A Formulation of the Channel Capacity of Multiple-Access Channel

Yoichiro Watanabe, *Member, IEEE* and Koichi Kamoi, *Member, IEEE*

Abstract— The necessary and sufficient condition of the channel capacity is rigorously formulated for the N-user discrete memoryless multiple-access channel (MAC). The essence of the formulation is to invoke an *elementary* MAC where sizes of input alphabets are not greater than the size of output alphabet. The main objective is to demonstrate that the channel capacity of an MAC is achieved by an elementary MAC included in the original MAC. The proof is quite straightforward by the very definition of the elementary MAC. Moreover it is proved that the Kuhn-Tucker conditions of the elementary MAC are strictly sufficient and obviously necessary for the channel capacity. The latter proof requires some steps such that for the elementary MAC every solution of the Kuhn-Tucker conditions reveals itself as local maximum on the domain of all possible input probability distributions and then it achieves the channel capacity. As a result, in respect of the channel capacity, the MAC in general can be regarded as an aggregate of a finite number of elementary MAC's.

Index Terms— multiple-access channel (MAC), elementary MAC, master elementary set, channel capacity, Kuhn-Tucker conditions, capacity region, boundary equation

I. INTRODUCTION

THE channel capacity is without question recognized as an essential subject of the (discrete memoryless) multiple-access channel (MAC) with N input-terminals and one output-terminal. Since it is defined as the maximum of the mutual information, we are familiar with the so-called Kuhn-Tucker conditions as necessary to achieve the channel capacity. Up to now, however, the Kuhn-Tucker conditions are not entirely examined as sufficient for the N-user MAC except for the simplest case of single user discrete memoryless channel (DMC). Thus it is natural to ask how the sufficiency could be formulated for the case of MAC in general.

In this paper, we demonstrate that there exists a non-trivial MAC where the Kuhn-Tucker conditions are strictly sufficient (and obviously necessary) for the channel capacity. We refer to it as *an elementary* MAC whose sizes of input alphabets are not greater than the size of output alphabet. Evidently the DMC is an elementary MAC. The most of this paper is devoted to the proof that the Kuhn-Tucker conditions are sufficient (the necessity is self-evident) for the channel capacity of the elementary MAC.

On the other hand, for any given N-user MAC we can uniquely determine a finite set of elementary MAC's. It is an aggregate of *the largest* possible elementary MAC's included in the given N-user MAC and is referred to as *the master*

elementary set to be denoted by Ω_N. We demonstrate that the channel capacity of the N-user MAC is achieved by the channel capacity of *an elementary* MAC of the set Ω_N. The proof here appears quite straightforward by merely appealing to the very definition of the elementary MAC without asking for any other features such that the Kuhn-Tucker conditions are sufficient.

Thus an MAC in general can be regarded as simply an aggregate of elementary MAC's where the Kuhn-Tucker conditions are necessary and sufficient for the channel capacity. Roughly speaking, an MAC comprises a finite number of elementary MAC's. This statement is a basic idea behind our formulation of this paper.

Here we must emphasize that several steps are required to prove the sufficiency of the Kuhn-Tucker conditions of the elementary MAC. In fact, we need to prove two distinctive features: The first is that for the elementary MAC every solution of the Kuhn-Tucker conditions is *local maximum* on the domain of all possible input probability distribution (IPD) (or the probability simplex, see Cover [1] for this terminology and we refer to them as IPD *vectors* for our purposes). The second is that for the elementary MAC a set of IPD vectors for which the value of the mutual information is not smaller than the arbitrary positive number is *connected* on the domain of all possible IPD vectors. To prove the second property of connectedness we require the first property of local maximum. Then it follows after a bit of procedures that solutions of the Kuhn-Tucker conditions are uniquely determined, that is, each solution takes the same value for the mutual information and therefore it achieves the channel capacity.

For the explicit description of our concept we take a logical stream as follows: After defining the elementary MAC and determining the set Ω_N, we first prove as the main theorem that the channel capacity of an N-user MAC is achieved by the channel capacity of an elementary MAC of Ω_N and then we prove as the second theorem that the Kuhn-Tucker conditions are sufficient for the channel capacity of the elementary MAC. These are the main objective of this paper.

After Shannon [2], the study of multiuser channel (multi-terminal network) has long been carried out in various fields including MAC, broadcast channel, relay channel, interference channel, two-way channel and so forth. The channel coding theorem was proved independently by Liao [3], Ahlswede [4] and Meulen [5]. These are followed by many authors ([6], [7], [8], [9], [10], [11], [12]) to provide a deeper insight into the capacity region. Recently, information-theoretic approach has been adopted to large scale networks, such that code division multiple-access channel, continuous time multiple-access channel and space-time multiple-access channel (e.g.,

Yoichiro Watanabe is with the Department of Intelligent Information Eng. and Sci., Doshisha University, Kyotanabe, Kyoto, 619-0321 Japan.

Koichi Kamoi is with Knowledge X Inc., Wako-Shi, Saitama, 351-0114 Japan.

[13], [14], [15], [16], [17]), as we know. Also a computation procedure for the channel capacity of MAC has been developed (e.g., [18]).

The purpose of the study of MAC is mostly to investigate the multiuser coding that retains both reliability and efficiency. The investigation has been carried out mostly on the computational calculations for practical applications. Not much has been made for the mathematical rigorousness of the formulation since it appears rather hard to solve a non-linear optimization problem of the mutual information with several variables under constraints. We have been highly expecting a theoretical foundation, in particular, for the rigorous evaluation of the channel capacity and the exact determination of the capacity region for the MAC in general. These can provide us with the mathematical essence as well as the fine structure inherent in the MAC. Also we believe that these can in part complement the computational approaches to various applications as well.

In the past, for the MAC of two-user and binary output [19], we have shown that the Kuhn-Tucker conditions are necessary and sufficient for the channel capacity. The basic idea was to identify the channel matrix of the MAC as *a linear* mapping from the convex-closure of IPD vectors to the range of output probability distributions. Now we expand the idea and remind a clear conception to describe the MAC as a pair of channel matrix P and domain X where X is a set of IPD vectors and P is interpreted as a mapping (*non-linear* in general) from IPD vectors to output probability distributions. Any quantity such as mutual information and so forth is considered as a function of IPD vectors defined on a restricted domain (a sub-set) of X. These are seemingly non-standard in contrast to the ordinary description of information theory as in [1]. However, we assure ourselves that these conceptions including the notation adopted in this paper are so successful to overcome some difficulties and cumbersome procedures underlying in the non-linear optimization problem relating to the mutual information of the MAC.

In Section II we describe some expressions and definitions to be used in this paper. In particular we introduce an elementary MAC and the master elementary set for the MAC. In Section III we prove the main theorem of this paper as Theorem 1 followed by indicating the value of this theorem. In Section IV we investigate distinctive features of the elementary MAC that are required to prove the succeeding Theorem 2. In Section V we investigate an special case of binary-inputs MAC. In Section VI we prove Theorem 2. In the last Section VII we summarize the paper with some comments.

II. ELEMENTARY MAC

In this section we introduce an elementary MAC with some expressions and definitions to be used in this paper.

An N-user MAC is specified by N input alphabets A_k with size of n_k, $k = 1, \cdots, N$, an output alphabet B with size of m, and an m by $(n_1 \times \cdots \times n_N)$ channel matrix $P = [P(j|i_1, \cdots, i_N)]$ of transmission probabilities $P(j|i_1, \cdots, i_N)$'s to be given *a priori* for the MAC, where $j=0, \cdots, m-1$, $i_k=0, \cdots, n_k - 1$, and $\sum_{j=0}^{m-1}$

$P(j|i_1, \cdots, i_N) = 1$. Assume that there is no zero row in P, $n_k \geq 2$, $m \geq 2$, and transmission is synchronized. The model thus defined is called an N-user MAC with a type $(n_1, \cdots, n_N; m)$.

The IPD vector $\boldsymbol{p}_k$ is assigned to an n_k-tuple column set of input probability $(p_k(0), \cdots, p_k(n_k - 1))^{\mathrm{T}}$, $k = 1, \cdots, N$, where $(\cdots)^{\mathrm{T}}$ implies a transposition, defined on A_k with the probability constraint $\sum_{i_k=0}^{n_k-1} p_k(i_k) = 1$. Thus each $\boldsymbol{p}_k$ is located on an $(n_k - 1)$-dimensional *simplex* X_k with n_k vertices $\boldsymbol{e}_{k\ell}$, $\ell = 0, \cdots, n_k - 1$. Here $\boldsymbol{e}_{k\ell}$ is a unit column vector and takes 1 in the ℓth column component and 0 elsewhere. Obviously each X_k is convex and is observed as *a domain* of $\boldsymbol{p}_k$.

The face F_k of X_k is defined by an $(f_k - 1)$-dimensional simplex whose f_k vertices are chosen from vertices $\boldsymbol{e}_{k0}, \cdots, \boldsymbol{e}_{k(n_k-1)}$ of X_k, where $f_k \leq n_k$. A set of f_k indices of F_k is denoted by $\Lambda(F_k) = \{\ell | \boldsymbol{e}_{k\ell} \in F_k\}$. There are several choices for f_k indices. Obviously $\Lambda(F_k) \subset \Lambda(X_k)$ and $\Lambda(X_k) = \{0, \cdots, n_k - 1\}$. Zero-dimensional faces are vertices, one-dimensional faces are lines, and so forth. If an IPD vector $\boldsymbol{p}_k$ is on the boundary of X_k, then there exists *a minimum* face F_k which contains $\boldsymbol{p}_k$ exactly inside (and not on the boundary of) F_k. Thus if $\boldsymbol{p}_k$ is $p_k(i_k) = 0$ for $i_k \notin \Lambda(F_k)$ and $p_k(i_k) > 0$ for $i_k \in \Lambda(F_k)$, then F_k for $\Lambda(F_k)$ is the minimum face which contains $\boldsymbol{p}_k$ and is uniquely determined. If $p_k(i_k) > 0$ for all i_k's, then the minimum face which contains the $\boldsymbol{p}_k$ is X_k itself. Here F_k is also referred to as *a sub-domain* of X_k. Also $\boldsymbol{p}_k$, with $p_k(i_k) > 0$ for $i_k \in \Lambda(F_k)$ and $p_k(i_k) = 0$ for $i_k \notin \Lambda(F_k)$, is naturally regarded as an $(f_k - 1)$-dimensional vector on F_k even it is still an $(n_k - 1)$-dimensional vector on the whole domain X_k.

The Kronecker product of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ is defined here by

$$\boldsymbol{p}_1 \times \boldsymbol{p}_2 \equiv \begin{bmatrix} p_1(0)\boldsymbol{p}_2 \\ \vdots \\ p_1(n_1 - 1)\boldsymbol{p}_2 \end{bmatrix}$$

and then the Kronecker product of $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$ is defined by induction: $\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_k \equiv (\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_{k-1}) \times \boldsymbol{p}_k$, $k = 3, \cdots, N$. In the same way we arrange the Kronecker product of $X_1, \cdots, X_N$ as

$$X \equiv X_1 \times \cdots \times X_N.$$

The set X is a domain of the IPD vector $\boldsymbol{p} = \boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N$ of the N-user MAC. Remark that X is not convex as a whole but each X_k is convex. Also we can set a Kronecker product of faces $F_1, \cdots, F_N$ as $F \equiv F_1 \times \cdots \times F_N$ which is a sub-domain of X. Obviously F is not convex as a whole even each F_k is convex.

A pair (P, X) is assigned to the N-user MAC to specify a channel matrix P and a domain X. Here P has columns $(P(0|i_1, \cdots, i_N), \cdots, P(m - 1|i_1, \cdots, i_N))^{\mathrm{T}}$ arranged in the order of the components of $\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N$. An MAC is denoted in more detail by N-user $(n_1, \cdots, n_N; m)$-MAC (P, X).

The mutual information of the N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is defined by

$$
\begin{aligned}
I(\boldsymbol{p}_1 &\times \cdots \times \boldsymbol{p}_N) \\
&= \sum_{j, i_1, \cdots, i_N} p_1(i_1) \cdots p_N(i_N) P(j|i_1, \cdots, i_N) \\
&\quad \cdot \log \frac{P(j|i_1, \cdots, i_N)}{q(j)}
\end{aligned} \tag{1}
$$

where $q(j) \equiv \sum_{i_1, \cdots, i_N} p_1(i_1) \cdots p_T(i_N) P(j|i_1, \cdots, i_N)$ is an output probability of the jth symbol of B and log is the natural logarithm. For any $\boldsymbol{p}', \boldsymbol{p}'' \in X$, a convex-linear combination $\lambda \boldsymbol{p}' + (1 - \lambda) \boldsymbol{p}''$, $0 \le \lambda \le 1$, does not always belong to X, since X is not convex except $N = 1$. Therefore, P is considered in general as *a non-linear* mapping from $\boldsymbol{p} \in X$ to $\boldsymbol{q} \equiv (q(0), \cdots, q(m-1))^{\mathrm{T}}$: $\boldsymbol{q} = P\boldsymbol{p} = P(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N)$. Also $I(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N)$ is regarded as a multi-variables function defined on the domain $X = X_1 \times \cdots \times X_N$ and is concave (convex-above) on each X_k, when $\boldsymbol{p}_\ell$'s, $\ell \ne k$, are fixed, but is not concave on the whole domain X.

The channel capacity of the N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is defined as usual by the maximum value of the mutual information (1):

$$
C = \max_{\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N \in X} I(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N). \tag{2}
$$

An IPD vector which achieves the channel capacity is referred to as *an optimal* IPD vector.

The Kuhn-Tucker conditions are introduced as the conditions to obtain the local extrema of a function of several variables subject to one or more constraints. For the mutual information (1) of the N-user $(n_1, \cdots, n_N; m)$-MAC (P, X), the conditions to take the maximum value (channel capacity) are stated as follows: If $\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N$ is optimal, then it satisfies

$$
\begin{aligned}
&J(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N; i_k) \begin{cases} = C, & p_k(i_k) > 0 \\ \le C, & p_k(i_k) = 0 \end{cases} \\
&i_k = 0, \cdots, n_k - 1, \quad k = 1, \cdots, N \\
&C = I(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N)
\end{aligned} \tag{3}
$$

where

$$
J(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N; i_k) \equiv \frac{\partial I(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N)}{\partial p_k(i_k)} + 1
$$

$$
\begin{aligned}
= \sum_{j, i_1, \cdots, i_{k-1}, i_{k+1} \cdots, i_N} & p_1(i_1) \cdots p_{k-1}(i_{k-1}) p_{k+1}(i_{k+1}) \\
& \cdots p_N(i_N) P(j|i_1, \cdots, i_N) \log \frac{P(j|i_1, \cdots, i_N)}{q(j)}.
\end{aligned}
$$

These equations (3) are collectively referred to as *the Kuhn-Tucker conditions* for the mutual information (1). These are quite easy to obtain, for example, by a method of Lagrange multipliers to maximize the mutual information (1) subject to the constraints of $\boldsymbol{p}_k$: $\sum_{i_k=0}^{n_k-1} p_k(i_k) = 1$, $k = 1, \cdots, N$. Remark that the Kuhn-Tucker conditions (3) are obviously necessary but not in general sufficient for the channel capacity of the MAC (P, X). In the case of DMC, however, they are necessary and sufficient for the channel capacity [20].

A sub-MAC (P, Y), or a sub-channel, of an N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is reasonably defined as an N-user MAC where the channel matrix is set to the same P and the domain is assigned to a non-empty subset Y of X. In the subsequent discussions we focus mostly on the sub-MAC (P, Y) where Y is restricted to a sub-domain $F \equiv F_1 \times \cdots \times F_N \subseteq X_k$. Here if $\boldsymbol{p}$ is an IPD vector of the sub-MAC (P, F), then each $\boldsymbol{p}_k$ of $\boldsymbol{p}$ acts as an $(f_k - 1)$-dimensional vector on F_k (i.e., $p_k(i_k) = 0$ for $i_k \notin \Lambda(F_k)$) even it is still an $(n_k - 1)$-dimensional vector on the whole domain X_k as mentioned before. The mutual information of an N-user sub-MAC (P, F) is given by $I(\boldsymbol{p} \in F)$ where the $(i_1, \cdots, i_k, \cdots, i_N)$th columns of P for $i_k \notin \Lambda(F_k)$ do not affect the mutual information (1). The Kuhn-Tucker conditions for an N-user sub-MAC (P, F) are also given by the expression (3) where $\boldsymbol{p} \in F$.

The elementary MAC now we define in general as follows: If N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) satisfies $n_k \le m$ for all $k = 1, \cdots, N$, then it is referred to as *an elementary* MAC. The elementary MAC is an MAC whose sizes of input alphabets are not greater than the size of output alphabet.

The elementary (face) set $\Phi_N^{(m)}$ of X is defined by the set of faces as follows: If $n_k \ge m$, then F_k is put to an $(m-1)$-dimensional face of X_k, and if $n_k < m$, then F_k is put to the $(n_k - 1)$-dimensional X_k itself. Thus the dimension of each F_k is less than or equal to $(m-1)$. If X is formed by $n_k \le m$ for all $k = 1, \cdots, N$, then $\Phi_N^{(m)} = \{X\}$.

A master (elementary) MAC (P, F) of an N-user MAC (P, X) is defined as the MAC with a domain $F \in \Phi_N^{(m)}$. Here each $\boldsymbol{p}_k \in F_k$ acts as an $(f_k - 1)$-dimensional vector as mentioned above, where $p_k(i_k) = 0$ for $i_k \notin \Lambda(F_k)$, $k = 1, \cdots, N$. Note that the master MAC $(P, F \in \Phi_N^{(m)})$ is regarded as *the largest* possible elementary MAC of (P, X) in the sense that there is no elementary MAC (P, F') such that (P, F) is an elementary sub-MAC of (P, F'). A set of all master MAC's is referred to as the master (elementary) set of the N-user MAC (P, X) and is denoted by Ω_N. Obviously Ω_N is finite and is uniquely determined. If an MAC (P, X) is itself elementary, then $\Omega_N = \{\mathrm{MAC}(P, X)\}$. The channel capacity of an MAC $(P, F) \in \Omega_N$ is denoted by $C(F)$.

In later discussions we investigate the IPD vector $\boldsymbol{p}$ which satisfies the Kuhn-Tucker conditions (3). If $p_k(i_k) > 0$ for all $i_k = 0, \cdots, n_k - 1$, $k = 1, \cdots, N$, then $\boldsymbol{p}$ is located exactly inside (not on the boundary of) X. If $p_k(i_k) = 0$ for $i_k \notin \Lambda(F_k)$ and $p_k(i_k) > 0$ for $i_k \in \Lambda(F_k)$, $k = 1, \cdots, N$, then the sub-domain $F = F_1 \times \cdots \times F_N$ of X formed by $\Lambda(F_k)$ is the minimum domain which contains $\boldsymbol{p}$ exactly inside F.

More importantly, the non-elementary MAC has in essence a *degenerate* property as follows: if $n_k > m$ for an N-user $(n_1, \cdots, n_N; m)$-MAC, then for a fixed $\boldsymbol{p}_k \in F_k \subseteq X_k$ with $f_k > m$, there exists an IPD vector $\boldsymbol{p}'_k \in F'_k$ where $\boldsymbol{p}' \ne \boldsymbol{p}$, $F'_k \subset (\ne) F_k$, $f'_k = m$, and $\boldsymbol{p} = \boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \boldsymbol{p}_N$, $\boldsymbol{p}' = \boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}'_k \times \cdots \times \boldsymbol{p}_N$, such that $\boldsymbol{q} = P\boldsymbol{p}' = P\boldsymbol{p}$. The elementary MAC has in general no such property. This notion is crucial to the subsequent discussions.

Finally for this section, we remark that we are going to investigate various types of MAC's. For example, we examine an MAC (P, Y) with a domain $Y = Y_1 \times \cdots \times Y_N \subset X$ where each Y_k, $k = 1, \cdots N$, is formed by *the line* segment of IPD vectors of X_k. Even then we can examine the Kuhn-Tucker

conditions in the same way as mentioned above.

III. Main Result

The master elementary set Ω_N as defined above has an intrinsic property with respect to the N-user $(n_1, \cdots, n_N; m)$-MAC (P, X). We can state it as a main theorem:

Theorem 1: The channel capacity C of an N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is achieved by the channel capacity $C(F)$ of an N-user elementary MAC $(P, F \in \Phi_N^{(m)})$ of Ω_N as follows:

$$C = \max_{F \in \Phi_N^{(m)}} C(F). \tag{4}$$

$\square$

Proof: It is sufficient to prove the case that the original MAC (P, X) is not elementary. Let $\bar{\boldsymbol{p}} = \bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_k \times \cdots \times \bar{\boldsymbol{p}}_N$ be an optimal IPD vector that achieves the channel capacity C. Let $\bar{F}_k$ be the minimum face of X_k which contains $\bar{\boldsymbol{p}}_k$ exactly inside $\bar{F}_k$, $k = 1, \cdots N$. It is sufficient to assume that $\bar{F}_k$ is the m or more dimensional face. Then by the degenerate property there exists an $(m-1)$-dimensional face $\tilde{F}_k \subset (\neq) \bar{F}_k$ such that for an IPD vector $\tilde{\boldsymbol{p}}_k \in \tilde{F}_k$,

$$P\bar{\boldsymbol{p}} = P(\bar{\boldsymbol{p}}_1 \times \cdots \times \tilde{\boldsymbol{p}}_k \times \cdots \times \bar{\boldsymbol{p}}_N). \tag{5}$$

Put

$$K(\theta) \equiv I(\bar{\boldsymbol{p}}_1 \times \cdots \times (\theta \bar{\boldsymbol{p}}_k + (1-\theta)\tilde{\boldsymbol{p}}_k) \times \cdots \times \bar{\boldsymbol{p}}_N)$$

for the mutual information of the original MAC (P, X), where $0 \le \theta \le 1$. The derivative $\partial K(\theta)/\partial \theta$ is constant by (5) and moreover $\partial K(\theta)/\partial \theta$ is equal to zero since $\bar{\boldsymbol{p}}$ is optimal. Then it holds

$$I(\bar{\boldsymbol{p}}) = I(\bar{\boldsymbol{p}}_1 \times \cdots \times \tilde{\boldsymbol{p}}_k \times \cdots \times \bar{\boldsymbol{p}}_N).$$

This implies that the optimal IPD vector exists in a domain $F = F_1 \times \cdots \times \tilde{F}_k \times \cdots \times F_N \in \Phi_N^{(m)}$. Thus Theorem 1 is proved. $\blacksquare$

Theorem 1 states that the channel capacity C of any N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is rigorously determined by the channel capacity $C(F)$ of an N-user master elementary MAC $(P, F) \in \Omega_N$. In other words, an optimal IPD vector exists at least on a domain $F \in \Phi_N^{(m)}$. However Theorem 1 does not guarantee that the optimal IPD vector exists only on a domain $F \in \Phi_N^{(m)}$, that is, there *might* exist in general an optimal IPD vector that is located exactly inside X and not on any $F \in \Phi_N^{(m)}$. Note that if the N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is elementary, then Theorem 1 appears self-evident since Ω_N contains only an MAC (P, X) itself.

In the remaining section of this paper we focus on the proof that the Kuhn-Tucker conditions of an elementary MAC (P, X) are necessary and sufficient for the channel capacity. We will state it in advance as a second theorem.

Theorem 2: The Kuhn-Tucker conditions for the channel capacity C of an N-user elementary $(n_1, \cdots, n_N; m)$-MAC (P, X), where $n_k \le m$ for all $k = 1, \cdots, N$, are necessary and sufficient. $\square$

It is sufficient to prove only the sufficiency since the necessity is self-evident. From these two theorems the MAC in general can be regarded as simply an aggregate of a finite number of elementary MAC's where the Kuhn-Tucker conditions for the channel capacity are necessary and sufficient.

IV. Features of Elementary MAC

In this section we prepare basic properties that are required to prove the sufficiency of Theorem 2.

The first property *A* is *the chain rules* [1]: We recall that the mutual information of an N-user MAC is in general decomposed into N components with $N!$ different decompositions by the chain rules.

The second property *B* is *the capacity region*: We describe that the capacity region of the N-user MAC is given by the convex-closure of all achievable rate regions of the $N!$ decompositions for the mutual information [4]. It is summarized as Proposition 1.

The third property *C* is *the boundary equations*: We investigate that a boundary of an achievable rate region satisfies by a method of Lagrange multipliers a set of conditions to be referred to as the boundary equations for the capacity region of the N-user MAC.

The fourth property *D* is *a relation between the Kuhn-Tucker equations and the boundary equations*: We prove as Proposition 2 that a solution of the Kuhn-Tucker conditions of an N-user MAC with some restrictions satisfies the boundary equations.

The fifth property *E* is *local maximum*: We prove as Proposition 3 that every solution of the Kuhn-Tucker conditions of an elementary MAC (P, X) is local maximum in the domain X. To prove Proposition 3 we need Proposition 2.

Finally, the sixth property *F* is *connectedness*: We prove as Proposition 4 that a set of IPD vectors of an elementary MAC (P, X), for which the value of the mutual information is not smaller than the arbitrary positive number, is connected in the domain X. To prove Proposition 4 we use Proposition 3.

We emphasize here that the last two properties, i.e. *local maximum* and *connectedness*, are the most distinctive features exclusive to the elementary MAC. However the first four properties, although they hold for any MAC in general, are required to step by step prove the last two.

A. Chain Rules

The mutual information of an N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is decomposed into N components by the chain rules [1]. For the IPD vectors $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_{k-1}, \boldsymbol{p}_k, \boldsymbol{p}_{k+1}, \cdots, \boldsymbol{p}_N$, let $\boldsymbol{\rho}_{\{u, \cdots, w\}}$ be a Kronecker product of $\boldsymbol{p}_k$, $k \notin \{u, \cdots, w\}$, and let $\boldsymbol{\sigma}_{\{u, \cdots, w\}}$ be a Kronecker product of $\boldsymbol{p}_k$, $k \in \{u, \cdots, w\}$. Obviously $\boldsymbol{\sigma}_{\{u\}} = \boldsymbol{p}_u$.

The mutual information (1) is decomposed into two components as

$$I(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N) = $$
$$I(\boldsymbol{\sigma}_{\{u\}} | \boldsymbol{\rho}_{\{u\}}) + I(\boldsymbol{\rho}_{\{u\}} / \boldsymbol{\sigma}_{\{u\}}).$$

Here

$$I(\boldsymbol{\sigma}_{\{u\}}|\boldsymbol{\rho}_{\{u\}}) =$$
$$\sum_{j,i_1,\cdots,i_N} p_1(i_1)\cdots p_N(i_N)P(j|i_1,\cdots,i_N)$$
$$\cdot\log\frac{P(j|i_1,\cdots,i_N)}{\sum_h p_u(h)P(j|i_1,\cdots,h,\cdots,i_N)}$$

which is the conditional mutual information of $\boldsymbol{p}_u$ with respect to $\boldsymbol{p}_1,\cdots,\boldsymbol{p}_{u-1},\boldsymbol{p}_{u+1}\cdots,\boldsymbol{p}_N$, and

$$I(\boldsymbol{\rho}_{\{u\}}/\boldsymbol{\sigma}_{\{u\}}) =$$
$$\sum_{j,i_1,\cdots,i_N} p_1(i_1)\cdots p_N(i_N)P(j|i_1,\cdots,i_N)$$
$$\cdot\log\frac{\sum_h p_u(h)P(j|i_1,\cdots,h,\cdots,i_N)}{q(j)}$$

which is the mutual information of an $(N-1)$-user MAC with the channel matrix $[\sum_h p_u(h)P(j|i_1,\cdots,h,\cdots,i_N)]$. Moreover we decompose the latter into

$$I(\boldsymbol{\rho}_{\{u\}}/\boldsymbol{\sigma}_{\{u\}}) =$$
$$I(\boldsymbol{p}_w|\boldsymbol{\rho}_{\{u,w\}}/\boldsymbol{\sigma}_{\{u\}}) + I(\boldsymbol{\rho}_{\{u,w\}}/\boldsymbol{\sigma}_{\{u,w\}}).$$

In general,

$$I(\boldsymbol{\rho}_{\{u,\cdots,w\}}/\boldsymbol{\sigma}_{\{u,\cdots,w\}}) =$$
$$I(\boldsymbol{p}_x|\boldsymbol{\rho}_{\{u,\cdots,x,\cdots,w\}}/\boldsymbol{\sigma}_{\{u,\cdots,w\}})$$
$$+I(\boldsymbol{\rho}_{\{u,\cdots,x,\cdots,w\}}/\boldsymbol{\sigma}_{\{u,\cdots,x\cdots,w\}}).$$

Here

$$I(\boldsymbol{p}_x|\boldsymbol{\rho}_{\{u,\cdots,x,\cdots,w\}}/\boldsymbol{\sigma}_{\{u,\cdots,w\}}) =$$
$$\sum_{j,i_1,\cdots,i_N} p_1(i_1)\cdots p_T(i_N)P(j|i_1,\cdots,i_N)$$
$$\cdot\log\frac{\langle\boldsymbol{\sigma}_{\{u,\cdots,w\}}\cdot P\rangle}{\langle\boldsymbol{\sigma}_{\{u,\cdots,x,\cdots,w\}}\cdot P\rangle}$$

$$I(\boldsymbol{\rho}_{\{u,\cdots,x,\cdots,w\}}/\boldsymbol{\sigma}_{\{u,\cdots,x\cdots,w\}}) =$$
$$\sum_{j,i_1,\cdots,i_N} p_1(i_1)\cdots p_N(i_N)P(j|i_1,\cdots,i_N)$$
$$\cdot\log\frac{\langle\boldsymbol{\sigma}_{\{u,\cdots,x,\cdots,w\}}\cdot P\rangle}{q(j)}$$

where

$$\langle\boldsymbol{\sigma}_{\{u,\cdots,w\}}\cdot P\rangle \equiv \sum_{h_u,\cdots,h_w} p_u(h_u)\cdots p_w(h_w)$$
$$\cdot P(j|i_1,\cdots,h_u,\cdots,h_w,\cdots,i_N).$$

If $\{u,\cdots,w\}$ is empty, then $\langle\boldsymbol{\sigma}_{\{u,\cdots,w\}}\cdot P\rangle$ reduces to $P(j|i_1,\cdots,i_N)$. Thus successively reducing the suffices $\{u,\cdots,w\}$ of $\boldsymbol{\rho}_{\{u,\cdots,w\}}$ up to $\{1,\cdots,k-1,k+1,\cdots,N\}$, $I(\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_N)$ is decomposed into N components. Note that there exist as a whole $N!$ different decompositions.

B. Capacity Region

A set of all achievable rates for an N-user $(n_1,\cdots,n_N;m)$-MAC (P,X) is called a capacity region (e.g., [1], [3], [4]). By a decomposition we obtain

$$I(\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_N) =$$
$$I(\boldsymbol{p}_1|\boldsymbol{p}_2\times\cdots\times\boldsymbol{p}_N) + I(\boldsymbol{p}_2|\boldsymbol{p}_3\times\cdots\times\boldsymbol{p}_N/\boldsymbol{p}_1)$$
$$+I(\boldsymbol{p}_3|\boldsymbol{p}_4\times\cdots\times\boldsymbol{p}_N/\boldsymbol{p}_1\times\boldsymbol{p}_2)+\cdots$$
$$\cdots+I(\boldsymbol{p}_N/\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_{N-1}). \qquad (6)$$

There exist as a whole $N!$ different decompositions as mentioned above. Define a sub-region G_1 as

$$\bigcup_{\boldsymbol{p}\in X}(I(\boldsymbol{p}_1|\boldsymbol{p}_2\times\cdots\times\boldsymbol{p}_N),\cdots,I(\boldsymbol{p}_N/\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_{N-1})).$$

This is identified as a set of achievable rates G_1 for the decomposition (6). Other $N!-1$ sets of achievable rates G_2, $\cdots$, $G_{N!}$ are also defined in the same way as G_1. Then the capacity region G is determined by those sub-regions G_i's as follows [Theorem 15.3.6 in [1]]:

Proposition 1: The capacity region of an N-user $(n_1,\cdots,n_N;m)$-MAC (P,X) is given by

$$G = \text{co}\bigcup_{i=1}^{N!} G_i \qquad (7)$$

where "co" implies the convex-closure. $\qquad\Box$

C. Boundary Equations

A boundary of each sub-region G_i, $i=1,\cdots,N!$, for an N-user $(n_1,\cdots,n_N;m)$-MAC (P,X), can be determined by a method of Lagrange multipliers. The boundary of G_1, for example, is evaluated by a Lagrange multiplier function,

$$L(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_N,\lambda_1,\cdots,\lambda_{N-1},\zeta_1,\cdots,\zeta_N) =$$
$$I(\boldsymbol{p}_1|\boldsymbol{p}_2\times\cdots\times\boldsymbol{p}_N)$$
$$-\lambda_1 I(\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_N)$$
$$-\lambda_2 I(\boldsymbol{p}_2|\boldsymbol{p}_3\times\cdots\times\boldsymbol{p}_N/\boldsymbol{p}_1)-\cdots$$
$$-\lambda_{N-1}I(\boldsymbol{p}_{N-1}|\boldsymbol{p}_N/\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_{N-2})$$
$$-\sum_{k=1}^{N}\zeta_k\sum_{i_k}p_k(i_k)$$

where $\lambda_1,\cdots,\lambda_{N-1}$ and $\zeta_1,\cdots,\zeta_N$ are so-called Lagrange multipliers. The conditions that an IPD vector $\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_N$ takes extremum (maximum or minimum) for G_1 are given by the equations (see Fig. 1 for $N=3$)

$$\det\begin{bmatrix}\dfrac{\tilde{\partial}I(\boldsymbol{p}_1|\boldsymbol{p}_2\times\cdots\times\boldsymbol{p}_N)}{\tilde{\partial}p_1(i_1)} & \dfrac{\tilde{\partial}I(\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_N)}{\tilde{\partial}p_1(i_1)} \\ \vdots & \vdots \\ \dfrac{\tilde{\partial}I(\boldsymbol{p}_1|\boldsymbol{p}_2\times\cdots\times\boldsymbol{p}_N)}{\tilde{\partial}p_N(i_N)} & \dfrac{\tilde{\partial}I(\boldsymbol{p}_1\times\cdots\times\boldsymbol{p}_N)}{\tilde{\partial}p_N(i_N)} \\ \dfrac{\tilde{\partial}I(\boldsymbol{p}_2|\boldsymbol{p}_3\times\cdots\times\boldsymbol{p}_N/\boldsymbol{p}_1)}{\tilde{\partial}p_1(i_1)} & \cdots \\ \vdots & \vdots \\ \dfrac{\tilde{\partial}I(\boldsymbol{p}_2|\boldsymbol{p}_3\times\cdots\times\boldsymbol{p}_N/\boldsymbol{p}_1)}{\tilde{\partial}p_N(i_N)} & \cdots \end{bmatrix}$$
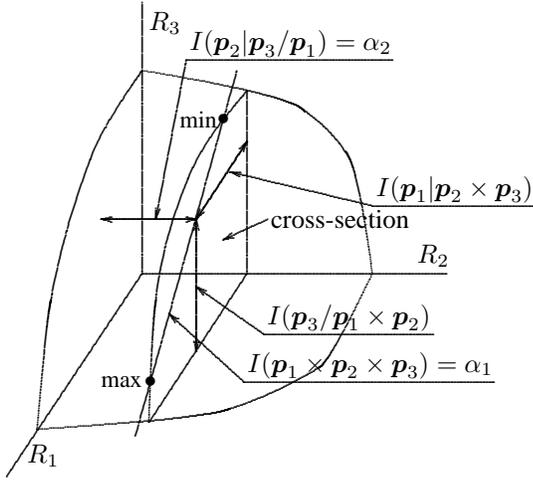
Fig. 1. Sub-region G_1 of three-user MAC.

$$\cdots \quad \frac{\tilde{\partial} I(\boldsymbol{p}_{N-1}|\boldsymbol{p}_N/\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_{N-2})}{\tilde{\partial} p_1(i_1)}$$
$$\vdots \qquad\qquad \vdots$$
$$\cdots \quad \frac{\tilde{\partial} I(\boldsymbol{p}_{N-1}|\boldsymbol{p}_N/\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_{N-2})}{\tilde{\partial} p_N(i_N)} \Bigg]$$
$$= 0, \quad i_k = 0, \cdots, n_k - 2, k = 1, \cdots, N. \qquad (8)$$

Here, we define partial derivatives as:

$$\frac{\tilde{\partial} I(\cdots)}{\tilde{\partial} p_k(i_k)} \equiv \frac{\partial I(\cdots)}{\partial p_k(i_k)} - \frac{\partial I(\cdots)}{\partial p_k(n_k - 1)}, i_k \neq n_k - 1.$$

Total $(n_1 - 1) \times \cdots \times (n_N - 1)$ equations (8) are collectively referred to as *the boundary equations* for G_1. Solutions of (8) include both maximization and minimization as usual. Successively we can set up the boundary equations for $G_2, \cdots, G_{N!}$ with totally the same form as (8). Note that the boundary equations have the same form as (8) for the different choices of starting Lagrange multiplier function.

D. A relation between the Kuhn-Tucker equations and the boundary equations

The boundary equations thus obtained have an important property which we state as a proposition:

Proposition 2: If a solution $\bar{\boldsymbol{p}} = \bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_N \in X$ of the Kuhn-Tucker conditions for the mutual information $I(\boldsymbol{p})$ of an N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) satisfies

$$J(\bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_N; i_k) = C,$$
$$i_k = 0, \cdots, n_k - 1, k = 1, \cdots, N$$
$$C = I(\bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_N) \qquad (9)$$

then $\bar{\boldsymbol{p}}$ is a solution of the boundary equations for sub-regions G_i, $i = 1, \cdots, N!$. □

Proof: It is sufficient to prove that $\bar{\boldsymbol{p}}$ satisfies the boundary equation (8) for G_1. By the assumption (9), it holds

$$\frac{\tilde{\partial} I(\boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N)}{\tilde{\partial} p_k(i_k)} \Bigg|_{\boldsymbol{p}_k = \bar{\boldsymbol{p}}_k, i_k = 0, \cdots, n_k - 2} =$$
$$J(\bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_N; i_k) - J(\bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_N; n_k - 1)$$
$$= 0.$$

Then the second column of (8) reduces to zeros. Therefore $\bar{\boldsymbol{p}}$ is a solution of the boundary equation (8). ■

Remark that Proposition 2 holds for any MAC including the elementary MAC if it satisfies the conditions (9).

E. Local Maximum

An IPD vector $\bar{\boldsymbol{p}}$ is called *a local maximum point* for the mutual information $I(\boldsymbol{p})$, if there exists a neighborhood $U_{\bar{\boldsymbol{p}}}$ of $\bar{\boldsymbol{p}}$ such that $I(\boldsymbol{p}) \leq I(\bar{\boldsymbol{p}})$ for any $\boldsymbol{p} \in U_{\bar{\boldsymbol{p}}}$. We prove here that for *the elementary* MAC every solution of the Kuhn-Tucker conditions is local maximum. We state it as a proposition:

Proposition 3: If an N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is elementary, i.e. $n_k \leq m$, $k = 1, \cdots, N$, then every solution $\boldsymbol{p}^* \equiv \boldsymbol{p}_1^* \times \cdots \times \boldsymbol{p}_N^* \in X$ of the Kuhn-Tucker conditions for the mutual information $I(\boldsymbol{p})$ is local maximum in X. □

Before proceeding we remark that Proposition 3 does not hold in general for the non-elementary MAC by the degenerate property as is stated in the beginning of the proof of Theorem 1. In fact, we note without proof that a non-elementary two-user $(3, 3; 2)$-MAC (P, X), for example, with $N = 2$, $n_1 = n_2 = 3$, $m = 2$, for some channel matrix P, has a solution of the Kuhn-Tucker conditions which is not local maximum in X.

Proof: Since the N-user MAC (P, X) is elementary, it is sufficient to investigate two cases for the solution $\boldsymbol{p}^*$ of the Kuhn-Tucker conditions: the first is that every $p_k^*(i_k)$ is non-zero and the second is that at least one of $p_k^*(i_k)$'s is zero.

In the first case, since $p_k^*(i_k) > 0$ for all components, $\boldsymbol{p}^*$ satisfies the Kuhn-Tucker conditions:

$$J(\boldsymbol{p}_1^* \times \cdots \times \boldsymbol{p}_N^*; i_k) = M_T$$
$$p_k^*(i_k) > 0, \quad i_k = 0, \cdots, n_k - 1, \quad k = 1, \cdots, N$$
$$M_T = I(\boldsymbol{p}_1^* \times \cdots \times \boldsymbol{p}_N^*). \qquad (10)$$

and there exist $\boldsymbol{p}'$ and $\boldsymbol{p}''$ in X such that

$$\boldsymbol{p}^* = (\theta_1^* \boldsymbol{p}_1'' + (1 - \theta_1^*)\boldsymbol{p}_1') \times \cdots \times (\theta_N^* \boldsymbol{p}_N'' + (1 - \theta_N^*)\boldsymbol{p}_N')$$

where $\boldsymbol{p}_k^* \neq \boldsymbol{p}_k'$, $\boldsymbol{p}_k^* \neq \boldsymbol{p}_k''$, and $0 < \theta_k^* < 1$.

Here we put by using θ_k, $0 \leq \theta_k \leq 1$, $k = 1, \cdots, N$,

$$\mathcal{K}(\theta_1, \cdots, \theta_N) =$$
$$I((\theta_1 \boldsymbol{p}_1'' + (1 - \theta_1)\boldsymbol{p}_1') \times \cdots \times (\theta_N \boldsymbol{p}_T'' + (1 - \theta_N)\boldsymbol{p}_N'))$$

and investigate the Kuhn-Tucker conditions and the boundary equations with respect to this $\mathcal{K}(\theta_1, \cdots, \theta_N)$. The Kuhn-Tucker conditions are simple to see as

$$\mathcal{K}_k(\theta_1, \cdots, \theta_N) = 0, \quad k = 1, \cdots, N \qquad (11)$$

where $\mathcal{K}_k(\theta_1, \cdots, \theta_N) \equiv \partial \mathcal{K}(\theta_1, \cdots, \theta_N)/\partial \theta_k$. Also by a decomposition

$$\mathcal{K}(\theta_1, \cdots, \theta_N) =$$
$$\mathcal{K}(\theta_1|\theta_2, \cdots, \theta_N) + \mathcal{K}(\theta_2|\theta_3, \cdots, \theta_N/\theta_1)$$
$$+ \mathcal{K}(\theta_3|\theta_4 \cdots, \theta_N/\theta_1, \theta_2) + \cdots$$
$$+ \mathcal{K}(\theta_N/\theta_1, \cdots, \theta_{N-1})$$
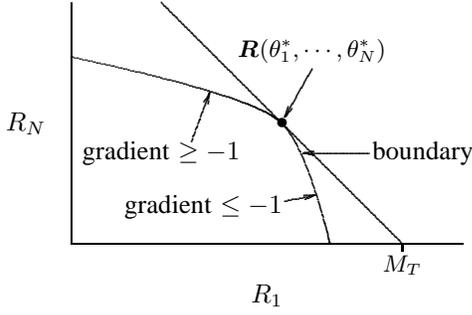
Fig. 2. Boundary of cross-section $\mathcal{G}_1(R_1, R_N)$ in R_1-R_N plain. Rates $R_2, \cdots, R_{N-1}$ are fixed as specified by (14).

we obtain a set of achievable rates

$$\mathcal{G}_1 = \bigcup_{\theta_1, \cdots, \theta_N} (\mathcal{K}(\theta_1|\theta_2, \cdots, \theta_N), \cdots, \mathcal{K}(\theta_N/\theta_1, \cdots, \theta_{N-1})) \tag{12}$$

which leads us to the boundary equation for $\mathcal{G}_1$ as follows:

$$\det \begin{bmatrix} \mathcal{K}_1(\theta_1|\theta_2, \cdots, \theta_N) & \mathcal{K}_1(\theta_1, \cdots, \theta_N) \\ \vdots & \vdots \\ \mathcal{K}_N(\theta_1|\theta_2, \cdots, \theta_N) & \mathcal{K}_N(\theta_1, \cdots, \theta_N) \end{bmatrix}$$

$$\begin{matrix} \mathcal{K}_1(\theta_2|\theta_3, \cdots, \theta_N/\theta_1) & \cdots \\ \vdots & \vdots \\ \mathcal{K}_N(\theta_2|\theta_3, \cdots, \theta_N/\theta_1) & \cdots \end{matrix}$$

$$\left. \begin{matrix} \mathcal{K}_1(\theta_{N-1}|\theta_N/\theta_1, \cdots, \theta_{N-2}) \\ \vdots \\ \mathcal{K}_N(\theta_{N-1}|\theta_N/\theta_1, \cdots, \theta_{N-2}) \end{matrix} \right] = 0 \tag{13}$$

where $\mathcal{K}_k(\cdots) \equiv \partial\mathcal{K}(\cdots)/\partial\theta_k$.

Since $\boldsymbol{p}^*$ satisfies (10), $(\theta_1^*, \cdots, \theta_N^*)$ is a solution of the Kuhn-Tucker conditions (11). Then by Proposition 2 it satisfies the boundary equation (13) and $M_T = \mathcal{K}(\theta_1^*, \cdots, \theta_N^*)$.

Now we examine a gradient of the boundary of $\mathcal{G}_1$ at $\theta^* \equiv (\theta_1^*, \cdots, \theta_N^*)$. Note that the solution $\theta \equiv (\theta_1, \cdots, \theta_N)$ of the boundary equation (13) around θ^* defines a set of achievable rates (12) as $\mathcal{G}_1(R_1, \cdots, R_N)$. Obviously $\mathcal{K}(\theta^*) = M_T$. At this step we investigate a cross-section of (12) subject to the restrictions such that

$$\mathcal{K}(\theta_2|\theta_3, \cdots, \theta_N/\theta_1) = \mathcal{K}(\theta_2^*|\theta_3^*, \cdots, \theta_N^*/\theta_1^*)$$
$$\vdots$$
$$\mathcal{K}(\theta_{N-1}|\theta_N/\theta_1, \cdots, \theta_{N-2}) = \mathcal{K}(\theta_{N-1}^*|\theta_N^*/\theta_1^*, \cdots, \theta_{N-2}^*). \tag{14}$$

We denote a cross-section (subset) of $\mathcal{G}_1$ subject to (14) as $\mathcal{G}_1(R_1, R_N)$. This is composed of

$$\boldsymbol{R}(\theta_1, \cdots, \theta_N) \equiv$$
$$(\mathcal{K}(\theta_1|\theta_2, \cdots, \theta_N), \mathcal{K}(\theta_2^*|\theta_3^*, \cdots, \theta_N^*/\theta_1^*), \cdots,$$
$$\mathcal{K}(\theta_{N-1}^*|\theta_{N-1}^*/\theta_1^*, \cdots, \theta_{N-2}^*), \mathcal{K}(\theta_N/\theta_1, \cdots, \theta_{N-1})).$$

The cross-section $\mathcal{G}_1(R_1, R_N)$ is a region in the two-dimensional $(R_1$-$R_N)$ plain as shown in Fig. 2. Since it holds

$$\mathcal{K}_1(\theta_2|\theta_3, \cdots, \theta_N/\theta_1) = 0$$
$$\vdots$$
$$\mathcal{K}_1(\theta_{N-1}|\theta_N/\theta_1, \cdots, \theta_{N-2}) = 0$$

by the restrictions (14), then we have $\mathcal{K}_1(\theta_1, \cdots, \theta_N) = \mathcal{K}_1(\theta_1|\theta_2, \cdots, \theta_{N-1}) + \mathcal{K}_1(\theta_N/\theta_1, \cdots, \theta_{N-1})$. Thus the gradient of the boundary of $\mathcal{G}_1(R_1, R_N)$ appears

$$\frac{\mathcal{K}_1(\theta_N/\theta_1, \cdots, \theta_{N-1})}{\mathcal{K}_1(\theta_1|\theta_2, \cdots, \theta_N)} = -1 + \frac{\mathcal{K}_1(\theta_1, \cdots, \theta_N)}{\mathcal{K}_1(\theta_1|\theta_2, \cdots, \theta_N)}. \tag{15}$$

The right-hand side of (15) is estimated as

$$-1 + \frac{\mathcal{K}_1(\theta_1, \cdots, \theta_N)}{\mathcal{K}_1(\theta_1|\theta_2, \cdots, \theta_N)} \le (\ge)1 \tag{16}$$

according to the maximization (minimization) conditions of $\mathcal{K}(\theta_1|\theta_2, \cdots, \theta_N)$ subject to (14) where it holds

$$\mathcal{K}_1(\theta_1, \cdots, \theta_N)\mathcal{K}_1(\theta_1|\theta_2, \cdots, \theta_N) \le (\ge) 0.$$

Also the gradient of the boundaries of any cross-section $\mathcal{G}_1(R_1, R_k)$ $(2 \le k \le N - 1)$ is given by (16).

For any region $\mathcal{G}_i$ of $N!$ decompositions, the gradient of the boundary of $\mathcal{G}_i$ at θ^* takes the same condition as that of $\mathcal{G}_1$.

Since the inequalities (16) are valid for any $\boldsymbol{p}_k'$, $\boldsymbol{p}_k''$, $k = 1, \cdots, N$, there exists a neighborhood $U_{\boldsymbol{p}^*}$ of $\boldsymbol{p}^*$ in X, such that $I(\boldsymbol{p}^*) \ge I(\boldsymbol{p} \in U_{\boldsymbol{p}^*})$. This means that $\boldsymbol{p}^*$ is local maximum in X.

In the second case, since at least one of $p_k^*(i_k)$'s is zero, $\boldsymbol{p}^*$ satisfies the Kuhn-Tucker conditions:

$$J(\boldsymbol{p}_1^* \times \cdots \times \boldsymbol{p}_N^*; i_k) \begin{cases} = M_T, & i_k \in \Lambda(F_k) \\ \le M_T, & i_k \notin \Lambda(F_k) \end{cases} \tag{17}$$
$$k = 1, \cdots, N, \quad M_T = I(\boldsymbol{p}_1^* \times \cdots \times \boldsymbol{p}_N^*)$$

where $p_k^*(i_k) > 0$ for $i_k \in \Lambda(F_k)$ and $p_k^*(i_k) = 0$ for $i_k \notin \Lambda(F_k)$. Thus there exists a sub-domain $F = F_1 \times \cdots \times F_N$ such that $\boldsymbol{p}^* \in F$. This implies that $\boldsymbol{p}^*$ is local maximum in F as described in the first case and there exists a neighborhood $U_{0\boldsymbol{p}^*} \subset F$ such that $I(\boldsymbol{p}^*) \ge I(\boldsymbol{p} \in U_{0\boldsymbol{p}^*})$.

For any $\boldsymbol{p}' = \boldsymbol{p}_1' \times \cdots \times \boldsymbol{p}_k' \times \cdots \times \boldsymbol{p}_N' \in U_{0\boldsymbol{p}^*}$, consider $\boldsymbol{p}'' = \boldsymbol{p}_1' \times \cdots \times \boldsymbol{p}_k'' \times \cdots \times \boldsymbol{p}_N'$, where $\boldsymbol{p}_k'' \in X_k$ and $\boldsymbol{p}_k'' \notin F_k$. Put for $0 \le \theta \le 1$

$$K(\theta) = I(\boldsymbol{p}_1' \times \cdots \times (\theta\boldsymbol{p}_k'' + (1 - \theta)\boldsymbol{p}_k') \times \cdots \times \boldsymbol{p}_N').$$

It holds $dK(\theta)/d\theta \mid_{\theta=0} \le 0$, since $K(\theta)$ is concave, differentiable, and $\boldsymbol{p}^*$ satisfies (17). Therefore $K(\theta)$ is monotone non-increasing for θ. Thus there exists $\theta' > 0$ such that $I(\boldsymbol{p}_1' \times \cdots \times (\theta'\boldsymbol{p}_k'' + (1 - \theta')\boldsymbol{p}_k') \times \cdots \times \boldsymbol{p}_N') < I(\boldsymbol{p}^*)$. Hence, there exists a neighborhood $U_{\boldsymbol{p}^*} \subset X$ of $\boldsymbol{p}^*$ such that $I(\boldsymbol{p}^*) \ge I(\boldsymbol{p} \in U_{\boldsymbol{p}^*})$. This means that $\boldsymbol{p}^*$ is local maximum in X.

By these two cases Proposition 3 is proved. ∎
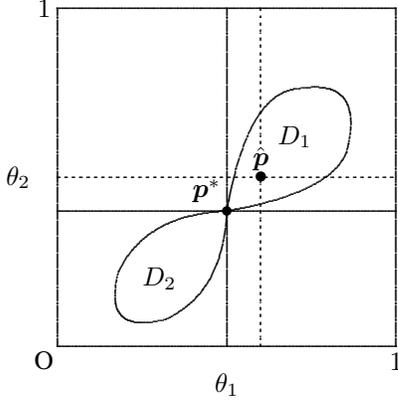
Fig. 3. Pattern of $D(a_0) = D_1 \cup D_2$ for the case of $N = 2$, $n_1 = n_2 = 2$, $p_1(0) = \theta_1$, $p_2(0) = \theta_2$.

F. Connectedness

Finally in this section, we prove the property of connectedness for the elementary MAC as a proposition:

Proposition 4: If an N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) is elementary, i.e. $n_k \leq m$, $k = 1, \cdots, N$, then the set

$$D(a) \equiv \{\boldsymbol{p} | I(\boldsymbol{p} \in X) \geq a\} \qquad (18)$$

is connected for any $a \geq 0$. $\square$

Proof: Assume that for any $\varepsilon > 0$, there exists $a_0 > 0$ such that $D(a_0)$ is connected and $D(a_0 + \varepsilon)$ is disconnected.

Since $I(\boldsymbol{p})$ is concave on each X_k, then there exist subsets D_1 and D_2 of $D(a_0)$ with properties as follows:

1) $D(a_0) = D_1 \cup D_2$, and $I(\boldsymbol{p}^*) = a_0$, for $\boldsymbol{p}^* \in D_1 \cap D_2$.
2) For any $\boldsymbol{p}'_1 \times \cdots \times \boldsymbol{p}'_k \times \cdots \times \boldsymbol{p}'_N \in D_1$, all IPD vectors $\boldsymbol{p}'_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \boldsymbol{p}'_N$, $\boldsymbol{p}_k \in X_k$, $k = 1, \cdots, N$, satisfying $I(\boldsymbol{p}'_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \boldsymbol{p}'_N) \geq a_0$ belongs to D_1, and also for any $\boldsymbol{p}''_1 \times \cdots \times \boldsymbol{p}''_k \times \cdots \times \boldsymbol{p}''_N \in D_2$, all IPD vectors $\boldsymbol{p}''_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \boldsymbol{p}''_N$, $\boldsymbol{p}_k \in X_k$, $k = 1, \cdots, N$, satisfying $I(\boldsymbol{p}''_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \boldsymbol{p}''_N) \geq a_0$ belongs to D_2 (cf. Fig. 3).

Thus for any $\varepsilon > 0$, $D(a_0 + \varepsilon)$ is separated into subsets $D'_1 \subset D_1$ and $D'_2 \subset D_2$ such that $D(a) = D'_1 \cup D'_2$ and $D'_1 \cap D'_2 = \phi$.

It is easy to see that for any $\boldsymbol{p}^* = \boldsymbol{p}^*_1 \times \cdots \times \boldsymbol{p}^*_k \times \cdots \times \boldsymbol{p}^*_N \in D_1 \cap D_2$, $k = 1, \cdots N$, it holds

$$I(\boldsymbol{p}^*_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \boldsymbol{p}^*_N) \leq a_0, \quad \boldsymbol{p}_k \in X_k \qquad (19)$$

since for any $\hat{\boldsymbol{p}} = \hat{\boldsymbol{p}}_1 \times \cdots \times \hat{\boldsymbol{p}}_N \in D_1$ (or $\hat{\boldsymbol{p}} \in D_2$), $\hat{\boldsymbol{p}} \notin D_1 \cap D_2$, every $\hat{\boldsymbol{p}}_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \hat{\boldsymbol{p}}_N$, $k = 1, \cdots, N$, satisfying

$$I(\hat{\boldsymbol{p}}_1 \times \cdots \times \boldsymbol{p}_k \times \cdots \times \hat{\boldsymbol{p}}_N) < a_0, \boldsymbol{p}_k \in X_k$$

belongs to neither D_1 nor D_2 by the property 2) (see Fig. 3).

Consider two cases: Every components of $\boldsymbol{p}^*$ is non-zero and at least a component of $\boldsymbol{p}^*$ is zero.

In the first case, it holds by (19) that $\boldsymbol{p}^* \in D_1 \cap D_2$ satisfies the Kuhn-Tucker conditions

$$
\begin{aligned}
J(\boldsymbol{p}^*_1 \times \cdots \times \boldsymbol{p}^*_N; i_k) &= a_0, \quad p^*_k(i_k) \neq 0 \\
& i_k = 0, \cdots, n_k - 1 \\
& k = 1, \cdots, N.
\end{aligned}
$$

Therefore $\boldsymbol{p}^*$ is local maximum for $\boldsymbol{p} \in X$ by Proposition 3 since $(n_1, \cdots, n_N; m)$-MAC (P, X) is elementary. Then there

exists a neighborhood $U_{\boldsymbol{p}^*}$ of $\boldsymbol{p}^*$ in X such that $I(\boldsymbol{p}) \leq a_0$ for any $\boldsymbol{p} \in U_{\boldsymbol{p}^*}$.

On the other hand, by the properties of D_1 and D_2 there exists $\boldsymbol{p}'$ in either $U_{\boldsymbol{p}^*} \cap D_1$ or $U_{\boldsymbol{p}^*} \cap D_2$ such that $I(\boldsymbol{p}') > a_0$. This is inconsistent with that $\boldsymbol{p}^*$ is local maximum. Therefore $D(a)$ is connected.

For the second case, consider the minimum domain $F \equiv F_1 \times \cdots \times F_N \subset X$ which contains the $\boldsymbol{p}^*$ exactly inside (and not on the boundary of) F_k, where $p_k(i_k) = 0$ for $i_k \notin \Lambda(F_k)$ and $p_k(i_k) > 0$ for $i_k \in \Lambda(F_k)$, $k = 1, \cdots, N$. In the same way as in the first case, it is proved that $D(a) \cap F$ is connected. Therefore $D(a)$ is connected. ∎

V. BINARY-INPUTS MAC

In this section, we investigate an N-user binary-inputs MAC (P, Y) of the N-user $(n_1, \cdots, n_N; m)$-MAC (P, X) where each Y_k of Y is formed by a line segment. For any given $\boldsymbol{\rho}'_k, \boldsymbol{\rho}''_k \in X_k$, $k = 1, \cdots, N$, define a line segment Y_k by

$$Y_k = \{\theta_k \boldsymbol{\rho}'_k + (1 - \theta_k)\boldsymbol{\rho}''_k | 0 \leq \theta_k \leq 1\}, \quad k = 1, \cdots, N$$

and denote $Y = Y_1 \times \cdots \times Y_N$. Reasonably we set $\boldsymbol{\theta} \equiv (\theta_1, \cdots, \theta_N)$ and write $\theta_k \in Y_k$, $\boldsymbol{\theta} \in Y$. Thus we can build up an N-user binary-inputs $(2, \cdots, 2; m)$-MAC (P, Y) whose channel matrix is P and domain is a subset Y of X. Obviously it is an elementary MAC since $m \geq 2$.

The mutual information of the N-user $(2, \cdots, 2; m)$-MAC (P, Y) is given by

$$
\begin{aligned}
\mathcal{I}(\theta_1, \cdots, \theta_N; \boldsymbol{\rho}', \boldsymbol{\rho}'') &\equiv I((\theta_1 \boldsymbol{\rho}'_1 + (1 - \theta_1)\boldsymbol{\rho}''_1) \times \\
& \cdots \times (\theta_N \boldsymbol{\rho}'_N + (1 - \theta_N)\boldsymbol{\rho}''_N))
\end{aligned} \qquad (20)
$$

where $0 \leq \theta_k \leq 1$, $k = 1, \cdots, N$, and $\boldsymbol{\rho}' = \boldsymbol{\rho}'_1 \times \cdots \times \boldsymbol{\rho}'_N$, $\boldsymbol{\rho}'' = \boldsymbol{\rho}''_1 \times \cdots \times \boldsymbol{\rho}''_N$. It depends on the choice of $\boldsymbol{\rho}', \boldsymbol{\rho}''$. The Kuhn-Tucker conditions for (20) are given by

$$
\begin{aligned}
\mathcal{I}_k(\theta_1, \cdots, \theta_N; \boldsymbol{\rho}', \boldsymbol{\rho}'') &= 0, \quad \theta_k > 0 \\
& \leq 0, \quad \theta_k = 0 \\
& k = 1, \cdots, N
\end{aligned} \qquad (21)
$$

where $\mathcal{I}_k(\cdots; \boldsymbol{\rho}', \boldsymbol{\rho}'') = \partial \mathcal{I}(\cdots; \boldsymbol{\rho}', \boldsymbol{\rho}'')/\partial \theta_k$. For simplicity we omit $\boldsymbol{\rho}', \boldsymbol{\rho}''$ from the expression and denote $\mathcal{I}(\boldsymbol{\theta}; \boldsymbol{\rho}', \boldsymbol{\rho}'') \equiv \mathcal{I}(\boldsymbol{\theta})$, in the subsequent discussions.

We prove the lemma to be used for the proof of Theorem 2 as follows:

Lemma 1: The Kuhn-Tucker conditions for the N-user binary $(2, \cdots, 2; m)$-MAC (P, Y) as defined above are necessary and sufficient for optimality. $\square$

Proof: It is sufficient to prove the sufficiency. Assume that there exist two solutions $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1, \cdots, \bar{\theta}_N)$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_N)$ of the Kuhn-Tucker conditions (21) such that $\mathcal{I}(\bar{\boldsymbol{\theta}}) \neq \mathcal{I}(\hat{\boldsymbol{\theta}})$. Without loss of generality, assume that $\mathcal{I}(\bar{\boldsymbol{\theta}}) > \mathcal{I}(\hat{\boldsymbol{\theta}})$.

Since the N-user binary $(2, \cdots, 2; m)$-MAC (P, Y) is elementary, by Proposition 3 the solution $\hat{\boldsymbol{\theta}}$ is local maximum in Y and there exists a neighborhood $U_{\hat{\boldsymbol{\theta}}}$ of $\hat{\boldsymbol{\theta}}$ such that $\mathcal{I}(\hat{\boldsymbol{\theta}}) \geq \mathcal{I}(\boldsymbol{\theta} \in U_{\hat{\boldsymbol{\theta}}})$. Also by Proposition 4 the set $D(\mathcal{I}(\hat{\boldsymbol{\theta}})) \equiv \{\boldsymbol{\theta} | \mathcal{I}(\boldsymbol{\theta}) \geq \mathcal{I}(\hat{\boldsymbol{\theta}}), \boldsymbol{\theta} \in Y\}$ is connected and includes both $\bar{\boldsymbol{\theta}}$
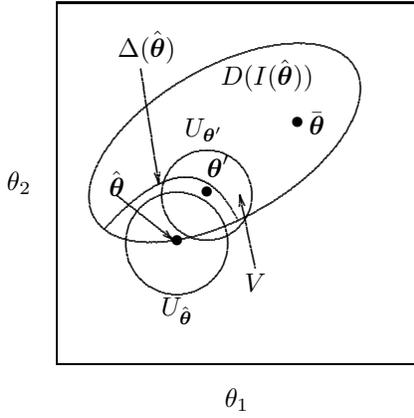
Fig. 4. Two solutions of the Kuhn-Tucker conditions for two-user case.

and $\hat{\boldsymbol{\theta}}$. Then for any $\boldsymbol{\theta} \in D(\mathcal{I}(\hat{\boldsymbol{\theta}})) \cap U_{\hat{\boldsymbol{\theta}}}$, it is easy to see $\mathcal{I}(\boldsymbol{\theta}) = \mathcal{I}(\hat{\boldsymbol{\theta}})$.

Let $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^\dagger$ be any points in $D(\mathcal{I}(\hat{\boldsymbol{\theta}})) \cap U_{\hat{\boldsymbol{\theta}}}$, and set $\mathcal{I}(\theta_1^*, \cdots, (\alpha\theta_k^* + (1-\alpha)\theta_k^\dagger), \cdots, \theta_N^*)$ as a function of the variable α. Since $\mathcal{I}(\boldsymbol{\theta})$ is concave for each variable θ_k and $\mathcal{I}(\boldsymbol{\theta}^*) = \mathcal{I}(\boldsymbol{\theta}^\dagger) = I(\hat{\boldsymbol{\theta}})$, we have $\mathcal{I}(\theta_1^*, \cdots, (\alpha\theta_k^* + (1-\alpha)\theta_k^\dagger), \cdots, \theta_N^*) = \mathcal{I}(\hat{\boldsymbol{\theta}})$ for $0 \le \alpha \le 1$. Therefore, it holds

$$\frac{d\mathcal{I}(\theta_1^*, \cdots, (\alpha\theta_k^* + (1-\alpha)\theta_k^\dagger), \cdots, \theta_N^*)}{d\alpha} =$$
$$(\theta_k^* - \theta_k^\dagger)\mathcal{I}_k(\theta_1^*, \cdots, (\alpha\theta_k^* + (1-\alpha)\theta_k^\dagger), \cdots, \theta_N^*)$$
$$= 0.$$

This implies that any $\boldsymbol{\theta} \in D(\mathcal{I}(\hat{\boldsymbol{\theta}})) \cap U_{\hat{\boldsymbol{\theta}}}$ satisfies the Kuhn-Tucker conditions (21): $\mathcal{I}_k(\theta_1, \cdots, \theta_k, \cdots, \theta_N) = 0$, $k = 1, \cdots, N$, even if $\hat{\boldsymbol{\theta}}$ is located on the boundary of Y.

Let $\Delta(\hat{\boldsymbol{\theta}})$ be a set

$$\Delta(\hat{\boldsymbol{\theta}}) \equiv \{\boldsymbol{\theta} | \mathcal{I}_k(\boldsymbol{\theta}) = 0, \ k = 1, \cdots, N,$$
$$\mathcal{I}(\boldsymbol{\theta}) = \mathcal{I}(\hat{\boldsymbol{\theta}}), \ \boldsymbol{\theta} \in D(I(\hat{\boldsymbol{\theta}}))\}.$$

Clearly, this includes $D(\mathcal{I}(\hat{\boldsymbol{\theta}})) \cap U_{\hat{\boldsymbol{\theta}}}$ and it holds $\mathcal{I}(\boldsymbol{\theta}) = \mathcal{I}(\hat{\boldsymbol{\theta}})$ for $\boldsymbol{\theta} \in \Delta(\hat{\boldsymbol{\theta}})$ (see Fig 4). Note that each point in $\Delta(\hat{\boldsymbol{\theta}})$ is local maximum. Then for any $\boldsymbol{\theta}' \in \Delta(\hat{\boldsymbol{\theta}})$, there exists a neighborhood $U_{\boldsymbol{\theta}'}$, such that $\mathcal{I}(\boldsymbol{\theta}) \le \mathcal{I}(\boldsymbol{\theta}')(= \mathcal{I}(\hat{\boldsymbol{\theta}}))$ for any $\boldsymbol{\theta} \in U_{\boldsymbol{\theta}'}$.

Here we define a subset of $U_{\boldsymbol{\theta}'}$ as $V \equiv \{\boldsymbol{\theta} | \boldsymbol{\theta} \in U_{\boldsymbol{\theta}'}, \boldsymbol{\theta} \notin \Delta(\hat{\boldsymbol{\theta}})\} \cap D(\mathcal{I}(\hat{\boldsymbol{\theta}}))$. Assume that V is non-empty. Then it holds $\mathcal{I}(\boldsymbol{\theta}) < \mathcal{I}(\hat{\boldsymbol{\theta}})$ for any $\boldsymbol{\theta} \in V$, since $\boldsymbol{\theta} \in \{\boldsymbol{\theta} | \boldsymbol{\theta} \in U_{\boldsymbol{\theta}'}, \boldsymbol{\theta} \notin \Delta(\hat{\boldsymbol{\theta}})\}$. On the other hand, it holds that $\mathcal{I}(\boldsymbol{\theta}) \ge \mathcal{I}(\hat{\boldsymbol{\theta}})$ for any $\boldsymbol{\theta} \in V$ since $\boldsymbol{\theta} \in D(\mathcal{I}(\hat{\boldsymbol{\theta}}))$ by the definition of V. This is inconsistent with the assumption that V is non-empty. Thus V is empty and $\Delta(\hat{\boldsymbol{\theta}}) = D(\mathcal{I}(\hat{\boldsymbol{\theta}}))$.

Since both $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ belong to $D(\mathcal{I}(\hat{\boldsymbol{\theta}}))$, it holds $\mathcal{I}(\hat{\boldsymbol{\theta}}) = \mathcal{I}(\bar{\boldsymbol{\theta}})$. Therefore the assumption $\mathcal{I}(\bar{\boldsymbol{\theta}}) > \mathcal{I}(\hat{\boldsymbol{\theta}})$ is invalid. This means that any solution $\boldsymbol{\theta}$ of the Kuhn-Tucker conditions (21) for (P, Y) gives the same value for $\mathcal{I}(\boldsymbol{\theta})$ and then it is optimal. Thus the sufficiency is proved. ∎

Note that the Lemma 1 holds for any domain Y of X formed by $\boldsymbol{\rho}'$ and $\boldsymbol{\rho}''$.

VI. PROOF OF THEOREM 2

In this section, we prove Theorem 2 by using Lemma 1. We state again Theorem 2:

Theorem 2: The Kuhn-Tucker conditions for the channel capacity C of an N-user elementary $(n_1, \cdots, n_N; m)$-MAC (P, X), where $n_k \le m$ for all $k = 1, \cdots, N$, are necessary and sufficient. ∎

Proof: It is sufficient to prove the sufficiency. Let $\bar{\boldsymbol{p}} = \bar{\boldsymbol{p}}_1 \times \cdots \times \bar{\boldsymbol{p}}_N$ a solution of the Kuhn-Tucker conditions (3) for the N-user elementary $(n_1, \cdots, n_N; m)$-MAC (P, X), where $n_k \le m$, $k = 1, \cdots, N$. We prove that $\bar{\boldsymbol{p}}$ is uniquely determined in the sense that any solution $\boldsymbol{p}$ of the Kuhn-Tucker conditions (3) gives the same for $I(\boldsymbol{p})$.

For an arbitrary $\boldsymbol{p}_k' \in X_k$, $k = 1, \cdots, N$, there exist $\boldsymbol{p}_k'' \in X_k$ and $\bar{\theta}_k$ such that

$$\bar{\boldsymbol{p}}_k = \bar{\theta}_k \boldsymbol{p}_k' + (1 - \bar{\theta}_k)\boldsymbol{p}_k'', \ \ 0 \le \bar{\theta}_k \le 1 \tag{22}$$

since X_k is simplex. Then $\bar{\boldsymbol{p}}$ is represented by

$$\bar{\boldsymbol{p}} = (\bar{\theta}_1 \boldsymbol{p}_1' + (1 - \bar{\theta}_1)\boldsymbol{p}_1'') \times$$
$$\cdots \times (\bar{\theta}_N \boldsymbol{p}_N' + (1 - \bar{\theta}_N)\boldsymbol{p}_N''). \tag{23}$$

Here we define a function of variables $(\theta_1, \cdots, \theta_N) \equiv \boldsymbol{\theta}$ $(0 \le \theta_k \le 1)$ by

$$\mathcal{I}(\boldsymbol{\theta}; \boldsymbol{p}', \boldsymbol{p}'') \equiv I((\theta_1 \boldsymbol{p}_1' + (1 - \theta_1)\boldsymbol{p}_1'') \times$$
$$\cdots \times (\theta_N \boldsymbol{p}_N' + (1 - \theta_N)\boldsymbol{p}_N'')) \tag{24}$$

where $\boldsymbol{p}' = \boldsymbol{p}_1' \times \cdots \times \boldsymbol{p}_N'$ and $\boldsymbol{p}'' = \boldsymbol{p}_1'' \times \cdots \times \boldsymbol{p}_N''$. The function (24) can be regarded as the mutual information of an N-user $(2, \cdots, 2; m)$-MAC (P, Y) with the domain $Y \equiv Y_1 \times \cdots \times Y_N$, where $Y_k \equiv \{\theta_k \boldsymbol{p}_k' + (1 - \theta_k)\boldsymbol{p}_k'' | 0 \le \theta_k \le 1\}$, $k = 1, \cdots, N$. The N-user $(2, \cdots, 2; m)$-MAC (P, Y) is denoted by $(P, Y)_{(\boldsymbol{p}', \boldsymbol{p}'')}$, since it depends on $\boldsymbol{p}', \boldsymbol{p}''$.

Since $\bar{\boldsymbol{p}}$ is a solution of the Kuhn-Tucker conditions (3) for (P, X), then $\bar{\boldsymbol{\theta}}$ is a solution of the Kuhn-Tucker conditions for the mutual information (24) of $(P, Y)_{(\boldsymbol{p}', \boldsymbol{p}'')}$:

$$\begin{aligned} \mathcal{I}_k(\boldsymbol{\theta}; \boldsymbol{p}', \boldsymbol{p}'') &= C, \ \theta_k > 0 \\ &\le C, \ \theta_k = 0 \end{aligned}$$

$$k = 1, \cdots, N, \ \ C = \mathcal{I}(\boldsymbol{\theta}; \boldsymbol{p}', \boldsymbol{p}'') \tag{25}$$

where $\mathcal{I}_k(\boldsymbol{\theta}; \boldsymbol{p}', \boldsymbol{p}'') = \partial\mathcal{I}(\boldsymbol{\theta}; \boldsymbol{p}', \boldsymbol{p}'')/\partial\theta_k$. Therefore, it follows from Lemma 1 that $\bar{\boldsymbol{\theta}}$ is optimal for $(P, Y)_{(\boldsymbol{p}', \boldsymbol{p}'')}$, which means

$$\mathcal{I}(\bar{\boldsymbol{\theta}}; \boldsymbol{p}', \boldsymbol{p}'') \ge \mathcal{I}(\boldsymbol{\theta}; \boldsymbol{p}', \boldsymbol{p}'') \tag{26}$$

for any $\boldsymbol{\theta} \in Y$.

Since $\bar{\boldsymbol{\theta}}$ is given by (23), it holds

$$\mathcal{I}(\bar{\boldsymbol{\theta}}; \boldsymbol{p}', \boldsymbol{p}'') = I(\bar{\boldsymbol{p}}) \tag{27}$$

for any $\boldsymbol{p}' \in X$, where $\boldsymbol{p}''$ satisfies (22). Thus since (26) and (27) are valid for any $\boldsymbol{p}' \in X$, it holds

$$I(\bar{\boldsymbol{p}}) \ge I(\boldsymbol{p})$$

on the whole domain X. This implies that $\bar{\boldsymbol{p}}$ is optimal.

Thus we proved the theorem. ∎

VII. CONCLUSIONS

After Shannon [2] multiuser channel has long been studied in various fields. However not much works have been made for the fundamental property of the channel capacity of an N-user $(n_1, \cdots, n_N)$-MAC (P, X) in general except for some specific cases.

We have shown that there exists a non-trivial MAC where the Kuhn-Tucker conditions are necessary and sufficient for the channel capacity. We called it as an *elementary* MAC that was defined by the MAC whose sizes of input alphabets must be not greater than the size of output alphabet. Obviously the N-user binary inputs $(2, \cdots, 2; m)$-MAC (P, X) is a typical example of the elementary MAC. Also the DMC is a trivial elementary MAC.

We believe that there is considerable merit in a concept of elementary MAC for which the channel capacity is evaluated precisely by the necessary and sufficient condition as in the case of DMC. In fact, we have proved as Theorem 1 that the channel capacity of any MAC is achieved by the channel capacity of *an* elementary MAC contained in the original MAC. Thus an MAC in general can be regarded as simply an aggregate of elementary MAC's. This statement is a basic idea behind our formulation of this paper.

The most of this paper was devoted to the proof of Theorem 2 such that the Kuhn-Tucker conditions are sufficient (the necessity is self-evident) for the channel capacity of the elementary MAC. We have shown as Proposition 2 that a solution of the Kuhn-Tucker conditions if it satisfies the equality portion of the conditions satisfies the boundary equations which define the boundary of the capacity region. Then we could prove the property of local maximum as Proposition 3 followed by the property of connectedness as Proposition 4. By using these two distinctive features we could prove that any solution of the Kuhn-Tucker conditions of the elementary MAC was uniquely determined, that is, each solution takes the same value for the the mutual information and therefore it achieves the channel capacity.

In this respect, we remark that the non-elementary MAC has a degenerate property as explained in Section II. If it exists, then it is difficult to identify which IPD vectors are exactly contributed to the mutual information of the MAC. However we overcome these difficulties by introducing the concept of elementary MAC where there exists no such degenerate property. Since the well-known DMC is elementary, then the elementary MAC is identified as an extension of the DMC.

Incidentally, our notation introduced in this paper seems rather non-standard including expressions of IPD vector $\boldsymbol{p}$, Kronecker products $\boldsymbol{p} = \boldsymbol{p}_1 \times \cdots \times \boldsymbol{p}_N$, the channel matrix P regarded as a non-linear mapping, domain X, face F, and so force. However we emphasize that the notation appears effective to resolve the cumbersome procedures relating to the extremum evaluation of the multi-variable mutual information with constraints for the MAC.

Before closing we remark that the very essence of information theory consists in two major subjects such as source coding and channel coding as we know. This paper seems to be quite effective in working out the subject of channel coding since we provide for a formalism to determine the channel capacity of the MAC. We are confident that two distinctive features of local maximum (Proposition 3) and connectedness (Proposition 4) represent an intrinsic structure of the MAC. However we are not content ourselves with this stage. We are expecting that our results will be a mathematical base for various subjects of the MAC including the numerical and/or exact evaluation of the capacity region, the analysis of the MAC with feedback as well as the structured approach to the multiuser coding, and so force.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory,* John Wiley & Sons, Inc., 2006.

[2] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.,* vol.1, pp.611-644, Berkeley Calif.:Univ. Calif. Press, 1961.

[3] H. Liao, "Multiple-access channels," Ph.D. dissertation, Dept. Elec., Univ. Hawaii, Honolulu, 1972.

[4] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. on Information Theory,* Tsahkadsor, Armenia S.S.R., 1971, Publishing House of the Hungarian Academy of Sciences, pp.23-52, 1973.

[5] E. C. van der Meulen, "The discrete memoryless channels with two senders and one receiver," in *Proc. 2nd Int. Symp. Inform. Theory,* Tsahkadsor, Armenia S.S.R., 1971, Publishing House of the Hungarian Academy of Sciences, pp.103-135, 1973.

[6] M. L. Ulrey, "The capacity region of a channel with s senders and r receivers," *Inform. and Control,* vol.29, pp.185-203, 1975.

[7] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inf. Theory*, vol.IT-21, No.1, pp.100-102, Jan. 1975.

[8] Shih-Chun Chang, "Coding for T-user multiple-access Channels," *IEEE Trans. Inf. Theory*, vol.IT-25, No.6, pp.684-691, Nov. 1979.

[9] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol.IT-27, No.3, pp.292-298, May 1981.

[10] J. Y. N. Hui and P. A. Humblet, "The Capacity region of the totally asynchronous multiple-access channel," *IEEE Trans. Inf. Theory*, vol.IT-31, No.2, pp.207-216, March 1985.

[11] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inf. Theory*, vol.IT-35, No.3, pp.605-619, May 1989.

[12] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. on Information Theory*, vol.49, no.1, pp.4-21, Jan. 2003.

[13] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels—Part I," *IEEE Trans. Inf. Theory*, vol.43, No.6, pp.1877-1894, Nov. 1997.

[14] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels—Part II," *IEEE Trans. Inf. Theory*, vol. 43, No. 6, pp.1895-1911, Nov. 1997.

[15] B. Suard G. Xu, H. Liu and T. Kailath, "Uplink channel capacity of space-division-multiple-access schemes," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1468-1476, July 1998.

[16] Wei Yu, J. M.Cioffi, "FDMA Capacity of Gaussian Multiple-Access Channel with ISI," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 102-111, Jan. 2002.

[17] P. Gupta and P. R. Kumar, "Towards an information theory of large Network: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp.1877-1894, Aug. 2003.

[18] M. Rezaeian and A. J. Grant, "Computation of total capacity for discrete memoryless multiple-access channel," *IEEE Trans. Inf. Theory*, vol. IT-50, no. 11, pp. 2779-2784, Nov., 2003.

[19] Y. Watanabe, "The total capacity of two-user multiple-access channel with binary output," *IEEE Trans. Inform. Theory*, vol. IT-42, no. 5, pp. 1453-1465, Sept., 1996.

[20] R. G. Gallager, *Information Theory and Reliable Communication,* New York, Wiley, 1968.