Supplemental Information: “Robust exponential memory in Hopfield networks”

Christopher Hillar*
Redwood Center for Theoretical Neuroscience, Berkeley, CA 94720

Ngoc M. Tran*
University of Texas, Austin, TX†

In this supplementary material, we elaborate on the mathematics involved in the claims of the main paper.

*chillar@msri.org
†tran.mai.ngoc@gmail.com
I. SYMMETRIC 3-PARAMETER \((x, y, z)\) NETWORKS

The first step of our construction is to exploit symmetry in the following set of linear inequalities:

\[ E_c - E_{c'} < 0, \quad (8) \]

where \(c\) runs over \(k\)-cliques and \(c'\) over vectors differing from \(c\) by a single bit flip. The space of solutions to (8) is the convex polyhedral cone of networks having each clique as a strict local minimum of the energy function, and thus a fixed-point of the dynamics.

The permutations \(P \in P_V\) of the vertices \(V\) act on a network by permuting the rows/columns of the weight matrix \((J \mapsto PJP^T)\) and thresholds \((\theta \mapsto P\theta)\), and this action on a network satisfying property (8) preserves that property. Consider the average \((\bar{J}, \bar{\theta})\) of a network over the group \(P_V\):

\[ \bar{J} = \frac{1}{v!} \sum_{P \in P_V} PJP^T, \quad \bar{\theta} = \frac{1}{v!} \sum_{P \in P_V} P\theta, \]

and note that if \((J, \theta)\) satisfies (8) then so does the highly symmetric object \((\bar{J}, \bar{\theta})\). To characterize \((\bar{J}, \bar{\theta})\), observe that \(P\bar{J}P^T = \bar{J}\) and \(P\bar{\theta} = \bar{\theta}\) for all \(P \in P_V\).

These strong symmetries imply there are \(x, y, z\) such that \(\bar{\theta} = (z, \ldots, z) \in \mathbb{R}^n\) and for each pair \(e \neq f\) of all possible edges:

\[ \bar{J}_{ef} := \begin{cases} x & \text{if } |e \cap f| = 1 \\ y & \text{if } |e \cap f| = 0, \end{cases} \]

where \(|e \cap f|\) is the number of vertices that \(e\) and \(f\) share.

Our next demonstration is an exact setting for weights in these Hopfield networks.

II. EXPONENTIAL STORAGE

For an integer \(r \geq 0\), we say that state \(x^*\) is \(r\)-stable if it is an attractor for all states with Hamming distance at most \(r\) from \(x^*\). Thus, if a state \(x^*\) is \(r\)-stably stored, the network is guaranteed to converge to \(x^*\) when exposed to any corrupted version not more than \(r\) bit flips away.

For positive integers \(k\) and \(r\), is there a Hopfield network on \(n = \binom{2k}{2}\) nodes storing all \(k\)-cliques \(r\)-stably? We necessarily have \(r \leq \lfloor k/2 \rfloor\), since \(2(\lfloor k/2 \rfloor + 1)\) is greater than or equal to the Hamming distance between two \(k\)-cliques that share a \((k - 1)\)-subclique. In fact, for any \(k > 3\), this upper bound is achievable by a sparsely-connected three-parameter network.
Proposition 1. There exists a family of three-parameter Hopfield networks with $z = 1$, $y = 0$ storing all $k$-cliques as $\lfloor k/2 \rfloor$-stable states.

The proof relies on the following lemma, which gives the precise condition for the three-parameter Hopfield network to store $k$-cliques as $r$-stable states for fixed $r$.

Lemma 1. Fix $k > 3$ and $0 \leq r < k$. The Hopfield network $(J(x, y), \theta(z))$ stores all $k$-cliques as $r$-stable states if and only if the parameters $x, y, z \in \mathbb{R}$ satisfy

$$M \cdot \begin{bmatrix} x \\ y \end{bmatrix} < \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix} z,$$

where

$$M = \begin{bmatrix} 4(2 - k) + 2r & (2 - k)(k - 3) \\ 4(2 - k) & (2 - k)(k - 3) - 2r \\ 2(k - 1) + 2r & (k - 1)(k - 2) \\ 2(k - 1) & (k - 1)(k - 2) - 2r \end{bmatrix}.$$

Furthermore, a pattern within Hamming distance $r$ of a $k$-clique converges after one iteration of dynamics.

Proof: For fixed $r$ and $k$-clique $x$, there are $2^r$ possible patterns within Hamming distance $r$ of $x$. Each of these patterns defines a pair of linear inequalities on the parameters $x, y, z$. However, only the inequalities from the following two extreme cases are active constraints. All the other inequalities are convex combinations of these.

1. $r$ edges in the clique with a common node $i$ are removed.

2. $r$ edges are added to a node $i$ not in the clique.

In the first case, there are two types of edges at risk of being mislabeled. The first are those of the form $ij$ for all nodes $j$ in the clique. Such an edge has $2(k - 2) - r$ neighbors and $\binom{k-2}{2}$ non-neighbors. Thus, each such edge will correctly be labeled as 1 after one network update if and only if $x, y, z$ satisfy

$$2(2k - r - 4)x + (k - 2)(k - 3)y > 2z.$$  \hspace{1cm} (9)

The other type are those of the form $\bar{i}j$ for all nodes $\bar{i} \neq i$ in the clique, and $j$ not in the clique. Assuming $r < k - 1$, such an edge has at most $k - 1$ neighbors and $\binom{k-1}{2} - r$
non-neighbors. Thus, each such edge will be correctly labeled as 0 if and only if
\[
2(k - 1)x + ((k - 1)(k - 2) - 2r)y < 2z. \tag{10}
\]
Rearranging equations (9) and (10) yield the first two rows of the matrix in the lemma. A similar argument applies for the second case, giving the last two inequalities.

From the derivation, it follows that if a pattern is within Hamming distance \( r \) of a \( k \)-clique, then all spurious edges are immediately deleted by case 1, all missing edges are immediately added by case 2, and thus the clique is recovered in precisely one iteration of the network dynamics. ■

Proof of Proposition 1: The matrix inequalities in Lemma 1 define a cone in \( \mathbb{R}^3 \), and the cases \( z = 1 \) or \( z = 0 \) correspond to two separate components of this cone. For the proof of Theorem 1 in the main article, we shall use the cone with \( z = 1 \). We further assume \( y = 0 \) to achieve a sparsely-connected matrix \( J \). In this case, the second and fourth constraints are dominated by the first and third. Thus, we need \( x \) that solves:
\[
\frac{1}{2(k - 1) - r} < x < \frac{1}{k - 1 + r}.
\]
There exists such a solution if and only if
\[
2(k - 1) - r > k - 1 + r \iff k > 2r + 1. \tag{11}
\]
The above equation is feasible if and only if \( r \leq \lfloor k/2 \rfloor \). ■

III. PROOFS OF THEOREMS 1, 2

Fix \( y = 0 \) and \( z = 1 \). We now tune \( x \) such that asymptotically the \( \alpha \)-robustness of our set of Hopfield networks storing \( k \)-cliques tends to 1/2 as \( n \to \infty \). By symmetry, it is sufficient to prove robustness for one fixed \( k \)-clique \( x \), say, the one with vertices \( \{1, 2, \ldots, k\} \). For \( 0 < p < 1 \), let \( x_p \) be the \( p \)-corruption of \( x \). For each node \( i \in \{1, \ldots, 2k\} \), let \( i_{in}, i_{out} \) denote the number of edges from \( i \) to other clique and non-clique nodes, respectively. With an abuse of notation, we write \( i \in x \) to mean a vertex \( i \) in the clique, that is, \( i \in \{1, \ldots, k\} \).

We need the following inequality originally due to Bernstein [24].

Lemma 2 (Bernstein’s inequality). Let \( S_i \) be independent Bernoulli random variables taking values +1 and −1 each with probability 1/2. For any \( \epsilon > 0 \), the following holds:
\[
P \left( \frac{1}{n} \sum_{i=1}^{n} S_i > \epsilon \right) \leq \exp \left( -\frac{n\epsilon^2}{2 + 2\epsilon/3} \right).
\]
The following fact is a fairly direct consequence of Lemma 2.

**Lemma 3.** Let $Y$ be an $n \times n$ symmetric matrix with zero diagonal, $Y_{ij} \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$. For each $i = 1, \ldots, n$, let $Y_i = \sum_j Y_{ij}$ be the $i$-th row sum. Let $M_n = \max_{1 \leq i \leq n} Y_i$, and $m_n = \min_{1 \leq i \leq n} Y_i$. Then for any constant $c > 0$, as $n \to \infty$, we have:

$$\mathbb{P}(|m_n - np| > c\sqrt{n \ln n}) \to 0$$

and

$$\mathbb{P}(|M_n - np| > c\sqrt{n \ln n}) \to 0.$$  

In particular, $|m_n - np|, |M_n - np| = o(\sqrt{n \ln n})$.

**Proof:** Fix $c > 0$. As a direct corollary of Bernstein’s inequality, for each $i$ and for any $\epsilon > 0$, we have

$$\mathbb{P}(Y_i - np > n\epsilon - (p + \epsilon)) \leq \exp\left(-\frac{(n - 1)\epsilon^2}{2 + 2\epsilon/3}\right).$$

It follows that

$$\mathbb{P}(Y_i - np > n\epsilon) \leq \exp\left(-\frac{n\epsilon^2}{4 + 4\epsilon/3}\right),$$

and thus from a union bound with $\epsilon = \frac{c\ln n}{\sqrt{n}}$, we have

$$\mathbb{P}(\max_i Y_i - np > c\sqrt{n \ln n}) \leq \exp\left(-\frac{n\epsilon^2}{4 + 4\epsilon/3} + \ln n\right) \leq \exp\left(-\frac{c^2 \ln^2 n}{4 + 4c} + \ln n\right).$$

Since this last bound converges to 0 with $n \to \infty$, we have proved the claim for $M_n$. Since $Y_i$ is symmetric about $np$, a similar inequality holds for $m_n$. □

**Corollary 1.** Let $M_{in} = \max_{i \in X} i_{in}$, $m_{in} = \min_{i \in X} i_{in}$, $M_{out} = \max_{i \in X} i_{out}$, $m_{out} = \min_{i \in X} i_{out}$, and $M_{between} = \max_{i \in X} i_{in}$. Then $M_{in} - k(1 - p)$, $m_{in} - k(1 - p)$, $M_{out} - kp$, $m_{out} - kp$, and $M_{between} - kp$ are all of order $o(\sqrt{k \ln k})$ as $k \to \infty$ almost surely.

**Proofs of Theorem 1, 2 (robustness):** Let $N(e)$ be the number of neighbors of edge $e$. For each $e$ in the clique:

$$N(e) \geq 2m_{in} + 2m_{out} \sim 2k + o(\sqrt{k \ln k}) \quad \text{w.h.p.}$$

To guarantee that all edges $e$ in the clique are labeled 1 after one dynamics update, we need

$$x > \frac{1}{N(e)}; \quad \text{that is},$$

$$x > \frac{1}{2k + o(\sqrt{k \ln k})}. \quad (12)$$
If $f$ is an edge with exactly one clique vertex, then

$$N(f) \leq M_{in} + M_{out} + 2M_{between}$$

$$\sim k(1 + 2p) + o(\sqrt{k \ln k}) \quad \text{w.h.p.}$$

To guarantee that $x_f = 0$ for all such edges $f$ after one iteration of the dynamics, we need $x < \frac{1}{N(f)}$; that is,

$$x < \frac{1}{k(1 + 2p) + o(\sqrt{k \ln k})}. \quad (13)$$

In particular, if $p = p(k) \sim \frac{1}{2} - k^{\delta-1/2}$ for some small $\delta \in (0,1/2)$, then taking $x = x(k) = \frac{1}{2}(\frac{1}{k(1 + 2p)} + \frac{1}{2k})$ would guarantee that for large $k$, both equations (12) and (13) are simultaneous satisfied. In this case, $\lim_{k \to \infty} p(k) = 1/2$, and thus the family of two-parameter Hopfield networks with $x(k), y = 0, z = 1$ has robustness index $\alpha = 1/2$. ■

**IV. CLIQUE RANGE STORAGE**

In this section, we give precise conditions for the existence of a Hopfield network on $\binom{v}{2}$ nodes that stores all $k$-cliques for $k$ in an interval $[m, M]$, $m \leq M \leq v$. We do not address the issue of robustness as the qualitative trade-off is clear: the more memories the network is required to store, the less robust it is. The trade-off can be analyzed by large deviation principles as in Theorem 2.

**Theorem 3.** Fix $m$ such that $3 \leq m < v$. For $M \geq m$, there exists a Hopfield network on $\binom{v}{2}$ nodes which stores all $k$-cliques in the range $[m, M]$ if and only if $M$ solves the implicit equation $x_M - x_m < 0$, where

$$x_m = -\frac{(4m - \sqrt{12m^2 - 52m + 57} - 7)}{2(m^2 - m - 2)},$$

$$x_M = -\frac{(4M + \sqrt{12M^2 - 52M + 57} - 7)}{2(M^2 - M - 2)}.$$  

**Proof of Theorem 3:** Fix $z = 1/2$ and $r = 0$ in Lemma 1. (We do not impose the constraint $y = 0$). Then the cone defined by the inequalities in Lemma 1 is in bijection with the polyhedron $\mathcal{I}_k \subseteq \mathbb{R}^2$ cut out by inequalities:

$$4(k - 2)x + (k - 2)(k - 3)y - 1 > 0,$$

$$2(k - 1)x + (k - 1)(k - 2)y - 1 < 0.$$
Let $R_k$ be the line $4(k-2)x + (k-2)(k-3)y - 1 = 0$, and $B_k$ be the line $2(k-1)x + (k-1)(k-2)y - 1 = 0$. By symmetry, there exists a Hopfield network which stores all $k$-cliques in the range $[m, M]$ if and only if $\bigcap_{k=m}^{M} T_k \neq \emptyset$. For a point $P \in \mathbb{R}^2$, write $x(P)$ for its $x$-coordinate. Note that for $k \geq 3$, the points $B_k \cap B_{k+1}$ lie on the following curve $Q$ implicitly parametrized by $k$:

$$Q = \left\{ \left( \frac{1}{k-1}, \frac{-1}{(k-1)(k-2)} \right) : k \geq 3 \right\}.$$ 

When the polytope $\bigcap_{k=m}^{M} T_k$ is nonempty, its vertices are the following points: $R_M \cap R_m$, $R_M \cap B_m$, $B_k \cap B_{k+1}$ for $m \leq k \leq M - 1$, and the points $B_M \cap R_m$. This defines a nonempty convex polytope if and only if

$$x_M := x(Q \cap R_M) < x_m := x(Q \cap R_m).$$

Direct computation gives the formulae for $x_M, x_M$ in Theorem 3. See Fig 6 for a visualization of the constraints of the feasible region. ■

Fixing the number of nodes and optimizing the range $M - m$ in Theorem 3, we obtain the following result.

**Theorem 4.** For large $v$, there is a Hopfield network on $n = \binom{v}{2}$ nodes that stores all $\approx 2^v(1 - e^{-Cv})$ cliques of size $k$ as memories, where $k$ is in the range:

$$m = \frac{1}{D}v \leq k \leq v = M,$$

for constants $C \approx 0.43, D \approx 13.93$. Moreover, this is the largest possible range of $k$ for any Hopfield network.

**Proof of Theorem 4:** From Theorem 3, for large $m, M$ and $v$, we have the approximations

$$x_m \approx \frac{\sqrt{12} - 4}{2m}, \quad x_M \approx \frac{-\sqrt{12} - 4}{2M}.$$ 

Hence $x_M - x_m < 0$ when $M \leq \frac{\sqrt{2} + \sqrt{3}}{2} = Dm$. Asymptotically for large $v$, the most cliques are stored when $M = Dm$ and $[m, M]$ contains $v/2$. Consider $m = \beta v$ so that $v \geq M = D\beta v \geq v/2$, and thus $1/D \geq \beta \geq 1/(2D)$. Next, set $u = v/2 - m = v(1/2 - \beta)$ and $w = M - v/2 = v(D\beta - 1/2)$ so that storing the most cliques becomes the problem of maximizing over admissible $\beta$ the quantity:

$$\max\{u, w\} = \max\{v(1/2 - \beta), v(D\beta - 1/2)\}.$$ 

One can now check that $\beta = 1/D$ gives the best value, producing the range in the statement of the theorem.
Next, note that \( \binom{v}{k} 2^{-v} \) is the fraction of \( k \)-cliques in all cliques on \( v \) vertices, which is also the probability of a \( \text{Binom}(v, 1/2) \) variable equaling \( k \). For large \( v \), approximating this variable with a normal distribution and then using Mill’s ratio to bound its tail c.d.f. \( \Phi \), we see that the proportion of cliques storable tends to

\[
1 - \Phi \left( \frac{D - 1}{D} \sqrt{v} \right) \approx 1 - \exp(-Cv),
\]

for some constant \( C \approx \frac{(D-1)^2}{2D^2} \approx 0.43 \).

\[ \blacksquare \]

V. HOPFIELD-PLATT NETWORKS

We prove the claim in the main text that the Hopfield-Platt network will not robustly store derangements (permutations without fixed-points). For large \( k \), the fraction of permutations that are derangements is known to be \( e^{-1} \approx 0.36 \). Fix a derangement \( \sigma \) on \( k \), represented as a binary vector \( x \) in \( \{0, 1\}^n \) for \( n = k(k-1) \). For each ordered pair \( (i, j) \), \( i \neq j \), \( j \neq \sigma(i) \), we construct a pattern \( y_{ij} \) that differs from \( x \) by exactly two bit flips:

1. Add the edge \( ij \).
2. Remove the edge \( i\sigma(i) \).

There are \( k(k-2) \) such pairs \( (i, j) \), and thus \( k(k-2) \) different patterns \( y_{ij} \). For each such pattern, we flip two more bits to obtain a new permutation \( x^{ij} \) as follows:

1. Remove the edge \( \sigma^{-1}(j)j \).
2. Add the edge \( \sigma^{-1}(j)\sigma(i) \).

It is easy to see that \( x^{ij} \) is a permutation on \( k \) letters with exactly two cycles determined by \( (i, j) \). Call the set of edges modified the critical edges of the pair \( (i, j) \). Note that \( x^{ij} \) are all distinct and have disjoint critical edges.

Each \( y_{ij} \) is exactly two bit flips away from \( x \) and \( x^{ij} \), both permutations on \( k \) letters. Starting from \( y_{ij} \), there is no binary Hopfield network storing all permutations that always correctly recovers the original state. In other words, for a binary Hopfield network, \( y_{ij} \) is an indistinguishable realization of a corrupted version of \( x \) and \( x^{ij} \).
We now prove that for each derangement $x$, with probability at least $1 - (1 - 4p^2)^{n/2}$, its $p$-corruption $x_p$ is indistinguishable from the $p$-corruption of some other permutation. This implies the statement in the main text.

For each pair $(i, j)$ as above, recall that $x_p$ and $x_p^{ij}$ are two random variables in $\{0, 1\}^n$ obtained by flipping each edge of $x$ (resp. $x^{ij}$) independently with probability $p$. We construct a coupling between them as follows: define the random variable $x'_p$ via

- For each non-critical edge, flip this edge on $x'_p$ and $x_p^{ij}$ with the same Bernoulli($p$).
- For each critical edge, flip them on $x'_p$ and $x_p^{ij}$ with independent Bernoulli($p$).

Then $x'_p \overset{d}{=} x_p$ have the same distribution, and $x'_p$ and $x_p^{ij}$ only differ in distribution on the four critical edges. Their marginal distributions on these four edges are two discrete variables on $2^4$ states, with total variation distance $1 - 4(1 - p)^2p^2$. Thus, there exists a random variable $x''_p$ such that $x''_p \overset{d}{=} x'_p \overset{d}{=} x_p$, and $\Pr(x''_p = x_p^{ij}) = 4(1 - p)^2p^2$.

In other words, given a realization of $x_p^{ij}$, with probability $4(1 - p)^2p^2$, this is equal to a realization from the distribution of $x_p$, and therefore no binary Hopfield network storing both $x^{ij}$ and $x$ can correctly recover the original state from such an input. An indistinguishable realization occurs when two of the four critical edges are flipped in a certain combination. For fixed $x$, there are $k(k - 2)$ such $x^{ij}$ where the critical edges are disjoint. Thus, the probability of $x_p$ being an indistinguishable realization from a realization of one of the $x^{ij}$ is at least

$$1 - (1 - 4(1 - p)^2p^2)^{k(k-2)} > 1 - (1 - 4p^2)^{n/2},$$

completing the proof of the claim. ■

VI. EXAMPLES OF CLIQUE STORAGE ROBUSTNESS

Finally, in Fig 5 below we present examples of robust storage of cliques for the networks in Fig. 4 of the main text.
FIG. 5. Examples of robustness for networks in Fig. 4 of main text with $v = 128$, $k = 64$, $n = 8128$. Adjacency matrices of noisy cliques (in red) have $1219/1625$ bits corrupted out of $8128$ ($p = .15/2$) from the original $64$-clique (in green). Images show result of dynamics applied to these noisy patterns using networks with All-to-all MPF parameters after L-BFGS training on $50000$ $64$-cliques ($\approx 2e-31\%$ of all $64$-cliques), Large Deviation parameters $(x, y, z) = (.0091, 0, 1)$, or MPF Theory parameters $(x, y, z) = (.0107, 0, 1)$ from expression (7) in the main text.
FIG. 6. Feasible exponential storage. a) The shaded region is the feasible polytope for network parameters giving clique storage for $5 \leq k \leq 15$. Black points are its vertices, the red $R_k$ and blue $B_k$ lines are linear constraints. b) Lines $R_k$ (red) and $B_k$ (blue) for $1000 \leq k \leq 5500$. Note the appearance of the smooth curve $Q$ enveloping the family $B_k$ in the figure.