Encoding of cellular positional information and maximum capacity of parallel coupled channels

Eduard A. Jorswieck*,†, Andreas I. Reppas†, Haralampos Hatzikirou†
Technische Universität Dresden
*School of Engineering Sciences
Communications Laboratory, Communications Theory
†Center for Advancing Electronics Dresden
{eduard.jorswieck,andreas.reppas,haralambos.hatzikirou}@tu-dresden.de

***Abstract*—Inspired by the parallels between information coding in morphogenesis and information coding in computer communication, we introduce a new model for coupled discrete memoryless channels in which the error probability of one channel depends on the output of the other channel. The model is motivated by a type of cell-cell communication. It is shown that coupling will lead to higher sum capacities with both optimal input distribution and with uniform input distribution under joint coding. Thereby, nature can achieve more than the sum of the individual capacities (synergistic effect). We compare this result with the maximum achievable sum capacity by arbitrary ideal coupling using Majorization theory. Finally, we illustrate the model with applications from wireless communications.**

I. Introduction

The development of multicellular organisms requires the cooperation and coordination of neighbouring cells [1]. Cell-cell coordination and specification depends on extracellular signals transmitted through signal transduction pathways [2]. Recently, it was observed in [3] that cell-cell signalling is close to the information theoretic model of communication which was developed in [4]. In this regard, several information theoretical based approaches were proposed to model, analyse an design the inter-cellular communications (see e.g. [5], [6], [7] for recent reviews).

Recently, parallels have been identified between cell fate decision making and information coding in computer communication [8]. An important concept is positional information that is necessary for pattern formation robustness. Signal transduction pathways process the received signal in such a way that increase the fidelity of cell fate specification [2], [9]. Positional information, in terms of the communication theory view, can be interpreted as the encoding of multiple time-varying signals measured in bits [10]. The latter coincides with the channel capacity as the maximum number of bits which can be reliably transmitted over an asymptotic large number of channels.

In this paper, we introduce a new model for coupled discrete memoryless channels (DMC) motivated by the interaction between different signal transduction pathways. We claim that this kind of cell-cell communication induces a synergistic effect that leads to increased information rates. We support the latter statement by using a specific cell fate specification pathway, the Notch-Delta.

In the proposed model, the error probability (or fidelity) of one channel depends on the output of the other channel. It is shown that coupling will lead to higher sum capacities with both optimal input distribution and with uniform input distribution. Thereby, we achieve more than the sum of the individual capacities.

Note that the channel models can be also formulated using continuous input and output distributions, e.g., by additive white Gaussian noise channels with different error variances as fidelity parameter. The choice depends on which model fits better to the noise process in signal transduction pathways.

Therefore, we study the achievable rates of a hybrid model with DMC and Gaussian channel as a special case. A similar synergistic effect can be observed here and the DMC can be used to 'activate' the Gaussian channel. The relation to resource allocation in heterogeneous wireless networks is also briefly mentioned.

A. Biological motivation: The Notch-Delta interaction

The Notch-Delta signalling pathway aims at the specification of a cell's fate. It is responsible for the mutual inactivation between interacting cells which results in a pattern formation in a multicellular context [11]. The whole process is based on a lateral inhibition mechanism where a cell tending to adopt a particular fate inhibits its interacting neighbours from following the same fate [12]. Thus, according to the above mechanism, a cell's fate affects the fate of its interconnected cells. In this paper, we explore, through a simple mathematical model of Notch-Delta communication with noise, the conditions which increase the fidelity of a cell to finally adopt a specific fate.

Mathematical models of Notch-Delta signalling have been introduced to investigate the underlying mechanisms of the interaction [11] as well as the pattern formation potential [13]. Here, we model the effective lateral inhibition of Notch-Delta interaction, similar to the one introduced in [13]: *Delta* in one

The authors gratefully acknowledge support from the German Excellence Initiative via the Cluster of Excellence EXC 1056 Center for Advancing Electronics Dresden (cfAED).

cell activates *Notch* in its neighboring cells [14]. *Notch*, inside a cell, is suppressed by *Delta* activity [15] (Figure 1).

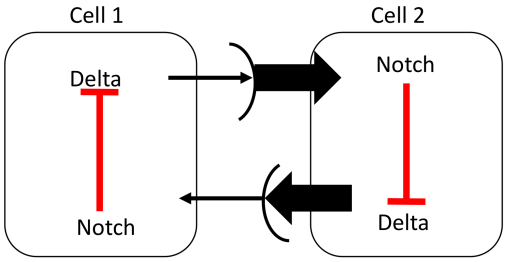

Fig. 1. Notch-Delta signalling pathway in neighboring cells.

Denoting the levels of *Notch* and *Delta* in a cell i by N_i and D_i respectively, we write

$$\frac{dN_i}{dt} = \alpha\bar{D} + \eta_{int}(t) - k_1 N_i, \tag{1a}$$

$$\frac{dD_i}{dt} = \frac{1}{1+\beta {N_i}^h} - k_2 D_i = f(N_i, D_i), \tag{1b}$$

where $\bar{D}$ denotes the mean value of the levels of *Delta* activity in the cells which interact with the cell i, $\bar{D} = \frac{1}{|\mathfrak{N}_i|}\sum_{j\in\mathfrak{N}_i} D_j$, where $\mathfrak{N}_i$ denotes the neighbourhood of cell i and $|\mathfrak{N}_i|$ specifies the number of its neighbours. Parameters k_1, k_2 are the intracellular degradation rates of *Notch* and *Delta* respectively and β, h are positive constants. Parameter α models the strength of the interaction which we assume it is a normal random variable written as: $\alpha = \lambda + \eta_{ext}(t)$, where $\eta_{ext}(t) \sim \mathcal{N}(0, \sigma^2_{ext})$ represents the extrinsic noise. Similarly, η_{int} represents the intrinsic noise of the interaction of Notch and Delta inside the cell which we also assume it is a normal random variable, $\eta_{int}(t) \sim \mathcal{N}(0, \sigma^2_{int})$. Let $k = ext, int$ we have

$$\mathbb{E}_t[\eta_k(t_1)\eta_k(t_2)] = \sigma_k^2\delta(t_1 - t_2), \tag{2}$$

where $\mathbb{E}_t$ denotes averaging over time.

Please note that both intrinsic and extrinsic noise can deviate from the assumption of normality. The extrinsic noise refers to the fluctuations in receiving free Delta ($\bar{D}$) by a cell. The intrinsic noise characterizes the involved stochasticity in the intracellular gene regulation of Notch-Delta expression.

Separating the deterministic and stochastic part of (1a) we derive the following equations

$$\frac{dN_i}{dt} = \lambda\bar{D} - k_1 N_i + \bar{D}\eta_{ext}(t) + \eta_{int}(t), \tag{3a}$$

$$\frac{dD_i}{dt} = f(N_i, D_i). \tag{3b}$$

As we can see from equation (3a), the total noise can be expressed as the sum of the intrinsic and extrinsic noise, $\eta(t) = \bar{D}\eta_{ext}(t) + \eta_{int}(t)$. The fidelity maximization (which here is equivalent with the total noise minimization) occurs only when the intrinsic and extrinsic noise are negatively correlated.

A plausible question is what biological mechanisms underlie this noise minimization. As it has been shown in [16], the negative correlation between the internal and external signalling has a characteristic time that is associated with a time delay in the inhibition process. In other words, when a cells detects an excess of free Delta, i.e. all Notch receptors are occupied at its membrane, then it produces more receptors. This state has been characterized as a receiver one. In the contrary, when Notch receptors are in excess the cell destroys part of them resulting in the production and release of Delta (sender state) [11]. This correlation has been also identified in further gene regulation systems [17].

The above provides an illustrative answer to the question how a cell that experiences noisy signals and process them in a stochastic way decides for its fate so robustly.

B. System Model

Based on the observation in the previous subsection, we present our system model in Figure 2 for the general communications scenario over K parallel coupled links.

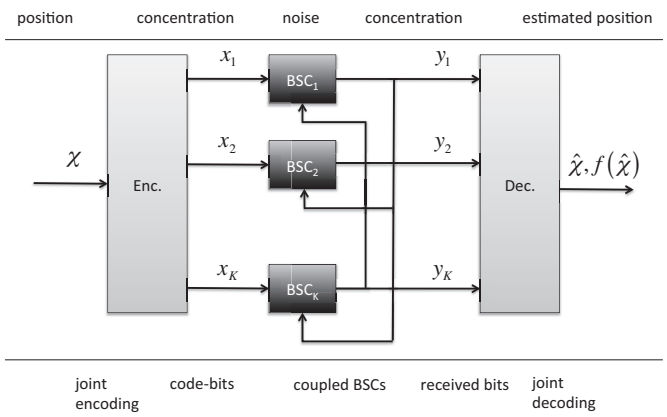

Fig. 2. General system model of K parallel coupled BSCs.

The model in Figure 2 is described on the upper boundary with its operational meaning from information coding in morphogenesis and pattern formation. On the lower boundary its operational meaning from information coding in computer communications is displayed.

The position to be encoded into K concentration expressions is χ. The corresponding concentrations (codewords) are $x_1, ..., x_K$. The encoding can be done jointly for all K parallel channels. They enter the set of parallel BSCs which introduce the noise during the reception of the concentrations. Their fidelities are coupled by the output concentrations $\boldsymbol{y} = [y_1, ..., y_K]$. The concentration y_k is then measured at the output of BSC k and a joint decoder produces an estimate of the position $\hat{\chi}$ or a function which corresponds to the position dependent response of the cell $f(\hat{\chi})$. In the following, we describe the operation of the coupling based on the multi-channel signal transduction.

II. Multi-Channel Signal Transduction

In general, multiple signal transduction pathways are active in parallel. Interestingly, they do influence each other by positive or negative effects on the receptor fidelity and the equivalent noise in the signal pathways [2]. In order to understand the design decision of such parallel multi-channel communications, we study the capacity of independent as well as coupled parallel channels. Then, we show that by smart influencing each others fidelity, the capacity of two channels can be more than the sum of their individual capacities.

A. Maximum Sum Capacity of Coupled Parallel BSCs

To motivate the model and problem statement, let us start with the analysis of K independent parallel binary symmetric channels (BSC) with error probability of $\epsilon_1, ..., \epsilon_K$. Collect the error probabilities in a vector $\boldsymbol{\epsilon} = [\epsilon_1, ..., \epsilon_K]$. The sum capacity of these channels is given by [18]

$$C(\boldsymbol{\epsilon}) = \sum_{k=1}^{K} C_{BSC}(\epsilon_k) = \sum_{k=1}^{K} \left(1 - \mathbb{H}_b(\epsilon_k)\right), \tag{4}$$

with binary entropy function $\mathbb{H}_b(x) = -x \log_2 x - (1-x)\log_2(1-x)$. The optimal input distribution to achieve the capacity in (4) is a uniform input distribution on each parallel channel. Furthermore, independent coding can be applied on each parallel channel.

In order to understand the impact of coupling, we allow to choose the error probabilities freely under a sum constraint, i.e., $\sum_{k=1}^{K} \epsilon_k = \text{const.}$ and $0 \leq \epsilon_k \leq 1/2$ and ask about the maximum and minimum achievable sum rate. The floor function is denoted by $\lfloor \cdot \rfloor$.

Proposition 1. *The sum capacity of K parallel BSCs with error probabilities $\boldsymbol{\epsilon} \in \mathcal{E}^{(c)} = \{\boldsymbol{\epsilon} \in \mathbb{R}_0^+ : 0 \leq \epsilon_k \leq 0.5, \sum_{k=1}^{K} \epsilon_k = c \leq 1/2K\}$ is upper and lower bounded by*

$$\begin{aligned} K\left(1 - \mathbb{H}_b\left(\frac{c}{K}\right)\right) &= C\left(\frac{c}{K}\mathbf{1}\right) \leq C(\boldsymbol{\epsilon}) \\ &\leq C([\underbrace{1/2, ..., 1/2}_{L}, c - L/2, 0..., 0]) \\ &= K - L + 1 - \mathbb{H}_b(c - L/2), \end{aligned} \tag{5}$$

where $L = \lfloor 2c \rfloor$.

As special cases of Proposition 1, the following two cases are:

- if $c = 1/2K$ then upper and lower bound are equal and the sum capacity is equal to zero, because $C(\frac{1}{2}\mathbf{1}) = C(1/2, ..., 1/2) = 0$.
- if $c \leq 1/2$ then the upper bound simplifies to $C(c, 0, ..., 0) = K - \mathbb{H}_b(c)$.

We observe from (5) that for K identical parallel BSCs the worst case lower bound is achieved. A much higher performance could be achieved if the fidelity of one channel is reduced and the fidelity of another channel is increased. Note that these upper and lower bounds are derived under the ideal assumption that the error probability of one BSC can be traded for the error probability of another parallel BSC.

In [19], the channel polarization technique is described for a set of N binary input channels. For growing N a similar observation as in the upper bound in (5) is obtained: A subset of channels approaches the rate one and the other subset approaches zero rate. The main difference is that there special channel codes are applied whereas in our model the coupling is created from the channel itself.

In the following subsections, we develop models for coupled BSC and test whether the achievable sum rates are close to the theoretical upper bound in (5).

B. Coupled BSC Model

We consider the following set of two coupled BSCs: BSC1 and BSC2 each with error probability ϵ. We have described in Subsection I-A, and it is observed in [2] and [9] , that the presence of one signal can influence the detection fidelity of the other signal.

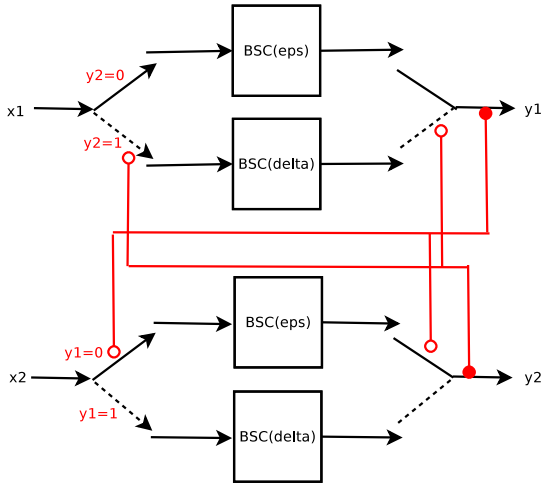

Fig. 3. The proposed coupled BSC model for the interaction between different signalling pathways.

We use the model in Figure 3: There are two different BSC available, a high fidelity and a low fidelity model. If the receiver one observes a signal from source two $y_2 = 1$, the fidelity for detecting a signal from source one x_1 is high, if the signal of source two is absent $y_2 = 0$, the fidelity for detecting the signal from source one is low. The corresponding transfer matrices for the BSC1 are

$$\begin{aligned} \mathcal{W}_1^{(y_2=0)} &= \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix} \\ \mathcal{W}_1^{(y_2=1)} &= \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix} \end{aligned} \tag{6}$$

with low fidelity ϵ and high fidelity δ, $0 < \delta < \epsilon \leq 1/2$. The same model applies for the second BSC

$$\mathcal{W}_2^{(y_1=0)} = \mathcal{W}_1^{(y_2=0)}, \qquad \mathcal{W}_2^{(y_1=1)} = \mathcal{W}_1^{(y_2=1)}. \tag{7}$$

We note that each individual channel considered separately belongs to the class of arbitrarily varying channels (AVC) [20] because the channel state y_2 for BSC1 changes from channel

use to channel use in an unknown manner. The state is not known a priori at the transmitter side when the codeword is constructed. The following results are known from [21] about the capacities of AVC: If the channel is *symmetrizable* the channel capacity is zero. Otherwise it is given by the following max-min problem

$$C = \max_{p(x)} \min_{p(y)} I(x_1; y_1) = C(\epsilon),$$

because both channels with low and high fidelity achieve capacity with uniform input distribution. We state this observation informally: *If independent coding is applied over both AVC, the benefits from coupling cannot be realized.*

C. Joint Coding Achieves Larger Sum Rates

The situation changes significantly if joint coding over both parallel channels is used. Therefore, let us consider the super channel $\mathcal{W}$ which is constructed as follows: The input is a vector of two bits x_1, x_2 (and has thereby four states) and the output is a vector of two bits y_1, y_2 (also four states). An input $(0,0)$ to this super channel leads with a probability of $(1-\epsilon)(1-\epsilon)$ to the output $(0,0)$ and with a probability of $(1-\epsilon)\epsilon$ to the output $(0,1)$. All other conditional probabilities are computed analogously. The resulting corresponding transfer matrix is illustrated in Table I.

The mutual information depends on the channel probabilities ϵ, δ and the input distribution $\boldsymbol{p} = [p_1, p_2, p_3, p_4]$ indicated by $C(\epsilon, \delta, \boldsymbol{p}) = I(\boldsymbol{x}; \boldsymbol{y})$. The maximum mutual information corresponding to the following programming problem

$$\max_{\boldsymbol{p} \geq 0, ||\boldsymbol{p}||=1} C(\epsilon, \delta, \boldsymbol{p}) \tag{8}$$

with

$$C(\epsilon, \delta, \boldsymbol{p}) = \sum_{k=1}^{4} p_k D(\boldsymbol{w}_k || \boldsymbol{p}\mathcal{W}) \tag{9}$$

with $\boldsymbol{w}_k$ as k-th row of the channel matrix $\mathcal{W}$. The capacity in (8) can be computed numerically by the Blahut-Arimoto algorithm [18]. To gain further insights into the optimal input distribution, we prove the following characterization.

Proposition 2. *The optimal input distribution for the DMC with channel transfer matrix in Table I has the following properties:*

1) *The inputs* $(0,1)$ *and* $(1,0)$ *have the same probability, i.e.,* $p_2 = p_{\boldsymbol{x}}(\boldsymbol{x} = [0,1]) = p_{\boldsymbol{x}}(\boldsymbol{x} = [1,0]) = p_3$.
2) *For large differences* $\Delta = 2(\epsilon - \delta) > \tilde{\Delta}$, *the input* $(0,0)$ *gets zero probability, i.e.,* $p_1 = p_{\boldsymbol{x}}(\boldsymbol{x} = [0,0]) = 0$.

The first statement in Proposition 2 follows from the symmetry of the second and third row in the channel matrix. The second statement is intuitive. It says that the larger the fidelity difference $\Delta = 2(\epsilon - \delta)$ is, the more likely it is to receive a $y_1 = 1$ or y_2. Since the input $(0,0)$ does only marginally contribute to the state $(1,1)$, it is switched off for large $\Delta > \tilde{\Delta}$. Note that the second statement is not a contradiction to the result in [22] because the input distribution between the remaining states $(0,1), (1,0), (1,1)$ is still to be optimized.

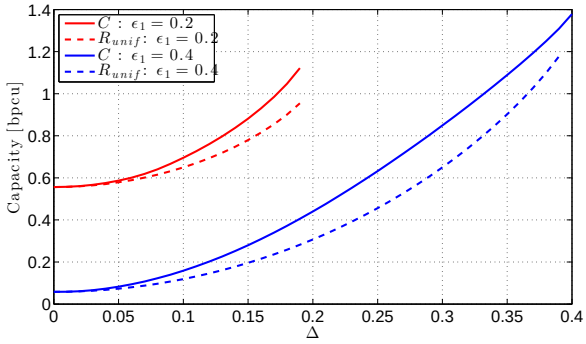

Fig. 4. Capacity and achievable rate with uniform distribution of two coupled BSC with $\epsilon = 0.4$ and $\epsilon_1 = 0.2$, respectively.

In Subsection II-E, numerical simulations illustrate the theoretical characterization from above and show the optimal power allocation found by the Blahut-Arimoto algorithm.

D. Uniform Input Distribution

As a special case, the input distribution to the coupled BSCs in Figure 3 is assumed to be uniform, i.e., $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$. This input distribution is robust against the channel parameters δ, ϵ. Then, the following transmission rates are achievable.

Proposition 3. *The sum rate achievable over the DMC characterized by the channel transfer matrix in Table I with uniform input distribution is the sum rate of the two channels* $\mathcal{W}_1^{y_2=0}$ *and* $\mathcal{W}_1^{y_2=1}$, *i.e.,*

$$C(\epsilon, \delta, 1/4 \cdot \mathbf{1}) = C(\epsilon) + C(\delta).$$

This shows that for the combined coupled BSC channels a uniform distribution is sufficient to obtain the sum rate results predicted and envisaged in Proposition 1 if ϵ and δ are sum constrained.

E. Numerical Illustration

For a fair comparison, we normalize the error probabilities ϵ and δ in the model above. In particular, we increase the fidelity in one channel state by the same amount Δ as we decrease the fidelity Δ in the other state, for BSC1 $\epsilon = \epsilon_1 + \Delta$ and $\delta = \epsilon_1 - \Delta$. Note that the channel in Table I is not symmetric. The resulting channel matrix as a function of ϵ and Δ is given by equation (10) on the top of the next page.

For $\Delta = 0$, the sum capacity of two independent parallel BSC is achieved. In Figure II-E, it can be observed that even with the suboptimal uniform input distribution, the achievable rate is increasing with the coupling Δ. The additional gain by an optimal input distribution increases with increasing Δ. Therefore, we conclude with the statement: *The capacity of coupled BSCs can be larger than the sum of the individual capacities.*

		0	1	2	3
	$(x_1,x_2)/(y_1,y_2)$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
0	$(0,0)$	$(1-\epsilon)^2$	$(1-\delta)\epsilon$	$(1-\delta)\epsilon$	δ^2
1	$(0,1)$	$(1-\epsilon)\epsilon$	$(1-\delta)(1-\epsilon)$	$\epsilon\delta$	$(1-\delta)\delta$
2	$(1,0)$	$(1-\epsilon)\epsilon$	$\epsilon\delta$	$(1-\epsilon)(1-\delta)$	$(1-\delta)\delta$
3	$(1,1)$	ϵ^2	$\delta(1-\epsilon)$	$\delta(1-\epsilon)$	$(1-\delta)^2$

TABLE I
CHANNEL MATRIX OF COMBINED COUPLED PARALLEL BSCS BY JOINT CODING.

$$\begin{pmatrix} (1-\epsilon_1-\Delta)^2 & (1-\epsilon_1+\Delta)(\epsilon_1+\Delta) & (1-\epsilon_1+\Delta)(\epsilon_1+\Delta) & (\epsilon_1-\Delta)^2 \\ (1-\epsilon_1-\Delta)(\epsilon_1+\Delta) & (1-\epsilon_1+\Delta)(1-\epsilon_1-\Delta) & (\epsilon_1-\Delta)(\epsilon_1+\Delta) & (1-\epsilon_1+\Delta)(\epsilon_1-\Delta) \\ (1-\epsilon_1-\Delta)(\epsilon_1+\Delta) & (\epsilon_1-\Delta)(\epsilon_1+\Delta) & (1-\epsilon_1+\Delta)(1-\epsilon_1-\Delta) & (1-\epsilon_1+\Delta)(\epsilon_1-\Delta) \\ (\epsilon_1+\Delta)^2 & (1-\epsilon_1-\Delta)(\epsilon_1-\Delta) & (1-\epsilon_1-\Delta)(\epsilon_1-\Delta) & (1-\epsilon_1+\Delta)^2 \end{pmatrix}. \quad (10)$$

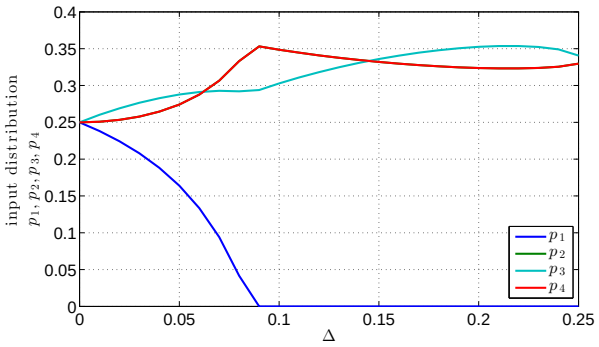

Fig. 5. Optimal input distribution for two coupled BSC with $\epsilon_1 = 0.25$ and $0 \leq \Delta \leq 0.25$, respectively. Note that p_2 and p_3 are equal and therefore their curves are on top of each other.

In Figure II-E, the optimal input distribution for the two coupled BSC with $\epsilon_1 = 0.25$ and varying δ is shown. For Δ larger than 0.09, the first input $(0,0)$ is switched off $p_1 = 0$ and only three inputs are active.

F. Extension to K coupled BSCs

For the extension to K parallel coupled BSCs, we model the fidelity of channel k as a function of the output of all other channels $l \neq k$ as follows

$$\epsilon_k(\boldsymbol{y}_{-k}) = \epsilon + \sum_{l \neq k} \Delta(1-2y_l) \quad (11)$$

where $\Delta > 0$ is chosen such that all ϵ_k are non-negative and smaller than or equal to $1/2$. This gives the set of possible fidelities

$$\mathcal{E} = \{\epsilon_1(\boldsymbol{y}_{-1}) : \boldsymbol{y}_{-1} \in \{0,1\}^{K-1}\}. \quad (12)$$

We can enumerate the K unique elements in $\mathcal{E} = \{\epsilon_1, ..., \epsilon_K\}$.

The model in (11) assumes that the quality of transmission and reception of channel k is improved when more "ones" of the other channels are received properly. For example with $K = 3$, the error probability of channel one is best, if on the other two channels a 1 is received correctly. Based on (11) for $K = 3$, we obtain the following three error probabilities (here for channel $k = 1$)

$$\epsilon_1(y_2, y_3) = \epsilon + \begin{cases} +2\Delta & y_2 = 0, y_1 = 0 \\ 0 & y_2 = 1, y_1 = 0 \vee y_2 = 0, y_1 = 1 \\ -2\Delta & y_2 = 1, y_1 = 1 \end{cases}.$$

This generalizes the case with two BSCs from the last subsections.

The resulting overall channel is described by the conditional probability distribution

$$p(\boldsymbol{y}|\boldsymbol{x}) = p(y_1, ..., y_K|x_1, ..., x_K) = \prod_{k=1}^{K}(1 - |y_k - x_k| + (-1)^{y_k - x_k}(\epsilon + \sum_{l \neq k} \Delta(1-2y_l))), \quad (13)$$

together with its input alphabet $\mathcal{X} = \{1, ..., 2^K\}$ and output alphabet $\mathcal{Y} = \{1, ..., 2^K\}$. Using the conditional probability in (13), the channel matrix for any K can be easily constructed. In order to obtain the channel matrix in (10), choose $K = 2$ and evaluate (13).

A generalization of Proposition 3 derives the sum rate for general K achieved with uniform input distribution.

Proposition 4. *The sum rate achievable over the coupled K BSCs (or the equivalent DMC $(\mathcal{X}, \mathcal{Y}, p(\boldsymbol{y}|\boldsymbol{x})$ in (13)) with uniform input distribution is given by*

$$R(K, \epsilon, \Delta) = \frac{2K}{2^K} \sum_{l=1}^{K} \binom{K-1}{l-1} \cdot C(\epsilon_l), \quad (14)$$

with ϵ_l in (11).

For $K = 2$, the result from Proposition 3 follows. It is interesting to note that for $K > 2$ the achievable sum rate does not grow as expected from Proposition 1. The first decreasing term $\frac{2K}{2^K}$ and the sum of binomial coefficients $\sum_{l=1}^{K} \binom{K-1}{l-1}$ converges for $K \to \infty$ to a constant value. This gives the BSCs with medium fidelity much higher weight and the overall performance converges to the average BSC.

Here, $\boldsymbol{y}_{-k}$ is the vector which consists of all components of the vector $\boldsymbol{y}$ except component k.

K	2	3	4	5
$R(K)$ with coupling	0.1984	0.2084	0.2188	0.2298
$R(K) - R_0$	0.01	0.02	0.03	0.04

TABLE II
COMPARISON OF DIFFERENT NUMBER OF COUPLED BSCS: $K = \{2, 3, 4, 5\}$ WITH $\epsilon = 0.25$ AND $\Delta = 0.05$.

For interpretation, this means that under the model assumption of (11), nature could improve the gain by coupling using more than two channels. Coupling more than two BSCs and using a joint coding with uniform input distribution does achieve higher coupling gains. However, there are overhead costs associated with joint coding which might lower the achieved gains.

In order to provide numerical evidence for the statement above, we list the sum rate for fixed initial fidelity $\epsilon = 0.25$ and fixed coupling $\Delta = 0.05$ for $K = 2, 3, 4, 5$ parallel links in Table II-F. Note that we have normalized the sum rate for a fair comparison. The rate achievable without coupling is $R_0 = 1 - \mathbb{H}_b(0.25) = 0.1887$ bits/s.

III. APPLICATIONS IN MOBILE COMMUNICATIONS

In wireless systems, it is usual to have different virtual channels with different properties for signalling, controlling and for data transmission. For example in the LTE standard there are DL-SCH: the DownLink Shared CHannel for downlink user data or radio resource control messages, PBCH: the Physical Broadcast CHannel which contains the Master Information Block, and others [23].

In addition to these rather standard control channels, in heterogeneous dense wireless networks, so called *activation channels* are used to initiate wake-up mechanisms for femtocell or cellular base stations [24]. In general, we can imagine a number of context information which are used on different time scales to flexible switch between different transceiver and radio access techniques. As one simple example, we consider the case of activation of a Gaussian channel by a BSC.

A. Two Coupled Channels: DMC and Gaussian

In this section, we study the following two-channel setup: one BSC is replaced by an additive white Gaussian channel and its noise variance is determined by the output of the BSC. Denote the input of the BSC as x_0 and its output as y_0. The bitflip probability is ϵ. Denote the input of the Gaussian channel as x and its output as y. The input variance of x is restricted to $\mathbb{E}[|x|^2] \leq P = 1$. The noise variance σ_n^2 depends on the BSC output y_0 and is given by

$$\sigma_n^2 = \begin{cases} \sigma_0^2 & y_0 = 0 \\ \sigma_1^2 & y_0 = 1 \end{cases}, \tag{15}$$

with instantaneous noise constraints

$$\underline{N} \leq \sigma_0^2 \leq \sigma_1^2 \leq \overline{N}, \tag{16}$$

and average noise variance

$$N_0 = \mathbb{E}_{y_0}[\sigma_n^2] = \frac{\underline{N} + \overline{N}}{2}. \tag{17}$$

Fix $0 \leq \epsilon \leq 1$ and $N_0, \underline{N}, \overline{N}$. The average rate maximization problem is formulated as

$$\begin{aligned} \max_{\pi, \sigma_0^2, \sigma_1^2} \pi \log\left(1 + \frac{1}{\sigma_0^2}\right) + (1-\pi)\log\left(1 + \frac{1}{\sigma_1^2}\right) \\ \text{s.t.} \quad \pi\sigma_0^2 + (1-\pi)\sigma_1^2 = N_0 \\ \text{and} \quad \epsilon \leq \pi_0 \leq 1 - \epsilon \\ \text{and} \quad \underline{N} \leq \sigma_0^2 \leq \sigma_1^2 \leq \overline{N}. \end{aligned} \tag{18}$$

The probability of output $y_0 = 1$ π_0 is related to the input distribution of the BSC via $\pi_0(p_0) = \epsilon + p_0(1 - 2\epsilon)$. from $0 \leq p_0 \leq 1$ follows $\epsilon \leq \pi_0 \leq 1 - \epsilon$. For solving (18), we note that the function $\log(1+1/x)$ is convex in x. Furthermore, the constraint set in (18) is convex and the objective linear in π and convex in σ_0^2, σ_1^2. Therefore, the optimal noise variances are $\sigma_0^2 = \underline{N}$ and $\sigma_1^2 = \overline{(N)}$. The optimal input distribution for the BSC is $p_0 = 1/2$ and hence $\pi_0 = 1/2$, too.

The resulting optimal average rate is given by

$$\bar{R} = \frac{1}{2}\log\left(1 + \frac{1}{\underline{N}}\right) + \frac{1}{2}\log\left(1 + \frac{1}{\overline{N}}\right) \tag{19}$$

compared to the achievable average rate with constant noise variance

$$\underline{R} = \log\left(1 + \frac{1}{N_0}\right). \tag{20}$$

The difference between $\bar{R}$ and $\underline{R}$ is the same as the gap applying Jensen's inequality.

This example illustrates that similar to the coupled BSCs in the previous section, a BSC with uniform input distribution can also be used to activate a Gaussian channel with different levels of noise variances. The large the noise variance difference is, the larger the coupling gain. In the ideal case, it could be infinity for $\underline{N} = 0$.

IV. CONCLUSIONS

In this work, we developed a new model for parallel coupled channels based on observations from signal pathways between interacting cells. The lateral inhibition mechanism results in a channel model, where the output of one channel influences the error performance of the other. In the ideal case, if we could design arbitrary coupling, the maximum achievable sum capacity is characterized based on Majorization theory (Proposition 1). In order to realize the capacity gain over two coupled channels, a joint coding over these two channels is necessary. We describe the optimal input distribution (Proposition 2) and compute the achievable sum capacity for uniform input distribution (Proposition 3). Finally, the extension to K parallel coupled channels is computed (Proposition 4). As one interesting application in wireless communications, we study the activation of a Gaussian channel via a parallel BSC.

There are many open challenges for future work. They include a coupled model for continuous (e.g. Gaussian) channels and the optimization of the corresponding input distribution. The scenario we considered is static. Therefore, another interesting direction is to model the dynamic behaviour of the

coupled channels and characterize the steady state. This model includes feedback because of the interactions between two cells.

In biological terms, this work is a step towards revealing the design principle of embryonic development. In particular, it is interesting in understanding of how multicellular systems process local information leading to emergent systems of higher organization, such as tissues, organs, etc.

ACKNOWLEDGEMENTS

Eduard Jorswieck thanks Martin Mittelbach and Christian Scheunert from the Communications Theory group, School of Engineering Sciences at TU Dresden, for interesting discussions about the information theoretic model and the achievable sum rates. Haralambos Hatzikirou and Andreas Reppas thank George Lolas for all the fruitful discussions regarding the biological motivation of the model.

APPENDIX A
PROOF OF PROPOSITION 1

Proof. Note that the binary entropy function $\mathbb{H}_b(x)$ (as the entropy itself) is a concave function of its argument. Therefore, the function $-\mathbb{H}_b(x)$ is convex in x. We need the following auxiliary result.

Lemma 1 (3.C.1 in [25])**.** *If $g : \mathbb{R} \to \mathbb{R}$ is convex and twice differentiable, then*

$$\phi(\boldsymbol{x}) = \sum_{k=1}^{n} g(x_k)$$

is Schur-convex.

From this follows that the sum capacity $C(\boldsymbol{\epsilon})$ is a Schur-convex function. Therefore, the maximum is achieved for the vector $\boldsymbol{\epsilon} = [1, 0, ..., 0]$ and the minimum is achieved for the uniform vector $\boldsymbol{\epsilon} = [1/K, 1/K, ..., 1/K]$. This explains the lower bound in (5).

To prove the upper bound, we consider two cases. If $c \leq 1/2$ then Schur-convexity directly implies the upper bound. If $c > 1/2$, then we define $L = \lfloor 2c \rfloor$ and argue that the vector $\boldsymbol{\epsilon}_1 = \{1/2, ..., 1/2, c_L/2, 0, ..., 0$ majorizes all other vectors $\boldsymbol{\epsilon} \in \mathcal{E}^{(c)}$.

□

APPENDIX B
PROOF OF PROPOSITION 2

Proof. For computation of the channel capacity, the following characterization of the optimal input distribution from [26, Proposition 5], which is based on the KKT optimality conditions, is useful. Here $\boldsymbol{w}_k$ is the k-th row of the channel matrix $\mathcal{W}$.

Proposition 5. *Let the channel matrix $\mathcal{W}$ satisfy the condition*

$$\max_{1 \leq i,j \leq 4} \sum_{k=1}^{4} |w_{ik} - w_{jk}| > 0 \tag{21}$$

then the capacity and the solution to (8) is attained by $\boldsymbol{p}^* = [p_1^*, p_2^*, p_3^*, p_4^*]$ if and only if

$$D(\boldsymbol{w}_i || \boldsymbol{p}^* \mathcal{W}) = \zeta, \tag{22}$$

for some $\zeta > 0$ and all i with $p_i^* > 0$. Moreover the capacity is equal to ζ.

The condition in (22) implies that for all active inputs, it holds $D(\boldsymbol{w}_1||\boldsymbol{p}^*\mathcal{W}) = D(\boldsymbol{w}_2||\boldsymbol{p}^*\mathcal{W}) = D(\boldsymbol{w}_3||\boldsymbol{p}^*\mathcal{W}) = D(\boldsymbol{w}_4||\boldsymbol{p}^*\mathcal{W})$. In particular, for the second and third row $D(\boldsymbol{w}_2||\boldsymbol{p}^*\mathcal{W}) = D(\boldsymbol{w}_3||\boldsymbol{p}^*\mathcal{W})$ this implies

$$\begin{aligned} D(\boldsymbol{w}_2||\boldsymbol{p}^*\mathcal{W}) &= \sum_{i=1}^{4} w_{2i} \log \frac{w_{2i}}{[\boldsymbol{p}^*\mathcal{W}]_i} \\ &= \sum_{i=1}^{4} w_{3i} \log \frac{w_{3i}}{[\boldsymbol{p}^*\mathcal{W}]_i} = D(\boldsymbol{w}_3||\boldsymbol{p}^*\mathcal{W}). \end{aligned} \tag{23}$$

From (23) with $w_{21} = w_{31}$ and $w_{24} = w_{34}$ it follows

$$\begin{aligned} &w_{22} \log \frac{w_{22}}{[\boldsymbol{p}^*\mathcal{W}]_2} + w_{23} \log \frac{w_{23}}{[\boldsymbol{p}^*\mathcal{W}]_3} \\ &= w_{32} \log \frac{w_{32}}{[\boldsymbol{p}^*\mathcal{W}]_2} + w_{33} \log \frac{w_{33}}{[\boldsymbol{p}^*\mathcal{W}]_3}. \end{aligned} \tag{24}$$

With $w_{22} = w_{33} = x$ and $w_{23} = w_{32} = y$ and $\phi_2 = [\boldsymbol{p}^*\mathcal{W}]_2$ and $\phi_3 = [\boldsymbol{p}^*\mathcal{W}]_3$ it follows from (24)

$$\begin{aligned} &x \log x - x \log \phi_2 + y \log - y \log \phi_3 \\ &= y \log y - y \log \phi_2 + x \log x - x \log \phi_3 \end{aligned} \tag{25}$$

and finally for arbitrary $x \neq y$ it follows $\phi_2 = \phi_3$ and thereby $p_2 = p_3$.

For the second part, we parametrize the input distribution $\boldsymbol{p}(t)$ by $0 \leq t \leq 1$ as

$$\boldsymbol{p}(t) = [t, x, x, (1 - 2x - t)],$$

with $0 \leq x \leq 1/4$. The mutual information is a concave function of $\boldsymbol{p}$ [21, Lemma 3.5]. Therefore, it suffices to show that

$$\left. \frac{\partial C(\epsilon, \delta, \boldsymbol{p}(t))}{\partial t} \right|_{t=0} \leq 0 \tag{26}$$

for sufficient small $\delta > 0$ to allocate zero probability to p_1. This is equivalent to

$$\begin{aligned} &D(\boldsymbol{w}_1||\boldsymbol{p}\mathcal{W}) \leq D(\boldsymbol{w}_4||\boldsymbol{p}\mathcal{W}) \quad \Longleftrightarrow \\ &D(\boldsymbol{w}_1||x(\boldsymbol{w}_2 + \boldsymbol{w}_3) + (1 - 2x)\boldsymbol{w}_4) \leq \\ &D(\boldsymbol{w}_4||x(\boldsymbol{w}_2 + \boldsymbol{w}_3) + (1 - 2x)\boldsymbol{w}_4). \end{aligned} \tag{27}$$

with $\boldsymbol{w}_1 = [(1-\epsilon)^2, (1-\epsilon)\epsilon, (1-\epsilon)\epsilon, \epsilon^2]$, $\boldsymbol{w}_2 = [(1-\delta)\epsilon, (1-\delta)(1-\epsilon), \epsilon\delta, \delta(1-\epsilon)]$, $\boldsymbol{w}_3 = [(1-\delta)\epsilon, \epsilon\delta, (1-\epsilon)(1-\delta), \delta(1-\epsilon)]$ and $\boldsymbol{w}_4 = [\delta^2, (1-\delta)\delta, (1-\delta)\delta, (1-\delta)^2]$. For sufficiently small δ, we verify that (27) holds. □

APPENDIX C
PROOF OF PROPOSITION 3

Proof. The mutual information $C(\epsilon, \delta, \boldsymbol{p})$ can be written as

$$C(\epsilon, \delta, \boldsymbol{p}) = H(\boldsymbol{p}\mathcal{W}) - \sum_{k=1}^{4} p_k H(\boldsymbol{w}_k). \tag{28}$$

A simple inspection and taking the column sum of the probabilities in Table I reveals for $\boldsymbol{p} = \frac{1}{4}\mathbf{1}$ that

$$H(1/4 \cdot \mathbf{1}\mathcal{W}) = H([1/4, 1/4, 1/4, 1/4]) = 2. \tag{29}$$

The second term in (28) can be computed as follows

$$\frac{1}{4}\sum_{k=1}^{4} H(\boldsymbol{w}_k) = \frac{1}{4}\sum_{k=1}^{4}\sum_{l=1}^{4} w_{kl} \log w_{kl}, \tag{30}$$

which is the complete sum over the entropies of all 16 entries of the channel matrix $\mathcal{W}$. When we flip the sum over k and l, we recognize the binary entropies $\mathbb{H}_b(\epsilon)$ and $\mathbb{H}_b(\delta)$, i.e.,

$$\begin{aligned} &\frac{1}{4}\sum_{l=1}^{4}\sum_{k=1}^{4} w_{kl} \log w_{kl} \\ &= \frac{1}{4}[2\mathbb{H}_b(\epsilon) + 2\mathbb{H}_b(\epsilon) + 2\mathbb{H}_b(\delta) + 2\mathbb{H}_b(\delta)] \\ &= \mathbb{H}_b(\epsilon) + \mathbb{H}_b(\delta), \end{aligned} \tag{31}$$

where the first equality follows from the additivity property of the entropy function, i.e., $H([ab, (1-a)b, a(1-b), (1-a)(1-b)]) = H(a, (1-a)) + H(b, (1-b))$. Together with the result in (29), we finally obtain

$$C(\epsilon, \delta, \boldsymbol{p}) = 2 - \mathbb{H}_b(\epsilon) + \mathbb{H}_b(\delta) = C(\epsilon) + C(\delta). \tag{32}$$

□

REFERENCES

[1] B. Alberts, A. Johnson, J. Lewis, and M. Raff, "Development of multicellular organisms," in *Molecular Biology of the Cell*. New York, NY: Garland Science, 2002, ch. 22. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK21046/

[2] P. Hayward, T. Kalmar, and A. M. Arias, "Wnt/Notch signalling and information processing during development." *Development (Cambridge, England)*, vol. 135, no. 3, pp. 411–24, Feb. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18192283

[3] Y. Morishita and Y. Iwasa, "Coding design of positional information for robust morphogenesis," *Biophysical Journal*, vol. 101, no. 10, pp. 2324–2335, 2011.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–656, 1948.

[5] I. S. Mian and C. Rose, "Communication theory and multicellular biology." *Integrative biology : quantitative biosciences from nano to macro*, vol. 3, no. 4, pp. 350–67, Apr. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21424025

[6] C. Waltermann and E. Klipp, "Information theory based approaches to cellular signaling." *Biochimica et biophysica acta*, vol. 1810, no. 10, pp. 924–32, Oct. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21798319

[7] A. Rhee, R. Cheong, and A. Levchenko, "The application of information theory to biochemical signaling systems." *Physical biology*, vol. 9, no. 4, p. 045011, Aug. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22872091

[8] K. Hironaka and Y. Morishita, "Encoding and decoding of positional information in morphogen-dependent patterning," *Current opinion in genetics and development*, vol. 22, pp. 553–561, 2012.

[9] E. Klipp and W. Liebermeister, *Mathematical modeling of intracellular signaling pathways*. BMC Neuroscience, 2006.

[10] J. O. Dubuis, E. F. Wieschaus, T. Gregor, and W. Bialek, "Positional information, in bits," *PNAS*, vol. 41, pp. 16 301–16 308, 2013. [Online]. Available: http://www.pnas.org/content/110/41/16301.full

[11] D. Sprinzak, A. Lakhanpal, L. LeBon, M. Fontes, G. Anderson, J. Garcia-Ojalvo, and M. Elowitz, "Cis-interactions between notch and delta generate mutually exclusive signalling states," *Nature*, vol. 465, pp. 86–90, 2010.

[12] H. Pascal and P. Simpson, "The choice of cell fate in the epidermis of drosophila," *Cell*, vol. 64, no. 6, pp. 1083–1092, 1991.

[13] J. Collier, N. Monk, P. Maini, and J. Lewis, "Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling," *Journal of Theoretical Biology*, vol. 183, no. 4, pp. 429–446, 1996.

[14] S. Artavanis-Tsakonas, K. Matsuno, and M. Fortini, "Notch signaling," *Science*, vol. 268, pp. 225–268, 1995.

[15] J. de Celis and S. Bray, "The abruptex domain of notch regulates negative interactions between notch, its ligands and fringe," *Development*, vol. 127, pp. 1291–1302, 2000.

[16] O. Barad, D. Rosin, E. Hornstein, and N. Barkai, "Error minimization in lateral inhibition circuits," *Science Signaling*, vol. 3, no. 129, p. ra51, 2010.

[17] Z. Hensel, H. Feng, B. Han, C. Hatem, J. Wang, and J. Xiao, "Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis," *Nature Structural & Molecular Biology*, vol. 19, pp. 797–802, 2012.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley & Sons, 1991.

[19] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[20] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, Sept. 1960.

[21] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[22] G. Kumar and A. Manolakos, "No input symbol should occur more frequently than 1-1/e," Jan. 2012, http://arxiv.org/abs/1201.6425.

[23] R. Love, R. Kuchibhotla, A. Ghosh, R. Ratasuk, B. Classon, and Y. Blankenship, "Downlink control channel design for 3gpp lte," in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, March 2008, pp. 813–818.

[24] L. Saker, S.-E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 664–672, April 2012.

[25] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications*. Springer Series in Statistics, 2011.

[26] R. Mathar and A. Schmeink, "Saddle-point properties and Nash equilibria for channel games," *EURASIP Journal on Advances in Signal Processing*, 2009.