i-1}^{t,\mathsf{c}}}, \frac{\epsilon}{3\hat{\alpha}_{j-1}^{t,\mathsf{c}}})$. Step (b) follows from the inductive hypothesis $\mathcal{B}_{t-1}(g)$ (Eq. (100)), $\mathcal{B}_t(f)$, and (189).

8.5 Step 4: Showing $\mathcal{H}_{t+1}$ holds

We prove the statements in $\mathcal{H}_{t+1}$ assuming that $\mathcal{B}_0, \ldots, \mathcal{B}_t$, and $\mathcal{H}_1, \ldots, \mathcal{H}_t$ hold due to the induction hypothesis. We begin with a lemma that is used to prove $\mathcal{H}_{t+1}(a)$. The lemma as well as other parts of $\mathcal{H}_{t+1}$ assume the invertibility of $\mathbb{Q}_1^{\mathsf{r}}, \ldots, \mathbb{Q}_t^{\mathsf{r}}$, but for the sake of brevity, we do not explicitly specify the conditioning.

Lemma 8.2. *Let* $\mathbf{v}^{r,c} := \frac{1}{L}(\boldsymbol{Y}_{t+1,r})^* \boldsymbol{m}_{\perp,r}^{t,c}$ *and* $\mathbb{Q}_{t+1}^r = \frac{1}{L}(\boldsymbol{Q}_{t+1}^r)^* \boldsymbol{Q}_{t+1}^r$. *If* $\mathbb{Q}_1^r, \ldots, \mathbb{Q}_{t+1}^r$ *are invertible, for* $0 \le j \le t-1$,

$$P\Big(\Big|\sum_{r\in[R]} \sqrt{W_{\mathsf{rc}}}[(\mathbb{Q}_{t+1}^r)^{-1}\mathbf{v}^{r,c}]_{j+1} + \alpha_j^{t,c}\Big| \ge \epsilon\Big) \le t^2 K K'_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^4}\kappa\kappa'_{t-1}\pi_{t-1}\epsilon^2\} \tag{191}$$

$$P\Big(\Big|\sum_{r\in[R]} \sqrt{W_{\mathsf{rc}}}[(\mathbb{Q}_{t+1}^r)^{-1}\mathbf{v}^{r,c}]_{t+1} - 1\Big| \ge \epsilon\Big) \le t^2 K K'_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^4}\kappa\kappa'_{t-1}\pi_{t-1}\epsilon^2\}. \tag{192}$$

Proof. By induction hypothesis $\mathcal{H}_t(g)$, element-wise $(\mathbb{Q}_{t+1}^{\mathsf{r}})^{-1}$ concentrates to $(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}$, so before proving (191) and (192), we prove intermediary results about element-wise concentration of $\mathbf{v}^{\mathsf{r},\mathsf{c}}$. For $0 \le k \le t$, define

$$\mathbb{E}_{k+1,\mathsf{r}} = \frac{N_{\mathsf{R}}\sqrt{W_{\mathsf{rc}}}}{L}\Big(S_{\mathsf{rc}}^t\sigma_{\mathsf{r}}^t - \hat{\alpha}_{t-1}^{t,\mathsf{c}}S_{\mathsf{rc}}^{t-1}\sigma_{\mathsf{r}}^{\max(k,t-1)}\Big) = \frac{N_{\mathsf{R}}\tau_{\mathsf{c}}^t\sqrt{W_{\mathsf{rc}}}}{L}\Big(\frac{\sigma_{\mathsf{r}}^t}{\phi_{\mathsf{r}}^t} - \frac{\sigma_{\mathsf{r}}^{\max(k,t-1)}}{\phi_{\mathsf{r}}^{t-1}}\Big). \tag{193}$$

For $\mathsf{B} > 0$, a universal constant, we will prove that

$$P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]} \sqrt{W_{\mathsf{rc}}}(\mathrm{v}_{k+1}^{\mathsf{r},\mathsf{c}} - \mathbb{E}_{k+1,\mathsf{r}})\Big| \ge \epsilon\Big) \le tKK'_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^2}\kappa\kappa'_{t-1}\pi_{t-1}\epsilon^2\}, \tag{194}$$

$$P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]} \sqrt{W_{\mathsf{rc}}}\,|\mathrm{v}_{k+1}^{\mathsf{r},\mathsf{c}}| \ge \mathsf{B}\Big) \le tKK'_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^2}\kappa\kappa'_{t-1}\pi_{t-1}\}. \tag{195}$$

We prove the lemma using (194) and (195), and then prove (194). We first claim that

$$\sum_{\mathsf{r}\in[\mathsf{R}]} \sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t+1}[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(j+1),k}\,\mathbb{E}_{k,\mathsf{r}} = -\hat{\alpha}_{t-1}^{t,\mathsf{c}}, \qquad 0 \le j \le (t-1), \tag{196}$$

$$\sum_{\mathsf{r}\in[\mathsf{R}]} \sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t+1}[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1),k}\,\mathbb{E}_{k,\mathsf{r}} = 1. \tag{197}$$

To show (197), noting from (193) that $\mathbb{E}_{1,\mathsf{r}} = \ldots = \mathbb{E}_{t,\mathsf{r}}$, we have

$$\sum_{\mathsf{r}\in[\mathsf{R}]} \sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t+1}[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\mathbb{E}_{k,\mathsf{r}} = \sum_{\mathsf{r}\in[\mathsf{R}]} \sqrt{W_{\mathsf{rc}}}\Big[\mathbb{E}_{t+1,\mathsf{r}}[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)(t+1)} + \mathbb{E}_{t,\mathsf{r}}\sum_{k=1}^{t}[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\Big]$$

$$\overset{(a)}{=} \sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\sqrt{W_{\mathsf{rc}}}}{\sigma_{\perp,\mathsf{r}}^t}\Big[\mathbb{E}_{t+1,\mathsf{r}} - \mathbb{E}_{t,\mathsf{r}}\hat{\gamma}_{t-1}^{t,\mathsf{r}}\Big] = \sum_{\mathsf{r}\in[\mathsf{R}]} \frac{\sqrt{W_{\mathsf{rc}}}}{\sigma_{\perp,\mathsf{r}}^t}\Big[\mathbb{E}_{t+1,\mathsf{r}} - \mathbb{E}_{t,\mathsf{r}}\hat{\gamma}_{t-1}^{t,\mathsf{r}}\Big] \overset{(b)}{=} \frac{N_{\mathsf{R}}\tau_{\mathsf{c}}^t}{L}\sum_{\mathsf{r}\in[\mathsf{R}]} \frac{W_{\mathsf{rc}}}{\phi_{\mathsf{r}}^t} = 1. \tag{198}$$

In the above, step (a) is obtained as follows using the definition of $\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}}$ in (65). Using block inversion as we did for $\breve{\boldsymbol{C}}^{t,\mathsf{c}}$ in (154), it is straightforward to show that if $\widetilde{\boldsymbol{C}}^{t,\mathsf{r}}$ is invertible,

$$(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1} = \begin{pmatrix} (\widetilde{\boldsymbol{C}}^{t,\mathsf{r}})^{-1} + (\sigma^t_{\perp,\mathsf{r}})^{-1}\hat{\boldsymbol{\gamma}}^{t,\mathsf{r}}(\hat{\boldsymbol{\gamma}}^{t,\mathsf{r}})^* & -(\sigma^t_{\perp,\mathsf{r}})^{-1}\hat{\boldsymbol{\gamma}}^{t,\mathsf{r}} \\ -(\sigma^t_{\perp,\mathsf{r}})^{-1}(\hat{\boldsymbol{\gamma}}^{t,\mathsf{r}})^* & (\sigma^t_{\perp,\mathsf{r}})^{-1} \end{pmatrix}. \tag{199}$$

Step (a) then follows since $\hat{\gamma}^{t,\mathsf{r}}_0 = \ldots = \hat{\gamma}^{t,\mathsf{r}}_{t-2} = 0$. Step (b) of (198) then follows by using the definitions of $\mathbb{E}_{t+1,\mathsf{r}}, \mathbb{E}_{t,\mathsf{r}}$ in (193) and of $\hat{\gamma}^{t,\mathsf{r}}_{t-1}$ in (66). The result in (196) can be shown similarly, using that for $0 \le j \le (t-1)$, $\sum_{k=1}^t [(\widetilde{\boldsymbol{C}}^{t,\mathsf{r}})^{-1}]_{(j+1)k} = [(\widetilde{\boldsymbol{C}}^{t,\mathsf{r}})^{-1}(1,\ldots,1)^*]_{j+1} = (\sigma^t_{\mathsf{r}})^{-1}\hat{\gamma}^{t,\mathsf{r}}_j$, where the last equality follows from the definition of $\hat{\gamma}^{t,\mathsf{r}}$ in (66).

We now use these results to prove (192). Using (197) and Lemma B.2, the LHS of (192) can be bounded as follows:

$$\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_{t+1} - 1\Big| \ge \epsilon\Big) = P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t+1}[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(t+1)k}\mathrm{v}^{\mathsf{r},\mathsf{c}}_k - 1\Big| \ge \epsilon\Big) \\
&= P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t+1}\Big[[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(t+1)k}\mathrm{v}^{\mathsf{r},\mathsf{c}}_k - [(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\,\mathbb{E}_{k,\mathsf{r}}\Big]\Big| \ge \epsilon\Big) \\
&\le \sum_{k=1}^{t+1} P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}[\mathrm{v}^{\mathsf{r},\mathsf{c}}_k - \mathbb{E}_{k,\mathsf{r}}]\Big| \ge \frac{\epsilon}{2(t+1)}\Big) \\
&\quad + \sum_{k=1}^{t+1} P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\mathrm{v}^{\mathsf{r},\mathsf{c}}_k\Big[[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(t+1)k} - [(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\Big]\Big| \ge \frac{\epsilon}{2(t+1)}\Big).
\end{aligned} \tag{200}$$

Label the terms on the RHS of (200) as T_1 and T_2. Note that term T_1 has the desired upper bound using (194) and the fact that the non-zero elements of $(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1} \in \Theta(1)$. Now consider T_2.

$$\begin{aligned}
T_2 &\le \sum_{k=1}^{t+1} P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\,|\mathrm{v}^{\mathsf{r},\mathsf{c}}_k|\,\Big|[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(t+1)k} - [(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\Big| \ge \frac{\epsilon}{2(t+1)}\Big) \\
&\le \sum_{k=1}^{t+1}\Big[P\Big(\cup_{\mathsf{r}\in[\mathsf{R}]}\Big|[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(t+1)k} - [(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\Big| \le \frac{\kappa\epsilon}{2t\mathsf{B}}\Big) + P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\,|\mathrm{v}^{\mathsf{r},\mathsf{c}}_k| \ge \mathsf{B}\Big)\Big] \\
&\le \sum_{k=1}^{t+1}\Big[\sum_{\mathsf{r}\in[\mathsf{R}]}P\Big(\Big|[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(t+1)k} - [(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(t+1)k}\Big| \ge \frac{\kappa\epsilon}{2t\mathsf{B}}\Big) + P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\,|\mathrm{v}^{\mathsf{r},\mathsf{c}}_k| \ge \mathsf{B}\Big)\Big] \\
&\le t\mathsf{R}KK'_{t-1}\Pi'_{t-2}\exp\Big\{-\frac{\kappa\kappa'_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}\epsilon^2}{t^2(\log M)^2}\Big\} + t^2KK'_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^2}\kappa\kappa'_{t-1}\pi_{t-1}\}.
\end{aligned} \tag{201}$$

The upper bound for the two terms of (201) follows by (195) and $\mathcal{H}_t(g)$. The bound in (192) follows since $\mathsf{R}\Pi'_{t-2} = \Pi_{t-1}$ and $(\omega/\mathsf{R})^2\pi'_{t-2}(\log M)^2 = \pi_{t-1}$.

We now show the bound in (191) for $0 \le j \le t-1$. By Lemma B.2,

$$P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_{j+1}+\alpha_j^{t,\mathsf{c}}\Big|\ge\epsilon\Big)$$
$$\le P\Big(|\alpha_j^{t,\mathsf{c}}-\hat\alpha_j^{t,\mathsf{c}}|\ge\frac{\epsilon}{2}\Big)+P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_{j+1}+\hat\alpha_j^{t,\mathsf{c}}\Big|\ge\frac{\epsilon}{2}\Big).$$

The first term of the above is upper bounded with $\mathcal{B}_t(f)$. Using (196), the second term above is

$$P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}[(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}\mathbf{v}^{\mathsf{r},\mathsf{c}}]_{j+1}+\hat\alpha_j^{t,\mathsf{c}}\Big|\ge\frac{\epsilon}{2}\Big)$$
$$=P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\sum_{k=1}^{t+1}\Big([(\mathbb{Q}^{\mathsf{r}}_{t+1})^{-1}]_{(j+1),k}\mathrm{v}_k^{\mathsf{r},\mathsf{c}}-[(\widetilde{\boldsymbol{C}}^{t+1,\mathsf{r}})^{-1}]_{(j+1)k}\,\mathbb{E}_{k,\mathsf{r}}\Big)\Big|\ge\frac{\epsilon}{2}\Big). \tag{202}$$

Now we can bound the term in (202) with work similar to that used for (200).

To complete the proof, it remains to show (194), (195). Using the definition of $\boldsymbol{Y}_{t+1}$ in (59),

$$\sqrt{W_{\mathsf{rc}}}\mathrm{v}_{k+1}^{\mathsf{r},\mathsf{c}}=\frac{\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^k+\upsilon_{\mathsf{r}}^k\breve{\boldsymbol{m}}_{\mathsf{r}}^{k-1})^*\,\boldsymbol{m}_{\perp,\mathsf{r}}^{t,\mathsf{c}}}{L}=\Big(\frac{\sqrt{W_{\mathsf{rc}}}\boldsymbol{b}_{\mathsf{r}}^k}{L}+\frac{\upsilon_{\mathsf{r}}^k\boldsymbol{m}_{\mathsf{r}}^{k-1,\mathsf{c}}}{LS_{\mathsf{rc}}^{k-1}}\Big)^*(\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}-\sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}}), \tag{203}$$

where we have used $\upsilon_{\mathsf{r}}^k=\sigma_{\mathsf{r}}^k/\phi_{\mathsf{r}}^{k-1}$ and $\boldsymbol{m}_{\perp,\mathsf{r}}^{t,\mathsf{c}}=\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}-\boldsymbol{m}_{\|,\mathsf{r}}^{t,\mathsf{c}}=\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}-\sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}}$.

We first prove (194). Using (203), we have the following bound for the probability in (194):

$$P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}(\mathrm{v}_{k+1}^{\mathsf{r},\mathsf{c}}-\mathbb{E}_{k+1,\mathsf{r}})\Big|\ge\epsilon\Big)$$
$$\overset{(a)}{\le}P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^k)^*(\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}-\sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}})}{L}-\frac{N_{\mathsf{R}}}{L}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}(S_{\mathsf{rc}}^t\sigma_{\mathsf{r}}^t-\hat\alpha_{t-1}^{t,\mathsf{c}}S_{\mathsf{rc}}^{t-1}\sigma_{\mathsf{r}}^{\max(k,t-1)})\Big|\ge\frac{\epsilon}{2}\Big)$$
$$+P\Big(\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{\upsilon_{\mathsf{r}}^k(\boldsymbol{m}_{\mathsf{r}}^{k-1,\mathsf{c}})^*\,(\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}-\sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}})}{LS_{\mathsf{rc}}^{k-1}}\Big|\ge\frac{\epsilon}{2}\Big). \tag{204}$$

The terms on the RHS of (204) are bounded via the inductive hypotheses $\mathcal{B}_t(d)$ and $\mathcal{B}_t(e)$ with $\mathsf{u}=1$ and $\mathcal{B}_t(f)$, similarly to the first term on the RHS of (159) (see (160)).

We now prove (195) using (158) and $\upsilon_{\mathsf{r}}^k/S_{\mathsf{rc}}^{k-1}=\sigma_{\mathsf{r}}^k/\tau_{\mathsf{c}}^{k-1}$. Recall that $\tau_{\mathsf{c}}^{k-1}\in\Theta(L/n)$, then

$$P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{W_{\mathsf{rc}}}\,|v_{k+1}^{\mathsf{r},\mathsf{c}}|\ge\mathsf{B}\Big)=P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]}\Big|\Big(\frac{\sqrt{W_{\mathsf{rc}}}\boldsymbol{b}_{\mathsf{r}}^k}{L}+\frac{\upsilon_r^k\boldsymbol{m}_{\mathsf{r}}^{k-1,\mathsf{c}}}{LS_{\mathsf{rc}}^{k-1}}\Big)^*(\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}-\sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}})\Big|\ge\mathsf{B}\Big)$$
$$\le P\Big(\sum_{\mathsf{r}\in[\mathsf{R}]}\Big|\frac{1}{L}\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^k)^*\,\boldsymbol{m}_{\mathsf{r}}^{t,\mathsf{c}}\Big|\ge\frac{\mathsf{B}}{4}\Big)+P\Big(\sum_{i=0}^{t-1}\Big|\alpha_i^{t,\mathsf{c}}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\Big|\frac{1}{L}\sqrt{W_{\mathsf{rc}}}(\boldsymbol{b}_{\mathsf{r}}^k)^*\,\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}}\Big|\ge\frac{\mathsf{B}}{4}\Big)$$
$$+P\Big(\frac{n}{L^2}\sum_{\mathsf{r}\in[\mathsf{R}]}\Big|(\boldsymbol{m}_{\mathsf{r}}^{k-1,\mathsf{c}})^*\,\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}}\Big|\ge\frac{\mathsf{B}}{4}\Big)+P\Big(\frac{n}{L^2}\sum_{i=0}^{t-1}\Big|\alpha_i^{t,\mathsf{c}}\Big|\sum_{\mathsf{r}\in[\mathsf{R}]}\Big|(\boldsymbol{m}_{\mathsf{r}}^{k-1,\mathsf{c}})^*\,\boldsymbol{m}_{\mathsf{r}}^{i,\mathsf{c}}\Big|\ge\frac{\mathsf{B}}{4}\Big). \tag{205}$$

We now provide upper bounds for the terms in (205), labelled $T_1 - T_4$, by taking

$$\mathsf{B} = 4\max\Big\{(\frac{n}{L}\tau_{\mathsf{c}}^0+1)(\hat{\alpha}_{t-1}^{t,\mathsf{c}}+1),\,(\frac{n}{L}\tau_{\mathsf{c}}^0+1)^{1/2}(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\sigma_{\mathsf{r}}^t+1)^{1/2}\Big\}. \tag{206}$$

First consider T_3. For any $0 \le j \le t$, using Cauchy-Schwarz $\sum_{\mathsf{r}\in[\mathsf{R}]}|(\boldsymbol{m}_{\mathsf{r}}^{k-1,\mathsf{c}})^*\,\boldsymbol{m}_{\mathsf{r}}^{j,\mathsf{c}}| \le \|\boldsymbol{m}^{k-1,\mathsf{c}}\|\|\boldsymbol{m}^{j,\mathsf{c}}\|$, and therefore using the fact that $\mathsf{B}^2 \ge 16(\frac{n}{L}\tau_{\mathsf{c}}^0+1)^2 \ge 16(\frac{n}{L}\tau_{\mathsf{c}}^{k-1}+1)(\frac{n}{L}\tau_{\mathsf{c}}^j+1)$, we have

$$\begin{aligned}
T_3 &\le P\Big(\frac{n}{L^2}\|\boldsymbol{m}^{k-1,\mathsf{c}}\|\,\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\| \ge \frac{\mathsf{B}}{4}\Big) \le P\Big(\frac{n}{L^2}\|\boldsymbol{m}^{k-1,\mathsf{c}}\|^2 \ge \frac{n}{L}\tau_{\mathsf{c}}^{k-1}+1\Big) + P\Big(\frac{n}{L^2}\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\|^2 \ge \frac{n}{L}\tau_{\mathsf{c}}^j+1\Big)\\
&\le P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{m}^{k-1,\mathsf{c}}\|^2 - \tau_{\mathsf{c}}^{k-1}\Big| \ge 1\Big) + P\Big(\frac{n}{L}\Big|\frac{1}{L}\|\boldsymbol{m}^{j,\mathsf{c}}\|^2 - \tau_{\mathsf{c}}^{j}\Big| \ge 1\Big).
\end{aligned} \tag{207}$$

This has the desired bound by $\mathcal{B}_t(e)$ with $\mathsf{u}=1$. Next we bound T_4 as follows using (207) and the fact that $\mathsf{B} > 4(\frac{n}{L}\tau_{\mathsf{c}}^0+1)(\hat{\alpha}_{t-1}^{t,\mathsf{c}}+1) > 4(\hat{\alpha}_{t-1}^{t,\mathsf{c}}+1)\sqrt{(\frac{n}{L}\tau_{\mathsf{c}}^{k-1}+1)(\frac{n}{L}\tau_{\mathsf{c}}^j+1)}$:

$$\begin{aligned}
T_4 &\le P\Big(\frac{n}{L^2}\|\boldsymbol{m}^{k-1,\mathsf{c}}\|\,\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\|\sum_{i=0}^{t-1}\Big|\alpha_i^{t,\mathsf{c}}\Big| \ge \frac{\mathsf{B}}{4}\Big)\\
&\le P\Big(\frac{n}{L^2}\|\boldsymbol{m}^{k-1,\mathsf{c}}\|\,\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\| \ge \sqrt{(\frac{n}{L}\tau_{\mathsf{c}}^{k-1}+1)(\frac{n}{L}\tau_{\mathsf{c}}^j+1)}\Big) + P\Big(\sum_{i=0}^{t-1}\Big|\alpha_i^{t,\mathsf{c}}\Big| \ge \sum_{i=0}^{t-1}\Big|\hat{\alpha}_i^{t,\mathsf{c}}\Big|+1\Big).
\end{aligned}$$

The first term in the above can be bounded as in (207) and the second as follows by $\mathcal{B}_t(f)$:

$$P\Big(\sum_{i=0}^{t-1}\Big|\alpha_i^{t,\mathsf{c}}\Big| \ge \sum_{i=0}^{t-1}\Big|\hat{\alpha}_i^{t,\mathsf{c}}\Big|+1\Big) \le \sum_{i=0}^{t-1}P\Big(|\alpha_i^{t,\mathsf{c}}-\hat{\alpha}_i^{t,\mathsf{c}}| \ge \frac{1}{t}\Big) \le tKK'_{t-1}\Pi_{t-1}e^{-\frac{1}{t^2}\kappa\kappa'_{t-1}\pi_{t-1}}. \tag{208}$$

Now we study term T_1. Using Cauchy-Schwarz, for any $0 \le j \le t$,

$$\sum_{\mathsf{r}\in[\mathsf{R}]}\frac{1}{L}\sqrt{W_{\mathsf{rc}}}|(\boldsymbol{b}_{\mathsf{r}}^k)^*\,\boldsymbol{m}_{\mathsf{r}}^{j,\mathsf{c}}| \le \sum_{\mathsf{r}\in[\mathsf{R}]}\sqrt{\frac{W_{\mathsf{rc}}}{n}}\|\boldsymbol{b}_{\mathsf{r}}^k\|\cdot\frac{\sqrt{n}}{L}\big\|\boldsymbol{m}_{\mathsf{r}}^{j,\mathsf{c}}\big\| \le \sqrt{\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\|\boldsymbol{b}_{\mathsf{r}}^k\|^2}\cdot\frac{\sqrt{n}}{L}\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\|. \tag{209}$$

Using the above in term T_1 we have,

$$\begin{aligned}
T_1 &\le P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\|\boldsymbol{b}_{\mathsf{r}}^k\|^2\cdot\frac{n}{L^2}\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\|^2 \ge \frac{\mathsf{B}^2}{16}\Big)\\
&\overset{(a)}{\le} P\Big(\frac{1}{n}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\|\boldsymbol{b}_{\mathsf{r}}^k\|^2 \ge \frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\sigma_{\mathsf{r}}^t+1\Big) + P\Big(\frac{n}{L^2}\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\|^2 \ge \frac{n}{L}\tau_{\mathsf{c}}^j+1\Big)\\
&\le P\Big(\Big|\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\Big(\frac{1}{N_{\mathsf{R}}}\|\boldsymbol{b}_{\mathsf{r}}^k\|^2-\sigma_{\mathsf{r}}^t\Big)\Big| \ge 1\Big) + P\Big(\frac{n}{L}\Big|\frac{1}{L}\big\|\boldsymbol{m}^{j,\mathsf{c}}\big\|^2-\tau_{\mathsf{c}}^j\Big| \ge 1\Big)\\
&\overset{(b)}{\le} KK'_{t-1}\Pi_{t-1}\exp\{-\kappa\kappa'_{t-1}\pi_{t-1}\} + KK'_{t-1}\Pi_{t-1}\exp\{-\kappa\kappa'_{t-1}\pi_{t-1}\}.
\end{aligned}$$

Step (a) uses $\frac{\mathsf{B}^2}{16} \ge (\frac{n}{L}\tau_{\mathsf{c}}^j+1)(\frac{1}{\mathsf{R}}\sum_{\mathsf{r}\in[\mathsf{R}]}W_{\mathsf{rc}}\sigma_{\mathsf{r}}^t+1)$ and (b) uses $\mathcal{B}_t(c)$ and $\mathcal{B}_t(e)$ both with $\mathsf{u}=1$.

Finally, using (209), we bound T_2:

$$T_2 \le P\Big(\sqrt{\frac{1}{n}\sum_{\mathsf{r}} W_{\mathsf{rc}}\|\boldsymbol{b}_{\mathsf{r}}^k\|^2}\cdot\frac{\sqrt{n}}{L}\left\|\boldsymbol{m}^{j,\mathsf{c}}\right\|\sum_{i=0}^{t-1}\left|\alpha_i^{t,\mathsf{c}}\right| \ge \frac{\mathsf{B}}{4}\Big)$$

$$\le P\Big(\frac{1}{n}\sum_{\mathsf{r}} W_{\mathsf{rc}}\|\boldsymbol{b}_{\mathsf{r}}^k\|^2\cdot\frac{n}{L^2}\left\|\boldsymbol{m}^{j,\mathsf{c}}\right\|^2 \ge (\frac{1}{\mathsf{R}}\sum_{\mathsf{r}} W_{\mathsf{rc}}\sigma_{\mathsf{r}}^t+1)(\frac{n}{L}\tau_{\mathsf{c}}^j+1)\Big) + P\Big(\sum_{i=0}^{t-1}\left|\alpha_i^{t,\mathsf{c}}\right| \ge \sum_{i=0}^{t-1}\left|\hat{\alpha}_i^{t,\mathsf{c}}\right|+1\Big).$$

Now the first term in the above can be bounded as the work in T_1 and the second term as in (208). This completes the proof of the lemma. ■

(a) To show the upper bound in (101), recall the definition of $\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}$ from Lemma 7.4 Eq. (82). First, using arguments similar to (167), we can show for a single entry $i \in \mathsf{c}$,

$$\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|}{\sqrt{L}}[\mathbf{P}^{\parallel}_{\boldsymbol{Q}_{t+1}^{\mathsf{r}'}}\boldsymbol{Z}_t^{\mathsf{r}'}]_{i\in\mathsf{c}} \overset{d}{=} Z\sqrt{\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|^2([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2}{L\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}}, \qquad Z\sim\mathcal{N}(0,1).$$

Therefore if we consider the maximum squared element for a section $\ell \in \mathsf{c}$, we have

$$\max_{i\in sec(\ell);\ell\in\mathsf{c}}\Big\{\Big(\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|}{\sqrt{L}}[\mathbf{P}^{\parallel}_{\boldsymbol{Q}_{t+1}^{\mathsf{r}'}}\boldsymbol{Z}_t^{\mathsf{r}'}]_{i\in\mathsf{c}}\Big)^2\Big\} = Z^2\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|^2}{L\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\max_{i\in sec(\ell)}([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2 \tag{210}$$

Now we simplify the final three terms of $\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}$ in (82), using the Lemma 8.2 notation $\mathbb{Q}_{t+1}^{\mathsf{r}}, \mathbf{v}^{\mathsf{rc}}$.

$$\sum_{\mathsf{r}'}\boldsymbol{Q}_{t+1,\mathsf{c}}^{\mathsf{r}'}(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}} + \sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\breve{\boldsymbol{q}}_{\mathsf{c}}^i - \breve{\boldsymbol{q}}_{\mathsf{c}}^t = \sum_{j=1}^{t+1}\sum_{\mathsf{r}'}\boldsymbol{q}_{\mathsf{c}}^{j-1,\mathsf{r}'}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_j + \sum_{i=0}^{t-1}\alpha_i^{t,\mathsf{c}}\breve{\boldsymbol{q}}_{\mathsf{c}}^i - \breve{\boldsymbol{q}}_{\mathsf{c}}^t$$

$$= \sum_{j=1}^{t}\breve{\boldsymbol{q}}_{\mathsf{c}}^{j-1}\Big[\sum_{\mathsf{r}'}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_j + \alpha_{j-1}^{t,\mathsf{c}}\Big] + \breve{\boldsymbol{q}}_{\mathsf{c}}^t\Big[\sum_{\mathsf{r}'}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_{t+1} - 1\Big].$$

Therefore for section $\ell \in \mathsf{c}$, using the triangle inequality, Lemma C.2, the fact that $\max_{j\in sec(\ell)}|[\breve{\boldsymbol{q}}_{\mathsf{c}}^s]_j|^2 \le 4$ for $0 \le s \le t$, and the definition of $\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}$ from (82), we have

$$\frac{1}{2(t+1)}\max_{j\in sec(\ell)}|[\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}]_j|^2 \le \sum_{i=0}^{t-2}|\alpha_i^{t,\mathsf{c}}-\hat{\alpha}_i^{t,\mathsf{c}}|^2\max_{j\in sec(\ell)}\left|[\boldsymbol{h}_{\mathsf{c}}^{i+1}]_j\right|^2$$

$$+ \Big[\frac{\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|}{\sqrt{L}} - \sqrt{\tau_{\perp,\mathsf{c}}^t}\Big]^2\max_{j\in sec(\ell)}|[\boldsymbol{Z}_{t,\mathsf{c}}]_j|^2 + Z^2\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|^2}{L\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\max_{i\in sec(\ell)}([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2$$

$$+ 4\sum_{j=1}^{t}\Big[\sum_{\mathsf{r}'\in[\mathsf{R}]}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_j + \alpha_{j-1}^{t,\mathsf{c}}\Big]^2 + 4\Big[\sum_{\mathsf{r}'\in[\mathsf{R}]}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_{t+1} - 1\Big]^2. \tag{211}$$

Using (211) and Lemma B.2, we have the following bound, where $\tilde{\epsilon}_t = \frac{\epsilon}{4(t+1)^2}$,

$$
\begin{aligned}
&P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\max_{j\in sec(\ell)} W_{\mathsf{rc}}^{2\mathsf{v}}\,|[\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}]_j|^2 \geq \epsilon\Big)\\
&\overset{(a)}{\leq} \sum_{i=0}^{t-1} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}\Big|\alpha_i^{t,\mathsf{c}} - \hat{\alpha}_i^{t,\mathsf{c}}\Big|^2 \sum_{\ell\in\mathsf{c}}\max_{j\in sec(\ell)}\big|[\boldsymbol{h}_{\mathsf{c}}^{i+1}]_j\big|^2 \geq \tilde{\epsilon}_t\Big)\\
&+ P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{L}{n}W_{\mathsf{rc}}^{2\mathsf{v}}\Big|\sqrt{\frac{n\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2}{L^2}} - \sqrt{\frac{n\tau_{\perp,c}^t}{L}}\Big|^2 \sum_{\ell\in\mathsf{c}}\max_{j\in sec(\ell)}|[\boldsymbol{Z}_{t,\mathsf{c}}]_j|^2 \geq \tilde{\epsilon}_t\Big)\\
&+ P\Big(\frac{Z^2}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|^2}{L\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2 \geq \tilde{\epsilon}_t\Big)\\
&+ P\Big(\frac{4}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}\Big|\sum_{\mathsf{r}'\in[\mathsf{R}]}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_t - 1\Big|^2 \geq \tilde{\epsilon}_t\Big)\\
&+ \sum_{i=1}^{t} P\Big(\frac{4}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}\Big|\sum_{\mathsf{r}'\in[\mathsf{R}]}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_i + \alpha_{i-1}^{t,\mathsf{c}}\Big|^2 \geq \tilde{\epsilon}_t\Big). \qquad (212)
\end{aligned}
$$

Label the terms on the RHS of (212) as $T_1 - T_5$. We show that each has the desired upper bound.

First consider T_1. Let κ_0 be defined via $(\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{rc}'}^{2\mathsf{v}})\tau_{\mathsf{c}}^t \log M = \kappa_0(\mathsf{R}/\omega)^{2\mathsf{v}}$, noting that $\kappa_0 = \Theta(1)$. Then,

$$
\begin{aligned}
T_1 &\leq \sum_{i=0}^{t-1}\Big[P\Big(\cap_{\mathsf{c}\in[\mathsf{C}]}\Big\{|\alpha_i^{t,\mathsf{c}} - \hat{\alpha}_i^{t,\mathsf{c}}|^2 \leq \frac{\tilde{\epsilon}_t}{(6\kappa_0(\mathsf{R}/\omega)^{2\mathsf{v}} + \epsilon)}\Big\}\Big)\\
&\quad + P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\boldsymbol{h}_{\mathsf{c}}^{i+1}]_j)^2 \leq 6(\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{rc}'}^{2\mathsf{v}})\tau_{\mathsf{c}}^t \log M + \epsilon\Big)\Big]\\
&\leq \sum_{i=0}^{t-1}\Big[\sum_{\mathsf{c}\in[\mathsf{C}]} P\Big(|\alpha_i^{t,\mathsf{c}} - \hat{\alpha}_i^{t,\mathsf{c}}|^2 \geq \frac{\tilde{\epsilon}_t}{(6\kappa_0(\mathsf{R}/\omega)^{2\mathsf{v}} + \epsilon)}\Big)\\
&\quad + P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\boldsymbol{h}_{\mathsf{c}}^{i+1}]_j)^2 \geq 6(\max_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{rc}'}^{2\mathsf{v}})\tau_{\mathsf{c}}^t \log M + \epsilon\Big)\Big].\\
&\leq t\mathsf{C}KK'_{t-1}\Pi_{t-1}\exp\{-\frac{1}{t^2}\kappa\kappa'_{t-1}(\omega/\mathsf{R})^{2\mathsf{v}}\pi_{t-1}\epsilon^2\} + tKK'_{t-1}\Pi'_{t-2}\exp\{-\kappa\kappa'_{t-1}(\omega/\mathsf{R})^{2\mathsf{v}}\pi'_{t-2}\epsilon^2\},
\end{aligned}
$$

The final inequality follows by $\mathcal{B}_t(f)$ and $\mathcal{H}_t(e)$.

Term T_2 of (212) has the desired upper bound by work similar to that used to bound the corresponding term in the $\mathcal{H}_1(a)$ step in equations (127) - (130).

Next consider term T_5 in (212). Using the union bound,

$$
\begin{aligned}
T_5 &\leq \sum_{i=1}^{t}\sum_{\mathsf{c}\in[\mathsf{C}]} P\Big(\Big|\sum_{\mathsf{r}'\in[\mathsf{R}]}\sqrt{W_{\mathsf{r}'\mathsf{c}}}[(\mathbb{Q}_{t+1}^{\mathsf{r}'})^{-1}\mathbf{v}^{\mathsf{r}',\mathsf{c}}]_i + \alpha_{i-1}^{t,\mathsf{c}}\Big| \geq \cdot\sqrt{\frac{\tilde{\epsilon}_t}{4(\sum_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{rc}'}^{2\mathsf{v}}/\mathsf{C})}}\Big)\\
&\leq t^3\mathsf{C}KK'_{t-1}\Pi_{t-1}\exp\Big\{\frac{-1}{t^6}\kappa\kappa'_{t-1}(\omega/\mathsf{R})^{(2\mathsf{v}-1)_+}\pi_{t-1}\epsilon\Big\},
\end{aligned} \qquad (213)
$$

where the second inequality follows from Lemma 8.2 and (88), which gives $\sum_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{rc}'}^{2\mathsf{v}}/\mathsf{C} \le \kappa(\mathsf{R}/\omega)^{(2\mathsf{v}-1)_+}$ for $\mathsf{v}\in\{0,1\}$. The bound for T_4 is obtained similarly to T_5.

Finally, we bound the third term T_3 in (212). For T_3, if all three of the following events hold:

$$\cap_{\mathsf{c}\in[\mathsf{C}]}\cap_{\mathsf{r}'\in[\mathsf{R}]}\cap_{j=0}^{t}\Big\{\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}\frac{([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2}{\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\le\frac{\kappa t\mathsf{R}}{\mathsf{C}\omega}\Big\},\qquad\Big\{Z^2\le\frac{N_{\mathsf{R}}\omega(\mathsf{R}/\omega)^{2\mathsf{v}}\,\tilde{\epsilon}_t}{t^2}\Big\},$$
$$\text{and }\Big\{\frac{(\omega/\mathsf{R})^{2\mathsf{v}}}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{nW_{\mathsf{rc}}^{2\mathsf{v}}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2}{L^2}\le\frac{n(\omega/\mathsf{R})^{2\mathsf{v}}}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}^{2\mathsf{v}}\tau_{\perp,\mathsf{c}}^{t}+1\Big\},$$

then

$$\frac{Z^2}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}^{2\mathsf{v}}\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}\frac{\|\boldsymbol{m}_{\perp,\mathsf{r}'}^{t,\mathsf{c}}\|^2}{L\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2\le\kappa'\tilde{\epsilon}_t,$$

where we have used that $\frac{n(\omega/\mathsf{R})^{2\mathsf{v}}}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}^{2\mathsf{v}}\tau_{\perp,\mathsf{c}}^{t}\le\kappa''$ using (88). Here, κ,κ', and κ'' are suitable universal positive constants. Therefore,

$$\begin{aligned}T_3\le&\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}P\Big(\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}\frac{([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2}{\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\ge\frac{\kappa t\mathsf{R}}{\mathsf{C}\omega}\Big)+P\Big(Z^2\ge\frac{N_{\mathsf{R}}\omega(\omega/\mathsf{R})^{2\mathsf{v}}\tilde{\epsilon}_t}{t^2}\Big)\\&+P\Big(\frac{(\omega/\mathsf{R})^{2\mathsf{v}}}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{nW_{\mathsf{rc}}^{2\mathsf{v}}\|\boldsymbol{m}_{\perp}^{t,\mathsf{c}}\|^2}{L^2}\ge\frac{n(\omega/\mathsf{R})^{2\mathsf{v}}}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}W_{\mathsf{rc}}^{2\mathsf{v}}\tau_{\perp,\mathsf{c}}^{t}+1\Big).\end{aligned}\tag{214}$$

We consider each of the terms of (214) separately. For the first term, using Lemma C.2 and the fact that $\max_{j\in sec(\ell)}|[\breve{\boldsymbol{q}}_{\mathsf{c}}^{s}]_j|^2\le 4$ for $0\le s\le t$, for any $\mathsf{c}\in[\mathsf{C}],\mathsf{r}'\in[\mathsf{R}]$ and $0\le j\le t$,

$$\begin{aligned}&\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2=W_{\mathsf{r}'\mathsf{c}}\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}([\breve{\boldsymbol{q}}_{\mathsf{c}}^{j}]_i-\sum_{k=0}^{j-1}\gamma_k^{j,\mathsf{r}'}[\breve{\boldsymbol{q}}_{\mathsf{c}}^{k}]_i)^2\\&\le 2tW_{\mathsf{r}'\mathsf{c}}\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}\Big\{([\breve{\boldsymbol{q}}_{\mathsf{c}}^{j}]_i)^2+\sum_{k=0}^{j-1}(\gamma_k^{j,\mathsf{r}'}[\breve{\boldsymbol{q}}_{\mathsf{c}}^{k}]_i)^2\Big\}\le\frac{4Pt(1+\sum_{k=0}^{j-1}(\gamma_k^{j,\mathsf{r}'})^2)L\mathsf{R}}{\mathsf{C}\omega},\end{aligned}$$

Therefore letting $\kappa:=\max_j\kappa_j$ with $\kappa_j:=4P(2+(\hat{\gamma}_{j-1}^{j,\mathsf{r}'})^2)((\sigma_{\perp,\mathsf{r}'}^{j+1})^{-1}+1)$ we have,

$$\begin{aligned}&\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}P\Big(\sum_{\ell\in\mathsf{c}}\max_{i\in sec(\ell)}\frac{([\boldsymbol{q}_{\perp,\mathsf{c}}^{j,\mathsf{r}'}]_i)^2}{\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\ge\frac{\kappa t\mathsf{R}}{\mathsf{C}\omega}\Big)\\&\le\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}P\Big(\frac{(1+\sum_{k=0}^{j-1}(\gamma_k^{j,\mathsf{r}'})^2)L}{\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}}\|^2}\ge(2+(\hat{\gamma}_{j-1}^{j,\mathsf{r}'})^2)\Big(\frac{1}{\sigma_{\perp,\mathsf{r}'}^{j+1}}+1\Big)\Big)\\&\le\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}'\in[\mathsf{R}]}\Big[P\Big(1+\sum_{k=0}^{j-1}(\gamma_k^{j,\mathsf{r}'})^2\ge 2+(\hat{\gamma}_{j-1}^{j,\mathsf{r}'})^2\Big)+P\Big(\frac{L}{\|\boldsymbol{q}_{\perp}^{j,\mathsf{r}'}\|^2}\ge\frac{1}{\sigma_{\perp,\mathsf{r}'}^{j+1}}+1\Big)\Big].\end{aligned}\tag{215}$$

For the second term on the RHS of (215), by Lemma B.7 and $\mathcal{H}_t(f)$,

$$\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}\in[\mathsf{R}]} P\Big(\frac{L}{\|\boldsymbol{q}_\perp^{j,\mathsf{r}}\|^2} \ge \frac{1}{\sigma_{\perp,\mathsf{r}}^{j+1}} + 1\Big) \le \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}\in[\mathsf{R}]} P\Big(\Big|\frac{L}{\|\boldsymbol{q}_\perp^{j,\mathsf{r}}\|^2} - \frac{1}{\sigma_{\perp,\mathsf{r}}^{j+1}}\Big| \ge 1\Big)$$
$$\le t\mathsf{R}\mathsf{C}KK'_{t-1}\Pi'_{t-2}\exp\Big\{-\frac{\kappa\kappa'_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}}{(\log M)^2}\Big\}.$$

For the first term on the RHS of (215), using Lemma B.2, the fact that $\hat{\gamma}_k^{t,\mathsf{r}} = 0$ for $0 \le k \le t-2$, along with Lemma B.6, and $\mathcal{H}_t(f)$,

$$\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}\in[\mathsf{R}]} P\Big(1+\sum_{k=0}^{j-1}(\gamma_k^{j,\mathsf{r}})^2 \ge 2 + (\hat{\gamma}_{j-1}^{j,\mathsf{r}})^2\Big) \le \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}\in[\mathsf{R}]} P\Big(\Big|\sum_{k=0}^{j-1}(\gamma_k^{j,\mathsf{r}})^2 - (\hat{\gamma}_{j-1}^{j,\mathsf{r}})^2\Big| \ge 1\Big)$$
$$\le \sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{j=0}^{t}\sum_{\mathsf{r}\in[\mathsf{R}]}\sum_{k=0}^{j-1} P\Big(\Big|(\gamma_k^{j,\mathsf{r}})^2 - (\hat{\gamma}_j^{j,\mathsf{r}})^2\Big| \ge \frac{1}{t}\Big) \le t\mathsf{R}\mathsf{C}KK'_{t-1}\Pi'_{t-2}\exp\Big\{\frac{-\kappa\kappa'_{t-1}(\omega/\mathsf{R})^2\pi'_{t-2}}{t^4(\log M)^2}\Big\}.$$

The second term of (214) is upper bounded by $K\exp\{-\kappa N_{\mathsf{R}}\omega(\omega/\mathsf{R})^{2\mathsf{v}}\epsilon^2/t^4\}$ using Lemma B.9. Finally, we bound the second term of (214).

$$P\Big(\frac{(\omega/\mathsf{R})^{2\mathsf{v}}}{\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{nW_{\mathsf{rc}}^{2\mathsf{v}}\|\boldsymbol{m}_\perp^{t,\mathsf{c}}\|^2}{L^2} \ge \frac{n(\omega/\mathsf{R})^{2\mathsf{v}}}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}\tau_{\perp,\mathsf{c}}^t + 1\Big)$$
$$\le P\Big(\frac{n(\omega/\mathsf{R})^{2\mathsf{v}}}{L\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}\Big(\frac{\|\boldsymbol{m}_\perp^{t,\mathsf{c}}\|^2}{L} - \tau_{\perp,\mathsf{c}}^t\Big)\Big| \ge 1\Big) \overset{(a)}{\le} P\Big(\frac{n}{L\mathsf{C}}\sum_{\mathsf{c}\in[\mathsf{C}]}\Big|\frac{\|\boldsymbol{m}_\perp^{t,\mathsf{c}}\|^2}{L} - \tau_{\perp,\mathsf{c}}^t\Big| \ge 1\Big)$$
$$\overset{(b)}{\le} \sum_{\mathsf{c}\in[\mathsf{C}]} P\Big(\frac{n}{L}\Big|\frac{\|\boldsymbol{m}_\perp^{t,\mathsf{c}}\|^2}{L} - \tau_{\perp,\mathsf{c}}^t\Big| \ge 1\Big) \overset{(c)}{\le} \mathsf{C}KK'_{t-1}\Pi_{t-1}\exp\{-\kappa\kappa'_{t-1}\pi_{t-1}\}.$$

In the above, step (a) follows since $(\omega W_{\mathsf{rc}}/\mathsf{R})^{2\mathsf{v}} \le ((1-\rho)P)^{2\mathsf{v}} \le \kappa$. Step (b) uses Lemma B.2 and step (c) from $\mathcal{B}_t(f)$. Note that $\sum_{\mathsf{c}'\in[\mathsf{C}]} W_{\mathsf{rc}'}/\mathsf{R} \le \kappa$.

We finally note that the desired upper bound in (101) follows since $\Pi'_{t-2} \le \Pi'_{t-1}$ and $\mathsf{R}\mathsf{C}\Pi'_{t-2} = \mathsf{C}\Pi_{t-1} = \Pi'_{t-1}$ along with the fact that $(\omega/\mathsf{R})^{(2\mathsf{v}-1)_+}\pi_{t-1}$, $(\omega/\mathsf{R})^{2\mathsf{v}}\pi'_{t-2}$, and $(\omega/\mathsf{R})^2\pi'_{t-2}/(\log M)^2$ are all lower bounded by $(\omega/\mathsf{R})^{2\mathsf{v}}\pi_{t-1} = (\omega/\mathsf{R})^{2\mathsf{v}}\pi'_{t-1}$.

(b) Using the conditional distribution of $\boldsymbol{h}^{t+1}$ from Lemma 7.4 Eq. (78), we have

$$P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^{t+1})^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}\Big| \ge \epsilon\Big)$$
$$= P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big(\frac{\tau_{\mathsf{c}}^t(\boldsymbol{h}_{\mathsf{c}}^t)^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}}{\tau_{\mathsf{c}}^{t-1}} + \sqrt{\tau_{\perp,\mathsf{c}}^t}\,\boldsymbol{Z}_{t,\mathsf{c}}^*\,\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}} + \boldsymbol{\Delta}_{t+1,t,\mathsf{c}}^*\,\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}\Big)\Big| \ge \epsilon\Big)$$
$$\le P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{\sqrt{W_{\mathsf{rc}}}\tau_{\mathsf{c}}^t(\boldsymbol{h}_{\mathsf{c}}^t)^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}}{\tau_{\mathsf{c}}^{t-1}}\Big| \ge \frac{\epsilon}{2}\Big) + P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big[\sqrt{\tau_{\perp,\mathsf{c}}^t}\,\boldsymbol{Z}_{t,\mathsf{c}}^*\,\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}} + \boldsymbol{\Delta}_{t+1,t,\mathsf{c}}^*\,\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}\Big]\Big| \ge \frac{\epsilon}{2}\Big).$$

The bound for first term above follows from the induction hypothesis $\mathcal{H}_t(b)$. The second term has the desired upper bound using work as in the proof of $\mathcal{H}_1(b)$ (Eqs. (132) - (133)).

(c) We show (103) when $0 \le s \le t+1$. For brevity, we write $\eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_0 - \boldsymbol{h}^s)$ to mean $\eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_{0,\mathsf{c}} - \boldsymbol{h}_{\mathsf{c}}^s)$, noting that $\eta_{\mathsf{c}}^{s-1}(\cdot)$ depends only on the elements of its input in column block c. Using the conditional distribution of $\boldsymbol{h}^{t+1}$ in Lemma 7.5 Eq. (83), we have

$$
\begin{aligned}
&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\Big[\frac{W_{\mathsf{rc}}^{\mathsf{v}}(\breve{\boldsymbol{q}}_{\mathsf{c}}^s)^*\breve{\boldsymbol{q}}_{\mathsf{c}}^{t+1}}{L} - \frac{W_{\mathsf{rc}}^{\mathsf{v}}\psi_{\mathsf{c}}^{t+1}}{\mathsf{C}}\Big]\Big| \ge \epsilon\Big) \\
&= P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{\mathsf{v}}\Big[\frac{\mathsf{C}}{L}(\eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^s - \tilde{\boldsymbol{\Delta}}_s) - \boldsymbol{\beta}_{0,\mathsf{c}})^*(\eta_{\mathsf{c}}^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1} - \tilde{\boldsymbol{\Delta}}_{t+1}) - \boldsymbol{\beta}_{0,\mathsf{c}}) - \psi_{\mathsf{c}}^{t+1}\Big]\Big| \ge \epsilon\Big) \\
&\le P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{\mathsf{v}}(\eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^s - \tilde{\boldsymbol{\Delta}}_s) - \boldsymbol{\beta}_{0,\mathsf{c}})^*\Big[\eta_{\mathsf{c}}^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1} - \tilde{\boldsymbol{\Delta}}_{t+1}) - \eta_{\mathsf{c}}^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1})\Big]\Big| \ge \frac{\epsilon}{3}\Big) \\
&+ P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{\mathsf{v}}\Big[\eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^s - \tilde{\boldsymbol{\Delta}}_s) - \eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^s)\Big]^*(\eta_{\mathsf{c}}^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1}) - \boldsymbol{\beta}_{0,\mathsf{c}})\Big| \ge \frac{\epsilon}{3}\Big) \\
&+ P\Big(\frac{1}{\mathsf{C}}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{\mathsf{v}}\Big[\frac{\mathsf{C}}{L}(\eta_{\mathsf{c}}^{s-1}(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^s) - \boldsymbol{\beta}_{0,\mathsf{c}})^*(\eta_{\mathsf{c}}^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1}) - \boldsymbol{\beta}_{0,\mathsf{c}}) - \psi_{\mathsf{c}}^{t+1}\Big]\Big| \ge \frac{\epsilon}{3}\Big).
\end{aligned}
\tag{216}
$$

Label the three terms on the RHS of (216) as $T_1 - T_3$ and provide bounds for each. First note the following bound for $\tilde{\boldsymbol{\Delta}}_{s,\mathsf{c}}$ (defined in Lemma 7.5) that will be used repeatedly. For $1 \le s \le t+1$,

$$
\begin{aligned}
&P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{\mathsf{v}} \max_{j\in sec(\ell)} |[\tilde{\boldsymbol{\Delta}}_{s,\mathsf{c}}]_j| \ge \epsilon\Big) = P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{\mathsf{v}} \max_{j\in sec(\ell)} \Big|\sum_{i=0}^{s-1}\Big(\frac{\tau_{\mathsf{c}}^{s-1}}{\tau_{\mathsf{c}}^i}\Big)[\boldsymbol{\Delta}_{i+1,i,\mathsf{c}}]_j\Big| \ge \epsilon\Big) \\
&\overset{(a)}{\le} \sum_{i=0}^{s-1} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}} \max_{j\in sec(\ell)} |[\boldsymbol{\Delta}_{t+1,t,\mathsf{c}}]_j|^2 \ge \frac{\epsilon^2}{s^2}\Big) \overset{(b)}{\le} t^4 K K'_{t-1}\Pi'_{t-1} e^{-\frac{1}{t^8}\kappa\kappa'_{t-1}(\omega/\mathsf{R})^{2\mathsf{v}}\pi'_{t-1}\epsilon^2}.
\end{aligned}
\tag{217}
$$

In the above, step (a) follows by using the triangle inequality, and then Cauchy-Schwarz, noting that $\tau_{\mathsf{c}}^{s-1}/\tau_{\mathsf{c}}^i \le 1$ for $i \le s-1$. Step (b) follows from $\mathcal{H}_{t+1}(a)$ and $s \le (t+1)$.

Consider T_1, the first term on the RHS of (216). Since $\max_j |\eta_j^{s-1}(\cdot) - \beta_{0,j}| \le 2$, we have

$$
\begin{aligned}
T_1 &\le P\Big(\frac{2}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\sum_{j\in\ell} W_{\mathsf{rc}}^{\mathsf{v}}\Big|\eta_j^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1} - \tilde{\boldsymbol{\Delta}}_{t+1}) - \eta_j^t(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{t+1})\Big| \ge \kappa\epsilon\Big) \\
&\overset{(a)}{\le} P\Big(\frac{4}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\frac{W_{\mathsf{rc}}^{\mathsf{v}}}{\tau_{\mathsf{c}}^t}\sum_{\ell\in\mathsf{c}} \max_{j\in sec(\ell)} |[\tilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}}]_j| \ge \kappa\epsilon\Big) \overset{(b)}{\le} P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{\mathsf{v}} \max_{j\in sec(\ell)} |[\tilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}}]_j| \ge \frac{\kappa\epsilon}{\log M}\Big) \\
&\overset{(c)}{\le} t^4 K K'_{t-1}\Pi'_{t-1}\exp\Big\{-\frac{\kappa\kappa'_{t-1}(\omega/\mathsf{R})^{2\mathsf{v}}\pi'_{t-1}\epsilon^2}{t^8(\log M)^2}\Big\}.
\end{aligned}
\tag{218}
$$

Step (a) follows by Lemma C.5 applied to each section, (b) by $\tau_{\mathsf{c}}^t = \Theta(L/n) = \Theta(1/\log M)$, and (c) by (217). The bound for T_1 in (218) also holds for T_2 of (216) and is shown similarly.

Finally, consider term T_3 of (216). Recalling from Lemma 7.5 that $\tilde{\boldsymbol{h}}_{\mathsf{c}}^s \overset{d}{=} \sqrt{\tau_{\mathsf{c}}^{s-1}}\tilde{\boldsymbol{Z}}_{s-1,\mathsf{c}}$ where

$\tilde{\boldsymbol{Z}}_{s-1,\mathsf{c}} \sim \mathcal{N}(0, \mathbb{I}_{N\mathsf{C}})$. Then, (dropping the c subscript when used inside the $\eta_{\mathsf{c}}^t(\cdot)$ function),

$$T_3 = P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{\mathsf{v}}\Big[(\eta_\ell^{s-1}(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^{s-1}}\tilde{\boldsymbol{Z}}_{s-1}) - \beta_{0,\ell})^*(\eta_\ell^t(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^t}\tilde{\boldsymbol{Z}}_t) - \beta_{0,\ell}) - \psi_{\mathsf{c}}^{t+1}\Big]\Big| \geq \kappa\epsilon\Big),$$

$$\overset{(a)}{\leq} 2\exp\Big\{\frac{-\kappa L^2\epsilon^2}{(\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}} W_{\mathsf{rc}}^{2\mathsf{v}})}\Big\} = 2\exp\Big\{\frac{-\kappa L\epsilon^2}{(\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}^{2\mathsf{v}}/\mathsf{C})}\Big\} \overset{(b)}{\leq} 2\exp\Big\{\frac{-\kappa n(\omega/N_{\mathsf{R}})^{(2\mathsf{v}-1)_+}\epsilon^2}{\log M}\Big\},$$

where (a) is obtained using Hoeffding's inequality (Lemma B.1) and (b) uses $nR = L\log M$. Verifying that the expectation of the random variable in T_3 is ψ_{c}^{t+1} is similar to (135).

(d) Using the representation of $\boldsymbol{h}^{s+1}$ in Lemma 7.5, we have

$$\begin{aligned}(\boldsymbol{h}_{\mathsf{c}}^{\tilde{s}+1})^*\boldsymbol{q}_{\mathsf{c}}^{s+1,\mathsf{r}} &= \sqrt{W_{\mathsf{rc}}}(\boldsymbol{h}_{\mathsf{c}}^{\tilde{s}+1})^*[\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \boldsymbol{h}^{s+1}) - \boldsymbol{\beta}_{0,\mathsf{c}}] \\ &= \sqrt{W_{\mathsf{rc}}}(\tilde{\boldsymbol{h}}_{\mathsf{c}}^{\tilde{s}+1} + \tilde{\boldsymbol{\Delta}}_{\tilde{s}+1,\mathsf{c}})^*\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1} - \tilde{\boldsymbol{\Delta}}_{s+1}) + (\boldsymbol{h}_{\mathsf{c}}^{\tilde{s}+1})^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}.\end{aligned} \tag{219}$$

Using (219) and Lemma B.2,

$$\begin{aligned}&P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big[\frac{(\boldsymbol{h}_{\mathsf{c}}^{\tilde{s}+1})^*\boldsymbol{q}_{\mathsf{c}}^{s+1,\mathsf{r}}}{L} + \frac{\sqrt{W_{\mathsf{rc}}}\,\psi_{\mathsf{c}}^{s+1}\tau_{\mathsf{c}}^{\max(\tilde{s},s)}}{\mathsf{C}\tau_{\mathsf{c}}^s}\Big]\Big| \geq \epsilon\Big) \\ &\leq P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big[\frac{(\tilde{\boldsymbol{h}}_{\mathsf{c}}^{\tilde{s}+1} + \tilde{\boldsymbol{\Delta}}_{\tilde{s}+1,\mathsf{c}})^*\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1} - \tilde{\boldsymbol{\Delta}}_{s+1})}{L} + \frac{\psi_{\mathsf{c}}^{s+1}\tau_{\mathsf{c}}^{\max(\tilde{s},s)}}{\mathsf{C}\tau_{\mathsf{c}}^s}\Big]\Big| \geq \frac{\epsilon}{2}\Big) \\ &\quad + P\Big(\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sqrt{W_{\mathsf{rc}}}\Big[\frac{(\boldsymbol{h}_{\mathsf{c}}^{\tilde{s}+1})^*\boldsymbol{q}_{\mathsf{c}}^{0,\mathsf{r}}}{L}\Big]\Big| \geq \frac{\epsilon}{2}\Big).\end{aligned} \tag{220}$$

The second term on the RHS of (220) has the desired upper bound by $\mathcal{H}_{t+1}(b)$. We now bound the first term of (220), denoted by T_1. Using Lemma B.2,

$$\begin{aligned}T_1 &\leq P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big[(\tilde{\boldsymbol{h}}_{\mathsf{c}}^{\tilde{s}+1})^*\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1}) + \frac{L\psi_{\mathsf{c}}^{s+1}\tau_{\mathsf{c}}^{\max(\tilde{s},s)}}{\mathsf{C}\tau_{\mathsf{c}}^s}\Big]\Big| \geq \epsilon/6\Big) \\ &\quad + P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big|(\tilde{\boldsymbol{\Delta}}_{\tilde{s}+1,\mathsf{c}})^*\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1} - \tilde{\boldsymbol{\Delta}}_{s+1})\Big| \geq \epsilon/6\Big) \\ &\quad + P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]} W_{\mathsf{rc}}\Big|(\tilde{\boldsymbol{h}}_{\mathsf{c}}^{\tilde{s}+1})^*[\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1} - \tilde{\boldsymbol{\Delta}}_{s+1}) - \eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1})]\Big| \geq \epsilon/6\Big).\end{aligned} \tag{221}$$

Label the terms of (221) as T_{1a}, T_{1b}, T_{1c}, and we bound each separately. First consider T_{1a},

$$\begin{aligned}&\sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}\,\mathbb{E}_{\tilde{\boldsymbol{Z}}_{\tilde{s}},\tilde{\boldsymbol{Z}}_s}\Big\{(\tilde{\boldsymbol{Z}}_{\tilde{s},\mathsf{c}})^*\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^s}\tilde{\boldsymbol{Z}}_s)\Big\} \overset{(a)}{=} \sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}\,\mathbb{E}_{\tilde{\boldsymbol{Z}}_{\tilde{s}},\tilde{\boldsymbol{Z}}_s,\boldsymbol{\beta}_0}\Big\{(\tilde{\boldsymbol{Z}}_{\tilde{s},\mathsf{c}})^*\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^s}\tilde{\boldsymbol{Z}}_s)\Big\} \\ &\overset{(b)}{=} \frac{\tau_{\mathsf{c}}^{\max\{\tilde{s},s\}}}{\tau_{\mathsf{c}}^s}\Big[\mathbb{E}_{\tilde{\boldsymbol{Z}}_s,\boldsymbol{\beta}}\Big\{\|\eta_{\mathsf{c}}^s(\boldsymbol{\beta} - \sqrt{\tau_{\mathsf{c}}^s}\tilde{\boldsymbol{Z}}_s)\|^2\Big\} - \frac{L}{\mathsf{C}}\Big] \overset{(c)}{=} \frac{-L\psi_{\mathsf{c}}^{s+1}\tau_{\mathsf{c}}^{\max\{\tilde{s},s\}}}{\mathsf{C}\tau_{\mathsf{c}}^s}.\end{aligned} \tag{222}$$

Step (a) follows for each $\boldsymbol{\beta}_0 \in \mathcal{B}_{M,L}$ because of the uniform distribution of the non-zero entry in each section of $\boldsymbol{\beta}_0$ over the M possible locations and the entry-wise i.i.d. distributions of $\tilde{\boldsymbol{Z}}_{\tilde{s}}$

and $\tilde{\boldsymbol{Z}}_s$, step (b) by Stein's Lemma (see [4, p.1491, Eqs. (102)–(104)] for details), and step (c) from Lemma C.4. Now recall from Lemma 7.5 that $\tilde{\boldsymbol{h}}_{\mathsf{c}}^{s+1} \stackrel{d}{=} \sqrt{\tau_{\mathsf{c}}^{s}}\tilde{\boldsymbol{Z}}_{s,\mathsf{c}}$ where $\tilde{\boldsymbol{Z}}_{s,\mathsf{c}} \sim \mathcal{N}(0, \mathbb{I}_{N_{\mathsf{C}}})$ with $\mathbb{E}\{[\tilde{\boldsymbol{Z}}_{\tilde{s},\mathsf{c}}]_j[\tilde{\boldsymbol{Z}}_{s,\mathsf{c}}]_j\} = \sqrt{\tau_{\mathsf{c}}^{\max\{s,\tilde{s}\}}/\tau_{\mathsf{c}}^{\min\{s,\tilde{s}\}}}$, for $j \in [N_{\mathsf{C}}]$. Using this representation of of $\tilde{\boldsymbol{h}}^{s}$ and (222), the first term in (221) can be bounded using Lemma C.7:

$$T_{1a} = P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}\sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}W_{\mathsf{rc}}\Big[(\tilde{\boldsymbol{Z}}_{\tilde{s},\ell})^*\eta_\ell^s(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^{s}}\tilde{\boldsymbol{Z}}_s) - \mathbb{E}_{\tilde{\boldsymbol{Z}}_{\tilde{s}},\tilde{\boldsymbol{Z}}_s,\boldsymbol{\beta}_0}\{(\tilde{\boldsymbol{Z}}_{\tilde{s},\ell})^*\eta_\ell^s(\boldsymbol{\beta}_0 - \sqrt{\tau_{\mathsf{c}}^{s}}\tilde{\boldsymbol{Z}}_s)\}\Big]\Big| \geq \epsilon/6\Big)$$
$$\leq \exp\{-\kappa_1 L(\omega/\mathsf{R})\epsilon^2\} = \exp\{-\kappa N_{\mathsf{R}}\omega\epsilon^2/\log M\}.$$

For T_{1b}, we recall that $\sum_{j\in sec(\ell)}\eta_j^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1} - \tilde{\boldsymbol{\Delta}}_{s+1}) = 1$ and use (217) with $\mathsf{v} = 1$ to obtain

$$T_{1b} \leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}\Big|(\tilde{\boldsymbol{\Delta}}_{\tilde{s}+1,\ell})^*\eta_\ell^s(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{h}}^{s+1} - \tilde{\boldsymbol{\Delta}}_{s+1})\Big| \geq \epsilon/6\Big)$$
$$\leq P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}\max_{j\in sec(\ell)}\Big|[\tilde{\boldsymbol{\Delta}}_{\tilde{s}+1,\mathsf{c}}]_j\Big| \geq \kappa\epsilon\Big) \leq t^4KK'_{t-1}\Pi'_{t-1}\exp\{-\frac{1}{t^8}\kappa\kappa'_{t-1}(\omega/\mathsf{R})^2\pi'_{t-1}\epsilon^2\}.$$

Finally for T_{1c}, using Lemma 7.5 Eq. (84) to write $\tilde{\boldsymbol{h}}_{\mathsf{c}}^{\tilde{s}+1} = \sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}\tilde{\boldsymbol{Z}}_{\tilde{s},\mathsf{c}}$, we can prove the bound as in the $\mathcal{H}_1(d)$ step in (142) - (143), appealing to Lemma C.5, Lemma B.9, and (217).
(e) Using the conditional distribution of $\boldsymbol{h}^{t+1}$ from Lemma 7.5 Eq. (83) and Eq. (84) we write, $\boldsymbol{h}_{\mathsf{c}}^{t+1} \stackrel{d}{=} \tilde{\boldsymbol{h}}_{\mathsf{c}}^{t+1} + \tilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}} \stackrel{d}{=} \sqrt{\tau_{\mathsf{c}}^{t}}\tilde{\boldsymbol{Z}}_{t,\mathsf{c}} + \tilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}}$. Then we have as in the $\mathcal{H}_1(e)$ proof, namely (144),

$$P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\boldsymbol{h}_{\mathsf{c}}^{t+1}]_j)^2 \geq 6(\max_{\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}})\tau_{\mathsf{c}}^{t}\log M + \epsilon\Big) \leq$$
$$P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\tilde{\boldsymbol{Z}}_{t,\mathsf{c}}]_j)^2 \geq 3(\max_{\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}})\log M\Big) + P\Big(\frac{1}{L}\sum_{\mathsf{c}\in[\mathsf{C}]}\sum_{\ell\in\mathsf{c}}W_{\mathsf{rc}}^{2\mathsf{v}}\max_{j\in sec(\ell)}([\tilde{\boldsymbol{\Delta}}_{t+1,\mathsf{c}}]_j)^2 \geq \frac{\epsilon}{2}\Big).$$

Then the desired bound follows by Lemma B.9 and (217).

(f) The proof of (106) follows similarly to the corresponding $\mathcal{B}_t(f)$ proof of (97) using inductive hypothesis $\mathcal{H}_t(g)$ result (109) and $\mathcal{H}_{t+1}(c)$. The proof of (107) follows similarly to the corresponding $\mathcal{B}_t(f)$ proof of (98) using $\mathcal{H}_{t+1}(c)$ and $\mathcal{H}_{t+1}(f)$ result (106).

(g) The proof of (108) follows similarly to the corresponding $\mathcal{B}_t(g)$ proof of (99) using Fact C.1 and inductive hypotheses $\mathcal{H}_1(g) - \mathcal{H}_{t+1}(g)$. Result (109) follows similarly to the corresponding $\mathcal{B}_t(g)$ proof of (97): using block inversion we can represent $(\widetilde{\boldsymbol{C}}^{t+2,\mathsf{r}})^{-1}$ and $(\mathbb{Q}_{t+2}^{\mathsf{r}})^{-1}$. From the representation of $(\widetilde{\boldsymbol{C}}^{t+2,\mathsf{r}})^{-1}$, we can see that each non-zero element is $\Theta(1)$, and then we can show the element-wise concentration with inductive hypothesis $\mathcal{H}_t(g)$ and $\mathcal{H}_{t+1}(f)$.

A Proofs of bounds on state evolution parameters

A.1 Proof of Lemma 4.1

Recall that $\psi_{\mathsf{c}}^{t+1} = 1 - \mathcal{E}(\tau_{\mathsf{c}}^t)$ where $\mathcal{E}(\tau_{\mathsf{c}}^t)$ is defined in (13). It is shown in [41, Appendix A] that for sufficiently large M and any $\delta \in (0, \frac{1}{2})$,

$$\mathcal{E}(\tau_{\mathsf{c}}^t) \geq \left(1 - \frac{M^{-k\delta^2}}{\delta\sqrt{\log M}}\right) \mathbb{I}\{\nu_{\mathsf{c}}^t > 2 + \delta\} \qquad \mathsf{c} \in [\mathsf{C}]. \tag{223}$$

The upper bound on ψ_{c}^t directly follows from the lower bound in (223).

We now lower bound ψ_{c}^t by obtaining an upper bound for $\mathcal{E}(\tau_{\mathsf{c}}^t)$. We will use the following concentration inequality the maximum of M i.i.d. standard Gaussian random variables $U_1, \ldots, U_M$. For any $\epsilon \in (0, 1)$,

$$P\Big(\max_{1\leq j\leq M} U_j < \sqrt{2\ln M}(1-\epsilon)\Big) \leq \exp\Big(\frac{-\kappa M^{\epsilon(2-\epsilon)}}{\sqrt{\ln M}}\Big), \tag{224}$$

where $\kappa > 0$ is a universal constant.

Consider a column block c such that $\nu_{\mathsf{c}}^t \leq 2 - \widetilde{\delta}$. Recall from (21) that $\nu_{\mathsf{c}}^t = (\tau_{\mathsf{c}}^t \ln M)^{-1}$ and due to the assumption on the base matrix, $\kappa_1 \leq \nu_{\mathsf{c}}^t \leq \kappa_2$ for some strictly positive constants κ_1, κ_2. Then, with positive constants $a, a' \in (0, 1)$ to be specified later, we have

$$\begin{aligned}
\mathcal{E}(\tau_{\mathsf{c}}^t) &= \mathbb{E}\left[\frac{1}{1 + e^{-\nu_{\mathsf{c}}^t \ln M} \sum_{j=2}^M e^{U_j\sqrt{\nu_{\mathsf{c}}^t \ln M}} \cdot e^{-U_1\sqrt{\nu_{\mathsf{c}}^t \ln M}}}\right] \\
&\leq P\left(\max_{2\leq j\leq M} U_j < \sqrt{2\ln M}(1 - a\widetilde{\delta})\right) \cdot 1 \\
&\quad + P\left(\max_{2\leq j\leq M} U_j \geq \sqrt{2\ln M}(1 - a\widetilde{\delta})\right) \mathbb{E}\left(1 + \exp\left(-\sqrt{\nu_{\mathsf{c}}^t \ln M}\, U_1\right) M^{-\nu_{\mathsf{c}}^t} \exp(\sqrt{2\nu_{\mathsf{c}}^t}(1 - a\widetilde{\delta}) \ln M)\right)^{-1} \\
&\overset{(a)}{\leq} \exp\left(\frac{-\kappa M^{a\widetilde{\delta}(2-a\widetilde{\delta})}}{\sqrt{\ln M}}\right) + \mathbb{E}\left(1 + \exp\left(-\sqrt{\nu_{\mathsf{c}}^t \ln M}\, U_1\right) M^{\sqrt{2\nu_{\mathsf{c}}^t}(1-a\widetilde{\delta}) - \nu_{\mathsf{c}}^t}\right)^{-1} \\
&\overset{(b)}{\leq} \exp\left(\frac{-\kappa M^{a\widetilde{\delta}(2-a\widetilde{\delta})}}{\sqrt{\ln M}}\right) + P\left(U_1 \leq \sqrt{2\ln M} a'\widetilde{\delta}\right) \cdot \frac{1}{1 + M^{\sqrt{2\nu_{\mathsf{c}}^t}(1-(a+a')\widetilde{\delta}) - \nu_{\mathsf{c}}^t}} + P\left(U_1 > \sqrt{2\ln M} a'\widetilde{\delta}\right) \\
&\overset{(c)}{\leq} \exp\left(\frac{-\kappa M^{a\widetilde{\delta}(2-a\widetilde{\delta})}}{\sqrt{\ln M}}\right) + M^{-\left(\sqrt{2(2-\widetilde{\delta})}(1-(a+a')\widetilde{\delta}) - 2 + \widetilde{\delta}\right)} + M^{-(a')^2\widetilde{\delta}^2}.
\end{aligned} \tag{225}$$

In the above sequence of inequalities, step (a) uses (224); step (b) holds because

$$\exp\left(-\sqrt{\nu_{\mathsf{c}}^t \ln M}\, U_1\right) \geq M^{-\sqrt{2\nu_{\mathsf{c}}^t} a'\widetilde{\delta}} \quad \text{when} \quad U_1 \leq \sqrt{2\ln M} a'\widetilde{\delta}.$$

In step (c), the last term is obtained using a standard Gaussian tail bound for U_1 (Lemma B.8); the second term is obtained by noting that $\sqrt{2\nu_{\mathsf{c}}^t}(1 - (a + a')\widetilde{\delta}) - \nu_{\mathsf{c}}^t$ is a concave function of ν_{c}^t which (for a, a' specified below) takes its minimum value in $\nu_{\mathsf{c}}^t \in [\kappa_1, 2 - \widetilde{\delta}]$ at the endpoint $\nu_{\mathsf{c}}^t = (2 - \widetilde{\delta})$.

Using the inequality $(1-x)^{1/2} \geq 1 - \frac{5x}{8}$ for $x \in [0, 1/2]$, we have the following lower bound for the exponent of the second term in (225):

$$\begin{aligned}\sqrt{2(2-\widetilde{\delta})}(1-(a+a')\widetilde{\delta}) - 2 + \widetilde{\delta} &\geq 2\Big(1 - \frac{5\widetilde{\delta}}{16}\Big)\Big(1-(a+a')\widetilde{\delta}\Big) - 2 + \widetilde{\delta} \\ &= \widetilde{\delta}\Big(\frac{3}{8} - 2(a+a')\Big) + \frac{5}{8}(a+a')\widetilde{\delta}^2.\end{aligned}$$

Taking $a = 1/64$ and $a' = 5/32$, we get the desired upper bound.

A.2 Proof of Proposition 4.1

Using the definition in (6), the state evolution equations (10)–(12) for the (ω, Λ, ρ) base matrix are as follows. With $\psi_{\mathsf{c}}^0 = 1$ for $\mathsf{c} \in [\mathsf{C}]$, for $t \geq 0$:

$$\phi_{\mathsf{r}}^t = \sigma^2\left[1 + \vartheta\,\mathsf{snr}\left(\frac{1-\rho}{\omega}\sum_{\mathsf{c}=\underline{\mathsf{c}}_{\mathsf{r}}}^{\bar{\mathsf{c}}_{\mathsf{r}}}\psi_{\mathsf{c}}^t + \frac{\rho}{\Lambda-1}\sum_{\mathsf{c}\in[\Lambda]\setminus\{\underline{\mathsf{c}}_{\mathsf{r}},\ldots,\bar{\mathsf{c}}_{\mathsf{r}}\}}\psi_{\mathsf{c}}^t\right)\right], \quad \mathsf{r} \in [\Lambda+\omega-1], \tag{226}$$

$$\nu_{\mathsf{c}}^t = \frac{1}{\tau_{\mathsf{c}}^t \ln M} = \frac{1}{R}\left[\frac{(1-\rho)\cdot P}{\omega}\sum_{\mathsf{r}=\mathsf{c}}^{\mathsf{c}+\omega-1}\frac{1}{\phi_{\mathsf{r}}^t} + \frac{\rho\cdot P}{\Lambda-1}\sum_{\mathsf{r}\in[\Lambda+\omega-1]\setminus\{\mathsf{c},\ldots,\mathsf{c}+\omega-1\}}\frac{1}{\phi_{\mathsf{r}}^t}\right], \tag{227}$$

$$\psi_{\mathsf{c}}^{t+1} = 1 - \mathcal{E}(\tau_{\mathsf{c}}^t). \tag{228}$$

Here $\mathcal{E}(\tau_{\mathsf{c}}^t)$ is defined in (13), and $\underline{\mathsf{c}}_{\mathsf{r}}, \bar{\mathsf{c}}_{\mathsf{r}}$ are defined (27).

Since the variables ψ_{c}^t for $\mathsf{c} \in [\Lambda]$ and $t \geq 0$ are symmetric about the center column index, i.e. $\psi_{\mathsf{c}}^t = \psi_{\Lambda-\mathsf{c}+1}^t$ for $\mathsf{c} \leq \lceil \frac{\Lambda}{2} \rceil$, we carry out the analysis for $\mathsf{c} \leq \lceil \frac{\Lambda}{2} \rceil$; the result for the other half then holds by symmetry. We will upper bound ψ_{c}^t using Lemma 4.1. Using (22), for the first iteration we will have $\psi_{\mathsf{c}}^1 \leq f_{M,\delta}$ for indices c for which $\{\nu_{\mathsf{c}}^0 > 2+\delta\}$. Letting $F_{\mathsf{c}}^0 := \nu_{\mathsf{c}}^0 R$, this corresponds to finding indices c for which $F_{\mathsf{c}}^0 > (2+\delta)R$. We now obtain a lower bound on F_{c}^0 for indices $\mathsf{c} < \omega$.

Using (227) we have

$$\begin{aligned}
F_{\mathsf{c}}^0 &= \frac{(1-\rho)P}{\omega}\sum_{\mathsf{r}=\mathsf{c}}^{\mathsf{c}+\omega-1}\frac{1}{\phi_{\mathsf{r}}^0} + \frac{\rho P}{\Lambda-1}\sum_{\mathsf{r}\in[\Lambda+\omega-1]\setminus\{\mathsf{c},\ldots,\mathsf{c}+\omega-1\}}\frac{1}{\phi_{\mathsf{r}}^0} \qquad (229)\\
&\geq \frac{(1-\rho)P}{\omega}\sum_{\mathsf{r}=\mathsf{c}}^{\mathsf{c}+\omega-1}\frac{1}{\phi_{\mathsf{r}}^0} \\
&\overset{\text{(i)}}{=} \frac{(1-\rho)\,\mathsf{snr}}{\omega}\left(\sum_{\mathsf{r}=\mathsf{c}}^{\omega-1}\frac{1}{1+\frac{(1-\rho)\,\vartheta\,\mathsf{snr}}{\omega}\mathsf{r} + \frac{\rho\,\vartheta\,\mathsf{snr}}{\Lambda-1}(\Lambda-\mathsf{r})} + \frac{\mathsf{c}}{1+(1-\rho)\,\vartheta\,\mathsf{snr}+\rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\omega}{\Lambda-1}}\right) \\
&\geq \frac{(1-\rho)\,\mathsf{snr}}{\omega}\sum_{\mathsf{r}=\mathsf{c}}^{\omega-1}\frac{1}{1+\rho\,\vartheta\,\mathsf{snr}+(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{r}}{\omega}} + \frac{\mathsf{c}}{\omega}\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\overset{\text{(ii)}}{\geq} \frac{1}{\vartheta}\left(\ln(1+\vartheta\,\mathsf{snr}) - \ln\left(1+\rho\,\vartheta\,\mathsf{snr}+(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{c}}{\omega}\right)\right) + \frac{\mathsf{c}}{\omega}\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\overset{\text{(iii)}}{\geq} \frac{1}{\vartheta}\ln(1+\vartheta\,\mathsf{snr}) - \rho\,\mathsf{snr} - (1-\rho)\,\mathsf{snr}\,\frac{\mathsf{c}}{\omega} + \frac{\mathsf{c}}{\omega}\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\geq \frac{1}{\vartheta}\ln(1+\vartheta\,\mathsf{snr}) - \rho\,\mathsf{snr} - \frac{\mathsf{c}}{\omega}\frac{\vartheta\,\mathsf{snr}^2}{1+\vartheta\,\mathsf{snr}},
\end{aligned}$$

where the labelled steps are obtained as follows: (i) using the expression for ϕ_{r}^t in (226) and the fact that $\mathsf{c} < \omega$, (ii) using a definite integral to lower bound the left Riemann sum of a positive decreasing function:

$$\frac{1}{\omega}\sum_{\mathsf{r}=\mathsf{c}}^{\omega-1}\frac{1}{\frac{1+\rho\,\vartheta\,\mathsf{snr}}{(1-\rho)\,\vartheta\,\mathsf{snr}}+\frac{\mathsf{r}}{\omega}} \geq \int_{\mathsf{c}/\omega}^{1}\frac{1}{\frac{1+\rho\,\vartheta\,\mathsf{snr}}{(1-\rho)\,\vartheta\,\mathsf{snr}}+x}\,\mathrm{d}x. \tag{230}$$

Inequality (iii) is obtained using $\ln(1+x) \leq x$.

Therefore, the condition $F_{\mathsf{c}}^0 > (2+\delta)R$ will be satisfied if

$$\frac{1}{\vartheta}\ln(1+\vartheta\,\mathsf{snr}) - \rho\,\mathsf{snr} - \frac{\mathsf{c}}{\omega}\frac{\vartheta\,\mathsf{snr}^2}{1+\vartheta\,\mathsf{snr}} \;>\; (2+\delta)R. \tag{231}$$

Rearranging (231) gives

$$\frac{\mathsf{c}}{\omega} \;<\; 2\left(1+\frac{1}{\vartheta\,\mathsf{snr}}\right)\frac{1}{\mathsf{snr}}\left[\frac{1}{2\vartheta}\ln\left(1+\vartheta\,\mathsf{snr}\right) - \frac{\rho\,\mathsf{snr}}{2} - R - \frac{\delta R}{2}\right]. \tag{232}$$

Note that the RHS of (232) is smaller than or equal to 1 if $R \geq \frac{(1-\rho)\mathsf{snr}}{(2+\delta)(1+\vartheta\,\mathsf{snr})}$. Using $\rho \leq \frac{\Delta}{3\mathsf{snr}}$ and $\delta < \frac{\Delta}{2R}$, the sufficient condition in (232) for $\psi_{\mathsf{c}}^1 \leq f_{M,\delta}$ can be weakened to $\mathsf{c} \leq g$ where g is defined in (34). Note that the condition (33) on ω guarantees that $g > 1$.

Notice from (229) that F_{c}^0 is decreasing in c for $\mathsf{c} \in [1,\omega]$ and is then constant for $\mathsf{c} \in [\omega, \lceil\frac{\Lambda}{2}\rceil]$. Therefore, for any $\delta \in (0,1)$, if $F_{\mathsf{c}}^0 > (2+\delta)R$ is satisfied for $\mathsf{c} = \omega$ then $\psi_{\mathsf{c}}^1 \leq f_{M,\delta}$ for all $\mathsf{c} \in [\Lambda]$. By using a similar analysis as above for lower bounding F_{c}^0, one can show that a sufficient condition for $F_{\omega}^0 > (2+\delta)R$ is $R < \frac{(1-\rho)\mathsf{snr}}{(2+\delta)(1+\vartheta\,\mathsf{snr})}$.

Next we consider subsequent iterations $t > 1$. Assume towards induction that

$$\psi_{\mathsf{c}}^t \leq f_{M,\delta}, \quad \text{for } \mathsf{c} \leq g_t, \tag{233}$$

where $g_t \geq tg$. We will prove that (233) implies $\psi_{\mathsf{c}}^{t+1} \leq f_{M,\delta}$ for $\mathsf{c} \leq g_t + g$. We prove the result for $g_t \geq \omega$, with the other case being similar. We wish to find column indices $\mathsf{c} \in (g_t, g_t+\omega)$ for which $\psi_{\mathsf{c}}^{t+1} \leq f_{M,\delta}$, or equivalently $F_{\mathsf{c}}^t > (2+\delta)R$. For brevity, we will use the shorthand $f := f_{M,\delta}$. Using the induction assumption (233) in (226), we deduce

$$\frac{\phi_{\mathsf{r}}^t}{\sigma^2} \leq \begin{cases} 1 + f\,(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{r}}{\omega} + f\,\rho\,\vartheta\,\mathsf{snr}\,\frac{g_t-\mathsf{r}}{\Lambda-1} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-g_t}{\Lambda-1}, & 1 \leq \mathsf{r} \leq \omega, \\ 1 + f\,(1-\rho)\,\vartheta\,\mathsf{snr} + f\,\rho\,\vartheta\,\mathsf{snr}\,\frac{g_t-\omega}{\Lambda-1} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-g_t}{\Lambda-1}, & \omega \leq \mathsf{r} \leq g_t, \\ 1 + \frac{(1-\rho)\,\vartheta\,\mathsf{snr}}{\omega}\left[f(\omega-(\mathsf{r}-g_t)) + (\mathsf{r}-g_t)\right] + f\,\rho\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{r}-\omega}{\Lambda-1} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\mathsf{r}}{\Lambda-1}, & g_t \leq \mathsf{r} < g_t+\omega, \\ 1 + (1-\rho)\,\vartheta\,\mathsf{snr} + f\,\rho\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{r}-\omega}{\Lambda-1} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\mathsf{r}}{\Lambda-1}, & \mathsf{r} \geq g_t+\omega. \end{cases} \tag{234}$$

For M sufficiently large (i.e., f sufficiently small), noting that $g_t \geq \omega$ we can simplify (234) to

$$\frac{\phi_{\mathsf{r}}^t}{\sigma^2} \leq \begin{cases} 1 + f\,(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{r}}{\omega} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\mathsf{r}}{\Lambda-1}, & 1 \leq \mathsf{r} \leq \omega, \\ 1 + f\,(1-\rho)\,\vartheta\,\mathsf{snr} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\omega}{\Lambda-1}, & \omega \leq \mathsf{r} \leq g_t, \\ 1 + \frac{(1-\rho)\,\vartheta\,\mathsf{snr}}{\omega}\left[f(\omega-(\mathsf{r}-g_t)) + (\mathsf{r}-g_t)\right] + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\omega}{\Lambda-1}, & g_t \leq \mathsf{r} < g_t+\omega, \\ 1 + (1-\rho)\,\vartheta\,\mathsf{snr} + \rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\omega}{\Lambda-1}, & \mathsf{r} \geq g_t+\omega. \end{cases} \tag{235}$$

We now obtain a lower bound on $F_{\mathsf{c}}^t := \nu_{\mathsf{c}}^t R$ for $g_t < \mathsf{c} < g_t + \omega$. Using (227) we have

$$
\begin{aligned}
F_{\mathsf{c}}^t &\geq \frac{(1-\rho)P}{\omega} \sum_{\mathsf{r}=\mathsf{c}}^{\mathsf{c}+\omega-1} \frac{1}{\phi_{\mathsf{r}}^t} \\
&\overset{\text{(i)}}{\geq} \frac{(1-\rho)\,\mathsf{snr}}{\omega} \sum_{\mathsf{r}=\mathsf{c}}^{g_t+\omega-1} \frac{1}{1+\frac{(1-\rho)\,\vartheta\,\mathsf{snr}}{\omega}\left[f(\omega-(r-g_t))+(r-g_t)\right]+\frac{\rho\,\vartheta\,\mathsf{snr}}{\Lambda-1}(\Lambda-\omega)} \\
&\qquad\qquad + \frac{\mathsf{c}-g_t}{\omega}\,\frac{(1-\rho)\,\mathsf{snr}}{1+(1-\rho)\,\vartheta\,\mathsf{snr}+\rho\,\vartheta\,\mathsf{snr}\,\frac{\Lambda-\omega}{\Lambda-1}} \\
&\geq \frac{(1-\rho)\,\mathsf{snr}}{\omega} \sum_{\mathsf{r}=\mathsf{c}}^{g_t+\omega-1} \frac{1}{1+\rho\,\vartheta\,\mathsf{snr}+f\,(1-\rho)\,\vartheta\,\mathsf{snr}+(1-f)(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{(\mathsf{r}-g_t)}{\omega}} + \frac{\mathsf{c}-g_t}{\omega}\,\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\geq \frac{(1-\rho)\,\mathsf{snr}}{\omega} \sum_{\mathsf{r}'=\mathsf{c}-g^t}^{\omega-1} \frac{1}{1+\rho\,\vartheta\,\mathsf{snr}+f\,(1-\rho)\,\vartheta\,\mathsf{snr}+(1-f)(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{r}'}{\omega}} + \frac{\mathsf{c}-g_t}{\omega}\,\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\overset{\text{(ii)}}{\geq} \frac{1}{1-f}\,\frac{1}{\vartheta}\left(\ln(1+\vartheta\,\mathsf{snr})-\ln\left(1+\rho\,\vartheta\,\mathsf{snr}+f\,(1-\rho)\,\vartheta\,\mathsf{snr}+(1-f)(1-\rho)\,\vartheta\,\mathsf{snr}\,\frac{\mathsf{c}-g_t}{\omega}\right)\right) \\
&\qquad + \frac{\mathsf{c}-g_t}{\omega}\,\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\overset{\text{(iii)}}{\geq} \frac{1}{\vartheta}\ln(1+\vartheta\,\mathsf{snr})-\rho\,\mathsf{snr}-f\,(1-\rho)\,\mathsf{snr}-(1-\rho)\,\mathsf{snr}\,\frac{\mathsf{c}-g_t}{\omega}+\frac{\mathsf{c}-g_t}{\omega}\,\frac{(1-\rho)\,\mathsf{snr}}{1+\vartheta\,\mathsf{snr}} \\
&\geq \frac{1}{\vartheta}\ln(1+\vartheta\,\mathsf{snr})-\rho\,\mathsf{snr}-f\,\mathsf{snr}-\frac{\mathsf{c}-g_t}{\omega}\,\frac{\vartheta\,\mathsf{snr}^2}{1+\vartheta\,\mathsf{snr}},
\end{aligned}
\tag{236}
$$

where the labelled steps are obtained as follows: (i) using the bounds for ϕ_{r}^t given in (235), (ii) using a definite integral to lower bound the left Riemann sum of a decreasing function, similar to (230), and (iii) using the inequalities $\ln(1+x) \leq x$ and $\frac{1}{1-f} \geq 1$.

Recall from Lemma 4.1 that $\psi_{\mathsf{c}}^{t+1} \leq f_{M,\delta}$ if $F_{\mathsf{c}}^t > (2+\delta)R$. From (236), this condition will be satisfied if

$$
\frac{1}{\vartheta}\ln(1+\vartheta\,\mathsf{snr})-\rho\,\mathsf{snr}-f\,\mathsf{snr}-\frac{\mathsf{c}-g_t}{\omega}\,\frac{\vartheta\,\mathsf{snr}^2}{1+\vartheta\,\mathsf{snr}} \;>\; (2+\delta)R. \tag{237}
$$

Rearranging (237) gives

$$
\frac{\mathsf{c}-g_t}{\omega} < 2\left(1+\frac{1}{\vartheta\,\mathsf{snr}}\right)\frac{1}{\mathsf{snr}}\left[\frac{1}{2\vartheta}\ln\left(1+\vartheta\,\mathsf{snr}\right)-\frac{\rho\,\mathsf{snr}}{2}-\frac{f\,\mathsf{snr}}{2}-R-\frac{\delta R}{2}\right]. \tag{238}
$$

Note that the RHS of (238) is smaller than or equal to 1 if $R \geq \frac{(1-\rho)\mathsf{snr}}{(2+\delta)(1+\vartheta\,\mathsf{snr})}$. Using $\rho \leq \frac{\Delta}{3\mathsf{snr}}$ and $\delta < \frac{\Delta}{2R}$ in (238), we obtain that a sufficient condition for $\psi_{\mathsf{c}}^{t+1} \leq f_{M,\delta}$ is

$$
\frac{\mathsf{c}-g_t}{\omega} < 2\left(1+\frac{1}{\vartheta\,\mathsf{snr}}\right)\frac{1}{\mathsf{snr}}\left[\frac{7\Delta}{12}-\frac{f\mathsf{snr}}{2}\right] \tag{239}
$$

For M sufficiently large, $f < \Delta/(6\mathsf{snr})$. Thus we conclude from (239) that $\psi_{\mathsf{c}}^{t+1} \leq f_{M,\delta}$ for $\mathsf{c}-g_t \leq g$, and hence for $\mathsf{c} \leq (t+1)g$ (since $g_t \geq tg$).

A.3 Proof of Lemma 7.1

We will prove the lemma by showing that the following inequalities hold $1 \le t < T$:

$$\sigma_{\mathsf{r}}^{t-1} - \sigma_{\mathsf{r}}^{t} \ge C_1 \left(\frac{\omega}{\Lambda}\right), \tag{240}$$

$$\nu_{\mathsf{c}}^{t} - \nu_{\mathsf{c}}^{t-1} \ge C_2 \left(\frac{\omega}{\Lambda}\right), \tag{241}$$

where we recall that $\nu_{\mathsf{c}}^t = \frac{1}{\tau_{\mathsf{c}}^t \ln M}$. Here the constant C_1 is defined in (246) below, and $C_2 = PC_1/(R\sigma^4)$. From these inequalities it follows that

$$\sigma_{\perp,\mathsf{r}}^{t} = \frac{\sigma_{\mathsf{r}}^{t}}{\sigma_{\mathsf{r}}^{t-1}} \left(\sigma_{\mathsf{r}}^{t-1} - \sigma_{\mathsf{r}}^{t}\right) \ge \frac{C_1^2}{\vartheta P} \left(\frac{\omega}{\Lambda}\right)^2, \tag{242}$$

$$\frac{n}{L}\tau_{\perp,\mathsf{c}}^{t} = \frac{(\nu_{\mathsf{c}}^{t-1} - \nu_{\mathsf{c}}^{t})}{R(\nu_{\mathsf{c}}^{t})^2} \ge \frac{RC_2}{\mathsf{snr}^2} \left(\frac{\omega}{\Lambda}\right). \tag{243}$$

The inequality in (242) is obtained using $\sigma_{\mathsf{r}}^t \ge C_1\omega/\Lambda$ and the fact that $\sigma_{\mathsf{r}}^{t-1} \le \vartheta P$ (from (226) since $\psi_{\mathsf{c}}^{t-1} \le 1$). The inequality in (243) is obtained by using $\nu_{\mathsf{c}}^t \ge \mathsf{snr}/R$ (from (227) since $\phi_{\mathsf{r}}^t \ge \sigma^2$).

We will prove (240) via induction. The result (241) can then be obtained from (240) as follows. Using (227), we write

$$\nu_{\mathsf{c}}^{t} - \nu_{\mathsf{c}}^{t-1} = \frac{1}{R}\left[\frac{(1-\rho)\cdot P}{\omega}\sum_{\mathsf{r}=\mathsf{c}}^{\mathsf{c}+\omega-1}\left(\frac{1}{\phi_{\mathsf{r}}^t} - \frac{1}{\phi_{\mathsf{r}}^{t-1}}\right) + \frac{\rho\cdot P}{\Lambda-1}\sum_{\mathsf{r}\in[\Lambda+\omega-1]\setminus\{\mathsf{c},\ldots,\mathsf{c}+\omega-1\}}\left(\frac{1}{\phi_{\mathsf{r}}^t} - \frac{1}{\phi_{\mathsf{r}}^{t-1}}\right)\right].$$

Using $\phi_{\mathsf{r}}^{t-1} \ge \phi_{\mathsf{r}}^t \ge \sigma^2$ and $(\phi_{\mathsf{r}}^{t-1} - \phi_{\mathsf{r}}^t) = (\sigma_{\mathsf{r}}^{t-1} - \sigma_{\mathsf{r}}^t) \ge C_1\omega/\Lambda$ by the induction assumption, the above equation yields

$$\nu_{\mathsf{c}}^{t} - \nu_{\mathsf{c}}^{t-1} \ge \frac{PC_1}{R\sigma^4}\frac{\omega}{\Lambda} = \frac{C_2\omega}{\Lambda}. \tag{244}$$

Next we show the lower bound for $\sigma_{\mathsf{r}}^t - \sigma_{\mathsf{r}}^{t+1}$ in (240). For $t = 1$, noting that $\psi_{\mathsf{c}} = 1$ for $\mathsf{c} \in [\mathsf{C}]$, we have

$$\sigma_{\mathsf{c}}^1 - \sigma_{\mathsf{c}}^0 = \frac{1}{\mathsf{C}}\sum_{c=1}^{\mathsf{C}} W_{\mathsf{rc}}(1 - \psi_{\mathsf{c}}^1). \tag{245}$$

From Proposition 4.1, we know that $\psi_{\mathsf{c}}^1 \le f_{M,\delta}$, for $1 \le \mathsf{c} \le g$ and $\Lambda - g + 1 \le \mathsf{c} \le \Lambda$. Since $W_{\mathsf{rc}} \ge \rho P \cdot \frac{\Lambda+\omega-1}{\Lambda-1}$ and $\mathsf{C} = \Lambda$, we have

$$\sigma_{\mathsf{c}}^1 - \sigma_{\mathsf{c}}^0 \ge \frac{1}{\Lambda} 2g\rho P \frac{\Lambda+\omega-1}{\Lambda-1}(1 - f_{M,\delta}) \ge \underbrace{\frac{\rho P(1+\vartheta\,\mathsf{snr})\Delta}{\vartheta\,\mathsf{snr}^2}}_{C_1} \frac{\omega}{\Lambda-1}, \tag{246}$$

where the last inequality holds for M sufficiently large. Assume towards induction that (240) holds for $\sigma_{\mathsf{r}}^{t-1} - \sigma_{\mathsf{r}}^t$, for $1 \le t \le (T-2)$. From (226) we have

$$\sigma_{\mathsf{r}}^{t} - \sigma_{\mathsf{r}}^{t+1} = \vartheta\, P\left(\frac{1-\rho}{\omega}\sum_{\mathsf{c}=\underline{\mathsf{c}}_{\mathsf{r}}}^{\bar{\mathsf{c}}_{\mathsf{r}}}(\psi_{\mathsf{c}}^t - \psi_{\mathsf{c}}^{t+1}) + \frac{\rho}{\Lambda-1}\sum_{\mathsf{c}\in[\Lambda]\setminus\{\underline{\mathsf{c}}_{\mathsf{r}},\ldots,\bar{\mathsf{c}}_{\mathsf{r}}\}}(\psi_{\mathsf{c}}^t - \psi_{\mathsf{c}}^{t+1})\right). \tag{247}$$

For $\delta, \widetilde{\delta} \in (0, \min\{\frac{1}{2}, \frac{\Delta}{2R}\})$, define the set

$$\mathcal{I}_t := \Big\{ \mathsf{c} \in [\Lambda] : \ \nu_{\mathsf{c}}^{t-1} \le 2 - \widetilde{\delta} \text{ and } \nu_{\mathsf{c}}^{t} > 2 + \delta \Big\} \tag{248}$$

For each $\mathsf{c} \in \mathcal{I}_t$, from Lemma 4.1 and Proposition 4.1 we have

$$\psi_{\mathsf{c}}^t \ge 1 - M^{-k_1 \tilde{\delta}^2}, \qquad \psi_{\mathsf{c}}^{t+1} \le \frac{M^{-k\delta^2}}{\delta\sqrt{\log M}}, \qquad \mathsf{c} \in \mathcal{I}_t. \tag{249}$$

We prove below that for sufficiently small choices of $\delta, \widetilde{\delta}$, we have $|\mathcal{I}_t| \ge 2(g-2)$, where g is defined in (34). Using this and (249) in (247), we conclude that

$$\begin{aligned} \sigma_{\mathsf{r}}^t - \sigma_{\mathsf{r}}^{t+1} &\ge \frac{\vartheta P \rho}{\Lambda - 1} 2(g-2) \left(1 - M^{-k_1 \tilde{\delta}^2} - \frac{M^{-k\delta^2}}{\delta\sqrt{\log M}} \right) \\ &\overset{(a)}{\ge} \frac{P \rho \, g}{\Lambda - 1} = \underbrace{\frac{P\rho(1 + \vartheta \, \mathsf{snr})\Delta}{\vartheta \, \mathsf{snr}^2}}_{C_1} \cdot \frac{\omega}{\Lambda - 1}, \end{aligned} \tag{250}$$

where the inequality (a) holds for M sufficiently large. It remains to show that $|\mathcal{I}_t| \ge 2(g-2)$ for suitably chosen $\delta, \widetilde{\delta}$. Let c_* denote the largest index $\mathsf{c} \le \Lambda/2$ such that $\nu_{\mathsf{c}}^{t-1} > 2$. That is,

$$\nu_{\mathsf{c}}^{t-1} > 2 \ \text{ for } \ \mathsf{c} \le \mathsf{c}_*, \qquad \nu_{\mathsf{c}_*+1}^{t-1} \le 2, \qquad \text{and } \ \nu_{\mathsf{c}}^{t-1} < 2 \ \text{ for } \ \mathsf{c}_* + 2 \le \mathsf{c} \le \lfloor \Lambda/2 \rfloor. \tag{251}$$

Let $\delta = \min\{\frac{(\nu_{\mathsf{c}_*}^{t-1} - 2)}{2}, \frac{\Delta}{2R}, \frac{1}{2}\}$ and $\widetilde{\delta} = \min\{\frac{(2 - \nu_{\mathsf{c}_*+2}^{t-1})}{2}, \frac{1}{3}\}$. For these choices, Lemma 4.1 guarantees that

$$\psi_{\mathsf{c}}^t < f_{M,\delta} := \frac{M^{-k\delta^2}}{\delta\sqrt{\log M}} \ \text{ for } \ \mathsf{c} \le \mathsf{c}_*, \quad \text{ and } \quad \psi_{\mathsf{c}}^t > 1 - M^{-k_1 \tilde{\delta}^2} \ \text{ for } \ \mathsf{c} \ge \mathsf{c}_* + 2, \tag{252}$$

Furthermore, Proposition 4.1 guarantees that $\psi_{\mathsf{c}}^{t+1} \le f_{M,\delta}$ for column indices $\mathsf{c}_* + 2 \le \mathsf{c} \le \mathsf{c}_* + g$. Therefore, all these indices belong to the set $\mathcal{I}_t$. Therefore, $\mathcal{I}_t$ contains at least $(g-2)$ column indices $\mathsf{c} \le \Lambda/2$, and by symmetry, contains an equal number of indices $\Lambda/2 < \mathsf{c} \le \Lambda$. This completes the proof of the lemma.

B Concentration lemmas

In the following $\epsilon > 0$ is assumed to be a generic constant, with additional conditions specified whenever needed.

Lemma B.1 (Hoeffding's inequality [49, Thm. 2.8])**.** *If $X_1, \ldots, X_n$ are independent, bounded random variables such that $a_i \le X_i \le b_i$, then for $\nu = 2[\sum_i (b_i - a_i)^2]^{-1}$*

$$P\Big(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}X_i) \ge \epsilon\Big) \le e^{-\nu n^2 \epsilon^2}, \quad P\Big(\Big|\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}X_i)\Big| \ge \epsilon\Big) \le 2e^{-\nu n^2 \epsilon^2}.$$

Lemma B.2 (Concentration of sums)**.** *If random variables $X_1, \dots, X_M$ satisfy $P(|X_i| \geq \epsilon) \leq e^{-r_i \epsilon^2}$ for $1 \leq i \leq M$ with r_i indicating the 'rate' of concentration of random variable X_i, then*

$$P\Big(\Big|\sum_{i=1}^{M} X_i\Big| \geq \epsilon\Big) \leq \sum_{i=1}^{M} P\Big(|X_i| \geq \frac{\epsilon}{M}\Big) \leq M e^{-(\min_i r_i)\epsilon^2/M^2}.$$

Moreover, for constants $\kappa_1, \dots, \kappa_M > 0$,

$$P\Big(\Big|\sum_{i=1}^{M} \kappa_i X_i\Big| \geq \epsilon\Big) \leq \sum_{i=1}^{M} P\Big(|X_i| \geq \frac{\epsilon}{\sum_{j=1}^{M} \kappa_j}\Big) \leq M e^{-(\min_i r_i)\epsilon^2/(\sum_{j=1}^{M} \kappa_j)^2}.$$

Lemma B.3 (Concentration of Products [44, Lemma 15])**.** *For random variables X, Y and non-zero constants c_X, c_Y, if $P(|X - c_X| \geq \epsilon) \leq K e^{-r_X \epsilon^2}$ and $P(|Y - c_Y| \geq \epsilon) \leq K e^{-r_Y \epsilon^2}$, then the probability $P(|XY - c_X c_Y| \geq \epsilon)$ is bounded by*

$$P(|X - c_X| \geq \min\Big\{\sqrt{\frac{\epsilon}{3}}, \frac{\epsilon}{3c_Y}\Big\}) + P(|Y - c_Y| \geq \min\Big\{\sqrt{\frac{\epsilon}{3}}, \frac{\epsilon}{3c_X}\Big\}) \leq 2K \exp\Big\{\frac{-\min\{r_X, r_Y\}\epsilon^2}{9\max(1, c_X^2, c_Y^2)}\Big\}.$$

Lemma B.4 (Concentration of Products)**.** *For random variables X, Y, and constant $c_X \neq 0$, if $P(|X - c_X| \geq \epsilon) \leq K e^{-r_X \epsilon^2}$ and $P(|Y| \geq \epsilon) \leq K e^{-r_Y \epsilon^2}$, then,*

$$P(|XY| \geq \sqrt{\epsilon}) \leq P(|X - c_X| \geq \sqrt{\epsilon}) + P\Big(|Y| \geq \frac{\epsilon}{2\max\{1, |c_X|\}}\Big) \leq 2K \exp\Big\{-\frac{\epsilon^2 \min\{r_Y, r_X\}}{4\max\{1, c_X^2\}}\Big\}.$$

Lemma B.5 (Concentration of square roots, [44, Lemma 16])**.** *Let $c \neq 0$. Then*

$$\text{If } P(|X_n^2 - c^2| \geq \epsilon) \leq e^{-\kappa n \epsilon^2}, \text{ then } P(||X_n| - |c|\,| \geq \epsilon) \leq e^{-\kappa n |c|^2 \epsilon^2}.$$

Lemma B.6 (Concentration of powers, [44, Lemma 17])**.** *Assume $c \neq 0$ and $0 < \epsilon \leq 1$. Then for any integer $k \geq 2$,*

$$\text{if } P(|X_n - c| \geq \epsilon) \leq e^{-\kappa n \epsilon^2}, \text{ then } P(|X_n^k - c^k| \geq \epsilon) \leq e^{-\kappa n \epsilon^2/[(1+|c|)^k - |c|^k]^2}.$$

Lemma B.7 (Concentration of Scalar Inverses, [44, Lemma 18])**.** *Assume $c \neq 0$ and $0 < \epsilon < 1$.*

$$\text{If } P(|X_n - c| \geq \epsilon) \leq e^{-\kappa n \epsilon^2}, \text{ then } P(|X_n^{-1} - c^{-1}| \geq \epsilon) \leq 2e^{-n\kappa\epsilon^2 c^2 \min\{c^2, 1\}/4}.$$

Lemma B.8. *For a standard Gaussian random variable Z and $\epsilon > 0$, $P(|Z| \geq \epsilon) \leq 2e^{-\frac{1}{2}\epsilon^2}$.*

Lemma B.9. *Let $Z_1, Z_2, \dots, Z_N$ and $\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_N$ be i.i.d. standard Gaussian random variables and $0 \leq \epsilon \leq 1$ and $\sigma_1, \sigma_2, \dots, \sigma_N$ be positive constants. Let $\sigma_{\max} = \max(\sigma_1, \sigma_2, \dots, \sigma_N)$. Then the following concentration results hold for $\epsilon \in (0, 1]$.*

$$P\Big(\Big|\frac{1}{N}\sum_{i=1}^{N} \sigma_i (Z_i^2 - 1)\Big| \geq \epsilon\Big) \leq 2\exp\Big\{\frac{-N\epsilon^2}{4\max\{\sum_{i=1}^{N} 2\sigma_i^2/N,\, \sigma_{\max},\, 1\}}\Big\}, \tag{253}$$

$$P\Big(\Big|\frac{1}{N}\sum_{i=1}^{N} \sigma_i Z_i \tilde{Z}_i\Big| \geq \epsilon\Big) \leq 2\exp\Big\{\frac{-N\epsilon^2}{4\max\{\sum_{i=1}^{N} \sigma_i^2/N, \sigma_{\max},\, 1\}}\Big\}, \tag{254}$$

$$P\Big(\frac{1}{L}\sum_{\ell=1}^{L} \sigma_\ell \max_{j \in sec(\ell)} Z_j^2 \geq 3\sigma_{\max}\log M + \epsilon\Big) \leq \exp\Big\{\frac{-L}{5}\Big(2\epsilon + \log\frac{M}{70}\Big)\Big\}. \tag{255}$$

Proof. Recall that a random variable X is sub-exponential with non-negative parameters (ν, b) if

$$\mathbb{E}[\exp(\lambda(X - \mathbb{E}X))] \leq \exp(\nu^2\lambda^2/2), \quad \text{for all } |\lambda| < 1/b. \tag{256}$$

Furthermore, if X is sub-exponential with parameters (ν, b), then [50, Proposition 2.9]

$$P(|X - \mathbb{E}X| \geq t) \leq \begin{cases} 2\,e^{-\frac{t^2}{2\nu^2}} & \text{for } 0 < t \leq \frac{\nu^2}{b}, \\ 2\,e^{-\frac{t}{b}} & \text{for } t > \frac{\nu^2}{b}. \end{cases} \tag{257}$$

Taking $X = \sum_{i=1}^N \sigma_i Z_i^2$, we will show that X is sub-exponential with $\nu^2 = 4\sum_{i=1}^N \sigma_i^2$ and $b = \max\{4\sigma_{\max}, 1\}$, from which it follows that (plugging in $t = N\epsilon$ in (257))

$$P\Big(\Big|\frac{1}{N}\sum_{i=1}^N \sigma_i(Z_i^2 - 1)\Big| \geq \epsilon\Big) \leq \begin{cases} 2\,\exp\Big\{-\frac{N\epsilon^2}{\frac{8}{N}\sum_{i=1}^N \sigma_i^2}\Big\} & \text{for } 0 < \epsilon \leq \frac{\frac{4}{N}\sum_{i=1}^N \sigma_i^2}{\max\{4\sigma_{\max}, 1\}}, \\ 2\,\exp\Big\{-\frac{N\epsilon}{\max\{4\sigma_{\max}, 1\}}\Big\} & \text{for } \epsilon > \frac{\frac{4}{N}\sum_{i=1}^N \sigma_i^2}{\max\{4\sigma_{\max}, 1\}}. \end{cases}$$

To finish the proof of (253), we now show that $X = \sum_{i=1}^N \sigma_i Z_i^2$ has sub-exponential parameters given above. Using the moment generating function of a chi-squared random variable, we have

$$\mathbb{E}[\exp(\lambda X)] = \prod_{i=1}^n \frac{1}{\sqrt{1 - 2\lambda\sigma_i}}, \quad \text{for } \lambda < \frac{1}{2\sigma_{\max}}.$$

Therefore, for $\lambda < \min\{\frac{1}{4\sigma_{\max}}, 1\}$, we show the desired sub-exponential parameters as follows:

$$\begin{aligned} \mathbb{E}[\exp(\lambda(X - \mathbb{E}X))] &= \exp\Big(-\frac{1}{2}\sum_{i=1}^N \ln(1 - 2\lambda\sigma_i) - \lambda\sum_{i=1}^N \sigma_i\Big) \\ &\overset{(a)}{\leq} \exp\Big(\sum_{i=1}^N \Big[\lambda\sigma_i + \frac{(\lambda\sigma_i)^2}{1 - 2\lambda\sigma_i}\Big] - \lambda\sum_{i=1}^N \sigma_i\Big) \overset{(b)}{\leq} \exp\Big(2\lambda^2\sum_{i=1}^N \sigma_i^2\Big). \end{aligned} \tag{258}$$

In step (a), we use that $-\log(1-u) \leq u + \frac{u^2}{2(1-u)}$ for $u \in [0, 1)$ and step (b) holds since for $\lambda < \frac{1}{4\sigma_{\max}}$, we have $1 - 2\lambda\sigma_i \geq 1 - \sigma_i/(2\sigma_{\max}) \geq 1/2$.

We prove (254) similarly. For $X = \sum_{i=1}^N \sigma_i Z_i \tilde{Z}_i$, the moment generating function is

$$\mathbb{E}[\exp(\lambda X)] = \prod_{i=1}^n (1 - \lambda^2\sigma_i^2)^{-1/2}, \quad \text{for } \lambda^2\sigma_i^2 < 1. \tag{259}$$

Using this and steps similar to (258), we can show that X is sub-exponential with parameters $\nu^2 = 2\sum_{i=1}^N \sigma_i^2$ and $b = \max\{\sigma_{\max}\sqrt{3/2}, 1\}$. Then, using $t = N\epsilon$ in (257), we obtain

$$P\Big(\Big|\frac{1}{N}\sum_{i=1}^N \sigma_i Z_i \tilde{Z}_i\Big| \geq \epsilon\Big) \leq \begin{cases} 2\,\exp\Big\{-\frac{N\epsilon^2}{\frac{4}{N}\sum_{i=1}^N \sigma_i^2}\Big\} & \text{for } 0 < \epsilon \leq \frac{\frac{2}{N}\sum_{i=1}^N \sigma_i^2}{\max\{\sigma_{\max}\sqrt{3/2}, 1\}.}, \\ 2\,\exp\Big\{-\frac{N\epsilon}{\max\{\sigma_{\max}\sqrt{3/2}, 1\}.}\Big\} & \text{for } \epsilon > \frac{\frac{2}{N}\sum_{i=1}^N \sigma_i^2}{\max\{\sigma_{\max}\sqrt{3/2}, 1\}.}. \end{cases}$$

This proves (254). The inequality (255) is shown in [41, Lemma 16]. ∎

C Other useful lemmas

Lemma C.1. *[14, Lemma 8] Let $\mathbf{v}_1, \ldots, \mathbf{v}_t$ be a sequence of vectors in $\mathbb{R}^n$ such that for $i \in [t]$, $\frac{1}{n}\|\mathbf{v}_i - \mathbf{P}^{\|}_{i-1}(\mathbf{v}_i)\|^2 \geq c$, where c is a positive constant that does not depend on n, and $\mathbf{P}^{\|}_{i-1}$ is the orthogonal projection onto the span of $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. Then the matrix $\mathbf{C} \in \mathbb{R}^{t\times t}$ with $C_{ij} = \mathbf{v}_i^* \mathbf{v}_j / n$ has minimum eigenvalue $\lambda_{\min} \geq c'_t$, where c'_t is a positive constant (not depending on n).*

Lemma C.2. *For any scalars $a_1, ..., a_t$ and positive integer m, we have $(|a_1| + \ldots + |a_t|)^m \leq t^{m-1} \sum_{i=1}^t |a_i|^m$. Consequently, for any vectors $\mathbf{u}_1, \ldots, \mathbf{u}_t \in \mathbb{R}^N$, $\left\|\sum_{k=1}^t \mathbf{u}_k\right\|^2 \leq t \sum_{k=1}^t \|\mathbf{u}_k\|^2$.*

Lemma C.3 (Stein's lemma)**.** *For zero-mean jointly Gaussian random variables Z_1, Z_2, and any function $f : \mathbb{R} \to \mathbb{R}$ for which $\mathbb{E}[Z_1 f(Z_2)]$ and $\mathbb{E}[f'(Z_2)]$ both exist, we have $\mathbb{E}[Z_1 f(Z_2)] = \mathbb{E}[Z_1 Z_2]\mathbb{E}[f'(Z_2)]$.*

Lemma C.4. *Let $\boldsymbol{Z}_{\tilde{s},\mathsf{c}}, \boldsymbol{Z}_{s,\mathsf{c}} \in \mathbb{R}^{N\mathsf{c}}$ be random vectors such that the pairs $(Z_{\tilde{s},i}, Z_{s,i})$, $i \in [M_C]$, are i.i.d. bivariate Gaussian with covariance $\mathbb{E}[Z_{\tilde{s},i} Z_{s,i}] = (\tau_c^s / \tau_c^{\tilde{s}})$. Then for $0 \leq \tilde{s} \leq s \leq T$,*

$$\frac{\mathsf{C}}{L}\mathbb{E}\{[\eta_c^{\tilde{s}}(\boldsymbol{\beta}_{0,c} - \sqrt{\tau_c^{\tilde{s}}}\boldsymbol{Z}_{\tilde{s},c})]^*[\eta_c^s(\boldsymbol{\beta}_{0,c} - \sqrt{\tau_c^s}\boldsymbol{Z}_{s,c})]\} = (1 - \psi_c^{\tilde{s}+1}), \tag{260}$$

$$\frac{\mathsf{C}}{L}\mathbb{E}\{[\eta_c^{\tilde{s}}(\boldsymbol{\beta}_{0,c} - \sqrt{\tau_c^{\tilde{s}}}\boldsymbol{Z}_{\tilde{s},c}) - \boldsymbol{\beta}_{0,\mathsf{c}}]^*[\eta_c^s(\boldsymbol{\beta}_{0,c} - \sqrt{\tau_c^s}\boldsymbol{Z}_{s,c}) - \boldsymbol{\beta}_{0,\mathsf{c}}]\} = \psi_c^{s+1}. \tag{261}$$

Proof. We will use the following fact, adapted from [4, Proposition 1]:

$$\mathbb{E}\Big\{\boldsymbol{\beta}_{0,\mathsf{c}}^* \eta_{\mathsf{c}}^s\Big(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^s}\boldsymbol{Z}_{s,\mathsf{c}}\Big)\Big\} = \frac{L}{\mathsf{C}}(1 - \psi_{\mathsf{c}}^{s+1}), \quad \text{for } 0 \leq s < T. \tag{262}$$

Let $\mathbf{u}^{\tilde{s}} = \boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}\boldsymbol{Z}_{\tilde{s},\mathsf{c}}$ and $\mathbf{u}^s = \boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^s}\boldsymbol{Z}_{s,\mathsf{c}}$. Then, $\eta_{\mathsf{c}}^{\tilde{s}}(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}\boldsymbol{Z}_{\tilde{s},\mathsf{c}}) = \mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^{\tilde{s}}]$ and $\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^s}\boldsymbol{Z}_{s,\mathsf{c}}) = \mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^s]$, and therefore, for $\tilde{s} \leq s$,

$$\begin{aligned}
\mathbb{E}\{[\eta_{\mathsf{c}}^{\tilde{s}}(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^{\tilde{s}}}\boldsymbol{Z}_{\tilde{s},\mathsf{c}})]^*[\eta_{\mathsf{c}}^s(\boldsymbol{\beta}_{0,\mathsf{c}} - \sqrt{\tau_{\mathsf{c}}^s}\boldsymbol{Z}_{s,\mathsf{c}})]\} &= \mathbb{E}\{[\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^{\tilde{s}}]]^*\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^s]\} \\
&\overset{(a)}{=} \mathbb{E}\{[\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^{\tilde{s}}]]^*[\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^s, \mathbf{u}^{\tilde{s}}] - \boldsymbol{\beta}_{0,\mathsf{c}} + \boldsymbol{\beta}_{0,\mathsf{c}}]\} \\
&\overset{(b)}{=} \mathbb{E}\{[\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^{\tilde{s}}]]^*\boldsymbol{\beta}_{0,\mathsf{c}}\} \overset{(c)}{=} \frac{L}{\mathsf{C}}(1 - \psi_{\mathsf{c}}^{\tilde{s}+1}).
\end{aligned}$$

In the above, step (a) holds since $\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^s, \mathbf{u}^r] = \mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}} \mid \mathbf{u}^s]$, which can be shown using steps similar to those in [41, Lemma 22]. Step (b) follows from the orthogonality property of conditional expectation: $\mathbb{E}\{(\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}}|\mathbf{u}^s, \mathbf{u}^{\tilde{s}}] - \boldsymbol{\beta}_{0,\mathsf{c}})^*\mathbb{E}[\boldsymbol{\beta}_{0,\mathsf{c}}|\mathbf{u}^{\tilde{s}}]\} = 0$ due to the orthogonality principle, and step (c) by (262). The result (261) follows from (260) and (262), noting that $\left\|\boldsymbol{\beta}_{0,\mathsf{c}}\right\|^2 = L/L_C$. ∎

Lemma C.5. *For the function $\eta^t : \mathbb{R}^{ML} \to \mathbb{R}^{ML}$ defined (16), $\mathbf{s}, \boldsymbol{\Delta} \in \mathbb{R}^{ML}$, and $sec(\ell) \in \mathsf{c}$,*

$$\sum_{i \in sec(\ell)} \left|\eta_i^t(\mathbf{s}) - \eta_i^t(\mathbf{s} + \boldsymbol{\Delta})\right| \leq (2/\tau_c^t) \max_{i \in sec(\ell)} |\Delta_i| .$$

Proof. From the multivariate version of Taylor's theorem, for any $i \in [ML]$ and for some $\kappa \in (0, 1)$,

$$\eta_i^t(\mathbf{s} + \boldsymbol{\Delta}) = \eta_i^t(\mathbf{s}) + \Delta^T \nabla \eta_i^t(\mathbf{s} + \kappa\boldsymbol{\Delta}). \tag{263}$$

For $i \in sec(\ell)$, as η_i^t depends only on the subset of its input also belonging to section ℓ, using (263),

$$\sum_{i\in sec(\ell)} \left|\eta_i^t(\mathbf{s}) - \eta_i^t(\mathbf{s}+\mathbf{\Delta})\right| = \sum_{i\in sec(\ell)} \Big| \sum_{j\in sec(\ell)} \Delta_j \frac{\partial}{\partial s_j}\eta_i^t(\mathbf{s}+\kappa\mathbf{\Delta})\Big|$$
$$\overset{(a)}{\leq} \frac{1}{\tau_{\mathsf{c}}^t} \sum_{i\in sec(\ell)} \Big|\Delta_i \eta_i^t(\mathbf{s}+\kappa\mathbf{\Delta})\Big| + \frac{1}{\tau_{\mathsf{c}}^t} \sum_{i\in sec(\ell)} \Big|\eta_i^t(\mathbf{s}+\kappa\mathbf{\Delta}) \sum_{j\in sec(\ell)} \Delta_j \eta_j^t(\mathbf{s}+\kappa\mathbf{\Delta})\Big| \overset{(b)}{\leq} \frac{2}{\tau_{\mathsf{c}}^t} \max_{i\in \mathrm{sec}(\ell)} |\Delta_i|\,,$$

where inequality (a) uses the fact that for $i, j \in [ML]$,

$$\frac{\partial}{\partial s_j}\eta_i^t(\mathbf{s}) = \frac{\eta_i^t(\mathbf{s})}{\tau_{\mathsf{c}}^t}[\mathbf{1}\{j=i\} - \eta_j^t(\mathbf{s})]\mathbf{1}\{i,j \in sec(\ell), \ell \in \mathsf{c}\}.$$

Inequality (b) uses the fact that $\sum_{j\in sec(\ell)} |\eta_j^t(\mathbf{s}+\kappa\mathbf{\Delta})| = \sum_{j\in sec(\ell)} \eta_j^t(\mathbf{s}+\kappa\mathbf{\Delta}) = 1$. ■

Lemma C.6. *Let $\mathcal{W}$ be a d-dimensional subspace of $\mathbb{R}^n$ for $d \leq n$ and let $\boldsymbol{Z} \sim \mathcal{N}(0, \boldsymbol{I}_n)$ be a standard Gaussian random vector. Let $(\boldsymbol{w}_1, ..., \boldsymbol{w}_d)$ be an orthonormal basis of $\mathcal{W}$ with $\|\boldsymbol{w}_i\|^2 = 1$ for $i \in [d]$, and let $\mathbf{P}_{\mathcal{W}}^{\|}$ denote the orthogonal projection operator onto $\mathcal{W}$. Then for $\mathbf{D} = [\boldsymbol{w}_1 \mid \ldots \mid \boldsymbol{w}_d]$, we have $\mathbf{P}_{\mathcal{W}}^{\|}\boldsymbol{Z} \overset{d}{=} \mathbf{D}\tilde{\boldsymbol{Z}}$ where $\tilde{\boldsymbol{Z}} \sim \mathcal{N}(0, \boldsymbol{I}_d)$ is independent of $\mathbf{D}$.*

Lemma C.7 ($\mathcal{H}(d)$ concentration)**.** *Let $\boldsymbol{Z} \sim \mathcal{N}(0, \mathbb{I}_{ML})$ and $\tilde{\boldsymbol{Z}} \sim \mathcal{N}(0, \mathbb{I}_{ML})$ such that $(Z_i, \tilde{Z}_i)$ are i.i.d. bivariate Gaussian, for $1 \leq i \leq ML$. For $\ell \in [L]$, let $\mathbf{Y}_\ell = \boldsymbol{Z}_\ell^* \eta_\ell(\boldsymbol{\beta}_0 - \sqrt{\tau}\tilde{\boldsymbol{Z}})$. Then, for a universal positive constant, κ, and $\lambda_{\mathsf{c}} \in \Theta(1)$ for each $\mathsf{c} \in [\mathsf{C}]$,*

$$P\Big(\frac{1}{L}\Big|\sum_{\mathsf{c}\in[\mathsf{C}]} \frac{\lambda_{\mathsf{c}} W_{rc}}{\sqrt{\log M}} \sum_{\ell\in\mathsf{c}} (\mathbf{Y}_\ell - \mathbb{E}[\mathbf{Y}_\ell])\Big| \geq \epsilon\Big) \leq \exp\{-\kappa_1 L(\omega/R)\epsilon^2\}.$$

Proof. The proof is along the same lines as that of Lemma 20 in [41], and is hence omitted. ■

References

[1] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2541–2557, May 2012.

[2] A. Joseph and A. R. Barron, "Fast sparse superposition codes have near exponential error probability for $R < \mathcal{C}$," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 919–942, 2014.

[3] S. Cho and A. R. Barron, "Approximate iterative Bayes optimal estimates for high-rate sparse superposition codes," in *Sixth Workshop on Inf. The. Methods in Sci. and Eng.*, pp. 35–42, 2013.

[4] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse superposition codes via approximate message passing decoding," *IEEE Trans. Inf. Theory*, vol. 63, pp. 1476–1500, March 2017.

[5] A. Greig and R. Venkataramanan, "Techniques for improving the finite length performance of sparse superposition codes," *IEEE Trans. Commun.*, vol. 66, pp. 905–917, March 2018.

[6] R. Venkataramanan, S. Tatikonda, and A. Barron, "Sparse regression codes," *Foundations and Trends in Communications and Information Theory*, vol. 15, no. 1-2, pp. 1–195, 2019.

[7] J. Barbier, C. Schülke, and F. Krzakala, "Approximate message-passing with spatially coupled structured operators, with applications to compressed sensing and sparse superposition codes," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 5, p. P05013, 2015.

[8] J. Barbier, M. Dia, and N. Macris, "Proof of threshold saturation for spatially coupled sparse superposition codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016.

[9] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity achieving sparse superposition codes," *IEEE Trans. Inf. Theory*, vol. 63, pp. 4894–4927, Aug 2017.

[10] J. Barbier, M. Dia, and N. Macris, "Universal sparse superposition codes with spatial coupling and GAMP decoding," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5618–5642, 2019.

[11] K. Hsieh, C. Rush, and R. Venkataramanan, "Spatially coupled sparse regression codes: Design and state evolution analysis," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018.

[12] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. D. Pfister, "A simple proof of Maxwell saturation for coupled scalar recursions," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6943–6965, 2014.

[13] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[14] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, pp. 764–785, Feb 2011.

[15] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[16] M. Bayati and A. Montanari, "The LASSO risk for Gaussian matrices," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 1997–2017, 2012.

[17] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2015.

[18] A. Fengler, P. Jung, and G. Caire, "SPARCs and AMP for unsourced random access," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019.

[19] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011.

[20] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Processing*, vol. 63, no. 4, pp. 1043–1055, 2014.

[21] D. Donoho and A. Montanari, "High dimensional robust M-estimation: Asymptotic variance via Approximate Message Passing," *Probab. Theory and Relat. Fields*, vol. 166, no. 3-4, pp. 935–969, 2016.

[22] J. Bradic, "Robustness in sparse high-dimensional linear models: Relative efficiency and robust approximate message passing," *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 3894–3944, 2016.

[23] S. Rangan and A. K. Fletcher, "Iterative estimation of constrained rank-one matrices in noise," in *Proc. Int. Symp. on Information Theory*, pp. 1246–1250, 2012.

[24] Y. Kabashima, F. Krzakala, M. Mézard, A. Sakata, and L. Zdeborová, "Phase transitions and sample complexity in Bayes-optimal matrix factorization," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4228–4265, 2016.

[25] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse PCA," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2197–2201, June 2014.

[26] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—part II: Applications," *IEEE Trans. Signal Processing*, vol. 62, no. 22, pp. 5854–5867, 2014.

[27] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Advances in Neural Information Processing Systems*, pp. 424–432, 2016.

[28] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *Annals of Statistics*, vol. 49, pp. 321–345, 2 2021.

[29] A. J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, pp. 2181–2191, Sep 1999.

[30] M. Lentmaier, A. Sridharan, D. J. Costello, and K. S. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5274–5289, 2010.

[31] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, pp. 803–834, Feb 2011.

[32] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, pp. 7761–7813, Dec 2013.

[33] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "Threshold saturation for spatially coupled LDPC and LDGM codes on BMS channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7389–7415, 2014.

[34] D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, "Spatially coupled LDPC codes constructed from protographs," *IEEE Trans. Inf. Theory*, vol. 61, pp. 4866–4889, Sept 2015.

[35] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. 48th Annu. Allerton Conf. Commun.,Control, Comput.*, pp. 347–353, 2010.

[36] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, no. 2, pp. 1–18, 2012.

[37] K. Takeuchi, T. Tanaka, and T. Kawabata, "A phenomenological study on threshold improvement via spatial coupling," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. 95, no. 5, pp. 974–977, 2012.

[38] K. Takeuchi, T. Tanaka, and T. Kawabata, "Performance improvement of iterative multiuser detection for large sparsely spread cdma systems by spatial coupling," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1768–1794, 2015.

[39] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, pp. 7434–7464, Nov. 2013.

[40] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Information and Inference: A Journal of the IMA*, vol. 2, no. 2, pp. 115–144, 2013.

[41] C. Rush and R. Venkataramanan, "The error probability of sparse superposition codes with approximate message passing decoding," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3278–3303, 2019.

[42] S. Liang, J. Ma, and L. Ping, "Clipping can improve the performance of spatially coupled sparse superposition codes," *IEEE Commun. Letters*, vol. 21, pp. 2578–2581, Dec. 2017.

[43] R. El-Khatib and N. Macris, "The velocity of the propagating wave for spatially coupled systems with applications to ldpc codes," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7113–7131, 2018.

[44] C. Rush and R. Venkataramanan, "Finite-sample analysis of approximate message passing algorithms," *IEEE Trans. Inf. Theory*, vol. 64, pp. 7264–7286, November 2018.

[45] K. Hsieh and R. Venkataramanan, "Modulated sparse superposition codes for the complex AWGN channel," *To appear in IEEE Trans. Inf. Theory*, 2021. https://arxiv.org/abs/2004.09549.

[46] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, May 2010.

[47] A. Cassagne, O. Hartmann, M. Léonardon, *et al.*, "Fast simulation and prototyping with AFF3CT," in *IEEE International Workshop on Signal Processing Systems (SiPS)*, Oct. 2017.

[48] K. Hsieh, "Python implementation of sparse regression codes." https://github.com/kuanhsieh/sparc_public, 2020.

[49] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[50] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48. Cambridge University Press, 2019.