# On Witsenhausen's zero-error rate for multiple sources

**Gábor Simonyi**[1]
Alfréd Rényi Institute of Mathematics,
Hungarian Academy of Sciences,
1364 Budapest, POB 127, Hungary.
simonyi@renyi.hu

October 1, 2001

[1]Research partially supported by the Hungarian Foundation for Scientific Research Grant (OTKA) Nos. F023442, T029255, and T032323.

**Abstract**

We investigate the problem of minimum rate zero-error source coding when there are several decoding terminals having different side information about the central source variable and each of them should decode in an error-free manner. For one decoder this problem was considered by Witsenhausen. The Witsenhausen rate of the investigated multiple source is the asymptotically achievable minimum rate. We prove that the Witsenhausen rate of a multiple source equals the Witsenhausen rate of its weakest element. The proof relies on a powerful result of Gargano, Körner, and Vaccaro about the zero-error capacity of the compound channel.

# 1 Introduction

Let $X, Y^{(1)}, Y^{(2)}, \ldots, Y^{(k)}$ be $k+1$ discrete random variables. Consider $X$ as a 'central' variable available for a transmitter $T$ and the $Y^{(i)}$'s as side information available for different stations $S^{(i)}, i = 1, 2, \ldots, k$, that are located at different places. The joint distribution is known for $X$ and $Y^{(i)}$ for every $i$. The task is that $T$ broadcast a message received by all $S^{(i)}$'s in such a way that learning this message all $S^{(i)}$'s should be able to determine $X$ in an error-free manner. The question is the minimum number of bits that should be used for this per transmission if block coding is allowed. This problem is considered for $k = 1$ by Witsenhausen in [20]. He translated the problem to a graph theoretic one and showed that block coding can indeed help in decreasing the (per transmission) number of possible messages that should be used. The optimal number of bits to be sent per transmission defines a graph parameter that is called *Witsenhausen's zero-error rate* in [1]. (We will write simply *Witsenhausen rate* in the sequel.) In this paper we define the Witsenhausen rate of a family of graphs. Our main result is that the Witsenhausen rate of a family of graphs equals its obvious lower bound: the largest Witsenhausen rate of the graphs in the family. This will easily follow from a powerful result of Gargano, Körner, and Vaccaro [8].

# 2 The graph theory model

For each $i = 1, 2, \ldots, k$ we define the following graph $G_i$. The vertex set $V(G_i) = \mathcal{X}$ is the support set of the variable $X$ for every $i$. Two elements, $a$ and $b$ of $\mathcal{X}$ form an edge in $G_i$ if and only if there exists some possible value $c$ of the variable $Y^{(i)}$ that is jointly possible with both $a$ and $b$, i.e., $Prob(c, a)Prob(c, b) > 0$. It is already explained in [20] that the minimum number of bits to be sent by $T$ to (one) $S^{(i)}$ for making it learn $X$ (for one instance) in an error-free manner is $\log_2 \chi(G_i)$, where $\chi(F)$ denotes the chromatic number of graph $F$. Indeed, if $T$ would use less bits, than there were some two elements of $\mathcal{X}$ that form an edge in $G_i$ and still $T$ sends the same message when one or the other appears as the actual value of $X$. Since they form an edge there is some possible value $c$ of $Y^{(i)}$ that is jointly possible with both, thus $S^{(i)}$ would not be able to decide which of them occured if it had $c$ as side information. (As in [20], we use the assumption, that the side information $Y^{(i)}$ is not available at $T$.) On the other hand, if a proper coloring of $G_i$ is given, then if $T$ sends the color of $X$ this will make $S^{(i)}$ learn $X$ using the side-information contained by $Y^{(i)}$.

All subsequent logarithms are on base two.

If block coding is allowed then the minimum number of bits to be transmitted to $S^{(i)}$ (not caring about the other $S^{(j)}$'s for the moment) by $T$ after observing the $n$-fold variable $(X_1, X_2, \ldots, X_n)$ will be $\log \chi(G_i^n)$ where $G_i^n$ is an appropriately defined power of the graph $G_i$.

**Definition 1** *Let $G = (V, E)$ be a graph. The $n^{th}$ normal (in [1] it is called 'anded') power of $G$ is the graph $G^n$ defined as follows. $V(G^n) = V^n$ and*

$$E(G^n) = \{\{\mathbf{u}, \mathbf{v}\} : \mathbf{u} \neq \mathbf{v}, \forall i \; x_i = y_i \text{ or } \{x_i, y_i\} \in E(G)\}.$$

*That is the vertices of $G^n$ are the $n$-length sequences over $V$ and two are adjacent iff they are adjacent at every coordinate where they are not equal.*

It is easy to see that two $n$-length sequences over $\mathcal{X}$ are jointly possible with some $n$-fold outcome $(Y_1^{(i)}, Y_2^{(i)}, \ldots, Y_n^{(i)})$ iff they are adjacent in $G_i^n$ (our sources are stationary and memoryless). Thus the previous argument gives that if we cared only about $S^{(i)}$ then $T$ should transmit one of $\chi(G_i^n)$ messages for making $S^{(i)}$ learn $(X_1, X_2, \ldots, X_n)$. Thus, in case $k = 1$ (and denoting $G_1$ by $G$) the value of interest is the Witsenhausen rate of $G$ defined as

$$R(G) = \lim_{n\to\infty} \frac{1}{n} \log(\chi(G^n)).$$

In our problem (when $k > 1$) the message sent by $T$ should be such that learning it should be enough for each $S^{(i)}$ to determine $X$ without error. Thus $T$ cannot send the same message for two $n$-fold outcome of the variable $X$ if they, as vertices of the graphs $G_i^n$, are adjacent *in any* $G_i^n$. On the other hand, if we colour the elements of $V^n = \mathcal{X}^n$ in such a way that elements adjacent in any $G_i^n$ get different color, then transmitting the color of the actual $(X_1, X_2, \ldots, X_n)$ will make all $S^{(i)}$'s able to determine $(X_1, X_2, \ldots, X_n)$. This justifies the following definition.

**Definition 2** *Let $\mathcal{G} = (G_1, \ldots, G_k)$ be a family of graphs all of which have the same vertex set $V$. The Witsenhausen rate of the family $\mathcal{G}$ is defined by*

$$R(\mathcal{G}) = \lim_{n\to\infty} \frac{1}{n} \log(\chi(\cup_i G_i^n)),$$

*where $\cup_i G_i^n$ is meant to be the graph on the common vertex set $V^n$ of the $G_i^n$ with edge set $\cup_i E(G_i^n)$.*

It is obvious from the definition that $R(\mathcal{G}) \geq \max\{R(G_i)\}$. Our main result is that this trivial estimation is sharp.

**Theorem 1** *If $\mathcal{G} = (G_1, \ldots, G_k)$ is a family of graphs on the same vertex set, then*

$$R(\mathcal{G}) = \max\{R(G_i)\}.$$

To appreciate the above statement consider the following.

*Example:* Let $|V| = v$ and $\cup_i G_i = K_v$, i.e, the complete graph an $v$ vertices such that each $G_i$ is bipartite. (This needs $k \geq \log v$.) Now for $n = 1$ we would be obliged to use

$\log v$ bits to make sure that each $S^{(i)}$ can decode the outcome of $X$ correctly. However, with block coding, the above theorem states that roughly one bit per source outcome is enough if we let $n$ go to infinity.

For proving Theorem 1 we have to introduce some other notions. This is done in the next section.

# 3 Probabilistic graph invariants

The proof of our theorem relies on a result that determines the zero-error capacity of a compound channel. Here we give our definitions already in graph terms, the translation is explained in detail in [3] where these investigations started, in [8], where the powerful result we are going to use was obtained, and also in the survey article [12].

**Definition 3** *Let $\mathcal{G} = (G_1, \ldots, G_k)$ be a family of graphs all of which have the same vertex set $V$. The capacity of the family $\mathcal{G}$ is defined by*

$$C(\mathcal{G}) = \lim_{n\to\infty} \frac{1}{n} \log(\alpha(\cup_i G_i^n)),$$

*where $\alpha$ stands for the independence number (size of largest edgeless subgraph) of a graph. If $\mathcal{G} = \{G\}$ we write $C(G)$ instead of $C(\mathcal{G})$.*

*Remark:* If $|\mathcal{G}| = 1$ then $C(\mathcal{G})$ becomes equivalent to the Shannon capacity of the graph $G$. It is not hard to see that the value $C(\mathcal{G})$ represents the zero-error capacity of the compound channel the individual channels in which are described by the graphs in the family. For a more detailed explanation, see [3], [8], [12]. We have to warn the reader, however, that several papers, including the ones just cited, use a complementary language and define $C(G)$ as our $C(\bar{G})$, while $C(\mathcal{G})$ is also defined via cliques instead of independent sets. The language we use here is the more traditional one (cf. [16], [14]), although the earlier cited papers have their good reason to do differently. (It has to do with a generalization to oriented graphs that we will not need here.)

It is obvious that $C(\mathcal{G}) \leq \min_i C(G_i)$. An easy but somewhat more sophisticated upper bound is obtained in [3]. This needs the following notion of *capacity within a given type* introduced by Csiszár and Körner [6]. First we need the concept of $(P, \epsilon)$-typical sequences, cf. [5].

**Definition 4** *Let $V$ be a finite set, $P$ a probability distribution on $V$, and $\epsilon > 0$. A sequence $\mathbf{x}$ in $V^n$ is said to be $(P, \epsilon)$-typical if for every $a \in V$ we have $|\frac{1}{n}N(a|\mathbf{x}) - P(a)| < \epsilon$, where $N(a|\mathbf{x}) = |\{i : x_i = a\}|$.*

**Definition 5** *Given graph $G$, probability distribution $P$, and $\epsilon > 0$, the graph $G^n_{P,\epsilon}$ is the graph induced in $G^n$ by the $(P, \epsilon)$-typical sequences.*

**Definition 6** *The Shannon capacity $C(G,P)$ of a graph $G$ within a given type $P$ is the value*

$$C(G,P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \alpha(G^n_{P,\epsilon}).$$

*In a similar manner we write*

$$C(\mathcal{G},P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \alpha(\cup_i G^n_{i;P,\epsilon}).$$

The upper bound shown in [3] states $C(\mathcal{G}) \leq \max_P \min_i C(G_i,P)$. Gargano, Körner, and Vaccaro proved the surprising result that this bound is sharp. This is a corollary of their more general result that we will also need.

**Theorem** (Gargano, Körner, Vaccaro [8]): *For any family of graphs $\mathcal{G} = \{G_1,\ldots,G_k\}$ and any probability distribution $P$ on the common vertex set of the $G_i$'s we have*

$$C(\mathcal{G},P) = \min_i C(G_i,P).$$

*Remark:* The exact statement proven in [8] (cf. also [7] for an important special case) is that $C(\mathcal{G}_1 \cup \mathcal{G}_2, P) = \min\{C(\mathcal{G}_1,P), C(\mathcal{G}_2,P)\}$ holds for any two graph families $\mathcal{G}_1$ and $\mathcal{G}_2$ with common vertex set. This easily implies the above by setting iteratively $\mathcal{G}_1 = \{G_1,G_2,\ldots,G_i\}$ and $\mathcal{G}_2 = \{G_{i+1}\}$ for all $i = 1,2,\ldots,k-1$.

To relate $C(G)$ and $R(G)$ we will use the "within a type version" of $R(G)$ which was already introduced in [11] by Körner and Longo in a different context under the name *complementary graph entropy*. (We use the name *co-entropy* as in [18] and [19].) Marton [15] investigated this functional further, while recent interest in it also occured in [9].

**Definition 7** *The co-entropy $\bar{H}(G,P)$ of a graph $G$ within a given type $P$ is the value*

$$\bar{H}(G,P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \chi(G^n_{P,\epsilon}).$$

*In a similar manner we write*

$$\bar{H}(\mathcal{G},P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \chi(\cup_i G^n_{i;P,\epsilon}).$$

*Remark:* It would be appropriate to denote the within a type version of $R(G)$ by $R(G,P)$. Here we keep the $\bar{H}(G,P)$ notation only to emphasize that this is not a new concept.

*Remark:* We use the name *co-entropy* to distinguish the above value from the related notion of graph entropy introduced by Körner [10]. For a detailed account of their relation see [15] or [18], [19]. For further relations of graph entropy and source coding, cf. also [2].

In [15] the following relation is proven.

**Lemma** (Marton [15]): *For any graph $G$ and probability distribution $P$ on its vertex set*

$$\bar{H}(G,P) = H(P) - C(G,P)$$

*where $H(P)$ is the Shannon entropy of the distribution $P$.*

We also need the following more general statement the proof of which is exactly the same as that of Marton's Lemma.

**Lemma 1** *For any family of graphs and any probability distribution on their common vertex set $V$ one has*

$$\bar{H}(\mathcal{G},P) = H(P) - C(\mathcal{G},P).$$

We sketch the proof of this lemma for the sake of completeness. Setting $\mathcal{G} = \{G\}$ it also implies Marton's result. The following lemma of Lovász [13] is needed.

**Lemma 2** *For any graph $G$*

$$\chi(G) \leq \chi^*(G)(1 + \ln \alpha(G)),$$

*where $\chi^*(G)$ is the fractional chromatic number of graph $G$.*

*Remark:* Lovász's result is formulated in a more general setting for the covering numbers of hypergraphs. The above statement is a straightforward corollary of that. For basic facts about the fractional chromatic number we refer the reader to [17]. One such fact we need is that for a vertex-transitive graph $G$ one always has $\chi^*(G) = \frac{|V(G)|}{\alpha(G)}$ (see [17] Proposition 3.1.1).

**Sketch of proof of Lemma 1.** Consider a sequence $P_n$ of probability distributions that converges to $P$ in the sense that $\forall \epsilon > 0$ $\exists n_0$ such that $n \geq n_0$ implies $\forall a \in V$ : $|P_n(a) - P(a)| < \epsilon$. As the number of possible types of an $n$-length sequence is only a polynomial function of $n$ (cf. Lemma 2.2 of Chapter 1 in [5]) we can write

$$\bar{H}(\mathcal{G},P) = \lim_{n\to\infty} \frac{1}{n} \log \chi(\cup_i G^n_{i;P_n,0}).$$

Since sequences of the same type are all permutations of each other, one easily sees that the graph $\cup_i G^n_{i;P_n,0}$ is vertex transitive for any $n$. Thus we can continue by

$$\lim_{n\to\infty} \frac{1}{n} \log \chi(\cup_i G^n_{i;P_n,0}) \leq \lim_{n\to\infty} \frac{1}{n} \log \frac{|V(\cup_i G^n_{i;P_n,0})|}{\alpha(\cup_i G^n_{i;P_n,0})} (1 + \ln \alpha(\cup_i G^n_{i;P_n,0})) =$$

$$H(P) - C(\mathcal{G},P) + \lim_{n\to\infty} \frac{1}{n} \log(1 + \ln \alpha(\cup_i G^n_{i;P_n,0})) = H(P) - C(\mathcal{G},P).$$

The opposite inequality is obvious as $\chi(F) \geq \frac{|V(F)|}{\alpha(F)}$ is trivially true for any graph $F$ and applying it for $F = \cup_i G^n_{i;P_n,0}$ we get what we need.

□

# 4 Proof of Theorem 1

For proving Theorem 1 we first need an easy lemma.

**Lemma 3**

$$R(\mathcal{G}) = \max_P \bar{H}(\mathcal{G}, P).$$

**Proof.** Using again the Type Counting Lemma (Lemma 2.2 on page 29) from Csiszár and Körner's book [5] we get that

$$\chi(\cup_i G_i^n) \leq (n+1)^{|V|} \max_P \chi(\cup_i G_{i;P,0}^n)$$

where the maximization is meant over those $P$'s that can be exact types of sequences of length $n$. (Equivalently, we can just think of $G_{i;P,0}^n$ as a graph with no vertices for other $P$'s). Since $\chi(\cup_i G_i^n \geq \chi(\cup_i G_{i;P,0}^n$ obviously holds for any $P$, taking the logarithm and let $n$ go to infinity in the earlier inequality we get the desired result.

□

**Proof of Theorem 1.** By the previous two lemmas and the Gargano-Körner-Vaccaro theorem we have

$$R(\mathcal{G}) = \max_P \bar{H}(\mathcal{G}, P) = \max_P (H(P) - C(\mathcal{G}, P)) =$$

$$= \max_P (H(P) - \min_{G_i \in \mathcal{G}} C(G_i, P)) = \max_P \max_{G_i} (H(P) - C(G_i, P)) =$$

$$= \max_{G_i} \max_P \bar{H}(G_i, P) = \max_{G_i} R(G_i)$$

giving the desired result.

□

*Remark:* It seems worth noting that while the proof of Theorem 1 needed separate investigation of the different types $P$ the statement itself does not contain any reference to types. This is not so in the original Gargano-Körner-Vaccaro result. Though the reason of this is a very simple technical difference (namely that the chromatic number is defined as a minimum, while the clique number and the independence number are appropriate maximums), we feel that this phenomenon makes Theorem 1 another good example of a result that demonstrates the power of the method of types, cf. [4].

# References


[1] N. Alon and A. Orlitsky, Repeated communication and Ramsey graphs, *IEEE Trans. Inform. Theory*, **41** (1995), 1276–1289.

[2] N. Alon and A. Orlitsky, Source coding and graph entropies, *IEEE Trans. Inform. Theory*, **42** (1996), 1329–1339.

[3] G. Cohen, J. Körner, and G. Simonyi, Zero-error capacities and very different sequences, in: *Sequences: combinatorics, compression, security and transmission*, R. M. Capocelli ed., Springer-Verlag, 144–155.

[4] I. Csiszár, The method of types, *IEEE Trans. Inform. Theory*, Vol. **44**, No. 6 (October 1998, commemorative issue), 2505–2523.

[5] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1982.

[6] I. Csiszár and J. Körner, On the capacity of the arbitrarily varying channel for maximum probability of error, *Z. Warsch. Verw. Gebiete* **57** (1981), 87–101.

[7] L. Gargano, J. Körner, and U. Vaccaro, Sperner capacities, *Graphs Combin.* **9** (1993), 31–46.

[8] L. Gargano, J. Körner, and U. Vaccaro, Capacities: from information theory to extremal set theory, *J. Combin. Theory Ser. A*, **68** (1994), 296–316.

[9] P. Koulgi, E. Tuncel, S. Regunathan, and K. Rose, On zero-error source coding with decoder side information, submitted to *IEEE Trans. Inform. Theory.*

[10] J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs, in: *Transactions of the 6th Prague Conference on Information Theory, etc.*, 1971, Academia, Prague, (1973), 411–425.

[11] J. Körner and G. Longo, Two-step encoding of finite memoryless sources, *IEEE Trans. Inform. Theory*, **19** (1973), 778-782.

[12] J. Körner and A. Orlitsky: Zero-error information theory, *IEEE Trans. Inform. Theory*, Vol. **44**, No. 6 (October 1998, commemorative issue), 2207–2229.

[13] L. Lovász, On the ratio of optimal integer and fractional covers, *Discrete Math.*, **13** (1975), 383–390.

[14] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **25** (1979), 1–7.

[15] K. Marton, On the Shannon capacity of probabilistic graphs, *J. Combin. Theory Ser. B*, **57** (1993), 183–195.

[16] C. E. Shannon, The zero-error capacity of a noisy channel, *IRE Transactions on Information Theory*, **2** (1956), 8–19. (Reprinted in: *Key papers in the Development of Information Theory*, D. Slepian ed., IEEE Press, New York 1974.)

[17] E. R. Scheinerman, D. H. Ullman, *Fractional Graph Theory*, Wiley, New York, 1997.

[18] G. Simonyi: Graph entropy: a survey, in: *Combinatorial Optimization*, (W. Cook, L. Lovász, and P. Seymour eds.), DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 20, AMS, 1995, 399-441.

[19] G. Simonyi, Perfect Graphs and Graph Entropy. An Updated Survey in: *Perfect Graphs* (J. Ramirez-Afonsín, B. Reed, eds.), Wiley, to appear.

[20] H. S. Witsenhausen, The zero-error side-information problem and chromatic numbers, *IEEE Trans. Inform. Theory*, **41** (1975), 1276–1289.