

Supplementary Information for "Fragility of Magic State Distillation under Imperfect Measurements"

Yunzhe Zheng,^{1,2} Yuanchen Zhao,^{1,3} and Dong E. Liu^{1,3}

¹*Department of Physics, Tsinghua University, Beijing, 100084, China*

²*Department of Applied Physics, Yale University, New Haven, Connecticut, 06520, USA*

³*Frontier Science Center for Quantum Information, Beijing 100184, China*

(Dated: July 16, 2026)

CONTENTS

I. Supplementary Note 1: Unnecessity for encoding in practical MSD	1
II. Supplementary Note 2: Description of $[[15, 1, 3]]$ and $[[14, 2, 2]]$ codes	2
III. Supplementary Note 3: Gaussian measurement error model and generality	3
IV. Supplementary Note 4: Used Lemmas	4
V. Supplementary Note 5: Proof of Theorems	6
VI. Supplementary Note 6: Eigenvalue of Jacobian matrix at fixed point	10
VII. Supplementary Note 7: Other choice of stabilizer generators	10
References	11

I. SUPPLEMENTARY NOTE 1: UNNECESSITY FOR ENCODING IN PRACTICAL MSD

It's unnecessary to encode the input states into a logical state in the framework of stabilizer reduction [1]. The conclusion holds for all stabilizer codes with transversal non-Clifford gates, but they only provide MSD protocols distilling into $\theta_k = (|0\rangle + e^{i\pi/2^k} |1\rangle)/\sqrt{2}$ (or their Clifford equivalence) due to the restriction on transversal logical gates [2].

To see this fact, we take the protocols for the $|T\rangle$ states and start with the canonical way for error corrected T gates as Figure S1(a). We prepare the logical $|+\rangle_L$ by initializing tensor product state $|+\rangle^{\otimes n}$ and measure every Z -type stabilizer generators. Gauge correction might be required depending on the measurement outcome. This preparation operation can be characterized by the projector $\bar{P}_Z = \prod_{g \in \mathcal{G}_Z} \frac{I+g}{2}$ where $\mathcal{G}_Z$ is the set of all Z -type stabilizer generators. As show later in Lemma 1, it commutes with the transversal $\bar{T}$ operations. Therefore, we can move the transversal T gates to the very beginning and merge the stabilizer measurements for X -type and Z -type stabilizer as Figure S1(b). Lastly, we should notice that the decoder operation convert the logical state in the codespace of $\mathcal{Q}$ back to physical state, so it will also propagate stabilizer generators back to the raw physical Z operators on every ancilla qubits. We can then postpone the measurement after decoding. The gauge correction can also be propagated after the decoder operation as it's a Clifford operation, but we only need to track the effect of propagated gauge correction on the decoded data qubits and applying it back to our physical state, as shown in Figure S1(c).

We emphasize that these gauge corrections are not needed in the stabilizer-reduction (SR) description used in the main text. The discussion above only shows the equivalence between the usual deterministic logical-state-preparation circuit and the SR circuit. In the SR protocol, the relevant map is the post-selected branch in which the measured stabilizer generators return the accepted, all-+1 syndrome. Therefore, no gauge correction is applied in the SR analysis. The subspace-transfer operators that appear later in the proofs should instead be understood as algebraic Pauli operators that relate a syndrome-subspace projector to the codespace projector, e.g., $P_{\mathbf{x}} = C_{\mathbf{x}} \bar{P} C_{\mathbf{x}}$, rather than as feed-forward gauge corrections performed by the protocol. Consequently, including or omitting gauge corrections in the circuit representation does not affect the noisy-projector analysis or the standard-form mitigation strategy.

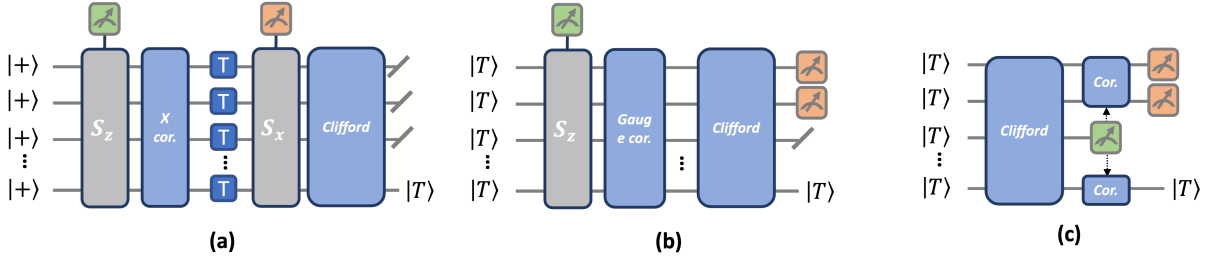


FIG. S1. Equivalence between code with transversal T gates and a stabilizer reduction protocol for magic state $|T\rangle$. (a) Distillation circuit with logical state preparation. We prepare the logical $|+\rangle_L$ by preparing tensor product of $|+\rangle^{\otimes n}$ and measuring all Z -type stabilizer generators. We then apply $\bar{T} = T^{\otimes 15}$ and decode it with error detection. (b) As $T^{\otimes n}$ commutes with the state preparation operator P_Z , we can bring it forward and act on the $|+\rangle^{\otimes n}$ directly. (c) We can then propagate the stabilizer measurement after the decoder circuit, which will be exactly measuring ancilla qubits.

II. SUPPLEMENTARY NOTE 2: DESCRIPTION OF $[[15, 1, 3]]$ AND $[[14, 2, 2]]$ CODES

Both the $[[15, 1, 3]]$ [3] and $[[14, 2, 2]]$ [4] codes have logical transversal T gates and therefore can be used as MSD protocol to distill the T states. We give their stabilizer description here as they are practical small examples to work with.

For concreteness, Figure S2 shows decoding circuits for the two protocols in the stabilizer-reduction implementation used in the main text.

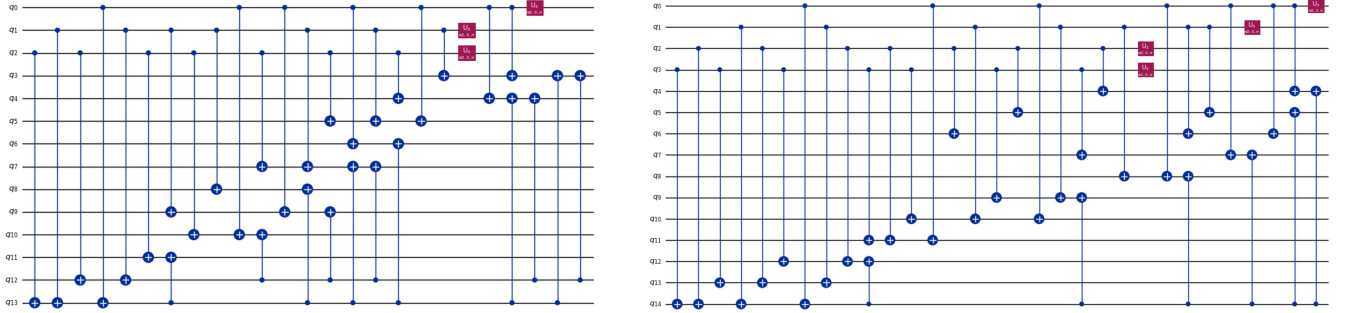


FIG. S2. **Decoding circuits for the two example MSD protocols.** The left panel shows the decoding circuit for the $[[15, 1, 3]]$ protocol and the right panel shows the decoding circuit for the $[[14, 2, 2]]$ protocol. The output qubits are placed at the bottom, and all other qubits are measured in the Z basis.

The $[[15, 1, 3]]$ code has 10 Z -type generators and 4 X -type generators. For the canonical generator choice, the parity check matrix is given by

$$H_X = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_Z = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \quad (S1)$$

Notably, $H_Z = [H_X, H'_Z]^T$. The logical operator is respectively $X_L = X^{\otimes 15}$, $Z_L = Z^{\otimes 15}$. For the parity-check matrices

in standard form, we have

$$H_X = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_Z = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad (S2)$$

and the logical operator is given by $X_L = IIIIXXIXXXIIXIX$ and $Z_L = ZZZZIIIIIIIIIZ$.

The $[[14, 2, 2]]$ code has 9 Z -type generators and 3 X -type generators. For the canonical generator choice, its parity check matrix is given by

$$H_X = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_Z = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (S3)$$

The logical operators are respectively $X_{L,1} = XXXXXXXXIIIIII$, $Z_{L,1} = ZZZZZZZIIIIII$, and $X_{L,2} = IIIIIIXXXXXXX$, $Z_{L,2} = IIIIIIZZZZZZ$. For the parity-check matrices in standard form,

$$H_X = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_Z = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad (S4)$$

and the logical operators are respectively $X_{L,1} = IIIIIXIXXXXXX$, $Z_{L,1} = IZZIIIIIIIIIZI$ and $X_{L,2} = IIIXIXXXIIXIX$, $Z_{L,2} = ZIIIIIIIIIIIZZ$.

III. SUPPLEMENTARY NOTE 3: GAUSSIAN MEASUREMENT ERROR MODEL AND GENERALITY

In the main text, we considered the discretized noisy measurement operators under Gaussian noise. In practice, the measurement outcome is continuous and the noisy measurement under Gaussian noise for observable g with outcome s is given by:

$$E(s) = (k/2\pi)^{1/2} \exp(-k(s-g)^2/2), \quad \int_{-\infty}^{+\infty} ds E(s) = I, \quad (S5)$$

and $E(s) = M^\dagger(s)M(s)$ where $M(s)$ is the weak measurement operator. The outcome provided by the detector is a real number s , but since the eigenvalue of g is ± 1 and we expect a binary outcome, in practice we often identify the $s > 0$ outcome as $\tilde{s} = +1$ and $s < 0$ outcomes as $\tilde{s} = -1$. This leads to the measurement probability,

$$P(\tilde{s} = \pm 1) = P(\pm s > 0) = \frac{1}{2} \left(1 \pm g \operatorname{Erf} \left(\sqrt{k/2} \right) \right), \quad (S6)$$

where Erf is the Gaussian error function. This probability is equivalent to that yielded by the binary model used in the simulation

$$E(\tilde{s}) \propto e^{-\tilde{s}\beta g}, \quad (\text{S7})$$

where we defined the measurement strength $\beta = \text{arctanh}(\text{Erf}(\sqrt{k/2}))$.

In the generic case, the format of $E(s)$ doesn't need to be Gaussian. However, whatever it turns to be, the discretized measurement operator will still be a similar form as the Gaussian case, except for a change of the coefficients. Particularly, the noisy measurement operator for a qubit Pauli observable g is given by

$$\tilde{M}_+ = f_1 P_+ + f_2 P_-, \quad \tilde{M}_- = h_1 P_- + h_2 P_+, \quad (\text{S8})$$

where $P_\pm = \frac{I \pm g}{2}$ is the projector for ± 1 eigenspace of g . The the normalization condition $\tilde{M}_+^\dagger \tilde{M}_+ + \tilde{M}_-^\dagger \tilde{M}_- = I$ demands

$$f_1^2 + h_2^2 = 1, \quad f_2^2 + h_1^2 = 1, \quad (\text{S9})$$

and the semi-positive property for measurement operators requires $f_{1,2}$ and $h_{1,2}$ to be non-negative. If we are under such a noise model, we can still use the conclusion and analysis for Gaussian noise by a parameter mapping. The mapping is given by

$$f_2/f_1 = \tanh \beta/2, \quad (\text{S10})$$

and we can relate the new defined noise parameter with the measurement strength for gaussian noise. For example, we may consider the following noisy measurements that differ from the Gaussian noise:

$$\tilde{M}_+ = \sqrt{\frac{1+\eta}{2}} P_+ + \sqrt{\frac{1-\eta}{2}} P_-, \quad \tilde{M}_- = \sqrt{\frac{1-\eta}{2}} P_+ + \sqrt{\frac{1+\eta}{2}} P_-, \quad (\text{S11})$$

where $\eta \in [0, 1]$ now is the new "measurement strength". We then have the equation

$$\sqrt{\frac{1-\eta}{1+\eta}} = \tanh \beta/2. \quad (\text{S12})$$

Now all the analysis showed with β can be showed using η by replacing the parameter using the above relation. For example, the first-order biased noise for the $[[15, 1, 3]]$ protocol now scales as $e^{-\tanh^{-1}(\sqrt{\frac{1-\eta}{1+\eta}})}$ rather than $e^{-2\beta}$.

IV. SUPPLEMENTARY NOTE 4: USED LEMMAS

Lemma 1. *If a n -qubit stabilizer code Q allows transversal implementation of logical K gates, then $K^{\otimes n}$ commutes with the codespace projector $\bar{P}$.*

Proof. For any n -qubit quantum states $|\psi\rangle$ and stabilizer code Q with codespace projector $\bar{P}$, we can decompose the state into components in/out of the codespace defined by Q as $|\psi\rangle = |\psi\rangle_i + |\psi\rangle_o$ such that $\bar{P}|\psi\rangle = |\psi\rangle_i$, $(I - \bar{P})|\psi\rangle = |\psi\rangle_o$. As logical operations should preserve the codespace as well as the non-codespace and $K^{\otimes n}$ is a logical operation, $\bar{P}K^{\otimes n}|\psi\rangle_o = 0$. Therefore,

$$\bar{P}K^{\otimes n}|\psi\rangle = \bar{P}(K^{\otimes n}|\psi\rangle_i + K^{\otimes n}|\psi\rangle_o) = K^{\otimes n}|\psi\rangle_i. \quad (\text{S13})$$

Besides, we can easily have

$$K^{\otimes n}\bar{P}|\psi\rangle = K^{\otimes n}|\psi\rangle_i. \quad (\text{S14})$$

As $K^{\otimes n}\bar{P}|\psi\rangle = \bar{P}K^{\otimes n}|\psi\rangle$ for any quantum states, $\bar{P}$ must commute with $K^{\otimes n}$. $\square$

Although we don't have any restriction for the transversal K gates, the only possible transversal gates are just $R_Z(\theta)$ gates (or their Pauli equivalence), with $\theta = \pi/2^k$ for k being a non-negative integer [2, 5].

Lemma 2. For a CSS stabilizer code $\mathcal{Q}$ with generator $\{g_i\}$, we can always find a pure X or Z Pauli subspace-transfer operator C_i , i.e. composed of only one type of Pauli operators (X, Y or Z) and I , to map its codespace to another subspace that is stabilized by $\{(-1)^{\alpha_i} g_i\}$ where $\alpha_i \in \{0, 1\}$ and $|\alpha| = 1$ with $\alpha = \{\alpha_1, \alpha_2, \dots\}$, while preserving the logical state.

Proof. Without loss of generality we consider the subspace $\tilde{\mathcal{Q}}$ where only the syndrome of X -type stabilizer g_k is flipped. The subspace projector associated with $\tilde{\mathcal{Q}}$ is now:

$$P_{\tilde{\mathcal{Q}}} = \frac{I - g_k}{2} \prod_{j=1, j \neq k}^{\bar{n}} \frac{I + g_j}{2}. \quad (\text{S15})$$

As the subspace-transfer operator C_k should map the codespace to the subspace, we have $C_k \bar{P} C_k = P_{\tilde{\mathcal{Q}}} = \frac{I - g_k}{2} \prod_{j=1, j \neq k}^{\bar{n}} \frac{I + g_j}{2}$. Therefore, C_k can be understood as a physical error that only triggers the g_k syndrome and we can restrict it to be Z -type errors. Furthermore, C_k will preserve logical state and therefore should commute with all logical operators. For CSS codes, we can choose logical $X(Z)$ to be pure $X(Z)$ operators. If C_k commutes with all logical X operators, then we are all set and the C_k is the given subspace-transfer operator. If not, we can multiply C_k with the logical Z operators associated with the non-commuting local X operators as the $\tilde{C}_k$. The $\tilde{C}_k$ can then be the subspace-transfer operator we want. $\square$

Lemma 3. For a stabilizer code whose generators are written in standard form, the subspace-transfer operators C_i in Lemma 2 can be chosen to have weight one, to have pairwise disjoint support, and to commute with all logical operators.

Proof. Write the stabilizer generators in standard form

$$[g_1, g_2, \dots, g_{\bar{n}}]^T = [H_X | H_Z] = \left[\begin{array}{ccc|ccc} I & A_1 & A_2 & B & 0 & C \\ 0 & 0 & 0 & D & I & E \end{array} \right]. \quad (\text{S16})$$

Here $r = \text{rank}(H_X)$. For the first r generators, choose $C_i = Z_i$. For the remaining generators, indexed as g_{r+a} with $1 \leq a \leq \bar{n} - r$, choose $C_{r+a} = X_{r+a}$. By the identity blocks in this standard-form representation, each chosen C_i anticommutes only with the corresponding generator g_i and commutes with all other generators. Hence C_i maps the codespace to the syndrome subspace in which only the i -th syndrome is flipped. The operators C_i are manifestly weight-one and have pairwise disjoint support.

The logical operators associated with this standard form may be chosen as [6]

$$L_X = [0 \ E^T \ I | C^T \ 0 \ 0], L_Z = [0 \ 0 \ 0 | A_2^T \ 0 \ I]. \quad (\text{S17})$$

The supports selected above lie in columns where these representatives have no anticommuting support. Therefore every C_i commutes with both L_X and L_Z , and hence with the logical Pauli group. $\square$

For example, in the $[[15, 1, 3]]$ standard-form matrix in Equation (S2), the four X -type generators have the identity block on the first four qubits, so one can choose the corresponding subspace-transfer operators as Z_1, Z_2, Z_3, Z_4 . The ten Z -type generators have the identity block on qubits 5 through 14, so the corresponding subspace-transfer operators can be chosen as $X_5, \dots, X_{14}$; these single-qubit operators have disjoint support and commute with the standard-form logical operators $X_L = IIIIXXIXXXIXIX$ and $Z_L = ZZZZIIIIIIIIIZ$.

Lemma 4. The subspace-transfer operators in Lemma 2 can always be chosen such that they don't support logical operators.

Proof. If the generators of the code are in standard form, then all subspace-transfer operators apparently don't support logical operators. Now let's consider an arbitrary CSS code with X -type and Z -type generators. Since the parity check matrix for a CSS code can always be converted into standard form by row adding and qubit permutation, we consider the effect of row adding on the subspace-transfer operator. Without loss of generality we consider X -type generator: As long as we just add X -type row with X -type row, the subspace-transfer operator will always support on the first r qubits. However, no logical operators will be supported on the first r qubits. Besides, qubit permutation doesn't modify the weight of subspace-transfer operators and logical operators. $\square$

V. SUPPLEMENTARY NOTE 5: PROOF OF THEOREMS

Theorem 1. *For MSD protocols based on $[[n, k, d]]$ ($d \geq 2$) CSS codes with transversal non-Clifford gates, the first-order effect of weak measurement for X -type generators won't affect the asymptotically distilled output states up to the first order of $e^{-2\beta}$. The effect of weak measurement for Z -type generators will be at most X noise on the asymptotically distilled output states up to the first order of $e^{-2\beta}$. Therefore, the first-order noise ($e^{-2\beta}$), if present, must be biased.*

Proof. Without loss of generality, we may assume the Gaussian noise to clearly denote the noise order. (Note the proof doesn't rely on this property) The noisy measurement operator for stabilizer g to measure $+1$ can be written as

$$M_+(g, \beta) = K(I + \tanh \frac{\beta}{2} g) = K(\frac{2}{1 + e^{-\beta}} P_+ + \frac{2}{1 + e^{\beta}} P_-), \quad (\text{S18})$$

where $P_{\pm} = (I \pm g)/2$. We define an $\bar{n}$ -bit binary string operator on $\mathbb{F}_2$: $\mathbf{x} = (x_1, x_2, \dots, x_{\bar{n}})$. For stabilizer code $\mathcal{Q}$ with $\bar{n}$ stabilizer generators, all the subspace projectors associated with stabilizer code $\mathcal{Q}$ are given by

$$\bar{P}_{\mathbf{x}} = \frac{1}{2^{\bar{n}}} \prod_{i=1}^{\bar{n}} (I + (-1)^{x_i} g_i). \quad (\text{S19})$$

It's easy to see we recover codespace projector $\bar{P}$ when $\mathbf{x} = \mathbf{0}$, i.e. $\bar{P} = \bar{P}_{\mathbf{0}}$. Therefore, we can rewrite the codespace measurement operator as a sum of subspace projector:

$$\tilde{M}_{\mathcal{Q}} = \prod_{i=1}^{\bar{n}} M_+(g_i, \beta) = K^{\bar{n}} \sum_{\mathbf{x} \in \mathbb{F}_2^{\bar{n}}} \gamma_{\mathbf{x}} \bar{P}_{\mathbf{x}}, \quad (\text{S20})$$

where

$$\gamma_{\mathbf{x}} = (\frac{2e^{\beta}}{1 + e^{\beta}})^{\bar{n} - |\mathbf{x}|} (\frac{2}{1 + e^{\beta}})^{|\mathbf{x}|} = (\frac{2}{1 + e^{\beta}})^{\bar{n}} e^{-\beta|\mathbf{x}|}. \quad (\text{S21})$$

The post-measurement state of ρ_{in} is then given by

$$\begin{aligned} \rho_{\text{p}} &\propto \tilde{M}_{\mathcal{Q}} \rho_{\text{in}} \tilde{M}_{\mathcal{Q}} \\ &\propto \sum_{\mathbf{x}, \mathbf{x}'} \gamma_{\mathbf{x}} \gamma_{\mathbf{x}'} \bar{P}_{\mathbf{x}} \rho_{\text{in}} \bar{P}_{\mathbf{x}'} \\ &\propto \sum_{\mathbf{x}, \mathbf{x}'} e^{-(|\mathbf{x}| + |\mathbf{x}'|)\beta} \bar{P}_{\mathbf{x}} \rho_{\text{in}} \bar{P}_{\mathbf{x}'}. \end{aligned} \quad (\text{S22})$$

Now we consider the final output state ρ_{o} decoded from the post-measurement state ρ_{p} . There are two procedure in the decoding process for MSD: First, we should apply the decoding circuit [1] (the inverse of encoding circuit found using Gottesman's algorithm [7]). Second, we should trace out all ancilla qubits and only keep these storing logical information. When the measurement is perfect, the logical qubits are fully unentangled with all other qubits and tracing out won't impact the output state. However, they are indeed entangled in the imperfect measurement case, and the tracing out operation will impact the qubits storing logical information. For simplicity we consider $k = 1$, then

$$\begin{aligned} \rho_{\text{o}} &= \text{Tr}_{\text{A}}[D_{\mathcal{Q}} \rho_{\text{p}} D_{\mathcal{Q}}^{\dagger}] \\ &\propto \sum_{\mathbf{x}, \mathbf{x}'} e^{-(|\mathbf{x}| + |\mathbf{x}'|)\beta} \text{Tr}_{\text{A}}[D_{\mathcal{Q}} \bar{P}_{\mathbf{x}} \rho_{\text{in}} \bar{P}_{\mathbf{x}'} D_{\mathcal{Q}}^{\dagger}] \\ &\propto \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}_{\text{A}}[D_{\mathcal{Q}} \bar{P}_{\mathbf{x}} \rho_{\text{in}} \bar{P}_{\mathbf{x}} D_{\mathcal{Q}}^{\dagger}] \end{aligned} \quad (\text{S23})$$

where Tr_{A} means tracing out all ancillary qubits and from the second row to third row we used the fact that non-diagonal terms don't contribute when the ancillary qubits will be traced out after decoding the logical states to physical states. Let's restrict $\rho_{\text{in}} = \rho_{\text{i}}^{\otimes n}$ to be a product state. For leading-order consideration, we have

$$\rho_{\text{o}}(\rho_{\text{i}}) \propto \text{Tr}_{\text{A}}[D_{\mathcal{Q}} \bar{P}_{\mathbf{0}} \rho_{\text{i}}^{\otimes n} \bar{P}_{\mathbf{0}} D_{\mathcal{Q}}^{\dagger}] + \sum_{|\mathbf{x}|=1} e^{-2\beta} \text{Tr}_{\text{A}}[D_{\mathcal{Q}} \bar{P}_{\mathbf{x}} \rho_{\text{i}}^{\otimes n} \bar{P}_{\mathbf{x}} D_{\mathcal{Q}}^{\dagger}]. \quad (\text{S24})$$

Let's consider when ρ_{in} is the ideal input state. Say non-Clifford gate $R_Z(\theta)$ is transversal for code $\mathcal{Q}$. The ideal input state can be written as

$$|\theta\rangle^{\otimes n} = R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n}. \quad (\text{S25})$$

The effect of codespace projector $\bar{P}_0$ on $|\theta\rangle^{\otimes n}$ is:

$$\bar{P}_0 |\theta\rangle^{\otimes n} = \bar{P}_0 R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n} = R_Z(\theta)^{\otimes n} \bar{P}_0 |+\rangle^{\otimes n} \propto \bar{R}_Z(\theta) |\bar{+}\rangle = |\bar{\theta}\rangle. \quad (\text{S26})$$

Therefore, the decoded state will be exactly $|\theta\rangle$, and the first term on in Equation (S24) is proportional to $|\theta\rangle \langle \theta|$.

As the next step, we show the effect of $\bar{P}_{\mathbf{x}}$ on $|\theta\rangle^{\otimes n}$ for $|\mathbf{x}| = 1$. Since the code is CSS code, the stabilizer generators should be either X -type or Z -type. If the subspace projector $\bar{P}_{g_{\mathbf{x}}}$ has syndrome -1 for an X -type generator $g_{\mathbf{x}}$ and +1 for all other stabilizers, then based on Lemma 2, we can always find a subspace-transfer operator $\tilde{Z}$ composed with only Pauli Z that maps the codespace to this syndrome subspace without affecting the logical state, i.e., $\bar{P}_{g_{\mathbf{x}}} = \tilde{Z} \bar{P} \tilde{Z}$. Notably, both $\bar{P}_{g_{\mathbf{x}}}$ and $\tilde{Z}$ depend on the choice of $g_{\mathbf{x}}$. However, $\bar{P}_{g_{\mathbf{x}}} |\theta\rangle^{\otimes n} = 0$. To see this,

$$\begin{aligned} \bar{P}_{g_{\mathbf{x}}} |\theta\rangle^{\otimes n} &= \tilde{Z} \bar{P} \tilde{Z} R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n} \\ &= \tilde{Z} \bar{P} R_Z(\theta)^{\otimes n} \tilde{Z} |+\rangle^{\otimes n} \\ &= \tilde{Z} R_Z(\theta)^{\otimes n} \bar{P} \tilde{Z} |+\rangle^{\otimes n} = 0, \end{aligned} \quad (\text{S27})$$

where we used Lemma 1 from the second row to the last row. The $\tilde{Z} |+\rangle^{\otimes n}$ is a tensor product state of $|+\rangle$ and $|-\rangle$, while $\bar{P}$ contains projectors for every X -type stabilizer. As $\tilde{Z}$ should not be a logical operator, the state $\tilde{Z} |+\rangle^{\otimes n}$ will be anti-stabilized by at least a X -type stabilizer, and $\bar{P} \tilde{Z} |+\rangle^{\otimes n} = 0$. Therefore, $\bar{P}_{\mathbf{x}}(|\theta\rangle \langle \theta|)^{\otimes n} \bar{P}_{\mathbf{x}} = 0$ for $|\mathbf{x}| = 1$ if the flipped stabilizer is X -type.

Now we consider the case when $\bar{P}_{\mathbf{x}}$ has syndrome -1 for an arbitrary Z -type stabilizer $g_{\mathbf{z}}$. We denote the corresponding subspace projector by $\bar{P}_{g_{\mathbf{z}}}$, and $\tilde{X}$ is the corresponding subspace-transfer operator between $\bar{P}_{g_{\mathbf{z}}}$ and $\bar{P}$ without affecting the logical state such that $\bar{P}_{g_{\mathbf{z}}} = \tilde{X} \bar{P} \tilde{X}$. We define the support of $\tilde{X}$ as $\{m_i\}$. To calculate $\bar{P}_{g_{\mathbf{z}}} |\theta\rangle^{\otimes n}$, we define Z_{m_i} as the Pauli operator Z and R_{Z,m_i} as the Pauli Z rotation gate on qubit m_i :

$$\begin{aligned} \bar{P}_{g_{\mathbf{z}}} |\theta\rangle^{\otimes n} &= \tilde{X} \bar{P} \tilde{X} R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n} \\ &= \tilde{X} \bar{P} \left(\bigotimes_{\{m_i\}} R_{Z,m_i}(-2\theta) \right) R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n} \\ &= \tilde{X} \bar{P} \left(\bigotimes_{\{m_i\}} (\cos \theta I + i \sin \theta Z_{m_i}) R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n} \right) \\ &\propto \tilde{X} \bar{P} |\theta\rangle^{\otimes n} \end{aligned} \quad (\text{S28})$$

where we used the fact that $X R_Z(\theta) = R_Z(-\theta) X$. From the second last row to the last row, we explicitly expanded $\bigotimes_{\{m_i\}} (\cos \theta I + i \sin \theta Z_{m_i})$ into sum of terms with pure Z -type Pauli operators. These who commute with every stabilizer will be either logical I or Z operators, but Lemma 4 prevents them to accumulate into logical Z . These that don't commute with every stabilizer will get eliminated by the codespace projector $\bar{P}$ as we just showed in Equation S27. Notice that $\tilde{X}$ must commute with the logical X operator of $\mathcal{Q}$, the effect of $\tilde{X}$ on the logical states after decoding can only be either a logical X error if it anticommutes with the logical Z operators or identity in the other case. Therefore, $\text{Tr}_A[D_{\mathcal{Q}} \bar{P}_{\mathbf{x}}(|\theta\rangle \langle \theta|)^{\otimes n} \bar{P}_{\mathbf{x}} D_{\mathcal{Q}}^\dagger]$ should be either $X |\theta\rangle \langle \theta| X$ or $|\theta\rangle \langle \theta|$. Summarizing all up, when ρ_{in} is the ideal input state, the output state ρ_o is given by

$$\rho_o(|\theta\rangle \langle \theta|) \propto p_0 |\theta\rangle \langle \theta| + \sum_{\{g_{\mathbf{z}}\}} e^{-2\beta} p_{g_{\mathbf{z}}} X |\theta\rangle \langle \theta| X + \sum_{\{g'_{\mathbf{z}}\}} e^{-2\beta} p_{g'_{\mathbf{z}}} |\theta\rangle \langle \theta| + O(e^{-4\beta}), \quad (\text{S29})$$

where $\{g_{\mathbf{z}}\}(\{g'_{\mathbf{z}}\})$ is the set of Z -type generators whose subspace-transfer operators anti-commute (commute) with logical Z . $p_0 = \text{Tr}[\bar{P}_0(|\theta\rangle \langle \theta|)^{\otimes n}]$ and $p_{g_{\mathbf{z}}} = \text{Tr}[\bar{P}_{g_{\mathbf{z}}}(|\theta\rangle \langle \theta|)^{\otimes n}]$ and $\bar{P}_{g_{\mathbf{z}}}$ is the subspace projector that is stabilized by every generator but antistabilized by $g_{\mathbf{z}}$.

Now we wanna find ρ_* such that $\rho_o(\rho_*) = \rho_*$. As we only have infidelity contribution from Z -type generator measurement that could lead to logical X noise, we may set $\rho_* = (1 - \alpha_*) |\theta\rangle \langle \theta| + \alpha_* X |\theta\rangle \langle \theta| X$, and α_* should be small when β is large. We can also consider the leading order of ρ_{in} :

$$\rho_{\text{in}} \approx (1 - n\alpha_*) (|\theta\rangle \langle \theta|)^{\otimes n} + \alpha_* \sum_{i=1}^n X_i (|\theta\rangle \langle \theta|)^{\otimes n} X_i. \quad (\text{S30})$$

Combine Equation (S30) and (S24) we have:

$$\begin{aligned} \rho_o(\rho_*) &\propto p_0^* |\theta\rangle \langle \theta| + (1 - n\alpha_*) \sum_{|\mathbf{x}|=1} e^{-2\beta} \text{Tr}_A[D_Q \bar{P}_\mathbf{x} |\theta\rangle \langle \theta| \bar{P}_\mathbf{x} D_Q^\dagger] + \alpha_* \sum_{|\mathbf{x}|=1} e^{-2\beta} \text{Tr}_A[D_Q \bar{P}_\mathbf{x} X_i |\theta\rangle \langle \theta| X_i \bar{P}_\mathbf{x} D_Q^\dagger] \\ &\approx p_0^* |\theta\rangle \langle \theta| + (1 - n\alpha_*) \left(\sum_{g_z} e^{-2\beta} p_{g_z} X |\theta\rangle \langle \theta| X + \sum_{g'_z} e^{-2\beta} p_{g'_z} |\theta\rangle \langle \theta| \right). \end{aligned} \quad (\text{S31})$$

As $\rho_o(\rho_*) = \rho_*$:

$$\frac{\alpha_*}{1 - \alpha_*} = \frac{(1 - n\alpha_*) \sum_{g_z} e^{-2\beta} p_{g_z}}{p_0^* + (1 - n\alpha_*) \sum_{g'_z} e^{-2\beta} p_{g'_z}}. \quad (\text{S32})$$

Define $A = \sum_{g_z} p_{g_z}$ and $B = \sum_{g'_z} p_{g'_z}$ for simplicity, from which we have

$$\alpha_* \approx \frac{e^{-2\beta} A}{p_0^* + (n+1)Ae^{-2\beta} + Be^{-2\beta}} \propto e^{-2\beta} \quad (\text{S33})$$

Therefore, the target state ρ_* is at most under a Pauli X noise with strength scales with $e^{-2\beta}$ if $A \neq 0$. $\square$

Theorem 2. For non-trivial MSD protocols based on $[[n, k, d]]$ ($d \geq 2$) CSS codes, the distillation convergence rate under finite weak measurement strength is linear.

Proof key: Assume the input state to be all the noisy target unless one with epsilon deviation. The final term in Eq (51) would then be easy to prove to be non-zero.

Proof. Again, we may assume the Gaussian noise to clearly denote the noise order. For simplicity we define a linear quantum map from n -qubit space to k -qubit space

$$\mathcal{F}_\mathbf{x}(\cdot) = \text{Tr}_A[D_Q \bar{P}_\mathbf{x} \cdot \bar{P}_\mathbf{x} D_Q^\dagger]. \quad (\text{S34})$$

We define another map $\mathcal{D}$ for MSD protocols based on $\mathcal{Q}$ under measurement strength β . Therefore, from Equation (S23) we have:

$$\rho_o = \mathcal{D}(\rho_i^{\otimes n}) = \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_\mathbf{x}(\rho_i^{\otimes n}) / \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_i^{\otimes n} P_\mathbf{x}] \propto \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_\mathbf{x}(\rho_i^{\otimes n}) \quad (\text{S35})$$

Without loss of generality we consider $k = 1$ first. Consider a fixed relatively large $\beta \gg 1$ with fixed point ρ_* , such that $\rho_* = \mathcal{D}(\rho_*)$. Then we can always write ρ_* in the following form:

$$\rho_* = (1 - m_x - m_y) |\theta\rangle \langle \theta| + m_x X |\theta\rangle \langle \theta| X + m_y Y |\theta\rangle \langle \theta| Y, \quad (\text{S36})$$

where $m_x \propto e^{-2q_x\beta}$ and $m_y \propto e^{-2q_y\beta}$. We assume the purity of ρ_* to be δ . Without loss of generality we assume $q_x \leq q_y$, so $q = q_x$. Now consider an input state ρ_ϵ such that $\|\rho_* - \rho_\epsilon\| \propto \epsilon$, we may assume

$$\rho_\epsilon = (1 - \epsilon)\rho_* + \epsilon(u_x X \rho_* X + u_y Y \rho_* Y + u_z Z \rho_* Z) \quad (\text{S37})$$

Define a channel for n -qubit states $\mathcal{N}(\rho) = \frac{1}{3n} \sum_{E \in \mathcal{P}_1} u_E E \rho E$ where $\mathcal{P}_1$ is the set for all non-trivial single-weight Pauli operators on n qubits. Therefore, $\rho_\epsilon^{\otimes n} \approx (1 - n\epsilon)\rho_*^{\otimes n} + n\epsilon\mathcal{N}(\rho_*^{\otimes n})$. Define $\tilde{\epsilon} = n\epsilon$:

$$\mathcal{D}(\rho_\epsilon^{\otimes n}) = \frac{(1 - \tilde{\epsilon}) \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_\mathbf{x}(\rho_*^{\otimes n}) + \tilde{\epsilon} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_\mathbf{x}(\mathcal{N}(\rho_*^{\otimes n}))}{(1 - \tilde{\epsilon}) \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_*^{\otimes n} P_\mathbf{x}] + \tilde{\epsilon} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\mathcal{N}(\rho_*^{\otimes n}) P_\mathbf{x}]} \quad (\text{S38})$$

$$\begin{aligned}
\|\mathcal{D}(\rho_\epsilon^{\otimes n}) - \mathcal{D}(\rho_*^{\otimes n})\| &= \left\| \frac{(1-\tilde{\epsilon}) \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\rho_*^{\otimes n}) + \tilde{\epsilon} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\mathcal{N}(\rho_*^{\otimes n}))}{(1-\tilde{\epsilon}) \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_*^{\otimes n} P_{\mathbf{x}}] + \tilde{\epsilon} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\mathcal{N}(\rho_*^{\otimes n}) P_{\mathbf{x}}]} - \frac{\sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\rho_*^{\otimes n})}{\sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_*^{\otimes n} P_{\mathbf{x}}]} \right\| \\
&\propto \left\| \frac{(1-\tilde{\epsilon}) \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\rho_*^{\otimes n}) + \tilde{\epsilon} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\mathcal{N}(\rho_*^{\otimes n}))}{(1-\tilde{\epsilon}) + \tilde{\epsilon} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\mathcal{N}(\rho_*^{\otimes n}) P_{\mathbf{x}}] / \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_*^{\otimes n} P_{\mathbf{x}}]} - \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\rho_*^{\otimes n}) \right\| \\
&= \tilde{\epsilon} \left\| - \frac{\sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\mathcal{N}(\rho_*^{\otimes n}) P_{\mathbf{x}}]}{\sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_*^{\otimes n} P_{\mathbf{x}}]} \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\rho_*^{\otimes n}) + \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(\mathcal{N}(\rho_*^{\otimes n})) \right\| + O(\tilde{\epsilon}^2) \\
&\propto \tilde{\epsilon} \|\mathcal{D}(\rho_*^{\otimes n}) - \mathcal{D}(\mathcal{N}(\rho_*^{\otimes n}))\| \quad (\text{if } \sum_{\mathbf{x}} e^{-2|\mathbf{x}|\beta} \text{Tr}[\mathcal{N}(\rho_*^{\otimes n}) P_{\mathbf{x}}] \neq 0) \\
&= \tilde{\epsilon} \left\| \rho_* - \frac{\sum_{\mathbf{x}} \sum_{E \in \mathcal{P}_1} u_E e^{-2|\mathbf{x}|\beta} F_{\mathbf{x}}(E \rho_*^{\otimes n} E)}{\sum_{\mathbf{x}} \sum_{E \in \mathcal{P}_1} u_E e^{-2|\mathbf{x}|\beta} \text{Tr}[\rho_*^{\otimes n} E P_{\mathbf{x}} E]} \right\| \\
&\propto \tilde{\epsilon} \left\| \sum_{\mathbf{x}} \sum_{E \in \mathcal{P}_1} u_E e^{-2|\mathbf{x}|\beta} [\rho_* * \text{Tr}[\rho_*^{\otimes n} E P_{\mathbf{x}} E] - F_{\mathbf{x}}(E \rho_*^{\otimes n} E)] \right\|
\end{aligned} \tag{S39}$$

Now we prove by contradiction: If the convergence rate is beyond linear order, the prefactor in the last row in (S39) must be zero. Notice that the convergence doesn't depend on specific direction, the prefactor must also be zero for all potential combination of u_E in the channel $\mathcal{N}$. Therefore, the prefactor should also be zero for any single Pauli error $E \in \mathcal{P}_1$. Besides, as we are in the large β regime, We only require the coefficient for the same order of $e^{-2\beta}$ to be zero. Based on these reasons, we must require

$$\left\| \sum_{|\mathbf{x}|=j} [\rho_* * \text{Tr}[\rho_*^{\otimes n} E P_{\mathbf{x}} E] - F_{\mathbf{x}}(E \rho_*^{\otimes n} E)] \right\| = 0, \quad \forall E \in \mathcal{P}_1, \quad 0 \leq j \leq \bar{n}. \tag{S40}$$

Notice that $F_{\mathbf{x}}(E \rho_*^{\otimes n} E) = C_E F_{\mathbf{y}(\mathbf{x}, E)}(\rho_*^{\otimes n}) C_E$, where C_E is a single-qubit Pauli associated with E_i :

$$C_E = \begin{cases} X & \text{if } [E, \bar{Z}] \neq 0, [E, \bar{X}] = 0 \\ Z & \text{if } [E, \bar{Z}] = 0, [E, \bar{X}] \neq 0 \\ Y & \text{if } [E, \bar{Z}] \neq 0, [E, \bar{X}] \neq 0 \\ I & \text{other cases} \end{cases} \tag{S41}$$

and $\mathbf{y}$ is another binary vector associated with $\mathbf{x}$ and E which strictly follows $\mathbf{y}(\mathbf{x}, E_i) \neq \mathbf{x}$ such that

$$(\mathbf{y}(\mathbf{x}, E) \oplus \mathbf{x})_k = \begin{cases} 0 & \text{if } [E, g_k] = 0 \\ 1 & \text{if } [E, g_k] \neq 0 \end{cases}. \tag{S42}$$

Therefore, the Eq (40) would be

$$\left\| \sum_{|\mathbf{x}|=j} [\rho_* * \text{Tr}[\rho_*^{\otimes n} P_{\mathbf{y}(\mathbf{x}, E)}] - C_E F_{\mathbf{y}(\mathbf{x}, E)}(\rho_*^{\otimes n}) C_E] \right\| = 0, \quad \forall E \in \mathcal{P}_1, \quad 0 \leq j \leq \bar{n}. \tag{S43}$$

Only consider $j = 0$: the above equation means the distillation protocol with ideal measurements should distill ρ_* at at least quadratic efficiency. However, as the distillation protocol already distills into the pure magic state $|\theta\rangle$ and ρ_* is a close mixed state, this is not possible. Therefore, even the coefficient for $j = 0$ won't be zero. As a result, the distillation under imperfect measurement can only exhibit linear convergence rate. $\square$

Theorem 3. *For MSD protocols based on $[[n, k, d]]$ ($d \geq 2$) CSS codes with transversal non-Clifford gates, we can choose the stabilizer generators in standard form, such that the asymptotically distilled states are robust to weak measurement up to order d .*

Proof. Let's consider the effect of $\bar{P}_{\mathbf{x}}$ with $|\mathbf{x}| \leq d - 1$. Then there exist some subspace-transfer operators $C_{\mathbf{x}}$ such that $\bar{P}_{\mathbf{x}} = C_{\mathbf{x}} \bar{P}_0 C_{\mathbf{x}}$. Importantly, $C_{\mathbf{x}} = \prod_{\{i: x_i \neq 0\}} C_i$ is a product of subspace-transfer operators for $|\mathbf{x}| = 1$. From Lemma 3 we see all C_i is single-weight and have different support. Therefore, $C_{\mathbf{x}}$ has weight smaller than d , and

$\bar{P}_0 C_{\mathbf{x}} |\bar{\psi}\rangle \propto |\bar{\psi}\rangle$ for any logical state $|\bar{\psi}\rangle$. Without loss of generality we might assume $C_{\mathbf{x}}$ on the first m qubits are Pauli X . Let's then consider the effect of $P_{\mathbf{x}}$ on $|\theta\rangle^{\otimes n}$:

$$\begin{aligned}
\bar{P}_{\mathbf{x}} |\theta\rangle^{\otimes n} &= C_{\mathbf{x}} \bar{P}_0 C_{\mathbf{x}} R_Z(\theta)^{\otimes n} |+\rangle^{\otimes n} \\
&= C_{\mathbf{x}} \bar{P}_0 R_Z(\theta)^{\otimes n} \bigotimes_{j=1}^m R_{Z,j}(-2\theta) C_{\mathbf{x}} |+\rangle^{\otimes n} \\
&= C_{\mathbf{x}} R_Z(\theta)^{\otimes n} \bar{P}_0 \left(\bigotimes_{j=1}^m R_{Z,j}(-2\theta) \right) \bar{C} |+\rangle^{\otimes n} \\
&\propto C_{\mathbf{x}} R_Z(\theta)^{\otimes n} \bar{P}_0 |+\rangle^{\otimes n} = C_{\mathbf{x}} \bar{P}_0 |\theta\rangle^{\otimes n}.
\end{aligned} \tag{S44}$$

From Lemma 3, we also know that $C_{\mathbf{x}}$ must commute with all logical operators. Therefore, it will act as if a logical identity and $C_{\mathbf{x}} \bar{P}_0 |\theta\rangle^{\otimes n}$ will give exactly the same state as $\bar{P}_0 |\theta\rangle^{\otimes n}$ after decoding.

Combining (S23), we have

$$\rho_0(|\theta\rangle\langle\theta|) \propto p_0 |\theta\rangle\langle\theta| + \sum_{|\mathbf{x}| \geq d} e^{-2|\mathbf{x}|\beta} \text{Tr}_A[D_Q \bar{P}_{\mathbf{x}} \rho_{\text{in}} \bar{P}_{\mathbf{x}} D_Q^\dagger] \tag{S45}$$

Therefore, we only have infidelity contribution with lead order $e^{-2d\beta}$. $\square$

VI. SUPPLEMENTARY NOTE 6: EIGENVALUE OF JACOBIAN MATRIX AT FIXED POINT

As shown in Figure S3, we calculate the eigenvalue of the Jacobian matrix at the non-trivial fixed points for the simulated MSD protocols. The largest eigenvalue increases when measurement strength decreases. When the largest eigenvalue reaches to 1, the fixed point is no longer stable and will disappear, which corresponds to the threshold of measurement strength.

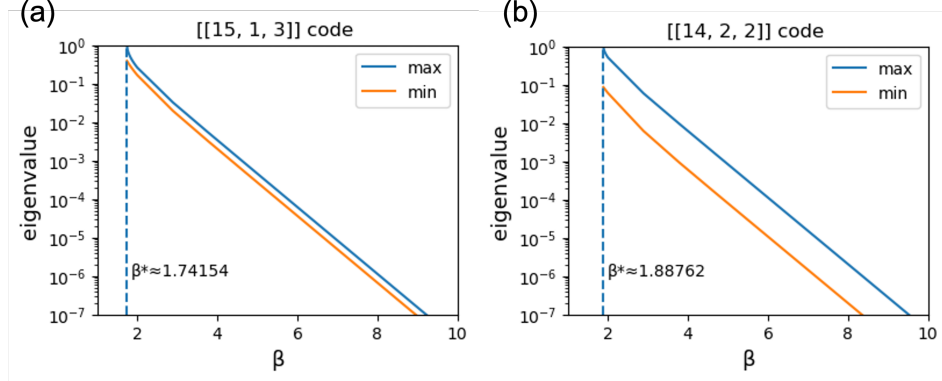


FIG. S3. **The Jacobian eigenvalue identifies the non-trivial fixed-point stability of the two example MSD protocols.** Results are shown for (a) the $[[15, 1, 3]]$ code and (b) the $[[14, 2, 2]]$ code, using the canonical generator sets.

VII. SUPPLEMENTARY NOTE 7: OTHER CHOICE OF STABILIZER GENERATORS

In this section, we again provide some examples that show that we can actually engineer the biased Pauli noise by switch the type of stabilizer generators. For the $[[15, 1, 3]]$ code, $H_X \subset H_Z$ (see Sec II). Therefore, we can modify all weight-8 Z -type generators to Y -type and with the codespace intact. In this generator choice, both the X -type and Y -type noise are first-order as shown in Figure S4(a,d). We can also replace all the X -type generators with Y -type generators that share the same support and the codespace is still exactly the same, which make us able to engineer the leading order biased noise in the output state from logical Pauli X to Pauli Y compared to the canonical generator choice as shown in Figure S4(b,e). Notably, the two examples don't conflict with our Theorem 1 as now the code is

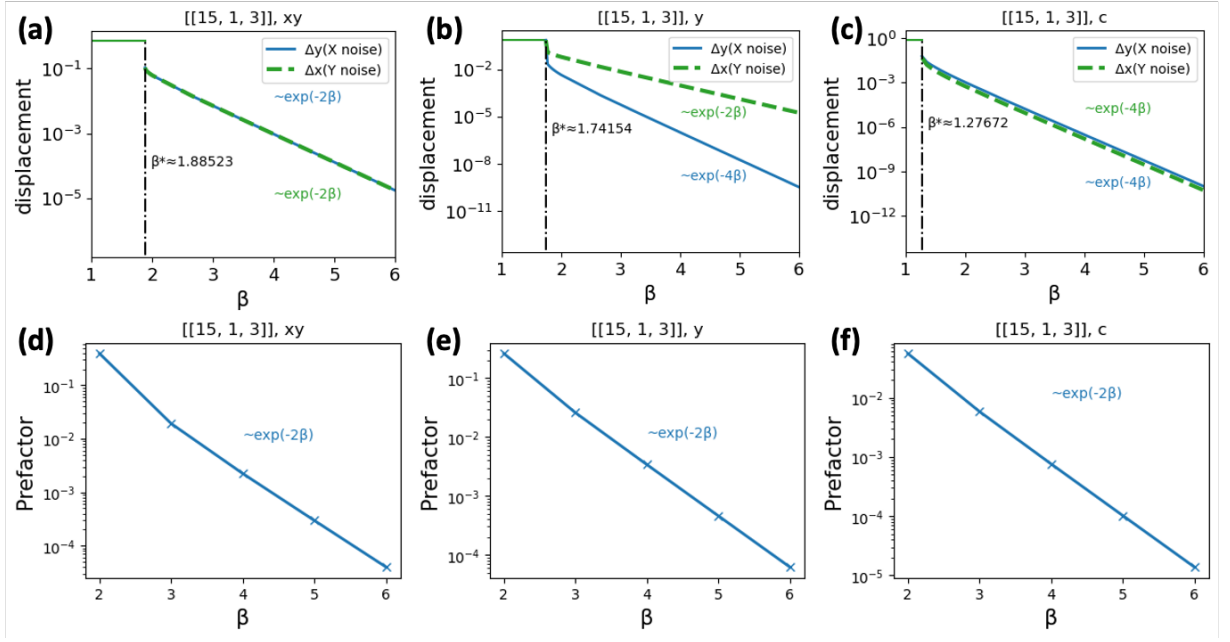


FIG. S4. **Different stabilizer-generator choices alter the performance of the $[[15, 1, 3]]$ protocol.** Panels (a) and (d) show the output-state displacement and corresponding prefactor for the XY -symmetric choice. Panels (b) and (e) show the same quantities after replacing all X -type generators with Y -type generators. Panels (c) and (f) show the results for the modified standard form.

no longer strictly CSS since Y -type generators are introduced. Further, we might consider a modified standard-form generator choice, such that H_Z is formed as:

$$H_Z = \begin{bmatrix} 0 & 0 & 0 \\ D' & I' & E' \end{bmatrix} \quad (\text{S46})$$

where I' is an upper triangular matrix with all upper element being one. This H_Z can be obtained by performing the reverse of gaussian elimination. This generator choice will give us performance as shown in Figure S4(c,f).

-
- [1] Y. Zheng and D. E. Liu, From Magic State Distillation to Dynamical Systems (2024), arXiv:2412.04402.
 - [2] J. T. Anderson and T. Jochym-O'Connor, Classification of transversal gates in qubit stabilizer codes (2014), arXiv:1409.8320 [math-ph, physics:quant-ph].
 - [3] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Physical Review A* **71**, 022316 (2005).
 - [4] S. Bravyi and J. Haah, Magic-state distillation with low overhead, *Physical Review A* **86**, 052329 (2012).
 - [5] B. Zeng, A. Cross, and I. L. Chuang, Transversality versus Universality for Additive Quantum Codes (2007), arXiv:0706.1382 [quant-ph].
 - [6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge university press, 2010).
 - [7] D. Gottesman, Stabilizer Codes and Quantum Error Correction (1997), arXiv:quant-ph/9705052.