Quantum rate-distortion coding

H. Barnum
I introduce rate-distortion theory for the coding of quantum information, and derive a lower bound, involving the coherent information, on the rate at which qubits must be used to store or compress an entangled quantum source with a given maximum level of distortion per source emission.

PACS numbers: 03.67-a, 03.67.Hk, 89.70.+c

I. INTRODUCTION

Recent discoveries about the power of quantum computation and other uses of quantum information processing underline the importance of understanding the limits on the compression of quantum information and on its transmission or storage in the presence of noise. But the noiseless and noisy channel coding theorems which address these issues use fidelity criteria for successful transmission which at first seem excessively stringent. As the size of the block of material produced by the source grows indefinitely, the fidelity of the entire block is required to approach one, a very strong requirement. This is analogous to the situation for classical information transmission, in which the criterion used in the coding theorems requires that the probability of an error in the entire block approach zero as the block length goes to infinity. A code with a constant nonzero error rate per symbol would fail this test miserably (error probability would go to one in the large block limit), but could still be perfectly acceptable as long as the error rate was sufficiently small. Most, if not all, noisy channel coding protocols used with real-world communications channels fail similarly. In the case of transmission of quantum information, we might be willing to tolerate a constant rate of bad Einstein-Podolsky-Rosen (EPR) pairs when transmitting or protecting entanglement, or, when sending pure states, a small expected deviation of the received state of each transmission from the transmitted one. A theory which tells us, given a level of distortion which we have decided we can tolerate, whether a given channel (noisy or noiseless) can achieve that error rate, would be decidedly useful. This is rate-distortion theory.

II. A QUANTUM VERSION OF RATE-DISTORTION

In generalizing classical rate-distortion theory to the quantum case, I will be concerned primarily with the transmission of quantum information. (For the rate-distortion theory of the transmission of classical information through a quantum channel, see [1].) For the purpose of initial definitions and the introduction of the notion of the rate-distortion function, I will consider both the transmission of ensembles of pure quantum states (which does have the transmission of classical information as a special case) and the transmission of a quantum system’s entanglement with some reference system. The main result of the paper, however, is derived for entanglement transmission, and thus definitely deals with quantum, rather than classical, information.

An ensemble $E$ of pure quantum states is a normalized probability measure $\mu_E(\rho)$ on Hilbert space. For pure state ensemble transmission, the criterion of fidelity I use is the average pure state fidelity of an operation $\mathcal{A}$, defined as:

$$ F(E, A) := \int d\rho \mu_E(\rho) \|\rho \mathcal{A} \rho^E \| $$ (2.1)

The operation $A$, like all operations or maps in this paper unless otherwise specified, is taken to be a linear, trace-preserving, completely positive map (see e.g. [2, 5]). For entanglement transmission, I will use the entanglement fidelity of [2]. If we conceive of the source density matrix $\rho$ as the partial trace of a pure state $|\psi\rangle_R$ in a larger Hilbert space, then the entanglement fidelity $F_e$ is the matrix element

$$ F_e(\rho, A) := \langle \psi_{RQ} | I \otimes A(|\psi_{RQ}\rangle\langle\psi_{RQ}|) |\psi_{RQ}\rangle $$ (2.2)

(Here $I$ is the identity operation on $R$.) Thus it measures how well the entangled pure state is preserved when the subsystem is subjected to the operation $A$. If the entanglement fidelity of a density operator $\rho$ is $F$, any ensemble $E$ of pure states whose density operator is $\rho$ has average pure-state fidelity higher than $F$ [2]. As measures of distortion I will use one minus the above fidelities. These must be evaluated for single transmissions, and then averaged over the block of $n$ transmissions. It is this procedure of evaluating marginal fidelity of individual transmissions, and then averaging over transmissions, which provides the weaker “distortion rate” criterion of fidelity, rather than the extremely strong criterion of fidelity of the total state of all transmissions.
which is used in the noiseless and noisy-channel coding theorems. The latter is so strong that a single transmission supported in a subspace orthogonal to the correct one gives zero total fidelity even if all other transmissions in the block come through perfectly.) In most of what follows, entanglement fidelity will be used; some definitions and results (noted explicitly) will apply also to pure state fidelity. Useful background may be found in [5]. Sources are sequences of density operators (or ensembles, when pure state transmission is considered) \( \rho^{(n)} \) (or \( E^{(n)} \)) on Hilbert spaces \( H^{\otimes n} \). I will confine myself to i.i.d. sources, defined as those for which \( \rho^{(n)} \equiv \rho^{\otimes n} \) (or \( E^{(n)} = E^{\otimes n} \)). While it is desirable to lift this restriction, it is not as strong as it might seem at first. I believe it to be technical in nature, and not much more restrictive in this context than the initial restriction of classical information theory to i.i.d. sources. Most of the classical theory was subsequently shown to be applicable to correlated sources under natural assumptions. In the quantum case, non-i.i.d. source density operators may exhibit correlation or even entanglement between transmissions. It might seem that to exclude the possibility of entanglement between successive transmissions excludes some of the most important quantum states. However, it should be clearly understood that the density operators describing a quantum source are assumed entangled with a reference system, and their entanglement is preserved by transmission. Transmission of such an entangled-reference-system source allows transmission of pure-state ensemble sources which have the same source density matrices when the pure states are averaged over (see [6], [2], [7]). These pure-state ensembles may of course be made up of states which themselves exhibit inter-transmission entanglement; it is only the density operator averaged over all these states which is assumed not to have this property. Moreover, no restriction at all is placed on the channel density operator of the coded states, which therefore may exhibit inter-transmission entanglement. Most work on quantum data compression and channel capacity has been in the context of i.i.d. sources; [8], [6], [9], [10], and [11] are some exceptions.

The channel will be taken to be noiseless; then a \((n, 2^R)\) rate-distortion code consists of a map \( \mathcal{E}^{(n)} \) from \( n \) copies of the source space to \( n \) copies of a channel space of dimension \( 2^R \), followed by a decoding \( D^{(n)} \) from \( n \) channels to \( n \) source spaces. The average entanglement distortion for an i.i.d. source can then be defined as:

\[
d(\rho, \mathcal{E}^{(n)}, D^{(n)}) := \frac{1}{n} \sum_{i=1}^{n} (1 - F_{\epsilon}(\rho, T_i^{(n)})),
\]

where \( T_i^{(n)} \) is the “marginal operation” on the \( i \)-th copy of the source space induced by the overall operation \( D^{(n)} \mathcal{E}^{(n)} \). More formally,

\[
T_i^{(n)}(\sigma) := \text{tr}_{Q_{i+1} \ldots Q_n} \text{tr}_{Q_i} \mathcal{E}^{(n)}(\rho \otimes \cdots \otimes \rho \otimes \sigma \otimes \rho \cdots \otimes \rho),
\]

where the \( \sigma \) in the input density operator is in the \( i \)-th position. (It is easily checked that this defines a trace-preserving operation.) The same definition, but with \( \overline{F}(E, T_i^{(n)}) \) as the fidelity criterion, defines the average pure-state distortion \( \overline{d} \).

\( R \) is said to be the rate of a rate-distortion code. (To avoid confusion, note that this is the inverse of a rate of information transmission.) The rate in rate-distortion is the rate at which the source is described, that is, the number of qubits, or the log of the number of Hilbert space dimensions, used to encode the source, per source emission. The goal of rate-distortion coding is to achieve low rates, i.e., to encode the source into as few qubits as possible per source emission, subject to the constraint that the distortion be no greater than some tolerable level.

A rate-distortion pair \((R, D)\) is achievable for a given source if there is a sequence of \((n, 2^R)\) rate-distortion codes \((\mathcal{E}^{(n)}, D^{(n)})\) such that

\[
\lim_{n \to \infty} d(\rho, \mathcal{E}^{(n)}, D^{(n)}) \leq D.
\]

Here \( d \) is whatever average distortion measure is used, e.g., \( \overline{d} \) or \( d \). The rate-distortion feasible set for an i.i.d. source \( \rho^{\otimes n} \) is the closure of the set of achievable rate-distortion pairs. The rate-distortion function \( R(D) \) is defined by

\[
R(D) := \inf_{\mathcal{E}, D} R(D) | \text{is achievable}.
\]

The rate-distortion frontier is the graph of the rate distortion function.

The coherent information [2] of a density operator \( \rho \) under an operation \( \mathcal{A} \) is defined as

\[
I_c(\rho, \mathcal{A}) := S(\mathcal{A}(\rho^Q)) - S(\mathcal{I} \otimes \mathcal{A}(\rho^{RQ})).
\]

That is, it is the entropy in the output state minus the entropy induced by the operation in an initially pure entangled state which purifies the input state. The term \( S(\mathcal{I} \otimes \mathcal{A}(\rho^{RQ})) \), which is also equal to the entropy introduced into a pure environment or apparatus which is used to implement the operation \( \mathcal{A} \) through a unitary interaction with the system, was dubbed in [2] the entropy exchange, \( S_e(\rho^Q, \mathcal{A}) \). Since it is the entropy of a density matrix, it is positive. The coherent information satisfies the data processing inequality [9]

\[
I_c(\rho, \mathcal{A}) \geq I_c(\rho, \mathcal{B} \mathcal{A}),
\]

where \( \mathcal{B} \) and \( \mathcal{A} \) are trace-preserving operations.

The coherent information is a good candidate to play the role, in quantum information theory, of the mutual information in classical information theory. In classical information theory [4], the rate-distortion function for a source \( X \) is given by

\[
R^f(D)_C := \min_{p(y|x) | d(p, x, p) \leq D} H(X, Y).
\]
That is, we minimize, over channels \( p(Y|X) \) relating outputs to inputs, the mutual information between output and input, subject to the constraint that average distortion induced by the channel be below \( d \). I use the coherent information to define a quantum analogue of the information rate-distortion function. The \textit{entanglement information rate-distortion function} \( R^e(D) \) for a source is defined by:

\[
R^e(D) := \min_{A|d(A) \leq D} I_e(\rho, A). 
\]  

(2.10)

I conjecture that, as in the classical case, the information rate-distortion function just defined is equal to the information rate-distortion function defined above, and thus that \( R^e(D) \) tells us the lowest rate at which we can use channel qubits to send a quantum source with entanglement distortion no greater than \( D \). In Section IV I will show that it is a lower bound to this rate. First, however, I will try to give an intuitive feel for why this holds true, and for why it is natural to conjecture that minimal coherent information gives the achievable information-distortion function, by comparing the classical and quantum situation.

III. A HEURISTIC DISCUSSION OF CLASSICAL AND QUANTUM RATE-DISTORTION

Why should the minimal coherent information be the achievable rate-distortion function? Some insight may be gained by considering the classical case. With a given input distribution \( X \) resulting in noisy outputs \( Y \), the noisy channels over which one minimizes in the expression for the information rate-distortion function can be made to behave as if they have a set of noise-free states of cardinality \( 2^n H(X,Y) \) given by the mutual information \( H(X,Y) \) between outputs and inputs. The other channel inputs are useless for conveying messages, even if a finite error probability is tolerated. (This follows from the strong converse of the noisy channel coding theorem, at least for sources maximizing the mutual information, and follows for general sources by reasoning similar to that in the proof of the strong converse.) The “intuition” (if it can be called that) behind rate-distortion theory is that the rest of a noisy channel is also useless for sending messages even at finite tolerable \textit{per-transmission} error probability. Rather than something intuitively clear in itself, this is probably best viewed as a rough \textit{reformulation} of rate-distortion theory, which helps us understand the role of mutual information in the result. The idea is that any source-channel pair acts, at best, like truncation to a subset of size \( 2^n H(X,Y) \); hence if the “room” in a given channel (as measured by its mutual information) can be used to send a given set of messages with some average distortion, those messages will undergo about the same distortion upon compression to a set of codewords whose number is given by the amount of room in the channel. Hence the smallest mutual information of a source-channel pair that achieves a given distortion level represents the smallest amount of room the messages can be compressed into, at that distortion level. Some source-channel pairs with a given mutual information may actually act worse, with respect to distortion, than truncation to a set of states of cardinality given by their mutual information. This is because the mutual information represents what can be done with optimal decoding, and the channel described may not include appropriate decoding. However, there always exists another operation that is the concatenation of the original channel with an optimal decoding: this channel has lower distortion and (by the classical data-processing inequality for mutual information) no higher mutual information, so it will be the one relevant in the minimization.

The quantum version of this involves noting that noisy channels (which can be represented by trace-preserving operations \( A \)) can be made to behave (by block coding schemes) as if they have a “good” subspace for information transmission, of dimension equal to the channel capacity. If the asymptotic maximal coherent information turns out to be the capacity ( [7] outlines a proof that it is, and see [6], [3], [8] for what would be a weak converse), and we have a strong converse, then we can also say that the channel behaves, from the point of view of asymptotic large-block fidelity, at best like truncation to a subspace of dimension given by the coherent information. More accurately but with no different effects on rate-distortion transmission, it behaves like decoherence of a set of subspaces the log of whose size is at most the coherent information. (Having a set of subspaces of size bounded by the coherent information rather than a single subspace of size given by the coherent information still does not help us send quantum information, since the best we can do is put the quantum information in the largest one). It is natural to suppose that quantum rate-distortion theory will tell us that for a per-transmission distortion measure, as well, channels acting on a given source density operator still behave, when optimally decoded, essentially like decoherence of a set of subspaces of size bounded by the coherent information (or at least, that they contain at least one “clean” subspace of that size, decohered from the rest of the channel input space).

IV. A LOWER BOUND ON THE ENTANGLEMENT RATE-DISTORTION FUNCTION

In this section, I will show that the rate-distortion function of a source for entanglement transmission is bounded below by the entanglement information rate-distortion function, i.e.

\[
R(D) \geq R^e(D). 
\]

(4.1)

This gives a lower bound on the required description rate for the source, for each level of tolerable distortion. One
might worry that this rate-distortion bound is not tight because of the same peculiarly quantum properties of coherent information which require modifications to the straightforward quantum analogue of the noisy coding theorem [5]. However, the fact that general encodings are used in deriving the bound makes me doubt that the failure of data pipelining is relevant. The possible relevance of the superadditivity of the coherent information is discussed in Appendix A.

To facilitate comparison with the classical case, I have endeavored to make the presentation parallel, as much as possible, that of the classical bound in [4]. The proof I will give uses two lemmas. First, we need the convexity of the information rate-distortion function:

**Lemma 1** $R^I(D)$ is a nonincreasing, convex function of $D$; that is, $D_1 < D_2 \Rightarrow R^I(D_1) \geq R^I(D_2)$ and $R^I(\lambda D_1 + (1 - \lambda)D_2) \leq \lambda R^I(D_1) + (1 - \lambda)R^I(D_2)$ where $0 \leq \lambda \leq 1$.

**Proof**: Nondecrease: As $D$ increases, the domain of the minimization in the definition of $R^I(D)$ becomes larger (or at least no smaller). Convexity: Let $(R_1, D_1)$ and $(R_2, D_2)$ be points on the information rate-distortion curve, and let $E_1$ and $E_2$ be operations achieving the minimum in the definition of $R^I(D)$ for $D = D_1$ and $D = D_2$ respectively. Consider $E_\lambda := \lambda E_1 + (1 - \lambda)E_2$. Since the entanglement distortion is linear in the operation, this operation has distortion $D_\lambda = \lambda d(E_1) + (1 - \lambda)d(E_2)$. Since $R^I(D_\lambda)$ is the minimum of the coherent information over operations, $R^I(D_\lambda) \leq I_c(\rho, E_\lambda)$. And since $I_c$ is convex in the operation, this is less than $\lambda I_c(\rho, E_1) + (1 - \lambda)I_c(\rho, E_2) = \lambda R^I(D_1) + (1 - \lambda)R^I(D_2)$.

The only property of the distortion that was used in this proof was the linearity of the distortion in the operation; hence it applies to any quantum rate-distortion function defined using a distortion measure with this property, including the one defined using average pure-state fidelity.

The second lemma we need is that the coherent information for a process on a composite state is greater than or equal to the total of the “marginal coherent informations” for the reductions of the process and the initial state to the subsystems.

**Lemma 2**

$$I_c(\rho^{(n)}, E^{(n)}) \geq \sum_i I_c(\rho_i, E_i^{(n)})$$

Here the definition of the reduced operation $E_i^{(n)}$ is the same as that of $T_i^{(n)}$ in (2.4), except that $D^{(n)}$ is omitted. $\rho_i$ is the marginal density operator of the $i$-th system.

**Proof**: The lemma obviously follows from the two-system case:

$$I_c(\rho^{(2)}, E^{(2)}) \geq I_c(\rho_1, E_1^{(2)}) + I_c(\rho_2, E_2^{(2)})$$

If we model this in the usual way, by purifying $Q_1$ into $R_1$ and $Q_2$ into $R_2$, adjoining an initially pure environment $E$ and effecting the operation $E^{(2)}$ by a unitary interaction $U^{Q_1 Q_2 E}$, this becomes:

$$S(\rho^{Q_1 Q_2}) - S(\rho^{R_1:Q_1,R_2:Q_2}) \geq S(\rho^{Q_2}) + S(\rho^{R_1:Q_1}) - S(\rho^{R_2:Q_2})$$

(4.3)

which may be rewritten

$$S(\rho^{R_1:Q_1}) + S(\rho^{R_2:Q_2}) - S(\rho^{Q_1 Q_2}) \geq S(\rho^{Q_1}) + S(\rho^{Q_2}) - S(\rho^{Q_1 Q_2})$$

(4.4)

The quantity appearing in this last form is the sum of the marginal entropies of two subsystems, minus the joint entropy of the composite system; it is a quantity which can be larger in quantum theory than it can in classical theory, due to entanglement. (It is often termed the quantum mutual information because of the formal similarity to the classical mutual information $I(A) + I(B) - I(AB)$ [12], [13], but should not be confused with the coherent information, which plays the quantum information-theoretic role analogous to mutual information in the present context.) In this form, the inequality says that this excess of marginal over joint entropies is reduced if we ignore (trace over) parts of each of the subsystems. This follows from strong subadditivity, as we may show by rewriting it yet again as:

$$S(\rho^{R_1:Q_1,R_2:Q_2}) + S(\rho^{Q_1}) + S(\rho^{Q_2}) \leq S(\rho^{Q_1 Q_2}) + S(\rho^{R_1:Q_1}) + S(\rho^{R_2:Q_2})$$

(4.5)

In this form, it follows from two applications of strong subadditivity [14]. We start with a case of strong subadditivity for the three systems $R_1$, $Q_1$, and $R_2:Q_2$:

$$S(\rho^{R_1:Q_1,R_2:Q_2}) + S(\rho^{Q_1}) \leq S(\rho^{R_1:Q_1}) + S(\rho^{R_2:Q_2})$$

Adding $S(\rho^{Q_2})$ to both sides gives:

$$S(\rho^{R_1:Q_1,R_2:Q_2}) + S(\rho^{Q_1}) + S(\rho^{Q_2}) \leq S(\rho^{R_1:Q_1}) + S(\rho^{R_2:Q_2}) + S(\rho^{Q_2})$$

The last two terms on the right hand side are then bounded above by another application of strong subadditivity in the form $S(\rho^{R_1:Q_1,R_2:Q_2}) + S(\rho^{R_2:Q_2}) \leq S(\rho^{Q_1 Q_2}) + S(\rho^{R_2:Q_2})$, giving (4.5).

**Theorem 1** Let $(E^{(n)}, D^{(n)})$ be a $(2^n R_n n)$ rate-distortion code with distortion $D$. Then $R \geq R^I(D)$.

**Proof**: I give the proof as a chain of inequalities and equalities, followed by notes justifying each inequality. I use the notation $\rho^{(n)} = E(\rho^{(n)})$.

$$nR \geq S(\rho^{(n)})$$

(4.6)

$$\geq S(\rho^{(n)}) - S_c(\rho^{(n)}, E^{(n)}) \equiv I_c(\rho^{(n)}, E^{(n)})$$

(4.7)

$$\geq I_c(\rho^{(n)}, D^{(n)} E^{(n)})$$

(4.8)
\[ \geq \sum_i I_c(\rho, T_i^{(n)}) \]  
\[ \geq \sum_i R^d(d(\rho, T_i^{(n)})) \equiv n \sum_i \frac{1}{n} R^d(d(\rho, T_i^{(n)})) \]  
\[ \geq n R^d(\sum_i \frac{1}{n} d(\rho, T_i^{(n)})) \equiv n R^d(D). \]  

(4.6) holds because \( nR \) is the log of the dimension of an \( n \)-block of channel Hilbert space, which constitutes an upper bound to the von Neumann entropy of a density operator on that space. (4.7) follows from the positivity of entropy exchange, (4.8) from the data processing inequality, (4.9) from Lemma 2, the superadditivity of coherent information compared to marginal coherent information, (4.10) follows from the definition of the entanglement information rate-distortion function, and (4.11) from Lemma 1, the convexity of the rate-distortion function.

\section*{V. CONCLUSIONS}

I have introduced quantum rate-distortion theory for the transmission of the entanglement of i.i.d. density operator sources, and established a lower bound on the rate at which such transmission can be accomplished, as a function of the per-transmission distortion rate tolerated. This bound bears the same relation to the coherent information as the similar classical bound, which is known to be achievable, bears to the mutual information. This is of great interest, since the coherent information is conjectured to play the same role (with the technical complication of superadditivity) in the theory of the capacity of the noisy quantum channel as the mutual information does in the theory of the capacity of the noisy classical channel. Heuristic arguments similar to the classical case therefore suggest that the quantum bound given by minimum coherent information may also be achievable. I have not shown the achievability of compression to the bound (2.10) in this note. I expect the techniques required for noisy channel coding may help, although rate-distortion may be more difficult as we cannot rely on bounds that only become tight for fidelities near one. In the actual transmission protocols considered in rate-distortion, the "noise"-like element is only truncation to a smaller space, so one might think this is likely to be much easier to deal with than a general channel operation. However, the expression (2.10) conjectured to give the achievable rate-distortion function involves a general operation. In the classical case the proof of achievability involves considerations very similar to those involved for noisy channel coding. I suspect the same will be true in the quantum case.

\section*{ACKNOWLEDGMENTS}

I thank the Office of Naval Research (grant No. N00014-93-1-0116), the U.S. NSF (grant PHY-972614), and the Institute for Scientific Interchange Foundation (I.S.I.), Turin, Italy, and Elsay-Bailey for financial support. I thank Michael Nielsen for valuable discussions.

\section*{APPENDIX A: ADDITIVITY OF THE INFORMATION RATE-DISTORTION FUNCTION}

The capacity of a noisy quantum channel for transmission of quantum sources entangled with a reference system, is likely to be given by the maximum, over input density operators, of the coherent information. This maximum, however, is known not to be additive over successive transmissions through the channel (when the input density matrices to the tensor product of successive channel uses are unrestricted). Because of this, the upper bound to channel capacity given by coherent information requires a large-block limit:

\[ \lim_{n \to \infty} (1/n) \max_{\rho^{(n)}} I_c(\rho^{(n)}, A^{(n)}). \]  

(1)

The "obvious" rate-distortion analogue of the superadditivity of the maximum coherent information would be the subadditivity of the minimum coherent information at given distortion, i.e. the subadditivity of the information rate-distortion (2.10). Thus one might conjecture that \( R(D) \) is actually given by

\[ \liminf_{n \to \infty} (1/n) \min_{A^{(n)} \leq D} I_c(\rho^{(n)}, A^{(n)}). \]  

(2)

Theorem 1 implies that (2.10) cannot be both subadditive and achievable. In fact, it is additive:

\textbf{Proposition 1}

\[ \min_{A^{(n)} \leq D} I_c(\rho^{(n)}, A^{(n)})/n = \min_{A \leq D} I_c(\rho, A). \]  

(3)

\textbf{Proof:} Lemma 2 implies that

\[ I_c(\rho^{(n)}, A^{(n)}) \geq \sum_i I_c(\rho, A_i^{(n)}). \]  

(4)

Also, the distortion caused by an operation \( A^{(n)} \) depends only on the marginal operations \( A_i \); it gives rise to. From these two facts it follows that only operations \( A^{(n)} \) having the tensor product form \( A_1 \otimes A_2 \otimes \cdots \otimes A_n \), with \( A_i \neq A_j \) need be considered in the minimization; a general block operation \( A^{(n)} \) has the same average distortion as the tensor product of its marginal operations, with coherent information at least as great as this tensor product’s. In other words, to show additivity we need only show
can be achieved with all $A_i = A_j$. That this can be done follows from the fact that given any set of operations $A_i, i = 1, \ldots, n$, satisfying the distortion constraint $\sum_i d(\rho, A_i) \leq D$ the operation

$$((1/n) \sum_{i=1}^n A_i)^\otimes n$$

will (by the linearity of the distortion in the operation) have the same distortion, but (by the convexity of the coherent information in the operation) lower coherent information.

Since the structure $T^{(n)} := D^{(n)} E^{(n)}$ was not used in them, lines (4.8)–(4.11) in the proof of Theorem 1 constitute an (essentially equivalent) proof of this.