Abstract—Sparse superposition (SS) codes were originally proposed as a capacity-achieving communication scheme over the additive white Gaussian noise channel (AWGNC) [1]. Very recently, it was discovered that these codes are universal, in the sense that they achieve capacity over any memoryless channel under generalized approximate message-passing (GAMP) decoding [2], although this decoder has never been stated for SS codes. In this contribution we introduce the GAMP decoder for SS codes, we confirm empirically the universality of this communication scheme through its study on various channels and we provide the main analysis tools: state evolution and potential. We also compare the performance of GAMP with the Bayes-optimal MMSE decoder. We empirically illustrate that despite the presence of a phase transition preventing GAMP to reach the optimal performance, spatial coupling allows to boost the performance that eventually tends to capacity in a proper limit. We also prove that, in contrast with the AWGNC case, SS codes for binary input channels have a vanishing error floor in the limit of large codewords even with finite sparsity. This means that when decoding is possible, optimal decoding is asymptotically perfect as well as GAMP decoding until some threshold, a very promising feature which is not present for the real-valued input AWGNC. Keeping in mind practicality, we focus our empirical study on Hadamard-based coding operators that allow to drastically reduce the encoding and decoding complexity, while maintaining good performance for moderate block-lengths [11].

Our experiments confirm that SE recursion of [2] accurately tracks GAMP. Using the potential of the code we also compare the performance of GAMP to the optimal MMSE decoder. In addition, our empirical study confirms the asymptotic results of [2]: the performance of SS codes under GAMP decoding can be significantly increased towards capacity using spatial coupling, as already observed for the AWGNC [12]. Moreover, we prove that for binary input channels, SS codes have a vanishing error floor in the limit of large codewords even with finite sparsity. This means that when decoding is possible, optimal decoding is asymptotically perfect as well as GAMP decoding until some threshold, a very promising feature which is not present for the real-valued input AWGNC. Keeping in mind practicality, we focus our empirical study on Hadamard-based coding operators that allow to drastically reduce the encoding and decoding complexity, while maintaining good performance for moderate block-lengths [11].
self can be interpreted as an effective memoryless channel
\[ P_{\text{out}}(y|Ax) = \prod_{m=1}^{M} P_{\text{out}}(y_m|[Ax]_m) \]. For the channels we focus on, \( P_{\text{out}}(y_m|[Ax]_m) \) is expressed as follows:

- **AWGNC**: \( N(\gamma_m[[Ax]_m], 1/snr) \)
- **BEC**: \( (1-\epsilon)\delta(y_m-\text{sign}([Ax]_m)) + \epsilon\delta(y_m) \)
- **BSC**: \( (1-\epsilon)(y_m-\text{sign}([Ax]_m)) + \epsilon(y_m + \text{sign}([Ax]_m)) \)
- **ZC**: \( \delta(\text{sign}([Ax]_m)+1)(\epsilon(y_m-1) + (1-\epsilon)(y_m+1)) + \delta(\text{sign}([Ax]_m)-1) - \epsilon(y_m) \)

where snr is the signal-to-noise of the AWGNC, \( \epsilon \) the erasure or flip probability of the BEC, ZC and BSC. The sign maps the Gaussian distributed codeword components onto the input alphabets of the binary input channels.

Note that for the asymmetric ZC, the symmetric map \( \text{sign}([Ax]_m) \) leads to a sub-optimal uniform input distribution. The **symmetric capacity** of the ZC differs from Shannon’s capacity but the difference is small, and similarly for the algorithmic threshold, see [2]. We thus consider this symmetric setting for the sake of simplicity. The other channels are symmetric, this map thus leads to the optimal input distribution.

### III. THE GAMP DECODER

We consider a Bayesian setting and associate to the message the posterior \( P(x|y,A) = P_{\text{out}}(y|Ax)P_{0}(x)/P(y|A) \). The hard constraints for the sections are enforced by the prior \( P_{0}(x) = \prod_{i=1}^{l} P_{0}(x_i) \) with \( P_{0}(x_i) = B^{-1}\sum_{\delta_{x_i}} \prod_{j\in[l], j\neq i} \delta_{x_j,0} \), where \( i \in l \) are the B scalar components indices of the section \( l \). The GAMP decoder aims at performing MMSE estimation by approximating the posterior mean of each section.

In the GAMP decoder Algorithm 1, \( \circ \) denotes element-wise operations. GAMP was originally derived for scalar estimation. In this generalization to the vectorial setting of SS codes, whose derivation is similar to the one of AMP for SS codes found in [12], only the input non-linear steps differ from canonical GAMP [14]: here the so-called *denoiser* \( g_m \) acts *sectionwise* instead of componentwise. In full generality, it is defined as [14] \( g_m(r,\tau) := \mathbb{E}[X|R = r] \) for the random variable \( R = X + Z \) with \( X \sim P_{0} \) and \( Z \sim N(0, \text{diag}(\tau)) \).

### IV. STATE EVOLUTION AND THE POTENTIAL

We now present the analysis tools of the \( L \to \infty \) performance of SS codes under GAMP and MMSE decoding when Gaussian matrices are used: state evolution and potential.

#### A. State evolution

The asymptotic performance of GAMP with Gaussian i.i.d coding matrices is tracked by SE, a scalar recursion [2, 9, 10, 14] analogous to density evolution for low-density parity-check codes. Note that although SE is not rigorous for vectorial setting, the rigorous analysis of [8] and the present empirical results strongly suggest that it is exact, which we conjecture.

The aim is to compute the asymptotic MSE of the GAMP estimate \( \hat{E}_{(t)} := \lim_{t \to \infty} \mathbb{E}[(\hat{X}_{(t)} - X)_i^2] \). It turns out that this is equivalent to recursively compute the MMSE \( T(E) := \)
TABLE I: The expressions for $g_{\text{out}}$, $-g_{\text{out}}'$ and $\mathcal{F}$.  

| General | $[g_{\text{out}}(p,y,t)]$, | $[-g_{\text{out}}(p,y,t)]$, | $\mathcal{F}(p,E)$ |
|----------|------------------|------------------|------------------|
| AWGNC    | $\frac{p_{\text{out}}}{1+p_{\text{out}}}$ | $\frac{p_{\text{out}}}{1+p_{\text{out}}}$ | $\mathcal{F}(p,E)$ |
| BEC      | $\frac{p_{z_1}}{1+p_{\text{out}}}$ | $\frac{p_{z_1}}{1+p_{\text{out}}}$ | $\mathcal{F}(p,E)$ |
| ZC       | $\frac{p_{z_1}}{1+p_{\text{out}}}$ | $\frac{p_{z_1}}{1+p_{\text{out}}}$ | $\mathcal{F}(p,E)$ |

The MMSE of the equivalent AWGNC is obtained after simple algebra [2] and reads

$$ T(E) = E_{\mathbb{Z}}[g_{\text{in}}(\Sigma(E), Z) - 1]^2 + (B-1)g_{\text{in}}(\Sigma(E), Z)^2]. $$

Here $g_{\text{in}}^{(1)}$ is interpreted as the posterior mean approximated by GAMP for the non-zero component in the transmitted section while $g_{\text{in}}^{(2)}$ corresponds to the remaining components. The SE recursion tracking the MSEE of GAMP is then

$$ E_{(t+1)} = T(E_{(t)}), \quad t \geq 0, $$

initialized with $E_{(0)} = 1$. Hence, the asymptotic MSE reached by GAMP upon convergence is $E_{\infty}$. Moreover, define the asymptotic $L \rightarrow \infty$ error floor of SS codes $E_s$ for the fixed point of SE (2) initialized from $E_{(0)} = 0$. Fig. 1 shows that SE properly tracks GAMP on the BEC. Note that the section error rate (SER) of GAMP, the fraction of wrongly decoded sections after hard thresholding of $\hat{X}(t)$, can also be asymptotically tracked thanks to SE through a simple one-to-one mapping between $E_{(t)}$ and the asymptotic SER at $t$ [7, 12].

Under GAMP decoding SS codes exhibit, as $L \rightarrow \infty$, a sharp phase transition at an algorithmic threshold $R_{\text{GAMP}}$ below Shannon’s capacity. $R_{\text{GAMP}}$ is defined as the highest rate such that for $R \leq R_{\text{GAMP}}$, (2) has a unique fixed point $E_{(\infty)} = E_s$ (see [2] for formal definitions). In this regime GAMP decodes well, see red and blue curves of Fig. 1. If $R > R_{\text{GAMP}}$ GAMP decoding fails, see green curve. As we will see in the next sections, spatial coupling may allow to boost the performance of the scheme by increasing the GAMP algorithmic threshold.

B. Potential formulation

The SE (2) is associated with a potential $F_u(E)$, whose stationary points correspond to the fixed points of SE: $\partial E F_u(E) = 0 \Rightarrow T(E_0) = E_0$. For SS codes it is [2]:

$$ F_u(E) := U_u(E) - S_u(\Sigma(E)) \quad U_u(E) := \frac{E}{B} \ln B - \frac{1}{B} E_{\mathbb{Z}}[dy \log \phi(\tilde{g}(\delta))], \quad S_u(\Sigma(E)) := E_{\mathbb{Z}}[\log B (1 + \sum_{i=1}^B e_i(\Sigma(E), B))], $$

where $\phi = \phi(y | Z, E) := \int dy P_{\text{out}}(y | s) N(s \sqrt{1-E}, E, Z \sim \sqrt{N(0,1)})$ and $e_i(\Sigma(E), B)$ is the hard phase, the potential possesses another local min. (red dot) and the corresponding “bad” fixed point of SE prevents GAMP to reach $\tilde{E}$; decoding fails (yellow curves). Finally, the rate at which the local and global min. switch roles is the potential threshold $R_{\text{pot}}$ (purple curves). Optimal decoding is possible as long as $R < R_{\text{pot}}$ as the MMSE switches at $R_{\text{pot}}$ from a “low” to a “high” value. At higher rates GAMP is again optimal but leads to poor results as decoding is impossible. Note that if $R < R_{\text{pot}}$, then $E_{(0)} \neq \tilde{E}$. 

\[
\mathbb{E}_S\mathbb{Z}[\|S - \mathbb{E}[X|S + (\Sigma(E)/B)\mathbb{Z}]\|_2^2] \text{ of a single section } (S \sim p_y) \text{ sent through an equivalent AWGNC } (Z \sim \mathbb{N}(0, 1)) \text{ of noise variance } (\Sigma(E)/B)^2, b^2 := \log_2(B). \]
Universal mechanism behind the excellent performances of coupled limit of infinite coupled chains. This phenomenon is referred to as threshold saturation and is understood as the generic mechanism behind the excellent performances of coupled codes [2, 18]. Moreover, a very interesting feature of SS codes is that $R_{\text{pot}}$ itself approaches the capacity as $B \to \infty$ [2]. These phenomena imply together that in these limits (infinite chain length and $B$), spatially coupled SS codes under GAMP decoding are universal in the sense that they achieve the Shannon capacity of all memoryless channels.

C. Vanishing error floor for binary input memoryless channels

Another promising feature of SS codes is related to their error floor. In the real-valued input AWGNC case, an error floor always exists but it can be made arbitrary small by increasing $B$ [2, 12]: $\lim_{B \to \infty} E_* = \lim_{B \to \infty} \text{SER}_* = 0$, the error floor in the SER sense. In contrast, in the BSC, ZC and BSC cases (more generally for binary input memoryless channels), we now prove that as $L \to \infty$ the error floor vanishes for any $e$ and $B$. This implies that when $E_* = \tilde{E}$ optimal decoding is asymptotically perfect, and thus GAMP decoding as well for $R \leq R_{\text{GAMP}}$. This is actually verified in practice for GAMP where perfect decoding is statistically possible even for moderate block-lengths, see blue and red curves of Fig. 1.

The proof of $E_* = 0$, i.e. the existence of the trivial fixed point $T(0) = 0$ of (2), does not guarantee that this is the global minimum of the potential in the hard phase; i.e it is a priori possible that $E_* \neq \tilde{E}$. Nevertheless, our careful numerical work indicates that there exist at most two fixed points of $\Sigma$ at the same time or equivalently two minima in the potential, namely $E_* = \tilde{E} \neq E_{(\infty)}$ if $R \in [R_{\text{GAMP}}, R_{\text{pot}}]$ or $E_* = \tilde{E} = E_{(\infty)}$ if $R > R_{\text{pot}}$ (at least for the studied cases), see Fig. 2. This also agrees with the $B \to \infty$ analysis of the potential [2, 12].

Let us now prove that $E_* = 0$ for the BSC: the proof for other binary input channels being similar. It starts by noticing, from the definition of $T(E)$ as the MMSE of an AWGNC with noise parameter $\Sigma(E)$, that a sufficient condition for $T(0) = 0$ is $\lim_{E \to 0} \Sigma(E) = 0$; indeed no noise implies vanishing MMSE. From (1) this condition is equivalent to $\lim_{E \to 0} I_E(E) = \infty$ that we now prove, where $I_A(E) := \int_A dp \, N(p)(0, 1-E) \, F(p|E)$. Consider instead $I_E(E)$ where $E := [E-\sqrt{E}, E+\sqrt{E}]$. Using Table I for the expression of $F(p|E)$ for the BEC, this restricted integral is

$$I_E(E) = \frac{1}{2}(e^{-\sqrt{E}} - e^{\sqrt{E}}) \int_{E} dp \frac{e^{-\pi^2 p^2}}{Q(p, E)(1-Q(p, E))}.$$  

Here $Q(p, E) \in \left[C_E, 1-C_E\right]$, with $\lim_{E \to 0} C_E > 0$ for $p \in \mathcal{E}$, $E \leq 1$. This implies that $K(E) := \max_{p \in \mathcal{E}} Q(p, E)(1-Q(p, E)) = O(1)$. Since the interval $\mathcal{E}$ is of size $2\sqrt{E}$, then

$$I_E(E) \geq \frac{1}{2}(e^{-\sqrt{E}} - e^{\sqrt{E}}) \int_{E} dp \frac{1}{Q(p, E)}.$$  

From this we can assert that $\lim_{E \to 0} I_E(E) = \infty$. Moreover $I_E(E) < I_E(E)$ as $F(p|E) \geq 0$ (recall it is a Fisher information) and thus $\lim_{E \to 0} I_E(E) = \infty$ which ends the proof.

For the BSC and ZC the proof is similar, the main ingredient being the squared Gaussian $Q^2$ at the numerator of $F(p|E)$, see Table I, which leads to similar expressions as (3) and thus the $1/\sqrt{E}$ divergence when $E \to 0$. We believe that the same mechanism holds for any binary input memoryless channel, implying a vanishing error floor as well as asymptotic perfect decoding of GAMP below the algorithmic threshold.

V. NUMERICAL EXPERIMENTS

In Fig. 3 we compare the optimal and GAMP performances in terms of attainable rate, denoted by $R_{\text{pot}}$ and $R_{\text{GAMP}}$ respectively. For all channels, there exists, as long as the noise is not “too high”, a hard phase where GAMP is sub-optimal. Moreover, the use of Hadamard-based operators have a performance cost w.r.t Gaussian ones but which vanishes as $B$ increases; they both have the same algorithmic threshold for $B$ large enough (but still practical, $B \geq 64$ being enough).

![Fig. 2: Potential for the AWGNC with $\text{snr} = 100$ (top) and the BEC with $e = 0.1$ (bottom), in both cases with $B = 2$. The MMSE is the argmin $F_0(E)$ (black dot). When the min. is unique (i.e $R < R_{\text{GAMP}}$, blue curve) or if the global min. is the rightmost one ($R > R_{\text{pot}}$, green curve), GAMP is asymptotically optimal, despite that if $R > R_{\text{pot}}$ it leads to poor results. The red dot is the local min., preventing GAMP to decode if $R \in [R_{\text{GAMP}}, R_{\text{pot}}]$ (yellow curve).](image-url)
Spatial coupling allows important improvements towards $R_{\text{pot}}$ even in practical settings, confirming the universality of coupled SS codes under GAMP decoding as $\lim_{\beta \to \infty} R_{\text{pot}} = C$ [2]. The mismatch between $R_{\text{pot}}$ and $R_{\text{GAMP}}$ is due to finite size effects which are more evident in coupled codes (both chain lengths and coupling windows should go to infinity after $L$ for $R_{\text{GAMP}}$ to saturate $R_{\text{pot}}$).

**ACKNOWLEDGMENTS**

We acknowledge Nicolas Macris, Florent Krzakala and Rüdiger Urbanke for helpful comments as well as Alper Kose and Berke Aral Sonmez for an early stage of study of GAMP. This work was funded by the SNSF grant no. 200021-156672.

**REFERENCES**

[1] A. Barron and A. Joseph, “Toward fast reliable communication at rates near capacity with gaussian noise,” in Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, June 2010, pp. 315–319.

[2] J. Barbier, M. Dia, and N. Macris, “Threshold Saturation of Spatially Coupled Sparse Superposition Codes for All Memoryless Channels,” in Information Theory workshop (ITW), 2016 IEEE, 2016.

[3] A. Joseph and A. R. Barron, “Fast sparse superposition codes have near exponential error probability for RSC,” IEEE Trans. on Information Theory, vol. 60, no. 2, pp. 919–942, 2014.

[4] A. R. Barron and S. Cho, “High-rate sparse superposition codes with iteratively optimal estimates,” in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on., IEEE, 2012, pp. 120–124.

[5] J. Barbier, “Approximate message-passing algorithms for compressed sensing,” Proceedings of the National Academy of Sciences, vol. 106, no. 45, pp. 18914–18919, 2009.

[6] 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. P08009, 2012.

[7] J. C. Condo and W. J. Gross, “Sparse superposition codes: A practical algorithm for compressed sensing,” Proceedings of the National Academy of Sciences, vol. 106, no. 25, pp. 10 256–10 260, 2009.

[8] J. Barbier and F. Krzakala, “Recall replica analysis and approximate message passing decoder for superposition codes,” in Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on, 2014.

[9] C. Rush, A. Greig, and R. Venkataramanan, “Capacity-achieving sparse superposition codes via approximate message passing decoding,” arXiv preprint arXiv:1501.05892, 2015.

[10] J. Barbier and F. Krzakala, “Approximate message-passing decoder for superposition codes,” in Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on, 2014.

[11] A. Joseph and A. R. Barron, “Fast sparse superposition codes have near exponential error probability for RSC,” IEEE Trans. on Information Theory, vol. 60, no. 2, pp. 919–942, 2014.

[12] 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, 2015.

[13] J. Barbier, M. Dia, and N. Macris, “Proof of Threshold Saturation for Spatially Coupled Sparse Superposition Codes,” in Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on, 2016.

[14] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” arXiv preprint arXiv:1001.5411, 2012.

[15] C. Condo and W. J. Gross, “Sparse superposition codes: A practical approach,” in Signal Processing Systems (SiPS), 2015 IEEE Workshop on, Oct 2015, pp. 1–6.

[16] J. Barbier, M. Dia, N. Macris, and F. Krzakala, “The Mutual Information in Random Linear Estimation,” in in the 54th Annual Allerton Conference on Communication, Control, and Computing, September 2016.

[17] G. Reeves and H. D. Pfister, “The replica-symmetric prediction for compressed sensing with gaussian matrices is exact,” in 2016 IEEE International Symposium on Information Theory (ISIT), July 2016.

[18] S. Kudekar, T. Richardson, and R. L. Urbanke, “Spatially coupled ensembles universally achieve capacity under belief propagation,” IEEE Trans. on Information Theory, vol. 59, no. 12, pp. 7761–7813, Dec 2013.