APPENDIX

6.1 PROOF OF THEOREM 2

Here we show how all incoming messages from outside the patch can be subsumed into an effective local field.

Proof. First let us revisit the update equation from X_i to $X_j : X_j \in \mathbf{X}_i$

$$\mu_{ij}^{n+1}(x_j) \propto \sum_{x_i \in \mathcal{X}} \Phi(x_i, x_j)\Phi(x_i) \prod_{X_k \in \{\partial(i) \setminus X_j\}} \mu_{ki}^n(x_i).$$

Now we group the incoming messages into two groups, i.e., messages coming from outside the patch and messages coming from inside the patch so that

$$\mu_{ij}^{n+1}(x_j) \propto \sum_{x_i \in \mathcal{X}} \Phi(x_i, x_j)\Phi(x_i) \prod_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \mu_{ki}^n(x_i) \prod_{X_k \in \{\mathbf{X}_i \cap \partial(i) \setminus X_j\}} \mu_{ki}^n(x_i).$$

Now we make use of the fact that we are only dealing with binary RVs for which $\mu_{ij}(X_j = -1) = (1 - \mu_{ij}(X_j = 1))$ and express the message explicitly by

$$\begin{aligned}\mu_{ij}^{n+1}(x_j) \propto& \exp(Jx_j)\exp(\theta_i) \prod_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \mu_{ki}^n(X_i = 1) \prod_{X_k \in \{\mathbf{X}_i \cap \partial(i) \setminus X_j\}} \mu_{ki}^n(X_i = 1) \\ &+ \exp(-Jx_j)\exp(-\theta_i) \prod_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} (1 - \mu_{ki}^n(X_i = 1)) \prod_{X_k \in \{\mathbf{X}_i \cap \partial(i) \setminus X_j\}} (1 - \mu_{ki}^n(X_i = 1)). \qquad (25)\end{aligned}$$

We now want to get rid of the first product and absorb the influence of these messages into the local field. In particular we aim to express it according to

$$\begin{aligned}\exp\big(\tilde{\theta}_i x_i\big) &= \Phi(x_i) \prod_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \mu_{ki}^n(x_i) \\ &= \exp\big((\theta_i + c)x_i\big) \cdot \exp(g) \qquad (26)\end{aligned}$$

In order to do so we take the logarithm of the product over all messages in (26) and put them into the exponent with the local field so that for $x_i = +1$

$$\begin{aligned}\exp(\tilde{\theta}_i) &= \exp\Bigg(\theta_i + \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \log\big(\mu_{ki}(X_i = 1)\big)\Bigg) \\ &= \exp\big(\theta_i + \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} c_i + \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} g_i\big), \qquad (27)\end{aligned}$$

and for $x_i = -1$

$$\begin{aligned}\exp(-\tilde{\theta}_i) &= \exp\Bigg(- \theta_i + \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \log\big(1 - (\mu_{ki}(X_i = 1))\big)\Bigg) \\ &= \exp\big(- \theta_i - \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} c_i + \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} g_i\big). \qquad (28)\end{aligned}$$

Equating the coefficients for c_i and g_i in (27) and (28) gives us the final results:

$$\begin{aligned}c_i &= \frac{1}{2}\Bigg(\log \mu_{ki}(X_i = 1) - \log\big(1 - \mu_{ki}(X_i = 1)\big)\Bigg), \\ c &= \sum_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \operatorname{atanh}\big(2\mu_{ki}(X_i = 1) - 1\big), \qquad (29) \\ g &= \frac{1}{2}\log \prod_{X_k \in \partial(i) \setminus \{X_j \cap \mathbf{X}_i\}} \Big(\mu_{ki}(X_i = 1) - \mu_{ki}(X_i = 1)^2\Big). \qquad (30)\end{aligned}$$

Note that we can further express the message for the second state $\mu_{ij}(X_j = -1)$ in a similar way, with the only difference that the values of the pairwise potentials change. Consequently we get exactly the same result for g again; this allows us to neglect the influence of g altogether, as it will be canceled out when normalizing the messages so that the sum up to one. □

6.2 PROOFS OF SECTION 5.3

The subsequent properties are a direct consequence of Definition 2, and are restricted to region (II). One crucial ingredient in many arguments will be the fact that every attractive model with vanishing or unidirectional local field has a stable unique or two stable fixed points (cf. Sec. 3).

6.2.1 PROOF OF THEOREM 3

We will first proof Thm. 3 that bounds the number of possible fixed point solutions inside (II).

Proof. First assume that a given patch $\mathcal{G}_i$ is flipped; then by the definition of the patch potential model (Def. 1) and by Thm. 2 it follows that effective field $\tilde{\theta}$ is aligned with the local field of the variables at any neighbor patch $\mathcal{G}_j$ so that

$$\operatorname{sgn}(\tilde{\theta_j}) = \operatorname{sgn}(\theta_j), \tag{31}$$

$$|\tilde{\theta_j}| > |\theta_j|. \tag{32}$$

Further let us recall the definition of (II) (Def. 2.2 in particular). It follows that the effective field stabilizes its neighbor patch $\mathcal{G}_j$, i.e., $J_A(\mathcal{G}_j, \tilde{\theta}) > J_A(\mathcal{G}_j, \theta)$ so that, according to the definition $J < J_A(\mathcal{G}_j, \tilde{\theta})$, and $\mathcal{G}_j$ admits only a unique solution.

Second assume that $\mathcal{G}_i$ is not flipped; then it follows by Thm. 2 that $\tilde{\theta}_i < \theta_i$ for X_i on the boundary of the neighbor patch $\mathcal{G}_j$. This decrease in the local field reduces the threshold for the existence of two solutions to smaller values of J (cf. Knoll and Pernkopf (2017)) so that

$$J_A(\mathcal{G}_j, \tilde{\theta}) < J_A(\mathcal{G}, \theta) < J. \tag{33}$$

By Def. 2.1 it follows that the neighbor patch $\mathcal{G}_j$ has two fixed points now, and it depends on the initialization to which one BP will converge.

Finally, we aim to show that the number of possible fixed points is bounded. Therefore we want to stress that the above arguments show how patches can either be aligned with the local potential or be flipped; it is crucial that every patch acts as one instance and that all variables belonging to one patch are aligned. Else the fixed point would be disordered which we rule out precisely by Definition 2. This and the fact that we are considering binary random variables limits the amount of possible solutions to

$$M \leq 2^{|\mathcal{G}_i|}, \tag{34}$$

where $|\mathcal{G}_i|$ denotes the overall number of patches. □

6.2.2 PROOF OF THEOREM 4

Next, we will guarantee that all existing fixed points are stable inside (II).

Proof. The proof relies on the definition of (II) again: this implies that every patch is effectively in one of two performance regions (cf. discussion in Sec. 5.3). We will discuss both performance regions one after another.

Either the neighbor patch $\mathcal{G}_j$ is flipped and stabilizes $\mathcal{G}_i$, i.e., $J < J_A(\mathcal{G}_i, \tilde{\theta})$ so that a unique fixed point exists; in that case it is a well-established fact that the fixed point is stable (Knoll and Pernkopf, 2017; Watanabe and Fukumizu, 2009; Ihler et al., 2005).

If however the neighbor patch is not flipped we have $J < J_A(\mathcal{G}_i, \tilde{\theta})$ in which case two stable fixed points exist (Knoll and Pernkopf, 2017) □

6.2.3 PROOF OF THEOREM 5

Proof. According to (15) we can express the exact solution by the convex combination of all fixed points. Consequently, using symmetry properties of the binary random variables, the error is given by

$$E_P(k) = \frac{2}{N} \sum_{X_i} \left| \frac{\sum_m \mathcal{Z}_B^m \tilde{P}_{X_i}^m}{\sum_m \mathcal{Z}_B^m} - \tilde{P}_{X_i}^k \right|^2 \tag{35}$$

$$= \frac{2}{N} \sum_{X_i} \left| \frac{\sum_m \mathcal{Z}_B^m \tilde{P}_{X_i}^m - \sum_m \mathcal{Z}_B^m \tilde{P}_{X_i}^k}{\sum_m \mathcal{Z}_B^m} \right|^2 \tag{36}$$

$$= \frac{2}{N\big(\sum_m \mathcal{Z}_B^m\big)^2} \sum_{X_i} \left| \sum_{m \backslash k} \mathcal{Z}_B^m \tilde{P}_{X_i}^m - \tilde{P}_{X_i}^k \right|^2 \tag{37}$$

where we first, bring everything on the same denominator so that the k^{th} contribution cancels out subsequently. □

6.2.4 EXPRESSING THE ERROR RATIO

We want to compare error of the state-preserving fixed point $E_P(p)$ to the error $E_P(q)$ of a fixed point that has all marginals biased towards one state. We discuss the error-ratio in a general manner in Sec. 6.2.5 and provide a more accessible proof for the special case of a model with two patches, i.e., for Example 1 in Sec. 6.2.6.

6.2.5 PROOF OF THEOREM 6

Proof. To show that, irrespective of the value of $\mathcal{Z}_B$, $\frac{E_P(p)}{E_P(q)} < 1$ we assume that the state preserving fixed point does not minimize the Bethe free energy, i.e., $\mathcal{Z}_B^q > \mathcal{Z}_B^p$.

Without loss of generality we make some prior assumptions on the model:

First, we assume that the overall number of variables with a positive local field equals the number of variables with a negative local field, i.e.,

$$|\{X_i : \theta_i = +\theta\}| = |\{X_j : \theta_i = -\theta\}|. \tag{38}$$

Second, we assume that all patches are of equal size. And finally, we group all possible fixed points according to their marginals and denote them as follows: The state-preserving fixed point is referred to as p; all fixed points that have more patches biased towards $X_i = +1$ are referred to as $s = 1, \dots, S$, with q being the fixed point that has all variables biased towards $X_i = +1$; all fixed points that have more patches biased towards $X_i = -1$ are referred to as $t = 1, \dots, T$, with r being the fixed point that has all variables biased towards $X_i = -1$.

This has some implications that will ease the subsequent analysis significantly. Specifically, the number of fixed points favoring one state equals the number of fixed points favoring the other state, i.e., $S = T$ by (38) and by Theorem 3.

Another important consequence of incorporating the interactions between patches into an effective field is that the number of variables that favor one state has an immediate influence on the value of the singleton marginals. It can be shown that the effective field, if stronger than the local field – note that the effective field is stronger than the local field whenever two neighboring patches are biased towards the same state – increases the bias of the variables. This intuitive statement is a consequence of the the Griffiths-Hurst-Sherman inequality (Griffiths et al., 1970) that can be extended to specific fixed points by straightforward manipulations (cf. Knoll et al. (2018a)). In essence this means that for all variables $X_i \in \mathbf{X}$ we have

$$\tilde{P}_{X_i}^q(X_i = 1) \geq \tilde{P}_{X_i}^s(X_i = 1) \geq \tilde{P}_{X_i}^p(X_i = 1) \geq \tilde{P}_{X_i}^t(X_i = 1) \geq \tilde{P}_{X_i}^r(X_i = 1). \tag{39}$$

Moreover, as all patches have equal size and because $J_{ij} = J$ as well as $\theta_i \in \{-\theta, \theta\}$ every fixed point s has a symmetric fixed point t that has the same value for the approximate partition function (cf. Proof of Theorem 7). That is, except for the state-preserving fixed point p all fixed points come in couples that satisfy

$$\mathcal{Z}_B^s = \mathcal{Z}_B^t. \tag{40}$$

We will further utilize the properties of the mismatch $Q_i(k,l)$; in particular the symmetry property

$$Q_i(k,l) = -Q_i(l,k), \tag{41}$$

and the expansion property

$$Q_i(k,l) = Q_i(k,m) + Q_i(m,l). \tag{42}$$

Finally, we express the error ration between the state-preserving fixed point p and the biased fixed point q according to Cor. 5.1 so that

$$\frac{E_P(p)}{E_P(q)} = \frac{\sum_{X_i} |\sum_{m\setminus p} \mathcal{Z}_B^m Q_i(m,p)|^2}{\sum_{X_i} |\sum_{m\setminus q} \mathcal{Z}_B^m Q_i(m,q)|^2} \tag{43}$$

$$\overset{(a)}{=} \frac{\sum_{X_i} |\sum_{s} \mathcal{Z}_B^s Q_i(s,p) + \sum_t \mathcal{Z}_B^t Q_i(t,p)|^2}{\sum_{X_i} |\sum_{s\setminus q} \mathcal{Z}_B^s Q_i(s,q) + \sum_t \mathcal{Z}_B^t Q_i(t,q) + \mathcal{Z}_B^p Q_i(p,q)|^2} \tag{44}$$

$$\overset{(b)}{=} \frac{\sum_{X_i} |\sum_{s\setminus q} \mathcal{Z}_B^s \big(Q_i(s,p) + Q_i(t,p)\big) + \mathcal{Z}_B^q \big(Q_i(q,p) + Q_i(r,p)\big)|^2}{\sum_{X_i} |\sum_{s\setminus q} \mathcal{Z}_B^s \big(Q_i(s,q) + Q_i(t,q)\big) + \mathcal{Z}_B^q Q_i(r,q) + \mathcal{Z}_B^p Q_i(p,q)|^2}, \tag{45}$$

where (a) follows from splitting the sum into the fixed points s that are more biased towards $X_i = 1$ and into the fixed points t that are more biased towards $X_i = -1$. Note that the state-preserving fixed point p does not belong to either set and is consequently expressed explicitly in the denominator. For (b) we make use of (40) and arrange the terms by making the dependence on q and r explicit so that the sum goes over the same terms in the numerator and in the denominator.

We can further express the error ratio and bound it using Jensen's inequality according to

$$\begin{aligned}\frac{E_P(p)}{E_P(q)} &\overset{(a)}{=} \frac{\sum_{X_i} |\sum_{s\setminus q} \mathcal{Z}_B^s \big(Q_i(s,p) + Q_i(t,p)\big) + \mathcal{Z}_B^q \big(Q_i(q,p) + Q_i(r,p)\big)|^2}{\sum_{X_i} \Big(\sum_{s\setminus q} \mathcal{Z}_B^s \big|\big(Q_i(s,q) + Q_i(t,q)\big)\big| + \mathcal{Z}_B^q \big|Q_i(r,q)\big| + \mathcal{Z}_B^p \big|Q_i(p,q)\big|\Big)^2} \\ &\leq \frac{\sum_{X_i} \Big(\sum_{s\setminus q} \mathcal{Z}_B^s \big|\big(Q_i(s,p) + Q_i(t,p)\big)\big| + \mathcal{Z}_B^q \big|\big(Q_i(q,p) + Q_i(r,p)\big)\big|\Big)^2}{\sum_{X_i} \Big(\sum_{s\setminus q} \mathcal{Z}_B^s \big|\big(Q_i(s,q) + Q_i(t,q)\big)\big| + \mathcal{Z}_B^q \big|Q_i(r,q)\big| + \mathcal{Z}_B^p \big|Q_i(p,q)\big|\Big)^2}\end{aligned} \tag{46}$$

where separating the norm does not change the result in (a) because of (39), which implies $Q_i(m,q) < 0$ for all fixed points $m \neq q$.

For completing the proof we make use of the symmetry property (41) and the expansion property (42) in order to rearrange the terms; in particular note that $Q_i(q,s) + Q_i(q,t) = Q_i(q,s) + Q_i(q,s) + Q_i(s,p) + Q_i(p,t)$, and that $Q_i(q,r) = Q_i(q,p) + Q_i(p,r)$ so that

$$\frac{E_P(p)}{E_P(q)} \leq \frac{\sum_{X_i} \Big(\sum_{s\setminus q} \mathcal{Z}_B^s \big|\big(Q_i(s,p) - Q_i(p,t)\big)\big| + \mathcal{Z}_B^q \big|\big(Q_i(q,p) - Q_i(p,r)\big)\big|\Big)^2}{\sum_{X_i} \Big(\sum_{s\setminus q} \mathcal{Z}_B^s \big|\big(Q_i(s,p) + Q_i(p,t) + Q_i(q,s) + Q_i(q,s)\big)\big| + \mathcal{Z}_B^q \big|Q_i(q,p) + Q_i(p,r)\big| + \mathcal{Z}_B^p \big|Q_i(q,p)\big|\Big)^2}.$$

Note that we have applied (41) so that every mismatch-term is strictly positive. It is thus straightforward to see, by comparing all terms, that the numerator is strictly smaller than the denominator for every variable $X_i \in \mathbf{X}$ so that

$$\frac{E_P(p)}{E_P(q)} < 1. \tag{47}$$

□

6.2.6 PROOF OF COROLLARY 6.1

Although Cor.6.1 is an immediate consequence of Thm. 6 we illustrate the proof that admits some intuitive arguments

Proof. For a symmetric model with two equal-sized patches we evaluate the error ratio between the state-preserving fixed point p and one of the fixed points that have all marginals biased towards one state, these are q and r and have *symmetric* marginals. Further assume that the fixed points q and r minimize the Bethe free energy, i.e.,

$$\mathcal{F}_B^q = \mathcal{F}_B^r < \mathcal{F}_B^p \tag{48}$$
$$\mathcal{Z}_B^q = \mathcal{Z}_B^r > \mathcal{Z}_B^p. \tag{49}$$

Then we want to show that (49) does not imply that $E_P(q) < E_P(p)$, i.e., we want to show that $\frac{E_P(p)}{E_P(q)} < 1$ despite (49); therefore, we express the ratio of the marginal errors according to

$$\frac{E_P(p)}{E_P(q)} = \frac{\sum_{X_i} |\sum_{m\backslash p} \mathcal{Z}_B^m Q_i(m,p)|^2}{\sum_{X_i} |\sum_{m\backslash q} \mathcal{Z}_B^m Q_i(m,q)|^2}$$
$$= \frac{\sum_{X_i} |\mathcal{Z}_B^q Q_i(q,p) + \mathcal{Z}_B^r Q_i(r,p)|^2}{\sum_{X_i} |\mathcal{Z}_B^q Q_i(r,q) + \mathcal{Z}_B^p Q_i(p,q)|^2} \tag{50}$$

Note that because of (49) we have

$$\frac{E_P(p)}{E_P(q)} = \frac{\sum_{X_i} |\mathcal{Z}_B^q \big(Q_i(q,p) + Q_i(r,p)\big)|^2}{\sum_{X_i} |\mathcal{Z}_B^q Q_i(r,q) + \mathcal{Z}_B^p Q_i(p,q)|^2} \tag{51}$$
$$\overset{(a)}{=} \frac{\sum_{X_i} |\mathcal{Z}_B^q \big(Q_i(q,p) + Q_i(r,p)\big)|^2}{\sum_{X_i} \big(\mathcal{Z}_B^q Q_i(q,r) + \mathcal{Z}_B^p Q_i(q,p)\big)^2} \tag{52}$$

where (a) follows from the fact that $Q_i(r,q) < 0$, $Q_i(p,q) < 0$, and the symmetry property (41). We further denote the constant difference between the biased fixed points by

$$0 < Q_i(q,r) = d < 1. \tag{53}$$

We can use the expansion property (42) to bound the numerator as $|Q_i(q,p) + Q_i(r,p)|^2 < Q_i(q,r)^2 = d^2$ so that

$$\frac{E_P(p)}{E_P(q)} < \frac{\sum_{X_i} \mathcal{Z}_B^q d^2}{\sum_{X_i} \big(\mathcal{Z}_B^q d + \mathcal{Z}_B^p Q_i(q,p)\big)^2}. \tag{54}$$

Which completes the proof as $Q_i(q,p) > 0$ □

6.2.7 PROOF OF THEOREM 7

Minimum of $\mathcal{F}_B$. Let us consider three different stationary points $\mathcal{F}_B^p$ (state-preserving), $\mathcal{F}_B^q$ (biased to one state), and $\mathcal{F}_B^m$ that has some patches flipped. Note that we consider $\mathcal{F}_B^q$ as a limiting case for $\mathcal{F}_B^m$. We will denote the number of variables that are aligned with the local field N_c and the number of flipped variables N_f.

The set of boundary edges that connects two patches is denoted by

$$\mathbf{E}_P = \{(i,j) \in \mathbf{E} : X_i \in \mathbf{X}_i, X_j \in \mathbf{X}_j \neq \mathbf{X}_i\}, \tag{55}$$

and the set of edges that connects two patches that have their variables not aligned is denoted by

$$\mathbf{E}_C = \{(i,j) \in \mathbf{E} : X_i \in \mathbf{X}_i, X_j \in \mathbf{X}_j \neq \mathbf{X}_i, \operatorname{sgn}\big(\tilde{P}_{X_i}(X_i = 1) - 0.5\big) \neq \operatorname{sgn}\big(\tilde{P}_{X_j}(X_j = 1) - 0.5\big)\}. \tag{56}$$

Note that $\mathbf{E}_P$ is constant for a specified model, whereas $\mathbf{E}_C$ depends on the specific fixed point. We consequently have $\mathbf{E}_C \leq \mathbf{E}_P$ with equality for the state preserving fixed point p.

Our analysis is restricted to $(\theta, J) \in (II)$ per definition: one crucial consequence is that $J > J_A(\theta)$ and that most marginals either have $\tilde{P}_{X_i}(x_i) \approx 1$ or $\tilde{P}_{X_i}(x_i) \approx 0$. We will exploit this fact and express all marginals according to $\tilde{P}_{X_i}(x_i) \in \{0, 1\}$ which allows us to simplify $\mathcal{F}_B$, as defined in Sec. 2.3, according to

$$\mathcal{F}_B^p = -N\theta - \big(|\mathbf{E}| - 2|\mathbf{E}_P|\big)J - S_B^p \tag{57}$$

$$\mathcal{F}_B^m = -\big(N_c - N_f\big)\theta - \big(|\mathbf{E}| - 2|\mathbf{E}_C|\big)J - S_B^m. \tag{58}$$

Let $\Delta S_B = S_B^p - S_B^m$ be the difference in the entropy, then we can express the conditions for the state-preserving fixed point to have a lower value $\mathcal{F}_B^p \leq \mathcal{F}_B^m$ according to

$$\begin{aligned} -N\theta - \big(|\mathbf{E}| - 2|\mathbf{E}_P|\big)J &\leq -\big(N_c - N_f\big)\theta - \big(|\mathbf{E}| - 2|\mathbf{E}_C|\big)J + \Delta S_B \\ 2J(-|\mathbf{E}_C| + |\mathbf{E}_P|) &\leq \theta(N - N_c + N_f) + \Delta S_B \end{aligned} \tag{59}$$

Now let us express (59) for the fixed points that has all variables biased to one state, i.e., $|\mathbf{E}_C| = 0$. Then, the state-preserving fixed point has a lower value $\mathcal{F}_B^p < \mathcal{F}_B^q$ if

$$2J|\mathbf{E}_P| \leq \theta(N - N_c + N_f) + \Delta S_B \tag{60}$$

□

6.2.8 COROLLARY 7.1

Proof. For the specific case of a grid graph with two equal-sized patches (Example 1) we can further simplify the condition from Thm. 7. Therefore, note that for $\mathcal{F}_B^p < \mathcal{F}_B^q$ to be satisfied, we have $N_c = N_f$ and $|\mathbf{E}_P| = \sqrt{N}$ so that (60) reduces to

$$2J\sqrt{N} \leq \theta N + \Delta S_B. \tag{61}$$

The definition of (II) requires strong interactions J so that the entropy terms in the free energy vanish. We can consequently approximate (61) by

$$2J\sqrt{N} \leq \theta N. \tag{62}$$

□