Deterministic Annealing Design and Analysis of the Noisy MIMO Interference Channel

Francesco Negro*, Irfan Ghauri†, Dirk T.M. Slock*
†Intel Mobile Communications, GAIA, 2600 Route des Crêtes, 06560 Sophia Antipolis Cedex, France
Email: irfan.ghauri@intel.com
*Mobile Communications Department, EURECOM, 2229 Route des Crêtes, BP 193, 06904 Sophia Antipolis Cedex, France
Email: francesco.negro@eurecom.fr, dirk.slock@eurecom.fr

***Abstract*—We consider the Noisy MIMO Interference Channel (IFC) with linear transmitters and receivers and full CSI. The maximization of the Weighted Sum Rate (WSR) or transceiver design for Interference Alignment (IA) lead to cost functions with many local optima. Deterministic annealing is an approach that allows to track the variation of the known solution of one version of the problem into the unknown solution of the desired version by a controlled variation of a parameter called temperature. When the temperature parameter is chosen as inverse SNR (or noise power), the transceiver design for maximum WSR is known at low SNR and can be tracked to any desired SNR, yielding an elegant technique to find the global optimum. The solution includes filter design for the progressive switching on of streams as the SNR increases. For IA on the other hand, IA feasibility is unchanged when the MIMO crosslink channel matrices have a reduced rank equal to the maximum of the number of streams passing through them in forward and dual IFC (this would correspond to LOS channels in the case of single streams). The rank reduction simplifies IA design and feasibility analysis, and allows in particular a counting of the number of IA solutions. By choosing now the temperature parameter to be a scale factor for the remaining channel singular values, the solution for reduced rank channels can be evolved into that for arbitrary channels.**

***Index Terms*—MIMO, MMSE, weighted sum rate, interference channel, linear transmitter, linear receiver, interference alignment, deterministic annealing**

I. Introduction

To achieve higher system capacity in modern cellular communication standards a frequency reuse factor of 1 is used. This increment in system performances determines, on the other hand, a drastical reduction of the capacity of the cell-edge users due to the fact that this aggressive frequency reuse factor increases the inter-cell interference.

To handle this problem current communication systems include different interference management solutions. Even if interference coming from out-of-cell transmission can be reduced using careful planning or introducing little cooperation among neighboring cells, such as smart user scheduling or soft handover, these techniques are sometimes not enough to guarantee high performance to cell-edge user. For that major standardization bodies are now including explicit interference coordination strategies in next generation cellular communication standards. These techniques are based on more interference-aware base station cooperation. A systematic study of the performance of cellular communication systems where each cell communicates multiple streams to its users while enduring/causing interference from/to neighboring cells due to transmission over a common shared resource comes under the purview of MIMO interference channels (MIMO IFC). A K-user MIMO-IFC models a network of K transmit-receive pairs where each transmitter communicates multiple data streams to its respective receiver. In doing so, it generates interference at all other receivers. While the interference channel has been the focus of intense research over the past few decades, its capacity in general remains an open problem and is not well understood even for simple cases. In [1] they show that even for the $2-$users system, the most studied case, to achieve the system capacity within one bit very complicated transmission schemes are required.

Recently, it was shown that the concept of interference alignment (IA) [2], allows each receiver to suppress more interfering streams than it could otherwise cancel in interference channels. This can be done using more simple linear transmitter and receiver filter. This makes IA a very attractive solution in practical systems. The focus of this paper is on the K-user frequency-flat MIMO IFC. In a frequency-flat MIMO IFC, the total number of streams contributing to the input signal at each receiver are, in general, greater than the number of antennas available at the transmitter or at the receiver. This would lead one to believe that, at least in the high-SNR regime, the network (comprising of K user pairs) performance can be maximized (i.e, the sum-rate can be maximized) using IA since aligning the streams at the transmitter will now allow the maximization of the capacity pre-log factor in a K-user IFC. The problem of determining whether an IA solution exists or not for a given antennas and stream distribution among the users for a K-user MIMO IFC it has been studied in [3] and [4]. In the former an extensive study of IA feasibility solution for the single stream case has been proposed. In the latter the authors propose a systematic method, and less computational expensive, to check feasibility regardless of the number of transmitted stream per user.

A distributed algorithm that exploits the reciprocity of the MIMO IFC to obtain the transmit and receiver filters in a K-user MIMO IFC was proposed in [5] (a similar algorithm has been proposed in [6]). It is was shown there that IA is a suboptimal strategy at finite SNRs. In the same paper, the authors propose a signal-to-interference-plus-noise-ratio (SINR) maximizing algorithm which outperforms the IA in finite SNRs and converges to the IA solution in the high SNR regime. However, this approach can be shown to be suboptimal for multiple stream transmission since it allocates equal power to all streams. In [7] the authors present an iterative algorithm that finds an IA solution that maximize the average sum-rate. At each step an IA solution is found using a technique proposed in [5] and then they move the solution along the direction of the gradient of the sum-rate w.r.t. the beamformers in the Grassmann manifold. Even though this algorithm performs better than traditional IA solutions in the High SNR regime it is highly sub-optimal, in terms of sum-rate, in medium SNR ranges. Thus an optimal solution for MIMO IFC at finite SNR remains an open problem.

Some early work on the MIMO IFC was reported in [8] by Ye and Blum for the asymptotic cases when the interference to noise ratio (INR) is extremely small or extremely large. It was shown there that a "greedy approach" where each transmitter attempts to maximize its individual rate regardless of its effect on other un-intended receivers is provably suboptimal.There have been some attempts to port the solution concepts of the MIMO BC and MIMO MAC to the MIMO IFC. For instance, the problem of joint transmitter and receiver design to minimize the sum-MSE of a multiuser MIMO uplink was considered in [9] where iterative algorithms that jointly optimize precoders and receivers were proposed. Subsequently [10] applied this algorithm to the MIMO IFC where each user transmits a single stream. In [11] the authors proposed an algorithm for finding the beamformer in the single stream K-user MIMO IFC that attempts to maximize the weighted sum rate (WSR). The beamforming vectors can be interpreted as a balance between an egoistic approach, where the transmitter tries to maximize its own rate and an altruistic approach where each beamformer put its effort to minimize the interference that it causes to the non intended receivers. The problem of a more general multi-stream MIMO IFC has been addressed in [12] where the objective of the algorithm proposed by the authors is to design a set of BF matrices in order to maximize the WSR. The algorithm proposed is based on a previous work on broadcast channel [13] but extended to a MIMO IFC and it has been further refined in [14].

The main problem with the maximization of the WSR is the highly non convexity of the cost function. This implies that even if it is possible to prove convergence of the proposed algorithms to a local optimal point convergence to global optima can not been shown. In addition convergence to local optimal solution is not a rare event if the initialization point of the algorithm is not carefully chosen. To avoid this situation several heuristic approach can be used. In [14] in order to avoid to converge to a local optima *Deterministic Annealing* (DA) has been proposed. DA is an heuristic approach based on *Simulated Annealing* (SA) where the basic principle is that the optimum of a problem in the next value of temperature (in our case SNR) is in the region of attraction of the solution of the problem in the previous temperature (more details will be provided further in the paper). In another independent work [15] the same principle has been explored but only for single stream MIMO systems.

The objective of this paper is to further study the concept of DA applied to WSR maximization in a general K-User MIMO interference channel where a general number of stream distribution is assumed. In addition the IA feasibility problem has been studied, in particular we propose a new approach to address the problem. It is based on the principle that IA feasibility is unchanged when the MIMO crosslink channel matrices have a reduced rank equal to the maximum of the number of streams passing through them in forward and dual IFC, then increasing constantly the rank of the channels the number of IA solutions will not decrease.

II. Signal Model

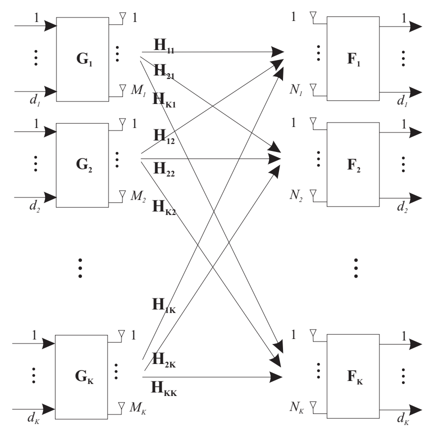

Fig. 1: MIMO Interference Channel

Fig. 1 depicts a K-user MIMO interference channel with K transmitter-receiver pairs. The k-th transmitter and its corresponding receiver are equipped with M_k and N_k antennas respectively. The k-th transmitter generates interference at all $l \neq k$ receivers. Assuming the communication channel to be frequency-flat, the $\mathbb{C}^{N_k \times 1}$ received signal $\mathbf{y}_k$ at the k-th receiver, can be represented as

$$\mathbf{y}_k = \mathbf{H}_{kk}\mathbf{x}_k + \sum_{\substack{l=1 \\ l \neq k}}^{K} \mathbf{H}_{kl}\mathbf{x}_l + \mathbf{n}_k \tag{1}$$

where $\mathbf{H}_{kl} \in \mathbb{C}^{N_k \times M_l}$ represents the channel matrix between the l-th transmitter and k-th receiver, $\mathbf{x}_k$ is the $\mathbb{C}^{M_k \times 1}$ transmit signal vector of the k-th transmitter and the $\mathbb{C}^{N_k \times 1}$ vector $\mathbf{n}_k$ represents (temporally white) AWGN with zero mean and covariance matrix $\mathbf{R}_{n_k n_k}$. Each entry of the channel matrix is a complex random variable drawn from a continuous distribution. It is assumed that each transmitter has complete knowledge of all channel matrices corresponding to its direct link and all the other cross-links in addition to the transmitter power constraints and the receiver noise covariances.

We denote by $\mathbf{G}_k$, the $\mathbb{C}^{M_k \times d_k}$ precoding matrix of the k-th transmitter. Thus $\mathbf{x}_k = \mathbf{G}_k\mathbf{s}_k$, where $\mathbf{s}_k$ is a $d_k \times 1$ vector representing the d_k independent symbol streams for the k-th user pair. We assume $\mathbf{s}_k$ to have a spatio-temporally white Gaussian distribution with zero mean and unit variance, $\mathbf{s}_k \sim \mathcal{N}(0, \mathbf{I}_{d_k})$. The k-th receiver applies $\mathbf{F}_k \in \mathbb{C}^{N_k \times d_k}$ to suppress interference and retrieve its d_k desired streams. The output of such a receive filter is then given by

$$\mathbf{r}_k = \mathbf{F}_k^H \mathbf{H}_{kk} \mathbf{G}_k \mathbf{s}_k + \sum_{\substack{l=1 \\ l \neq k}}^{K} \mathbf{F}_k^H \mathbf{H}_{kl} \mathbf{G}_l \mathbf{s}_l + \mathbf{F}_k^H \mathbf{n}_k$$

Note that $\mathbf{F}_k$ does not represent the whole receiver but only the reduction from a N_k-dimensional received signal $\mathbf{y}_k$ to a d_k-dimensional signal $\mathbf{r}_k$, to which further (possibly optimal) receive processing is applied.

III. Weighted sum rate maximization for the MIMO IFC

The stated objective of our investigation is the maximization of the WSR of MIMO IFC. In this paper we consider the weighted sum rate maximization problem for a K-user frequency-flat MIMO IFC and propose an iterative algorithm for linear precoder/receiver design. With full CSIT, but only knowledge of $\mathbf{s}_k$ at transmitter k, it is expected that linear processing at the transmitter should be sufficient. On the receive side however, optimal WSR approaches may involve joint detection of the signals from multiple transmitters. In this paper we propose to limit receiver complexity by restricting the modeling of the signals arriving from interfering transmitters as colored noise (which is Gaussian if we consider Gaussian codebooks at the transmitters). As a result, linear receivers are sufficient. For the MIMO IFC, one approach to linear transmit precoder design is the joint design of precoding matrices to be applied at each transmitter based on channel state information (CSI) of all users. Such a *centralized* approach [8] requires (channel) information exchange among transmitters. Nevertheless, studying such systems can provide valuable insights into the limits of perhaps more practical *distributed* algorithms [16] [17] that do not require any information transfer among transmitters.

A. Per-User WSR maximization

The WSR maximization problem can be mathematically expressed as follows.

$$\{\mathbf{G}_k^\star, \mathbf{F}_k^\star\} = \arg \min_{\{\mathbf{G}_k, \mathbf{F}_k\}} \mathcal{R} \quad \text{s.t} \ \operatorname{Tr}(\mathbf{G}_k^H \mathbf{G}_k) = P_k \ \ \forall k \tag{2}$$

where $\mathcal{R} = \sum_k -u_k R_k$
with $u_k \geq 0$ denoting the weight assigned to the k-th user's rate and P_k it's transmit power constraint. We use the notation $\{\mathbf{G}_k,\ \mathbf{F}_k\}$ to compactly represent the candidate set of transmitters $\mathbf{G}_k$ and receivers $\mathbf{F}_k$ $\forall k \in \{1, \ldots, K\}$ and the corresponding set of optimum transmitters and receivers is represented by $\{\mathbf{G}_k^\star, \mathbf{F}_k^\star\}$. Assuming Gaussian signaling, the k-th user's achievable rate is given by

$$\begin{aligned} R_k &= \log |\mathbf{E}_k|, \\ \mathbf{E}_k &= \mathbf{I}_k + \mathbf{F}_k^H \mathbf{H}_{kk} \mathbf{G}_k (\mathbf{F}_k^H \mathbf{H}_{kk} \mathbf{G}_k)^H (\mathbf{F}_k^H \mathbf{R}_{\overline{k}} \mathbf{F}_k)^{-1} \end{aligned} \tag{3}$$

where the interference plus noise covariance matrix $\mathbf{R}_{\overline{k}}$ is:

$$\mathbf{R}_{\overline{k}} = \mathbf{R}_{n_k n_k} + \sum_{l \neq k} \mathbf{H}_{kl} \mathbf{G}_l \mathbf{G}_l^H \mathbf{H}_{kl}^H.$$

We use here the standard notation $|\ .\ |$ to denote the determinant of a matrix. The MIMO IFC rate region is known to be non-convex. The presence of multiple local optima complicates the computation of optimum precoding matrices to be applied at the transmitter in order to maximize the weighted sum rate. What is known however, is that, for a given set of precoders, linear minimum mean squared error (LMMSE) receivers are optimal in terms of interference suppression. In addition we can extend this concept saying that, for a given set of linear beamforming filters applied at the transmitters, the LMMSE interference-suppressing filter applied at the receiver does not lose any information of the desired signal in the process of reducing the N_k dimensional $\mathbf{y}_k$ to a d_k dimensional vector $\mathbf{r}_k$. This is of course under the assumption that all interfering signals can be treated as Gaussian noise. In other words, the linear MMSE interference suppressor filter is information lossless and is thus optimal in terms of maximizing the WSR. Thanks to this property of the LMMSE Rx filter, we consider a (more tractable) optimization problem where MMSE processing at the receiver is implicitly assumed. The WSR maximization problem in (2) that we consider becomes:

$$\{\mathbf{G}_k^\star\} = \arg \min_{\{\mathbf{G}_k\}} \sum_{k=1}^{K} -u_k \log |\mathbf{E}_k^{-1}| \ \ \text{s.t} \ \operatorname{Tr}(\mathbf{G}_k^H \mathbf{G}_k) \leq P_k \ \forall k \tag{4}$$

where $\mathbf{E}_k$ is given by

$$\mathbf{E}_k = (\mathbf{I} + \mathbf{G}_k^H \mathbf{H}_{kk}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{kk} \mathbf{G}_k)^{-1}. \tag{5}$$

This problem in non convex and hence finding a solution is a complex task. In order to obtain the stationary points for the optimization problem (4), we solve the Lagrangian:

$$L\left(\{\mathbf{G}_k, \lambda_k\}\right) = \sum_{k=1}^{K} -u_k \log |\mathbf{E}_k^{-1}| + \lambda_k (\operatorname{Tr}\{\mathbf{G}_k^H \mathbf{G}_k\} - P_k)$$

Now setting the gradient of the Lagrangian w.r.t. the transmit filter $\mathbf{G}_k$ to zero, we have:

$$\begin{gathered} \frac{\partial J(\{\mathbf{G}_k, \lambda_k\})}{\partial \mathbf{G}_k^*} = 0 \\ \textstyle\sum_{l \neq k} u_l \mathbf{H}_{lk}^H \mathbf{R}_{\overline{l}}^{-1} \mathbf{H}_{ll} \mathbf{G}_l \mathbf{E}_l \mathbf{G}_l^H \mathbf{H}_{ll}^H \mathbf{R}_{\overline{l}}^{-1} \mathbf{H}_{lk} \mathbf{G}_k \\ -u_k \mathbf{H}_{kk}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{kk} \mathbf{G}_k \mathbf{E}_k + \lambda_k \mathbf{G}_k = 0 \end{gathered} \tag{6}$$

Our approach to the design of the WSR maximizing transmit filters for the MIMO IFC is based on introducing an augmented cost function in which two additional optimization variables appear [13]. The optimization problem that we consider now is

$$\begin{aligned} &\{\mathbf{G}_k^\star,\ \mathbf{F}_k^\star,\ \mathbf{W}_k^\star\} = \\ &\arg \max_{\{\mathbf{G}_k, \mathbf{F}_k, \mathbf{W}_k\}} \sum_k -u_k (\operatorname{Tr}(\mathbf{W}_k \boldsymbol{\mathcal{E}}_k) - \log |\mathbf{W}_k| - d_k^{max}) \\ &\qquad \text{s.t} \sum_k \operatorname{Tr}(\mathbf{G}_k \mathbf{G}_k^H) \leq P_k. \end{aligned} \tag{7}$$

where $d_k^{max} \leq \min\{N_k, M_k\}$ represents the maximum number of independent data streams that can be transmitted to user k. This cost function is concave or even quadratic in one set of variables, keeping the other two fixed. Hence we shall optimize it using alternating maximization. Assuming $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_{d_k}$, the MSE covariance matrix for general Tx and Rx filters is

$$\begin{aligned}\boldsymbol{\mathcal{E}}_k = & \mathbb{E}[(\mathbf{s}-\mathbf{F}_k^H\mathbf{y}_k)(\mathbf{s}-\mathbf{F}_k^H\mathbf{y}_k)^H]\} \\ = & \mathbf{I}-\mathbf{G}_k^H\mathbf{H}_{kk}^H\mathbf{F}_k-\mathbf{F}_k^H\mathbf{H}_{kk}\mathbf{G}_k \\ & +\mathbf{F}_k^H\mathbf{H}_{kk}\mathbf{G}_k\mathbf{G}_k^H\mathbf{H}_{kk}^H\mathbf{F}_k \\ & +\textstyle\sum_{l\neq k}\mathbf{F}_k^H\mathbf{H}_{kl}\mathbf{G}_l\mathbf{G}_l^H\mathbf{H}_{kl}^H\mathbf{F}_k+\mathbf{F}_k^H\mathbf{R}_{n_k n_k}\mathbf{F}_k\end{aligned} \tag{8}$$

The corresponding Lagrangian can be written as:

$$\begin{aligned}J(\{\mathbf{G}_k,\mathbf{F}_k,\mathbf{W}_k,\lambda_k\}) = & -\sum_k \lambda_k(\mathrm{Tr}\{\mathbf{G}_k^H\mathbf{G}_k\}-P_k) \\ & -\sum_k u_k(\mathrm{Tr}(\mathbf{W}_k\boldsymbol{\mathcal{E}}_k)-\log|\mathbf{W}_k|-d_k^{max})\end{aligned} \tag{9}$$

This new cost function will be optimized w.r.t. one set of variables, keeping the other two fixed. The first step in our optimization process is the calculation of the optimal Rx filters assuming fixed the matrices $\mathbf{G}_k$ and $\mathbf{W}_k$. It can easily be seen that the optimal Rx filter is an MMSE filter:

$$\mathbf{F}_k^{LMMSE} = (\mathbf{R}_{\overline{k}}+\mathbf{H}_{kk}\mathbf{G}_k\mathbf{G}_k^H\mathbf{H}_{kk}^H)^{-1}\mathbf{H}_{kk}\mathbf{G}_k \tag{10}$$

The following step in the optimization procedure is the determination of the optimal expression for the matrix $\mathbf{W}_k$ while keeping the other two variable sets fixed.
What we get is:

$$\mathbf{W}_k = \mathbf{E}_k^{-1} \tag{11}$$

The final step is the maximization of the given cost function w.r.t. the BF matrix. To accomplish this task we derive the Lagrangian w.r.t. the matrix $\mathbf{G}_k$ and equate it to zero:

$$\begin{aligned}&\frac{\partial \mathrm{J}(\{\mathbf{G}_k,\lambda_k\})}{\partial \mathbf{G}_k^*} = \\ & u_k\mathbf{H}_{kk}^H\mathbf{F}_k\mathbf{W}_k-\lambda_k\mathbf{G}_k-\textstyle\sum_{l=1}^K u_l\mathbf{H}_{lk}^H\mathbf{F}_l\mathbf{W}_l\mathbf{F}_l^H\mathbf{H}_{lk}\mathbf{G}_k = 0.\end{aligned} \tag{12}$$

This leads to the following expression for the optimizing BF:

$$\mathbf{G}_k = \left(\sum_{l=1}^K u_l\mathbf{H}_{lk}^H\mathbf{F}_l\mathbf{W}_l\mathbf{F}_l^H\mathbf{H}_{lk}+\lambda_k\mathbf{I}\right)^{-1}\mathbf{H}_{kk}^H\mathbf{F}_k\mathbf{W}_k u_k \tag{13}$$

The only variable that still needs to be optimized is the Lagrange multiplier λ_k. First check if $\mathrm{Tr}(\mathbf{G}_k^H\mathbf{G}_k) \leq P_k$ for $\lambda_k = 0$. If yes, than $\lambda_k = 0$. If not, the Tx power equality constraint is active. In this case to determine the optimal value of the lagrange multiplier λ_k we consider equation (12) that for the optimality of the BF matrix it is satisfied. In addition pre-multiplying the derivative of the cost function w.r.t. the BF matrix by $\mathbf{G}_k^H$ and taking the trace the product is still equal to zero:

$$\mathrm{Tr}\left\{\mathbf{G}_k^H\frac{\partial \mathrm{J}(\{\mathbf{G}_k,\lambda_k\})}{\partial \mathbf{G}_k^*}\right\} = 0$$

$$\begin{aligned}&\mathrm{Tr}\left\{u_k\mathbf{G}_k^H\mathbf{H}_{kk}^H\mathbf{F}_k\mathbf{W}_k\right\}-\lambda_k\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\} \\ &-\sum_{l=1}^K u_l\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{H}_{lk}^H\mathbf{F}_l\mathbf{W}_l\mathbf{F}_l^H\mathbf{H}_{lk}\mathbf{G}_k\right\} = 0.\end{aligned} \tag{14}$$

In equation (14) we impose the power constraint to be satisfied with equality hence the contribution $\lambda_k Tr\left\{\mathbf{G}_k^H\mathbf{G}_k\right\} = \lambda_k P_k$. Using the definition of the MMSE Rx filter we get the following expression for the Lagrange multiplier:

$$\begin{aligned}\lambda_k = & -\tfrac{1}{P_k}\left(\sum_{l\neq k} u_l\mathrm{Tr}\{\mathbf{W}_l\mathbf{F}_l^H\mathbf{H}_{lk}\mathbf{G}_k(\mathbf{F}_l^H\mathbf{H}_{lk}\mathbf{G}_k)^H\}\right) \\ & +\tfrac{1}{P_k}\left(\sum_{l\neq k} u_k\mathrm{Tr}\{\mathbf{W}_k\mathbf{F}_k^H\mathbf{H}_{kl}\mathbf{G}_l(\mathbf{F}_k^H\mathbf{H}_{kl}\mathbf{G}_l)^H\}\right) \\ & +\tfrac{u_k}{P_k}\left(\mathrm{Tr}\{\mathbf{W}_k\mathbf{F}_k^H\mathbf{R}_{n_k n_k}\mathbf{F}_k\}\right).\end{aligned} \tag{15}$$

With this value of the Lagrange multiplier the final expression for the BF becomes (16). The algorithm proposed in [13] was developed for a MIMO broadcast channel, where only an overall Tx power constraint is applied on the system and, in addition, maximizing the WSR automatically requires to transmit with full power. On the other hand in the MIMO IFC the WSR maximization may require some links to transmit with a power less than the maximum power available at that links.
At low SNR regime the maximization of the WSR leads to activate only one stream per link, allocating full power on the best singular mode of the direct channel $\mathbf{H}_{kk}$.
For SNR values sufficiently high the maximization of the sum rate converges to an IA solution. IA feasibility may imply zero streams for some links. Here we propose to determine the optimal value of $\lambda_k \geq 0$ using a linear search algorithm.

Grouping together all the optimization steps that describe our maximization procedure we have the following two-steps iterative algorithm to compute the precoders that maximize the weighted sum rate for a given MIMO IFC (c.f Table **Algorithm 1**). Introducing the augmented cost function, for

---

Algorithm 1 MWSR Algorithm for MIMO IFC

Fix an arbitrary initial set of precoding matrices $\mathbf{G}_k$, $\forall \in k = \{1, 2 \ldots K\}$
set $n = 0$
repeat
  $n = n + 1$
  Given $\mathbf{G}_k^{(n-1)}$, compute $\mathbf{F}_k^{(n)}$ and $\mathbf{W}_k^{(n)}$ from (10) and (11) respectively $\forall k$
  Given $\mathbf{F}_k^{(n)}$ and $\mathbf{W}_k^{(n)}$, compute $\mathbf{G}_k^{(n)}$ $\forall k$ using (13)
until convergence

---

the calculation of the optimal BF matrix that maximize the WSR, we are able to determine an iterative algorithm that can be easily proved to converge to a local optima that corresponds also to an extremum of the original cost function (4).
Each step of our iterative algorithm increases the cost function, which is bounded above (e.g. by cooperative WSR), and hence convergence is guaranteed. In addition the augmented cost function once we substitute $\mathbf{W}_k$ and $\mathbf{F}_k$ with their optimal values, becomes exactly the original WSR cost function (4). Finally using matrix inversion lemma[1] it is possible to rewrite

[1]If $\mathbf{P}$ and $\mathbf{R}$ are positive definite the following relation is true:

$$\mathbf{P}\mathbf{B}^T(\mathbf{B}\mathbf{P}\mathbf{B}^T+\mathbf{R})^{-1} = (\mathbf{P}^{-1}+\mathbf{B}^T\mathbf{R}^{-1}\mathbf{B})\mathbf{B}^T\mathbf{R}^{-1}$$

the expression of the MMSE (10) as

$$\mathbf{F}_k = \mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{kk}\mathbf{G}_k\mathbf{E}_k.$$

With this representation of the Rx filters it is possible to interpret some quantities in the gradient of the WSR (6) as Rx filters and hence the expression that comes out of this elaboration is the same as the gradient of the augmented cost function w.r.t. the BF matrix (12). This implies that a stationary point of the original cost function is also a stationary point of the augmented cost function.

A final remark can be made about the dimensions imposed on the beamforming matrix. In particular at high SNR we can put $d_k = M_k$ if we want the algorithm to figure out the feasible set of $\{d_k\}$, in this case all IA-feasible solutions represent local optima. Another possible choice is to use d_k that corresponds to an IA-feasible solution if we want to focus on that particular stream distribution.
At low or medium SNR regime a possible choice is $d_k = \max\{1, d_k^{IA}\}$ where the set $\{d_k^{IA}\}$ form a IA-feasible set.

B. Per-Stream WSR maximization

In the algorithm presented so far the stream of each user are correlated to each other. It is possible to show that modifying the BF in order to decorrelate the stream of each user does not reduce the overall sum rate. Using a per-stream approach leads to a solution in which the MMSE matrix is diagonal. This property will be explored further later in the paper. The cost function proposed in this paper for the per-user approach can be written in the per-stream case as:

$$\begin{aligned}\mathcal{O} = &-\sum_{k=1}^{K} u_k \sum_{n=1}^{d_k}(-\ln(w_{kn}) - 1\\ &+w_{kn}(1-\mathbf{f}_{kn}^H\mathbf{H}_{kk}\mathbf{g}_{kn})(1-\mathbf{f}_{kn}^H\mathbf{H}_{kk}\mathbf{g}_{kn})^H\\ &+w_{kn}\mathbf{f}_{kn}^H\underbrace{(\mathbf{R}_{n_kn_k} + \sum_{(im)\neq(kn)}\mathbf{H}_{ki}\mathbf{g}_{im}\mathbf{g}_{im}^H\mathbf{H}_{ki}^H)}_{\mathbf{R}_{\overline{kn}}}\mathbf{f}_{kn}).\end{aligned} \tag{17}$$

The optimization problem when we work per stream becomes:

$$\begin{array}{cc}\max\limits_{\mathbf{f}_{kn},\mathbf{g}_{kn},w_{kn}} & \mathcal{O}\\ \text{s.t.} & \sum_n^{d_k}\mathbf{g}_{kn}^H\mathbf{g}_{kn} \leq P_k \ \ \forall k\end{array} \tag{18}$$

and the corresponding lagrangian is:

$$J = \mathcal{O} + \sum_{k=1}^{K}\lambda_k\left(P_k - \sum_{n=1}^{d_k}\mathbf{g}_{kn}^H\mathbf{g}_{kn}\right) \tag{19}$$

To solve the given optimization problem we use alternating optimization. As first step we determine the Rx filter assuming all the other optimization variables to be fixed. Deriving the cost function above w.r.t. the Rx filter we obtain an MMSE receiver per stream:

$$\mathbf{f}_{kn} = (\mathbf{H}_{kk}\mathbf{g}_{kn}\mathbf{g}_{kn}^H\mathbf{H}_{kk}^H + \mathbf{R}_{\overline{kn}})^{-1}\mathbf{H}_{kk}\mathbf{g}_{kn} \tag{20}$$

Given the optimal Rx filter we derive (19) w.r.t. the scalar weight and we find:

$$w_{kn} = e_{kn}^{-1} \tag{21}$$

where $e_{kn} = (1 + \mathbf{g}_{kn}^H\mathbf{H}_{kk}^H\mathbf{R}_{\overline{kn}}^{-1}\mathbf{H}_{kk}\mathbf{g}_{kn})^{-1}$. The third step is the optimization of the beamforming vectors:

$$\mathbf{g}_{kn} = \left[\sum_{l=1}^{K}\sum_{j=1}^{d_l} u_l\mathbf{H}_{lk}^H\mathbf{f}_{lj}w_{lj}\mathbf{f}_{lj}^H\mathbf{H}_{lk} + \lambda_k\mathbf{I}\right]^{-1}\mathbf{H}_{kk}^H\mathbf{f}_{kn}w_{kn}u_k \tag{22}$$

To determine the optimal value of the lagrange multiplier λ_k we can multiply the derivative of the lagrangian w.r.t $\mathbf{g}_{kn}$ by the BF vector hence the following holds true:

$$\sum_{n=1}^{d_k}\left[\mathbf{g}_{kn}^H\frac{\partial J}{\partial \mathbf{g}_{kn}^*}\right] = 0$$

solving the equation above w.r.t. the lagrange multiplier we get:

$$\begin{aligned}\lambda_k = &\ \frac{1}{P_k}\left[\sum_{n=1}^{d_k}\mathbf{g}_{kn}^H\mathbf{H}_{kk}^H\mathbf{f}_{kn}w_{kn}u_k\right]\\ &-\frac{1}{P_k}\left[\sum_{n=1}^{d_k}\sum_{l=1}^{K}\sum_{j=1}^{d_l}u_l\mathbf{g}_{kn}^H\mathbf{H}_{lk}^H\mathbf{f}_{lj}w_{lj}\mathbf{f}_{lj}^H\mathbf{H}_{lk}\mathbf{g}_{kn}\right]\end{aligned} \tag{23}$$

The final algorithm (PS-MWSR algorithm in Table **Algorithm 2**) for the per-stream optimization requires the iteration of the three steps for the optimization of Rx filters, weights, Tx beamforming vectors, in the prescribed order, untill convergence.

Algorithm 2 PS-MWSR Per-Stream Algorithm for MIMO IFC

Fix an arbitrary initial set of precoding matrices $\mathbf{G}_k$, $\forall \in k = \{1, 2 \ldots K\}$
set $n = 0$
repeat
  $n = n + 1$
  for $k = 1$ to K **do**
    Given $\mathbf{g}_i^{(n-1)}$ $\forall i$, compute $\mathbf{f}_{kl}^{(n)}$ and $w_{kl}^{(n)}$ from (20) and (21) respectively for $l = 1, \ldots, d_k$
    Given $\mathbf{f}_{kl}^{(n)}$ and $w_{kl}^{(n)}$ for $l = 1, \ldots, d_k$, compute $\mathbf{g}_{kl}^{(n)}$ for $l = 1, \ldots, d_k$ using (22)
  end for
until convergence

C. Rate Duality in MIMO IFC

In the previous section the expressions of the beamformer (22) and the MMSE Rx filter (20) are given when we assume to work per stream. Looking deeper at the expression of the cost function (17) it is possible to establish a duality relationship between the DL IFC considered and a dual UL IFC:

- The DL channel matrix $\mathbf{H}_{kl}$ becomes $\tilde{\mathbf{H}}_{lk}^H$ in the dual UL

$$\mathbf{G}_k = \left(\sum_{l=1}^{K} \mathbf{H}_{lk}^H \mathbf{F}_l \mathbf{W}_l \mathbf{F}_l^H \mathbf{H}_{lk} - \frac{1}{P_k} \left(\left(\sum_{l \neq k} \mathrm{Tr}\{\mathbf{W}_l \mathbf{J}_l^{(k)}\} - \mathrm{Tr}\{\mathbf{W}_k \mathbf{J}_k^{(l)}\} \right) - \mathrm{Tr}\{\mathbf{W}_k \mathbf{N}_k\} \right) \mathbf{I} \right)^{-1} \mathbf{H}_{kk}^H \mathbf{F}_k \mathbf{W}_k \tag{16}$$

$$\mathbf{J}_l^{(k)} = \mathbf{F}_l^H \mathbf{H}_{lk} \mathbf{G}_k \mathbf{G}_k^H \mathbf{H}_{lk}^H \mathbf{F}_l; \quad \mathbf{J}_k^{(l)} = \mathbf{F}_k^H \mathbf{H}_{kl} \mathbf{G}_l \mathbf{G}_l^H \mathbf{H}_{kl}^H \mathbf{F}_k; \quad \mathbf{N}_k = \mathbf{F}_k^H \mathbf{R}_{n_k n_k} \mathbf{F}_k$$

- The Rx (Tx) filter in the DL (UL) $\mathbf{f}_{kn}$ ($\mathbf{g}_{kn}$) becomes the Tx (Rx) filter in the UL (DL) $\tilde{\mathbf{g}}_{kn}^H$ ($\tilde{\mathbf{f}}_{kn}^H$)
- The unit DL Tx signal variance for stream (k, n) becomes $u_k w_{kn}$ in the dual UL channel
- DL noise covariance matrix $\mathbf{R}_{n_k n_k} = \sigma_k^2 \mathbf{I}$ becomes $\lambda_k \mathbf{I}$ in the UL.

With this relationship we can interpret the BF filter in the DL as an MMSE Rx filter in the virtual UL IFC.
A similar reasoning can be naturally extended to the per-user approach discussed in section III-A

IV. Deterministic Annealing to Avoid Local Optima

In the previous section we have described an alternating optimization algorithm that designs BF and RX filters in order to maximize the WSR in a K-user MIMO IFC. As already remarked, the WSR cost function is a non convex function and this makes the optimization troublesome due to the presence of many local optima. In optimization, a number of heuristic approaches exist to handle non convex optimization problems. Some examples of such methods are: genetic algorithms, ant colony optimization or simulated annealing (SA). We will describe briefly the SA approach. This method takes its name from the physical annealing process in which a system is first “melted” and then slowly cooled down in order to allow the atoms in the system to find a state with lower energy until the system is “frozen” in a globally optimum state.
In SA the problem is optimized using a sequence of random moves, the size of which reduces as a parameter called temperature decreases. The random moves would allow the optimization process to get out of local optima. In a certain sense, the randomness tend sto convexify the problem. Cooling protocols have been derived to allow ending up in the global optimum with high probability. *Deterministic Annealing* (DA) is a related technique but does not involve any randomness, see e.g. [18]. In DA, an increase of the temperature parameter allows to convexify the problem: the temperature parameter transforms (deterministically) the originally non-convex cost function into a convex cost function (convex should be replaced by concave in the case of maximization). So, at high temperature, there is no problem in finding the global optimum. Then gradually the temperature gets reduced, making the problem increasingly non-convex. However, if the temperature variation is sufficiently small, the gobal optimum at the previous higher temperature will be in the region of attraction of the global optimum at the next lower temperature and the global optimum remains tracked in this way.
As in physical systems, also in the optimization problem it can happen that phase transitions occur as the temperature cools down [18]. A phase transition corresponds to a split of the trajectory (as a function of temperature) of the global optimum into several trajectories. From a mathematical perspective a phase transition is characterized by the Hessian of the problem becoming singular at a critical temperature (hence being positive semidefinite instead of postive definite). In our problem the cost function is the WSR, a highly non convex function, and the annealing parameter is related to the noise variance, $t \propto \sigma^2$ (or the inverse of the SNR).

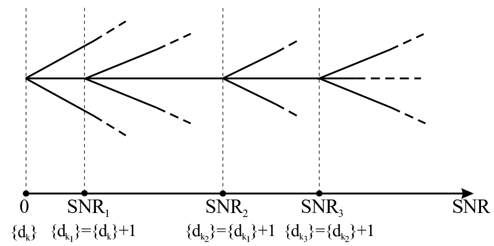

Fig. 2: Phase transitions representation

Interestingly also in WSR maximization in a K-user MIMO IFC, phase transitions can appear. At low SNR (high noise variance), any interference is negligible compared to the noise. Hence, all links can be considered decoupled, and, like in single-user MIMO, rate maximization becomes SNR maximization for a single stream to which all transmit power is devoted. Hence in link k, the optimal Tx and Rx filters correspond to the left and right singular vectors corresponding to the largest singular value of $\mathbf{H}_{kk}$. Hence, as the SNR goes to zero, the globally optimum solution is clear. However, zero SNR itself is already a phase transition because as soon as the SNR becomes positive, a multitude of local optima may exist that we shall interpret below. As the SNR increases further, at some point another phase transition may occur, at which point a second stream needs to be introduced in one of the links. We shall see that at such a phase transition, it is possible to determine the filters corresponding to the new stream. However, as soon as the SNR increases further, many further local optima get introduced due to the appearance of the additional stream. Then, as the SNR increases further, another phase transition can occur, with the introduction of one more stream at one of the transmitters. This process goes on until a stream distribution is reached, at some higher SNR, corresponding to a maximal stream distribution for which interference alignment is feasible. Indeed, at very high SNR, the Tx and Rx filters coverge to the (max WSR-)IA solution, and the sum rate prelog is maximized if the number of streams is maximized (see [14]). This whole process is depicted schematically in Fig.2. One interesting observation is that it is fairly straightforward to check that all extrema of the WSR correspond to local maxima. So, whereas the Hessian is in general indefinite, reflecting the non-concavity of the WSR cost function, it turns out that the Hessian is always negative definite when evaluated at an extremum (or semidefinite at the phase transitions).

Whereas DA is about tracking of a global optimum, the tracking of extrema, the zeros of the KKT conditions, is actually called a homotopy method. So in DA, going from one phase transition to the next and tracking the (appropriate) extremum, this could be considered a homotopy method.

A. Homotopy Methods

Homotopy methods [19] are used to find the roots of a non-linear system of equations $\mathcal{F}(x) = 0$. A homotopy transformation is such that it starts from a trivial system $\mathcal{G}(x)$, with known solution, and it evolves towards the target system $\mathcal{F}(x)$ via continuous deformations according to the homotopy parameter $t = 0 \to 1$:

$$\mathcal{H}(x,t) = (1-t)\,\mathcal{G}(x) + t\mathcal{F}(x)$$

Predicting the solution at the next value of $t^{(i+1)} = t^{(i)} + \Delta t$ is called an Euler prediction step; a solution at $t^{(i+1)}$ can be refined using a Newton correction step for fixed t. A property of Homotopy continuation methods for the solution of system of equation is that the number of solutions in the target system is at most equal to the number of solutions in the trivial system. The number of solutions with varying t remains constant as long as the Jacobian (w.r.t. x and t jointly) is full rank. So as t reaches 1, it can happen that the Jacobian becomes singular.

B. Homotopy Applied to IA

Homotopy method can be applied to the IA problem, in particular here it is not really suggested for computing IA solutions, but for counting number of solutions. The objective in IA is to design Tx and Rx filters that satisfy the ZF conditions

$$\mathbf{F}_k^H \mathbf{H}_{kl} \mathbf{G}_l = \mathbf{0} \quad \forall l \neq k \tag{24}$$

and the rank conditions

$$\mathrm{rank}(\mathbf{F}_k^H \mathbf{H}_{kk} \mathbf{G}_k) = d_k \quad \forall k \in \{1, 2, \ldots, K\} \tag{25}$$

which correspond to the traditional single user MIMO constraint $d_k \le \min(M_k, N_k)$ for d_k streams to be able to pass over the k-th link. The main constraints are the n ZF conditions in (24). These conditions are bilinear equations in the Tx and Rx filters, hence they are of second order. As a result, the overall order of the ZF conditions jointly is 2^n, which is also the maximum number of solutions. It turns out that due to the particular structure of the ZF conditions (in a given ZF condition only one Tx and Rx filter appear), the actual number of solutions is much more limited. To analyze the number of IA solutions, the following approach has been proposed in [20]. Instead of choosing the homotopy parameter to be related to SNR, we choose it here to attenuate the MIMO channel singular values beyond the main ones:

$$H_{ji} = \sum_{k=1}^{d} \sigma_{jik} \mathbf{u}_{jik} \mathbf{v}_{jik}^H + t \sum_{k=d+1} \sigma_{jik} \mathbf{u}_{jik} \mathbf{v}_{jik}^H \quad .$$

The IA Jacobian is still full rank if we reduce $\mathrm{rank}(H_{ji})$ to $\max(d_j, d_i)$. Hence we can still count the same number of IA solutions when $t = 0$. The case of $d_k \equiv d = 1$ is considered here. Then finding the IA solutions at $t = 0$ becomes trivial. Indeed, IA requires

$$\mathbf{f}_j^H \mathbf{u}_{ji1} \mathbf{v}_{ji1}^H \mathbf{g}_i = 0$$

or hence either $\mathbf{f}_j^H \mathbf{u}_{ji1} = 0$ or $\mathbf{v}_{ji1}^H \mathbf{g}_i = 0$. The joint Tx-Rx ZF is achieved by either the Tx or the Rx suppressing the particular interfering stream. This analysis supports a suggestion provided in [4] which states that it should be possible to check IA feasibility and count the number of IA solutions by verifying if the ZF task can be properly distributed over Tx and Rx filters. The idea is that a stream transmitted from TX k and causes interference to the non intended RX j needs to be suppressed at either the Tx or at the Rx. Denoting with t_{kj} the size of the subset of streams d_k, that are received at Rx j that the k-th Tx suppresses, and with r_{kj} the size of the subset of streams d_k, that are received at Rx j, that the j-th Rx suppresses, the sum of these two quantities should be: $t_{kj} + r_{kj} \ge d_k$. The total number of streams that Tx k can suppress is at most $M_k - d_k$ and the total number of streams that the j-th Rx can suppress is not greater than $N_j - d_j$. Therefore, to check the feasibility of an interference alignment solution, the following conditions should be satisfied:

$$\begin{array}{ll} \sum_{j \neq k} t_{kj} & \le M_k - d_k \\ \sum_{k \neq j} r_{kj} & \le N_j - d_j \end{array} \tag{26}$$

$$\forall t_{kj}, r_{kj} \in \{0, 1 \ldots, d_k\}, \text{ and } t_{kj} + r_{kj} = d_k$$

$$\max_{k \neq j}(d_j - [M_k - N_j]) \le (N_j - d_j) \forall j \in \{1, \ldots, K\}$$

As before, due to alignment duality, IA must be checked also when the sets of M_k and N_k are interchanged (the dual channel case). One possible way to verify if all this inequalities are satisfied or not is to check all the possible $\prod_{k=1}^{K} (d_k + 1)^{K-1}$ combination of t_{kj} and r_{kj}. So, the homotopy method allows to substantiate this approach, at least in the single stream per link case.

More generally, determining IA solutions by continuation methods can be obtained by perturbing the ZF conditions up to first order

$$(\mathbf{F}_j^H + d\mathbf{F}_j^H)(\mathbf{H}_{ji} + d\mathbf{H}_{ji})(\mathbf{G}_i + d\mathbf{G}_i) = 0$$

Assuming that an IA solution for channel $\mathbf{H}_{ji},\ \forall (i,j)$ has already been determined using filters $\mathbf{F}_j$ and $\mathbf{G}_i$ then considering only the terms up to first order in the product above we get:

$$\mathbf{F}_j^H \mathbf{H}_{ji} d\mathbf{G}_i + d\mathbf{F}_j^H \mathbf{H}_{ji} \mathbf{G}_i = -\mathbf{F}_j^H d\mathbf{H}_{ji} \mathbf{G}_i .$$

To find the IA solution for channel $(\mathbf{H}_{ji} + d\mathbf{H}_{ji})$ we determine the matrices $d\mathbf{F}_j^H$ and $d\mathbf{G}_i\ \forall i, j$ by solving linear equations.

C. Homotopy Applied to WSR

As remarked previously, maximizing WSR at very high SNR corresponds to determining IA solutions, as can be seen immediately from the augmented WSR cost function. Any IA solution leads to a local maximum of WSR. Now, consider again the low rank channels considered above, in which we can discribe and count the number of IA solutions. Instead of increasing the channel rank first, we shall lower the SNR (or increase the noise variance). Note that we can even consider linear homotopy here by using t to multiply the transmit powers or the noise variances, since the augmented WSR cost function is linear in transmit powers or noise covariances. By non-singularity of the Jacobian, the various IA solutions will each get transformed into a local WSR maximum as the SNR lowers. Until a phase transition is reached in which some stream gets switched off. This will eliminate a subset of the IA possibilities and hence a subset of the local WSR maxima. This process continues until at low SNR there is one stream per link. For any given SNR, the low rank channel can also be transformed until the original full rank channel, without affecting the number of local maxima.

V. Deterministic Annealing for WSR Maximization

What we propose in this paper is to extend the MWSR algorithm presented before in order to include DA and hence reduce the probability to be trapped in local optima. So we consider again DA for the original full rank channels, for SNR increasing from zero. To modify the algorithm proposed in Table Algorithm 1 to include DA we only need to run the algorithm for each SNR point initializing the algorithm with the optimal beamformers found at the previous SNR iteration. However, this does not handle phase transitions, corresponding to the introduction of a new stream. Hence, at every SNR increment, we need to try adding a stream to each of the K links (one at a time). It is possible to find the proper initialization for the Tx and Rx filters of the new stream analytically.

A. Initialization at Phase Transitions

To find the direction of the BF vector corresponding to the new stream, indexed as (k, n), we need to optimize our per-stream cost function (17) w.r.t. the quantities corresponding to the new allocated stream. Note that the new stream, if it should be switched on, will be switched on with very small power. Hence the new stream will barely perturb the existing streams.
For the moment we do not include in the optimization function the power constraint, so we need to find the Tx and Rx filter that minimize the MSE for stream (k, n). The derivative of the MSE w.r.t. the Rx filter is:

$$\frac{\partial \mathcal{O}}{\partial \mathbf{f}_{kn}} = -\mathbf{g}_{kn}^H \mathbf{H}_{kk}^H + \mathbf{f}_{kn}^H \mathbf{H}_{kk} \mathbf{g}_{kn} \mathbf{g}_{kn}^H \mathbf{H}_{kk}^H + \mathbf{f}_{kn}^H \mathbf{R}_{\overline{kn}} \tag{27}$$

considering only the terms up to first order in $\mathbf{g}_{kn}$ the expression for the receiver is $\mathbf{f}_{kn} = \mathbf{R}_{\overline{kn}}^{-1} \mathbf{H}_{kk} \mathbf{g}_{kn}$ that has an expression like matched filter (MF) in colored noise. Consider a parametrization of the BF vector in direction vector and power allocation like: $\mathbf{g}_{kn} = \overline{\mathbf{g}}_{kn} \sqrt{p_{kn}}$ and define $x_{kn} = \overline{\mathbf{g}}_{kn}^H \mathbf{H}_{kk}^H \mathbf{H}_{kk} \overline{\mathbf{g}}_{kn}$. Substituting the Rx filter with its expression in function of the BF, the MSE cost function can be written as:

$$e_{kn} = 1 - p_{kn} x_{kn} + (p_{kn} x_{kn})^2$$

Considering only the contribution up to first order in x_{kn} the minimization of the MSE leads to the maximization of x_{kn} and hence the optimal BF vector direction is

$$\overline{\mathbf{g}}_{kn} = v_{max}(\mathbf{H}_{kk}^H \mathbf{R}_{\overline{kn}}^{-1} \mathbf{H}_{kk}) \tag{28}$$

where $v_{max}(\mathbf{A})$ represents the eigenvector corresponding to the maximum eigenvalue of matrix $\mathbf{A}$. Once we have the direction of the BF associated to the new stream we need to determine the corresponding power.
Consider $\mathbf{G}_k$ the BF matrix obtained untill the current SNR point for link k and its decomposition as $\mathbf{G}_k = \overline{\mathbf{G}}_k \mathbf{P}_k^{1/2}$, where $\overline{\mathbf{G}}_k$ has normalized columns and $\mathbf{P}_k^{1/2}$ is the power allocation matrix. For the per-stream approach the MMSE is diagonal and hence:

$$\mathbf{E}_k^{-1} = \mathbf{I} + \mathbf{G}_k^H \mathbf{H}_{kk}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}^H \mathbf{G}_k^H = \mathbf{I} + \mathbf{D}\mathbf{P}_k$$

Introducing the additional stream we obtain the following matrix :

$$\mathbf{X} = [\mathbf{G}_k \; \mathbf{g}_{kn}]^H \mathbf{H}_{kk}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{kk}^H [\mathbf{G}_k \; \mathbf{g}_{kn}] = \begin{bmatrix} \mathbf{D}\mathbf{P}_k & \sqrt{p_{kn}}\mathbf{u} \\ \sqrt{p_{kn}}\mathbf{u}^H & a p_{kn} \end{bmatrix}$$

where $\mathbf{u} = \mathbf{G}_k^H \mathbf{H}_{kk}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{kk} \overline{\mathbf{g}}_{kn}$ and $a = \overline{\mathbf{g}}_{kn}^H \mathbf{H}_{kk}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{kk} \overline{\mathbf{g}}_{kn}$. The corresponding rate for user k is

$$\ln |\mathbf{E}_k^{-1}| = \ln |\mathbf{I} + \mathbf{X}| = \ln |\mathbf{I} + \mathbf{D}\mathbf{P}_k| + \ln(1 + p_{kn} d_{kn})$$

$d_{kn} = a - \mathbf{u}^H (\mathbf{I} + \mathbf{D}\mathbf{P}_k)^{-1} \mathbf{u}$.
Finally to find the power allocation among different streams of user k we propose the following.

1) Jammer Water-Filling (JWF) algorithm: Include in the matrix $\mathbf{P}_k$ the power allocated to the new stream p_{kn} and in the diagonal matrix $\mathbf{D}$ include the element d_{kn} associated to the new stream. To find the power allocation matrix we take the original per-stream cost function (17) and optimize it with respect to (and then eliminate) the weights w_{kn} for link k. After this, the terms in the WSR affected by $\mathbf{P}_k$ are

$$\mathcal{O} = \ln |\mathbf{I} + \mathbf{D}\mathbf{P}_k| - Tr\{\mathbf{P}_k \boldsymbol{\Delta}\} - \lambda_k (Tr\{\mathbf{P}_k\} - P_k)$$

where $Tr\{\mathbf{P}_k \boldsymbol{\Delta}\}$ takes into account the interference power generated to the non intended receivers (for this reason we called this algorithm Jammer WF):

$$Tr\{\mathbf{P}_k \boldsymbol{\Delta}\} = \sum_i p_{ki} \underbrace{\sum_{l \neq k} \frac{u_l}{u_k} \sum_{m=1}^{d_l} w_{lm} |\mathbf{f}_{lm}^H \mathbf{H}_{lk} \overline{\mathbf{g}}_{ki}|^2}_{\boldsymbol{\Delta}_{ki}} .$$

Deriving the cost function above w.r.t. p_{ki} the expression for the power allocation is:

$$p_{ki} = \left[\frac{1}{\lambda_k + \boldsymbol{\Delta}_{ki}} - \frac{1}{d_{ki}} \right]_+ \tag{29}$$

where $[(.)]_+ = max((.), 0)$. To find the optimal value of λ_k we first check if the power constraint is inactive. In particular we determine the powers using (29) assuming $\lambda_k = 0$ and we verify if the transmitted power is less then the power constraint. If the power constraint is not satisfied we determine λ_k using a bisection method. Consider the following function of the lagrange multiplier

$$\mathcal{T}(\lambda_k) = \sum_i \left[\frac{1}{\lambda_k + \mathbf{\Delta}_{ki}} - \frac{1}{d_{ki}} \right]_+ - P_k$$

as we can see $\mathcal{T}(\lambda_k)$ is a decreasing function of λ_k. In particualr for $\lambda_k^0 = 0$ $\mathcal{T}(\lambda_k) > 0$ while for λ_k^1, determined as water-level of a tradition WF algorithm on $\mathcal{T}(\lambda_k)$ when $\mathbf{\Delta}_{ki} = 0$, $\forall i$, the function $\mathcal{T}(\lambda_k) < 0$. The optimal value $\lambda_k^\star$ can be found using a bisection algorithm to solve $\mathcal{T}(\lambda_k) = 0$. The final extended BF matrix $\mathbf{G}_k = [\mathbf{G}_k \ \ \mathbf{g}_{kn}]$ obtained using the procedure described so far is used as initialization of the DA-WSR for the following SNR point.

Algorithm 3 DA-MWSR Algorithm for MIMO IFC

set $t = 0$
Fix the initial set of precoding matrices $\mathbf{G}_k$, $\forall \in k = \{1, 2 \dots K\}$
repeat
  increment SNR: $t^{(i+1)} = t^{(i)} + \delta t$
  Augment $\mathbf{G}$
  repeat
    Given $\mathbf{G}_k$ compute $\mathbf{F}_k$ and $\mathbf{W}_k$, $\forall k$
    Given $\mathbf{F}_k$, $\mathbf{W}_k$, compute $\mathbf{G}_k$ $\forall k$
  until convergence
until target SNR is reach

A remaining open question is now the following: at a phase transition, even if we are able to determine the solution analytically, the global maximum splits up into a whole set of local maxima trajectories. The question is whether the algorithm above will in fact track the global maximum. The answer is yes. Indeed it turns out that an alternating optimization approach as the one considered here (or also the one used in [15]), in spite of the non-concavity of the problem, optimizes (globally) the WSR up to second order in transmit power (or SNR). Indeed, we are able to determine analytically the optimal Tx and Rx filters up to zeroth order in Tx power, the one iteration of an alternating optimization approach will provide the optimal Tx and Rx filters up to first order in Tx power, which maximize WSR up to second order in Tx power. In other words, the alternating optimization approach inherently sets course on the trajectory of the global optimum.

VI. Simulation Results

We provide here some simulation results to compare the performance of the proposed max-WSR algorithm (DA-MWSR) where we deterministic annealing is used to avoid local optimal point. i.i.d Gaussian channels (direct and cross links) are generated for each user. For a fixed channel realization transmit and receiver filters are computed based on IA algorithm and DA-MWSR algorithm over multiple SNR points. The resulting sum rate (SR) is averaged over 50 channel

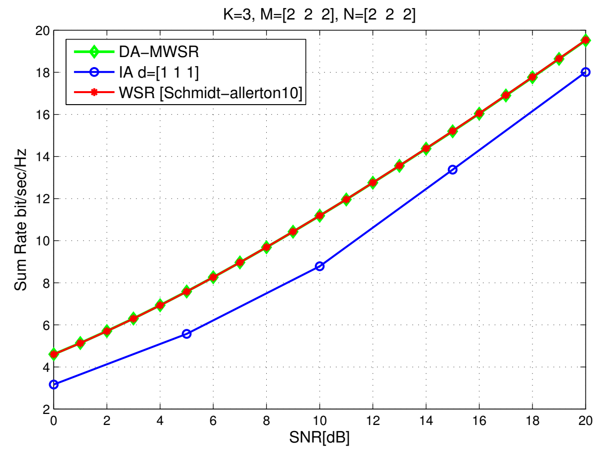

Fig. 3: WSR for $K = 3$, $M_k = 2$, $N_k = 2$

realizations. In Fig. 3 we compare the SR obtained using three different algorithms. In particular we compare our algorithm DA-MWSR with IA algorithm proposed in [5] and another WSR algorithm recently proposed in [15] where also a numerical continuation method is used to find the BF to maximize the WSR. This algorithm works only for single stream transmissions. As we can see both algorithms that maximize the WSR outperform IA in all SNR regimes. On the other hand there is no difference between the proposed algorithm and the one in [15].

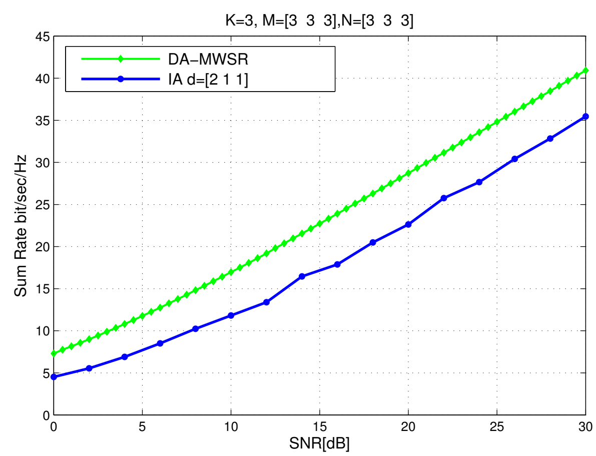

Fig. 4: WSR for $K = 3$, $M_k = 3$, $N_k = 3$

In Fig. 4 we report the SR for a $K = 3$ users IFC where each Tx and Rx are equipped with $M_k = N_k = 3$ antennas. According to IA the total maximum number of streams that can be transmitted in the system is $d = 4$. We determine the IA beamformers and receiver filters using the algorithm in [5] for a stream distribution $d_1 = 2, d_2 = d_3 = 1$. We compare the performance of IA with our algorithm where the annealing parameter, noise variance, has been increased of $\delta t = 0.5$ dB. As we can see the proposed algorithm outperforms IA even at high SNR regime. The slope of the sum rate obtained using our algorithm is the same of the IA curve. This shows that the correct number of streams has been sent.

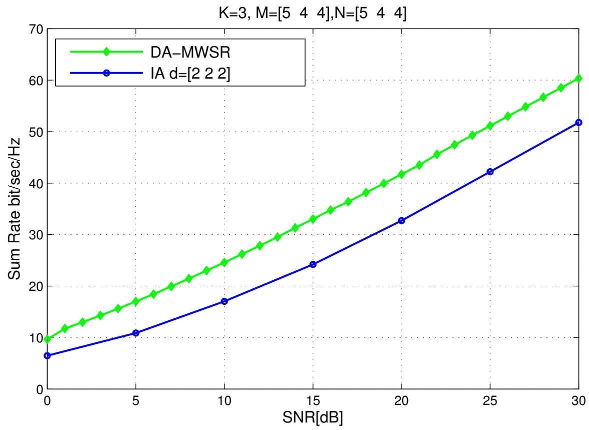

Fig. 5: WSR for $K = 3,\ M_1 = N_1 = 5,\ M_i = N_i = 4,\ i = 2, 3,\ d_k = 2\ \forall k$

Finally Fig. 5 depicts the performances of the proposed algorithm, WSR DA, in comparison with IA for a $K = 3$ user IFC with an asymmetric antennas distribution. We assume that $M_1 = N_1 = 5,\ M_i = N_i = 4\ i = 2, 3$, the stream distribution, according to IA is $d_k = 2\ \forall k$. As we can see also in this case the proposed algorithm outperform IA keeping the same slope in the high SNR regime.

VII. Conclusions

In this paper we addressed maximization of the weighted sum rate for the MIMO IFC. We introduced an iterative algorithm to solve this optimization problem that is characterized by the presence of several possible local minima. To avoid to be stack in one suboptimal stationary point we propose to introduce Deterministic Annealing. This approach allows to track the variation of the known solution of one version of the problem into the unknown solution of the desired version by a controlled variation of a parameter called temperature. In our problem the temperature is related to the inverse of the SNR. The proposed algorithm include filter design for the progressive switching on of streams as the SNR increases

In the second part of the paper we study IA feasibility. Exploring the fact that IA feasibility is unchanged when the MIMO crosslink channel matrices have a reduced rank, equal to the maximum of the number of streams passing through them we propose a new way to study the problem using numerical continuation method. The rank reduction simplifies IA design and feasibility analysis, and allows in particular a counting of the number of IA solutions. In this approach the temperature parameter is a scale factor for the remaining channel singular values, the solution for reduced rank channels can be evolved into that for arbitrary channels.

VIII. Acknowledgment

EURECOM's research is partially supported by its industrial members: BMW Group, Swisscom, Cisco, ORANGE, SFR, ST Ericsson, Thales, Symantec, SAP, Monaco Telecom. The research of EURECOM and Intel Mobile Communications France is also supported in part by the EU FP7 projects CROWN, SACRA and WHERE2.

References

[1] R. H. Etkin, D. N. C. Tse, and Hua Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. on Inform. Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[2] V.R. Cadambe and S.A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. on Inform. Theory*, vol. 54, no. 8, pp. 3425 –3441, Aug. 2008.

[3] C.M. Yetis, Tiangao Gou, S.A. Jafar, and A.H. Kayran, "On feasibility of interference alignment in mimo interference networks," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4771 –4782, 2010.

[4] F. Negro, S. P. Shenoy, I. Ghauri, and D.T.M. Slock, "Interference alignment feasibility in constant coefficients MIMO interference channel," *in Proc. 11th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2010)*, June 2010.

[5] K. Gomadam, V.R. Cadambe, and S.A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, Dec 2008.

[6] Steven W. Peters and Robert W. Heath, "Interference alignment via alternating minimization," in *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009, pp. 2445–2448.

[7] I. Santamaria, O. Gonzalez, R.W. Heath, and S.W. Peters, "Maximum sum-rate interference alignment algorithms for mimo channels," in *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, 2010, pp. 1 –6.

[8] Sigen Ye and R.S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. on Signal Processing*, vol. 51, no. 11, pp. 2839–2848, Nov 2003.

[9] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. on Signal Processing*, vol. 52, no. 1, pp. 214 –226, Jan. 2004.

[10] David A. Schmidt, Shi Changxin, Randall A. Berry, Michael L. Honig, and Wolfgang Utschick, "Minimum mean squared error interference alignment," in *Proc. 43rd IEEE Annual Asilomar Conference on Signals, Systems & Computers*, Pacific Grove,California, USA, Nov 2009.

[11] Z.K.M. Ho and D. Gesbert, "Balancing egoism and altruism on interference channel: The mimo case," in *Communications (ICC), 2010 IEEE International Conference on*, May 2010, pp. 1 –5.

[12] F. Negro, S.P. Shenoy, I. Ghauri, and D.T.M. Slock, "On the mimo interference channel," in *Information Theory and Applications Workshop (ITA), 2010*, 31 2010-feb. 5 2010, pp. 1 –9.

[13] S.S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[14] Francesco Negro, Irfan Ghauri, and Dirk T.M. Slock, "Optimizing the noisy mimo interference channel at high snr," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, 29 2010-oct. 1 2010, pp. 254 –261.

[15] David A. Schmidt, Wolfgang Utschick, and Michael L. Honig, "Beamforming techniques for single-beam mimo interference networks," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, 29 2010-oct. 1 2010, pp. 1182 –1187.

[16] Scutari Gesualdo, Daniel P Palomar, and Sergio Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. on Signal Processing*, vol. 57, no. 5, pp. 1917–1935, May 2009.

[17] Shi Changxin, David A. Schmidt, Randall A. Berry, Michael L. Honig, and Wolfgang Utschick, "Distributed interference pricing for the MIMO interference channel," in *Proc. IEEE Int'l Communications Conf. (ICC)*, Dresden, Germany, 2009.

[18] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210 –2239, nov 1998.

[19] E.L. Allgower and K. Georg, "Introduction to numerical continuation methods," *Philadelphia, PA: Society for Industrial and Applied Mathematics*, 2003.

[20] O.. Gonzalez and I. Santamaria, "Interference Alignment in Single-Beam MIMO Networks via Homotopy Continuation," in *IEEE Int'l Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, Prague, 2011.