COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 1, No. 1, pp. 87–100, January 2001

006

A CLASS OF NON-SHANNON-TYPE INFORMATION INEQUALITIES AND THEIR APPLICATIONS*

RAYMOND W. YEUNG† AND ZHEN ZHANG‡

Abstract. Information inequalities form the most important set of tools for proving converse coding theorems in information theory. They are sometimes referred to as the "Laws of Information Theory," because they govern the impossibilities in information theory. For a long time, all information inequalities we knew were nothing but simple consequences of the nonnegativity of Shannon's information measures. Owing to the recent discovery of a few so-called non-Shannon-type information inequalities, it is now known that there are laws in information theory beyond those laid down by Shannon. In this paper, we show that the unconditional inequality discovered by the authors in 1998 in fact implies a class of 2^{14} non-Shannon-type inequalities, and we show possible applications of these inequalities in information theory problems. The results thus obtained are not possible otherwise.

Keywords: Non-Shannon-Type Information Inequalities, Entropy Functions.

1. Introduction. Entropy is the most fundamental measure of information in information theory. We begin by first defining the entropy of a random variable and then the joint entropy of a pair of random variables. In this paper, all random variables are discrete, and all logarithms are in base 2.

DEFINITION 1.1. *Let X be a random variable with support set $\mathcal{S}_X$ and probability mass function $p(x) = \Pr(X = x), x \in \mathcal{S}_X$. The entropy $H(X)$ of X is defined by*

$$H(X) = -\sum_{x \in \mathcal{S}_X} p(x) \log p(x).$$

DEFINITION 1.2. *Let (X, Y) be a pair of random variables with support set $\mathcal{S}_{XY}$ and joint probability mass function $p(x, y) = \Pr(X = x, Y = y)$, $(x, y) \in S_{XY}$. The joint entropy $H(X, Y)$ of (X, Y) is defined by*

$$H(X, Y) = -\sum_{(x,y) \in \mathcal{S}_{XY}} p(x, y) \log p(x, y).$$

The joint entropy for three or more random variables can be defined likewise. Let X, Y, and Z be random variables. From the (joint) entropies, we can define the conditional entropy of X given Y as

$$H(X|Y) = H(X, Y) - H(Y),$$

the mutual information between X and Y as

$$I(X; Y) = H(X) - H(X|Y),$$

*Invited paper; received June 13, 2000; accepted for publication August 28, 2000.

† Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; email: whyeung@ie.cuhk.edu.hk

‡Communication Sciences Institute, Department of Electrical Engineering-systems, University of Southern California, Los Angeles, California, 90089-2565; email: zzhang@milly.usc.edu

and the conditional mutual information between X and Y given Z as

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z).$$

Entropies, conditional entropies, mutual informations, and conditional mutual informations are called *Shannon's information measures.*

Let $\mathcal{N} = \{1, \ldots, n\}$ and let $\Theta = \{X_i, i \in \mathcal{N}\}$ be any collection of n random variables. Associated with $\{X_i, i \in \mathcal{N}\}$ are $2^n - 1$ joint entropies. For example, when $n = 3$, the 7 joint entropies are

$$H(X_1), H(X_2), H(X_3), H(X_1, X_2), H(X_2, X_3), H(X_1, X_3), H(X_1, X_2, X_3).$$

Note that all other types of Shannon's information measures, namely mutual informations, conditional entropies, and conditional mutual informations are all linear combinations of entropies.

For any subset α of $\mathcal{N}$, let $X_\alpha = (X_i, i \in \alpha)$ and $H_\Theta(\alpha) = H(X_\alpha)$. For fixed Θ, one can then view H_Θ as a set function from $2^\mathcal{N}$ to $\mathbb{R}$ with $H_\Theta(\phi) = 0$, i.e., the entropy of the empty set of random variables is equal to zero. For this reason, we call H_Θ the entropy function of Θ.

It is well-known that all Shannon's information measures are always nonnegative, i.e., they are nonnegative for all joint probability mass functions of the random variables involved (see for example [17, Chapter 2]). These are called the *basic inequalities* [5]. In Appendix A, we will formally show that the basic inequalities are equivalent to the following set of constraints on the entropy functions: for any Θ and all $\alpha, \beta \subset \mathcal{N}$, H_Θ satisfies

(P1) $H_\Theta(\phi) = 0$;

(P2) $H_\Theta(\alpha) \leq H_\Theta(\beta)$ if $\alpha \subset \beta$;

(P3) $H_\Theta(\alpha) + H_\Theta(\beta) \geq H_\Theta(\alpha \cap \beta) + H_\Theta(\alpha \cup \beta)$.

(P1)-(P3) are called the *polymatriod axioms.*

In the rest of the paper, we will refer to inequalities (identities) involving only Shannon's information measures as information inequalities (identities). In fact, an information identity can be regarded as two information inequalities (see Section 2.1).

Information inequalities form the most important set of tools for proving converse coding theorems in information theory. They govern the impossibilities in information theory. In the 1986 SPOC Conference, N. Pippenger gave a talk in which he referred to constraints on entropies as the "laws of information theory" [1]. He asked whether there is any constraint on entropy functions in addition to the polymatroid axioms.

During the last ten years, a number of researchers have made much progress in understanding the properties of entropy functions. So far, these results not only have revealed the set-theoretic structure of Shannon's information measures [3] (see Section 2.3), but also have made machine-proving of information inequalities possible [5] (see Section 2.2). In particular, owing to the discovery of a so-called *non-Shannon-type* information inequality [7], Pippenger's open problem is finally settled: the polymatroid axioms form an incomplete set of constraints on entropy functions.

Recent findings show that information inequalities not only are intimately related to certain multiterminal source coding problems [8], but they also have fundamental

implications beyond information theory, for example, in conditional independence of random variables and in group theory. We refer the reader to [14] for a list of reference along this line.

Although non-Shannon-type inequalities originated in information theory, so far there has not been any application of these inequalities in information theory problems. Some researchers even doubt whether non-Shannon-type inequalities have any implication at all in information theory. The current work is a first attempt to explore possible applications of these inequalities in information theory problems.

The unconditional non-Shannon-type inequality reported in [7] was presented in more than one form. In this paper, we focus on the following form which is symmetrical both in X_1 and X_2 and in X_3 and X_4, and we will denote it by ZY98.

$$I(X_3; X_4) \leq I(X_3; X_4|X_1) + I(X_3; X_4|X_2) + 0.5I(X_1; X_2)$$

$$+0.25I(X_1; X_3, X_4) + 0.25I(X_2; X_3, X_4). \qquad (ZY98)$$

The results in this paper apply equally well to the other forms of the inequality in [7].

In the next section, we will present the preliminaries for the results to be discussed in Sections 3 and 4. In Section 3, we will show that ZY98 in fact implies $2^{14} - 1$ conditional non-Shannon-type inequalities. Together with ZY98, they form a class of 2^{14} non-Shannon-type information inequalities. In Section 4, we will discuss possible applications of this class of inequalities. Concluding remarks are in Section 5.

2. Preliminaries.

2.1. Entropy Functions and Information Inequalities. In this subsection, we present the framework for information inequalities in [5]. Recall that H_Θ is a function from $2^\mathcal{N}$ to $\mathbb{R}$ with $H_\Theta(\phi) = 0$. Let $k = 2^n - 1$. Labeling the coordinates of $\mathbb{R}^k$ by $h_\alpha, \alpha \in 2^\mathcal{N} \backslash \phi$, where h_α corresponds to the value of $H_\Theta(\alpha)$, an entropy function H_Θ can be represented by a vector in $\mathbb{R}^k$. On the other hand, a vector $h \in \mathbb{R}^k$ is called *constructible* if h represents the entropy function of some collection of n random variables. Define the following region in $\mathbb{R}^k$:

$$\Gamma_n^* = \{h \in \mathbb{R}^k : h \text{ is constructible}\}.$$

For example, when $n = 3$, the coordinates of $\mathbb{R}^7$ are labeled by

$$h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123},$$

and Γ_3^* is the region in $\mathbb{R}^7$ of all entropy functions of 3 random variables.

An information inequality (linear or nonlinear) has the form $f(h) \geq 0$, where $f : \mathbb{R}^k \to \mathbb{R}$. We consider non-strict inequalities only because these are usually the inequalities of concern in information theory. For example, the inequality $I(X_1; X_2) \geq 0$ is written as $h_1 + h_2 - h_{12} \geq 0$. Since an information inequality involving n random variables *always holds* if and only if it is satisfied by the entropy function of any collection of n random variables, we have the following geometric interpretation of an information inequality:

$f(h) \geq 0$ always holds if and only if $\Gamma_n^* \subset \{h \in \mathbb{R}^k : f(h) \geq 0\}$.

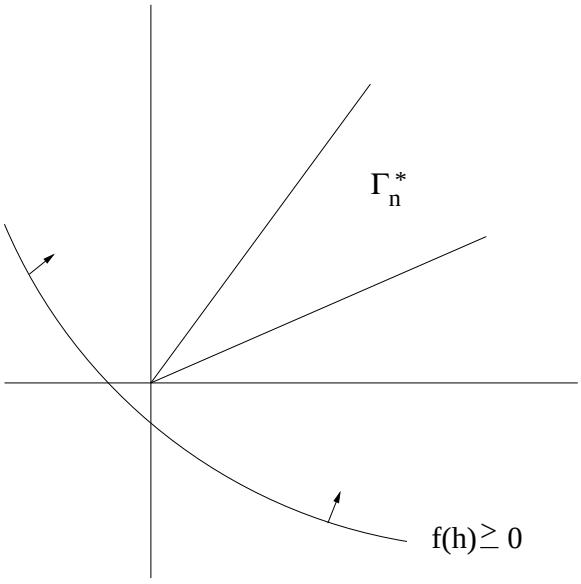

FIG. 2.1. $f(h) \geq 0$ *always holds.*

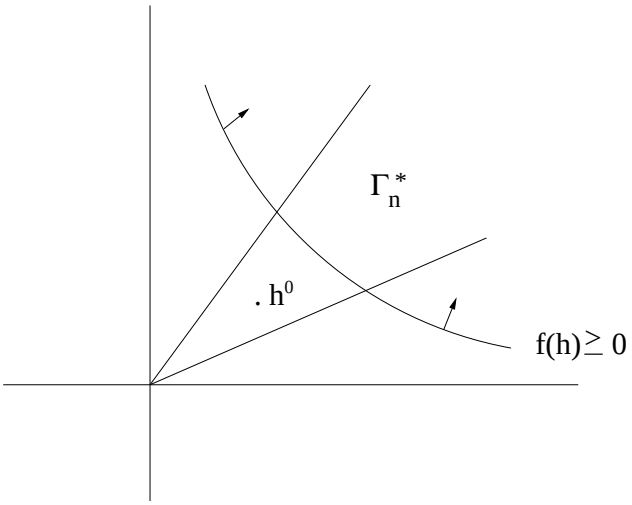

FIG. 2.2. $f(h) \geq 0$ *does not always hold.*

The two possible cases for $f(h) \geq 0$ are illustrated in Figure 2.1 and Figure 2.2.

Note that Γ_n^* obviously contains the origin, which is the entropy function of the collection of n degenerate random variables. In Figure 2.1, Γ_n^* is completely included in the region $\{h \in \mathbb{R}^k : f(h) \geq 0\}$, so $f(h) \geq 0$ always holds. In Figure 2.2, there exists a vector h^0 which corresponds to some entropy function H_Θ such that $f(h^0) < 0$. Thus the inequality $f(h) \geq 0$ does not always hold. If Γ_n^* is known, we in principle can determine whether any information inequality always holds.

In information theory, we very often deal with information inequalities with cer-

tain constraints on the random variables involved. These are called conditional information inequalities. Such constraints on the random variables can usually be expressed as linear constraints on the entropies. The following are such examples:

1. X_1, X_2 and X_3 are mutually independent if and only if $H(X_1, X_2, X_3) = H(X_1) + H(X_2) + H(X_3)$.
2. X_1 is a function of X_2 if and only if $H(X_1|X_2) = 0$.
3. The Markov chain $X_1 - X_2 - X_3 - X_4$ is equivalent to $I(X_1; X_3|X_2) = 0$ and $I(X_1, X_2; X_4|X_3) = 0$.

It turns out that Γ_n^* not only characterizes all unconditional information inequalities, but also all conditional information inequalities. This is seen by observing that each linear constraint on the entropies is a hyperplane in $\mathbb{R}^k$. In general, linear constraints on the entropies can be expressed as a set of homogeneous linear equations $Qh = 0$. Defining the linear subspace

$$\Phi = \{h \in \mathbb{R}^k : Qh = 0\} \tag{2.1}$$

and generalizing our interpretation of unconditional information inequalities, we have

Under the constraint $Qh = 0$, $f(h) \geq 0$ always holds if and only if $(\Gamma_n^* \cap \Phi) \subset \{h \in \mathbb{R}^k : f(h) \geq 0\}$.

An information identity $f(h) = 0$ always holds if and only if both $f(h) \geq 0$ and $f(h) \leq 0$ always hold. Then we have the following interpretation of a conditional information identity:

Under the constraint $Qh = 0$, $f(h) = 0$ always holds if and only if $(\Gamma_n^* \cap \Phi) \subset \{h \in \mathbb{R}^k : f(h) = 0\}$.

Unfortunately, Γ_n^* is extremely difficult to characterize, and only partial characterizations of the region have been possible. Let us now define Γ_n as the set of all $h \in \mathbb{R}^k$ which satisfy the following properties for all $\alpha, \beta \subset \mathcal{N}$:

1. $h_\alpha \leq h_\beta$ if $\phi \neq \alpha \subset \beta$;
2. $h_\alpha + h_\beta \geq h_{\alpha\cap\beta} + h_{\alpha\cup\beta}$.

These are precisely the polymatroid axioms except that the coordinate h_ϕ is degenerated since $H_\Theta(\phi)$ is always equal to 0. Note that Γ_n is also the set of all vectors in $\mathbb{R}^k$ which satisfy the basic inequalities (cf. Appendix A). Since the basic inequalities are observed by all entropy functions, we immediately see that Γ_n is an outer bound on Γ_n^*. The question is whether this outer bound is tight. It turns out that $\Gamma_2^* = \Gamma_2$, but for $n \geq 3$, $\Gamma_n^* \neq \Gamma_n$. In fact, it has been found that Γ_3^* is not even closed [6]!

As Γ_n^* cannot be fully characterized, a more manageable task is to characterize $\overline{\Gamma}_n^*$, the closure of Γ_n^*. If one is interested in unconditional linear information inequalities, then it suffices to consider $\overline{\Gamma}_n^*$ because

$$\Gamma_n^* \subset \{h \in \mathbb{R}^k : f(h) \geq 0\}$$

if and only if

$$\overline{\Gamma}_n^* \subset \{h \in \mathbb{R}^k : f(h) \geq 0\}.$$

This follows from the fact that the region $\{h \in \mathbb{R}^k : f(h) \geq 0\}$ is closed. However, if one is interested in conditional inequalities, a more detailed characterization of Γ_n^* is necessary.

The authors proved in [6] that $\overline{\Gamma}_n^*$ is in general a convex cone. In the same paper, we further proved that $\overline{\Gamma}_3^* = \Gamma_3$ (also see [9]). In other words, every unconditional inequality involving 3 random variables can be proved by invoking the basic inequalities.

Now write ZY98 as $g(h) \geq 0$. This inequality is called non-Shannon-type because it is not a consequence of the basic inequalities, i.e.,

$$\Gamma_4 \not\subset \{h \in \mathbb{R}^{15} : g(h) \geq 0\}.$$

On the other hand, since $g(h) \geq 0$ always holds,

$$\Gamma_4^* \subset \{h \in \mathbb{R}^{15} : g(h) \geq 0\}.$$

Taking closure on both sides, we have

$$\overline{\Gamma}_4^* \subset \{h \in \mathbb{R}^{15} : g(h) \geq 0\}.$$

Thus we conclude that $\overline{\Gamma}_4^*$ is a proper subset of Γ_4. This is illustrated in Figure 2.3. From this figure, we see that ZY98 together with the basic inequalities form a tighter outer bound on Γ_4^* than the basic inequalities alone.

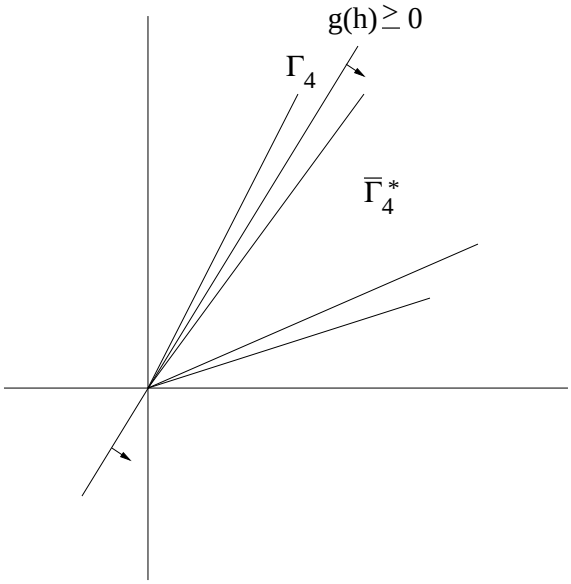

FIG. 2.3. *An illustration of* $\overline{\Gamma}_4^*$, Γ_4, *and* $g(h) \geq 0$.

With the discovery of ZY98, Pippenger's problem is finally settled. However, a physical interpretation of this inequality is yet to be obtained.

We now summarize what is known about the relation among Γ_n^*, $\overline{\Gamma}_n^*$, and Γ_n. For $n = 2$,

$$\Gamma_2^* = \overline{\Gamma}_2^* = \Gamma_2.$$

For $n = 3$,

$$\Gamma_3^* \subsetneq \overline{\Gamma}_3^* = \Gamma_3.$$

For $n \geq 4$,

$$\Gamma_n^* \subsetneq \overline{\Gamma}_n^* \subsetneq \Gamma_n.$$

Actually, the first non-Shannon-type inequality, which is a conditional inequality, was reported earlier by the authors in [6]. We showed that if $I(X_1; X_2) = 0$ and $I(X_1; X_2|X_3) = 0$[1], then

$$I(X_1; X_2|X_3, X_4) \leq I(X_1; X_2|X_4) + I(X_3; X_4|X_1, X_2). \tag{2.2}$$

Since the constraints on the above inequality are obtained by setting two basic inequalities to equality, this inequality means that there is a certain region on the boundary of Γ_4 which is not constructible. However, this is not strong enough to imply that $\overline{\Gamma}_4^* \neq \Gamma_4$.

For the inequality in (2.2), if we further impose the condition $I(X_1; X_2|X_4) = I(X_3; X_4|X_1, X_2) = 0$, then we immediately have $I(X_1; X_2|X_3, X_4) = 0$ because it is always nonnegative. That is, for 4 random variables X_1, X_2, X_3, and X_4, if 1) X_1 and X_2 are independent, 2) X_1 and X_2 are independent given X_3, 3) X_1 and X_2 are independent given X_4, and 4) X_3 and X_4 are independent given X_1 and X_2, then X_1 and X_2 given X_3 and X_4 are independent. This is a constraint on conditional independence relations for 4 random variables which are not implied by the basic inequalities.

Subsequent to [6] and [7], the open problem of the conditional independence structure for 4 random variables was finally settled by Matúš [12] by means of a conditional non-Shannon-type inequality involving 4 random variables which is different from the one in (2.2).

2.2. ITIP. In the past, information inequalities had to be proved by hand. This is done by successive invocations of the basic inequalities. When a certain inequality cannot be proved, we do not know whether the inequality is incorrect, or we just have not invoked the right basic inequality at the right step. Of course, we now know that there exist non-Shannon-type inequalities which cannot be proved by this method.

Now information inequalities can be proved by a software called ITIP [18]. The current version of ITIP runs on MATLAB™ on the Unix System. It can prove all inequalities involving a definite number of random variables which are implied by the basic inequalities, namely those provable by the method we used to know. In fact, it was shown in [5] that all these inequalities are nothing but linear combinations of the basic inequalities with nonnegative coefficients.

Using ITIP is simple and intuitive. The following examples illustrate the use of ITIP:

1.
```
> ITIP('H(XYZ) <= H(X) + H(Y) + H(Z)')
> True
```
2.
```
> ITIP('I(Y;Z) >= I(X;U)', 'I(X;Z|Y) = 0',
  'I(XY;U|Z) = 0')
> True
```

[1] $I(X_1; X_2) = 0$ and $I(X_1; X_2|X_3) = 0$ do not imply each other.

```
3. > ITIP('I(Z;U) - I(Z;U|X) - I(Z;U|Y) <= 0.5 I(X;Y)
     + 0.25 I(X;ZU) + 0.25 I(Y;ZU)')
   > Not provable by ITIP
```

In the first example, we prove an unconditional inequality. In the second example, we prove the celebrated Data Processing Theorem (see for example [16, p. 80]). In this example, the first inequality is what we want to prove, while the second and the third equalities are the conditions which specify the Markov chain $X - Y - Z - U$. In the third example, we try to prove the inequality ZY98. When ITIP returns the clause "Not provable by ITIP," it means that the inequality may be true but it cannot be proved by ITIP. But of course, ZY98 is a non-Shannon-type inequality which always holds.

ITIP results from the framework for information inequalities presented in the last subsection. Basically, the geometrical interpretation of information inequalities allows one to formulate the problem of proving these inequalities (both unconditional and conditional) as a linear programming problem. We refer the reader to [5] for the details.

2.3. I-Measure. In this subsection, we give a review of the main results regarding I-Measure. For a detailed discussion of the theory, we refer the reader to [3] (also see the tutorial [2]). Further results on I-Measure can be found in [4] and [15].

Let $X_i, i \in \mathcal{N} = \{1, \ldots, n\}$ be n jointly distributed random variables, and $\tilde{X}$ be a set variable corresponding to a random variable X. Define the universal set Ω to be $\cup_{i\in\mathcal{N}}\tilde{X}_i$ and let $\mathcal{F}$ be the σ-field generated by $\{\tilde{X}_i, i \in \mathcal{N}\}$. The atoms of $\mathcal{F}$ have the form $\cap_{i\in\mathcal{N}}Y_i$, where Y_i is either $\tilde{X}_i$ or $\tilde{X}_i^c$. Let $\mathcal{A} \subset \mathcal{F}$ be the set of all the atoms of $\mathcal{F}$ except for $\cap_{i\in\mathcal{N}}\tilde{X}_i^c$, which is equal to the empty set by construction because

$$\bigcap_{i\in\mathcal{N}} \tilde{X}_i^c = \left(\bigcup_{i\in\mathcal{N}} \tilde{X}_i\right)^c = \Omega^c = \phi.$$

Note that $|\mathcal{A}| = 2^n - 1$. In this subsection, when we refer to an atom of $\mathcal{F}$, we always mean an atom of $\mathcal{F}$ in $\mathcal{A}$.

To simplify notations, we will use X_U to denote $(X_i, i \in U)$ and $\tilde{X}_U$ to denote $\cup_{i\in U}\tilde{X}_i$ for any $U \subset \mathcal{N}$. It was shown in [3] that there exists a unique *signed* measure μ^* on $\mathcal{F}$, called the I-Measure, which is consistent with all Shannon's information measures via the following formal substitution of symbols:

$$\begin{aligned} H/I &\to \mu^* \\ , &\to \cup \\ ; &\to \cap \\ | &\to - \end{aligned}$$

($X - Y = X \cap Y^c$), i.e., for any (not necessarily disjoint) $U, U', U'' \subset \mathcal{N}$:

$$\mu^*(\tilde{X}_U \cap \tilde{X}_{U'} - \tilde{X}_{U''}) = I(X_U; X_{U'}|X_{U''}). \tag{2.3}$$

When $U'' = \phi$, we interpret (2.3) as

$$\mu^*(\tilde{X}_U \cap \tilde{X}_{U'}) = I(X_U; X_{U'}).$$

When $U = U'$, (2.3) becomes

$$\mu^*(\tilde{X}_U - \tilde{X}_{U''}) = H(X_U|X_{U''}).$$

When $U = U'$ and $U'' = \phi$, (2.3) becomes

$$\mu^*(\tilde{X}_U) = H(X_U). \tag{2.4}$$

Thus (2.3) covers all the cases of Shannon's information measures.

We now show how μ^* is defined. Let

$$\mathcal{D} = \left\{ D \in \mathcal{F} \mid D = \cup_{i \in U} \tilde{X}_i \text{ for some } U \subset \mathcal{N},\ U \neq \phi \right\}$$

be the set consisting of the unions formed from $\tilde{X}_i, 1 \leq i \leq n$. Note that $|\mathcal{D}| = |\mathcal{A}| = 2^n - 1$. Let $k = 2^n - 1$. Define arbitrary one-to-one mappings

$$\rho : \{1, 2, \ldots, k\} \to \mathcal{A}$$

$$\sigma : \{1, 2, \ldots, k\} \to \mathcal{D},$$

and let

$$\mathbf{u} = [\ u_1\ \cdots\ u_k\ ]^T$$

$$\mathbf{v} = [\ v_1\ \cdots\ v_k\ ]^T$$

where $u_j = \mu^*(\rho(j))$ and $v_l = \mu^*(\sigma(l))$ for $1 \leq j, l \leq k$. Note that $u_j, 1 \leq j \leq k$ are the values of μ^* on all the atoms of $\mathcal{F}$, and $v_l, 1 \leq l \leq k$ are the values of μ^* on all the unions formed from $\tilde{X}_i, 1 \leq i \leq n$, or equivalently, all the joint entropies involving the random variables $X_1, \ldots, X_n$ by (2.4). Then

$$\mathbf{v} = \mathbf{C}\mathbf{u},$$

where $\mathbf{C} = [c_{lj}]$ is a unique $k \times k$ matrix (independent of μ^*) with

$$c_{lj} = \begin{cases} 1 & \text{if } \rho(j) \subset \sigma(l) \\ 0 & \text{if } \rho(j) \not\subset \sigma(l). \end{cases}$$

An important characteristic of $\mathbf{C}$ is that it is invertible [3], so we can write

$$\mathbf{u} = \mathbf{C}^{-1}\mathbf{v}.$$

In other words, μ^* is completely specified by the set of values $\mu^*(D)$, $D \in \mathcal{D}$, namely all the joint entropies involving $X_1, \ldots, X_n$, and by virtue of (2.4), μ^* is the unique measure on $\mathcal{F}$ which is consistent with all Shannon's information measures. Note that μ^* in general is not nonnegative. However, if $X_1, \ldots, X_n$ form a Markov chain, μ^* is always nonnegative [4].

To conclude, the theory of I-Measure enables the use of the language and the rich set of tools in set theory to study the structure of Shannon's information measures. As a consequence of this theory, the information diagram was introduced as a tool to visualize the relationship among information measures [3]. An information diagram is a special case of a Venn diagram with $\Omega = \cup_{i \in \mathcal{N}} \tilde{X}_i$. Figure 2.4 shows the information diagram for random variables X_1, X_2, X_3. Examples of applications of information diagrams can be found in [3], [4], [2], [11] and [13].

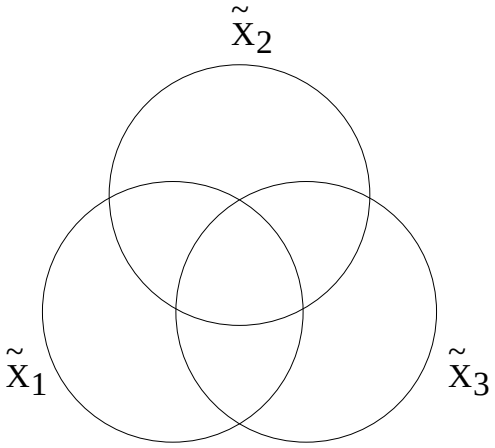

FIG. 2.4. *The information diagram for random variables X_1, X_2 and X_3.*

3. A Class of non-Shannon-type Information Inequalities. Recall from Section 2.1 that for 4 random variables, an unconditional information inequality $f(h) \geq 0$ (which always holds) is of non-Shannon-type if and only if

$$\Gamma_4 \not\subset \{h \in \mathbb{R}^{15} : f(h) \geq 0\}.$$

Likewise, under the constraint of a linear subspace Φ on the entropies (cf. (2.1)), an information inequality $f(h) \geq 0$ is of non-Shannon-type if and only if

$$(\Gamma_4 \cap \Phi) \not\subset \{h \in \mathbb{R}^k : f(h) \geq 0\}.$$

Since ZY98 (written as $g(h) \geq 0$) always holds, it remains valid when certain linear constraints on the entropies are imposed. The question is whether under these additional constraints ZY98 continues to be of non-Shannon-type. As we will see shortly, this would be the case if the additional constraints are chosen carefully.

To fix ideas, we now give an example for which ZY98 is a Shannon-type inequality when a certain linear constraint on the entropies is imposed. Suppose $I(X_3; X_4) = 0$. Then the left hand side of ZY98 as displayed in Section 1 becomes 0, and the inequality is trivially implied by the basic inequalities because all the terms on the right hand side are Shannon's information measures.

In the course of proving that ZY98 is of non-Shannon-type, it was shown in [7] that there exists an $h^1 \in \Gamma_4$ which does not satisfy ZY98, where h^1 is defined by

$$h^1_1 = h^1_2 = h^1_3 = h^1_4 = 2a$$

$$h^1_{12} = 4a, h^1_{13} = h^1_{14} = h^1_{23} = h^1_{24} = h^1_{34} = 3a$$

$$h^1_{123} = h^1_{124} = h^1_{134} = h^1_{234} = h^1_{1234} = 4a$$

with $a > 0$. Since ZY98 is satisfied by all entropy functions, h^1 is not an entropy function, or $h^1 \notin \Gamma^*_4$. From the theory of I-Measure, we can obtain the set-theoretic

structure of h^1, which is shown in Figure 3.1. (Here we use the measure in Figure 5 to illustrate the set-theoretic structure of h^1, but this measure is actually not a valid I-Measure because h^1 is not an entropy function.) It is easy to verify from Figure 3.1 that h^1 lies in exactly 14 hyperplanes defining the boundary of Γ_4, which correspond to setting the following 14 Shannon's measures to 0:

$$I(X_1;X_2), I(X_1;X_2|X_3), I(X_1;X_2|X_4), I(X_1;X_3|X_4), I(X_1;X_4|X_3),$$

$$I(X_2;X_3|X_4), I(X_2;X_4|X_3), I(X_3;X_4|X_1), I(X_3;X_4|X_2), I(X_3;X_4|X_1,X_2),$$

$$H(X_1|X_2,X_3,X_4), H(X_2|X_1,X_3,X_4), H(X_3|X_1,X_2,X_4), H(X_4|X_1,X_2,X_3).$$

Since Γ_4 is in $\mathbb{R}^{15}$, h^1 is along an extreme direction of Γ_4.

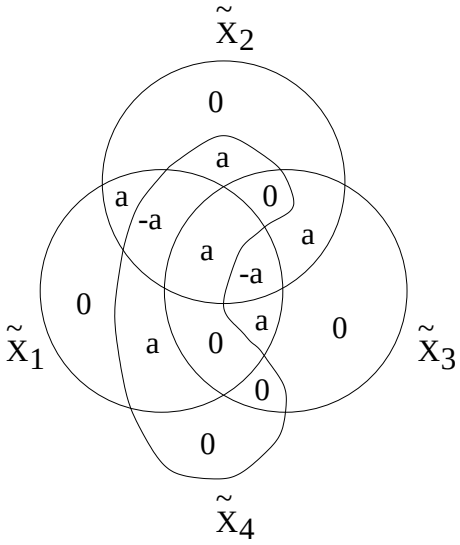

FIG. 3.1. *The set-theoretic structure of* h^1.

Now for any linear subspace Φ in $\mathbb{R}^{15}$ containing h^1, we have

$$h^1 \in \Gamma_4 \cap \Phi$$

and h^1 does not satisfy ZY98. Therefore,

$$(\Gamma_4 \cap \Phi) \not\subset \{h \in \mathbb{R}^k : g(h) \geq 0\}.$$

This means that ZY98 is a non-Shannon-type inequality under the constraint Φ. From the above, we see that Φ can be the intersection of any nonempty subset of the 14 hyperplanes containing h^1. Thus ZY98 is a non-Shannon-type inequality conditioning on any nonempty subset of the above 14 Shannon's measures equal 0. Hence, ZY98 implies a class of $2^{14}-1$ conditional non-Shannon-type inequalities.

4. Applications. Although non-Shannon-type inequalities originated in information theory, so far there has not been any application of these inequalities in information theory problems. In this section, we give two examples of application of the class of non-Shannon-type inequalities implied by ZY98. The results obtained in these two examples are not possible otherwise.

EXAMPLE 1. *Consider a fault-tolerant distributed database system consisting of random variables X_1, X_2, X_3, X_4 such that any three random variables can recover the remaining one, i.e.,*

$$H(X_i|X_j, j \neq i) = 0, \quad 1 \leq i, j \leq 4. \tag{4.1}$$

We are interested in the set of all entropy functions subject to these constraints, denoted by Υ, which characterizes the amount of joint information which can possibly be stored in such a database system. The set Υ is given as the intersection of Γ_4^* and the 4 hyperplanes corresponding to the 4 constraints in (4.1). Since each constraint in (4.1) is one of the 14 constraints specified in the last section, ZY98 is a non-Shannon-type inequality under the constraints in (4.1).

When ZY98 is written in terms of unconditional joint entropies, it becomes

$$\begin{array}{rcl} H(X_1)+H(X_2)+4H(X_3) & & 4H(X_1,X_3)+4H(X_1,X_4) \\ +4H(X_4)+2H(X_1,X_2) & \leq & +4H(X_2,X_3)+4H(X_2,X_4) \\ +5H(X_1,X_3,X_4)+5H(X_2,X_3,X_4) & & +6H(X_3,X_4). \end{array}$$

Upon invoking $H(X_1|X_2,X_3,X_4) = H(X_2|X_1,X_3,X_4) = 0$ from (4.1) so that

$$H(X_1,X_3,X_4) = H(X_2,X_3,X_4) = H(X_1,X_2,X_3,X_4),$$

ZY98 becomes

$$\begin{array}{rcl} H(X_1)+H(X_2)+4H(X_3) & & 4H(X_1,X_3)+4H(X_1,X_4) \\ +4H(X_4)+2H(X_1,X_2) & \leq & +4H(X_2,X_3)+4H(X_2,X_4) \\ +10H(X_1,X_2,X_3,X_4) & & +6H(X_3,X_4). \end{array}$$

Since this inequality is symmetrical in X_1 and X_2 and in X_3 and X_4, by permuting the indices, we can obtain five other distinct inequalities. These 6 inequalities together with the basic inequalities give a tighter bound on Υ than the basic inequalities alone.

ZY98 conditioning on (4.1) cannot be proved by ITIP. This is consistent with our claim that ZY98 is of non-Shannon-type conditioning on (4.1).

EXAMPLE 2. *Consider 4 random variables X_1, X_2, X_3, X_4 such that $X_3 - (X_1, X_2) - X_4$ form a Markov chain. This Markov condition is equivalent to*

$$I(X_3; X_4|X_1, X_2) = 0$$

which is one of the 14 constraints specified in the last section. Therefore, ZY98 is a non-Shannon-type inequality under this condition.

It can be proved by invoking the basic inequalities (using ITIP [18]) that

$$I(X_3; X_4) \leq I(X_3; X_4|X_1) + I(X_3; X_4|X_2) + 0.5I(X_1; X_2)$$

$$+cI(X_1; X_3, X_4) + (1-c)I(X_2; X_3, X_4)$$

where $0.25 \leq c \leq 0.75$ *(this is the best possible). However, from ZY98, the last two terms above can be sharpened to* $0.25I(X_1; X_3, X_4) + 0.25I(X_2; X_3, X_4)$.

The Markov chain $X_3 - (X_1, X_2) - X_4$ *arises in many communication situations. As an example, consider a person listening to an audio source. Then the situation can be modeled by this Markov chain with* X_3 *being the sound wave generated at the source,* X_1 *and* X_2 *being the sound waves received at the two ear drums, and* X_4 *being the nerve impulses which eventually arrive at the brain. The inequality ZY98 gives a tighter upper bound on* $I(X_3; X_4)$ *(tighter than what can be implied by the basic inequalities), which appears to be fundamental. This bound may be useful in proving certain converse coding theorems in multiterminal information theory.*

There is some resemblance between the conditional form of ZY98 discussed in this example and the Data Processing Theorem, but there does not seem to be any direct relation between them.

5. Concluding Remarks. Owing to the recent discovery of a few so-called non-Shannon-type information inequalities, it is now known that there are laws in information theory beyond those laid down by Shannon. Since there exist non-Shannon-type inequalities for as few as 4 random variables, it is believed that there are many more such inequalities yet to be discovered. In this paper, we have derived a class of non-Shannon-type inequalities from the inequality discovered by the authors in 1998, and we have shown possible application of these inequalities in information theory problems. The results thus obtained are not possible otherwise.

It is straightforward to see that each of the conditional non-Shannon-type inequalities reported in [6] and [12] implies a class of conditional non-Shannon-type inequalities by means of a slight modification of the arguments in this paper. It is conceivable that some of these inequalities have applications in information theory problems.

Our results have shed some light on the role of non-Shannon-type inequalities in information theory. Further investigation along this line may lead to new territories in information theory. In particular, the solutions of certain open problems in multiterminal information theory may be made possible by some non-Shannon-type inequalities.

Appendix A. The proof for the equivalence of the Polymatriod Axioms and the Basic Inequalities. We first show that the polymatroid axioms imply the basic inequalities. Obviously, (P1) and (P2) imply all entropies are nonnegative. For (P2), by letting $\gamma = \beta \backslash \alpha$, we have $H_\alpha \leq H_{\alpha \cup \gamma}$, or $H(X_\gamma | X_\alpha) \geq 0$. Here, γ and α are non-overlapping subsets of $\mathcal{N}$. For (P3), by letting $\gamma = \beta \backslash \alpha$, $\delta = \alpha \cap \beta$, and $\sigma = \alpha \backslash \beta$, we have $H_{\sigma \cup \delta} + H_{\gamma \cup \delta} \geq H_\delta + H_{\sigma \cup \delta \cup \gamma}$, or $I(X_\sigma; X_\gamma | X_\delta) \geq 0$. Again, σ, δ, and γ are non-overlapping subsets of $\mathcal{N}$. When $\delta = \phi$, from (P3), we have $I(X_\sigma; X_\gamma) \geq 0$. Thus, (P1)-(P3) imply that all entropies are nonnegative, and that all conditional entropies, mutual informations, and conditional mutual informations are nonnegative provided that the subsets of random variables involved do not overlap. However, for any conditional entropy, mutual information, or conditional mutual information, even if the subsets of random variables involved are overlapping, it can always be

written as a linear combination with nonnegative coefficients of entropies, conditional entropies, mutual informations, and conditional mutual informations for which the subsets of random variables involved in any of the latter three types of information measures do not overlap. For example, $I(X_1, X_2; X_1, X_3, X_5|X_3, X_4)$ can be written as $H(X_1|X_3, X_4) + I(X_1, X_2; X_5|X_1, X_3, X_4)$. This shows that (P1)-(P3) imply the basic inequalities.

The converse is trivial and its proof is omitted.

REFERENCES

[1] N. Pippenger, *What are the laws of information theory?* 1986 Specific Problems on Communication and Computation Conference.

[2] R. W. Yeung, *A tutorial on entropy, information inequality and I-Measure,* http://www.ie.cuhk.edu.hk/whyeung.

[3] R. W. Yeung, *A new outlook on Shannon's information measures,* IEEE Trans. Inform. Theory, 37(1991), pp. 466–474.

[4] T. Kawabata and R. W. Yeung, *The structure of the I-Measure of a Markov chain,* IEEE Transactions on Information Theory, 38(1992), pp. 1146–1149.

[5] R. W. Yeung, *A framework for linear information inequalities,* 43(1997), pp. 1924–1934.

[6] Z. Zhang and R. W. Yeung, *A non-Shannon-type conditional information inequality,* IEEE Trans. Inform. Theory, 43(1997), pp. 1982–1986.

[7] Z. Zhang and R. W. Yeung, *On Characterization of entropy function via information inequalities,* IEEE Trans. Inform. Theory, 44(1998), pp. 1440–1452.

[8] R. W. Yeung and Z. Zhang, *Distributed source coding for satellite communications,* IEEE Transactions on Information Theory, 45(1999), pp.1111–1120.

[9] F. Matúš, *Probabilistic conditional independence structures and matroid theory: Background,* Int. J. General Systems, 22(1994), pp. 185–196.

[10] F. Matúš and M. Studený, *Conditional independences among four random variables I,* Combinatorics, Probability and Computing, 4:3(1995), pp. 267–278.

[11] R. W. Yeung, *Multilevel diversity coding with distortion,* IEEE Transactions on Information Theory, Vol. IT-41(1995), pp. 412–422.

[12] F. Matúš, *Conditional independence among four random variables III: Final conclusion,* Combinatoric, Probability and Computing, 8(1999), pp. 269–276.

[13] R. W. Yeung and Z. Zhang, *Distributed source coding for satellite communications,* IEEE Trans. Inform. Theory, 45(1999), pp. 1111–1120.

[14] H.-L. Chan and R. W. Yeung, *Information Inequalities, Conditional Independence and Group,* Proceedings of the 2000 Conference on Information Sciences and Systems, Princeton University, March 15-17, 2000, pp. WP6:23–26.

[15] R. W. Yeung, T. T. Lee, and Z. Ye, *An information-theoretic characterization of Markov random fields and its applications,* in 1998 IEEE International Symposium on Information Theory, MIT, Cambridge, MA, Aug 16-21, 1998.

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

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, New York, 1991.

[18] R. W. Yeung and Y.-O. Yan, *Information-Theoretic Inequality Prover,* http://www.ie.cuhk.edu.hk/~ITIP, or http://www.itsoc.org (mirror site).