Error exponent bounds and practical short-length coding schemes for Distributed Hypothesis Testing (DHT)

Ismaila Salihou Adamou

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPERIEURE MINES-TELECOM ATLANTIQUE BRETAGNE PAYS DE LA LOIRE – IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 648
Sciences pour l'Ingénieur et le Numérique
Spécialité : *Télécommunication*

Par
Ismaila SALIHOU ADAMOU

Error Exponent Bounds and Practical Short-Length Coding Schemes for Distributed Hypothesis Testing (DHT)

Thèse présentée et soutenue à IMT Atlantique, Brest, le 9 Décembre 2024
Unité de recherche : Lab-STICC
Thèse N° : 2024IMTA0446

Rapporteurs avant soutenance :

Maël LETREUST	Chargé de recherche CNRS, INRIA Rennes
Jean-Marc BROSSIER	Professeur, Grenoble INP

Composition du Jury :

Président :	Philippe MARY	Professeur, INSA Rennes
Rapporteurs :	Maël LETREUST	Chargé de recherche CNRS, INRIA Rennes
	Jean-Marc BROSSIER	Professeur, Grenoble INP
Examinateurs :	Michèle WIGGER	Professeur, Télécom Paris
	Meryem BENAMMAR	Maîtresse de conférence, ISAE Supaéro Toulouse
Dir. de thèse :	Elsa DUPRAZ	Maîtresse de conférence HDR, IMT Atlantique

Remerciements

I would like to express my deepest gratitude to everyone who supported me throughout my Ph.D. journey, especially to my family for their immense moral support.

First and foremost, my sincere thanks go to my advisor, Elsa Dupraz, for the guidance she provided over these three years. Thank you for your advice, feedback, expertise, explanations, corrections, and suggestions, and especially for your patience throughout the entire supervision process. I also appreciate your valuable advice regarding my career after the thesis. I have personally learned so much from your way of supervising, your availability, and most importantly, your quick responsiveness in addressing the technical challenges of my thesis.

I would also like to thank the members of our IoTAD-CEO project team, especially Prof. Tad Matsumoto, for his perspectives on my research, his advice, and his leadership. Thank you as well for welcoming me to three different teams in Japan. I thoroughly enjoyed the visit and the various meetings we had.

My gratitude extends to my thesis jury: Philippe MARY, who served not only as the president of my CSI but also as the president of my thesis jury, with whom I also had many technical discussions, and to the rapporteurs Maël LETREUST and Jean-Marc BROSSIER, as well as the examiners Michèle WIGGER and Meryem BENAMMAR. A special thanks to Michèle for her insightful ideas and the technical discussions we had during various conferences.

Lastly, a big thank you to all my fellow Ph.D. students and postdocs in the MEE department, with whom I shared many conversations during lunches and coffee breaks. Among them: Ahcen, Anthony, Nga, Aeref, Daniel, Oscar, Ahmed, Solène, Dereck, Prince, and many others.

RÉSUMÉ EN FRANÇAIS

0.1 Introduction

Dans les réseaux de communication distribués, les données sont collectées, compressées, et transmises depuis des nœuds distants vers un serveur central pour un traitement ultérieur. Cependant, l'objectif du serveur n'est pas toujours de reconstruire les données originales, mais plutôt de prendre des décisions à partir des données codées reçues. Dans ce contexte, le Test d'Hypothèses Distribué se concentre sur le cas particulier de deux sources et vise à effectuer une prise de décision directement à partir des données compressées, sans passer par une reconstruction préalable. Comme dans le test d'hypothèses classique, deux types d'erreurs sont pris en compte pour évaluer les performances : l'erreur de Type I (fausse alarme) et l'erreur de Type II (décision manquée). Le test d'hypothèses distribué prend en compte une contrainte de débit sur le lien de communication, et l'objectif est de concevoir un schéma de codage afin de maximiser la décroissance exponentielle, appelée exposant d'erreur, de la probabilité d'erreur de Type II, tout en maintenant la probabilité d'erreur de Type I en dessous d'un seuil spécifié. Dans la littérature, ce problème a principalement été étudié en utlisant la théorie de l'information, et la plupart des travaux existants analysent les performances des schémas du test d'hypothèses distribué en supposant des sources independantes et identiquement distribuées (i.i.d.).

Dans la première partie de cette thèse, nous abordons un modèle plus réaliste et général de sources non-i.i.d. Ce modèle englobe des sources non stationnaires et non ergodiques, reflétant mieux les scénarios réels par rapport au cas i.i.d. Pour ce modèle général de sources, nous dérivons des bornes génériques sur l'exposant d'erreur pour le test d'hypothèses distribué à l'aide d'outils du spectre de l'information. Nous montrons la cohérence de ces bornes avec le cas i.i.d. et les caractérisons plus précisément pour deux modèles spécifiques de sources : les sources gaussiennes non-i.i.d., et les sources de type Gilbert-Elliot.

De plus, l'étude du test d'hypothèses distribué ne devrait pas se limiter seulement à l'analyse des limites théoriques, mais aussi inclure le développement de schémas de codage pratiques pour ce cadre. Ainsi, dans la deuxième partie de cette thèse, nous développons et implémentons des schémas de codage pratiques de courte longueur, spécialement conçus pour le test d'hypothèses distribué, qui n'avaient pas encore été étudiés dans la littérature. Ces schémas de codage sont basés sur des codes linéaires en blocs et visent des longueurs très courtes, appropriées pour le test d'hypothèses distribué (moins de 100 bits). En outre, nous fournissons des expressions analytiques exactes pour les probabilités d'erreurs de Type I et Type II pour chaque schéma de

codage proposé, offrant ainsi des outils utiles pour la conception optimale future de codes pour le test d'hypothèses distribué.

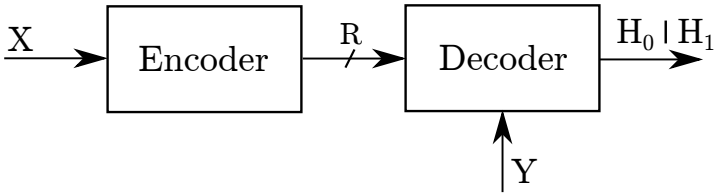

FIGURE 1 – Test d'hpothèses distribué.
Source : © 2023 IEEE. Reproduced with permission from [1].

0.2 Etat de l'art

Dans cette partie, nous passons en revue la littérature existante sur le test d'hypothèses distribué. Le cas le plus traité est celui d'un encodeur et un centre de décision avec information adjacente, appelé configuration asymétrique, comme illlustré à la Figure 1. Pour cette configuration, nous présentons les schémas de codage issus de la théorie de l'information et leurs bornes correspondantes sur les exposants d'erreur atteignables.

Énoncé du Problème

Dans la configuration asymétrique du test d'hypothèses distribué, un encodeur observe une séquence source X^n et transmet sa version encodée au décodeur disposant d'une information adjacente Y^n. Le décodeur doit décider entre deux hypothèses H_0 et H_1, avec :

$$H_0 : (X, Y) \sim P_{XY}, \quad H_1 : (X, Y) \sim P_{\bar{X}\bar{Y}}. \tag{1}$$

Les probabilités d'erreur de Type-I et de Type-II, α_n et β_n, sont définies comme suit [17] :

$$\alpha_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(X^n\right), Y^n\right) = H_1 \mid H_0\right] \tag{2}$$

$$\beta_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(X^n\right), Y^n\right)\right) = H_0 \mid H_1\right]. \tag{3}$$

Ici, $f^{(n)}$ and $g^{(n)}$ son respectivement les fonctions d'encodage et de decodage. L'objectif est de minimiser β_n tout en maintenant $\alpha_n \leq \epsilon$, avec un exposant d'erreur défini comme :

$$\theta = \limsup_{n\to\infty} -\frac{1}{n} \log \beta_n. \tag{4}$$

Schéma d'Ahlswede et Csiszár

Ce schéma repose sur la quantification de X^n et l'envoi de son type. L'exposant d'erreur atteint est donné par [18] :

$$\theta_{AC}(R) \geq D(P_X||P_{\bar{X}}) + \sup_{P_{U|X}: I(X;U)\leq R} D(P_{UY}||P_{\bar{U}\bar{Y}}), \tag{5}$$

où $U - X - Y$ et $U - \bar{X} - \bar{Y}$ forment des chaînes de Markov.

Dans le cas particulier des tests contre l'indépendance ($P_{\bar{X}\bar{Y}} = P_X P_Y$), l'exposant se simplifie en [18] :

$$\theta_{AC}(R) = \max_{P_{U|X}: I(X;U)\leq R} I(U;Y). \tag{6}$$

Schéma de Han

Han améliore ce résultat en incluant le test sur le type conjoint (X, U). L'exposant d'erreur atteint est donné par [21] :

$$\theta_{HAN}(R) \geq \sup_{\substack{P_{U|X}: \\ I(X,U)\leq R}} \min_{P_{\tilde{U}\tilde{X}\tilde{Y}}\in\mathcal{P}_{\text{HAN}}} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\bar{U}\bar{X}\bar{Y}}\right), \tag{7}$$

avec

$$\mathcal{P}_{\text{HAN}} = \left\{P_{\tilde{U}\tilde{X}\tilde{Y}} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{U}\tilde{Y}} = P_{UY}\right\}. \tag{8}$$

Schéma de Shimokawa, Han et Amari (SHA)

Le schéma de quantification-binning ou de SHA exploite la corrélation entre X^n et Y^n pour réduire le taux de compression. L'exposant d'erreur obtenu est donné par [22] :

$$\theta_{SHA}(R) \geq \sup_{\substack{P_{U|X}: \\ I(U;X|Y)<R<I(U;X)}} \min \Bigg[\min_{P_{\tilde{U}\tilde{X}\tilde{Y}}\in\mathcal{P}_{\text{SHA}}\left(P_{U|X}\right)} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\bar{U}\bar{X}\bar{Y}}\right) + R - I(U;X \mid Y),$$
$$\min_{P_{\tilde{U}\tilde{X}\tilde{Y}}\in\mathcal{P}_{\text{HAN}}} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\bar{U}\bar{X}\bar{Y}}\right) \Bigg], \tag{9}$$

où

$$\mathcal{P}_{\text{SHA}}\left(P_{U|X}\right) := \{P_{\tilde{U}\tilde{X}\tilde{Y}} :$$
$$P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{Y}} = P_Y, H(\tilde{U} \mid \tilde{Y})_e \geq H(U \mid Y)\Big\}. \tag{10}$$

Le schéma de SHA est optimal dans certains cas particuliers, comme le test conditionnel contre l'indépendance [23].

Améliorations récentes

Kochman et Wang [25] ont élargi les contraintes sur $P_{U|X}$, ce qui améliore les performances du schéma de SHA en considérant plus de distributions possibles. Cependant, Watanabe [24] a montré des cas spécifiques où le schéma de SHA est sous-optimal.

0.3 Test d'hypothèses distribué pour des modèles de sources générales non-iid, non-stationnaires, et non-ergodiques

Cette partie traite du test d'hypothèses distribué pour des modèles de sources générales $\mathbf{X}$ et $\mathbf{Y}$, et propose un schéma réalisable fournissant une borne générique sur l'exposant d'erreur applicable à une large gamme de modèles de sources, pas nécessairement i.i.d. Notre schéma réalisable s'appuie sur les méthodes du spectre d'information [42] pour traiter des sources générales. Il fournit une borne inférieure sur l'exposant d'erreur général, qui est relativement facile à calculer pour des sources i.i.d. et/ou stationnaires gaussiennes.

Modèle de sources générales

Nous adoptons la définition de [42] pour le modèle de sources générales. Les sources $\mathbf{X}$ et $\mathbf{Y}$ génèrent deux suites infinies :

$$\begin{aligned} &\{\mathbf{X}^n = (X_1, X_2, \cdots, X_n)\}_{n=1}^{\infty}, \\ &\{\mathbf{Y}^n = (Y_1, Y_2, \cdots, Y_n)\}_{n=1}^{\infty} \end{aligned} \tag{11}$$

de variables aléatoires $\mathbf{X}^n, \mathbf{Y}^n$, chacune de dimension n. De plus, les symboles X_i, Y_i prennent leurs valeurs dans des alphabets $\mathcal{X}, \mathcal{Y}$, respectivement. Le modèle de [57] suppose ensuite que les distributions de probabilités jointes $P_{\mathbf{X}^n\mathbf{Y}^n}$ sont connues mais quelconques (pas d'hypothèse de stationnarité ou d'ergodicité).

Test d'hypothèses distribué

Nous définissons maintenant le problème du test d'hypothèses distribué pour le modèle de sources générales que nous avons présenté précédemment. Nous supposons que la loi jointe du couple $(\mathbf{X}^n, \mathbf{Y}^n)$ dépend des hypothèses sous-jacentes $\mathcal{H}_0$ et $\mathcal{H}_1$, définies comme suit :

$$\mathcal{H}_0 : (\mathbf{X}^n, \mathbf{Y}^n) \sim P_{\mathbf{X}^n\mathbf{Y}^n}, \tag{12}$$

$$\mathcal{H}_1 : (\mathbf{X}^n, \mathbf{Y}^n) \sim P_{\overline{\mathbf{X}}^n\overline{\mathbf{Y}}^n}. \tag{13}$$

Nous rappelons d'abord les définitions des outils de l'"information spectrum" de [42] qui seront utiles pour notre analyse.

Définitions

Tout d'abord, nous définissons respectivement la lim sup et la lim inf en probabilité d'une suite de variables aléatoires $\{Z_n\}_{n=1}^{\infty}$ comme suit [42] :

$$\mathrm{p}-\limsup_{n\to\infty} Z_n = \inf\left\{\alpha \mid \lim_{n\to+\infty} \mathbb{P}\left(Z_n > \alpha\right) = 0\right\}$$
$$\mathrm{p}-\liminf_{n\to\infty} Z_n = \sup\left\{\alpha \mid \lim_{n\to+\infty} \mathbb{P}\left(Z_n < \alpha\right) = 0\right\}.$$

Ensuite, l'information mutuelle spectrale supérieure $\overline{I}(\mathbf{X};\mathbf{U})$, l'information mutuelle spectrale inférieure $\underline{I}(\mathbf{U};\mathbf{Y})$, la divergence spectrale inférieure $\underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right)$, et la divergence spectrale supérieure $\overline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right)$ sont respectivement définies comme [42] :

$$\bar{I}(\mathbf{X};\mathbf{U}) = \mathrm{p}-\limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{X}^n}\left(\mathbf{U}^n \mid \mathbf{X}^n\right)}{P_{\mathbf{U}^n}\left(\mathbf{U}^n\right)}, \tag{14}$$
$$\underline{I}(\mathbf{U};\mathbf{Y}) = \mathrm{p}-\liminf_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{Y}^n}\left(\mathbf{U}^n \mid \mathbf{Y}^n\right)}{P_{\mathbf{U}^n}\left(\mathbf{U}^n\right)}, \tag{15}$$

$$\underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right) = \mathrm{p}-\liminf_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n\mathbf{Y}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}, \tag{16}$$
$$\overline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right) = \mathrm{p}-\limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n\mathbf{Y}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}. \tag{17}$$

Dans le cas i.i.d., on retrouve les définitions classiques de l'information mutuelle et de la divergence.

Exposant d'erreur atteignable pour des sources générales

Dans cette partie, nous présentons notre résultat principal sur l'exposant d'erreur de Type-II réalisable θ pour les sources générales définies precedement.

Theorem 0.1 *L'exposant d'erreur θ suivant est atteignable pour des sources générales définies par* (11) *:*

$$\theta \geq \min\left\{R - \left(\overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{U};\mathbf{Y})\right), \underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right) + \left(\underline{I}(\mathbf{X};\mathbf{U}) - \overline{I}(\mathbf{X};\mathbf{U})\right)\right\},$$

avec $\mathbf{U}$*, variable aléatoire auxiliaire telle que la chaîne de Markov* $\mathbf{U}\to\mathbf{X}\to\mathbf{Y}$ *est satisfaite à la fois sous* H_0 *et sous* H_1*.* $P_{\mathbf{UY}}$ *et* $P_{\overline{\mathbf{UY}}}$ *sont les distributions jointes de* $(\mathbf{U}^n,\mathbf{Y}^n)$ *sous* H_0 *et* H_1*, respectivement, et* $R \geq \underline{I}(\mathbf{U};\mathbf{X}\mid\mathbf{Y})$*.*

On remarque que lorsque les sources $\mathbf{X}$,$\mathbf{Y}$ et $\mathbf{U}$ sont i.i.d., notre exposant d'erreur se résume à celui trouvé dans [35]. Ceci montre la cohérence de notre analyse.

0.4 Exemple : Sources Gaussiennes

Nous appliquons maintenant le Théorème 0.1 à des modèles de sources Gaussiennes non i.i.d. mais stationnaires et ergodiques, telles que $\mathbf{X} \sim \mathcal{N}(0, \mathbf{K_X})$ et $\mathbf{Y} \sim \mathcal{N}(0, \mathbf{K_Y})$, où $\mathbf{K_X}$ et $\mathbf{K_Y}$ sont les matrices de covariance de $\mathbf{X}$ et $\mathbf{Y}$, respectivement.

Proposition 0.1 *Si les sources* $\mathbf{X}$ *et* $\mathbf{Y}$ *sont gaussiennes, stationnaires et ergodiques sous* H_0 *et* H_1*, l'exposant d'erreur dans le Théorème 0.1 devient :*

$$\theta \geq \min \left\{ R - \lim_{n\to\infty} \left[\frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{Y}^n\right) - \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{X}^n\right) \right], \lim_{n\to\infty} \frac{1}{n} D\left(P_{\mathbf{U}^n\mathbf{Y}^n} \| P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right) \right\}.$$

Cette proposition découle de la propriété *"strong converse property"* [42, Page 48-49]. De plus :

$$\lim_{n\to\infty} \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{Y}^n\right) - \lim_{n\to\infty} \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{X}^n\right) = \lim_{n\to\infty} \frac{1}{2n} \sum_{i=1}^{n} \log \frac{\lambda_i^{(X|Y)} + \kappa}{\kappa}, \tag{18}$$

$$\lim_{n\to\infty} \frac{1}{n} D\left(P_{\mathbf{U}^n\mathbf{Y}^n} \| P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right\} = \lim_{n\to\infty} \frac{1}{2n} \left[\log \frac{\left|\overline{\mathbf{\Sigma}}\right|}{|\mathbf{\Sigma}|} - 2n + \mathrm{tr}\left\{\overline{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\right\} \right], \tag{19}$$

où $\mathbf{\Sigma}$ et $\overline{\mathbf{\Sigma}}$ sont les matrices de covariance conjointes de $\mathbf{U}$ et $\mathbf{Y}$ sous H_0 et H_1, respectivement, $|.|$ représente un déterminant, et tr(.) représente la trace. Les termes donnés par (18) and (19) sont obtenus en considérant que la source $\mathbf{U}$ est gaussienne de sorte que $\mathbf{U} = \mathbf{X} + \mathbf{Z}$. Avec $\mathbf{Z} \sim \mathcal{N}(0, \kappa\mathbf{I}_n)$, indépendant de $\mathbf{X}$ et $\mathbf{I}_n$ est la matrice identité de dimension $n \times n$. Les matrices de covariance $\mathbf{\Sigma}$ et $\overline{\mathbf{\Sigma}}$ sont alors définies comme $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{K_U} & \mathbf{K_{UY}} \\ \mathbf{K_{YU}} & \mathrm{K_Y} \end{bmatrix}$ et $\overline{\mathbf{\Sigma}} = \begin{bmatrix} \mathbf{K_U} & \overline{\mathbf{K}}_{\mathbf{UY}} \\ \overline{\mathbf{K}}_{\mathbf{YU}} & \mathrm{K_Y} \end{bmatrix}$. On note que les matrices $\mathbf{\Sigma}$ et $\overline{\mathbf{\Sigma}}$ sont de dimension $(2n) \times (2n)$, avec $n \to \infty$ dans les équations précédentes.

0.4.1 Exemple : Modèle de Gilbet-Elliot (GE)

Dans cette partie, les séquences binaires générées par $\mathbf{X}$ et $\mathbf{Y}$ sont respectivement notées $\{X_k\}_{k=1}^{+\infty}$ et $\{Y_k\}_{k=1}^{+\infty}$. Nous supposons que la source $\mathbf{X}$ est i.i.d., telle que pour tout $k \geq 1$, X_k suit une distribution de Bernoulli Bern(p). De plus, $\mathbf{X}$ et $\mathbf{Y}$ sont corrélés tel que :

$$Y_k = X_k \oplus Z_k.$$

Ici, la source $\mathbf{Z}$, qui génère la séquence binaire $\{Z_k\}_{k=1}^{+\infty}$, est indépendante de $\mathbf{X}$ et suit un

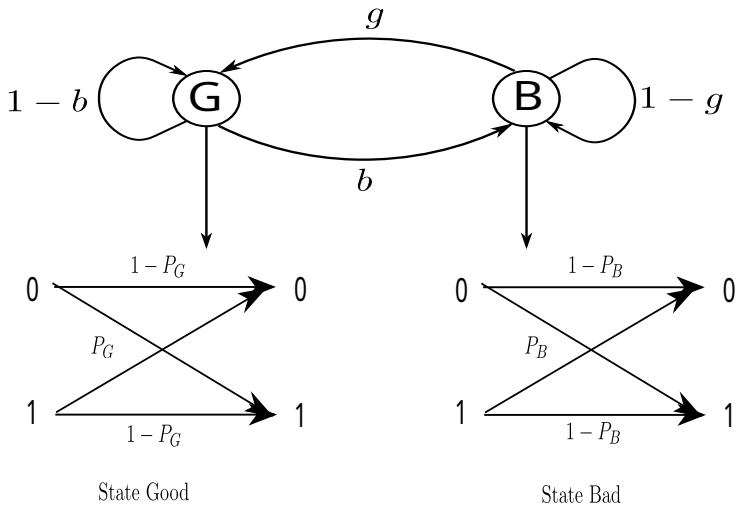

FIGURE 2 – Test d'hpothèses distribué

modèle GE [65] avec un état caché $\mathbf{S}$. La séquence des états binaires cachés $\{S_k\}_{k=1}^{+\infty}$ est telle que $S_k \in \{G, B\}$ comme illustré à la Figure 2. De plus, chaque symbole Z_k prend la valeur 0 ou 1 selon la valeur de l'état caché $S_k = s$, tel que :

$$P\left(Z_k = 1 \mid S_k = s\right) = p_s, \quad s \in \{G, B\},$$

où p_G et p_B représentent respectivement les probabilités de transition dans les états G (forte corrélation) et B (faible corrélation). Les hypothèses H_0 et H_1 sont définies par :

$$H_0 : (p_G, p_B), \tag{20}$$
$$H_1 : (\bar{p}_G, \bar{p}_B), \tag{21}$$

où p_G, p_B et $\bar{p}_G$, $\bar{p}_B$ représentent les probabilités de transition dans les états G et B sous les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$, respectivement.

Nous spécifions ici la borne générale de l'exposant d'erreur donnée dans le Théorème 0.1 pour les modèles ergodiques de Gilbert-Elliot (GE).

Proposition 0.2 *Si les sources* $\mathbf{X}$ *et* $\mathbf{Y}$ *sont corrélées selon le modèle GE sous les hypothèses* H_0 *et* H_1*, la borne générale de l'exposant d'erreur dans le Théorème 0.1 se réduit à :*

$$\theta \geq \sup_{P_{\mathbf{U}|\mathbf{X}}} \min\left\{R - \left[H_s\left(\mathbf{U} \mid \mathbf{Y}\right) - H_s\left(\mathbf{U} \mid \mathbf{X}\right)\right], D_s\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)\right\}, \tag{22}$$

où le sup *est pris sur toutes les distributions conditionnelles* $P_{\mathbf{U}|\mathbf{X}}$ *satisfaisant la contrainte* $R \geq I_s(\mathbf{U}; \mathbf{X} \mid \mathbf{Y})$.

De plus,

$$H_s(\mathbf{U}|\mathbf{Y}) = p - \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P(\mathbf{U}^n|\mathbf{Y}^n)}, \tag{23}$$

et

$$D_s(P_{\mathbf{U},\mathbf{Y}}||P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}}) = p - \lim_{n\to\infty} \frac{1}{n} \log \frac{P(\mathbf{U}^n, \mathbf{Y}^n)}{P(\overline{\mathbf{U}}^n, \overline{\mathbf{Y}}^n)}. \tag{24}$$

Pour évaluer les termes $H_s(\mathbf{U}|\mathbf{Y})$ et $D_s(P_{\mathbf{U},\mathbf{Y}}||P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}})$, nous avons proposé d'utiliser des estimateurs pour lesquels, pour une grande valeur de n, nous générons aléatoirement des échantillons $(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n)$ selon le modèle de Gilbet-Elliot. Nous calculons ensuite les probabilités $P(\mathbf{u}^n|\mathbf{y}^n)$, $P(\mathbf{u}^n)$, $P(\mathbf{u}^n, \mathbf{y}^n)$, et $P(\overline{\mathbf{u}}^n, \overline{\mathbf{y}}^n)$ en utlisant un algorithme recursif tel quel le BCJR. Cette méthodologie est similaire à celle utilisée dans [69] pour évaluer numériquement la capacité d'un canal de Gilbert-Elliot.

0.5 Schémas pratiques à courte longueur pour le test d'hypothèses distribué

Cette partie traite de la conception de schémas de codage à courte longueur pour le test d'hypothèses distribué avec des sources binaires. Les preuves d'atteignabilité issues de la théorie de l'information suggèrent d'envisager des schémas de codage basés uniquement sur la quantification [21] ou sur des schémas de quantification-binning [22] pour le test d'hypothèses distribué. Notre implémentation pratique de schémas de codage suit de près l'approche décrite dans les preuves théoriques en théorie de l'information. Contrairement aux travaux existants, qui se concentrent sur des séquences très longues (souvent supérieures à 10^5 bits), cette partie propose des solutions adaptées à des séquences courtes, où quelques dizaines de bits suffisent pour une prise de décision correcte. Nous avons donc proposé des implémentations pratiques de schémas de quantification et de quantification-binning. Ces derniers utilisent des codes linéaires en blocs binaires à courte longueur, spécifiquement adaptés au problème de prise de décision. Des expressions analytiques précises des probabilités d'erreur de Type-I et Type-II sont également dérivées pour la configuration asymétrique, permettant une optimisation et une comparaison des performances des différents schémas.

Test d'hypothèses distribué pour les sources binaires

Dans ce résumé, nous présentons les résultats des schémas pratiques uniquement pour le cas asymétrique, comme illustré en Figure 1. Cependant, le cas symétrique est également traité dans cette thèse. Nous supposons que les n symboles des séquences $\mathbf{X}^n$ et $\mathbf{Y}^n$ sont i.i.d. et générés selon les variables aléatoires X et Y, respectivement. De plus, X et Y sont distribués conjointement selon le modèle $Y = X \oplus E$, où E est une variable aléatoire binaire indépendante de X, avec

$P(X=1)=1/2$. Nous notons également $p=\mathbb{P}(E=1)$ avec $0<p\leq 1/2$. Les deux hypothèses sont exprimées comme suit :

$$\begin{cases} H_0: & p=p_0, \\ H_1: & p=p_1. \end{cases} \tag{25}$$

Nous supposons, sans perte de généralité, que $p_0 < p_1$. Il est important de noter que les distributions de probabilité de X et Y sont indépendantes de l'hypothèse choisie, étant donné que $P(X=1)=1/2$.

Schéma de quantification

L'objectif est de concevoir une implémentation pratique d'un schéma de quantification pour des séquences binaires de courte longueur, basé sur des codes linéaires en blocs. Pour cela, pour une séquence binaire $\mathbf{x}^n$ de longueur n, l'encodeur utlise une matrice génératrice G_q d'un code linéaire en bloc pour produire une séquence quantifiée $\mathbf{z}_q^m$ tel que :

$$\mathbf{z}_q^m = \arg\min_{\mathbf{z}^m} \; d(G_q\mathbf{z}^m, \mathbf{x}^n), \tag{26}$$

où $d(.,.)$ représente la distance de Hamming. La séquence $\mathbf{z}_q^m$ est transmise au décodeur avec un taux $R=m/n$.

Au décodeur, la séquence quantifiée $\mathbf{x}_q^n = G_q\mathbf{z}_q^m$ est reconstruite. Pour décider entre les hypothèses, on applique le test de Neyman-Pearson suivant :

$$\sum_{i=1}^{n}(x_{q,i}\oplus y_i) < \lambda_q, \tag{27}$$

où λ_q est un seuil déterminé.

Les expressions analytiques exactes pour les probabilités d'erreur de Type-I ($\alpha_n^{(q)}$) et de Type-II ($\beta_n^{(q)}$) pour le schéma de quantification sont exprimées comme suit :

$$\alpha_n^{(q)} = 1 - \frac{1}{N_0^{(q)}}\sum_{\lambda=0}^{\lambda_q}\sum_{\gamma=0}^{d_{\max}^{(q)}}\sum_{j=0}^{n} E_\gamma^{(q)}\Gamma_{\lambda,j,\gamma}p_0^j(1-p_0)^{n-j}, \tag{28}$$

$$\beta_n^{(q)} = \frac{1}{N_0^{(q)}}\sum_{\lambda=0}^{\lambda_q}\sum_{\gamma=0}^{d_{\max}^{(q)}}\sum_{j=0}^{n} E_\gamma^{(q)}\Gamma_{\lambda,j,\gamma}p_1^j(1-p_1)^{n-j}. \tag{29}$$

Ici, $E_\gamma^{(q)}$ représente le nombre de mots $\mathbf{x}^n$ de poids de Hamming γ appartenant à la région de décision $\mathcal{C}_0^{(q)}$ telle que $\mathbf{x}_q^n=\mathbf{0}^n$. De plus, nous définissons $N_0^{(q)}=\sum_{\gamma=0}^{d_{\max}^{(q)}}E_\gamma^{(q)}$, où $d_{\max}^{(q)}$ est le poids de Hamming maximal des mots de code dans la région $\mathcal{C}_0^{(q)}$.

Schéma de quantification-binning

Ici, nous proposons une solution pratique pour le schéma de quantification-binning. Comme précédemment, nous considérons une matrice génératrice G_q de dimensions $n \times m$. Nous utilisons également une matrice de contrôle de parité H_b de dimensions $k \times m$ issue d'un autre code linéaire en blocs. Après avoir utilisé G_q pour effectuer la quantification binaire, comme décrit précédemment, l'encodeur utilise la matrice H_b pour calculer :

$$\mathbf{u}^k = H_b \mathbf{z}_q^m. \tag{30}$$

Le syndrome $\mathbf{u}^k$ est ensuite transmis au décodeur. Dans ce cas, le taux de codage est donné par $R = k/n$.

Au décodeur, pour appliquer le test de Neyman-Pearson, nous identifions d'abord un vecteur $\hat{\mathbf{z}}_q^m$ par recherche exhaustive :

$$\hat{\mathbf{z}}_q^m = \arg\min_{\mathbf{z}^m} \; d(G_q \mathbf{z}^m, \mathbf{y}^n) \text{ s.t. } H_b \mathbf{z}^m = \mathbf{u}^k. \tag{31}$$

Ensuite, le test suivant est appliqué :

$$\sum_{i=1}^{n} (\hat{x}_{q,i} \oplus y_i) < \lambda_{qb}, \tag{32}$$

où $\hat{\mathbf{x}}_q^n = G_q \hat{\mathbf{z}}_q^m$, et λ_{qb} est un seuil entier.

Les probabilités d'erreur de Type-I ($\alpha_n^{(qb)}$) et de Type-II ($\beta_n^{(qb)}$) pour la quantification-binning sont exprimées comme suit :

$$\alpha_n^{(qb)} = 1 - \mathbb{P}_B(p_0) - \mathbb{P}_{\bar{B}}(p_0), \tag{33}$$

$$\beta_n^{(qb)} = \mathbb{P}_B(p_1) + \mathbb{P}_{\bar{B}}(p_1), \tag{34}$$

où :

$$\mathbb{P}_B(\delta) = \sum_{\nu=0}^{\min(d_{\max}^{(qb)}, \lambda_{qb})} \frac{E_\nu^{(qb)}}{\binom{n}{\nu}} \sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \Gamma_{\nu,j,\gamma} \delta^j (1-\delta)^{n-j}, \tag{35}$$

$$\mathbb{P}_{\bar{B}}(\delta) = \sum_{i=0}^{n} \left[\left(\sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \Gamma_{i,j,w} \delta^j (1-\delta)^{n-j} \right) \times \left(\sum_{t=1}^{n} \sum_{\nu=0}^{\lambda_{qb}} \frac{E_\nu^{(qb)}}{\binom{n}{\nu}} \frac{A_t^{(qb)} \Gamma_{i,\nu,t}}{\binom{n}{i}} \right) \right]. \tag{36}$$

Ici, $E_\nu^{(qb)}$ est le nombre de mots $\mathbf{y}^n$ de poids de Hamming ν appartenant à la région de décision $\mathcal{C}_0^{(qb)}$. De plus, $\{A_t^{(qb)}\}_{t \in [\![0,n]\!]}$ répresente le nombre de mots $\mathbf{x}_q^n$ de poids de Hamming t tels qu'il existe $\mathbf{z}_q^m$ satisfaisant $\mathbf{x}_q^n = G_q \mathbf{z}_q^n$ et $H_b \mathbf{z}_q^m = \mathbf{0}^k$.

Résultats numériques et conclusions

Nos résultats numériques mettent en évidence l'efficacité des schémas de quantification et de quantification-binning par rapport aux schémas non codés. En outre, ces résultats valident l'exactitude des probabilités d'erreur analytiques en démontrant leur cohérence avec les simulations Monte-Carlo.

0.6 Conclusion Générale

Dans cette thèse, nous avons étendu l'étude du test d'hypothèses distribué à des modèles de sources plus généraux, dépassant les hypothèses classiques d'indépendance et d'identiquement distribués (i.i.d.). Nous avons analysé les performances du test d'hypothèses distribué pour ces modèles et dérivé des exposants d'erreur atteignables en proposant un schéma de codage basé sur l'approche du spectre de l'information, introduite par Han [42]. Contrairement aux schémas existants pour les sources i.i.d., qui reposent sur la méthode des types, notre approche offre une borne inférieure générale sur l'exposant d'erreur. Notamment, pour le cas particulier des sources i.i.d., notre exposant général correspond aux résultats bien établis de [2].

Nous avons ensuite démontré l'applicabilité de notre analyse à divers modèles de sources d'intérêt, tels que les sources gaussiennes stationnaires et ergodiques, ainsi que le modèle de Gilbert-Elliot (GE). En particulier, pour le modèle GE, nous avons introduit une méthode efficace pour estimer l'exposant d'erreur, en utilisant la récursion avant des modèles de Markov cachés. Les résultats numériques ont permis d'évaluer l'impact des paramètres du modèle sur l'exposant d'erreur et le compromis entre l'erreur de test et l'erreur de binning.

Nous nous sommes ensuite concentrés sur le développement de schémas de codage pratiques. Plus précisément, nous avons proposé des implémentations à courte longueur des schémas de quantification et de quantification-binning, construits à l'aide de codes linéaires en blocs. Pour ces deux schémas, nous avons abordé la manière d'effectuer le test d'hypothèses dans des scénarios pratiques. En plus des constructions pratiques, nous avons dérivé des expressions théoriques des probabilités d'erreur de Type-I et de Type-II pour chaque schéma proposé. Les résultats numériques ont montré que nos implémentations pratiques offrent des améliorations notables en termes de performance par rapport aux schémas non codés de référence, où seule une partie des bits est transmise sans codage.

Enfin, bien que les preuves issues de la théorie de l'information aient constitué une base pour le développement de nos schémas pratiques, les enseignements tirés de la conception pratique peuvent fournir des orientations précieuses pour de futurs travaux théoriques sur le test d'hypothèses distribué, en particulier pour la configuration symétrique. En plus, le travail réalisé dans cette thèse pourrait servir de base à l'investigation théorique et pratique de schémas de codage dédiés à des tâches d'apprentissage plus complexes, telles que la classification.

Contents

Liste des acronymes

DHT	Distributed Hypothesis Testing
i.i.d.	Independent and identically distributed
SHA	Shimokawa, Han, and Amari
RW	Rahman and Wagner
KP	Katz and Piantanida
IS	Information Spectrum
p.m.f	probability mass function
HMM	Hidden Markov Models
NP	Neyman-Pearson
ROC	Receiver Operating Characteristic
LDGM	Low-Density Generator Matrices

TABLE DES FIGURES

Chapitre 1

Introduction

1.1 Background and Motivation

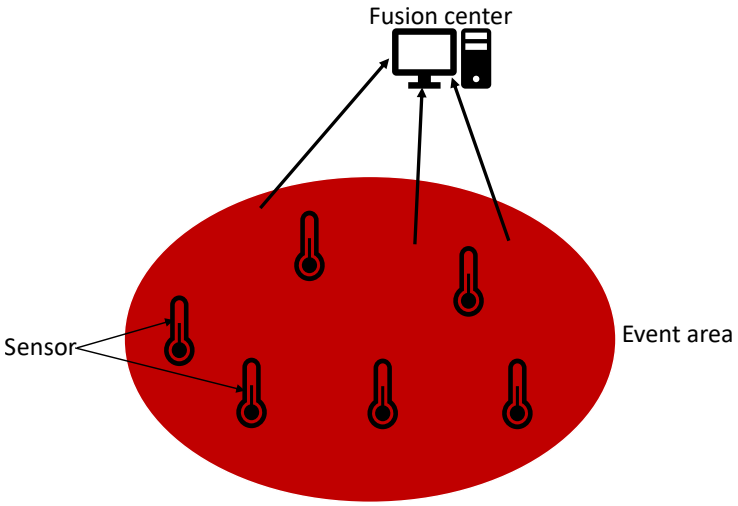

Figure 1.1 – Distributed sensors network

In the era of 5G and Beyond 5G technologies, the paradigm of communication systems is shifting to address emerging challenges and requirements driven by practical concerns, such as energy consumption and system complexity. One of the key current challenges lies in distributed communication networks, where the data is not centralized at one location. It is rather collected in a distributed manner at several locations and gathered at a fusion center for further processing. As a toy example, we can consider a network of sensors for measuring the temperature at different locations, as illustrated in Figure 1.1. Due to the energy cost of wireless communication, each sensor is constrained in its ability to communicate its measurements to the fusion center. However, given that the sensors observe the same phenomenon, their measurements are often highly correlated. Distributed source coding [4, 5, 6] is a compression technique that exploits such correlation to significantly reduce the amount of information that each sensor needs to transmit to the fusion center.

1.1.1 Distributed source coding

In distributed source coding, sensors measurements are compressed independently and later jointly decompressed at the fusion center. This concept was first introduced as an information-theoretic problem in the seminal work of Slepian and Wolf [7]. A simple setup was considered, where two sources, $\mathbf{X}$ and $\mathbf{Y}$, are encoded independently and decoded jointly. The results of [7] provide an achievable rate region and show that, asymptotically, separate encoding can achieve the same performance as joint encoding.

While the initial work in [7] only considered lossless source coding where the sources $\mathbf{X}$ and $\mathbf{Y}$ need to be reconstructed exactly, the problem was then extended to many setups. Notably, Wyner and Ziv [8] considered a lossy version of the asymmetric Slepian-Wolf problem, where one source, $\mathbf{Y}$, is available at the decoder as side information, and only the source $\mathbf{X}$ needs to be encoded. In such a setup, the decoder reconstructs the source $\mathbf{X}$ with the help of this side information, ensuring that the average distortion does not exceed a specified threshold. An example of this setup consists of transmitting an image at high quality to a decoder that already has access to a noisy or lower-quality version of the image. Although the asymmetric setup is commonly studied in information theory for its apparent simplicity, the symmetric setup becomes more relevant in practical scenarios where the side information $\mathbf{Y}$ is also encoded.

1.1.2 Goal-oriented communications

In the previous distributed source coding problem, the main objective of the fusion center is to reproduce the original data, mostly focusing on minimizing error probability or distortion between original and reconstructed data [8, 9]. However, in modern communication systems, the focus may shift toward addressing a specific task. Especially, in the emerging field of goal-oriented communications [10, 11, 12], the objective is no longer to only reconstruct the data but rather to enable the fusion center to apply specific tasks, such as classification, decision-making, or semantic analysis, directly on the received data. By transmitting only task-relevant information, this approach reduces the communication rate and may even improve the performance of the target task. For instance, a fire detector targeting to decide between the presence or absence of a fire may only need a few bits rather than the full observation to make the decision. In the same way, in scenarios like image compression for classification purposes [13, 14], the goal is to transmit only the information necessary for accurate classification, rather than the entire raw data.

1.1.3 Decision-making over coded data

In this thesis, we focus on the specific case of decision-making, where the objective of the fusion center is to make decisions directly from the received data. As examples of applications,

one may consider embedded sensors on the human body for health disease detection, underwater activity monitoring, traffic jam detection from route planning of autonomous vehicles, or fire alarm detection [15]. Another relevant example is smart farming systems, where sensors and drones transmit information in real-time to a server for decision-making regarding resource management, such as water and fertilizer levels [16]. In information theory, the problem of decision-making over coded data was formulated as distributed hypothesis testing (DHT) [17].

1.2 Distributed Hypothesis Testing (DHT)

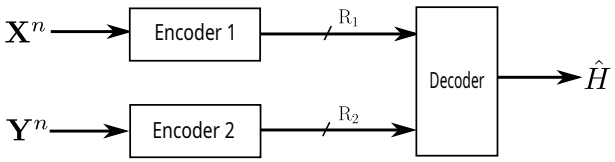

FIGURE 1.2 – Distributed hypothesis testing

The DHT setup focuses on binary hypothesis testing where the decoder aims to distinguish between two possible hypotheses called null hypothesis denoted by $\mathcal{H}_0$ and alternative hypothesis denoted by $\mathcal{H}_1$.

1.2.1 System model

Consider a DHT problem involving two separate terminals, one observing a source $\mathbf{X}$, and the other a source $\mathbf{Y}$, as illustrated in Figure 1.2. We distinguished two different setups:

1. Symmetric setup: here, both $\mathbf{X}$ and $\mathbf{Y}$ are encoded at rates R_1 and R_2, respectively.
2. Asymmetric setup: in this setup, $\mathbf{Y}$ is fully available at the decoder, i.e., $R_2 = \infty$, as in the Wyner-Ziv setup described above [8].

We assume that the sources $\mathbf{X}$ and $\mathbf{Y}$ follow a joint probability distribution determined by one of two hypotheses, $\mathcal{H}_0$ or $\mathcal{H}_1$ [18, 19, 20]. For instance, in the special case of testing against independence, $\mathbf{X}$ and $\mathbf{Y}$ are independent under the alternative hypothesis $\mathcal{H}_1$. The objective of the decoder is to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$.

The performance of DHT is characterized by two error probabilities referred to as Type-I and Type-II error probabilities, denoted by α_n and β_n, , where n is the source length. Type-I error occurs when $\mathcal{H}_1$ is chosen while hypothesis $\mathcal{H}_0$ is true, whereas Type-II error is when $\mathcal{H}_0$ is selected under hypothesis $\mathcal{H}_1$. In this thesis, in order to fully address DHT, we investigate both information-theoretic performance limits and practical coding schemes for this setup.

1.2.2 Information-theoretic analysis of DHT

In the information-theoretic framework of DHT, the Type-I error probability must remain below a prescribed threshold. The requirement for the Type-II error probability is to decay exponentially to zero, with the decay rate defined as the error exponent [21]. The primary objective is to determine the achievable error exponent [18, 19].

In the asymmetric case, the DHT problem has been well studied in the literature for i.i.d. sources. Different achievable coding schemes have been proposed, providing increasingly precise bounds on the error exponent, as detailed in Chapter 2. Ahlswede and Csiszár introduced the so-called quantization scheme [18], which provides a lower bound on the error exponent and is optimal for specific cases such as testing against independence. Han later refined this scheme, yielding a tighter lower bound [21]. To better exploit source correlation, Shimokawa et al. introduced the quantize-binning scheme [22], which further reduces the coding rate. The quantize-binning scheme closely resembles the Wyner-Ziv scheme [8] but is specifically tailored for hypothesis testing, focusing on the analysis of Type-I and Type-II errors rather than distortion. Moreover, for a given coding rate, the quantize-binning scheme achieves a tighter error exponent bound compared to the quantization scheme. However, this scheme is not always optimal, and its optimality and sub-optimality have been analyzed in several works [23, 2, 24], with further improvements presented in [25]. The quantize-binning scheme has also been extended to more complex setups, including discrete memoryless channels [26], multiple-access channels [27], and two-hop relay networks [28].

While error exponent bounds for DHT have been explored in the asymmetric setup, the symmetric setup has received significant attention in the specific case of zero-rate compression, where one or both coding rates asymptotically approach zero [21, 29, 30, 31, 32]. Although this scenario is not relevant for usual lossless or lossy data compression [7, 8], it has important applications in statistics [21]. Since no explicit coding scheme is required when considering zero-rate compression, research has primarily focused on the design of testing schemes [32, 33] and the characterization of achievable error exponents [21, 29, 30, 31, 32, 33].

1.3 Limitations of Previous Work on DHT

The achievable coding schemes for DHT, described previously, assume that the sources $\mathbf{X}$ and $\mathbf{Y}$ generate independent and identically distributed (i.i.d.) pairs of symbols (X_t, Y_t), $t \in [\![1, n]\!]$[2, 27, 34, 23, 35], or block-i.i.d. vectors $(\mathbf{X}_t^M, \mathbf{Y}_t^M)$ [36, 37]. However, i.i.d. and block-i.i.d. models are often inadequate for capturing the statistics of signals like time series or videos, which cannot be decomposed into fixed-length independent blocks and are frequently non-stationary and/or non-ergodic. Additionally, the proofs for achievable error exponents in these models often rely on the method of types [38], which cannot be applied to non-i.i.d. sources.

This means that the derived error exponent bounds are also restricted to i.i.d. sources. The key question is then whether it is possible to develop a more general information-theoretic analysis that does not rely on assumptions of independence between symbols, stationarity, or ergodicity. Therefore, the first objective of this thesis is to investigate DHT using a generic sources model, that is non-i.i.d. and can account for non-stationary and non-ergodic signals, while still encompassing the i.i.d. models as particular instances.

In addition, while the DHT information-theoretic performance is known in the i.i.d. case, the problem of designing practical short-length coding schemes for this problem has been by far less investigated. Some existing works have already introduced practical binary quantizers [39], binning schemes [4], and quantize-binning schemes [40] for Wyner-Ziv coding, all constructed with linear block codes. However, these constructions were designed for the purpose of source reconstruction and typically involve very long source sequences, often exceeding 10^5 bits. Moreover, they rely on message-passing algorithms that perform poorly with shorter sequences. In contrast, DHT inherently deals with short-length sequences, where just a few dozen bits may suffice to make a correct decision. Notably, no previous work has proposed practical implementation of the quantization and quantize-binning schemes dedicated to DHT. The construction of efficient short-length linear block codes is often known to be a challenging problem [41]. Consequently, an important question arises regarding whether binary quantizers and quantize-binning schemes are efficient structures for practical short-length DHT. Another relevant challenge lies in how to perform the hypothesis test, particularly since the strategies proposed in information theory proofs are not directly implementable in practice. In this thesis, we address all these key points.

1.4 Main contributions

This thesis presents two main contributions, as outlined in the following sections: DHT for general sources and design of practical short-length coding schemes for DHT.

1.4.1 DHT for general non-i.i.d. sources model

This thesis first investigates DHT using a generic sources model, that is non-i.i.d. and can account for non-stationary and non-ergodic signals. We investigate the performance of DHT for these source models, specifically deriving a generic formula for the achievable error exponent. For this, we propose an information-theoretic coding scheme that achieves a general lower bound on the error exponent. Our achievability proof is based on the information spectrum approach, originally introduced by Han in [42].

To validate our general analysis, we show that when applied to the specific case of i.i.d. sources, our derived error exponent aligns with the well-established results found in [2]. We then extend the analysis to other source models of interest, including the stationary and ergodic Gaussian

source model, as well as the Gilbert-Elliot (GE) source model. The GE model is particularly relevant for practical applications, such as video coding [43], link quality estimation [44], and packet loss analysis [45], and has not been previously studied in the context of DHT. For this model, we propose an efficient numerical method to evaluate the error exponent and explore the impact of different GE model parameters on the error exponent. This provides valuable insights for the design of practical coding schemes for DHT.

1.4.2 Practical short-length coding schemes for DHT

As a second main contribution, we propose practical quantization and quantize-binning schemes specifically tailored for DHT in both symmetric and asymmetric setups. Similar to previous code designs for Wyner-Ziv coding, our schemes are built using binary linear block codes but are tailored for short block lengths (less than 100 bits). For each proposed scheme, we first address how to perform the hypothesis test in the practical case. Then, to evaluate the code performance, we derive exact analytical expressions for the Type-I and Type-II error probabilities of each scheme. These tools are novel, and enable the optimization and comparison of the proposed schemes across a broad range of source and code parameters. Our simulation results demonstrate the superior performance of our schemes compared to the baseline uncoded schemes.

1.5 Organization of the thesis

The organization of the thesis is as follows. Chapter 2 provides a review of the existing literature on DHT. Chapter 3 focuses on DHT for the general sources model and presents the achievable bound for the error exponent. Chapter 4 applies the general error exponent bound to specific cases, including the stationary and ergodic Gaussian source model and the GE source model. Chapter 5 introduces practical short-length coding schemes for DHT and evaluates their performance. Finally, Chapter 6 summarizes the findings and presents some perspectives.

Publications

— I. Salihou Adamou, E. Dupraz, T. Matsumoto, "An Information-Spectrum Approach to Distributed Hypothesis Testing for General Sources," in *Proc. Int. Zurich Seminar on Information and Communication (IZS)*, pp. 144–148, Zurich, Switzerland, 2024.

— E. Dupraz, I. S. Adamou, R. Asvadi, and T. Matsumoto, "Practical Short-Length Coding Schemes for Binary Distributed Hypothesis Testing," *International Symposium on Information Theory (ISIT)*, 2024.

— I. S. Adamou, E. Dupraz, Z. Amin, and T. Matsumoto, "Error-Exponent of Distributed Hypothesis Testing for Gilbert-Elliot Source Models," in *International Symposium on Topics in Coding (ISTC)*, 2023.

— I. Salihou Adamou, E. Dupraz, T. Matsumoto, "Test d'hypothèses distribuées pour des modèles de sources générales non-iid, non-stationnaires, et non-ergodiques", in *GRETSI 2023 : XXIXème Colloque Francophone de Traitement du Signal et des Images*, pp. 1061–1064, 2023.

— I. Salihou Adamou, E. Dupraz, T. Matsumoto, "Practical Binary Quantizers and Quantize-Binning Schemes for Two-Encoders Distributed Hypothesis Testing", To be submitted at *IEEE Transactions on Communications (TCOM)*

Chapitre 2

State of the Art

2.1 Introduction

In this chapter, we review the existing literature on Distributed Hypothesis Testing (DHT). Section 2.3 begins with a description of the system model involving one encoder and one decision center with side information, referred to as the asymmetric setup. For this setup, we present the existing information-theoretic coding schemes and their corresponding achievable error exponent bounds. Section 2.4 focuses on the symmetric setup of DHT, which involves two encoders. Section 2.5 presents DHT over more complex scenarios including noisy channels and multi-hop networks.

2.2 Notation

The following notation will be used throughout this thesis. The set of integers between 1 and M is denoted by $[\![1, M]\!]$. The cardinality of a finite set A and of the range of a function f are denoted by $|A|$, and $\|f\|$, respectively. Calligraphic letter $\mathcal{X}$ represents a finite set. Random variables are indicated by capital letters, such as X, while their specific realizations are represented by lowercase letters, like x. The set of probability distributions on $\mathcal{X}$ is denoted by $\mathcal{P}(X)$. Random sequences of length n are denoted with bold capital letters $\mathbf{X}^n = (X_1, X_2, \ldots, X_n)$, with their realizations denoted as bold lowercase letters $\mathbf{x}^n$. For any $\mu \geq 0$, the set of all μ-typical sequences for X is denoted by $T_\mu^n(X)$.

The Kullback–Leibler divergence between two distributions P and Q is represented by $D(P||Q)$. The mutual information between X and Y is denoted by $I(X;Y)$, and the entropy of a source X is denoted by $H(X)$. We also denote by $I(\mathbf{u}^n;\mathbf{u}^n)_e$ and $H(\mathbf{x}^n)_e$, the empirical mutual information between X and Y, and the empirical entropy of a source X, respectively. Additionally, H_2 represents the binary entropy function. The discrete probability distribution of a Bernoulli random variable is denoted by $\mathrm{Bern}(p)$.

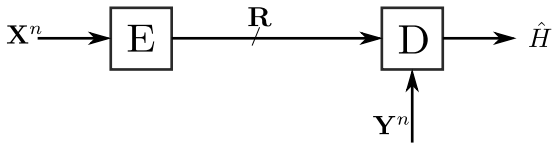

FIGURE 2.1 – Distributed Hypothesis testing with side information.

2.3 Distributed Hypothesis Testing with Side Information (asymmetric setup)

DHT, first introduced by Berger in [17] and later considered in [18, 21, 22], extends the Wyner-Ziv setup [8], in which the decoder aims to reconstruct a source within a specified distortion using side information available only to the decoder. In the DHT setup, the objective of the decoder shifts to making a decision between two hypotheses.

2.3.1 Information-theoretic formulation of DHT

Consider a two-terminal problem involving an encoder that observes a source sequence $\mathbf{X}^n$ and a decoder that observes a side information sequence $\mathbf{Y}^n$, as illustrated in Figure 2.1. The encoder sends a coded version of $\mathbf{X}^n$ while the decoder aims to decide between two hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, based on both $\mathbf{Y}^n$ and the received coded data. Since most of existing works focus on the independent and identically distributed (i.i.d.) source model, we also assume in this chapter that the n symbols pair (X_t, Y_t) for $t \in [\![1,\ n]\!]$, of the sequences $\mathbf{X}^n$ and $\mathbf{Y}^n$ are i.i.d. Moreover, the joint probability mass function (p.m.f.) of the tuple (X, Y) depends on the underlying hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. These hypotheses are defined as

$$\mathcal{H}_0 : (X, Y) \sim P_{XY}, \tag{2.1}$$

$$\mathcal{H}_1 : (X, Y) \sim P_{\bar{X}\bar{Y}}, \tag{2.2}$$

where the marginal probability distributions P_X and P_Y do not depend on the hypothesis. Hypothesis testing against independence [46, 47] is a special case of (3.14) and (3.15), in which X and Y are assumed to be independent under the alternative hypothesis $\mathcal{H}_1$, i.e., $P_{\bar{X}\bar{Y}} = P_X P_Y$.

We consider the following usual coding scheme defined in the literature on DHT [19, 35].

Definition 2.1 *Given a rate parameter $R \geq 0$, consider a sequence $(f^{(n)}, g^{(n)})_{n\in\mathbb{N}}$ of encoding and decoding functions, defined for each $n \in \mathbb{N}$ by*

$$f^{(n)} : \mathcal{X}^n \longrightarrow \mathcal{M}_n = [\![1, M_n]\!], \tag{2.3}$$

$$g^{(n)} : \mathcal{M}_n \times \mathcal{Y}^n \longrightarrow \mathcal{H} = \{\mathcal{H}_0, \mathcal{H}_1\}, \tag{2.4}$$

where the sequence $(M_n)_{n\in\mathbb{N}}$ *satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R. \tag{2.5}$$

Given that the decoder decides between the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, we can define two types of errors: Type-I error probability α_n, which occurs when $\mathcal{H}_1$ is chosen while hypothesis $\mathcal{H}_0$ is true, and Type-II error probability β_n, which occurs when $\mathcal{H}_0$ is selected under hypothesis $\mathcal{H}_1$ [19, 35]. More formally:

Definition 2.2 *The Type-I and Type-II error probabilities,* α_n *and* β_n*, for each* $n \in \mathbb{N}$*, are defined as*

$$\alpha_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^n\right), \mathbf{Y}^n\right) = \mathcal{H}_1 \mid \mathcal{H}_0 \text{ is true}\right], \tag{2.6}$$

$$\beta_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^n\right), \mathbf{Y}^n\right) = \mathcal{H}_0 \mid \mathcal{H}_1 \text{ is true}\right]. \tag{2.7}$$

In DHT setups, it is common to require the Type-I error α_n to be below a specified threshold, while focusing on the exponential decay of the Type-II error β_n, which is sometimes called "Stein regime" [18, 19]. The objective of the information-theoretic analysis is then to characterize the achievable Type-II error exponent.

Definition 2.3 (Achievability under rate constraints) *A Type-II error exponent* θ *is said to be achievable for a given* $R \geq 0$*, if for each* $\epsilon > 0$ *and for a large blocklength* n*, there exists a sequence* $\left(f^{(n)}, g^{(n)}\right)_{n\in\mathbb{N}}$ *of encoding and decoding functions such that the Type-I and Type-II error probabilities* α_n *and* β_n *satisfy*

$$\alpha_n \leq \epsilon, \tag{2.8}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta. \tag{2.9}$$

In the literature, various achievable coding schemes have been proposed to characterize the achievable error exponent, θ. In the following sections, we present the existing achievable coding schemes and their corresponding error exponents' bounds for DHT with side information.

2.3.2 Ahlswede and Csiszár' scheme

In [18], Ahlswede and Csiszár proposed an achievable scheme in which the encoder transmits both a quantized version of X and its type (which is sent with zero rate asymptotically) to the decoder. The decoder then uses the received type and the quantized value of X to perform the

test with the help of Y. The following theorem provides a lower bound on the achievable error exponent for this scheme:

Theorem 2.1 ([18, Theorem 5] Ahlswede and Csiszár error exponent) *For every $R > 0$,*

$$\theta_{AC}(R) \geq D\left(P_X \| P_{\bar{X}}\right) + \sup_{\substack{P_{U|X}: \\ I(X,U) \leq R}} D\left(P_{UY} \| P_{\bar{U}\bar{Y}}\right), \tag{2.10}$$

where the Markov chains $U - X - Y$, and $U - \bar{X} - \bar{Y}$ hold.

The proof is provided in [18]. The first term of (2.10) represents the contribution of the type of X and is equal to 0 when the marginal distribution of X is the same under $\mathcal{H}_0$ and $\mathcal{H}_1$. The second one comes from the fact that a lossy description U of X was sent with rate R.

The scheme of [18], although suboptimal in general, is proven to be optimal for the problem of testing against independence [18]. In this particular case, the error exponent (2.10) simplifies to:

Theorem 2.2 ([18, Theorem 2] Error exponent when testing against independence) *For every $R > 0$, when $P_{\bar{X}\bar{Y}} = P_X \times P_Y$:*

$$\theta_{AC}(R) = \max_{\substack{P_{U|X}: \\ I(X,U) \leq R \\ |\mathcal{U}| \leq |\mathcal{X}|+1}} I(U;Y). \tag{2.11}$$

Note that, the first term of (2.10) does not appear in (2.11) because it is assumed in [18] that the marginal distribution of X, P_X under $\mathcal{H}_0$ is the same as the marginal distribution $P_{\bar{X}}$ under $\mathcal{H}_1$. In addition, the Kullback-Leibler divergence term of (2.10) simplifies to the mutual information expression in (2.11).

2.3.3 Han scheme

Han improved upon the quantization scheme of Ahlswede and Csiszár by transmitting both the quantized version of X (represented by U) and the joint type (which is sent with zero rate asymptotically) of X and U to the decoder. At the decoder, by using the side information $\mathbf{y}^n$ and the received quantized vector $\mathbf{u}^n$, the decoder decides $\mathcal{H}_0$ if $(\mathbf{u^n}, \mathbf{x^n})$ and $(\mathbf{u^n}, \mathbf{y^n})$ are jointly typical under $\mathcal{H}_0$. This leads to the following lower bound on the achievable error exponent [19].

Theorem 2.3 ([19, Theorems 2,3] Han lower bound) *For every $R > 0$,*

$$\theta_{HAN}(R) \geq \sup_{\substack{P_{U|X}: \\ I(X,U) \leq R}} \min_{P_{\tilde{U}\tilde{X}\tilde{Y}} \in \mathcal{P}_{HAN}} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\bar{U}\bar{X}\bar{Y}}\right), \tag{2.12}$$

$$\mathcal{P}_{HAN} = \left\{P_{\tilde{U}\tilde{X}\tilde{Y}} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{U}\tilde{Y}} = P_{UY}\right\} \tag{2.13}$$

Here, $\mathcal{P}_{\text{HAN}}$ represents the set of the all possible joint types $P_{\tilde{U}\tilde{X}\tilde{Y}}$ such that the joint type $P_{\tilde{U}\tilde{X}}$ coincides with the true distribution P_{UX} under $\mathcal{H}_0$, and the joint type $P_{\tilde{U}\tilde{Y}}$ coincide with the true distribution P_{UY} under $\mathcal{H}_0$. The detailed proof of (2.12) is provided in [19], using the method of types [38].

2.3.4 Shimokawa et al. scheme

Later on, Shimokawa et al. [22] proposed a quantize-binning scheme to achieve an error exponent tighter than (2.12). In this scheme, the encoder quantizes and then bins its observation $\mathbf{x}^n$, and the receiver performs the test directly using the received bin index and the side information $\mathbf{y}^n$. Binning, similar to Slepian-Wolf coding [7], takes advantage of the correlation between the sources X and Y, thereby reducing the compression rate.

The quantize-binning scheme operates as follows. To construct the codebook, we first generate $2^{nR'}$ sequences $\mathbf{u}^n$ randomly generated according to a pre-defined distribution $P_{U|X}$, and then distribute them uniformly into 2^{nR} bins, with $R < R'$. The codebook and the bin assignment are revealed to the encoder and the decoder. To encode $\mathbf{x}^n$, the encoder first quantizes $\mathbf{x}^n$ by selecting a codeword $\mathbf{u}^n$ that is jointly typical with $\mathbf{x}^n$ and sends to the decoder the index of the bin to which the sequence $\mathbf{u}^n$ belongs. The joint type of $(\mathbf{x}^n, \mathbf{u}^n)$ is also sent to the decoder, which requires zero additional rate asymptotically. In the next paragraphs, we will describe the testing strategy introduced by Shimokawa et al.[22].

The scheme of Shimokawa et al. was later considered by many works [15, 2, 24, 25], where lower and upper bounds on the error exponents were established. Before discussing the lower bound achieved by Shimokawa et al., we first present a simple lower bound on the error exponent from the literature, provided by Katz et al. [2]. Although the result of [2] was obtained long after the initial work [22], presenting first this result is insightful for understanding the interest of the achievable coding scheme of Shimokawa et al.

2.3.4.1 Katz and Piantanida lower bound on the error exponent for the quantize-binning scheme

The bound on the error exponent in [2] is achieved by adopting a simple solution at the decoder. From the received bin index and side information $\mathbf{y}^n$, the decoder goes over all sequences within the bin. For each sequence $\mathbf{u}^n$ in the bin, the decoder assumes it is the correct one and decides $\mathcal{H}_0$ if $(\mathbf{u}^n, \mathbf{y}^n)$ is jointly typical [2]. The following lower bound on the error exponent is achieved from this scheme [2].

Theorem 2.4 ([2, Proposition 4] Simpler bound on the error exponent) *For any* $R >$

$$\theta_{KP}(R) \geq \sup_{\substack{P_{U|X}: \\ I(U;X|Y)<R<I(U;X)}} \min \left[R - [I(X;U) - I(U;Y)],\ \mathcal{D}\left(P_{UY} \| P_{\bar{U}\bar{Y}}\right) \right] \tag{2.14}$$

The proof is provided in [2]. On the right-hand side of (2.14), the first term reflects the binning error and the second one represents the testing error. The resulting error exponent therefore corresponds to a trade-off between the two errors due. A precise description of X using U allows the decoder to perform hypothesis testing with high accuracy. However, this approach results in a large codebook and larger bins, making the binning error more likely. On the other hand, a low description leads to a smaller codebook and bins, reducing the risk of binning errors. However, the retrieved sequence is less effective for hypothesis testing, making the testing error more likely. This trade-off between the two error events was properly addressed in [2].

2.3.4.2 Shimokawa, Han, and Amari lower bound

The lower bound in (2.14) is simple to compute and was obtained through a straightforward decoding approach, where $\mathcal{H}_0$ is accepted if there exists a pair $(\mathbf{u}^n, \mathbf{y}^n)$ which is jointly typical under $\mathcal{H}_0$, assuming $\mathbf{u}^n$ is the correct sequence. However, this approach may not be optimal, as the joint probability P_{XY} between the source X and side information Y depends on the hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$. Therefore, Shimokawa, Han, and Amari (SHA) [22] had introduced a more refined decoding strategy. Instead of assuming that $\mathbf{u}^n$ is the correct sequence, their scheme employs minimal empirical entropy check as follows. From the received bin index and the side information $\mathbf{y}^n$, it selects $\hat{\mathbf{u}}^n$ if:

$$H_e(\hat{\mathbf{u}}^n|\mathbf{y}^n) < H_e(\tilde{\mathbf{u}}^n|\mathbf{y}^n), \quad \text{for all } \tilde{\mathbf{u}}^n \neq \hat{\mathbf{u}}^n. \tag{2.15}$$

It then performs the test by selecting $\mathcal{H}_0$ if the extracted sequence $\hat{\mathbf{u}}^n$ and the side information $\mathbf{y}^n$ are jointly typical under $\mathcal{H}_0$ [22, 23]. This scheme achieves a tighter lower bound on the error exponent compared to (2.14).

Theorem 2.5 ([20, Theorem 4.3] Shimokawa, Han and Amari (SHA) lower bound) *For any $R > 0$,*

$$\theta_{SHA}(R) \geq \sup_{\substack{P_{U|X}: \\ I(U;X|Y)<R<I(U;X)}} \min \Bigg[\min_{P_{\tilde{U}\tilde{X}\tilde{Y}} \in \mathcal{P}_{SHA}\left(P_{U|X}\right)} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\bar{U}\bar{X}\bar{Y}}\right) + R - I(U;X \mid Y),$$
$$\min_{P_{\tilde{U}\tilde{X}\tilde{Y}} \in \mathcal{P}_{HAN}} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\bar{U}\bar{X}\bar{Y}}\right) \Bigg], \tag{2.16}$$

where

$$\mathcal{P}_{SHA}\left(P_{U|X}\right) := \{P_{\tilde{U}\tilde{X}\tilde{Y}} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{Y}} = P_Y, H(\tilde{U} \mid \tilde{Y})_e \geq H(U \mid Y)\Big\}, \tag{2.17}$$

In (2.16), the first two terms represent the binning error. Unlike (2.14), the Kullback-Leibler divergence in the binning error of (2.16) arises from the minimal empirical entropy decoding, which is absent in (2.14). The third term of (2.16) represents the testing error, which is achieved by Han's scheme. It is evident that the lower bound (2.16) is tighter than (2.12). Finally, note that, in (2.17), $H(\tilde{U} \mid \tilde{Y})_e \geq H(U \mid Y)$ is equivalent to $I(\tilde{U};\tilde{Y})_e \leq I(U;Y)$.

2.3.5 Kochman and Wang improvement

The empirical entropy minimization in the SHA scheme is restricted to $P_{\tilde{U}\tilde{X}\tilde{Y}}$ satisfying $I(\tilde{U};\tilde{Y})_e \leq I(U;Y)$ as shown in (2.17). However, recently, Kochman and Wang have shown that a better error exponent bound for the quantize-binning scheme can be achieved by replacing the condition $I(\tilde{U};\tilde{Y})_e \leq I(U;Y)$ in (2.17) by $I(\tilde{U};\tilde{Y})_e \leq R'$, where R' corresponds to the size of a bin. This choice enlarges the set of possible empirical distributions compared to (2.17), therefore potentially reducing the minimization in (2.16) and improving the error exponent. In fact, the error exponent derived in [25] is the same as the SHA error exponent (2.16), except that in (2.17), $I(U;Y)$ is replaced by R' (see [25], for more details).

2.3.6 Optimality of SHA scheme

It is worth noting that the quantize-binning scheme is not optimal in general. Rahman and Wagner have shown in [23] that the error exponent (2.16) is optimal when the side information Y is replaced by a tuple of sources (Y, Z) such that X and Y are conditionally independent knowing Z under $\mathcal{H}_1$ [23]. More formally, in Fig.2.1, the decoder observes the sequences $\mathbf{Y}^n$ and $\mathbf{Z}^n$ instead of $\mathbf{Y}^n$ only. Therefore, the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are now defined as

$$\mathcal{H}_0 : (X, Y, Z) \sim P_{XYZ}, \tag{2.18}$$

$$\mathcal{H}_1 : (X, Y, Z) \sim P_{\bar{X}\bar{Y}\bar{Z}} = P_Z P_{X|Z} P_{Y|Z}. \tag{2.19}$$

This problem is the conditional version of the test against independence studied by Ahlswede and Csiszár [18].

Theorem 2.6 ([23, Theorem 3] Error exponent for testing against conditional independence)

For any $R > 0$,

$$\theta_{RW}(R) = \sup_{\substack{U: I(U;X|Z) \leq R \\ U \to X \to Z \\ |\mathcal{U}| \leq |\mathcal{X}|+1}} I(U;Y \mid Z) \tag{2.20}$$

The proof is provided in [23]. In addition, Rahaman and Wagner provided both inner and outer bounds for the error exponent of the quantize-binning scheme for the problem of L-encoder hypothesis testing against conditional independence. In the special case of $L = 1$, as defined in (2.18) and (2.19), they established that the inner and outer bounds coincide with (2.20), which proves that the quantize-binning scheme is optimal for this problem. They also proved that in this case, the SHA lower bound (2.16) satisfies $\theta_{SHA}(R) \geq \theta_{RW}(R)$, establishing that the SHA scheme is also optimal in this case.

However, Watanabe provided in [24] an example where the SHA scheme is sub-optimal, which we now describe in detail.

2.3.7 Sub-optimality of the quantize-binning scheme

In this section, we present the example of [24], which illustrates the sub-optimality of the SHA scheme in other cases than testing against conditional independence. Let $X \sim \text{Bern}\left(\frac{1}{2}\right)$, and $Y = X \oplus E$, where E is a binary random variable independent of X, and we denote $p = \mathbb{P}(E = 1)$ with $0 < p \leq 1/2$. The two hypotheses are expressed as:

$$\begin{cases} \mathcal{H}_0 : & p = p_0, \\ \mathcal{H}_1 : & p = p_1, \end{cases} \tag{2.21}$$

with $p_0 < p_1$. For this example, Figure 2.2 shows the error exponents of the HAN scheme (2.12) and the SHA scheme (2.16), as achieved by the critical rate defined by Watanabe in [24]. The critical rate is the minimum communication rate required to achieve the Stein exponent [24]:

$$R_{\text{cr}} = \inf \{R : E(R) = D(P_{XY} \| Q_{XY})\}. \tag{2.22}$$

Watanabe showed that to attain the Stein exponent, one should take $U = X$ in the quantize-binning bound, i.e., no quantization is used (see [24, Prop.2]). To achieve (2.22), using Han's scheme requires a rate $R_{cr} = I(X;X) = 1$ bit. Since (2.16) lies below (2.12) for all $R < R_{cr}$, the SHA scheme has worse performance than the Han scheme.

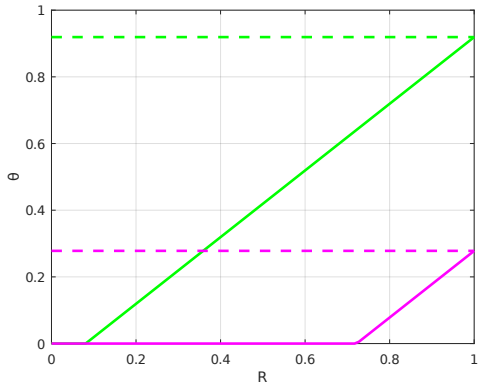

FIGURE 2.2 – Error exponents (2.10) (dashed curves) and (2.16) (plain curves) for $U = X$ for the values of $p = 0.01$ (green) and $p = 0.1$(magenta).

2.4 Distributed Hypothesis Testing with two encoders (Symmetric setup)

In the previous sections, we focused on the asymmetric setup of DHT, where the source X is encoded and the side information Y is fully observed at the decoder. Another relevant setup of DHT, which is referred to as symmetric setup, is the case where both X and Y are encoded at rates R_1 and R_2, respectively. The definitions of the encoding and decoding functions of the asymmetric setup can be straightforwardly extended to the symmetric setup as follows.

Definition 2.4 *Given rate parameters $R_1 \geq 0$ and $R_2 \geq 0$, consider a sequence $(f_1^{(n)}, f_2^{(n)}, g^{(n)})_{n\in\mathbb{N}}$ of encoding and decoding functions in the symmetric setup, defined for each blocklenght $n \in \mathbb{N}$ by*

$$f_1^{(n)} : \mathcal{X}^n \longrightarrow \mathcal{M}_n = [\![1, M_n]\!], \tag{2.23}$$

$$f_2^{(n)} : \mathcal{Y}^n \longrightarrow \mathcal{N}_n = [\![1, N_n]\!], \tag{2.24}$$

$$g^{(n)} : \mathcal{M}_n \times \mathcal{N}_n \longrightarrow \mathcal{H} = \{\mathcal{H}_0, \mathcal{H}_1\}, \tag{2.25}$$

where the sequence $(M_n)_{n\in\mathbb{N}}$ and $(N_n)_{n\in\mathbb{N}}$ satisfy

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R_1, \tag{2.26}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log N_n \leq R_2, \tag{2.27}$$

respectively, and M_n and N_n are the cardinalities of the alphabet sets $\mathcal{M}_n$ and $\mathcal{N}_n$, respectively.

2.4.1 Existing works on the error exponent bounds for the symmetric setup

The primary challenge in the symmetric setup is to determine the optimal error exponent [20]. However, Han [21] provided a lower bound on the achievable error exponent for the quantization scheme.

Theorem 2.7 ([21, Theorem 6] Han Lower bound for the symmetric setup) *For every* $R_1 > 0$, *and* $R_2 > 0$,

$$\theta_{HAN}(R_1, R_2) \geq \sup_{\substack{U,V: \\ I(X,U)\leq R_1 \\ I(Y,V)\leq R_2 \\ U\to X\to Y\to V}} \min_{P_{\tilde{U}\tilde{X}\tilde{Y}\tilde{U}}\in\mathcal{P}} D\left(P_{\tilde{U}\tilde{X}\tilde{Y}\tilde{V}} \| P_{\bar{U}\bar{X}\bar{Y}\bar{V}}\right), \tag{2.28}$$

$$\mathcal{P} = \left\{P_{\tilde{U}\tilde{X}\tilde{Y}} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{U}\tilde{Y}} = P_{UY}, P_{\tilde{U}\tilde{V}} = P_{UV}\right\}, \tag{2.29}$$

where $\bar{U}, \bar{V}$ *are the random variables such that* $\bar{U} \to \bar{X} \to \bar{Y} \to \bar{V}$.

The error exponent (2.28) is a generalization of the error exponent (2.12) achieved for the quantization scheme for the asymmetric setup. The proof is provided in [21].

In the specific case of binary sources, Haim and Kochman [33] utilized the Körner-Marton decoder [48] to establish error exponents for Type-I and Type-II error probabilities in the symmetric setup for the binning-only scheme. However, to the best of our knowledge, no previous work has explicitly determined an achievable error bound for the quantize-binning scheme in the symmetric setup. The most investigated case for the symmetric setup is the zero-rate hypothesis testing problem which we now briefly describe.

2.4.2 Zero-rate Hypothesis Testing Problem

The zero-rate hypothesis testing involves the two terminals transmitting their messages to the decoder at zero rate. More formally, the rate constraints (2.26) and (2.27) satisfy

$$R_1 = \frac{1}{n}\log M_n \to 0, \quad R_2 = \frac{1}{n}\log N_n \to 0, \tag{2.30}$$

asymptotically. The objective of the decoder is still to decide between hypothesis $\mathcal{H}_0$ or the alternative hypothesis $\mathcal{H}_1$.

In the case of one-bit data compression where the encoders $f_1^{(n)}$ and $f_2^{(n)}$ are such that $\left\|f_1^{(n)}\right\| = \left\|f_2^{(n)}\right\| = 2$, independently from n, Han established the optimal error exponent for this setup through a single-letter characterization [21].

Theorem 2.8 ([21, Theorem 8] Optimal error exponent for one-bit compression)

$$\theta_{HAN}(1,1) \geq \min_{\substack{P_{\tilde{X}\tilde{Y}}: \\ P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D\left(P_{\tilde{X}\tilde{Y}} \| P_{\bar{X}\bar{Y}}\right). \tag{2.31}$$

To achieve (2.31), the encoding functions $f_1^{(n)}$ and $f_2^{(n)}$ are defined such that $f_1^{(n)}(\mathbf{x}^n) = \mathcal{H}_0$ if $\mathbf{x}^n \in T_\mu^n(X)$ and $f_1^{(n)}(\mathbf{x}^n) = \mathcal{H}_1$, otherwise; $f_2^{(n)}(\mathbf{y}^n) = \mathcal{H}_0$ if $\mathbf{y}^n \in T_\mu^n(Y)$ and $f_2^{(n)}(\mathbf{y}^n) = \mathcal{H}_1$, otherwise (see [21] for more details on the proof). Here $T_\mu^n(X)$ and $T_\mu^n(Y)$ are the μ-typical sets for X and for Y, respectively. In addition, Shalaby and Papamarcou [29] proved that Han's exponent (2.31) is tight under the positivity condition $P_{\bar{X}\bar{Y}} > 0, \quad (x,y) \in \mathcal{X} \times \mathcal{Y}$. Note that the Type-II error exponent (2.31) is derived with the constraint of $\alpha_n \leq \epsilon$. Han and Kobayashi have considered the same one-bit compression scheme with an exponential constraint $\alpha_n \leq e^{-nr}$ ($r > 0$) on the Type-I error probability [30]. They obtained a more general result than (2.31).

In the general case of zero-rate compression, the encoders $f_1^{(n)}$ and $f_2^{(n)}$ are such that $\left\|f_1^{(n)}\right\| = k_n$, and $\left\|f_2^{(n)}\right\| = h_n$, such that

$$k_n \to \infty, \quad h_n \to \infty \text{ as } n \to \infty \tag{2.32}$$

but

$$R_1 = \frac{1}{n}\log k_n \to 0, \quad R_2 = \frac{1}{n}\log h_n \to 0. \tag{2.33}$$

For this setup, Han and Kobayashi established the following optimal error exponent among the class of all zero-rate testing schemes [30, 31].

Theorem 2.9 ([20, Theorem 5.4]Optimal error exponent for zero-rate compression) *for any* $r > 0$,

$$\theta(0,0) \geq \min_{\substack{P_{\tilde{X}\tilde{Y}}: \\ P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y \\ D\left(P_{\tilde{X}\tilde{Y}} \| P_{\bar{X}\bar{Y}}\right) \leq r}} D\left(P_{\tilde{X}\tilde{Y}} \| P_{\bar{X}\bar{Y}}\right). \tag{2.34}$$

The error exponent (2.34) is achieved by the following test, as detailed in [30, 32]: after observing $\mathbf{x}^n$ and $\mathbf{y}^n$, the encoders transmit their respective types. Upon receiving the marginal types (t_x, t_y), the decoder calculates the projected relative entropy $E(t_x \times t_y || P_{XY})$[32]. If this value is below a certain threshold r, the decoder outputs $\mathcal{H}_1$; otherwise, it outputs $\mathcal{H}_0$.

Recently, Watanabe addressed the non-asymptotic performance of the zero-rate compression setup in [32]. He proposed a Neyman-Pearson-like test tailored for short block lengths, which outperforms the test proposed in [30]. However, the optimality of Watanabe's test for a given block length remains an open question.

2.5 Distributed hypothesis testing in more complex scenarios

DHT has also been extended to various network scenarios. For instance, it was extended to take into account noisy channels [27, 49], Discrete Memoryless Channels (DMC) [26], Multiple Access Channels (MACs) [28], and Broadcast channel [50]. In these works, the authors have developed coding schemes that either combine hypothesis testing and channel coding jointly or apply them separately. The resulting error exponent in these schemes is characterized by competing terms coming from hypothesis testing and channel coding [26, 50]. Notably, joint approaches consistently outperform separation-based methods by achieving higher error exponents [26]. However, the trade-offs between these competing exponents require further investigation. Additionally, the complexity of joint hypothesis testing and channel coding schemes must be carefully evaluated for practical implementation.

Some other works explored DHT with interactive terminals [51, 52]. In these setups, two nodes can interactively communicate over a noiseless, bidirectional link before one of them performs hypothesis testing. In this setup, the optimal error exponent for testing against independence was established in [51]. For more general hypotheses (not only testing against independence), a new achievable error exponent was provided in [52], which was shown to be optimal, given that it matches the previously known result of [51] when testing against independence. However, in these works, the coding scheme used for the interactive hypothesis testing problem is joint typicality encoding, similar to [18], without using binning techniques. Therefore, the investigation of quantize-binning schemes for interactive hypothesis testing remains an open question.

Another setup corresponds to DHT with multi-hop network [28, 53, 54, 55, 16], which has potential applications in the Internet of Things (IoT) and for sensor networks. In this setup, the transmitter communicates directly with a relay over a noise-free link but cannot communicate directly with the receiver. Furthermore, DHT under privacy constraints has also been considered [56].

2.6 Summary and Discussion

This chapter reviewed the literature on DHT. We first presented the asymmetric setup involving a single sensor and a single decision center with side information. For this setup, we described achievable coding schemes, including quantization and quantize-binning schemes. The existing results demonstrate that the quantize-binning scheme achieves a better lower bound on the error exponent compared to the quantization one. Secondly, we reviewed the DHT in the symmetric setup involving two encoders transmitting their compressed version at the decoder.

It is important to note that achievable error exponents in all the reviewed works are characterized only for i.i.d. source models. However, i.i.d. models are often inadequate for capturing

the statistics of signals like time series or videos. Consequently, in Chapter 3, we will investigate a more general non-i.i.d. models, that can account for non-stationary and non-ergodic signals.

In addition to the information-theoretic analysis of DHT, we will also investigate practical short-block length coding schemes for DHT. While previous works have introduced practical binary quantizers, binning schemes, and quantize-binning schemes using linear block codes, these were primarily designed for Wyner-Ziv coding aimed at source reconstruction, typically requiring very long source sequences (over 10^5 bits) [4, 39, 40]. These schemes rely on message-passing algorithms, which do not perform well for shorter sequences, particularly in DHT. Notably, no previous work has proposed practical implementation of the quantization and quantize-binning schemes dedicated to DHT. Therefore, in Chapter 4, we address the practical short-length coding schemes dedicated to DHT.

Chapitre 3

Distributed Hypothesis Testing For General non-i.i.d. Sources

3.1 Introduction

In this chapter, we focus on the asymmetric setup of DHT, which involves an encoder observing the source $\mathbf{X}$ and a decoder observing the side information $\mathbf{Y}$. The source $\mathbf{X}$ and the side information $\mathbf{Y}$ generate sequence of random variables $\mathbf{X}^n$ and $\mathbf{Y}^n$, respectively. Previous studies typically assume that $\mathbf{X}^n$ and $\mathbf{Y}^n$ are i.i.d. pairs of symbols (X_t, Y_t) for $t \in [\![1, n]\!]$ [17, 18, 22, 27, 34, 23, 2]. Some more complex source models have been investigated in [36, 37], which assume that the sources $\mathbf{X}^n$ and $\mathbf{Y}^n$ generate pairs of Gaussian vectors $(\mathbf{X}_t^M, \mathbf{Y}_t^M)$ for $t \in [\![1, n]\!]$, with statistical dependencies within each vector $\mathbf{X}_t^M$, $\mathbf{Y}_t^M$, respectively, and between the two vectors. However, the models of [36, 37] are block-i.i.d. in the sense that the successive pairs $(\mathbf{X}_t^M, \mathbf{Y}_t^M)$ are assumed to be i.i.d. with t.

Nevertheless, i.i.d. and block-i.i.d. models are often inadequate for capturing the statistics of signals like time series or videos, which cannot be decomposed into fixed-length independent blocks and are frequently non-stationary and/or non-ergodic. As a result, the objective of this chapter is to consider a more general source model that is non-i.i.d. and can account for non-stationary and non-ergodic signals, while still encompassing the i.i.d. models as particular instances. To investigate DHT under these conditions, we utilize information spectrum tools, which were first introduced in [42] and generally provide information theory results that are applicable to a broad range of source models. It should be noted that information spectrum has been previously used for hypothesis testing in [57], but only for the encoding of a source $\mathbf{X}$ alone, without the use of side information $\mathbf{Y}$.

Therefore, this chapter addresses DHT using general source models for $\mathbf{X}$ and $\mathbf{Y}$, as defined in [42], and provides an achievability scheme that yields a generic error exponent bound applicable to a wide range of source models, not necessarily i.i.d. Our achievability scheme utilizes information spectrum methods [42] to handle general sources. It provides a simple lower bound on the general error exponent that is relatively straightforward to compute for i.i.d. and/or stationary Gaussian sources. We show that, when applied to the specific case of i.i.d. sources, our general

error exponent aligns with the established results from [2], demonstrating the consistency of our broader analysis.

3.2 General sources model

This section provides definitions for general sources, which are presented in [42] for the information-theoretic analysis of any source model, not necessarily stationary or ergodic.

3.2.1 Model definition

We define general sources $\mathbf{X}$ and $\mathbf{Y}$ as two infinite sequences [42]

$$\begin{aligned}\mathbf{X} &= \{\mathbf{X}^n = (X_1, X_2, \cdots, X_n)\}_{n=1}^{\infty}, \\ \mathbf{Y} &= \{\mathbf{Y}^n = (Y_1, Y_2, \cdots, Y_n)\}_{n=1}^{\infty}\end{aligned} \tag{3.1}$$

of n-dimensional random variables $\mathbf{X}^n, \mathbf{Y}^n$, respectively. Each component random variable X_i, Y_i, $i \in [\![1, n]\!]$, takes values in a finite source alphabet $\mathcal{X}, \mathcal{Y}$, respectively. Next, $P_{\mathbf{X}^n}$ is the probability distribution of the length-n vector $\mathbf{X}^n$, and $P_{\mathbf{X}} = \{P_{\mathbf{X}^n}\}_{n=1}^{\infty}$ is the collection of all probability distributions $P_{\mathbf{X}^n}$. The same holds for the source $\mathbf{Y}$. In the above definition, we assume that the sequences $\mathbf{X}$ and $\mathbf{Y}$ satisfy the consistency condition in the sense that for any integers m, l such that $m < l$, the first m components of the sequence $\mathbf{X}^l$ are equal to the components of the sequence $\mathbf{X}^m$ [57]. The same holds for the sequence $\mathbf{Y}$.

We now describe two particular cases of the model described in (3.1). The first one consists of the usual scalar i.i.d. model in which the sequences $\mathbf{X}^n$ and $\mathbf{Y}^n$ come from two i.i.d. sources, *i.e.*, the successive pairs of symbols (X_n, Y_n) are independent and distributed according to the same joint distribution P_{XY}. The second case still relies on an i.i.d. model but for source vectors. In this case, the source sequences $\mathbf{X}^n$ and $\mathbf{Y}^n$ are defined as [36, 37]

$$\mathbf{X}^n = \left\{\mathbf{X}_t^M\right\}_{t=1}^n, \ \mathbf{Y}^n = \left\{\mathbf{Y}_t^M\right\}_{t=1}^n, \tag{3.2}$$

where $\{\mathbf{X}_t^M\}_{t=1}^n$ and $\{\mathbf{Y}_t^M\}_{t=1}^n$ are sequences of i.i.d. M-dimensional random vectors and the successive pairs $\left(\mathbf{X}_t^M, \mathbf{Y}_t^M\right)$ are distributed according to the same joint distribution $P_{\mathbf{X}^M\mathbf{Y}^M}$. The i.i.d. property of the successive M-length vectors simplifies the DHT analysis by enabling the application of an orthogonal transform, such as the Karhunen-Loève Transform (KLT), to the independent blocks $\mathbf{X}_t^M$ and $\mathbf{Y}_t^M$ [36, 37]. Our model described in (3.1) is more general since it considers infinite sequences without the i.i.d. assumption.

3.2.2 Information spectrum terms

In this section, we introduce information spectrum terms that will allow us to investigate general sources from an information-theoretic analysis. Before introducing these terms, we justify why conventional definitions of e.g., entropy are not sufficient to address general sources. Consider a sequence of random variables $\{Z_n\}_{n=1}^{\infty}$. We are interested in the asymptotic behavior of the normalized information rate, defined as

$$-\frac{1}{n}\log P(Z_1, Z_2, \ldots, Z_n). \tag{3.3}$$

For i.i.d. sources, the joint probability $P(Z_1, Z_2, \ldots, Z_n)$ can be expressed as $P(Z_1, Z_2, \ldots, Z_n) = \prod_{i=1}^{n} P(Z_i)$. Consequently, the normalized information rate can be expressed as

$$-\frac{1}{n}\log P(Z_1, Z_2, \ldots, Z_n) = -\frac{1}{n}\sum_{i=1}^{n}\log P(Z_i). \tag{3.4}$$

By the Law of Large Numbers, the average $-\frac{1}{n}\sum_{i=1}^{n}\log P(Z_i)$ converges almost surely to the expectation $\mathbb{E}[-\log P(Z)]$, which is the entropy $H(Z)$:

$$-\frac{1}{n}\log P(Z_1, Z_2, \ldots, Z_n) \to H(Z), \text{ as } n \to \infty. \tag{3.5}$$

This result demonstrates that, for i.i.d. sources, the asymptotic behavior of the normalized information rate in (3.3) is deterministic and converges to a single value.

However, for general sources, we can not apply the Law of Large Numbers, in addition, the sequence $-\frac{1}{n}\log P(Z_1, Z_2, \ldots, Z_n)$ does not necessarily converge in the classical sense. Non-stationarity, non-ergodicity, or dependencies within the source can lead to significant fluctuations, making it challenging to describe the asymptotic behavior of the sequence using a single deterministic value.

To address this, we use information-spectrum terms which characterize the probabilistic bounds of the sequence. These terms rely on the notions of lim sup and lim inf in probability. Specifically, we define the spectral inf-entropy rate and spectral sup-entropy rate as

$$\underline{H}(\mathbf{X}) = \mathrm{p} - \liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{P(Z_1, Z_2, \ldots, Z_n)} \tag{3.6}$$

$$\overline{H}(\mathbf{X}) = \mathrm{p} - \limsup_{n\to\infty}\frac{1}{n}\log\frac{1}{P(Z_1, Z_2, \ldots, Z_n)}, \tag{3.7}$$

where the lim sup and lim inf in probability of a sequence $\{Z_n\}_{n=1}^{\infty}$ are, respectively, defined as

[42]

$$\mathrm{p}-\limsup_{n\to\infty} Z_n = \inf\left\{\alpha \mid \lim_{n\to+\infty} \mathbb{P}\left(Z_n > \alpha\right) = 0\right\}, \tag{3.8}$$

$$\mathrm{p}-\liminf_{n\to\infty} Z_n = \sup\left\{\alpha \mid \lim_{n\to+\infty} \mathbb{P}\left(Z_n < \alpha\right) = 0\right\}. \tag{3.9}$$

Similarly, the spectral sup-mutual information $\overline{I}(\mathbf{X};\mathbf{U})$, the spectral inf-mutual information $\underline{I}(\mathbf{U};\mathbf{Y})$, the spectral inf-divergence rate $\underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right)$, and the spectral sup-divergence rate $\overline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right)$ are, respectively, defined as [42]

$$\bar{I}(\mathbf{X};\mathbf{U}) = \mathrm{p}-\limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{X}^n}\left(\mathbf{U}^n \mid \mathbf{X}^n\right)}{P_{\mathbf{U}^n}\left(\mathbf{U}^n\right)}, \tag{3.10}$$

$$\underline{I}(\mathbf{U};\mathbf{Y}) = \mathrm{p}-\liminf_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{Y}^n}\left(\mathbf{U}^n \mid \mathbf{Y}^n\right)}{P_{\mathbf{U}^n}\left(\mathbf{U}^n\right)}, \tag{3.11}$$

$$\underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right) = \mathrm{p}-\liminf_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n\mathbf{Y}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}, \tag{3.12}$$

$$\overline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right) = \mathrm{p}-\limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n\mathbf{Y}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\left(\mathbf{U}^n,\mathbf{Y}^n\right)}. \tag{3.13}$$

It is straightforward to show that, for i.i.d. sources $\mathbf{U}$ and $\mathbf{X}$, the spectral mutual information terms in (3.11) and (3.10) converge to the mutual information $I(U;X)$. Similarity, for i.i.d. sources $\mathbf{U}$ and $\mathbf{Y}$, the spectral divergence terms in (3.13) and (3.12) converge the Kullback-Leibler divergence $D\left(P_{UY}\|P_{\overline{UY}}\right)$. Next, we define the DHT problem for general sources.

3.3 DHT for general sources

In what follows, we consider that the joint distribution of the sequence pair $\{(\mathbf{X}^n,\mathbf{Y}^n)\}_{n=1}^{\infty}$ depends on the underlying hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ defined for a given $n\in\mathbb{N}$ as

$$\mathcal{H}_0 : (\mathbf{X}^n,\mathbf{Y}^n) \sim P_{\mathbf{X}^n\mathbf{Y}^n}, \tag{3.14}$$

$$\mathcal{H}_1 : (\mathbf{X}^n,\mathbf{Y}^n) \sim P_{\overline{\mathbf{X}}^n\overline{\mathbf{Y}}^n}. \tag{3.15}$$

where the marginal probability distributions $P_{\mathbf{X}^n}$ and $P_{\mathbf{Y}^n}$ do not depend on the hypothesis.

We consider the following usual coding scheme defined in the literature on DHT [19, 35].

Definition 3.1 *Given a rate $R\geq 0$, consider a sequence $(f^{(n)},g^{(n)})_{n\in\mathbb{N}}$ of encoding and decoding*

functions, defined for each blocklenght $n \in \mathbb{N}$, such that

$$f^{(n)} : \mathcal{X}^n \longrightarrow \mathcal{M}_n = [\![1, M_n]\!], \tag{3.16}$$

$$g^{(n)} : \mathcal{M}_n \times \mathcal{Y}^n \longrightarrow \mathcal{H} = \{\mathcal{H}_0, \mathcal{H}_1\}, \tag{3.17}$$

such that $\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R$*, where* R *is the rate and* M_n *is the cardinality of the alphabet set* $\mathcal{M}_n$*.*

Definition 3.2 *The Type-I and Type-II error probabilities* α_n *and* β_n *are defined for each* $n \in \mathbb{N}$ *as*

$$\alpha_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^n\right), \mathbf{Y}^n\right) = \mathcal{H}_1 \mid \mathcal{H}_0 \text{ is true}\right], \tag{3.18}$$

$$\beta_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^n\right), \mathbf{Y}^n\right) = \mathcal{H}_0 \mid \mathcal{H}_1 \text{ is true}\right]. \tag{3.19}$$

Definition 3.3 *A Type-II error exponent* θ *is said to be achievable for a given* $R \geq 0$*, if for each* $\epsilon > 0$ *and for a large blocklength* n*, there exists a sequence* $\left(f^{(n)}, g^{(n)}\right)_{n\in\mathbb{N}}$ *of encoding and decoding functions such that the Type-I and Type-II error probabilities* α_n *and* β_n *satisfy*

$$\alpha_n \leq \epsilon, \tag{3.20}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta \tag{3.21}$$

for any $\epsilon > 0$*.*

In the following, we aim to determine the achievable Type-II error exponent θ for general sources.

3.4 Error exponent bound for general sources

Theorem 3.1 *The following lower bound on the error exponent* θ *is achievable for general sources*:

$$\theta \geq \sup_{P_{\mathbf{U}|\mathbf{X}}} \min\left\{R - \left(\overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{U};\mathbf{Y})\right), \underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right) + \left(\underline{I}(\mathbf{X};\mathbf{U}) - \overline{I}(\mathbf{X};\mathbf{U})\right)\right\}, \tag{3.22}$$

where $\mathbf{U}$ *is an auxiliary random variable with same conditional distribution* $P_{\mathbf{U}^n|\mathbf{X}^n} = P_{\overline{\mathbf{U}}^n|\overline{\mathbf{X}}^n}$ *under* $\mathcal{H}_0$ *and* $\mathcal{H}_1$ *and such that the Markov chain* $\mathbf{U}^n \to \mathbf{X}^n \to \mathbf{Y}^n$ *is satisfied under both* $\mathcal{H}_0$ *and* $\mathcal{H}_1$ *for all* n*. In addition,* $P_{\mathbf{UY}}$*, and* $P_{\overline{\mathbf{UY}}}$ *are the collection of the joint distributions of* $(\mathbf{U}^n, \mathbf{Y}^n)$ *under* H_0 *and* H_1 *respectively. Moreover, the* sup *is taken over all the conditional distributions* $P_{\mathbf{U}|\mathbf{X}}$ *such that the rate constraint* $R \geq \underline{I}(\mathbf{U};\mathbf{X} \mid \mathbf{Y})$ *is satisfied.*

The proof of Theorem 3.1 is provided in section 3.6.

As expected, we find that our error exponent is consistent with that provided in [2] for the i.i.d. case. The error exponent (3.22) is the result of a trade-off between the binning error (the first term in the right-hand side of (3.22)) and the decision error (the remaining term in the right-hand side of (3.22)), as in the i.i.d. case [23, 2]. This tradeoff does not appear in the hypothesis testing problem without coding for general sources of [57]. In addition, the decision error, *e.g.*, the second term in (3.22), not only contains a divergence term that appears in [2] and related works but also an additional term that is the difference $\underline{I}(\mathbf{X};\mathbf{U}) - \overline{I}(\mathbf{X};\mathbf{U})$ between the spectral inf-mutual information and the spectral sup-mutual information of $\mathbf{X}$ and $\mathbf{U}$. Especially, if the term $\frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{U}^n|\mathbf{X}^n)}{P_{\mathbf{U}^n}(\mathbf{U}^n)}$ does not converge in probability, then the two mutual information terms differ, inducing a penalty in the error exponent. For stationary and ergodic sources, this term converges and there is no such penalty.

3.5 Discussion

Our general error exponent bound in (3.22) is derived using a coding scheme similar to the one presented in [2], with several adaptations to accommodate general source models. As a result, we obtain a bound comparable to that of [2]. The primary advantage of our bound is that it is simpler to evaluate than the bound of [22]. Indeed, it avoids the minimization over the joint distribution required in the bound of Shimokawa et al. [22], shown in (2.16). However, it may not be as tight as the bound by Shimokawa et al., which, although more difficult to evaluate, is more precise. Consequently, building upon the Shimokawa et al. scheme for general sources is currently under investigation.

3.6 Proof of Theorem 3.1

We first restate the following lemma from [58], which will be useful in the derivation of Type-I error probability.

Lemma 3.1 ([58]) *Let $\mathbf{Z}^n, \mathbf{X}^n, \mathbf{U}^n$, be random sequences which take values in finite sets $\mathcal{Z}^n, \mathcal{X}^n, \mathcal{U}^n$, respectively, and satisfy the Markov condition $\mathbf{U}^n \to \mathbf{X}^n \to \mathbf{Z}^n$. Let $\{\Psi_n\}_{n=1}^{\infty}$ be a sequence of mappings such that $\Psi_n : \mathcal{Z}^n \times \mathcal{U}^n \to \{0,1\}$, and*

$$\lim_{n\to\infty} \mathbb{P}\left(\Psi_n(\mathbf{Z}^n, \mathbf{U}^n) = 1\right) = 0. \tag{3.23}$$

Then, $\forall \varepsilon > 0$, there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of mappings $f_n : \mathcal{X}^n \to \{\mathbf{u}_i^n\}_{i=1}^{M} \subset \mathcal{U}^n$ such that $M = \lceil e^{n(\overline{I}(\mathbf{U};\mathbf{X})+\varepsilon)} \rceil$ and

$$\lim_{n\to\infty} \mathbb{P}\left(\Psi_n(\mathbf{Z}^n, f_n(\mathbf{X}^n)) = 1\right) = 0. \tag{3.24}$$

3.6.1 Coding scheme

Random codebook generation: Generate $\mathrm{M}_1 = e^{n\bar{r}_0}$ sequences $\mathbf{u}_i^n$ randomly according to $P_{\mathbf{U}^n}$, where $P_{\mathbf{U}^n}$ is derived from a fixed distribution $P_{\mathbf{U}^n|\mathbf{X}^n}$. Assign randomly each $\mathbf{u}_i^n$ to one of $\mathrm{M}_2 = e^{nR}$ bins according to a uniform distribution over $[\![1, \mathrm{M}_2]\!]$. Let $\boldsymbol{B}(\mathbf{u}_i^n) \in [\![1, \mathrm{M}_2]\!]$ denote the index of the bin to which $\mathbf{u}_i^n$ belongs to.

Encoder: Given the sequence $\mathbf{x}^n$, the encoder uses a pre-defined mapping $f_n : \mathcal{X}^n \to \{\mathbf{u}_i^n\}_{i=1}^{\mathrm{M}_1}$ to output a certain sequence $\mathbf{u}_i^n = f_n(\mathbf{x}^n)$ and checks if the condition $(\mathbf{x}^n, \mathbf{u}_\mathbf{i}^\mathbf{n}) \in T_n^{(1)}$ is satisfied, where

$$T_n^{(1)} = \left\{ (\mathbf{x}^n, \mathbf{u}^n) \text{ s.t. } \underline{r}_0 - \epsilon < \frac{1}{n} \log \frac{P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{u}^n \mid \mathbf{x}^n)}{P_{\mathbf{U}^n}(\mathbf{u}^n)} < \overline{r}_0 + \epsilon \right\} \tag{3.25}$$

where $\underline{r}_0, \overline{r}_0 \in \mathbb{R}$. If such a sequence is found, the encoder sends the bin index $\boldsymbol{B}(\mathbf{u}_i^n)$. Otherwise, it sends an error message.

Decoder: The decoder first looks for a sequence in the bin according to the joint distribution $P_{\mathbf{U}^n\mathbf{Y}^n}$ under $\mathcal{H}_0$. Given the received bin index and the side information $\mathbf{y}^n$, going over the sequences $\mathbf{u}^n$ in the bin one by one, the decoder checks whether $(\mathbf{y}^n, \mathbf{u}^n) \in T_n^{(2)}$ with

$$T_n^{(2)} = \left\{ (\mathbf{y}^n, \mathbf{u}^n) \text{ s.t. } \frac{1}{n} \log \frac{P_{\mathbf{U}^n|\mathbf{Y}^n}(\mathbf{u}^n \mid \mathbf{y}^n)}{P_{\mathbf{U}^n}(\mathbf{u}^n)} > r^{'} - \epsilon \right\}, \tag{3.26}$$

with $r' \in \mathbb{R}$. The decoder declares $\mathcal{H}_1$ if no such sequence is found in the bin or if it receives an error message from the encoder. Otherwise, it declares $\mathcal{H}_0$ if the sequence $\mathbf{u}^n$ extracted from the bin belongs to the acceptance region $\mathcal{A}_n$ defined as

$$\mathcal{A}_n = \left\{ (\mathbf{y}^n, \mathbf{u}^n) \text{ s.t. } \frac{1}{n} \log \frac{P_{\mathbf{U}^n\mathbf{Y}^n}(\mathbf{u}^n, \mathbf{y}^n)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}(\mathbf{u}^n, \mathbf{y}^n)} > S - \epsilon \right\}, \tag{3.27}$$

where $S \in \mathbb{R}$ is the decision threshold ; if otherwise, it declares $\mathcal{H}_1$. The sets $T_n^{(1)}$, $T_n^{(2)}$, and $\mathcal{A}_n$ can be seen as decision regions depending on threshold values $\overline{r}_0, \bar{r}_0, r'$ and S. It is important to note that these sets are need to be defined so that the constraint on the Type-I error is satisfied. Specifically, the parameters $\overline{r}_0, \bar{r}_0, r'$ and S will be chosen such that $\alpha_n \leq \epsilon$, for any $\epsilon > 0$.

3.6.2 Error probabilities analysis

Type-I error α_n: The error events with which the decoder declares $\mathcal{H}_1$ under $\mathcal{H}_0$ are as follows

$$E_{11} = \left\{ \nexists \mathbf{u}^n \text{ s.t. } (\mathbf{X}^n, \mathbf{u}^n) \in T_n^{(1)}, (\mathbf{Y}^n, \mathbf{u}^n) \in T_n^{(2)}, (\mathbf{Y}^n, \mathbf{u}^n) \in \mathcal{A}_n \right\}, \tag{3.28}$$

$$E_{12} = \left\{ \exists \mathbf{u}'^n \neq \mathbf{u}^n \text{ s.t. } \boldsymbol{B}(\mathbf{u}'^n) = \boldsymbol{B}(\mathbf{u}^n), (\mathbf{Y}^n, \mathbf{u}'^n) \in T_n^{(2)}, \text{ but } (\mathbf{Y}^n, \mathbf{u}'^n) \notin \mathcal{A}_n \right\}. \tag{3.29}$$

The first event E_{11} is when there is an error either in the encoding, during debinning, or when taking the decision. The second event E_{12} corresponds to a debinning error, where a wrong sequence is extracted from the bin. By the union-bound, the Type-I error probability α_n can be upper bounded as

$$\alpha_n \leq \mathbb{P}\left(E_{11}\right) + \mathbb{P}\left(E_{12}\right). \tag{3.30}$$

Regarding the first error event, for $\underline{r}_0 = \underline{I}(\mathbf{X};\mathbf{U})$, $\overline{r}_0 = \overline{I}(\mathbf{X};\mathbf{U})$, and from the definitions of $\underline{I}(\mathbf{X};\mathbf{U})$ and $\overline{I}(\mathbf{X};\mathbf{U})$ in (3.10) and (3.11), we have

$$\lim_{n\to\infty} \mathbb{P}\left((\mathbf{X}^n, \mathbf{U}^n) \notin T_n^{(1)}\right) = 0.$$

In addition, according to the definition of $\underline{I}(\mathbf{Y};\mathbf{U})$ in (3.11), and setting $r^{'} = \underline{I}(\mathbf{Y};\mathbf{U})$, we also have

$$\lim_{n\to\infty} \mathbb{P}\left((\mathbf{Y}^n, \mathbf{U}^n) \notin T_n^{(2)}\right) = 0. \tag{3.31}$$

Finally, when $S = \underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$ and from the definition of $\underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$, we have

$$\lim_{n\to\infty} \mathbb{P}\left((\mathbf{Y}^n, \mathbf{U}^n) \notin \mathcal{A}_n\right) = 0.$$

Thus, by defining

$$\Psi_n(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n) = \tag{3.32}$$
$$\begin{cases} 0, & \text{if } (\mathbf{x}^n, \mathbf{u}^n) \in T_n^{(1)}, (\mathbf{y}^n, \mathbf{u}^n) \in T_n^{(2)} \text{ and} \\ & (\mathbf{y}^n, \mathbf{u}^n) \in \mathcal{A}_n, \\ 1, & \text{otherwise.} \end{cases}$$

we get that $\mathbb{P}(\Psi_n(\mathbf{X}^n, \mathbf{Y}^n, \mathbf{U}^n) = 1) \to 0$ as $n \to \infty$. Then, given that $\mathbf{U}^n \to \mathbf{X}^n \to \mathbf{Y}^n$ forms a Markov chain, applying Lemma 3.1 allows to show that there exists a sequence of functions f_n such that $\mathbb{P}(E_{11}) \to 0$ as $n \to \infty$.

Then, the error probability $\mathbb{P}(E_{12})$ can be expressed as

$$\begin{aligned}\mathbb{P}(E_{12}) &\leq \sum_{\mathbf{y}^n} P_{\mathbf{Y}^n}(\mathbf{y}^n) \sum_{\substack{\mathbf{u}'^n:\mathbf{u}'^n\neq\mathbf{u}^n \\ (\mathbf{y}^n,\mathbf{u}'^n)\in\mathcal{T}_n^{(2)}\cap\overline{\mathcal{A}_n}}} \mathbb{P}\left(\boldsymbol{B}(\mathbf{u}'^n)=\boldsymbol{B}(\mathbf{u}^n)\right) \\ &\leq \sum_{\mathbf{y}^n} P_{\mathbf{Y}^n}(\mathbf{y}^n) \sum_{\substack{\mathbf{u}'^n:\mathbf{u}'^n\neq\mathbf{u}^n \\ (\mathbf{y}^n,\mathbf{u}'^n)\in\mathcal{T}_n^{(2)}}} e^{-nR} \end{aligned} \tag{3.33}$$

From (3.26), for $(\mathbf{y}^n,\mathbf{u}'^n)\in T_n^{(2)}$ we get

$$P_{\mathbf{Y}^n}(\mathbf{y}^n) < P_{\mathbf{Y}^n|\mathbf{U}^n}(\mathbf{y}^n \mid \mathbf{u}'^n)\, e^{-n(r'-\epsilon)},$$

which allows us to write

$$\begin{aligned}\mathbb{P}(E_{12}) &\leq \sum_{\mathbf{u}'^n} \sum_{\mathbf{y}^n:(\mathbf{y}^n,\mathbf{u}'^n)\in\mathcal{T}_n^{(2)}} P_{\mathbf{Y}^n|\mathbf{U}^n}(\mathbf{y}^n \mid \mathbf{u}'^n)\, e^{-n(R+r'-\epsilon)} \\ &\leq e^{-n(R+r'-\overline{r}_0-\epsilon)} \end{aligned} \tag{3.34}$$

where $e^{n\overline{r}_0}$ is the number of sequences $\mathbf{u}^n$ in the codebook. Therefore, from the condition $R\geq \overline{r}_0 - r' + \epsilon = \overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{Y};\mathbf{U}) + \epsilon$, we get that $\mathbb{P}(E_{21})\to 0$ as $n\to\infty$.

Type-II error β_n: A Type-II error occurs when the decoder declares $\mathcal{H}_0$ although $\mathcal{H}_1$ is the true hypothesis. The corresponding error events are:

$$\begin{aligned}E_{21} &= \left\{\exists\tilde{\mathbf{u}}^n\neq\mathbf{u}^n : \boldsymbol{B}(\tilde{\mathbf{u}}^n)=\boldsymbol{B}(\mathbf{u}^n), \left(\overline{\mathbf{Y}}^n,\tilde{\mathbf{u}}^n\right)\in T_n^{(2)}, \text{ and } \left(\overline{\mathbf{Y}}^n,\tilde{\mathbf{u}}^n\right)\in\mathcal{A}_n\right\}, \\ E_{22} &= \left\{(\overline{\mathbf{Y}}^n,\mathbf{u}^n)\in T_n^{(2)}, (\overline{\mathbf{Y}}^n,\mathbf{u}^n)\in\mathcal{A}_n\right\}. \end{aligned} \tag{3.35}$$

The first event E_{21} is a debinning error and the second event E_{22} is the testing error. By the union bound, we get

$$\beta_n \leq \mathbb{P}(E_{21}) + \mathbb{P}(E_{22}). \tag{3.36}$$

Since the marginal probability distribution $P_{\mathbf{Y}^n}$ does not depend on the hypothesis, the probability $\mathbb{P}(E_{21})$ can be expressed by following the same steps as for $\mathbb{P}(E_{12})$. Given that $\overline{r}_0 = \overline{I}(\mathbf{X};\mathbf{U})$ and $r^{'} = \underline{I}(\mathbf{Y};\mathbf{U})$, we get

$$\mathbb{P}(E_{21}) \leq e^{-n\left(R-\left(\overline{I}(\mathbf{X};\mathbf{U})-\underline{I}(\mathbf{Y};\mathbf{U})\right)-\varepsilon\right)}. \tag{3.37}$$

Next, the probability $\mathbb{P}(E_{22})$ can be expressed as

$$\begin{aligned}\mathbb{P}(E_{22}) &\leq \sum_{(\mathbf{x}^n,\mathbf{y}^n)} P_{\overline{\mathbf{X}}^n\overline{\mathbf{Y}}^n}(\mathbf{x}^n,\mathbf{y}^n) \sum_{\substack{\mathbf{u}^n\in[\![1,\mathrm{M}_1]\!],\\ (\mathbf{x}^n,\mathbf{u}^n)\in\mathcal{T}_n^{(1)}}} \mathbb{P}\left((\mathbf{y}^n,\mathbf{u}^n)\in\mathcal{A}_n\right)\\ &\leq e^{n\overline{r}_0} \sum_{(\mathbf{x}^n,\mathbf{y}^n)} P_{\overline{\mathbf{X}}^n\overline{\mathbf{Y}}^n}(\mathbf{x}^n,\mathbf{y}^n) \sum_{\substack{\mathbf{u}^n:\\ (\mathbf{x}^n,\mathbf{u}^n)\in\mathcal{T}_n^{(1)}\\ (\mathbf{y}^n,\mathbf{u}^n)\in\mathcal{A}_n}} P_{\mathbf{U}^n}(\mathbf{u}^n)\end{aligned}$$

Since $(\mathbf{x}^n,\mathbf{u}^n)\in T_n^{(1)}$,

$$P_{\mathbf{U}^n}(\mathbf{u}^n) < P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{u}^n \mid \mathbf{x}^n) e^{-n(\underline{r}_0-\epsilon)}.$$

In addition, the conditional distributions $P_{\mathbf{U}^n|\mathbf{X}^n}$ and $P_{\overline{\mathbf{U}}^n|\overline{\mathbf{X}}^n}$ are the same, and the Markov chain $\mathbf{U}^n\to\mathbf{X}^n\to\mathbf{Y}^n$ is satisfied. Thus, $P_{\mathbf{U}^n|\mathbf{X}^n} = P_{\overline{\mathbf{U}}^n|\overline{\mathbf{X}}^n,\overline{Y}^n}$, and

$$\mathbb{P}(E_{22}) \leq e^{n(\overline{r}_0-\underline{r}_0+\epsilon)} \sum_{\mathbf{u}^n:(\mathbf{y}^n,\mathbf{u}^n)\in\mathcal{A}_n} P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}(\mathbf{u}^n,\mathbf{y}^n). \tag{3.38}$$

For $(\mathbf{y}^n,\mathbf{u}^n)\in\mathcal{A}_n$, we have

$$P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}(\mathbf{u}^n,\mathbf{y}^n) < P_{\mathbf{U}^n\mathbf{Y}^n}(\mathbf{u}^n,\mathbf{y}^n)\, e^{-n(S-\epsilon)}. \tag{3.39}$$

Combining this with (3.38) gives that

$$\mathbb{P}(E_{22}) \leq e^{-n(\underline{r}_0-\overline{r_0}+S-2\epsilon)} \tag{3.40}$$

Now, substituting (3.37) and (3.40) into (3.36), with $S=\underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right)$, the Type-II error is upper-bounded as

$$\beta_n \leq e^{-n\left(R-\left(\overline{I}(\mathbf{X};\mathbf{U})-\underline{I}(\mathbf{Y};\mathbf{U})\right)-\epsilon\right)} + e^{-n\left(\underline{I}(\mathbf{X};\mathbf{U})-\overline{I}(\mathbf{X};\mathbf{U})+\underline{D}\left(P_{\mathbf{UY}}\|P_{\overline{\mathbf{UY}}}\right)-2\epsilon\right)}.$$

Finally, from the definition of the error exponent θ given by (3.21), we show that (3.22) is achievable, which proves Theorem 3.1.

3.7 Summary and Discussion

In this chapter, we studied DHT for general source models. We presented an achievable coding scheme that provides a generic error exponent bound applicable to a wide range of sources, not limited to i.i.d. models. The achievability proof relies on information-spectrum methods [42]. We

demonstrated that, when applied to the specific case of i.i.d. sources, our general error exponent aligns with the established results in the i.i.d. case of [2].

However, the obtained bound may not be as tight as the well-known bound by Shimokawa et al. [22]. Furthermore, recent work by [25] demonstrated an enhancement of the achievable error exponent presented by Shimokawa et al. [22]. In future work, we aim to incorporate ideas from the achievable coding scheme of [22] as well as the improvements from [25] into our coding scheme, which may lead to a tighter general error exponent.

By considering general source models, our goal is to establish an error exponent bound that is broadly applicable to various source models. Therefore, in the next chapter, we will apply this general bound to specific source models, such as Gaussian stationary sources and the Gilbert-Elliot (GE) model. Additionally, we will introduce an efficient method to estimate the Type-II error exponent for the GE model.

Chapitre 4

Error exponent for stationary and ergodic Gaussian and Gilbert-Elliot sources models

4.1 Introduction

In the previous chapter, we have derived a generic error exponent bound for general sources, applicable to a wide range of source models. To show the consistency of our general analysis, and to progress toward the development of practical DHT coding schemes, it is crucial to investigate a broader range of source models of interest. In this chapter, we will apply our general error bound in (3.22) to two specific sources models: the stationary and ergodic Gaussian sources model, and the Gilbert-Elliot (GE) sources model. The stationary and ergodic Gaussian model is often used to model various real-world processes. For instance, in subspace techniques for array signal processing, it is common to assume that both signals and noise are modeled as stationary and ergodic Gaussian processes [59]. Additionally, i.i.d. and block-i.i.d. Gaussian models have been considered in the context of DHT, as explored in [60, 61, 62]. For stationary and ergodic Gaussian source models, not necessarily block-i.i.d., we derive closed-form expressions of the error exponent using the general bound provided in (3.22).

The GE model finds applications in many domains such as video coding [43], link quality estimation [44], or packet loss analysis [45]. The GE model has not been previously investigated in the context of DHT. Since closed-form expressions for the error exponents of the GE model do not exist, we propose a novel method to evaluate these error exponents, by formulating the terms involved in the error exponents as estimators. Subsequently, we provide an efficient numerical method to evaluate these estimators, using forward recursions proposed for Hidden Markov Models (HMM) in [63]. To provide valuable insights for the design of practical coding schemes for DHT, we present numerical results that explore: (i) the impact of different GE model parameters on the error-exponent, and (ii) the tradeoff between the binning error and the testing error.

4.2 Stationary and ergodic Gaussian sources

We first introduce the following notation which will be useful in this chapter. The conditional differential entropy of $\mathbf{X}$ given $\mathbf{Y}$ is denoted by $h(\mathbf{X} \mid \mathbf{Y})$. The covariance of a zero-mean random vector $\mathbf{X}$ is denoted by $\Sigma_{\mathbf{x}} = \mathbb{E}[\mathbf{X}\mathbf{X}^{\dagger}]$. The cross-correlation of two zero-mean vectors $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{\Sigma_{xy}} = \mathbb{E}\left[\mathbf{X}\mathbf{Y}^{\dagger}\right]$. The conditional correlation matrix of $\mathbf{X}$ given $\mathbf{Y}$ is denoted by $\mathbf{\Sigma_{x|y}} = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}])(\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}])^{\dagger}\right]$, which simplifies to $\mathbf{\Sigma_{x|y}} = \mathbf{\Sigma_x} - \mathbf{\Sigma_{xy}}\mathbf{\Sigma_y^{-1}}\mathbf{\Sigma_{yx}}$. The trace of a covariance matrix $\mathbf{\Sigma}$ is denoted by $\mathrm{tr}\{\mathbf{\Sigma}\}$. The limit in probability γ of a sequence of random variables $\{A_n\}_{n=1}^{+\infty}$ is denoted by $\gamma = \mathrm{p}-\lim_{n\to\infty} A_n$ and it verifies

$$\lim_{n\to\infty} P\left(|A_n - \gamma| > \epsilon\right) = 0 \tag{4.1}$$

for all $\epsilon > 0$.

4.2.1 Definitions

Here, we recall the definitions of two commonly used concepts: stationarity and ergodicity. These definitions are provided in [64]. Following that, we introduce some useful notations.

Definition 4.1 (Stationarity) *Let $\mathbf{Z}$ be a source generating a sequence of random variables $\{Z_n\}_{n=1}^{+\infty}$. The source $\mathbf{Z}$ is stationary if, for all $n, L \in \mathbb{N}$ and for all $(z'_1, \ldots, z'_n) \in \mathcal{Z}^n$,*

$$P\left(Z_1 = z'_1, \ldots, Z_n = z'_n\right) = P\left(Z_{1+L} = z'_1, \ldots, Z_{n+L} = z'_n\right).$$

According to this definition, a source is stationary if the probability of any given vector $(z'_1, \ldots, z'_n)$ does not depend on the position at which it is evaluated. In other words, this probability is independent of the time origin.

Definition 4.2 (Ergodicity) *Let $\mathbf{Z}$ be a source generating a sequence of random variables $\{Z_n\}_{n=-\infty}^{+\infty}$. Let f_n be a function of the sequence $\mathbf{z} = \{z_n\}_{n=-\infty}^{+\infty}$ that depends only on the n components $z_1, \ldots, z_n$. We denote by T^{ℓ} an operator such that $T^{\ell}\mathbf{z}$ represents the same sequence shifted by ℓ symbols. The source $\mathbf{Z}$ is said to be ergodic if for every integrable function f_n,*

$$\lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} f_n\left(T^{\ell}\mathbf{z}\right) = \mathbb{E}_{\mathbf{Z}}\left[f_n(\mathbf{Z})\right] \quad a.e.$$

(a.e.: almost everywhere).

According to this definition, a source is ergodic if its statistical characteristics (such as its mean) are independent of the specific realization, i.e., the particular sequence $\mathbf{z}$ being considered.

In the case where sources $\mathbf{X}$ and $\mathbf{Y}$ are stationary and ergodic Gaussian sources (not necessarily i.i.d.), the general error exponent bound in (3.22) can be simplified as follows.

Proposition 4.1 *If the sources* $\mathbf{X}$ *and* $\mathbf{Y}$ *are Gaussian, stationary and ergodic under both* $\mathcal{H}_0$ *and* $\mathcal{H}_1$*, the error exponent in* (3.22) *becomes*:

$$\theta \geq \sup_{P_{\mathbf{U}|\mathbf{X}}} \min \left\{ \lim_{n\to\infty} R - \left[\frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{Y}^n\right) - \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{X}^n\right) \right], \lim_{n\to\infty} \frac{1}{n} D\left(P_{\mathbf{U}^n\mathbf{Y}^n} \| P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right) \right\} \quad (4.2)$$

This proposition is due to the *strong converse property* under the same condition as in [42, Theorem 1.5.1].

4.2.2 Error exponent for stationary and ergodic Gaussian sources

Let $\mathbf{X}$ and $\mathbf{Y}$ be two stationary and ergodic sources distributed according to Gaussian distributions $\mathcal{N}\left(\mu_{\mathbf{X}}, \mathbf{K}_{\mathbf{X}}\right)$ and $\mathcal{N}\left(\mu_{\mathbf{Y}}, \mathbf{K}_{\mathbf{Y}}\right)$, with covariance matrices $\mathbf{K}_{\mathbf{X}}$ and $\mathbf{K}_{\mathbf{Y}}$, respectively. The two hypotheses are formulated as

$$\mathcal{H}_0 : \begin{pmatrix} \mathbf{X}^n \\ \mathbf{Y}^n \end{pmatrix} \sim \mathcal{N}(\mu_{\mathbf{XY}}, \mathbf{K}), \quad (4.3)$$

$$\mathcal{H}_1 : \begin{pmatrix} \mathbf{X}^n \\ \mathbf{Y}^n \end{pmatrix} \sim \mathcal{N}(\overline{\mu}_{\mathbf{XY}}, \overline{\mathbf{K}}). \quad (4.4)$$

In the expressions (4.3) and (4.4), $\mu_{\mathbf{XY}}$ is defined as a block vector $[\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}]^T$. In addition, $\mathbf{K}$ and $\overline{\mathbf{K}}$ are the joint covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ defined as

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{X}} & \mathbf{K}_{\mathbf{XY}} \\ \mathbf{K}_{\mathbf{YX}} & \mathbf{K}_{\mathbf{Y}} \end{bmatrix}, \overline{\mathbf{K}} = \begin{bmatrix} \mathbf{K}_{\mathbf{X}} & \overline{\mathbf{K}}_{\mathbf{XY}} \\ \overline{\mathbf{K}}_{\mathbf{YX}} & \mathbf{K}_{\mathbf{Y}} \end{bmatrix}. \quad (4.5)$$

Although not explicit in our notation, we here consider that the vectors $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$ are of length n, and that the covariance matrices $\mathbf{K}$ and $\overline{\mathbf{K}}$ are of size $2n \times 2n$, where n will tend to infinity in the subsequent analysis. We assume that all the matrices $\mathbf{K}_{\mathbf{X}}$, $\mathbf{K}_{\mathbf{Y}}$, $\overline{\mathbf{K}}_{\mathbf{Y}}$, $\mathbf{K}_{\mathbf{XY}}$, and $\overline{\mathbf{K}}_{\mathbf{XY}}$ are positive-definite. We also denote the conditional covariance matrix of $\mathbf{X}^n$ given $\mathbf{Y}^n$ by

$$\mathbf{K}_{\mathbf{X}|\mathbf{Y}} = \mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{XY}} \mathbf{K}_{\mathbf{Y}}^{-1} \mathbf{K}_{\mathbf{XY}}. \quad (4.6)$$

The eigenvalues of $\mathbf{K}_{\mathbf{X}|\mathbf{Y}}$ are further denoted by $\lambda_i^{(X|Y)}$.

For the hypothesis problem formulated in (4.3) and (4.4), the terms in (4.2) reduce to

$$\lim_{n\to\infty} \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{Y}^n\right) - \lim_{n\to\infty} \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{X}^n\right) = \lim_{n\to\infty} \frac{1}{2n} \sum_{i=1}^{n} \log \frac{\lambda_i^{(X|Y)} + \kappa}{\kappa}, \quad (4.7)$$

and

$$\lim_{n\to\infty}\frac{1}{n}D\left(P_{\mathbf{U}^n\mathbf{Y}^n}\|P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right\} = \lim_{n\to\infty}\frac{1}{2n}\left[\log\frac{\left|\overline{\mathbf{\Sigma}}\right|}{|\mathbf{\Sigma}|} - 2n \ + \left(\overline{\mu}_{\mathbf{UY}}-\mu_{\mathbf{UY}}\right)^T\overline{\mathbf{\Sigma}}^{-1}\left(\overline{\mu}_{\mathbf{UY}}-\mu_{\mathbf{UY}}\right) + \right.$$
$$\left. \operatorname{tr}\left\{\overline{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\right\}\right], \tag{4.8}$$

where $\mathbf{\Sigma}$ and $\overline{\mathbf{\Sigma}}$ are the joint covariance matrices of $\mathbf{U}$ and $\mathbf{Y}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

The terms given by (4.7) and (4.8) are obtained by considering that the source $\mathbf{U}$ is Gaussian such that $\mathbf{U}=\mathbf{X}+\mathbf{Z}$, where $\mathbf{Z}\sim\mathcal{N}(0,\kappa\mathbf{I}_n)$ is independent of $\mathbf{X}$, and $\mathbf{I}_n$ is the identity matrix of dimension $n\times n$. The covariance matrices $\mathbf{\Sigma}$ and $\overline{\mathbf{\Sigma}}$ are then defined as

$$\mathbf{\Sigma}=\begin{bmatrix}\mathbf{K_U} & \mathbf{K_{UY}}\\ \mathbf{K_{YU}} & \mathrm{K_Y}\end{bmatrix}, \overline{\mathbf{\Sigma}}=\begin{bmatrix}\mathbf{K_U} & \overline{\mathbf{K}}_{\mathbf{UY}}\\ \overline{\mathbf{K}}_{\mathbf{YU}} & \mathrm{K_Y}\end{bmatrix}. \tag{4.9}$$

We now consider the case where the pair $(\mathbf{U},\mathbf{Y})$ has different covariance matrices, $\mathbf{\Sigma}$ under H_0 and $\overline{\mathbf{\Sigma}}$ under H_1. We also assume that all the Gaussian vectors are zero-centered. We then define $\mathcal{H}_0$ and $\mathcal{H}_1$ as

$$\mathcal{H}_0 : \begin{pmatrix}\mathbf{X}^n\\ \mathbf{Y}^n\end{pmatrix}\sim\mathcal{N}(\mathbf{0},\mathbf{K}), \tag{4.10}$$

$$\mathcal{H}_1 : \begin{pmatrix}\mathbf{X}^n\\ \mathbf{Y}^n\end{pmatrix}\sim\mathcal{N}(\mathbf{0},\overline{\mathbf{K}}). \tag{4.11}$$

In this case, it can be shown that the expression (4.7) remains the same, while the expression (4.8) reduces to

$$\lim_{n\to\infty}\frac{1}{n}D\left(P_{\mathbf{U}^n\mathbf{Y}^n}\|P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right\} = \lim_{n\to\infty}\frac{1}{2n}\left[\log\frac{\left|\overline{\mathbf{\Sigma}}\right|}{|\mathbf{\Sigma}|} - 2n + \operatorname{tr}\left\{\overline{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\right\}\right]. \tag{4.12}$$

Note that the matrices $\mathbf{\Sigma}$ and $\overline{\mathbf{\Sigma}}$ are of length $2n\times 2n$, where n tends to infinity. Therefore, to specify the previous result to some specific Gaussian sources, one needs to study the convergence of the determinants $|\mathbf{\Sigma}|$ and $\left|\overline{\mathbf{\Sigma}}\right|$, and also of the trace $\operatorname{tr}\left\{\overline{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\right\}$. These closed-form expressions demonstrate that our general error exponent bound in (3.22) can be extended to a broader range of source models. Furthermore, in subspace techniques for array signal processing [59], a relevant question is on the accuracy of the estimated signal subspace dimension, which inherently leads to a hypothesis testing problem. In this context, our DHT framework for Gaussian sources could provide valuable insights.

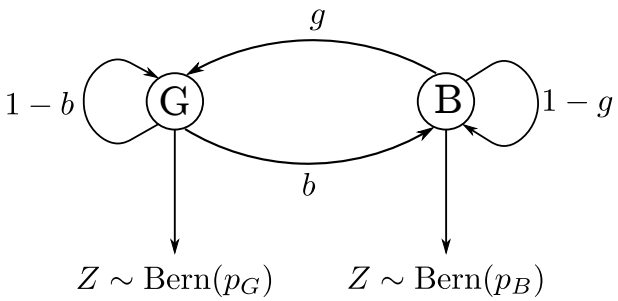

FIGURE 4.1 – Gilbert-Elliot model for the correlation noise Z under hypothesis $\mathcal{H}_0$. Under hypothesis $\mathcal{H}_1$, the crossover probabilities become $\overline{p}_G$ and $\overline{p}_B$, while the parameters g and b remain the same. Source : © 2023 IEEE. Reproduced with permission from [1].

4.3 Error exponent for Gilbert-Elliot (GE) sources model

The GE model is a non-i.i.d. time-varying binary model which involves a good state (G) and a bad state (B) [65], as illustrated in Figure 4.1. The transition between these two states is modeled by a Markov chain, thus accounting for memory in the successive state values. For instance, in sensor networks, state G (respectively state B) represents high correlation (respectively low correlation) between sensors measurements.

4.3.1 GE model definition

We consider two correlated sources, $\mathbf{X}$ and $\mathbf{Y}$, where $\mathbf{X}$ is the source to be encoded, and $\mathbf{Y}$ is the side information available at the decoder. The binary sequences generated by $\mathbf{X}$ and $\mathbf{Y}$ are denoted as $\{X_k\}_{k=1}^{+\infty}$ and $\{Y_k\}_{k=1}^{+\infty}$, respectively. We assume that the source $\mathbf{X}$ is i.i.d., such that for all $k \geq 1$, X_k follows a Bernoulli distribution Bern(p) with parameter p, constant across the hypotheses. On the other hand, the source $\mathbf{Y}$ is not i.i.d. and its probability distribution depends on the hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$, as described below:

$$\mathcal{H}_0 : Y_k = X_k \oplus Z_k \tag{4.13}$$

$$\mathcal{H}_1 : \overline{Y}_k = X_k \oplus \overline{Z}_k. \tag{4.14}$$

With a slight abuse of notation, we use $\overline{Y}_k$ to denote the side information symbols generated under $\mathcal{H}_1$, to make it clear that they have a different probability distribution than under $\mathcal{H}_0$. The sources $\mathbf{Z}$ and $\overline{\mathbf{Z}}$ which generate sequences of binary symbols $\{Z_k\}_{k=1}^{+\infty}$ and $\{\overline{Z}_k\}_{k=1}^{+\infty}$, respectively, are independent of $\mathbf{X}$ and follow GE models described below.

Under $\mathcal{H}_0$, the source $\mathbf{Z}$ follows a GE model [65] with hidden state $\mathbf{S}$ depicted in Figure 4.1. The sequence output from the binary hidden states $\{S_k\}_{k=1}^{+\infty}$ is such that $S_k \in \{G, B\}$, and it

follows a Markov model described with the following state transition probabilities:

$$P\left(S_k = G \mid S_{k-1} = B\right) = g, \tag{4.15}$$
$$P\left(S_k = B \mid S_{k-1} = G\right) = b. \tag{4.16}$$

Due to the Markov property,

$$P(S_k|S_{k-1}, S_{k-2} \cdots S_1) = P(S_k|S_{k-1}). \tag{4.17}$$

Each symbol Z_k takes value 0 or 1 depending on the hidden state value $S_k = s$ such that

$$P\left(Z_k = 1 \mid S_k = G\right) = p_G, \tag{4.18}$$
$$P\left(Z_k = 1 \mid S_k = B\right) = p_B, \tag{4.19}$$

where p_G and p_B are crossover probabilities. We often consider that $p_G < p_B$, so that the state G corresponds to high correlation between $\mathbf{X}$ and $\mathbf{Y}$ while the state B corresponds to low correlation. In addition, the GE model assumes that

$$P(Z_k|Z_1, \cdots Z_n, S_1, \cdots S_n) = P(Z_k|S_k). \tag{4.20}$$

Under hypothesis $\mathcal{H}_1$, the source $\overline{\mathbf{Z}}_k$ also follows a GE model, with the same hidden state $\mathbf{S}$ described by equations (4.15) and (4.16) and same values g and b as under $\mathcal{H}_0$. However, the crossover probabilities differ from those specified by (4.18) and (4.19). They are now denoted by $\overline{p}_G$ and $\overline{p}_B$ with

$$P\left(\overline{Z}_k = 1 \mid S_k = G\right) = \overline{p}_G, \tag{4.21}$$
$$P\left(\overline{Z}_k = 1 \mid S_k = B\right) = \overline{p}_B. \tag{4.22}$$

4.3.2 Information-spectrum terms for ergodic GE models

The general error exponent bound given in (3.22) relies on information-spectrum terms [42], which are defined from limits inferior and limits superior in probability. These information spectrum terms have simplified expressions for the GE model, given that the underlying Markov chain is ergodic, where ergodic means that it admits a unique stationary distribution as the sequence length n tends to infinity [66]. The GE model described in Section 4.3.1 is ergodic given that the conditions $0 < g < 1$ and $0 < b < 1$ are satisfied [66].

For ergodic GE models for some sources $\mathbf{U}, \mathbf{Y}, \overline{\mathbf{U}}, \overline{\mathbf{Y}}$, combining the information-spectrum definitions from [42] with the convergence proofs from [66] allows us to show that the spectral

conditional entropy has expression

$$H_s(\mathbf{U}|\mathbf{Y}) = p-\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P(\mathbf{U}^n|\mathbf{Y}^n)}, \tag{4.23}$$

the conditional spectral mutual information between $\mathbf{U}$ and $\mathbf{X}$, conditioned on $\mathbf{Y}$ has expression

$$I(\mathbf{U}^n; \mathbf{X}^n \mid \mathbf{Y}^n) = p-\lim_{n\to\infty} \frac{1}{n} \log \frac{P_{\mathbf{U}^n,\mathbf{X}^n|\mathbf{Y}^n}(\mathbf{U}^n, \mathbf{X}^n \mid \mathbf{Y}^n)}{P_{\mathbf{U}^n|\mathbf{Y}^n}(\mathbf{U}^n \mid \mathbf{Y}^n) P_{\mathbf{X}^n|\mathbf{Y}^n}(\mathbf{X}^n \mid \mathbf{Y}^n)}, \tag{4.24}$$

and the spectral divergence has expression

$$D_s(P_{\mathbf{U},\mathbf{Y}}||P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}}) = p-\lim_{n\to\infty} \frac{1}{n} \log \frac{P(\mathbf{U}^n, \mathbf{Y}^n)}{P(\overline{\mathbf{U}}^n, \overline{\mathbf{Y}}^n)}. \tag{4.25}$$

4.3.3 Error exponent for the GE model

We now specify the general error exponent bound in (3.22) to ergodic GE models.

Proposition 4.2 *If the sources* $\mathbf{X}$ *and* $\mathbf{Y}$ *are correlated given the GE model under both* $\mathcal{H}_0$ *and* $\mathcal{H}_1$*, the general error exponent in* (3.22) *reduces to*

$$\theta \geq \sup_{P_{\mathbf{U}|\mathbf{X}}} \min \left\{\theta_{bin}, \theta_{test}\right\}, \tag{4.26}$$

where

$$\theta_{bin} = R - \left[H_s\left(\mathbf{U} \mid \mathbf{Y}\right) - H_s\left(\mathbf{U} \mid \mathbf{X}\right)\right] \tag{4.27}$$

$$\theta_{test} = D_s\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right) \tag{4.28}$$

and the sup *is taken over all the conditional distributions* $P_{\mathbf{U}|\mathbf{X}}$ *such that the rate constraint* $R \geq I_s(\mathbf{U}; \mathbf{X} \mid \mathbf{Y})$ *is satisfied.*

The error exponent in (4.26) arises from a tradeoff between the binning error, denoted by θ_{bin} in (4.27), and the testing error, represented by θ_{test} in (4.28). This tradeoff is achieved through a quantize-binning scheme, which is used to derive the general bound in (3.22).

The terms θ_{bin} and θ_{test} depend on an auxiliary source $\mathbf{U}$ which has to satisfy the Markov chain $\mathbf{U} \to \mathbf{X} \to \mathbf{Y}$ [67]. In what follows, we assume that this auxiliary source generates a sequence of symbols $\{U_k\}_{k=1}^{+\infty}$ such that

$$U_k = X_k \oplus \phi_k, \tag{4.29}$$

where $\phi_k \sim \text{Bern}\left(\delta\right)$, and X_k and ϕ_k are independent. The parameter δ is key as it addresses the tradeoff between the binning error and the testing error. Indeed, a small value of δ means that

the sequence $\mathbf{u}^n$ selected by the encoder will be close to $\mathbf{x}^n$, which reduces the testing error. But this increases the binning error at the same time, as it leads to a larger number of sequences in each bin in order to meet the rate constraint for a given R.

We choose definition (4.29) for the auxiliary source $\mathbf{U}$ because due to the Markov chain, $\mathbf{U}$ needs to be expressed from $\mathbf{X}$ which is itself Bernoulli and i.i.d. From this choice of $\mathbf{U}$, the term $H_s(\mathbf{U}|\mathbf{X})$ in (4.27) can be expressed as

$$H_s(\mathbf{U}|\mathbf{X}) = -\delta \log \delta - (1-\delta)\log(1-\delta), \tag{4.30}$$

which is the conventional entropy of a Bernoulli source. We leave to future work the investigation of if (4.29) is the optimal choice for the auxiliary source $\mathbf{U}$. However, since there are no known analytical expressions for the terms $H_s\left(\mathbf{U} \mid \mathbf{Y}\right)$ in (4.27) and $D_s\left(P_{\mathbf{UY}}|P_{\overline{\mathbf{UY}}}\right)$ in (4.28), we now describe our numerical evaluation procedure for these terms.

4.3.4 Statistical evaluation of the error exponent for the GE model

This section provides a statistical method to evaluate the error exponent for the GE model described in Section 4.3.1.

4.3.4.1 Estimators of spectral information-theory terms

Given that the spectral conditional entropy in (4.23) and the spectral divergence in (4.25), are both defined from a limit in probability, the terms

$$\widehat{H}_s\left(\mathbf{U} \mid \mathbf{Y}\right) = \frac{1}{n} \log \frac{1}{P(\mathbf{u}^n|\mathbf{y}^n)}, \tag{4.31}$$

$$\widehat{D}_s(\mathbf{U},\mathbf{Y}||P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}}) = \frac{1}{n} \log \frac{P(\mathbf{u}^n,\mathbf{y}^n)}{P(\overline{\mathbf{u}}^n,\overline{\mathbf{y}}^n)}, \tag{4.32}$$

are consistent estimators [68, Section 1.8] of $H_s\left(\mathbf{U} \mid \mathbf{Y}\right)$ and $D_s(P_{\mathbf{U},\mathbf{Y}}||P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}})$, respectively. To evaluate these estimators, we propose to use a large value of n, and to randomly generate samples $(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n)$ according to the GE model described in section 4.3.1, and to the definition of $\mathbf{U}$ in (5.8). We then calculate the probability terms $P(\mathbf{u}^n|\mathbf{y}^n)$, $P(\mathbf{u}^n)$, $P(\mathbf{u}^n,\mathbf{y}^n)$, and, $P(\overline{\mathbf{u}}^n,\overline{\mathbf{y}}^n)$ involved in (4.31) or (4.32), which allows evaluating the previous estimators, and then the error exponent (4.26).

A similar methodology was employed in [69] to numerically evaluate the capacity of a Gilbert-Elliot channel, from an estimator defined as the log-probability of certain random vectors. In [69], the estimator was defined from the notion of information-rate, while here, we rely on the definition of information spectrum terms. Moreover, since the capacity of a channel only involves computing the mutual information between the channel input and output, the probability computation

in [69] is simpler. In this section, we evaluate the probability terms involved in (4.31) and (4.32) from forward recursions which we now describe.

4.3.4.2 Forward recursions

In this section, we describe the computation of all the probabilities involved in the estimators (4.31) and (4.32). We first calculate the joint probability $p(\mathbf{u}^n, \mathbf{y}^n)$ by evaluating for $s \in \{G, B\}$, $\alpha_n^{(u,y)}(s) = P(\mathbf{u}^n, \mathbf{y}^n, S_n = s)$. This probability can be computed efficiently from the HMM forward recursion described in [63]. This recursion is initialized for $s \in \{G, B\}$ as

$$\alpha_1^{(u,y)}(s) = P\left(S_1 = s\right) P\left(y_1 \mid S_1 = s\right) P\left(u_1 \mid y_1, S_1 = s\right), \tag{4.33}$$

and then defined for all $k \in [\![2, n-1]\!]$ as

$$\begin{aligned}\alpha_{k+1}^{(u,y)}(s) = \left[\sum_{s' \in \{G,B\}} \alpha_k^{(u,y)}(s') P\left(S_{k+1} = s \mid S_k = s'\right)\right] \ldots \\ P\left(y_{k+1} \mid S_{k+1} = s\right) P\left(u_{k+1} \mid y_{k+1}, S_{k+1} = s\right).\end{aligned} \tag{4.34}$$

We them compute

$$P(\mathbf{u}^n, \mathbf{y}^n) = \sum_{s \in \{G,B\}} \alpha_n^{(u,y)}(s). \tag{4.35}$$

The marginal probability $P(\mathbf{y}^n)$ can also be evaluated from a forward recursion initialized as

$$\alpha_1^{(y)}(s) = P\left(S_1 = s\right) P\left(y_1 \mid S_1 = s\right), \tag{4.36}$$

and defined for all $k \in [\![2, n-1]\!]$ as

$$\alpha_{k+1}^{(y)}(s) = \left[\sum_{s' \in \{G,B\}} \alpha_k^{(y)}(s') P\left(S_{k+1} = s \mid S_k = s'\right)\right] P\left(y_{k+1} \mid S_{k+1} = s\right). \tag{4.37}$$

Then,

$$P(\mathbf{y}^n) = \sum_{s \in \{G,B\}} \alpha_n^{(y)}(s). \tag{4.38}$$

This allows us to calculate the conditional probability $P\left(\mathbf{u}^n \mid \mathbf{y}^n\right)$ using the formula

$$P\left(\mathbf{u}^n \mid \mathbf{y}^n\right) = \frac{P\left(\mathbf{u}^n, \mathbf{y}^n\right)}{P\left(\mathbf{y}^n\right)} = \frac{\sum_{s \in \{G,B\}} \alpha_n^{(u,y)}(s)}{\sum_{s \in \{G,B\}} \alpha_n^{(y)}(s)}. \tag{4.39}$$

We apply a similar recursion to calculate the joint probability $P\left(\bar{\mathbf{u}}^n \mid \bar{\mathbf{y}}^n\right)$ which appears in (4.32). To avoid numerical issues, we compute all the previous terms in log.

4.3.5 Numerical results

In this section, we aim to numerically evaluate the DHT error exponents with the GE model, by applying the method described in section 4.3.4. We first investigate the convergence of the proposed estimators with respect to the sequence length n. Figure 4.2 shows the estimated spectral information terms as functions of n, for a set of parameters given in the caption of the figure. We consider a maximum value $n = 15000$, and represent two types of curves: one obtained from a single set of source vectors $(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n)$, and the other obtained by averaging over $K = 15$ realizations of the source vectors. It is clear that averaging improves the estimation quality, especially for the testing error.

Next, we consider a large value of $n = 70000$, and investigate the effect of the model parameters onto the error exponent. First of all, Figure 4.3 shows the two error exponents θ_{bin} and θ_{test} as functions of p_G with p_B values as a parameter, where p_G and p_B are the crossover probabilities under the hypothesis $\mathcal{H}_0$. All other parameters are fixed, and indicated in the caption of the figure. We see that both error exponents decrease as p_G and p_B increase, due to the fact that increasing these parameters makes the source closer to uniform. This, in turn, causes $\mathcal{H}_0$ and $\mathcal{H}_1$ to become more similar, making it challenging for the decoder to accurately distinguish between them. In addition, Figure 4.4 shows the two error exponents with respect to a parameter $\mu = 1 - b - g$. While the binning error only slightly varies with μ, the testing error varies with large values of μ, especially when b is fixed.

Finally, we aim to investigate the tradeoff between the testing error and the binning error. Figure 4.5 shows the error exponents θ_{bin} and θ_{test} as function of δ, which is the parameter that defines the auxiliary source $\mathbf{U}$. The other parameters are fixed and indicated in the figure. It is found that for a small value range of δ, the binning error is smaller than the testing error, and therefore is a dominating factor of the Type-II error. However, when δ increases, the testing error tends to become the dominant error event. Finally, as pointed out in [35] for i.i.d. binary sources, the best tradeoff between the two error events is provided by the value δ at which the two curves intersect, *e.g*, $\delta \approx 0.25$ for this set of parameters.

4.4 Summary and Discussion

In this chapter, we specified our general error exponent bound in (3.22) to the stationary and ergodic Gaussian sources model and GE sources model. For the Gaussian sources, we derived closed-form expressions for the error exponent. For the GE model, we introduced an efficient method to estimate the error exponent, which utilizes the forward recursion of HMMs. Our numerical results have evaluated the effects of the model parameters onto the error exponent, as well as the tradeoff between the testing error and the binning error. These insights provide valuable guidance for the design of practical DHT coding schemes for binary sources, which will

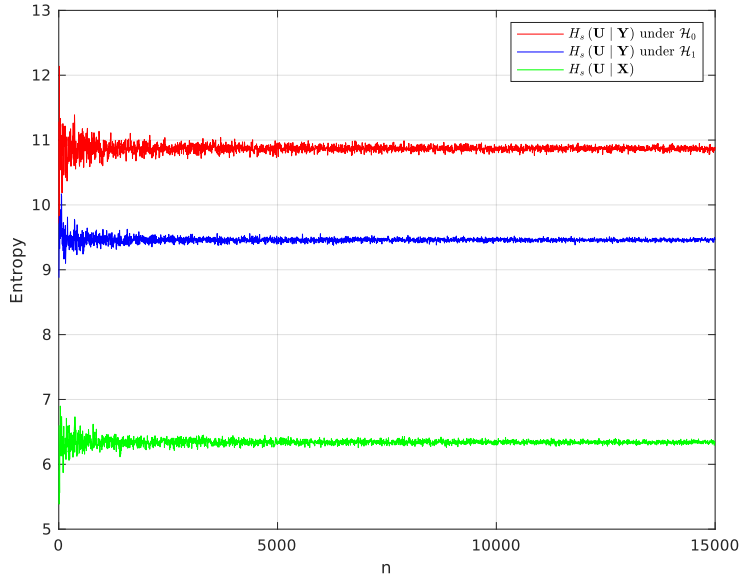

FIGURE 4.2 – Estimated spectral information terms as functions of n for $p_G = 0.05$, $p_B = 0.03$, $\bar{p}_G = 0.3$, $\bar{p}_B = 0.5$, $g = 0.001$, $b = 0.002$, $p = 0.2$, $\delta = 0.15$, $R = 0.4$. The red and purple curves are averaged over $K = 15$ sequence realizations. Source : © 2023 IEEE. Reproduced with permission from [1].

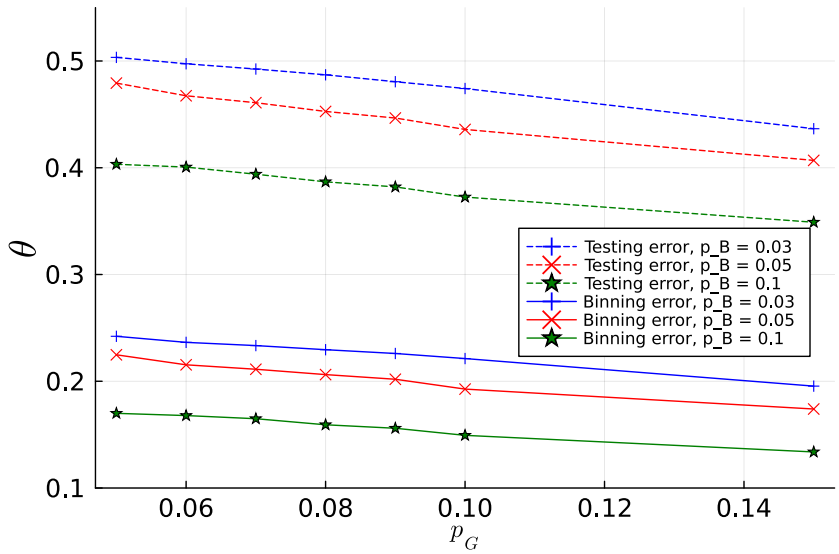

FIGURE 4.3 – Error exponents as functions of p_g, for various values of p_b and for $\bar{p}_G = 0.3$, $\bar{p}_B = 0.5$, $g = 0.001$, $b = 0.002$, $p = 0.2$, $\delta = 0.001$. Source : © 2023 IEEE. Reproduced with permission from [1].

be the focus of the next chapter. Future works will be dedicated to the analysis of more generic Hidden Markov Models as well as Gauss Markov models. Additionally, it is worth noting that for the considered source models, the limsup and liminf in probability converge to the same limit.

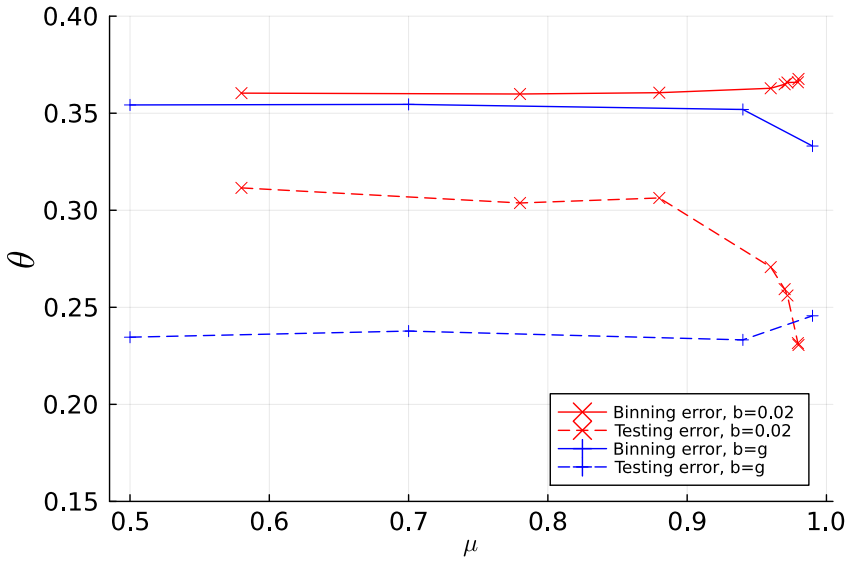

FIGURE 4.4 – Error exponents as functions of $\mu = 1 - g - b$, for $\bar{p}_G = 0.3$, $\bar{p}_B = 0.5$, $p = 0.2$, $\delta = 0.15$. Source : © 2023 IEEE. Reproduced with permission from [1].

However, one may consider other source models where the limsup and liminf do not converge to the same value. An example of such models is the class of mixed sources, which are stationary but non-ergodic [42]. We leave for future works the investigation of such models.

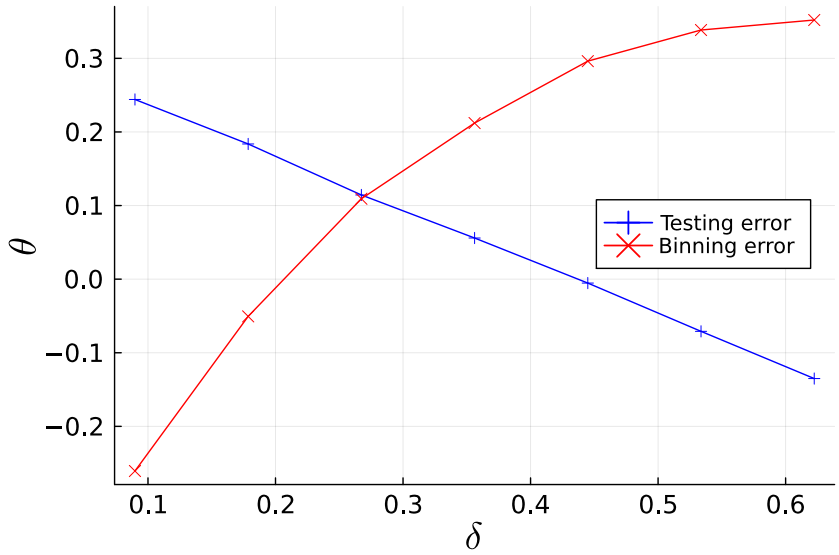

FIGURE 4.5 – Error exponents as functions of δ for $p_G = 0.5$, $p_B = 0.3$, $\bar{p}_G = 0.05$, $\bar{p}_B = 0.03$, $p = 0.2$, $R = 0.4$.

Chapitre 5

Practical short-length schemes for distributed hypothesis testing

5.1 Introduction

In this chapter, we focus on the design of practical short-length coding schemes for DHT for binary sources. Achievability proofs presented in Chapters 2 and 3 suggest considering either quantizer-alone or quantize-binning coding schemes for DHT. Interestingly, these achievability proofs consider coding schemes similar to what was proposed for the Wyner-Ziv problem ; with additional mechanisms specific to hypothesis testing. Some existing works have already introduced practical binary quantizers [39], binning schemes [4], and quantize-binning schemes [40], for Wyner-Ziv coding, all constructed with linear block codes. However, these constructions were originally designed with the aim of source reconstruction and typically involve very long source sequences, often exceeding 10^5 bits. In addition, the encoding and decoding rely on message-passing algorithms that perform poorly with shorter sequences. In contrast, DHT inherently deals with short-length sequences, where just a few dozen bits may suffice to make a correct decision. Therefore, it is essential to design short-length coding schemes dedicated to DHT. An additional challenge lies in how to perform the hypothesis test, particularly since the strategies proposed in information theory proofs, such as minimal entropy checks, are not directly implementable in practice. In this chapter, we address these key points.

We consider two different setups of the DHT problem. The first setup is the symmetric setup, where both $\mathbf{X}^n$ and $\mathbf{Y}^n$ are encoded, as illustrated in Figure 5.1. The second setup, referred to as the asymmetric setup, involves $\mathbf{Y}^n$ being fully available at the decoder and used as side information. For both symmetric and asymmetric setups, we begin by analyzing the truncation scheme, where only a portion of $\ell < n$ of the bits from the sequences are transmitted without any coding. This scheme serves as a baseline for performance comparison. Next, we propose practical implementations of quantization and quantize-binning schemes, both constructed with short-length binary linear block codes and specifically adapted to the decision-making problem. For the proposed practical scheme, we also make a parallel with the information-theoretic achievable coding schemes, emphasizing similarities and differences. The performance of the proposed

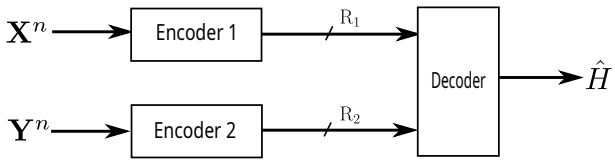

FIGURE 5.1 – Distributed hypothesis testing scheme

schemes will be compared to the baseline truncation scheme. Additionally, in the asymmetric setup, we derive precise analytical expressions for the Type-I and Type-II error probabilities for the proposed schemes. These new analytical tools should enable the optimization and comparison of the proposed practical schemes across a wide range of source and code parameters. Note that these expressions are derived only for the asymmetric case, which is simpler to handle compared to the symmetric case.

Our numerical results, reveal the efficiency of the quantize-binning scheme compared to the quantization and uncoded schemes, particularly when parameters are carefully chosen. Additionally, these numerical result validate the accuracy of the analytical error probabilities by demonstrating their consistency with Monte-Carlo simulations.

5.2 Notation

In addition to what was defined in Section 2.2, the following specific notations will be used throughout this chapter. We use $w(\mathbf{x}^n)$ to denote the Hamming weight of the vector $\mathbf{x}^n$, and $d(\mathbf{x}^n, \mathbf{y}^n)$ to denote the Hamming distance between $\mathbf{x}^n$ and $\mathbf{y}^n$. The binomial coefficient of the pair of integers (n, k) with $k \leq n$ is expressed as $\binom{n}{k}$.

5.3 System model

We consider the DHT problem depicted in Figure 5.1, where $\mathbf{X}^n$ and $\mathbf{Y}^n$ are two sequences of length n. Encoder 1 and Encoder 2 send coded versions of $\mathbf{X}^n$ and $\mathbf{Y}^n$ at rates R_1, and R_2, respectively, while the decoder aims to make a decision between two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, based on the received coded data. We consider two distinct setups:

1. Symmetric setup: here, both $\mathbf{X}^n$ and $\mathbf{Y}^n$ are encoded at rates R_1 and R_2, respectively.
2. Asymmetric setup: in this setup, $\mathbf{Y}^n$ is fully available at the decoder, i.e., $R_2 = \infty$.

In what follows, we focus only on binary sources as a starting point toward developing practical DHT schemes. This setup allows us to design and analyze practical DHT schemes in a first simplified setting, which can later be extended to more complex and realistic scenarios.

5.3.1 DHT for Binary Sources

We assume that the n symbols of the sequences $\mathbf{X}^n$ and $\mathbf{Y}^n$ are i.i.d. realizations of some random variables X and Y, respectively. Furthermore, X and Y are jointly distributed according to the model $Y = X \oplus Z$, where Z is a binary random variable independent of X, and we denote $p = \mathbb{P}(X = 1)$, and $c = \mathbb{P}(Z = 1)$ with $0 < c \leq 1/2$. The two hypotheses are expressed as:

$$\begin{cases} \mathcal{H}_0 : & (p = p_0, c = c_0), \\ \mathcal{H}_1 : & (p = p_1, c = c_1). \end{cases} \tag{5.1}$$

We assume, without loss of generality, that $p_0 < p_1$, and $c_0 < c_1$. This model was investigated from an information-theoretic perspective for instance in [2, 33]. Furthermore, when $p_0 = p_1$, and $c_1 = 1/2$, the problem (5.1) reduces to testing against independence [18]. However, it is important to note that, in the subsequent analysis, we do not restrict our attention to testing against independence.

5.3.2 Error exponent for binary DHT

Following the approach used in the information theory definitions in Chapter 3, we define the encoding functions for binary sources as follows:

$$f_1^{(n)} : \{0,1\}^n \to [\![1, 2^{v_1}]\!], \tag{5.2}$$

$$f_2^{(n)} : \{0,1\}^n \to [\![1, 2^{v_2}]\!], \tag{5.3}$$

and a decision function

$$g^{(n)} : [\![1, 2^{v_2}]\!] \times [\![1, 2^{v_2}]\!] \times \to \{0,1\} \tag{5.4}$$

We consider a rate-limited setup in which $v_1/n \leq R_1$, and $v_2/n \leq R_2$.

For given functions $(f_1^{(n)}, f_2^{(n)}, g^{(n)})$, we define Type-I error probability α_n and Type-II error probability β_n as [33]

$$\alpha_n = \mathbb{P}\left(g^{(n)}(f_1^{(n)}(\mathbf{X}^n), f_2^{(n)}(\mathbf{Y}^n)) = 1 | \mathcal{H}_0\right), \tag{5.5}$$

$$\beta_n = \mathbb{P}\left(g^{(n)}(f_1^{(n)}(\mathbf{X}^n), f_2^{(n)}(\mathbf{Y}^n)) = 0 | \mathcal{H}_1\right). \tag{5.6}$$

For a given value $\epsilon \in (0,1)$ such that $\alpha_n < \epsilon$, Type-II error exponent θ is defined as [33]

$$\lim_{n\to\infty} \sup \frac{1}{n} \log_2 \frac{1}{\beta_n} \geq \theta. \tag{5.7}$$

In Chapter 3, we established a general bound on the error exponent θ, as expressed in equation (3.22) for the asymmetric setup. This error exponent depends on an auxiliary source $\mathbf{U}$ which

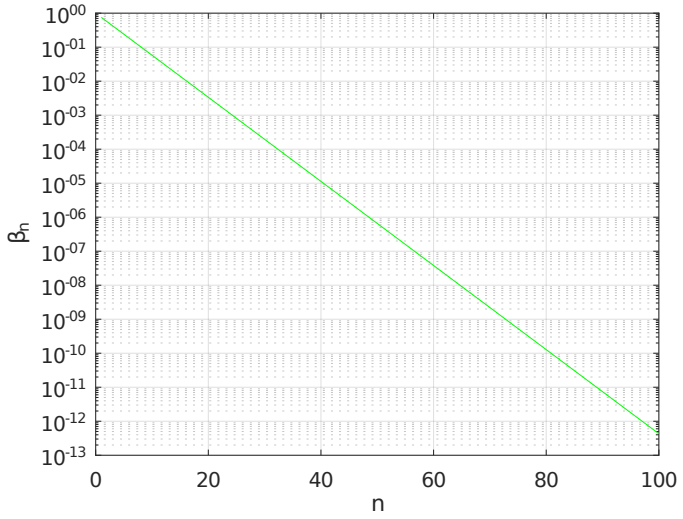

FIGURE 5.2 – Type-II error probability as function of n

has to satisfy the Markov chain $\mathbf{U} \to \mathbf{X} \to \mathbf{Y}$ [67]. In what follows, we assume that this auxiliary source generates a sequence of symbols $\{U_k\}_{k=1}^{+\infty}$ such that

$$U_k = X_k \oplus \phi_k, \tag{5.8}$$

where $\phi_k \sim \text{Bern}(\delta)$, and X_k and ϕ_k are independent. Specifying the error exponent bound in (3.22) for the hypothesis testing problem defined in (5.1) leads to

$$\theta \geq \sup_{\delta \in [0,1]} \min \Big\{ R - [H_2(p_0 * \delta) - H_2(\delta)], \\ (p_0 * \delta) \log \frac{p_0 * \delta}{p_1 * \delta} + (1 - (p_0 * \delta)) \log \frac{1 - (p_0 * \delta)}{1 - (p_1 * \delta)} \Big\}. \tag{5.9}$$

Here, H_2 is the binary entropy function, and $*$ is the binary convolution operator defined as $x * y = (1-x)y + (1-y)x$, with $0 \leq x, y \leq 1$.

5.3.3 Short-length nature of DHT

While the expression of the error exponent θ in (5.9) provides a scaling law for Type-II error probability β_n, it remains an asymptotic result due to the limit as n tends to infinity in definition (5.7). Nevertheless, it confirms the intuition that the problem inherently involves short sequences. For instance, consider parameters $p_0 = 0.05$, $p_1 = 0.5$, $\delta = 0.1$, $r = 0.4$. In Figure 5.2, we evaluate the quantity $e^{-n\theta}$ for these parameters with $n = 100$, the result is approximately 10^{-12}, and for $n = 50$, it yields approximately 10^{-6}. This strongly suggests that practical schemes

should focus on values of n less than 50. Hence, we now introduce practical coding schemes for DHT tailored for such short sequence lengths.

5.4 Uncoded schemes

When considering DHT, there is no need to reconstruct all the exact values of the source bits $\mathbf{x}^n$, and $\mathbf{y}^n$. Therefore, in this section, we describe two schemes which do not involve any coding. In the first scheme, called separate scheme, the decision is taken directly at the encoders and then transmitted to the decoder, which consists of 1 bit of information per decision. The second scheme, referred to as the truncation scheme, involves the encoders transmitting only the first $\ell < n$ bits of their respective sequences, with the decision being made at the decoder. While these schemes do not constitute novelty in themselves, they serve as baselines for performance comparison with the proposed practical quantization and quantize-binning schemes in the subsequent sections. Additionally, the separate scheme achieves a very low transmission rate by making the decision at the encoders and sending only a single bit to the decoder at the price of a loss in testing performance. We will compare the performance of the separate and the truncation schemes at the end of the section.

5.4.1 Separate scheme

When the marginal distributions of $\mathbf{X}^n$ or $\mathbf{Y}^n$ depend on the hypothesis, a test can be constructed at the encoders based only on their observations. Therefore, the separate scheme we describe in this section is relevant only when $p_0 \neq p_1$. In this scheme, Encoder 1 and Encoder 2 independently make local decisions based on their respective observations $\mathbf{x}^n$ and $\mathbf{y}^n$. Each encoder sends only a single bit to inform the decoder of its decision. The decoder then makes the final decision based on the received bits from the encoders. While this setup may not be optimal in terms of test, it offers the advantage of achieving very low communication rates by transmitting just one bit to the decoder.

In this scheme, the encoding functions $f_1^{(n)}$ and $f_2^{(n)}$ for X and Y, respectively are described as

$$f_1^{(n)} : \{0,1\}^n \to \{0,1\}, \tag{5.10}$$

$$f_2^{(n)} : \{0,1\}^n \to \{0,1\}, \tag{5.11}$$

and the decision function is given by

$$g^{(n)} : \{0,1\} \times \{0,1\} \to \{\mathcal{H}_0, \mathcal{H}_1\}. \tag{5.12}$$

The coding rates are given by $R_1 = R_2 = 1/n$.

5.4.1.1 Construction of the encoding functions of the separate scheme

The encoding functions $f_1^{(n)}$ and $f_2^{(n)}$ in (5.10) and (5.11) are respectively constructed based on a Neyman-Pearson (NP) test [70] on the observations $\mathbf{x}^n$, and $\mathbf{y}^n$, respectively. Under certain constraints $\alpha_n^{(x)} < \epsilon$, and $\alpha_n^{(y)} < \epsilon$ on Type-I error probabilities for Encoder 1 and Encoder 2, respectively, the NP lemma [70] states that the following tests at Encoder 1 and Encoder 2:

$$\mathbb{P}_1(\mathbf{x}^n) < \mu_1 \mathbb{P}_0(\mathbf{x}^n), \tag{5.13}$$

$$\mathbb{P}_1(\mathbf{y}^n) < \mu_2 \mathbb{P}_0(\mathbf{y}^n), \tag{5.14}$$

minimize Type-II error probabilities $\beta_n^{(x)}$, and $\beta_n^{(y)}$, respectively, where μ_1, and μ_1 are threshold values chosen to satisfy the Type-I error constraints. $\mathbb{P}_0$, and $\mathbb{P}_1$ are the marginal distributions of $\mathbf{x}^n$ ($\mathbf{y}^n$) under hypothesis $\mathcal{H}_0$ and under hypothesis $\mathcal{H}_1$, respectively.

Given that $p_0 < p_1$, and $c_0 < c_1$ it is shown in [70] that the tests described by (5.13) and (5.14) are equivalent respectively to the conditions:

$$w(\mathbf{x}^n) < \lambda_1, \tag{5.15}$$

$$w(\mathbf{y}^n) < \lambda_2 \tag{5.16}$$

where λ_1 and $\lambda_2 \in \mathbb{N}$ are integer threshold values chosen so as to satisfy the constraints $\alpha_n^{(x)} < \epsilon$, and $\alpha_n^{(y)} < \epsilon$, respectively.

5.4.1.2 Construction of the decision function of the separate scheme

The decision function $g^{(n)}$ in (5.12) is described as follows. Let us denote $b_1 = f_1^{(n)}(\mathbf{x}^n)$ and $b_2 = f_2^{(n)}(\mathbf{y}^n)$ as the 1-bit produced by Encoder 1 and Encoder 2, respectively. Upon receiving b_1 and b_2, the decoder decides that $g^{(n)}(b_1, b_2) = \mathcal{H}_i$ if $b_1 = i$ and $b_2 = i$ for $i = 0, 1$. Otherwise, the decoder prefers to rely on the decision of Encoder 1. This test is motivated by the fact that Y is a noisy version of X, and that $\mathbb{P}(Y = 1) > \mathbb{P}(X = 1)$ under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ as given by (5.1). Other strategies may be considered depending on the assumptions on p_0, p_1, c_0, c_1, but for simplicity, we focus only on this one here.

The next proposition provides analytical expressions of Type-I and Type-II error probabilities of the separate scheme.

Proposition 5.1 *The Type-I and Type-II error probabilities for NP test of the separate scheme*

are given by

$$\alpha_s = 1 - \sum_{k=0}^{\gamma_s} \binom{n}{k} p_0^k (1-p_0)^{n-k}, \tag{5.17}$$

$$\beta_s = \sum_{k=0}^{\gamma_s} \binom{n}{k} p_1^k (1-p_1)^{n-k}, \tag{5.18}$$

where $\gamma_s \in \mathbb{N}$ is an integer threshold value chosen to satisfy the constraint on the Type-I error of the NP test:

$$w(\mathbf{x}^n) < \gamma_s. \tag{5.19}$$

Proof. For the separate scheme described above, the Type-I error can be evaluated as

$$\begin{aligned} \alpha_s &= \mathbb{P}_0(b_1 = 1, b_2 = 1) + \mathbb{P}_0(b_1 = 1, b_2 = 0) \\ &= \mathbb{P}_0(b_1 = 1) \\ &= \mathbb{P}_0(w(\mathbf{x}^n) > \gamma_s) \\ &= 1 - \sum_{k=0}^{\gamma_s} \mathbb{P}_0(w(\mathbf{x}^n = k)) \\ &= 1 - \sum_{k=0}^{\gamma_s} \binom{n}{k} p_0^k (1-p_0)^{n-k}. \end{aligned} \tag{5.20}$$

This gives (5.17). In (5.20), $\mathbb{P}_0$ represents the distribution under hypothesis $\mathcal{H}_0$. To obtain (5.18), we follow the same steps as in (5.20) by replacing p_0 by p_1, which ends the proof.

The expressions (5.17) and (5.18) are also applicable to the asymmetric setup, where Y serves as side information at the decoder.

5.4.2 Truncation scheme

In the truncation scheme, the encoders transmit $\ell < n$ of their observations to the decoder which proceeds to the decision.

5.4.2.1 Code construction of the truncation scheme

The truncation scheme consists of sending the first ℓ symbols of the source vector $\mathbf{x}^n$ and $\mathbf{y}^n$ at the coding rate $R_1 = R_2 = \ell/n$ at the decoder. The decoder can then perform a standard NP test [70] on the pair $(\mathbf{x}^\ell, \mathbf{y}^\ell)$. Under a certain constraint $\alpha_n^t < \epsilon$ on Type-I error probability for the truncation scheme, the NP lemma [70] states that the following test:

$$\mathbb{P}_1(\mathbf{x}^\ell, \mathbf{y}^\ell) < \mu \mathbb{P}_0(\mathbf{x}^\ell, \mathbf{y}^\ell), \tag{5.21}$$

minimizes Type-II error probability β_n^t, where μ is a threshold value chosen to satisfy the Type-I error constraint. In (5.21), $\mathbb{P}_0$ and $\mathbb{P}_1$ are the joint probability distributions of $(\mathbf{x}^\ell, \mathbf{y}^\ell)$ under hypothesis $\mathcal{H}_0$ and under hypothesis $\mathcal{H}_1$, respectively.

5.4.2.2 Theoretical analysis of the truncation scheme

The next proposition provides analytical expressions of Type-I and Type-II error probabilities of the truncation scheme.

Proposition 5.2 *For the truncation scheme, given that $p_0 < p_1$ and $c_0 < c_1$, the analytical expressions of Type-I and Type-II errors are given by*

$$\alpha_n^t = \sum_{\substack{(\lambda,j):\\ T_{\lambda,j} \geq \tau_t}} \binom{\ell}{\lambda} p_0^\lambda (1-p_0)^{\ell-\lambda} \binom{n}{j} c_0^j (1-c_0)^{\ell-j} \tag{5.22}$$

$$\beta_n^t = \sum_{\substack{(\lambda,j):\\ T_{\lambda,j} \leq \tau_t}} \binom{\ell}{\lambda} p_1^\lambda (1-p_1)^{n-\lambda} \binom{\ell}{j} c_1^j (1-c_1)^{\ell-j} \tag{5.23}$$

where $T_{\lambda,j} = \mu \log_2 \frac{p_1(1-p_0)}{p_0(1-p_1)} + j \log_2 \frac{c_1(1-c_0)}{c_0(1-c_1)}$, and $\tau_t = \log_2 \mu + n \log_2 \frac{(1-p_0)(1-c_0)}{(1-p_1)(1-c_1)}$.

Proof. For the truncation scheme in the symmetric setup, the NP test (5.21) is equivalent to

$$\begin{aligned} \mathbb{P}_1\left(\mathbf{x}^\ell\right) \mathbb{P}_1\left(\mathbf{y}^\ell \mid \mathbf{x}^\ell\right) &\leq \mu\, \mathbb{P}_0\left(\mathbf{x}^\ell\right) \mathbb{P}_0\left(\mathbf{y}^\ell \mid \mathbf{x}^\ell\right) \\ \mathbb{P}_1\left(\mathbf{x}^\ell\right) \mathbb{P}_1\left(\mathbf{z}^\ell\right) &\leq \mu\, \mathbb{P}_0\left(\mathbf{x}^\ell\right) \mathbb{P}_0\left(\mathbf{z}^\ell\right), \end{aligned} \tag{5.24}$$

where $\mathbf{z}^\ell = \mathbf{x}^\ell \oplus \mathbf{y}^\ell$. After simplification and passing through the logarithms, given that $p_0 < p_1$ and $c_0 < c_1$, we obtain

$$w(\mathbf{x}^\ell) \log_2 \frac{p_1(1-p_0)}{p_0(1-p_1)} + w(\mathbf{z}^\ell) \log_2 \frac{c_1(1-c_0)}{c_0(1-c_1)} \leq \tau_t, \tag{5.25}$$

where $\tau_t = \log_2 \mu + n \log_2 \frac{(1-p_0)(1-c_0)}{(1-p_1)(1-c_1)}$. From (5.25), we remark that the Type-I can be evaluated

as

$$\begin{aligned}\alpha_n^t &= \mathbb{P}_0\left(w(\mathbf{x}^\ell)\log_2\frac{p_1(1-p_0)}{p_0(1-p_1)} + w(\mathbf{z}^\ell)\log_2\frac{c_1(1-c_0)}{c_0(1-c_1)} \geq \tau_t\right)\\ &= \sum_{\substack{(\lambda,j):\\ T_{\lambda,j}\geq\tau_t}} \mathbb{P}_0\left(w(\mathbf{x}^\ell)=\lambda, w(\mathbf{z}^\ell)=j\right) \qquad (5.26)\\ &= \sum_{\substack{(\lambda,j):\\ T_{\lambda,j}\geq\tau_t}} \mathbb{P}_0\left(w(\mathbf{x}^\ell)=\lambda\right)\mathbb{P}_0\left(w(\mathbf{z}^\ell)=j\right)\\ &= \sum_{\substack{(\lambda,j):\\ T_{\lambda,j}\geq\tau_t}} \binom{\ell}{\lambda} p_0^\lambda (1-p_0)^{\ell-\lambda}\binom{n}{j} c_0^j (1-c_0)^{\ell-j}, \qquad (5.27)\end{aligned}$$

where $T_{\lambda,j} = \mu\log_2\frac{p_1(1-p_0)}{p_0(1-p_1)} + j\log_2\frac{c_1(1-c_0)}{c_0(1-c_1)}$. This gives (5.22). Similarly,

$$\beta_n^t = \mathbb{P}_1\left(w(\mathbf{x}^\ell)\log_2\frac{p_1(1-p_0)}{p_0(1-p_1)} + w(\mathbf{z}^\ell)\log_2\frac{c_1(1-c_0)}{c_0(1-c_1)} \leq \tau_t\right),$$

we can obtain (5.23) by following the same steps, which ends the proof.

The expressions (5.22) and (5.23) for the truncation scheme are also applicable to the asymmetric setup. The only difference is that in the asymmetric case, there is no need to transmit anything for Y; it is fully available at the decoder.

We expect the truncation scheme to be more efficient in terms of testing than the separate one while necessitating a larger rate ℓ/n. To verify this, we now compare the performance of both schemes in terms of Type-I and Type-II error probabilities.

5.4.3 Separate scheme versus truncation scheme

In this part, we aim to compare the separate scheme to the truncation scheme. To do this, we rely on Receiver Operating Characteristic (ROC) curves, a standard tool for evaluating hypothesis test performance. The ROC curve illustrates the trade-off between the Type-II and Type-I error probabilities. Each point of the curves is obtained for a different value of the decision threshold.

In Figure 5.3, we set $n = 30$, $c_0 = 0.1$, $p_1 = 0.5$ and $c_1 = 0.35$. We then compare the Type-II versus Type-I performance of the separate scheme (as discussed in Section 5.4.1) and of the truncation scheme (outlined in Section 5.4.2) for different values on p_0. The results show that the separate scheme outperforms the truncation scheme for small p_0. On the other hand, as p_0 increases and approaches p_1, the truncation scheme gains an advantage compared to the separate one. This is because when p_0 approaches p_1, it becomes very challenging for Encoder 1

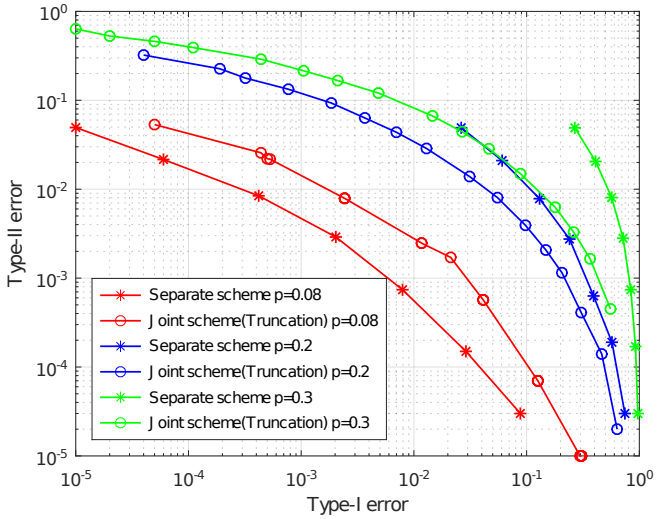

FIGURE 5.3 – ROC curve for separate scheme compared to truncation scheme.

to distinguish accurately between $\mathcal{H}_0$ and $\mathcal{H}_1$ only based on its local observations. However, the truncation scheme exploits the joint information of both encoders to improve decision accuracy.

In conclusion, the separate scheme excels in rate efficiency by transmitting only 1-bit, but its decision accuracy is compromised when p_0 increases and approaches p_1. However, the truncation scheme sacrifices rate efficiency, as encoders transmit truncated data at a rate $R_1 = R_2 = R = \ell/n$, yet achieve higher decision accuracy through joint decoding when p_0 increases.

However, these schemes do not exploit coding. Information-theoretic results suggest that coding can improve performance for DHT. Therefore, we now turn our focus to code design for this problem.

5.5 Quantization scheme

In their seminal work [18], Ahlswede and Csiszár introduced the first DHT scheme based only on a quantizer from the information-theoretic perspective, as discussed in Chapter 2, Section 2.3.2. Here, we present a practical implementation of this scheme for short-length sequences by utilizing linear block codes.

5.5.1 Code construction of the quantization scheme for the symmetric setup

To practically implement binary quantization for the symmetric setup, we follow the approach of [71] and consider a generator matrix G_q with dimension $n \times m$ of a linear code. For given

source sequences $\mathbf{x}^n$ and $\mathbf{y}^n$, the encoders produce vectors $\mathbf{z}_{q,1}^m$ and $\mathbf{z}_{q,2}^m$ of length m bits as [72]

$$\mathbf{z}_{q,1}^m = \arg\min_{\mathbf{z}^m} d\left(G_q\mathbf{z}^m, \mathbf{x}^n\right) \tag{5.28}$$

and

$$\mathbf{z}_{q,2}^m = \arg\min_{\mathbf{z}^m} d\left(G_q\mathbf{z}^m, \mathbf{y}^n\right). \tag{5.29}$$

In (5.28), and (5.29), the challenge lies in determining the quantized vectors $\mathbf{z}_{q,1}^m$ and $\mathbf{z}_{q,2}^m$ that achieve the minimum Hamming distance. In [39, 40], it is proposed to build efficient binary quantizers using low-density generator matrices (LDGM). LDGM codes were considered so as to develop a low complexity message-passing algorithm called Bias-Propagation to solve (5.28) and (5.29). However, the schemes introduced in [39, 40] consider very long codes (more than 10^5 bits). Here, due to the short-length nature of the problem, we cannot use the Bias-Propagation algorithm since it leads to an important loss in performance on the considered codes. Instead, we will solve (5.28) and (5.29) exactly by exhaustive search. Therefore, we consider any generator matrix G_q, not necessarily obtained from an LDGM code. For example, in our simulations, we will consider BCH and Reed-Muller codes.

The codewords $\mathbf{z}_{q,1}^m$, and $\mathbf{z}_{q,2}^m$ are then transmitted to the decoder at code rates $R_1 = R_2 = m/n$. The decoder first computes the quantized vectors $\mathbf{x}_q^n = G_q\mathbf{z}_{q,1}^m$ and $\mathbf{y}_q^n = G_q\mathbf{z}_{q,2}^m$ and then decides between $\mathcal{H}_0$ and $\mathcal{H}_1$ based on the NP test

$$\mathbb{P}_1\left(\mathbf{x}_q^n, \mathbf{y}_q^n\right) \leq \mu_q \mathbb{P}_0\left(\mathbf{x}_q^n, \mathbf{y}_q^n\right), \tag{5.30}$$

where μ_q is an integer threshold. The test (5.30) is equivalent to

$$w(\mathbf{x}_q^n)\log_2\frac{\hat{p}_1(1-\hat{p}_0)}{\hat{p}_0(1-\hat{p}_1)} + w(\mathbf{v}_q^n)\log_2\frac{\hat{c}_1(1-\hat{c}_0)}{\hat{c}_0(1-\hat{c}_1)} \leq \tau_q, \tag{5.31}$$

where $\tau_q = \log_2\mu_q + n\log_2\frac{(1-\hat{p}_0)(1-\hat{c}_0)}{(1-\hat{p}_1)(1-\hat{c}_1)}$, and $\mathbf{v}_q = \mathbf{x}_q \oplus \mathbf{y}_q$. In (5.31), $(\hat{p}_0, \hat{c}_0)$ and $(\hat{p}_1, \hat{c}_1)$ are estimated values of p, and c, respectively, under $\mathcal{H}_0$, and $\mathcal{H}_1$ through Monte-Carlo simulations.

Note that computing the joint distribution distributions $\mathbb{P}_0(\mathbf{x}_q^n, \mathbf{y}_q^n)$ and $\mathbb{P}_1(\mathbf{x}_q^n, \mathbf{y}_q^n)$ in (5.30) is not straightforward in the symmetric setup. To facilitate the transition from (5.30) to (5.31), we introduced the following assumptions. First, given that $Y = X \oplus Z$, the conditional probability $\mathbb{P}(\mathbf{y}^n \mid \mathbf{x}^n)$ is given by $\mathbb{P}(\mathbf{z}^n)$, where, $\mathbf{x}^n$, $\mathbf{y}^n$, and $\mathbf{z}^n = \mathbf{x}^n \oplus \mathbf{y}^n$ represent the realizations of the random variables X, Y, and Z, respectively. Therefore, given that $\mathbf{x}^n$ and $\mathbf{z}^n$ are the realizations of i.i.d. random variables $X \sim \text{Bern}(p)$ and $Z \sim \text{Bern}(c)$, respectively, we extend this assumption to the quantized variables. Specifically, we assume that $\mathbf{x}_q^n$ and $\mathbf{v}_q^n = \mathbf{x}_q^n \oplus \mathbf{y}_q^n$ are the realizations of i.i.d. random variables $X_q \sim \text{Bern}(\hat{p})$ (with $\hat{p} = \hat{p}_0$ under $\mathcal{H}_0$, and $\hat{p} = \hat{p}_1$ under $\mathcal{H}_1$) and $V_q \sim \text{Bern}(\hat{c})$ (with $\hat{c} = \hat{c}_0$ under $\mathcal{H}_0$, and $\hat{c} = \hat{c}_1$ under $\mathcal{H}_1$), with $V_q = X_q \oplus Y_q$.

5.5.2 Code construction of the quantization scheme for the asymmetric setup

In the asymmetric setup, the coding scheme is obtained in a straightforward manner from the previous one. In this case, only $\mathbf{x}^n$ is quantized and transmitted at rate $R_1 = m/n$, while $\mathbf{y^n}$ is available at the decoder. Therefore, the decoder first computes the quantized vector $\mathbf{x}_q^n = G_q \mathbf{z}_{q,1}^m$. Then, since by the previous assumptions, $\mathbb{P}(\mathbf{z}_{q,1}^m, \mathbf{y}^n) = \mathbb{P}(\mathbf{x}_q^n, \mathbf{y}^n)$, the NP test (5.30) in the asymmetric setup reduces to

$$\sum_{i=1}^{n} (x_{q,i} \oplus y_i) < \lambda_q, \tag{5.32}$$

where λ_q is an integer threshold.

5.5.3 Comparison with information-theoretic scheme

We now discuss the construction of this practical scheme compared to the information-theoretic one of [21]. Our practical implementation of the quantization scheme closely follows the approach outlined in the information-theoretic proofs [21]. In information theory, the encoder selects a sequence $\mathbf{z}^n$ from a generated codebook as a quantized version of $\mathbf{x}^n$ if $(\mathbf{z}^n, \mathbf{x}^n)$ are jointly typical. In our practical implementation, we use linear codes for quantization, where the quantized version of $\mathbf{x}^n$ is the one that minimizes the Hamming distance, as defined in (5.28). On the decoder side, while the information-theoretic method accepts $\mathcal{H}_0$ if $\mathbf{z}^n$ and $\mathbf{y}^n$ are jointly typical under $\mathcal{H}_0$, our practical implementation uses the NP test (5.30) to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$. Since the NP test is known to be optimal in hypothesis testing for the problem stated in (5.1) [70], we are sure not to lose any performance compared to the information-theoretic approach. Moreover, the NP test is simple to manipulate for our specific source models.

5.5.4 Theoretical analysis of the quantization scheme

In this section, we provide a theoretical analysis of the practical quantization scheme, for the asymmetric setup only. As for the separate and truncation coding schemes in Sections 5.4.2 and 5.4.2, we provide exact analytical expressions of the Type-I and Type-II error probabilities of the proposed scheme for the considered generator matrix G_q. The extension to the symmetric setup is very complex and is left for future work.

In the NP test (5.32) of the quantization scheme in the asymmetric setup, the decision is made on the vectors $\mathbf{x}_q^n$ and $\mathbf{y}^n$, with larger dimensions $n > \ell$, compared to the baseline truncation scheme. However, the vector $\mathbf{x}_q^n$ contains quantization errors with respect to the original $\mathbf{x}^n$. This is why, in what follows, we aim to provide a theoretical analysis to compare the impact of quantization on decision performance by providing closed-form expressions of the Type-I and Type-II error probabilities.

We first introduce the following notation related to the considered code with generator

matrix G_q. Consider the set of integers $\{E_\gamma^{(q)}\}_{\gamma\in[\![0,d_{\max}^{(q)}]\!]}$, where $E_\gamma^{(q)}$ is the number of words $\mathbf{x}^n$ of Hamming weight γ that belong to the decision region $\mathcal{C}_0^{(q)}$ of $\mathbf{x}_q^n = \mathbf{0}^n$. In other words, $\mathbf{x}^n \in \mathcal{C}_0^{(q)}$ means that the solution of (5.28) for $\mathbf{x}^n$ is $\mathbf{0}^m$. We further denote $N_0^{(q)} = \sum_{\gamma=0}^{d_{\max}^{(q)}} E_\gamma^{(q)}$.

Proposition 5.3 *For the quantization scheme and a threshold value λ_q, Type-I and Type-II error probabilities are given by*

$$\alpha_n^{(q)} = 1 - \frac{1}{N_0^{(q)}} \sum_{\lambda=0}^{\lambda_q} \sum_{\gamma=0}^{d_{max}^{(q)}} \sum_{j=0}^{n} E_\gamma^{(q)} \Delta_{\lambda,j,\gamma} \binom{n}{j} c_0^j (1-c_0)^{n-j}, \tag{5.33}$$

$$\beta_n^{(q)} = \frac{1}{N_0^{(q)}} \sum_{\lambda=0}^{\lambda_q} \sum_{\gamma=0}^{d_{max}^{(q)}} \sum_{j=0}^{n} E_\gamma^{(q)} \Delta_{\lambda,j,\gamma} \binom{n}{j} c_1^j (1-c_1)^{n-j}, \tag{5.34}$$

where

$$\Delta_{\lambda,j,\gamma} = \frac{\Gamma_{\lambda,j,\gamma}}{\sum_{i=0}^{max(\lambda,j)} \binom{\lambda}{i}\binom{n-\lambda}{j-i}} \tag{5.35}$$

and, for $\gamma = j + \lambda - 2u$ and $0 \le u \le \min(\lambda, j) \le n$,

$$\Gamma_{\lambda,j,\gamma} = \binom{\lambda}{u}\binom{n-\lambda}{j-u}. \tag{5.36}$$

Proof. Since by symmetry, the quantizer error probability is independent of the transmitted codeword [71], we consider the all-zero codeword $\mathbf{x}_q^n = \mathbf{0}$. From (5.32), we develop

$$\begin{aligned}
\alpha_n^{(q)} &= 1 - \sum_{\lambda=0}^{\lambda_q} \mathbb{P}_0(w(\mathbf{Y}^n) = \lambda) \\
&= 1 - \sum_{\lambda=0}^{\lambda_q} \sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \mathbb{P}_0(w(\mathbf{Y}^n) = \lambda | w(\mathbf{X}^n) = \gamma) \\
&= 1 - \sum_{\lambda=0}^{\lambda_q} \sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \mathbb{P}_0(d(\mathbf{X}^n, \mathbf{Y}^n) = j) \Delta_{\lambda,j,\gamma} \\
&= 1 - \sum_{\lambda=0}^{\lambda_q} \sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \binom{n}{j} c_0^j (1-c_0)^{n-j} \Delta_{\lambda,j,\gamma}.
\end{aligned} \tag{5.37}$$

This gives (5.33). To obtain (5.34), we remark that $\beta_n^{(q)} = \sum_{\lambda=0}^{\lambda_q} \mathbb{P}_1(w(\mathbf{Y}^n) = \lambda)$ and follow the same steps as in (5.37), by replacing c_0 by c_1, which ends the proof.

These theoretical results are novel and differ from both the information-theoretic analysis of DHT and existing results in channel coding. Especially, while error probability expressions exist for linear block codes dedicated to channel coding, such results had not been established for the

DHT problem. These new analytical allow us to predict decision performance without relying on Monte-Carlo simulations and facilitate code design by considering parameters such as $E_\gamma^{(q)}$. For instance, optimizing $E_\gamma^{(q)}$ can yield an optimal quantizer for DHT, as explored in [73].

5.6 Quantize-binning scheme

In their seminal work [22], Shimokawa et al. introduced the quantize-binning scheme for DHT. This scheme leverages the correlation between sources to reduce the compression rate. Furthermore, for a given compression rate, the quantize-binning scheme achieves lower error rates compared to the quantization scheme, as discussed in Section 2.3.4. We now introduce a practical short-length implementation of this scheme by using linear block codes.

5.6.1 Code construction of the quantize-binning scheme for the symmetric setup

To practically implement the quantize-binning scheme, we consider as before a generator matrix G_q of size $n \times m$. We also resort to the parity check matrix H_b of size $k \times m$ of another linear block code. In the symmetric setup, given the source vectors $\mathbf{x}^n$ and $\mathbf{y}^n$, the encoders employ the quantization method described in (5.28) and (5.29) to obtain sequences $\mathbf{z}_{q,1}^m$ and $\mathbf{z}_{q,2}^m$ for $\mathbf{x}^n$ and $\mathbf{y}^n$, respectively. Then, the encoders utilize the parity check matrix H_b of size $m \times k$ from another linear code to compute the syndromes

$$\mathbf{u}_1^k = H_b \mathbf{z}_{q,1}^m, \tag{5.38}$$

and

$$\mathbf{u}_2^k = H_b \mathbf{z}_{q,2}^m. \tag{5.39}$$

The syndromes $\mathbf{u}_1^k$ and $\mathbf{u}_2^k$ are then transmitted to the decoder at rates $R_1 = R_2 = k/n$. At the decoder, as discussed in Section 5.5, we avoid using message-passing algorithms since they do not perform well with finite-length sequences. Instead, we opt for an exhaustive search. Therefore, the decoder first identifies by exhaustive search, vectors $\hat{\mathbf{z}}_{q,1}^m$ and $\hat{\mathbf{z}}_{q,2}^m$ as

$$\hat{\mathbf{z}}_{q,1}^m, \hat{\mathbf{z}}_{q,2}^m = \arg \min_{\mathbf{z}_1^m, \mathbf{z}_2^m} d\left(G_q \mathbf{z}_1^m, G_q \mathbf{z}_2^m\right) \text{ s.t. } H_b \mathbf{z}_1^m = \mathbf{u}_1^k, \text{ and } H_b \mathbf{z}_2^m = \mathbf{u}_2^k. \tag{5.40}$$

We then compute $\mathbf{x}_{q,b}^n = G_q \hat{\mathbf{z}}_{q,1}^m$ and $\mathbf{y}_{q,b}^n = G \hat{\mathbf{z}}_{q,2}^m$, and, apply the following NP test

$$\mathbb{P}_1\left(\mathbf{x}_{q,b}^n, \mathbf{y}_{q,b}^n\right) \leq \mu_{q,b} \mathbb{P}_0\left(\mathbf{x}_{q,b}^n, \mathbf{y}_{q,b}^n\right), \tag{5.41}$$

where $\mu_{q,b}$ is an integer threshold. The NP test (5.41) for the quantize-binning is equivalent to the following condition

$$w(\mathbf{x}_{q,b}^n)\log_2\frac{\hat{p}_{1,b}(1-\hat{p}_{0,b})}{\hat{p}_{0,b}(1-\hat{p}_{1,b})}+w(\mathbf{v}_{q,b}^n)\log_2\frac{\hat{c}_{0,b}(1-\hat{c}_{0,b})}{\hat{c}_{0,b}(1-\hat{c}_{1,b})}\leq\tau_{q,b}, \tag{5.42}$$

where $\tau_{q,b}=\log_2\mu_{q,b}+n\log_2\frac{(1-\hat{p}_{0,b})(1-\hat{c}_{0,b})}{(1-\hat{p}_{1,b})(1-\hat{c}_{1,b})}$, and $\mathbf{v}_{q,b}^n=\mathbf{x}_{q,b}^n\oplus\mathbf{y}_{q,b}^n$. Here $\hat{p}_{0,b}$, $\hat{c}_{0,b}$, $\hat{p}_{1,b}$, and $\hat{c}_{1,b}$ are also estimated through Monte-Carlo Simulations.

As in Section 5.5.1, we note that computing the joint distributions $\mathbb{P}_0\left(\mathbf{x}_{q,b}^n,\mathbf{y}_{q,b}^n\right)$ and $\mathbb{P}_1\left(\mathbf{x}_{q,b}^n,\mathbf{y}_{q,b}^n\right)$ in (5.41) for the quantize-binning scheme in the symmetric setup is even more complex. To facilitate the transition from (5.41) to (5.42), we adopt the same assumptions as in Section 5.5.1. Specifically, we assume that $\mathbf{x}_{q,b}^n$ and $\mathbf{v}_{q,b}^n=\mathbf{x}_{q,b}^n\oplus\mathbf{y}_{q,b}^n$ are the realizations of i.i.d. random variables of some vector random variables $X_{q,b}\sim\text{Bern}(\hat{p}_b)$ (with $\hat{p}_b=\hat{p}_{0,b}$ under $\mathcal{H}_0$, and $\hat{p}_b=\hat{p}_{1,b}$, under $\mathcal{H}_1$), and $V_{q,b}\sim\text{Bern}(\hat{c}_b)$ (with $\hat{c}_b=\hat{c}_{0,b}$ under $\mathcal{H}_0$, and $\hat{c}_b=\hat{c}_{1,b}$, under $\mathcal{H}_1$), with $V_{q,b}=X_{q,b}\oplus Y_{q,b}$. Here $\hat{p}_{0,b}$, $\hat{c}_{0,b}$, $\hat{p}_{1,b}$, and $\hat{c}_{1,b}$ are also estimated through Monte-Carlo simulations.

5.6.2 Code construction of the quantize-binning scheme for the asymmetric setup

In the asymmetric case, only $\mathbf{x}^n$ is quantized and binned as $\mathbf{u}_1^n$, and transmitted at rate $R_1=k/n$, while $\mathbf{y}^n$ serves as side information. Therefore, the receiver first identifies by exhaustive search vector $\hat{\mathbf{z}}_1^m$ as

$$\hat{\mathbf{z}}_1^m=\arg\min_{\mathbf{z}^m}\ d(G_q\mathbf{z}^m,\mathbf{y}^n)\text{ s.t. }H_b\mathbf{z}^m=\mathbf{u}_1^k. \tag{5.43}$$

Therefore, in the asymmetric setup, the NP test (5.41), in which $\mathbf{y}_{q,b}^n$ is replaced by $\mathbf{y}^n$, reduces to

$$\sum_{i=1}^n(\hat{x}_{q,i}\oplus y_i)<\lambda_{qb}, \tag{5.44}$$

where $\hat{\mathbf{x}}_q^n=G_q\hat{\mathbf{z}}_1^m$, and λ_{qb} is an integer threshold.

5.6.3 Comparison with information-theoretic scheme

We now discuss the construction of this practical scheme compared to the information-theoretic one of [22]. As detailed in Section 5.6.2, the practical implementation of the quantization part closely follows the approach outlined in the information-theoretic proofs. For the binning part, the information-theoretic method involves random binning, where the quantized vectors are partitioned into bins randomly, with the number of bins being smaller than the number of quantized vectors. In contrast, our practical implementation uses syndrome-based binning. For each quantized vector, we compute a syndrome $\mathbf{u}^k$ using a parity check matrix of a linear code, as

formulated in (5.38). In addition, at the decoder side, the information-theoretic approach uses a minimal entropy check for decoding and decides $\mathcal{H}_0$ if the output sequence and the side information are jointly typical under $\mathcal{H}_0$. In our practical implementation, to decide between $\mathcal{H}_0$ or $\mathcal{H}_1$, we applied the NP test (5.41), which, by definition is known to be optimal in hypothesis testing problem [70]. Therefore, we do not lose any performance compared to the information-theoretic approach which could be more difficult to apply.

5.6.4 Theoretical analysis of the quantize-binning scheme

The NP test for the quantize-binning scheme in the asymmetric setup is also given by (5.44). The binning allows us to leverage the side information vector $\mathbf{y}^n$ so as to further reduce the coding rate. However, it also introduces a binning error which can impact Type-I and Type-II error probabilities [2]. We now provide exact analytical expressions of Type-I and Type-II error probabilities for the quantize-binning scheme in the asymmetric setup.

We consider the decision region $\mathcal{C}_0^{(qb)}$ for the all-zero codeword of the quantize-binning scheme. Especially, a side information vector $\mathbf{y}^n$ belongs to $\mathcal{C}_0^{(qb)}$ if the solution of (5.43) for this vector is $\hat{\mathbf{z}}_q^m = \mathbf{0}^m$. We then define the set of integers $\{E_\nu^{(qb)}\}_{\nu\in[\![0,d_{\max}^{(qb)}]\!]}$, where $E_\nu^{(qb)}$ is the number of words $\mathbf{y}^n$ of Hamming weight ν that belong to the decision region $\mathcal{C}_0^{(qb)}$. We also define the set of integers $\{A_t^{(qb)}\}_{t\in[\![0,n]\!]}$, where $A_t^{(qb)}$ is the number of codewords $\mathbf{x}_q^n$ of Hamming weight t such that there exists $\mathbf{z}_q^m$ that satisfies $\mathbf{x}_q^n = G_q\mathbf{z}_q^n$, and $H_b\mathbf{z}_q^m = \mathbf{0}^k$. As a result, the set $\{A_t^{(qb)}\}_{t\in[\![0,n]\!]}$ is the code weight distribution of the concatenated code.

Proposition 5.4 *For the quantize-binning scheme and for a threshold value λ_{qb}, Type-I and Type-II error probabilities are given by*

$$\alpha_n^{(qb)} = 1 - \mathbb{P}_B(c_0) - \mathbb{P}_{\bar{B}}(c_0), \tag{5.45}$$

$$\beta_n^{(qb)} = \mathbb{P}_B(c_1) + \mathbb{P}_{\bar{B}}(c_1), \tag{5.46}$$

where

$$\mathbb{P}_B(\delta) = \sum_{\nu=0}^{\min(d_{\max}^{(qb)},\lambda_{qb})} \frac{E_\nu^{(qb)}}{\binom{n}{\nu}} \sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \Gamma_{\nu,j,\gamma}\delta^j(1-\delta)^{n-j}, \tag{5.47}$$

$$\mathbb{P}_{\bar{B}}(\delta) = \sum_{i=0}^{n}\left[\left(\sum_{\gamma=0}^{d_{max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \Gamma_{i,j,w}\delta^j(1-\delta)^{n-j}\right)\left(\sum_{t=1}^{n}\sum_{\nu=0}^{\lambda_{qb}} \frac{E_\nu^{(qb)}}{\binom{n}{\nu}} \frac{A_t^{(qb)}\Gamma_{i,\nu,t}}{\binom{n}{i}}\right)\right]. \tag{5.48}$$

Proof. We consider the all-zero codeword $\mathbf{x}_q^n = \mathbf{0}$. Under the hypothesis $\mathcal{H}_0$, we express

$$\alpha_n^{(qb)} = 1 - \mathbb{P}_0(\widehat{\mathcal{H}}_0, B) - \mathbb{P}_0(\widehat{\mathcal{H}}_0, \bar{B}). \tag{5.49}$$

In this expression, B is the event that the correct sequence $\hat{\mathbf{x}}_q = \mathbf{x}_q$ was retrieved at the decoder, while $\bar{B}$ is the event that an incorrect sequence $\hat{\mathbf{x}}_q \neq \mathbf{x}_q$ was output by the decoder. In addition, $\widehat{\mathcal{H}}_0$ is the event that hypothesis $\mathcal{H}_0$ was decided at the decoder. We further denote $\mathbb{P}_B(p_0) = \mathbb{P}_0(\widehat{\mathcal{H}}_0, B)$ and $\mathbb{P}_{\bar{B}}(p_0) = \mathbb{P}_0(\widehat{\mathcal{H}}_0, \bar{B})$. We then express

$$\mathbb{P}_B(p_0) = \sum_{\nu=0}^{n} \mathbb{P}_0(w(\mathbf{Y}^n) = \nu)\mathbb{P}_0(\widehat{\mathcal{H}}_0, B|w(\mathbf{Y}^n) = \nu) \tag{5.50}$$

$$= \sum_{\nu=0}^{\min(d_{\max}^{(qb)}, \lambda_{qb})} \mathbb{P}_0(w(\mathbf{Y}^n) = \nu)\frac{E_\nu^{(qb)}}{\binom{n}{\nu}}. \tag{5.51}$$

Next, by following the same steps as in the proof of Proposition 5.3, we show that

$$\mathbb{P}_0(w(\mathbf{Y}^n) = \nu) = \sum_{\gamma=0}^{d_{\max}^{(q)}} \frac{E_\gamma^{(q)}}{N_0^{(q)}} \sum_{j=0}^{n} \Gamma_{\nu,j,\gamma} c_0^j (1-c_0)^{n-j}, \tag{5.52}$$

which provides (5.47). We then write

$$\mathbb{P}_{\bar{B}}(p_0) = \sum_{i=0}^{n} \mathbb{P}_0(w(\mathbf{Y}^n) = i)\mathbb{P}_0(\widehat{\mathcal{H}}_0, \bar{B}|w(\mathbf{Y}^n) = i), \tag{5.53}$$

where $\mathbb{P}_0(w(\mathbf{Y}^n) = i)$ is given by (5.52). Next, we develop

$$\mathbb{P}_0(\widehat{\mathcal{H}}_0, \bar{B}|w(\mathbf{Y}^n) = i)$$

$$= \sum_{t=1}^{n} \sum_{\nu=0}^{\lambda_{qb}} \mathbb{P}_0(w(\hat{\mathbf{X}}_q^n) = t, d(\hat{\mathbf{X}}_q^n, \mathbf{Y}^n) = \nu|w(\mathbf{Y}^n) = i) \tag{5.54}$$

$$= \sum_{t=1}^{n} \sum_{\nu=0}^{\lambda_{qb}} \frac{E_\nu^{(qb)}}{\binom{n}{\nu}} \frac{A_t^{(qb)} \Gamma_{i,\nu,t}}{\binom{n}{i}}. \tag{5.55}$$

This provides the expression of $P_{\bar{B}}(c_0)$ in (5.48). We obtain the expression of $\beta_n^{(qb)}$ from the previous equations by noticing that $\beta_n^{(qb)} = \mathbb{P}_B(c_1) + \mathbb{P}_{\bar{B}}(c_1)$. This ends the proof.

These theoretical results are novel and have been derived only for the asymmetric setup so far, as extending them to the symmetric setup becomes highly complex and intractable. They also facilitate the optimization and comparison of the proposed practical schemes across a wide range of source and code parameters. For instance, optimizing $E_\gamma^{(q)}$, $E_\nu^{(qb)}$, and $A_t^{(qb)}$ may lead to an optimal quantize-binning scheme for DHT, which we leave for future work.

5.7 Numerical Results

We now compare the different proposed coding schemes. To do this, we rely on ROC curve to show the Type-II error probability versus the Type-I error probability by varying the decision threshold for each of the considered coding schemes.

5.7.1 Truncation versus quantization

Here, we evaluate the performance of the quantization scheme against the truncation scheme for both symmetric and asymmetric setups. We fix $c_0 = 0.1$, $c_1 = 0.35$ and $p_1 = 0.5$. For both schemes, we evaluate the Type-II error with respect to the Type-I error.

In the symmetric setup, for the quantization (as described in Section 5.5), each encoder utilizes the BCH $(31, 16)$-code with minimum distance $d_{\min} = 7$. As a result, after applying the quantizer, $m = 16$ bits are sent to the decoder. Each encoder transmits $\ell = 16$ bits in the truncation scheme. Therefore, we consider for comparison the truncation scheme with $\ell = 16$ bits. Figure 5.4 illustrates the results of Monte-Carlo simulations averaged over 10000 trials, showing that the quantization scheme outperforms the truncation scheme in the symmetric setup for the considered values of p_0.

Similarly, in the asymmetric setup, we consider the same BCH $(31, 16)$-code for the quantization and the same truncation scheme with $\ell = 16$ bits for comparison. The only difference is that in the asymmetric setup, the BCH code is only used to quantize the observation $\mathbf{x}^n$, while $\mathbf{y}^n$ is available as side information. In Fig. 5.5, the Monte-Carlo simulations, averaged over 10000 trials, also reveal that the quantization scheme outperforms the truncation scheme in the asymmetric setup for the considered values of p_0. In addition, we observe that the theoretical Type-I and Type-II error probabilities are closely consistent with the Monte Carlo results. This is because the error probability expressions take into account the considered code through the terms $E_\gamma^{(q)}$. As a result, the theoretical expressions are found to be relevant tools for the DHT code design.

5.7.2 Truncation versus quantize-binning

We now evaluate the performance of the quantize-binning scheme for both symmetric and asymmetric setups. We set $c_0 = 0.1$, $c_1 = 0.35$ and $p_1 = 0.5$.

In the asymmetric setup, for the quantize-binning scheme (as described in Section 5.6), we consider the BCH $(31, 16)$-code for the quantizers, and the Reed-Muller $(16, 5)$ code with $d_{\min} = 8$ for the binning part. As a result, each encoder send only $k = 8$ coded bits to the receiver. Therefore, for comparison, we consider the truncation scheme with $\ell = 8$. Fig. 5.6 shows the Monte-Carlo simulations, averaged over 10000 trials, for both schemes. We observe that the quantize-binning scheme performs better than the truncation scheme, despite the fact that both the quantization and binning process can introduce errors.

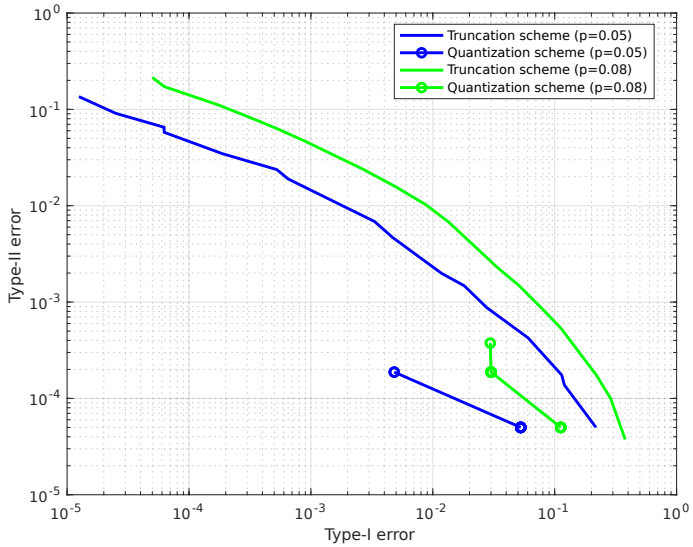

FIGURE 5.4 – ROC curve for the BCH code $(31, 16, 7)$ used as a quantizer, compared to the truncation scheme in the symmetric setup.

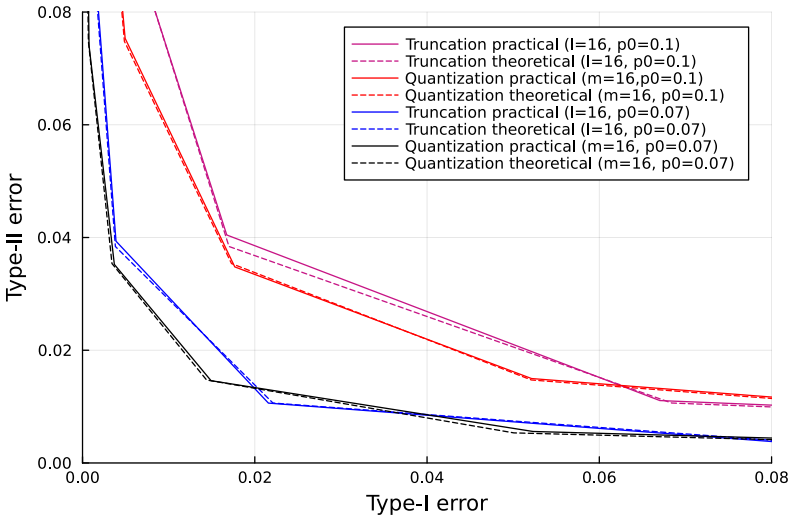

FIGURE 5.5 – ROC curve for the BCH code $(31, 16, 7)$ used as a quantizer, compared to the truncation scheme in the asymmetric setup. Source : © 2024 IEEE. Reproduced with permission from [3].

In the asymmetric setup, only the observation $\mathbf{x}^n$ is quantized and binned, while $\mathbf{y}^n$ serves as side information. We utilise the same BCH $(31, 16)$-code, Reed-Muller $(16, 5)$ for the quantize-binning and the same truncation scheme with $\ell = 8$ than the symmetric setup. Again, in Fig. 5.7, the Monte-Carlo simulations show superior performance of the quantize-binning scheme over the truncation scheme in the asymmetric setup. We also observe that the theoretical Type-I and Type-II error probabilities are closely consistent with the practical performance.

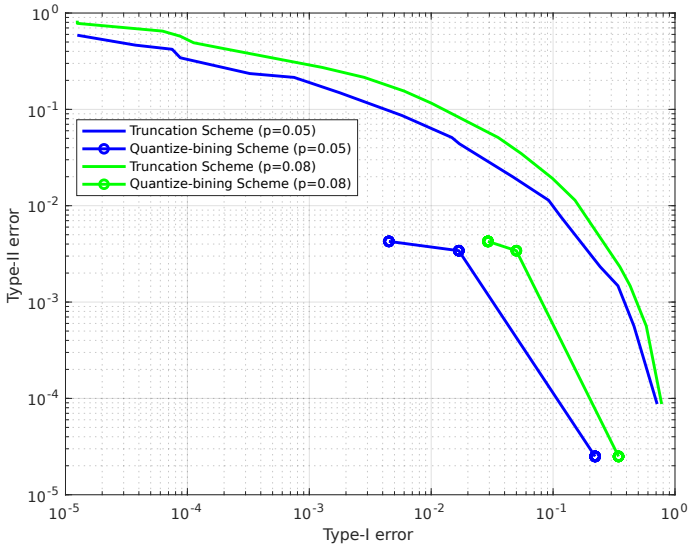

FIGURE 5.6 – ROC curve for the quantize-binning scheme built from the BCH code $(31, 16, 7)$ for quantization combined with the Reed-Muller code $(16, 5, 8)$ for binning in the symmetric case.

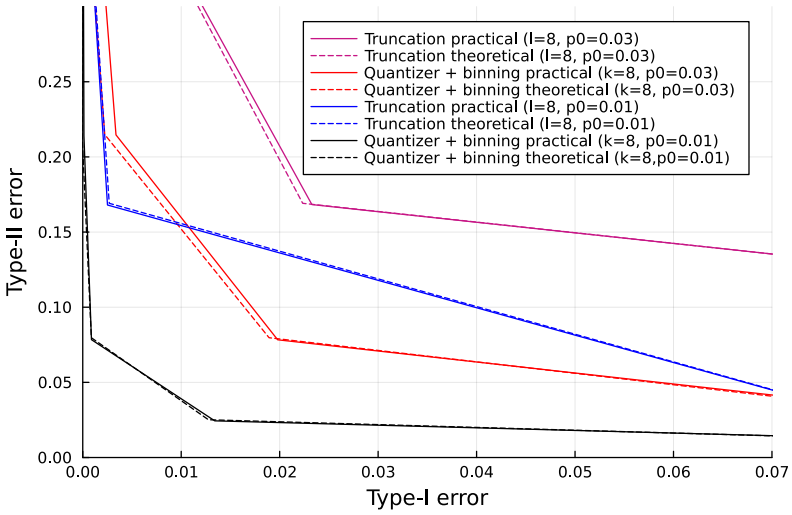

FIGURE 5.7 – ROC curve for the quantize-binning scheme built from the BCH code $(31, 16, 7)$ for quantization combined with the Reed-Muller code $(16, 5, 8)$ for binning in the asymmetric case. Source : © 2024 IEEE. Reproduced with permission from [3].

5.8 Summary and Discussion

In this chapter, we proposed practical short-length coding schemes for binary DHT. We first analyzed and compared two uncoded schemes for both symmetric and asymmetric setups of DHT. The first uncoded scheme, known as the separate scheme, involves each encoder independently

making local decisions based on their observations and transmitting a single bit to the receiver for the final decision. The second uncoded scheme, referred to as the truncation scheme, involves transmitting the first ℓ bits of the sources to be encoded. This scheme served as a performance baseline in our numerical results. The separate scheme is efficient in terms of transmission rate, sending only 1 bit, but its decision accuracy is compromised when p_0 increases. In contrast, the truncation scheme, while less efficient due to transmitting truncated data at a rate $R = \ell/n$, provides better decision accuracy through joint decoding when p_0 increases.

We then introduced two schemes using coding, one built with binary quantizer, and the other built as a quantize-binning scheme. Both schemes were designed from short linear block codes. For each considered scheme, in addition to practical constructions, we derived theoretical expressions of Type-I and Type-II error probabilities in the asymmetric case. Simulation results demonstrated the superiority of the proposed quantization and quantize-binning schemes compared to the baseline truncation scheme and also showed the accuracy of the proposed theoretical expressions. In addition, the performance of the quantize-binning scheme provides more gains in the symmetric setup compared to the asymmetric setup. This is a notable insight, as the symmetric setup has received relatively little attention in the DHT literature.

From a practical point of view, future works will include an interleaver design to improve the performance of the concatenated construction, as well as complexity reduction of the decoders so as to allow for larger code length to be considered. Another interesting analysis is to compare our proposed separate scheme in the symmetric case to the Watanabe's test [32] that is based on fixed-length coding.

Chapitre 6

Conclusion and perspectives

6.1 Conclusion

In this thesis, we first extended the study of DHT to more general source models, moving beyond the traditional i.i.d. assumptions. We analyzed the performance of DHT for these models and derived achievable error exponents. To achieve this, we proposed a coding scheme that provides a general lower bound on the error exponent. Unlike existing coding schemes for i.i.d. sources, which rely on the method of types, our proposed scheme is based on the information spectrum approach introduced by Han [42]. Notably, when applied to the specific case of i.i.d. sources, our general error exponent aligns with well-established results of [2].

We then demonstrated the applicability of our general analysis to various sources of interest, including stationary and ergodic Gaussian sources, and the Gilbert-Elliot (GE) sources model. Particularly, for the GE model, we introduced an efficient method to estimate the error exponent, which utilizes the forward recursion of HMMs. Our numerical results have evaluated the effects of the model parameters on the error exponent, as well as the tradeoff between the testing error and the binning error.

We then focused on the development of practical coding schemes. Specifically, we proposed short-length implementations of quantization and quantize-binning schemes, both built using linear block codes. For both schemes, we first addressed how to perform the hypothesis test in practical scenarios. In addition to practical constructions, we also derived theoretical expressions of Type-I and Type-II error probabilities for each proposed scheme. Numerical results showed that our practical implementations exhibited notable performance improvements compared to baseline uncoded schemes, where only a portion of the bits were transmitted without coding. While information-theoretic proofs have provided a basis for the development of our practical coding schemes, insights from the practical design could offer valuable guidance for future theoretical work on DHT, particularly in the symmetric setup.

6.2 Perspectives

6.2.1 Information spectrum method

In Chapter 3, we used the information spectrum approach to derive a generic bound on the error exponent. This bound is applicable to a broad range of source models, extending beyond the usual i.i.d. assumptions. However, when applied to the i.i.d. source model, our general error exponent does not achieve the optimal lower bound established by Shimokawa et al. [22]. Additionally, recent work by Kochman et al. [25] demonstrated an enhancement of the achievable error exponent presented by Shimokawa et al. As future work, we aim to incorporate the improvements from [25] into our coding scheme for general sources. Specifically, in the decoding process, we will explicitly include the minimal entropy check, as in [25], but adapted to general sources, which may lead to a tighter general error exponent.

The information spectrum approach is a powerful tool for developing a general theory of information, as emphasized by Han in his book [42]. Although this method has been used in hypothesis testing before, notably by Han [57], it was applied only to a single source, without considering distributed setups and coding. In [57], Han's approach relied on large deviation theory and on an information-spectrum slicing procedure, which partitions the acceptance region into subsets of equal width to bound the error exponent. As a perspective for future work, incorporating the information spectrum slicing method into the coding schemes for DHT is promising. It might simplify proofs and potentially yield tighter bounds on the error exponent.

6.2.2 Error exponent expressions for specific source models

In Chapter 4, we applied our general bound to two specific source models: the stationary and ergodic Gaussian sources model, and the GE sources model. For the stationary and ergodic Gaussian sources models, we derived closed-form expressions. Explicit expressions for the optimal error exponent in the case of vector Gaussian sources have also been found in works such as [60, 61, 62]. As a perspective for future work, our derived error exponents can be compared with those presented in these studies. Furthermore, in subspace techniques for array signal processing [59], a relevant question is the accuracy of the estimated signal subspace dimension, which inherently leads to a hypothesis testing problem. Our DHT framework for Gaussian sources could provide valuable insights in this context.

Regarding the GE model, we provided an efficient numerical method to evaluate the error exponent, using forward recursions proposed for Hidden Markov Models (HMM) in [63]. Future works will be dedicated to analyzing more generic Hidden Markov Models and Gauss Markov models.

6.2.3 Practical short-length coding schemes and theoretical analysis of short-length coding regime

The construction of efficient short-length linear block codes is often known to be a challenging problem [41]. Despite this, in Chapter 5, we introduced two coded schemes: one based on a binary quantizer and the other utilizing a quantize-binning scheme. Both schemes were constructed using short linear block codes. In addition to their practical implementations, we derived theoretical expressions for the Type-I and Type-II error probabilities in the asymmetric case.

Another coding scheme to explore is the truncation-binning scheme, where the first $\ell < n$ bits of the sequence $\mathbf{x}^n$ are selected, followed by applying a parity-check matrix from a linear block code for the binning step. This scheme can be compared to the quantize-binning scheme. The advantage of the truncation-binning scheme lies in the fact that the truncation process does not introduce any coding error, unlike the quantization step in the quantize-binning scheme. This makes truncation-binning a potentially efficient approach to explore. In addition, future works will include the design of interleavers to improve the performance of the concatenated construction. Another important research question is on the complexity reduction of the decoders to allow for larger code length to be considered. Another interesting study would be to compare our proposed separate scheme in the symmetric case to Watanabe's test [32] which is based on fixed-length coding.

Additionally, the theoretical expressions we derived for both the quantization and quantize-binning schemes are novel, providing a framework for optimizing and comparing these practical schemes across various source and code parameters. Future work could focus on optimizing the parameters such as $E_\gamma^{(q)}$, $E_\nu^{(qb)}$, and $A_t^{(qb)}$ to develop optimal quantization and quantize-binning schemes for DHT. For instance, the work in [73] demonstrates that optimizing $E_\gamma^{(q)}$ can lead to an optimal quantizer for DHT. Extending this optimization approach to the quantize-binning scheme could be promising. In addition, these theoretical expressions have only been derived for the asymmetric setup. Therefore, another interesting future work could be to derive these expressions for the symmetric setup, although this could lead to very complex expressions.

Since our work is just the beginning of designing short-length coding schemes for DHT, another interesting direction would be to explore the characterization of an optimal error exponent for finite-block length. This remains a challenging problem and an open question.

6.2.4 Universal coding schemes for Goal-oriented communication

Goal-oriented communication is an evolving field in modern communications, where data transmission is designed for specific tasks such as training machine learning models, decision-making, or semantic analysis [10, 11, 12]. In this thesis, we focused on the particular case of

decision-making, with a special emphasis on DHT. We provided both an information-theoretic framework and practical short-length coding schemes tailored to DHT.

Another interesting research direction is: can we design a universal coding scheme that would allow for different learning tasks to be performed on the same compressed data ? To address this, we first need to establish the theoretical limits of coding schemes designed for specific tasks as well as develop practical coding schemes tailored to those tasks. By analyzing the theoretical and practical performance of coding schemes for various tasks, we can potentially identify common features that could guide the development of universal coding schemes.

BIBLIOGRAPHIE

[1] I. S. Adamou, E. Dupraz, A. Zribi, and T. Matsumoto, "Error-exponent of distributed hypothesis testing for gilbert-elliot source models," in *2023 12th International Symposium on Topics in Coding (ISTC)*, 2023, pp. 1–5.

[2] G. Katz, P. Piantanida, and M. Debbah, "Distributed Binary Detection with Lossy Data Compression," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5207–5227, 2017.

[3] E. Dupraz, I. S. Adamou, R. Asvadi, and T. Matsumoto, "Practical short-length coding schemes for binary distributed hypothesis testing," *in Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2024.

[4] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–94, 2004.

[5] J. Barros and S. D. Servetto, "Network information flow with correlated sources," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 155–170, 2005.

[6] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication : Np-completeness and algorithms," *IEEE/ACM Transactions On Networking*, vol. 14, no. 1, pp. 41–54, 2006.

[7] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[8] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[9] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE transactions on information theory*, vol. 44, no. 6, pp. 2325–2383, 1998.

[10] E. C. Strinati and S. Barbarossa, "6g networks : Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.

[11] P. A. Stavrou and M. Kountouris, "A rate distortion approach to goal-oriented communication," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 590–595.

[12] H. Zou, C. Zhang, S. Lasaulce, L. Saludjian, and H. V. Poor, "Goal-oriented quantization : Analysis, design, and application to resource allocation," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 42–54, 2022.

[13] R. Piau, T. Maugey, and A. Roumy, "Learning on entropy coded images with cnn," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[14] A. Aliouat and E. Dupraz, "Learning on jpeg-ldpc compressed images : Classifying with syndromes," *arXiv preprint arXiv :2403.10202*, 2024.

[15] M. S. Rahman, "Distributed vector Gaussian source-coding and distributed hypothesis testing," no. January, 2012.

[16] M. Hamad, M. Wigger, and M. Sarkiss, "Multi-hop network with multiple decision centers under expected-rate constraints," *IEEE Transactions on Information Theory*, vol. 69, no. 7, pp. 4255–4283, 2023.

[17] T. Berger, "Decentralized estimation and decision theory," in *IEEE Seven Springs Workshop on Information Theory, Mt. Kisco, NY*, 1979.

[18] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, 1986.

[19] T. S. Han, "Hypothesis Testing with Multiterminal Data Compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, 1987.

[20] S. Amari *et al.*, "Statistical inference under multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, 1998.

[21] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE transactions on information theory*, vol. 33, no. 6, pp. 759–772, 1987.

[22] H. Shimokawa, S. Amari *et al.*, "Error bound of hypothesis testing with data compression," in *Proceedings of 1994 IEEE International Symposium on Information Theory*. IEEE, 1994, p. 114.

[23] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, 2012.

[24] S. Watanabe, "On sub-optimality of random binning for distributed hypothesis testing," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 2708–2713.

[25] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," *arXiv preprint arXiv :2306.14499*, 2023.

[26] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2044–2066, 2019.

[27] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over multi-access channels," in *IEEE Global Communications Conference (Globecom)*, 2018, pp. 1–6.

[28] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4411–4433, 2019.

[29] H. M. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, 1992.

[30] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 2006.

[31] S.-I. Amari and T. S. Han, "Statistical inference under multiterminal rate restrictions : A differential geometric approach," *IEEE Transactions on Information Theory*, vol. 35, no. 2, pp. 217–227, 1989.

[32] S. Watanabe, "Neyman–pearson test for zero-rate multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, 2017.

[33] E. Haim and Y. Kochman, "On Binary Distributed Hypothesis Testing," pp. 1–37, 2017. http://arxiv.org/abs/1801.00310

[34] S. Sreekumar and D. Gunduz, "Distributed Hypothesis Testing over Discrete Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2044–2066, 2020.

[35] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "On the necessity of binning for the distributed hypothesis testing problem," *IEEE Int. Symp. Inf. Theory - Proc.*, pp. 2797–2801, 2015.

[36] M. S. Rahman and A. B. Wagner, "Vector gaussian hypothesis testing and lossy one-helper problem," *IEEE Int. Symp. Inf. Theory - Proc.*, pp. 968–972, 2009.

[37] P. Escamilla, A. Zaidi, and M. Wigger, "Some Results on the Vector Gaussian Hypothesis Testing Problem," *IEEE Int. Symp. Inf. Theory - Proc.*, no. May, pp. 2421–2425, 2020.

[38] I. Csiszár, "The method of types [information theory]," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[39] J. Fridrich and T. Filler, "Binary quantization using belief propagation with decimation over factor graphs of ldgm codes," in *Proceedings of the 45th Allerton Conference on Coding, Communication, and Control*, 2007, pp. 495–501.

[40] M. J. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1061–1079, 2009.

[41] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient error-correcting codes in the short blocklength regime," *Physical Communication*, vol. 34, pp. 66–79, 2019.

[42] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, Ed., 2003. link:https://www.google.fr/books/edition/Information_Spectrum_Methods_in_Informat/LZBHu1Vrc6kC?hl=fr&gbpv=0

[43] V. Toto-Zarasoa, A. Roumy, and C. Guillemot, "Hidden markov model for distributed video coding," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 3709–3712.

[44] A. Bildea, O. Alphand, F. Rousseau, and A. Duda, "Link quality estimation with the gilbert-elliot model for wireless sensor networks," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 2049–2054.

[45] C. A. G. da Silva and C. M. Pedroso, "Packet loss characterization using cross layer information and hmm for wi-fi networks," *Sensors*, vol. 22, no. 22, p. 8592, 2022.

[46] W. Zhao and L. Lai, "Distributed testing against independence with conferencing encoders," *ITW 2015 - 2015 IEEE Inf. Theory Work.*, pp. 19–23, 2015.

[47] A. Zaidi, "Hypothesis Testing Against Independence Under Gaussian Noise," *IEEE Int. Symp. Inf. Theory - Proc.*, vol. 2020-June, pp. 1289–1294, 2020.

[48] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via körner-marton coding," in *2016 IEEE Information Theory Workshop (ITW)*. IEEE, 2016, pp. 146–150.

[49] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 983–987.

[50] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over noisy broadcast channels," *Information*, vol. 12, no. 7, p. 268, 2021.

[51] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," *IEEE Int. Symp. Inf. Theory - Proc.*, pp. 2840–2844, 2013.

[52] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing with general hypotheses," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 1705–1709.

[53] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detections," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 166–170.

[54] D. Cao, L. Zhou, and V. Y. Tan, "A strong converse theorem for hypothesis testing against independence over a two-hop network," *Entropy*, vol. 21, no. 12, p. 1171, 2019.

[55] M. Hamad, M. Wigger, and M. Sarkiss, "Two-hop network with multiple decision centers under expected-rate constraints," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[56] J. Liao, L. Sankar, V. Y. Tan, and F. du Pin Calmon, "Hypothesis testing under mutual information privacy constraints in the high privacy regime," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 1058–1071, 2017.

[57] T. S. Han, "Hypothesis testing with the general source," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2415–2427, 2000.

[58] K.-i. Iwata and J. Muramatsu, "An information-spectrum approach to rate-distortion function with side information," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 85, no. 6, pp. 1387–1395, 2002.

[59] R. J. Vaccaro, "The role of subspace estimation in array signal processing," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1566–1572.

[60] M. S. Rahman and A. B. Wagner, "Vector gaussian hypothesis testing and lossy one-helper problem," in *2009 IEEE International Symposium on Information Theory*. IEEE, 2009, pp. 968–972.

[61] A. Zaidi and I. E. Aguerri, "Optimal rate-exponent region for a class of hypothesis testing against conditional independence problems," in *2019 IEEE Information Theory Workshop (ITW)*. IEEE, 2019, pp. 1–5.

[62] P. Escamilla, A. Zaidi, and M. Wigger, "Some results on the vector gaussian hypothesis testing problem," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2421–2425.

[63] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[64] E. Dupraz, "Codage de sources avec information adjacente et connaissance incertaine des corrélations," Ph.D. dissertation, Université Paris Sud-Paris XI, 2013.

[65] M. Mushkin and I. Bar-David, "Capacity and coding for the gilbert-elliott channels," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1277–1290, 1989.

[66] S. Biswas, "Various proofs of the fundamental theorem of markov chains," *arXiv preprint arXiv :2204.00784*, 2022.

[67] I. S. Adamou, E. Dupraz, and T. Matsumoto, "An information-spectrum approach to distributed hypothesis testing for general sources," *arXiv preprint arXiv :2305.06887*, 2023.

[68] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 2006.

[69] M. Rezaeian, "Computation of capacity for gilbert-elliott channels, using a statistical method," in *2005 Australian Communications Theory Workshop*. IEEE, 2005, pp. 56–61.

[70] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing statistical hypotheses*. Springer, 2005, vol. 3.

[71] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[72] V. Chandar, E. Martinian, and G. W. Wornell, "Information embedding codes on graphs with iterative encoding and decoding," in *2006 IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 866–870.

[73] E. D. Fatemeh Khaledian, Reza Asvadi and T. Matsumoto, "Covering codes as near optimal quantizers for distributed hypothesis testing against independence," in *2024 IEEE Information Theorie Workshop (ITW)*. IEEE, 2024, pp. 1–6.

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Bornes sur l'exposant d'erreur et schémas de codage pratiques pour le test d'hypothèse distribué

Mots clés : théorie de l'information, test d'hypothèse distribué, spectre d'information, sources générales, exposant d'erreur, quantification, codes linéaires par bloc.

Résumé : Dans les réseaux de communication distribués, les données sont collectées, compressées, et transmises depuis des nœuds distants vers un serveur central pour un traitement ultérieur. Cependant, l'objectif du serveur n'est pas toujours de reconstruire les données originales, mais plutôt de prendre des décisions à partir des données codées reçues. Dans ce contexte, le Test d'Hypothèses Distribué se concentre sur le cas particulier de deux sources et vise à effectuer une prise de décision directement à partir des données compressées, sans passer par une reconstruction préalable. Comme dans le test d'hypothèses classique, deux types d'erreurs sont pris en compte pour évaluer les performances : l'erreur de Type I (fausse alarme) et l'erreur de Type II (décision manquée). Le Test d'Hypothèses Distribué prend en compte une contrainte de débit sur le lien de communication, et l'objectif est de concevoir un schéma de codage afin de maximiser la décroissance exponentielle, appelée exposant d'erreur, de la probabilité d'erreur de Type II, tout en maintenant la probabilité d'erreur de Type I en dessous d'un seuil spécifié. Dans la littérature, ce cadre a principalement été étudié sous un angle théorique de l'information, et la plupart des travaux existants analysent les performances des schémas du Test d'Hypothèses Distribué en supposant des sources indépendantes et identiquement distribuées (i.i.d.). Dans la première partie de cette thèse, nous abordons un modèle plus réaliste et général de sources non-i.i.d. Ce modèle englobe des sources non stationnaires et non ergodiques, reflétant mieux les scénarios réels par rapport au cas i.i.d. Nous dérivons des bornes génériques sur l'exposant d'erreur pour le Test d'Hypothèses Distribué à l'aide d'outils du spectre de l'information pour ce modèle général de sources. Nous montrons la cohérence de ces bornes avec le cas i.i.d. et les caractérisons plus précisément pour deux modèles spécifiques de sources : les sources gaussiennes non-i.i.d., et les sources de type Gilbert-Elliot. De plus, l'étude du Test d'Hypothèses Distribué ne se limite pas à l'analyse des limites théoriques de l'information, mais inclut aussi le développement de schémas de codage pratiques pour ce cadre. Ainsi, dans la deuxième partie de cette thèse, nous développons et implémentons des schémas de codage pratiques de courte longueur, spécialement conçus pour le Test d'Hypothèses Distribué, qui n'avaient pas encore été étudiés dans la littérature. Ces schémas de codage sont basés sur des codes linéaires en blocs et visent des longueurs très courtes, appropriées pour le Test d'Hypothèses Distribué (moins de 100 bits). En outre, nous fournissons des expressions analytiques exactes pour les probabilités d'erreurs de Type I et Type II pour chaque schéma de codage proposé, offrant ainsi des outils utiles pour la conception optimale future de codes DHT. Le travail réalisé dans cette thèse pourrait servir de base à l'investigation théorique et pratique de schémas de codage dédiés à des tâches d'apprentissage plus complexes, telles que la classification.

Title : Error Exponent Bounds and Practical Short-Length Coding Schemes for Distributed Hypothesis Testing (DHT)

Keywords : information theory, distributed hypothesis testing, information spectrum, general sources, error exponent, quantization, quantize-binning, linear block codes.

Abstract : In distributed communication networks, data is gathered, compressed, and transmitted from remote nodes to a central server for further processing. However, often, the objective of the server is not to reconstruct the original data, but rather to make decisions based on the received coded data. In this context, Distributed Hypothesis Testing (DHT) focuses on the particular case of two sources and addresses decision-making directly from compressed data without prior reconstruction.
As in conventional hypothesis testing, two types of errors are considered for performance evaluation: Type-I error (false alarm) and Type-II error (missed detection). DHT considers a rate-limited communication link, and the objective is to design a coding scheme so as to maximize the exponential decay, termed error exponent, of Type-II error probability, while keeping Type-I error probability below a specified threshold. In the literature, this setup was mostly investigated from an information-theoretic perspective, and most existing work analyze the performance of DHT schemes under the assumption of i.i.d. sources. In the first part of this PhD thesis, we address a more realistic and general model of non-i.i.d. sources. This model encompasses non-stationary and non-ergodic sources, and better reflects real-world scenarios compared to the i.i.d. case.
We derive generic error exponent DHT bounds using information spectrum tools for the considered general source model. We show the consistency of these bounds with the i.i.d. case and further characterize these bounds for two specific source models: non-i.i.d. Gaussian sources, and Gilbert-Elliot sources. In addition, addressing DHT requires not only the investigation of information-theoretic limits, but also the development of practical coding schemes for this setup. Therefore, in the second part of this thesis, we develop and implement practical short-length coding schemes specifically for DHT, which had not yet been investigated in the literature. These coding schemes are based on linear block codes, and they target very short length which are appropriate for DHT (less than 100 bits). Additionally, we provide tights analytical expressions for the Type-I and Type-II error probabilities for each proposed coding scheme, which provides useful tools for further optimal DHT code designs. The work carried out in this PhD may serve as a basis for the theoretical and practical investigation of coding schemes dedicated to more complex learning tasks such as classification.