Introduzione alla teoria dell'informazione classica e quantistica

Simone Bisogno, Francesco Mogavero

19 febbraio 2020

Indice

1 Introduzione

Una delle più grandi innovazioni del secolo scorso è legata alla formulazione di un modello matematico che descrivesse il processo di codifica, decodifica e trasmissione dell'informazione, sia che quest'ultima avvenga su un canale senza rumore, cosiddetto *ideale*, sia che essa avvenga su un canale rumoroso. Lo studio di quella che è poi diventata la disciplina scientifica della *teoria dell'informazione* ha permesso lo sviluppo, fondato su principi matematici, di moderni algoritmi di compressione e di trasmissione dell'informazione; ciò ha portato ad una memorizzazione di dati efficiente che minimizzasse lo spazio occupato nei dispositivi di memoria di massa e una trasmissione dell'informazione che ottimizzasse l'uso del canale fisico di telecomunicazione.
Claude Shannon può considerarsi il pioniere in questo settore di ricerca, avendo definito formalmente il concetto di *entropia di una sorgente*[1], legato alla quantità di informazione attesa dalla prossima trasmissione di un carattere, ed avendo contestualmente dimostrato come tale valore si leghi alla *minima* quantità di bit da usare per codificare in binario ogni carattere, in modo da ottenere una codifica che rappresenti con elevata *fedeltà* quasi tutte le sequenze emettibili. Nel 1995, il fisico e teorico dell'informazione Benjamin Schumacher ha inoltre dimostrato come tali concetti si applicano anche al mondo quantistico, qualora la sorgente emetta dell'informazione rappresentata da stati quantici piuttosto che da caratteri di un alfabeto e l'alfabeto della codifica sia composto da *qubit* (quantum bit) invece che da bit.

Il lavoro proposto cerca di ripercorrere tali eventi storico-scientifici. Partendo dal mondo classico, nel Capitolo 2 sono definite l'entropia associata a emissioni provenienti da una singola variabile aleatoria e l'entropia congiunta associata a due o più variabili, assieme ad alcuni concetti legati a quest'ultima, come l'entropia condizionata e la mutua informazione. Per ogni concetto, accanto alla definizione formale, ne sono enunciate le principali proprietà.
Nel Capitolo 3 si dimostra il teorema di Shannon, il quale ha lo scopo di trovare una limitazione superiore ed inferiore al numero di bit necessari per codificare le sequenze emettibili da M sorgenti i.i.d.[2]. Shannon giunse alla conclusione che se ci si limita a voler codificare e decodificare *con successo* solo le sequenze emanabili con maggiore probabilità (ignorando la quasi totalità delle altre) allora *esiste* almeno una codifica a *lunghezza fissa* per la quale è *necessario* e *sufficiente* un numero di bit che è circa M volte l'entropia di una singola sorgente. L'aver ignorato quelle che si considerano come le sequenze *meno probabili* produce un errore di trasmissione associata all'impossibilità di codificare e decodificare *correttamente* tali sequenze: nel caso classico ciò viene definito come *errore di codifica-decodifica*. La probabilità di occorrenza di quest'ultimo è limitato da un *upper bound*, ed inoltre, per sequenze sufficientemente lunghe, tende a 0.
In aggiunta, si dimostra che non è possibile definire alcun algoritmo di codifica che, considerando come insieme delle sequenze *rilevanti* un *qualunque* sottoinsieme delle sequenze emanabili da M sorgenti i.i.d., riesca contemporaneamente ad

[1] Per sorgente è intesa una singola variabile aleatoria composta da t valori discreti emessi in accordo ad una distribuzione di probabilità nota sia al codificatore che al decodificatore.

[2] Indipendenti ed Identicamente Distribuite.

utilizzare un numero di bit inferiore alla quantità definita precedentemente ed a mantenere la probabilità di occorrenza dell'errore di codifica-decodifica al di sotto di un upper bound. Al contrario, per sequenze sufficientemente lunghe l'errore di codifica-decodifica si verifica con probabilità tendente ad 1.
Nel Capitolo 4 si studia l'analogo quantistico di tutto ciò che è stato enunciato e dimostrato nel Capitolo 2: nel dettaglio, si definisce formalmente l'*entropia di von Neumann*, l'analogo quantistico dell'entropia associata ad una singola variabile aleatoria, e gli equivalenti quantistici dell'entropia classica congiunta, condizionata e della mutua informazione classica. Anche in questo caso, oltre ai concetti si provvede all'illustrazione delle principali proprietà, ponendo accento sulle analogie e differenze con il mondo classico. Si descrive in seguito una possibile interpretazione degli algoritmi di codifica e di decodifica nel mondo quantistico: la codifica può essere vista come una *porta logica* che opera su M registri di input rappresentanti la sequenza emessa dalle sorgenti; la decodifica è la porta logica inversa. In questo caso è possibile considerare gli M registri come l'analogo quantistico delle M variabili, o sorgenti, i.i.d. del caso classico; ciascun registro è una *sorgente quantistica*.
Infine, nel Capitolo 5 si dimostra il teorema di Schumacher. Esso parte dall'analogo quantistico dell'assunzione fatta da Shannon: si va a considerare il sottospazio di Hilbert composto dagli stati quantici che saranno trasmessi con maggiore probabilità. Schumacher giunse alla conclusione che se si considerano M sorgenti quantistiche i.i.d. allora è possibile comprimere gli M stati quantici che esse trasmetteranno in un numero di qubit che è circa M volte l'entropia di von Neumann associata alla singola sorgente. Analogamente al caso classico, anche qui vi è un errore di codifica-decodifica associato alla trasmissione di uno stato situato nel *complemento ortogonale* del sottospazio considerato: tale errore è inversamente legato al concetto di *quantum fidelity* tra il valore dei registri dati in input alla porta logica di codifica (ovvero lo stato da codificare) e quello dei registri contenenti l'ouput della porta logica di decodifica (ovvero lo stato decodificato). Come nel caso classico, qualora siano usati un numero di qubit pari ad almeno M volte l'entropia di von Neumann della sorgente allora esiste almeno una procedura che codifica e decodifica senza errori tutti gli stati quantici del sottospazio che contiene tutti gli stati che saranno trasmessi con maggiore *probabilità* e per la quale la fidelity è limitata inferiormente. Al contrario, se si intende utilizzare un numero di qubit inferiore a tale quantità, si dimostra che per M sufficientemente grande non esiste alcun algoritmo capace di codificare con successo un *generico* sottospazio dello spazio di Hilbert di definizione dei registri mantenendo una fidelity sufficientemente elevata. In quest'ultimo caso, al crescere di M la quantum fidelity si riduce asintoticamente a 0.

2 Entropia classica

L'entropia classica è stata definita da Claude Shannon nell'articolo "*A mathematical Theory of Communication*" [1]. Intuitivamente, essa rappresenta la misura dell'incertezza della trasmissione dell'informazione – rappresentata come una sequenza di caratteri – tra un dispositivo emittente (sorgente) e un dispositivo ricevente (ricevitore).

Fissato $t \geq 1$, sia X una sorgente (o registro) che emette caratteri associati ad un alfabeto sorgente $\mathcal{X} = \{x_1, x_2, \ldots, x_t\}$, con una distribuzione di probabilità $P(\mathcal{X}) = \{p(x_1), p(x_2), \ldots, p(x_t)\}$ associata all'emissione di ciascun carattere. L'entropia della sorgente X (o *entropia di Shannon*) è definita come

$$H(X) \stackrel{\text{def}}{=} - \sum_{\substack{i=1 \\ p(x_i)>0}}^{t} p(x_i) \log_2 p(x_i). \tag{1}$$

Si dimostra che $\forall X \; H(X) \in [0, log_2 t]$. La minima entropia si ottiene con una sorgente *costante* (o *certa*); al contrario, una sorgente *uniforme* produce la massima entropia.

2.1 Entropia relativa

Date due distribuzioni di probabilità, $P(\mathcal{X}) = \{p(x_1), p(x_2), \ldots, p(x_t)\}$ e $Q(\mathcal{X}) = \{q(x_1), q(x_2), \ldots, q(x_t) \colon q(x_i) \geq 0 \;\; \forall i \in \{1, \ldots, t\}\}$, si definisce l'*entropia relativa di $P(\mathcal{X})$ rispetto a $Q(\mathcal{X})$* (conosciuta anche come *divergenza informazionale* o *divergenza di Kullback–Leibler*) come

$$D(p \mid\mid q) \stackrel{\text{def}}{=} \sum_{\substack{i=1 \\ p(x_i)>0}}^{t} p(x_i) \log_2 \frac{p(x_i)}{q(x_i)}. \tag{2}$$

Semanticamente, questo valore indica quanto due distribuzioni di probabilità siano diverse. Esso soddisfa le seguenti due proprietà:

$$\begin{aligned} &D(p \mid\mid q) \geq 0 \;\; \forall P(\mathcal{X}), Q(\mathcal{X}), \\ &D(p \mid\mid q) = 0 \iff p(x_i) = q(x_i) \;\; \forall i \in \{1, 2, \ldots, t\}. \end{aligned} \tag{3}$$

La divergenza informazionale non è una metrica, poiché non rispetta né la disuguaglianza triangolare né la relazione di simmetria. Pertanto, date due generiche distribuzioni $P(\mathcal{X})$ e $Q(\mathcal{X})$ non è detto che $D(p \mid\mid q) = D(q \mid\mid p)$.

Dal punto di vista della teoria dell'informazione, $D(p \mid\mid q)$ è il valore del numero di *bit extra* usati in media per la codifica di ogni carattere sorgente quando un

algoritmo di codifica entropico (come ad esempio l'algoritmo di Huffman o l'arithmetic coding) riceve in input una scorretta distribuzione di probabilità, $P(\mathcal{X})$, associata alla sorgente invece della reale distribuzione di probabilità, $Q(\mathcal{X})$, con la quale sono emessi i simboli dell'alfabeto sorgente.

Si dimostra, infatti, che tali algoritmi sono *asintoticamente ottimali* (dal punto di vista della compressione) quando ricevono in input l'esatta distribuzione di probabilità legata alle emissioni della sorgente ed una sequenza di caratteri da codificare sufficientemente lunga, mentre l'*overhead* medio nella lunghezza della codifica, causato da un errore nella scelta della distribuzione di probabilità, è rappresentato proprio dalla divergenza informazionale.

2.2 Entropia congiunta, entropia condizionata e mutua informazione

Scelti $l, p \geq 1$, si considerino ora due sorgenti X e Y, con alfabeto sorgente rispettivamente $\mathcal{X} = \{x_1, x_2, \ldots, x_l\}$ e $\mathcal{Y} = \{y_1, y_2, \ldots, y_p\}$. Sia $p(x_i, y_j)$ la probabilità congiunta che $X = x_i$ e $Y = y_j$ $\forall i \in \{1, \ldots, l\}$ $\forall j \in \{1, \ldots, p\}$. Si definisce *entropia congiunta* la quantità

$$H(X, Y) \stackrel{\text{def}}{=} -\sum_x \sum_y p(x, y) \cdot \log_2 p(x, y). \tag{4}$$

Essa è il valore dell'incertezza legato alla trasmissione di due caratteri, il primo proveniente da X, il secondo da Y. Dal fatto che la probabilità congiunta è simmetrica ($p(x, y) = p(y, x)$) ne consegue che anche l'entropia congiunta rispetta tale proprietà, come sarà dimostrato in seguito.

In un contesto in cui si hanno $N \geq 2$ emissioni da parte di N sorgenti non indipendenti, è necessario considerare l'*entropia condizionata* e la *mutua informazione*. Sia $N = 2$, consideriamo sorgenti X e Y definite per la (4), con X legata alla prima emissione ed Y, dipendente da X, legata alla seconda. L'entropia di Y *condizionata* da X, definita come

$$H(Y \mid X) \stackrel{\text{def}}{=} H(Y, X) - H(X), \tag{5}$$

rappresenta la stima dell'incertezza condizionata del ricevitore su Y.

Facendo riferimento alle medesime sorgenti, la *mutua informazione* tra X ed Y è il valore

$$I(X : Y) \stackrel{\text{def}}{=} H(X) + H(Y) - H(X, Y). \tag{6}$$

Equivalentemente, può essere espressa come:

$$
\begin{aligned}
I(X:Y) &= H(X) - H(X \mid Y), \\
I(X:Y) &= H(Y) - H(Y \mid X).
\end{aligned}
\tag{7}
$$

La mutua informazione indica quanta informazione su X si ottiene dal ricevere un carattere proveniente da Y e viceversa.

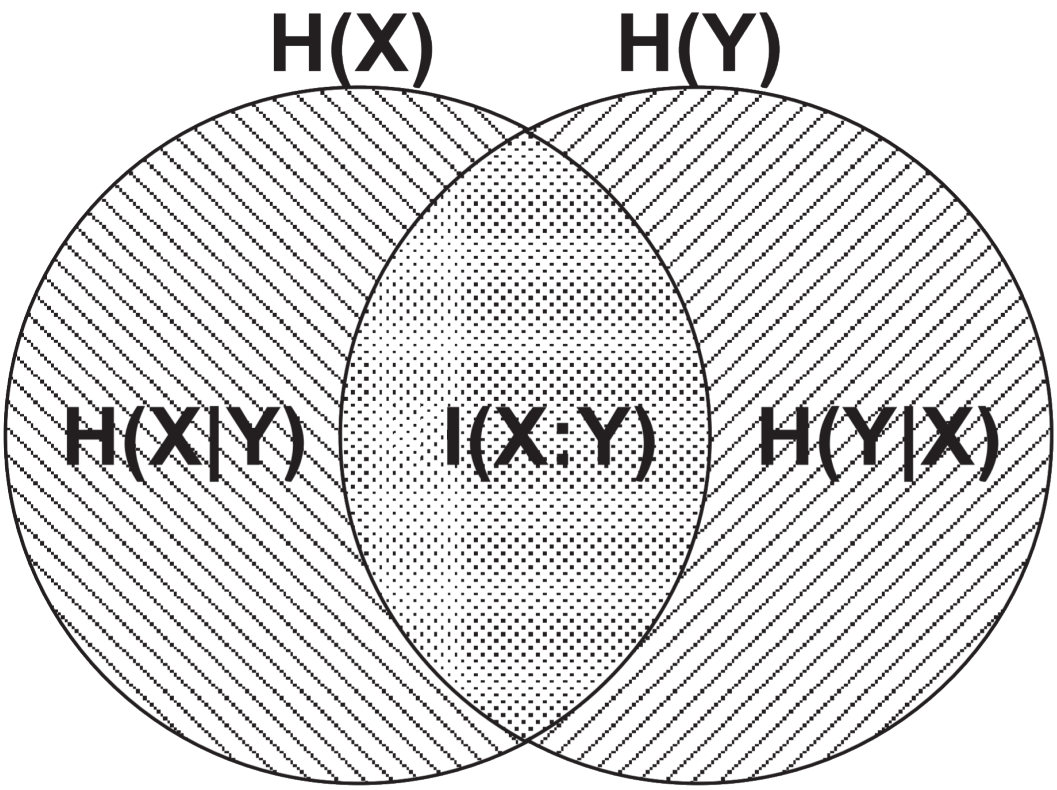

Figura 1: Relazione tra le diverse entropie.

2.3 Proprietà

L'entropia classica gode di alcune proprietà:

1) **Riflessività**: $H(X,Y) = H(Y,X), I(X:Y) = I(Y:X)$

2) **Non-negatività**: $H(Y \mid X) \geq 0 \implies I(X:Y) \leq H(Y)$ con l'uguaglianza se e solo se Y sia funzione di X

3) $H(X) \leq H(X,Y)$ con l'uguaglianza se e solo se Y è funzione di X

4) **Subadditività**: $H(X,Y) \leq H(X) + H(Y)$ con l'uguaglianza se e solo se X e Y sono due variabili aleatorie indipendenti

5) $H(X \mid Y) \leq H(Y)$ quindi $I(X:Y) \geq 0$ con l'uguaglianza se e solo se X e Y sono due variabili aleatorie indipendenti

6) **Subadditività forte**: $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$ con l'uguaglianza se e solo se $Z \to X \to Y$ forma una catena di Markov

7) **il condizionamento riduce l'entropia**: $H(X \mid Y,Z) \leq H(X \mid Y)$

Dimostriamo brevemente le proprietà enunciate.

1) Tenuto conto del fatto che la probabilità congiunta è riflessiva, si ha

$$\begin{aligned} & H(X,Y) = \\ & = -\textstyle\sum_x \sum_y p(x,y) \log_2 p(x,y) = \\ & = -\textstyle\sum_x \sum_y p(y,x) \log_2 p(y,x) = \\ & = H(Y,X). \end{aligned}$$

2) Poiché $p(x,y) = p(x)p(y \mid x)$ si ha

$$\begin{aligned} & H(X,Y) = \\ & = -\textstyle\sum_x \sum_y p(x,y) \log_2 p(x)p(y \mid x) = \\ & = -\textstyle\sum_x p(x) \log_2 p(x) - \sum_{xy} p(x,y) \log_2 p(y \mid x) = \\ & = H(X) - \textstyle\sum_{xy} p(x,y) \log_2 p(y \mid x). \end{aligned}$$

Quindi $H(Y \mid X) = -\sum_{xy} p(x,y) \log_2 p(y \mid x)$. Vale ovviamente che $-\log_2 p(y \mid x) \geq 0$ quindi $H(Y \mid X) \geq 0$; l'uguaglianza è soddisfatta se e solo se Y è funzione deterministica di X.

3) Segue da 2).

4) Per dimostrare la subadditività ci avvaliamo ancora del fatto che $\log_2 x \leq (x-1)/\ln 2, \forall x \in \mathbb{R}^+$, con l'uguaglianza se solo se $x = 1$. Scopriamo che

$$\begin{aligned} & \textstyle\sum_{x,y} p(x,y) \log_2 \frac{p(x)p(x)}{p(x,y)} \leq \\ & \leq \textstyle\frac{1}{\ln 2} \sum_{x,y} p(x,y) \left(\frac{p(x)p(y)}{p(x,y)} - 1 \right) = \\ & = \textstyle\frac{1}{\ln 2} \sum_{x,y} p(x)p(y) - p(x,y) = \\ & = \textstyle\frac{1-1}{\ln 2} = \\ & = 0 \end{aligned}$$

da cui segue la subadditività. Notiamo che l'uguaglianza si raggiunge se e solo se $p(x,y) = p(x)p(y), \forall x, y$, ossia la disuguaglianza triangolare si satura se e solo se X e Y sono indipendenti.

5) Segue dalla subadditività e dalle definizioni.

6) Per la dimostrazione della subadditività forte si usa la stessa tecnica vista in 4), con un livello di difficoltà leggermente superiore.

7) Intuitivamente, ci aspettiamo che l'incertezza su X, noti i valori di Y e Z, sia inferiore all'incertezza su X, noto soltanto Y; formalmente, il risultato per cui il condizionamento riduce l'entropia è equivalente a

$$H(X,Y,Z) - H(Y,Z) \leq H(X,Y) - H(Y)$$

che è una riformulazione della disuguaglianza di subadditività forte.

3 Teorema della codifica sorgente

3.1 Il problema della codifica sorgente

Nello stesso articolo ove ha definito la nozione di entropia, Claude Shannon ne ha mostrato anche un primo, e fondamentale, utilizzo pratico. L'obiettivo perseguito e raggiunto da Shannon è stato la minimizzazione del numero di bit necessari per codificare in binario le sequenze emettibili da M sorgenti i.i.d., per poi trasmetterle su un canale binario senza rumore, o *ideale*, e decodificarle lato ricevitore. Formalmente, si consideri $t \geq 1$. Il teorema rappresenta una sorgente con una variabile aleatoria X:

$$X \stackrel{\text{def}}{=} \left(\begin{array}{c} x_1, \quad x_2, \ \ldots \ , x_t \\ p(x_1), \ p(x_2), \ \ldots, p(x_t) \end{array} : t \geq 1 \right). \tag{8}$$

L'insieme $\mathcal{X} = \{x_1, x_2, \ldots x_t\}$ è l'alfabeto della sorgente, mentre $\{p(x_1), p(x_2), \ldots p(x_t)\}$ è una distribuzione di probabilità per $\mathcal{X}$. Osservando M emissioni da parte di $M \geq 1$ variabili aleatorie i.i.d. definite in accordo a (8), è possibile considerare la sequenza degli M caratteri così ottenuti come un *messaggio* lungo M:

$$w = X_1 \ldots X_M. \tag{9}$$

Per semplificare il calcolo, si assume s.l.g.[3] che tutte le variabili aleatorie siano uguali:

$$X_1 = X_2 = \cdots = X_M \stackrel{\text{def}}{=} X. \tag{10}$$

Dunque il generico messaggio emesso risulta essere:

$$w = \underbrace{X \ldots X}_{\text{M volte}}. \tag{11}$$

Chiarita la composizione delle sorgenti, l'attenzione si sposta alla definizione di una coppia di funzioni, c e d. La funzione di codifica c si occupa di rappresentare messaggi lunghi M caratteri sorgente in N bit, mentre la funzione di decodifica d effettua l'operazione inversa:

$$\begin{aligned} c\colon \mathcal{X}^M &\to \{0,1\}^N \\ w &\mapsto c(w), \end{aligned} \tag{12}$$

$$\begin{aligned} d\colon \{0,1\}^N &\to \mathcal{X}^M \\ \gamma &\mapsto d(\gamma). \end{aligned} \tag{13}$$

La funzione di codifica presa in esame indica un generico codificatore a *lunghezza fissa*. Qualora debbano essere codificati dei messaggi di lunghezza inferiore ad

[3] Senza ledere generalità.

M sarà aggiunto del padding; al contrario, quando devono essere rappresentati messaggi di lunghezza superiore ad M si provvederà a suddividerli in blocchi di taglia M. Affinché le operazioni di codifica e decodifica avvengano correttamente, è necessaria la condizione

$$\forall w \in \mathcal{X}^M \;\; d(c(w)) = w. \tag{14}$$

Volendo rappresentare in binario tutti i possibili messaggi, ogni algoritmo di codifica necessita di almeno $N = \left\lceil \log_2 |\mathcal{X}|^M \right\rceil = \lceil M \log_2 |\mathcal{X}| \rceil$ bit. In altri termini, devono essere usati in media $\lceil \log_2 |\mathcal{X}| \rceil$ bit per la codifica di ognuno degli M caratteri di un generico messaggio. Con il suo lavoro Shannon ha dimostrato che è possibile ridurre il valore di N fino al limite inferiore $N \simeq M \cdot H(X)$, se si è disposti ad introdurre un errore di codifica-decodifica. La probabilità di occorrenza di tale errore resta comunque limitata superiormente e al crescere di M tende esponenzialmente a 0. Ogni carattere viene in tal modo codificato in media con $H(X)$ bit.

Essendo $\forall X \; H(X) \in [0, log_2|\mathcal{X}|]$, ne consegue che il risparmio in termini di numero di bit utilizzati per codificare messaggi lunghi M può essere sostanziale, ed è *adattivo* rispetto alla casualità della sorgente; usando invece un numero di bit inferiore alla quantità $M \cdot H(X)$, si dimostra che non è possibile limitare superiormente la probabilità di occorrenza dell'errore, che al crescere di M diventa 1.

3.2 Dimostrazione dell'enunciato

Dimostrazione. Si parte con la definizione di un sottoinsieme delle $|\mathcal{X}|^M$ sequenze emettibili dalla sorgente, detto l'insieme delle *sequenze ϵ-tipiche relative ad X*. Sia $\epsilon > 0$:

$$\mathcal{T}_{M,\epsilon}(X) \stackrel{\text{def}}{=} \{w \in \mathcal{X}^M \colon 2^{-M(H(X)+\epsilon)} \leq p(w) \leq 2^{-M(H(X)-\epsilon)}\} \tag{15}$$

da cui consegue la definizione equivalente:

$$\begin{aligned}
&2^{-M(H(X)+\epsilon)} \leq p(w) \leq 2^{-M(H(X)-\epsilon)} \implies \\
&-M(H(X)+\epsilon) \leq \log_2 \, p(w) \leq -M(H(X)-\epsilon) \implies \\
&H(X) - \epsilon \leq \tfrac{1}{M} \log_2 \, \tfrac{1}{p(w)} \leq H(X) + \epsilon \implies \\
&-\epsilon \leq \tfrac{1}{M} \log_2 \, \tfrac{1}{p(w)} - H(X) \leq \epsilon \implies \\
&|\tfrac{1}{M} \log_2 \, \tfrac{1}{p(w)} - H(X)| \leq \epsilon \implies \\
&|\tfrac{1}{M} \log_2 \, \tfrac{1}{\prod_{i=1}^{M} p(x_i)} - H(X)| \leq \epsilon \implies \\
&|\tfrac{1}{M} \, \textstyle\sum_{i=1}^{M} \log_2 \tfrac{1}{p(x_i)} - H(X)| \leq \epsilon,
\end{aligned}$$

$$\mathcal{T}_{M,\epsilon}(X) \stackrel{\text{def}}{=} \{w \in \mathcal{X}^M \colon \Big|\frac{1}{M}\sum_{i=1}^{M} \log_2 \frac{1}{p(x_i)} - H(X)\Big| \leq \epsilon\} \tag{16}$$

□

Sono ora enunciate e successivamente dimostrate due interessanti proprietà legate a $\mathcal{T}_{M,\epsilon}(X)$.

1) Sia $\epsilon > 0$. Per ogni $\delta > 0$ e per M intero sufficientemente grande:

$$P(\mathcal{T}_{M,\epsilon}(X)) \stackrel{\text{def}}{=} \sum_{w \in \mathcal{T}_{M,\epsilon}(X)} p(w) \geq 1 - \delta \tag{17}$$

o equivalentemente

$$P(\overline{\mathcal{T}_{M,\epsilon}(X)}) \stackrel{\text{def}}{=} \sum_{w \notin \mathcal{T}_{M,\epsilon}(X)} p(w) \leq \delta. \tag{18}$$

2) Sia $\epsilon > 0$ e $\delta > 0$. Per M sufficientemente grande:

$$(1-\delta)2^{n(H(X)-\epsilon)} \leq |\mathcal{T}_{M,\epsilon}(X)| \leq 2^{n(H(X)+\epsilon)} \tag{19}$$

Dimostrazione. 1) Si consideri una variabile aleatoria Y legata alla X definita in (8).

$$Y \stackrel{\text{def}}{=} \left(\begin{array}{c} -\log_2 x_1, -\log_2 x_2, \ldots, -\log_2 x_t \\ p(x_1),\ p(x_2),\ \ldots,\ p(x_t) \end{array} : t \geq 1 \right). \tag{20}$$

Risulta $\mathbf{E}[Y] = H(X)$. Inoltre $\frac{1}{M}\sum_{i=1}^{M} log_2 \frac{1}{p(x_i)}$ è la media aritmetica relativa ad M esperimenti i.i.d. relativi alla variabile Y.
Dato $w = x_1 \ldots x_M$, applicando la *legge debole dei grandi numeri* si ottiene:

$$P\Big(\lim_{M\to\infty} \mathcal{T}_{M,\epsilon}(X)\Big) = P\Big(\lim_{M\to\infty} \Big|\sum_{i=1}^{M} \log_2 \frac{1}{p(x_i)} - H(X)\Big| \leq \epsilon\Big) = 1 \quad \forall \epsilon > 0 \tag{21}$$

o, equivalentemente

$$\lim_{M\to\infty} P\Big(\mathcal{T}_{M,\epsilon}(X)\Big) = \lim_{M\to\infty} P\Big(\Big|\sum_{i=1}^{M} \log_2 \frac{1}{p(x_i)} - H(X)\Big| \leq \epsilon\Big) = 1 \quad \forall \epsilon > 0. \tag{22}$$

Dato $\epsilon > 0$ ed $M \to \infty$ il valore di δ è 0.
Partendo dall'equazione (22) è possibile effettuare considerazioni algebriche riguardanti la definizione di limite convergente ad un valore finito. Si fissi $\epsilon > 0$; sfruttando la definizione di limite, vale

$$\forall \delta > 0 \, \exists M_0 : \forall M > M_0 \colon \left| P\Big(\Big| \sum_{i=1}^{M} log_2 \frac{1}{p(x_i)} - H(X) \Big| \leq \epsilon \Big) - 1 \right| \leq \delta. \tag{23}$$

Risulta dunque:

$$\forall \delta > 0 \, \exists M_0 : \forall M > M_0 \colon P\Big(\Big| \sum_{i=1}^{M} \log_2 \frac{1}{p(x_i)} - H(X) \Big| \leq \epsilon \Big) \geq 1 - \delta \tag{24}$$

il che completa la prova. □

Conseguenze. È stata provata la (17), mostrando come al crescere di M la generica sequenza emessa dalle M sorgenti i.i.d. è quasi certamente ϵ-*tipica*, indipendentemente dal valore di ϵ piccolo a piacere. Ciò significa che, sotto le medesime condizioni, la δ delle equazioni (17) e (18) tende a 0. In generale, per ogni possibile valore di ϵ e di δ è possibile trovare un M abbastanza grande in modo tale che la probabilità di estrazione di una sequenza lunga M che non sia ϵ-*tipica* sia al più δ.

Dimostrazione. 2) Si cerca un limite superiore alla cardinalità dell'insieme delle sequenze ϵ-*tipiche*.

$$\begin{aligned} & 1 \geq P(\mathcal{T}_{M,\epsilon}(X)) = \\ & = \textstyle\sum_{w \in \mathcal{T}_{\mathcal{M},\epsilon}} p(w) \geq \sum_{w \in \mathcal{T}_{\mathcal{M},\epsilon}} 2^{-M(H(X)+\epsilon)} = \\ & = 2^{-M(H(X)+\epsilon)} \textstyle\sum_{w \in \mathcal{T}_{\mathcal{M},\epsilon}} 1 = \\ & = 2^{-M(H(X)+\epsilon)} |\mathcal{T}_{M,\epsilon}(X)| \end{aligned}$$

Risulta quindi:

$$|\mathcal{T}_{M,\epsilon}(X)| \leq 2^{M(H(X)+\epsilon)} \tag{25}$$

Viene ora dimostrata l'esistenza di un limite inferiore alla medesima quantità:

$$\begin{aligned} & 1 - \delta \leq P(\mathcal{T}_{M,\epsilon}(X)) = \\ & \textstyle\sum_{w \in \mathcal{T}_{\mathcal{M},\epsilon}} p(w) \leq \sum_{w \in \mathcal{T}_{\mathcal{M},\epsilon}} 2^{-M(H(X)-\epsilon)} = \\ & = 2^{-M(H(X)-\epsilon)} \textstyle\sum_{w \in \mathcal{T}_{\mathcal{M},\epsilon}} 1 = \\ & = 2^{-M(H(X)-\epsilon)} |\mathcal{T}_{M,\epsilon}(X)| \end{aligned}$$

Dove la prima diseguaglianza segue dalla (17) (con M sufficientemente grande, come mostrato in (24)). Risulta dunque il limite inferiore:

$$|\mathcal{T}_{M,\epsilon}(X)| \geq (1-\delta)2^{M(H(X)-\epsilon)} \tag{26}$$

In conclusione, dalle equazioni (25) e (26), segue:

$$(1-\delta)2^{M(H(X)-\epsilon)} \leq |\mathcal{T}_{M,\epsilon}(X)| \leq 2^{M(H(X)+\epsilon)} \tag{27}$$

Il che completa la prova. □

Per minimizzare il valore di N, l'idea consiste nel fissare $\epsilon > 0$ e $\delta > 0$: il secondo parametro rappresenta la probabilità di errore massimale desiderata. In base a quanto scelto, vi è un valore M_0, ottenuto dalla (24), stante ad indicare a quanto debba essere il minimo valore di M da usare per il generico codificatore a lunghezza fissa definito in (12). Si decide di costruire la funzione di codifica c opportunamente, affinché la funzione parziale limitata alle sottosequenze *ϵ-tipiche* contenute nell'insieme $\mathcal{T}_{M,\epsilon}(X)$ (con $M \geq M_0$) debba essere iniettiva. Alla *quasi totalità delle sequenze* presenti nel complemento viene invece assegnata una codifica binaria già usata per codificare un'altra sequenza. Chiaramente, tali sequenze, una volta codificate, non sono decodificabili correttamente, in quanto non vale la (14): la decodifica della codifica restituirà un valore diverso dalla sequenza originale.

Considerando un canale *ideale*, non sono previsti errori dovuti a fattori esterni, come ad esempio gli errori di *trasmissione*; pertanto ogni eventuale errore può essere associato univocamente all'emissione di una sequenza per cui si verifica l'errore di codifica-decodifica precedentemente definito. Da tutto ciò ne consegue che

$$P(e_{cd}(c)) \leq P\Big(\overline{\mathcal{T}_{M,\epsilon}(X)}\Big) \tag{28}$$

dove con $e_{cd}(c)$ è indicato l'errore di codifica-decodifica associato all'algoritmo di codifica c.

Dalla (18) si deduce che $P(e_{cd}(c))$ è limitata superiormente dal δ scelto. In aggiunta, dalla (25) si ottiene che sono sufficienti $N = M(H(x)+\epsilon) \simeq MH(X)$ bit per codificare in maniera iniettiva tutto l'insieme $\mathcal{T}_{M,\epsilon}(X)$. In altri termini, in media, la codifica di ogni carattere di una sequenza di lunghezza M può essere composta da $H(X)$ bit.
In maniera più formale, viene ora dimostrato quanto manca per completare il teorema; in particolare, viene dapprima mostrato come, qualora si scelgano ϵ, δ ed $M > M_0$ come descritto, siano effettivamente sufficienti $MH(X)$ bit per codificare tutto l'insieme $\mathcal{T}_{M,\epsilon}(X)$ mantenendo valida la condizione

$$P(e_{cd}(c)) \leq \delta. \tag{29}$$

In seguito si dimostra che con lo stesso valore di ϵ, qualora si intenda costruire un algoritmo che usi un numero medio di bit per codificare ciascun carattere inferiore a $H(X)$ bit, non è possibile limitare superiormente la probabilità di occorrenza dell'errore di codifica-decodifica; al contrario, al crescere di M, questa tende ad 1.

1) **Definizione formale di un algoritmo di codifica.** Dopo aver scelto opportunamente ϵ, δ ed $M > M_0$, si scelga un numero medio di bit per la codifica di ogni carattere pari ad $\mathcal{A} \geq H(X)+2\epsilon$. Il numero di bit usati dalla codifica è, per definizione, $N \stackrel{\text{def}}{=} \lfloor M \cdot \mathcal{A} \rfloor > M(H(X) + \epsilon)$. Considerando la (25), vale

$$|\mathcal{T}_{M,\epsilon}(X)| \leq 2^{M(H(X)+\epsilon)} \leq 2^N. \tag{30}$$

Quindi è sufficiente usare un siffatto numero N di bit per codificare tutte le sequenze ϵ-*tipiche* mantenendo una probabilità di errore limitata. Si costruisca, dunque, c in modo tale che a ciascuna sequenza ϵ-*tipica* corrisponda una stringa binaria univoca. Si consideri, inoltre, l'insieme $\mathcal{G}_M$, definito come l'insieme delle al più 2^N sequenze sorgenti che possono essere *mappate* con distinte stringhe binarie dall'algoritmo appena definito su codominio $\{0,1\}^N$ e che, pertanto, possono essere decodificate senza errore:

$$\mathcal{G}_M \stackrel{\text{def}}{=} \{w \in \mathcal{X}^M \colon d(c(w)) = w\}. \tag{31}$$

Per costruzione di c risulta $\mathcal{T}_{M,\epsilon}(X) \subseteq \mathcal{G}_M$. La probabilità di errore di codifica-decodifica relativa all'algoritmo appena costruito è pertanto

$$P(e_{cd}(c)) \stackrel{\text{def}}{=} P(\overline{\mathcal{G}_M}) = \sum_{w \notin \mathcal{G}_M} p(w). \tag{32}$$

Da $\mathcal{T}_{M,\epsilon}(X) \subseteq \mathcal{G}_M$ segue che

$$P(\mathcal{G}_M) \geq P(\mathcal{T}_{M,\epsilon}(X)), \tag{33}$$

$$P(\overline{\mathcal{G}_M}) \leq P(\overline{\mathcal{T}_{M,\epsilon}(X)}). \tag{34}$$

Sfruttando (18) e (34) si ottiene infine

$$P(\overline{\mathcal{G}_M}) \leq \delta \tag{35}$$

il che completa la prova.

Il concetto chiave è che, scelto ϵ, qualora si voglia mantenere un errore di codifica-decodifica ampio al più δ, è possibile considerare un opportuno M sufficientemente grande e adottare la strategia discussa. Si dimostra ora che la codifica proposta è *ottima*: scelto ϵ per M sufficientemente grande non risulta possibile definire un limite superiore alla probabilità di errore di codifica, che cresce al crescere di M fino ad assumere valore 1.

2) **Dimostrazione dell'ottimalità.** Sia invece $\mathcal{A} < H(X) - \epsilon$.
Sia c un qualunque algoritmo di codifica c che rappresenti in maniera iniettiva un *qualunque* sottoinsieme $\mathcal{S}$ di $\mathcal{X}^M$. Si osservi innanzitutto che per $\mathcal{G}_M$ vale ($\forall \epsilon > 0$, $\forall M$ intero positivo)

$$\mathcal{G}_M \subseteq (\Sigma^M \setminus \mathcal{T}_{M,\epsilon}(X)) \cup (\mathcal{G}_M \cap \mathcal{T}_{M,\epsilon}(X)). \tag{36}$$

Di conseguenza

$$\begin{aligned} &P(\mathcal{G}_M) = \\ &= \textstyle\sum_{w \in \mathcal{G}_M} p(w) = \\ &= \textstyle\sum_{w=a_1 \dots a_M \in \mathcal{G}_M} p(a_1) \dots p(a_M) \leq \\ &\leq (1 - \textstyle\sum_{w=a_1 \dots a_M \in \mathcal{T}_{M,\epsilon}(X)} p(a_1) \dots p(a_M)) + \\ &(\textstyle\sum_{w=a_1 \dots a_M \in \mathcal{G}_M \cap \mathcal{T}_{M,\epsilon}(X)} p(a_1) \dots p(a_M)). \end{aligned} \tag{37}$$

Si studi il secondo termine ottenuto: per la definizione di $\mathcal{T}_{M,\epsilon}(X)$ vale

$$\begin{aligned} &\textstyle\sum_{w=a_1 \dots a_M \in \mathcal{G}_M \cap \mathcal{T}_{M,\epsilon}(X)} p(a_1) \dots p(a_M) \leq \\ &\leq \textstyle\sum_{w=a_1 \dots a_M \in \mathcal{G}_M \cap \mathcal{T}_{M,\epsilon}(X)} 2^{-M(H(X)-\epsilon)} \leq \\ &\leq \textstyle\sum_{w=a_1 \dots a_M \in \mathcal{G}_M} 2^{-M(H(X)-\epsilon)} = \\ &= 2^{-M(H(X)-\epsilon)} |\mathcal{G}_M|. \end{aligned}$$

Si ottiene dunque

$$P(\mathcal{G}_M) \leq (1 - \sum_{w=a_1 \dots a_M \in \mathcal{T}_{M,\epsilon}(X)} p(a_1) \dots p(a_M)) + 2^{-M(H(X)-\epsilon)} |\mathcal{G}_M|. \tag{38}$$

Dalla definizione di $\mathcal{G}_M$ (31):

$$|\mathcal{G}_M| \leq 2^N. \tag{39}$$

Poiché $N = \lfloor M \cdot \mathcal{A} \rfloor$, e per scelta di $\mathcal{A}$, segue

$$|\mathcal{G}_M| \leq 2^N = 2^{\lfloor M \cdot \mathcal{A} \rfloor} < 2^{M(H(X)-\epsilon)}. \tag{40}$$

Si studi ora il comportamento di (38) al crescere di M. Come già osservato, la (38) rappresenta la probabilità di codifica-decodifica avvenuta senza errore secondo un generico algoritmo c. Dalla (40) si ottiene

$$\lim_{M \to \infty} 2^{-M(H(X)-\epsilon)} |\mathcal{G}_M| = 0, \tag{41}$$

dunque

$$\begin{aligned} &\lim_{M\to\infty} P(\mathcal{G}_M) \leq \\ &\leq \lim_{M\to\infty}(1 - \sum_{w=a_1\dots a_M\in\mathcal{T}_{M,\epsilon}(X)} p(a_1)\dots p(a_M)) = \\ &= 1 - \lim_{M\to\infty} \sum_{w=a_1\dots a_M\in\mathcal{T}_{M,\epsilon}(X)} p(a_1)\dots p(a_M)). \end{aligned} \tag{42}$$

Per la definizione di $\mathcal{T}_{M,\epsilon}(X)$ in (16), per la (17) e la (21) si sa che, al crescere di M, le sequenze emesse sono quasi certamente ϵ-*tipiche*. Formalmente:

$$\lim_{M\to\infty} \sum_{w=a_1\dots a_M\in\mathcal{T}_{M,\epsilon}(X)} p(a_1)\dots p(a_M) = 1. \tag{43}$$

In conclusione:

$$\lim_{M\to\infty} P(\mathcal{G}_M) = 0. \tag{44}$$

È stato dunque dimostrato che scegliere un numero medio A di bit inferiore ad $H(X)$ non è sufficiente, al crescere di M, per codificare un qualunque sottoinsieme delle sequenze emettibili mantenendo la probabilità di occorrenza dell'errore di codifica-decodifica limitata.
Si può concludere che $N = H(X)$ rappresenta il valore ideale per ottenere una codifica a lunghezza con errore limitato (che all'infinito è 0) e che minimizzi quanto più possibile la quantità di bit necessaria.

4 Entropia quantistica

L'entropia di Shannon misura l'incertezza relativa a una distribuzione di probabilità classica; gli stati quantistici sono descritti in maniera simile, con gli operatori di densità al posto delle distribuzioni di probabilità. In questa sezione, generalizziamo la definizione di entropia di Shannon per gli stati quantistici.

Von Neumann definisce l'*entropia* di uno stato quantistico ρ come

$$S(\rho) = -\operatorname{Tr}(\rho \log_2 \rho). \tag{45}$$

Se λ_x sono gli autovalori di ρ, allora la (45) può essere riscritta come

$$S(\rho) = -\sum_x \lambda_x \log_2 \lambda_x, \tag{46}$$

in cui poniamo per definizione $0 \log_2 0 \equiv 0$, come per l'entropia di Shannon; la formulazione (46) ben si presta ai calcoli. Per esempio, l'operatore densità misto in uno spazio d-dimensionale, I/d, è $\log_2 d$.

Da qui in poi, quando si parlerà di entropia, sarà chiaro dal contesto se si farà riferimento all'entropia di Shannon o a quella di von Neumann.

4.1 Entropia relativa quantistica

È utile definire la versione quantistica dell'entropia relativa, come fatto per quella di Shannon. Supponiamo ρ e σ operatori densità; l'*entropia relativa* di ρ a σ è definita come

$$S(\rho \,||\, \sigma) \equiv \operatorname{Tr}(\rho \log_2 \rho) - \operatorname{Tr}(\sigma \log_2 \sigma). \tag{47}$$

Come per l'entropia relativa classica, quella quantistica può essere, a volte, infinita; in particolare, è non-negativa: questo concetto è a volte noto come *disuguaglianza di Klein*:

Teorema (Disuguaglianza di Klein)**.** *L'entropia relativa quantistica è non negativa,*

$$S(\rho \,||\, \sigma) \geq 0, \tag{48}$$

con l'uguaglianza se e solo se $\rho = \sigma$.

Il contenuto di questo capitolo è tratto dal Capitolo 11 di [2].

4.2 Proprietà

L'entropia quantistica gode anch'essa di alcune proprietà:

1) L'entropia è non negativa, al più nulla se lo stato è puro.

2) In uno spazio di Hilbert d-dimensionale, l'entropia è al più $\log_2 d$; è uguale a $\log_2 d$ se e solo se il sistema è in uno stato completamente misto $\rho = I/d$[4].

3) Relazione con l'entropia di Shannon. Sia ρ una matrice densità e $P = \{p_1, \ldots, p_t\} : t \geq 1$ la distribuzione di probabilità classica associata a ρ. Risulta:
$$S(\rho) = -\operatorname{Tr}(\rho \log_2 \rho) = H(P) \tag{49}$$

4) Supponiamo che un sistema composto AB sia in stato puro; allora $S(A) = S(B)$

5) Supponiamo p_i probabilità e gli stati ρ_i con supporto su sottospazi ortogonali; allora
$$S\left(\sum_i p_i \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i) \tag{50}$$

6) **Entropia congiunta**: supponiamo p_i probabilità, $\mid i \rangle$ stati ortogonali di un sistema A e ρ_i un qualsiasi insieme di operatori densità per un altro sistema B; allora
$$S\left(\sum_i p_i \mid i \bigtimes i \mid \otimes \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i) \tag{51}$$

Dimostrazione. 1) Segue dalla definizione.

2) Il risultato segue dalla non-negatività dell'entropia relativa, $0 \leq S(\rho \mid\mid I/d) = S(\rho) + \log_2 d$.

3) Dalla decomposizione di Schmidt, sappiamo che gli autovalori degli operatori densità di due sistemi A e B sono gli stessi; l'entropia è determinata completamente dagli autovalori, per cui $S(A) = S(B)$.

4) Siano λ_i^j ed $\mid e_i^j \rangle$ rispettivamente autovalori e autovettori di ρ_i; si osservi che $p_i \lambda_i^j$ e $\mid e_i^j \rangle$ sono rispettivamente autovalori e autovettori di $\sum_i p_i \rho_i$, quindi
$$\begin{aligned} &S\left(\textstyle\sum_i p_i \rho_i\right) = \\ &= -\textstyle\sum_i p_i \log_2 p_i - \sum_i p_i \sum_j \lambda_i^j \log_2 \lambda_i^j = \\ &= H(p_i) + \textstyle\sum_i p_i S(\rho_i). \end{aligned}$$

[4]Assumendo che ρ sia rappresentata in una base in cui sia diagonale.

5) Segue immediatamente dal risultato precedente.

□

Decomposizione di Schmidt

Supponiamo $\mid \psi \rangle$ stato puro di un sistema composto AB; allora, esistono gli stati ortonormali $\mid i_A \rangle$ per il sistema A e gli stati ortonormali $\mid i_B \rangle$ per il sistema B tali che

$$\mid \psi \rangle = \sum_i \lambda_i \mid i_A \rangle \mid i_B \rangle \tag{52}$$

dove i λ_i sono numeri reali non-negativi che soddisfano $\sum_i \lambda_i^2 = 1$ noti come *coefficienti di Schmidt*.
Questo risultato è molto utile. Si consideri la seguente conseguenza: sia $\mid \psi \rangle$ stato puro di un sistema composto AB; allora, per la decomposizione di Schmidt, $\rho^A = \sum_i \lambda_i^2 \mid i_A \bigtimes i_A \mid$ e $\rho^B = \sum_i \lambda_i^2 \mid i_B \bigtimes i_B \mid$ così che gli autovalori di ρ^A e di ρ^B siano identici (λ_i^2 per entrambi). Molte proprietà dei sistemi quantistici sono determinate completamente dagli autovalori dell'operatore densità ridotto del sistema, così che per lo stato puro di un sistema composto, tali proprietà siano le stesse per entrambi i sistemi.

4.3 Entropia congiunta, condizionale e mutua informazione quantistiche

Analogamente all'entropia di Shannon, è possibile definire sia entropia congiunta sia entropia condizionale sia mutua informazione, nel caso quantistico, per sistemi composti.

L'entropia congiunta $S(A, B)$ per un sistema composto di due componenti A e B è definito come

$$S(A, B) \equiv -\operatorname{Tr}(\rho^{AB} \log_2 \rho^{AB}) \tag{53}$$

dove ρ^{AB} è la matrice densità del sistema AB.

Definiamo l'entropia condizionale come

$$S(A \mid B) \equiv S(A, B) - S(B) \tag{54}$$

e la mutua informazione come

$$\begin{aligned} &S(A\colon B) \equiv S(A) + S(B) - S(A, B) = \\ &= S(A) - S(A \mid B) = \\ &= S(B) - S(B \mid A) \end{aligned} \tag{55}$$

Alcune proprietà dell'entropia di Shannon non valgono per quella di von Neumann, con conseguenze interessanti nella teoria dell'informazione quantistica. Per esempio, per due variabili aleatorie X e Y, la disuguaglianza $H(X) \leq H(X,Y)$ vale: intuitivamente, ciò ha senso perché non possiamo essere più incerti sullo stato di X di quanto non lo siamo dello stato congiunto di X e Y. Tuttavia, nel caso quantistico, quest'intuizione non vale: consideriamo un sistema AB di due qubit in uno stato entangled $(\mid 00 \rangle + \mid 11 \rangle)/\sqrt{2}$. Questo stato è puro, quindi $S(A,B) = 0$; d'altra parte, il sistema A ha come operatore densità $I/2$, quindi ha entropia pari a 1. Un altro modo per esporre questo risultato è che, per questo sistema, la quantità $S(B \mid A) = S(A,B) - S(A)$ è negativa.

4.4 Misura ed entropia

Come si comporta l'entropia di un sistema quantistico se vi eseguiamo una misura? In maniera non molto sorprendente, la risposta dipende dal tipo di misura eseguita; in ogni caso, ci sono assunti generali sul comportamento dell'entropia al riguardo.

Supponiamo, per esempio, che una misura proiettiva descritta dai proiettori P_i sia eseguita su un sistema quantistico ma non ne leggiamo mai il risultato: se lo stato del sistema prima della misurazione era ρ, allora lo stato dopo la misurazione è dato da

$$\rho' = \sum_i P_i \rho P_i. \tag{56}$$

Il risultato seguente mostra che l'entropia non è mai decrementata da questo processo ma rimane costante solo se lo stato non è cambiato dalla misurazione.
Teorema (La misura proiettiva aumenta l'entropia)**.** *Supponiamo P_i insieme completo di proiettori ortogonali e ρ operatore densità. L'entropia dello stato $\rho' = \sum_i P_i \rho P_i$ del sistema dopo la misura è almeno grande quanto quella originale*

$$S(\rho') \geq S(\rho) \tag{57}$$

con uguaglianza se e solo se $\rho = \rho'$.

4.5 Subadditività

Supponiamo di avere due sistemi quantistici distinti A e B con stato congiunto ρ^{AB}; allora, l'entropia congiunta per i due sistemi soddisfa le disuguaglianze

$$S(A,B) \leq S(A) + S(B) \tag{58}$$

$$S(A,B) \geq |S(A) - S(B)|. \tag{59}$$

La prima prende il nome di *subadditività* per l'entropia di von Neumann e vale l'uguaglianza se e solo se i due sistemi A e B non sono correlati, ossia $\rho^{AB} = \rho^A \otimes \rho^B$.

La seconda prende il nome di *disuguaglianza triangolare* (o, a volte, disuguaglianza di *Araki-Lieb*): è l'analogo quantistico della disuguaglianza $H(X,Y) \geq H(X)$ per l'entropia di Shannon.

4.6 Subadditività forte

Le disuguaglianze subadditiva e triangolare per due sistemi quantistici possono essere estese a tre sistemi. Il risultato è noto come *subadditività forte* ed è uno dei risultati più importanti e utili della teoria dell'informazione quantistica. La disuguaglianza afferma che per tre sistemi quantistici A, B, C allora

$$S(A,B,C) + S(B) \leq S(A,B) + S(B,C) \tag{60}$$

5 Teorema quantistico della codifica sorgente

5.1 Il problema quantistico della codifica sorgente

L'analogo quantistico del teorema presentato da Shannon nel 1948 è stato definito da Benjamin Schumacher nel 1995 [4]. Con il suo lavoro, l'autore ha provato l'esistenza di una stretta correlazione tra il problema della codifica classico e l'analogo quantistico. Quest'ultimo consiste, proprio come nel caso classico, nella realizzazione di una coppia di algoritmi: un algoritmo Γ di codifica e un algoritmo Π di decodifica. Nel dettaglio, Γ effettua la codifica di sequenze composte da M stati quantistici prodotti da $M \geq 1$ *sistemi quantistici i.i.d.* in $N \geq 1$ qubit, cercando di minimizzare il valore di N al fine di ottimizzare l'utilizzo del *canale quantistico* di comunicazione; l'algoritmo Π, al contrario, riporta l'insieme di N qubit che riceve nello stato e nello spazio originali.

La più piccola quantità d'informazione *quantistica* codificata in qubit e trasmessa non è più un carattere di un alfabeto (informazione *classica*) ma consiste in uno stato *puro* di un sistema quantistico sorgente.

Formalmente, dato l'ensemble $\{p_x, \mid \psi_x \rangle\}$ (con $\{p_x\}$ distribuzione di probabilità), una sorgente *quantistica* (o sistema quantistico sorgente) è una matrice densità ρ definita in uno spazio di Hilbert $\mathcal{H}_{Sorg}$

$$\rho \stackrel{\text{def}}{=} \sum_x p_x \mid \psi_x \rangle\langle \psi_x \mid . \tag{61}$$

Una sorgente emette un solo stato; in particolare, lo stato $\mid \psi_x \rangle$ viene emesso con probabilità classica p_x. Nel teorema di Schumacher sono considerate M sorgenti i.i.d. $\rho_1 \dots \rho_M$: in altri termini, esse condividono una comune distribuzione di probabilità classica $\{p_x\}$.

Similmente al caso classico, anche quantisticamente viene assunto s.l.g. che le M sorgenti siano uguali: $\rho_1 = \cdots = \rho_M$. La generica sequenza di M stati (o *messaggio*) da codificare è rappresentabile come uno stato *puro* e *prodotto* di $\mathcal{H}_{Sorg}^{\otimes M}$

$$\underbrace{\mid \psi \rangle \otimes \cdots \otimes \mid \psi \rangle}_{M volte} = \mid \psi \rangle^{\otimes M}. \tag{62}$$

La matrice densità associata all'emissione di un messaggio di lunghezza M tra tutti i messaggi possibili è

Il contenuto del quinto capitolo è tratto da [2], [3], [4].

$$\rho^{\otimes M} = \underbrace{\rho \otimes \cdots \otimes \rho}_{M volte}. \tag{63}$$

Matematicamente, è possibile definire le funzioni di codifica e decodifica, Γ e Π, come

$$\begin{aligned} &\Gamma \colon \mathcal{H}_{Sorg}^{\otimes M} \to \mathcal{H}^{\otimes N} \\ &\quad | \ \psi^{\otimes M} \ \rangle \mapsto \Gamma(| \ \psi^{\otimes M} \ \rangle) \end{aligned} \tag{64}$$

$$\begin{aligned} &\Pi \colon \mathcal{H}^{\otimes N} \to \mathcal{H}_{Sorg}^{\otimes M} \\ &\quad | \ \gamma \ \rangle^{\otimes N} \mapsto \Pi(| \ \gamma \ \rangle^{\otimes N}). \end{aligned} \tag{65}$$

ove $\mathcal{H}^{\otimes N}$ è lo spazio degli N qubit sfruttati per la codifica e $| \ \gamma \ \rangle^{\otimes N}$ è uno stato *puro* nello spazio dei qubit $\mathcal{H}^{\otimes N}$. Ragionando in termini di circuiti logici, è possibile considerare la codifica Γ e la decodifica Π come due porte logiche. La prima ha in input M *registri* in parallelo, preparati in modo che lo stato del registro i sia "scelto" in accordo a ρ_i, ed effettua una codifica arbitraria ed iniettiva in uno stato composto da N qubit. La seconda ha in input dei registri in N qubit ed esegue l'operazione inversa. La dimensione dello spazio di Hilbert della sorgente è pari a $dim(\mathcal{H}_{Sorg})^M$ mentre quella dello spazio di Hilbert dei qubit è 2^N.

Per rappresentare (codificare) l'intero spazio $\mathcal{H}_{Sorg}^{M}$ in qubit in maniera *iniettiva* e *invertibile* sono necessari almeno $N \geq \lceil M \log_2 dim(H_\rho) \rceil$ qubit. Il teorema di Schumacher, similmente al teorema di Shannon del mondo classico, prova che è possibile abbassare il numero di qubit necessari per la codifica sino a $N \simeq M \cdot S(\rho)$ se si accetta una probabilità di errore di codifica-decodifica arbitrariamente bassa. Nel caso quantistico, tale probabilità consiste in un abbassamento della *fedeltà* (*quantum fidelity*) del *canale quantistico* $\Phi = \Pi \, \Gamma$.

Per codificare in tal modo, si cerca il *sottospazio ϵ-tipico* di $\mathcal{H}_{Sorg}^{\otimes M}$ relativo a ρ, ove sono contenute tutte le sequenze di stati emittibili con maggiore probabilità. L'algoritmo di codifica, opportunamente realizzato, trasforma in maniera iniettiva ogni sequenza in tale sottospazio in una successione di N qubit distinta; per le sequenze di stati situate nel *complemento ortogonale*[5] del sottospazio individuato viene invece usata una successione di qubit già sfruttata: ciò renderà impossibile decodificare tali sequenze, diminuendo la fedeltà del canale.

In sintesi, si dimostra che è possibile definire un algoritmo di codifica con $N \simeq M \cdot S(\rho)$ qubit utilizzati con *perdita di fedeltà* limitata superiormente da un parametro δ; in tal modo, ogni registro viene rappresentato in media con circa $S(\rho)$ qubit. Il teorema conclude, inoltre, che tutti gli algoritmi di codifica che usano un numero di qubit inferiore a $M \cdot S(\rho)$ ottengono, per M sufficientemente grande, un valore di fedeltà tendente a 0: si ha, dunque, una totale differenza tra lo stato originale e quello ricavato dal ricevitore.

[5] I due sottospazi sono in *somma diretta*.

5.2 Fedeltà quantistica

5.2.1 Fedeltà tra due stati quantistici

La fedeltà tra due stati quantistici (o *quantum state fidelity*) è una delle possibili misure (pur non essendo una metrica) legata alla loro similarità.

Formalmente, dati due operatori densità ρ e σ definiti all'interno di uno stesso spazio di Hilbert $\mathcal{H}$, la *fedeltà* $F(\rho, \sigma)$ tra ρ e σ si definisce come

$$F(\rho, \sigma) \stackrel{\text{def}}{=} \|\sqrt{\rho}\sqrt{\sigma}\|. \tag{66}$$

Nella (66) sono presenti le nozioni di radice quadrata di un operatore densità e di *norma di traccia* ($\|\bullet\|$) di una matrice.

Radice quadrata di un operatore densità

Data una matrice M rappresentante un'operatore densità, $Y = \sqrt{M}$ denota, per definizione, l'unica radice quadrata di M *semidefinita positiva*. A sua volta, una matrice B è detta operatore semidefinito positivo se vale la seguente condizione:

$$\exists \; operatore\, lineare \; T \colon B = T^\dagger T. \tag{67}$$

Ogni operatore semidefinito positivo ha autovalori positivi. Inoltre, esso è hermitiano, infatti:

$$\begin{aligned} &[B, B^\dagger] = \\ &= BB^\dagger - B^\dagger B = \\ &= (T^\dagger T)(T^\dagger T)^\dagger - (T^\dagger T)^\dagger (T^\dagger T) = \\ &= 0 \end{aligned} \tag{68}$$

ove l'ultima uguaglianza vale poiché $(AB)^\dagger = B^\dagger A^\dagger$.
In conclusione, $\sqrt{\rho}$ e $\sqrt{\sigma}$ sono hermitiane ed hanno, a loro volta, una radice quadrata ben definita. Ponendo sia $\sqrt{\rho}$ che $\sqrt{\sigma}$ in forma diagonale risulta evidente che i loro autovalori sono, rispettivamente, le radici quadrate degli autovalori di ρ e σ.

Norma di traccia di una matrice

L'operatore $\|\bullet\|$ è la *norma di traccia* della matrice, la cui definizione è:

$$\|S\| \stackrel{\text{def}}{=} \text{Tr}\,\sqrt{(S^\dagger \cdot S)} \tag{69}$$

È possibile correlare il concetto di norma di traccia di una matrice a quello di norma di un vettore di numeri complessi: infatti, considerando S in forma diagonale, si giunge a conclusione che $\|S\|$ è il valore *reale* ottenuto dalla radice quadrata della somma del modulo quadro di ciascun autovalore di S.

Una diversa equazione per la fedeltà di uno stato quantistico è direttamente derivabile dalla (66) tramite la definizione della norma di traccia (69).

$$\begin{aligned} & F(\rho, \sigma) = \\ & = \| \sqrt{\rho}\sqrt{\sigma} \| = \\ & = \mathrm{Tr} \sqrt{(\sqrt{\rho}\sqrt{\sigma})^{\dagger}(\sqrt{\rho}\sqrt{\sigma})} = \\ & = \mathrm{Tr} \sqrt{\sqrt{\sigma}\sqrt{\rho}\,\sqrt{\rho}\,\sqrt{\sigma}} = \\ & = \mathrm{Tr} \sqrt{\sqrt{\rho}\,\sqrt{\sigma}\sqrt{\sigma}\sqrt{\rho}} = \\ & = \mathrm{Tr} \sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}} \end{aligned} \tag{70}$$

dove la quarta uguaglianza vale per via dell'invarianza della traccia rispetto ad una permutazione ciclica del prodotto delle matrici.
Alcuni autori considerano invece la fedeltà come il quadrato di tale valore:

$$F(\rho, \sigma) = \left(\mathrm{Tr} \sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}} \right)^2 \tag{71}$$

D'ora in poi, per la definizione di fedeltà tra due stati quantistici si farà riferimento a (71).

La fedeltà gode di alcune proprietà fondamentali, tra cui si elencano le più significative.

1) Simmetria:

$$\begin{aligned} & F(\sigma, \rho) = \\ & = \left(\mathrm{Tr} \sqrt{\sqrt{\sigma}\rho\,\sqrt{\sigma}} \right)^2 = \\ & = \left(\mathrm{Tr} \sqrt{\sqrt{\sigma}\sqrt{\rho}\,\sqrt{\rho}\,\sqrt{\sigma}} \right)^2 = \\ & = \left(\mathrm{Tr} \sqrt{\sqrt{\rho}\,\sqrt{\sigma}\sqrt{\sigma}\sqrt{\rho}} \right)^2 = \\ & = F(\rho, \sigma) \end{aligned} \tag{72}$$

   dove la terza uguaglianza vale per via dell'invarianza della traccia rispetto ad una permutazione ciclica del prodotto delle matrici.

2) Dalla (66) e (69) si ricava che la fedeltà tra due matrici densità è un valore reale. In più, è possibile dimostrare che, dati due operatori densità ρ e σ, $0 \leq F(\rho, \sigma) \leq 1$; in particolare, l'uguaglianza a 0 si ha se e solo se ρ e σ definiscono due sottoinsiemi di vettori mutualmente ortogonali di $\mathcal{H}$, mentre l'uguaglianza a 1 se e solo se $\rho = \sigma$.

3) Se $\rho = \mid \psi \bigtimes \psi \mid$, ossia ρ è uno stato puro, allora

$$F(\mid \psi \bigtimes \psi \mid, \sigma) = \langle \psi \mid \sigma \mid \psi \rangle \tag{73}$$

Quando ρ è uno stato puro, si farà riferimento a $F(\mid \psi \bigtimes \psi \mid, \sigma)$ anche con la notazione $F(\mid \psi \rangle, \sigma)$

Dimostrazione.

$$
\begin{aligned}
& F(\mid \psi \bigtimes \psi \mid, \sigma) = \\
& = \left(\mathrm{Tr} \sqrt{\sqrt{\mid \psi \bigtimes \psi \mid} \;\; \sigma \;\; \sqrt{\mid \psi \bigtimes \psi \mid}} \right)^2 = \\
& = \left(\mathrm{Tr} \sqrt{\sqrt{\mid \psi \bigtimes \psi \mid} \;\; \sqrt{\mid \psi \bigtimes \psi \mid} \;\; \sigma} \right)^2 = \\
& = \left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid \;\; \sigma} \right)^2 = \\
& = \left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid \;\; \mid \psi \bigtimes \psi \mid \;\; \sigma} \right)^2 = \\
& = \left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid \; \sigma \; \mid \psi \bigtimes \psi \mid} \right)^2 = \\
& = \left(\mathrm{Tr} \sqrt{\mid \psi \rangle \langle \psi \mid \sigma \mid \psi \rangle \langle \psi \mid} \right)^2 = \\
& = \left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid \;\; \langle \psi \mid \sigma \mid \psi \rangle} \right)^2
\end{aligned}
$$

dove la seconda e la quinta uguaglianza valgono grazie all'invarianza della traccia rispetto ad una permutazione ciclica del prodotto delle matrici mentre la quarta per l'idempotenza dell'operatore di proiezione.
In quanto operatore densità, σ è hermitiano. Lo scalare $\langle \psi \mid \sigma \mid \psi \rangle$ è il *valore atteso*, *reale*, della misurazione dell'hermitiana σ quando il sistema quantistico è nello stato puro ψ. Pertanto

$$
\left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid \;\; \langle \psi \mid \sigma \mid \psi \rangle} \right)^2 = \langle \psi \mid \sigma \mid \psi \rangle \left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid} \right)^2 .
$$

Ci si focalizza ora su $\left(\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid} \right)^2$, dimostrando che l'espressione ha valore 1; Per far ciò, sono necessarie alcune considerazioni algebriche.

Per costruzione, gli autovalori di $\sqrt{\mid \psi \bigtimes \psi \mid}$ sono la radice quadrate degli autovalori di $\mid \psi \bigtimes \psi \mid$ (avendo considerato, per definizione, $\sqrt{\mid \psi \bigtimes \psi \mid}$ come l'unica radice semidefinita positiva di $\mid \psi \bigtimes \psi \mid$). Ogni autovalore di un proiettore è l'intero 0 oppure 1; la somma degli autovalori, ossia la traccia dell'operatore di proiezione, è pari alla dimensione del sottospazio a cui l'operatore proietta. Nel caso in esame, la matrice $\mid \psi \bigtimes \psi \mid$ proietta su un asse: la cardinalità di tale sottospazio è 1. Si può concludere che $\mid \psi \bigtimes \psi \mid$ ha $dim(\mathcal{H}) - 1$ autovalori con valore 0 ed un solo autovalore pari ad 1; la medesima situazione si presenta anche per $\sqrt{\mid \psi \bigtimes \psi \mid}$ per quanto detto in precedenza. Dunque $\mathrm{Tr} \sqrt{\mid \psi \bigtimes \psi \mid} = 1$, il che completa la dimostrazione. □

5.2.2 Fedeltà di un canale

Il concetto di fedeltà si estende anche ad un *canale quantistico*. Dati due spazi di Hilbert $\mathcal{H}_A$ e $\mathcal{H}_B$, un canale quantistico Φ è, algebricamente, una *funzione lineare* tra i due spazi:

$$\Phi\colon \mathcal{H}_A \to \mathcal{H}_B \tag{74}$$

Dal punto di vista dell'algebra delle matrici, Φ è una matrice avente $dim(\mathcal{H}_B)$ righe e $dim(\mathcal{H}_A)$ colonne. Formalmente il canale quantistico è un *operatore lineare* avente uno *scopo* aggiuntivo: trasmettere informazione quantistica tra un mittente ed un ricevitore. Questi ultimi possiedono, rispettivamente, un sistema quantistico nello spazio $\mathcal{H}_A$ ed un sistema quantistico nello spazio $\mathcal{H}_B$. Ogni canale Φ soddisfa due condizioni:

1) Φ è una funzione lineare completamente positiva;

2) Φ preserva la traccia.

Sia $H_A = H_B \stackrel{\text{def}}{=} \mathcal{H}_1$, $\Phi\colon \mathcal{H}_1 \to \mathcal{H}_1$ un canale quantistico e ρ_1 (rappresentato da una matrice densità) uno stato di $\mathcal{H}_1$ che descrive una sorgente quantistica (analogamente a quelle descritte in precedenza) che emette un singolo stato quantistico. L'assunzione fondamentale effettuata dal teorema di Schumacher è che lo stato ρ_1 è *potenzialmente* misto a causa di una *possibile* precedente relazione di *entanglement* con un sistema esterno (per esempio, uno o più sistemi quantistici, un bagno, ecc...). Si indichi con $\mathcal{H}_2$ lo spazio di Hilbert che racchiude tutti i sistemi esterni con il quale $\mathcal{H}_1$ era in interazione. Lo stato generico $\mid \psi \rangle_{12} \in \mathcal{H}_1 \otimes \mathcal{H}_2$ è *puro* e *potenzialmente entangled*; ρ_1 è invece la *matrice densità ridotta* ottenuta dall'applicazione della *traccia parziale* di $\mathcal{H}_2$ su $\mid \psi \rangle_{12}$.

$$\rho_1 = \mathrm{Tr}_2(\mid \psi \rangle_{12}) \tag{75}$$

Si assume il sottosistema $\mathcal{H}_2$ in comune sia al mittente sia al ricevitore. Il mittente deve dunque inviare uno stato $\mid \psi \rangle_1 \in \mathcal{H}_1$ (in accordo a ρ_1) attraverso il canale di comunicazione Φ. Lo scopo del mittente è quello di inviare dell'informazione relativa a ρ_1 in maniera più fedele possibile: si desidera una $F(\rho_1, \Phi\rho_1)$ massimale.

Dal punto di vista del sistema composto $\mathcal{H}_1 \otimes \mathcal{H}_2$, il canale di comunicazione è l'operatore $\Phi \otimes \mathcal{I}_{H_2}$, lo stato *puro* del sistema è $\mid \psi \rangle_{12}$. Per quanto detto, la fedeltà del sistema composto si esprime come $F(\mid \psi \rangle_{12}\ ,\ (\Phi \otimes \mathcal{I}_{H_2}) \mid \psi \rangle_{12})$. Essendo il sottosistema $\mathcal{H}_2$ non alterato dalla trasmissione, la massimizzazione della fedeltà consiste nel trasmettere ρ_1 mantenendo inalterato (*conservando*) il più possibile l'*entanglement* tra i due sottosistemi. Formalmente, si definisce *fedeltà dell'entanglement* del canale Φ rispetto a $\mid \psi \rangle_{12}$ la quantità $\mathcal{F}(\Phi, \mid \psi \rangle_{12})$ definita come

$$\mathcal{F}\left(\Phi, \mid \psi \rangle_{12}\right) \stackrel{\text{def}}{=} F\left(\mid \psi \rangle_{12}\,,\, \left(\Phi \otimes \mathcal{I}_{H_2}\right) \mid \psi \rangle_{12}\right). \tag{76}$$

Per la (73) tale valore è esprimibile, in modo equivalente, come

$$\mathcal{F}\left(\Phi, \mid \psi \rangle_{12}\right) = {}_{12}\langle \psi \mid \ \left[\left(\Phi \otimes \mathcal{I}_{H_2}\right) \mid \psi \rangle_{12}\right] \ \mid \psi \rangle_{12}. \tag{77}$$

Accanto alla nozione di *fedeltà di entanglement*, vi è quella di *fedeltà media dell'ensemble* del canale Φ rispetto a ρ_1, la cui notazione è $\hat{F}\left(\Phi, \rho_1\right)$. Quest'ultima definizione considera unicamente lo spazio $\mathcal{H}_1$. Dato l'ensemble relativo a ρ_1, $\{p_i, \mid \psi_i \rangle\} \ : i \in \{1 \ldots k\}$ con k pari al numero di vettori dell'ensemble, la *fedeltà media dell'ensemble* si esprime come segue

$$\hat{F}\left(\Phi, \rho_1\right) \stackrel{\text{def}}{=} \sum_{i=1}^{k} p_i \, F\left(\mid \psi_i \rangle_1\,,\, \Phi \mid \psi_i \rangle_1\right) \tag{78}$$

o, in maniera analoga, per la (73):

$$\hat{F}\left(\Phi, \rho_1\right) = \sum_{i=1}^{k} \ p_i \, {}_1\langle \psi_i \mid \ \left[\Phi \mid \psi_i \rangle_1\right] \ \mid \psi_i \rangle_1 \tag{79}$$

È possibile dimostrare come la fedeltà di entanglement del canale di trasmissione $\Phi \otimes \mathcal{I}_{H_2}$ rispetto allo stato puro $\mid \psi \rangle_{12}$ del sistema composto $\mathcal{H}_1 \otimes \mathcal{H}_2$ rappresenti una limitazione inferiore per la fedeltà media dell'ensemble del canale di trasmissione Φ rispetto allo stato potenzialmente misto ρ_1 del sistema $\mathcal{H}_1$

$$F\left(\Phi, \rho_1\right) \leq \hat{F}\left(\Phi, \rho_1\right). \tag{80}$$

La dimostrazione (omessa) segue dal fatto che la fedeltà dell'entanglement è una funzione convessa e dall'applicazione della disuguaglianza di Jensen.

5.3 Dimostrazione dell'enunciato

Si dimostra ora l'enunciato del teorema quantistico di codifica sorgente elaborato da Schumacher. Prima di passare al formalismo, si effettua un richiamo dei concetti utili alla formalizzazione del problema, già affrontati nei paragrafi precedenti. Si considerino messaggi composti da $M \geq 1$ stati quantistici puri, ove ciascuno stato è ottenuto dalla matrice densità ρ derivata dall'ensemble $\{p_x, \mid \psi_x \rangle\}$. Una sorgente quantistica che emette un singolo stato può essere espressa come

$$\rho = \sum_{x} p_x \mid \psi_x \bigtimes \psi_x \mid . \tag{81}$$

Per semplicità, si diagonalizzi ρ:

$$\rho = \sum_{i=1}^{t} p_i \mid \psi_i \rangle\langle \psi_i \mid \tag{82}$$

dove $t = dim(\mathcal{H}_p)$. In tal modo, gli autovettori di ψ_i sono una base ortogonale per lo spazio, mentre gli autovalori $\{p_i\}$ sono una distribuzione di probabilità classica. D'ora in poi, per definizione di ρ si sottintende la (82): tale è, al tempo stesso, la definizione di sorgente quantistica che emette uno stato quantistico.

Il teorema considera sequenze di M stati quantistici, con $M \geq 1$ valore fissato. Ciascuno stato di ogni sequenza è emesso da una sorgente i.i.d. rispetto a tutte le altre. Inoltre, si assume s.l.g. che tutte le sorgenti siano uguali. Da un punto di vista dei circuiti logici ciò equivale ad avere M registri paralleli in input, ciascuno dei quali preparato, indipendentemente dagli altri, in accordo ad una stessa matrice densità ρ.
La matrice densità associata alla sorgente che emette sequenze di M stati (un messaggio di lunghezza M) è dunque:

$$\rho^{\otimes M} = \underbrace{\rho \otimes \cdots \otimes \rho}_{M volte}. \tag{83}$$

Dalla definizione (83) si ricava che i $t \cdot M$ autovalori di $\rho^{\otimes M}$ sono tutti i possibili valori $\lambda = \lambda_1 \cdot \ldots \cdot \lambda_t$, con λ_i autovalore di ρ_i. I $t \cdot M$ autovettori, invece, sono tutti i possibili vettori avente forma $\mid \psi \rangle^{\otimes M} = \mid \psi \rangle_1 \otimes \ldots \otimes \mid \psi \rangle_t$, con $\mid \psi \rangle_i$ autovettore di ρ_i. In breve, $\rho^{\otimes M}$ è una matrice densità nello spazio ottenuto dal prodotto tensore degli spazi dei singoli caratteri della sequenza: ciascuno dei suoi autovalori di $\rho^{\otimes M}$ corrisponde alla probabilità che la corrispondente sequenza (autovettore di $\rho^{\otimes M}$) venga emessa dalla sorgente.

Come già accennato nel paragrafo 5.1, per la codifica l'autore decide di rappresentare iniettivamente in qubit tutte le sequenze presenti nel sottospazio tipico; per quelle nel complemento ortogonale viene invece usata una sequenza già sfruttata. In tal modo, come sarà dimostrato nel paragrafo 5.3.3, egli riesce a diminuire il numero di qubit necessari fino a $N \simeq M \cdot S(\rho)$ mantenendo la fedeltà media d'ensemble limitata inferiormente. Inoltre, si dimostra nel paragrafo 5.3.4 che con un numero di qubit inferiore a tale soglia la medesima fidelity, al crescere di M, si abbassa fino a 0. Nel seguente paragrafo viene esplicato il concetto di sottospazio ϵ-*tipico*, mostrando alcune proprietà interessanti ed una correlazione con le sequenze ϵ-*tipiche* del teorema classico.

5.3.1 Il sottospazio ϵ-*tipico*

Dati $\epsilon > 0$ ed $M \geq 1$, il sottospazio ϵ-*tipico* di $\mathcal{H}_{Sorg}^{\otimes M}$ relativo a ρ è quel sottospazio di $\mathcal{H}_{Sorg}^{\otimes M}$ avente come *base* l'insieme degli autovettori di $\rho^{\otimes M}$ il cui autovalore λ soddisfa l'equazione

$$2^{-M(S(\rho)+\epsilon)} \leq \lambda \leq 2^{-M(S(\rho)-\epsilon)}. \tag{84}$$

Sia $\mathcal{T}_{M,\epsilon}(\rho) \subseteq \mathcal{H}_{Sorg}^{\otimes M}$ il sottospazio ϵ-*tipico*. Dati i vettori di base del sottospazio ϵ-*tipico*, denotati con $\{|\ v_1^{\otimes M}\ \rangle, \ldots, |\ v_l^{\otimes M}\ \rangle : l \geq 1\}$, si definisce il proiettore $\mathcal{P}_{M,\epsilon}(\rho)$ al medesimo sottospazio

$$\mathcal{P}_{M,\epsilon}(\rho) \stackrel{\text{def}}{=} |\ v_1^{\otimes M} \bigtimes v_1^{\otimes M}\ | + \cdots + |\ v_l^{\otimes M} \bigtimes v_l^{\otimes M}\ |\ . \tag{85}$$

Vale

$$l \stackrel{\text{def}}{=} dim(\mathcal{T}_{M,\epsilon}(\rho)) \leq t \cdot M. \tag{86}$$

Si intende ora stabilire, analogamente al caso classico, la probabilità che un vettore in $\mathcal{H}_{Sorg}^{\otimes M}$ sia ϵ-*tipico* e trovare una limitazione inferiore ed uno superiore alla cardinalità del sottospazio (86). Quest'ultima non potrà essere più grande di $t \cdot M$.

Esaminando l'analogia tra entropia classica e quantistica (49) illustrata nel paragrafo 4.2, risulta esserci un'equivalenza tra l'entropia quantistica di una matrice densità e l'entropia classica della distribuzione di probabilità associata alla medesima matrice. Pertanto, riferendosi ad una sorgente quantistica quale quella rappresentata nell'equazione (82), è possibile sfruttare sia l'equazione di Shannon (sulla distribuzione di probabilità relativa a ρ) che quella di von Neumann (sull'intera matrice densità ρ) come definizione di entropia.

Per le considerazioni appena effettuate, la stessa sorgente quantistica può essere vista come l'analogo di una sorgente classica che emette un carattere di un alfabeto classico con una distribuzione di probabilità pari all'insieme degli autovalori di ρ. Diviene dunque possibile l'uso, anche nel caso quantistico, di alcune considerazioni legate ai concetti di entropia e di sorgente valevoli nel mondo classico ed analizzate nel capitolo 3.

Un primo risultato preliminare e di fondamentale importanza è il seguente. Sia $\epsilon > 0$ ed $M \geq 1$. Data la sorgente $\rho^{\otimes M}$ ed il proiettore al sottospazio ϵ-*tipico* relativo a ρ, $\mathcal{P}_{M,\epsilon}(\rho)$, vale

$$\mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right) = \sum_{|w\ \rangle^{\otimes M} = |w_1\ \rangle \otimes \cdots \otimes |w_M\ \rangle \in \mathcal{T}_{M,\epsilon}} p_{w_1} \ldots p_{w_M}. \tag{87}$$

Dimostrazione.

$$\begin{aligned}
&\mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\, \rho^{\otimes M}\right) = \\
&= \mathrm{Tr}\left(\ |\ v_1^{\otimes M} \bigtimes v_1^{\otimes M}\ |\ \rho^{\otimes M} + \cdots + |\ v_l^{\otimes M} \bigtimes v_l^{\otimes M}\ |\ \rho^{\otimes M}\right) = \\
&= \textstyle\sum_{i=1}^{l} \mathrm{Tr}\left(|\ v_i^{\otimes M} \bigtimes v_i^{\otimes M}\ |\ \rho^{\otimes M}\right) = \sum_{i=1}^{l} \mathrm{Tr}\left(\rho^{\otimes M}\ |\ v_i^{\otimes M} \bigtimes v_i^{\otimes M}\ |\right) = \\
&= \textstyle\sum_{i=1}^{l} \mathrm{Tr}\left(\lambda_i\ |\ v_i^{\otimes M} \bigtimes v_i^{\otimes M}\ |\right) = \sum_{i=1}^{l} \lambda_i \,\mathrm{Tr}\left(\ |\ v_i^{\otimes M} \bigtimes v_i^{\otimes M}\ |\right) = \\
&= \textstyle\sum_{i=1}^{l} \lambda_i
\end{aligned}$$

dove la prima uguaglianza vale per la (85), la seconda e la quinta per la linearità della traccia, la terza per l'invarianza della traccia rispetto ad una permutazione ciclica del prodotto delle matrici, la quarta dall'applicazione dell'equazione agli autovalori di $\rho^{\otimes M}$ (applicata su ciascun autovettore che ha in comune con $\mathcal{P}_{M,\epsilon}(\rho)$), la sesta dal fatto che il proiettore ad un asse ha traccia pari ad 1. Per la relazione (citata ad inizio paragrafo) in essere tra gli autovettori di $\rho^{\otimes M}$ e di autovettori di $\mathcal{P}_{M,\epsilon}(\rho)$ (questi ultimi sono i vettori di base di $\mathcal{T}_{M,\epsilon}(\rho)$), la proposizione è provata. □

Dati M, ϵ e ρ, per la (87) per quanto appena mostrato risulta possibile associare al valore $\mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right)$ la probabilità che uno stato di $\mathcal{H}_{Sorg}^{\otimes M}$ sia nel sottospazio ϵ-*tipico* $\mathcal{T}_{M,\epsilon}(\rho)$. In altri termini, $\mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right)$ è l'analogo quantistico dell'equazione (17) del paragrafo 3.2. Denotando con $P(\mathcal{T}_{M,\epsilon}(\rho))$ la probabilità che un vettore nello spazio $\mathcal{H}_{Sorg}^{\otimes M}$ sia ϵ-*tipico* relativo a ρ, vale

$$P(\mathcal{T}_{M,\epsilon}(\rho)) = \mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right). \tag{88}$$

Data la corrispondenza definita in precedenza tra la generica sorgente quantistica ed un'analoga classica, dalla (17) del paragrafo 3.2 e dalla (88) è possibile concludere che, dato $\epsilon > 0$, $\forall \delta > 0$ e per M intero sufficientemente grande $(\exists M_0 \in \mathcal{N} :\ \forall M \in \mathcal{N},\ M \geq M_0)$

$$\mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right) \geq 1 - \delta \tag{89}$$

o, equivalentemente, sotto le medesime condizioni

$$P(\overline{\mathcal{T}_{M,\epsilon}(\rho)}) = \mathrm{Tr}\left(\overline{\mathcal{P}_{M,\epsilon}(\rho)}\,\rho^{\otimes M}\right) \leq \delta \tag{90}$$

ove $\overline{\mathcal{P}_{M,\epsilon}(\rho)} = \mathcal{I}_{\mathcal{H}_{Sorg}^{\otimes M}} - \mathcal{P}_{M,\epsilon}(\rho)$ è il proiettore al complemento ortogonale di $\mathcal{T}_{M,\epsilon}(\rho)$: $\overline{\mathcal{T}_{M,\epsilon}(\rho)}$. L'equazione (90) è la corrispondente quantistica della (18) del paragrafo 3.2. La dimostrazione di (89) e di (90) è omessa in quanto equivalente a quella del caso classico per via delle già discusse equivalenze tra le definizioni classiche e quantistiche di sorgente, di entropia e di distribuzioni di probabilità associate alle sorgenti.

Come nel caso classico (21) e (22) del paragrafo 3.2, si può provare (la dimostrazione è omessa) che anche quantisticamente che, dato $\epsilon > 0$ ed $M \to \infty$ risulti

$$\lim_{M\to\infty} P\Big(\mathcal{T}_{M,\epsilon}(\rho)\Big) = \lim_{M\to\infty} \mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\,\rho^{\otimes M}\right) = 1 \tag{91}$$

Al tendere di M all'infinito, dunque, tutte le sequenze di lunghezza M rientrano nel sottospazio ϵ-*tipico* $\mathcal{T}_{M,\epsilon}(\rho)$ di $\mathcal{H}_{Sorg}^{\otimes M}$ rispetto a ρ. In altri termini, $\delta = 0$.

Dimostrazione. Si definiscono ora i limiti alla cardinalità del sottospazio ϵ-*tipico* (86).

Limitazione superiore:

$$\begin{aligned}
&1 \geq P(\mathcal{T}_{M,\epsilon}(\rho)) \underbrace{=}_{(88)} \\
&= \mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right) \underbrace{=}_{(87)} \\
&= \textstyle\sum_{v\in\mathcal{T}_{\mathcal{M},\epsilon}} \lambda_v \underbrace{\geq}_{(84)} \sum_{w\in\mathcal{T}_{\mathcal{M},\epsilon}} 2^{-M(S(\rho)+\epsilon)} = \\
&= 2^{-M(S(\rho)+\epsilon)} \textstyle\sum_{v\in\mathcal{T}_{\mathcal{M},\epsilon}} 1 = \\
&= 2^{-M(S(\rho)+\epsilon)} \cdot dim(\mathcal{T}_{M,\epsilon}(\rho)).
\end{aligned}$$

Risulta quindi

$$l = dim(\mathcal{T}_{M,\epsilon}(\rho)) \leq 2^{M(S(\rho)+\epsilon)}. \tag{92}$$

Viene ora calcolato una limitazione inferiore:

$$\begin{aligned}
&1-\delta \leq P(\mathcal{T}_{M,\epsilon}(\rho)) \underbrace{=}_{(88)} \\
&= \mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}\right) \underbrace{=}_{(87)} \\
&= \textstyle\sum_{v\in\mathcal{T}_{\mathcal{M},\epsilon}} \lambda_v \underbrace{\leq}_{(84)} \sum_{w\in\mathcal{T}_{\mathcal{M},\epsilon}} 2^{-M(S(\rho)-\epsilon)} = \\
&= 2^{-M(S(\rho)-\epsilon)} \textstyle\sum_{w\in\mathcal{T}_{\mathcal{M},\epsilon}} 1
\end{aligned}$$

ove la prima disuguaglianza segue dalla (89) ed è valida, dati $\epsilon > 0$ ed $\delta > 0$, per M sufficientemente grande ($M \geq M_0$). Risulta dunque, sotto le medesime condizioni, la limitazione inferiore:

$$l = dim(\mathcal{T}_{M,\epsilon}(\rho)) \geq (1-\delta)2^{M(S(\rho)-\epsilon)}. \tag{93}$$

In conclusione, dalle equazioni (92) e (93) segue, per parametri opportunamente scelti, che

$$(1-\delta)2^{M(S(\rho)-\epsilon)} \leq dim(\mathcal{T}_{M,\epsilon}(\rho)) \leq 2^{M(S(\rho)+\epsilon)}. \tag{94}$$

□

Si giunge, anche in questo caso, ad una stretta correlazione col mondo classico ((27) paragrafo 3.2).

5.3.2 Definizione formale degli algoritmi di codifica e di decodifica

Formalmente, il processo di codifica e di decodifica (o canale Φ di codifica-decodifica) consiste in una coppia di algoritmi Γ e Π, come definiti in (64) e (65). A loro volta Γ e Π possono essere implementate, dal punto di vista dei circuiti logici, come due porte logiche: una di codifica e l'altra di decodifica. L'algoritmo di codifica, Γ, riceve in input M registri in parallelo, ciascuno dei quali preparato in accordo a ρ. Esso opera effettuando dapprima una misurazione proiettiva avente lo scopo di accertare se lo stato da codificare $\mid \psi_i^{\otimes M} \rangle$ è nel sottospazio ϵ-*tipico*. Tale misurazione viene realizzata in modo da proiettare lo stato in $\mathcal{T}_{M,\epsilon}(\rho)$ o in $\overline{\mathcal{T}_{M,\epsilon}(\rho)}$ (sottospazi in somma diretta). Algebricamente, la misurazione proietta mediante $\mathcal{P}_{M,\epsilon}(\rho)$ e verifica se, dato uno stato puro $\mid \psi_i^{\otimes M} \rangle$

$$\mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle \stackrel{?}{=} \mid \psi_i^{\otimes M} \rangle \tag{95}$$

o, equivalentemente

$$\mathrm{Tr}\left(\mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid\right) \stackrel{?}{=} 1. \tag{96}$$

Poiché $\mathrm{Tr}\left(\mid a \bigtimes b \mid\right) = \langle\, b \mid a \,\rangle$, ponendo $\mid a \,\rangle = (\mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle)$ e $\langle\, b \mid = \langle\, \psi_i^{\otimes M} \mid$, la (96) si può esprimere come

$$\langle\, \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \,\rangle \stackrel{?}{=} 1. \tag{97}$$

Qualora la misurazione determini che lo stato è ϵ-*tipico* la sequenza viene *mappata* in uno stato (composto da N qubit) non usato in precedenza; nel caso opposto, lo stato del sistema viene sostituito con un altro stato puro $\mid \psi_i^{\otimes M} \rangle'$, presente nel sottospazio ϵ-*tipico*, che viene infine *mappato* in N qubit. L'algoritmo di decodifica, Π, riceve in input N registri, ciascuno dei quali rappresenta un qubit, ed effettua il *mapping* inverso.
Nello scenario in cui lo stato iniziale $\mid \psi_i^{\otimes M} \rangle$ non risulti essere nel sottospazio tipico è impossibile ottenere dalla decodifica lo stato originale. La probabilità che le equazioni (96) e (97) siano soddisfatte è descritta in (88) e (89).

In aggiunta, essendo $\mathcal{T}_{M,\epsilon}(\rho)$ e $\overline{\mathcal{T}_{M,\epsilon}(\rho)}$ in somma diretta, si conclude che ogni vettore di $\mathcal{H}_{Sorg}^{\otimes M}$ ha tutte le sue componenti nel sottospazio ϵ-*tipico* oppure nel complemento. Di conseguenza le equazioni (96) e (97) hanno solo due possibili valori di output: 1 se e solo se lo stato originale è ϵ-*tipico*, 0 se le ha tutte nel complemento ortogonale. Per la stessa motivazione l'equazione (**??**) restituisce come possibili output solo due stati: lo stato originale (se esso è nel sottospazio tipico) altrimenti il vettore nullo.

Nei successivi paragrafi viene provata la correttezza e l'ottimalità dell'algoritmo (canale) $\Phi = \Pi\,\Gamma$ proposto. In dettaglio, viene dimostrato come tale algoritmo funzioni ed è ottimo se e solo se vengono usati almeno $N \geq M(S(\rho) + \epsilon)$ qubit per la codifica. In tal modo, infatti, risulta possibile codificare in maniera

iniettiva ciascuna sequenza ϵ-*tipica*, con una fedeltà media dell'ensemble che è limitata inferiormente e, al crescere di M, tende ad 1. Al contrario, sfruttando $N \leq M(S(\rho) - \epsilon)$ qubit, si dimostra che come non possa esistere alcun algoritmo di codifica che, al crescere di M, riesca a mantenere un valore accettabile per la fedeltà media dell'ensemble, la quale diminuisce (al crescere di M) fino a diventare 0.
In tutti i casi si considera Φ come canale *ideale*. Vi è dunque assenza di rumore: ogni alterazione tra lo stato trasmesso dal trasmittente e quello ottenuto dal ricevitore è da imputare alla costruzione del canale Φ appena stabilita.

5.3.3 Dimostrazione della correttezza

Viene ora dimostrata formalmente la correttezza dell'algoritmo proposto; in particolare, viene dimostrato come è possibile limitare inferiormente la fedeltà media dell'ensemble se vengono scelti gli opportuni valori per ciascun parametro. Si proverà come per ogni possibile scelta dei parametri $\epsilon > 0$ e $\delta > 0$ esiste un intero M_0 per il quale, se

1) si pone $M \geq M_0$ come parametro stante ad indicare la lunghezza fissa delle sequenze da codificare

2) si usano $N \geq M \cdot S(\rho) + \epsilon$ qubit per la codifica

allora la *fedeltà media dell'ensemble* rispetto a $\rho^{\otimes M}$ del canale di codifica-decodifica proposto è limitata inferiormente dal valore $1 - 2 \cdot \delta$. Quest'ultima al crescere di M (e per gli stessi valori di ϵ, δ ed M_0) tende ad 1. Detto in altri termini, ciò significa che, sotto le medesime condizioni, la perdita di fedeltà media dell'ensemble è limitata superiormente dal doppio del parametro δ.
La prima delle due condizioni poste deriva dalla (88) e dalla (89) ed è importante poiché, dati ϵ e δ, scegliere un opportuno M grande fa si che quasi tutte le sequenze risultano nel sottospazio ϵ-*tipico* (89). La seconda viene invece descritta in seguito. Entrambe le condizioni sono le analoghe quantistiche di richieste poste nel caso classico dal teorema di Shannon (paragrafo 3.2).

Si richiami ora la definizione di fedeltà media dell'ensemble (79) del canale di codifica-decodifica $\Phi = \Pi\,\Gamma$ rispetto alla sorgente $\rho^{\otimes M}$. La fedeltà viene, in questo contesto, calcolata sull'ensemble composto da stati mutualmente ortogonali $\{p_i, \mid \psi_i \rangle\}$ $i \in \{1, \ldots, t \cdot M\}$

$$\hat{F}\left(\Phi, \rho^{\otimes M}\right) = \sum_{i=1}^{t \cdot M} p_i \ \langle\, \psi_i^{\otimes M} \mid \ \left(\Phi \mid \psi_i^{\otimes M} \rangle\right) \ \mid \psi_i^{\otimes M} \rangle \tag{98}$$

ove $\Phi \mid \psi_i^{\otimes M} \rangle$ rappresenta l'input che il ricevitore ottiene dal canale, ossia l'output del processo di codifica-decodifica.

Essendo la cardinalità del sottospazio ϵ-*tipico* limitata superiormente da $2^{M(S(\rho)+\epsilon)}$ (92) e per via della scelta di N ne consegue che è possibile codificare ogni vettore del sottospazio ϵ-*tipico* in maniera iniettiva. Per la costruzione del canale di codifica-decodifica, risulta

$$\begin{aligned}\Phi \mid \psi_i^{\otimes M} \rangle = \mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid \langle \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle + \\ + \rho_{i,junk} \langle \psi_i^{\otimes M} \mid \overline{\mathcal{P}_{M,\epsilon}(\rho)} \mid \psi_i^{\otimes M} \rangle\end{aligned} \tag{99}$$

ove $\rho_{i,junk}$ è lo stato del sottospazio ϵ-*tipico* che viene scelto per sostituire $\mid \psi_i^{\otimes M} \rangle$ quando la misurazione iniziale non classifica $\mid \psi_i^{\otimes M} \rangle$ nel sottospazio tipico. È importante notare che i calcoli che seguono sono indipendenti dalla scelta di $\rho_{i,junk}$. Lo scalare $\langle \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle$ ha due soli possibili valori, come già specificato per l'equazione (97). Il valore di $\langle \psi_i^{\otimes M} \mid \overline{\mathcal{P}_{M,\epsilon}(\rho)} \mid \psi_i^{\otimes M} \rangle$ è l'opposto di quello assunto dal precedente scalare.
Focalizzandosi sul primo addendo dell'equazione (99), ovvero $\mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid \langle \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle$, è possibile effettuare alcuni passaggi algebrici

$$\begin{aligned}&\mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid \left[\langle \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle\right] = \\ &= \mid \psi_i^{\otimes M} \rangle \langle \psi_i^{\otimes M} \mid \left[\langle \psi_i^{\otimes M} \mid (\mathcal{P}_{M,\epsilon}(\rho))^2 \mid \psi_i^{\otimes M} \rangle\right] = \\ &= \mid \psi_i^{\otimes M} \rangle \left[\langle \psi_i^{\otimes M} \mid (\mathcal{P}_{M,\epsilon}(\rho))^2 \mid \psi_i^{\otimes M} \rangle\right] \langle \psi_i^{\otimes M} \mid = \\ &= \mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid (\mathcal{P}_{M,\epsilon}(\rho))^2 \mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid = \\ &= \mathcal{P}_{M,\epsilon}(\rho) \left(\mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid\right)^2 \mathcal{P}_{M,\epsilon}(\rho) = \\ &= \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho)\end{aligned}$$

ove la prima e la quinta uguaglianza valgono per l'idempotenza dell'operatore di proiezione, mentre la quarta per la *commutazione* tra $\mathcal{P}_{M,\epsilon}(\rho)$ ed il proiettore $\mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid$. È possibile dunque riscrivere la (99) come

$$\begin{aligned}\Phi \mid \psi_i^{\otimes M} \rangle = \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \bigtimes \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) + \\ + \rho_{i,junk} \langle \psi_i^{\otimes M} \mid \overline{\mathcal{P}_{M,\epsilon}(\rho)} \mid \psi_i^{\otimes M} \rangle.\end{aligned} \tag{100}$$

Il primo addendo denota la *forma matriciale* dello stato ottenuto dal ricevitore quando $\mid \psi_i^{\otimes M} \rangle$ è ϵ-*tipico* ($\forall i \in \{1 \dots t \cdot M\}$); il secondo addendo rappresenta il medesimo stato quando $\mid \psi_i^{\otimes M} \rangle$ è nel complemento ortogonale.

Tornando alla (98), è possibile sostituire $\Phi \mid \psi_i^{\otimes M} \rangle$ con l'espressione a destra nella (100):

$$\begin{aligned}&\hat{F}\left(\Phi, \rho^{\otimes M}\right) = \\ &= \sum_{i=1}^{t \cdot M} p_i \langle \psi_i^{\otimes M} \mid \left(\Phi \mid \psi_i^{\otimes M} \rangle\right) \mid \psi_i^{\otimes M} \rangle = \\ &= \sum_{i=1}^{t \cdot M} p_i \langle \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle \langle \psi_i^{\otimes M} \mid \mathcal{P}_{M,\epsilon}(\rho) \mid \psi_i^{\otimes M} \rangle + \\ &+ \sum_{i=1}^{t \cdot M} p_i \langle \psi_i^{\otimes M} \mid \rho_{i,junk} \mid \psi_i^{\otimes M} \rangle \langle \psi_i^{\otimes M} \mid \overline{\mathcal{P}_{M,\epsilon}(\rho)} \mid \psi_i^{\otimes M} \rangle.\end{aligned}$$

La seconda sommatoria è non negativa, dunque

$$\begin{array}{l}\sum_{i=1}^{t\cdot M}\ p_i\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\ \mid \psi_i^{\otimes M}\ \rangle + \\ +\ \sum_{i=1}^{t\cdot M}\ p_i\ \langle\ \psi_i^{\otimes M}\mid\ \rho_{i,junk}\mid \psi_i^{\otimes M}\ \rangle\ \langle\ \psi_i^{\otimes M}\mid \overline{\mathcal{P}_{M,\epsilon}(\rho)}\mid \psi_i^{\otimes M}\ \rangle \geq \\ \geq \sum_{i=1}^{t\cdot M}\ p_i\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\ \mid \psi_i^{\otimes M}\ \rangle\ = \\ =\ \sum_{i=1}^{t\cdot M}\ p_i\ \mid\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\mid\mid^4\end{array}$$

ove $\mid\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\mid\mid$ è la *norma* del vettore $\Big(\mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\Big)$. Poiché

$$\forall x\in\mathcal{R}\ \ (x-1)^2\geq 0,$$

ponendo $x=\mid\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\mid\mid^2$ risulta

$$\begin{array}{l}\mid\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\mid\mid^4 \geq \\ \geq \mid\mid 2\,\mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\mid\mid^2 - 1 = \\ = 2\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\ \mid \psi_i^{\otimes M}\ \rangle - 1\end{array}$$

di conseguenza

$$\begin{array}{l}\hat{F}\left(\Phi,\rho^{\otimes M}\right)\geq \\ \geq \sum_{i=1}^{t\cdot M}\ \left[\,p_i\ \mid\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\mid\mid^4\right]\ \geq \\ \geq \sum_{i=1}^{t\cdot M}\ p_i\big(\ 2\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle - 1\big) = \\ =\ 2\ \sum_{i=1}^{t\cdot M}\ \left[\,p_i\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\right] - \sum_{i=1}^{t\cdot M} p_i = \\ =\ 2\ \sum_{i=1}^{t\cdot M}\ \left[\,p_i\ \langle\ \psi_i^{\otimes M}\mid \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\ \rangle\right] - 1 = \\ =\ 2\ \sum_{i=1}^{t\cdot M}\ \left[\,p_i\ \mathrm{Tr}(\ \mathcal{P}_{M,\epsilon}(\rho)\mid \psi_i^{\otimes M}\bigtimes\psi_i^{\otimes M}\mid)\right] - 1 = \\ =\ 2\ \sum_{i=1}^{t\cdot M}\ \left[\ \mathrm{Tr}(\ \mathcal{P}_{M,\epsilon}(\rho)\ \cdot p_i\cdot\mid \psi_i^{\otimes M}\bigtimes\psi_i^{\otimes M}\mid)\right] - 1 = \\ =\ 2\ \mathrm{Tr}(\ \sum_{i=1}^{t\cdot M}\left[\mathcal{P}_{M,\epsilon}(\rho)\ \cdot p_i\cdot\mid \psi_i^{\otimes M}\bigtimes\psi_i^{\otimes M}\mid\ \right]) - 1 = \\ =\ 2\ \mathrm{Tr}(\ \mathcal{P}_{M,\epsilon}(\rho)\ \sum_{i=1}^{t\cdot M}\ \left[p_i\cdot\mid \psi_i^{\otimes M}\bigtimes\psi_i^{\otimes M}\mid\ \right]) - 1 = \\ =\ 2\ \mathrm{Tr}(\ \mathcal{P}_{M,\epsilon}(\rho)\ \rho^{\otimes M}) - 1.\end{array}$$

Considerando solo le uguaglianze, si ha che la terza vale poiché, ponendo $\mid a\ \rangle = (\mathcal{P}_{M,\epsilon}(\rho)\ \mid \psi_i^{\otimes M}\ \rangle)$ e $\langle\ b\mid = \langle\ \psi_i^{\otimes M}\mid$, risulta $\mathrm{Tr}\left(\mid a\bigtimes b\mid\right) = \langle\ b\mid a\ \rangle$, la quarta e la quinta valgono per via della linearità della traccia, la sesta per l'associatività rispetto alla somma del prodotto tra matrici, la settima dalla definizione di $\rho^{\otimes M}$ (83). Dalla (89) si ottiene che, dato $\epsilon>0$, $\delta>0$, $\exists M_0\in\mathcal{N}:\ \forall M\geq M_0$:

$$\hat{F}\left(\Phi,\rho^{\otimes M}\right)\geq 2\ \mathrm{Tr}(\mathcal{P}_{M,\epsilon}(\rho)\,\rho^{\otimes M}) - 1\geq 2(1-\delta) - 1 = 1 - 2\,\delta \tag{101}$$

il che completa la prova di correttezza dell'algoritmo proposto. Dalla (91) segue inoltre che, per $\epsilon>0$ e $M\to\infty$ risulta $\lim_{M\to\infty}\hat{F}\left(\Phi,\rho^{\otimes M}\right)=1$.

In sintesi, è stato dimostrato come, dopo aver scelto i parametri $\epsilon > 0$, $\delta > 0$ e $M \geq M_0$, codificando la sequenza sorgente con almeno $N = M(S(\rho)) + \epsilon$ qubit è possibile limitare l'abbassamento della *fedeltà media d'ensemble* ad un valore superiore parametrizzata da δ, raggiungendo un risultato analogo al teorema di Shannon.

5.3.4 Dimostrazione di ottimalità

Scelti $\epsilon > 0$, $\delta > 0$ ed $M \geq M_0$, si supponga ora di voler codificare con un numero di qubit $N < M(S(\rho) - \epsilon)$. Analogamente al caso classico, anche quantisticamente viene dimostrato un risultato forte: non esiste alcun canale *ideale* di codifica-decodifica $\Phi' = \Pi' \Gamma'$ che, codificando arbitrariamente ed in maniera iniettiva un qualunque sottospazio $\mathcal{A} \subseteq \mathcal{H}_{Sorg}^{\otimes M}$ (non necessariamente $\mathcal{T}_{M,\epsilon}(\rho)$), riesca a mantenere una fedeltà media dell'ensemble sufficientemente elevata. Al contrario, quest'ultima, al crescere di M, tende a 0.
Scegliendo un siffatto N, per ogni canale Φ' il sottospazio $\mathcal{A}$ degli stati codificati in modo iniettivo ha dimensione

$$l' = dim(\mathcal{A}) \leq 2^{M(S(P)-\epsilon)} \tag{102}$$

dove gli l' vettori che formano la base di $\mathcal{A}$ sono scelti arbitrariamente tra gli autovettori di $\rho^{\otimes M}$. Dopo aver codificato e decodificato uno stato puro del sistema $\mid \psi_i^{\otimes M} \rangle$ (ossia, in seguito alla trasmissione attraverso il canale di codifica-decodifica) il ricevitore ottiene uno stato rappresentato dalla matrice densità γ_i

$$\gamma_i = \Phi' \mid \psi_i^{\otimes M} \rangle = \sum_{j=1}^{l'} \lambda_j \mid j^{\otimes M} \bigtimes j^{\otimes M} \mid \tag{103}$$

con $\mid j^{\otimes M} \bigtimes j^{\otimes M} \mid : \ j \in \{1, \ldots, l'\}$ base di $\mathcal{A}$. Sia

$$\mathcal{P}_{\mathcal{A}} \stackrel{\text{def}}{=} \sum_{j=1}^{l'} \mid j^{\otimes M} \bigtimes j^{\otimes M} \mid \tag{104}$$

il proiettore al sottospazio $\mathcal{A}$. La *fedeltà media dell'ensemble* del canale Φ' rispetto alla sorgente $\rho^{\otimes M}$ è

$$
\begin{aligned}
&\hat{F}\left(\Phi', \rho^{\otimes M}\right) = \\
&= \textstyle\sum_{i=1}^{t\cdot M} \; p_i \; \langle \; \psi_i^{\otimes M} \mid \; \left[\Phi' \mid \psi_i^{\otimes M} \; \rangle\right] \; \mid \psi_i^{\otimes M} \; \rangle = \\
&= \textstyle\sum_{i=1}^{t\cdot M} \; p_i \; \langle \; \psi_i^{\otimes M} \mid \; \left[\sum_{j=1}^{l'} \lambda_j \mid j^{\otimes M} \bigtimes j^{\otimes M} \mid \right] \; \mid \psi_i^{\otimes M} \; \rangle \leq \\
&\leq \textstyle\sum_{i=1}^{t\cdot M} \; p_i \; \langle \; \psi_i^{\otimes M} \mid \; \left[\mid j^{\otimes M} \bigtimes j^{\otimes M} \mid \right] \; \mid \psi_i^{\otimes M} \; \rangle = \\
&= \textstyle\sum_{i=1}^{t\cdot M} \; p_i \; \langle \; \psi_i^{\otimes M} \mid \; \mathcal{P}_{\mathcal{A}} \; \mid \psi_i^{\otimes M} \; \rangle = \\
&= \mathrm{Tr}\left(\rho^{\otimes M} \mathcal{P}_{\mathcal{A}}\right).
\end{aligned}
\tag{105}
$$

Per completare l'enunciato si prova che, scelti i parametri come illustrato in precedenza, risulta

$$
\mathrm{Tr}\left(\rho^{\otimes M} \mathcal{P}_{\mathcal{A}}\right) \leq \delta. \tag{106}
$$

Dimostrazione. Innanzitutto, si scompone $\rho^{\otimes M}\mathcal{P}_{\mathcal{A}}$ nelle sue componenti presenti nel sottospazio ϵ-*tipico* e nel complemento ortogonale di quest'ultimo.

$$
\begin{aligned}
&\mathrm{Tr}\left(\rho^{\otimes M} \; \mathcal{P}_{\mathcal{A}}\right) = \\
&= \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \rho^{\otimes M}\right) = \\
&= \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \rho^{\otimes M} \; \mathcal{P}_{M,\epsilon}(\rho)\right) + \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \rho^{\otimes M} \; \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right)
\end{aligned}
\tag{107}
$$

Si consideri ora il primo addendo, $\mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \rho^{\otimes M} \; \mathcal{P}_{M,\epsilon}(\rho)\right)$. Si assuma s.l.g. di operare nella base in cui $\rho^{\otimes M}$ e $\mathcal{P}_{M,\epsilon}(\rho)$ sono diagonali: ciò renderà più evidente quanto segue.

$$
\begin{aligned}
&\mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \rho^{\otimes M} \; \mathcal{P}_{M,\epsilon}(\rho)\right) \; = \\
&= \; \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \rho^{\otimes M} \; \textstyle\sum_{i=1}^{l} \mid v_i^{\otimes M} \bigtimes v_i^{\otimes M} \mid\right) \; = \\
&= \; \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \textstyle\sum_{i=1}^{l} \; \rho^{\otimes M} \; \mid v_i^{\otimes M} \bigtimes v_i^{\otimes M} \mid\right) \; = \\
&= \; \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \textstyle\sum_{i=1}^{l} \; \lambda_i \; \mid v_i^{\otimes M} \bigtimes v_i^{\otimes M} \mid\right) \; \leq \\
&\leq \; \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \textstyle\sum_{i=1}^{l} \; 2^{-M(S(\rho)-\epsilon)} \; \mid v_i^{\otimes M} \bigtimes v_i^{\otimes M} \mid\right) \; = \\
&= \; 2^{-M(S(\rho)-\epsilon)} \; \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \textstyle\sum_{i=1}^{l} \; \mid v_i^{\otimes M} \bigtimes v_i^{\otimes M} \mid\right) \; = \\
&= \; 2^{-M(S(\rho)-\epsilon)} \; \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}} \; \mathcal{P}_{M,\epsilon}(\rho)\right)
\end{aligned}
$$

ove la prima e la quinta uguaglianza valgono per la definizione di $\mathcal{P}_{M,\epsilon}(\rho)$ (85), la seconda per l'associatività rispetto alla somma del prodotto tra matrici, la terza per l'applicazione dell'equazione agli autovalori di $\rho^{\otimes M}$ applicata sugli autovettori che ha in comune con $\mathcal{P}_{M,\epsilon}(\rho)$, la quarta per la linearità della traccia. La diseguaglianza invece deriva dalla corrispondenza tra autovettori di $\rho^{\otimes M}$ ed i vettori di base di $\mathcal{T}_{M,\epsilon}(\rho)$ (autovettori di $\mathcal{P}_{M,\epsilon}(\rho)$) citata nel paragrafo 5.3.1 e dalla conseguente relazione valida per gli autovalori di $\rho^{\otimes M}$ relativi ai vettori in comune (84). Inoltre poiché gli autovalori di ogni operatore di proiezione assumono solo valore

0 oppure 1 (e dal fatto che $\mathcal{P}_{M,\epsilon}(\rho)$ è in forma diagonale) segue che, sostituendo il valore 1 al posto di $\mathcal{P}_{M,\epsilon}(\rho)$, si ottiene una maggiorazione:

$$\begin{aligned}
&\mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right)\ \leq \\
&\leq 2^{-M(S(\rho)-\epsilon)}\ \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \mathcal{P}_{M,\epsilon}(\rho)\right)\ \leq \\
&\leq 2^{-M(S(\rho)-\epsilon)}\ \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\right) = \\
&=\ 2^{-M(S(\rho)-\epsilon)} \cdot l'
\end{aligned}$$

ove l'ultima uguaglianza vale poiché il valore della traccia di un proiettore equivale alla dimensione del sottospazio a cui quest'ultimo proietta. Dunque

$$\mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right)\ \leq\ 2^{-M(S(\rho)-\epsilon)} \cdot l' \tag{108}$$

Al crescere di M, per via della definizione di l (102), risulta

$$\lim_{M\to\infty} \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right)\ =\ \lim_{M\to\infty} 2^{-M(S(\rho)-\epsilon)} \cdot l'\ =\ 0 \tag{109}$$

Ci si concentra ora sul secondo addendo, $\mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \rho^{\otimes M}\ \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right)$. Si assuma s.l.g. di operare nella base in cui $\mathcal{P}_{\mathcal{A}}$ è diagonale. Ragionando sugli autovalori di $\mathcal{P}_{\mathcal{A}}$ analogamente a quanto fatto in precedenza con $\mathcal{P}_{M,\epsilon}(\rho)$ si ottiene una maggiorazione dalla sostituzione del valore 1 al posto di $\mathcal{P}_{\mathcal{A}}$

$$\mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \rho^{\otimes M}\ \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right) \leq \mathrm{Tr}\left(\rho^{\otimes M}\ \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right) \leq \delta \tag{110}$$

ove l'ultima disuguaglianza vale per la (90) con parametri opportunamente scelti $\epsilon > 0$, $\delta > 0$ ed $M \geq M_0$.

Effettuando ulteriori passaggi algebrici

$$\begin{aligned}
&\mathrm{Tr}\left(\rho^{\otimes M}\ \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right) = \mathrm{Tr}\left(\rho^{\otimes M}\ \left(\mathcal{I}_{\mathcal{H}_{Sorg}^{\otimes M}} - \mathcal{P}_{M,\epsilon}(\rho)\right)\right) = \\
&= \mathrm{Tr}\left(\rho^{\otimes M}\ -\ \rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right) = \mathrm{Tr}\,\rho^{\otimes M}\ -\ \mathrm{Tr}\left(\rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right) = \\
&= 1 -\ \mathrm{Tr}\left(\rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right)
\end{aligned}$$

Dalla (91) segue infine che, per $\epsilon > 0$ e $M \to \infty$

$$\lim_{M\to\infty} \mathrm{Tr}\left(\rho^{\otimes M}\ \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right) = 1 - \lim_{M\to\infty} \mathrm{Tr}\left(\rho^{\otimes M}\ \mathcal{P}_{M,\epsilon}(\rho)\right) = 0$$

Utilizzando quest'ultimo risultato nella (110) risulta, per $\epsilon > 0$ e $M \to \infty$, che

$$\lim_{M\to\infty} \mathrm{Tr}\left(\mathcal{P}_{\mathcal{A}}\ \rho^{\otimes M}\ \overline{\mathcal{P}_{M,\epsilon}(\rho)}\right) = 0 \tag{111}$$

il che completa la prova. □

In conclusione, è stata dimostrata la parte conclusiva del teorema di Schumacher. È stato provato come, dati i parametri $\epsilon > 0$ e $\delta > 0$, al crescere di M non siano sufficienti $N \leq M(S(\rho) - \epsilon)$ qubit per realizzare alcun canale di codifica-decodifica che mantenga un livello di fedeltà media d'ensemble sufficientemente elevato. Al contrario, quest'ultimo valore al crescere di M tende asintoticamente a 0.

Riferimenti bibliografici

[1] C.E. Shannon. 'A Mathematical Theory of Communication'. eng. In: *Bell System Technical Journal* 27.3 (1948).

[2] CHUANG Isaac L. NIELSEN Michael A. *Quantum computation and quantum information*. eng. 10.th Anniversary edition. Cambridge: Cambridge University Press, 2016.

[3] John Watrous. *The Theory of Quantum Information*. 1st. New York, NY, USA: Cambridge University Press, 2018.

[4] B. Schumacher. 'Quantum coding'. eng. In: *Physical review. A, Atomic, molecular, and optical physics* 51.4 (1995).