UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : STITS

Laboratoire des Signaux et Systèmes (LSS)

DISCIPLINE : Physique

THÈSE DE DOCTORAT

soutenue le 03/12/2013

par

Elsa DUPRAZ

Codage de Sources avec Information Adjacente et Connaissance Incertaine des Corrélations

Directeur de thèse :	Michel Kieffer	Professeur Université Paris-Sud
Co-directeur de thèse :	Aline Roumy	Chargé de Recherche INRIA

Composition du jury :

Président du jury :	Enrico Magli	Associate Professor Politecnico di Torino
Rapporteurs :	Charly Poulliat	Professeur INP - ENSEEIHT
	Vladimir Stankovic	Senior Lecturer University of Strathclyde
Examinateurs :	Pierre Duhamel	Directeur de Recherche CNRS
	Claudio Weidmann	Maître de Conférence Université Cergy-Pontoise

La foi transporte les montagnes

Remerciements

Je souhaite commencer par remercier mes deux encadrants de thèse. L'histoire a commencé avec Michel Kieffer, qui avait su me convaincre, par son enthousiasme et ses qualités pédagogiques, de faire une thèse avec lui, puis a continué à s'écrire avec Aline Roumy, qui nous a rejoint petit à petit au cours de ma première année. J'ai particulièrement apprécié les nombreuses discussions téléphoniques que nous avons eu à cette époque, qui ont permis de brasser beaucoup d'idées et de comprendre beaucoup de choses. Cette habitude prise de discuter tous les trois s'est poursuivie plus tard, nottament lors de la rédaction de papiers, et nos nombreux débats sérieux et acharnés ont permis, je pense, de systématiquement améliorer la qualité de mes travaux de thèse. Je remercie également Aline pour son excellent accueil lors de mon séjour à Rennes, dont je garde vraiment un très bon souvenir, et pour les nombreuses conversations que nous avons eu par la suite.

Je voulais ensuite remercier les autres personnes avec qui j'ai eu l'occasion de travailler pendant la thèse, et tout d'abord Thomas Rodet, pour son aide sur les méthodes MCMC, pour sa patience et la qualité de ses explications sur la méthode qu'il m'avait proposé. Je commence par lui à cause d'un souvenir plus ancien : mon TER, qu'il avait encadré en fin de M1. C'est à cette occasion, à un moment où j'en doutais particulièrement, que je m'étais rendue compte que faire de la recherche pouvait me plaire. Je remercie ensuite Francesca Bassi, qui m'avait encadré pendant mon stage de M2, pendant lequel j'avais appris énormément de chose et beaucoup apprécié la richesse du sujet proposé. Francesca a continué à suivre épisodiquement mes travaux de thèse et les discussions avec elle ont toujours été une grande source de motivation. je remercie Valentin Savin, qui a toujours répondu à mes questions avec

beaucoup de précision, et dont la qualité des commentaires m'a beaucoup aidé à faire murir certains de mes travaux. Je remercie enfin Aurélia Fraysse, non seulement pour le travail que nous avons effectué ensemble, mais aussi pour son soutien au cours de la thèse.

Je tiens ensuite à remercier les membres de mon jury de thèse, à commencer par les deux rapporteurs, Charly Poulliat et Vladimir Stankovic, qui ont courageusement relu mon rapport et m'ont chacun apporté leur expertise dans leur domaine de compétence. Je remercie également Enrico Magli, qui a présidé mon jury avec beaucoup d'élégance et de pertinence. Je remercie enfin mes deux autres examinateurs, qui d'ailleurs ont d'ailleurs eu une importante particulière au cours de ma thèse. Pierre Duhamel a toujours encouragé les initiatives doctorantes au sein du laboratoire. Claudio Weidmann, qui suit mon travail depuis mon stage de M2 et avec qui j'ai eu l'occasion de discuter régulièrement, propose toujours une vision très intéréssante de son domaine de recherche.

Je remercie aussi les personnes avec qui j'ai enseigné à l'école Polytechique : Alain Louis-Joseph et Yvan Bonnassieux. J'ai toujours beaucoup apprécié les moments que j'y ai passé, et ces deux personnes y ont grandement contribué.

Je voulais ensuite dire un mot pour l'ensemble de mes collègues et amis doctorants : Benjamin, Olivier, Thang, Zeina, Amadou, Achal, Vineeth, Chengfang, Jean-François, Etienne, Leyla, Sarra, Jinane, Alex, Nabil, Axel, Adrien, Diane. Sans rentrer dans les détails de ceux que chacun d'eux m'a apporté à titre individuel, ils ont toujours representé un soutien important tout au long de mes trois années de thèse. Merci aussi à Maryvonne Giron, qui apporte beaucoup au labo par sa gentillesse et sa bienveillance. Je remercie également Lana, avec qui j'ai beaucoup apprécié d'être représentante des doctorants, ainsi que François, mon courageux camarade de M2 puis co-bureau, à qui je souhaite beaucoup de réussite pour la suite. Je remercie aussi José, pour les nombreuses heures que nous avons passé à discuter de tous les sujets possibles et inimaginables.

Je remercie les quelques personnes avec qui j'ai crée l'association des doctorants à Supélec : Mathieu, qui a apporté beaucoup pars son dynamisme, Daniel et son en-

thousiasme, Najett, avec qui j'ai beaucoup apprécié les interminables discussions que nous avons eu, Maud, pour ses nombreuses idées, Meryem, qui a su être à l'initiative du projet. Créer cette association a été un contre-poids appréciable à la solitude que l'on ressent parfois lors du travail de recherche. J'ai tiré beaucoup d'énergie de ce beau projet qui, je l'espère, continuera à se développer.

Enfin, je fais de grosses bises à la famille Dupraz : Alain, Claudette et Louise, qui représentent un soutien de poids dans les moments importants. Et pour finir, je remercie Pierre, pour tout ce que nous avons fait ensemble, pour son amitié et pour son amour qui m'ont souvent beaucoup portés.

Table des matières

Chapitre 1

Introduction

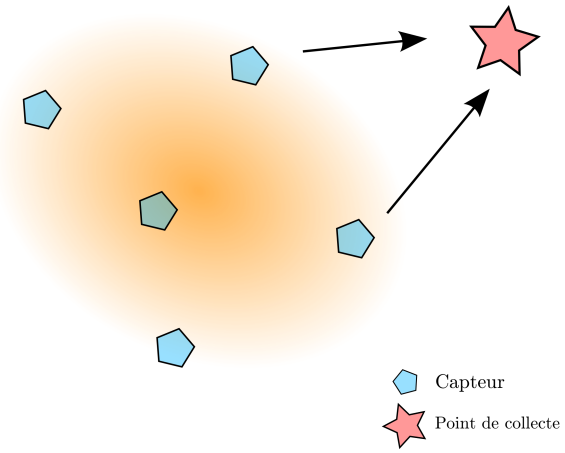

FIGURE 1.1 – Réseau de capteurs

Considérons un réseau de capteurs dans lequel les éléments de mesure prélèvent des mesures d'un phénomène physique (température, pression , hydrométrie, etc.) et doivent transmettre leurs observations à un point de collecte chargé de reconstituer l'ensemble des informations disponibles dans le réseau (voir la Figure 1.1) Il peut, par exemple, s'agir de prévenir les incendies dans une forêt en mesurant les températures à différents emplacements, ou d'effectuer des mesures pour du contrôle de trafic de véhicules.

Les capteurs sont des petits éléments qui doivent être autonomes et consommer aussi peu d'énergie que possible. L'échange d'informations dans le réseau se fait par des techniques de transmission sans fil, par exemple à l'aide de Zigbee [BPC+07].

FIGURE 1.2 – Codage de source avec information adjacente disponible au décodeur uniquement

Pour limiter la consommation d'énergie, il faut donc essayer de restreindre à la fois les opérations à effectuer par les capteurs, et les transmissions sur les liens sans fil. D'autre part, étant donné que les capteurs observent le même phénomène, leurs mesures sont vraisemblablement corrélées. Il devrait donc être possible d'exploiter cette corrélation pour compresser efficacement les données et ainsi diminuer la quantité d'information à transmettre dans le réseau. Cependant, pour les deux raisons évoquées précédemment, il n'est pas souhaitable que que les capteurs s'échangent leurs mesures respectives pour effectuer la compression. En revanche, on peut effectuer ce que l'on appelle du *codage de sources distribué* [BS06, CBLV05, XLC04] Avec ce type de codage, les capteurs peuvent compresser indépendamment leurs données au point de collecte, avec une performance égale de celle qu'ils auraient s'ils avaient communiqués entre eux. A charge ensuite pour le point de collecte qui reçoit toutes les données compressées de reconstruire l'ensemble des mesures relevées par les capteurs. L'un des objectifs de cette thèse est d'étudier ce problème de codage de sources distribué.

Ici, nous allons nous intéresser particulièrement au cas plus simple de deux capteurs X et Y. De plus, nous supposerons que les mesures de Y sont directement disponibles au au point de collecte et qu'il s'agit pour un capteur d'effectuer la compression de X sans connaître Y. Ce schéma de codage est représenté sur la Figure 1.2. On appelle X la *source* et Y *l'information adjacente*. Ce cadre simplifié est certes moins général, mais il constitue une première approche au problème de la compression distribuée. En particulier, nous n'aborderons pas ici les problèmes tels que le routage des informations dans le réseau, et nous focaliserons cette thèse sur les aspects liés

au codage de sources.

Pour ce problème, nous allons nous intéresser aux performances théoriques atteignables au sens de la théorie de l'information, et à la construction de schémas de codages performants. Dans le cas où la distribution de probabilité jointe $P(X, Y)$ est parfaitement connue, les performances théoriques ont été obtenues et des schémas pratiques efficaces ont été proposés, comme nous le verrons dans la suite [GD05, SW73]. Cependant, dans beaucoup de situations pratiques, la distribution jointe $P(X, Y)$ est souvent mal connue et peut même varier au cours du temps. Elle peut par exemple dépendre des conditions climatiques. De même, dans de nombreux cas, les mesures des capteurs sont fortement corrélées, mais un départ d'incendie va provoquer une brusque augmentation des températures dans une zone restreinte et seules les mesures d'un petit nombre de capteurs vont être influencées par l'événement. La corrélation entre les mesures de ces capteurs et les mesures des autres capteurs va donc fortement diminuer. De plus, si l'on n'autorise pas les communications entre capteurs, il est difficile pour eux d'apprendre les niveaux de dépendances entre les données qu'ils mesurent. Dans la suite, nous allons donc chercher à prendre en compte une possible mauvaise connaissance de la distribution de probabilité $P(X, Y)$. En particulier, nous nous interrogerons (i) la manière de décrire cette incertitude, (ii) l'étude des performances, (iii) la réalisation du décodage robuste à l'incertitude.

Dans cette thèse, nous introduirons donc tout d'abord des modèles de sources qui prennent en compte l'incertitude sur la distribution de probabilité. Pour cela, nous décrirons tout d'abord les performances et les schémas de codages utilisés dans le cas d'une distribution jointe bien connue (partie 2.1), Puis nous étudierons les différentes manières de prendre en compte cette incertitude (partie 2.2). Ensuite, nous décrirons les modèles considérés (partie 3.1) et nous considérerons une situation de codage sans pertes. Pour cette situation, nous effectuerons l'analyse de performance pour nos modèles de sources (partie 3.2) et présenterons les schémas de codage que nous avons proposés (partie 3.3). Les schémas de codage que nous proposons s'appuient tous sur des codes LDPC non binaires. Pour que ces schémas soient efficaces, il faut donc utiliser des codes LDPC performants. C'est pourquoi nous proposons une solution

de type évolution de densité pour réaliser la construction de bons codes LDPC non-binaires pour notre problème de codage (partie 3.4). Enfin, nous nous intéresserons à une situation de codage avec pertes. Nous considérerons cette fois un modèle de sources avec mémoire sur les symboles successifs. Pour ce modèle, nous effectuerons une analyse de performance et proposerons un schéma de codage pratique (partie 3.5). Enfin, les conclusions et perspectives de la thèse seront décrites dans la partie 4.

En annexe, nous fournissons également les articles soumis et rapports techniques en lien avec les contributions précédentes.

— L'annexe A correspond à l'article soumis suivant, qui traite de l'analyse de performance du schéma de codage sans pertes pour nos modèles de sources.

  Elsa Dupraz, Aline Roumy, Michel Kieffer *Coding strategies for source coding with side information and uncertain knowledge of the correlation* Submitted at IEEE Transactions on Information Theory, March 2013

— L'annexe B correspond à l'article soumis suivant, qui propose des schémas de codages sans pertes pour nos modèles de sources.

  Elsa Dupraz, Aline Roumy, Michel Kieffer *Source coding with side information at the decoder and uncertain knowledge of the correlation* Submitted at IEEE Transactions on Communications, February 2013, major revision

— L'annexe C correspond au rapport technique suivant, qui présente l'évolution de densité pour la construction de bons codes LDPC non binaires pour le problème de codage sans pertes.

  Elsa Dupraz, Aline Roumy, Michel Kieffer *Density Evolution for the Design of Non-Binary Low Density Parity Check Codes for Slepian-Wolf Coding* Technical report 2013

— L'annexe D correspond au rapport technique suivant, qui traite le cas de modèles avec mémoires.

  Elsa Dupraz, Francesca Bassi, Thomas Rodet, Michel Kieffer *Source Coding with Side Information at the Decoder and Uncertain Knowledge of the Correlation* Technical report 2013

Chapitre 2

État de l'art

Dans la suite, les variables aléatoires telles que X sont en majuscules, leurs réalisations x sont en minuscules et les vecteurs $\mathbf{X}^n$ sont en gras.

Nous rappellerons tout d'abord les résultats existants dans le cas où $P(X,Y)$ est bien connu et où les symboles de source sont indépendants et identiquement distribués ($i.i.d$) (Partie 2.1). Ensuite, nous supposerons que $P(X,Y)$ n'est pas bien connu, et nous nous interrogerons sur les manières de représenter et prendre en compte cette incertitude sur $P(X,Y)$ (Partie 2.2).

2.1 Sources *i.i.d.* et distribution jointe bien connue

Dans cette partie, la source X et l'information adjacente Y génèrent conjointement des suites de symboles $\{(X_n,Y_n)\}_{n=1}^{+\infty}$, *i.i.d.*, à valeurs dans des alphabets quelconques $\mathcal{X}$ et $\mathcal{Y}$, respectivement. La distribution de probabilité jointe $P(X,Y)$ est supposée parfaitement connue. Nous étudions tout d'abord les performances de codage de sources (Partie 2.1.1), puis décrivons les schémas de pratiques qui ont été proposés dans cette situation (Partie 2.1.2).

2.1.1 Performances de codage

Nous nous interrogeons ici sur la performance du schéma de codage avec information adjacente au décodeur seulement. Intuitivement, on pourrait se dire que ce

FIGURE 2.1 – Codage de la source X avec l'information adjacente Y disponible au codeur et au décodeur

schéma est beaucoup moins efficace en terme de compression que celui disposant de l'information adjacente également au codeur (voir Figure 2.1). Dans cette partie, nous comparons donc les performances de ces deux schémas de codage dans deux cas : le codage sans pertes et le codage avec pertes.

Codage de sources sans pertes

Dans cette partie, nous supposerons que les alphabets $\mathcal{X}$ et $\mathcal{Y}$ sont dénombrables. En codage de sources *sans pertes*, on veut reconstruire sans erreurs au décodeur la suite de symboles $\{X_n\}_{n=1}^{+\infty}$. Dans ce cas, on appelle *codage conditionnel* [Gucsel72] le schéma dans lequel Y est connu à la fois au codeur et au décodeur, et *codage de Slepian Wolf* (SW) [SW73] celui où Y n'est disponible qu'au décodeur. En réalité le codage de SW tel que défini dans [SW73] correspond au cas plus général de la compression distribuée de deux sources X et Y. Ici, par simplicité, l'appellation codage de SW désignera cependant le cas asymétrique, où seul X est compressé et où Y est directemment disponible au décodeur. On appelle $R^{\mathrm{c}}_{X|Y}$ (codage conditionnel) et $R^{\mathrm{SW}}_{X|Y}$ les débits minimums auxquels il est possible de coder X pour assurer une reconstruction sans pertes de la source. D'après [Gucsel72] et [SW73], ces débits sont donnés par

$$R^{\mathrm{c}}_{X|Y} = R^{\mathrm{SW}}_{X|Y} = H(X|Y) \tag{2.1}$$

où $H(X|Y)$ est l'entropie conditionnelle de X sachant Y. Il n'y a donc pas de perte de performance dans le cas où l'information adjacente est disponible uniquement au décodeur. Ce résultat étonnant a justifié l'intérêt pour le codage de sources avec information adjacente au décodeur et plus généralement pour le codage de sources

distribué.

Codage de sources avec pertes

En codage de sources *avec pertes*, on tolère un certain écart entre la source et sa reconstruction. L'écart moyen maximum est spécifié par une contrainte fixée de qualité de reconstruction appelée contrainte de distortion. Dans ce cas, on appelle *codage de Wyner-Ziv* (WZ) [WZ76] le schéma où Y est connu au décodeur seulement, et *codage conditionnel avec pertes* le cas où Y est connu au codeur et au décodeur.

Soit $d(.,.)$ une mesure de distortion et D la contrainte de distortion telle que $E[d(X,\hat{X})] \leq D$. D'après [WZ76], pour du codage de WZ, le débit minimum atteignable pour une contrainte de distortion D est

$$R^{\mathrm{WZ}}_{X|Y}(D) = \inf I(X;U|Y) \tag{2.2}$$

où le inf est sur l'ensemble des distributions de probabilité $f_{U|X}$ pour lesquelles $U \leftrightarrow X \leftrightarrow Y$ forment une chaîne de Markov et telles qu'il existe une fonction de reconstruction $F : \mathcal{U} \times \mathcal{Y} \to \mathcal{X}$ telle que $E\left[d(X,F(U,Y))\right] \leq D$. En codage conditionnel avec pertes, on a d'après [Gucsel72],

$$R^{\mathrm{c}}_{X|Y}(D) = \inf I(X;U|Y) \tag{2.3}$$

où le inf est sur les distributions de probabilité $f_{U|X,Y}$ telles que $E\left[d(X,U)\right] \leq D$.

Les fonctions $R^{\mathrm{WZ}}_{X|Y}(D)$ et $R^{\mathrm{c}}_{X|Y}(D)$ sont appelées *fonctions débit-distortion*. Étant donné que l'ensemble de minimisation pour le cas WZ est plus restreint que pour le cas conditionnel, [Zam96] montre que $R^{\mathrm{c}}_{X|Y}(D) \leq R^{\mathrm{WZ}}_{X|Y}(D)$, avec égalité seulement dans certains cas tels que le cas Gaussien. [Zam96] propose également une comparaison précise des performances. Il s'agit ensuite de déterminer l'expression explicite de ces fonctions débit-distortion, pour des modèles donnés de sources. Cela peut être un problème difficile, puisqu'il s'agit de faire une recherche sur les fonctions contenues dans les ensembles de minimisation de (2.2) et (2.3). Le cas Gaussien a été traité dans [Wyn78], et [BKW10] propose une analyse pour des mélanges de Gaussiennes.

De multiples variations du problème de codage de WZ ont été proposées et étudiées. Ainsi, [Gas04, GDV06, Ooh99] présentent le cas plus général de plusieurs

sources à compresser et [VP10] traite le problème de la sécurité dans un problème de codage avec information adjacente. Dans [FE06], deux informations adjacentes sont disponibles : l'une des deux est présente à la fois au codeur et au décodeur, l'autre n'est observée qu'au décodeur. Dans [MWZ08] le codeur reçoit de l'information sur la distortion entre la source et l'information adjacente présente au décodeur. De plus, la question des délais de décodage a également été abordée. En ce sens, [Ten04] s'intéresse à la construction de codeurs WZ temps réel, et [SP13, WG06] supposent que l'information adjacente arrive avec du retard au décodeur. Enfin, [Pra04, VP07] abordent ce problème des délais d'un point de vue légèrement différent. Dans ces travaux, lors du décodage d'un symbole X_n, le décodeur a accès à tous les symboles précédents $X_1 \dots X_{n-1}$.

2.1.2 Schémas de codage

Nous décrivons ici quelques solutions existantes pour le codage sans pertes puis pour le codage avec pertes.

Solutions pratiques pour le codage sans pertes

La plupart des schémas de codage sans pertes [GD05, XLC04] s'appuient sur des codes de canal tels que les turbo-codes [CPR03, SLXG04], et les codes LDPC (Low Density Parity Check codes) [EY05a, LXG02, MUM10a, WLZ08]. Le choix des codes LDPC présente plusieurs intérêts. Tout d'abord, [MUM10a] montre que les codes LDPC permettent d'atteindre l'entropie conditionnelle (2.1), à condition toutefois de construire le code avec précaution [CHJ09b]. Ensuite, de nombreux outils ont été développés pour les codes LDPC en codage de canal : algorithmes de décodage [CBMW10, CF02b, DM98, DF05, DF07, GCGD07, LGTD11, Sav08a, Wei08], analyse de performances [AK04,BB06,CRU01,FFR06,LFK09,Ric03,PRW06,WKP05, RSU01, RU01] et construction de codes [PFD08] notamment. En codage de SW, il est donc possible d'utiliser ces outils, soit directement, soit après un travail d'adaptation. Ces arguments s'appliqueraient également pour des turbo-codes, mais, comme

indiqué dans [XLC04], les méthodes de construction de codes performants sont plus simples et plus flexibles pour les codes LDPC que pour des turbo-codes.

Une majorité des schémas proposés précédemment [CF02a,CPR03,SLXG04,EY05a, LXG02,MUM10a] s'appuie sur des codes LDPC binaires. Or, dans un problème de codage de sources, les symboles à coder sont souvent non-binaires, comme par exemple les pixels d'une image ou les coefficients de leur transformée en cosinus discrète (DCT) ou en ondelette. Dans ce cas, une solution fréquemment employée est de transformer les symboles en plans de bits qui seront codés indépendamment à l'aide de codes LDPC binaires [EY05a, LXG02, VL12]. Il existe cependant une dépendance statistique entre les plans de bits, qui doit être prise en compte lors du décodage si l'on veut éviter une perte de performance [LW08, TZRG10a, VMFG07, VCFG08]. Mais les méthodes proposées pour tenir compte de cette dépendance induisent une complexité supplémentaire de décodage et, surtout, prennent difficilement en compte l'ensemble des dépendances. C'est pourquoi nous nous intéresserons ici à des codes LDPC travaillant directement avec des symboles non-binaires, c'est-à-dire dans GF(q), le corps de Galois de dimension q [MS77].

Concrètement, le codeur réalise le codage du vecteur de source $\mathbf{x}^n$ à l'aide d'une matrice creuse H de taille $n \times m$ ($m < n$) et dont les coefficients non nuls sont dans GF(q). Les proportions de coefficients non nuls par ligne et par colonne de H sont déterminées par les distributions de degrés $\lambda(x)$ et $\rho(x)$ [RSU01]. Le décodeur dispose du mot de code $\mathbf{u}^m = H^T\mathbf{x}^n$ et d'un vecteur d'information adjacente $\mathbf{y}^n$, à partir desquels il doit reconstruire le vecteur de source $\mathbf{x}^n$. En codage de canal, plusieurs algorithmes de décodage LDPC ont été introduits. Un aperçu de ces algorithmes est disponible dans [RU01]. On peut citer le décodage de type hard [LGTD11], le décodage avec des niveaux quantifiés [PDDV12], les algorithmes de type minimum-somme [DF05, Sav08a, Sav08c, CLXZS13] et les algorithmes de type somme-produit [DM98, GCGD07, Mac99]. Ces derniers, qui sont les plus performants, réalisent une estimation approchée au sens du maximum *a posteriori* et sont systématiquement utilisés en codage de SW. La performance de tous ces algorithmes dépend beaucoup des distributions des degrés $\lambda(x)$ et $\rho(x)$ du code. C'est pourquoi des méthodes appelées

évolution de densité [BB06, LFK09, WKP05, RSU01, RU01, SRA08, RU05, Sav08b] ont été introduites pour évaluer la performance asymptotique d'un code en fonction de ses distributions de degrés. Ensuite, une fois qu'une bonne distribution de degrés est obtenue, il s'agit de construire soigneusement la matrice H à longueur finie [PFD08, VCNP09].

Décodeur LDPC somme-produit

Les schémas de codage qui ont été proposés dans la thèse ont tous pour base des codes LDPC non binaires avec un décodage de type somme-produit. C'est pourquoi dans cette partie nous décrivons cet algorithme de décodage somme-produit.

A partir du vecteur de sources $\mathbf{x}^n$, on obtient donc le mot de code $\mathbf{u}^m = H^T\mathbf{x}^n$. On souhaite réaliser une estimation approchée de $\mathbf{x}^n$ à partir de $\mathbf{u}^m$ et $\mathbf{y}^n$ au sens du maximum *a posteriori* à l'aide d'un algorithme somme-produit [KFL01] que nous décrivons ici. Pour le cas non binaire et en codage de canal, cet algorithme est présenté dans [LFK09]. Nous en avons repris les notations. Nous exprimons simplement ici la version SW de cet algorithme. On peut représenter par un graphe $\mathcal{G}$ les dépendances entre les différentes variables aléatoires du problème (voir un exemple sur la Figure 2.2). Dans ce graphe, on appelle nœuds variables (NV) les sommets représentant les variables $X_1 \dots X_n$ et nœuds de parité (NP) les sommets représentant les composantes $U_1 \dots U_m$ du mot de code. Une arête relie un NV i à un NP j si et seulement si $H_{i,j} \neq 0$. On note $\mathcal{N}_{\mathrm{P}}(i)$ l'ensemble des NP connectés à un NV i et $\mathcal{N}_{\mathrm{V}}(j)$ l'ensemble des NV connectés à un NP j. La distribution de degrés des NV est donnée par $\lambda(x) = \sum_{i>1} \lambda_i x^{i-1}$, où λ_i représente la proportion d'arêtes de degré i, c'est-à-dire provenant de NV reliés à i NP. De même, la distribution de degrés des NP est donnée par $\rho(x) = \sum_{j>1} \lambda_j x^{j-1}$. Le débit d'un code de distributions de degrés $(\lambda(x), \rho(x))$ est donné par $r(\lambda, \rho) = \frac{M}{N} = \frac{\sum_{i\geq 2} \rho_i/i}{\sum_{i\geq 2} \lambda_i/i}$.

L'algorithme somme-produit est un algorithme de passage de messages dans $\mathcal{G}$. Les messages sont des vecteurs de dimension q calculés itérativement et initialisés à chaque NV i avec

$$m_k^{(0)}(i) = \log \frac{P(X_i = 0 | Y_i = y_i)}{P(X_i = k | Y_i = y_i)} \ . \tag{2.4}$$

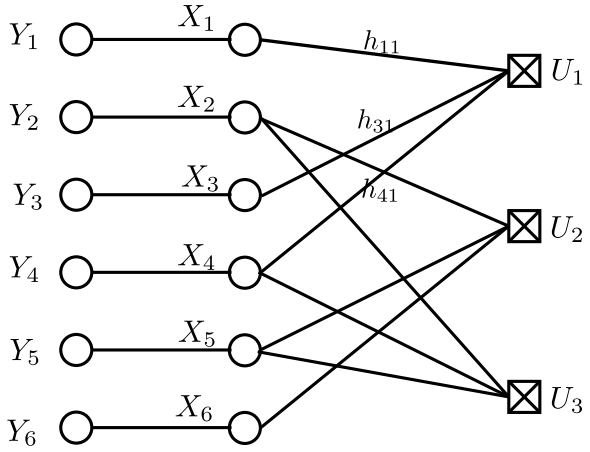

FIGURE 2.2 – Exemple de graphe de dépendances entre variables

À l'itération ℓ, on calcule les messages des NP j vers les NV i

$$\mathbf{m}^{(\ell)}(j,i) = \mathcal{A}[\overline{u}_j]\mathcal{F}^{-1}\left(\prod_{i' \in \mathcal{N}_{\mathrm{V}}(j)\backslash i} \mathcal{F}\left(W\left[\overline{H}_{i'j}\right]\mathbf{m}^{(\ell-1)}(i',j)\right)\right) \tag{2.5}$$

avec $\bar{s}_j = \ominus s_j \oslash H_{i,j}$, $\overline{H}_{i'j} = \ominus H_{i',j} \oslash H_{i,j}$ et $W[a]$ est une matrice $q \times q$ telle que $W[a]_{k,i} = \delta(a \otimes i \ominus k), \forall 0 \leq k, i \leq q-1$, où δ est la fonction de Kronecker. $\mathcal{A}[k]$ est une matrice $q \times q$ qui transforme un message $\mathbf{m}$ en un message $\mathbf{l} = \mathcal{A}[k]\mathbf{m}$ avec $l_j = m_{j \oplus k} - m_k$. $\mathcal{F}$ est une transformée de Fourier particulière dont la k-ème composante a pour expression

$$\mathcal{F}_k(\mathbf{m}) = \sum_{j=0}^{q-1} \frac{r^{k \otimes j} e^{-m_j}}{\sum_{j=0}^{q-1} e^{-m_j}} \tag{2.6}$$

où r est la racine de l'unité associée à $\mathrm{GF}(q)$. Ensuite, chaque NV i envoie un message $\mathbf{m}^{(\ell)}(i,j,y_i)$ aux NP j auquel il est relié et calcule un message *a posteriori* $\tilde{\mathbf{m}}^{(\ell)}(i,y_i)$. Ces messages sont donnés par

$$\mathbf{m}^{(\ell)}(i,j) = \sum_{j' \in \mathcal{N}_{\mathrm{P}}(i)\backslash j} \mathbf{m}^{(\ell)}(j',i) + \mathbf{m}^{(0)}(i) \ , \tag{2.7}$$

$$\tilde{\mathbf{m}}^{(\ell)}(i) = \sum_{j' \in \mathcal{N}_{\mathrm{P}}(j)} \mathbf{m}^{(\ell)}(j',i) + \mathbf{m}^{(0)}(i,y_i) \ . \tag{2.8}$$

A partir de (2.8), chaque NV i calcule une estimée x_i telle que

$$\widehat{x}_i^{(\ell)} = \arg \max_{k \in \mathrm{GF}(q)} \tilde{m}_k^{(\ell)}(i). \tag{2.9}$$

L'algorithme se termine si $\mathbf{u} = H^T\widehat{\mathbf{x}}^{(\ell)}$ ou si $l = L_{\max}$, le nombre maximum d'itérations. L'algorithme somme-produit est initialisé grâce aux probabilités conditionnelles $P(X|Y)$, ce qui nécessite une bonne connaissance de ces probabilités pour que l'algorithme donne des performances satisfaisantes.

Solutions pratiques pour le codage avec pertes

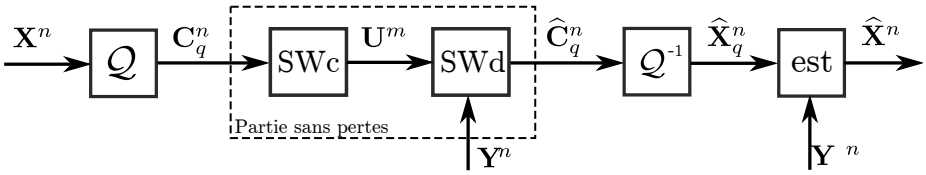

FIGURE 2.3 – Schéma de codage avec pertes

La figure 2.3 représente le schéma de codage de WZ classique, comme décrit dans [XLC04] et utilisé dans [AZG02, CMRTZ11, MWGL05, PR03, PR02, VAG06]. Ce schéma se compose d'une couche de quantification ($\mathcal{Q}$, $\mathcal{Q}^{-1}$), d'une chaîne de codage sans pertes (SWc, SWd), et de la reconstruction du vecteur de source en exploitant l'information adjacente (est). Différentes techniques ont ensuite été proposées pour chaque partie du schéma.

Tout d'abord, la quantification peut-être scalaire uniforme [RMRAG06], éventuellement précédée par un *dither* [Bas10, Section 1.3.2]. D'après [RMRAG06], par rapport à la borne de WZ (2.2), une quantification scalaire uniforme suivie d'une chaîne de SW idéale souffre d'une perte de distortion $1,53$ dB. Ce résultat s'étend à n'importe quel débit si on ajoute le dither [Bas10, Section 1.3.2]. La quantification peut aussi se faire avec des réseaux de points (*lattice*) en haute dimension [LCLX06, ZSE02], ce qui permet d'atteindre la borne de WZ mais est difficile à réaliser en pratique. Une dernière possibilité est d'utiliser une quantification en treillis (TCQ) [CGM03, YXZ09]. D'après les expériences réalisées dans [YXZ09], cette technique souffre d'une perte de seulement $0,5$ et 1 dB par rapport à la borne de WZ et est réalisable en pratique. La partie sans pertes est ensuite conçue comme on l'a décrit dans la 2.1.2, c'est-à-dire avec des codes LDPC ou avec des turbo-codes. Enfin,

pour la partie reconstruction, un choix optimal est d'utiliser un estimateur MMSE (minimum mean square error), qui minimise la distortion [BKW10]. Cependant, cet estimateur peut-être difficile à obtenir en expression explicite, et est parfois remplacé par un estimateur LMMSE (linear MMSE) [Zam96]. D'autres techniques telles qu'une simple interpolation [AZG02] peuvent aussi être employées.

2.2 Distribution de probabilité mal connue

Dans l'analyse de performance comme dans la description des schémas de codage, nous avons jusqu'à maintenant supposé que la distribution jointe $P(X, Y)$ était parfaitement connue. Nous nous intéressons maintenant aux situations où cette hypothèse n'est pas vérifiée. L'objectif ici est d'avoir un aperçu des différentes solutions qui permettent de prendre en compte l'incertitude sur la distribution de probabilité jointe.

En réalité, on peut aborder ce problème de deux manière distinctes. Une première possibilité est de travailler directement sur une modélisation des sources qui prendrait en compte l'incertitude. Pour cela, on peut par exemple choisir d'introduire des paramètres inconnus dans les distributions de probabilité. Une deuxième option est de travailler sur la formulation même du problème de codage. On peut ainsi s'interroger sur le cas de plusieurs décodeurs qui auraient accès à des informations adjacentes de qualité différente. Dans cette partie, nous présentons donc ces deux points de vue, les notions qui y sont rattachées, et les connections entre elles. Nous commençons par la modélisation des sources (Partie 2.2.1), qui semble le point de vue le plus naturel, puis nous abordons la question de la formulation des problèmes de codage (Partie 2.2.2).

Dans la suite, la source X et l'information adjacente Y produisent des couples de symboles $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ qui ne sont plus nécessairement indépendant ou identiquement distribués. En conséquence, on note $P(X_n, Y_n)$ la distribution jointe de (X_n, Y_n) et $P(\mathbf{X}^n, \mathbf{Y}^n)$ la distribution de probabilité de vecteurs de taille n.

2.2.1 Modélisation des sources

Nous nous intéresserons tout d'abord à deux notions que sont la stationnarité et l'ergodicité, qui permettent de classer les sources en différentes catégories, chacune avec leurs résultats particuliers. Nous commençons donc par rappeler leurs définitions et les résultats associés, avant de présenter deux modélisations particulières de sources : les sources générales et les sources variant arbitrairement. Les premières peuvent représenter un grand nombre de sources, pas nécessairement stationnaires et ergodiques, tandis que les deuxièmes ne sont ni stationnaires, ni ergodiques, et n'entrent même pas dans la catégorie des sources générales.

Dans la suite, les définitions sont données pour une source quelconque Z qui produit des symboles à valeurs dans un alphabet quelconque $\mathcal{Z}$. Les suites de symboles peuvent être données indifféremment par $\{Z_n\}_{n=1}^{+\infty}$ ou $\{Z_n\}_{n=-\infty}^{+\infty}$, suivant les besoins des définitions. Les résultats des définitions sont ensuite applicables directement au couple de sources (X, Y).

Stationnarité et Ergodicité

Nous rappelons ici les définitions de deux notions fréquemment employées de stationnarité et d'ergodicité. Ces définitions sont données dans [Gal68, Chapitre 3].

Définition. (Stationnarité) *Soit Z une source quelconque générant des suites de variables aléatoires $\{Z_n\}_{n=1}^{+\infty}$. Z est une source* stationnaire *si $\forall n, L \in \mathbb{N}$ et $\forall (z'_1, \ldots, z'_n) \in \mathcal{Z}^n$,*

$$P(Z_1 = z'_1, \ldots, Z_n = z'_n) = P(Z_{1+L} = z'_1, \ldots, Z_{n+L} = z'_n) \ . \tag{2.10}$$

D'après cette définition, une source est stationnaire si la probabilité d'un vecteur quelconque $(z'_1 \ldots z'_n)$ ne dépend pas de la position à laquelle on l'évalue. On dit que cette probabilité est indépendante de l'origine des temps. Les caractéristiques statistiques de la source ne varient pas dans le temps, et on peut donc appliquer le même traitement (estimation, compression, etc.) à chaque instant. Pour le cas particulier d'une chaîne de Markov, supposons que Z est une source discrète telle que $\forall n \in \mathbb{N}$, $P(Z_n | Z_{n-1} \ldots Z_1) = P(Z_n | Z_{n-1})$ et on note $P(Z_n = k | Z_{n-1} = j) = P^n_{j,k}$, avec

$0 < P^n_{j,k} < 1$. Si $P^n_{j,k}$ dépend de n, la source n'est pas stationnaire. Si $P^n_{j,k}$ ne dépend pas de n, on note $P^n_{j,k} = P_{j,k}$ et la source peut être, ou non, stationnaire, suivant les caractéristiques des $P_{j,k}$ et de la distribution de probabilités initiale $P(Z_0)$.

Définition. (Ergodicité) *Soit Z une source quelconque générant des suites de variables aléatoires $\{Z_n\}_{n=-\infty}^{+\infty}$. Soit f_n une fonction de la suite $\mathbf{z} = \{z_n\}_{n=-\infty}^{+\infty}$ qui ne dépend que des n composantes $z_1 \dots z_n$. On note T^ℓ un opérateur tel que $T^\ell \mathbf{z}$ représente la même suite décalée de ℓ symboles. La source Z est dite ergodique si pour toute fonction f_n intégrable,*

$$\lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} f_n(T^\ell \mathbf{z}) = E_{\mathbf{Z}}\left[f_n(\mathbf{Z})\right] \quad p.p. \tag{2.11}$$

(p.p. : presque partout).

D'après cette définition, une source est ergodique si ses caractéristiques statistiques (comme sa moyenne) sont indépendantes de la réalisation, c'est-à-dire de la suite $\mathbf{z}$ considérée. Si la source est ergodique, le même traitement peut être appliquée à toutes les suites $\{z_n\}_{n=1}^{+\infty}$ que la source pourrait générer.

Par exemple, supposons que les variables aléatoires de la suite $\{Z_n\}_{n=-\infty}^{+\infty}$ sont binaires, indépendantes, et que $P(Z_n = 1) = \theta$. Le paramètre θ est fixe pour la suite $\{Z_n\}_{n=-\infty}^{+\infty}$, il peut varier de suite en suite et correspond à la réalisation d'une variable aléatoire Θ de distribution de probabilité P_Θ. Une telle source n'est donc pas ergodique. En revanche, d'après [GD74], elle peut être décomposée en composantes ergodiques, c'est-à-dire en ensembles de probabilité non-nulle de suites obtenues à partir du même θ.

Pour finir, si on reprend les deux exemples fournis précédemment pour la stationnarité et l'ergodicité, on peut voir que la chaîne de Markov pour laquelle $P^n_{j,k}$ varie avec n est une source non stationnaire mais ergodique, et que la source binaire décrite par θ est une source non ergodique mais stationnaire.

Le débit minimum atteignable pour le problème de codage de SW pour une source ergodique est donné dans [Cov75], et les performances du schéma de codage de WZ peuvent également être obtenues relativement facilement. Cette analyse pour des

sources ergodiques est possible, puisque les caractéristiques statistiques des suites de symboles $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ sont toutes les mêmes. En revanche, il n'y a pas de résultats sur les performances pour des sources stationnaires, puisque la catégorie des sources stationnaires regroupe des situations très variées. De même, il n'existe pas de schémas de codages généraux qui s'appliqueraient de la même manière à toutes les sources stationnaires ou à toutes les sources ergodiques, même si ces deux notions peuvent aider à définir des principes de codage (même traitement sur toute la suite pour une source stationnaire par exemple).

Sources générales

Les sources générales sont présentées dans [Han03]. Lorsqu'elles ont été introduites, l'objectif était de proposer des outils d'analyses de performance qui puissent s'appliquer à n'importe quel type de source, pas nécessairement stationnaire ou ergodique.

Définition. (Source générale) *Une source générale Z produit des suites de vecteurs aléatoires $\left\{\mathbf{Z}^n = (Z_1^{(n)}, \ldots, Z_n^{(n)})\right\}_{n=1}^{+\infty}$. Les variables aléatoires $Z_k^{(n)}$ prennent leurs valeurs dans un alphabet arbitraire $\mathcal{Z}$ et les vecteurs aléatoires $\mathbf{Z}^n$ sont distribués suivant des distributions de probabilité $P(\mathbf{Z}^n)$ arbitraires.*

D'après l'expression de la suite $\left\{\mathbf{Z}^n = (Z_1^{(n)}, \ldots, Z_n^{(n)})\right\}_{n=1}^{+\infty}$, on voit que la source ainsi définie n'est pas nécessairement *consistante*. En effet, l'exposant (n) signifie que les réalisations de $\mathbf{Z}^n$ peuvent être différentes à chaque longueur n : les $n-1$ premières composantes de $\mathbf{z}^n$ peuvent être différentes de $\mathbf{z}^{n-1}$. Bien que le cas de sources non-consistantes soit rare, la formulation précédente des sources générales les prend en compte et une analyse de performance peut donc être obtenue dans ce cas. Dans le cas plus classique où la source est consistante, elle produit simplement des suites de variables aléatoires $\{Z_n\}_{n=1}^{+\infty}$.

Le débit atteignable et la fonction débit-distortion ont été obtenus pour des sources générales pour les problèmes de SW [Han03, Section 7.2] et de WZ [Iwa02, YZQ07]. Nous en rappelons ici les principaux résultats. Définissons tout d'abord la

lim sup en probabilités d'une suite quelconque de variables aléatoires $\{A_n\}_{n=1}^{+\infty}$ à valeurs dans $\mathbb{R}$ comme

$$\mathrm{P}-\limsup_{n\to\infty} A_n = \inf\left\{\alpha | \lim_{n\to\infty} P(A_n > \alpha) = 0\right\} . \tag{2.12}$$

On définit ensuite l'entropie spectrale conditionnelle [Han03, Section 7.2]

$$\overline{H}(\mathbf{X}|\mathbf{Y}) = \mathrm{P}-\limsup_{n\to\infty} \frac{1}{n}\log\frac{1}{P(\mathbf{X}^n|\mathbf{Y}^n)} \tag{2.13}$$

D'après [Han03, Section 7.2] le débit minimum atteignable pour le codage de SW pour une source générale est

$$R_{X|Y}^{\mathrm{SW}} = \overline{H}(\mathbf{X}|\mathbf{Y}) . \tag{2.14}$$

Pour obtenir la fonction débit-distortion du problème de WZ, on spécifie une suite de mesures de distortions normalisées $\left\{\frac{1}{n}d_n(\mathbf{X}^n, \widehat{\mathbf{X}}^n)\right\}_{n=1}^{+\infty}$ et on définit l'information mutuelle spectrale [Iwa02]

$$\overline{I}(\mathbf{X};\mathbf{U}|\mathbf{Y}) = \mathrm{P}-\limsup_{n\to\infty} \frac{1}{n}\log\frac{P(\mathbf{U}^n|\mathbf{X}^n,\mathbf{Y}^n)}{P(\mathbf{U}^n|\mathbf{Y}^n)} . \tag{2.15}$$

La fonction débit-distortion pour le problème de WZ pour une source générale est alors [Iwa02]

$$R_{X|Y}^{\mathrm{WZ}}(D) = \inf \overline{I}(\mathbf{X};\mathbf{U}|\mathbf{Y}) \tag{2.16}$$

où le inf est sur les suites de distributions de probabilités $\{f_{\mathbf{U}^n|\mathbf{X}^n}\}_{n=1}^{+\infty}$ pour lesquelles $\forall n \in \mathbb{N}$, $\mathbf{U}^n \leftrightarrow \mathbf{X}^n \leftrightarrow \mathbf{Y}^n$ forment une chaîne de Markov et telles qu'il existe une suite de fonctions de reconstruction $\{F_n\}_{n=1}^{+\infty}$ avec $F_n : \mathcal{U}^n \times \mathcal{Y}^n \to \mathcal{X}^n$ telles que $\mathrm{P}-\limsup_{n\to\infty} \frac{1}{n}d_n(\mathbf{X}^n, F_n(\mathbf{U}^n,\mathbf{Y}^n)) \leq D$.

La difficulté est ensuite d'obtenir les expressions explicites de (2.14) et de (2.16) pour des sources particulières que l'on souhaite étudier. Il n'existe bien entendu pas de schémas de codage pour des sources générales qui puissent être utilisés quelle que soit la source. D'autre part, nous verrons dans la suite que certains objets que l'on appelle habituellement "sources" n'entrent pas dans le cadre des sources générales et nécessitent des études particulières.

Sources variant arbitrairement

Les sources variant arbitrairement ont été introduites dans [Ahl79a, Ahl79b] et [Ber71]. Ces travaux sont assez anciens, mais les sources variant arbitrairement ont été récemment remises au goût du jour dans [TL11, SG12].

Définition. (Source variant arbitrairement) *Soit $\mathcal{P}_\pi$ un ensemble discret et $\mathcal{P} = \{P(Z|\pi), \pi \in \mathcal{P}_\pi\}$ un ensemble de distributions de probabilités conditionnelles. Une source Z variant arbitrairement produit des suites de variables aléatoires $\{Z_n\}_{n=1}^{+\infty}$ pour lesquelles la distribution de probabilité $P(\mathbf{Z}^n)$ d'un vecteur $\mathbf{Z}^n$ appartient à un ensemble $\mathcal{P}^n = \{P(\mathbf{Z}^n|\boldsymbol{\pi}^n), \boldsymbol{\pi}^n \in \mathcal{P}_\pi^n\}$ tel que*

$$P(\mathbf{Z}^n|\boldsymbol{\pi}^n) = \prod_{k=1}^{n} P(Z_n|\pi_n) \tag{2.17}$$

et $P(Z_n|\pi_n) \in \mathcal{P}$.

Une interprétation possible de cette définition est qu'il existe une suite d'états cachés $\{\pi_n\}_{n=1}^{+\infty}$ et que, à chaque instant, la distribution de probabilité de Z_n est donnée par la valeur de l'état caché π_n. Si les π_n sont connus, la source est stationnaire (mais pas nécessairement ergodique) puisque $P(Z_n = z|\pi_n)$ ne dépend pas de n. En revanche, si les π_n ne sont pas connus, on a une source non stationnaire, non ergodiques, et même pas générale au sens de [Han03]. En effet, avec une telle définition, on ne dispose d'aucune information sur le choix et les variations des π_n. Il peut s'agir de paramètres déterministes, ou de paramètres aléatoires dont la distribution serait inconnue. En particulier, si on note N_π^n le nombre d'apparitions de $\pi \in \mathcal{P}_\pi$ dans $(\pi_1 \dots \pi_n)$, les suites $\left\{\frac{\mathrm{N}_\pi^n}{n}\right\}_{n=1}^{+\infty}$ peuvent ne pas converger. De même, la distribution de probabilité $P(\mathbf{Z}^n)$ n'a pas de sens seule sans les π_n (en particulier, on ne peut pas marginaliser par rapport aux π_n). Pour ces raisons, la définition d'une source générale au sens de [Han03] ne s'applique pas ici. Pour s'en sortir, on pourrait dire qu'il ne s'agit pas d'une source, mais d'une combinaison de sources, et qu'à chaque instant, le symbole observé provient d'une de ces sources. La variation arbitraire serait ainsi sur la provenance des symboles de la suite $\{Z_n\}_{n=1}^{+\infty}$.

Pour le codage de SW, une analyse de performance est fournie dans [Ahl79a]. Si on suppose que les π_n sont disponibles au décodeur, on a

$$R^{\mathrm{SW}}_{X|Y,\pi} = \max_{\pi\in\mathcal{P}_\pi} H(X|Y,\pi)\ . \tag{2.18}$$

Si au contraire les π_n sont inconnus, d'après [Ahl79a],

$$R_{X|Y} = \sup_{P(X,Y)\in\mathrm{Conv}(\{P(X,Y|\pi)\}_{\pi\in\mathcal{P}_\pi})} H(X|Y) \tag{2.19}$$

où $\mathrm{Conv}(\{P(X,Y|\pi)\}_{\pi\in\mathcal{P}_\mathrm{D}})$ est l'enveloppe convexe des éléments de $\{P(X,Y|\pi)\}_{\pi\in\mathcal{P}_\pi}$. En revanche, aucun résultat n'a été obtenu pour le codage de WZ. La situation du codage avec pertes a bien été abordée dans [Ber71], mais pour un cas sans information adjacente.

En ce qui concerne les schémas pratiques, si les π_n sont connus au décodeur, on voit facilement que l'on peut réaliser la partie sans pertes avec un code LDPC classique. Au codeur, le débit du code est choisi en fonction de (2.18) (pire cas sur les π). Au décodeur, un algorithme de décodage somme-produit initialisé avec les $P(X_n|Y_n,\pi_n)$ est utilisé. Si les π_n ne sont pas connus au décodeur, nous ne connaissons pas de solution évidente ni de schéma de codage qui aurait été proposé. De même pour le cas avec pertes.

2.2.2 Problèmes de codage

Jusqu'à maintenant, nous avons défini les caractéristiques statistiques des sources et ensuite cherché à déterminer les débits atteignables en fonction de ces caractéristiques. L'incertitude était donc prise en compte directement dans les définitions des caractéristiques statistiques des sources.

Dans cette partie, nous allons étudier des problèmes de codage particuliers qui prennent directement en compte l'incertitude. Par exemple, pour le problème de codage de sources avec plusieurs décodeurs, chaque décodeur k a accès à une information adjacente différente $Y^{(k)}$. Les distributions jointes $P(X,Y^{(k)})$ sont cette fois supposées parfaitement connues, mais le codeur doit construire un seul mot de code qui permette de satisfaire l'ensemble des décodeurs. Ce schéma prend donc en compte une

possible incertitude directement dans la formulation du problème de codage. Dans cette partie, nous nous intéresserons également à deux autres formulations : le codage universel et le codage avec canal de retour.

Codage universel

En codage universel [CT06, Section 11.3], [Csi82, Dav73, BF12, MUM10b, Uye01b], on suppose que le débit R est fixé, et on se demande quelles sources, ou quel ensemble de sources, on peut coder avec ce débit. Les caractéristiques statistiques de ces sources ne sont pas nécessairement connues, et pour un code universel, les fonctions de codage et de décodage ne doivent pas dépendre de ces caractéristiques statistiques. D'après [Uye01b], on peut construire un code universel de débit R capable de réaliser le codage sans pertes de n'importe quelle source de distribution jointe $P(X, Y)$ telle que $R > H(X|Y)$. De plus, [MUM10b, SRA08] montrent que le codage universel peut être réalisé avec des codes LDPC. Le problème de codage universel de WZ est, lui, étudié dans [WK13]. En réalité, d'autres travaux tels que [JVW10, MZ06] proposent des schémas de codage de WZ dits "universels", mais dans ces travaux, seule $P(X)$ est inconnue.

On trouve aussi régulièrement un autre sens au codage universel. En codage de sources sans information adjacente, on dit qu'un code est universel [CT06, chapitre 13] si le débit de codage est donné par $H(X)$, la vraie entropie de la source, même si la distribution $P(X)$ est inconnue. En ce sens, les codes de Lempel-Ziv donc des codes universels. De tels codes ont été obtenus pour le codage conditionnel [UK03], mais il n'en existe pas ni pour le codage de SW, ni pour le codage de WZ. En effet, en codage de sources sans information adjacente ou en codage de sources conditionnel, le codeur s'appuie sur l'observation complète des suites de symboles pour construire le mot de code. Il peut par exemple apprendre les paramètres de la distribution de probabilité jointe et choisir le débit de codage en fonction de cet apprentissage. Une telle opération n'est pas possible en codage de SW et en codage de WZ, puisque le codeur à accès uniquement à $\{X_n\}_{n=1}^{+\infty}$. Pour bien différencier les deux situations, nous désignerons ce dernier cas sous le nom de *codage universel à débit variable*.

Codage avec plusieurs décodeurs

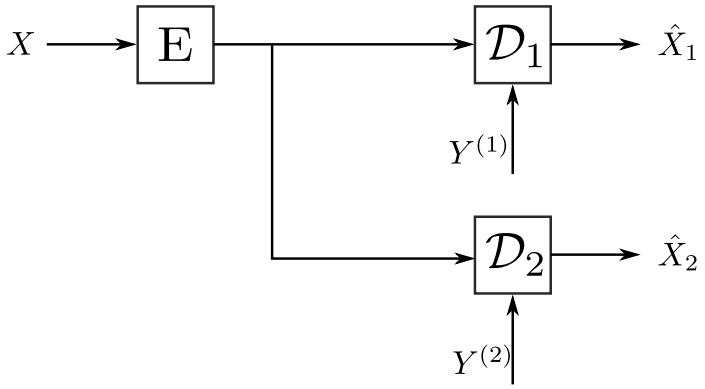

FIGURE 2.4 – Codage de sources avec plusieurs décodeurs

Le schéma de codage avec plusieurs décodeurs est représenté sur la Figure 2.4 (cas avec deux décodeurs). Le codeur doit produire un seul mot de code et satisfaire l'ensemble des décodeurs $\mathcal{D}_1$, $\mathcal{D}_2 \ldots \mathcal{D}_K$, qui observent chacun une information adjacente $Y^{(k)}$ de caractéristique statistique différente. Les distributions de probabilités jointes $P(X, Y^{(k)})$ (cas *i.i.d.*) sont supposées parfaitement connues. On peut donc considérer qu'il s'agit d'un problème dans lequel le codeur a une mauvaise connaissance de la distribution de probabilités jointe. Le codeur sait seulement que les distributions de probabilités appartiennent à l'ensemble $\{P(X, Y^{(k)}\}_{k=1}^{K}$ et doit choisir le débit et le mot de code en conséquence. Le décodeur k possède, lui, la connaissance complète de $P(X, Y^{(k)})$ et peut s'en servir pour reconstruire les symboles de source.

Dans le cas sans pertes, [Sga77], montre que le débit minimum atteignable est donné par

$$R^{\mathrm{p}}_{X|Y} \geq \max_{k \in \{1 \ldots K\}} H(X|Y^{(k)}) \ . \tag{2.20}$$

D'après ce résultat, on doit donc choisir le débit pour la pire qualité possible d'information adjacente. La réalisation d'un schéma de codage pour cette situation est relativement simple. Il suffit de choisir le débit pour le pire cas comme indiqué dans (2.20). Ensuite, le codage peut être réalisé avec un code LDPC, et le décodage se fait avec un algorithme somme-produit initialisé avec $P(X|Y^{(k)})$ connu au décodeur k.

Dans le cas avec pertes, un problème proche a été traité dans [HB85, Kas94]. Dans

ces travaux, le schéma de codage est constitué d'un seul codeur et de deux décodeurs ; l'un des codeurs a accès à l'information adjacente Y, l'autre décodeur n'a accès à aucune information adjacente. Les deux articles fournissent les fonctions débit-distortion pour cette situation et [RW11] propose un schéma de codage. Pour le cas plus général de plusieurs décodeurs avec des information adjacentes différentes, [KKU10] propose un schéma de codage mais ne donne pas les performances théoriques débit-distortion.

Codage avec canal de retour

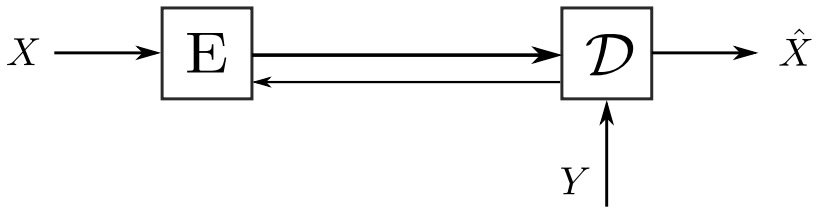

FIGURE 2.5 – Codage avec canal de retour

Nous nous intéressons ici au schéma de codage avec canal de retour, aussi appelé codage interactif dans [YH10] et représenté sur la Figure 2.5. Le terme interactif reflète en effet l'idée qu'il peut y avoir plusieurs échanges successifs par le lien direct et par le canal de retour avant de produire une reconstruction de la source. Le canal de retour doit donc permettre de minimiser la mauvaise influence de l'incertitude sur la distribution de probabilité.

Le cas sans pertes est présenté dans [YH10]. On considère une source (X, Y) stationnaire, non ergodique. D'après [GD74], toute source non ergodique peut être décomposée en composantes ergodiques, c'est-à-dire qu'il existe une variable aléatoire Θ à valeurs dans $\mathcal{P}_\Theta$ et distribué suivant P_Θ tel que $\forall n \in \mathbb{N}$,

$$P(\mathbf{X}^n, \mathbf{Y}^n) = \int_{\theta \in \mathcal{P}_\theta} P(\mathbf{X}^n, \mathbf{Y}^n | \Theta = \theta) P_\Theta(\theta) d\theta \ . \tag{2.21}$$

On note

$$H_\theta(\mathbf{X}|\mathbf{Y}) = \lim_{n \to \infty} \frac{1}{n} H(\mathbf{X}^n | \mathbf{Y}^n, \Theta = \theta), \tag{2.22}$$

dans lequel la limite existe puisque θ définit une composante ergodique de la source. D'après [YH10], avec un codage interactif, le débit minimum atteignable est donné

par

$$R^{\mathrm{f}}_{X|Y} = \tilde{H}(\mathbf{X}|\mathbf{Y}) := \int_{\theta\in\mathcal{P}_\theta} H_\theta(\mathbf{X}|\mathbf{Y}) P_\Theta(\theta) d\theta \ . \tag{2.23}$$

Cela correspond à la moyenne de l'entropie conditionnelle sur les différentes composantes ergodiques. Dans l'évaluation de $R^{\mathrm{f}}_{X|Y}$, le débit transitant sur le canal de retour est bien entendu pris en compte. En réalité, ce résultat montre qu'une suite de symboles $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ liée à la composante ergodique θ peut être transmise avec un débit $H_\theta(\mathbf{X}|\mathbf{Y})$. Le canal de retour permet de réaliser un codage universel à débit variable.

Ces résultats sont confirmés par les schémas de codage pratiques avec canal de retour [AZG02, EY05b, VAG06, VMFG07, VCFG08]. Dans ces articles, on utilise un code LDPC pour construire un mot de code $\mathbf{u}^m = H^T\mathbf{x}^n$. Le codeur commence par transmettre un morceau de $\mathbf{u}^{m_1}$ de $\mathbf{u}^m$ ($m_1 < m$) et le décodeur tente de reconstruire $\mathbf{x}^n$ à partir de $\mathbf{u}^{m_1}$ et $\mathbf{y}^n$. Si le décodage échoue, le décodeur le signale au codeur par l'intermédiaire du canal de retour, et le codeur transmet un deuxième morceau de mot de code. Le processus continue jusqu'à ce que le décodage réussisse. Toute la difficulté ici est de construire une matrice de codage H qui donne de bonnes performances de décodage même quand une partie seulement du mot de code est reçu. Pour cela, [HKKM07] propose une méthode de décodage adaptée à ces codes et [GSD10, KHM08, PY04, ZMZ+12] présentent et évaluent différentes méthodes de construction de codes. Aucune analyse de performance n'a été proposée dans le cas avec pertes, mais les schémas de codage [AZG02, VAG06, VMFG07, VCFG08] sont en réalité construits pour le cas avec pertes et exploitent le canal de retour uniquement pour la partie sans pertes. En revanche, [PL13] propose un schéma de codage pour lequel le canal de retour est utilisé à la fois pour la partie sans pertes et pour la partie avec pertes.

Toutefois, malgré ses avantages en terme de débit, une solution avec canal de retour présente deux inconvénients majeurs. Tout d'abord, le décodage peut prendre beaucoup de temps, puisque l'algorithme somme-produit est appliqué plusieurs fois. Ensuite, un canal de retour peut être difficile à mettre en place dans une situation pratique comme un réseau de capteurs. C'est pourquoi, dans la suite, nous ne

considérerons pas la possibilité de mettre en place un canal de retour.

Chapitre 3

Contributions

Dans les parties précédentes, nous avons décrit le problème de codage d'une source X avec information adjacente Y, puis présenté des analyses de performances et des schémas de codage dans le cas où la distribution de probabilité jointe $P(X, Y)$ était parfaitement connue. Nous avons ensuite passé en revue un certain nombre de manières de tenir compte d'une mauvaise connaissance de cette distribution de probabilité. Les notions précédentes telles que la stationnarité et l'ergodicité, les sources générales, le codage universel, permettent de caractériser et d'étudier les sources, mais il s'agit ensuite de spécifier des situations et des modèles d'intérêt, pour lesquels nous pourrons exploiter tous ces outils. Ici en particulier, les modèles que nous allons introduire et étudier permettent chacun de prendre en compte des dynamiques différentes de variation des caractéristiques statistiques des sources. Pour ces modèles, nous suivons les démarches précédentes et proposons des analyses de performance et des schémas de codage. Les contributions que nous décrivons dans cette partie sont donc les suivantes.

1. (Partie 3.1) Nous introduisons cinq modèles de sources de dynamiques différentes. Pour deux de ces modèles, les caractéristiques statistiques de sources varient lentement ; pour deux autres, elles peuvent varier très rapidement. Le dernier modèle présente un cas intermédiaire.
2. (Partie 3.2) Nous étudions les performances du schéma de codage de SW pour les quatre premiers modèles. En particulier, nous comparons ces performances

à celles obtenues dans le cas du codage conditionnel.

3. (Partie 3.3) Pour ces quatre mêmes modèles, nous proposons des schémas pratiques de codage de SW, capables de tenir compte de l'incertitude sur les caractéristiques statistiques des sources. Ces schémas s'appuient sur des codes LDPC non-binaires.
4. (Partie 3.4) Étant donné que pour être efficaces, les schémas de codage précédents doivent s'appuyer sur des codes LDPC performants, nous proposons une méthode de type évolution de densité pour déterminer des bonnes distributions de degrés pour des codes LDPC non-binaires pour du codage de SW.
5. (Partie 3.5) Nous étudions les performances du schéma de codage de WZ pour le dernier modèle de sources, et proposons un schéma de codage pratique pour ce modèle.

3.1 Modèles de sources

Cette partie présente les modèles de sources auxquels nous allons nous intéresser dans la suite. Quand on parle de distribution de probabilité $P(X, Y)$ mal connue, une première possibilité est de supposer que la distribution $P(X)$ de la source n'est pas bien caractérisée, mais que le *canal de corrélation* $P(Y|X)$ est parfaitement connu. Cependant même si $P(X)$ est complètement inconnu, [JVW10] montre que la fonction débit-distortion pour le codage de WZ est toujours donnée par (2.2). Ceci est confirmé par [TZRG10a, TZRG10b] qui proposent des schémas de codage qui ne souffrent pas de perte de performance comparé au cas où $P(X)$ est connu. Un tel résultat est finalement assez intuitif. En effet, la source X est complètement observée par le codeur, qui peut donc apprendre ses statistiques et les utiliser directement. C'est pourquoi dans la suite, nous supposerons au contraire que $P(X)$ est parfaitement déterminé, et que ce sont les caractéristiques du *canal de corrélation* $P(Y|X)$ qui sont mal connues.

De plus, dans beaucoup de problèmes, on peut supposer que le canal de corrélation possède une forme bien bien déterminée (Gaussienne, binaire symétrique, etc.). En

codage vidéo, par exemple, [BAP06] montre que le canal de corrélation peut être bien représenté par une distribution Laplacienne ; [MGPPG10] propose un modèle exponentiel un peu plus général. L'incertitude est alors sur les paramètres du modèle, comme la variance dans l'exemple précédent. Nous supposerons donc que $P(Y|X)$ appartient à une classe bien définie de modèles, mais que ses paramètres sont inconnus. Ici, nous allons nous intéresser aux dynamiques de variation de ces paramètres inconnus. On peut en effet supposer que ces paramètres sont fixés (voir partie 3.1.1), ou au contraire qu'ils peuvent varier dans le temps. Cette variation peut se faire sans mémoire (voir partie 3.1.2), ou avec mémoire (voir partie 3.1.3), c'est-à-dire que les paramètres à un instant donné dépendent des paramètres à l'instant précédent.

3.1.1 Paramètres fixés

Ici, nous supposerons que $P(Y|X)$ est paramétré par un vecteur $\boldsymbol{\theta}$ inconnu mais fixe pour la suite $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ et qui peut varier de suite en suite. Le vecteur $\boldsymbol{\theta}$ appartient à un ensemble de paramètres possibles $\mathcal{P}_\theta$ bien déterminé. On note $P(Y|X, \boldsymbol{\theta})$ la distribution de probabilité conditionnelle pour un $\boldsymbol{\theta}$ donné. On peut ensuite faire l'hypothèse que $\boldsymbol{\theta}$ est la réalisation d'une variable aléatoire $\boldsymbol{\Theta}$ distribuée suivant une distribution *a priori* $P_{\boldsymbol{\Theta}}$ connue. Dans ce cas, on appelle cette source une source *statique avec a priori*, ou *source SP*. Ce modèle tombe dans la catégorie des sources générales (voir partie 2.2.1). Si on retire cette hypothèse de distribution *a priori*, le modèle est en plus non stationnaire et ne peut pas être représenté par une source générale. On appelle cette source une source *statique sans a priori*, ou *source SwP*. En réalité, comme pour une source variant arbitrairement, une bonne interprétation serait de dire que le modèle SwP ne correspond pas à une source unique, mais à plusieurs sources que l'on doit être capables de coder.

Ces deux modèles sont non-ergodiques (voir partie 2.2.1), et seule la source SP est stationnaire. Avec ces modèles, on suppose que les variations des caractéristiques statistiques sont lentes. Cela correspond par exemple au cas d'un réseau de capteurs, dans lequel la corrélation entre les mesures effectuées varierait avec les conditions climatiques.

3.1.2 Paramètres variables sans mémoire

Pour deux autres modèles, on suppose que le canal de corrélation dépend d'un paramètre $\boldsymbol{\pi}_n$ inconnu et qui peut varier pour chaque couple (X_n, Y_n). Les $\boldsymbol{\pi}_n$ appartiennent à un ensemble $\mathcal{P}_{\boldsymbol{\pi}}$ bien déterminé. On note $P(Y_n|X_n, \boldsymbol{\pi}_n)$ la distribution conditionnelle pour un $\boldsymbol{\pi}_n$ donné. Si $\boldsymbol{\pi}_n$ est la réalisation d'une variable aléatoire $\boldsymbol{\Pi}_n$ de distribution de probabilité $P_{\boldsymbol{\Pi}}$ connue, le modèle est stationnaire et ergodique. Si ce n'est pas le cas, on a une source variant arbitrairement (voir partie 2.2.1). Nous appellerons les deux sources définies ici *source dynamique avec a priori*, ou *source DP*, et *source dynamique sans a priori*, ou *source DwP*. Avec ces modèles, on suppose que la corrélation entre les sources peut varier rapidement. Par exemple, en codage vidéo, des événement tels que l'apparition et la disparition objets peuvent faire varier le niveau de corrélation entre la source et l'information adjacente. La source DP produit en réalité une suite de symboles *i.i.d.*, et n'est donc pas difficile à étudier et à caractériser. Le modèle DwP, en revanche, correspond exactement à une source variant arbitrairement.

Dans le cas de caractéristiques statistiques mal connues, la distinction entre un paramètre fixe (slow fading) et un paramètre variable (fast fading) a été étudiée également en codage de canal [PMD09].

3.1.3 Paramètres variables avec mémoire

Une troisième possibilité consiste à supposer que les paramètres varient par blocs, comme cela a été proposé dans [CWC09]. Dans ce cas, la difficulté est alors de fixer la longueur des blocs. Pour éviter ce problème, nous allons considérer un modèle avec mémoire. On reprend le modèle DP et on suppose que la variation des paramètres inconnus $\boldsymbol{\Pi}_n$ est modélisée par une chaîne de Markov, c'est-à-dire que

$$P(\boldsymbol{\Pi}_n|\boldsymbol{\Pi}_{n-1}\dots\boldsymbol{\Pi}_1) = P(\boldsymbol{\Pi}_n|\boldsymbol{\Pi}_{n-1}) \ . \tag{3.1}$$

On fait également l'hypothèse que les $(Y_n|X_n, \boldsymbol{\Pi}_n)$ sont indépendants et on obtient ainsi un modèle de Markov caché [Rab89]. Avec ce modèle, on suppose bien que le canal de corrélation peut varier dans le temps, mais on suppose également que cette

variation a une mémoire : un ensemble de symboles faiblement corrélés peut être suivi par un ensemble de symboles fortement corrélés, par exemple.

3.2 Analyse de performance

Dans cette partie, nous proposons une analyse de performance pour le codage de SW pour les modèles SP, SwP, DP et DwP. L'article correspondant est disponible en Annexe A et contient également une définition formelle des modèles présentés dans la Partie 3.1. L'un des objectifs est en particulier de comparer les performances du schéma de codage de SW aux performances du schéma de codage conditionnel pour nos modèles de sources. En effet, intuitivement, dans le schéma de codage conditionnel, étant donné que la source X et l'information adjacente Y sont toutes les deux disponibles au codeur, il devrait être possible d'apprendre leurs caractéristiques statistiques et d'adapter le débit de codage à ces caractéristiques. D'autre part, les définitions particulières de nos sources induisent des problèmes particuliers de codage. Tout d'abord, on peut supposer que des estimées $\hat{\boldsymbol{\theta}}$ ou $\hat{\boldsymbol{\pi}}$ des paramètres sont disponibles au décodeur. Ensuite, étant donné que certains paramètres contenus dans $\mathcal{P}_{\boldsymbol{\theta}}$ peuvent réclamer un débit de codage important, on peut autoriser une *coupure* ou *outage*, c'est-à-dire que l'on accepte que le décodage puisse échouer pour une proportion γ des paramètres de $\mathcal{P}_{\boldsymbol{\theta}}$. Ces problèmes particuliers de codage ont déjà été abordés en codage de canal, voir [BRG02, CS99, Med00, PMD09, PSD09, SSZ02] pour les paramètres estimés, et [PMD09, PSD09] pour l'analyse outage. Nous les étudions ici pour le problème de codage de SW Nos contributions sont donc les suivantes :

1. (*Partie 3.2.1*) Pour les quatre modèles, nous effectuons une synthèse des résultats sur les débits atteignables en codage de conditionnel et en codage de SW. Certains des cas ont en effet déjà été étudiés par ailleurs avec des formulations différentes. Nous les rappellons ici et complétons avec les résultats manquants. A partir de cette synthèse, nous comparons les débits pour le codage conditionnel et pour le codage de SW.
2. (*Partie 3.2.2*) Nous étudions le schéma de codage de SW dans lequel des

estimées $\hat{\boldsymbol{\theta}}$ ou $\hat{\boldsymbol{\pi}}$ sont disponibles au décodeur et fournissons les débits atteignables pour ces situations.

3. (*Partie 3.2.3*) Nous proposons une formulation de l'analyse outage pour le codage de SW pour les modèles SP et SwP et fournissons les débits atteignables correspondants.
4. (*Partie 3.2.4*) Nous illustrons les résultats précédents dans le cas d'un réseau à 3 noeuds.

3.2.1 Codage conditionnel et codage de SW

Pour cette partie, les définitions complètes et les résultats détaillés sont disponibles en Annexe A, parties III et IV. Nous nous intéressons ici aux deux notions que sont le codage à longueur fixe et le codage à longueur variable [Han03]. En codage à longueur fixe, le débit est fixé avant le processus de codage et ne change pas, quelle que soit la suite $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ pour laquelle on doit réaliser le codage. A l'inverse, en codage à longueur variable, le débit peut varier de suite en suite, et donc éventuellement s'adapter aux caractéristiques statistiques de la suite $\{X_n, Y_n\}_{n=1}^{+\infty}$ courante.

Pour le codage à longueur fixe, les débits minima atteignables sont les mêmes pour le codage SW et pour le codage conditionnel. On a

1. pour la source SP [Han03, Théorème 7.3.4],

$$R_{X|Y}^{c,SP} = R_{X|Y}^{SW,SP} = P_{\boldsymbol{\Theta}}\text{-ess. sup}\, H(X|Y, \boldsymbol{\Theta}) \tag{3.2}$$

dans lequel

$$P_{\Theta}\text{-ess. sup}\, H(X|Y, \boldsymbol{\Theta}) = \inf_{\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}} \{\alpha | Pr(H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta}) > \alpha) = 0\} \tag{3.3}$$

2. pour la source SwP [Uye01a],

$$R_{X|Y}^{c,SwP} = R_{X|Y}^{SW,SwP} = \sup_{\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}} H(X|Y, \boldsymbol{\theta}) \tag{3.4}$$

3. pour la source DP [SW73],

$$R_{X|Y}^{c,DP} = R_{X|Y}^{SW,DP} = H(X|Y) \tag{3.5}$$

,

4. pour la source DwP [Ahl79a],

$$R_{X|Y}^{c,DwP} = R_{X|Y}^{SW,DwP} = \sup_{q \in \mathrm{Conv}(\{P(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}})} H(X|Y,q)\ . \tag{3.6}$$

On voit que ces résultats correspondent aux pires cas pour les paramètres et que la forme du pire cas varie suivant le modèle considéré. En codage à longueur variable, en revanche, il y a une différence de débit entre le cas SW et le cas conditionnel, sauf pour la source DP. En codage de SW à longueur variable, les débits minima atteignables sont les mêmes que précédemment. En revanche, pour le codage conditionnel à longueur variable, ces débits dépendent des valeurs des paramètres $\boldsymbol{\theta}$ ou $\{\boldsymbol{\pi}_n\}_{n=1}^{+\infty}$ en cours et on montre que

1. pour la source SP,

$$R_{X|Y}^{c,\mathrm{SP}}(\boldsymbol{\theta}) = H(X|Y,\boldsymbol{\Theta}=\boldsymbol{\theta}) \tag{3.7}$$

2. pour la source SwP,

$$R_{X|Y}^{c,\mathrm{SwP}}(\boldsymbol{\theta}) = H(X|Y,\boldsymbol{\theta}) \tag{3.8}$$

3. Pour la source DP,

$$R_{X|Y}^{c,\mathrm{SwP}} = H(X|Y) \tag{3.9}$$

4. pour la source DwP,

$$R_{X|Y}^{c,\mathrm{DwP}}(\{\boldsymbol{\pi}^n\}_{n=1}^{+\infty}) = H(\mathcal{X}|\mathcal{Y},\{\boldsymbol{\pi}^n\}_{n=1}^{+\infty}) \tag{3.10}$$

où

$$H(\mathcal{X}|\mathcal{Y},\{\boldsymbol{\pi}^n\}_{n=1}^{+\infty}) = \mathrm{P} - \limsup_{n\to\infty} -\frac{1}{n}\log\frac{P(\mathbf{X}^n,\mathbf{Y}^n|\{\boldsymbol{\pi}_k\}_{k=1}^n)}{\sum_{\mathbf{x}\in\mathcal{X}} P(\mathbf{x},\mathbf{Y}|\{\boldsymbol{\pi}_k\}_{k=1}^n)}. \tag{3.11}$$

Dans ces résultats, on retrouve l'idée intuitive de l'adaptation aux vraies caractéristiques statistiques de la source. Bien entendu, une telle adaptation n'est pas possible en codage de SW, qui souffre donc d'une perte de performance par rapport au codage conditionnel. Nous illustrerons les effets de cette perte en étudiant le cas d'un réseau à 3 nœuds dans la partie 3.2.4.

3.2.2 Paramètres estimés

Pour cette partie, les définitions complètes et les résultats détaillés sont disponibles en Annexe A, partie V. Pour les sources décrites précédemment, on étudie ici la situation où le décodeur aurait accès à des estimées $\widehat{\boldsymbol{\theta}}$ ou $\widehat{\boldsymbol{\pi}}_n$ des vrais paramètres. Ces estimées peuvent pas exemple avoir été obtenus grâce à une séquence d'apprentissage envoyée au préalable. Dans ce cas (pour le cas du codage de SW uniquement), on montre que :

1. pour la source SP,

$$R_{X|Y}^{SW,SP} = P_{\hat{\boldsymbol{\Theta}}}\text{-ess. sup } P_{\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}=\hat{\boldsymbol{\theta}}}\text{-ess. sup } H(X|Y,\boldsymbol{\Theta}) \tag{3.12}$$

2. pour la source SwP,

$$R_{X|Y}^{SW,SwP} = \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(X|Y,\boldsymbol{\theta}) \tag{3.13}$$

3. pour la source DP [SW73],

$$R_{X|Y}^{SW,DP} = H(X|Y,\hat{\boldsymbol{\Pi}}) \tag{3.14}$$

4. pour la source DwP [Ahl79a],

$$R_{X|Y}^{SW,DwP} = \sup_{p\in\text{Conv}(\{p(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}})} H(X|Y,\hat{\boldsymbol{\Pi}},p)\ . \tag{3.15}$$

Pour les modèles SP et SwP, on voit qu'il n'y a pas de gain en débit quand les paramètres estimés sont disponibles au décodeur. En effet, le codeur n'a pas accès aux paramètres estimés, et doit donc choisir un débit adapté au pire cas. En revanche, ces résultats ne traduisent pas le fait que qu'une estimée des paramètres disponibles au décodeur peut fournir des avantages pratiques comme la diminution du temps de décodage. A l'opposé, pour les modèles DPet DwP, la présence de paramètres estimés au décodeur permet bien de réduire le débit.

3.2.3 L'analyse outage

Pour cette partie, les définitions complètes et les résultats détaillés sont disponibles en Annexe A, partie VI. Nous nous intéressons ici seulement aux modèles SP et

SwP. Nous avons vu que pour le problème de codage de SW, les débits minimums atteignables étaient donnés par les pires cas sur les paramètres. Cependant, dans l'ensemble $\mathcal{P}_{\boldsymbol{\theta}}$, certains paramètres peuvent induire des débits importants. Ici, nous allons donc autoriser le décodeur a échouer pour une proportion γ des paramètres et exprimer les débits minima atteignables en prenant en compte cette contrainte. Pour la source SP, on note $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$ les ensembles tels que $P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}) \geq 1-\gamma$ et on montre que

$$R_{f,SW}^{SP}(\gamma) = \inf_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} P_{\boldsymbol{\Theta}|\boldsymbol{\Theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} - \text{ess. sup}\, H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\Theta}) \ . \tag{3.16}$$

Pour la source SwP, on note $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$ les ensembles tels que $\frac{\int_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} d\boldsymbol{\theta}}{\int_{\mathcal{P}_{\boldsymbol{\theta}}} d\boldsymbol{\theta}} \geq 1-\gamma$ et on montre que

$$R_{f,SW}^{SwP}(\gamma) = \inf_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} H(X|Y, \boldsymbol{\theta}). \tag{3.17}$$

On peut également formuler le problème autrement. On impose un débit R et on recherche la proportion γ de paramètres que l'on ne peut pas coder avec ce débit. Il s'agit d'une formulation proche du codage universel. On note $\overline{\mathcal{P}}_{\boldsymbol{\theta}}^{R} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$ l'ensemble des $\boldsymbol{\theta}$ pour lesquels $R > H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$ (source SP) ou $R > H(X|Y, \boldsymbol{\theta})$ (source SwP) et on montre que l'on a $\gamma_{f,SW}^{SP}(R) = 1 - P(\boldsymbol{\Theta} \in \overline{\mathcal{P}}_{\boldsymbol{\theta}}^{R})$.

3.2.4 Application au cas d'un réseau à trois nœuds

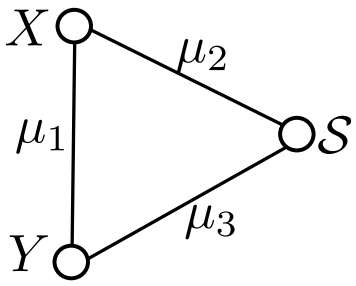

FIGURE 3.1 – Réseau à 3 nœuds

Pour cette partie, les résultats détaillés sont disponibles en Annexe A, partie VII. Nous présentons ici l'étude d'un réseau à trois nœuds pour la source SP. X, Y et $\mathcal{S}$ peuvent communiquer entre eux selon le schéma de communication représenté sur la

figure 3.1. Les deux sources X, Y doivent transmettre leurs informations au point de collecte $\mathcal{S}$ et on souhaite minimiser les coûts de transmission de ces informations. Les arêtes représentent les liens de communication entre les nœuds et sont associées à des poids μ_1, μ_2 et μ_3. Le coût de transmission à un débit R sur le lien i est $\mu_i R$. La détermination des μ_i résulte des problèmes de codage de canal associés au réseau. Nous ne les analysons pas ici. En revanche, nous nous intéressons au problème de codage des sources et évaluons le coût des transmissions point-à-point pour des μ_i fixés. Nous exprimons ces coûts pour différentes stratégies de codage, et proposons des outils de comparaison de ces stratégies.

Codage conjoint ou Codage séparé

Nous comparons plusieurs stratégies de codage. Avec la première stratégie de codage *conjoint*, X transmet $\{X_n\}_{n=1}^{+\infty}$ à Y et à $\mathcal{S}$, et Y transmet $\{Y_n\}_{n=1}^{+\infty}$ à $\mathcal{S}$ en utilisant $\{X_n\}_{n=1}^{+\infty}$ comme information adjacente disponible à la fois au codeur et au décodeur. Avec la deuxième stratégie de codage conjoint, les rôles de X et Y sont inversés. On note $m_c^{(X)}(\boldsymbol{\theta})$ et $m_c^{(Y)}(\boldsymbol{\theta})$ les coûts moyens par symbole correspondant aux deux stratégies pour un $\boldsymbol{\theta}$ donné. On suppose ensuite que $\mu_3 \geq \mu_2$, et on considère une seule stratégie de codage séparé. X transmet $\{X_n\}_{n=1}^{+\infty}$ uniquement à $\mathcal{S}$, et Y transmet $\{Y_n\}_{n=1}^{+\infty}$ à $\mathcal{S}$ en utilisant $\{X_n\}_{n=1}^{+\infty}$ comme information adjacente disponible seulement au décodeur. on note $m_s^{(X)}$ le coût moyen par symbole associé à cette stratégie. On a alors

$$m_{\mathrm{c}}^{(X)}(\boldsymbol{\theta}) = (\mu_1+\mu_2)H(X) + \mu_3 H(Y|X,\boldsymbol{\theta}) \tag{3.18}$$

$$m_{\mathrm{c}}^{(Y)}(\boldsymbol{\theta}) = (\mu_1+\mu_3)H(Y|\boldsymbol{\theta}) + \mu_2 H(X|Y,\boldsymbol{\theta}) \tag{3.19}$$

$$m_{\mathrm{s}}^{(X)} = \mu_2 H(X) + \mu_3 \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta})\ . \tag{3.20}$$

Pour comparer les deux stratégies de codage conjoint, considérons la différence de coût $\Delta_c(\boldsymbol{\theta}) = m_{\mathrm{c}}^{(X)}(\boldsymbol{\theta}) - m_{\mathrm{c}}^{(Y)}(\boldsymbol{\theta})$. Soit $\mathcal{P}_{\boldsymbol{\theta}}^c$ l'ensemble des $\boldsymbol{\theta}$ pour lesquels $\Delta_c(\boldsymbol{\theta}) \geq 0$ et soit $\gamma_c = 1 - \int_{\mathcal{P}_{\boldsymbol{\theta}}^c} d\boldsymbol{\theta} / \int_{\mathcal{P}_{\boldsymbol{\theta}}} d\boldsymbol{\theta}$ la mesure associée à $\mathcal{P}_{\boldsymbol{\theta}}^c$. On conserve la stratégie où Y est l'information adjacente si $\gamma_c > 0.5$, on conserve l'autre stratégie sinon. Ensuite, on compare la stratégie conservée à la stratégie de codage séparé. Si $\gamma_c > 0.5$, on

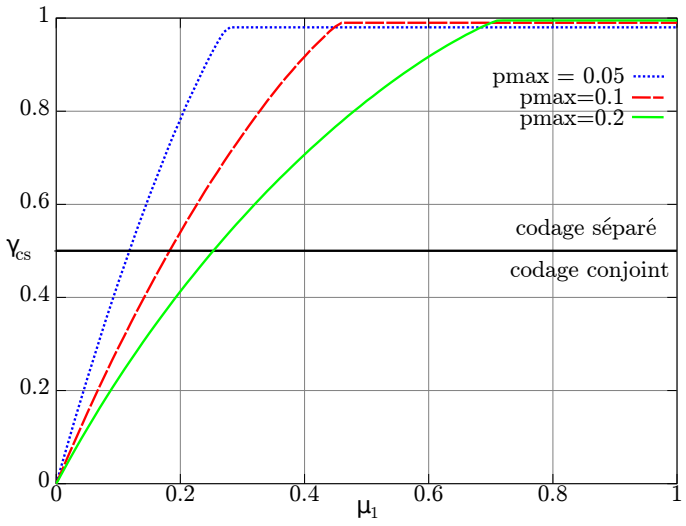

FIGURE 3.2 – Exemple pour des sources binaires. X est distribué uniformément, le canal de corrélation est un canal binaire symétrique de probabilité de transition p inconnu, $p \in [0, \text{pmax}]$. On fixe $\mu_2 = \mu_3 = 1$. On trace γ_{cs} en fonction de μ_1 pour comparer les stratégies codage séparé et codage conjoint quand X est l'information adjacente. Tant que $\gamma_{cs} > 0.5$, la stratégie codage séparé est conservée.

calcule $\Delta_{cs}(\boldsymbol{\theta}) = m_{\text{c}}^{(Y)}(\boldsymbol{\theta}) - m_{\text{s}}^{(X)}$, on définit $\mathcal{P}_{\boldsymbol{\theta}}^{cs}$ comme l'ensemble des $\boldsymbol{\theta}$ tels que $\Delta_{cs}(\boldsymbol{\theta}) \geq 0$ et on note γ_{cs} la mesure associée. On choisit l'une ou l'autre des stratégies en conséquence.

Un exemple de comparaison des stratégies de codage séparé et de codage conjoint est représenté sur la Figure 3.2. On peut voir que, en fonction de valeur de μ_1, qui représente en quelque sorte le coût de coopération entre les sources, on ne choisit pas toujours la même stratégie. Ainsi, une valeur faible de μ_1 signifie que les capteurs peuvent communiquer facilement entre eux, et donc la stratégie de codage conjoint sera conservée. Le choix de la stratégie dépend aussi du pire des paramètres possible. En effet, si ce pire paramètre est particulièrement défavorable, il induira un débit élevé.

Estimation de paramètres

Nous étudions ici l'intérêt pour les sources de transmettre au préalable des séquences d'apprentissage, pour permettre au décodeur d'estimer $\boldsymbol{\theta}$. Dans cette partie, nous supposons que la stratégie optimale est la solution séparée, qui est la solution privilégiée lorsque μ_1 est grand devant μ_2 et μ_3. On étudie deux stratégies possibles. Soit $\{u_N\}_{N\in\mathbb{N}}$ une suite d'entiers tels que $\lim_{N\to\infty} u_N = +\infty$ et $\lim_{N\to\infty} \frac{u_N}{N} = 0$. Pour la stratégie *avec* séquence d'apprentissage, X transmet $\{X_n\}_{n=1}^{N}$ à $\mathcal{S}$ à un débit $H(X)$, Y transmet une séquence d'apprentissage $\{Y_n\}_{n=1}^{u_N}$ de longueur u_N à $\mathcal{S}$ à un débit $H(Y)$ puis transmet les $\{Y_n\}_{n=u_N+1}^{N}$ restants en utilisant $\{Y_n\}_{n=u_N+1}^{N}$ comme information adjacente présente uniquement au décodeur. La stratégie *sans* séquence d'apprentissage correspond à la stratégie de codage séparé décrite dans la section précédente. Dans ce cas, le décodeur ne connaît pas $\boldsymbol{\theta}$, ce qui induit une difficulté supplémentaire lors du décodage qui doit être prise en compte. Lors de l'évaluation du coût, on ajoutera donc un terme $\alpha(\boldsymbol{\theta})$ représentant cette difficulté. Par exemple, si le décodeur doit à la fois reconstruire la source et estimer $\boldsymbol{\theta}$, le coût supplémentaire correspond à un temps de décodage plus grand, et $\alpha(\boldsymbol{\theta})$ est proportionnel à un temps.

On note $m_{\text{l}}(\boldsymbol{\theta})$ et $m_{\text{wl}}(\boldsymbol{\theta})$ les coûts moyens par symbole correspondant à chaque stratégie pour un $\boldsymbol{\theta}$ donné, et on a

$$m_{\text{l}}(\boldsymbol{\theta}) = \mu_2 H(X) + \mu_3 \left(\frac{u_N}{N} H(Y|\boldsymbol{\theta}) + \frac{N-u_N}{N} \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) \right) \quad (3.21)$$

$$m_{\text{wl}}(\boldsymbol{\theta}) = \mu_2 H(X) + \mu_3 \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) + \alpha(\boldsymbol{\theta}) \ . \quad (3.22)$$

Soit $\Delta(u_N,\theta)$ la perte en débit de la solution avec séquence d'apprentissage en fonction de u_N et de $\boldsymbol{\theta}$:

$$\Delta(u_N,\boldsymbol{\theta}) = \frac{m_{\text{l}} - m_{\text{wl}}}{\mu_3} = \frac{u_N}{N} \left(H(Y|\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) \right) - \frac{\alpha(\boldsymbol{\theta})}{\mu_3} \ . \quad (3.23)$$

Il s'agira ensuite de choisir une fonction α qui représente le fonctionnement du décodeur.

3.3 Schémas de codage de SW

Suite à l'analyse de performance effectuée précédemment, on souhaite obtenir des schémas pratiques de codage pour le problème de SW et pour les quatre modèles de source. L'article correspondant est disponible en Annexe B. La source et l'information adjacente prennent leurs valeurs dans GF(q), le corps de Galois de dimension q. Nous supposerons ici que le modèle est additif, c'est-à-dire qu'il existe une variable aléatoire Z indépendante de X et telle que $Y = X \oplus Z$. La même étude peut être réalisée pour le modèle inversé, c'est-à-dire pour lequel il existe une variable aléatoire Z indépendante de Y et telle que $X = Y \oplus Z$.

Le schéma proposé s'appuie sur des codes LDPC non binaires. Pour ce problème, on pourrait choisir d'utiliser des algorithmes de décodage LDPC de type hard [LGTD11, RSU01], ou min-sum [CF02b, Sav08a], qui ne nécessitent pas la connaissance de la distribution de probabilité de la source. Mais le premier souffre d'une perte de performance importante par rapport à l'algorithme de décodage somme-produit, et le second ne peut être exprimé pour notre cas, sauf si la source X est distribuée uniformément. Nous utiliserons donc un algorithme de décodage de type somme-produit.

D'autres travaux ont proposés un schéma de codage s'appuyant sur des codes LDPC avec décodage somme-produit, dans le cas où le canal de corrélation n'est pas bien connu. Pour une source binaire, le cas d'un canal de corrélation binaire symétrique de probabilité de transition $p = P(Y = 1|X = 0) = P(Y = 0|X = 1)$ inconnue a été traité dans [CWC09, CWC12, TZRG11, ZRS07]. D'autres travaux se sont intéressés au cas d'une source binaire et d'une information adjacente continue : le canal de corrélation est alors additif et Gaussien [SSWC11] ou Laplacien [BAP06, MWGL05, VCFG08, WCS$^+$12], et le paramètre inconnu est la variance du bruit de corrélation. Dans certains cas [TZRG11, VCFG08, ZRS07], le paramètre est fixe pour la suite de symboles $\{(X_n, Y_n)\}_{n=1}^{+\infty}$. Dans les autres, il peut varier sur des blocs de longueur fixée.

Dans notre cas non-binaire, pour le modèle DP, le décodeur LDPC somme-produit peut être utilisé directement puisque les probabilités conditionnelles $P(X_n|Y_n)$ sont

connues. En revanche, pour les autres modèles, ces probabilités conditionnelles ne peuvent pas être exprimées directement, et l'algorithme somme-produit ne peut pas être initialisé correctement. Dans ce cas, le code LDPC peut être dimensionné en utilisant les résultats de la partie 3.2.1 et il reste à construire un algorithme de décodage LDPC qui fonctionne correctement malgré le manque d'information sur les probabilités conditionnelles. Nos contributions sont donc les suivantes :

1. Nous proposons des algorithmes de décodage pour les modèles SP, SwP et DwP. Ces algorithmes estiment conjointement le vecteur de source $\mathbf{X}^n$ et les paramètres inconnus à l'aide d'un algorithme de type Expectation-Maximization (EM) que nous explicitons pour nos cas.
2. Etant donné que l'algorithme EM est très sensible à son initialisation, nous proposons également une méthode pour initialiser l'algorithme correctement.

Nous décrivons l'algorithme de décodage pour le modèle SwP dans la partie 2.1.2. Les autres cas s'en déduisent assez directement. Nous présentons ensuite nos résultats de simulations dans la partie 3.3.2.

3.3.1 Schéma de codage pour le modèle SwP

Le codage LDPC est réalisé comme décrit dans la partie 2.1.2. Pour cette partie, les résultats complets et les détails des calculs sont disponibles en Annexe B, partie V. Le décodeur doit donc réaliser l'estimation conjointe du vecteur de source $\mathbf{x}^n$ et du vecteur de paramètres $\boldsymbol{\theta}$ à partir du mot de code $\mathbf{u}^m$ et du vecteur d'information adjacente $\mathbf{y}^n$. On utilise pour cela un algorithme EM. Pour le modèle additif décrit précédemment, on note $P(Z = k) = \theta_k$. L'algorithme EM est un algorithme itératif qui produit des estimées de $\mathbf{x}^n$ et de $\boldsymbol{\theta}$ à chaque itération ℓ.

On cherche tout d'abord à produire une première estimée $\hat{\boldsymbol{\theta}}^{(0)}$ de $\boldsymbol{\theta}$ pour initialiser l'algorithme EM. Pour cela, on montre que l'on peut initialiser l'algorithme avec le vecteur $\boldsymbol{\theta}$ qui maximise la fonction de log-vraisemblance

$$L(\boldsymbol{\theta}) = \sum_{m=1}^{M} \log \mathcal{F}_{u_m}^{-1} \left(\prod_{j=1}^{dc} \mathcal{F}(W[h_j^{(m)}]\boldsymbol{\theta}) \right) . \tag{3.24}$$

Les transformations W, $\mathcal{F}$ et $\mathcal{F}^{-1}$ ont été définies dans la partie 2.1.2.

Ensuite, à l'itération $\ell+1$, on montre que les équations de mise à jour de l'algorithme EM sont

$$\forall k \in \mathrm{GF}(q), \quad \theta_k^{(\ell+1)} = \frac{\sum_{n=1}^{N} P_{y_n \ominus k,n}^{(\ell)}}{\sum_{n=1}^{N} \sum_{k'=0}^{q-1} P_{y_n \ominus k',n}^{(\ell)}} \tag{3.25}$$

dans lequel $P_{k,n}^{(\ell)} = P(X_n = k | y_n, \mathbf{s}, \boldsymbol{\theta}^{(\ell)})$. Les probabilités $P_{k,n}^{(\ell)}$ correspondent à la sortie de l'algorithme somme-produit initialisé avec l'estimée $\boldsymbol{\theta}^{(\ell)}$ obtenue à l'itération précédente. Les $P_{k,n}^{(\ell)}$ permettent également de réaliser l'estimation au sens du maximum *a posteriori* du vecteur de source $\mathbf{x}^n$ et donc d'obtenir l'estimée $\mathbf{x}^{n,(\ell)}$ à l'itération ℓ.

3.3.2 Résultats de simulations

Nous décrivons ici une partie des résultats de simulations pour les sources SP et SwP. Ls résultats complets sont disponibles en Annexe B, partie VI. Les symboles de sources prennent leurs valeurs dans GF(4). X est distribué uniformément et on a $P(Z = k) = \theta_k$, inconnu. On choisit un code LDPC de distribution de degrés $\lambda(x) = x^2$ et $\rho(x) = 0.5038x^2 + 0.2383x^3 + 0.0035x^4 + 0.00354x^5 + 0.0033x^{10} + 0.1252x^{11} + 0.0256x^{12} + 0.0089x^{18} + 0.0260x^{19} + 0.0301x^{20}$ et de débit $R = 1.5$ bit/symbole (voir partie 3.4 pour la sélection des distributions de degrés). Pour la source SwP, on compare quatre situations de codage. Pour chacune des situations testées, on fait la moyenne du taux d'erreur sur 1000 vecteurs de dimension 10000. Pour chaque vecteur testé, on génère un $\boldsymbol{\theta}$ aléatoirement et uniformément dans l'ensemble des $\boldsymbol{\theta}$ tels que $\theta_0 > p$, et p est fixé. On fait varier p de 0.67 (entropie de 1.42 bit/symbole) à 0.71 (entropie de 1.33 bit/symbol). On choisit d'effectuer 20 itérations pour le décodeur LDPC et 3 itérations pour l'algorithme EM (si nécessaire).

La première situation de codage correspond au cas déterministe, c'est-à-dire que le vecteur de paramètres est fixé à $\boldsymbol{\theta} = [1-p, (1-p)/3, (1-p)/3, (1-p)/3]$. De plus, on donne ce vecteur de paramètres au décodeur. On teste ensuite le décodeur dans le cas où $\boldsymbol{\theta}$ est généré aléatoirement, mais donné au décodeur (situation genie-aided). Dans le troisième cas, on utilise un algorithme EM initialisé aléatoirement. Le

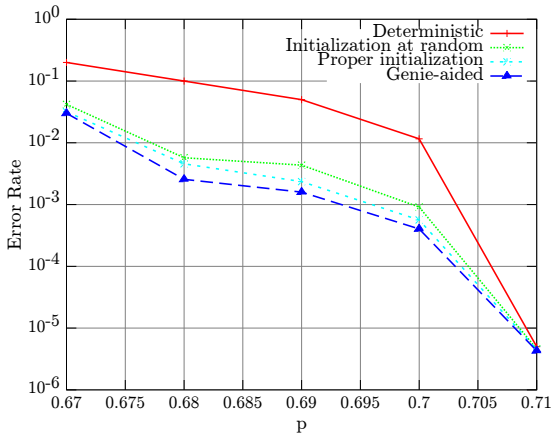

FIGURE 3.3 – Taux d'erreur pour la source SwP

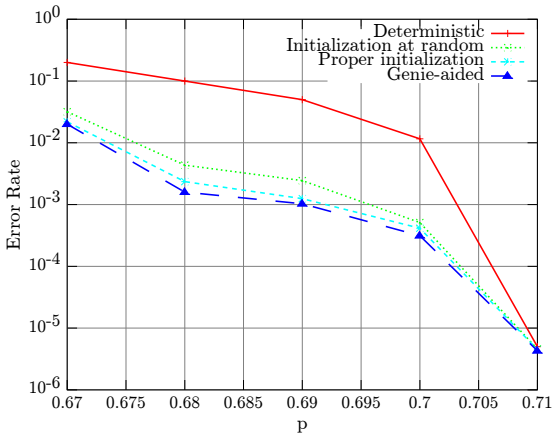

FIGURE 3.4 – Taux d'erreur pour la source SP

quatrième cas correspond à la méthode décrite précédemment, avec initialisation de l'algorithme EM. Les résultats sont représentés sur la Figure 3.3. Le cas déterministe donne des performances moins bonnes, puisque $\boldsymbol{\theta}$ est fixé au pire cas, alors que pour les trois autres situations, le vecteur de paramètres varie et peut donner des cas plus favorables. Ensuite, on voit que le taux d'erreur pour notre méthode est très proche de celui pour la situation genie-aided, et que l'initialisation aléatoire donne un taux d'erreur plus élevé. De plus, dans ce dernier cas, le temps de calcul est augmenté d'un facteur 1.5 environ.

Pour évaluer les performances dans le cas de la source SP, on choisit une distribution *a priori* sur θ_0. La distribution est triangulaire, centrée sur $p + 1/2(1-p)$. L'algorithme est évalué de la même manière et les résultats sont présentés sur la Figure 3.4. On obtient les mêmes conclusions que pour le cas de source SwP.

3.4 Design de codes LDPC non-binaires

Dans la partie précédente, nous avons décrit un schéma de codage qui s'appuie sur des codes LDPC non-binaires. Or, pour que le schéma proposé soit performant, il faut pouvoir construire des bons codes LDPC, c'est-à-dire des codes qui, à un débit proche de l'entropie conditionnelle, permettent d'assurer une probabilité d'erreur faible. Ici,

nous nous intéressons donc à la construction de bons codes LDPC non-binaires pour le problème de codage de SW. Dans cette partie, nous supposerons que les paramètres du canal de corrélation sont parfaitement connus. L'article correspondant est disponible en Annexe C.

Lorsque l'on cherche à construire des codes LDPC performants, une première étape consiste dans la recherche de bonnes distributions de degrés $(\lambda(x), \rho(x))$. Cela peut être effectué à partir de méthodes dites d'évolution de densité, introduites à l'origine dans [RSU01, RU01] pour des codes de canal binaires. A partir d'une analyse asymptotique, l'évolution de densité permet d'évaluer la probabilité d'erreur d'un (λ, ρ)-code pour un canal $P(W|U)$ donné. L'un des intérêts de cette méthode est que l'on montre que si le canal est symétrique, la probabilité d'erreur ne dépend pas du mot de code à l'entrée du canal. On peut donc supposer que le mot de code ne contenant que des zéros a été transmis, ce qui simplifie beaucoup les calculs.

Lorsque l'on veut effectuer une évolution de densité pour le codage de SW, une première idée serait d'identifier le canal de corrélation $P(Y|X)$ et de lui appliquer les techniques d'évolution de densité développées pour le problème de codage de canal [BB06, LFK09]. Malheureusement, comme expliqué dans [CHJ09b], un bon code LDPC pour un canal donné $P(W|U)$ en codage de canal n'est pas nécessairement un bon code pour un problème de codage de SW de canal de corrélation donné par $P(W|U)$. La raison principale est que, en codage de canal, pour un canal discret symétrique, on montre qu'on atteint la capacité de Shannon si les symboles d'entrée U sont distribués uniformément. A l'inverse, en codage de SW, on ne maîtrise pas la distribution des symboles de source, et cette distribution n'est donc pas nécessairement uniforme. En particulier, l'hypothèse du mot de code ne contenant que des zéros, très pratique en codage de canal, ne peut pas s'appliquer ici, à cause de la distribution possiblement non-uniforme de X.

Heureusement, [CHJ09b] montre que dans le cas binaire, pour tout canal de corrélation $P(Y|X)$, pas nécessairement symétrique, il existe un canal équivalent $P(W|U)$ symétrique et pour lequel on peut effectuer l'hypothèse du mot de code ne contenant que des zéros. Le terme équivalent signifie que l'évolution de densité est la

même pour les deux canaux, et que la probabilité d'erreur du code est également la même. De plus, $H(U|W) = H(X|Y)$. Ici, nous généralisons ce résultat à des canaux non-binaires. Les contributions de cette partie sont donc les suivantes

1. Pour le problème de codage de canal, nous obtenons une expression analytique récursive de l'évolution de densité pour un canal symétrique. Cette expression ne peut pas être utilisée pour réaliser l'évolution de densité en pratique, puisque qu'elle n'est pas explicite. En revanche, elle nous sera utile dans la suite pour exprimer l'équivalence.
2. Pour le problème de codage de SW, nous obtenons une expression analytique récursive de l'évolution de densité pour n'importe quel canal de corrélation. De même que pour le codage de canal, cette expression n'est pas utilisable en pratique.
3. A partir des deux récursions précédentes et des propriétés d'un canal symétrique, nous exprimons l'équivalence.

Dans la suite, nous commençons par présenter le décodeur LDPC somme-produit sous une forme plus pratique pour réaliser l'évolution de densité (Partie 3.4.1). Nous exprimons ensuite les deux récursions et l'équivalence (Partie 3.4.2). Nous présentons enfin un exemple d'utilisation de la méthode proposée (Partie 3.4.3).

3.4.1 Décodage LDPC

Ici, nous considérerons un code LDPC non-binaire et un décodage de type somme-produit. Les opérations de codage et de décodage ont été décrites dans la partie 2.1.2. Pour les besoins de l'évolution de densité, introduisons la fonction $\boldsymbol{\gamma}$ suivante, qui s'applique sur des vecteurs de taille q. Sa k-ème composante $\gamma_k : \mathbb{C} \to \mathbb{R} \times [-\pi, \pi]$ est donnée par

$$\gamma_k(f_k) = \begin{cases} \left(\frac{1}{2}\log(x_k^2 + y_k^2), \arctan\frac{y_k}{x_k}\right) & \text{si } x_k \geq 0, y_k \neq 0 \\ \left(\frac{1}{2}\log(x_k^2 + y_k^2), \arctan\frac{y_k}{x_k} + \pi\right) & \text{si } x_k \leq 0 \ , \ y_k \geq 0 \\ \left(\frac{1}{2}\log(x_k^2 + y_k^2), \arctan\frac{y_k}{x_k} - \pi\right) & \text{si } x_k \leq 0 \ , \ y_k < 0 \end{cases} \tag{3.26}$$

où x_j et y_j sont la partie réelle et la partie imaginaire de f_k. A partir de cette définition et de la fonction inverse $\boldsymbol{\gamma}^{-1}$, on obtient une autre expression du passage de messages des NP vers les NV (2.5) :

$$\mathbf{m}^{(\ell)}(m,n) = \mathcal{A}[\overline{s}_m]\tilde{\mathcal{F}}^{-1}\left(\boldsymbol{\gamma}^{-1}\left(\sum_{n'\in\mathcal{N}(m)\setminus n}\boldsymbol{\gamma}\left(\tilde{\mathcal{F}}\left(W\left[\overline{g}_{n'm}\right]\mathbf{m}^{(\ell-1)}(n',m)\right)\right)\right)\right) . \tag{3.27}$$

L'intérêt de cette expression est qu'elle transforme en somme le produit présent dans (2.5). En effet, dans l'évolution de densité, on part de la densité de probabilité des messages à l'itération ℓ, et on cherche à exprimer la densité de probabilité des messages à l'itération $\ell+1$, à partir des passages de messages (2.7) et (3.27). Il est relativement simple d'exprimer probabilité d'une somme de variables aléatoires, (il suffit d'exprimer un produit de convolution), alors qu'évaluer la probabilité d'un produit est beaucoup plus compliqué. C'est pourquoi nous avons introduit la fonction $\boldsymbol{\gamma}$.

3.4.2 Évolution de densité

Pour cette partie, les définitions et les résultats détaillés sont disponibles en Annexe C, partie III.

Ici, nous nous intéresserons tout d'abord au problème de codage de canal. Pour ce problème, nous supposons que le canal est symétrique et nous effectuons les deux hypothèses suivantes : (i) Les messages sont indépendants [WKP05], (ii) le mot de code zéro a été transmis [RSU01]. On note $P^{(\ell)}$ la densité de probabilité des messages des NV aux NP à l'itération ℓ, et $Q^{(\ell)}$ la densité de probabilité des messages des NP aux NV. Ces densités de probabilités sont exprimées sachant que le mot de code zéro a été transmis. D'après [LFK09], la probabilité d'erreur de l'algorithme somme-produit à l'itération ℓ est donnée par

$$p_e^{(\ell)} = 1 - \int_{\mathbf{m}\in\mathbb{R}_+^q} P^{(\ell)}(\mathbf{m})d\mathbf{m} . \tag{3.28}$$

De plus, on exprime la récursion suivante sur $P^{(\ell)}$:

$$P^{(\ell)}(\mathbf{m}) = P^{(0)}(\mathbf{m}) \bigotimes \lambda \left(\Gamma_d^{-1} \left(\frac{1}{q-1} \sum_{g=1}^{q} \rho \left(\Gamma_c^g \left(P^{(\ell-1)} \right) \right) \right) \right). \tag{3.29}$$

Dans cette expression, Γ_d^{-1} et Γ_c^g sont des opérateurs de transformation de densité (voir Annexe C, partie III).

Pour le problème de codage SW, on s'intéresse maintenant à un canal de corrélation qui n'est pas nécessairement symétrique. On note $P_k^{(\ell)}$ la densité de probabilité des messages des NV aux NP à l'itération ℓ sachant que $X = k$, et $Q_k^{(\ell)}$ la densité de probabilité des messages des NP aux NV. On définit

$$\langle P^{(\ell)} \rangle(\mathbf{m}) = \sum_{k=0}^{q-1} P(X = k) P_k^{(\ell)} \circ \mathcal{A}[-k](\mathbf{m}) \tag{3.30}$$

et on montre que la probabilité d'erreur est donnée par

$$p_e^{(\ell)} = 1 - \int_{\mathbf{m} \in \mathbb{R}_+^q} \langle P^{(\ell)} \rangle(\mathbf{m}) d\mathbf{m}. \tag{3.31}$$

On exprime ensuite la récursion suivante sur $\langle P^{(\ell)} \rangle(\mathbf{m})$:

$$\langle P^{(\ell)} \rangle(\mathbf{m}) = \langle P^{(0)} \rangle \bigotimes \lambda \left(\Gamma_d^{-1} \left(\frac{1}{q-1} \sum_{g=1}^{q} \rho \left(\Gamma_c^g \left(\langle P^{(\ell-1)} \rangle \right) \right) \right) \right) (\mathbf{m}). \tag{3.32}$$

On voit que les récursions (3.29) et (3.32) sont les mêmes. La seule différence est que la première s'exprime sur les $P^{(\ell)}$, et la deuxième sur les $\langle P^{(\ell)} \rangle$. Si ces deux récursions sont initialisées de la même manière, elles donnent donc exactement les mêmes expressions de densité de probabilité à l'itération ℓ. Il suffit alors de travailler sur une équivalence au niveau des densités initiales $P^{(0)}$ et $\langle P^{(0)} \rangle$. A partir de ce résultat, nous démontrons le théorème suivant :

Théorème. *Soit $P(Y|X)$ un canal de corrélation q-aire. On note $\langle P^{(0)} \rangle$ sa densité de probabilité initiale pour l'évolution de densité. Il existe un canal* équivalent *$P(W|U)$ symétrique de densité de probabilité initiale $P^{(0)} = \langle P^{(0)} \rangle$. De plus, $P(Y|X)$ et $P(W|U)$ ont les mêmes équations d'évolution de densité et $H(X|Y) = H(U|W)$.*

3.4.3 Exemples

Pour cette partie, les détails des calculs sont disponibles en Annexe C, partie IV. On s'intéresse au cas d'une source X à valeurs dans $\mathrm{GF}(q)$, et on note $P(X = k) = p_k$. On suppose que le canal de corrélation est un canal q-aire symétrique tel quel

$$\begin{aligned} \forall y \neq k,\ P(Y = y|X = k) =& \frac{p}{q-1} \\ P(Y = k|X = k) =& 1 - p \end{aligned} \tag{3.33}$$

avec $0 < p < 1$. On montre (voir Annexe C, partie IV) que le canal $P(W|U)$ équivalent à $P(Y|X)$ est un canal à q entrées et q^2 sorties. On donne les probabilités de transition du canal pour l'entrée $U = 0$. Les autres peuvent être obtenues par symétrie, mais pour l'évolution de densité, seule l'entrée $U = 0$ est importante (hypothèse du mot de code zéro). Chaque entrée $k \in \mathrm{GF}(q)$ donne une sortie telle que

$$P(W = w|U = 0) = p_k(1 - p)$$

et $q - 1$ sorties telles que

$$P(W = w|U = 0) = p_k \frac{p}{q-1}.$$

Ici, nous supposerons que les p_k sont fixés, et que l'on veut calculer une valeur approchée du seuil du code par rapport au paramètre p. On appelle *seuil* d'un code le plus grand paramètre p possible pour lequel le code à une probabilité d'erreur inférieure à une valeur ϵ, que l'on fixe ici à 10^{-5}. Les valeurs approchées des seuils des codes sont obtenues grâce à la méthode MCMC décrite dans [GSD10] et appliquée au canal équivalent sur des vecteurs de taille 100000. Nous donnons ici les seuils approchés de deux codes d'efficacité 1/2. Le premier code est le code régulier avec $d_v = 2$, $d_c = 4$. Le second code à une distribution de degrés $\lambda(x) = x^2$, $\rho(x) = 0.0110735x^2 + 0.1073487x^3 + 0.2583159x^4 + 0.4296047x^5 + 0.0115064x^6 + 0.1592504x^7 + 0.0166657x^8 + 0.0046979x^{25} + 0.0015367x^{26}$. La distributions de degrés est ici exprimée d'un point de vue nœuds (*i.e.* λ_i correspond à la proportion de nœuds de degré i). Dans chacun des cas considérés, nous précisons le corps de Galois, la distribution de X et $\overline{p}$, le plus grand paramètre (approché) tel que $H(X|Y) \leq 1/2$).

Dans GF(4), X distribuée uniformément, $\overline{p} = 0.189$ Pour le code régulier, un seuil approché est donné par $p = 0.071$ ($H(p) = 0.24$ bit/symbole). Pour le code irrégulier, un seuil approché est donné par $p = 0.159$ ($H(p) = 0.44$ bit/symbole).

Dans GF(4), X de distribution $[0.5, 0.25, 0.125, 0.125]$, $\overline{p} = 0.225$ Pour le code régulier, un seuil approché est donné par $p = 0.083$ ($H(p) = 0.25$ bit/symbole). Pour le code irrégulier, un seuil approché est donné par $p = 0.192$ ($H(p) = 0.45$ bit/symbole).

Dans GF(16), X distribué uniformément, $\overline{p} = 0.289$ Pour le code régulier, un seuil approché est donné par $p = 0.149$ ($H(p) = 0.30$ bit/symbole). Pour le code irrégulier, un seuil approché est donné par $p = 0.248$ ($H(p) = 0.44$ bit/symbole).

Dans GF(16), X de distribution $[0.4, 0.04, \ldots, 0.04]$, $\overline{p} = 0.367$ Pour le code régulier, le seuil est donné par $p = 0.198$ ($H(p) = 0.32$ bit/symbole). Pour le code irrégulier, le seuil est donné par $p = 0.318$ ($H(p) = 0.45$ bit/symbole).

On voit que les mêmes codes dans différentes situations donnent des seuils différents, et que ces seuils correspondent à des valeurs différentes d'entropie.

3.5 Codage de WZ pour un canal de corrélation distribué suivant un modèle de Markov caché

L'article correspondant est disponible en Annexe D. Dans cette partie, nous souhaitons étudier le schéma de codage de WZ pour un modèle de corrélation qui varie dans le temps. Nous nous intéressons donc au modèle suivant. La dépendance entre les variables X_k et Y_k est représentée par le modèle additif $Y_k = X_k + Z_k$ dans lequel Z_k est une variable aléatoire indépendante de X_k. Les variables aléatoires de la suite $\{X_k\}_{k=1}^{+\infty}$ sont *i.i.d.*, distribués suivant $\mathcal{N}(0, \sigma_x^2)$. Les variables aléatoires de la suite $\{Z_k\}_{k=1}^{+\infty}$ sont, elles, distribuées suivant un modèle de Markov caché [Rab89] à émissions Gaussiennes et état caché S_n. L'état caché S_k prend ses valeurs dans l'al-

phabet $\mathcal{S} = \{0, 1\}$, et la variable aléatoire $(Z_k|S_k = s)$ est distribuée suivant $\mathcal{N}(0, \sigma_s^2)$. La suite $\{S_k\}_{k=1}^{+\infty}$ est distribuée suivant un processus de Markov invariant d'ordre 1 et de matrice de transition P telle que

$$P_{s',s} = \Pr(S_n = s|S_{n-1} = s') > 0 \quad \forall (s', s) \in \mathcal{S}^2 \ . \tag{3.34}$$

Les probabilités initiales sont notées $p_s^{(1)} = P(S_1 = s)$. On choisit une mesure de distortion quadratique $d(\mathbf{X}^n, \hat{\mathbf{X}}^n) = \|\mathbf{X}^n - \hat{\mathbf{X}}^n\|^2$. Des cas similaires ont été étudiés dans [MBD89], qui propose une analyse de performance pour un canal binaire avec mémoire et dans [CX, GDV06] qui traitent le cas d'échantillons Gaussiens corrélés, sans état caché.

Pour notre modèle avec mémoire, nous souhaitons effectuer l'analyse de performance et proposer un schéma de codage. Le schéma de codage que nous considérerons s'appuie sur la structure Quantification + code LDPC non-binaire + estimateur MMSE. Ce schéma de codage doit notamment exploiter le modèle à mémoire sur l'état caché. Nos contributions sont les suivantes

1. Nous effectuons l'analyse de performance pour notre source. Le modèle défini ici rentre dans le cadre des sources générales. Cependant, l'expression explicite de (2.16) est difficile à déterminer pour notre modèle. C'est pourquoi nous fournissons simplement des bornes de la fonction débit-distortion.

2. Nous proposons un algorithme de décodage de type somme-produit pour des codes LDPC non-binaire. L'algorithme proposé prend en compte la mémoire sur les états cachés.

3. Comme l'estimateur MMSE ne peut pas être exprimé de manière explicite pour notre modèle, nous proposons une méthode MCMC qui permet de réaliser cette opération d'estimation. Cette méthode tient compte de la mémoire sur les états cachés.

Dans la suite, la partie 3.5.1 présente les résultats de l'analyse de performance, la partie 3.5.2 décrit le schéma de codage que nous proposons, et la partie 3.5.3 présente quelques résultats de simulations.

3.5.1 Analyse de performance

Les définitions complètes et les résultats détaillés ici sont disponibles en Annexe D, partie III. Dans cette partie, nous caractérisons deux fonctions débit-distortion. La première, $R^{\mathrm{WZ}}_{X|Y,S}(D)$, correspond au cas où les états cachés S_k sont donnés au décodeur. On sait que les variables aléatoires successives $(X_k, Y_k|S_k = s_k)$ sont indépendantes, Gaussiennes, de variance connue. Le calcul de $R^{\mathrm{WZ}}_{X|Y,S}(D)$ correspond donc à l'évaluation de la fonction débit-distortion de WZ pour une source Gaussienne, sans mémoire. $R^{\mathrm{WZ}}_{X|Y,S}(D)$ constitue une borne inférieure pour $R^{\mathrm{WZ}}_{X|Y}(D)$ qui correspond à la fonction débit-distortion de Wyner-Ziv pour notre modèle. On montre que

$$R^{\mathrm{WZ}}_{X|Y,S}(D) = \sum_{s\in\mathcal{S}} p_s \max\left(0, \frac{1}{2}\log_2 \frac{\sigma^2_{X|Y,s}}{D'}\right), \tag{3.35}$$

avec D' tel que $\sum_{s\in\mathcal{S}} p_s \min(D', \sigma^2_{X|Y,s}) \leq D$. On obtient également les bornes suivantes sur $R^{\mathrm{WZ}}_{X|Y}(D)$:

$$R^{\mathrm{WZ}}_{X|Y,S}(D) \leq R^{\mathrm{WZ}}_{X|Y}(D) \leq R^{\mathrm{WZ}}_{X|Y,S}(D) + L^{\mathrm{WZ}}_{X|Y}(D) + \Lambda^{\mathrm{WZ}}_{X|Y} \tag{3.36}$$

avec

$$L^{\mathrm{WZ}}_{X|Y}(D) = \frac{1}{2}\log_2\left(1 + \frac{D}{\sigma^2_{X|Y,0}}\right) \tag{3.37}$$

$$\Lambda^{\mathrm{WZ}}_{X|Y} = \min\big(\lim_{k\to\infty} H(S_k|S_{k-1}),\ \mathrm{h}(\mathcal{Z}) - \lim_{k\to\infty} h(Z_k|S_k)\big), \tag{3.38}$$

et $h(\mathcal{Z}) = \lim_{n\to\infty} \frac{1}{n} h(\mathbf{Z}^n)$. La borne supérieure dans (3.36) dépend de $R^{\mathrm{WZ}}_{X|Y,S}(D)$ et de deux termes de perte. Le premier, $L^{\mathrm{WZ}}_{X|Y}(D)$, dépend de la distortion, et on a $\lim_{D\to 0} L^{\mathrm{WZ}}_{X|Y}(D) = 0$. Le second, $\Lambda^{\mathrm{WZ}}_{X|Y}$, ne dépend pas de la distortion et reflète bien le fait que les S_k sont inconnus.

3.5.2 Schéma de codage

Pour cette partie, les détails des résultats et des calculs sont disponibles en Annexe D, partie IV. Le schéma de codage proposé est représenté sur la Figure 3.5. Tout d'abord, les symboles X_n sont quantifiés à l'aide d'une quantification scalaire

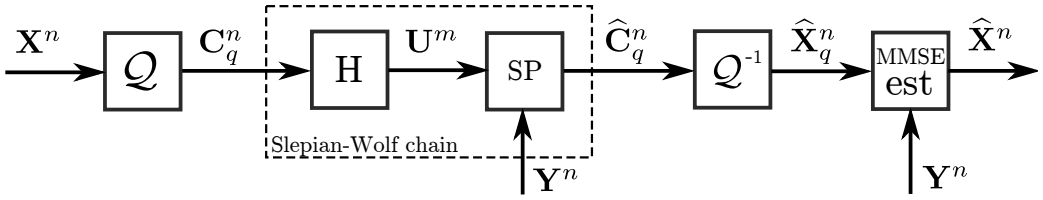

FIGURE 3.5 – Schéma de codage

uniforme à q niveaux et avec dithering préalable. Ici, on ne peut pas utiliser une quantification en treillis, parce que le décodeur LDPC qui suit a besoin des probabilités *a priori* des symboles quantifiés, ce qui est difficile à obtenir avec la quantification en treillis. Ensuite, les symboles quantifiés sont transposés dans GF(q), ce qui donne le vecteur d'éléments discrets $\mathbf{C}_q^n$.

Ce vecteur est transmis au moyen d'une chaîne de codage de SW, réalisée à l'aide d'un code LDPC dans GF(q). On souhaite réaliser le décodage du mot de code $\mathbf{U}^m$ avec un algorithme de décodage LDPC de type somme-produit. En particulier, l'algorithme de décodage doit prendre en compte la mémoire sur les états cachés. Dans le cas binaire, [EKP05, GF04] proposent un algorithme de décodage qui tient compte de la mémoire. Nous proposons donc une généralisation de ce décodeur au cas non-binaire. Pour cela, nous partons de l'algorithme somme-produit [RU08, Chapitre 2], [KFL01] et explicitons les équations de mise à jour de l'algorithme pour notre modèle. Les équations de l'algorithme sont disponibles en Annexe D, partie IV.2. On obtient un vecteur estimé $\hat{\mathbf{C}}_q^n$ que l'on retranspose dans l'espace des niveaux de quantification.

Pour finir, nous réalisons l'estimation MMSE du vecteur de source à partir des symboles décodés $\widehat{\mathbf{X}}_q^n$ et de l'information adjacente $\mathbf{Y}^n$. Les équations de l'estimateur MMSE ne peuvent être exprimées en expression explicite pour notre modèle. C'est pourquoi nous proposons de réaliser l'estimation à l'aide d'une méthode MCMC qui s'appuie sur la technique d'échantillonnage de Gibbs [C.P00]. On pourrait s'intéresser à d'autres solutions pour réaliser l'estimation MMSE. On pourrait par exemple chercher à adapter [SPZ08] au cas avec mémoire, mais cette méthode réalise une estimation MMSE trop grossière et ne fonctionne pas bien pour notre modèle. Pour l'estimation de $\mathbf{x}^n$, une méthode MCMC consiste dans la production d'un grand

nombre d'échantillons aléatoires de $\mathbf{X}^n$ suivant la distribution de probabilité *a posteriori* $P(\mathbf{X}^n|\mathbf{Y}^n, \hat{\mathbf{X}}_q^n)$. On réalise ensuite l'estimation en calculant la moyenne des échantillons aléatoires. Malheureusement, il est difficile de générer des échantillons aléatoires à partir de $P(\mathbf{X}^n|\mathbf{Y}^n, \hat{\mathbf{X}}_q^n)$ puisque cette distribution ne peut être obtenue qu'en marginalisant par rapport à $\mathbf{S}^n$. Il est en revanche beaucoup plus simple d'effectuer l'échantillonnage à partir de $P(\mathbf{X}^n|\mathbf{Y}^n, \hat{\mathbf{X}}_q^n, \mathbf{S}^n)$ puisqu'il s'agit d'une distribution Gaussienne.

Le principe de l'échantillonnage de Gibbs est donc le suivant. On produit des échantillons aléatoires de $\mathbf{X}^n$ et $\mathbf{S}^n$ à partir des distributions $P(\mathbf{X}^n|\mathbf{Y}^n, \hat{\mathbf{X}}_q^n, \mathbf{S}^n)$ et $P(\mathbf{S}^n|\mathbf{Y}^n, \hat{\mathbf{X}}_q^n, \mathbf{X}^n)$ alternativement et itérativement. Les résultats de [C.P00] montrent ensuite que lorsque le nombre d'itérations tend vers l'infini, l'échantillonnage successif produit des échantillons aléatoires suivant la distribution $P(\mathbf{X}^n|\mathbf{Y}^n, \hat{\mathbf{X}}_q^n)$. L'algorithme est donc le suivant.

1. On génère des échantillons de $\mathbf{X}^{n,(\ell)}$ à partir de

$$P\left(\mathbf{x}^n|\mathbf{s}^{n,(l-1)}, \bar{\mathbf{x}}_q^n, \bar{\mathbf{y}}^n\right) = \frac{1}{(2\pi)^{n/2} R_x^{1/2}} \exp\left(-\frac{(\mathbf{x}^n - \mathbf{m}_x^n)^T R_x^{-1} (\mathbf{x}^n - \mathbf{m}_x^n)}{2}\right) \tag{3.39}$$

dans lequel

$$\begin{aligned} R_x &= \left(I_n\left(\frac{1}{\Delta^2/12} + \frac{1}{\sigma_x^2}\right) + R_s^{-1}\right)^{-1} \\ \mathbf{m}_x^n &= R_x\left(\frac{\bar{\mathbf{x}}_q^n}{\Delta^2/12} + R_s^{-1}\bar{\mathbf{y}}^n\right) \end{aligned} \tag{3.40}$$

2. On génère des échantillons de $\mathbf{X}^{n,(\ell)}$ à partir de

$$P\left(S_1^{(\ell)} = 1|x_1^{(\ell)}, y_1, x_{q,1}\right) = \left(1 + \frac{P(S_1=1)}{P(S_1=0)}\sqrt{\frac{\sigma_0^2}{\sigma_1^2}}\ e^{\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)(x_1^{(\ell)} - \bar{y}_1)^2\right)}\right)^{-1} \tag{3.41}$$

et $\forall k = 2 \ldots n$,

$$P\left(S_k^{(l)} = 1|x_k^{(\ell)}, y_k, x_{q,k}, s_{k-1}^{(\ell)}\right) = \left(1 + \frac{P(S_k=1|s_{k-1}^{(\ell)})}{P(S_k=0|s_{k-1}^{(\ell)})}\sqrt{\frac{\sigma_0^2}{\sigma_1^2}}\ e^{\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)(x_k^{(\ell)} - \bar{y}_k)^2\right)}\right)^{-1} \tag{3.42}$$

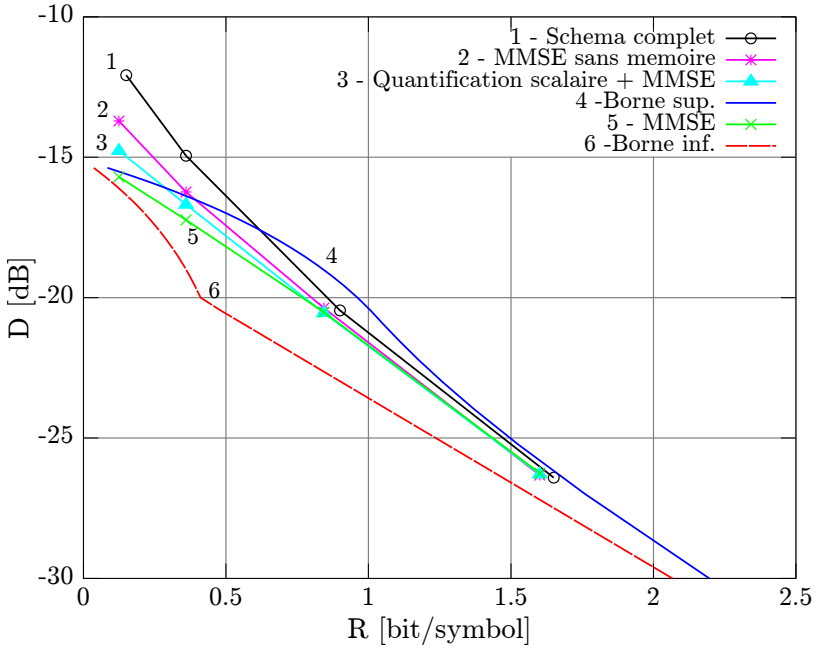

FIGURE 3.6 – Courbes débit-distortion pour les méthodes évaluées

3.5.3 Résultats de simulation

Pour cette partie, les résultats détaillés sont disponibles en Annexe D, partie V. On choisit les paramètres suivants : $\mu = 1 - p_{01} - p_{10} = 0.98$, voir [MBD89], avec $p_{01} = \Pr(S_k = 1|S_{k-1} = 0) = 0.0156$ et $p_{10} = \Pr(S_k = 0|S_{k-1} = 1) = 0.0044$. La variance des X_k est fixée à $\sigma_x^2 = 1$ et $\sigma_0^2 = 0.1,\ \sigma_1^2 = 0.01,\ p_0 = 0.9844$. Les résultats sont obtenus sur des blocs de longueur $n = 10000$ et sur 100 réalisations (sauf pour les bornes inf. et sup., qui sont calculées à partir de leurs expressions analytiques). Toutes les courbes débit-distortion pour les méthodes évaluées sont représentées sur la Figure 3.6.

On représente tout d'abord la bornes inférieure (courbe 6) et la borne supérieure (courbe 4). Ensuite, on évalue la performance de la méthode MCMC pour la reconstruction MMSE, en supposant que la chaîne de SW est idéale. On évalue alors trois configurations. Dans tous les cas, on génère 400 échantillons pour la méthode MCMC. Dans la première configuration (MMSE, courbe 5), on génère les $\mathbf{C}_q^n$ directement à partir du modèle $\mathbf{C}_q^n = \mathbf{X}_q^n + \mathbf{B}_q^n$, et on effectue l'estimation MMSE à partir de la

méthode proposée dans la partie 3.5.2. Dans la seconde configuration (MMSE sans mémoire, courbe 2), on génère de même les $\mathbf{C}_q^n$ à partir du modèle Gaussien, et on applique la méthode MCMC, mais sans tenir compte de la mémoire sur les états. On observe une perte d'environ 2dB à bas débit. Dans la troisième configuration (quantification scalaire + MMSE, courbe 3), on obtient les $\mathbf{C}_q^n$ à partir d'une quantification scalaire uniforme. On observe cette fois une perte d'environ 1dB, ce qui montre que le modèle Gaussien précédent est cohérent. Dans ces cas, on voit qu'à haut débit, les courbes sont confondues. En effet, à haut débit, le bruit de quantification est extrêmement faible, et donc le décodeur tient plus compte des symboles quantifiés (sans mémoire) que de l'information adjacente (qui contient de la mémoire).

Enfin, on évalue le schéma complet quantification scalaire + codes LDPC + estimation MMSE (courbe 1). On choisit d'effectuer 40 itérations pour le décodeur LDPC et on considère les codes LDPC suivants (distributions exprimées d'un point de vue nœuds). Pour 8 niveaux de quantification, $\lambda(x) = x$, $\rho(x) = x^{39}$. Pour 16 niveaux, $\lambda(x) = x$, $\rho(x) = 0.78x^{21} + 0.22x^{22}$. Pour 32 niveaux, $\lambda(x) = x$, $\rho(x) = 0.01x^9 + 0.88x^{10} + 0.11x^{11}$. Pour 64 niveaux, $\lambda(x) = x$, $\rho(x) = 0.01x^5 + 0.72x^6 + 0.27x^7$. On observe une perte par rapport au cas quantification scalaire + MMSE. Cette perte est due à la fois aux erreurs introduites par le décodeur LDPC, et à la construction des matrices de codage LDPC.

Chapitre 4

Conclusion et perspectives

Dans cette thèse, nous avons considéré des modèles de sources de caractéristiques incertaines. Pour ces modèles, nous avons étudié les performances du schéma de codage de SW et proposé un schéma de codage pratique s'appuyant sur des codes LDPC non binaires. Nous avons également proposé une méthode pour effectuer le design de codes LDPC non binaires pour le schéma de codage de SW dans le cas où la distribution des sources est bien connue. Enfin, nous avons considéré un modèle de sources avec mémoire. Nous avons effectué l'analyse des performances et proposé un schéma de codage pratique pour le problème de codage de WZ.

Nous décrivons ici les perspectives liées à ces résultats. Ces perspectives peuvent se séparer en trois catégories. Il s'agira tout d'abord de réfléchir à l'amélioration des schémas de codage qui ont été proposés précédemment (partie 4.1). Ensuite, nous décrirons un certain nombre de problèmes liés aux incertitudes sur les modèles de sources (partie 4.2). Enfin, nous considérerons le cas de réseaux plus complexes que ceux décrits précédemment (partie 4.3).

4.1 Schémas de codage

Nous décrivons ici des perspectives liées aux schémas de codages de SW décrits dans le thèse.

4.1.1 Décodage de type minimum-somme pour le codage de SW

Dans le cas où les distributions des sources sont mal connues, nous avons proposé un schéma de codage utilisant des codes LDPC non binaires et un algorithme EM pour estimer conjointement le vecteur de source et les paramètres. Pour gagner du temps de décodage en évitant d'utiliser un algorithme EM (et aussi pour éviter les éventuels échecs d'estimation), on pourrait réfléchir à l'utilisation d'un algorithme de décodage LDPC de type minimum-somme proposé à l'origine pour des problèmes de codage de canal [DF05, Sav08c, CLXZS13]. Au prix d'une légère perte de performance, le décodage de type minimum-somme n'a pas besoin de la connaissance des probabilités conditionnelles pour fonctionner. Deux points en particulier pourraient donc être étudiés.

Tout d'abord, l'algorithme minimum-somme doit être initialisé correctement pour éviter une perte de performance sensible par rapport à l'algorithme somme-produit. Il s'agirait donc de réfléchir à ces problèmes d'initialisations et en particulier de voir si les solutions proposées dans la partie 3.3 peuvent s'appliquer.

Ensuite, en réalité, on ne peut pas exprimer les équations de l'algorithme minimum-somme si les sources ne sont pas distribuées uniformément. Une première idée serait d'appliquer un mélange aux symboles d'entrée avant compression, pour les rendre uniforme. Mais dans ce cas, l'expression du canal de corrélation pourrait devenir particulièrement complexe.

Une deuxième idée consister à transformer les éléments de GF(q) en bits, la distribution de probabilité des symboles binaires est quasiment uniforme, mais on perd beaucoup en terme de débit. En revanche, en partant d'un corps de Galois assez grand (64 ou 128 par exemple), on pourrait exprimer les symboles dans un corps plus petit (16 ou 8 par exemple). Les symboles d'entrée seraient également plus ou moins distribués uniformément mais on perdrait moins en débit. Cela permettrait en plus d'utiliser des codes LDPC dans des corps plus petits, et donc d'accélérer le décodage. Il s'agirait alors d'exprimer le compromis entre perte de débit et temps de décodage.

4.1.2 Évolution de densité pour le cas incertain

Ici, nous avons décrit l'évolution de densité pour le codage de SW dans le cas où la distribution des sources était parfaitement connue. Il pourrait être intéressant de réfléchir au design de codes LDPC pour le cas où l'incertitude sur la distribution des sources est prise en compte explicitement. Le problème a été étudié dans [CHJ09a, SB09] dans le cas binaire, et nous nous intéressons ici au cas non binaire. Si l'ensemble des distributions des sources constitue un ensemble dégradé, l'évolution de densité que nous avons proposée permet d'étudier les performances des codes pour l'ensemble en entier : il suffit de trouver un code qui fonctionne bien pour le pire cas [RU01]. Si cette propriété n'est pas respectée (par exemple, si l'ensemble des paramètres n'est pas connexe, ou s'il y a plusieurs pires cas), il s'agira de réfléchir à une manière d'étudier les performances du code malgré tout.

4.2 Incertitude sur les modèles

Les modèles de sources que nous avons introduits permettent de tenir compte d'une incertitude. D'autres problèmes en lien avec cette incertitude peuvent aussi être étudiés.

4.2.1 Sélection de modèles pour des applications particulières

Il s'agirait ici de travailler à la construction de classes réalistes de modèles pour des problèmes de codage particuliers tels que la vidéo. Dans ce genre de problème, de nombreuses questions peuvent se poser : type de modèle (Gaussien, Laplacien, mixte), nombre d'éléments s'il s'agit d'un mélange, prise en compte ou non de la mémoire. Ainsi, [BAP06] montre qu'en codage vidéo, un modèle de corrélation Laplacien est plus pertinent qu'un modèle Gaussien. Tout cela a des conséquences en terme de débit et de traitement des données. En effet, un modèle plus complexe peut permettre de représenter plus fidèlement les données, et donc de diminuer le débit de codage, mais il faut pouvoir estimer tous les paramètres inconnus à partir des

données compressées. L'idée ici serait donc de proposer des critères de sélection de modèle adaptés aux problèmes de codage. Par exemple, [EKP07] étudie la complexité des modèles en fonction de la complexité du décodeur LDPC, mais d'autres critères tels que le minimum description length [HY01] pourraient être étudiés.

4.2.2 Modèles plus complexes

On pourrait s'intéresser à des modèles de corrélation plus complexes que ceux que nous avons considéré jusqu'à maintenant. Par exemple, on pourrait imaginer que le modèle de corrélation est un mélange de Gaussiennes, mais qu'on ne connaît pas le nombre et les éléments qui composent le mélange. Ce type de modèle et les méthodes d'apprentissage associées (données non compressées) sont présentées dans [HTF09]. Il s'agirait alors de mettre en œuvre des techniques d'apprentissage pour estimer ces paramètres à partir des données compressées.

4.2.3 Incertitude sur les modèles de sources pour des problèmes particuliers de codage

Nous avons étudié des modèles de sources de caractéristiques incertaines principalement pour le problème de codage de SW. Nous pourrions nous intéresser maintenant à d'autres problèmes de codage. Pour ces problèmes de codage, il s'agirait alors de réfléchir à nouveau sur les notions de codage universel, sur l'étude de performance et sur la construction de schémas pratiques.

Codage de Wyner-Ziv Tout d'abord, en codage de Wyner-Ziv, aucune analyse de performance n'a été proposée pour les modèles que nous avons considérés. Seul le cas sans information adjacente a été étudié dans [Ber71]. Pour garantir une contrainte de distortion donnée, il faudrait probablement là-aussi choisir le débit pour le pire cas sur les paramètres. Cependant, intuitivement, si le vrai paramètre en cours est plus favorable que le pire cas, on devrait pouvoir diminuer la distortion. On souhaiterait donc vérifier cette propriété en exprimant les fonctions débit-distortion et les fonctions

distortion-débit dans le cas de sources incertaines. Il s'agirait ensuite de proposer un schéma de codage qui puisse atteindre les performances théoriques obtenues. Par exemple, un code LDPC standard sans canal de retour ne devrait pas permettre de réduire la distortion dans un cas favorable. Une idée serait peut-être de travailler sur un schéma quantification + codage sans perte + MMSE complètement intégré, à l'aide d'un algorithme de type somme-produit général.

Codage à description multiple En codage à descriptions multiples [EC82], le codeur doit construire plusieurs descriptions de la source et les différents décodeurs ont accès chacun à un nombre différent de ces descriptions. Plus un décodeur reçoit de descriptions, plus sa qualité de reconstruction doit être améliorée. Ici, on pourrait supposer en plus que les décodeurs ont accès à des informations adjacentes de qualité différente, et que le codeur ne connaît pas la destination de ses descriptions. Il s'agirait alors d'effectuer l'analyse de performance et de proposer un schéma de codage pour ce problème. En codage vidéo, cette question a été abordée d'un point de vue pratique dans [CPPG10].

Codage de canal avec des relais Le problème de l'incertitude peut également se poser dans le domaine du codage de canal. On pourrait par exemple imaginer le cas où un relai est disponible entre l'émetteur et le récepteur [NBK04]. Le relais, l'émetteur et le récepteur peuvent avoir des connaissances différentes des canaux en jeu. Il s'agirait alors d'étudier comment ces différentes connaissances ou méconnaissances peuvent dégrader les performance. Là aussi, il s'agira ensuite de construire le schéma de codage adapté à ce problème.

4.3 Généralisation à un réseau de capteurs

Jusqu'à maintenant, nous avons principalement considérés le cas du codage de sources avec information adjacente au décodeur. Il s'agirait d'étudier plus généralement le problème de codage de sources distribué. Nous avons effectué une analyse simple dans le cas d'un réseau avec deux capteurs et un point de collecte, et il s'agirait de

généraliser cette étude. Ce problème comporte plusieurs aspects que nous décrivons ici.

4.3.1 Analyse de performance

Le papier original de Slepian-Wolf [SW73] décrit les performances de codage pour le cas sans pertes, pour un nombre arbitraire de sources et un seul décodeur. Le cas de plusieurs décodeurs a ensuite été étudié dans [CBLV05] Il s'agirait donc dans un premier temps d'obtenir ces performances dans le cas de sources incertaines, pour un seul décodeur puis pour plusieurs décodeurs.

4.3.2 Autres stratégies de codage

Lorsque nous avons étudié le cas d'un réseau à trois nœuds (voir partie 3.2.4), nous nous sommes intéressés à plusieurs stratégies de codage : codage séparé ou codage conjoint, séquence d'apprentissage ou non. Dans ce cas simple, on pourrait étudier également d'autres stratégies, telles que le relayage ou le codage avec canal de retour. L'idée serait de voir si ces stratégies peuvent avoir un intérêt, et de réfléchir à la manière de les étudier (par exemple, comment pénaliser le canal de retour). De plus, nous avons jusqu'à maintenant supposé que les coûts de communication étaient additifs et calculés à partir de coefficients multiplicatifs. D'autres modèles de communication (cas broadcast, fonctions non-linéaires, etc.) pourraient également être étudiés. Ainsi, [Rub76] propose une fonction de coût qui prend en compte les délais de transmission.

4.3.3 Construction de codes

Toujours dans le cas simple d'un réseau à trois nœuds, on pourrait réfléchir aux schémas de codage que l'on pourrait implémenter pour les différentes stratégies : codage conjoint, codage séparé, séquence d'apprentissage, relayage, canal de retour. Par exemple, pour le codage séparé, [EY05a] traite le cas d'un réseau simple avec deux sources, et [CG12] considère le cas plus général d'un réseau avec un nombre

arbitraire de nœuds. Mais dans les deux cas, les distributions jointes de sources sont supposées connues. Il s'agirait de voir quelles sont les performances de codes que l'on pourrait construire en pratique dans le cas incertain, et de ne plus se limiter à l'analyse théorique précédente. Cette étude pourrait permettre d'affiner le calcul du coût de chaque stratégie. En effet, ces coûts pourraient éventuellement prendre en compte une dégradation de performances pour certains schémas, ou des critères particuliers comme la complexité de décodage.

4.3.4 Réseaux plus grands

Ensuite, on pourrait réfléchir au cas d'un réseau plus grand, c'est-à-dire avec un nombre arbitraire de sources et de points de collectes. La situation où les distributions jointes des différentes sources appliquées sont parfaitement connues a notamment été abordée dans [CBLV05, CBLVW06, HME+04, RJCE06, VAR10]. Ce problème impliquerait de prendre en compte le routage des informations dans le réseau. Étant donné que ce problème est particulièrement complexe, il faudrait réfléchir à des manières de le simplifier. En particulier, pour éviter d'avoir à optimiser conjointement le codage et le routage, on pourrait s'interroger sur la séparabilité des deux problèmes, par exemple en fonction des modèles de transmission (calcul des coûts), ou des stratégies considérées. Si les deux problèmes ne sont effectivement pas séparables, on pourrait chercher à évaluer la perte induite par une optimisation séparée. On pourrait également travailler sur le problème de l'optimisation centralisée ou distribuée, et chercher à proposer des heuristiques pour réaliser cette optimisation. Cette étude pourrait être réalisée par stratégie (par exemple, codage conditionnel seulement), ou en autorisant des stratégies mixtes (codage conditionnel partiel et codage de SW), ou en comparant plusieurs stratégies mais en n'en choisissant à la fin qu'une seule et même pour tout le réseau.

4.3.5 Une seule source mais plusieurs capteurs

Un problème un peu différent correspondrait au cas de plusieurs capteurs qui effectueraient des mesures différentes (avec des modèles différents) d'une seule et même source. Il s'agirait ici également de transmettre le minimum d'information concernant la source à travers le réseau, mais suffisamment pour que le point de collecte puisse la reconstituer. Ce problème a été introduit dans [BZV96] dans le cas où la distribution de la source est bien connue. Ce problème pourrait également être appliqué au cas où on ne chercherait pas à estimer la source, mais, par exemple, ses caractéristiques statistiques.

Bibliographie

[Ahl79a] R. Ahlswede. Coloring hypergraphs : A new approach to multi-user source coding-1. *Journal of Combinatorics*, 4(1) :76–115, 1979.

[Ahl79b] R. Ahlswede. Coloring hypergraphs : A new approach to multi-user source coding-2. *Information and System Sciences*, 4(1) :76–115, 1979.

[AK04] M. Ardakani and F.R. Kschischang. A more accurate one-dimensional analysis and design of irregular LDPC codes. *IEEE Transactions on Communications*, 52(12) :2106–2114, 2004.

[AZG02] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv coding of motion video. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 240–244, 2002.

[BAP06] C. Brites, J. Ascenso, and F. Pereira. Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding. In *Proc. IEEE International Conference on Image Processing*, pages 273–276, 2006.

[Bas10] F. Bassi. *Wyner-Ziv coding with uncertain side information quality.* PhD thesis, Université Paris-Sud, 2010.

[BB06] A. Bennatan and D. Burshtein. Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels. *IEEE Transactions on Information Theory*, 52(2) :549–583, 2006.

[Ber71] T. Berger. The source coding game. *IEEE Transactions on Information Theory*, 17(1) :71–76, 1971.

[BF12] A. Beirami and F. Fekri. On lossless universal compression of distributed identical sources. In *Proc. IEEE International Symposium on Information Theory*, pages 561–565, 2012.

[BKW10] F. Bassi, M. Kieffer, and C. Weidmann. Wyner-Ziv Coding with uncertain side information quality. *Proceedings EUSIPCO*, 2010.

[BPC+07] P. Baronti, P. Pillai, V. Chook, S. Chessa, A. Gotta, and Y. Hu. Wireless sensor networks : A survey on the state of the art and the 802.15. 4 and ZigBee standards. *Computer communications*, 30(7) :1655–1695, 2007.

[BRG02] R. Berry, A. Randall, and R.G. Gallager. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, 48(5) :1135–1149, 2002.

[BS06] J. Barros and S.D. Servetto. Network information flow with correlated sources. *IEEE Transactions on Information Theory*, 52(1) :155–170, 2006.

[BZV96] T. Berger, Z. Zhang, and H. Viswanathan. The ceo problem [multiterminal source coding]. *IEEE Transactions on Information Theory*, 42(3) :887–902, 1996.

[CBLV05] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. Networked Slepian-Wolf : theory, algorithms, and scaling laws. *IEEE Transactions on Information Theory*, 51(12) :4057–4073, 2005.

[CBLVW06] 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*, 14(1) :41–54, 2006.

[CBMW10] C. Chen, B. Bai, X. Ma, and X. Wang. A symbol-reliability based message-passing decoding algorithm for nonbinary LDPC codes over finite fields. In *Proc. 6th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, pages 251–255. IEEE, 2010.

[CF02a] J. Chen and M. Fossorier. Density evolution for BP-based decoding algorithms of LDPC codes and their quantized versions. In *Proc. Global Telecommunications Conference, GLOBECOM*, volume 2, pages 1378–1382, 2002.

[CF02b] J. Chen and M.P.C. Fossorier. Near optimum universal belief propagation based decoding of Low-Density Parity Check codes. *IEEE Transactions on Communications*, 50(3) :406–414, 2002.

[CG12] A. Cano and G. Giannakis. Distributed belief propagation using sensor networks with correlated observations. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2841–2844. IEEE, 2012.

[CGM03] V. Chappelier, C. Guillemot, and S. Marinkovic. Turbo trellis coded quantization. In *Proc. of the Intl. symp. on turbo codes*, 2003.

[CHJ09a] J. Chen, D. He, and A. Jagmohan. On the duality between Slepian–Wolf coding and channel coding under mismatched decoding. *IEEE Transactions on Information Theory*, 55(9) :4006–4018, 2009.

[CHJ09b] J. Chen, D.K. He, and A. Jagmohan. The equivalence between Slepian-Wolf coding and channel coding under density evolution. *IEEE Transactions on Communications*, 57(9) :2534–2540, 2009.

[CLXZS13] W. Chung-Li, C. Xiaoheng, L. Zongwang, and Y. Shaohua. A Simplified Min-Sum Decoding Algorithm for Non-Binary LDPC Codes. *IEEE Transactions on Communications*, 61(1) :24–32, 2013.

[CMRTZ11] G. Coluccia, E. Magli, A. Roumy, and V. Toto-Zarasoa. Lossy Compression of Distributed Sparse Sources : a Practical Scheme. In *Proc. European Signal Processing Conference*, 2011.

[Cov75] T. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Transactions on Information Theory*, 21(2) :226–228, 1975.

[C.P00] C.P. Robert and G. Casella. *Monte-Carlo statistical methods.* Springer Texts in Statistics. Springer, New York, 2000.

[CPPG10] O. Crave, B. Pesquet-Popescu, and C. Guillemot. Robust video coding based on multiple description scalar quantization with side information. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(6) :769–779, 2010.

[CPR03] J. Chou, S. Pradhan, and K. Ramchandran. Turbo and trellis-based constructions for source coding with side information. In *Proc. Data Compression Conference*, pages 33–42, 2003.

[CRU01] S.Y. Chung, T.J. Richardson, and R.L. Urbanke. Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation. *IEEE Transactions on Information Theory*, 47(2) :657–670, 2001.

[CS99] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45(6) :2007–2019, 1999.

[Csi82] I. Csiszar. Linear codes for sources and source networks : Error exponents, universal coding. *IEEE Transactions on Information Theory*, 28(4) :585–592, 1982.

[CT06] T.M. Cover and J.A. Thomas. *Elements of information theory, second Edition.* Wiley, 2006.

[CWC09] S. Cheng, S. Wang, and L. Cui. Adaptive Slepian-Wolf decoding using particle filtering based belief propagation. In *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing*, pages 607–612, 2009.

[CWC12] L. Cui, S. Wang, and S. Cheng. Adaptive Slepian-Wolf decoding based on expectation propagation. *IEEE Communications Letters*, 16(2) :252–255, 2012.

[CX] T. Chu and Z. Xiong. Coding of Gauss-Markov sources with side information at the decoder. In *Proc. of WSSP 2009*, pages 26–29.

[Dav73] L. Davisson. Universal noiseless coding. *IEEE Transactions on Information Theory*, 19(6) :783–795, 1973.

[DF05] D. Declercq and M. Fossorier. Extended minsum algorithm for decoding LDPC codes over GF(q). In *Proc. International Symposium on Information Theory*, pages 464–468, 2005.

[DF07] D. Declercq and M. Fossorier. Decoding Algorithms for Nonbinary LDPC Codes Over GF(q). *IEEE Transactions on Communications*, 55(4) :633–643, 2007.

[DM98] M.C. Davey and D.J.C. MacKay. Low Density Parity Check codes over GF(q). In *Proc. Information Theory Workshop*, pages 70–71, 1998.

[EC82] A. El Gamal and T.M. Cover. Achievable rates for multiple descriptions. *IEEE transaction on Information Theory*, 28(6) :851–557, Nov 1982.

[EKP05] A.W. Eckford, F.R. Kschischang, and S. Pasupathy. Analysis of low-density parity-check codes for the Gilbert-Elliott channel. *IEEE Transactions on Information Theory*, 51(11) :3872–3889, 2005.

[EKP07] A. Eckford, F. Kschischang, and S. Pasupathy. A partial ordering of general finite-state markov channels under ldpc decoding. *IEEE Transactions on Information Theory*, 53(6) :2072–2087, 2007.

[EY05a] A.W. Eckford and W. Yu. Density evolution for the simultaneous decoding of LDPC-based Slepian-Wolf source codes. In *Proceedings of the International Symposium on Information Theory*, pages 1401–1405. IEEE, 2005.

[EY05b] A.W. Eckford and W. Yu. Rateless Slepian-Wolf Codes. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, pages 1757 – 1761, 2005.

[FE06] M. Fleming and M. Effros. On rate-distortion with mixed types of side information. *IEEE Transactions on Information Theory*, 52(4) :1698–1705, 2006.

[FFR06] M. Franceschini, G. Ferrari, and R. Raheli. Does the performance of LDPC codes depend on the channel ? *IEEE Transactions on Communications*, 54(12) :2129–2132, 2006.

[Gal68] R.G. Gallager. *Information theory and reliable communication*. Wiley, 1968.

[Gas04] M. Gastpar. The Wyner-Ziv problem with multiple sources. *IEEE Transactions on Information Theory*, 50(11) :2762–2768, 2004.

[GCGD07] A. Goupil, M. Colas, G. Gelle, and D. Declercq. FFT-based BP decoding of general LDPC codes over Abelian groups. *IEEE Transactions on Communications*, 55(4) :644–649, 2007.

[GD74] R. Gray and L. Davisson. The ergodic decomposition of stationary discrete random processes. *IEEE Transactions on Information Theory*, 20(5) :625–636, 1974.

[GD05] N. Gehrig and P.L. Dragotti. Symmetric and asymmetric Slepian-Wolf codes with systematic and nonsystematic linear codes. *Communications Letters, IEEE*, 9(1) :61–63, 2005.

[GDV06] M. Gastpar, P.L. Dragotti, and M. Vetterli. The distributed Karhunen-Loeve transform. *IEEE Trans. Inf. Th.*, 52(12) :1–10, 2006.

[GF04] J. Garcia-Frias. Decoding of low-density parity-check codes over finite-state binary Markov channels. *IEEE Transactions on Communications*, 52(11) :1840–1843, 2004.

[GSD10] M. Gorgoglione, V. Savin, and D. Declercq. Optimized puncturing distributions for irregular non-binary LDPC codes. In *Proc. International Symposium on Information Theory and its Applications (ISITA)*, pages 400–405. IEEE, 2010.

[Gucsel72] R.M. Gray and Stanford univ calif stanford electronics labs. Conditional rate-distortion theory, 1972.

[Han03] T.S. Han. *Information-spectrum methods in information theory*. Springer, 2003.

[HB85] C. Heegard and T. Berger. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory*, 31(6) :727–734, Nov 1985.

[HKKM07] J. Ha, D. Klinc, J. Kwon, and S.W. McLaughlin. Layered BP decoding for rate-compatible punctured LDPC codes. *IEEE Communications Letters*, 11(5) :440–442, 2007.

[HME+04] T. Ho, M. Médard, M. Effros, R. Koetter, and D.R. Karge. Network coding for correlated sources. In *Proceedings of Conference on Information Sciences and Systems*, 2004.

[HTF09] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning : data mining, inference and prediction*. Springer, 2009.

[HY01] Mark H Hansen and Bin Yu. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454) :746–774, 2001.

[Iwa02] K. Iwata. An information-spectrum approach to rate-distortion function with side information. In *Proc. IEEE International Symposium on Information Theory.*, page 156, 2002.

[JVW10] S. Jalali, S. Verdú, and T. Weissman. A universal scheme for Wyner-Ziv coding of discrete sources. *IEEE Transactions on Information Theory*, 56(4) :1737–1750, 2010.

[Kas94] A.H. Kaspi. Rate-distortion function when side-information may be present at the decoder. *IEEE Transactions on Information Theory*, 40(6) :2031–2034, Nov 1994.

[KFL01] F.R. Kschischang, B.J. Frey, and H-A. Loeliger. Factor graphs and the Sum-Product algorithm. *IEEE Transactions on Information Theory*, 47(2) :498–519, 2001.

[KHM08] D. Klinc, J. Ha, and S.W. McLaughlin. On rate-adaptability of non-binary LDPC codes. In *Proc. 5th International Symposium on Turbo Codes and Related Topics*, pages 231–236, 2008.

[KKU10] S. Kuzuoka, A. Kimura, and T. Uyematsu. Universal source coding for multiple decoders with side information. In *IEEE International Symposium on Information Theory Proceedings*, pages 1–5. IEEE, 2010.

[LCLX06] Z. Liu, S. Cheng, A.D. Liveris, and Z. Xiong. Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding : High-rate performance analysis and code design. *IEEE Transactions on Information Theory*, 52(10) :4358–4379, 2006.

[LFK09] G. Li, I.J. Fair, and W.A. Krzymien. Density evolution for nonbinary LDPC codes under Gaussian approximation. *IEEE Transactions on Information Theory*, 55(3) :997–1015, 2009.

[LGTD11] B. Liu, J. Gao, W. Tao, and G. Dou. Weighted symbol-flipping decoding algorithm for nonbinary LDPC codes with flipping patterns. *Journal of Systems Engineering and Electronics*, 22(5) :848–855, 2011.

[LW08] G. Lechner and C. Weidmann. Optimization of binary LDPC codes for the q-ary symmetric channel with moderate q. In *Proc. International Symposium on Turbo Codes and Related Topics*, pages 221–224, 2008.

[LXG02] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6 :440–442, 2002.

[Mac99] D.J. C. MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45(2) :399–431, 1999.

[MBD89] M. Mushkin and I. Bar-David. Capacity and coding for the Gilbert-Elliot channels. *IEEE Transactions on Information Theory*, 35(6) :1277–1290, 1989.

[Med00] M. Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, 46(3) :933–946, 2000.

[MGPPG10] T. Maugey, J. Gauthier, B. Pesquet-Popescu, and C. Guillemot. Using an exponential power model for Wyner Ziv video coding. In *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, pages 2338–2341, 2010.

[MS77] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-correcting Codes*, volume 16. Elsevier, 1977.

[MUM10a] T. Matsuta, T. Uyematsu, and R. Matsumoto. Universal Slepian-Wolf source codes using Low-Density Parity-Check matrices. In *Proc. IEEE International Symposium on Information Theory*, pages 186–190, june 2010.

[MUM10b] T. Matsuta, T. Uyematsu, and R. Matsumoto. Universal Slepian-Wolf source codes using low-density parity-check matrices. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 93(11) :1878–1888, 2010.

[MWGL05] P.F.A. Meyer, R.P. Westerlaken, R.K. Gunnewiek, and R.L. Lagendijk. Distributed source coding of video with non-stationary side-information. In *Proceedings of SPIE*, volume 5960, page 59602J, 2005.

[MWZ08] E. Martinian, G.W. Wornell, and R. Zamir. Source coding with distortion side information. *IEEE Transactions on Information Theory*, 54(10) :4638–4665, 2008.

[MZ06] N. Merhav and J. Ziv. On the Wyner-Ziv problem for individual sequences. *IEEE Transactions on Information Theory*, 52(3) :867–873, 2006.

[NBK04] R. Nabar, H. Bolcskei, and F. Kneubuhler. Fading relay channels : Performance limits and space-time signal design. *IEEE Journal on Selected Areas in Communications*, 22(6) :1099–1109, 2004.

[Ooh99] Y. Oohama. Multiterminal source coding for correlated memoryless Gaussian sources with several side information at the decoder. In *Proc. Information Theory and Communications Workshop*, page 100, 1999.

[PDDV12] S.K. Planjery, D. Declercq, L. Danjean, and B. Vasic. Finite Alphabet Iterative Decoders, Part I : Decoding Beyond Belief Propagation on BSC. *arXiv preprint arXiv :1207.4800*, 2012.

[PFD08] C. Poulliat, M. Fossorier, and D. Declercq. Design of regular $(2, d_c)$-LDPC codes over GF(q) using their binary images. *IEEE Transactions on Communications*, 56(10) :1626–1635, october 2008.

[PL13] R. Parseh and F. Lahouti. Multi-mode nested quantization in presence of uncertain side information and feedback. *IEEE Transactions on Communications*, 61(2) :743–752, 2013.

[PMD09] P. Piantanida, G. Matz, and P. Duhamel. Outage behavior of discrete memoryless channels under channel estimation errors. *IEEE Transactions on Information Theory*, 55(9) :4221–4239, Sep 2009.

[PR02] R. Puri and K. Ramchandran. PRISM : A new robust video coding architecture based on distributed compression principles. In *Proc. Annual Allerton Conference on Communications Control and Computing*, volume 40, pages 586–595, 2002.

[PR03] S.P. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS) : Design and construction. *IEEE Transactions on Information Theory*, 49(3) :626–643, 2003.

[Pra04] S. Pradhan. Source coding with feedforward : Gaussian sources. In *Proc. IEEE International Symposium on Information Theory*, pages 212–212, 2004.

[PRW06] F. Peng, W. Ryan, and R.D. Wesel. Surrogate-channel design of universal LDPC codes. *IEEE Communications Letters*, 10(6) :480–482, 2006.

[PSD09] P. Piantanida, S. Sadough, and P. Duhamel. On the outage capacity of a practical decoder accounting for channel estimation inaccuracies. *IEEE Transactions on Communications*, 57(5) :1341–1350, 2009.

[PY04] R. Palanki and J.S. Yedidia. Rateless codes on noisy channels. In *Proceedings. International Symposium on Information Theory*, page 37, 2004.

[Rab89] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2) :257–286, February 1989.

[Ric03] T. Richardson. Error floors of LDPC codes. In *Proceedings of the annual Allerton conference on communication control and computing*, volume 41, pages 1426–1435. The University ; 1998, 2003.

[RJCE06] A. Ramamoorthy, K. Jain, P.A. Chou, and M. Effros. Separating distributed source coding from network coding. *IEEE/ACM Transactions on Networking (TON)*, 14(SI) :2785–2795, 2006.

[RMRAG06] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod. High-rate quantization and transform coding with side information at the decoder. *EURASIP Journal on Applied Signal Processing (Special Issue on Distributed Source Coding)*, 86(11) :3160–3179, 2006.

[RSU01] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke. Design of capacity-approaching irregular Low-Density Parity-Check codes. *IEEE Transactions on Information Theory*, 47(2) :619–637, 2001.

[RU01] T.J. Richardson and R.L. Urbanke. The capacity of Low-Density Parity-Check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2) :599–618, 2001.

[RU05] V. Rathi and R. Urbanke. Density evolution, thresholds and the stability condition for non-binary LDPC codes. In *IEEE Transactions on Communications*, volume 152, pages 1069–1074, 2005.

[RU08] T.J. Richardson and R.L. Urbanke. *Modern coding theory*. Cambridge Univ Press, 2008.

[Rub76] I. Rubin. Data compression for communication networks : The delay-distortion function. *IEEE Transactions on Information Theory*, 22(6) :655–665, 1976.

[RW11] S. Ramanan and J.M. Walsh. Practical codes for lossy compression when side information may be absent. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3048–3051, 2011.

[Sav08a] V. Savin. Min-Max decoding for non binary LDPC codes. In *Proc. IEEE International Symposium on Information Theory*, pages 960–964, 2008.

[Sav08b] V. Savin. Non binary LDPC codes over the binary erasure channel : density evolution analysis. In *First International Symposium on Applied Sciences on Biomedical and Communication Technologies*, pages 1–5, 2008.

[Sav08c] V. Savin. Self-corrected Min-Sum decoding of LDPC codes. In *Proc. IEEE International Symposium on Information Theory*, pages 146–150, 2008.

[SB09] H. Saeedi and A. Banihashemi. Design of irregular ldpc codes for biawgn channels with snr mismatch. *IEEE Transactions on Communications*, 57(1) :6–11, 2009.

[SG12] A.D. Sarwate and M. Gastpar. Relaxing the Gaussian AVC. *arXiv preprint arXiv :1209.2755*, 2012.

[Sga77] A. Sgarro. Source coding with side information at several decoders. *IEEE Transactions on Information Theory*, 23(2) :179–182, 1977.

[SLXG04] V. Stankovic, A.D. Liveris, Z. Xiong, and C.N. Georghiades. Design of Slepian-Wolf codes by channel code partitioning. In *Proc. Data Compression Conference*, pages 302 – 311, 2004.

[SP13] O. Simeone and H.H. Permuter. Source coding when the side information may be delayed. *IEEE Transactions on Information Theory*, 59(6) :3607–3618, 2013.

[SPZ08] P. Schniter, L.C. Potter, and J. Ziniel. Fast Bayesian matching pursuit. In *Proc. Information Theory and Applications Workshop*, pages 326–333, 2008.

[SRA08] A. Sanaei, M. Ramezani, and M. Ardakani. On the design of universal LDPC codes. In *Proc. IEEE International Symposium on nformation Theory*, pages 802–806, 2008.

[SSWC11] L. Stankovic, V. Stankovic, S. Wang, and S. Cheng. Correlation estimation with particle-based belief propagation for distributed video coding. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1505–1508. IEEE, 2011.

[SSZ02] I. Sutskover, S. Shamai, and J. Ziv. Iterative decoding of low-density parity check codes over compound channels. In *Proc. of the 22nd Convention of Electrical and Electronics Engineers in Israel*, pages 265–267, 2002.

[SW73] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4) :471–480, July 1973.

[Ten04] D. Teneketzis. Optimal real-time encoding-decoding of markov sources in noisy environments. *Proc. Math. Theory of Networks and Sys.(MTNS), Leuwen, Belgium*, 2004.

[TL11] M. Le Treust and S. Lasaulce. Resilient source coding. In *Proc. 5th International Conference on Network Games, Control and Optimization*, pages 1–6, 2011.

[TZRG10a] V. Toto-Zarasoa, A. Roumy, and C. Guillemot. Hidden Markov model for distributed video coding. *Proc. International Conference on Image Processing*, pages 3709–3712, 2010.

[TZRG10b] V. Toto-Zarasoa, A. Roumy, and C. Guillemot. Non-uniform source modeling for distributed video coding. In *Proc. 18th European Signal Processing Conference*, 2010.

[TZRG11] V. Toto-Zarasoa, A. Roumy, and C. Guillemot. Maximum Likelihood BSC Parameter Estimation for the Slepian-Wolf Problem. *IEEE Communications Letters*, 15(2) :232 –234, 2011.

[UK03] T. Uyematsu and S. Kuzuoka. Conditional Lempel-Ziv complexity and its application to source coding theorem with side information. In *Proc. IEEE International Symposium on Information Theory*, pages 142–145, 2003.

[Uye01a] T. Uyematsu. An algebraic construction of codes for Slepian-Wolf source networks. *IEEE Transactions on Information Theory*, 47(7) :3082–3088, 2001.

[Uye01b] T. Uyematsu. Universal coding for correlated sources with memory. In *Proc. Canadian Workshop Information Theory*. Citeseer, 2001.

[VAG06] D. Varodayan, A. Aaron, and B. Girod. Rate-adaptive codes for distributed source coding. *EURASIP Signal Processing*, 86(11) :3123–3130, 2006.

[VAR10] K. Viswanatha, E. Akyol, and K. Rose. Towards optimum cost in multi-hop networks with arbitrary network demands. In *Proc. IEEE International Symposium on Information Theory Proceedings*, pages 1833–1837. IEEE, 2010.

[VCFG08] D. Varodayan, D. Chen, M. Flierl, and B. Girod. Wyner–Ziv coding of video with unsupervised motion vector learning. *Signal Processing : Image Communication*, 23(5) :369–378, 2008.

[VCNP09] B. Vasic, S.K. Chilappagari, D.V. Nguyen, and S.K. Planjery. Trapping set ontology. In *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing*, pages 1–7, 2009.

[VL12] M. Vaezi and F. Labeau. Improved modeling of the correlation between continuous-valued sources in LDPC-based DSC. In *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1659–1663, 2012.

[VMFG07] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod. Distributed grayscale stereo image coding with unsupervised learning of disparity. In *Proc. Data Compression Conference*, pages 143–152, 2007.

[VP07] R. Venkataramanan and S. Pradhan. Source coding with feed-forward : Rate-distortion theorems and error exponents for a general source. *IEEE Transactions on Information Theory*, 53(6) :2154–2179, 2007.

[VP10] J. Villard and P. Piantanida. Secure lossy source coding with side information at the decoders. In *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing*, pages 733–739. IEEE, 2010.

[WCS+12] C. Wang, L. Cui, L. Stankovic, V. Stankovic, and S. Cheng. Adaptive correlation estimation with particle filtering for distributed video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(5) :649–658, 2012.

[Wei08] C. Weidmann. Coding for the q-ary symmetric channel with moderate q. In *Proc. IEEE International Symposium on Information Theory*, pages 2156–2159, 2008.

[WG06] T. Weissman and A. El Gamal. Source coding with limited-look-ahead side information at the decoder. *IEEE Transactions on Information Theory*, 52(12) :5218–5239, 2006.

[WK13] S. Watanabe and S. Kuzuoka. Universal Wyner-Ziv coding for distortion constrained general side-information. *arXiv preprint arXiv :1302.0050*, 2013.

[WKP05] C.C. Wang, S.R. Kulkarni, and H.V. Poor. Density evolution for asymmetric memoryless channels. *IEEE Transactions on Information Theory*, 51(12) :4216–4236, 2005.

[WLZ08] Z. Wang, X. Li, and M. Zhao. Distributed coding of Gaussian correlated sources using non-binary LDPC. In *Congress on Image and Signal Processing*, volume 2, pages 214–218. IEEE, 2008.

[Wyn78] A.D. Wyner. The rate-distortion function for source coding with side information at the decoder\ 3-II : General sources. *Information and Control*, 38(1) :60–80, 1978.

[WZ76] A. Wyner and J. Ziv. The rate-distorsion function for source coding with side information at the decoder. *IEEE Trans. on Inf. Th.*, 22(1) :1–10, Jan 1976.

[XLC04] Z. Xiong, A.D. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5) :80–94, Sep 2004.

[YH10] E.H. Yang and D.K. He. Interactive encoding and decoding for one way learning : Near lossless recovery with side information at the decoder. *IEEE Transactions on Information Theory*, 56(4) :1808–1824, 2010.

[YXZ09] Y. Yang, S. Cheng Z. Xiong, and W. Zhao. Wyner-Ziv coding based on TCQ and LDPC codes. *IEEE Transactions on Communications*, 57(2) :376–387, 2009.

[YZQ07] S. Yang, M. Zhao, and P. Qiu. On Wyner-Ziv Problem for general sources with average distortion criterion. *Journal of Zhejiang University science A*, 8(8) :1263–1270, 2007.

[Zam96] R. Zamir. The rate-loss in the Wyner-Ziv problem. *IEEE Trans. on Inf. Th.*, 42(6) :2073 – 2084, Nov 1996.

[ZMZ+12] K. Zhang, X. Ma, S. Zhao, B. Bai, and X. Zhang. A new ensemble of rate-compatible LDPC codes. In *Proc. IEEE International Symposium on Information Theory*, pages 2536–2540, 2012.

[ZRS07] A. Zia, J.P. Reilly, and S. Shirani. Distributed parameter estimation with side information : A factor graph approach. In *Proc. IEEE International Symposium on Information Theory*, pages 2556–2560, 2007.

[ZSE02] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Transactions on Information Theory*, 48(6) :1250–1276, 2002.

Annexes

Annexe A

Analyse de performances

Elsa Dupraz, Aline Roumy, Michel Kieffer *Coding strategies for source coding with side information and uncertain knowledge of the correlation* Submitted at IEEE Transactions on Information Theory, March 2013

Une partie de cet article a été publié dans

Elsa Dupraz, Aline Roumy, Michel Kieffer *Codage distribué dans des réseaux de capteurs avec connaissance incertaine des corrélations.* Accepted at GRETSI 2013

Coding Strategies for Source Coding with Side Information and Uncertain Knowledge of the Correlation

Elsa DUPRAZ[1], Aline ROUMY[2] and Michel KIEFFER[1,3,4]

[1] L2S - CNRS - SUPELEC - Univ Paris-Sud, 91192 Gif-sur-Yvette, France

[2] INRIA, Campus de Beaulieu, 35042 Rennes, France

[3] Partly on leave at LTCI – CNRS Télécom ParisTech, 75013 Paris, France

[4] Institut Universitaire de France

Abstract

In this paper, lossless source coding with side information at the decoder is studied when the joint distribution of the sources is not perfectly known. Four uncertainty models are introduced that differ through the possible variations (instantaneous or quasi-static) of the statistics and through the amount of *a priori* on the law. It is shown that classical results on distributed source coding no longer holds: when the source distribution is not perfectly known, there might be a discrepancy between the schemes with and without side information at the encoder. To bridge the gap between the two strategies, estimation of the parameters is considered, and it is shown that, only for fast varying statistics, the discrepancy might be reduced. Moreover, outage strategies and their design are proposed to mitigate the large increase in the coding rate of non-cooperative schemes. Finally, these results are applied to a sensor network where they can help in choosing the source coding strategy that minimizes the total transmission cost.

I. INTRODUCTION

The problem of lossless source coding with side information (SI) at the decoder only, also called the Slepian-Wolf (SW) problem, has been well investigated when the correlation between the source X and the SI Y is perfectly known. In this case, there is no loss in performance compared to the conditional

setup, *i.e.*, the setup where the SI is also known at the encoder [23]. This justified the interest for SW coding and led to the development of many practical coding solutions, see *e.g.*, [18], [19], [21], [26].

Nonetheless, most of these works assume perfect knowledge of the joint distribution $P(X,Y)$ between the source and the SI. In [17], it is shown that the performance of the SW coding scheme remains the same if the probability distribution $P(X)$ is unknown. Here we mainly focus on uncertain correlation channels $P(Y|X)$, which corresponds to the case where the correlation is more difficult to obtain. For instance, when considering data collection in a sensor network, the correlation depends on the position of the sensors, the diffusion of the measured field, etc. In this way, [22] considers that the correlation channel is given to the decoder but is not perfectly known at the encoder. Here we assume that the correlation channel is uncertain at both the encoder and the decoder. A usual solution to address this problem is to use a feedback channel [1] or to allow interactions between the encoder and the decoder [27]. However, these solutions are difficult to implement in practical situations such as sensor networks. Here we consider coding without feedback.

Four signal models are introduced in this paper to take into account the uncertainty on the knowledge of the correlation channel and to capture its variations. For two of these models, the correlation channel is parametrized by an unknown parameter $\boldsymbol{\theta}$, fixed for a sequence $\{(X_n,Y_n)\}_{n=1}^{+\infty}$ but allowed to vary from sequence to sequence. One of these two models assumes the knowledge of a prior distribution $P_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$. These models are called respectively Static with Prior (SP) and Static without Prior (SwP). For the two other models, the joint distribution of source symbols (X_n,Y_n) is parametrized by an unknown parameter $\boldsymbol{\pi}_n$, varying from symbol to symbol. Again the two models differ through the knowledge of a prior for $\boldsymbol{\pi}_n$. These models called respectively Dynamic with Prior (DP) and Dynamic without Prior (DwP) can account for the instantaneous variations of the statistics of the sources. The distinction between a fixed parameter $\boldsymbol{\theta}$ and varying parameters $\boldsymbol{\pi}_n$ has been presented earlier in [20] in the case of channel coding.

Most of the introduced signal models are non-ergodic (SP, SwP, DwP) and/or non-stationary (SwP, DwP). As the DP-Source is stationary and ergodic, classical results from information theory such as the original theorem of Slepian and Wolf [23] apply. The SP-Source is non-ergodic but can be separated into ergodic components, from the source definition and results of [14]. Moreover, the DP- and SP-Sources fall in the context of general sources [16], unlike the SwP- and DwP-Sources. Indeed, for the last two

sources, there is no probability distribution to represent the choice of the unknown parameters $\boldsymbol{\theta}$ or $\boldsymbol{\pi}_n$. Thus, the probability model describing the sources (see [14] for an example of source definition) is not completely defined. Another closely related notion is universal coding [13], [19]. Universal coding results show that for a given rate R, there exists a code such that successful decoding is guaranteed for every source (X, Y) of conditional entropy less than R. Furthermore, this code does not require the knowledge of the joint distribution $P(X, Y)$ to perform good. On the opposite, in our context, the source models are chosen to represent the uncertainty on the correlation between X and Y, and we look for the infimum of achievable rates R for these models. To finish, the DwP-Sources is an Arbitrarily Varying Source as introduced in [2], [3].

Different coding strategies can be compared for the previously defined models: fixed-length (FL) *versus* variable-length (VL) coding, conditional *versus* SW coding, estimation of the parameters or not, authorization for the decoder to fail for some parameter values. The contribution of the paper are threefold. First, we survey, while adapting to our context, information theoretical results scattered throughout the existing literature. Second, we complete the analysis of the missing cases. Finally, these results are applied to optimize the coding strategy in a sensor network. More precisely, we first compare FL and VL coding for the SW setup. In VL coding, the coding rate can depend on the input sequence, whereas in FL coding, the rate is fixed before the coding process and therefore remains unchanged for all possible source sequences. In classical SW, the joint distribution is known, and FL and VL codings lead to the same coding rate. For SW coding with uncertain joint distribution knowledge, we show that the equivalence of FL and VL coding remains for all proposed models. By contrast, FL and VL provide different coding rates in the conditional scheme, except for the DP-Model. This is explained by the fact that the latter model is ergodic. Thus, only in the ergodic DP-Model, the SW and VL conditional coding schemes achieve the same compression ratio. For all the other non-ergodic models, the compression ratios of SW coding are given by worst cases defined on the set of possible parameters. Instead, the conditional VL coding scheme reaches the same rate as if the true joint distribution were known.

To bridge the gap between SW and conditional schemes, we then consider that estimates $\hat{\boldsymbol{\theta}}$ or $\hat{\boldsymbol{\pi}}$ of the parameters are available at the decoder. Interestingly, we show that having estimates may reduce the SW coding rate only if the statistics of the source vary at each symbol (Dynamic models). Second, as

the infimums of achievable rates are given by worst cases on the sets of parameters, if the sets are big, the coding rate difference between SW and conditional schemes can be important. Thus, one can tolerate some outage level γ, *i.e.*, authorize the decoder to fail for a proportion γ of source parameters and evaluate the impact of γ on the infimum of achievable rates R. The converse problem is also considered: for a given rate constraint R, one determines the proportion γ of parameters for which the decoder fails. Here, the tradeoff between the outage level γ and the infimum of achievable rates R is expressed in the two cases.

Finally, we propose an application of the results of the paper to a three-node network. Several coding strategies are analyzed and tools to compare their performance are provided. In particular, we explain how to choose a coding strategy (SW or conditional, with or without learning sequence to estimate the parameters) that minimizes the total transmission cost of the network.

The paper is organized as follows. Section II gives formal definitions of the source models we consider. Section III introduces the notions of FL coding and VL coding and give the SW performance for the four models in each case. Section IV, compares the conditional and SW setups. Section V provides an analysis of the case where estimated parameters are available at the decoder. Section VI expresses the tradeoff between outage level and achievable rate. Section VII presents the three-node network example and shows how the tools of the paper can be helpful to determine an optimum coding strategy.

II. Signal Model

The source X to be compressed and the SI Y available at the decoder produce sequences of symbols $\{X_n\}_{n=1}^{+\infty}$ and $\{Y_n\}_{n=1}^{+\infty}$, respectively. $\mathcal{X}$ and $\mathcal{Y}$ denote the source and SI discrete alphabets. The n-th extension of a set $\mathcal{S}$ is denoted $\mathcal{S}^n$. Bold upper case letters $\mathbf{X}_1^N = \{X_n\}_{n=1}^N$, denote random vectors, whereas bold lower case letters, $\mathbf{x}_1^N = \{x_n\}_{n=1}^N$, represent their realizations. When it is clear from the context that the distribution of a random variable X_n does not depend on n, the index n is omitted.

The goal of this section is to model the uncertainty on the correlation channel $P(Y|X)$. Each of the four proposed models consists of a family of parametric distributions. In each case, the source distribution $P(X)$ is assumed to be perfectly known and does not depend on the uncertain parameters. The first two models consider a time-invariant parameter.

Definition 1. (SP-Source) *A* Static with Prior source (X, Y) (SP-Source) *produces a sequence of* ***independent*** *symbols* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ *drawn from a distribution belonging to a family* $\{P(X, Y|\mathbf{\Theta} = \boldsymbol{\theta}) = P(X)P(Y|X, \mathbf{\Theta} = \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *parametrized by a* ***random vector*** $\mathbf{\Theta}$*. The random vector* $\mathbf{\Theta}$*, with distribution* $P_{\mathbf{\Theta}}(\boldsymbol{\theta})$*, takes its value in a set* $\mathcal{P}_{\boldsymbol{\theta}}$ *that is either* ***discrete or continuous****. The source symbols* X *and* Y *take their values in the discrete sets* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively.. Moreover, the* ***realization*** $\boldsymbol{\theta}$ ***of the parameter*** $\mathbf{\Theta}$ ***is fixed*** *for the sequence* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$.

The SP-source, determined by $\mathcal{P}_{\boldsymbol{\theta}}$, $P_{\mathbf{\Theta}}(\boldsymbol{\theta})$, and $\{P(X, Y|\mathbf{\Theta} = \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$, is stationary but non-ergodic [12, Section 3.5].

Definition 2. (SwP-Source). *A* Static without Prior source (X, Y) (SwP-Source) *produces a sequence of* ***independent*** *symbols* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ *drawn from a distribution belonging to a family* $\{P(X, Y|\boldsymbol{\theta}) = P(X)P(Y|X, \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *parametrized by a vector* $\boldsymbol{\theta}$*. The vector* $\boldsymbol{\theta}$ *takes its value in a set* $\mathcal{P}_{\boldsymbol{\theta}}$ *that is either* ***discrete or continuous****. The source symbols* X *and* Y *take their values in the discrete sets* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively. Moreover, the* ***parameter*** $\boldsymbol{\theta}$ ***is fixed*** *for the sequence* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$.

The SwP-source, completely determined by $\mathcal{P}_{\boldsymbol{\theta}}$ and $\{P(X, Y|\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$, is non-stationary and non-ergodic [12, Section 3.5]. The only difference between the SP- and SwP-Sources lies in the definition of $\boldsymbol{\theta}$ (no distribution for $\boldsymbol{\theta}$ is specified in the SwP-Model).

The last two models allow parameter variations from symbol to symbol.

Definition 3. (DP-Source). *A* Dynamic with Prior source (X, Y)*, or* DP-Source*, produces a sequence of* ***independent*** *symbols* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ *, drawn* $\forall n$ *from* $P(X_n, Y_n)$ *that belongs to a family of distributions* $\{P(X, Y|\mathbf{\Pi} = \boldsymbol{\pi}) = P(X)P(Y|X, \mathbf{\Pi} = \boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}}$ ***parametrized by a random vector*** $\mathbf{\Pi}_n$*. The* $\{\mathbf{\Pi}_n\}_{n=1}^{+\infty}$ *are i.i.d. with distribution* $P(\mathbf{\Pi})$ *and take their values in a* ***discrete set*** $\mathcal{P}_{\boldsymbol{\pi}}$*. The source symbols* X_n *and* Y_n *take their values in the discrete sets* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively.*

The DP-Source, completely determined by $\mathcal{P}_{\boldsymbol{\pi}}$, $P(\mathbf{\Pi})$, and $\{P(X, Y|\mathbf{\Pi} = \boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}}$, is stationary and ergodic, see [12, Section 3.5].

Definition 4. (DwP-Source). *A* Dynamic without Prior source (X, Y)*, or* DwP-Source*, produces a sequence of* ***independent*** *symbols* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ *drawn* $\forall n$ *from* $P(X_n, Y_n)$ *that belongs to a family of*

distributions $\{P(X,Y|\boldsymbol{\pi}) = P(X)P(Y|X,\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}}$ ***parametrized by a vector*** $\boldsymbol{\pi}_n$*. Each* $\boldsymbol{\pi}_n$ *takes its values in a* ***discrete set*** $\mathcal{P}_{\boldsymbol{\pi}}$*. The source symbols* X_n *and* Y_n *take their values in the discrete sets* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively.*

The DwP-Source, determined by $\mathcal{P}_{\boldsymbol{\pi}}$ and $\{P(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}}$, is non-stationary and non-ergodic [12, Section 3.5]. The only difference between the DP- and DwP-Sources lies in the definition of the parameters $\boldsymbol{\pi}_n$. In the DwP-Source, no distribution for $\boldsymbol{\pi}_n$ is specified, either because its distribution is not known or because $\boldsymbol{\pi}_n$ is not modeled as a random variable.

III. FL AND VL CODING

This section considers the notions of FL and VL coding and gives the performance of the SW setup in both cases. In FL coding, the rate of the code is fixed and does not depend on the realization of the source sequence $\{X_n\}_{n=1}^{+\infty}$ to encode. Instead in VL coding, the rate of the code can depend on the source sequence to encode. For a stationary and ergodic model representing the correlation between X and Y, [4] shows that in SW coding, the optimal performance is the same in both FL and VL coding. This is also true for our models, as will be seen.

We now look for the minimum rate needed to transmit a source sequence $\{X_n\}_{n=1}^{+\infty}$ with vanishing error probability, when $\{Y_n\}_{n=1}^{+\infty}$ is available at the decoder. We first give a preliminary definition for the DwP-Source and then define achievable rates for FL and VL coding for the SW setup for the four models.

Definition 5. (Preliminary definition) *Let* (X,Y) *be a DwP-Source. A sequence of distributions* $\{p^n\}_{n\in\mathbb{N}}$ *for* (X,Y) *is called* possible*, if there exists a sequence of parameters* $\{\boldsymbol{\pi}_n\}_{n\in\mathbb{N}}$*,* $\boldsymbol{\pi}_n \in \mathcal{P}_{\boldsymbol{\pi}}$ *such that*

$$\forall n \in \mathbb{N}, \quad p^n(\mathbf{X},\mathbf{Y}) = P(\mathbf{X},\mathbf{Y}|p^n) = \prod_{k=1}^{n} P(X_k,Y_k|\boldsymbol{\pi}_k). \tag{1}$$

The set of all possible distributions p^n *at length* n *is denoted* $\mathcal{P}_n$*, and the set of all possible sequences* $\{p^n\}_{n\in\mathbb{N}}$*, is denoted* $\mathcal{P}^\star$*.*

Note that, if the sequence $\{p^n\}_{n\in\mathbb{N}}$ is possible, then p^{n-1} is obtained by marginalizing p^n over (X_n,Y_n).

Definition 6. (Achievable rates for the SW setup, FL coding) *Let* $\mathcal{M}_n = \{1\dots|\mathcal{M}_n|\}$ *be a set of integers. A rate* R *is said to be* achievable *for the SW setup if there exists two sequences of mappings*

$\{\phi_n : \mathcal{X}^n \to \mathcal{M}_n\}_{n\in\mathbb{N}}$ *and* $\{\psi_n : \mathcal{M}_n \times \mathcal{Y}^n \to \mathcal{X}^n\}_{n\in\mathbb{N}}$ *such that* $\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{M}_n| \leq R$ *and the error probabilities satisfy*

1) *for the SP-Source and the DP-Source,*

$$\lim_{n\to\infty} P_e^n = \lim_{n\to\infty} P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{X}), \mathbf{Y})\right) = 0, \tag{2}$$

2) *for the SwP-Source,*

$$\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}, \quad \lim_{n\to\infty} P_e^n(\boldsymbol{\theta}) = \lim_{n\to\infty} P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{X}), \mathbf{Y})|\boldsymbol{\theta}\right) = 0, \tag{3}$$

3) *for the DwP-Source,*

$$\forall \{p^n\}_{n\in\mathbb{N}} \in \mathcal{P}^\star, \quad \lim_{n\to\infty} P_e^n(p^n) = P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{Y}), \mathbf{Y})|p^n\right) = 0. \tag{4}$$

Denote the infimum of achievable rates for the SW setup with FL coding respectively $R_{f,SW}^{SP}$, $R_{f,SW}^{DP}$, $R_{f,SW}^{SwP}$, *and* $R_{f,SW}^{DwP}$.

The definitions for VL coding are now provided.

Definition 7. (Achievable rates for the SW setup, VL coding) *Let* $\mathcal{U} = \{1 \dots K\}$ *be a set of integers and denote by* $\mathcal{U}^\star$ *the set of all the strings of finite length over* $\mathcal{U}$. *Let* $\phi_n : \mathcal{X}^n \to \mathcal{U}^\star$ *and* $\psi_n : \mathcal{U}^\star \times \mathcal{Y}^n \to \mathcal{X}^n$. *Denote by* $E_{\boldsymbol{\theta}}[.]$ *the expectation with respect to the true parameter* $\boldsymbol{\theta}$. *Then*

1) *for the SP-Source, a set of rates* $\{R(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *is said to be* achievable *if there exists two sequences of mappings* $\{\phi_n, \psi_n\}_{n\in\mathbb{N}}$ *such that* $\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}$, *(i.e. whatever the true parameter* $\boldsymbol{\theta}$ *is),* $\limsup_{n\to\infty} \frac{1}{n} E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X})|] \leq R(\boldsymbol{\theta})$ *and*

$$\lim_{n\to\infty} P_e^n = \lim_{n\to\infty} P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{X}), \mathbf{Y})\right) = 0, \tag{5}$$

2) *for the SwP-Source, a set of rates* $\{R(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *is said to be* achievable *if there exists two sequences of mappings* $\{\phi_n, \psi_n\}_{n\in\mathbb{N}}$ *such that* $\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}$, $\limsup_{n\to\infty} \frac{1}{n} E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X})|] \leq R(\boldsymbol{\theta})$ *and*

$$\lim_{n\to\infty} P_e^n(\boldsymbol{\theta}) = \lim_{n\to\infty} P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{X}), \mathbf{Y})|\boldsymbol{\theta}\right) = 0, \tag{6}$$

3) *for the DP-Source, a rate* R *is said to be* achievable *if there exists two sequences of mappings* $\{\phi_n, \psi_n\}_{n\in\mathbb{N}}$ *such that* $\limsup_{n\to\infty} \frac{1}{n} E[|\phi_n(\mathbf{X})|] \leq R$ *and*

$$\lim_{n\to\infty} P_e^n = \lim_{n\to\infty} P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{X}), \mathbf{Y})\right) = 0, \tag{7}$$

4) *for the DwP-Source, a set of rates* $\{R(\{p^n\}_{n\in\mathbb{N}})\}$ *is said to be* achievable *if there exists two sequences of mappings* $\{\phi_n, \psi_n\}_{n\in\mathbb{N}}$ *such that* $\forall\{p^n\}_{n\in\mathbb{N}} \in \mathcal{P}^\star$ *(see Definition 5),* $\limsup_{n\to\infty} \frac{1}{n} E_{p^n}[|\phi_n(\mathbf{X})|] \leq R(\{p^n\}_{n\in\mathbb{N}})$ *and*

$$\limsup_{n\to\infty} P_e^n(p^n) = \limsup_{n\to\infty} P\left(\mathbf{X} \neq \psi_n(\phi_n(\mathbf{X}), \mathbf{Y})|p^n\right) = 0. \tag{8}$$

For the DP-Source, the infimum of achievable rates is denoted $R_{v,SW}^{DP}$.

For the SP-, SwP- and DwP-Sources, a set of infimums of achievable rates, if exists, is a set of achievable rates in which each component takes its infimum possible value. The sets of infimums of achievable rates are denoted respectively $\{R_{v,SW}^{SP}(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$, $\{R_{v,SW}^{SwP}(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *and* $\{R_{v,SW}^{DwP}(\{p^n\}_{n\in\mathbb{N}})\}$.

In the following theorem, we compare FL and VL coding, when the sources are SW encoded. We give the infimums of achievable rates for each model and show that there is no difference between FL and VL coding. The definition of the essential sup, that will appear in the formulation of the theorem, is given first.

Definition 8. (Essential sup [16]) *Let* (X, Y) *be a SP-Source and let* $Z(\boldsymbol{\Theta})$ *be a random variable taking its values in the set* $\mathcal{P}_Z = \{H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *with probability distribution* $P_{\boldsymbol{\Theta}}$. *Then the essential sup of* $Z(\boldsymbol{\Theta})$ *with respect to* $P_{\boldsymbol{\Theta}}$ *is*

$$P_{\boldsymbol{\Theta}}\text{-}ess.\sup Z(\boldsymbol{\Theta}) = \inf_{z\in\mathcal{P}_Z} \{z|P(Z(\boldsymbol{\Theta}) > z) = 0\}\ . \tag{9}$$

Theorem 1. *In the SW setup, the sets of infimums of achievable rates in VL coding, introduced in Definition 7, are given by*

1) *for the SP-Source [16, Theorem 7.3.4],*

$$\forall\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}},\quad R_{v,SW}^{SP}(\boldsymbol{\theta}) = P_{\boldsymbol{\Theta}}\text{-}ess.\sup Z(\boldsymbol{\Theta}) = R_{v,SW}^{SP} \tag{10}$$

where $Z(\boldsymbol{\Theta})$ *is introduced in Definition 8,*

2) *for the SwP-Source [24],*

$$\forall\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}},\quad R_{v,SW}^{SwP}(\boldsymbol{\theta}) = \sup_{\boldsymbol{\alpha}\in\mathcal{P}_{\boldsymbol{\theta}}} H(X|Y, \boldsymbol{\alpha}) = R_{v,SW}^{SwP} \tag{11}$$

3) *for the DP-Source [23],* $R_{v,SW}^{DP} = H(X|Y)$,

4) *for the DwP-Source [2],*

$$\forall \{p^n\}_{n\in\mathbb{N}} \in \mathcal{P}^\star, \quad R_{v,SW}^{DwP}(\{p^n\}_{n\in\mathbb{N}}) = \sup_{q\in \mathrm{Conv}(\{P(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}})} H(X|Y,q) = R_{v,SW}^{DwP} \tag{12}$$

where $\mathrm{Conv}\left(\{P(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}}\right)$ *is the convex hull of the elements of* $\{P(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}}$.

For the four models, the infimums of achievable rates in FL coding are equal to the infimums of achievable rates in VL coding, i.e.*,* $R_{v,SW}^{SP} = R_{f,SW}^{SP}$ *([27] for the VL part),* $R_{v,SW}^{SwP} = R_{f,SW}^{SwP}$*,* $R_{v,SW}^{DP} = R_{f,SW}^{DP}$ *([27] for the VL part), and* $R_{v,SW}^{DwP} = R_{f,SW}^{DwP}$.

Proof: For the SwP-Source, for VL coding, the achievable part is given by the FL setup. For the converse part, first fix a rate R. Then, for every $\boldsymbol{\theta}$ such that $H(X|Y,\boldsymbol{\theta}) > R$, the probability of error does not go to 0 as n goes to infinity, even if $\boldsymbol{\theta}$ is known, from classical SW results. Thus R must be larger than $\sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(X|Y,\boldsymbol{\theta})$. The same reasoning applies for the DwP-Source. ■

We now provide an interpretation of Theorem 1. The goal is to compare VL with FL coding. Note that, by construction, VL coding adapts the coding rate to the input sequence. This adaptation is made explicit in Definition 7, where the coding rate depends on the parameter that characterizes the joint distribution. Indeed, the non-ergodic SP-, SwP-, DwP-Sources can be decomposed into several ergodic components [14], where each component induces a rate value, and is completely determined by a parameter of the joint distribution. It is therefore sufficient to write the rate as a function of this parameter to show the adaptation of the coding rate to the input sequence. More precisely, the rates in Definition 7 depend on the parameter $\boldsymbol{\theta}$ (for the SP-, SwP-Sources) or on p^n (for the DwP-Source). By contrast, the rate for the DP-Source does not depend on any parameter. This is because the DP-Source is ergodic and the entropy can converge to only a single rate. Then, Theorem 1 states that the VL coding rates of all the considered sources are independent of the true parameter value. This can be explained by the fact that in the SW setup, the adaptation of the VL encoder is based on the input sequence $\mathbf{x}^n$. However, this input sequence does not depend on the parameter $\boldsymbol{\theta}$ or on p^n (see Definitions 1 to 4). As a consequence, VL and FL coding are equivalent in the SW setup. By contrast, in the conditional setup, the encoder observes the source X and the side information Y. Therefore, the rate in VL coding can be adapted to the true but unknown $\boldsymbol{\theta}$ or p^n, and we expect that VL and FL coding will lead to different rates.

IV. CONDITIONAL AND SW CODING

The comparison between conditional and SW coding has been studied for stationary and ergodic correlation noise between X and Y. In [6] (FL coding) and [5] (VL coding), it is shown that the knowledge of the SI at the encoder does not allow to increase the compression efficiency. We now extend this study to the case of the SP-, SwP- and DwP-Sources, which are non ergodic, and show that this property is no more valid. First, we introduce the definitions of achievable rates and of sets of achievable rates for conditional coding.

Definition 9. (Achievable rates for the conditional setup, FL coding) *For the conditional setup in FL coding, the definitions of achievable rates for the SP-, the SwP-, the DP- and the DwP-Source denoted respectively* $R_{f,c}^{SP}$,$R_{f,c}^{DP}$, $R_{f,c}^{SwP}$, $R_{f,c}^{DwP}$ *are the ones of Definition 6 except that* ϕ_n *is now defined as* $\phi_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{M}_n$.

Definition 10. (Achievable rates for the conditional setup, VL coding) *For the conditional setup with VL coding, the definition of infimums achievable rates for the DP-Source denoted* $R_{v,SW}^{DP}$ *and the definitions of sets of infimums achievable rates for the SP-, the SwP- and the DwP-Sources denoted respectively* $\{R_{v,SW}^{SP}(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$, $\{R_{v,SW}^{SwP}(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$, $\{R_{v,SW}^{DwP}(\{p^n\}_{n\in\mathbb{N}})\}$ *are the ones of Definition 6 except that* ϕ_n *is now defined as* $\phi_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{U}^\star$.

The comparison of the conditional and SW setups is given in the following theorem. For VL coding, rate losses depending on the true unknown parameters are expressed for three models.

Theorem 2. *In FL coding, the infimums of achievable rates introduced in Definition 9 are given by* $R_{f,c}^{SP} = R_{f,SW}^{SP} = R_{f,c}^{SwP} = R_{f,SW}^{SwP}$, $R_{f,c}^{DP} = R_{f,SW}^{DP}$, $R_{f,c}^{DwP} = R_{f,SW}^{DwP}$.

In VL coding, from Definition 10, for the DP-Source $R_{v,c}^{DP} = R_{v,SW}^{DP}$ *(from [15] for the conditional part) and for the other models, the rate loss between the conditional setup and the SW setup is given by*

1) *for the SP-Source,*

$$\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}, \quad \Delta^{SP}(\boldsymbol{\theta}) = P_{\boldsymbol{\Theta}}\text{-}ess.\sup Z(\boldsymbol{\Theta}) - H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta}) \tag{13}$$

where $Z(\boldsymbol{\Theta})$ *is introduced in Definition 8,*

2) *for the SwP-Source,*

$$\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}, \quad \Delta^{SwP}(\boldsymbol{\theta}) = \sup_{\boldsymbol{\alpha} \in \mathcal{P}_{\boldsymbol{\theta}}} H(X|Y, \boldsymbol{\alpha}) - H(X|Y, \boldsymbol{\theta}), \tag{14}$$

3) *for the DwP-Source,*

$$\forall \{p^n\}_{n\in\mathbb{N}} \in \mathcal{P}^\star, \quad \Delta^{DwP}(\{p^n\}_{n\in\mathbb{N}}) = \sup_{q \in Conv(\{P(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}})} H(X|Y,q) - H(\mathcal{X}|\mathcal{Y}, \{p^n\}_{n\in\mathbb{N}}) \tag{15}$$

where

$$H(\mathcal{X}|\mathcal{Y}, \{p^n\}_{n\in\mathbb{N}}) = \boldsymbol{P} - \limsup_{n\to\infty} -\frac{1}{n} \log \frac{p^n(\mathbf{X}, \mathbf{Y})}{\sum_{\mathbf{x}\in\mathcal{X}} p^n(\mathbf{x}, \mathbf{Y})} \tag{16}$$

and $\boldsymbol{P} - \limsup$ *is the* $\limsup$ *in probability.*

Proof: The SW rates come from Theorem 1. The conditional rates can be obtained as follows.

For FL coding, for the SP-,SwP- and DwP-Sources, achievable parts are given by the SW setup. The converse parts are adapted from the converse part for the SwP-Source in Theorem 1. For VL coding, see Appendix A (SP- and SwP-Sources) and Appendix B (DwP-Source).

■

Theorem 2 first states that the SW and conditional schemes are equivalent if FL encoding is used. This is an intuitive result in that FL encoding fixes the rate independently of the source realization. Therefore, even in the conditional setup where the true parameter could be estimated, this estimate cannot be used to choose the rate in the FL scheme.

Then, VL coding is studied. When the side information Y is available at the encoder, and when VL coding is used, the parameter of the joint distribution can be learned (at the encoder) and the rate can be adapted to the true distribution of the source. This should allow to achieve a lower encoding rate than the SW scheme, where the joint distribution of the source cannot be learned. Note that this behavior (lower coding rate for the conditional than the SW scheme) occurs for three of our source models: SP-, SwP- and DwP-Source. Instead, for the DP-Source, both SW and conditional schemes are equivalent. Taking a closer look at this result, we note that the DP-Source is ergodic. Therefore, a single rate is possible, which is chosen by the SW and the conditional schemes.

V. Estimated parameters

In this section, we consider SW coding and assume that estimates of the parameters $\boldsymbol{\theta}$ or $\boldsymbol{\pi}$ are available at the decoder only. This setup is of interest because it is easier to estimate the parameters at the decoder, for example with a learning sequence, than at the encoder. Indeed, estimation at the encoder requires a communication between the encoder and the side information, or with the decoder (feedback). Due to the equivalence of VL and FL coding (see Section III), we consider FL coding only in the following. In this case, the definitions of the infimums of achievable rates are as follows.

Definition 11. (Achievable rates with parameter estimation at the decoder) *Let $\hat{\boldsymbol{\Theta}}$ be a random variable taking its values in a set $\mathcal{P}_{\hat{\boldsymbol{\theta}}} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$. For the SP-Source, assume that $\hat{\boldsymbol{\Theta}}$ is distributed according to the known distributions $P_{\hat{\boldsymbol{\Theta}}|\boldsymbol{\Theta}=\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ and the Markov chain $\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} - (X,Y)$ holds. For the SwP-Source, assume that $\hat{\boldsymbol{\Theta}}$ is distributed according to the known distributions $P_{\hat{\boldsymbol{\Theta}}|\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ and one has $P(\hat{\boldsymbol{\Theta}}|\boldsymbol{\theta}, X, Y) = P(\hat{\boldsymbol{\Theta}}|\boldsymbol{\theta})$ and $P(X,Y|\hat{\boldsymbol{\Theta}}, \boldsymbol{\theta}) = P(X,Y|\boldsymbol{\theta})$ (Markov property when $\boldsymbol{\theta}$ is not random). In the SW setup for FL coding when $\hat{\boldsymbol{\Theta}}$ is available at the decoder, a rate is said to be* achievable *if it satisfies the conditions of Definition 6, where ψ_n is now defined as $\psi_n : \mathcal{Y}^n \times \mathcal{P}_{\hat{\boldsymbol{\theta}}} \to \mathcal{X}^n$. The infimums of achievable rates are denoted $R_{f,SW}^{SP,e}$ and $R_{f,SW}^{SwP,e}$.*

Let $\hat{\boldsymbol{\Pi}}$ be a random variable taking its values in a set $\mathcal{P}_{\hat{\boldsymbol{\pi}}} \subseteq \mathcal{P}_{\boldsymbol{\pi}}$. For the DP-Source, assume that $\hat{\boldsymbol{\Pi}}$ is distributed according to the known distributions $P(\hat{\boldsymbol{\Pi}}|\boldsymbol{\Pi} = \boldsymbol{\pi})$ and the Markov chain $\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi} - (X,Y)$ holds. For the DwP-Source, assume that $\hat{\boldsymbol{\Pi}}$ is distributed according to the known distributions $P(\hat{\boldsymbol{\Pi}}|\boldsymbol{\pi})$ and one has $P(\hat{\boldsymbol{\Pi}}|X,Y,\boldsymbol{\pi}) = P(\hat{\boldsymbol{\Pi}}|\boldsymbol{\pi})$ and $P(X,Y|\hat{\boldsymbol{\Pi}},\boldsymbol{\pi}) = P(X,Y|\boldsymbol{\pi})$ (Markov property). In the SW setup for FL coding when $\hat{\boldsymbol{\Pi}}$ is available at the decoder, a rate is said to be achievable *if it satisfies the conditions of Definition 6, where ψ_n is now defined as $\psi_n : \mathcal{Y}^n \times \mathcal{P}_{\hat{\boldsymbol{\pi}}} \to \mathcal{X}^n$. The infimums of achievable rates are denoted $R_{f,SW}^{DP,e}$ and $R_{f,SW}^{DwP,e}$.*

The following theorem gives the infimums of achievable rates for the four sources.

Theorem 3. *For SW FL coding, the infimum of achievable rates introduced in Definition 11 are given by*

1) *for the SP-Source, $R_{f,SW}^{SP,e} = P_{\hat{\boldsymbol{\Theta}}}$-ess. $\sup P_{\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}=\hat{\boldsymbol{\theta}}}$-ess. $\sup Z(\boldsymbol{\Theta})$, where $Z(\boldsymbol{\Theta})$ is introduced in Definition 8,*
2) *for the SwP-Source, $R_{f,SW}^{SwP,e} = \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(X|Y,\boldsymbol{\theta})$,*

3) *for the DP-Source [23],* $R_{f,SW}^{DP,e} = H(X|Y,\hat{\mathbf{\Pi}})$,

4) *for the DwP-Source [2],* $R_{f,SW}^{DwP,e} = \sup_{p \in Conv(\{p(X,Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\boldsymbol{\pi}}})} H(X|Y,\hat{\mathbf{\Pi}},p)$.

Proof: For the SwP-Source, The coding scheme can at least achieve the coding performance of a scheme without parameter estimate $\hat{\boldsymbol{\theta}}$, giving the achievable part. Moreover, if the encoder transmits a source sequence with rate $R < \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(X|Y,\mathbf{\Theta}=\boldsymbol{\theta})$, then the decoder fails to decode all sequences with parameter $\boldsymbol{\theta}$ such that $H(X|Y,\boldsymbol{\theta}) > R$, giving the converse part. The same reasoning applies for the SP-Source. ■

Theorem 3 first shows that, if an estimate of $\boldsymbol{\theta}$ or $\boldsymbol{\pi}_n$ is available at the decoder, the achievable rate of the SP- and SwP-Sources cannot be reduced. This is an intuitive result, since the coding rate is determined by the encoder, where the parameter estimate is not available. Still, a parameter estimate at the decoder is helpful, as it allows to use a standard and low-complexity decoder [10]. A less intuitive result is that a parameter estimate at the decoder may reduce the rate of the DP- and DwP-Sources. This can be explained by the fact that estimates of the parameters are available at each time instant n, and thus one random variable $\hat{\mathbf{\Pi}}_n$ can be exploited for each (X_n, Y_n). However, in fairness to static sources, one should notice that estimating $\boldsymbol{\pi}_n$ varying at each time instant is much more difficult than estimating the time-invariant $\boldsymbol{\theta}$.

VI. Outage Analysis

Section III shows that in the SW setup, the infimums of achievable rates are given by worst cases on the sets of parameters, which can lead to large rate values. Consequently, in this section, we assume that the decoder is authorized to fail for a proportion γ of the parameters. In the first theorem of this section, γ is fixed and the infimum of achievable rates is expressed with respect to γ. In the second theorem, a rate condition R is imposed, and the resulting γ is expressed with respect to R. The outage analysis is limited to the SP- and SwP-Sources. The outage analysis could also be provided for the DwP-Source, but, in this case, the outage set would depend of the possible joint distributions p^n, which are much more complex to manipulate.

Definition 12. (Achievable rate under outage constraint) *Denote* $0 \leq \gamma \leq 1$ *the outage constraint. Let* $\mathcal{M}_n = \{1 \ldots |\mathcal{M}_n|\}$ *be a set of integers. A rate* R *is said to be* achievable *if there exists two sequences of*

mappings $\{\phi_n : \mathcal{X}^n \to \mathcal{M}_n\}_{n\in\mathbb{N}}$ *and* $\{\psi_n : \mathcal{M}_n \times \mathcal{Y}^n \to \mathcal{X}^n\}_{n\in\mathbb{N}}$ *such that* $\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{M}_n| \leq R$ *and*

1) *for the SP-Source, there exists a set* $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$ *such that* $P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}) \geq 1-\gamma$ *and*

$$\lim_{n\to\infty} P(\psi_n(\mathbf{Y}^n, \phi_n(\mathbf{X}^n)) \neq \mathbf{X}^n | \boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}) = 0, \tag{17}$$

2) *for the SwP-Source, there exists a set* $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$ *such that* $\frac{\int_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} d\boldsymbol{\theta}}{\int_{\mathcal{P}_{\boldsymbol{\theta}}} d\boldsymbol{\theta}} \geq 1-\gamma$ *and*

$$\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}, \quad \lim_{n\to\infty} P(\psi_n(\mathbf{Y}^n, \phi_n(\mathbf{X}^n)) \neq \mathbf{X}^n | \boldsymbol{\theta}) = 0. \tag{18}$$

Denote respectively $R_{f,SW}^{SP}(\gamma)$ *and* $R_{f,SW}^{SwP}(\gamma)$ *the infimums of achievable rates with outage constraint* γ*.*

The following theorem gives the infimum of achievable rates with respect to γ for both sources.

Theorem 4. *The infimum of achievable rates introduced in Definition 12 with respect to an outage constraint* γ *is given by*

1) *for the SP-Source,*

$$R_{f,SW}^{SP}(\gamma) = \inf_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} P_{\boldsymbol{\Theta}|\boldsymbol{\Theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} - \textit{ess.}\sup H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\Theta}), \tag{19}$$

where $P_{\boldsymbol{\Theta}|\boldsymbol{\Theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} - \textit{ess.}\sup H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$ *is the ess.*$\sup$ *with respect to the distribution* $P_{\boldsymbol{\Theta}|\boldsymbol{\Theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}}$*,* i.e.*, knowing that* $\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}$*.*

2) *for the SwP-Source,*

$$R_{f,SW}^{SwP}(\gamma) = \inf_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} H(X|Y, \boldsymbol{\theta}). \tag{20}$$

Proof: First, for the SP-Source, denote $\mathcal{P}^{\gamma} = \{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma} \subseteq P_{\boldsymbol{\theta}} : P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}) \geq 1-\gamma\}$, the set of all possible sets of $\boldsymbol{\theta}$ such that $P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}) \geq 1-\gamma$. For every $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma} \in \mathcal{P}^{\gamma}$, from Theorem 1, the infimum of achievable rates is $P_{\boldsymbol{\Theta}|\boldsymbol{\Theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} - \text{ess.}\sup H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$, because only the $\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{\gamma}$ have to be considered. Thus the infimum of achievable rates in this setup is attained by the $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}$ for which the infimum of achievable rates is the smallest possible, *i.e.*, $\inf_{\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}} H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$. Note that the infimum can be achieved by several different $\mathcal{P}_{\boldsymbol{\theta}}^{\gamma}$. The same reasoning applies for the SwP-Source. ■

Now, a rate constraint R is imposed and we look at the level of outage γ as a function of R.

Definition 13. (Achievable outage level under rate constraint) *Let* R *be a rate constraint and let* $\mathcal{M}_n = \{1 \dots |\mathcal{M}_n|\}$ *be a set of integers. An outage level* γ *is said to be achievable if there exists a set* $\mathcal{P}_{\boldsymbol{\theta}}^{R} \subseteq \mathcal{P}_{\boldsymbol{\theta}}$

with $1 - P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^R) \leq \gamma$ *for the SP-Source,* $1 - \frac{\int_{\mathcal{P}_{\boldsymbol{\theta}}^\gamma} d\boldsymbol{\theta}}{\int_{\mathcal{P}_{\boldsymbol{\theta}}} d\boldsymbol{\theta}} \leq \gamma$ *for the SwP-Source, and such that there exists two mappings* $\phi_n : \mathcal{X}^n \to \mathcal{M}_n$ *and* $\psi_n : \mathcal{M}_n \times \mathcal{Y}^n \to \mathcal{X}^n$ *such that* $\limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{M}_n| \leq R$ *and*

1) *for the SP-Source,*

$$\lim_{n\to\infty} P(\psi_n(\mathbf{Y}^n, \phi_n(\mathbf{X}^n)) \neq \mathbf{X}^n | \boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^R) = 0 \tag{21}$$

2) *for the SwP-Source,*

$$\forall \boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}^R, \quad \lim_{n\to\infty} P(\psi_n(\mathbf{Y}^n, \phi_n(\mathbf{X}^n)) \neq \mathbf{X}^n | \boldsymbol{\theta}) = 0 \tag{22}$$

Denote respectively $\gamma_{f,SW}^{SP}(R)$ *and* $\gamma_{f,SW}^{SwP}(R)$ *the infimums of possible outage levels with respect to the rate constraint* R*.*

The following theorem gives the infimum of achievable γ in each case.

Theorem 5. *For a rate constraint* R*, denote* $\overline{\mathcal{P}}_{\boldsymbol{\theta}}^R \in \mathcal{P}_{\boldsymbol{\theta}}$ *the set of* $\boldsymbol{\theta}$ *such that* $R > H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$ *for the SP-Source,* $R > H(X|Y, \boldsymbol{\theta})$ *for the SwP-Source. Then, the infimums of outage values introduced in Definition 13 are given by* $\gamma_{f,SW}^{SP}(R) = 1 - P(\boldsymbol{\Theta} \in \overline{\mathcal{P}}_{\boldsymbol{\theta}}^R)$ *for the SP-Source, and by* $\gamma_{f,SW}^{SwP}(R) = 1 - \frac{\int_{\overline{\mathcal{P}}_{\boldsymbol{\theta}}^R} d\boldsymbol{\theta}}{\int_{\overline{\mathcal{P}}_{\boldsymbol{\theta}}} d\boldsymbol{\theta}}$ *for the SwP-Source.*

Proof: Apply Theorem 1 to $\overline{\mathcal{P}}_{\boldsymbol{\theta}}^R$. ■

When an outage constraint γ is imposed, the infimum of achievable rates is obtained as follows. We consider all the sets $\mathcal{P}_{\boldsymbol{\theta}}^\gamma$ of measure γ. For each set, the infimum of achievable rates is given by the conditional entropy for the worst possible $\boldsymbol{\theta}$ on the set. Then, the set for which this worst case induces the smallest rate is retained. When a rate constraint R is considered, the condition to obtain γ is much simpler to express because we know that the probability of error cannot go to zero for every $\boldsymbol{\theta}$ such that $H(X|Y, \boldsymbol{\theta}) > R$.

VII. Application

In this section, we apply the information theoretical results presented in the previous sections to optimize the source coding strategy in a sensor network. Consider the three-node fully-connected network depicted in Figure 1. Two correlated sources X and Y have to transmit sequences of n symbols to a receiver S.

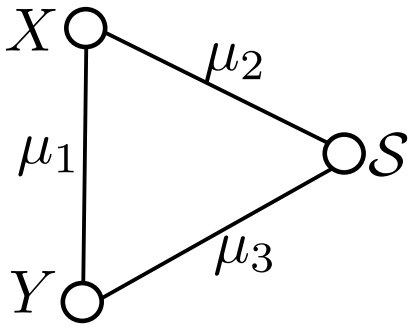

Fig. 1. Three-node network

Following the model proposed in [7], each communication link is associated to weights μ_1, μ_2, and μ_3. The cost of transmitting at rate R on a link i is $\mu_i R$. The channel coding aspects are taken into account in μ_i. But here, we do not detail how the μ_is are computed since we focus on source coding aspects. More precisely, we optimize the source coding strategy, under a fixed transmission scenario. Note that a similar analysis can be carried out for multicast transmission. In this case, the cost of transmitting at rate R on both links i and j could for example be given by $\max(\mu_i, \mu_j)R$ instead of $(\mu_i + \mu_j)R$. Several coding strategies are analyzed and compared.

A. Conditional and SW coding

Assume that the correlation between X and Y is modeled as a SP-Source characterized by $\boldsymbol{\Theta}$ and $\mathcal{P}_{\boldsymbol{\theta}}$. Four coding strategies are compared. The analysis here is asymptotic, meaning that the sequence length n goes to infinity and the coding rates are given by the infimums of achievable rates.

In the first conditional strategy, Y transmits its sequence to S and X. X transmits its sequence to S with VL conditional coding. In the second conditional strategy, the roles of X and Y are inverted. Denote $m_{\mathrm{c}}^{(X)}(\boldsymbol{\theta})$ and $m_{\mathrm{c}}^{(Y)}(\boldsymbol{\theta})$ the costs associated to the two conditional setups for a given $\boldsymbol{\theta}$. In the first SW strategy, Y transmits its sequence to S and X uses SW coding to transmit its sequence to S with Y as side information. In the second SW strategy, the roles of X and Y are again inverted. For the SW strategies, only corner points of the SW region are considered, because [7] shows that only these cases can be optimal. Moreover, [7] also shows that if $\mu_2 \leq \mu_3$ the case where X is side information gives lower cost. This is the only case we consider here, without loss of generality. Denote $m_{\mathrm{s}}^{(X)}$ the cost of this setup.

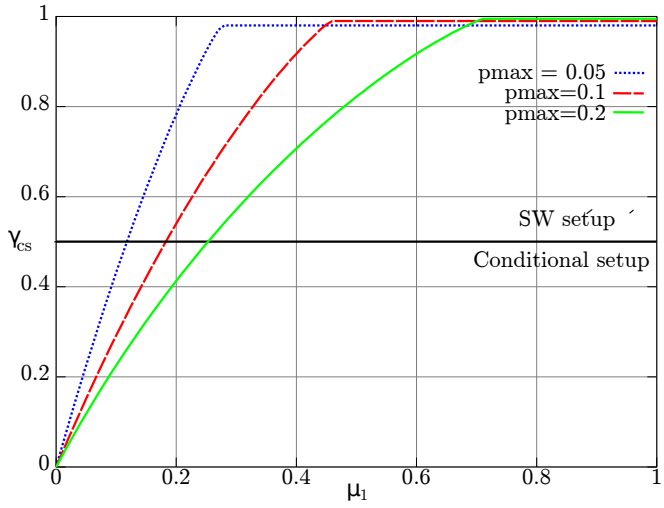

Fig. 2. Example for binary sources. X is distributed uniformly and the correlation channel is binary symmetric with unknown probability transition p unknown. We set $p \in [0, \text{pmax}]$ and $\mu_2 = \mu_3 = 1$. We plot γ_{cs} with respect to μ_1 in order to compare the conditional setup and the SW setup when X is the side information. If $\gamma_{cs} > 0.5$, the conditional coding strategy is kept.

The three considered strategies have costs

$$m_{\text{c}}^{(X)}(\boldsymbol{\theta}) = (\mu_1 + \mu_2)H(X) + \mu_3 H(Y|X, \boldsymbol{\Theta} = \boldsymbol{\theta}) \tag{23}$$

$$m_{\text{c}}^{(Y)}(\boldsymbol{\theta}) = (\mu_1 + \mu_3)H(Y|\boldsymbol{\Theta} = \boldsymbol{\theta}) + \mu_2 H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta}) \tag{24}$$

$$m_{\text{s}}^{(X)} = \mu_2 H(X) + \mu_3 \sup_{\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}} H(Y|X, \boldsymbol{\Theta} = \boldsymbol{\theta}) \ . \tag{25}$$

The strategies are then compared two by two, beginning with the comparison of the two conditional setups. First, compute the cost difference $\Delta_c(\boldsymbol{\theta}) = m_{\text{c}}^{(X)}(\boldsymbol{\theta}) - m_{\text{c}}^{(Y)}(\boldsymbol{\theta})$. Denote $\mathcal{P}_{\boldsymbol{\theta}}^c$ the set of $\boldsymbol{\theta}$ such that $\Delta_c(\boldsymbol{\theta}) \geq 0$ and denote $\gamma_c = 1 - P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^c)$. Then, the strategy where Y is the side information is kept if and only if $\gamma_c > 0.5$. Note that this is equivalent to comparing $E[m_{\text{c}}^{(X)}(\boldsymbol{\theta})]$ with $E[m_{\text{c}}^{(Y)}(\boldsymbol{\theta})]$. The retained conditional strategy is then compared to the SW strategy. Without loss of generality, we assume that $\gamma_c > 0.5$. Then, we compute $\Delta_{cs}(\boldsymbol{\theta}) = m_{\text{c}}^{(Y)}(\boldsymbol{\theta}) - m_{\text{s}}^{(X)}$. Denote $\mathcal{P}_{\boldsymbol{\theta}}^{cs}$ the set of $\boldsymbol{\theta}$ such that $\Delta_{cs}(\boldsymbol{\theta}) \geq 0$ and denote $\gamma_{cs} = 1 - P(\boldsymbol{\Theta} \in \mathcal{P}_{\boldsymbol{\theta}}^{cs})$. One or the other strategy is chosen consequently. For example, when $\mu_1 \ll \mu_3$ or $\mu_1 \ll \mu_2$, the strategy with minimum cost is always the conditional one. An other example is represented in Figure 2.

The same analysis can be carried out for the SP- and DwP-Sources, replacing $\sup_{\boldsymbol{\theta} \in \mathcal{P}_{\boldsymbol{\theta}}} H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$ in (23) and $H(X|Y, \boldsymbol{\Theta} = \boldsymbol{\theta})$ in (24) by their respective expressions from Theorem 2 and by adapting the expressions of γ_c and γ_{cs}. Same conclusions are reached. By contrast, the SW coding is always the best

strategy for the DP-Source.

B. Estimated parameters

Although estimated parameters were shown not to decrease the coding rate in Section V, they can help the decoding operation [10]. In this section, we express the cost of a learning sequence sent to the decoder to learn the parameters. Consider a SwP-Source and assume that μ_1 is large enough compared to μ_2 and μ_3 so that the SW coding strategy is chosen. Without loss of generality, we consider the case where $\mu_2 \leq \mu_3$ so that X is the side information. Now, finite-length coding is considered and two strategies are compared. In both cases, X transmits its sequence of length n to S. In the first strategy (learning sequence), Y transmits first a learning sequence of length u_n to S. The sequence $\{u_n\}_{n\in\mathbb{N}}$ is such that $\lim_{n\to\infty} u_n = +\infty$ and $\lim_{n\to\infty} \frac{u_n}{n} = 0$. Furthermore, we assume that, to perform well, the decoder requires a mean-squared error $E\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2\right]$ smaller than a fixed value ε_n. It will be taken into account for the choice of u_n. In the second strategy, no learning sequence is transmitted, and the decoder has to deal with the uncertainty on the parameters. Consequently, the cost of this strategy is supplemented with a positive term $\alpha(\boldsymbol{\theta})$. For example, [11] shows that the strategy without learning sequence induced an increased decoding time. Thus, $\alpha(\boldsymbol{\theta})$ may be proportional to this time increase, with respect to $\boldsymbol{\theta}$. Denote $m_{\text{l}}(\boldsymbol{\theta})$ and $m_{\text{wl}}(\boldsymbol{\theta})$ the respective costs for each strategy for a given $\boldsymbol{\theta}$. One has

$$m_{\text{l}}(\boldsymbol{\theta}) = \mu_2 H(X) + \mu_3 \left(\frac{u_n}{n} H(Y|\boldsymbol{\theta}) + \frac{n-u_n}{n} \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) \right) \tag{26}$$

$$m_{\text{wl}}(\boldsymbol{\theta}) = \mu_2 H(X) + \mu_3 \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) + \alpha(\boldsymbol{\theta}) \ . \tag{27}$$

When $n \to \infty$, from the conditions on u_n, $m_{\text{l}}(\boldsymbol{\theta})$ tends to $m_{\text{wl}}(\boldsymbol{\theta}) - \alpha(\boldsymbol{\theta})$ and a consistent estimator enables to obtain an estimate $\hat{\boldsymbol{\theta}}$ arbitrarily close to the true $\boldsymbol{\theta}$. Consequently, in this case, the strategy with learning sequence always gives lower cost.

But at finite length, a rate loss can be calculated as

$$\Delta(u_n, \boldsymbol{\theta}) = \frac{m_{\text{l}}(\boldsymbol{\theta}) - m_{\text{wl}}(\boldsymbol{\theta})}{\mu_3} = \frac{u_n}{n} \left(H(Y|\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) \right) - \frac{\alpha(\boldsymbol{\theta})}{\mu_3} \ . \tag{28}$$

For Maximum Likelihood estimation, if K is the number of parameters, then the mean squared error is bounded by K/u_n. By setting $K/u_n = \varepsilon_n$, we get

$$\Delta(u_n, \boldsymbol{\theta}) = \frac{\varepsilon_n}{Kn} \left(\sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(Y|X,\boldsymbol{\theta}) - H(Y|\boldsymbol{\theta}) \right) - \frac{\alpha(\boldsymbol{\theta})}{\mu_3} \ . \tag{29}$$

To finish, one has to evaluate ε_n and $\alpha(\boldsymbol{\theta})$ and can calculate the resulting $\Delta(u_n, \boldsymbol{\theta})$. The two strategies can then be compared as the conditional and SW setups in the previous section.

VIII. Conclusion

This paper addressed the problem of lossless source coding with SI at the decoder when the joint distribution of the source and side information is only partially known. Four parametric models have been considered. Unlike classical results in distributed source coding, it was shown that the source can be further compressed if the side information is also available at the encoder. To reduce the discrepancy between the cooperative and non-cooperative schemes, coding strategies based on parameter estimation and outage analysis have been proposed. Finally, methods to compare and choose the source coding strategy that minimizes the overall transmission cost in a sensor network has been introduced.

Appendix

A. SP and SwP-Sources, conditional setup, variable-length coding

The proof is the same for both the SP-Source and the SwP-Source. The basic idea of the proof was proposed in [9] in the case without SI. Here the source X has to be encoded conditionally to Y. For the achievable part, define a coding process as follows. For a source sequence $\mathbf{x}$ of length n to be encoded with the help of the SI sequence $\mathbf{y}$, the encoder first determines the conditional type $\hat{P}_{\mathbf{y}}$ of $\mathbf{x}$ with respect to $\mathbf{y}$. Denote $\mathcal{T}^n_{\hat{P}}(\mathbf{y})$ the set of sequences $\mathbf{x}$ of conditional type $\hat{P}_{\mathbf{y}}$ *i.e.* [8, Definition 2.4],

$$\mathcal{T}^n_{\hat{P}}(\mathbf{y}) = \left\{ \mathbf{x} \in \mathcal{X}^n : (a,b) \in \mathcal{X} \times \mathcal{Y} : N(a,b|\mathbf{x},\mathbf{y}) = N(b|\mathbf{y}) \hat{P}_{\mathbf{y}}(a|b) \right\} \tag{30}$$

where $N(b|\mathbf{y})$ is the number of occurrences of the symbol b in $\mathbf{y}$. Each $\mathbf{x} \in \mathcal{T}^n_{\hat{P}}(\mathbf{y})$ is indexed and the encoder transmits the index of the actual source sequence. Successful decoding is guaranteed by the fact that each sequence in $\mathcal{T}^n_{\hat{P}}(\mathbf{y})$ is indexed.

We now evaluate the coding rate $\frac{1}{n} E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X},\mathbf{Y})|]$. As there are $n^{|\mathcal{X}||\mathcal{Y}|}$ possible types, the per-symbol rate needed to transmit $\hat{P}_{\mathbf{y}}$ is $|\mathcal{X}||\mathcal{Y}|(\log n)/n$. From the coding process definition,

$$\begin{aligned} \frac{1}{n} E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X},\mathbf{Y})|] &= \frac{1}{n} \sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) |\phi_n(\mathbf{x},\mathbf{y})| \\ &= |\mathcal{X}||\mathcal{Y}| \frac{\log n}{n} + \frac{1}{n} \sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) \log |\mathcal{T}^n_{\hat{P}}(\mathbf{y})| \end{aligned} \tag{31}$$

Let $\{\delta_n\}_{n\in\mathbb{N}}$ be a sequence of $\delta_n \in \mathbb{R}$ such that $\lim_{n\to\infty} \delta_n = 0$, $\lim_{n\to\infty} \sqrt{n}\delta_n = +\infty$. Denote $\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})$ the set of sequences $\mathbf{x}$ conditionally typical with $\mathbf{y}$ for the true (unknown) parameter $\boldsymbol{\theta}$, *i.e.* [8, Definition 2.9],

$$\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y}) = \left\{ \mathbf{x} \in \mathcal{X}^n : \forall (a,b) \in \mathcal{X}\times\mathcal{Y}, \left| \frac{1}{n} N(a,b|\mathbf{x},\mathbf{y}) - \frac{1}{n} N(b|\mathbf{y}) P(a|b,\boldsymbol{\theta}) \right| \leq \delta_n \right\} \tag{32}$$

One has

$$\begin{aligned}
\frac{1}{n} E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X},\mathbf{Y})|] &= |\mathcal{X}||\mathcal{Y}| \frac{\log n}{n} + \frac{1}{n} \sum_{(\mathbf{x},\mathbf{y}):\mathbf{x}\in\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})} P(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) \log |\mathcal{T}^n_{\hat{P}}(\mathbf{y})| \\
&\quad + \frac{1}{n} \sum_{(\mathbf{x},\mathbf{y}):\mathbf{x}\notin\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})} P(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) \log |\mathcal{T}^n_{\hat{P}}(\mathbf{y})| \\
&\leq |\mathcal{X}||\mathcal{Y}| \frac{\log n}{n} + \frac{1}{n} \sum_{(\mathbf{x},\mathbf{y}):\mathbf{x}\in\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})} P(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) \log |\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})| \\
&\quad + \sum_{(\mathbf{x},\mathbf{y}):\mathbf{x}\notin\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})} P(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) \log |\mathcal{X}| \ .
\end{aligned} \tag{33}$$

From [8, Lemma 2.13], there exists a sequence $\{\epsilon_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \epsilon_n = 0$ and $\forall \mathbf{y} \in \mathcal{Y}^n$, $|\mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})| \leq \exp(nH(X|Y,\boldsymbol{\theta}) + \epsilon_n)$. Thus

$$\begin{aligned}
\frac{1}{n} E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X},\mathbf{Y})|] \leq |\mathcal{X}||\mathcal{Y}| \frac{\log n}{n} &+ (H(X|Y,\boldsymbol{\theta}) + \epsilon_n)\, P\left((\mathbf{x},\mathbf{y}) : \mathbf{x} \in \mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})\right) \\
&+ \log|\mathcal{X}| P\left((\mathbf{x},\mathbf{y}) : \mathbf{x} \notin \mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})\right) \ .
\end{aligned} \tag{34}$$

From [8, Lemma 2.12], $\lim_{n\to\infty} P\left((\mathbf{x},\mathbf{y}) : \mathbf{x} \in \mathcal{T}^n_{[\boldsymbol{\theta}]\delta_n}(\mathbf{y})\right) = 1$, giving

$$\lim_{n\to\infty} \sup E_{\boldsymbol{\theta}}[|\phi_n(\mathbf{X},\mathbf{Y})|] \leq H(X|Y,\boldsymbol{\theta}) \tag{35}$$

and the rate $H(X|Y,\boldsymbol{\theta})$ is achievable. Note that the achievable part could also be proven from the Lempel Ziv complexity results [25].

Conversely, if for a true parameter $\boldsymbol{\theta}$, the coding rate is less than $H(X|Y,\boldsymbol{\theta})$, then from [15] the decoder fails with non-zero probability even if $\boldsymbol{\theta}$ is known.

B. DwP-Sources, conditional setup, variable-length coding

The method described in the previous proof cannot be applied here. Indeed, the probability distribution of a sequence $(\mathbf{x},\mathbf{y})$ depends on the parameter sequence $\{\boldsymbol{\pi}_k\}_{k=1}^n$ and thus the conditional type of $\mathbf{x}$ knowing $\mathbf{y}$ does not necessarily converge. Consequently, [8, Lemma 2.12] do not apply. For the achievable part,

we consider conditional LZ78 coding as defined in [25]. The coding process is the same and we keep the notations, but the main difference is that [25] considers stationary and ergodic sources. Consequently, the rate evaluation needs to be adapted.

A string $(\mathbf{x}, \mathbf{y})$ of length n observed at the encoder is parsed into c distinct phrases as

$$(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})_1^{n_1} (\mathbf{x}, \mathbf{y})_{n_1+1}^{n_2} \dots (\mathbf{x}, \mathbf{y})_{n_{c-1}+1}^{n_c} \tag{36}$$

where the n_i are the accumulated length of the i first phrases, and $n_c = n$. We now consider the individual parsing of $\mathbf{y}$ resulting from the joint parsing of $(\mathbf{x}, \mathbf{y})$, *i.e.*, $\mathbf{y}_1^{n_1}, \mathbf{y}_{n_1+1}^{n_2} \dots \mathbf{y}_{n_{c-1}+1}^{n_c}$. Denote $c(\mathbf{y})$ the number of *distinct* phrases in the individual parsing of $\mathbf{y}$ and denote by $\mathbf{y}(l)$ the l-th distinct phrase, with $l = 1 \dots c(\mathbf{y})$. Note that each distinct $\mathbf{y}(l)$ can appear several times in the parsing of $\mathbf{y}$. Then, denote $c_l(\mathbf{x}|\mathbf{y})$ the number of distinct phrases from $\mathbf{x}$ that appears jointly with each $\mathbf{y}(l)$. At length n the rate needed to transmit $\mathbf{x}$ with this parsing is [25]

$$R_n(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{l=1}^{c(\mathbf{y})} c_l(\mathbf{x}|\mathbf{y}) \left(\log(c_l(\mathbf{x}|\mathbf{y})) + 1 \right) \ . \tag{37}$$

Indeed, for each of the $c(\mathbf{y})$ distinct phrases for $\mathbf{y}$, $c_l(\mathbf{x}|\mathbf{y})$ distinct phrases from $\mathbf{x}$ have to be encoded, each at rate $\log(c_l(\mathbf{x}|\mathbf{y})) + 1$.

Before giving the main result, we state the following lemma, that is basically the Conditional Ziv's inequality applied to the DwP-Source.

Lemma 1. *(Conditional Ziv's inequality for the DwP-Source [25]) For any sequences $(\mathbf{x}, \mathbf{y})$ of length n obtained from a DwP-Source with true distribution $p^n \in \mathcal{P}_n$ (see Definition 5),*

$$\log P(\mathbf{x}|\mathbf{y}, p^n) \leq - \sum_{l=1}^{c(\mathbf{y})} c_l(\mathbf{x}|\mathbf{y}) \log(c_l(\mathbf{x}|\mathbf{y})) \ . \tag{38}$$

Proof: Apply [25, Lemma 1] to a source (X, Y) producing independent symbols. ■

Note that the Conditional Ziv's inequality comes from the Jensen's inequality and does not depend on the stationary or ergodic nature of the source. However, the formulation of the lemma as well as the proof are simpler than in [25] because in our case, the source symbols are assumed independent.

Lemma 2. *For any sequences $\{\mathbf{X}, \mathbf{Y}\}_{n=1}^{+\infty}$ obtained from a DwP-Source with true sequence of distributions*

$\{p^n\}_{n=1}^{+\infty} \in \mathcal{P}^\star$ *(see Definition 5),*

$$\lim_{n\to\infty} P\left(\frac{1}{n}\sum_{l=1}^{c(\mathbf{Y})} c_l(\mathbf{X}|\mathbf{Y})\log(c_l(\mathbf{X}|\mathbf{Y})) > H\left(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}}\right)\right) = 0\ . \tag{39}$$

Proof: From Lemma 1,

$$-\frac{1}{n}\log P(\mathbf{x}|\mathbf{y},p^n) \geq \frac{1}{n}\sum_{l=1}^{c(\mathbf{y})} c_l(\mathbf{x}|\mathbf{y})\log(c_l(\mathbf{x}|\mathbf{y})) \tag{40}$$

Then, from the definition of $H\left(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}}\right)$ in (16),

$$\lim_{n\to\infty} P\left(-\frac{1}{n}\log P(\mathbf{x}|\mathbf{y},p^n) > H\left(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}}\right)\right) = 0, \tag{41}$$

giving (39). ■

We are now ready to evaluate $E_{p^n}[|\phi_n(\mathbf{X},\mathbf{Y})|]$. One has

$$\begin{aligned}
\frac{1}{n}E_{p^n}[|\phi_n(\mathbf{X},\mathbf{Y})|] &= \frac{1}{n}\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y}|p^n)|\phi_n(\mathbf{x},\mathbf{y})| \\
&= \sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y}|p^n)R_n(\mathbf{x},\mathbf{y}) \\
&\leq \sum_{\substack{(\mathbf{x},\mathbf{y}):\\ R_n(\mathbf{x},\mathbf{y})\leq H(\mathcal{X}|\mathcal{Y}\{p^n\}_{n\in\mathbb{N}})}} P(\mathbf{x},\mathbf{y}|p^n)R_n(\mathbf{x},\mathbf{y}) \\
&\qquad + \sum_{\substack{(\mathbf{x},\mathbf{y}):\\ R_n(\mathbf{x},\mathbf{y})> H(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}})}} P(\mathbf{x},\mathbf{y}|p^n)\log|\mathcal{X}| \\
&\leq H(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}})P\big(R_n(\mathbf{X},\mathbf{Y}) \leq H(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}})\big) \\
&\qquad + \log|\mathcal{X}|P\big(R_n(\mathbf{X},\mathbf{Y}) > H(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}})\big)
\end{aligned} \tag{42}$$

To finish, from (37) and Lemma 2, $\lim_{n\to\infty} P\big(R_n(\mathbf{X},\mathbf{Y}) > H(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}})\big) = 0$ and then $\limsup_{n\to\infty} E_{p^n}[|\phi_n(\mathbf{X}$ $H\left(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}}\right)$ and the rate $H\left(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}}\right)$ is achievable.

Conversely, if for a true sequence of distributions $\{p^n\}$, the coding rate is less than $(\mathcal{X}|\mathcal{Y},\{p^n\}_{n\in\mathbb{N}})$, then from [16] the decoder fails with non-zero probability even if $\{p^n\}$ is known.

REFERENCES

[1] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv coding of motion video. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, volume 1, pages 240–244, 2002.

[2] R. Ahlswede. Coloring hypergraphs: A new approach to multi-user source coding-1. *Journal of Combinatorics*, 4(1):76–115, 1979.

[3] T. Berger. The source coding game. *IEEE Transactions on Information Theory*, 17(1):71–76, 1971.

[4] J. Chen, D. He, and A. Jagmohan. On the duality between Slepian–Wolf coding and channel coding under mismatched decoding. *IEEE Transactions on Information Theory*, 55(9):4006–4018, 2009.

[5] J. Chen, D.K. He, and E. Yang. Slepian–Wolf coding for general sources and its duality with channel coding. In *10th Canadian Workshop on Information Theory, CWIT '07*, pages 57–60, 2007.

[6] T.M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources (Corresp.). *IEEE Transactions on Information Theory*, 21(2):226–228, 1975.

[7] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. Networked Slepian–Wolf: theory, algorithms, and scaling laws. *IEEE Transactions on Information Theory*, 51(12):4057–4073, 2005.

[8] I. Csiszar and J. Korner. *Information theory: coding theorems for discrete memoryless systems*. Hardcover, 1997.

[9] L. Davisson. Universal noiseless coding. *IEEE Transactions on Information Theory*, 19(6):783–795, 1973.

[10] E. Dupraz, A. Roumy, and M. Kieffer. Source coding with side information at the decoder: models with uncertainty, performance bounds, and practical coding schemes. In *Proceedings of the International Symposium on Information Theory and its Applications*, pages 1–5, 2012.

[11] E. Dupraz, A. Roumy, and M. Kieffer. Practical coding scheme for universal source coding with side information at the decoder. In *Accepted to the Data Compression Conference, Snowbird*, March 2013.

[12] R.G. Gallager. *Information theory and reliable communication*. Wiley, 1968.

[13] R.G. Gallager. Source coding with side information and universal coding. 1979.

[14] R.M. Gray and L.D. Davisson. The ergodic decomposition of stationary discrete random processes. *IEEE Transactions on Information Theory*, 20(5):625–636, 1974.

[15] R.M. Gray and Stanford Univ Calif Stanford Electronics Labs. Conditional rate-distortion theory, 1972.

[16] T.S. Han. *Information-spectrum methods in information theory*. Springer, 2003.

[17] S. Jalali, S. Verdú, and T. Weissman. A Universal Scheme for Wyner–Ziv Coding of Discrete Sources. *IEEE Transactions on Information Theory*, 56(4):1737–1750, 2010.

[18] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6:440–442, 2002.

[19] T. Matsuta, T. Uyematsu, and R. Matsumoto. Universal Slepian–Wolf source codes using Low-Density Parity-Check matrices. In *IEEE International Symposium on Information Theory Proceedings (ISIT), 2010*, pages 186–190, 2010.

[20] P. Piantanida, G. Matz, and P. Duhamel. Outage behavior of discrete memoryless channels under channel estimation errors. *IEEE Transactions on Information Theory*, 55(9):4221–4239, 2009.

[21] R. Puri and K. Ramchandran. PRISM: A new robust video coding architecture based on distributed compression principles. In *Proc. of the Annual Allerton Conference on Communication, Control, and Computing*, volume 40, pages 586–595. Citeseer, 2002.

[22] A. Sgarro. Source coding with side information at several decoders. *IEEE Transactions on Information Theory*, 23(2):179–182, 1977.

[23] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, 1973.

[24] T. Uyematsu. An algebraic construction of codes for Slepian–Wolf source networks. *IEEE Transactions on Information Theory*, 47(7):3082–3088, 2001.

[25] T. Uyematsu and S. Kuzuoka. Conditional Lempel-Ziv complexity and its application to source coding theorem with side information. In *IEEE International Symposium on Information Theory, 2003. Proceedings.*, pages 142–145, 2003.

[26] Z. Xiong, A.D. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5):80–94, 2004.

[27] E.H. Yang and D. He. Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder. *IEEE Transactions on Information Theory*, 56(4):1808–1824, 2010.

Annexe B

Schémas de codage de SW

Elsa Dupraz, Aline Roumy, Michel Kieffer *Source coding with side information at the decoder and uncertain knowledge of the correlation* Submitted at IEEE Transactions on Communications, February 2013, major revision

Certaines parties de cet article ont été publiées dans

Elsa Dupraz, Aline Roumy, Michel Kieffer *Practical coding scheme for universal source coding with side information at the decoder.* Oral presentation at DCC 2013, Snowird, Utah, United States

Source Coding with Side Information at the Decoder and Uncertain Knowledge of the Correlation

Elsa DUPRAZ[1], Aline ROUMY[2] and Michel KIEFFER[1,3,4]

[1] L2S - CNRS - SUPELEC - Univ Paris-Sud, 91192 Gif-sur-Yvette, France

[2] INRIA, Campus de Beaulieu, 35042 Rennes, France

[3] Partly on leave at LTCI – CNRS Télécom ParisTech, 75013 Paris, France

[4] Institut Universitaire de France

Abstract

This paper considers the problem of lossless source coding with side information at the decoder, when the correlation model between the source and the side information is uncertain. Four parametrized models representing the correlation between the source and the side information are introduced. The uncertainty on the correlation appears through the lack of knowledge on the value of the parameters.

For each model, we propose a practical coding scheme based on non-binary Low Density Parity Check Codes and able to deal with the parameter uncertainty. At the encoder, the choice of the coding rate results from an information theoretical analysis. Then we propose decoding algorithms that jointly estimate the source vector and the parameters. As the proposed decoder is based on the Expectation-Maximization algorithm, which is very sensitive to initialization, we also propose a method to produce first a coarse estimate of the parameters.

Part of this paper was presented at the Data Compression Conference (DCC) 2013.

This work was partly supported by the ANR-09-VERS-019-02 grant (ARSSO project).

I. INTRODUCTION

The problem of lossless source coding with side information at the decoder has been well investigated when the correlation model between the source X and the side information (SI) Y is perfectly known. Slepian and Wolf showed that this case induces no loss in performance compared to the conditional setup, *i.e.*, the setup where the side information is also known at the encoder [43]. Following this principle, several works, see, *e.g.*, [38], [45], [54], propose practical coding schemes for the Slepian-Wolf (SW) problem. Most of them are based on channel codes [46], and particularly Low Density Parity Check (LDPC) codes [32], [34]. This approach allows to leverage on many results on LDPC codes for the code construction and optimization [30], [40] even if there is a need to adapt the developed algorithms to the case of SW coding [7].

Nonetheless, most of these works assume perfect knowledge of the joint distribution $P(X, Y)$. In [28], it is shown that the performance of the SW coding scheme remains the same if $P(X)$ is unknown. Here we consider the case where the characteristics of the *correlation channel* $P(Y|X)$ are uncertain because they are in general more difficult to obtain in practical situations. In this way, [42] considers the case where $P(Y|X)$ is given to the decoder but not perfectly known at the encoder. Here we assume that $P(Y|X)$ is uncertain at both the encoder and the decoder. A usual solution to address this problem is to use a feedback channel [1], [17], [51], or to allow interactions between the encoder and the decoder [55]. The advantage of the feedback channel is that the rate is adapted to the true characteristics of the source. However, a feedback channel can be difficult to implement in many practical situations such as sensor networks. Moreover, the feedback channel is in general used by the decoder to ask for additional packets to the encoder or to stop the transmission. Each time a new packet is received, the decoder processes again all the received packets to try to reconstruct the source. This can result in huge decoding delays.

When no feedback is allowed, several practical solutions based on LDPC codes and proposed

for channel coding may be adapted to the SW problem. When hard decoding is performed, as proposed in [31], [39] for channel coding, only symbol values are used, at the price of an important loss in performance. An alternative solution is the min-sum decoding algorithm proposed in [6], [41] for channel coding, respectively for binary and non-binary sources. The min-sum algorithm uses soft information for decoding, but does not require the knowledge of the correlation channel. However, in SW coding, if the source X is not distributed uniformly, the min-sum equations cannot be derived.

In many applications, it is possible to restrict the correlation channel model to a given class (*e.g.*, binary symmetric, Gaussian, etc.) due to the nature of the problem. Consequently, in this paper, we introduce four correlation channel models. Each model assumes that the correlation channel belongs to a given class and is parametrized by some unknown parameter vector. For two of the models, the correlation channel between source symbols (X_n, Y_n) is parametrized by an unknown parameter vector $\boldsymbol{\pi}_n$, varying from symbol to symbol. One of these two models assumes the knowledge of a prior distribution $P_{\boldsymbol{\Pi}}(\boldsymbol{\pi}_n)$ for $\boldsymbol{\pi}_n$. The case where no prior on $\boldsymbol{\pi}_n$ is known corresponds to *arbitrarily varying sources* [2], [4]. For the two other models, the correlation channel is parametrized by an unknown parameter vector $\boldsymbol{\theta}$, fixed for the sequence $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ but allowed to vary from sequence to sequence. This corresponds to universal source coding [21]. The distinction between the two models is also in the knowledge of a prior for $\boldsymbol{\theta}$. The distinction between varying parameters $\boldsymbol{\pi}_n$ and a fixed parameter $\boldsymbol{\theta}$ has been proposed earlier in [37] in the case of channel coding.

The coding scheme we propose is based on non-binary LDPC codes and assumes additive correlation channel. Hard and min-sum LDPC decoding are not able to exploit the knowledge of the structure of the class. Therefore, the sum-product LDPC decoding algorithm is considered. From an analysis of the performance bounds, we explain for each model how to choose the coding rate and the LDPC coding matrix. Then, we show that the classical sum-product LDPC decoding algorithm can be used for only one model. For the three other models, we propose a decoding

algorithm that jointly estimates the source vector and the parameters. As the method is based on the EM (Expectation Maximization) algorithm [25], which is very sensitive to initialization, we also propose a method to obtain first coarse estimates of the values of the parameters.

The paper is organized as follows. Section II presents the related works. In Section III, the four signal models we consider are described formally. Section IV explains how to choose the coding rates and to design the LDPC coding matrices. Section V proposes a decoding method adapted to each model. Finally, Section VI presents simulation results.

II. Related Works

In Slepian-Wolf coding, the issue of estimating jointly the correlation parameters and the LDPC encoded source vector was addressed in [8], [10], [18], [49], [48], [50], [56]. All the papers consider the case of a Binary Symmetric Channel (BSC) which is an additive model. In [48], [49], the probability transition of the BSC is known, but the source distribution is unknown. On the contrary, in [8], [10], [50], [56], the source distribution is known but the probability transition is unknown. Some of these works, *e.g.*, [50], [56], assume that the probability transition is fixed for the whole source vector. In this case, the joint estimation is performed with an EM algorithm. The other works [8], [10], [18], allow the parameter to vary by blocs of fixed length inside the source vector. The parameter estimation can then be realized with Particle Filtering [8], Expected Propagation [10], or Sliding-Window Belief Propagation [18]. However, as pointed out in [10], Particle Filtering is an MCMC based method and induces an important decoding complexity. On the other hand, the two other methods are less complex but require the knowledge of a prior distribution for the parameters. Furthermore, this prior distribution is required to be conjugate exponential for computational reasons.

The parameter estimation was also discussed in the area of Distributed Video Coding (DVC) [5], [27], [35], [44], [52], [53]. Indeed, in DVC, the lossless part of the transmission is realized with a SW chain based on binary LDPC codes. The correlation channel is assumed to be additive, with

Gaussian noise [44], or Laplacian noise [5], [35], [52], [53], for more accuracy. In both cases, the unknown parameter is the noise variance. Due to the particular Gaussian or Laplacian additive model, it is possible to realize the parameter estimation from the side information only [5], [27], [35]. Otherwise, for more accuracy, the estimation can be done with an EM algorithm [52], or with particle filtering [44], [53], with the same remarks as for SW coding. Moreover, in DVC the non-binary source symbols (the pixels or the DCT coefficients) are transformed into bits and transmitted independently by bit planes with binary LDPC codes. Consequently, the decoding algorithm has to take the dependencies between bit planes into account, as proposed in [44], [52], [53], at the price of a complexity increase.

In this paper, we focus on the SW coding aspects for discrete source and side information symbols. The correlation model is assumed to be additive. However, unlike the previously mentioned works, we consider non-binary source symbols and non-binary LDPC codes. From arguments given in introduction, the source distribution $P(X)$ does not depend on the unknown parameters. The choice of the parameters can be either deterministic or random, with respect to any kind of prior distribution. For the two models with fixed parameters, the EM algorithm is considered, as previously suggested. We derive the EM equations for our case, and propose a method to initialize the EM algorithm properly. On the other hand, we consider the case where the parameters may vary from symbol to symbol. We explain how to perform the decoding despite this possible important variability.

III. Signal Model

The source X to be compressed and the SI Y available at the decoder produce sequences of symbols $\{X_n\}_{n=1}^{+\infty}$ and $\{Y_n\}_{n=1}^{+\infty}$, respectively. $\mathcal{X}$ and $\mathcal{Y}$ denote the source and SI discrete alphabets. Bold upper case letters, *e.g.*, $\mathbf{X}_1^N = \{X_n\}_{n=1}^N$, denote random vectors, whereas bold lower case letters, $\mathbf{x}_1^N = \{x_n\}_{n=1}^N$, represent their realizations. When it is clear from the context that the distribution of a random variable X_n does not depend on n, the index n is omitted.

The goal of this section is to model the uncertainty on the correlation channel $P(Y|X)$. Each of the four proposed models consists of a family of parametric distributions[1]. In every case, the source distribution $P(X)$ is assumed perfectly known and does not depend on the uncertain parameters. The first two models allow parameter variations from symbol to symbol.

Definition 1. (DP-Source). *A Dynamic with Prior source (X, Y), or DP-Source, produces a sequence of **independent** symbols $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ drawn $\forall n$ from $P(X_n, Y_n)$ that belongs to a family of distributions $\{P(X, Y|\mathbf{\Pi} = \boldsymbol{\pi}) = P(X)P(Y|X, \mathbf{\Pi} = \boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_D}$ **parametrized by a random vector** $\mathbf{\Pi}_n$. The $\{\mathbf{\Pi}_n\}_{n=1}^{+\infty}$ are i.i.d. with distribution $P(\mathbf{\Pi})$ and take their values in a **discrete set** $\mathcal{P}_D$. The source symbols X_n and Y_n take their values in the discrete sets $\mathcal{X}$ and $\mathcal{Y}$, respectively.*

The DP-Source, completely determined by $\mathcal{P}_{\mathrm{D}}$, $P(\mathbf{\Pi})$, and $\{P(X, Y|\mathbf{\Pi} = \boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\mathrm{D}}}$, is stationary and ergodic, see [20, Section 3.5].

Definition 2. (DwP-Source). *A Dynamic without Prior source (X, Y), or DwP-Source, produces a sequence of **independent** symbols $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ drawn $\forall n$ from $P(X_n, Y_n)$ that belongs to a family of distributions $\{P(X, Y|\boldsymbol{\pi}) = P(X)P(Y|X, \boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_D}$ **parametrized by a vector** $\boldsymbol{\pi}_n$. Each $\boldsymbol{\pi}_n$ takes its values in a **discrete set** $\mathcal{P}_D$. The source symbols X_n and Y_n take their values in the discrete sets $\mathcal{X}$ and $\mathcal{Y}$, respectively.*

The DwP-Source, determined by $\mathcal{P}_{\mathrm{D}}$ and $\{P(X, Y|\boldsymbol{\pi})\}_{\boldsymbol{\pi}\in\mathcal{P}_{\mathrm{D}}}$, is non-stationary and non-ergodic [20, Section 3.5]. The only difference between the DP- and DwP-Sources lies in the definition of the parameters $\boldsymbol{\pi}_n$. In the DwP-Source, no distribution for $\boldsymbol{\pi}_n$ is specified, either because its

[1]The four models defined in this section were also introduced with different names in two papers [12], [15], of the same authors. M-Source was for DP-Source, WPM-Source for DwP-Source, P-Source for SP-Source, WP-Source for SwP-Source. The names were changed for the sake of clarity.

distribution is not known or because $\boldsymbol{\pi}_n$ is not modeled as a random variable.

The following models consider a time-invariant parameter vector.

Definition 3. (SP-Source) *A Static with Prior source (X, Y) (SP-Source) produces a sequence of **independent** symbols $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ drawn from a distribution belonging to a family $\{P(X, Y|\boldsymbol{\Theta} = \boldsymbol{\theta}) = P(X)P(Y|X, \boldsymbol{\Theta} = \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_S}$ parametrized by a **random vector** $\boldsymbol{\Theta}$. The random vector $\boldsymbol{\Theta}$, with distribution $P_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, takes its value in a set $\mathcal{P}_S$ that is either **discrete or continuous**. The source symbols X and Y take their values in the discrete sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Moreover, the **realization of the parameter $\boldsymbol{\theta}$ is fixed** for the sequence $\{(X_n, Y_n)\}_{n=1}^{+\infty}$.*

The SP-source, determined by $\mathcal{P}_{\mathrm{S}}$, $P_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, and $\{P(X, Y|\boldsymbol{\Theta} = \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\mathrm{S}}}$, is stationary but non-ergodic [20, Section 3.5].

Definition 4. (SwP-Source). *A Static without Prior source (X, Y) (SwP-Source) produces a sequence of **independent** symbols $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ drawn from a distribution belonging to a family $\{P(X, Y|\boldsymbol{\theta}) = P(X)P(Y|X, \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_S}$ parametrized by a vector $\boldsymbol{\theta}$. The vector $\boldsymbol{\theta}$ takes its value in a set $\mathcal{P}_S$ that is either **discrete or continuous**. The source symbols X and Y take their values in the discrete sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Moreover, the **parameter $\boldsymbol{\theta}$ is fixed** for the sequence $\{(X_n, Y_n)\}_{n=1}^{+\infty}$.*

The SwP-source, completely determined by $\mathcal{P}_{\mathrm{S}}$ and $\{P(X, Y|\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\mathrm{S}}}$, is stationary but non-ergodic [20, Section 3.5]. The only difference between the SP- and SwP-Sources lies in the definition of $\boldsymbol{\theta}$ (no distribution for $\boldsymbol{\theta}$ is specified in the SwP-Model). Note that both the encoder and the decoder are aware of the model characteristics given in Definitions 1 to 4.

In the SW setup, the infimum of achievable rates for our models are given by

1) for the DP-Source [43],

$$R = H(X|Y) \tag{1}$$

where $H(X|Y)$ is calculated from $P(X = x|Y = y) = \sum_{\pi\in\mathcal{P}_{\mathrm{D}}} P(\pi)P(X = x|, Y = y, \pi)$.

2) for the DwP-Source [2],

$$R = \sup_{P(X,Y)\in \text{Conv}\left(\{P(X,Y|\pi)\}_{\pi\in\mathcal{P}_\text{D}}\right)} H(X|Y) \tag{2}$$

where $\text{Conv}(\{P(X,Y|\pi)\}_{\pi\in\mathcal{P}_\text{D}})$ is the convex hull of the elements of $\{P(X,Y|\pi)\}_{\pi\in\mathcal{P}_\text{D}}$,

3) for the SP-Source [24, Theorem 7.3.4],

$$R = P_\Theta\text{-ess.}\sup H(X|Y,\Theta=\theta), \tag{3}$$

where $P_\Theta\text{-ess.}\sup$ is the essential sup (the sup on the support of the distribution) with respect to the prior distribution P_Θ,

4) for the SwP-Source [9],

$$R = \sup_{\theta\in\mathcal{P}_\text{S}} H(X|Y,\theta)\ . \tag{4}$$

We see that for the DwP-Model, the SP-Model, and the SwP-Model, the infimum of achievable rates are given by worst cases defined on the set of values the parameters may take (SP- and SwP-Models), or on the convex hull of this set of values (DwP-Model).

The sets $\mathcal{P}_\text{S}$ and $\mathcal{P}_\text{D}$ may contain some elements inducing an important rate. In this case, one should think of allowing some outage event, *i.e.*, the decoder may be authorized to fail for a given proportion γ of the parameters. From this condition, the failure set should be chosen carefully. In this case, the infimum of achievable rates is simply the worst case rate over the set of conserved parameters. Such an issue was discussed in [14] (achievable rates) and in [12] (design of binary LDPC codes) for the construction of sets of parameters satisfying the outage condition. Here, however, we implicitly assume that the sets $\mathcal{P}_\text{S}$ and $\mathcal{P}_\text{D}$ were already carefully designed, possibly considering an outage constraint.

IV. Encoding

The coding schemes we propose are based on LDPC codes for SW coding. As suggested by [32], [34], LDPC codes initially introduced for channel coding can also be used for SW

coding, after adaptation of the coding process and the decoding algorithm. In channel coding, LDPC codes were proposed for binary-input channels [19] and generalized to non-binary input channels in [11]. The adaptation to the SW setup is described in [32] for the binary case. In this paper, we propose a generalization of this adaptation to the non-binary case. This section describes the encoding part and introduces the involved notations. Note that the encoding part is as in the binary case, except that, now, the encoding operations are performed in GF(q). There are more differences in the decoding part.

We assume that the source symbols X are discrete and belong to GF(q). The SW coding of a source vector $\mathbf{x}$ of length N is performed by producing a vector $\mathbf{s} = H^T\mathbf{x}$ of length $M < N$. The matrix H is sparse, with non-zero coefficients uniformly distributed in GF(q)$\backslash\{0\}$. In the following, $\oplus$, $\ominus$, $\otimes$, $\oslash$ are the addition, subtraction, multiplication and division operators in GF(q), see [33, Chapter 4]. In the bipartite graph representing the dependencies between the random variables of $\mathbf{X}$ and $\mathbf{S}$, the entries of $\mathbf{X}$ are represented by Variable Nodes (VN) and the entries of $\mathbf{S}$ are represented by Check Nodes (CN). The set of CN connected to a VN n is denoted $\mathcal{N}(n)$ and the set of VN connected to a CN m is denoted $\mathcal{N}(m)$. The sparsity of H is determined by the VN degree distribution $\lambda(x) = \sum_{i\geq 2} \lambda_i x^{i-1}$ and the CN degree distribution $\rho(x) = \sum_{i\geq 2} \rho_i x^{i-1}$ with $\sum_{i\geq 2} \lambda_i = 1$ and $\sum_{i\geq 2} \rho_i = 1$. In SW coding, the rate $r(\lambda, \rho)$ of a code is given by $r(\lambda, \rho) = \frac{M}{N} = \frac{\sum_{i\geq 2} \rho_i/i}{\sum_{i\geq 2} \lambda_i/i}$.

In order to perform the encoding of a source vector $\mathbf{X}$, one needs to choose properly the coding rate and to design the LDPC coding matrix, *i.e.*, to impose good degree distributions $(\lambda(x), \rho(x))$ [30], [39]. The performance analysis of Section III suggests the following approach. For the DP-Source, the LDPC coding matrix is designed for the known distribution $P(X|Y)$. For the three other models, the LDPC coding matrix is designed for the worst cases defined by (2)-(4).

V. Decoding algorithm

This section introduces LDPC-based decoding algorithms capable of dealing with the uncertainty on the value of the parameters of the models. For the DP-Source, the decoding algorithm is the sum-product LDPC decoder adapted to SW coding. For the other sources, the LDPC decoding algorithm cannot be used directly because of the lack of knowledge on the parameters. We thus propose to jointly estimate the encoded source sequence $\mathbf{X}_1^N$ and the unknown parameters. This joint estimation is performed with an EM algorithm [25]. A method producing a first coarse estimate of the parameters is also presented to properly initialize the EM algorithm.

A. DP-Source: Standard LDPC decoding

In [32] the standard sum-product LDPC decoding algorithm has been adapted to SW coding of binary sources with perfect correlation channel knowledge. This section generalizes the adaptation of the decoding algorithm to non-binary SW coding. Indeed, in the SW case, one needs to take into account both the probability distribution of X and of the received codeword $\mathbf{s}$. For the DP-Source, the conditional distribution is perfectly determined as

$$P(X_n = k|Y_n = y_n) = \sum_{\boldsymbol{\pi}\in\mathcal{P}_{\mathrm{D}}} P(\boldsymbol{\pi})P(X_n = k|Y_n = y_n, \boldsymbol{\pi}) \ . \tag{5}$$

The sum-product decoder performs an approximate Maximum *A Posteriori* (MAP) estimation of $\mathbf{x}$ from the received codeword $\mathbf{s}$ and the observed side information $\mathbf{y}$. The messages exchanged in the dependency graph are vectors of length q. The initial messages for each VN n are denoted $\mathbf{m}^{(0)}(n, y_n)$, with components

$$m_k^{(0)}(n, y_n) = \log \frac{P(X_n = 0|Y_n = y_n)}{P(X_n = k|Y_n = y_n)}, \ k = 0 \dots q-1 \ . \tag{6}$$

The messages from CN to VN are computed with the help of a particular Fourier Transform (FT), denoted $\mathcal{F}(\mathbf{m})$. Denoting r the unit root associated to GF(q), the i-th component of the FT is given by [30] as $\mathcal{F}_i(\mathbf{m}) = \sum_{j=0}^{q-1} r^{i\otimes j} e^{-m_j} / \sum_{j=0}^{q-1} e^{-m_j}$.

At iteration ℓ, the message $\mathbf{m}^{(\ell)}(m,n,s_m)$ from CN m to VN n is

$$\mathbf{m}^{(\ell)}(m,n,s_m) = \mathcal{A}[\overline{s}_m]\mathcal{F}^{-1}\left(\prod_{n'\in\mathcal{N}(m)\backslash n} \mathcal{F}\left(W\left[\overline{H}_{n'm}\right]\mathbf{m}^{(\ell-1)}(n',m,y_{n'})\right)\right) \tag{7}$$

where $\bar{s}_m = \ominus s_m \oslash H_{n,m}$, $\overline{H}_{n'm} = \ominus H_{n',m} \oslash H_{n,m}$ and $W\left[a\right]$ is the $q\times q$ matrix such that $W\left[a\right]_{k,n} = \delta(a\otimes n\ominus k), 0\leqslant k,n\leqslant q-1$, where $\delta(x)=1$ if $x=0$, $\delta(x)=0$ otherwise. $\mathcal{A}[k]$ is a $q\times q$ matrix that maps a vector message $\mathbf{m}$ into a vector message $\mathbf{l}=\mathcal{A}[k]\mathbf{m}$ with $l_j = m_{j\oplus k} - m_k$. Note that $\mathcal{A}[k]$ does not appear in the channel coding version of the algorithm and is specific to SW coding. The derivation of (7) is shown in the appendix. At a VN n, a message $\mathbf{m}^{(\ell)}(n,m,y_i)$ is sent to the CN m and an *a posteriori* message $\tilde{\mathbf{m}}^{(\ell)}(n,y_n)$ is computed. They both satisfy

$$\mathbf{m}^{(\ell)}(n,m,y_n) = \sum_{m'\in\mathcal{N}(n)\backslash m} \mathbf{m}^{(\ell)}(m',n,s_{m'}) + \mathbf{m}^{(0)}(n,y_n)\ , \tag{8}$$

$$\tilde{\mathbf{m}}^{(\ell)}(n,y_n) = \sum_{m'\in\mathcal{N}(n)} \mathbf{m}^{(\ell)}(m',n,s_{m'}) + \mathbf{m}^{(0)}(n,y_n)\ . \tag{9}$$

From (9), each VN n produces an estimate $\widehat{x}_n^{(\ell)} = \arg\max_k \tilde{m}_k^{(\ell)}(n,y_n)$ of x_n. The algorithm ends if $\mathbf{s} = H^{\mathrm{T}}\widehat{\mathbf{x}}^{(\ell)}$ or if $\ell = L_{\max}$, the maximum number of iterations.

When the conditional distribution $P(Y|X)$ is uncertain, the previously described decoding algorithm cannot be applied directly, because the initial messages (6) cannot be evaluated accurately.

B. SwP-Source: EM algorithm

We first consider the SwP-Source and then extend the proposed algorithm to the cases of the DwP- and SP-Sources. For the SwP-Source, one needs the actual value of the parameter vector $\boldsymbol{\theta}$ because the sum-product LDPC decoder requires the knowledge of the conditional distribution $P(X|Y)$. The EM algorithm is thus used to estimate jointly the source sequence $\mathbf{X}$ and the parameter $\boldsymbol{\theta}$. A method to produce coarse estimates of the parameters is also described.

1) Joint estimation of $\boldsymbol{\theta}$ and $\mathbf{x}$: The joint estimation of the source vector $\mathbf{x}$ and of the parameter $\boldsymbol{\theta}$ from the observed vectors $\mathbf{y}$ and $\mathbf{s}$ is performed via the EM algorithm [25]. Knowing some estimate $\boldsymbol{\theta}^{(\ell)}$ obtained at iteration ℓ, the EM algorithm maximizes, with respect to $\boldsymbol{\theta}$,

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\ell)}) &= E_{\mathbf{X}|\mathbf{y},\mathbf{s},\boldsymbol{\theta}^{(\ell)}} \left[\log P(\mathbf{y}|\mathbf{X}, \mathbf{s}, \boldsymbol{\theta})\right] && (10) \\ &= \sum_{\mathbf{x} \in \mathrm{GF}(q)^n} P(\mathbf{x}|\mathbf{y}, \mathbf{s}, \boldsymbol{\theta}^{(\ell)}) \log P(\mathbf{y}|\mathbf{x}, \mathbf{s}, \boldsymbol{\theta}) && (11) \\ &= \sum_{n=1}^{N} \sum_{k=0}^{q-1} P(X_n = k|y_n, \mathbf{s}, \boldsymbol{\theta}^{(\ell)}) \log P(y_n|X_n = k, \boldsymbol{\theta}) \ . \end{aligned}$$

Solving this maximization problem gives the update equations detailed in Lemma 1. For simplicity, the correlation model between X and Y is assumed to be additive, *i.e.*, there exists a random variable Z such that $Y = X \oplus Z$ and $\boldsymbol{\theta}$ parametrizes the distribution of Z. The Binary Symmetric correlation Channel (BSC) of unknown transition probability $\theta = P(Y = 1|X = 0) = P(Y = 0|X = 1)$ is a special case, where Z is a binary random variable such that $P(Z = 1) = \theta$.

Lemma 1. *Let (X, Y) be a binary SwP-Source. Let the correlation channel be a Binary Symmetric channel (BSC) with parameter $\theta = P(Y = 0|X = 1) = P(Y = 1|X = 0)$, $\theta \in [0, 1]$. The update equation for the EM algorithm is [50]*

$$\theta^{(\ell+1)} = \frac{1}{N} \sum_{n=1}^{N} |y_n - p_n^{(\ell)}| \tag{12}$$

where $p_n^{(\ell)} = P(X_n = 1|y_n, \mathbf{s}, \theta^{(\ell)})$.

Let (X, Y) be a SwP-Source that generates symbols in GF(q). Let the correlation channel be such that $Y = X \oplus Z$, where Z is a random variable in GF(q), and $P(Z = k) = \theta_k$. The update equations for the EM algorithm are

$$\forall k \in \mathit{GF}(q), \ \theta_k^{(\ell+1)} = \frac{\sum_{n=1}^{N} P_{y_n \ominus k, n}^{(\ell)}}{\sum_{n=1}^{N} \sum_{k'=0}^{q-1} P_{y_n \ominus k', n}^{(\ell)}} \tag{13}$$

where $P_{k,n}^{(\ell)} = P(X_n = k|y_n, \mathbf{s}, \boldsymbol{\theta}^{(\ell)})$.

Proof: The binary case is provided by [50]. In the non-binary case, the updated estimate is obtained by maximizing (10) taking into account the constraints $0 \leq \theta_k \leq 1$ and $\sum_{k=0}^{q-1} \theta_k = 1$. ■

Note that $P_{k,n}^{(\ell)} = P(X_n = k | y_n, \mathbf{s}, \boldsymbol{\theta}^{(\ell)})$ in (13) can be estimated with a sum-product algorithm that assumes that the true parameter is $\boldsymbol{\theta}^{(\ell)}$.

2) Initialization of the EM algorithm: We now propose an efficient initialization of the EM algorithm valid for irregular codes and for sources X and Y taking values in GF(q). This generalizes the method proposed in [50] for regular and binary codes. The rationale is to derive a Maximum Likelihood (ML) estimate of $\boldsymbol{\theta}$ from a function $\mathbf{u} = H^{\mathrm{T}}\mathbf{x} \oplus H^{\mathrm{T}}\mathbf{y}$ of the observed data $H^{\mathrm{T}}\mathbf{x}$ and $\mathbf{y}$.

a) The BSC with irregular codes: In this case, each binary random variable U_m is the sum of random variables of $\mathbf{Z}$. Although each Z_n appears in several sums, the following assumption is made in this section.

Assumption 1. *Each U_m is obtained from* i.i.d. *random variables $Z_j^{(m)}$.*

The validity of this assumption depends on the choice of the matrix H and is not true in general. Although it produces an approximate solution, this choice may lead to a reasonable initialization for the EM algorithm. Furthermore, the number of terms in the sum for U_m depends on the degree of the CN m. The maximum possible CN degree is denoted d_c. One can use the CN degree distribution $\rho(x)$ as a probability distribution for the degrees, or decide to take into account the knowledge of the CN degrees. Both cases lead to a probability model for the U_m and enable to obtain an ML estimate for θ, as described in the two following lemmas. Note that the lemmas of this section describe the parameter estimation for generic random variables U and Z following Assumption 1.

Lemma 2. *Let* $\mathbf{U}$ *be a binary random vector of length M. Each U_m is the sum of J_m identically*

distributed binary random variables $Z_j^{(m)}$, i.e., $U_m = \sum_{j=1}^{J_m} Z_j^{(m)}$, *where the* $Z_j^{(m)}$ *are independent* $\forall j, m$. $\{J_m\}_{m=1}^M$ *are* i.i.d. *random variables taking their values in* $\{2, \ldots, d_c\}$ *with known probability* $P(J = j) = \rho_j$. *Denote* $\theta = P(Z = 1)$, $\alpha = P(U = 1)$ *and assume that* θ *and* α *are unknown. Then their ML estimates* $\widehat{\theta}$ *and* $\widehat{\alpha}$ *from an observed vector* $\mathbf{u}$ *satisfy* $\widehat{\alpha} = \frac{1}{M}\sum_{m=1}^M u_m$ *and* $\widehat{\theta} = f^{-1}(\widehat{\alpha})$, *where* f *is the invertible function* $f(\theta) = \frac{1}{2} - \frac{1}{2}\sum_{j=2}^{d_c} \rho_j (1-2\theta)^j$, $\forall \theta \in [0, \frac{1}{2}]$.

Proof: The random variables U_m are independent (sums of independent variables). They are identically distributed because the J_m and the $Z_j^{(m)}$ are identically distributed. $\alpha = P(U = 1) = \sum_{j=2}^{d_c} \rho_j P(U = 1|J = j)$. Then, from [50], $P(U = 1|J = j) = \sum_{i=1, i \text{ odd}}^{j} \binom{j}{i} \theta^i (1-\theta)^{j-i}$ and from [19, Section 3.8], $P(U = 1|J = j) = \frac{1}{2} - \frac{1}{2}(1-2\theta)^j$. Thus $\alpha = f(\theta)$. The ML estimate $\widehat{\alpha}$ of α given $\mathbf{u}$ is $\widehat{\alpha} = \frac{1}{M}\sum_{m=1}^M u_m$. Finally, as f is invertible for $\theta \in \left[0, \frac{1}{2}\right]$, then from [29, Theorem 7.2], the ML estimate of θ is given by $\widehat{\theta} = f^{-1}(\widehat{\alpha})$. ■

Lemma 3. *Let* $\mathbf{U}$ *be a binary random vector of length* M. *Each* U_m *is the sum of* j_m *identically distributed binary random variables* $Z_j^{(m)}$, i.e., $U_m = \sum_{j=1}^{j_m} Z_j^{(m)}$, *where* $Z_j^{(m)}$ *are independent* $\forall j, m$. *The values of* j_m *are known and belong to* $\{2, \ldots, d_c\}$. *Denote* $\theta = P(Z = 1)$ *and assume that* θ *is unknown. Then the entries of* $\mathbf{U}$ *are independent and the ML estimate* $\widehat{\theta}$ *from an observed vector* $\mathbf{u}$ *is the argument of the maximum of*

$$L(\theta) = \sum_{j=2}^{d_c} \mathbb{N}_{1,j}(\mathbf{u}) \log\left(\frac{1}{2} - \frac{1}{2}(1-2\theta)^j\right) + \sum_{j=2}^{d_c} \mathbb{N}_{0,j}(\mathbf{u}) \log\left(\frac{1}{2} + \frac{1}{2}(1-2\theta)^j\right) \tag{14}$$

where $\mathbb{N}_{1,j}(\mathbf{u})$ *and* $\mathbb{N}_{0,j}(\mathbf{u})$ *are the number of symbols in* $\mathbf{u}$ *obtained from the sum of* j *elements and respectively equal to* 1 *and* 0.

Proof: The random variables U_m are independent (sums of independent variables). Therefore, the likelihood function satisfy $L(\theta) = \log P(\mathbf{u}|\theta) = \sum_{m=1}^M \log P(u_m|j_m, \theta)$. Then, as in the proof of Lemma 2, we obtain (14). ■

The method of Lemma 2 is simpler to implement than the one of Lemma 3 but does not take into account the actual matrix H, at the price of a small loss in performance.

b) The non-binary discrete case: Only the case of regular codes is presented here, but the method can be generalized to irregular codes (see the previous section). Assumption 1 also holds in this case. Now, the probability mass function of Z is given by $\boldsymbol{\theta} = [\theta_0 \dots \theta_{q-1}]$ with $\theta_k = P(Z = k)\ \forall k \in \text{GF}(q)$. Now, each U_m is the sum of symbols of $\mathbf{Z}$, weighted by the coefficients contained in H. A first solution does not exploit the knowledge of these coefficients, but uses the fact that the non-zero coefficients of H are distributed uniformly in GF(q)$\backslash\{0\}$ (Lemma 4). A second solution takes into account the knowledge of the coefficients (Lemma 5).

Lemma 4. *Let* $\mathbf{U}$ *be a length* M *random vector with entries in GF(q) such that each* U_m *is the sum of* d_c i.i.d. *products of random variables,* i.e., $U_m = \sum_{j=1}^{d_c} H_j^{(m)} Z_j^{(m)}$. *The* $Z_j^{(m)}$ *and* $H_j^{(m)}$ *are identically distributed random variables, mutually and individually independent* $\forall j, m$. *The* $H_j^{(m)}$ *are uniformly distributed in GF(q)*$\backslash\{0\}$. *The* $Z_j^{(m)}$ *take their values in GF(q). Denote* $\theta_k = P(Z = k)$, $\alpha_k = P(U = k)$ *and assume that* $\boldsymbol{\theta} = [\theta_0 \dots \theta_{q-1}]$ *and* $\boldsymbol{\alpha} = [\alpha_0 \dots \alpha_{q-1}]$ *are unknown. Then the random variables of* $\mathbf{U}$ *are independent and the parameters satisfy* $\boldsymbol{\alpha} = f(\boldsymbol{\theta})$, *with*

$$f(\boldsymbol{\theta}) = \sum_{n_1,\dots,n_{q-1}} \binom{d_c}{n_1, \dots, n_{q-1}} \left(\frac{1}{q-1}\right)^{d_c} \mathcal{F}^{-1}\left(\prod_{j=0}^{q-1} \left(\mathcal{F}\left(W[j]\boldsymbol{\theta}\right)\right)^{n_j}\right) \tag{15}$$

where the sum is over all the possible combinations of integers $n_1, \dots, n_{q-1}$ *such that* $0 \leq n_k \leq d_c$ *and* $\sum_{k=1}^{q-1} n_k = d_c$ *and* $\binom{d_c}{n_1,\dots,n_{q-1}}$ *is a multinomial coefficient.*

Denote $\widehat{\boldsymbol{\theta}}$ *and* $\widehat{\boldsymbol{\alpha}}$ *the ML estimates of* $\boldsymbol{\theta}$ *and* $\boldsymbol{\alpha}$, *obtained from an observed vector* $\mathbf{u}$, *with* $\widehat{\alpha}_k = \frac{\mathbb{N}_k(\mathbf{u})}{M}$ *where* $\mathbb{N}_k(\mathbf{u})$ *is the number of occurrences of* k *in the vector* $\mathbf{u}$. *Then, if* f *is invertible,* $\widehat{\boldsymbol{\theta}} = f^{-1}(\widehat{\boldsymbol{\alpha}})$.

Proof: The random variables U_m are independent (sums of independent variables). Then, $\alpha_k = P(U = k) = \sum_{\{h_j\}_{j=1}^{d_c}} P(\{h_j\}_{j=1}^{d_c}) P(U = k | \{h_j\}_{j=1}^{d_c})$ in which the sum is on all the possible combinations of coefficients $\{h_j\}_{j=1}^{d_c}$. This can be simplified as $\alpha_k = \sum_{n_1,\dots,n_{q-1}} P(N_1 = n_1, \dots, N_{q-1} = n_{q-1}) P(U = k | n_1, \dots, n_{q-1})$ where n_k is the number of occurrences of k in

$\{h_j\}_{j=1}^{d_c}$. One has $P(N_1 = n_1, \ldots, N_{q-1} = n_{q-1}) = \binom{d_c}{n_1,\ldots,n_{q-1}} \left(\frac{1}{q-1}\right)^{d_c}$. Then, the vector denoted

$$P_{\mathbf{U}|n_1,\ldots,n_{q-1}} = [P(U = 0|n_1, \ldots, n_{q-1}) \ldots P(U = q-1|n_1, \ldots, n_{q-1})] \tag{16}$$

can be expressed as $P_{\mathbf{U}|n_1,\ldots,n_{q-1}} = \mathcal{F}^{-1}\left(\prod_{j=1}^{q-1} \left(\mathcal{F}\left(W[j]\boldsymbol{\theta}\right)\right)^{n_j}\right)$. Therefore,

$$\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{q-1}] = \sum_{n_1,\ldots,n_{q-1}} \binom{d_c}{n_1, \ldots, n_{q-1}} \left(\frac{1}{q-1}\right)^{d_c} \mathcal{F}^{-1}\left(\prod_{j=1}^{q-1} \left(\mathcal{F}\left(W[j]\boldsymbol{\theta}\right)\right)^{n_j}\right) . \tag{17}$$

The ML estimates $\widehat{\alpha}_k$ of α_k are $\widehat{\alpha}_k = \frac{\mathbb{N}_k(\mathbf{u})}{M}$. Finally, if f is invertible, then from [29, Theorem 7.2], the ML estimate of $\boldsymbol{\theta}$ is given by $\widehat{\boldsymbol{\theta}} = f^{-1}(\widehat{\boldsymbol{\alpha}})$. ■

Lemma 5. *Let* $\mathbf{U}$ *be a length* M *random vector with entries in GF(q) such that each* U_m *is the sum of* d_c i.i.d. *random variables,* i.e.*,* $U_m = \sum_{j=1}^{d_c} h_j^{(m)} Z_j^{(m)}$. *The* $Z_j^{(m)}$ *are independent* $\forall j, m$*, and identically distributed random variables taking their values in GF(q). The values of the coefficients* $h_j^{(m)}$ *are known and belong to GF(q)*$\backslash\{0\}$. *Denote* $\theta_k = P(Z = k)$*,* $\alpha_k = P(U = k)$ *and assume that* $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_{q-1}]$ *and* $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{q-1}]$ *are unknown. Then the random variables of* $\mathbf{U}$ *are independent and the ML estimate* $\widehat{\boldsymbol{\theta}}$ *from an observed vector* $\mathbf{u}$ *maximizes*

$$L(\boldsymbol{\theta}) = \sum_{m=1}^{M} \log \mathcal{F}_{u_m}^{-1}\left(\prod_{j=1}^{dc} \mathcal{F}(W[h_j^{(m)}]\boldsymbol{\theta})\right) \tag{18}$$

under the constraints $0 \leq \theta_k \leq 1$ *and* $\sum_{k=0}^{q-1} \theta_k = 1$.

Proof: The random variables U_m are independent (sums of independent variables). The ML estimate $\widehat{\boldsymbol{\theta}}$ is the value that maximizes the likelihood function given by

$$\begin{aligned} L(\boldsymbol{\theta}) &= \log P(\mathbf{u}|\boldsymbol{\theta}, \{h_j^{(m)}\}_{j=1,m=1}^{d_c,M}) && (19)\\ &= \sum_{m=1}^{M} \log P(u_m|\boldsymbol{\theta}, \{h_j^{(m)}\}_{j=1}^{d_c}) && (20) \end{aligned}$$

under the constraint that $0 \leq \theta_k \leq 1$ and $\sum_{k=0}^{q-1} \theta_k = 1$. The second equality (20) comes from the independence assumption. Following the steps of Lemma 4, we show that (20) becomes $L(\boldsymbol{\theta}) = \sum_{m=1}^{M} \log \mathcal{F}_{u_m}^{-1}\left(\prod_{j=1}^{d_c} \mathcal{F}(W[h_j^{(m)}]\boldsymbol{\theta})\right)$. ■

C. DwP-Source

The DwP-Source is non-stationary. Consequently, if one assumes a stationary model such that

$$P(X_n = k|Y_n = m) = \alpha_{k,m} \tag{21}$$

and tries to produce an estimate $\hat{\alpha}_{k,m}^{(n)}$ from observed sequences $(\mathbf{x}, \mathbf{y})$ of length n, then the sequence of estimates $\hat{\alpha}_{k,m}^{(n)}$ does not necessarily converge as n goes to infinity. However, such an estimate is well defined for a fixed length n. Thus, we apply the procedure defined for the SwP-Source to get $\hat{\alpha}_{k,m}^{(n)}$ from $\mathbf{y}$ and $\mathbf{u}$.

D. SP-Source: MAP with EM

For the SP-Source, the distribution $P_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is available and one can perform the MAP estimation of $\boldsymbol{\Theta}$. Then, the EM equation (10) for the MAP estimation becomes [3]

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\ell)}) = E_{\mathbf{X}|\mathbf{y},\mathbf{s},\boldsymbol{\theta}^{(\ell)}} \left[\log P(\mathbf{X}|\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})\right] + \log P_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \ . \tag{22}$$

Knowing some estimate $\boldsymbol{\theta}^{(\ell)}$ of $\boldsymbol{\theta}$ at iteration ℓ, one has to maximize (22) with respect to $\boldsymbol{\theta}$ to obtain $\boldsymbol{\theta}^{(\ell+1)}$. As for the SwP-Source, the LDPC decoding algorithm initialized with $\boldsymbol{\theta}^{(l)}$ provides an approximate version of $P(\mathbf{X}|\mathbf{y}, \mathbf{s}, \boldsymbol{\theta}^{(\ell)})$, required to perform the MAP estimation of $\boldsymbol{\theta}^{(l+1)}$.

A coarse estimation of $\boldsymbol{\theta}$ can be obtained from $\mathbf{u} = H^{\mathrm{T}}\mathbf{x} + H^{\mathrm{T}}\mathbf{y}$ as

$$\boldsymbol{\theta}^{(0)} = \arg\max_{\boldsymbol{\theta} \in \mathcal{P}_{\mathrm{S}}} \log P_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) + \log P(\mathbf{u}|H, \boldsymbol{\theta}) \tag{23}$$

in order to initialize the EM algorithm. In the binary case and from the assumptions of Lemma 3 this corresponds to maximizing

$$L_{\mathrm{MAP}}(\theta) = \log P_{\Theta}(\theta) + \sum_{j=2}^{d_c} \mathbb{N}_{1,j}(\mathbf{u}) \log\left(\frac{1}{2} - \frac{1}{2}(1-2\theta)^j\right) + \sum_{j=2}^{d_c} \mathbb{N}_{0,j}(\mathbf{u}) \log\left(\frac{1}{2} + \frac{1}{2}(1-2\theta)^j\right) \tag{24}$$

with respect to θ. In the non-binary case and from the assumptions of Lemma 5 this corresponds to maximizing

$$L_{\mathrm{MAP}}(\boldsymbol{\theta}) = \log P_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) + \sum_{m=1}^{M} \log \mathcal{F}_{u_m}^{-1}\left(\prod_{j=1}^{dc} \mathcal{F}(W[h_{s_{zm},j}]\boldsymbol{\theta})\right) \tag{25}$$

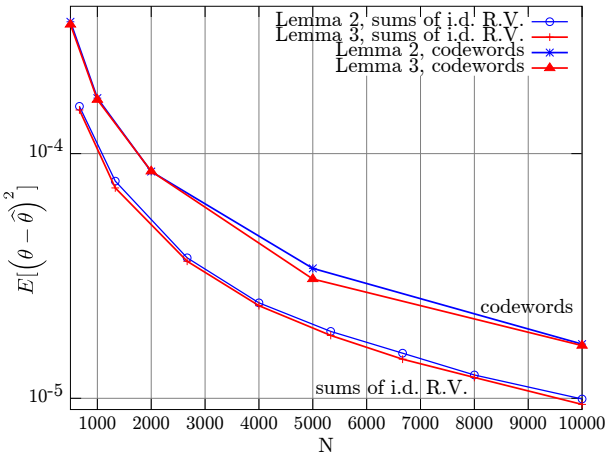

Fig. 1. MSE of the estimators for the binary case. (SwP-Source)

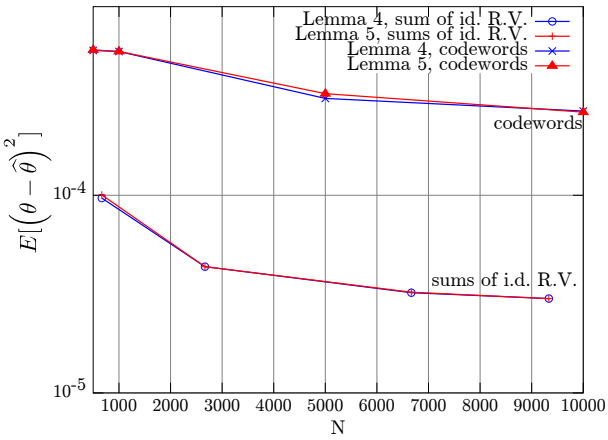

Fig. 2. MSE of the estimators for the non-binary case. (SwP-Source)

under the constraints $0 \leq \theta_k \leq 1$ and $\sum_{k=0}^{q-1} \theta_k = 1$.

However, this approach does not fully exploit the density over $\boldsymbol{\theta}$ but only its mode, because a hard value of $\boldsymbol{\theta}$ is estimated at each iteration and used for the following iterations. To deal with this problem, one could think of using Variational Bayesian Expectation Maximization (VBEM) [3]. Unfortunately, the VBEM equations are intractable for most of the distributions, particularly in the discrete case. The discrete additive model considered here is not a conjugate exponential model, for which a tractable implementation exists.

VI. Simulations

The performance of the initialization techniques obtained in Lemmas 2 to 5 are first compared. Then, we evaluate the joint estimation methods proposed for the various models introduced in Section III. The correlation model is such that there exists a random variable Z with $Y = X \oplus Z$. and X is distributed uniformly.

A. Performance of the initialization techniques (SwP-Model)

The binary case is considered first. X is distributed uniformly and Z is such that $P(Z = 1) = \theta$, $\theta \in \mathcal{P}_S = [0, 0.18]$. The worst case $\overline{\theta}$ gives $H(X|Y, \overline{\theta}) = 0.68$ bit/symbol. We choose a code

of rate $R = 0.75$ bit/symbol and of edge-perspective degree distributions $\lambda(x) = x^2$ and $\rho(x) = 0.4823x^2 + 0.2701x^3 + 0.0057x^4 + 0.0718x^5 + 0.0602x^{16} + 0.0732x^{17} + 0.0075x^{35} + 0.0292x^{36}$, designed for the worst possible parameter $\theta = 0.18$ and obtained from a code optimization based on density evolution and realized with a differential evolution algorithm [47]. Here, as X is assumed uniformly distributed, density evolution for channel coding can be directly used by the optimization algorithm. If the source is not distributed uniformly, density evolution has to be performed on an equivalent channel, as described in [7]. The initialization methods of Lemmas 2 and 3 are evaluated and compared through two experiments. Indeed, the models defined in the formulations of the lemmas are supposed to represent the behavior of the LDPC encoding using Assumption 1. In this section, we want to determine whether this assumption is meaningful.

First, we wish to evaluate the performance of the estimation methods on simulated codewords, *i.e.*, generated at random from the models as they are defined in the formulations of the lemmas. For that purpose, 10000 vectors $\mathbf{U}$ of length M are generated according to the models introduced in Lemmas 2 and 3, for $\theta = 0.12$. Assumption 1 is taken into account and the symbols U_m are drawn as sums of independent random variables. Then, the two proposed estimation methods are applied and the Mean Squared Error (MSE) $E\left[(\theta - \widehat{\theta})^2\right]$ is evaluated as a function of $N = \frac{M}{R}$. The estimated parameters are obtained numerically from a gradient descent initialized at random in $\mathcal{P}_{\mathrm{S}}$. This gives the two superposed lower curves of Figure 1, showing that the methods of the two lemmas provide similar performance.

Second, as the models introduced in the lemmas are supposed to represent the effects of the LDPC encoding, we also evaluate the performance of the estimators on actual codewords, *i.e.*, obtained from LDPC coding. Consequently, 10000 vectors $\mathbf{z}$ of length N are generated considering $\theta = 0.12$. Note that the estimation method requires the knowledge of $\mathbf{u} = H^{\mathrm{T}}\mathbf{y} \ominus H^{\mathrm{T}}\mathbf{x} = H^{\mathrm{T}}\mathbf{z}$ and thus vectors $\mathbf{z}$ are generated directly. The vectors $\mathbf{u}$ are then obtained by multiplying $\mathbf{z}$ by a matrix H of the considered code. The two proposed estimation methods are then applied to each realization to evaluate the MSE. This gives the two superposed upper curves

of Figure 1. As before the two methods give the same performance. However, we observe a loss of a factor 10 in MSE compared to the ideal case, due to the fact that the entries of $\mathbf{U}$ are not independent. Nevertheless, the performance seems sufficient for the initialization of the EM algorithm.

For the non-binary case, X is distributed uniformly and the probability distribution of Z is given by $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_3]$ where $P(Z = k) = \theta_k$. The set $\mathcal{P}_{\mathrm{S}}$ is such that $\forall \boldsymbol{\theta} \in \mathcal{P}_{\mathrm{S}}$, $\theta_0 \geq \overline{\theta}$ and $\overline{\theta}$ is fixed. We choose a code with edge perspective degree distributions $\lambda(x) = x^2$ and $\rho(x) = 0.5038x^2 + 0.2383x^3 + 0.0035x^4 + 0.00354x^5 + 0.0033x^{10} + 0.1252x^{11} + 0.0256x^{12} + 0.0089x^{18} + 0.0260x^{19} + 0.0301x^{20}$, giving $R = 1.5$ bit/symbol. In this case, the code was tuned for the worst case $\overline{\boldsymbol{\theta}} = [\overline{\theta}, (1-\overline{\theta})/3, (1-\overline{\theta})/3, (1-\overline{\theta})/3]$ where we consider the particular case $\overline{\theta} = 0.7$ giving entropy $H(X|Y, \overline{\boldsymbol{\theta}}) = 1.36$ bit/symbol. As the source symbols are distributed uniformly, the code optimization is realized from a channel coding density evolution technique realized with MCMC simulations as described in [22]. If the source symbols were not distributed uniformly, one could not simply apply the channel coding density evolution to the correlation channel $P(Y|X)$. In fact, in channel coding, the inputs of the channel are distributed uniformly. However, density evolution could be applied on a particular transformed channel with the same performance as for $P(Y|X)$. This transformed channel is determined in [7] for the binary case, and in [16] for the non-binary case. The code has then been constructed with an LDPC PEG (Progressive Edge Growth) algorithm [26]. Note that although the density evolution exhibits good performance for the selected degree distribution, the code construction at finite length introduces a loss in performance because of the cycles appearing in the decoding graph [36].

The experiments described in the binary case are repeated for the methods proposed in Lemmas 4 and 5. The parameter estimates are now obtained from a projected gradient descent. Figure 2 shows the MSE of the two cases obtained by averaging over 10000 vectors of different length N, generated from $\boldsymbol{\theta} = [0.79\ 0.07\ 0.07\ 0.07]$. The conclusions of the binary case hold also in this setup and in the following, the method of Lemma 4 is used since it is less complex.

Note that other cases may be considered. For example, we could assume that $\forall \boldsymbol{\theta} \in \mathcal{P}_{\mathrm{S}}$, $\theta_i \geq \overline{\theta}$, for some $i \neq 0$. We could also assume a combination of cases, such as $\theta_0 > \overline{\theta}$ or $\theta_1 > \overline{\theta}$. Indeed, all these cases give the same achievable rate in (2). The propose decoding method was shown to perform good as well for these cases (after a slight adaptation), see [13]. Otherwise, in this case, the set of channels is not degraded anymore, and thus it becomes more difficult to design good degree distributions. Indeed, it is shown that for a set of degraded channels, if a code of given degree distributions is good for the worst channel (*i.e.* sufficiently low error probability), it is also good for any channel in the set. On the opposite, if the set of possible channels is not degraded, one has to ensure that the code perform good for any individual channel in the set.

B. Complete coding scheme for the SwP- and SP-Sources

The performance of the complete scheme is now evaluated, in the non-binary case.

As for the initialization technique, X is distributed uniformly and the probability distribution of Z is given by $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2, \theta_3]$. The case of the SwP-Source is treated first, and four setups are compared. In each setup, 1000 source vectors of length 10000 are generated. The evaluation procedure is as follows. We choose three codes of different rates, obtained from the previously mentioned code optimization. The codes have the following edge-perspective degree distributions $\lambda(x) = x^2$ and $\rho(x) = 0.5038x^2 + 0.2383x^3 + 0.0035x^4 + 0.00354x^5 + 0.0033x^{10} + 0.1252x^{11} + 0.0256x^{12} + 0.0089x^{18} + 0.0260x^{19} + 0.0301x^{20}$, giving $R = 1.5$ bit/symbol. (1.5 bit/symbol). For each realization, $\boldsymbol{\theta}$ is generated randomly from the set $\mathcal{P}_{\mathrm{S}}$ such that $\theta_0 > p$, where p is fixed. For every described setup, p varies from 0.67 (entropy of 1.42 bit/symbol) to 0.71 (entropy of 1.33 bit/symbol). We set 20 iterations for the LDPC decoder and 3 iterations for the EM algorithm (when required). The results are represented in Figure 3.

In the deterministic setup, $\boldsymbol{\theta}$ is fixed and equal to $[1-p, (1-p)/3, (1-p)/3, (1-p)/3]$. The distribution is given to the decoder. This gives the error floor of the chosen code. Note that the error is high compared to the other setups, because in the other setups, $\boldsymbol{\theta}$ is generated at

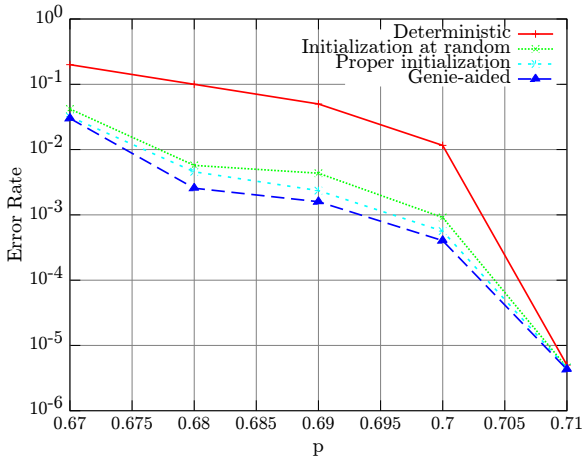

Fig. 3. Error rate with respect to p for the SwP-Source

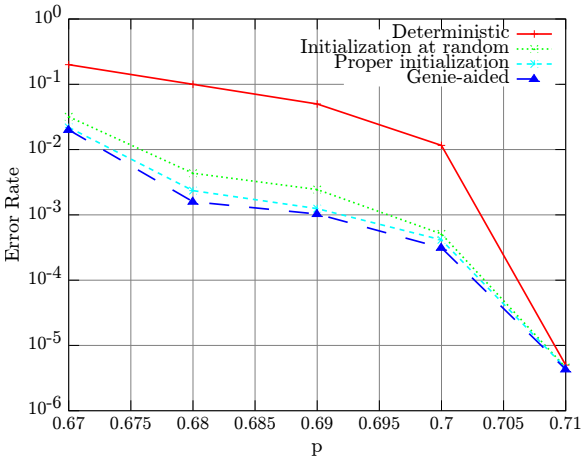

Fig. 4. Error rate with respect to p for the SP-Source

random and more favorable cases appear. For the genie-aided setup, $\boldsymbol{\theta}$ is given to the decoder. In the third setup, the EM algorithm is initialized at random. The fourth setup corresponds to the method presented in the paper. Coarse estimate of $\boldsymbol{\theta}$ obtained from Lemma 4 initializes the EM algorithm. We see that the EM initialized at random gives better result. Furthermore, the mean decoding time increases by a factor 1.5 when $\boldsymbol{\theta}$ is initialized at random. We see that this method increases the decoding time and produces poor performance.

For the SP-Source, the same model, codes and procedure are considered. The prior distribution on θ_0 is a triangle distribution centered on $\overline{\theta} + 1/2(1-\overline{\theta})$. The other components are distributed uniformly according to the probability distribution constraints. The three setups: genie-aided, EM initialized at random, method described in the paper, are tested again over 1000 source vectors of length 10000. The results are represented in Figure 4.

C. Comparison to a solution with feedback

In this section we compare our no-feedback coding approach with a 1-bit feedback transmission for a source generated from the SwP-Model of Section VI-B. The 1-bit feedback is sent by the receiver to the encoder to ask for additional packets or stop the transmission. The goal is to

save rate by avoiding sending data at the worst rate as in the no-feedback method. However, it results in multiple decoding trials and thus potentially large delays. It is therefore of interest to study the rate/decoding delay tradeoff.

Only an evaluation of the achievable rate and estimated mean-time decoding are provided. They are sufficient to determine the advantages and the drawbacks of the solution with feedback. In the solution with feedback, when the decoder cannot decode with the received codeword, it requests more check equations via the feedback channel. Each time it receives new equations, the decoder tries to reconstruct the source vector with the use of a sum-product LDPC decoder.

Denote N the length of the source vector and assume that $\boldsymbol{\theta}$ is of the form $\boldsymbol{\theta} = [1-3\theta, \theta, \theta, \theta]$ where $\theta \leq \overline{\theta} = 0.08$. Consider K rate levels $R_1, \ldots, R_K$ associated to K intervals $I_1 = [0, \overline{\theta}/K] \ldots I_K = [\overline{\theta}(K-1)/K, \overline{\theta}]$. The coding system processes as follows. The encoder first sends nR_1 symbols to the decoder. The decoder tries to reconstruct the source, assuming the true parameter is $\overline{\theta}/K$. If it fails, it sends a request via the feedback channel and the encoder sends $N(R_2 - R_1)$ new symbols. The decoder then tries to reconstruct the source from the nR_2 received symbols, assuming the true parameter is $2 \times \overline{\theta}/K$. The process continues until the source vector has been decoded. Note that here, it is assumed that the I_k intervals are small enough to allow the decoder to perform well with a parameter that is not exactly the true one.

Five setups are compared, in terms or achievable rate (R) and of estimated mean decoding time (T). The results are shown in Figure 5. Denote t the decoding time of one LDPC decoder iteration and N_{it} the required number of iterations. In the following, we set $K = 8$, $N_{\text{it}} = 20$ and choose $t = 4s$ from the previous experiments. Ddenote $h(\theta) = H(X|Y, \theta)$. In each case, we assume that a code or a sequence of codes reaching the entropy can be constructed.

For the solution with feedback, assume that we can construct a sequence of codes such that $R_1 = h(\overline{\theta}/K), \ldots R_K, = h(\overline{\theta})$ and achieving small probability of error respectively for $\theta \in I_1, \ldots, \theta \in I_K$. Thus, $\theta \in I_k$, $R_k = h(k\overline{\theta}/K)$. We also assume that the delay induced by the feedback is negligible compared to the decoding time. Then, for $\theta \in I_k$ the mean decoding

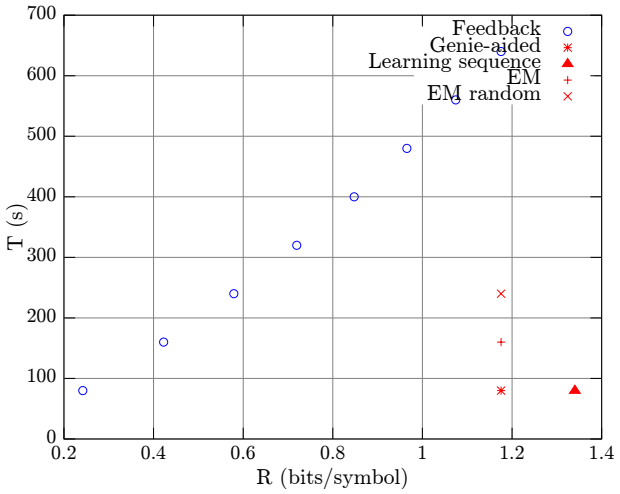

Fig. 5. Rate/Time performance of a solution with feedback

time is estimated as $T_k = t \times N_{\text{it}} \times k$. In the curve of Figure 5, the circles represent the various (R_k, T_k).

For the genie-aided setup, the rate is dimensioned for the worst case, *i.e.*, $R = h(\overline{\theta})$ and an approximation of the mean decoding time is calculated as $T = N_{\text{it}} \times t$. For the setup with learning sequence, assuming a sequence length representing a fraction $1/5$ of the total length, $R = 4/5h(\overline{\theta}) + 1/5H(X)$ and $T = N_{\text{it}} \times t$. For the coding scheme described in the paper, $R = h(\overline{\theta})$ and we approximate $T = 2 \times N_{\text{it}} \times t$, assuming that 2 iterations of the EM algorithm are required. For the coding scheme with EM initialized at random, $R = h(\overline{\theta})$ and we approximate $T = 4 \times N_{\text{it}} \times t$, assuming that 4 iterations of the EM algorithm are required.

When the parameter θ is small, the solution with feedback induces a significant rate gain. However, when θ increases, the price to pay for adapted rate is a very large decoding delay. The choice of the parameter K is important: if K decreases, the size of the intervals I_k increases which reduces the mean decoding time. On the other hand, as for $\theta \in I_k$, the effective coding rate is R_k, the rate needed to decode for θ can increase.

TABLE I

SETUP COMPARISON FOR THE DWP-SOURCE

m	1	100	500	1000
Err (setup 1)	1.91×10^{-3}	0.242×10^{-2}	0.247×10^{-2}	0.32×10^{-2}
Err (setup 2)	4.31×10^{-5}	5.22×10^{-5}	5.23×10^{-5}	5.3×10^{-5}

D. DwP-Source

The solution proposed for the DwP-Source is now evaluated in the non-binary case. The distribution of Z is given by $\boldsymbol{\pi} = [\pi, (1-\pi)/3, (1-\pi)/3, (1-\pi)/3]$. Two setups are considered. In setup 1, π can take the values $\{0.67, 0.7, 0.73\}$. In setup 2, π can take the values $\{0.7, 0.73, 0.76\}$ We now consider source vectors of length 10000 and fix a block length m. For each block of length m in a vector, a probability distribution for the states is generated uniformly at random. The values $m = 1, 100, 500$ and 1000 are tested. The method proposed for the SwP-Source is then applied with the same code over 1000 realizations for each m. The complete decoding technique described for the SwP-Source is used: coarse estimate of the parameter from Lemma 4 followed by EM algorithm. The results are presented in Table I. Compared to a case where $\boldsymbol{\theta}$ is fixed, we see that there is a loss in performance.

VII. CONCLUSION

This paper introduced four signal models modeling the uncertainty on the correlation channel between the source and the SI. Practical coding schemes based on non-binary LDPC codes were proposed for the SW setup and for the four models. Simulation results exhibit good performance in terms of probability of error, rate, or decoding delay, compared to the solution with a learning sequence or the solution with an EM algorithm initialized at random.

Here, only the additive case was considered. In fact, if the correlation channel is not additive, it may be described by an unknown (or partly unknown) probability transition matrix P with

$P_{i,j} = P(Y = j|X = i)$. The EM equations of Lemma 1 can be restated in this case but the problem is on the initialization of the EM algorithm. Indeed, the defined matrix P can cover a wide range of situations. For example, the set $\mathcal{P}_{\mathrm{S}}$ may be such that $P_{i,1} > 0.7\ \forall i$, or such that $P_{i,i} > 0.7$, a combination of these cases or anything else. If the EM algorithm is not initialized with the proper form of P, it will not be able to converge. Unfortunately, as pointed out in [13], the initialization method proposed here does not enable to produce a reasonable initial estimate of P, because it cannot make a distinction between the possible matrix structures.

Future works will be on the design of good degree distributions for our models with non-binary symbols, and on the extension to the lossy case. We will also investigate correlation model selection, *i.e.*, the choice of one of the four source correlation models and of the structure of the family distribution for the model.

ACKNOWLEDGMENTS

The authors would like to thank Valentin Savin for the helpful discussions and advices.

APPENDIX

In this appendix, we detail the derivation of the update rule (7) at a CN for the SW problem, when the LDPC code is non binary and the decoder is the sum-product algorithm. This update rule derives from the parity check equation at CN m, given by $\sum_{n' \in \mathcal{N}(m)} H_{m,n'} \otimes x_{n'} = s_m$, that can be restated as

$$x_n = s_m \oslash H_{m,n} \ominus \sum_{n' \in \mathcal{N}(m) \setminus n} (H_{m,n'} \oslash H_{m,n}) \otimes x_{n'} \ . \tag{26}$$

The update rule at a CN, and for the sum-product algorithm, consists in computing the reliability information on the variable x_n as a function of the reliability information on the variables $x_{n'}$, denoted $\mathbf{m}^{(\ell-1)}(n', m, y_{n'})$. Thus, the k-th component of the CN message m to VN n (7) is

$$\log \frac{P\left(X_n = 0 | s_m, \{\mathbf{m}^{(\ell-1)}(n', m, y_{n'})\}_{n' \in \mathcal{N}(m) \setminus n}\right)}{P\left(X_n = k | s_m, \{\mathbf{m}^{(\ell-1)}(n', m, y_{n'})\}_{n' \in \mathcal{N}(m) \setminus n}\right)} \tag{27}$$

Annexe C

Design de codes LDPC non-binaires

Elsa Dupraz, Aline Roumy, Michel Kieffer *Density Evolution for the Design of Non-Binary Low Density Parity Check Codes for Slepian-Wolf Coding* Technical report 2013

Density Evolution for the Design of Non-Binary Low Density Parity Check Codes for Slepian-Wolf Coding

Elsa DUPRAZ[1], Aline ROUMY[2] and Michel KIEFFER[1,3,4]

[1] L2S - CNRS - SUPELEC - Univ Paris-Sud, 91192 Gif-sur-Yvette, France

[2] INRIA, Campus de Beaulieu, 35042 Rennes, France

[3] Partly on leave at LTCI – CNRS Télécom ParisTech, 75013 Paris, France

[4] Institut Universitaire de France

Abstract

In this paper, we investigate the problem of designing good LDPC codes for Slepian-Wolf coding. More specificaly, we consider an asymptotic performance analysis called density evolution. In channel coding, if the channel is symmetric, density evolution can be performed assuming that the all-zero codeword was transmitted. Such an assumption does not hold in Slepian-Wolf coding, even if the correlation channel is symmetric, because of the non-uniform source distribution. In this paper, we show that any, even non-symmetric correlation channel in Slepian-Wolf coding is equivalent under density evolution to a symmetric channel in channel coding. Consequently, density evolution with the all-zero codeword assumption can be performed on this equivalent channel in order to obtain the performance of a code in Slepian-Wolf coding.

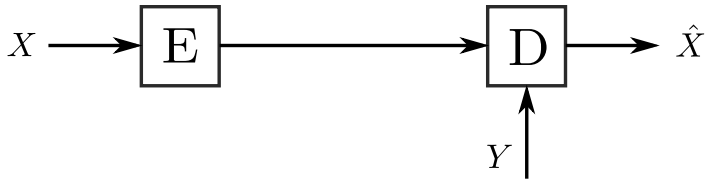

Fig. 1. Asymetric Slepian-Wolf coding

I. Introduction

In this paper, we consider the lossless coding of a source X with the help of some side information Y available at the decoder only (see Figure 1). This setup is called asymmetric Slepian-Wolf (SW) coding [26]. Here, for simplicity, it is referred to as SW coding. For this problem, it is well known that the infimum of achievable rates is given by $H(X|Y)$, the conditional entropy of X knowing Y and several practical coding schemes have been proposed [8], [22], [33]. Most of them are based on channel codes [7], [27], and particularly Low Density Parity Check (LDPC) codes [10], [17], [20]. In source coding, the source symbols are in general non-binary (for example the pixels of an image). A usual coding solution is to transform the non-binary symbols into bits and to encode the bit planes independently with binary LDPC codes. To avoid a performance loss, the dependency between bit planes has to be taken into account at the decoder [15], [32], at the price of a complexity increase. In this paper, in order to avoid this operation, we consider directly non-binary LDPC codes [11].

Many efforts have been made in channel coding for the design of good LDPC codes. In particular, density evolution techniques have been developed both for binary [23], [24], [30] and non-binary [1], [16] codes. A code is described by its variable and check node degree distributions $\lambda(x)$ and $\rho(x)$. From an asymptotic analysis, density evolution gives an evaluation of the error probability of a (λ, ρ)-code for a given channel of input U

and output W described by its conditional distribution $P(W|U)$. Optimization techniques such as differential evolution [29] can then be implemented in order to obtain good degree distributions for the considered channel. Although the issue of constructing properly the coding matrix at finite length remains [21], it can constitute a good starting point to the code design.

In SW coding, one could think of identifying the correlation channel $P(Y|X)$ and then simply applying the standard density evolution derived for channel coding. Unfortunately, as pointed out in [2], [4], a good LDPC code for channel coding is not necessarily good for SW coding. As an example, consider the case of a Binary Symmetric Channel (BSC) defined by $P(W = 1|U = 0) = P(W = 0|U = 1) = p$. The capacity of the BSC is $C_{W|U} = \max_{p(u)} I(U;W) = 1 - H(p)$ [9], *i.e.*, the max is achieved when U is distributed uniformly. In SW coding, consider as well a binary symmetric correlation channel $P(Y = 1|X = 0) = P(Y = 0|X = 1) = p$. Unfortunately in source coding, the source X is not necessarily uniformly distributed and $H(X|Y) \leq H(p)$ with equality if and only if X is distributed uniformly. Consequently, the channel coding scheme and the SW coding scheme require codes of different rate. On the other hand, in channel coding, the decoding error probability for a symmetric channel does not depend on the input codeword [16], [23]. This allows to assume that the all-zero codeword was transmitted and greatly simplifies the density evolution. In SW coding, this result does not hold because of the non-uniform source distribution.

For binary LDPC codes, [4] shows that for every even non-symmetric correlation channel $P(Y|X)$, there exists an equivalent symmetric channel $P(\tilde{W}|\tilde{U})$ for which the all-zero codeword assumption holds. The equivalent channel is such that a (λ, ρ)-code gives the same error probability on both channels, and $C_{\tilde{W}|\tilde{U}} = 1 - H(X|Y)$. Thus for any correlation

channel in SW coding, it suffices to identify the equivalent channel and then to perform traditional density evolution for channel coding. In this paper, we generalize this result to codes in GF(q), the Galois Field of size q. In particular, we derive the equivalence and explain how to design good non-binary LDPC codes. More in details, the contributions of the papers are as follows.

1) In channel coding, we derive a recursive expression of the density evolution for symmetric channels. The obtained analytic expression is difficult to express in an explicit form expression, and is not convenient for practical density evolution. Otherwise, it will be used to express the equivalence.
2) We also derive a recursive expression of the density evolution in SW coding for correlation channels $P(Y|X)$ that are not necessarily symmetric.
3) From the two previous recursions, we derive the channel $P(W|U)$ equivalent to $P(Y|X)$ under density evolution.
4) We present the example of a q-ary symmetric correlation channel $P(Y|X)$ where X is not necessarily distributed uniformly. We derive the equivalent channel and give some results on the performance of some codes.

The paper is organized as follows. Section II presents the related works. Section III introduces the notations and recalls some results on Galois Fields. Section IV restates the non-binary LDPC decoding algorithm for SW coding. Section V expresses the density evolution for SW coding and derives the channel equivalence in the non-binary case. Section VI presents the example of the q-ary symmetric channel..

II. Related Works

First, binary LDPC codes have been used for SW coding in [3], [5], [28], [18], [20] and references therein. In both cases, the LDPC decoder consists of some message passing procedure that is the sum-product algorithm. In the same way, [31] proposes to use non-binary LDPC codes and derives the decoding algorithm expressions. But these works do not provide a solution for the design of good non-binary LDPC codes for SW coding.

On the other hand, density evolution was initially introduced in [23], [24] for binary symmetric channels The case of binary non-symmetric channels was further investigated in [30]. All these works give a recursive expression of the density evolution. Furthermore, [6] proposed to approximate the messages involved in the decoding by Gaussian random variables, which greatly simplifies the density evolution computation. Then, [4] considered density evolution for binary SW coding and non-symmetric channels. In [4], an equivalence between SW coding and channel coding under density evolution is derived.

Density evolution for non-binary LDPC codes and symmetric channels has been investigated in [16]. In particular, it is shown that density evolution in channel coding can be performed from a Gaussian approximation. However, except with the Gaussian approximation, no recursive expression of the density evolution is provided. The particular case of the non-binary erasure channels is described in [25]. Then, [1] considered density evolution for coset non-binary LDPC codes. In this case, the channels are not necessarily symmetric, because it is shown that the coset has a symmetrizing effect. As before, no recursive expression of the density evolution is given, except with the Gaussian approximation. SW codes can be seen as particular coset LDPC codes, but, [1] considers channel coding, with fixed input symbols distribution. To finish, [13] shows that, if the all-zero codeword assumption holds, density evolution in channel coding can be performed with MCMC methods.

III. NOTATIONS AND PRELIMINARIES

In the following, upper case letters, *e.g.*, X, denote random variables whereas lower case letters, x, represent their realizations. Vectors, *e.g.*, $\mathbf{X} = \{X_k\}_{k=1}^n$, are in bold. When it is clear from the context that the distribution of a random variable X_k does not depend on k, the index k is omitted. The imaginary unit is denoted i. The Kronecker function is denoted $\delta(x)$, *i.e.*, $\delta(x) = 1$ if $x = 0$, $\delta(x) = 0$ otherwise. In the following, $\bigotimes$ stands for the convolution product (to avoid the confusion with $\otimes$, the multiplicative operator in GF(q)) and $\circ$ is the composition operator. In SW coding (see Figure 1), the source X to be compressed and the SI Y available at the decoder produce sequences of independent and identically distributed (*i.i.d.*) discrete symbols $\{X_n\}_{n=1}^{+\infty}$ and $\{Y_n\}_{n=1}^{+\infty}$ respectively. The realizations of the random variable X belong to GF(q) with $q = \kappa^\alpha$ and κ is prime. The realizations of Y belong to a discrete alphabet $\mathcal{Y}$. Denote $P(X = x) = p_x$ where $0 < p_x < 1$ and assume $\forall (x, y)$, $0 < P(Y = y|X = x) < 1$.

A. Operations in GF(q)

In the following, $\oplus$, $\ominus$, $\otimes$, $\oslash$ are the usual operators in GF(q) (see [19, Chapter 4] for more details on Galois Fields). Denote $r = \exp(\frac{2\mathrm{i}\pi}{\kappa})$ the unit root associated to GF(q). Roughly speaking, the Galois Field GF(q) is composed by q elements that are polynomials of order $\alpha - 1$ with coefficients in $\mathbb{Z}/\kappa\mathbb{Z}$. However, in the original coding problem, the source alphabet is composed by q possible symbols that are integers between 0 and $q - 1$. Thus each possible integer $a \in \{0, \ldots, q-1\}$ has to be mapped to a polynomial of GF(q). Remarking that a can be decomposed as

$$a = a_0 + a_1\kappa + \ldots a_{\alpha-1}\kappa^{\alpha-1} \tag{1}$$

where $a_k \in \{0 \dots \kappa - 1\}$, a is by convention associated to the polynomial

$$P_a(D) = a_0 + a_1 D + \dots a_{\alpha-1} D^{\alpha-1}. \tag{2}$$

With an abuse of notation, we denote $a \in \text{GF}(q)$, where a both refer to the integer and to the polynomial. As an example, r^a is evaluated from the integer version of a, but in the expression $r^{a \oplus b}$, $a \oplus b$ is performed within the particular polynomial algebra of GF(q). Furthermore, respecting the convention, one can show that $r^{a \oplus b} = r^a r^b$.

B. Probability evaluation in GF(q)

Let Z be a random variable with values in GF(q). Denote $\mathbf{p}$ the probability vector of size q with k-th component $p_k = P(Z = k)$ and $0 < p_k < 1$. Denote $\mathbf{m}$ the message vector of size q with k-th component $m_k = \log \frac{p_0}{p_k} = \log \frac{P(Z=0)}{P(Z=k)}$. Moreover, from the previous expression, $p_k = \frac{e^{-m_k}}{\sum_{k'=0}^{q-1} e^{-m'_k}}$. As part of the LDPC decoder consists of the evaluation of the probability of linear combinations of random variables, we wish to express the probabilities of $Z \oplus a$, $Z \otimes a$, where $a \in \text{GF}(q)$, and $Z_1 \oplus Z_2$. Note that the operators we describe here to realize these evaluations were initially introduced in [1] and [16]. We restate them here for the sake of clarity.

Denote $\mathbf{p}^{\times a}$ and $\mathbf{m}^{\times a}$ ($\forall a \in \text{GF}(q) \backslash \{0\}$), $\mathbf{p}^{+a}$ and $\mathbf{m}^{+a}$ ($\forall a \in \text{GF}(q)$) the probability and message vectors of $Z \otimes a$ and $Z \oplus a$. By definition, $\forall a \neq 0$, $p_k^{\times a} = P(Z \otimes a = k) = P(Z = k \oslash a)$ and

$$m_k^{\times a} = \log \frac{P(Z \otimes a = 0)}{P(Z \otimes a = k)} = \log \frac{P(Z = 0)}{P(Z = k \oslash a)} \ . \tag{3}$$

Let $W[a]$ be a $q \times q$ matrix such that $\forall k, j = 0, \dots, q-1$, $W_{k,j}[a] = \delta(a \otimes j \ominus k)$. Then, $\mathbf{p}^{\times a} = W[a]\mathbf{p}$ and $\mathbf{m}^{\times a} = W[a]\mathbf{m}$. On the other hand, $p_k^{+a} = P(Z \oplus a = k) = P(Z = k \ominus a)$ and

$$m_k^{+a} = \log \frac{P(Z \oplus a = 0)}{P(Z \oplus a = k)} = \log \frac{P(Z = \ominus a)}{P(Z = k \ominus a)} \tag{4}$$

Denote respectively $R[a]$ and $\mathcal{A}[a]$ the $q \times q$ matrices such that $\forall k, j = 0, \ldots, q-1$, $R_{k,j}[a] = \delta(a \oplus k \ominus j)$ and $\mathcal{A}_{k,j}[a] = \delta(a \oplus k \ominus j) - \delta(a \ominus j)$. Then, $\mathbf{p}^{+a} = R[a]\mathbf{p}$ and $\mathbf{m}^{+a} = \mathcal{A}[-a]\mathbf{m}$. Here, two different transforms are needed because of the numerator in (4). The notations $\mathbf{m}^{\times a}$ and $\mathbf{m}^{+a}$ come from [1] while $W[a]$ and $\mathcal{A}[a]$ come from [16].

Now, let Z_1 and Z_2 be two random variables with realizations in GF(q) and probability vectors $\mathbf{p}_1$ and $\mathbf{p}_2$. Then,

$$P(Z_1 \oplus Z_2 = k) = \sum_{j=0}^{q-1} P(Z_1 = j) P(Z_1 \oplus Z_2 = k | Z_1 = j) = \sum_{j=0}^{q-1} p_{1,j} p_{2,k \ominus j} \tag{5}$$

$$:= (\mathbf{p}_1 \overline{\otimes} \mathbf{p}_2)_k \ . \tag{6}$$

The defined operator $\overline{\otimes}$ represents a discrete convolution product but *does not* correspond to the classical circular convolution product. Consequently, as pointed out in [14], the usual discrete Fourier Transform cannot be used for the evaluation of (5) and there is a need to define an adapted Fourier-like transform $\mathcal{F}$. Let $\mathbf{f} = \mathcal{F}(\mathbf{p})$ and $\mathbf{p} = \mathcal{F}^{-1}(\mathbf{f})$ with from [16],

$$f_j = \sum_{k=0}^{q-1} r^{k \otimes j} p_k \ , \quad p_k = \frac{1}{q} \sum_{j=0}^{q-1} r^{-k \otimes j} f_j \ . \tag{7}$$

Then

$$P(Z_1 \oplus Z_2 = k) = \left(\mathcal{F}^{-1} \left(\mathcal{F}(\mathbf{p}_1) \mathcal{F}(\mathbf{p}_2) \right) \right)_k \ . \tag{8}$$

This expression can easily be generalized to a sum of K elements. A message version of the Fourier-like transform can also be defined as $\mathbf{f} = \tilde{\mathcal{F}}(\mathbf{m})$ and $\mathbf{m} = \tilde{\mathcal{F}}^{-1}(\mathbf{f})$ with

$$f_j = \sum_{k=0}^{q-1} r^{k \otimes j} \frac{e^{-m_k}}{\sum_{k'=0}^{q-1} e^{-m'_k}} \ , \quad m_k = \log \frac{\sum_{j=0}^{q-1} f_j}{\sum_{j=0}^{q-1} r^{-k \otimes j} f_j} \ . \tag{9}$$

IV. LDPC ENCODING AND DECODING

LDPC codes initially introduced for channel coding can also be used for SW coding, after adaptation of the coding process and the decoding algorithm [17], [20]. For non-binary

channel coding version The SW coding of a source vector $\mathbf{x}$ of length n is performed by producing a vector $\mathbf{s} = H^T\mathbf{x}$ of length $m < n$. The matrix H is sparse, with non-zero coefficients in GF(q)$\backslash\{0\}$. In the bipartite graph representing the dependencies between the random variables of $\mathbf{X}$ and $\mathbf{S}$, the entries of $\mathbf{X}$ are represented by Variable Nodes (VN) and the entries of $\mathbf{S}$ are represented by Check Nodes (CN). The set of CN connected to a VN n is denoted $\mathcal{N}_{\mathrm{C}}(n)$ and the set of VN connected to a CN m is denoted $\mathcal{N}_{\mathrm{V}}(m)$. The sparsity of H is determined by the edge-perspective VN degree distribution $\lambda(x) = \sum_{k\geq 2}\lambda_k x^{k-1}$ and CN degree distribution $\rho(x) = \sum_{j\geq 2}\rho_j x^{j-1}$. The constant $0 \leq \lambda_k \leq 1$ is the proportion of edges emanating from a VN of degree k and $0 \leq \rho_j \leq 1$ is the proportion of edges emanating from a CN of degree j. In SW coding, the coding efficiency $r(\lambda,\rho)$ of a code is given by $r(\lambda,\rho) = \frac{m}{n} = \frac{\sum_{j\geq 2}\rho_j/j}{\sum_{k\geq 2}\lambda_k/k}$. A code is said to be regular if the VN and CN have constant degrees d_v and d_c. In this case, $r(d_v,d_c) = \frac{d_v}{d_c}$.

The sum-product LDPC decoder performs an approximate Maximum *A Posteriori* (MAP) estimation of $\mathbf{x}$ from the received codeword $\mathbf{s}$ and the observed side information $\mathbf{y}$ by the mean of message exchange in the bipartite graph. In non-binary channel coding, the sum-product LDPC decoder is described in [16]. We expressed the SW version of the algorithm in [12] and restate it here for the sake of completness. The initial messages for a VN n is denoted $\mathbf{m}^{(0)}(n)$, and have k-th component

$$m_k^{(0)}(n) = \log\frac{P(X_n = 0|Y_n = y_n)}{P(X_n = k|Y_n = y_n)},\ k = 0\dots q-1\ . \tag{10}$$

At iteration ℓ, the message $\mathbf{m}^{(\ell)}(m,n)$ from CN m to VN n is

$$\mathbf{m}^{(\ell)}(m,n) = \mathcal{A}[\overline{s}_m]\tilde{\mathcal{F}}^{-1}\left(\prod_{n'\in\mathcal{N}_{\mathrm{V}}(m)\backslash n}\tilde{\mathcal{F}}\left(W\left[\overline{g}_{n'm}\right]\mathbf{m}^{(\ell-1)}(n',m)\right)\right) \tag{11}$$

where the product is componentwise, $\bar{s}_m = \ominus s_m \oslash H_{n,m}$, and $\overline{g}_{n'm} = \ominus H_{n',m} \oslash H_{n,m}$. Note that $\mathcal{A}[\overline{s}_m]$ does not appear in the channel coding version of the algorithm and is specific

to SW coding. At a VN n, a message $\mathbf{m}^{(\ell)}(n,m)$ is sent to the CN m and an *a posteriori* message $\tilde{\mathbf{m}}^{(\ell)}(n)$ is computed. They both satisfy

$$\mathbf{m}^{(\ell)}(n,m) = \sum_{m' \in \mathcal{N}_{\mathrm{C}}(n) \setminus m} \mathbf{m}^{(\ell)}(m',n) + \mathbf{m}^{(0)}(n) \ , \tag{12}$$

$$\tilde{\mathbf{m}}^{(\ell)}(n) = \sum_{m' \in \mathcal{N}_{\mathrm{C}}(n)} \mathbf{m}^{(\ell)}(m',n) + \mathbf{m}^{(0)}(n) \ . \tag{13}$$

The channel version of the algorithm has the same VN message computation. From (13), each VN n produces an estimate $\widehat{x}_n^{(\ell)} = \arg\max_k \tilde{m}_k^{(\ell)}(n)$ of x_n. The algorithm ends if $H^{\mathrm{T}}\widehat{\mathbf{x}}^{(\ell)} = \mathbf{s}$ or if $\ell = L_{\max}$, the maximum number of iterations.

The CN message (11) is calculated from linear operators and a componentwise product. Since the probability density of these products may be difficult to derive, we introduce the following transform $\boldsymbol{\gamma}$. The function $\boldsymbol{\gamma}$ applies on vectors of size q and has $k-$th component $\gamma_k : \mathbb{C} \to \mathbb{R} \times [-\pi, \pi]$ with

$$\gamma_k(f_k) = (z_k, t_k) = \begin{cases} \left(\frac{1}{2}\log(x_k^2 + y_k^2), \arctan\frac{y_k}{x_k}\right) & \text{if } x_k \geq 0, y_k \neq 0 \\ \left(\frac{1}{2}\log(x_k^2 + y_k^2), \arctan\frac{y_k}{x_k} + \pi\right) & \text{if } x_k \leq 0 \ , \ y_k \geq 0 \\ \left(\frac{1}{2}\log(x_k^2 + y_k^2), \arctan\frac{y_k}{x_k} - \pi\right) & \text{if } x_k \leq 0 \ , \ y_k < 0 \ . \end{cases} \tag{14}$$

where x_j and y_j are the real part and the imaginary part of f_k. Note that γ_k can also be seen as a function from $\mathbb{R}^2$ to $\mathbb{R} \times [-\pi, \pi]$. We complete the definition of γ_k by assuming that when $f_k = 0$, the value of t_k is given by the realization of a random variable Θ taking its values in $[0, 2\pi]$ and with probability density function $f_\Theta(\theta) = \frac{1}{2\pi}$. The inverse function $\boldsymbol{\gamma}^{-1}$ applies on vectors of size q and has $j-$th component $\gamma_j^{-1} : \mathbb{R} \times [-\pi, \pi] \to \mathbb{C}$ with

$$\gamma_j^{-1}(z_j, t_j) = \exp(z_j)\cos t_j + \mathrm{i}\exp(z_j)\sin t_j \ . \tag{15}$$

The CN to VN equation (11) can then be restated as

$$\mathbf{m}^{(\ell)}(m,n) = \mathcal{A}[\overline{s}_m]\tilde{\mathcal{F}}^{-1}\left(\boldsymbol{\gamma}^{-1}\left(\sum_{n' \in \mathcal{N}(m) \setminus n} \boldsymbol{\gamma}\left(\tilde{\mathcal{F}}\left(W\,[\overline{g}_{n'm}]\,\mathbf{m}^{(\ell-1)}(n',m)\right)\right)\right)\right) \ . \tag{16}$$

Density evolution consists in the evaluation of the probability densities of the messages at each iteration. The decoding error probability can then be calculated from the probability of the messages giving false estimates $\hat{x}_k$. In this way, the probability density of $\mathbf{m}^{(\ell)}(n,m)$ in (12) is easy to evaluate from the probability densities of the $\mathbf{m}^{(\ell)}(m',n)$ (assuming the $\mathbf{m}^{(\ell)}(m',n)$ are realizations of independent random variables). On the opposite, the probability density of $\mathbf{m}^{(\ell)}(m,n)$ in (11) is difficult to derive because of the componentwise product. That is the reason why we introduced the function $\boldsymbol{\gamma}$ that transforms the product in (11) into a sum. Then, the following section evaluates the probability densities of the messages in channel coding and in SW coding.

V. Density evolution

The messages exchanged in the graph during the decoding can be seen as random variables. From the density of the initial messages (10), we want to calculate recursively the probability density of the messages at iteration ℓ, exploiting (12) and (16). For this, several symplifying assumptions can be performed. First, it is assumed that the messages arriving at a node at iteration ℓ are independent. The so-called independence assumption was originally discussed in [24] and proved formaly to be reasonable in [30]. The main idea is that the messages are independent if they have been calculated on independent subtrees of the bipartite graph. It is called the cycle-free case. In [30], it is shown that this cycle-free case happens with probability arbitrarily closed to 1 when $n \to \infty$.

The second simplifying assumption is called the all-zero codeword assumption. In channel coding, it is shown to apply only for symmetric channel. Thus, before explaining the assumption, we restate the definition of a symmetric channel and give some examples.

Definition 1. [16] *Let $P(W|U)$ be a channel with q-ary input U and output W. Denote $\mathcal{I}[a]$*

the $(q-1)\times(q-1)$ *diagonal matrix with* $\mathcal{I}[a]_{i,i}=r^{i\otimes a}$, $i=1,\ldots,(q-1)$. *The channel* $P(W|U)$ *is said to be* q-ary input symmetric-output *if the possible values of* W *can be relabeled into length* $(q-1)$ *complex-valued vectors* $\tilde{\mathbf{W}}$ *such that*

$$\forall a\in\{0\ldots(q-1)\}, P(\tilde{\mathbf{W}}=\tilde{\mathbf{w}}|U=a)=P(\tilde{\mathbf{W}}=\mathcal{I}[a]\tilde{\mathbf{w}}|U=0)\ . \tag{17}$$

Example. *In this example, both* U *and* W *take their values in GF(*q*). Assume that the channel is additive,* i.e.*, there exists a random variable* Z*, independent of* U*, such that* $W=U\oplus Z$ *and* $P(Z=k)=p_k$, $\forall k=\{0,\ldots,(q-1)\}$.

First consider $q=3$ *(*i.e.*,* q *is prime). Relabel* $(w=0)$ *into* $\tilde{\mathbf{w}}^{(0)}=[1,0]$, $(w=1)$ *into* $\tilde{\mathbf{w}}^{(1)}=\left[\exp(\frac{4i\pi}{3}),0\right]$, *and* $(w=2)$ *into* $\tilde{\mathbf{w}}^{(2)}=\left[\exp(\frac{2i\pi}{3}),0\right]$. *Then*

$$p_{k\ominus a}=P(W=k|U=a)=P(\tilde{\mathbf{W}}=\tilde{\mathbf{w}}^{(k)}|U=a)=P(\tilde{\mathbf{W}}=I[a]\tilde{\mathbf{w}}^{(k)}|U=0). \tag{18}$$

Now consider $q=3^2$. *The possible values of* W *may be relabeled into vectors of size* 8 *with only* 2 *non-zero components. For simplicity, we thus express reduced version (of length* 2*) of* $\tilde{\mathbf{W}}$ *and consider as well reduced versions of the* $\mathcal{I}[a]$. *We relabel* $(w=0)$ *into* $\tilde{\mathbf{w}}^{(0)}=[1,1]$ *and* $\forall k\neq 0$, $(w=k)$ *into* $\tilde{\mathbf{w}}^{(k)}=\mathcal{I}[-k]\tilde{\mathbf{w}}^{(0)}$. *Then*

$$p_{k\ominus a}=P(W=k|U=a)=P(\tilde{\mathbf{W}}=\mathcal{I}[-k]\tilde{\mathbf{w}}^{(0)}|U=a)=P(\tilde{\mathbf{W}}=\mathcal{I}[a\ominus k]\tilde{\mathbf{w}}^{(0)}|U=0)\ ,$$

showing the channel symmetry.

In channel coding, [16, Proposition 2] shows that for symmetric channels, the error probability of the decoding algorithm is independent of the transmitted codeword. Consequently, the recursion on the probability density is calculated assuming the all-zero codeword was transmitted. Unfortunately, this result does not apply in SW coding (except if X is uniformly distributed), even if $P(Y|X)$ is symmetric, because of the non-uniform source distribution. Nonetheless, as originally proposed by [4] for the binary case, we show that for every, even

non-symmetric, correlation channel $P(Y|X)$, there exists an equivalent symmetric channel $P(W|U)$. Equivalent means the two channels induce the same density evolution equations and have the same conditional entropy.

In the following, we first express recursions on the probability densities of the messages in channel coding for symmetric channels. Then, we express the recursion for SW coding, for any channel. At the end, remarking similarities between the two recursions, we derive the equivalence.

A. Density evolution in channel coding for symmetric channels

Here, the case of a symmetric channel is considered. As the channel is symmetric, the probability densities of the messages exchanged in the graph do not depend on the transmitted codeword [16]. Consequently, we assume that the all-zero codeword was transmitted and express the density evolution with this assumption. First, denote $P^{(\ell)}$ the probability density of the messages from VN to CN at iteration ℓ with respect to the all-zero codeword assumption. It is shown in [16] that the error probability of the sum-product LDPC decoder at iteration ℓ can be calculated as

$$p_e^{(\ell)} = 1 - \int_{\mathbf{m}\in\mathbb{R}_+^q} P^{(\ell)}(\mathbf{m}) d\mathbf{m} \tag{19}$$

where $\mathbb{R}_+^q$ is the set of length q real-valued vectors with positive components only. It thus suffices to express $P^{(\ell)}$ at each iteration to obtain the error probability. The following proposition gives a recursive expression of this quantity.

Proposition 1. *Consider a q-ary input symmetric-output channel $P(W|U)$, an LDPC code of VN and CN degree distributions $\lambda(x)$ and $\rho(x)$, and sum-product LDPC decoding. Assume that the decoding graph is cycle-free and that the all-zero codeword was transmitted. Denote*

$P^{(\ell)}$ the probability density of the messages at iteration ℓ from VN to CN, and $Q^{(\ell)}$ the probability density of the messages from CN to VN. Then, in channel coding,

$$Q^{(\ell)}(\mathbf{m}) = \Gamma_d^{-1}\left(\frac{1}{q-1}\sum_{g=1}^{q}\rho\left(\Gamma_c^g(P^{(\ell-1)})\right)\right)(\mathbf{m}) \tag{20}$$

$$P^{(\ell)}(\mathbf{m}) = P^{(0)}\bigotimes\lambda\left(Q^{(\ell)}\right)(\mathbf{m}) \tag{21}$$

where Γ_d^{-1} and Γ_c^h are density transform operators defined in Appendix A. Consequently,

$$P^{(\ell)}(\mathbf{m}) = P^{(0)}\bigotimes\lambda\left(\Gamma_d^{-1}\left(\frac{1}{q-1}\sum_{g=1}^{q}\rho\left(\Gamma_c^g\left(P^{(\ell-1)}\right)\right)\right)\right)(\mathbf{m}). \tag{22}$$

Proof: The channel version of the message computation from VN to CN is given by (12). Consequently, the recursive expression (20) is obtained directly from (12) (sum of *i.i.d.* random variables of probability distribution $P^{(\ell-1)}$ and marginalization according to the VN degree distribution). The channel version of the message computation from CN to VN is given by (16) from which $\mathcal{A}[\overline{s}_m]$ is removed. Denote $\bar{G}$ a random variables taking its values in GF(q). For any message $\mathbf{m}$, the density of $W[\bar{G}]\mathbf{m}$, where can be obtained by marginalizing according to $\bar{G}$. From the density transform operator obtained in Appendix A1, it is

$$\frac{1}{q-1}\sum_{\bar{g}=1}^{q-1}\Gamma_W^{\bar{g}}(P^{(\ell-1)})(\mathbf{m}). \tag{23}$$

Furthermore, from the density transform operators $\Gamma_{\mathbf{m}}$, $\Gamma_{\mathcal{F}}$, $\Gamma_{\boldsymbol{\gamma}}$ obtained respectively for the transform of $\mathbf{m}$ into $\mathbf{p}$ (see Appendix A2), for the Fourier Transform (Appendix A3), and for $\boldsymbol{\gamma}$ (Appendix A4), the density of $\boldsymbol{\gamma}\left(\tilde{\mathcal{F}}\left(W\left[\overline{G}\right]\mathbf{m}\right)\right)$ is given by

$$\frac{1}{q-1}\sum_{\bar{g}=1}^{q-1}\Gamma_c^{\bar{g}}(P^{(\ell-1)})(\mathbf{m}) \tag{24}$$

where $\Gamma_c^{\bar{g}} = \Gamma_{\boldsymbol{\gamma}}\Gamma_{\mathcal{F}}\Gamma_{\mathbf{m}}\Gamma_W^{\bar{g}}$ and by the linearity of the density transform operators. To finish, from the density transform operators $\Gamma_{\mathbf{p}}$, $\Gamma_{\mathcal{F}^{-1}}$, $\Gamma_{\boldsymbol{\gamma}^{-1}}$ obtained respectively for the

transformation of $\mathbf{p}$ into $\mathbf{m}$ (see Appendix A2), for the inverse Fourier Transform (see Appendix A3), and for $\boldsymbol{\gamma}^{-1}$ (see Appendix A4), we get (21) where $\Gamma_d^{-1} = \Gamma_{\mathbf{p}}\Gamma_{\boldsymbol{\gamma}^{-1}}\Gamma_{\mathcal{F}^{-1}}$. Finally combining (20) and (21) gives (22). ■

The initial $P^{(0)}$ is obtained by evaluating the probability density of (10) conditionned on the fact that $X = 0$. Note that the recursive formula (22) is not conveniant for practical density evolution (see the expressions of the operators in Appendix). The objective here is only to express a recursion in order to show that a related form is obtained in SW coding.

B. Density evolution in SW coding

As explained in introduction, the symmetry property does not hold in general in SW coding, even if the correlation channel is itself symmetric. Consequently, the all-zero codeword transmission cannot be assumed anymore. Denote respectively $P_k^{(\ell)}$ and $Q_k^{(\ell)}$ the probability densities of the messages from VN to CN and from CN to VN conditioned on the fact that $X = k$. The following proposition gives the expression of the error probability of the sum-product LDPC decoder.

Proposition 2. *Consider a q-ary input correlation channel $P(Y|X)$, LDPC encoding and sum-product LDPC decoding. Let $P_k^{(\ell)}$ be the probability density of the messages from VN to CN conditioned on the fact that $X = k$ and define*

$$\langle P^{(\ell)} \rangle(\mathbf{m}) = \sum_{k=0}^{q-1} P(X = k) P_k^{(\ell)} \circ \mathcal{A}[-k](\mathbf{m}) \tag{25}$$

In SW coding, the error probability of the LDPC decoder at iteration ℓ is given by

$$p_e^{(\ell)} = 1 - \int_{\mathbf{m} \in \mathbb{R}_+^q} \langle P^{(\ell)} \rangle(\mathbf{m}) d\mathbf{m}. \tag{26}$$

See Appendix B1 for the proof.

The proposition can be interpreted as follows. For a randomly selected edge of the bipartite graph (see Section IV), the probability of error at iteration ℓ, $p_e^{(\ell)}$, is the probability for a message transmitted on the edge to produce a false estimate of the symbol value at the variable node. For example, in the binary case, if $X = 0$ but the scalar message $m^{(\ell)} < 0$, a false estimate of X is produced. Consequently, in the non-binary case, the error probability can be obtained by marginalizing according to $k = 0, \ldots, (q-1)$ and, for each k, by integrating $P_k^{(\ell)}$ over the set of messages producing an error. For $X = k$, this corresponds to the set of messages $\mathbf{m}$ such that there exists $i \neq k$ such that $m_i < m_k$. The marginalization operation appears in (25). Moreover, the operators $\mathcal{A}[-k]$ realize the projection of the space $\mathbb{R}_-^q$ on the set of messages producing an error, thus giving (26).

The following proposition gives the recursion on $\langle P^{(\ell)} \rangle$ obtained in SW coding.

Proposition 3. *Consider a q-ary input correlation channel $P(Y|X)$, an LDPC code of VN and CN degree distributions $\lambda(x)$ and $\rho(x)$, and sum-product LDPC decoding. Assume that the decoding graph is cycle-free. Denote $P_k^{(\ell)}$ the probability density of the messages at iteration ℓ conditionned on the fact that $X = k$ and denote $\langle P^{(\ell)} \rangle(\mathbf{m}) = \sum_{k=0}^{q-1} P(X = k) P_k^{(\ell)} \circ \mathcal{A}[-k](\mathbf{m})$. In SW coding, the following recursion holds*

$$\langle P^{(\ell)} \rangle(\mathbf{m}) = \langle P^{(0)} \rangle \bigotimes \lambda \left(\Gamma_d^{-1} \left(\frac{1}{q-1} \sum_{g=1}^{q} \rho \left(\Gamma_c^g \left(\langle P^{(\ell-1)} \rangle \right) \right) \right) \right) (\mathbf{m}). \tag{27}$$

where Γ_d^{-1} and Γ_c^g are density transform operators defined in Appendix A.

See Appendix B for the proof. The initial density is given by

$$\langle P^{(0)} \rangle = \sum_{k=0}^{q-1} P(X = k) P_k^{(0)} \circ \mathcal{A}[-k](\mathbf{m}) \tag{28}$$

where the $P_k^{(0)}$ are calculated from the expression of the initial messages (10).

We see that the recursion in SW coding is exactly that obtained in channel coding, except that it now applies on $\langle P^{(\ell)} \rangle$. Consequently, the only difference is on the initial $\langle P^{(0)} \rangle$

which, as expected, takes into account the probability distribution of X. Consequently, we see that if a correlation channel $P(Y|X)$ and a channel $P(W|U)$ have initial probability densities respectively $\langle P^{(0)} \rangle$ and $P^{(0)}$ such that $\langle P^{(0)} \rangle = P^{(0)}$, then they have the same density evolution equations. The source channel equivalence is derived from this remark.

C. The source-channel equivalence

From Proposition 3, we would like to identify a channel $P(W|U)$ with initial probability density $P^{(0)}$ such that $P^{(0)} = \langle P^{(0)} \rangle$. In this section, we give the expression of the equivalent channel and show it is symmetric.

Definition 2. [16] *The probability density P of a length q random vector $\mathbf{Z}$ is said to be* symmetric *if and only if* $\forall k \in \{0, \ldots, (q-1)\}$, $P(\mathcal{A}[k]\mathbf{z}) = \exp(-z_k)P(\mathbf{z})$.

Note that the symmetry of the probability density is different from the channel symmetry defined in Definition 1. However, a symmetric channel is shown to induce symmetric probability densities $P^{(\ell)}$ in the density evolution.

Proposition 4. *Let Q be a symmetric probability density applying on random vectors of size q and with r probability mass points. There exists a q-ary input symmetric output channel $P(W|U)$ for which the initial density probability $P^{(0)}$ of the density evolution is such that $P^{(0)} = Q$.*

Proof: Let Q be a density with r probability mass points $\boldsymbol{\alpha}_0, \ldots \boldsymbol{\alpha}_{r-1}$, such that $Q(\boldsymbol{\alpha}_i) = \beta_i$. From, Definition 2, the symmetry condition gives $Q(\boldsymbol{\alpha}_i) = \exp(\alpha_{i,k})Q(\mathcal{A}[-k]\boldsymbol{\alpha}_i)$. Consequently, if $\boldsymbol{\alpha}_i$ is a mass point, $\mathcal{A}[-k]\boldsymbol{\alpha}_i$ is also a mass point, $\forall k = 0 \ldots (q-1)$. Denote $Q(\mathcal{A}[-k]\boldsymbol{\alpha}_i) = \beta_{k\times i}$. Then from the symmetry condition, $\alpha_{i,k} = \log \frac{\beta_i}{\beta_{k\times i}}$ and consequently

$$Q\left(\left[\log \frac{\beta_i}{\beta_i}, \ldots, \log \frac{\beta_i}{\beta_{k\times i}}, \ldots\right]\right) = \beta_i \ . \tag{29}$$

On the other hand, for any channel with uniformly distributed input U and discrete symmetric output W, the initial messages (10) are such that

$$\log \frac{P(U=0|W=i)}{P(U=k|W=i)} = \log \frac{P(W=i|U=0)}{P(W=i|U=k)} \tag{30}$$

and their probability density conditionned on the fact that $U=0$ is such that

$$P\left(\log \frac{P(W=i|U=0)}{P(W=i|U=0)}, \ldots, \log \frac{P(W=i|U=0)}{P(W=i|U=k)}, \cdots \mid U=0\right) = P(W=i|U=0) \ . \tag{31}$$

Then from (29) and (31), the channel equivalent to Q is such that $P(W=i|U=0)=\beta_i$. The other probabilities $P(W=i|U=k)$ can be obtained by symmetry.

■

The previous proposition makes a link between symmetric probability density and a symmetric channel. Indeed, it shows that for every symmetric distribution Q, there exists a symmetric channel such that $P^{(0)}=Q$. Thus it remains to show that the initial probability distribution $\rangle P^{(0)} \langle$ for $P(Y|X)$ is symmetric.

Proposition 5. *Consider a q-ary input correlation channel $P(Y|X)$, LDPC encoding and sum-product LDPC decoding. Let $P_k^{(0)}$ be the probability density of the initial messages from VN to CN conditioned on the fact that $X=k$ for density evolution. Then the probability density $\langle P^{(0)} \rangle(\mathbf{m}) = \sum_{k=0}^{q-1} P(X=k) P_k^{(0)} \circ \mathcal{A}[-k](\mathbf{m})$ is symmetric.*

See Appendix C for the proof.

From Propositions 4 and 5, we see that for any correlation channel $P(Y|X)$, there exists a symmetric channel $P(W|U)$ such that the initial probability density $\langle P^{(0)} \rangle$ for $P(Y|X)$ is equal to the initial probability $P^{(0)}$ for $P(W|U)$. Consequently, from Propositions 1 and 3, the two channels have exactly the same density evolution equations. We say that $P(Y|X)$ and

$P(W|U)$ are equivalent under density evolution. This result is synthetized in the following theorem.

Theorem 1. *Consider a q-ary input discrete-output correlation channel $P(Y|X)$, LDPC encoding and sum-product LDPC decoding. Denote $\langle P^{(0)} \rangle$ its initial density under density evolution. Then there exists an equivalent symmetric channel $P(W|U)$ of initial density under density evolution given by $P^{(0)} = \langle P^{(0)} \rangle$. Furthermore, $P(Y|X)$ and $P(W|U)$ have the same density evolution equations and $H(X|Y) = H(U|W)$.*

Proof: The first part of the theorem come from Propositions 4 and 5. The second part of the theorem (*i.e.* the entropy equality) is given in Appendix D. ■

The result of Theorem 1 can be linked to the results of [1]. Indeed, in [1], it is shown that for non-binary *coset* LDPC codes, density evolution can be performed with the all-zero codeword assumption, whatever the channel. In fact, the coset is shown to have some symmetrizing effect and as a consequence, the all-zero codeword assumption holds even if the channel was not originally symmetric. A SW LDPC code can be considered as a particular coset LDPC codes. Note however that the results of [1] cannot be applied directly here because of the source distribution in SW coding. But the results of Theorem 1 exhibit the same symmetrizing effect.

VI. Example

Consider a source X taking its values in GF(q) and such that $P(X = k) = p_k$. The correlation channel between X and Y is described by a q-ary symmetric channel in GF(q)

with

$$\begin{aligned} \forall y \neq k,\ P(Y=y|X=k) =& \frac{p}{q-1} \\ P(Y=k|X=k) =& 1-p \end{aligned} \tag{32}$$

where $0 < p < 1$. From (10), the k-th component of an initial message $m_k^{(0)}(y)$ is given by

$$\begin{aligned} \forall y,\ m_0^{(0)}(y) =& 0 \\ \forall k \neq 0,\ m_k^{(0)}(0) =& \log \frac{p_0}{p_k} + \log \frac{(1-p)(q-1))}{p} \\ \forall k \neq 0,\ m_k^{(0)}(k) =& \log \frac{p_0}{p_k} + \log \frac{p}{(1-p)(q-1)} \\ \forall k \neq 0, \forall y \neq x, y \neq 0,\ m_k^{(0)}(y) =& \log \frac{p_0}{p_k}. \end{aligned} \tag{33}$$

The support of the probability densities $P_k^{(0)}$ is composed by the message vectors obtained from (33) and the support of $\langle P^{(0)} \rangle$ is given by the vectors $\mathcal{A}[k]\mathbf{m}^{(0)}(y)$. For $y = k$, one has

$$\forall i = 1 \ldots (q-1),\ (\mathcal{A}[k]\mathbf{m}^{(0)}(k))_i = \log \frac{p_k}{p_{k \oplus i}} + \log \frac{(1-p)(q-1)}{p} \tag{34}$$

and

$$\langle P^{(0)} \rangle (\mathcal{A}[k]\mathbf{m}^{(0)}(k)) = p_k(1-p) \tag{35}$$

For $y \neq k$,

$$\forall i \neq (y \ominus k),\ (\mathcal{A}[k]\mathbf{m}^{(0)}(y))_i = \log \frac{p_k}{p_{k \oplus i}} \tag{36}$$

$$i = y \ominus k,\ (\mathcal{A}[k]\mathbf{m}^{(0)}(y))_i = \log \frac{p_k}{p_{k \oplus i}} + \log \frac{p}{(q-1)(1-p)} \tag{37}$$

and

$$\langle P^{(0)} \rangle (\mathcal{A}[k]\mathbf{m}^{(0)}(y)) = p_k \frac{p}{q-1}. \tag{38}$$

From Section V-C, the channel $P(W|U)$ equivalent to $P(Y|X)$ has q inputs and q^2 outputs. We describe the channel only for the input $U = 0$ (the only useful for density evolution wit

the all-zero codeword assumption). The others can be obtained by symmetry. Each possible $k \in \text{GF}(q)$ gives 1 output such that

$$P(W = w|U = 0) = p_k(1 - p)$$

and $q - 1$ outputs such that

$$P(W = w|U = 0) = p_k \frac{p}{q-1}.$$

Now, we wish to design good LDPC codes for the equivalent channel. Here, the distribution of X is fixed and the threshold of a code is only on p. Here, we consider two codes of rate $1/2$. The first code is regular with degrees $d_v = 2$, $d_c = 4$. The second is irregular with node-perspective degree distributions $\lambda(x) = x^2$, $\rho(x) = 0.0110735x^2 + 0.1073487x^3 + 0.2583159x^4 + 0.4296047x^5 + 0.0115064x^6 + 0.1592504x^7 + 0.0166657x^8 + 0.0046979x^{25} + 0.0015367x^{26}$. For each case of interest, we give the corresponding Galois Field, the distribution of X, and $\overline{p}$, the approximate maximum parameter that can be coded with a code of rate $1/2$ (*i.e.* for which $H(X|Y) \leq 1/2$). Note that in this preliminary version of the paper, we only provide the approximate thresholds for some codes, but do not perform code optimization. The approximate thresholds are obtained from the MCMC method described in [13], applied on vectors of length 100000 and with target error probability 10^{-5}. We see that the approximate thresholds and the associated entropy values differ with the situations.

a) GF(4*),* X *distributed uniformly,* $\overline{p} = 0.189$*:* For the regular code, the approximate threshold is given by $p = 0.071$ ($H(p) = 0.24$ bit/symbol). For the irregular code, the approximate threshold is given by $p = 0.159$ ($H(p) = 0.44$ bit/symbol).

b) GF(4*),* X *with distribution* $[0.5, 0.25, 0.125, 0.125]$*,* $\overline{p} = 0.225$ *:* For the regular code, the approximate threshold is given by $p = 0.083$ ($H(p) = 0.25$ bit/symbol). For the irregular code, the approximate threshold is given by $p = 0.192$ ($H(p) = 0.45$ bit/symbol).

*c) GF(*16*), X distributed uniformly, $\overline{p} = 0.289$:* For the regular code, the approximate threshold is given by $p = 0.149$ ($H(p) = 0.30$ bit/symbol). For the irregular code, the approximate threshold is given by $p = 0.248$ ($H(p) = 0.44$ bit/symbol).

*d) GF(*16*), X with distribution* $[0.4, 0.04, \ldots, 0.04]$*, $\overline{p} = 0.367$:* For the regular code, the approximate threshold is given by $p = 0.198$ ($H(p) = 0.32$ bit/symbol). For the irregular code, the approximate threshold is given by $p = 0.318$ ($H(p) = 0.45$ bit/symbol).

VII. Conclusion

In this paper, we derived an equivalence between SW coding and channel coding under density evolution. From this equivalence, density evolution can be performed for any, even non-symmetric, correlation channel, assuming that the all-zero codeword was transmitted. Future works will be dedicated to the case where the correlation channel $P(Y|X)$ is not perfectly known.

Appendix

A. Recursion for channel coding

We look for recursive expressions of $Q^{(\ell)}$ from $P^{(\ell)}$ from (12) and (16). For this, we express the probability density transformations of the operators involved in (16).

1) $\mathcal{W}[g]$ and $R[s]$: In the following, $g \in \text{GF}(q)\backslash\{0\}$ and $s \in \text{GF}(q)$. Let $\mathbf{m}$ be a real-valued vector of size q and $\mathbf{l} = W[g]\mathbf{m}$. Denote $P_{\mathbf{M}}$ and $P_{\mathbf{L}}$ their respective probability densities and define $\varphi(\mathbf{l}) = W[g^{-1}]\mathbf{l}$. The function φ is invertible, and both φ and its inverse φ^{-1} are $\mathcal{C}^1$. The Jacobian matrix of φ is $J_\varphi = W[g^{-1}]$ and $\det(J_\varphi) \neq 0$. Consequently, φ is a $\mathcal{C}^1$-diffeomorphism. By expressing $E[f(\mathbf{L})]$ for any $\mathcal{L}^1$ function f and by variable change we get

$$P_{\mathbf{L}}(\mathbf{l}) = \det(J_\varphi) P_{\mathbf{M}}(W[g^{-1}]\mathbf{l}) = \Gamma_W^g(P_{\mathbf{M}})(\mathbf{l}) \tag{39}$$

where Γ_W^g is the density transform operator.

From the same arguments, a density transform operator Γ_R^s can be obtained for $R[s]$.

2) From LLR to probability representation: Define $\mathcal{P}$ as the set of vectors of q components such that $\forall k = 0 \dots q-1$, $0 < p_k < 1$ and $\sum_{k=0}^{q-1} p_k = 1$. Let $\mathbf{m} \in \{0\} \times \mathbb{R}^{q-1}$ and $\mathbf{p} \in \mathcal{P}$ be vectors of size q. The probability densities of $\mathbf{m}$ and $\mathbf{p}$ are denoted respectively $P_{\mathbf{M}}$ and $P_{\mathbf{P}}$. Define the function $\varphi : \{0\} \times \mathbb{R}^{q-1} \to \mathcal{P}$ with $\varphi(\mathbf{m}) = (\varphi_0(\mathbf{m}), \dots, \varphi_{q-1}(\mathbf{m}))$ and $\forall k = 0 \dots q-1$,

$$\varphi_k(\mathbf{m}) = \frac{\exp(-m_k)}{\sum_{k'=0}^{q-1} \exp(-m'_k)} \ . \tag{40}$$

The function φ is invertible with inverse $\varphi^{-1} : \mathcal{P} \to \{0\} \times \mathbb{R}^{q-1}$ with $\varphi^{-1}(\mathbf{p}) = (\phi_0(\mathbf{p}), \dots, \phi_{q-1}(\mathbf{p}))$ and $\forall j = 0 \dots q-1$,

$$\phi_j(\mathbf{p}) = \log \frac{1 - \sum_{j'=1}^{q-1} p'_j}{p_j} \ . \tag{41}$$

Both φ and φ^{-1} are $\mathcal{C}^1$. The Jacobian matrix J_φ of φ is given by

$$\begin{aligned} (J_\varphi(\mathbf{m}))_{k,k} &= -\exp(-m_k) \left(\sum_{k'=0, k' \neq k}^{q-1} \exp(-m'_k) \right) / \left(\sum_{k=0}^{q-1} \exp(-m'_k) \right)^2 \\ j \neq k : (J_\varphi(\mathbf{m}))_{j,k} &= \exp(-m_k) \exp(-m_j) / \left(\sum_{k=0}^{q-1} \exp(-m'_k) \right)^2 \end{aligned} \tag{42}$$

and $\det(J_\varphi(\mathbf{m})) \neq 0$. Consequently φ is a $\mathcal{C}^1$-diffeomorphism and by variable change in $E[f(\mathbf{M})]$ for every $\mathcal{L}^1$ function f,

$$P_{\mathbf{M}}(\mathbf{m}) = \det(J_\varphi(\mathbf{m})) P_{\mathbf{P}}(\varphi_1(\mathbf{m}) \dots \varphi_{q-1}(\mathbf{m})) = \Gamma_{\mathbf{m}}(P_{\mathbf{P}})(\mathbf{m}) \tag{43}$$

where $\Gamma_{\mathbf{m}}$ is the density transform operator. On the other hand, the Jacobian matrix $J_{\varphi^{-1}}$ of φ^{-1} is given by

$$\forall j \neq 0 : (J_{\varphi}^{-1}(\mathbf{p}))_{j,j} = -\frac{1}{p_j} - \frac{1}{\sum_{j'=1}^{q-1} p'_j} \tag{44}$$

$$\forall j \neq 0 : (J_{\varphi}^{-1}(\mathbf{p}))_{j,0} = 0 \tag{45}$$

$$\forall j \neq 0 : (J_{\varphi}^{-1}(\mathbf{p}))_{0,j} = -\frac{1}{\sum_{j'=1}^{q-1} p'_j} \tag{46}$$

$$\forall j,k \neq 0 : (J_{\varphi}^{-1}(\mathbf{p}))_{j,k} = -\frac{1}{\sum_{j'=1}^{q-1} p'_j} \tag{47}$$

$$(J_{\varphi}^{-1}(\mathbf{p}))_{0,0} = -\frac{1}{\sum_{j'=1}^{q-1} p'_j} \tag{48}$$

$$\tag{49}$$

Thus $\det(J_{\varphi}^{-1}(\mathbf{p})) \neq 0$ and from the same arguments as before, a density transform operator $\Gamma_{\mathbf{p}}$ can be obtained for the transformation of $\mathbf{m}$ into $\mathbf{p}$.

3) Fourier Transform and inverse Fourier Transform: We consider the Fourier Transform $\mathbf{f} = \mathcal{F}(\mathbf{p})$ of a vector $\mathbf{p}$. As $\mathcal{F}$ is an invertible linear application, by variable change and from the arguments of Appendix A1, we show that

$$P_{\mathbf{F}}(\mathbf{f}) = \det(J_{\mathcal{F}^{-1}}) P_{\mathbf{P}}(\mathcal{F}^{-1}(\mathbf{f})) = \Gamma_{\mathcal{F}}(P_{\mathbf{P}})(\mathbf{f}) \tag{50}$$

where $J_{\mathcal{F}^{-1}}$ is the Jacobian of $\mathcal{F}^{-1}$ and $\Gamma_{\mathcal{F}}$ is the defined density transform operator. A density transform operator $\Gamma_{\mathcal{F}^{-1}}$ can also be obtained from the inverse Fourier transform $\mathbf{p} = \mathcal{F}^{-1}(\mathbf{f})$.

4) γ transform: Define the restricted equivalent function $\tilde{\gamma} : \mathbb{R}^2 \backslash \{0,0\} \to \mathbb{R} \times [-\pi, \pi]$ and

$$\tilde{\gamma}(x,y) = \begin{cases} \left(\frac{1}{2}\log(x^2+y^2), \arctan\frac{y}{x}\right) & \text{if } x \geq 0 \\ \left(\frac{1}{2}\log(x^2+y^2), \arctan\frac{y}{x} + \pi\right) & \text{if } x < 0 \text{ , } y \geq 0 \\ \left(\frac{1}{2}\log(x^2+y^2), \arctan\frac{y}{x} - \pi\right) & \text{if } x < 0 \text{ , } y < 0 \text{ .} \end{cases} \tag{51}$$

We show that $\tilde{\gamma}$ is $\mathcal{C}^1$ over its interval definition even in the particular points $(x,0)\ \forall x \neq 0$ and $(0,y)\ \forall y \neq 0$. Its inverse application is $\tilde{\gamma}^{-1} : \mathbb{R} \times [-\pi, \pi] \to \mathbb{R}^2 \backslash \{0,0\}$ and $\gamma^{-1}(z,t) = (\exp(z)\cos t, \exp(z)\sin t)$. The determinants of the Jacobian matrices $J_{\tilde{\gamma}}$ of $\tilde{\gamma}$ and $J_{\tilde{\gamma}^{-1}}$ of $\tilde{\gamma}^{-1}$ are given by

$$\det(J_{\tilde{\gamma}}(x,y)) = \frac{1}{x^2+y^2} > 0 \quad , \quad \det(J_{\tilde{\gamma}^{-1}}(z,t)) = \exp(2z) > 0\ . \tag{52}$$

Consequently, $\tilde{\gamma}$ and $\tilde{\gamma}^{-1}$ are $\mathcal{C}^1-$diffeomorphisms. Denote $P_{X,Y}$ and $P_{Z,T}$ the probability densities associated to random variables (X,Y) and (Z,T). By expressing $E[f(X,Y)]$ and $E[f(Z,T)]$ for every $\mathcal{L}^1$ function f and by variable change, we show that density transform operators can be obtained $\forall (x,y) \in \mathbb{R}^2 \backslash \{0,0\}$ and $\forall (z,t) \in \mathbb{R} \times [-\pi, \pi]$ as

$$\tilde{P}_{X,Y}(x,y) = \Gamma_\gamma(P_{Z,T})(x,y) = \frac{1}{x^2+y^2} P_{Z,T} \circ \tilde{\gamma}(x,y) \tag{53}$$

$$\tilde{P}_{Z,T}(z,t) = \Gamma_{\gamma^{-1}}(P_{X,Y})(z,t) = \exp(z) P_{X,Y} \circ \tilde{\gamma}^{-1}(z,t)\ . \tag{54}$$

The density cannot be obtained in $(0,0)$ by the same method because γ is not continuous in $(0,0)$. However, the probability density functions have to be completed. Using the improper notation $P_{Z,T}(-\infty, t)$, we get

$$\tilde{P}_{Z,T}(-\infty, t) = \frac{1}{2\pi} P_{X,Y}(0,0) \tag{55}$$

$$\tilde{P}_{X,Y}(0,0) = P_{Z,T}(-\infty, t) = P_Z(-\infty) \tag{56}$$

where $P_{Z,T}(-\infty, t)$ does not depend on t and P_Z is the marginal density of the random variable Z.

Note that in (14), a transformation $\boldsymbol{\gamma}$ applying on vectors of size $q-1$ is defined. Its components γ_j, $j = 1 \ldots q-1$ apply independently on the components of the input vector (not necessarily composed by independent random variables). Consequently, the transforms defined in (53) can be directly generalized to the vector version.

B. Recursion for Slepian-Wolf coding

1) Expression of the error probability: The error probability $p_e^{(\ell)}$ can be developed as

$$p_e^{(\ell)} = 1 - \sum_{k=0}^{q-1} P(X=k) \int_{\mathbf{m}\in\Omega_k} P_k^{(\ell)}(\mathbf{m}) d\mathbf{m} \tag{57}$$

where $\Omega_k = \{\mathbf{m} \in \mathbb{R}^q : \forall k' \neq k : m_{k'} > m_k\}$ is the set of messages giving the right value of X. The function $\tilde{\mathbf{m}} \to \mathcal{A}[-k]\tilde{\mathbf{m}}$ is invertible, $\mathcal{C}_1$, and its inverse is also $\mathcal{C}_1$. The Jacobian of the application is $\mathcal{A}[-k]$ and $\det(\mathcal{A}[-k]) \neq 0$. Thus the application is a $\mathcal{C}_1$-diffeomorphism. By variable change,

$$p_e^{(\ell)} = 1 - \sum_{k=0}^{q-1} P(X=k) \int_{\tilde{\mathbf{m}}\in\mathbb{R}_+^q} P_k^{(\ell)}(\mathcal{A}[-k]\tilde{\mathbf{m}}) d\tilde{\mathbf{m}} \tag{58}$$

To finish (and by replacing $\tilde{\mathbf{m}}$ by $\mathbf{m}$),

$$p_e^{(\ell)} = 1 - \int_{\mathbf{m}\in\mathbb{R}_+^q} \langle P^{(\ell)} \rangle(\mathbf{m}) d\mathbf{m}. \tag{59}$$

2) Multinomial formula: The multinomial formula is restated here because it will be useful for the proof of the recursion. Let $(x_1 \dots x_m)$ be m scalar values. The multinomial formula gives

$$\left(\sum_{k=1}^{m} x_k\right)^n = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1, \dots, k_m} \prod_{i=1}^{m} x_i^{k_i} \tag{60}$$

where $\binom{n}{k_1,\dots,k_m} = \frac{n!}{k_1!\dots k_m!}$ is the multinomial coefficient. On the other hand, denote $\mathcal{X} = \{x_1, \dots, x_m\}$. One can show that the multinomial formula (60) gives also

$$\left(\sum_{k=1}^{m} x_k\right)^n = \sum_{(x'_1\dots x'_m)\in\mathcal{X}^n} \prod_{i=1}^{m} x'_i \ . \tag{61}$$

3) Recursion: For the sake of simplicity, the code is assumed regular with degrees d_v and d_c. The irregular version of the recursion is directly obtained by marginalizing according to the degree distributions.

The expression of the density $P_x^{(\ell)}$ is directly obtained from (12) (sum of random variables) as

$$P_x^{(\ell)}(\mathbf{m}) = P_x^{(0)} \bigotimes (Q_x^{(\ell-1)})^{\otimes(d_v-1)}(\mathbf{m}). \tag{62}$$

On the other hand, $Q_x^{(\ell)}(\mathbf{m})$ can be developed as

$$Q_x^{(\ell)}(\mathbf{m}) = \sum_{\bar{g}_1 \dots \bar{g}_{d_c-1}} \sum_{x_1 \dots x_{d_c-1}} \left(\prod_{i=1}^{d_c-1} \frac{p_{x_i}}{q-1} \right) P(\mathbf{m}|x, x_1 \dots x_{d_c-1}, \bar{g}_1 \dots \bar{g}_{d_c-1}) \tag{63}$$

$$P(\mathbf{m}|x, x_1 \dots x_{d_c-1}, \bar{g}_1 \dots \bar{g}_{d_c-1}) = \Gamma_d^{-1} \left(\bigotimes_{i=1}^{d_c-1} \Gamma_c^{\bar{g}_i}(P_{x_i}^{(\ell-1)}) \right) \circ \mathcal{A}[-\bar{s}](\mathbf{m}) \tag{64}$$

where $\bar{s} = x + \sum_{i=1}^{d_c-1} \bar{g}_i x_i$ and (64) is obtained from (21) completed with $\mathcal{A}$ and from the multinomial formula. Furthermore, $\mathcal{A}[c \oplus b]\mathbf{m} = \mathcal{A}[c]\mathcal{A}[b]\mathbf{m}$ and from (63),

$$Q_a^{(\ell)}(\mathbf{m}) = Q_b^{(\ell)} \circ \mathcal{A}[a \ominus b](\mathbf{m}) \tag{65}$$

Moreover,

$$Q_0^{(\ell)}(\mathbf{m}) = \sum_{\bar{g}_1 \dots \bar{g}_{d_c-1}} \sum_{x_1 \dots x_{d_c-1}} \left(\prod_{i=1}^{d_c-1} \frac{p_{x_i}}{q-1} \right) \Gamma_d^{-1} \left(\bigotimes_{i=1}^{d_c-1} \Gamma_c^{\bar{g}_i} \left(P_{x_i}^{(\ell-1)} \circ \mathcal{A}[-x_i] \right) \right) (\mathbf{m}) \tag{66}$$

$$= \Gamma_d^{-1} \left(\left(\sum_{\bar{g}=1}^{q-1} \sum_{x=0}^{q-1} \frac{p_x}{q-1} \Gamma_c^{\bar{g}} \left(P_x^{(\ell-1)} \circ \mathcal{A}[-x] \right) \right)^{\otimes(d_c-1)} \right) (\mathbf{m}) \tag{67}$$

by the multinomial formula. Finally, by linearity of the density transform operators

$$Q_0^{(\ell)}(\mathbf{m}) = \Gamma_d^{-1} \left(\left(\frac{1}{q-1} \sum_{\bar{g}=1}^{q-1} \Gamma_c^{\bar{g}} \left(\sum_{x=0}^{q-1} p_x P_x^{(\ell-1)} \circ \mathcal{A}[-x] \right) \right)^{\otimes(d_c-1)} \right) (\mathbf{m}) \tag{68}$$

$$= \Gamma_d^{-1} \left(\left(\frac{1}{q-1} \sum_{\bar{g}=1}^{q-1} \Gamma_c^{\bar{g}} \left(\langle P^{(\ell-1)} \rangle \right) \right)^{\otimes(d_c-1)} \right) (\mathbf{m}). \tag{69}$$

Then from (62)

$$\langle P^{(\ell)} \rangle(\mathbf{m}) = \sum_{x=0}^{q-1} p_x \left(P_x^{(0)} \bigotimes \left(Q_x^{(\ell-1)} \right)^{\otimes(d_v-1)} \right) \circ \mathcal{A}[-x](\mathbf{m}) \tag{70}$$

$$= \sum_{x=0}^{q-1} p_x \left(P_x^{(0)} \circ \mathcal{A}[-x] \right) \bigotimes \left(Q_x^{(\ell-1)} \circ \mathcal{A}[-x] \right)^{\otimes(d_v-1)} (\mathbf{m}) \tag{71}$$

by property of the convolution product. Furthermore, from (65),

$$\langle P^{(\ell)}\rangle = \langle P^{(0)}\rangle \bigotimes \left(Q_0^{(\ell-1)}\right)^{\otimes(d_v-1)}(\mathbf{m}). \tag{72}$$

To finish, replacing $Q_0^{(\ell-1)}$ from (69) gives (27).

C. Symmetry of $\langle P^{(0)}\rangle$

For any function f for which the integral exists,

$$\int_{\mathbb{R}^q} f(\mathbf{m})\langle P^{(0)}\rangle(\mathbf{m})d\mathbf{m} = \sum_{k=0}^{q-1} P(X=k)\int_{\mathbb{R}^q} f(\mathbf{m})P_k^{(0)}(\mathcal{A}[-k]\mathbf{m})d\mathbf{m} \tag{73}$$

By the change of variable $\mathbf{l} = \mathcal{A}[-k]\mathbf{m}$,

$$\begin{aligned}
\int_{\mathbb{R}^q} f(\mathbf{m})\langle P^{(0)}\rangle(\mathbf{m})d\mathbf{m} &= \sum_{k=0}^{q-1} P(X=k)E_{\mathbf{L}|X=k}\left[f(\mathcal{A}[k]\mathbf{L})\right] \qquad (74)\\
&= \sum_{k=0}^{q-1} P(X=k)E_{Y|X=k}\left[f(\mathcal{A}[k]\mathbf{l}(Y))\right]\\
&= \sum_{k=0}^{q-1} E_{P(Y|X=i\oplus k)}\left[P(X=k)\frac{P(Y|X=k)}{P(Y|X=i\oplus k)}f(\mathcal{A}[k]\mathbf{l}(Y))\right]\\
&= \sum_{k=0}^{q-1} E_{P(Y|X=i\oplus k)}\left[P(X=i\oplus k)\frac{P(X=k|Y)}{P(X=i\oplus k|Y)}f(\mathcal{A}[k]\mathbf{l}(Y))\right]\\
&= \sum_{k=0}^{q-1} E_{P_{k\oplus i}^{(0)}}\left[P(X=i\oplus k)e^{l_{i\oplus k}-l_k}f(\mathcal{A}[k]\mathbf{L})\right]\\
&= \sum_{k=0}^{q-1} P(X=i\oplus k)\int e^{l_{i\oplus k}-l_k}f(\mathcal{A}[k]\mathbf{l})P_{k\oplus i}^{(0)}(\mathbf{l})d\mathbf{l}
\end{aligned}$$

By the change of variable $\mathbf{m} = \mathcal{A}[i\oplus k]\mathbf{l}$,

$$l_{i\oplus k} = m_0 - m_{\ominus(i\oplus k)} \tag{75}$$

$$l_k = m_{\ominus i} - m_{\ominus(i\oplus k)} \tag{76}$$

and then $l_{i\oplus k} - l_k = m_{\ominus i}$. Consequently

$$\begin{aligned}\int_{\mathbb{R}^q} f(\mathbf{m})\langle P^{(0)}\rangle(\mathbf{m})d\mathbf{m} &= \sum_{k=0}^{q-1} P(X=i\oplus k)\int_{\mathbb{R}^q} e^{m_{\ominus i}} f(\mathcal{A}[-i]\mathbf{m})P^{(0)}_{k\oplus i}(\mathbf{m})d\mathbf{m} \\ &= \int_{\mathbb{R}^q} e^{m_{\ominus i}} f(\mathcal{A}[-i]\mathbf{m})\langle P^{(0)}\rangle(\mathbf{m})d\mathbf{m} \end{aligned} \tag{77}$$

Thus by definition, the density $\langle P^{(0)}\rangle$ is symmetric.

D. Proof of Theorem 1

First, from (4) and (40),

$$P(X=j|Y=y) = \frac{\exp(-m_j(y))}{\sum_{j'=0}^{q-1}\exp(-m_{j'}(y))} = \frac{\exp(-m^{+k}_{j\oplus k}(y))}{\sum_{j'=0}^{q-1}\exp(-m^{+k}_{j'\oplus k}(y))},\ \forall k = 0\dots q-1 \tag{78}$$

in which we denote $m_j(y) = \frac{\log P(X=0|Y=y)}{P(X=j|Y=y)}$. Then, the conditional entropy $H(X|Y)$ can be developed as

$$H(X|Y) = \sum_{k=0}^{q-1} P(X=k)\sum_{y\in\mathcal{Y}} P(Y=y|X=k)H(X|Y=y)\ . \tag{79}$$

Denote

$$H(X|Y=y) = h\left(P(X=0|Y=y),\dots,P(X=q-1|Y=y)\right)\ . \tag{80}$$

Indeed, $H(X|Y=y)$ can be seen as a function of the probabilities $P(X=k|Y=y)$. Then, from (78), developping $\sum_{y\in\mathcal{Y}} P(Y=y|X=k)H(X|Y=y)$ yields

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}} P(Y=y|X=k)H(X|Y=y) \\
&\quad= \sum_{y\in\mathcal{Y}} P(Y=y|X=k) h\left(\frac{\exp(-m_0^{+k}(y))}{\sum_{j'=0}^{q-1}\exp(-m_{j'\oplus k}^{+k}(y))},\ldots,\frac{\exp(-m_{(q-1)\oplus k}^{+k}(y))}{\sum_{j'=0}^{q-1}\exp(-m_{j'\oplus k}^{+k}(y))}\right) \\
&\quad= E_{P(Y|X=k)}\left[h\left(\frac{\exp(-m_0^{+k}(Y))}{\sum_{j'=0}^{q-1}\exp(-m_{j'\oplus k}^{+k}(Y))},\ldots,\frac{\exp(-m_{(q-1)\oplus k}^{+k}(Y))}{\sum_{j'=0}^{q-1}\exp(-m_{j'\oplus k}^{+k}(Y))}\right)\right] \\
&\quad= E_{P_k^{(0)}}\left[h\left(\frac{\exp(-M_0^{+k})}{\sum_{j'=0}^{q-1}\exp(-M_{j'\oplus k}^{+k})},\ldots,\frac{\exp(-M_{(q-1)\oplus k}^{+k})}{\sum_{j'=0}^{q-1}\exp(-M_{j'\oplus k}^{+k})}\right)\right] \\
&\quad= E_{P_k^{(0)}\circ\mathcal{A}[-k]}\left[h\left(\frac{\exp(-M_0)}{\sum_{j'=0}^{q-1}\exp(-M_{j'})},\ldots,\frac{\exp(-M_{q-1})}{\sum_{j'=0}^{q-1}\exp(-M_{j'})}\right)\right] \qquad (81)
\end{aligned}$$

To finish, from (79) and (81),

$$H(X|Y) = E_{\langle P^{(0)}\rangle}\left[h\left(\frac{\exp(-M_0)}{\sum_{j'=0}^{q-1}\exp(-M_{j'})},\ldots,\frac{\exp(-M_{q-1})}{\sum_{j'=0}^{q-1}\exp(-M_{j'})}\right)\right]. \qquad (82)$$

Taking back the previous equations, we show that

$$\begin{aligned}
H(U|W) &= E_{P^{(0)}}\left[h\left(\frac{\exp(-M_0)}{\sum_{j'=0}^{q-1}\exp(-M_{j'})},\ldots,\frac{\exp(-M_{q-1})}{\sum_{j'=0}^{q-1}\exp(-M_{j'})}\right)\right] \\
&= E_{\langle P^{(0)}\rangle}\left[h\left(\frac{\exp(-M_0)}{\sum_{j'=0}^{q-1}\exp(-M_{j'})},\ldots,\frac{\exp(-M_{q-1})}{\sum_{j'=0}^{q-1}\exp(-M_{j'})}\right)\right] \qquad (83)
\end{aligned}$$

and finally $H(X|Y) = H(U|W)$.

REFERENCES

[1] A. Bennatan and D. Burshtein. Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels. *IEEE Transactions on Information Theory*, 52(2):549–583, 2006.

[2] R.K. Bhattar, K.R. Ramakrishnan, and K.S. Dasgupta. Density Evolution Technique for LDPC Codes in Slepian-Wolf Coding of Nonuniform Sources. *International Journal of Computer Applications IJCA*, 7(8):1–7, 2010.

[3] J. Chen and M. Fossorier. Density evolution for BP-based decoding algorithms of LDPC codes and their quantized versions. In *Global Telecommunications Conference, GLOBECOM*, volume 2, pages 1378–1382. IEEE, 2002.

[4] J. Chen, D.K. He, and A. Jagmohan. The equivalence between Slepian-Wolf coding and channel coding under density evolution. *IEEE Transactions on Communications*, 57(9):2534–2540, 2009.

[5] J. Chou, S. Pradhan, and K. Ramchandran. Turbo and trellis-based constructions for source coding with side information. In *Proc. Data Compression Conference*, pages 33–42. IEEE, 2003.

[6] S.Y. Chung, T.J. Richardson, and R.L. Urbanke. Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation. *IEEE Transactions on Information Theory*, 47(2):657–670, 2001.

[7] T.P. Coleman, A.H. Lee, M. Medard, and M. Effros. On some new approaches to practical slepian-wolf compression inspired by channel coding. In *Data Compression Conference*, pages 282–291. IEEE, 2004.

[8] T.P. Coleman, M. Medard, and M. Effros. Towards practical minimum-entropy universal decoding. In *Data Compression Conference*, pages 33–42. IEEE, 2005.

[9] T.M. Cover and J.A. Thomas. *Elements of information theory, second Edition*. Wiley, 2006.

[10] L. Cui, S. Wang, S. Cheng, and M. Yeary. Adaptive binary Slepian-Wolf decoding using particle based belief propagation. *IEEE Transactions on Communications*, 59(9):2337–2342, 2011.

[11] M.C. Davey and D.J.C. MacKay. Low Density Parity Check codes over GF (q). In *Information Theory Workshop*, pages 70–71. IEEE, 1998.

[12] E. Dupraz, A. Roumy, and M. Kieffer. Practical coding scheme for universal source coding with side information at the decoder. *Proceedings of the Data Compression Conference*, pages 401–410, 2013.

[13] M. Gorgoglione, V. Savin, and D. Declercq. Optimized puncturing distributions for irregular non-binary LDPC codes. In *Proc. International Symposium on Information Theory and its Applications (ISITA)*, pages 400–405. IEEE, 2010.

[14] A. Goupil, M. Colas, G. Gelle, and D. Declercq. FFT-based BP decoding of general LDPC codes over Abelian groups. *IEEE Transactions on Communications*, 55(4):644–649, 2007.

[15] G. Lechner and C. Weidmann. Optimization of binary LDPC codes for the q-ary symmetric channel with moderate q. In *Proceedings of the International Symposium on Turbo Codes and Related Topics*, pages 221–224, 2008.

[16] G. Li, I.J. Fair, and W.A. Krzymien. Density evolution for nonbinary LDPC codes under Gaussian approximation. *IEEE Transactions on Information Theory*, 55(3):997–1015, 2009.

[17] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6:440–442, 2002.

[18] A.D. Liveris, Z. Xiong, and C.N. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6(10):440–442, 2002.

[19] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-correcting Codes*, volume 16. Elsevier, 1977.

[20] T. Matsuta, T. Uyematsu, and R. Matsumoto. Universal Slepian-Wolf source codes using Low-Density Parity-Check matrices. In *IEEE International Symposium on Information Theory, Proceedings.*, pages 186–190, june 2010.

[21] C. Poulliat, M. Fossorier, and D. Declercq. Design of regular (2, d/sub c/)-LDPC codes over GF (q) using their binary images. *IEEE Transactions on Communications*, 56(10):1626–1635, 2008.

[22] R. Puri and K. Ramchandran. PRISM: A new robust video coding architecture based on distributed compression principles. In *Annual Allerton Conference on Communications Control and Computing, Proceedings.*, volume 40, pages 586–595, 2002.

[23] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke. Design of capacity-approaching irregular Low-Density Parity-Check codes. *IEEE Transactions on Information Theory*, 47(2):619–637, 2001.

[24] T.J. Richardson and R.L. Urbanke. The capacity of Low-Density Parity-Check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2):599–618, 2001.

[25] V. Savin. Non binary LDPC codes over the binary erasure channel: density evolution analysis. In *First International Symposium on Applied Sciences on Biomedical and Communication Technologies*, pages 1–5. IEEE, 2008.

[26] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.

[27] V. Stankovic, A.D.Liveris, Z. Xiong, and C.N. Georghiades. On code design for the Slepian-Wolf problem and lossless multiterminal networks. *IEEE Transactions on Information Theory*, 52(4):1495 –1507, april 2006.

[28] V. Stankovic, A.D. Liveris, Z. Xiong, and C.N. Georghiades. Design of Slepian-Wolf codes by channel code partitioning. In *Data Compression Conference, Proceedings.*, pages 302 – 311, march 2004.

[29] R. Storn and K. Price. Differential evolution– a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[30] C.C. Wang, S.R. Kulkarni, and H.V. Poor. Density evolution for asymmetric memoryless channels. *IEEE Transactions on Information Theory*, 51(12):4216–4236, 2005.

[31] Z. Wang, X. Li, and M. Zhao. Distributed coding of Gaussian correlated sources using non-binary LDPC. In *Congress on Image and Signal Processing*, volume 2, pages 214–218. IEEE, 2008.

[32] Z-L. Wang, X-M Li, and Y. Xu. An Improved Decoding Algorithm for Distributed Video Coding. In *2nd International Congress on Image and Signal Processing, 2009. CISP'09.*, pages 1–4. IEEE, 2009.

[33] Z. Xiong, A.D. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5):80–94, Sep 2004.

Annexe D

Codage de WZ pour un canal de corrélation distribué suivant un modèle de Markov caché

Elsa Dupraz, Francesca Bassi, Thomas Rodet, Michel Kieffer *Source Coding with Side Information at the Decoder and Uncertain Knowledge of the Correlation* Technical report 2013

Certaines parties de cet article ont été publiées dans

Elsa Dupraz, Francesca Bassi, Thomas Rodet, Michel Kieffer *Distributed coding of sources with bursty correlation.* ICASSP 2012 : 2973-2976, Kyoto, Japan

Source Coding with Side Information at the Decoder and Uncertain Knowledge of the Correlation

Elsa Dupraz[1] Francesca Bassi[1], Thomas Rodet[1], Michel Kieffer[1,2,3]

[1] L2S - CNRS - SUPELEC - Univ Paris-Sud, 91192 Gif-sur-Yvette, France

[2] Partly on leave at LTCI – CNRS Télécom ParisTech, 75013 Paris, France

[3] Institut Universitaire de France

Abstract

This paper focuses on the performance of a Wyner-Ziv coding scheme for which the correlation between the source and the side information is modeled by a hidden Markov model with Gaussian emission. Such a signal model takes the memory of the correlation into account and is hence able to describe the bursty nature of the correlation between sources in applications such as sensor networks, video coding etc.

This paper provides bounds on the rate-distortion performance of a Wyner-Ziv coding scheme for such model. It proposes a practical coding scheme able to exploit the memory in the correlation. The coding scheme we consider is composed by a uniform scalar quantizer, followed by a Slepian-Wolf chain and at the end by some MMSE reconstruction. The Slepian-Wolf chain is realized with non-binary LDPC codes and we propose a sum-product LDPC decoder able to take into account the hidden Markov Models. We also introduce an MCMC method in order to realize the MMSE reconstruction. Finally, we compare the performance of the proposed coding scheme to the obtained bounds, and analyze the rate loss induced by each part of the scheme.

I. Introduction

In a network of sensors [36], the sensors have to transmit their measurements to a data collection node in charge of the reconstruction of all the information available in the network. As the sensors observe

Parts of this paper were presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013.

the same phenomenon, their measurements are correlated and each sensor can exploit the correlation in order to compress its data more efficiently [6], [32]. The vector of the measurements is referred to as the source vector. The correlation between the source symbols may vary with time [3], [7] and in particular, depending on external conditions, bursts of source samples with strong correlation may be followed by bursts of samples with low correlation level. Consequently, models assuming memoryless source symbols [1], [15], [30] may be unable to take the bursty aspect into account.

In this paper, the simpler case of two sources X and Y producing symbol sequences $\{X_k\}_{k=1}^{+\infty}$ and $\{Y_k\}_{k=1}^{+\infty}$ is considered. In Distributed Source Coding [36], the sensors cannot communicate with each other but are both aware of the correlation model between their measurements. They have to exploit this knowledge in order to compress their data more efficiently. In lossless coding [26], the decoder has to reconstruct the sequences $\{X_k\}_{k=1}^{+\infty}$ and $\{Y_k\}_{k=1}^{+\infty}$ without error, whereas in lossy coding [12], some distortion is tolerated.

Here, the coding setup is assumed asymmetric, *i.e.*, $\{Y_k\}_{k=1}^{+\infty}$ is directly available at the decoder (Wyner-Ziv coding [35]) and we are interested in the lossy case. The instantaneous correlation between X_k and Y_k is modeled introducing a hidden state S_k. The sequence of random variables $\{S_k\}_{k=1}^{+\infty}$ is generated according to a first-order Markov chain. The hidden state produces two correlation levels and the Markov chain represents the memory on the correlation level. The source symbols (X_k, Y_k) are jointly Gaussian with variance given by the realization of S_k. Furthermore, the knowledge of the sequence $\{S_k\}_{k=1}^{+\infty}$ breaks the dependency between the successive (X_k, Y_k). This corresponds to a Hidden Markov Model (HMM) [23]. In this paper, we are interested in the performance analysis and the construction of an asymmetric distributed source coding scheme able to exploit the memory on the hidden states.

For memoryless sources, the correlation was assumed simply Gaussian [34] or Gaussian conditionally to a memoryless state S_k giving the correlation level [3]. In the case of continuous sources with memory, Gauss-Markov models [4] and Gaussian correlation between the successive (X_k, Y_k) [12] were considered. The dependency between binary sources with memory has also been modeled by a Gilbert-Elliot channel [21]. But to the best of our knowledge no performance analysis has been provided for the model of the paper. The expression of the Wyner-Ziv rate-distortion function for general sources has been obtained in implicit form in [14], but the closed-form expression is difficult to derive for our model. Consequently

in this paper, bounds on the rate-distortion function are provided for the defined model.

On the other hand, several practical distributed coding schemes have been proposed for lossless coding in the memoryless case. They are based mainly on channel codes [29], and particularly Low Density Parity Check (LDPC) codes [20], [33]. Moreover, lossless schemes for binary sources and correlation described by a Gilbert-Elliot channel has been proposed in [9], [27] and a scheme for memoryless binary sources and bursty correlation is presented in [7]. In lossy source coding, [1] proposes a coding scheme for memoryless sources and [12] describes the case of jointly Gaussian sources. Here, the practical coding scheme we propose is composed by a Uniform Scalar Quantizer (USQ) followed by the lossless coding of the quantized symbols with the use of an LDPC code [8]. To finish, a Minimum Mean Square Error (MMSE) estimator [16] has to reconstruct the source symbols. The LDPC decoder as well as the MMSE estimator have to take the hidden states and the memory into account. For the LDPC part, a usual solution is to transform the non-binary source symbols into bits and to transmit the bit planes losslessly [27], [28]. The bit plane decomposition is easy at encoding, but complicates significantly the decoder that has to exploit the dependencies between bit planes [18], [31]. Consequently, in order to avoid this problem, the LDPC codes are directly in GF(q), the Galois Field of order q. LDPC decoders able to take the memory on the source symbols into account have been proposed [10], [11] for channel coding in the binary case. Here, these decoders are generalized to the non-binary case and adapted to our source model. Finally, as the MMSE estimator cannot be given in closed-form, a numerical sampling method based on Monte Carlo Markov Chains (MCMC) [5] and realizing the MMSE estimation is introduced.

The paper is organized as follows. Section II introduces the considered signal model. Section III gives the performance bounds and Section IV presents the proposed practical coding scheme. To finish, Section V shows simulation results.

II. Signal model

In the following, upper-case letters *e.g.*, Z denote random variables whereas lower-case letters z denote their realizations. In order to avoid a mix up with generic realizations, an observed variable is denoted $\bar{z}$.

The source X and side information Y (see Figure 1) generate real-valued symbols, corresponding to the realizations of the random sequences $\{X_k\}_{k=1}^{+\infty}$ and $\{Y_k\}_{k=1}^{+\infty}$. The instantaneous dependency is

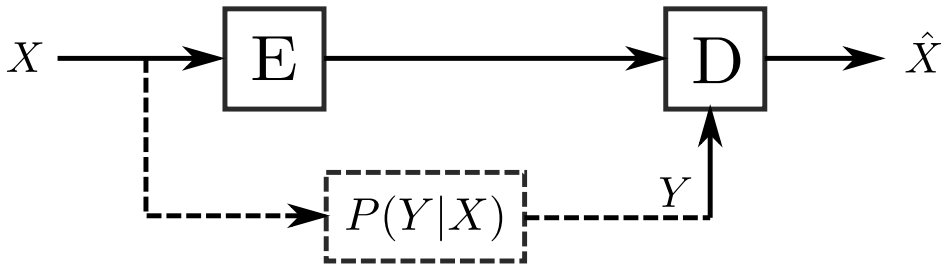

Fig. 1. Source coding with side information

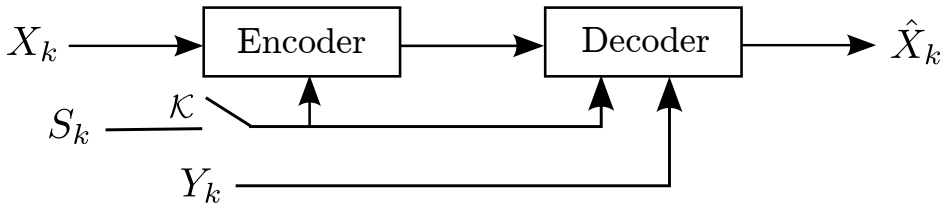

Fig. 2. Setup 1 ($\mathcal{K}$ closed) and Setup 2 ($\mathcal{K}$ open).

modeled by the additive channel $Y_k = X_k + Z_k$. The correlation noise Z_k and the source symbols X_k are assumed independent. The symbols in $\{X_k\}_{k=1}^{+\infty}$ are independent, identically distributed (i.i.d.) according to $\mathcal{N}(0, \sigma_x^2)$. The correlation noise sequence $\{Z_k\}_{k=1}^{+\infty}$ is distributed according to a hidden Markov model, with hidden state S_k and Gaussian emission. The state variable S_k takes values in $\mathcal{S} = \{0, 1\}$. Let σ_s^2 be the variance associated with the realization $S = s$, so that $(Z_k|S_k = s)$ is distributed according to $\mathcal{N}(0, \sigma_s^2)$. It is assumed $\sigma_0^2 < \sigma_1^2$. Denote also $\sigma_{X|Y,s}^2 = \text{Var}(X_n|Y_n, S_n = s) = \frac{\sigma_x^2 \sigma_s^2}{\sigma_x^2 + \sigma_s^2}$. The sequence $\{S_k\}_{k=1}^{+\infty}$ is a time-invariant Markov process with transition probability matrix P, with

$$P_{s',s} = \Pr(S_k = s | S_{k-1} = s') > 0 \quad \forall (s', s) \in \mathcal{S}^2, \tag{1}$$

and initial probabilities $p_s^{(1)} = P(S_1 = s) > 0$. The successive marginal probabilities are given by $P(S_k = s) = p_s^{(k)}$ and the sequence of marginals $\{p_s^{(k)}\}_{k=1}^{+\infty}$ converges to the stationary probabilities p_s (from the non-zero probability conditions). Note that if $\forall s \in \mathcal{S}$, $p_s^{(1)} = p_s$, then all the marginal probabilities $p_s^{(k)}$ are equal to p_s.

In the following, let $\mathbf{X}^n = (X_1 \dots X_n)$. We consider the quadratic distortion measure $d(\mathbf{X}^n, \hat{\mathbf{X}}^n) = \|\mathbf{X}^n - \hat{\mathbf{X}}^n\|^2$ and specify a distortion constraint D such as $E\left[\frac{1}{n} d(\mathbf{X}^n, \hat{\mathbf{X}}^n)\right] \leq D$.

III. PERFORMANCE

This section provided an analysis of the Wyner-Ziv (WZ) rate-distortion function $R_{X|Y}(D)$ for the signal model presented in section II. The WZ setup (Setup 2) is depicted in Figure 2, for Switch $\mathcal{K}$ open. Since $R_{X|Y}(D)$ cannot be expressed in closed form, we characterize the upper bound to the rate loss in $R_{X|Y}(D)$ with respect to the theoretical performance $R_{X|Y,S}(D)$ of a genie-aided setup, where the realization of S_k is made available both at the encoder and at the decoder. The genie-aided setup (Setup 1) is depicted in Figure 2, for Switch $\mathcal{K}$ closed.

Remarking that the random sequences defined in Section II are ergodic processes, the per-symbol rate-distortion function $R_{X|Y,S}(D)$ for Setup 1 is derived from [14] as

$$R_{X|Y,S}(D) = \lim_{n\to\infty} \frac{1}{n} R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D), \tag{2}$$

where $R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D)$ is defined by

$$R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D) = \inf \; I(\mathbf{X}^n; \mathbf{U}^n|\mathbf{Y}^n, \mathbf{S}^n). \tag{3}$$

The minimization in (3) is on the auxiliary random vectors $\mathbf{U}^n$ such that the Markov chain $\mathbf{U}^n \leftrightarrow (\mathbf{X}^n, \mathbf{S}^n) \leftrightarrow (\mathbf{Y}^n, \mathbf{S}^n)$ is satisfied and there exists some reconstruction function $f_n : \mathcal{U}^n \times \mathcal{Y}^n \times \mathcal{S}^n \to \mathcal{X}^n$ such that $E\left[\frac{1}{n} d(\mathbf{X}^n, f_n(\mathbf{U}^n, \mathbf{Y}^n, \mathbf{S}^n))\right] \leq D$ holds. Let $R^{\text{i.i.d.}}_{X|Y,S}(D)$ be the rate-distortion function associated to Setup 1 when the state variables S_k are assumed i.i.d. with probability distribution $P(S_k = s) = p_s$, $\forall k > 0$, as in [2].

Similarly, the rate-distortion function $R_{X|Y}(D)$ for Setup 2 is defined as

$$R_{X|Y}(D) = \lim_{n\to\infty} \frac{1}{n} R_{\mathbf{X}^n|\mathbf{Y}^n}(D) \tag{4}$$

where

$$R_{\mathbf{X}^n|\mathbf{Y}^n}(D) = \inf \; I(\mathbf{X}^n; \mathbf{U}^n|\mathbf{Y}^n). \tag{5}$$

The minimization in (5) is on the auxiliary random vectors $\mathbf{U}^n$ such that the Markov chain $\mathbf{U}^n \leftrightarrow \mathbf{X}^n \leftrightarrow \mathbf{Y}^n$ is satisfied and there exists some reconstruction function $f_n : \mathcal{U}^n \times \mathcal{Y}^n \to \mathcal{X}^n$ such that $E\left[\frac{1}{n} d(\mathbf{X}^n, f_n(\mathbf{U}^n, \mathbf{Y}^n))\right] \leq D$ holds.

A. Setup 1, closed-form expression of $R_{X|Y,S}(D)$

Proposition 1. *For the signal model defined in Section II,*

$$R_{X|Y,S}(D) = \sum_{s\in\mathcal{S}} p_s \max\left(0, \frac{1}{2}\log_2 \frac{\sigma^2_{X|Y,s}}{D'}\right), \tag{6}$$

where D' is such that $\sum_{s\in\mathcal{S}} p_s \min(D', \sigma^2_{X|Y,s}) \leq D$.

See Appendix A for the proof.

From Proposition 1 and [2], $R_{X|Y,S}(D) = R^{\text{i.i.d.}}_{X|Y,S}(D)$. Consequently, this setup does not suffer rate loss compared to the joint coding case. The rate distortion function $R_{X|Y,S}(D)$ can be expressed in closed-form expression because the variables X_k and Y_k are jointly Gaussian conditionally to S_k, and because the knowledge of S_k breaks the dependencies from time to time. Such a result is difficult to derive for Setup 2. Indeed, $R_{X|Y}(D)$ is difficult to obtain in explicit expression both because of the form of the minimization set in (5) and because the mutual information term has to be evaluated on sequences of dependent symbols. Nonetheless, bounds on $R_{X|Y}(D)$ can be derived as follows.

B. Setup 2, rate loss for $R_{X|Y}(D)$

Proposition 2. *For the signal model defined in Section II,*

$$R_{X|Y,S}(D) \leq R_{X|Y}(D) \leq R_{X|Y,S}(D) + L_{X|Y}(D) + \Lambda_{X|Y} \tag{7}$$

with

$$L_{X|Y}(D) = \frac{1}{2}\log_2\left(1 + \frac{D}{\sigma^2_{X|Y,0}}\right) \tag{8}$$

$$\Lambda_{X|Y} = \min\big(\lim_{k\to\infty} H(S_k|S_{k-1}), \quad h(\mathcal{Z}) - \lim_{k\to\infty} h(Z_k|S_k)\big), \tag{9}$$

and $h(\mathcal{Z})$ is the differential entropy rate, $h(\mathcal{Z}) = \lim_{n\to\infty} \frac{1}{n} h(\mathbf{Z}^n)$.

See Appendix A for the proof.

A lower bound to the rate-distortion function $R_{X|Y}(D)$ for Setup 2 is naturally provided by the performance of the genie-aided setup (Setup 1), $R_{X|Y,S}(D)$ because of the additional information given by the knowledge of $\{S_k\}_{k=1}^{+\infty}$. The upper bound is evaluated from a test-channel, *i.e.* particular choices of $\mathbf{U}^n$ and f_n from the minimization set in (5). This choice is not necessarily the optimal one, and hence

only an upper bound is obtained. The rate loss term (8) vanishes as $D \to 0$, and is due to the fact that no rate adaptation is possible at the encoder when the realization of S_k is not available. The rate loss term (9) is due to the uncertainty on S_k at the decoder side.

IV. Practical coding scheme

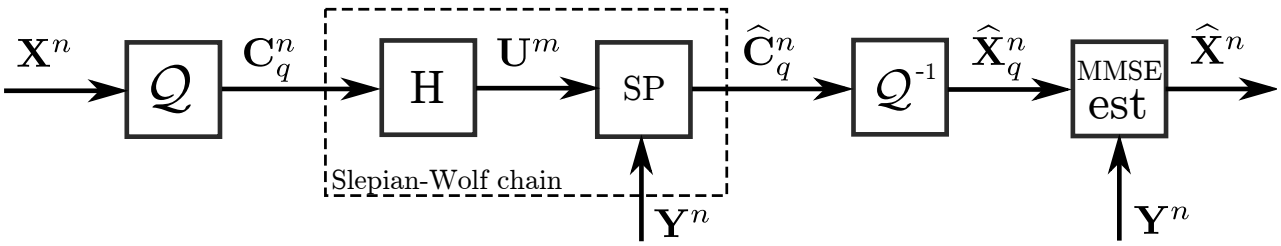

Fig. 3. Proposed WZ coding scheme

As described in Fig 3, the coding scheme we propose consists of a Dithered Uniform Scalar Quantizer (DUSQ, represented by $\mathcal{Q}$ on the scheme), a Slepian-Wolf (SW) chain, and a Minimum Mean Square Error (MMSE) reconstruction. More precisely, each symbol x_k of the sequence $\mathbf{x}^n$ is first quantized with a DUSQ with q quantization levels (Section IV-A) and the quantized symbol is mapped in GF(q), giving $\mathbf{c}_q^n$. Then, the quantized symbols are transmitted losslessly to the decoder, with the help of a Slepian-Wolf (SW) chain realized with an LDPC code of parity check matrix H in GF(q) and a Sum-Product (SP) LDPC decoder (Section IV-B). To finish, source reconstruction is performed via MMSE estimation of $\mathbf{x}^n$ (Section IV-C). Both the LDPC decoder and the MMSE estimator take the memory on the hidden states into account.

A. Uniform Scalar Quantization

From [24], DUSQ followed by SW encoding suffers only a 1.53 dB loss in the high rate regime compared to the case of an ideal quantizer. For this part, Trellis-Coded Quantization (TCQ) may also be employed, as suggested by [37]. Indeed, TCQ followed by SW encoding is shown experimentally to suffer only 0.22 dB loss compared to the ideal quantizer [37]. However, as pointed out in [37], TCQ makes difficult the evaluation of the probability distribution of the quantized symbols. In our case, as the LDPC decoder has to take the memory on the hidden states into account, there is a need to the probability distribution conditionaly to the hidden states. Hence the choice of a DUSQ.

The uniform scalar quantizer has $q = 2^l$ quantization levels. Denote Δ, q_j, and $(b_j,\ b_{j+1})$ respectively the size of the quantization cells, the level, and the boundaries of the j-th quantization level. Before quantization, a dither D_k distributed uniformly in $[0, \Delta]$ is added to each X_k. Then the successive $(x_k + d_k)$ are quantized and the quantized symbols are mapped one to one in GF(q), giving a sequence $\mathbf{c}_q^n$. As each $C_{\mathrm{q},k}$ is obtained from X_k only, the successive couples of random variables $(C_{\mathrm{q},k}, Y_k)$ are independent conditionally to S_k. Consequently, the correlation channel between $C_{q,k}$ and Y_k is driven by a HMM of hidden states S_k, and $P(C_{q,k} = j | Y_k = y_k, S_k = s)$ can be computed from the model and quantizer parameters. Finally, at the decoder side, $\mathcal{Q}^{-1}$ maps the symbols $\widehat{c}_{q,k}$ into their corresponding real-valued levels and subtracts the d_k, giving $\widehat{\mathbf{x}}_q^n$.

B. The LDPC decoder

The lossless compression of $\mathbf{c}_q^n$ with $\mathbf{y}^n$ available at the decoder is realized with LDPC codes. The encoding operation is realized by an LDPC coding matrix and the non-binary LDPC decoding algorithm has to be adapted in order to take the memory on the states into account. Note that a sum-product LDPC decoding algorithm has already been proposed in [10] for the Gilbert-Elliot channel (binary case). Here, the same decoding algorithm is derived for our model and in the non-binary case. As the LDPC decoder is a SP algorithm, we come back to the SP expressions [17] in order to derive the update equations for the LDPC decoding in our case. On the other side, expressing the equations resulting from the SP algorithm in an LDPC formalism allows to benefit from the tools already developed in this context.

The SW LDPC coding of a vector $\mathbf{c}_q^n$ is performed by producing a vector $\mathbf{u}^m$ of length $m < n$ as [20]

$$\mathbf{u}^m = H^{\mathrm{T}} \mathbf{c}_q^n. \tag{10}$$

The matrix H is sparse, with non-zero coefficients uniformly distributed in GF(q)$\backslash\{0\}$. In the following, $\oplus$, $\ominus$, $\otimes$, $\oslash$ are the usual operators in GF(q). The graph representing the dependencies between the random variables involved in (10) is called the bipartite graph. In this graph, the entries of $\mathbf{C}_q^n$ and $\mathbf{U}^m$ are represented respectively by Symbols Nodes (SN) and Check Nodes (CN). The set of CN connected to a SN C_k is denoted $\mathcal{N}_u(k)$ and the set of SN connected to a CN U_j is denoted $\mathcal{N}_c(j)$. The sparsity of H is determined by the degree distributions $\lambda(x) = \sum_{i\geq 2} \lambda_i x^{i-1}$ for the SN and $\rho(x) = \sum_{i\geq 2} \rho_i x^{i-1}$ for the CN, with $\sum_{i\geq 2} \lambda_i = 1$ and $\sum_{i\geq 2} \rho_i = 1$. The constant $0 \leq \lambda_i \leq 1$ is the proportion of edges emanating

from a SN of degree i and $0 \leq \rho_i \leq 1$ is the proportion of edges emanating from a CN of degree i [25]. In SW coding, the rate $r(\lambda, \rho)$ of a code is given by $r(\lambda, \rho) = \frac{m}{n} = \frac{\sum_{i \geq 2} \rho_i / i}{\sum_{i \geq 2} \lambda_i / i}$ [20].

We want to realize the Maximum *a posteriori* estimation of the source symbols $C_{q,k}$ from the side information vector $\bar{\mathbf{y}}^n$ and the received codeword $\bar{\mathbf{u}}^m$ as

$$\arg \max_{c \in \mathrm{GF}(q)} P(C_{q,k} = c | \bar{\mathbf{y}}^n, \bar{\mathbf{u}}^m) = \arg \max_{c \in \mathrm{GF}(q)} \sum_{\mathbf{s}^n, \mathbf{c}^n_{q \setminus k}} P(C_{q,k} = c, \mathbf{c}^n_{q \setminus k}, \mathbf{s}^n | \bar{\mathbf{y}}^n, \bar{\mathbf{u}}^m) \tag{11}$$

$$= \arg \max_{c \in \mathrm{GF}(q)} \sum_{\mathbf{s}^n, \mathbf{c}^n_{\setminus k}} P(C_{q,k} = c, \mathbf{c}^n_{q, \setminus k}, \mathbf{s}^n, \bar{\mathbf{y}}^n, \bar{\mathbf{u}}^m) \tag{12}$$

where $\mathbf{c}^n_{q \setminus k}$ means all the vector $\mathbf{c}^n_q$, except its k-th component. Note that the MAP estimation is realized only on the marginal probabilities of the $C_{q,k}$. This approach is suboptimal but much simpler than realizing the MAP of the whole sequence $\mathbf{C}^n_q$ from the joint probability distribution $P(\mathbf{C}_q | \bar{\mathbf{y}}^n, \bar{\mathbf{u}}^m)$. The summation in (12) can be obtained by the use of a SP algorithm. The joint probability distribution can be developed as

$$P(\mathbf{s}^n, \mathbf{c}^n_q, \bar{\mathbf{y}}^n, \bar{\mathbf{u}}^m) = \frac{1}{\kappa} \prod_{j=1}^{m} \delta \left(\sum_{k \in \mathcal{N}_c(j)} H_{kj} c_{q,k} - \bar{u}_j \right) \prod_{k=1}^{n} P(c_{q,k} | \bar{y}_k, s_k) P(\bar{y}_k | s_k) \prod_{k=2}^{n} P(s_k | s_{k-1}) P(s_1) \tag{13}$$

where κ is a normalization constant. From the decomposition, one can extend the bipartite graph between $\mathbf{C}^n_q$ and $\mathbf{U}^m$ into a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ representing the dependencies between all the involved random variables. The set of vertices $\mathcal{V}$ is composed by two types of nodes that are the Variables Nodes (VN) ($\mathbf{C}^n_q$, $\mathbf{S}^n$) and the Factor Nodes (FN) representing the relations between them. From the joint distribution decomposition (13), factors can be identified as $\Psi_k(c, s) = P(C_{q,k} = c | Y_k = \bar{y}_k, S_k = s)$, $\Psi_k(s) = P(Y_k = \bar{y}_k | S_k = s)$, $\Psi_{k-1,k}(s', s) = P(S_k = s | S_{k-1} = s')$, $\Psi_{j, \mathcal{N}_c(j)}(c_1 \dots c_{|\mathcal{N}_c(j)|}) = \delta \left(\sum_{k=1}^{|\mathcal{N}_c(j)|} H_{kj} c_{q,k} - \overline{u}_j \right)$ where $|\mathcal{N}_c(j)|$ is the number of elements in $\mathcal{N}_c(j)$. For simplicity the vertices of the graph have the names of the corresponding random variables. Note also that the random variables of $\mathbf{Z}^n$ do not appear in $\mathcal{G}$ because all the probability distributions involved in (13) can be expressed without marginalizing with respect to $\mathbf{Z}^n$.

We wish to perform the SP algorithm in $\mathcal{G}$ in order to obtain (12). In the following, we give the corresponding iterative message exchange. The message expressions are derived from the SP algorithm

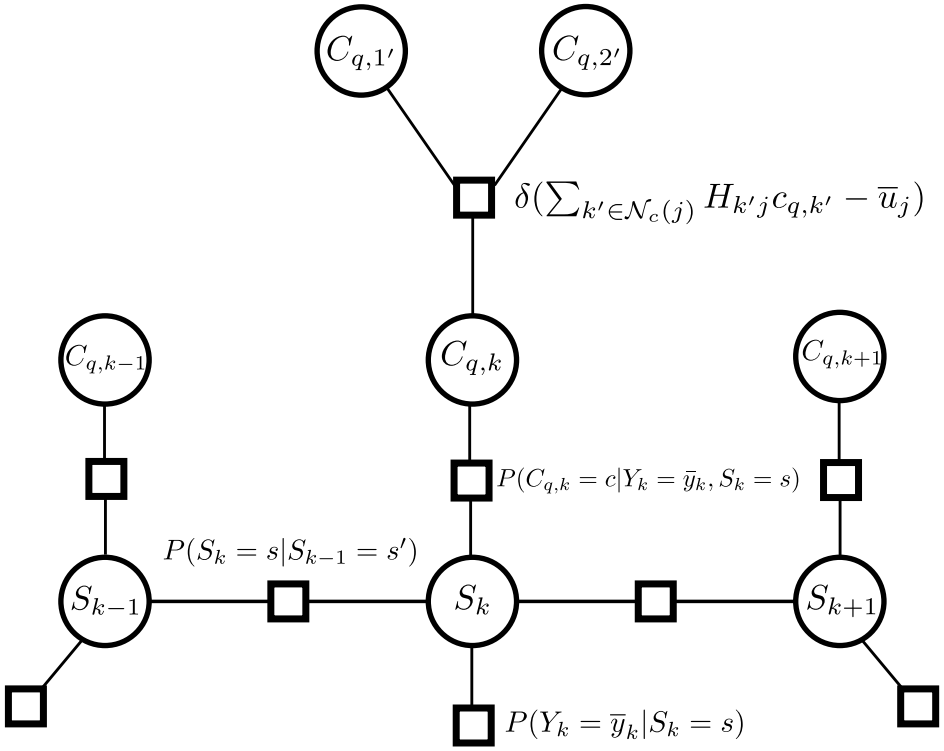

Fig. 4. A portion of the graph $\mathcal{G}$ representing the dependencies between variables

described in [17] for a general graph. At iteration ℓ, a message from an edge A to an edge B evaluated for a random variable C taking its values in $\mathcal{C}$ is denoted $p^{(\ell)}_{A\to B}(c)$. It is such that $0 \leq p^{(\ell)}_{A\to B}(c) \leq 1$ and $\sum_{c\in\mathcal{C}} p^{(\ell)}_{A\to B}(c) = 1$. The same message in the log-likelihood ratio (LLR) domain is denoted $m^{(\ell)}_{A\to B}(c)$ and can be computed as

$$m^{(\ell)}_{A\to B}(c) = \log \frac{p^{(\ell)}_{A\to B}(0)}{p^{(\ell)}_{A\to B}(c)}. \tag{14}$$

The message passing is described directly from VN to VN without the intermediate FN processing, except for the CN because they have a particular message computation.

a) Initialization: The messages from S_{k-1} to S_k are initialized $\forall s \in \mathcal{S}$ by $p^{(0)}_{S_{k-1}\to S_k}(s) = P(S_k = s)$. All the other messages are initialized by 1.

b) Message exchange at iteration ℓ: At iteration ℓ, the message from $C_{q,k}$ to S_k is computed $\forall s \in \mathcal{S}$ as

$$p^{(\ell)}_{C_{q,k}\to S_k}(s) = \sum_{c\in \mathrm{GF}(q)} P(C_{q,k} = c | Y_k = \bar{y}_k, S_k = s) \prod_{j\in\mathcal{N}_u(k)} p^{(\ell-1)}_{U_j\to C_k}(x) \tag{15}$$

The messages from S_{k-1} to S_k and the messages from S_{k+1} to S_k are calculated $\forall s \in \mathcal{S}$ as

$$p^{(\ell)}_{S_{k-1}\to S_k}(s) = \sum_{s'\in\mathcal{S}} P(S_k = s|S_{k-1} = s')P(Y_{k-1} = \bar{y}_{k-1}|S_{k-1} = s')p^{(\ell-1)}_{S_{k-2}\to S_{k-1}}(s')p^{(\ell-1)}_{C_{k-1}\to S_{k-1}}(s')$$

$$p^{(\ell)}_{S_{k+1}\to S_k}(s) = \sum_{s'\in\mathcal{S}} P(S_{k+1} = s'|S_k = s)P(Y_{k+1} = \bar{y}_{k+1}|S_{k+1} = s')p^{(\ell-1)}_{S_{k+2}\to S_{k+1}}(s')p^{(\ell-1)}_{C_{k+1}\to S_{k+1}}(s').$$

The messages from S_k to $C_{q,k}$ are calculated $\forall c \in$ GF(q) as

$$p^{(\ell)}_{S_k\to C_{q,k}}(c) = \sum_{s\in\mathcal{S}} P(C_{q,k} = c|Y_k = \bar{y}_k, S_k = s)P(Y_k = \bar{y}_k|S_k = s)p^{(\ell-1)}_{S_{k-1}\to S_k}(s)p^{(\ell-1)}_{S_{k+1}\to S_k}(s).$$

The messages from the SN $C_{q,k}$ to the connected CN U_j are computed $\forall c \in$ GF(q) as

$$p^{(\ell)}_{C_{q,k}\to U_j}(c) = p^{(\ell-1)}_{S_k\to C_{q,k}}(x) \prod_{j'\in\mathcal{N}_u(k)\backslash j} p^{(\ell-1)}_{U_{j'}\to,C_{q,k}}(x) \tag{16}$$

The messages from the CN U_j to the SN $C_{q,k}$ are computed for every $c \in$ GF(q) as

$$p_{U_j\to C_{q,k}}(c) = \sum\cdots\sum\left(\prod_{k'\in\mathcal{N}_c(j)\backslash k} p^{(\ell)}_{C_{q,k'}\to U_j}(c_{q,k'})\right)\delta\left(\sum_{k'\in\mathcal{N}_c(j)\backslash k} H_{k'j}c_{q,k'} + H_{kj}c - \bar{u}_j\right) \tag{17}$$

where the multiple summation is on all the values $c_{q,k'}$ that can be taken by the $C_{q,k'}$ such that $k' \in \mathcal{N}_c(j)\backslash k$.

The two last messages can be computed directly in LLR representation in order to be consistent with the classical LDPC message updates. In particular, the LLR messages from CN to SN are computed with the help of a Fourier-like transform, denoted $\mathcal{F}(\mathbf{m})$. Denoting r the unit root associated to GF(q), the i-th component of the transform is given by [19] as $\mathcal{F}_i(\mathbf{m}) = \sum_{j=0}^{q-1} r^{i\otimes j}e^{-m_j} / \sum_{j=0}^{q-1} e^{-m_j}$. One has

$$m^{(\ell)}_{C_{q,k}\to U_j}(c) = \sum_{j'\in\mathcal{N}_u(k)\backslash j} m^{(\ell-1)}_{U_{j'}\to C_{q,k}}(x) + m^{(\ell-1)}_{S_k\to C_{q,k}}(c) \tag{18}$$

and

$$\mathbf{m}^{(\ell)}_{U_j\to C_{q,k}} = \mathcal{A}[\tilde{u}_j]\mathcal{F}^{-1}\left(\prod_{k'\in\mathcal{N}_c(j)\backslash k} \mathcal{F}\left(W\left[\tilde{H}_{k'j}\right]\mathbf{m}^{(\ell-1)}_{C_{q,k'}\to U_j}\right)\right) \tag{19}$$

where $\mathbf{m}^{(\ell)}_{U_j\to C_{q,k}}$ and $\mathbf{m}^{(\ell-1)}_{C_{q,k'}\to U_j}$ are vectors of messages. In (19), $\tilde{u}_j = \ominus\bar{u}_j \oslash H_{kj}$, $\tilde{H}_{k'j} = \ominus H_{k'j} \oslash H_{kj}$, $W[a]$ is a $q \times q$ matrix such that $W_{i,j}[a] = \delta(a \otimes j \ominus i), 0 \leqslant i, j \leqslant q-1$ and $\mathcal{A}[k]$ is a $q \times q$ matrix that maps a message vector $\mathbf{m}$ into a message vector $\boldsymbol{\ell} = \mathcal{A}[k]\mathbf{m}$ with $\ell_j = m_{j\oplus k} - m_k$. Note that the whole message exchange may be described in LLR representation. Otherwise, the probability representation is mainly used because it gives much simpler message expressions.

c) Estimation of the VN values: Each VN computes its posterior probability $\forall c \in \text{GF}(q)$ as

$$p^{(\ell)}_{C_{q,k}}(c) = p^{(\ell)}_{S_k \to C_{q,k}}(c) \prod_{j \in \mathcal{N}_u(k)} p^{(\ell)}_{U_j \to C_{q,k}}(c) \tag{20}$$

It corresponds to the posterior probability that $C_{q,k} = c$ The decoder then produces an estimate of $c_{q,k}$ as

$$\hat{c}^{(\ell)}_{q,k} = \arg \max_{c \in \text{GF}(q)} p^{(\ell)}_{C_{q,k}}(c) \tag{21}$$

The algorithms stops if the produced $\hat{\mathbf{c}}^{(\ell)}_q$ meets the CN constraints or if the maximum number of iteration is reached.

Define the subgraph $\mathcal{N}^{2\ell}_k = \left\{ v \in \mathcal{V} : \bar{d}(v, C_{q,k}) \leq 2\ell \right\} \subseteq \mathcal{G}$ of depth 2ℓ centered on $C_{q,k}$. For example, $\mathcal{N}^2_k$ contains $C_{q,k}$, S_k, $\mathcal{N}_u(k)$, and $\forall j \in \mathcal{N}_u(k)$, $\mathcal{N}_c(j)$. At iteration ℓ, every $\hat{c}_{q,k}$ (21) is in fact computed from the messages exchanged in $\mathcal{N}^{2\ell}_k$. As an example, one can show that at the end of iteration 1, the described message exchange give the SP message passing in all the elementary subgraphs $\mathcal{N}^2_k$, $\forall k = 1 \dots n$. At iteration 2, the messages computed at iteration 1 (in $\mathcal{N}^2_{k-1}$, $\mathcal{N}^2_{k+1}$,etc.) are used to extend the SP message passing to all the subgraphs $\mathcal{N}^4_k$, and so on. Furthermore, from [17], if $\mathcal{N}^2_k$ is a tree, then $\hat{c}_k$ is the exact MAP estimate of $c_{q,k}$ from the variables observed in $\mathcal{N}^{2\ell}_k$. Interestingly, [10, Theorem 1] shows the following property

$$\forall \ell \in \mathbb{N}, P\left(\mathcal{N}^{2\ell}_k \text{ is a tree}\right) \to 1 \text{ as } n \to \infty \ . \tag{22}$$

Consequently, the SP algorithm gives reasonable approximate MAP estimates of the $c_{q,k}$.

C. The MMSE reconstruction

In this section we focus on the design of the decoder which outputs the reconstruction $\widehat{\mathbf{x}}^n$ of the source sequence. Since the distortion measure is quadratic, we consider MMSE estimation. Although the MMSE estimator cannot be characterized in analytical form, an asymptotically optimal implementation can be achieved via MCMC based methods [13, Chapter 8]. Assume that $\widehat{\mathbf{c}}^n_q$ is the perfect reconstruction of $\mathbf{c}^n_q$ and denote simply $\mathbf{x}^n_q$ the sequence obtained from this assumption. From [2, Section 1.3.2], the quantization error between X_k and $X_{q,k}$ is independent of X_k. Consequently, assume that there exists a sequence of independent random variables $\{B_{q,k}\}^n_{k=1}$ with $B_{\text{q},k} \sim \mathcal{N}(0, \Delta^2/12)$ such that $X_{\text{q},k} = X_k + B_{\text{q},k}$ where X_k and the quantization noise $B_{q,k}$ are independent.

We now describe the MMSE estimation process of $\mathbf{x}^n$ from the observed $\bar{\mathbf{x}}_q^n$ and $\bar{\mathbf{y}}^n$. The joint posterior distribution between the unknown variables can be expressed as

$$P\left(\mathbf{x}^n, \mathbf{s}^n | \bar{\mathbf{y}}^n, \bar{\mathbf{x}}_q^n\right) \;\propto\; P\left(\bar{\mathbf{x}}_q^n | \mathbf{x}^n\right) P\left(\bar{\mathbf{y}}^n | \mathbf{x}^n, \mathbf{s}^n\right) P\left(\mathbf{x}^n\right) P\left(\mathbf{s}^n\right) \tag{23}$$

where $\propto$ stands for *proportional to* and

$$\left(\mathbf{X}_q^n | \mathbf{x}^n\right) \sim \mathcal{N}\left(\mathbf{x}^n, \frac{\Delta^2}{12} I_n\right) \tag{24}$$

$$\left(\mathbf{Y}^n | \mathbf{x}^n, \mathbf{s}^n\right) \sim \mathcal{N}\left(\mathbf{x}^n, R_s\right) \tag{25}$$

$$\mathbf{X}^n \sim \mathcal{N}(0, \sigma_x^2 I_n) \tag{26}$$

$$P\left(\mathbf{s}^n\right) = P(s_1) \prod_{k=2}^{n} P(s_k | s_{k-1}) \tag{27}$$

where I_n is the identity matrix of size n. and $R_s = E\left[\mathbf{Z}^n (\mathbf{Z}^n)^{\mathrm{T}} | \mathbf{S}^n\right]$.

An MCMC method consists of sampling source vectors $\mathbf{x}^n$ from the posterior probability distribution $P(\mathbf{x}^n | \bar{\mathbf{y}}^n, \bar{\mathbf{x}}_q^n)$. The posterior mean can then be estimated by averaging over L samples. Unfortunately, $P(\mathbf{x}^n | \bar{\mathbf{y}}^n, \bar{\mathbf{x}}_q^n)$ is too complex to sample with, because its expression can only be obtained by marginalizing according to $\mathbf{s}^n$. On the other hand, sampling under $P(\mathbf{x}^n | \mathbf{s}^n, \bar{\mathbf{y}}^n, \bar{\mathbf{x}}_q^n)$ is much simpler, because it consists of sampling under a jointly Gaussian distribution. Thus we use a Gibbs sampler algorithm [5]. The principle of the Gibbs sampler is to sample under the conditional distributions $P(\mathbf{x}^n | \mathbf{s}^n, \bar{\mathbf{y}}^n, \bar{\mathbf{x}}_q^n)$ and $P(\mathbf{s}^n | \mathbf{x}^n, \bar{\mathbf{y}}^n, \bar{\mathbf{x}}_q^n)$ alternatively and iteratively. It is shown in [5] that when the number of iterations goes to infinity, the algorithm provides samples under the true joint distribution (23). Each iteration ℓ of the algorithm can be decomposed into two steps described as follows.

1. Sample $\mathbf{x}^{n,(\ell)}$ according to

$$P\left(\mathbf{x}^n | \mathbf{s}^{n,(l-1)}, \bar{\mathbf{x}}_q^n, \bar{\mathbf{y}}^n\right) = \frac{1}{(2\pi)^{n/2} R_x^{1/2}} \exp\left(-\frac{(\mathbf{x}^n - \mathbf{m}_x^n)^T R_x^{-1} (\mathbf{x}^n - \mathbf{m}_x^n)}{2}\right) \tag{28}$$

with

$$\begin{aligned} R_x &= \left(I_n \left(\frac{1}{\Delta^2/12} + \frac{1}{\sigma_x^2}\right) + R_s^{-1}\right)^{-1} \\ \mathbf{m}_x^n &= R_x \left(\frac{\bar{\mathbf{x}}_q^n}{\Delta^2/12} + R_s^{-1} \bar{\mathbf{y}}^n\right) \end{aligned} \tag{29}$$

2. Sample $\mathbf{s}^{n,(\ell)}$ sequentially according to

$$P\left(S_1^{(\ell)} = 1 | x_1^{(\ell)}, y_1, x_{q,1}\right) = \left(1 + \frac{P(S_1 = 1)}{P(S_1 = 0)} \sqrt{\frac{\sigma_0^2}{\sigma_1^2}} \;\; e^{\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)(x_1^{(\ell)} - \bar{y}_1)^2\right)}\right)^{-1} \tag{30}$$

and $\forall k = 2 \dots n$,

$$P\left(S_k^{(l)}=1|x_k^{(\ell)}, y_k, x_{q,k}, s_{k-1}^{(\ell)}\right) = \left(1 + \frac{P(S_k=1|s_{k-1}^{(\ell)})}{P(S_k=0|s_{k-1}^{(\ell)})}\sqrt{\frac{\sigma_0^2}{\sigma_1^2}}\ \ e^{\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)(x_k^{(\ell)}-\bar{y}_k)^2\right)}\right)^{-1} \tag{31}$$

in order to use the Markov chain property of $\mathbf{s}^n$.

The algorithm is initialized with a state sequence $\mathbf{s}^{n,(0)}$ sampled symbol by symbol from the *a priori* distribution of $\mathbf{s}^n$. To finish an approximation of the MMSE estimate $\widehat{\mathbf{x}}^n$ is calculated from the sampled sequences $\mathbf{x}^{n,(\ell)}$ as $\widehat{\mathbf{x}}^n = \frac{1}{L-b}\sum_{\ell=b+1}^{L}\mathbf{x}^{n,(\ell)}$ where b is the end of a burning period.

V. Experimental results

This section evaluates the performance of the proposed coding scheme. The memory of the state variable S is characterized by $\mu = 1 - p_{01} - p_{10} = 0.98$, see [21], where $p_{01} = \Pr(S_k = 1|S_{k-1} = 0) = 0.0156$ and $p_{10} = \Pr(S_k = 0|S_{k-1} = 1) = 0.0044$. The variance of X_k, is chosen as $\sigma_x^2 = 1$ and the parameters associated to source S are $\sigma_0^2 = 0.1,\ \sigma_1^2 = 0.01,\ p_0 = 0.9844$. Blocks of $N = 10000$ source samples are considered. We represent the lower bound and the upper bound and evaluate the performance of several setups. All the resulting curves are represented in Figure 5.

A. Lower and upper bounds

We first represent the lower bound (curve 6 in Figure 5) and the upper bound (curve 4 in Figure 5) to the rate-distortion function. In (9), $\mathrm{h}(\mathcal{Z})$ is evaluated numerically considering that the S_k are memoryless. At low rate, the rate evaluation is as follows. We assume a time-sharing system in which during a proportion $0 \leq \alpha \leq 1$ of the time, a rate $R(\overline{D})$ is transmitted and during the remaining $1-\alpha$ a rate 0 is transmitted. This gives a distortion $D = \alpha\overline{D} + (1-\alpha)\sigma_z^2$. The constant α is chosen in order to minimize the rate $\alpha R(\alpha\overline{D} + (1-\alpha)\sigma_z^2)$.

B. MMSE decoder with ideal SW chain

Then the performance of the MMSE decoder is evaluated. We assume an ideal SW chain achieving the theoretical performance for ergodic sources. We generate 100 realizations of $\mathbf{X}^n$, $\mathbf{Y}^n$ and $\mathbf{B_q}^n$, and we decode the sequence with the proposed MCMC method to obtain $\hat{\mathbf{X}}^n$. The distortion is given by the

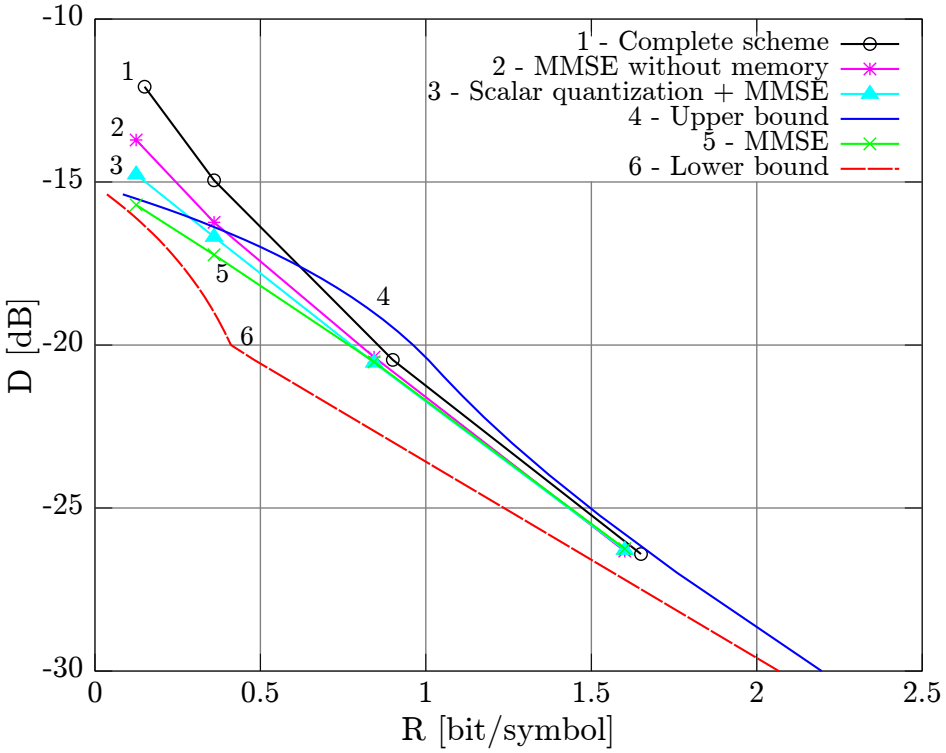

Fig. 5. Rate-distortion curves for the considered setups

mean-squared error $\frac{1}{n}\|\mathbf{X}^n - \hat{\mathbf{X}}^n\|^2$ averaged over all the realizations. Under the assumption of ideal SW chain, the rate needed to transmit $\mathbf{C}_q^n$ losslessly is given by [2]

$$R(\sigma_\phi^2) \;=\; \frac{1}{n}h(\mathbf{C}_q^n|\mathbf{Y}^n) - \frac{1}{2}\log_2\left(2\pi e\frac{\Delta^2}{12}\right) \tag{32}$$

where $h(\mathbf{C}_q^n|\mathbf{Y}^n)$ is evaluated numerically. For the MCMC method, 400 samples are generated. We consider the following numbers of quantization levels : 8, 16, 32, 64. The dynamic of the quantizer is given by $[-5\sigma_x^2, 5\sigma_x^2]$. Three setups are evaluated here. In the first setup (MMSE, curve 5), the $\mathbf{C}_q^n$ used by the MMSE estimator are directly generated according to the model $\mathbf{C}_q^n = \mathbf{X}_q^n + \mathbf{B}_q^n$ (see Section IV-C). In the second setup (MMSE without memory, curve 2), the $\mathbf{C}_q^n$ are generated as for the first setup, but the memory is not taken into account in the MMSE reconstruction. We see that if the memory is not taken into account, there is a loss of approximately 2 dB at low rate. In the third setup, (scalar quantization + MMSE, curve 3), the $\mathbf{C}_q^n$ are obtained from the scalar quantization of the $\mathbf{X}^n$. At low rate, there is a loss of approximately 1 dB compared to the first setup. We see that the Gaussian model is relevant

to represent the scalar quantization. In every case, when the rate increases, the three methods give the same performance. Indeed, in high rate, the quantization noise is low and the decoder relies more on the quantized symbols than on the side information. Furthermore, the memory is only on the side information, which explains that at high rate, the proposed method does not gain anymore.

C. Complete scheme

Here the complete scheme, *e.g.*, scalar quantizer + LDPC code + MMSE reconstruction is evaluated. The parameters for the scalar quantizer and MMSE reconstruction are the same as before. The following LDPC codes are considered (node-perspective degrees). For 8 quantization levels, $\lambda(x) = x$, $\rho(x) = x^{39}$. For 16 quantization levels, $\lambda(x) = x$, $\rho(x) = 0.78x^{21} + 0.22x^{22}$. For 32 quantization levels, $\lambda(x) = x$, $\rho(x) = 0.01x^9 + 0.88x^{10} + 0.11x^{11}$. For 64 quantization levels, $\lambda(x) = x$, $\rho(x) = 0.01x^5 + 0.72x^6 + 0.27x^7$. Note that, only code of SN degree 2 are considered, as suggested by [22] for the construction of non-binary LDPC codes. From these degree distributions, the LDPC coding matrices are generated from an LDPC PEG method. We set 40 iterations for the LDPC decoding. We observe a loss compared to the ideal MMSE case. First, a part of the loss is because the LDPC decoding is not perfect, which increases the distortion. A second part of the loss is due to the code construction.

VI. Conclusion

This work introduces an HMM driven correlation model in the context of lossy source coding with side information at the decoder. This model may capture the bursty nature of source correlation, *e.g.*, in sensor networks. Bounds on the rate-distortion function are obtained for this model, and a practical coding scheme is proposed.

Moreover, until now, only the cases with either perfect estimate of the states or no knowledge of the states have been considered. The next step may be to extend the results to imperfect knowledge of the states, or imperfect knowledge of the source correlation distribution.

APPENDIX

APPENDIX

If the initial probabilities $p_s^{(1)}$ are equal to the stationary probabilities p_s, it is easy to show that $p_s^{(k)} = p_s$, $\forall k > 0$. In this case, $\frac{1}{n} R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D) = R^{\text{i.i.d.}}_{X|Y,S}(D)$ because the knowledge of S_k breaks the temporal dependency and the S_k are identically distributed. Consequently in this case, we have directly $R_{X|Y,S}(D) = R^{\text{i.i.d.}}_{X|Y,S}(D)$. On the contrary, if $p_s^{(1)} \neq p_s$, $p_s^{(k)}$ varies with k, and the S_k are not identically distributed anymore. In this case, we have to show that $\frac{1}{n} R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D)$ converges to $R^{\text{i.i.d.}}_{X|Y,S}(D)$.

Define a new source as follows. Let $\{\tilde{X}_k\}_{k=1}^{+\infty}$, $\{\tilde{Y}_k\}_{k=1}^{+\infty}$ be two random sequences such that $\tilde{Y}_k = \tilde{X}_k + \tilde{Z}_k$, where $\tilde{X}_k$ and $\tilde{Z}_k$ are independent. $\tilde{X}_k$ is defined as X_k in Section II. The symbols of the sequence $\{\tilde{Z}_k\}_{k=1}^{+\infty}$ are independent, distributed according to a sequence of independent hidden states $\{\tilde{S}_k\}_{k=1}^{+\infty}$, taking their values in $\mathcal{S}$, with $P(\tilde{S}_k = s) = P(S_k = s)$ and $(\tilde{Z}_k | S_k = s) \sim \mathcal{N}(0, \sigma_s^2)$. In this model, the symbols $\tilde{S}_k$ are memoryless, with the marginal probabilities of the model of the paper.

Consider the per-symbol rate-distortion function $R_{\tilde{X}|\tilde{Y},\tilde{S}}(D)$ of the source in Setup 1, *i.e.* when $\{\tilde{S}_k\}_{k=1}^{+\infty}$ is available at the decoder. From [14], one has

$$R_{\tilde{X}|\tilde{Y},\tilde{S}}(D) = \lim_{n\to\infty} \frac{1}{n} R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D), \tag{33}$$

where $R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D)$ is defined by

$$R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D) = \inf \; I(\tilde{\mathbf{X}}^n; \tilde{\mathbf{U}}^n | \tilde{\mathbf{Y}}^n, \tilde{\mathbf{S}}^n). \tag{34}$$

The minimization in (34) is on the auxiliary random vectors $\tilde{\mathbf{U}}^n$ such that the Markov chain $\tilde{\mathbf{U}}^n \leftrightarrow (\tilde{\mathbf{X}}^n, \tilde{\mathbf{S}}^n) \leftrightarrow (\tilde{\mathbf{Y}}^n, \tilde{\mathbf{S}}^n)$ is satisfied and there exists some reconstruction function $f_n : \tilde{\mathcal{U}}^n \times \tilde{\mathcal{Y}}^n \times \tilde{\mathcal{S}}^n \to \tilde{\mathcal{X}}^n$ such that $E\left[\frac{1}{n} d(\tilde{\mathbf{X}}^n, f_n(\tilde{\mathbf{U}}^n, \tilde{\mathbf{Y}}^n, \tilde{\mathbf{S}}^n))\right] \leq D$ holds.

Proposition 3. $R_{\tilde{X}|\tilde{Y},\tilde{S}}(D) = R^{i.i.d.}_{X|Y,S}(D)$.

Proof: The mutual information $I(\tilde{\mathbf{X}}^n; \tilde{\mathbf{U}}^n | \tilde{\mathbf{Y}}^n, \tilde{\mathbf{S}}^n)$ can be expressed as

$$I(\tilde{\mathbf{X}}^n; \tilde{\mathbf{U}}^n | \tilde{\mathbf{Y}}^n, \tilde{\mathbf{S}}^n) = \sum_{k=1}^{n} \sum_{s\in\mathcal{S}} P(S_k = s) I(\tilde{X}_k; \tilde{U}_k | \tilde{Y}_k, \tilde{S}_k = s) \tag{35}$$

by independence of the involved random variables. The distortion constraint can be restated as

$$E\left[\frac{1}{n} d(\tilde{\mathbf{X}}^n, f_n(\mathbf{U}^n, \tilde{\mathbf{Y}}^n, \tilde{\mathbf{S}}^n))\right] = \frac{1}{n} \sum_{k=1}^{n} \sum_{s\in\mathcal{S}} P(S_k = s) \delta_{k,s} \leq D. \tag{36}$$

Remarking that the random variables $(X_k, Y_k|S_k = s)$ are jointly Gaussian, $\frac{1}{n}R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D)$ is of the form

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{s\in\mathcal{S}} P(S_k = s)I(\tilde{X}_k; \tilde{U}_k|\tilde{Y}_k, \tilde{S}_k = s) = \frac{1}{n}\sum_{k=1}^{n}\sum_{s\in\mathcal{S}} P(S_k = s)\frac{1}{2}\log_2\frac{\sigma^2_{X|Y,s}}{\delta_{k,s}}. \tag{37}$$

This equation has to be minimized with respect to $\delta_{k,s}$ taking into account the constraint (36). By using a Lagrange multiplier, we show that this gives

$$\frac{1}{n}R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D) = \sum_{s\in\mathcal{S}}\left(\frac{1}{n}\sum_{k=1}^{n}P(S_k = s)\right)\max\left(0, \frac{1}{2}\log_2\left(\frac{\sigma^2_{X|Y,s}}{D'}\right)\right), \tag{38}$$

where D' is such that $\sum_{s\in\mathcal{S}}\left(\frac{1}{n}\sum_{k=1}^{n}P(S_k = s)\right)\min(D', \sigma^2_{X|Y,s}) \leq D$. By ergodicity, $\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n}P(S_k = s)\right)$ p_s, the stationary probability. Finally, we get

$$\lim_{n\to\infty}\frac{1}{n}R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D) = R^{\text{i.i.d.}}_{X|Y,S}(D) \tag{39}$$

where the expression of $R^{\text{i.i.d.}}_{X|Y,S}(D)$ comes from [2].

■

We first derive a lower bound on $R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D)$. One has

$$I(\mathbf{X}^n; \mathbf{U}^n|\mathbf{Y}^n, \mathbf{S}^n) = \sum_{k=1}^{n}\left(h(X_k|\mathbf{X}^{k-1}, \mathbf{Y}^n, \mathbf{S}^n) - h(X_k|\mathbf{X}^{k-1}, \mathbf{U}^n, \mathbf{Y}^n, \mathbf{S}^n)\right)$$

by the chain rule, and

$$I(\mathbf{X}^n; \mathbf{U}^n|\mathbf{Y}^n, \mathbf{S}^n) \geq \sum_{k=1}^{n}\left(h(X_k|Y_k, S_k) - h(X_k|U_k, Y_k, S_k)\right) \tag{40}$$

by the knowledge of S_k (left entropy term) and the fact that conditioning reduces entropy (right entropy term). Consequently,

$$I(\mathbf{X}^n; \mathbf{U}^n|\mathbf{Y}^n, \mathbf{S}^n) \geq \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} P(S_k = s)I(X_k; U_k|Y_k, S_k = s)$$

Finally, by applying the $\inf$ on both parts of the inequality,

$$\frac{1}{n}R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D) \geq \frac{1}{n}R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D). \tag{41}$$

On the other side, an achievable rate for Setup 2 is given by a particular choice of $\mathbf{U}^n$ and f_n from the minimization set in (5). We choose $\mathbf{U}^n$ and f_n for which the memory on S_k is not taken into account. The evaluation of the mutual information term in (5) for this particular choice gives $R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D)$. This

choice is not necessarily optimal, *e.g.* a better reconstruction function might exist. Consequently it gives only an upper bound on the rate-distortion function as

$$\frac{1}{n}R_{\mathbf{X}^n|\mathbf{Y}^n,\mathbf{S}^n}(D) \leq \frac{1}{n}R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D). \tag{42}$$

To finish, by taking the limit when $n \to \infty$ in (41) and (42) and from Proposition 3, $R_{X|Y,S}(D) = R^{\text{i.i.d.}}_{X|Y,S}(D)$.

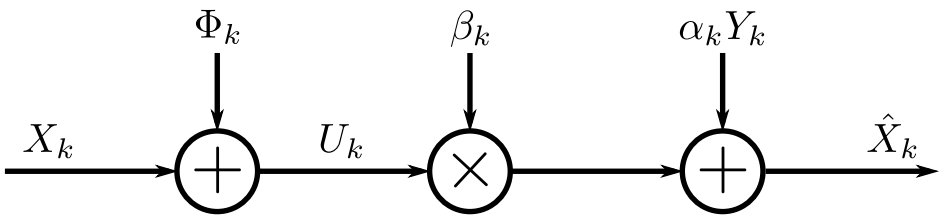

Fig. 6. Test channel

Fix a sequence length n. A test channel consists of particular choices of auxiliary random vector $\mathbf{U}^n$ and reconstruction function f_n, for which $I(\mathbf{X}^n;\mathbf{U}^n|\mathbf{Y}^n)$ is evaluated. The test channel we consider is depicted in Figure 6. The random variable Φ_k is distributed as $\mathcal{N}(0,\sigma^2_{\phi_k})$, $U_k = X_n + \Phi_k$ and Φ_k is independent of X_k. $\hat{X}_k = \alpha_k Y_k + \beta_k U_k$ is the LMMSE estimate of X_k from Y_k and U_k as where

$$\alpha_k = \frac{1}{\sigma^2_{z_k}}\left(\frac{1}{\sigma^2_{z_k}} + \frac{1}{\sigma^2_{\phi_k}} + \frac{1}{\sigma^2_x}\right)^{-1}, \quad \beta_k = \frac{1}{\sigma^2_{\phi_k}}\left(\frac{1}{\sigma^2_{z_k}} + \frac{1}{\sigma^2_{\phi_k}} + \frac{1}{\sigma^2_x}\right)^{-1}. \tag{43}$$

The distortion constraint can be restated as $E[\frac{1}{n}d(\mathbf{X}^n,\hat{\mathbf{X}}^n)] = \frac{1}{n}\sum_{k=1}^n \delta_k \leq D$ where the distortion δ_k on the n-th component is given by

$$\delta_k = \left(\frac{1}{\sigma^2_{z_k}} + \frac{1}{\sigma^2_{\phi_k}} + \frac{1}{\sigma^2_x}\right)^{-1}. \tag{44}$$

From (5), the rate of the test channel can be evaluated as

$$R_0 = I(\mathbf{X}^n;\mathbf{U}^n|\mathbf{Y}^n) \tag{45}$$

$$= h(\mathbf{U}^n) - h(\mathbf{U}^n|\mathbf{X}^n) - h(\mathbf{Y}^n) + h(\mathbf{Y}^n|\mathbf{U}^n) \tag{46}$$

by developing the entropy terms and from the Markov Chain $\mathbf{U}^n \leftrightarrow \mathbf{X}^n \leftrightarrow \mathbf{Y}^n$. Developing $h(\mathbf{Y}^n)$ and $h(\mathbf{Y}^n|\mathbf{U}^n)$ yields

$$\begin{aligned} h(\mathbf{Y}^n) &= h(\mathbf{Y}^n|\mathbf{S}^n) - H(\mathbf{S}^n) + H(\mathbf{S}^n|\mathbf{Y}^n) \\ h(\mathbf{Y}^n|\mathbf{U}^n) &= h(\mathbf{Y}^n|\mathbf{U}^n,\mathbf{S}^n) - H(\mathbf{S}^n) + H(\mathbf{S}^n|\mathbf{U}^n,\mathbf{Y}^n)\ . \end{aligned} \tag{47}$$

Furthermore,

$$\begin{aligned} h(\mathbf{Y}^n|\mathbf{U}^n,\mathbf{S}^n) =& h(\mathbf{X}^n+\mathbf{Z}^n|\mathbf{X}^n+\mathbf{\Phi}^n,\mathbf{S}^n) \\ =& \sum_{k=1}^{n} h(X_k+Z_k|X_k+\Phi_k,S_k) \end{aligned} \tag{48}$$

because the components are conditionally independent with respect to S_k. Finally, (48) gives

$$h(\mathbf{Y}^n|\mathbf{U}^n,\mathbf{S}^n) = \sum_{k=1}^{n} h((1-\gamma_k)X_k+Z_k-\gamma_k\Phi_k|S_k) \tag{49}$$

where $\gamma_k = \frac{\sigma_x^2}{\sigma_x^2+\sigma_{\Phi_k}^2}$ is the coefficient of the LMMSE estimator of X_k+Z_k from $X_k+\Phi_k$ and the equality holds because the random variables are Gaussian conditionally to S_k.

Replacing $h(\mathbf{Y}^n)$ and $h(\mathbf{Y}^n|\mathbf{U}^n)$ by their expressions from (47) and (49) in (45) gives

$$\begin{aligned} R_0 =& \sum_{k=1}^{n}\frac{1}{2}\log_2\left(\frac{\sigma_x^2+\sigma_{\Phi_k}^2}{\sigma_{\Phi_k}^2}\right) + \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} p(S_k=s)\frac{1}{2}\log_2\left(\frac{(1-\gamma_k)\sigma_x^2+\sigma_s^2+\gamma_k^2\sigma_{\Phi_k}^2}{\sigma_x^2+\sigma_s^2}\right) \\ &+ I(\mathbf{S}^n;\mathbf{U}^n|\mathbf{Y}^n) \\ =& \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} p(S_k=s)\frac{1}{2}\log_2\left(\frac{\sigma_x^2\sigma_s^2+\sigma_x^2\sigma_{\Phi_k}^2+\sigma_s^2\sigma_{\Phi_k}^2}{\sigma_{\Phi_k}^2(\sigma_x^2+\sigma_s^2)}\right) + I(\mathbf{S}^n;\mathbf{U}^n|\mathbf{Y}^n) \\ =& \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} p(S_k=s)\frac{1}{2}\log_2\left(\frac{\sigma_{X|Y,s}^2+\sigma_{\Phi_k}^2}{\sigma_{\Phi_k}^2}\right) + I(\mathbf{S}^n;\mathbf{U}^n|\mathbf{Y}^n) \\ =& \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} p(S_k=s)\frac{1}{2}\log_2\frac{\sigma_{X|Y,s}^2}{\delta_{k,s}} + \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} p(S_k=s)\frac{1}{2}\log_2\left(\frac{\delta_{k,s}}{\sigma_{X|Y,s}^2}\frac{\sigma_{X|Y,s}^2+\sigma_{\Phi_k}^2}{\sigma_{\Phi_k}^2}\right) \\ &+ I(\mathbf{S}^n;\mathbf{U}^n|\mathbf{Y}^n) \\ =& R_{\tilde{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n,\tilde{\mathbf{S}}^n}(D) + L_{\mathbf{X}^n|\mathbf{Y}^n}(D) + \Lambda_{\mathbf{X}^n|\mathbf{Y}^n} \end{aligned} \tag{50}$$

where

$$L_{\mathbf{X}^n|\mathbf{Y}^n}(D) = \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} P(S_k=s)\frac{1}{2}\log_2\left(\frac{\delta_{k,s}}{\sigma_{X|Y,s}^2}\frac{\sigma_{X|Y,s}^2+\sigma_{\Phi_k}^2}{\sigma_{\Phi_k}^2}\right) \tag{51}$$

$$\Lambda_{\mathbf{X}^n|\mathbf{Y}^n} = I(\mathbf{S}^n;\mathbf{U}^n|\mathbf{Y}^n). \tag{52}$$

The two terms $L_{\mathbf{X}^n|\mathbf{Y}^n}(D)$ and $\Lambda_{\mathbf{X}^n|\mathbf{Y}^n}$ are then bounded separately.

$L_{\mathbf{X}^n|\mathbf{Y}^n}(D)$ can be restated as

$$\begin{aligned} L_{\mathbf{X}^n|\mathbf{Y}^n}(D) =& \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} P(S_k=s)\frac{1}{2}\log_2\delta_{k,s} + \sum_{k=1}^{n}\sum_{s\in\mathcal{S}} P(S_k=s)\log_2\left(\frac{1}{\sigma_{\Phi_k}^2}+\frac{1}{\sigma_{X|Y,s}^2}\right) \\ \leq& \frac{1}{2}\log_2 D + \sum_{n=1}^{k}\log_2\left(\frac{1}{\sigma_{\Phi_k}^2}+\frac{1}{\sigma_{X|Y,0}^2}\right) \end{aligned} \tag{53}$$

from Jensen's inequality and the fact that $\sigma^2_{X|Y,0}$ is the smallest variance. Moreover, since $\delta_k \leq \sigma^2_{\Phi_k}$, one has

$$\frac{1}{n} L_{\mathbf{X}^n|\mathbf{Y}^n}(D) \leq \frac{1}{n} \sum_{k=1}^{n} \frac{1}{2} \log_2 \left(1 + \frac{\delta_k}{\sigma^2_{X|Y,0}} \right) \leq \log_2 \left(1 + \frac{D}{\sigma^2_{X|Y,0}} \right) \tag{54}$$

by concavity of the $\log$.

We now express a first bound on $\Lambda_{\mathbf{X}^n|\mathbf{Y}^n}$ as

$$\Lambda_{\mathbf{X}^n|\mathbf{Y}^n} = H(\mathbf{S}^n|\mathbf{Y}^n) - H(\mathbf{S}^n|\mathbf{U}^n, \mathbf{Y}^n) \leq H(\mathbf{S}^n) \tag{55}$$

by the positivity of the entropy terms and the fact that conditioning reduces entropy. In another way,

$$\begin{aligned} \Lambda_{\mathbf{X}^n|\mathbf{Y}^n} &\leq H(\mathbf{S}^n|\mathbf{Y}^n) - H(\mathbf{S}^n|\mathbf{U}^n, \mathbf{Y}^n, \mathbf{Z}^n) = H(\mathbf{S}^n|\mathbf{Y}^n) - H(\mathbf{S}^n|\mathbf{Z}^n) \\ &\leq I(\mathbf{S}^n; \mathbf{Z}^n) - I(\mathbf{S}^n; \mathbf{Y}^n) \leq h(\mathbf{Z}^n) - h(\mathbf{Z}^n|\mathbf{S}^n). \end{aligned} \tag{56}$$

Consequently,

$$\frac{1}{n} \Lambda_{\mathbf{X}^n|\mathbf{Y}^n} \leq \frac{1}{n} \min \left(H(\mathbf{S}^n), h(\mathbf{Z}^n) - h(\mathbf{Z}^n|\mathbf{S}^n) \right). \tag{57}$$

It is easy to show that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} H(\mathbf{S}^n) &= \lim_{k\to\infty} H(S_k|S_{k-1}) = \sum_{s' \in \mathcal{S}} p_{s'} H(S_k|S_{k-1} = s') \\ \lim_{n\to\infty} \frac{1}{n} H(\mathbf{Z}^n|\mathbf{S}^n) &= \lim_{k\to\infty} H(Z_k|S_k) = \sum_{s \in \mathcal{S}} p_s H(Z_k|S_k = s) \end{aligned} \tag{58}$$

where $H(S_k|S_{k-1} = s')$ and $H(Z_k|S_k = s)$ do not depend on k. Thus the limits in (58) exist and are finite. When $n \to \infty$, from (50), Proposition 3, (54) and (57), we get

$$R_{X|Y,S}(D) \leq R^{\text{i.i.d.}}_{X|Y,S}(D) + L_{X|Y}(D) + \Lambda_{X|Y} \tag{59}$$

where

$$L_{X|Y}(D) = \frac{1}{2} \log_2 \left(1 + \frac{D}{\sigma^2_{X|Y,0}} \right) \tag{60}$$

$$\Lambda_{X|Y} \leq \min \left(\lim_{k\to\infty} H(S_k|S_{k-1}), h(\mathcal{Z}) - \lim_{k\to\infty} h(Z_k|S_k) \right). \tag{61}$$

REFERENCES

[1] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv coding of motion video. In *Conference Record of the 36th Asilomar Conference on Signal, Systems and Computers*, volume 1, pages 240–244, 2002.

[2] F. Bassi. *Wyner-Ziv coding with uncertain side information quality*. PhD thesis, Université Paris-Sud, 2010.

[3] F. Bassi, M. Kieffer, and C. Weidmann. Wyner-Ziv coding with uncertain side information quality. In *Proceedings of the European Signal Processing Conference*, pages 2141–2145, 2010.

[4] T. Chu and Z. Xiong. Coding of Gauss-Markov sources with side information at the decoder. *IEEE Workshop on Statistical Signal processing*, 2003.

[5] C.P. Robert and G. Casella. *Monte-Carlo statistical methods*. Springer Texts in Statistics. Springer, New York, 2000.

[6] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. Networked Slepian-Wolf: theory, algorithms, and scaling laws. *IEEE Transactions on Information Theory*, 51(12):4057–4073, 2005.

[7] L. Cui, S. Wang, S. Cheng, and M. Yeary. Adaptive binary Slepian-Wolf decoding using particle based belief propagation. *IEEE Transactions on Communications*, 59(9):2337–2342, 2011.

[8] M.C. Davey and D. MacKay. Low-density parity check codes over GF(q). *IEEE Communications Letters*, 2(6):165–167, 1998.

[9] J. Del Ser, P.M. Crespo, and O. Galdos. Asymmetric joint source-channel coding for correlated sources with blind HMM estimation at the receiver. *EURASIP Journal on wireless communications and networking*, 2005(4):483–492, 2005.

[10] A.W. Eckford, F.R. Kschischang, and S. Pasupathy. Analysis of low-density parity-check codes for the Gilbert-Elliott channel. *IEEE Transactions on Information Theory*, 51(11):3872–3889, 2005.

[11] J. Garcia-Frias. Decoding of low-density parity-check codes over finite-state binary Markov channels. *IEEE Transactions on Communications*, 52(11):1840–1843, 2004.

[12] M. Gastpar, P.L. Dragotti, and M. Vetterli. The distributed Karhunen-Loeve transform. *IEEE Trans. Inf. Th.*, 52(12):1–10, 2006.

[13] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2009.

[14] K. Iwata. An information-spectrum approach to rate-distortion function with side information. In *Proceedings of the IEEE International Symposium on Information Theory*, page 156, 2002.

[15] S. Jalali, S. Verdú, and T. Weissman. A universal scheme for Wyner-Ziv coding of discrete sources. *IEEE Transactions Information Theory*, 56(4):1737–1750, 2010.

[16] S.M. Kay. *Fundamentals of Statistical Signal Processing, Estimation theory*. Prentice Hall PTR, 1993.

[17] F.R. Kschischang, B.J. Frey, and H-A. Loeliger. Factor graphs and the Sum-Product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

[18] G. Lechner and C. Weidmann. Optimization of binary LDPC codes for the q-ary symmetric channel with moderate q. In *Proceedings of the International Symposium on Turbo Codes and Related Topics*, pages 221–224, 2008.

[19] G. Li, I.J. Fair, and W.A. Krzymien. Density evolution for nonbinary LDPC codes under Gaussian approximation. *IEEE Trans. Inf. Th.*, 55(3):997–1015, 2009.

[20] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6(10):440–442, 2002.

[21] M. Mushkin and I. Bar-David. Capacity and coding for the Gilbert-Elliot channels. *IEEE Transactions on Information Theory*, 35(6):1277–1290, 1989.

[22] C. Poulliat, M. Fossorier, and D. Declercq. Design of regular $(2, d_c)$-LDPC codes over GF(q) using their binary images. *IEEE Transactions on Communications*, 56(10):1626–1635, october 2008.

[23] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[24] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod. High-rate quantization and transform coding with side information at the decoder. 86(11):3160–3179, 2006.

[25] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke. Design of capacity-approaching irregular Low-Density Parity-Check codes. *IEEE Trans. on Inf. Th.*, 47(2):619–637, 2001.

[26] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. on Inf. Th.*, 19(4):471–480, July 1973.

[27] V. Toto-Zarasoa, A. Roumy, and C. Guillemot. Hidden Markov Model for distributed video coding. In *Proceedings of the European Signal Processing Conference*, pages 1889–1893, 2010.

[28] V. Toto-Zarasoa, A. Roumy, and C. Guillemot. Maximum likelihood BSC parameter estimation for the Slepian-Wolf problem. *IEEE Communications Letters*, 15(2):232–234, 2011.

[29] V. Stankovic and A.D.Liveris and Z. Xiong and C.N. Georghiades. On code design for the Slepian-Wolf problem and lossless multiterminal networks. *IEEE Trans. on Inf. Th.*, 52(4):1495–1507, 2006.

[30] M. Vaezi and F. Labeau. Improved modeling of the correlation between continuous-valued sources in LDPC-based DSC. In *Conference Record of the 46th Asilomar Conference on Signal, Systems and Computers* , pages 1659–1663, 2012.

[31] D. Varodayan, D. Chen, M. Flierl, and B. Girod. Wyner–Ziv coding of video with unsupervised motion vector learning. *Signal Processing: Image Communication*, 23(5):369–378, 2008.

[32] K. Viswanatha, E. Akyol, and K. Rose. Towards optimum cost in multi-hop networks with arbitrary network demands. In *Proceedings of the International Symposium on Information Theory*, pages 1833–1837, 2010.

[33] Z. Wang, X. Li, and M. Zhao. Distributed coding of Gaussian correlated sources using non-binary LDPC. In *Congress on Image and Signal Processing*, volume 2, pages 214–218. IEEE, 2008.

[34] A.D. Wyner. The rate-distortion function for source coding with side information at the decoder\ 3-II: General sources. *Information and Control*, 38(1):60–80, 1978.

[35] A.D. Wyner and J. Ziv. The rate-distorsion function for source coding with side information at the decoder. *IEEE Transactions Information Theory*, 22(1):1–10, 1976.

[36] Z. Xiong, A.D. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5):80–94, 2004.

[37] Y. Yang, S. Cheng Z. Xiong, and W. Zhao. Wyner-Ziv coding based on TCQ and LDPC codes. *IEEE Transactions on Communications*, 57(2):376–387, 2009.