# One-Shot Mutual Covering Lemma and Marton's Inner Bound with a Common Message 

Jingbo Liu Paul Cuff Sergio Verdú<br>Dept. of Electrical Eng., Princeton University, NJ 08544<br>\{jingbo,cuff,verdu\} @ princeton.edu


#### Abstract

By developing one-shot mutual covering lemmas, we derive a one-shot achievability bound for broadcast with a common message which recovers Marton's inner bound (with three auxiliary random variables) in the i.i.d. case. The encoder employed is deterministic. Relationship between the mutual covering lemma and a new type of channel resolvability problem is discussed.


## I. Introduction

While many problems in network information theory have been successfully solved in the discrete memoryless case using either the method of types or weak/strong typicality (e.g., [1] or [2]), there is a gap between the i.i.d. assumptions underlying these methods and the nature of sources and channels arising from real world applications. Recent works (e.g. [3][4]) have developed new methods to derive tight one-shot achievability bounds for specialized problems. Meanwhile, a natural question is whether there exist general techniques to attach non-asymptotic fundamental limits in multiuser information theory. By developing one-shot versions of covering and packing lemmas, [5] successfully obtained one-shot achievability bounds for various problems (multiaccess, Slepian-Wolf, Gelfand-Pinsker, Wyner-Ziv, AhlsedeKörner, and broadcast of private messages) which lead to such non-asymptotic bounds and recover known results in the i.i.d. case. However the proof of Marton's inner bound of broadcast without public/common messages (the two auxiliary version, [1, Theorem 8.3]) proceeded by showing the achievability of each corner point, which requires time sharing to recover the full rate region in the i.i.d. case.

In this paper we develop a one-shot mutual covering lemma so that a one-shot version of Marton's inner bound with a common message (the three auxiliary version originally due to Liang et al. [6]; see also [1, Theorem 8.4]) can be obtained without time sharing.

Time sharing may not be satisfactory since it is meaningless in the one-shot case. This is keenly noted by the authors of [7], who also observed that the mutual covering lemma [1, Lemma 8.1], a technique for avoiding time sharing in the i.i.d. case, does not seem to have a one-shot counterpart:
... This is helpful because no one-shot extension of the mutual covering and packing lemma exists... For this reason Verdú's result seems to be weaker than ours.
The present paper gives a single-shot mutual covering lemma, thereby filling the gap noted by the above remark. We will
provide two derivations, one of them is based on a recent generalization of channel resolvability [8].

Partly motivated by the desirability of achievability bounds without time sharing, [7] (see also [9]) proposed a new technique for deriving one-shot achievability bounds using stochastic encoders, thus avoiding the mutual covering lemma and yielding a one-shot bound which recovers the two auxiliaryvariable version of Marton's inner bound. However that version of Marton's bound is known to not be tight, whereas the three auxiliary-variable version (Liang-Kramer bound) is still a candidate for the capacity region [1, Section 8.4]. In [9] a one-shot achievability bound with three auxiliaries is stated without a proof ${ }^{1}$, which is not easily comparable with our bound. However the main probability terms from the two bounds are the same, so they are equivalent in the second order rate analysis.

## II. One-shot Mutual Covering Lemma

We develop one-shot mutual covering lemma(s) in this section, which require a different proof idea than the i.i.d. case [1, Appendix 8A] because the empirical distribution is meaningless in the non-block coding case. The main device in our proof is the introduction of an auxiliary random variable $\tilde{V}$, which can be viewed as being "typical" with a random codeword from the $U$ codebook.

To begin with, let us introduce the notations of the relative information

$$
\begin{equation*}
\imath_{P \| Q}(x):=\log \frac{\mathrm{d} P}{\mathrm{~d} Q}(x) \tag{1}
\end{equation*}
$$

where $P$ and $Q$ are distributions on a same alphabet $\mathcal{X}$ with $\frac{\mathrm{d} P}{\mathrm{~d} Q}(x)$ being the Radon-Nikodym derivative, and the information density

$$
\begin{equation*}
\imath_{U ; V}(u ; v):=\imath_{P_{V \mid U=u} \| P_{V}}(v) \tag{2}
\end{equation*}
$$

Lemma 1. Fix $P_{U V}$ and let

$$
\begin{equation*}
P_{U^{M} V^{L}}:=P_{U} \times \cdots \times P_{U} \times P_{V} \times \cdots \times P_{V} \tag{3}
\end{equation*}
$$

[^0]then
\[

$$
\begin{align*}
& \mathbb{P}\left[\bigcap_{m=1, l=1}^{M, L}\left\{\left(U_{m}, V_{l}\right) \notin \mathcal{F}\right\}\right] \\
& \leq \mathbb{P}[(U, V) \notin \mathcal{F}]+\mathbb{P}\left[\exp \left(\imath_{U ; V}(U ; V)\right)>M L \exp (-\gamma)-\delta\right] \\
& \quad+\frac{\min \{M, L\}-1}{\delta}+e^{-\exp (\gamma)} \tag{4}
\end{align*}
$$
\]

for all $\delta, \gamma>0$ and event $\mathcal{F}$.
Proof: Let $\left(\tilde{V}_{1}, \ldots, \tilde{V}_{M}\right)$ be random variables satisfying

$$
\begin{equation*}
P_{U^{M} \tilde{V}^{M}}=P_{U V} \times \cdots \times P_{U V} \tag{5}
\end{equation*}
$$

and define

$$
\begin{equation*}
\tilde{V}:=\tilde{V}_{W} \tag{6}
\end{equation*}
$$

where $W$ is equiprobable on $\{1, \ldots, M\}$ and independent of $\left(U^{M}, \tilde{V}^{M}\right)$. By independence,

$$
\begin{equation*}
\mathbb{P}\left[\bigcap_{m=1, l=1}^{M, L}\left\{\left(U_{m}, V_{l}\right) \notin \mathcal{F}\right\}\right]=\mathbb{E} \mathbb{P}\left[\bigcap_{m=1}^{M}\left\{\left(U_{m}, V_{1}\right) \notin \mathcal{F}\right\} \mid U^{M}\right]^{L} \tag{7}
\end{equation*}
$$

Define the functions

$$
\begin{align*}
A_{u^{M}}:= & \mathbb{P}\left[\bigcup_{m=1}^{M}\left\{\left(u_{m}, V_{1}\right) \in \mathcal{F}\right\}\right]  \tag{8}\\
\pi\left(u^{M} ; v\right):= & 1\left\{\bigcup_{m=1}^{M}\left\{\left(u_{m}, v\right) \in \mathcal{F}\right\}\right\}  \tag{9}\\
& \cdot 1\left\{\imath_{U^{M} ; \tilde{V}}\left(u^{M} ; v\right) \leq \log L-\gamma\right\} \tag{10}
\end{align*}
$$

The second term in (10) will play a role in the change of measure step. We then obtain

$$
\begin{align*}
& {\left[1-A_{U^{M}}\right]^{L} }  \tag{11}\\
= & {\left[1-\frac{\left(L \exp (-\gamma) A_{U^{M}}\right) \exp (\gamma)}{L}\right]^{L} }  \tag{12}\\
\leq & {\left[1-\frac{\exp (\gamma)}{L} \mathbb{E}\left[\pi\left(U^{M} ; \tilde{V}\right) \mid U^{M}\right]\right]^{L} }  \tag{13}\\
\leq & 1-\mathbb{E}\left[\pi\left(U^{M} ; \tilde{V}\right) \mid U^{M}\right]+e^{-\exp (\gamma)} \tag{14}
\end{align*}
$$

where

- (13) is from the fact that (by change of measure)

$$
\begin{align*}
L \exp (-\gamma) A_{U^{M}} & \geq L \exp (-\gamma) \mathbb{E}\left[\pi\left(U^{M} ; V_{1}\right) \mid U^{M}\right]  \tag{15}\\
& \geq \mathbb{E}\left[\pi\left(U^{M} ; \tilde{V}\right) \mid U^{M}\right] \tag{16}
\end{align*}
$$

- (14) uses the inequality

$$
\begin{equation*}
\left(1-\frac{p \alpha}{M}\right)^{M} \leq 1-p+e^{-\alpha} \tag{17}
\end{equation*}
$$

for $M, \alpha>0$ and $0 \leq p \leq 1$.

Thus

$$
\begin{align*}
& \mathbb{E}\left[1-A_{U^{M}}\right]^{L}  \tag{18}\\
& \leq \mathbb{E}\left[1-\pi\left(U^{M} ; \tilde{V}\right)\right]+e^{-\exp (\gamma)}  \tag{19}\\
&= \frac{1}{M} \sum_{w=1}^{M} \mathbb{E}\left[1-\pi\left(U^{M} ; \tilde{V}_{W}\right) \mid W=w\right]+e^{-\exp (\gamma)}  \tag{20}\\
&= \mathbb{E}\left[1-\pi\left(U^{M} ; \tilde{V}_{1}\right)\right]+e^{-\exp (\gamma)}  \tag{21}\\
& \leq \mathbb{P}\left[\left(U_{1}, \tilde{V}_{1}\right) \notin \mathcal{F}, \text { or } \imath_{U^{M} ; \tilde{V}}\left(U^{M} ; \tilde{V}_{1}\right)>\log L-\gamma\right]+e^{-\exp (\gamma)}  \tag{22}\\
& \leq \mathbb{P}[(U, V) \notin \mathcal{F}] \\
&+\mathbb{P}\left[\frac{1}{M} \sum_{m=1}^{M} \frac{\mathrm{~d} P_{V \mid U=U_{m}}}{\mathrm{~d} P_{V}}\left(\tilde{V}_{1}\right)>L \exp (-\gamma)\right]+e^{-\exp (\gamma)}  \tag{23}\\
& \leq \mathbb{P}[(U, V) \notin \mathcal{F}]+\mathbb{P}\left[\frac{\mathrm{d} P_{V \mid U=U_{1}}}{\mathrm{~d} P_{V}}\left(\tilde{V}_{1}\right)>M L \exp (-\gamma)-\delta\right] \\
&+\mathbb{P}\left[\sum_{m=2}^{M} \frac{\left.\mathrm{~d} P_{V \mid U=U_{m}}^{\mathrm{d} P_{V}}\left(\tilde{V}_{1}\right)>\delta\right]+e^{-\exp (\gamma)}}{\leq}\right.  \tag{24}\\
& \mathbb{P}[(U, V) \notin \mathcal{F}]+\mathbb{P}\left[\frac{\mathrm{d} P_{V \mid U=U_{1}}}{\mathrm{~d} P_{V}}\left(\tilde{V}_{1}\right)>M L \exp (-\gamma)-\delta\right] \\
&+\frac{M-1}{\delta}+e^{-\exp (\gamma)} \tag{25}
\end{align*}
$$

where

- (21) is because by symmetry, all summands in (20) are equal.
- (23) is because by definition, $P_{U_{1} \tilde{V}_{1}}=P_{U V}$.
- (24) uses the elementary inequality $1\{x+y>a+b\} \leq$ $1\{x>a\}+1\{y>b\}$.
- (25) uses the Markov inequality.

Remark 2. Supposing that $M \leq L$, an optimal choice of $(\gamma, \delta)$ is the solution to the optimization problem

$$
\begin{align*}
\operatorname{minimize} & f(\gamma, \delta):=\frac{M-1}{\delta}+e^{-\exp (\gamma)}  \tag{26}\\
\text { subject to } & g(\gamma, \delta):=M L \exp (-\gamma)-\delta=\lambda \tag{27}
\end{align*}
$$

for some constant $\lambda \in \mathbb{R}$. Using Lagrange multipliers we can show that the optimal value of $\delta$ for fixed $\gamma$ is:

$$
\begin{equation*}
\delta=\sqrt{(M-1) M L} \exp (-\gamma) e^{\frac{1}{2} \exp (\gamma)} \tag{28}
\end{equation*}
$$

Note that Lemma 1 recovers the covering lemma in [5] when $\min \{M, L\}=1$. Also, Lemma 1 can be weakened to the following simpler bound by setting $\delta=M L(\exp (-\gamma)-$ $\exp (-2 \gamma))$.
Lemma 3. Under the same assumptions of Lemma 1,

$$
\begin{align*}
& \mathbb{P}\left[\bigcap_{m=1, l=1}^{M, L}\left\{\left(U_{m}, V_{l}\right) \notin \mathcal{F}\right\}\right] \\
& \leq \mathbb{P}[(U, V) \notin \mathcal{F}]+\mathbb{P}\left[\imath_{U ; V}(U ; V)>\log (M L)-2 \gamma\right] \\
&  \tag{29}\\
& \quad+\frac{\min \{M, L\}-1}{M L(\exp (-\gamma)-\exp (-2 \gamma))}+e^{-\exp (\gamma)}
\end{align*}
$$

for all $\gamma>0$ and event $\mathcal{F}$.
Remark 4. In all versions of the mutual covering lemmas given above, the sum of the two probability terms can be strengthened to the probability of the union of two events, if the union bound is not applied to simplify (22).

A conditional version of the mutual covering lemma (c.f. Lemma 8.1 in [1]) also follows by averaging:

Lemma 5. Fix $P_{U S T}$ and let

$$
\begin{equation*}
P_{U S^{M} T^{L}}:=P_{U} \times P_{S \mid U} \times \cdots \times P_{S \mid U} \times P_{T \mid U} \times \cdots \times P_{T \mid U} \tag{30}
\end{equation*}
$$

then

$$
\begin{align*}
& \mathbb{P}\left[\bigcap_{m=1, l=1}^{M, L}\left\{\left(U, S_{m}, T_{l}\right) \notin \mathcal{F}\right\}\right] \\
& \leq \mathbb{P}\left[\{(U, S, T) \notin \mathcal{F}\} \cup\left\{\imath_{S ; T \mid U}(S ; T \mid U)>\log (M L)-2 \gamma\right\}\right] \\
& \quad+\frac{\min \{M, L\}-1}{M L(\exp (-\gamma)-\exp (-2 \gamma))}+e^{-\exp (\gamma)} \tag{31}
\end{align*}
$$

for all $\gamma>0$ and event $\mathcal{F}$.

## III. Mutual Covering from Resolvability

While the mutual covering lemma in Section II suffices the purpose of Section IV, in this section we provide a simple alternative derivation based on a recent result on resolvability in the excess information metric [8], which illustrates the interesting connection between resolvability and the mutual covering lemma ${ }^{2}$.

Lemma 6. [8] Fix $P_{U V}=P_{U} P_{V \mid U}$. Let $\mathbf{c}=\left[c_{1}, \ldots, c_{M}\right]$ be i.i.d. according to $P_{U}$. Define

$$
\begin{equation*}
\hat{P}_{V}:=\frac{1}{M} \sum_{m=1}^{M} P_{V \mid U=c_{m}} \tag{32}
\end{equation*}
$$

Then for any $\lambda>2$,

$$
\begin{align*}
& \mathbb{E P}\left[\imath_{\hat{P}_{V} \| P_{V}}(\hat{V})>\log \lambda\right] \\
\leq & \mathbb{P}\left[\imath_{V ; U}(V ; U) \geq \log \frac{M \lambda}{2}\right]+\frac{2}{\lambda} \tag{33}
\end{align*}
$$

where the expectation is with respect to the codebook realization, $\hat{V} \sim \hat{P}_{V}$ and $(U, V) \sim P_{U V}$.

The above lemma follows by setting $\gamma=\lambda, \sigma=\epsilon \uparrow \frac{\lambda}{2}$, $\delta \uparrow 1, Q_{X U}=P_{V U}, \pi_{X}=P_{V}, L=M$ and $P_{X}=\hat{P}_{V}$ in [8, Remark 3]. We then have

Lemma 7. Under the same assumptions of Lemma 1,

$$
\begin{align*}
& \mathbb{P}\left[\bigcap_{m=1, l=1}^{M, L}\left\{\left(U_{m}, V_{l}\right) \notin \mathcal{F}\right\}\right] \\
& \leq \mathbb{P}[(U, V) \notin \mathcal{F}]+\mathbb{P}\left[l_{U ; V}(U ; V) \geq \log M L-\gamma\right] \\
& \quad+\frac{\exp (\gamma)}{\max \{M, L\}}+e^{-\frac{1}{2} \exp (\gamma)} \tag{34}
\end{align*}
$$

[^1]Remark 8. The bound in Lemma 7 appears similar to and slightly simpler than Lemma 3. However the sum of the two probabilities in (34) cannot be easily strengthened to the probability of a union (see Remark 4), which is important in the second order rate analysis.
Remark 9. In both derivations of the one-shot mutual covering lemma, the role of $U$ and $V$ are asymmetric. Moreover, the two methods are not readily extendable to obtain a one-shot version of the multivariate covering lemma [13, Lemma 8.2].

Proof of Lemma 7: Assume without loss of generality that $L \geq M$. Define the sets

$$
\begin{align*}
\mathcal{F}_{u} & :=\{v:(u, v) \in \mathcal{F}\}  \tag{35}\\
\mathcal{A}_{u^{M}} & :=\bigcup_{m=1}^{M} \mathcal{F}_{u} \tag{36}
\end{align*}
$$

For fixed $\mathbf{c}=c^{M}$ in Lemma 6,

$$
\begin{align*}
\hat{P}_{V}\left(\mathcal{A}_{\mathbf{c}}\right)-\lambda P_{V}\left(\mathcal{A}_{\mathbf{c}}\right) & \leq \sup _{\mathcal{A}}\left[\hat{P}_{V}(\mathcal{A})-\lambda P_{V}(\mathcal{A})\right]  \tag{37}\\
& =\mathbb{P}\left[\frac{\mathrm{d} \hat{P}_{V}}{\mathrm{~d} P_{V}}(\hat{V})>\lambda\right]-\lambda \mathbb{P}\left[\frac{\mathrm{d} \hat{P}_{V}}{\mathrm{~d} P_{V}}(V)>\lambda\right]  \tag{38}\\
& \leq \mathbb{P}\left[\frac{\mathrm{d} \hat{P}_{V}}{\mathrm{~d} P_{V}}(\hat{V})>\lambda\right] \tag{39}
\end{align*}
$$

where (38) is from Neyman-Pearson lemma. Thus

$$
\begin{align*}
\lambda P_{V}\left(\mathcal{A}_{\mathbf{c}}\right) & \geq \hat{P}_{V}\left(\mathcal{A}_{\mathbf{c}}\right)-\mathbb{P}\left[\frac{\mathrm{d} \hat{P}_{V}}{\mathrm{~d} P_{V}}(\hat{V})>\lambda\right]  \tag{40}\\
& \geq \frac{1}{M} \sum_{m=1}^{M} P_{V \mid U=c_{m}}\left(\mathcal{F}_{c_{m}}\right)-\mathbb{P}\left[\frac{\mathrm{d} \hat{P}_{V}}{\mathrm{~d} P_{V}}(\hat{V})>\lambda\right] \tag{41}
\end{align*}
$$

where (41) is from $\mathcal{F}_{c_{m}} \subseteq \mathcal{A}_{\mathbf{c}}$. Denote by $p_{\mathrm{c}}$ the right hand side of (41). Then $p_{\mathbf{c}} \leq 1$, and setting $\mathbf{c}=U^{M}$ we obtain

$$
\begin{align*}
\mathbb{P}\left[\bigcap_{l=1}^{L}\left\{\left(U_{m}, V_{l}\right) \notin \mathcal{F}\right\} \mid U^{M}\right] & =\left[1-P_{V}\left(\mathcal{A}_{U^{M}}\right)\right]^{L} \\
& \leq\left[1-\frac{p_{\mathbf{c}} \frac{L}{\lambda}}{L}\right]^{L}  \tag{42}\\
& \leq 1-p_{\mathbf{c}}+e^{-\frac{L}{\lambda}} \tag{43}
\end{align*}
$$

where (43) uses (17). Then the result follows by unconditioning $U^{M}$ on both sides of (43), applying (41) and Lemma 6, and setting $\lambda=2 L \exp (-\gamma)$.

While the derivations of the one-shot mutual covering lemma in Sections II and III follow different routes, their correspondences can be seen in the following ways:

1) The auxiliary random variable $\tilde{V}$, which is the main device in the first proof, has the distribution $\hat{P}_{V}$ as in (32) conditioned on $U^{M}=c^{M}$.
2) The change of measure steps (16) can be related to the Neyman-Pearson lemma (37)-(39).
3) Both derivations relies on the inequality (17) which also appeared in the proof of the standard covering lemma.

Although Lemma 6 implies the one-shot mutual covering lemma, the reverse implication does not seem to follow directly. Thus resolvability in the excess information is a stronger result than the mutual covering lemma.

## IV. Inner Bound with a Common Message

We prove a single shot version of the asymptotic achievability result of Liang-Kramer [6, Theorem 5] (see also [1, Theorem 8.4]). This region is equivalent to an inner bound obtained by Gelfand and Pinsker [14] upon optimization (see [15] or [1, Remark 8.6]).

Theorem 10. Fix arbitrary distributions $P_{Y_{1} Y_{2} \mid X}, P_{U S T}$, a map $x: \mathcal{U} \times \mathcal{S} \times \mathcal{T} \rightarrow \mathcal{X}$, and integers $M_{0}, M_{10}, M_{20}, N$, $L, \hat{N}$ and $\hat{L}$. Set

$$
\begin{align*}
M & :=M_{0} M_{10} M_{20}  \tag{44}\\
M_{1} & :=M_{10} N  \tag{45}\\
M_{2} & :=M_{20} L  \tag{46}\\
\tilde{N} & :=\hat{N} N  \tag{47}\\
\tilde{L} & :=\hat{L} L \tag{48}
\end{align*}
$$

Then there exists an $\left(M_{0}, M_{1}, M_{2}, \epsilon_{1}, \epsilon_{2}\right)$ code with

$$
\begin{align*}
& \max \left\{\epsilon_{1}, \epsilon_{2}\right\} \\
& \leq 2 \exp (-\gamma)+e^{-\exp (\gamma)}  \tag{49}\\
& +\mathbb{P}\left[\left\{\imath_{U S ; Y_{1}}\left(U S ; Y_{1}\right) \leq \log M \tilde{N}+\gamma\right\}\right. \\
& \quad \cup\left\{\imath_{U T ; Y_{2}}\left(U T ; Y_{2}\right) \leq \log M \tilde{L}+\gamma\right\} \\
& \quad \cup\left\{\imath_{S ; Y_{1} \mid U}\left(S ; Y_{1} \mid U\right) \leq \log \tilde{N}+\gamma\right\} \\
& \quad \cup\left\{\imath_{T ; Y_{2} \mid U}\left(T ; Y_{2} \mid U\right) \leq \log \tilde{L}+\gamma\right\} \\
& \\
& \left.\cup\left\{\imath_{S ; T \mid U}(S ; T \mid U)>\log (\hat{N} \hat{L})-2 \gamma\right\}\right]  \tag{50}\\
& +
\end{align*} \frac{\min \{\hat{N}, \hat{L}\}-1}{\hat{N} \hat{L}(\exp (-\gamma)-\exp (-2 \gamma))}
$$

where $P_{U S T X Y_{1} Y_{2}}:=P_{U S T} P_{X \mid U S T} P_{Y_{1} Y_{2} \mid X}$.
As in [1, Theorem 8.4], the private messages are decomposed into a public part and an individual part:

$$
\begin{equation*}
W_{i}=\left(W_{i 0}, W_{i i}\right) \tag{51}
\end{equation*}
$$

where $i=1,2$ and $W_{i 0}$ is supposed to be decodable by both users.

Proof:

- Codebook Generation: generate

$$
\begin{equation*}
\mathbf{u}=\left[\mathbf{u}_{1}, \ldots, \mathbf{u}_{M}\right] \tag{52}
\end{equation*}
$$

according to the distribution $P_{U}^{\otimes M}$. Also for each $1 \leq$ $i \leq M$, generate

$$
\begin{equation*}
\mathbf{s}_{i}=\left[\mathbf{s}_{i}(n, \hat{n})\right]_{1 \leq n \leq N, 1 \leq \hat{n} \leq \hat{N}} \tag{53}
\end{equation*}
$$

according to $P_{S \mid U=\mathrm{u}_{i}}^{\otimes N \hat{N}}$ and

$$
\begin{equation*}
\mathbf{t}_{i}=\left[\mathbf{t}_{i}(l, \hat{l})\right]_{1 \leq l \leq L, 1 \leq \hat{l} \leq \hat{L}} \tag{54}
\end{equation*}
$$

according to $P_{T \mid U=u_{i}}^{\otimes L \hat{L}}$. (In other words, for each $i$ we construct a codebook similar to [1, Figure 8.8], where each small rectangle has size $\hat{L} \times \hat{N}$.)

- Encoding: we may assume that the public message is $w_{0}$ and the private messages for the two users are $\left(w_{10}, a\right)$ and $\left(w_{20}, b\right)$ respectively, where

$$
\begin{align*}
w_{0} & \in\left\{1, \ldots, M_{0}\right\}  \tag{55}\\
w_{i 0} & \in\left\{1, \ldots, M_{i 0}\right\}  \tag{56}\\
a & \in\{1, \ldots, N\}  \tag{57}\\
b & \in\{1, \ldots, L\} \tag{58}
\end{align*}
$$

for $i=1,2$. Then the index $m=\mathrm{m}\left(w_{0}, w_{10}, w_{20}\right)$ is selected for the lower-layer codebook where m is a bijection between $\left[M_{0}\right] \times\left[M_{10}\right] \times\left[M_{20}\right]$ and $[M]$. The encoder then finds $\hat{a} \in\{1, \ldots, \hat{N}\}$ and $\hat{b} \in\{1, \ldots, \hat{L}\}$ that minimize

$$
\begin{equation*}
\zeta\left(\mathbf{u}_{m}, \mathbf{s}_{m}(a, \hat{a}), \mathrm{t}_{m}(b, \hat{b})\right) \tag{59}
\end{equation*}
$$

where

$$
\begin{equation*}
\zeta(u, s, t):=P_{Y_{1} Y_{2} \mid X=x(u, s, t)}\left(\mathcal{Y}_{u, s, t}\right) \tag{60}
\end{equation*}
$$

and

$$
\begin{align*}
\mathcal{Y}_{u, s, t} & :=\left\{\left(y_{1}, y_{2}\right) \in \mathcal{Y}_{1} \times \mathcal{Y}_{2}: \imath_{U S ; Y_{1}}\left(u s ; y_{1}\right) \leq \log M \tilde{N}+\gamma\right. \\
& \text { or } \imath_{U T ; Y_{2}}\left(u t ; y_{2}\right) \leq \log M \tilde{L}+\gamma \\
& \text { or } \imath_{S ; Y_{1} \mid U}\left(s ; y_{1} \mid u\right) \leq \log \hat{N}+\gamma \\
& \text { or } \left.\imath_{T ; Y_{2} \mid U}\left(t ; y_{2} \mid u\right) \leq \log \hat{L}+\gamma\right\} \tag{61}
\end{align*}
$$

The signal transmitted is then $x\left(\mathbf{u}_{m}, \mathbf{s}_{m}(a, \hat{a}), \mathbf{t}_{m}(b, \hat{b})\right)$.

- Decoder: assume that $y_{1}$ and $y_{2}$ are observed by the two receivers, respectively. The decoder of the first receiver (Decoder 1) finds the unique $m$ such that

$$
\begin{equation*}
\exists d, \hat{d}, \quad \imath_{U S ; Y_{1}}\left(\mathbf{u}_{m} \mathbf{s}_{m}(d, \hat{d}) ; y_{1}\right)>\log M \tilde{N}+\gamma \tag{62}
\end{equation*}
$$

and set $\left(m_{0}, m_{10}, m_{20}\right)=\mathrm{m}^{-1}(m)$, or declares an error if not possible. Decoder 1 then finds the unique $(c, \hat{c})$ such that

$$
\begin{equation*}
\imath_{S ; Y_{1} \mid U}\left(\mathbf{S}_{m}(c, \hat{c}) ; y_{1} \mid \mathbf{u}_{m}\right)>\log \tilde{N}+\gamma \tag{63}
\end{equation*}
$$

or declares an error if not possible. The output public message is then $m_{0}$ and the output private message is $\left(m_{10}, c\right)$.
Decoder 2 performs similar operations as Decoder 1.

- Error Analysis: By symmetry of the codebook, we may assume without loss of generality that

$$
\begin{equation*}
w_{0}=w_{10}=w_{20}=w_{11}=w_{22}=1 \tag{64}
\end{equation*}
$$

is sent and $m(1,1,1)=1$. Also it suffices to prove the bound for Decoder 1 only in view of the symmetry of the bound (50). Assume that ( $\hat{a}, \hat{b}$ ) minimizes (59) (and
is selected by the encoder). Decoder 1 fails only if one or more of the following events occur:

$$
\begin{align*}
& \mathcal{E}_{1}: \imath_{U S ; Y_{1}}\left(\mathbf{u}_{1} \mathbf{s}_{1}(1, \hat{a}) ; y_{1}\right) \leq \log M \tilde{N}+\gamma ;  \tag{65}\\
& \mathcal{E}_{2}: \exists m \neq 1, d, \hat{d}, \\
&{ }^{\imath} U S ; Y_{1}\left(\mathbf{u}_{m} \mathbf{s}_{m}(d, \hat{d}) ; y_{1}\right)>\log M \tilde{N}+\gamma ;  \tag{66}\\
& \mathcal{E}_{3}:{ }^{\imath} S ; Y_{1} \mid U \\
&\left(\mathbf{s}_{1}(1, \hat{a}) ; y_{1} \mid \mathbf{u}_{1}\right) \leq \log \tilde{N}+\gamma ;
\end{align*}
$$

$\mathcal{E}_{4}: \quad \exists n \neq 1, \hat{n}, \quad{ }^{\imath} S ; Y_{1} \mid U\left(\mathbf{s}_{1}(n, \hat{n}) ; y_{1} \mid \mathbf{u}_{1}\right)>\log \tilde{N}+\gamma$.

Denote by $S^{*}$ and $T^{*}$ the coefficients selected by the encoder and $Y_{1}^{*}, Y_{2}^{*}$ the corresponding outputs. Averaged over the codebook, we can bound

$$
\begin{align*}
& \mathbb{P}\left[\mathcal{E}_{2}\right] \\
= & \mathbb{P}\left[\bigcup_{m \neq 1, d, \hat{d}}\left\{\imath_{U S ; Y_{1}}\left(U_{m} S_{m}(d, \hat{d}) ; Y_{1}^{*}\right)>\log M \tilde{N}+\gamma\right\}\right]  \tag{69}\\
\leq & (M-1) \tilde{N} \mathbb{E} \mathbb{P}\left[\imath_{U S ; Y_{1}}\left(U_{2} S_{2}(1,1) ; Y_{1}^{*}\right)>\log M \tilde{N}+\gamma \mid Y_{1}^{*}\right]  \tag{70}\\
\leq & \frac{M-1}{M \exp (\gamma)} \tag{71}
\end{align*}
$$

by change of measure, where the joint probability of $\left\{U_{m}, S_{m}(d, d)\right\}_{m, d, \hat{d}}$ and $Y_{1}^{*}$ satisfies $P_{U_{m} S_{m}(d, \hat{d}) Y_{1}^{*}}=$ $P_{U S} \times P_{Y_{1}^{*}}$. Similarly

$$
\begin{align*}
& \mathbb{P}\left[\mathcal{E}_{4}\right] \\
= & \mathbb{P}\left[\bigcup_{n \neq 1,1 \leq \hat{n} \leq \hat{N}}\left\{\imath_{S ; Y_{1} \mid U}\left(S_{1}(n, \hat{n}) ; Y_{1}^{*} \mid U_{1}\right)>\log \tilde{N}+\gamma\right\}\right] \tag{72}
\end{align*}
$$

$$
\leq(N-1) \hat{N} \mathbb{E P}\left[\imath_{S ; Y_{1} \mid U}\left(S_{1}(2,1) ; Y_{1}^{*} \mid U_{1}\right)\right.
$$

$$
\begin{equation*}
\left.>\log \tilde{N}+\gamma \mid Y_{1}^{*} U_{1}\right] \tag{73}
\end{equation*}
$$

$$
\begin{equation*}
\leq \frac{(N-1) \hat{N}}{\tilde{N} \exp (\gamma)} \tag{74}
\end{equation*}
$$

where $P_{S_{1}(n, \hat{n}) Y_{1}^{*} U_{1}}=P_{U} P_{S \mid U} P_{Y_{1}^{*} \mid U}$ for $n \neq 1$. Next, notice that

$$
\begin{align*}
\mathbb{P}\left[\mathcal{E}_{1} \cup \mathcal{E}_{3}\right] & \leq \mathbb{E} \zeta\left(U_{1}, S^{*}, T^{*}\right)  \tag{75}\\
& =\int_{0}^{1} \mathbb{P}\left[\zeta\left(U_{1}, S^{*}, T^{*}\right)>v\right] \mathrm{d} v \tag{76}
\end{align*}
$$

Note that $\left(S^{*}, T^{*}\right)$ has a complicated distribution since the coefficients are selected through the minimization of (59). To tackle this, let

$$
\begin{equation*}
\mathcal{F}_{v}:=\{(u, s, t) \in \mathcal{U} \times \mathcal{S} \times \mathcal{T}: \zeta(u, s, t) \leq v\} \tag{77}
\end{equation*}
$$

Then we can bound

$$
\begin{align*}
& \mathbb{P}\left[\zeta\left(U_{1}, S^{*}, T^{*}\right)>v\right] \\
= & \mathbb{P}\left[\bigcap_{1 \leq \hat{n} \leq \hat{N}, 1 \leq \hat{l} \leq \hat{L}}\left\{\left(U_{1}, S_{1}(1, \hat{n}), T_{1}(1, \hat{l})\right) \notin \mathcal{F}_{v}\right\}\right]  \tag{78}\\
\leq & \mathbb{P}\left[(U, S, T) \notin \mathcal{F}_{v} \cup \mathcal{S}\right] \\
& +\frac{\min \{\hat{N}, \hat{L}\}-1}{\hat{N} \hat{L}(\exp (-\gamma)-\exp (-2 \gamma))}+e^{-\exp (\gamma)} \tag{79}
\end{align*}
$$

where we invoked Lemma 5 in (79), and defined the set

$$
\begin{equation*}
\mathcal{S}:=\left\{(u, s, t): \imath_{S ; T \mid U}(s ; t \mid u)>\log \hat{N} \hat{L}-2 \gamma\right\} \tag{80}
\end{equation*}
$$

Combining (76) and (79) we have

$$
\begin{align*}
& \mathbb{P}\left[\mathcal{E}_{1} \bigcup \mathcal{E}_{3}\right] \\
\leq & \int_{0}^{1} \mathbb{P}\left[(U, S, T) \notin \mathcal{F}_{v} \cup \mathcal{S}\right] \mathrm{d} v \\
& +\frac{\min \{\hat{N}, \hat{L}\}-1}{\hat{N} \hat{L}(\exp (-\gamma)-\exp (-2 \gamma))}+e^{-\exp (\gamma)} \tag{81}
\end{align*}
$$

but by definitions (60) and (77),

$$
\begin{align*}
& \int_{0}^{1} \mathbb{P}\left[(U, S, T) \notin \mathcal{F}_{v} \cup \mathcal{S}\right] \mathrm{d} v \\
= & \int_{0}^{1} \mathbb{E} 1\left\{(U, S, T) \in \mathcal{F}_{v}^{c} \cup \mathcal{S}\right\} \mathrm{d} v  \tag{82}\\
= & \mathbb{E}[1\{(U, S, T) \in \mathcal{S}\}+\zeta 1\{(U, S, T) \notin \mathcal{S}\}]  \tag{83}\\
= & \mathbb{P}[(U, S, T) \in \mathcal{S}] \\
& +\int \mathrm{d} P_{U S T} P_{Y_{1} Y_{2} \mid U S T}\left(\mathcal{Y}_{U S T}\right) 1\{(U, S, T) \notin \mathcal{S}\}  \tag{84}\\
= & \mathbb{P}[(U, S, T) \in \mathcal{S}] \\
& +\int \mathrm{d} P_{Y_{1} Y_{2} U S T} 1\left\{\left(Y_{1}, Y_{2}\right) \in \mathcal{Y}_{U S T} \cup(U, S, T) \notin \mathcal{S}\right\}  \tag{85}\\
= & \mathbb{P}\left[(U, S, T) \in \mathcal{S} \cup\left(Y_{1}, Y_{2}\right) \in \mathcal{Y}_{U S T}\right]  \tag{86}\\
= & \mathbb{P}\left[\left\{\imath_{S ; T \mid U}(S ; T \mid U)>\log \hat{N} \hat{L}-2 \gamma\right\}\right. \\
& \cup\left\{\imath_{U S ; Y_{1}}\left(U S ; Y_{1}\right) \leq \log M \tilde{N}+\gamma\right\} \\
& \cup\left\{\imath_{U T ; Y_{2}}\left(U T ; Y_{2}\right) \leq \log M \tilde{L}+\gamma\right\} \\
& \cup\left\{\imath_{S ; Y_{1} \mid U}\left(S ; Y_{1} \mid U\right) \leq \log \tilde{N}+\gamma\right\} \\
& \left.\cup\left\{\imath_{T ; Y_{2} \mid U}\left(T ; Y_{2} \mid U\right) \leq \log \tilde{L}+\gamma\right\}\right] \tag{87}
\end{align*}
$$

Finally, the proof is accomplished by substituting (87) into (81) and then applying the union bound with (71), (74) and (81).

Remark 11. In the i.i.d. setting Theorem 10 readily gives the
following achievable region

$$
\left\{\begin{align*}
R_{i i} & \leq R_{i}  \tag{88}\\
\hat{R}_{1}+\hat{R}_{2} & \geq I\left(\mathrm{U}_{1} ; \mathrm{U}_{2} \mid \mathrm{U}_{0}\right) \\
\sum_{i}\left(R_{i}-R_{i i}\right)+R_{0}+R_{i i}+\hat{R}_{i} & \leq I\left(\mathrm{U}_{0} \mathrm{U}_{i} ; \mathrm{Y}_{i}\right) \\
R_{i i}+\hat{R}_{i} & \leq I\left(\mathrm{U}_{i} ; \mathrm{Y}_{i} \mid \mathrm{U}_{0}\right)
\end{align*}\right.
$$

for some $R_{i i}, \hat{R}_{i}>0, i=1,2, P_{\mathrm{U}_{0} \mathrm{U}_{1} \mathrm{U}_{2}}$ and function $x$ : $\mathcal{U}_{0} \times \mathcal{U}_{1} \times \mathcal{U}_{2} \rightarrow \mathcal{X}$. Fourier-Motzkin elimination gives the same region as in [6]; see also [1, Theorem 8.4].

## V. DISCUSSION

In contrast to the one-shot mutual covering lemma, a oneshot version of mutual packing lemma [1, Lemma 12.2] holds trivially, because the key step union bound in the proof of [5, Lemma 2] does not require independence among the pairs.

Lemma 12. Fix $\left(P_{X Y}, M, N, \gamma\right)$, then

$$
\begin{equation*}
\mathbb{P}\left[\max _{m, n} \imath_{X ; Y}\left(X_{m} ; Y_{n}\right) \geq \log M N+\gamma\right] \leq \exp (-\gamma) \tag{89}
\end{equation*}
$$

where for all $1 \leq m \leq M, 1 \leq n \leq N, P_{X_{m} Y_{n}}=P_{X} \times P_{Y}$.
This can be used to derive a one-shot version of BergerTung inner bound without time sharing or stochastic decoders.

## ACKNOWLEDGMENT

This work was supported by NSF under Grants CCF1350595, CCF-1116013, CCF-1319299, CCF-1319304, and the Air Force Office of Scientific Research under Grant FA9550-12-1-0196.

## REFERENCES

[1] A. El Gamal and Y.-H. Kim, Network information theory. Cambridge University Press, 2011.
[2] I. Csiszár and J. Körner, Information theory: coding theorems for discrete memoryless systems. Academic Press, 1981.
[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307-2359, 2010.
[4] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2545-2575, 2013.
[5] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1-8, 2012.
[6] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," IEEE Transactions on Information Theory, vol. 53, no. 10, pp. 35173535, 2007.
[7] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in Proceedings of IEEE International Symposium on Information Theory, (Istanbul, Turkey), pp. 1287-1291, Jul. 2013.
[8] J. Liu, P. Cuff, and S. Verdú, "Resolvability in $E_{\gamma}$ metric with applications to lossy compression and wiretap channels," in Proceedings of IEEE International Symposium on Information Theory, accepted, 2015.
[9] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," arXiv:1303.0696 [cs.IT], 4 Mar 2013.
[10] P. Cuff, "Distributed channel synthesis," IEEE Transactions on Information Theory, vol. 59, pp. 7071-7096, Nov. 2013.
[11] T. Han and S. Verdu, "Approximation theory of output statistics," IEEE Transactions on Information Theory, vol. 39, no. 3, pp. 752-772, 1993.
[12] A. D. Wyner, "The common information of two dependent random variables," IEEE Transactions on Information Theory, vol. 21, no. 2, pp. 163-179, 1975.
[13] A. A. El Gamal and E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel.," IEEE Transactions on Information Theory, vol. 27, no. 1, pp. 120-122, 1981.
[14] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," Problemy Peredachi Informatsii, vol. 16, no. 1, pp. 24-34, 1980.
[15] Y. Liang, G. Kramer, and H. V. Poor, "On the equivalence of two achievable regions for the broadcast channel," IEEE Transactions on Information Theory, vol. 57, no. 1, pp. 95-100, 2011.


[^0]:    ${ }^{1}$ The authors of [9] announced that the proof will be included in their future draft.

[^1]:    ${ }^{2}$ Indeed, [10] uses the term "soft-covering" for the achievability part of channel resolvability [11], an idea that traces back to Wyner [12].

