Threshold Saturation of Spatially Coupled Sparse Superposition Codes for All Memoryless Channels

Jean Barbier, Member IEEE, Mohamad Dia and Nicolas Macris, Member IEEE.

Laboratoire de Théorie des Communications, Ecole Polytechnique Fédérale de Lausanne.
{jean.barbier, mohamad.dia, nicolas.macris}@epfl.ch

Abstract—We recently proved threshold saturation for spatially coupled sparse superposition codes on the additive white Gaussian noise channel [1]. Here we generalize our analysis to a much broader setting. We show for any memoryless channel that spatial coupling allows generalized approximate message-passing (GAMP) decoding to reach the potential (or Bayes optimal) threshold of the code ensemble. Moreover in the large input alphabet size limit: i) the GAMP algorithmic threshold of the underlying (or uncoupled) code ensemble is simply expressed as a Fisher information; ii) the potential threshold tends to Shannon’s capacity. Although we focus on coding for sake of coherence with our previous results, the framework and methods are very general and hold for a wide class of generalized estimation problems with random linear mixing.

I. INTRODUCTION

Spatial superposition (SS) codes were developed for reliable communication over the additive white Gaussian noise (AWGN) channel [2] and were proven to be capacity-achieving for this channel when power allocation and iterative decoding are employed [3, 4]. Later on, the approximate message-passing (AMP) decoder was introduced in [5] and spatial coupling (SC) constructions (also combined with efficient Hadamard-based operators) were presented in [6, 7]. These SC constructions have many similarities with those introduced in the context of compressed sensing [8, 9], the first successful application of SC to dense systems. An independent line of work also studying the AMP decoder, but using power allocation instead of SC, is presented in [10].

It appears that SS-SC codes have much better performances than power allocated ones [7]. This motivated the initiation of their rigorous study [1] using the potential method, originally developed for low density parity check codes [11–13]. In [1] we showed that i) threshold saturation occurs, i.e. minimum mean square error (MMSE) performance is reached using SC and AMP decoding, and ii) the potential threshold (above which AMP decoding is not possible without using SC or power allocation) tends to capacity in the large alphabet size limit, and this even without power allocation.

These encouraging results (obtained for the AWGN) naturally led us to study a general setting that includes all memoryless channels and any input signal model that factorizes over B-dimensional (B-d) sections \( p_0(s) = \prod_{i=1}^{L} p_0(s_i), s_i \in \mathbb{R}^B \).

The present analysis is also based on the potential method. The correct potential and associated state evolution (SE) for the present setting can be “guessed” using the replica method. Alternatively, one can “integrate” the SE associated with the GAMP algorithm in the vectorial setting. The GAMP equations were originally derived for scalar estimation [14], but their extension to the present vectorial setting is immediate.

II. CODE ENSEMBLES

In the sequel, the shorthands \([a_1 : a_n]\) and \([a_1, a_2]\) refer to \([a_1, \ldots, a_n]\) and \([a_1, \ldots, a_n]\) respectively. The probability distribution of a Gaussian random variable \(x\) with mean \(m\) and variance \(\sigma^2\) is denoted \(\mathcal{N}(x|m, \sigma^2)\).

Let us start defining the underlying ensemble of SS codes for transmission over a generic memoryless channel. The information word or message is a vector made of \(L\) sections, \(s = [s_1 : s_L]\). Each section is a \(B\)-d vector with a single non-zero component equal to 1. \(B\) is the section size (or alphabet size) and we set \(N = LB\). For example if \((B = 3, L = 4)\), then a valid message could be \(s = [001, 100, 100, 010]\). We consider random linear codes generated by a fixed coding matrix \(F \in \mathbb{R}^{MB \times N}\) drawn from the ensemble of random matrices with i.i.d Gaussian entries with distribution \(\mathcal{N}(0, 1/L)\). The codeword \(F_s \in \mathbb{R}^L\) and the cardinality of the code is \(B^L\). Hence, the (design) rate is \(R = L \log_2(B)/M = N \log_2(B)/(MB)\). The code is thus specified by \((M, R, B)\). The rate \(R\) can be linked to the “measurement rate” \(\alpha\), used in the compressive sensing literature [8], by \(\alpha := M/N = \log_2(B)/(BR)\).

We want to communicate through a known memoryless channel \(W\). This requires to map the codeword components \(\mathbb{R}^{M \times N}\) onto the input alphabet of \(W\). Call \(\pi\) this map (see Sec. V for various examples). The concatenation of \(\pi\) and \(W\) can be seen as an effective memoryless channel \(P_{\text{out}}\), such that \(P_{\text{out}}(y|F_s) = W(y|\pi(F_s))\). In the present framework, it is more convenient to work with this effective memoryless channel \(P_{\text{out}}(y|F_s) = \prod_{\mu=1}^{M} P_{\text{out}}(y|\pi(F_s))\), from which the receiver obtains the noisy channel observation \(y\).

We now present the spatially coupled ensemble of SS codes. We consider SC codes based on coding matrices in \(\mathbb{R}^{M \times N}\) made of \(\Gamma \times \Gamma\) blocks indexed by \((r, c)\), each with \(N/\Gamma\) columns and \(M/\Gamma = \alpha N/\Gamma\) rows. This ensemble of matrices is parametrized by \((M, R, B, \Gamma, w, g_w)\), where \(w\) is the coupling window and \(g_w\) is the design function. This is any function verifying \(g_w(x) = 0\) if \(|x| > 1\) and \(g_w(x) \geq g_0 > 0\) else, which is Lipschitz continuous on its support with Lipschitz constant \(g_\star\) independent of \(w\). From
ward application of the data processing inequality for Fisher
is induced by the code construction and depends on the
sent through an effective AWGN channel with a noise
s
i
where the expectation
ˆ
induces homogeneous power over the codeword components,
∑
J

g

r,c

1
1
r
. This normalization
in the message,
∑
J

J

r,c

r,c
= 1 \forall r. This normalization
induces a block structure in the message, 
| 1 : \Gamma
. In each of these blocks there is
L/\Gamma
sections. We assume that the sections in the first and last
4w
blocks of the message are known by the decoder. This
seed initiates a decoding wave in the SC code that propagates
inward through the entire message. The seed induces a rate
loss in the effective rate
R_{\text{eff}} = R(1 - \frac{8w}{T}) of the code, but this loss vanishes as 
\Gamma \rightarrow \infty.

III. STATE EVOLUTION AND POTENTIAL FORMULATION

The decoder is the GAMP algorithm, a generalization of
AMP to generic memoryless channels, introduced for esti-
mation of scalar signals with i.i.d components [14]. In the
present context the message components are correlated through
\rho_0(s_i), therefore we extend GAMP to cover this vectorial
setting (similarly to [7] for AMP). We first give the SE
equations associated with the underlying and SC ensembles.
SE is conjectured to track the performance of the vectorial
extension of the GAMP decoder (see Sec. VI). We then define
an appropriate potential function for each ensemble.

A. State evolution

The goal is to iteratively compute the average mean square
error (MSE) \hat{E}^{(t)} := \mathbb{E}_{s,y} \left[ \frac{1}{T} \sum_{t=1}^{T} ||s^{(t)} - \hat{s}||^2 \right] of the GAMP
estimate \hat{s}^{(t)} at iteration \( t \). We first need some definitions.

Definition 3.1 (Effective noise): Let us define the effective
noise variance \Sigma(E)^2 by the relation
\Sigma(E)^{-2} := \mathbb{E}_{p,E}[F(p|E)]\frac{R}{R}, \quad (1)
where the expectation \mathbb{E}_{p,E} is w.r.t. \mathbb{N}(0,1 - E)
and
\[ F(p|E) := \int df(y|p,E) (\partial_x \ln f(y|x,E))^2_{x=p} \quad (2) \]
is the Fisher information of \( p \) associated with the distribution
\( f(y|p,E) := \int du P_{\text{out}}(y|u) \mathbb{N}(u|p,E) \).

Lemma 3.2: \Sigma(E)^2 is non negative and increasing with E.

Proof: Positivity of the Fisher information implies
\Sigma(E)^2 \geq 0. The proof that it is increasing is a straightfor-
ward application of the data processing inequality for Fisher
information (Corollary 6 in [15]).

From now on, \( s \sim \rho_0(s) \) and \( z \sim \mathbb{N}(0,1) \) are
\( B \)-d random vectors and \( z \sim \mathbb{N}(\hat{z}|0,1) \), with expectations noted \( \mathbb{E}_s, \mathbb{E}_z \).

Definition 3.3 (Denoiser): The denoiser \( g_{m,i}(s,z) \) is
the MMSE estimator of the \( i \)-th component of a section
\( s \) sent through an effective AWGN channel with a noise
\( \mathbb{N}(0,1/E^2 \Sigma_{\text{GAMP}}(B)) \). Note that the effective AWGN
channel is induced by the code construction and depends on the
effective channel \( P_{\text{out}} \) only through \( \Sigma \). For any \( B \)-d prior, we
have for \( i \in \{1 : B\} \)
\[ g_{m,i}(s,z) := \int dx \rho_0(x) \theta(x,s,z,\Sigma) x_i \int dx \rho(x) \theta(x,s,z,\Sigma), \quad (3) \]
where \( \theta(x,s,z,\Sigma) := \exp \left( - \frac{||x-(s+i\Sigma)/\sqrt{\log_2(B)}||^2}{2\Sigma} \right) \). Using the
prior \( \rho_0(x) = \frac{1}{\sqrt{B}} \sum_{j \neq i} \delta_{rj}, \) \( i \) recoveres the
denoiser of SS codes [1].

Definition 3.4 (SE of the underlying system): The SE
operator of the underlying system is the average MSE associated
with the MMSE estimator of the effective channel,
\[ T_u(E) := \mathbb{E}_{s,z} \left[ \sum_{i=1}^B \left( g_{m,i}(s,z,\Sigma(E)) - s_i \right)^2 \right]. \quad (4) \]
The SE tracking the performance of the GAMP decoder is
\( \hat{E}^{(t+1)} = T_u(\hat{E}^{(t)}) \) for \( t > 0 \) and is initialized with \( \hat{E}^{(0)} = 1 \).
The existence of a fixed point is ensured by the monotonicity
and boundedness of the SE iterations, see Sec. IV.

Definition 3.5 (MSE Floor): The MSE floor \( E_0 \) is the fixed
point reached from trivial initial condition, \( E_0 = T_u(\infty)(0) \).

Definition 3.6 (Bassin of attraction): The basin of attraction
of the MSE floor \( E_0 \) is \( \mathbb{Y}_0 := \{ E | T_u(\infty)(E) = E_0 \} \).

Definition 3.7 (Threshold of underlying ensemble): The
GAMP threshold is \( R_u := \sup \{ \hat{R} > 0 | T_u(\infty)(1) = E_0 \} \).
For the present system, one can show that the only two possible
fixed points are \( T_u(\infty)(0) \) and \( T_u(\infty)(1) \). For \( R < R_u \), there is only one fixed point, namely the “good” one
\( T_u(\infty)(0) = E_0 \), and as the section size \( B \) increases \( E_0 \) and
the section error rate (that is the fraction of wrongly decoded
sections) vanish. Instead if \( R > R_u \), the GAMP decoder is
blocked by the “bad” fixed point \( T_u(\infty)(1) \neq E_0 \).

For a SC system, the performance of GAMP is described by
an average MSE profile \( \hat{E}^{(t)}(c = 1 : \Gamma) \) along the “spatial
dimension” indexed by the blocks of the message. To reflect
the seeding at the boundaries, we enforce the pinning condition
\( \hat{E}_c^{(t)} = 0 \) for \( c \in \{1 : 2w\} \cup \{\Gamma - 2w + 1 : \Gamma\} \), at all times.
Elsewhere, \( \hat{E}_c^{(t)} := \mathbb{E}_{s,y} \left[ \sum_{t=1}^{T} ||s^{(t)} - \hat{s}||^2 \right] \), where the sum
\( l \in c \) is over the set of indices of the \( L/\Gamma \) sections composing
the \( c \)-th block of \( s \). It turns out that the change of variables
\( E_r^{(t)} := \frac{1}{\Gamma} \sum_{t=1}^{T} \hat{E}_c^{(t)} \) makes the problem mathematically
more tractable. \( E \) is called a profile. The pinning condition
becomes \( E_r^{(t)} = 0 \) for \( r \in \mathbb{R} := \{1 : 3w\} \cup \{\Gamma - 3w + 1 : \Gamma\} \),
and at all times. In order to define the SE of the SC system,
we need first the following definition.

Definition 3.8 (Per-block effective noise): The per-block
effective noise variance \Sigma_c(E)^2 is \( \forall c \in \{1 : \Gamma\} \) defined by
\[ \Sigma_c(E)^{-2} := \sum_{r=1}^{\Gamma} \frac{J_{r,c}}{\Gamma \Sigma_c(E)^2} = \sum_{r=1}^{\Gamma} \frac{J_{r,c}}{\Gamma \Sigma_c(E)^2} \mathbb{E}_{s,y} \left[ F(p|E_r) \right]. \quad (5) \]

Definition 3.9 (SE of the coupled system): The vector valued
coupled SE operator is defined componentwise as
\[ [T_c(E)]_r := \sum_{c=1}^{\Gamma} \frac{J_{r,c}}{\Gamma} \mathbb{E}_{s,z} \left[ \sum_{i=1}^B (g_{m,i}(s,z,\Sigma_c(E)) - s_i)^2 \right]. \quad (6) \]
The SE for \( r \notin \mathbb{R} \) then reads \( E_r(t+1) = [T_c(E(t))]_r \) for \( t \geq 0 \). 
For \( r \in \mathbb{R} \), the pinning condition \( E_r(0) = 0 \) is enforced at all times. 
SE is initialized with \( E_r(0) = 1 \) for \( r \notin \mathbb{R} \).
Let \( E_0 := [E_r = E_0 | r = 1 : \Gamma] \) be the MSE floor profile.

**Definition 3.10 (Threshold of coupled ensemble):** The GAMP threshold of the SC system is defined as \( R_c := \lim \inf_{r \to \infty} \sup \{R > 0 | T_c(\infty)(1) < E_0 \} \) where \( 1 \) is the all ones vector. Here the limit \( \lim \inf_{r \to \infty} \) is taken along sequences where first \( \Gamma \to \infty \) and then \( w \to \infty \) (see Definition 4.1 for the meaning of \( \sim \)).

**B. Potential formulation**

The fixed point equations associated with SE can be reformulated as stationary point equations of potential functions (obtained from the replica method [5] or integrating SE). 

**Definition 3.11 (Potentials):** The potential of the underlying ensemble is \( F_u(E) := U_u(E) - S_u(\Sigma(E)) \), with
\[
U_u(E) := \frac{E}{2 \ln(2) |E|} - \frac{1}{R} E [\int dx P_{\text{out}}(y|x) N(x|\sqrt{1 - E}, E)]
\]
and the potential of the SC ensemble is \( F_c(E) := U_c(E) - S_c(\Sigma(E)) \), where \( U_c(E) := \sum_{r} U_u(E_r) \) and \( S_c(\Sigma(E)) := \sum_{z} S_u(\Sigma(z)) \).

**Definition 3.12 (Free energy gap):** The free energy gap is \( \Delta F_u := \inf_{E \in \mathbb{E}_0} F_u(E) - F_u (E_0) \), with the convention that the infimum over the empty set is \( \infty \) (i.e., when \( R < R_c \)).

**Definition 3.13 (Potential threshold):** The potential threshold is \( R_{\text{pot}} := \sup \{R > 0 | \Delta F_u > 0\} \).

The next Lemma links the potential and SE formulations.

**Lemma 3.14:** One can show that if \( T_u(E) = \hat{E} \), then \( \frac{\partial E_r}{\partial E} \bigg|_{E} = 0 \). Similarly for the SC system, if \( T_c(E) = \hat{E} \), \( \forall r \in \mathbb{R}^c = \{3w + 1 : \Gamma - 3w\} \) then \( \frac{\partial E_r}{\partial E} \bigg|_{E} = 0 \forall r \in \mathbb{R}^c \).

We end this section by pointing out that the terms composing the potentials have natural interpretations in terms of effective channels. The term \( \mathbb{E}_z[\int dx P_{\text{out}}(y|x) N(x|\sqrt{1 - E}, E)] \) is the conditional entropy \( H(Y|Z) \) for the concatenation of the channels \( N(x|\sqrt{1 - E}, E) \) and \( P_{\text{out}}(y|x) \) with a standardised input \( N(z|0, 1) \). The term \( S_u(\Sigma(E)) \) \( \log_2(B) \) is equal to minus the mutual information \( I(S; Y) \) for the Gaussian channel \( N(y|z, \Sigma)/\log_2(B) \) and input distribution \( p_0(s) \), up to a constant factor \(-2 \ln(2)\)^{-1}.

**IV. Sketch of the Proof of Threshold Saturation**

Monotonicity properties of the SE operators \( T_u \) and \( T_c \) are key elements in the analysis.

**Definition 4.1 (Degradation):** A profile \( E \) is degraded (resp. strictly degraded) w.r.t another one \( G \), denoted as \( E \succ G \) (resp. \( E \succ \succ G \)), if \( E_r \geq G_r \forall r \) (resp. if \( E \geq G \) and there exists some \( r \) such that \( E_r > G_r \)).

**Lemma 4.2:** The SE operator of the coupled system maintains degradation in space, i.e. if \( E \succ G \), then \( T_c(E) \succ T_c(G) \). It also maintains degradation in time, i.e. \( T_c(E(t)) \preceq E(t) \Rightarrow T_c(E(t+1)) \preceq E(t+1) \). Similarly \( T_u(E(t)) \preceq E(t) \Rightarrow T_u(E(t+1)) \preceq E(t+1) \). Furthermore, the limiting profile \( E(\infty) := T_c(\infty)(E(0)) \) exists. These properties are verified by \( T_u \) for a scalar error as well.

**Proof:** Combining Lemma 3.2 with (5) implies that if \( E \geq G \), then \( \Sigma_c(E) \geq \Sigma_c(G) \). The rest of the proof is similar to the one of Lemma 4.2 and Corollary 4.3 in [1].

The pinning condition together with the monotonicity properties of the coupled SE imply that its fixed point profile \( E^* \) must adopt a shape similar to Fig. 1. We associate to \( E^* \) a saturated profile \( E \) (see Fig. 1) that verifies by construction \( E \geq E^* \). Thus \( E \) serves as an upper bound in our proof.

**Definition 4.3 (Shift operator):** The shift operator is defined componentwise as \( S(E) := E_0, S(E)|_r := E_r - 1 \).

**Lemma 4.4:** Let \( E \) be a saturated profile. Then the coupled potential verifies \( |F_c(S(E)) - F_c(E)| < K/w \), where \( K \) is independent of \( w \) and \( \Gamma \).

**Proof:** The proof uses Lemmas 5.2, 5.3 and 5.4 of [1], where Lemma 5.2 is implied by the present Lemma 3.14 and Lemma 5.3 remains valid as it depends only on the SC construction. Lemma 5.4 can be shown to be true for any memoryless channel \( P_{\text{out}} \) such that the function \( g_{\text{out}} := \frac{\partial}{\partial y} \ln(\int dx P_{\text{out}}(y|x) N(x|\rho, v)) \) [14] is Lipschitz continuous in \( p \) with Lipschitz constant independent of the coupling window.

**Lemma 4.5:** Let \( E \) be a saturated profile such that \( E \succ E_0 \). Then \( F_c(S(E)) - F_c(E) \leq -\Delta F_u \).

**Proof:** See the proof of Lemma 5.6 in [1].

**Theorem 4.6:** Assume a spatially coupled SS code ensemble is used for communication through a memoryless channel. Fix \( R < R_{\text{pot}}, w > K/\Delta F_u \) (\( K \) is independent of \( w \) and \( \Gamma \)) and \( \Gamma > 8w \) (such that the code is well defined). Then any fixed point profile \( E^* \) of the coupled SE satisfies \( E^* \prec E_0 \).

**Proof:** It follows from Lemma 4.4 and 4.5 as in [1].

**Corollary 4.7:** By first taking \( \Gamma \to \infty \) and then \( w \to \infty \), the GAMP threshold of the coupled SS codes saturates to the potential threshold.

We emphasize that Theorem 4.6 and Corollary 4.7 hold for a large class of estimation problems with random linear mixing [14]. Both the SE and potential formulations of Sec. III as well as the proof sketched in the present section are not restricted to SS codes. Indeed all the definitions and results are obtained for any memoryless channel \( P_{\text{out}} \) and any factorizable (over \( B \)-d sections, \( B \in \mathbb{N} \)) prior over the message (or signal) \( s \).
Let \( A \) and \( B \) be the input and output alphabet of \( W \) respectively, where \( A, B \subseteq \mathbb{R} \) are defined over discrete or continuous supports. Call \( P \) the capacity-achieving input distribution associated with \( W \). Choose \( \pi : \mathbb{R} \to A \) such that i) \( P_{out}(y|z) = W(y|\pi(z)) \) and ii) if \( z \sim \mathcal{N}(z|0,1) \), then \( \pi(z) \sim P \). This map converts a standard Gaussian random variable \( z \) onto a channel-input random variable \( \pi(z) = a \) with capacity-achieving distribution \( P(a) \). Note that \( \pi \) can be viewed equivalently as part of the code or of the channel.

Now using the relation \( \int Dz P_{out}(y|z) = \int Dz W(y|\pi(z)) = \int daP(a)W(y|a) \) (9) can be expressed equivalently as

\[
R_{\text{pot}}^\infty = -\int dydaP(a)W(y|a) \log_2 \left( \int daP(a)W(y|a) \right)
+ \int dydaP(a)W(y|a) \log_2 \left( W(y|a) \right).
\]

The first term in (10) is nothing but the Shannon entropy \( H(Y) \) of the channel output-distribution, while the second term is the negative of the conditional entropy \( H(Y|A) \) of the channel-output distribution given the input \( A = \pi(Z) \), that has capacity-achieving distribution. Thus, \( R_{\text{pot}}^\infty \) is the Shannon capacity of \( W \). Combining this result with Corollary 4.7, we can assert that SC-SS codes allow to communicate reliably up to Shannon’s capacity over any memoryless channel under low complexity GAMP decoding.

But how to find the proper map \( \pi \) for a given memoryless channel? In the case of discrete input memoryless symmetric channels, Shannon’s capacity can be attained by inducing a uniform input distribution \( P = \mathcal{U}_A \). Let us call \( q \) the cardinality of \( A = \{a_1 : a_q\} \). In this case the mapping \( \pi \) is simply \( \pi(z) = a_i \) if \( z \in [z_{i-1}/q, z_i/q) \), where \( z_i/q \) is the \( i \)th \( q \)-quantile of the Gaussian distribution, with \( z_0 = -\infty, z_1 = \infty \). For asymmetric channels, one can use some standard methods such as Gallager’s mapping or more advanced ones [17] that introduce bias in the channel-input distribution in order to match the capacity-achieving one. We now illustrate these findings, depicted for various channels in Fig. 3 and Fig. 4.

**AWGN channel:** We start showing that our results for the AWGN channel [1] are a special case of the present general framework. No map \( \pi \) is required and the Shannon capacity is directly obtained from (9) because the capacity-achieving input distribution for the AWGN channel is Gaussian. Thus, by plugging \( P_{out}(y|z) = \mathcal{N}(y|z,1/\text{snr}) \) in (9), one recovers

\[
R_{\text{pot}}^\infty = -\int dy\mathcal{N}(y|0,1) \log_2 \left( \int dy\mathcal{N}(y|0,1) \right)
+ \int dy\mathcal{N}(y|0,1) \log_2 \left( \mathcal{N}(y|0,1) \right),
\]

and this is the Shannon capacity of the AWGN channel.

---

**Fig. 2.** The large alphabet potential \( \varphi_u(E) \) (7) as a function of the MSE for the BSC (left) and AWGN (right) channels with \( \epsilon = 0.1 \) and \( \text{snr} = 10 \) respectively. \( \varphi_u(E) \) is scaled such that \( \varphi_u(0) = 0 \). For \( R < R_u \), there is a unique minimum at \( E = 0 \) while just above \( R_u \), this minimum coexists with a local one at \( E = 1 \). At the optimal threshold of the code, that coincides with the Shannon capacity, the two minima are equal. Then, for \( R > C \) the minimum at \( E = 1 \) becomes the global one, and thus decoding is impossible.

**Fig. 3.** Large alphabet limits of the capacities and GAMP thresholds for the BSC (left) and AWGN (right) channels.
the Shannon capacity $R^\infty_{\text{pot}} = \frac{1}{2} \log_2(1 + \text{snr})$. Furthermore, one obtains $R^\infty_u = [2 \ln(2)/(1 + 1/\text{snr})]^{-1}$.

**BSC channel:** The binary symmetric channel (BSC) with flip probability $\epsilon$ has transition probability $W(y|a) = (1 - \epsilon)\delta(y - a) + \epsilon \delta(y + a)$, where both $y, a \in \{-1, 1\}$. The proper map is $\pi(z) = \text{sign}(z)$ since it induces uniform input distribution $\mathcal{U}_a = 1/2$. So by plugging $W$ and $\mathcal{U}_a$ in (10), or equivalently $P_{\text{out}}(y|z) = (1 - \epsilon)\delta(y - \pi(z)) + \epsilon \delta(y + \pi(z))$ into (9), one obtains the Shannon capacity of the BSC channel $R^\infty_{\text{pot}} = 1 - h_2(\epsilon)$ where $h_2$ is the binary entropy function. This map also gives $R^\infty_u = (\pi \ln(2))^{-1}((1 - 2\epsilon^2)^{-1}$.

**BEC channel:** Note that the binary erasure channel (BEC) is also symmetric. Therefore, the same mapping $\pi(z) = \text{sign}(z)$ is used and leads to the Shannon capacity $R^\infty_{\text{pot}} = 1 - \epsilon$, where $\epsilon$ is the erasure probability, and $R^\infty_u = \pi \ln(2)^{-1}(1 - \epsilon)$. The Z channel is the extremal discrete asymmetric channel, in the sense that it represents the “worst” one. It has binary input and output in $\{-1, 1\}$ with transition probability $W(y|a) = \delta(a - 1)\delta(y - a) + \delta(a + 1)(1 - \epsilon)\delta(y - a) + \epsilon \delta(y + a)$, where $\epsilon$ is the flip probability of the $-1$ input. The map $\pi(z) = \text{sign}(z)$ leads to the symmetric capacity of the Z channel $R^\infty_{\text{pot}} = h_2((1 - \epsilon)/2) - h_2(\epsilon/2)$, that is the input-output mutual information when the input is uniformly distributed, and $R^\infty_u = [\pi \ln(2)]^{-1}(1 - \epsilon)$. This expression differs from Shannon’s capacity. However, one can introduce bias in the input distribution and hence match the capacity-achieving one. To do so, the proper map defined in terms of the $Q$-function is $\pi(z) = \text{sign}(z - Q^{-1}(p_1))$, where $p_1$ is the input probability of the bit 1. By optimizing over $p_1$, one can obtain the Shannon’s capacity of the Z channel for $p^*_1 = 1 - [(1 - \epsilon)(1 + \sqrt{2} \epsilon)/(1 - \epsilon)]^{-1}$.

**VI. OPEN CHALLENGES**

We end up pointing some open problems. In order to have a fully rigorous capacity achieving scheme over any memoryless channel, using SC-SS codes and GAMP decoding, it must be shown that the SE tracks the asymptotic performance of GAMP. We conjecture that it is indeed the case and that the proof follows from the method of [18], then extended in [10] for power allocated SS codes. It is also desirable to consider practical coding schemes, using Hadamard-based operators or more generally, row-orthogonal matrices. Another important point is to estimate at what rate the error floor vanishes when $B$ increases. Finally, the finite size effects should be considered in order to assess the real potential of these codes. We plan to settle these questions in future works.

**ACKNOWLEDGMENTS**

J.B and M.D acknowledge funding from the Swiss National Science Foundation grant num. 200021-156672. We thank Florent Krzakala, Rüdiger Urbanke and Christophe Schüllke for helpful discussions.

**REFERENCES**

[1] J. Barbier, M. Dia, and N. Macris, “Proof of Threshold Saturation for Spatially Coupled Sparse Superposition Codes,” ArXiv e-prints, Mar. 2016. [Online]. Available: http://arxiv.org/pdf/1603.08171.pdf

[2] 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.

[3] A. Joseph and A. R. Barron, “Fast sparse superposition codes have near exponential error probability for ReC,” 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 and F. Krzakala, “Replica analysis and approximate message passing decoder for superposition codes,” in Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on, 2014.

[6] 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.

[7] J. Barbier and F. Krzakala, “Approximate message-passing decoder and capacity-achieving sparse superposition codes,” 2015. [Online]. Available: http://arxiv.org/abs/1503.08040

[8] F. Krzakala, M. Ménard, 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.

[9] F. Caltagirone and L. Zdeborová, “Properties of spatial coupling in compressed sensing,” CoRR, vol.abs/1401.6380, 2014.

[10] C. Rush, A. Greig, and R. Venkataramanan, “Capacity-achieving sparse regression codes via approximate message passing decoding,” in Information Theory (ISIT), 2015 IEEE International Symposium on, June 2015, pp. 2016–2020.

[11] J. Barbier, C. Schülke, and F. Krzakala, “Approximate message-passing decoding for spatially-coupled ldpc and ldgm codes on bms channels,” IEEE Trans. on Information Theory, vol. 60, pp. 7389–7415, 2013.

[12] A. Yedla, Y.-Y. Jian, P. Nguyen, and H. Pfister, “A simple proof of threshold saturation for coupled scalar recursions,” in 7th International Symposium on Turbo Codes and Iterative Information Processing (ISTC), 2012, pp. 51–55.

[13] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, “Threshold saturation for spatially-coupled ldpc and ldgm codes on bms channels,” IEEE Trans. on Information Theory, vol. 60, pp. 7389–7415, 2013.

[14] A. Yedla, Y.-Y. Jian, P. Nguyen, and H. Pfister, “A simple proof of maxwell saturation for coupled scalar recursions,” Information Theory, IEEE Trans. on, vol. 60, no. 11, pp. 6943–6965, 2014.

[15] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on. IEEE, 2011, pp. 2168–2172.

[16] P. Zegers, “Fisher information properties,” Entropy, vol. 17, no. 7, p. 4918, 2015. [Online]. Available: http://mdpi.com/1099-4304/17/7/4918

[17] J. Barbier, “Statistical physics and approximate message-passing algorithms for sparse linear estimation problems in signal processing and coding theory,” Ph.D. dissertation, Université Paris Diderot, 2015. [Online]. Available: http://arxiv.org/abs/1511.01650

[18] M. Mondelli, R. Urbanke, and S. H. Hassani, “How to achieve the capacity of asymmetric channels,” in Communication, Control, and Computing, 2014 Allerton Conference on. IEEE, 2014, pp. 789–796.

[19] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” IEEE Trans. on Information Theory, vol. 57, no. 2, pp. 764–785, 2011.