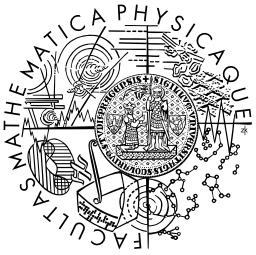


MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Filip Šohajek

Polární kódy

Katedra algebry

Vedoucí bakalářské práce: doc. Mgr. Štěpán Holub, Ph.D.

Studijní program: Obecná matematika

Praha 2024

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ....

Podpis autora

Rád bych poděkoval vedoucímu mé práce, docentu Štěpánu Holubovi, za přínosné vedení a podnětné rady při její tvorbě. Dále bych chtěl poděkovat své rodině za plnou podporu při studiu a psaní této práce.

Název práce: Polární kódy

Autor: Filip Šohajek

Katedra: Katedra algebry

Vedoucí bakalářské práce: doc. Mgr. Štěpán Holub, Ph.D., Katedra algebry

Abstrakt: Práce se zabývá strukturou a vlastnostmi polárních kódů, které jsou první deterministickou konstrukcí kapacity dosahujících kódů pro binární symetrické bezpaměťové kanály, pro kterou existuje efektivní dekódování. V práci nejprve představíme kosetové kódy, následně ukážeme algoritmus pro jejich dekódování a poté se budeme zabývat pro polární kódy klíčovým jevem polarizace kanálu. Na základě těchto výsledků pak nakonec zavedeme polární kódy.

Klíčová slova: polární kódy SC dekódování kapacity dosahující kódy teorie informace

Title: Polar codes

Author: Filip Šohajek

Department: Department of Algebra

Supervisor: doc. Mgr. Štěpán Holub, Ph.D., Department of Algebra

Abstract: This thesis outlines the structure and properties of polar coding, which is the first published deterministic construction of capacity-achieving efficiently-decodable codes for binary memoryless symmetric channels. We start by defining coset codes and a suitable decoding algorithm. Following that, we investigate the channel polarization effect, which is key to the construction. Finally, we use these results to construct polar codes themselves.

Keywords: polar codes SC decoding capacity-achieving codes information theory

Obsah

Úvod **2**

1 Základní definice **3**
1.1 Lineární algebra 3
1.2 Teorie pravděpodobnosti 4
1.3 Teorie informace 6

2 SC dekódování a kosetové kódy **11**
2.1 Symetrické kanály 14

3 Polarizace kanálu **16**
3.1 Rychlost polarizace 21

4 Polární kódy **25**

Závěr **30**

Seznam použité literatury **31**

Úvod

Základy matematické teorie přenosu informace položil Claude E. Shannon v roce 1948. Ve svém článku [7] formalizoval pojem komunikačního kanálu a následně ukázal, že každým kanálem lze vhodnou volbou kódu přenést jednoznačně definované množství informace, zároveň však toto množství nelze překročit. Bohužel známé důkazy této věty tradičně využívají pravděpodobnostního argumentu, který tyto kódy neumožňuje explicitně konstruovat, navíc v důkazu využívá metodu jejich dekódování, která je obecně výpočetně obtížně proveditelná. Tyto nedostatky daly vzniknout disciplíně samoopravných kódů, která se hledáním dobrých kódů zabývá. Pouhé nalezení dobrého kódu je však pouze část problému – pro praktické použití je nutné nalézt kód, pro který zároveň existuje výpočetně přístupný dekódovací algoritmus.

Polární kódy byly první popsány E. Arıkanem ve svém článku [1]. Jedná se o první deterministickou konstrukci, která pro daný symetrický kanál tvoří posloupnost efektivně (v log-lineárním čase v délce kódového slova) dekódovatelných lineárních kódů asymptoticky nabývajících Shannonovy kapacity. V této práci tuto konstrukci popíšeme, společně s klíčovými výsledky o nabývání kapacity a rychlosti konvergence.

Jakkoliv jsou polární kódy bez pochyby průlomovým výsledkem, je nutné zdůraznit, čím nejsou. Zejména se nejedná o první konstrukci kapacity dosahujících kódů – tu popsal D. Forney již v roce 1967, jejím nedostatkem je však dekódovací algoritmus, jehož složitost je závislá na pravděpodobnosti dekódovací chyby ϵ a to asymptoticky řádu $\frac{1}{\epsilon}$. Zde popsané výsledky také nic neříkají o vlastnostech polárních kódů konečné délky – zde popíšeme pouze asymptotické výsledky. Současné výsledky ukazují, že polární kódy dosahují při konečných délkách obdobných výsledků jako ostatní metody (například LDPC či Turbo kódy).

Zbytek práce je strukturován následovně:

- V Kapitole 1 shrneme některé potřebné definice a výsledky, které budeme využívat – zejména z teorie informace a pravděpodobnosti.
- V Kapitole 2 definujeme třídu afinních kódů do které polární kódy spadají (tzv. kosetových kódů), popíšeme pro ně dekódovací algoritmus a dokážeme odhad pravděpodobnosti chyby dekódování tohoto algoritmu.
- Kapitola 3 se zabývá jevem polarizace kanálu týkajícím se kapacit kanálů vzniklých rekurzivní aplikací určité transformace.
- V poslední Kapitole 4 využijeme polarizaci ke konstrukci polárních kódů jako kosetových kódů a ukážeme, že takto vzniklé kosetové kódy lze efektivně kódovat a dekódovat.

V celé práci budeme čerpat zejména z původních Arıkanových článků [1] [2] a z článku [5], který původní konstrukci rozšiřuje.

1. Základní definice

V následující kapitole stručně shrneme některé koncepty zejména z teorie pravděpodobnosti a teorie informace. Většinu z nich uvádíme bez důkazu – kde je důkaz netriviální, uvádíme odkaz na příslušnou literaturu.

1.1 Lineární algebra

V celém textu bude $\mathbb{F}_q$ označovat obecné q-prvkové těleso, $\mathbb{F}$ obecné konečné těleso a $\mathbb{F}_2$ dvouprvkové těleso $\mathbb{Z}_2 = \{0, 1\}$ – jeho operací sčítání budeme značit $\oplus$. Prvky aritmetických vektorových prostorů $\mathbb{F}^n$ jsou řádkové vektory $\mathbf{v} = (\mathbf{v}_1\ \mathbf{v}_2 \dots\ \mathbf{v}_n)$, nulový vektor v tomto prostoru značíme $\mathbf{0}_n$ (popřípadě pouze $\mathbf{0}$). Dále pro indexovou množinu $I = \{i_1, \dots, i_m\} \subseteq \mathbb{N}$, $i_1 \leq \dots \leq i_m$ a vektor $\mathbf{v} \in \mathbb{F}^n$ budeme rozumět $\mathbf{v}_I$ vektor $(\mathbf{v}_{i_1}\ \dots\ \mathbf{v}_{i_m}) \in \mathbb{F}^m$, pro $k \leq l$ pak $\mathbf{v}_k^l = (\mathbf{v}_k\ \mathbf{v}_{k+1}\ \dots\ \mathbf{v}_l)$.

Definice 1. ***Permutační maticí*** *příslušnou permutaci σ na množině $\{1, \dots, n\}$ rozumíme matici $P \in \{0, 1\}^{n \times n}$ splňující*

$$P_{ij} = 1 \iff \sigma(i) = j.$$

Lemma 1. *Buď $A \in \mathbb{F}^{n \times n}$ regulární. Pak existuje $P \in \mathbb{F}^{n \times n}$ permutační taková, že FP má na diagonále nenulové prvky.*

Důkaz. Z charakterizace regulárních matic máme, že $\det A \neq 0$, tedy

$$0 \neq \sum_{\sigma \in S^n} a_{1,\sigma(1)} a_{2,\sigma(2)} \dots a_{n,\sigma(n)}$$

a existuje $\sigma \in S^n$ splňující $a_{1,\sigma(1)} a_{2,\sigma(2)} \dots a_{n,\sigma(n)} \neq 0$. Uvažujme permutační matici $P \in \mathbb{F}^{n \times n}$ příslušnou permutaci σ. Pak pro všechna $1 \leq i \leq n$ platí $(FP)_{ii} = F_{i,\sigma(i)} \neq 0$.

□

Definice 2. *Buď $A \in \mathbb{F}^{m_1 \times n_1}$ a $B \in \mathbb{F}^{m_2 \times n_2}$. Pak definujeme* ***Kroneckerův součin*** *$A \otimes B$ jako matici $m_1 m_2 \times n_1 n_2$ s blokovými prvky*

$$A = \begin{pmatrix} A_{11}B & A_{12}B \dots & A_{1n_1}B \\ A_{21}B & A_{22}B \dots & A_{2n_1}B \\ \vdots & & \\ A_{m_1 1}B & A_{m_1 2}B \dots & A_{m_1 n_1} \end{pmatrix}.$$

Pro $n \in \mathbb{N}$ budeme značit $F^{\otimes n} = \underbrace{F \otimes F \otimes \dots \otimes F}_{n\text{-krát}}$

Tvrzení 2. *Buď A, B, C, D matice takových rozměrů, že jsou dobře definované součiny AC a BD. Pak*

$$(AC \otimes BD) = (A \otimes B)(C \otimes D).$$

Pro úplnost zde uvedeme základní definice z teorie lineárních kódů.

Definice 3. *Buď $q, K, N \in \mathbb{N}$, $K \leq N$ a $G \in \mathbb{F}_q^{K\times N}$ taková, že G má plnou sloupcovou hodnost. Pak definujeme* ***lineární*** *$[N, K]_q$-**kód** s generující maticí G jako kód přiřazující vstupnímu slovu $\mathbf{x} \in \mathbb{F}_q^K$ kódové slovo $\mathbf{u} = \mathbf{x}G$. Číslo N nazýváme* ***délkou*** *kódu, K pak jeho* ***dimenzí****. Množinu všech kódových slov takového kódu budeme značit jako $\mathcal{C}(G) = \{\mathbf{x}G : \mathbf{x} \in \mathbb{F}^K\}$.*

Z definice plyne, že $[N, K]_q$-kód je podprostorem vektorového prostoru $\mathbb{F}_q^N$ dimenze K. Linearita kódu je žádoucí vlastnost a teorie samoopravných kódů se lineárními kódy zabývá téměř výlučně. Klasický přístup ke konstrukci lineárních kódů hledá kódy s nejmenší vzdáleností danou následující definicí.

Definice 4. *Buď $n \in \mathbb{N}$.* ***Hammingovou vzdáleností*** *vektorů $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ rozumíme hodnotu $d_H(\mathbf{u}, \mathbf{v}) = |\{i \in \{1, \ldots, n\} : \mathbf{u}_i \neq \mathbf{v}_i\}|$. Pro $V \subseteq \mathbb{F}^n$ dále definujeme $d_H(\mathbf{u}, V) = \min_{\substack{\mathbf{v}\in V\\ \mathbf{v}\neq\mathbf{u}}} d_H(\mathbf{u}, \mathbf{v})$ a $d_H(V) = \min_{\mathbf{u}\in V} d_H(\mathbf{u}, V)$.* ***Hammingovou vahou*** *$\mathbf{v}$ rozumíme hodnotu $w_H(\mathbf{v}) = d_H(\mathbf{v}, \mathbf{0}_n)$.*
Vzdálenost lineárního kódu *s generující maticí G definujme jako $d(G) = d(\mathcal{C}(G))$.*

Vzdálenost kódu určuje jeho schopnost opravit chyby v nejhorším případě. Uvidíme, že polární kódy jsou lineárními kódy, ale jejich efektivita není založena na minimální vzdálenosti, nýbrž na dobré schopnosti opravit průměrnou chybu na daném kanálu ve spojení s vhodným dekódovacím algoritmem.

1.2 Teorie pravděpodobnosti

Pro důkaz polarizačních vět v Kapitole 3 je zapotřebí některých vět z teorie diskrétních martingálů. Znění definic a vět jsou adaptací z [4], kde lze také nalézt jejich důkazy.

Definice 5. *Buď $X : \Omega \longrightarrow \mathbb{R}^m$ a $Y : \Omega \longrightarrow \mathbb{R}^n$ náhodné vektory na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, \mathsf{P})$. Řekneme, že X je Y-měřitelná, pokud existuje měřitelná funkce $h : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ splňující $X = h(Y)$ skoro jistě.*

Definice 6. *Pro náhodné vektory $X : \Omega \longrightarrow \mathbb{R}^m$ a $Y : \Omega \longrightarrow \mathbb{R}^n$ definované na stejném pravděpodobnostním prostoru $(\Omega, \mathcal{F}, \mathsf{P})$ splňující $\mathsf{E}|X| < \infty$ definujeme podmíněnou střední hodnotu X za podmínky Y jako náhodnou veličinu $\mathsf{E}[X|Y] : \Omega \longrightarrow \mathbb{R}^m$ splňující*

1. *$\mathsf{E}[X|Y]$ je Y-měřitelná*
2. *Pro každou $B \in \mathbb{R}^m$ borelovskou platí $\mathsf{E}[\mathsf{E}[X|Y]\mathbb{1}_{[Y\in B]}] = \mathsf{E}[X\mathbb{1}_{[Y\in B]}]$*

Podmíněná střední hodnota $\mathsf{E}[X|Y]$ je nejlepší aproximací náhodné veličiny X pomocí náhodné veličiny Y v tom smyslu, že na množinách $[Y \in B]$ má stejné hodnoty integrálů jako X (podmíněnou střední hodnotu si lze představovat jako projekci X na podprostor generovaný náhodnou veličinou Y). Shrneme některé její základní vlastnosti:

Tvrzení 3. *Buď X, Y, Z reálné náhodné vektory na stejném pravděpodobnostním prostoru splňující $\mathsf{E}|X|$, $\mathsf{E}|Y| < \infty$. Pak*

1. $\mathsf{E}[\alpha X + \beta Y|Z] = \alpha\mathsf{E}[X|Z] + \beta\mathsf{E}[Y|Z]$ *pro všechna* $\alpha, \beta \in \mathbb{R}$

2. *Pokud je* X Z*-měřitelná, pak* $\mathsf{E}[XY|Z] = X\mathsf{E}[Y|Z]$

3. *Pokud jsou* X *a* Z *vzájemně nezávislé, pak* $\mathsf{E}[X|Z] = \mathsf{E}X$

Definice 7. *Buď* $(X_n)_{n\in\mathbb{N}_0}$ *a* $(A_n)_{n\in\mathbb{N}_0}$ *posloupnosti reálných náhodných veličin na stejném pravděpodobnostním prostoru splňující* $\mathsf{E}|X_n| < \infty$ *pro všechna* $n \in \mathbb{N}_0$*. Řekneme, že* $(X_n)_{n\in\mathbb{N}_0}$ *tvoří* ***(diskrétní) martingál vzhledem k*** $(A_n)_{n\in\mathbb{N}_0}$*, pokud pro všechna* $n \in \mathbb{N}_0$ *platí* $\mathsf{E}[X_{n+1}|A_1, \ldots, A_n] = X_n$*.*

Z Definice 7 plyne, že náhodná veličina X_n musí být $(A_1, \ldots, A_n)$-měřitelná. Posloupnost $(A_n)_{n\in\mathbb{N}_0}$ hraje v definici roli pomocného procesu charakterizující informaci o procesu $(X_n)_{n\in\mathbb{N}_0}$. Definiční podmínka na podmíněnou střední hodnotu pak říká, že nejlepším odhadem X_{n+1} na základě $A_1, \ldots, A_n$ je právě X_n. Pro naši konstrukci využijeme toho, že martingály za určitých podmínek konvergují k limitní náhodné veličině.

Definice 8. *Řekneme, že posloupnost* $(X_n)_{n\in\mathbb{N}_0}$ *náhodných veličin na stejném pravděpodobnostním prostoru* $(\Omega, \mathcal{A}, \mathsf{P})$ *konverguje*

- ***skoro jistě*** *k náhodné veličině* X*, pokud* $\mathsf{P}(\lim_{n\to\infty} X_n = X) = 1$*,*
- ***v*** L_1 *k náhodné veličině* X*, pokud* $\mathsf{E}|X_n|, \mathsf{E}|X| < \infty$ *a* $\lim_{n\to\infty} \mathsf{E}|X_n - X| = 0$*.*

Konvergence výše značíme jako $X_n \xrightarrow[n\to\infty]{s.j.} X_\infty$*, popř.* $X_n \xrightarrow[n\to\infty]{L_1} X_\infty$*.*

Tvrzení 4. *Pokud náhodné veličiny* $(X_n)_{n\in\mathbb{N}_0}$ *konvergují v* L_1 *k náhodné veličině* X*, pak je posloupnost* $(X_n)_{n\in\mathbb{N}_0}$ L_1*-cauchyovská, neboli*

$$\forall \epsilon > 0\ \exists n_0 \in \mathbb{N}_0\ \forall m, n \geq n_0\ \mathsf{E}|X_m - X_n| < \epsilon.$$

Důkaz. Viz [4] (*Theorem 6.25*).

□

Věta 5. *Buď* $(X_n)_{n\in\mathbb{N}_0}$ *nezáporný martingál takový, že* $\sup_{n\in\mathbb{N}_0} \mathsf{E}|X_n| < \infty$*. Pak existuje náhodná veličina* X_∞ *taková, že* $X_n \xrightarrow[n\to\infty]{s.j.} X_\infty$*,* $X_n \xrightarrow[n\to\infty]{L_1} X_\infty$ *a splňující* $\mathsf{E}X_\infty = \mathsf{E}X_0$*.*

Důkaz. Viz [4] (*Theorem 11.7*).

□

Pro důkaz rychlosti konvergence chyby dekódování budeme dále potřebovat některé varianty zákonů velkých čísel.

Tvrzení 6. *Buď* $(X_n)_{n\in\mathbb{N}_0}$ *posloupnost stejně rozdělených nezávislých náhodných veličin splňujících* $\mathsf{E}|X_1| < \infty$*. Pak pro každé* $\varepsilon > 0$ *platí*

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathsf{E}X_1\right| < \varepsilon\right) \xrightarrow[n\to\infty]{} 1$$

Důkaz. Viz [4] (*Theorem 5.17*).

□

Tvrzení 7. *Buď $X_1, \ldots, X_n$ nezávislé náhodné veličiny splňující $a \leq X_i \leq b$ pro nějaká $a < b$ reálná. Pak pro všechna $t \geq 0$ platí*

$$\mathsf{P}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathsf{E}X_i) \leq -t\right) \leq e^{-\frac{2nt^2}{(b-a)^2}}.$$

Důkaz. Je přímým důsledkem Azumovy nerovnosti ([4], *Exercise 9.2.4*).

□

Tvrzení 7 je také známé jako Hoeffdingova nerovnost.

1.3 Teorie informace

Definice a výsledky níže pochází primárně z [3].

Definice 9. *Nechť $\mathcal{X}$ je konečná množina a $\mathcal{Y}$ je libovolná množina.* ***Kanálem*** *W rozumíme množinu pravděpodobnostních rozdělení $\{\mathrm{W}_x\}_{x\in\mathcal{X}}$, všechna na množině $\mathcal{Y}$. Množinu $\mathcal{X}$ nazveme* ***vstupní abecedou*** *a množinu $\mathcal{Y}$* ***výstupní abecedou*** *(píšeme $W : \mathcal{X} \longrightarrow \mathcal{Y}$).*

Pro náhodnou veličinu X nabývající skoro jistě hodnot na $\mathcal{X}$ řekneme, že náhodná veličina Y je ***výstup*** *W* ***za vstupu*** *X, pokud W_x je podmíněné rozdělení Y za jevu $[X = x]$ pro všechna $x \in \mathcal{X}$.*

Řekneme, že kanál W je

- ***binární****, pokud je jeho vstupní abecedou $\mathbb{F}_2$,*
- ***diskrétní****, pokud je jeho výstupní abeceda konečná,*
- ***symetrický****, pokud je diskrétní a pro každé x_1, $x_2 \in \mathcal{X}$ existuje permutace $\pi_{x_1x_2}$ na množině $\mathcal{Y}$ taková, že $W(\pi_{x_1x_2}(y)|x_1) = W(y|x_2)$ pro všechna $y \in \mathcal{Y}$.*

Poznámka. Protože je výstup kanálu za dané vstupní náhodné veličiny X určen skoro jistě jednoznačně, budeme $W(X)$ rozumět náhodnou veličinu, která je výstupem W za vstupu X.

V dalším textu budeme pojmem kanál implicitně rozumět diskrétní kanál. Tímto omezením se vyhneme některým problémům definic kanálů se spojitou výstupní abecedou, zejména existenci hustot a definici kapacity. Pro diskrétní kanál W budeme psát $W(y|x) = W_x(\{y\})$, tedy pokud $Y = W(X)$, pak $W(y|x) = \mathsf{P}(Y = y|X = x)$. Diskrétní kanál je těmito hodnotami pro všechna $(x,y) \in \mathcal{X}\times\mathcal{Y}$ určen jednoznačně.

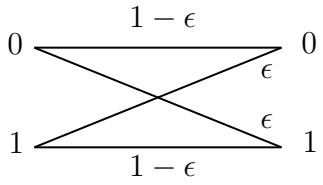
V literatuře je možné setkat se s jinou definicí kanálu, která za kanál považuje přímo pár náhodných veličin (X, Y) s daným podmíněným rozdělením. V naší definici kanál určujeme přímo podmíněným rozdělením a ne jeho realizací. Tato volba vede místy na formulačně komplikovanější tvrzení, ale umožňuje jednodušeji provádět transformace kanálu, jak uvidíme v Kapitole 3.

Příklad 1. *Buď $\epsilon > 0$.* ***Binární symetrický kanál***[1] *s pravděpodobností chyby ϵ je symetrický binární kanál $W = \mathsf{BSC}(\epsilon)$ s výstupní abecedou $\mathbb{F}_2$ definovaný jako*

$$W(0|0) = W(1|1) = 1 - \epsilon \qquad W(1|0) = W(0|1) = \epsilon,$$

[1] Je nutné rozlišovat pojmy *symetrický binární kanál* a *binární symetrický kanál* – první zmiňovaný označuje třídu binárních kanálů, které jsou symetrické a druhý specifický kanál popsaný v Příkladu.

tedy při přenosu přes W dojde k chybě s pravděpodobností ϵ.



Obrázek 1.1: Grafické znázornění kanálu $\mathsf{BSC}(\epsilon)$.

Jedná se o nejjednodušší netriviální kanál, zato je však velmi důležitým modelem – lze dokonce ukázat, že každý symetrický binární kanál (s obecnou výstupní abecedou) je konvexní kombinací právě binárních symetrických kanálů (ve smyslu konvexní kombinace hustot, resp. směsovitého rozdělení).

Definice 10. *Buď $W : \mathcal{X} \longrightarrow \mathcal{Y}$ diskrétní kanál a $n \in \mathbb{N}$. Pak definujeme **n-tou mocninu kanálu** W jako kanál $W^n : \mathcal{X}^n \longrightarrow \mathcal{Y}^n$ splňující*

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(\mathbf{y}_i|\mathbf{x}_i),$$

tedy jako kanál odpovídající n nezávislým kopiím kanálu W.

Definice 11. *Řekneme, že kanály $W : \mathcal{X} \longrightarrow \mathcal{Y}$ a $W' : \mathcal{X} \longrightarrow \mathcal{Y}'$ jsou **izomorfní**, pokud existuje bijekce $f : \mathcal{Y} \longrightarrow \mathcal{Y}'$ taková, že*

$$W(y|x) = W'(f(y)|x)$$

pro každé $x \in \mathcal{X}$ a $y \in \mathcal{Y}$.

Pokud jsou dva kanály izomorfní, mají stejné klíčové charakteristiky (viz Tvrzení 10(4)) – jedná se pouze o přejmenování výstupní abecedy. Proto budeme v případech, kdy nezáleží na konkrétních výstupních slovech brát za kanál W jeho třídu ekvivalence vůči relaci kanálového izomorfismu (zejména pak v Definici 16).

Definice 12. *Buďte X, Y, Z diskrétní náhodné veličiny. Pak definujeme*

- ***entropii** X jako*

$$H(X) = -\sum_{\mathsf{P}(X=x)>0} \mathsf{P}(X = x) \cdot \log_2 \mathsf{P}(X = x),$$

- ***podmíněnou entropii** X **při** Y jako*

$$H(X|Y) = -\sum_{\mathsf{P}(X=x,Y=y)>0} \mathsf{P}(X = x, Y = y) \cdot \log_2 \mathsf{P}(X = x|Y = y),$$

- ***vzájemnou informaci** X **a** Y jako*

$$I(X : Y) = H(X) - H(X|Y).$$

Tvrzení 8. *Buď $X_1, X_2, \ldots, X_n, Y$ diskrétní náhodné veličiny. Pak*

$$H(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|X_{i-1}, X_{i-2}, \ldots, X_1, Y).$$

Důkaz. Viz [3] (*Theorem 2.2.1*).

□

Definice 13. *Buď W kanál. Pak definujeme jeho*

- ***Kapacitu*** *jako $C(W) = \sup_X I(X : W(X))$, kde supremum uvažujeme přes všechna rozdělení náhodné veličiny X,*
- ***Symetrickou kapacitu*** *$I(W) = I(X : W(X))$, kde X je náhodná veličina s rovnoměrným rozdělením na $\mathcal{X}$.*

Pokud je W binární, pak dále definujeme ***Bhattacharyyův parametr***

$$Z(W) = \mathsf{E}\sqrt{\frac{W(Y|X \oplus 1)}{W(Y|X)}},$$

kde X je náhodná veličina s rovnoměrným rozdělením na $\mathbb{F}_2$ a $Y = W(X)$.

Kapacita udává míru informace mezi vstupem a výstupem kanálu. Tento výrok formálně vyjadřuje známá Shannonova věta o nabývání kapacity, která říká, že se kapacitě můžeme vhodnou volbou kódu přiblížit libovolně blízko.

Věta 9. *Buď $W : \mathcal{X} \longrightarrow \mathcal{Y}$ (diskrétní) kanál. Pak pro každé $0 < R < C(W)$ existuje posloupnost množin kódových slov $\mathcal{C}_n \subseteq \mathcal{X}^n$ velikostí $|\mathcal{C}_n| = |\mathcal{X}|^{nR}$ a dekódovacích zobrazení $d_n : \mathcal{Y}^n \to \mathcal{C}_n$ takových, že pokud je $\mathbf{X}^{(n)}$ náhodná veličina s rovnoměrným rozdělením na $\mathcal{C}_n$, pak*

$$\mathsf{P}(d_n(W^n(\mathbf{X}^{(n)})) \neq \mathbf{X}^{(n)}) \xrightarrow[n\to\infty]{} 0.$$

Důkaz. Viz [3] (*Theorem 7.7.1*).

□

Klasické důkazy Shannonovy věty probíhají volbou kódu $\mathcal{C}_n$ tak, aby rozdělení složek kódových slov odpovídalo rozdělení vstupu maximalizujícímu vzájemnou informaci v definici $C(W)$. Funkce d_n se pak realizuje obvykle jako tzv. sdruženě typické dekódování, popřípadě tzv. MAP (maximální aposteriorní) dekódování. Je důležité si uvědomit, že kvalita kódu je nutně spjatá s použitým dekódovacím algoritmem. A právě v tom spočívá problém v aplikovatelnosti Shannonovy věty na praktickou konstrukci dobrých kódů – jak sdruženě typické dekódování, tak MAP dekódování nelze obecně realizovat efektivně ve smyslu časové složitosti v n (jedná se o NP-těžké problémy). Polární kódy tento problém řeší pro symetrické binární kanály.

Níže shrneme některé vlastnosti $I(W)$, $Z(W)$ a $C(W)$, které budeme v textu používat.

Tvrzení 10. *Buď W binární kanál. Pak*

1. *$I(W) = 1 - H(X|W(X))$, kde X je náhodná veličina s rovnoměrným rozdělením na $\mathbb{F}_2$*
2. *$0 \leq I(W) \leq 1$ a $0 \leq Z(W) \leq 1$*
3. *$I(W)^2 + Z(W)^2 \leq 1$ a $I(W) \geq \log \frac{2}{1+Z(W)}$*

4. *Pokud je W' izomorfní kanálu W, pak $I(W) = I(W')$ a $Z(W) = Z(W')$*

5. *Pro $n \in \mathbb{N}$ je $C(W^n) = nC(W)$*

Důkaz. Body 1, 2, 4 a 5 plynou jednoduše z definic. Bod 3 je dokázán v [1]. □

Symetrická kapacita $I(W)$ odpovídá kapacitě kanálu W, pokud je W symetrický – na symetrických kanálech se kapacity nabývá rovnoměrně rozděleným vstupem. To je intuitivně zřejmé i z definice symetrického kanálu, kde změna vstupního symbolu odpovídá pouze fixní permutaci výstupních symbolů – kanál se proto pro žádný ze vstupních symbolů nemůže chovat odlišně.

Bhattacharyyův parametr je speciálním případem tzv. Bhattacharyyovy vzdálenosti dvou pravděpodobnostních rozdělení, zde konkrétně W_0 a W_1 z definice kanálu W (tj. rozdělení $W(X)$ za $[X = 0]$, respektive $[X = 1]$). Malé hodnoty Bhattacharyyovy vzdálenosti odpovídají vzdáleným rozdělením a naopak. Mohli bychom tedy očekávat, že méně chybovým kanálům budou příslušet nízké hodnoty $Z(W)$ (podmíněná rozdělení jsou od sebe dobře rozlišitelná) a analogicky naopak u více chybových kanálů. Tuto domněnku potvrzuje následující tvrzení.

Tvrzení 11. *Buď W binární kanál, X náhodná veličina s rovnoměrným rozdělením na $\mathbb{F}_2$ a $Y = W(X)$. Definujme ML dekódované slovo jako $\hat{X}_{ML} = \arg\max_{x \in \mathbb{F}_2} W(Y|x)$. Pak platí*

$$\mathsf{P}(\hat{X}_{ML} \neq X) \leq Z(W).$$

Důkaz. Pro jev $[\hat{X}_{ML} \neq X]$ platí

$$[\hat{X}_{ML} \neq X] = [W(Y|X \oplus 1) \geq W(Y|X)] = \left[\frac{W(Y|X \oplus 1)}{W(Y|X)} \geq 1\right]$$

Z toho dostáváme:

$$\mathsf{P}(\hat{X}_{ML} \neq X) = \mathsf{E}\mathbb{1}_{\left[\frac{W(Y|X\oplus 1)}{W(Y|X)} \geq 1\right]} \leq \mathsf{E}\sqrt{\frac{W(Y|X \oplus 1)}{W(Y|X)}} = Z(W),$$

neboť X má z předpokladu rovnoměrné rozdělení na $\mathbb{F}_2$. □

Za předpokladu rovnoměrně rozložených vstupů tedy odhaduje shora Bhattacharyův parametr pravděpodobnost dekódovací chyby při tzv. ML dekódování. Této vlastnosti využijeme při konstrukci polárních kódů a budeme se z nezávislých realizací daného kanálu vytvořit transformací kanály s nízkou hodnotou $Z(W)$.

Tato transformace se zakládá na provázání vstupů několika realizací kanálů lineárním zobrazením. Budeme tedy potřebovat odhadnout, jak se při lineární transformaci vstupů mění parametry vzniklých kanálů. To ukazuje Lemma 13, které je důsledkem výsledku z [10] známého jako *Mrs. Gerbers' Lemma*.

Lemma 12. *Buď $h(a) = -a \log a - (1-a) \log (1-a)$ pro $a \notin \{0,1\}$ a $h(0) = h(1) = 0$. Nechť h^{-1} je inverzní funkce k h na $0 \leq a \leq \frac{1}{2}$. Pak pro $0 < p_0 \leq \frac{1}{2}$ je funkce $f(u) = h(p_0 * h^{-1}(u))$ striktně konvexní na $0 \leq u \leq 1$, kde $x * y = x(1-y) + (1-x)y$.*

Důkaz. Důkaz lze nalézt v [10]. □

Následující lemma a jeho důkaz jsou adaptací výsledků z článku [8].

Lemma 13. *Buď X_1, X_2 nezávislé náhodné veličiny s nedegenerovaným rozdělením na $\mathbb{F}_2$. Pokud jsou Y_1, Y_2 náhodné veličiny takové, že (X_1, Y_1) a (X_2, Y_2) jsou nezávislé náhodné vektory, pak*

$$H(X_1 \oplus X_2|Y_1, Y_2) \geq h(h^{-1}(H(X_1|Y_1)) * h^{-1}(H(X_2|Y_2))).$$

Pokud dále $H(X_1|Y_1) = H(X_2|Y_2)$, pak platí

$$\exists \delta \in (0,1) : H(X_1|Y_1) \in (\delta, 1-\delta) \iff \exists \eta_\delta : H(X_1 \oplus X_2|Y_1, Y_2) - H(X_1|Y_1) \geq \eta_\delta.$$

Důkaz. Pro náhodnou veličinu X na $\mathbb{F}_2$ a libovolnou náhodnou veličinu Y platí z definice h, že $H(X|Y=y) = h(\mathsf{P}(X=1|Y=y))$. Tedy máme

$$\begin{aligned}
H(X_1 \oplus X_2|Y_1, Y_2) &= \sum_{y_{1,2}} \mathsf{P}(Y_1 = y_1, Y_2 = y_2) H(X_1 \oplus X_2|Y_1 = y_1, Y_2 = y_2) \\
&= \sum_{y_{1,2}} \mathsf{P}(Y_1 = y_1, Y_2 = y_2) h(\mathsf{P}(X_1 \oplus X_2 = 1|Y_1 = y_1, Y_2 = y_2)) \\
&= \sum_{y_{1,2}} \mathsf{P}(Y_1 = y_1, Y_2 = y_2) h(\mathsf{P}(X_1 = 1|Y_1 = y_1) \\
&\qquad * \mathsf{P}(X_2 = 1|Y_2 = y_2)) \\
&= \sum_{y_{1,2}} \mathsf{P}(Y_1 = y_1, Y_2 = y_2) h(h^{-1}(H(X_1|Y_1 = y_1)) \\
&\qquad * h^{-1}(H(X_2|Y_2 = y_2))) \\
&\geq \sum_{y_2} \mathsf{P}(Y_2 = y_2) h(h^{-1}(\sum_{y_1} \mathsf{P}(Y_1 = y_1) H(X_1|Y_1 = y_1)) \\
&\qquad * h^{-1}(H(X_2|Y_2 = y_2))) \\
&\geq h(h^{-1}(H(X_1|Y_1)) * h^{-1}(\sum_{y_2} \mathsf{P}(Y_2 = y_2) H(X_2|Y_2 = y_2))) \\
&= h(h^{-1}(H(X_1|Y_1)) * h^{-1}(H(X_2|Y_2))),
\end{aligned}$$

kde v nerovnostech jsme použili Lemma 12 a z něj se také rovností nabývá právě tehdy, když příslušně platí $H(X_1|Y_1) \in \{0,1\}$, respektive $H(X_2|Y_2) \in \{0,1\}$.

Druhá část tvrzení pak plyne ze spojitosti h a h^{-1}, společně s

$$h(h^{-1}(H(X_1|Y_1)) * h^{-1}(H(X_1|Y_1))) \geq h(h^{-1}(H(X_1|Y_1))) = H(X_1|Y_1),$$

kde rovnosti se nabývá právě tehdy, když $H(X_1|Y_1) \in \{0,1\}$. □

2. SC dekódování a kosetové kódy

Hledání dobrých lineárních kódů je kompromisem – na jednu stranu se snažíme získat kód s co největší dimenzí, na stranu druhou však kódy vysoké dimenze mají nízkou vzdálenost a jsou tedy špatně dekódovatelné (pravděpodobnost neopravitelné chyby je vysoká). Zvyšování délky kódového slova pak umožňuje dimenzi zvyšovat s menším dopadem na vzdálenost kódu.

Tuto konstrukci si můžeme také představit následovně – začneme s lineárním kódem, jehož generující matice je regulární. Takový kód má nejlepší možný kódový poměr (podíl dimenze a délky), zato je ale špatně dekódovatelný, protože jeho vzdálenost je rovna jedné. Výše zmíněný kompromis nyní můžeme realizovat tak, že z této generující matice odstraníme některé řádky, čímž získáme kód menší dimenze, ale potenciálně lepší vzdálenosti. Jiný pohled na tento postup spočívá v tom, že používáme původní regulární generující matici, ale zafixujeme hodnoty některých pozic vstupního slova.

Definice 14. *Buď $K \leq N$ přirozená a $G \in \mathbb{F}_2^{N\times N}$ regulární. Nechť $\mathcal{A}$ je libovolná podmnožina $\{1 \ldots, N\}$ velikosti K a $\mathbf{u}_{\mathcal{A}^c}$ fixní.* ***Kosetovým kódem s parametry*** *$(G, N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ rozumíme kód, který vstupnímu slovu $u_{\mathcal{A}}$ přiřadí kódové slovo*

$$\mathbf{u}_{\mathcal{A}}G_{\mathcal{A}} \oplus \mathbf{u}_{\mathcal{A}^c}G_{\mathcal{A}^c},$$

kde $G_{\mathcal{A}}$ (respektive $G_{\mathcal{A}^c}$) je matice vzniklá výběrem řádků z G s indexy z $\mathcal{A}$ (respektive $\mathcal{A}^c$). Množinu $\mathcal{A}$ nazýváme ***informační množinou*** *kódu.*

Kosetový kód je obecně afinním kódem (tj. lineárním kódem s fixním posunutím kódových slov) – pokud je však $\mathbf{u}_{\mathcal{A}^c} = \mathbf{0}_{N-K}$, pak je toto posunutí nulové a příslušný kód je lineární.

Výhody definice výše začnou být patrné při spojení s vhodným dekódovacím algoritmem. Za předpokladu rovnoměrného rozdělení vstupních slov, je pro symetrický kanál W a lineární kód s generující maticí G optimálním dekódovacím pravidlem (tj. pravidlem minimalizující pravděpodobnost chyby) maximálně věrohodné dekódování (zkráceně ML), které přijatému slovu $\mathbf{y}$ vzniklého přenosem slova $\mathbf{u}G$ přiřadí slovo

$$\hat{\mathbf{u}}_{\mathrm{ML}} = \arg\max_{\mathbf{u}\in\mathbb{F}_2^n} W(\mathbf{y}|\mathbf{u}G).$$

Algoritmus pro dekódování kosetových kódů níže (původně vyložený E. Arıkanem v [1]) je ML dekódování podobný, při dekódování ale postupuje po jednotlivých pozicích vstupního slova. V každém kroku předpokládá, že hodnoty na předchozích pozicích byly dekódovány správně a s tímto předpokladem dekóduje aktuální pozici pomocí ML dekódování na kanálu odpovídajícímu příslušnému podmíněnému rozdělení. V pozicích mimo informační množinu pak využívá znalosti zafixovaných hodnot $\mathbf{u}_{\mathcal{A}^c}$. Lze snadno vidět, že pro $\mathcal{A}^c = \emptyset$ se tento algoritmus redukuje na ML dekódování.

Definice 15. *Buď $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ binární kanál. Pro matici $G \in \mathbb{F}_2^{N\times N}$ definujeme* ***kódovací kanál*** *$W_G : \mathbb{F}_2^N \longrightarrow \mathcal{Y}^N$ předpisem*

$$W_G(\mathbf{y}|\mathbf{u}) = W^N(\mathbf{y}|\mathbf{u}G)$$

a pro $i \in \{1, \ldots, N\}$ ***i*-tý dekódovací kanál** $W_G^{(i)} : \mathbb{F}_2 \longrightarrow \mathcal{Y}^N \times \mathbb{F}_2^{i-1}$ *předpisem*

$$W_G^{(i)}(\mathbf{y}, \mathbf{u}_1^{i-1} | \mathbf{u}_i) = \frac{1}{2^{N-i}} \sum_{\mathbf{u}_{i+1}^N} W_N(\mathbf{y}|\mathbf{u}).$$

Algorithm 1 SC dekódování $(G, N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$-kosetového kódu

Vstup: Přijaté slovo $\mathbf{y} \in \mathbb{F}_2^N$
Výstup: Dekódované slovo $\hat{\mathbf{u}} \in \mathbb{F}_2^N$

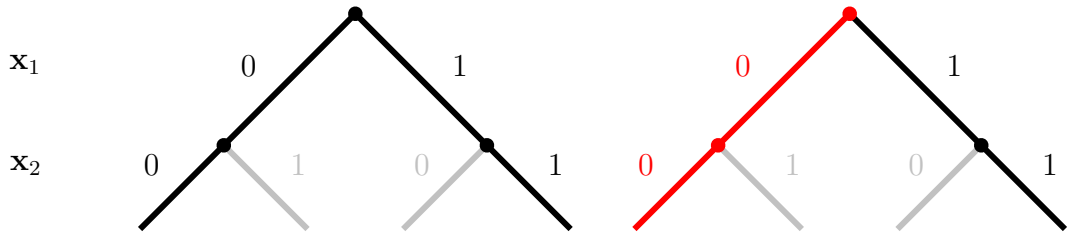
```
for i = 1, ..., N do
    if i ∉ A then
        û_i ← u_i
    else
        if W_G^(i)(y, û_1^{i-1}|0) ≥ W_G^(i)(y, û_1^{i-1}|1) then
            û_i ← 0
        else
            û_i ← 1
        end if
    end if
end for
return û
```

Ačkoliv $W_G^{(i)}$ výše jsou kanály, není vhodné je chápat jako modely přenosu informace reprezentující například nějaké fyzické médium, neboť z definice jejich výstup obsahuje vstupy předchozích kanálů. Než o kanály se tady spíše jedná o podmíněná rozdělení charakterizující stav SC dekódování v i-tém kroku.

Příklad 2. *Uvažujme* $(G, 2, 1, \{2\}, 0)$*-kosetový kód s generující maticí* $G = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ *a přenos přes symetrický binární kanál* W*. Předpokládejme, že jsme zakódovali slovo* $\mathbf{u}_{\mathcal{A}} = (0)$*, výsledné slovo poslali přes kanál* W *a výstupem kanálu bylo slovo* $\mathbf{y} = (1\ 0)$*. Konkrétně je* $\mathbf{y}$ *výstupem kanálu* W *za vstupu* $\mathbf{x} = \mathbf{u}G = (\mathbf{u}_1 \oplus \mathbf{u}_2\ \mathbf{u}_2)$*. Uvažujme nyní proces SC dekódování na tomto kanálu:*

- *Jelikož je* $1 \in \mathcal{A}^c$*, pak známe hodnotu* $\mathbf{u}_1 = 0$ *a tedy známe součet* $\mathbf{x}_1 \oplus \mathbf{x}_2 = \hat{\mathbf{u}}_1 = 0$*.*
- *Hodnotu* $\mathbf{u}_2$ *neznáme, jelikož* $2 \in \mathcal{A}$*, ale ze znalosti součtu vstupů víme, že obě kopie kanálu* W *měly stejný vstup, tedy* $\mathbf{x}_1 = \mathbf{x}_2$*. Jelikož ale* $\mathbf{y}_1 \neq \mathbf{y}_2$*, pak víme, že při přenosu nastala chyba – původní vstup ale neumíme určit, protože existují dvě stejně pravděpodobné varianty (*W *je symetrický) – buď* $\mathbf{x} = (0\ 0)$*, nebo* $\mathbf{x} = (1\ 1)$*. Algoritmus 1 v tomto případě rozhoduje ve prospěch* $\hat{\mathbf{u}}_2 = 0$*.*

Postup dekódování si lze představit také jako hledání cesty v rozhodovacím stromu na hodnotách $\mathbf{x}_i$ *– takto je znázorněn na Obrázku 2.1. Každá hrana ve stromu má stejnou váhu, neboť použité ML dekódování předpokládá, že rozdělení vstupních bitů je rovnoměrné. Provázání* $\mathbf{x}_i$ *podmínkami na hodnoty bitů* $\mathbf{u}_j$ *mimo informační množinu pak umožňuje efektivní prořezávání stromu.*



Obrázek 2.1: Postup SC dekódování kódu z příkladu znázorněný na hodnotách $\mathbf{x}$. Šedě jsou vyznačeny větve nesplňující podmínku $\mathbf{x}_1 \oplus \mathbf{x}_2 = 0$. Červeně je vyznačena trajektorie dekódování.

SC dekódování kosetových kódů není obecně optimální právě kvůli zafixování bitů mimo informační množinu. Nabízí se otázka, zda touto neoptimalitou získáváme nějakou žádoucí vlastnost. Jednou z těch potenciálně zajímavých je naděje na snížení výpočetní složitosti dekódování – ML dekódování je pro obecné lineární kódy NP-těžkým problémem. Z algoritmu výše je ale patrné, že pokud by se podařilo vhodnou volbou matice G zefektivnit výpočet rozdělení dekódovacích kanálů $W_G^{(i)}$ (například zavedením rekurzivní struktury), pak by takové zefektivnění mohlo vést na lepší časovou složitost algoritmu.

Druhou výhodou SC dekódování je možnost lépe kontrolovat pravděpodobnost dekódovací chyby, jak ukazuje následující tvrzení převzaté z [1]:

Tvrzení 14. *Algoritmus 1 vrátí správné vstupní slovo s průměrnou pravděpodobností chyby přes všechny volby* $\mathbf{u}_{\mathcal{A}^c}$ *nejvýše* $\sum_{i\in\mathcal{A}} Z(W_G^{(i)})$.

Důkaz. Buď $\mathbf{U}$ náhodný vektor s nezávislými binárními složkami a rovnoměrným rozdělením na $\mathbb{F}_2^N$. Nechť $\mathbf{Y}$ je náhodná veličina taková, že $\mathbf{Y} = W_G(\mathbf{U})$. Pak lze výsledek SC dekódování značit jako $\hat{\mathbf{U}}$, kde

$$\hat{\mathbf{U}}_i = \begin{cases} \mathbf{U}_i & i \in \mathcal{A}^c \\ \arg\max_{u\in\mathbb{F}_2} W_G^{(i)}(\mathbf{Y}, \hat{\mathbf{U}}_1^{i-1}|u) & i \in \mathcal{A} \end{cases}$$

Jev, že nastala chyba dekódování můžeme zapsat jako $\mathcal{B} = [\mathbf{U} \neq \hat{\mathbf{U}}]$. Definujme jev $\mathcal{E}_i = [\hat{\mathbf{U}}_1^{i-1} = \mathbf{U}_1^{i-1}] \cap [\hat{\mathbf{U}}_i \neq \mathbf{U}_i]$ že chyba dekódování nastala poprvé v i-tém kroku. Pak platí

$$\mathcal{B} = \bigcup_{i=1}^{N} \mathcal{E}_i = \bigcup_{i\in\mathcal{A}} \mathcal{E}_i \subseteq \bigcup_{i\in\mathcal{A}} [\hat{\mathbf{U}}_i \neq \mathbf{U}_i],$$

kde jsme využili toho, že $\hat{\mathbf{U}}_i = \mathbf{U}_i$ pro všechna $i \in \mathcal{A}^c$. Pravděpodobnost $\mathcal{B}$ pak pomocí Tvrzení 11 odhadneme shora jako

$$\mathsf{P}(\mathcal{B}) \leq \sum_{i\in\mathcal{A}} \mathsf{P}(\hat{\mathbf{U}}_i \neq \mathbf{U}_i) \leq \sum_{i\in\mathcal{A}} Z(W_G^{(i)}).$$

□

Tvrzení výše realizuje kompromis zmíněný v úvodu kapitoly – přidáváním indexů do informační množiny zvyšujeme dimenzi kódu (a tedy množství přenášené informace), ale zhoršujeme horní odhad pravděpodobnosti dekódovací chyby. Klíčové je, jak jsou hodnoty $Z(W_G^{(i)})$ rozděleny – pokud jsou všechny blízko sebe, není

možné dosáhnout tak optimálního kompromisu jako když jsou hodnoty rozděleny na relativně menší a relativně větší.

Smyslem konstrukce polárních kódů je nalezení takové generující matice G, aby příslušné dekódovací kanály měly hodnoty $Z(W_G^{(i)})$ limitně blízko buď 0, nebo 1. Klíčovým pak je, že poměr bezchybových kanálů je asymptoticky roven $I(W)$, čili výběrem vhodné $I(W)$-části indexů obdržíme kód dosahující symetrické kapacity.

2.1 Symetrické kanály

Věta 14 dává pouze průměrný odhad chyby přes všechny volby hodnot $\mathbf{u}_{\mathcal{A}^c}$ bitů mimo informační množinu. Jak však ukazuje následující tvrzení, pro symetrické kanály na této volbě nezáleží (což potvrzuje i intuice, jelikož nenulové $\mathbf{u}_{\mathcal{A}^c}$ odpovídá fixnímu posunutí kódu a u symetrických kanálů jsou všechny vstupní symboly v určitém smyslu rovnocenné). Budeme postupovat stejně jako v [1]. První ukážeme, že dekódovací kanály vzniklé ze symetrického kanálu splňují o něco silnější podmínku symetrie.

Lemma 15. *Nechť W je symetrický binární kanál a $G \in \mathbb{F}_2^{N\times N}$. Pak pro každé $\mathbf{u} \in \mathbb{F}_2^N$ existuje $\pi_{\mathbf{u}}$ permutace na množině $\mathcal{Y}^N$ splňující*

$$W_G^{(i)}(\mathbf{y}, \mathbf{u}_1^{i-1}|\mathbf{u}_i) = W_G^{(i)}(\pi_{\mathbf{u}}(\mathbf{y}_1^N), \mathbf{0}_{i-1}|0)$$

pro všechna $i \in \{1, \dots, N\}$.

Důkaz. Označme výstupní abecedu kanálu W jako $\mathcal{Y}$ a buď σ permutace na $\mathcal{Y}$ splňující $W(y|1) = W(\sigma(y)|0)$ pro všechna $y \in \mathcal{Y}$ (taková existuje ze symetrie W). Pak pro libovolné $\mathbf{a}, \mathbf{u} \in \mathbb{F}_2^N$ a $G \in \mathbb{F}_2^{N\times N}$ regulární z definice mocniny kanálu dostáváme

$$\begin{aligned} W_G(\mathbf{y}|\mathbf{u}) &= W^N(\mathbf{y}|\mathbf{u}G) = W^N(\pi_{\mathbf{u}}(\mathbf{y})|(\mathbf{u}\oplus\mathbf{a})G) \\ &= W_G(\pi_{\mathbf{u}}(\mathbf{y})|\mathbf{u}\oplus\mathbf{a}), \end{aligned}$$

kde $\pi_{\mathbf{a}}$ je permutace na množině $\mathcal{Y}^N$ definovaná předpisem

$$\pi_{\mathbf{a}}(\mathbf{y})_j = \begin{cases} \sigma(\mathbf{y}_j) & (\mathbf{a}G)_j = 1, \\ \mathbf{y}_j & (\mathbf{a}G)_j = 0. \end{cases}$$

Pro $i \in \{1, \dots, N\}$ pak platí

$$\begin{aligned} W_G^{(i)}(\mathbf{y}, \mathbf{u}_1^{i-1}|\mathbf{u}_i) &= \frac{1}{2^{N-i}} \sum_{\mathbf{v}} W_G(\mathbf{y}|(\mathbf{u}_1^i\ \mathbf{v})) \\ &= \frac{1}{2^{N-i}} \sum_{\mathbf{v}} W_G(\pi_{\mathbf{u}}(\mathbf{y})|(\mathbf{0}_{i-1}\ \mathbf{v}\oplus\mathbf{u}_{i+1}^N)) \\ &= \frac{1}{2^{N-i}} \sum_{\mathbf{v}} W_G(\pi_{\mathbf{u}}(\mathbf{y})|(\mathbf{0}_{i-1}\ \mathbf{v})) \\ &= W_G^{(i)}(\pi_{\mathbf{u}}(\mathbf{y}), \mathbf{0}_i|0). \end{aligned}$$

□

Dokázáná vlastnost pak přímo vede na důkaz následujícího tvrzení.

Tvrzení 16. *Nechť W je symetrický binární kanál. Pak Algoritmus 1 vrátí správné vstupní slovo s pravděpodobností chyby nejvýše $\sum_{i\in\mathcal{A}} Z(W_G^{(i)})$.*

Důkaz. Buď $\mathcal{E}_i$ jako v důkazu Věty 14 a $\mathbf{u} \in \mathbb{F}_2^N$ libovolné. Ukážeme, že jevy $\mathcal{E}_i$ a $[\mathbf{U} = \mathbf{u}]$ jsou nezávislé. Skutečně, buď $\pi_{\mathbf{u}}$ permutace z Lemmatu 15. Pak platí

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_i|\mathbf{U}=\mathbf{u}) &= \sum_{\mathbf{y}} \mathsf{P}(\mathcal{E}_i, \mathbf{Y}=\mathbf{y}|\mathbf{U}=\mathbf{u}) \\
&= \sum_{\mathbf{y}} \mathsf{P}(W_G^{(i)}(\mathbf{Y}, \mathbf{U}_1^{i-1}|\mathbf{U}_i \oplus 1) \geq W_G^{(i)}(\mathbf{Y}, \mathbf{U}_1^{i-1}|\mathbf{U}_i), \mathbf{Y}=\mathbf{y}|\mathbf{U}=\mathbf{u}) \\
&= \sum_{\mathbf{y}} \mathsf{P}(W_G^{(i)}(\pi_{\mathbf{u}}(\mathbf{y}), \mathbf{0}_1^{i-1}|1) \geq W_G^{(i)}(\pi_{\mathbf{u}}(\mathbf{y}), \mathbf{0}_1^{i-1}|0), \mathbf{Y}=\mathbf{y}|\mathbf{U}=\mathbf{u}) \\
&= \sum_{\mathbf{y}} \mathsf{P}(W_G^{(i)}(\pi_{\mathbf{u}}(\mathbf{y}), \mathbf{0}_1^{i-1}|1) \geq W_G^{(i)}(\pi_{\mathbf{u}}(\mathbf{y}), \mathbf{0}_1^{i-1}|0), \mathbf{Y}=\pi_{\mathbf{u}}(\mathbf{y})|\mathbf{U}=\mathbf{0}_N) \\
&= \mathsf{P}(\mathcal{E}_i|\mathbf{U}=\mathbf{0}_N),
\end{aligned}$$

kde ve třetí a čtvrté rovnosti využíváme Lemma 15 společně s tím, že $(\mathbf{Y}_i, \mathbf{U}_1^{i-1})$ je výstupem $W_G^{(i)}$ za vstupu $\mathbf{U}_i$. Z nezávislosti $\mathcal{E}_i$ a $[\mathbf{U} = \mathbf{u}]$ a z argumentů v důkazu Věty 14 plyne pro pravděpodobnost jevu $\mathcal{E} = \cup_{i=1}^N \mathcal{E}_i$ chyby dekódování za fixního $\mathbf{u}_{\mathcal{A}^c}$

$$\mathsf{P}(\mathcal{E}|\mathbf{U}_{\mathcal{A}^c}=\mathbf{u}_{\mathcal{A}^c}) \leq \sum_{i\in\mathcal{A}} \mathsf{P}(\mathcal{E}_i|\mathbf{U}_{\mathcal{A}^c}=\mathbf{u}_{\mathcal{A}^c}) = \sum_{i\in\mathcal{A}} \mathsf{P}(\mathcal{E}_i) \leq \sum_{i\in\mathcal{A}} Z(W_G^{(i)}).$$

□

3. Polarizace kanálu

V předchozí kapitole jsme pomocí kosetových kódů a SC dekódování redukovali problém hledání dobrých lineárních kódů na volbu matice G takové, aby výsledné dekódovací kanály $W_G^{(i)}$ měly co nejmenší hodnotu Bhattacharyyova parametru.

Nyní se zaměříme právě na dekódovací kanály a budeme studovat vlastnosti transformace, kterou tyto kanály vznikají. Konkrétně ukážeme, že opakované použití této transformace často vede na kanály s nízkým Bhattacharyyovým parametrem, což v další kapitole využijeme ke konstrukci kapacity dosahujících kódů. Myšlenky vět a důkazů níže pochází z [5], který rozšiřuje původní Arıkanovy články [1] a [2] v nichž jsou výsledky níže ukázány pouze pro specifickou volbu matice $G = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$.

Definice 16. *Buď $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ kanál a $F \in \mathbb{F}_2^{\ell\times\ell}$ regulární. Pak řekneme, že $\left(W^{(1)}, \ldots, W^{(\ell)}\right)$ je **F-transformací kanálu** W, pokud existuje bijekce f na množině $\mathcal{Y}^\ell$ taková, že pro všechna $i \in \{1, \ldots, \ell\}$ je $W^{(i)} : \mathbb{F}_2 \longrightarrow \mathcal{Y}^\ell \times \mathbb{F}_2^{i-1}$ binární kanál splňující*

$$W^{(i)}(f(\mathbf{y}), \mathbf{u}_1^{i-1}|\mathbf{u}_i) = \frac{1}{2^{\ell-i}} \sum_{\mathbf{u}_{i+1}^\ell} W^\ell(\mathbf{y}|\mathbf{u}F).$$

*Kanál $W^{(i)}$ nazveme **i-tou F-transformací kanálu** W.*

Poznámka. Jak jsme poznamenali dříve (poznámka pod Definicí 11), kanálem $W^{(i)}$ rozumíme v kontextech, kde nezáleží na konkrétních výstupních slovech, třídu ekvivalence všech kanálů izomorfních $W^{(i)}$. Zejména budeme v důkazech níže bez újmy na obecnosti předpokládat $f = \mathrm{id}$.

Povšimněme si, že F-transformace transformuje l nezávislých kopií kanálu W – důležitou vlastností je, že výsledné kanály jsou opět binární, ale jsou navzájem provázané. Právě tímto provázáním se mění hodnoty parametrů kanálu, z nichž jsou pro nás zajímavé zejména Bhattacharyyův parametr a symetrická kapacita. Mohli bychom očekávat, že jelikož je F regulární, pak bude F-transformace celkovou symetrickou kapacitu zachovávat.

Tvrzení 17. *Buď $(W^{(1)}, \ldots, W^{(\ell)})$ F-transformací binárního kanálu W, kde $F \in \mathbb{F}_2^{\ell\times\ell}$ je regulární. Pak $\sum_{i=1}^{\ell} I(W^{(i)}) = \ell I(W)$.*

Důkaz. Buď $\mathbf{X} \in \mathbb{F}_2^\ell$ náhodný vektor s nezávislými rovnoměrně rozdělenými binárními složkami a $\mathbf{Y} \in \mathbb{F}_2^\ell$ výstupem kanálu W^ℓ za vstupu $\mathbf{X}$. Pak pro $\mathbf{U} = \mathbf{X}F^{-1}$ je $(f(\mathbf{Y}), \mathbf{U}_1^{i-1})$ výstupem kanálu $W^{(i)}$ za vstupu $\mathbf{U}_i$. Z poznámky pod Definicí 16 budeme bez újmy na obecnosti předpokládat, že $f = \mathrm{id}$.

Z Tvrzení 8 dostáváme přímo

$$\begin{aligned}
\sum_{i=1}^{\ell} I(W^{(i)}) &= \sum_{i=1}^{\ell} (1 - H(\mathbf{U}_i|\mathbf{U}_1^{i-1}, \mathbf{Y})) \\
&= \ell - H(\mathbf{U}|\mathbf{Y}) \\
&= \ell - H(\mathbf{X}|\mathbf{Y}) \\
&= \ell - \ell H(\mathbf{X}_1|\mathbf{Y}_1) \\
&= \ell I(W),
\end{aligned}$$

kde druhá rovnost plyne z Tvrzení 8, třetí z regularity F a čtvrtá z Tvrzení 10(5). □

Jak se symetrická kapacita mění na jednotlivé hodnoty $I(W^{(i)})$ částečně objasňuje následující klíčové tvrzení, které je adaptací Lemmatu 3 z [5].

Tvrzení 18. *Buď $(W^{(1)}, \ldots, W^{(\ell)})$ F-transformací binárního kanálu W, kde $\ell \geq 2$ je přirozené a $F \in \mathbb{F}_2^{\ell \times \ell}$ regulární. Pak nastane právě jedna z následujících možností:*

1. *Existuje $P \in \mathbb{F}_2^{\ell \times \ell}$ permutační taková, že FP je horní trojúhelníková. Potom pro všechna $i \in \{1, \ldots, \ell\}$ platí $I(W^{(i)}) = I(W)$.*
2. *FP není horní trojúhelníková pro žádnou $P \in \mathbb{F}_2^{\ell \times \ell}$ permutační. Pak existuje $i \in \{1, \ldots, \ell\}$ takové, že*

$$\exists \delta > 0 : I(W) \in (\delta, 1-\delta) \iff \exists \eta_\delta > 0 : I(W) - I(W^{(i)}) > \eta_\delta$$

a toto η_δ závisí pouze na hodnotě δ.

Důkaz. Nechť $\mathbf{X}, \mathbf{Y}, \mathbf{U} \in \mathbb{F}_2^\ell$ jsou náhodné vektory jako v důkazu Tvrzení 17.

Všimněme si, že právě v horní trojúhelníkové matici lze pomocí prvních i sloupců vyjádřit prvních i kanonických bázových vektorů. Díky tomu můžeme pomocí prvních i sloupců vyjádřit pozici i, čímž se kapacita i-té F-transformace kanálu W redukuje na kapacitu kanálu W. Jednotlivé případy rozebereme podrobně zvlášť:

- Nechť platí první možnost a FP je horní trojúhelníková pro nějakou $P \in \mathbb{F}_2^{\ell \times \ell}$ permutační. Buď $i \in \{1, \ldots, \ell\}$. Pak platí:

$$\begin{aligned} I(W^{(i)}) &= 1 - H(\mathbf{U}_i | \mathbf{Y}, \mathbf{U}_1^{i-1}) \\ &= 1 - H((\mathbf{X}F^{-1})_i | \mathbf{Y}, (\mathbf{X}F^{-1})_1^{i-1}) \\ &= 1 - H((\mathbf{X}P(FP)^{-1})_i | \mathbf{Y}, (\mathbf{X}P(FP)^{-1})_1^{i-1}) \\ &= 1 - H((\mathbf{X}P)_i | \mathbf{Y}, (\mathbf{X}P(FP)^{-1})_1^{i-1}) \\ &= 1 - H(\mathbf{X}_1 | \mathbf{Y}) = 1 - H(\mathbf{X}_1 | \mathbf{Y}_1) \\ &= I(W), \end{aligned}$$

  kde čtvrtá a pátá rovnost plyne z toho, že FP je horní trojúhelníková a složky $\mathbf{X}$ jsou nezávislé – pak je totiž i $(FP)^{-1}$ horní trojúhelníková a $(\mathbf{X}P(FP)^{-1})_1^{i-1}$ je bijektivní funkcí pouze $(\mathbf{X}P)_1^{i-1}$, zatímco $(\mathbf{X}P(FP)^{-1})_i$ je lineární kombinací pouze $(\mathbf{X}P)_1^i$, čili podmíněné pravděpodobnosti v definici podmíněné entropie se redukují pouze na podmíněné pravděpodobnosti $(\mathbf{X}P)_i$. Šestá rovnost pak plyne z nezávislosti $\mathbf{X}_i$.

- Nyní nechť platí druhá možnost. Protože je F regulární, existuje z Lemmatu 1 permutační matice $P \in \mathbb{F}_2^{\ell \times \ell}$ taková, že $(FP)^{-1}$ má na diagonále nenulové prvky. Z předpokladů však není FP horní trojúhelníková, tedy $(FP)^{-1}$ není horní trojúhelníková a označíme-li (i,j)-prvek matice $(FP)^{-1}$ jako $(FP)^{-1}_{ij}$,

pak existují $i \in \{1, \ldots, \ell\}$ a $k \in \{i+1, \ldots, \ell\}$ takové, že $(FP)^{-1}_{ki} \neq 0$. Volme nejmenší takové i. Pak

$$\begin{aligned} I(W^{(i)}) &= 1 - H(\mathbf{U}_i|\mathbf{Y}, \mathbf{U}_1^{i-1}) \\ &= 1 - H((\mathbf{X}P(FP)^{-1})_i|\mathbf{Y}, (\mathbf{X}P(FP)^{-1})_1^{i-1}) \\ &= 1 - H((\mathbf{X}P)_i \oplus \sum_{j=i+1}^{l} (FP)^{-1}_{ji}(\mathbf{X}P)_j|\mathbf{Y}, (\mathbf{X}P(FP)^{-1})_1^{i-1}) \\ &= 1 - H((\mathbf{X}P)_i \oplus \sum_{j=i+1}^{l} (FP)^{-1}_{ji}(\mathbf{X}P)_j|\mathbf{Y}), \end{aligned}$$

kde používáme stejný argument jako v předchozím bodě (i-tá rohová čtvercová podmatice $(FP)^{-1}$ je z volby i horní trojúhelníková s nenulovými prvky na diagonále – tedy regulární). Tvrzení pak plyne přímo opakovaným použitím druhé části Lemmatu 13 (náhodné vektory $(\mathbf{X}_i, \mathbf{Y}_i)$ jsou vzájemně nezávislé).

□

Tato charakterizace ukazuje, že pro relativně velkou třídu regulárních matic F má jeden z výsledných kanálů F-transformace symetrickou kapacitu ostře menší, než kanál původní. Jelikož F-transformace symetrickou kapacitu zachovává (viz Tvrzení 17), pak má jeden z výsledných kanálů symetrickou kapacitu ostře větší – F-transformace v tomto případě štěpí kanál W na lepší a horší.

Z tohoto chování F-transformace se nabízí nápad ji znovu rekurzivně opakovat na vzniklé $W^{(i)}$ – jelikož má v jistém smyslu F-transformace fixní body právě v kanálech s $I(W) \in \{0, 1\}$, mohli bychom očekávat, že v limitě se budou těmto hodnotám blížit i symetrické kapacity takto vzniklých kanálů. Naopak pro matice, které tuto podmínku nesplňují, budou mít výsledné kanály stále stejnou kapacitu a k polarizaci nedojde. Toto chování ukazuje Věta 19 – její důkaz provedeme zavedením náhodného procesu W_n s hodnotami v kanálech, kdy v každém čase bude následující kanál jednou z F-transformací současného se stejnou pravděpodobností. Tento proces si můžeme představovat jako náhodnou procházku po stromu kanálů, jehož kořenem je kanál W a každý uzel má ℓ podstromů příslušných ℓ různým F-transformacím – toto ukazuje Obrázek 3.1.

Definice 17. *Řekneme, že matice $F \in \mathbb{F}_2^{\ell \times \ell}$ je **polarizující**, pokud je regulární a zároveň neexistuje permutační matice P taková, že FP je horní trojúhelníková.*

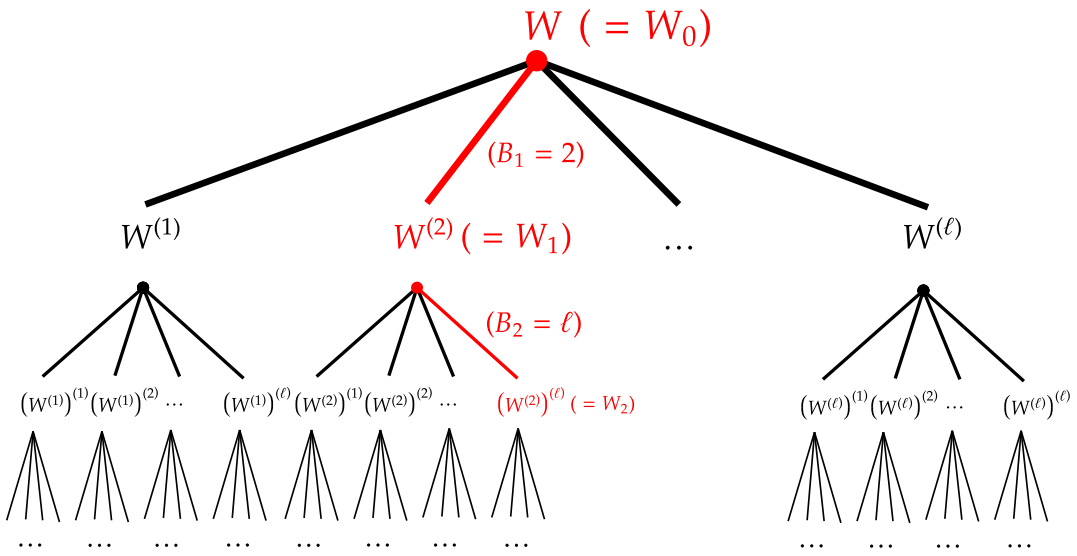
Tvrzení 19. *Buď $F \in \mathbb{F}_2^{\ell \times \ell}$ a W binární kanál. Nechť $(B_n)_{n \in \mathbb{N}}$ je posloupnost nezávislých náhodných veličin na stejném pravděpodobnostním prostoru $(\Omega, \mathcal{F}, \mathsf{P})$ s rovnoměrným rozdělením na $\{1, \ldots, \ell\}$. Definujme náhodné veličiny*

$$W_0 = W \quad W_n = W^{(B_n)} \quad I_n = I(W_n) \quad Z_n = Z(W_n),$$

kde $W^{(B_n)}$ je B_n-tá F-transformace kanálu W. Pak pokud je F polarizující, tak platí

- *$I_n \xrightarrow[n\to\infty]{s.j.} I_\infty$, kde I_∞ je náhodná veličina nabývající skoro jistě hodnot z $\{0, 1\}$, splňující $\mathsf{P}(I_\infty = 1) = I(W)$.*

- $Z_n \xrightarrow[n\to\infty]{s.j.} Z_\infty$, *kde* Z_∞ *je náhodná veličina nabývající skoro jistě hodnot z* $\{0,1\}$, *splňující* $\mathsf{P}(Z_\infty = 0) = I(W)$.



Obrázek 3.1: Schéma kanálového stromu. Červeně je vyznačen příklad trajektorie procesu W_n.

Důkaz.

- **Konvergence** $(I_n)_{n\in\mathbb{N}_0}$
  Pro všechna $n \in \mathbb{N}$ platí

$$\begin{aligned}
\mathsf{E}[I_{n+1}|B_1,\dots,B_n] &= \mathsf{E}[\sum_{i=1}^{\ell} I(W_n^{(i)})\mathbb{1}_{[B_{n+1}=i]}|B_1,\dots,B_n] \\
&= \sum_{i=1}^{\ell} \mathsf{E}[I(W_n^{(i)})\mathbb{1}_{[B_{n+1}=i]}|B_1,\dots,B_n] && \text{Tvrzení 3(1)} \\
&= \sum_{i=1}^{\ell} I(W_n^{(i)})\mathsf{E}[\mathbb{1}_{[B_{n+1}=i]}|B_1,\dots,B_n] && \text{Tvrzení 3(2)} \\
&= \sum_{i=1}^{\ell} I(W_n^{(i)})\mathsf{E}[\mathbb{1}_{[B_{n+1}=i]}] && \text{Tvrzení 3(3)} \\
&= \frac{1}{\ell}\sum_{i=1}^{\ell} I(W_n^{(i)}) && \text{rozdělení } B_{n+1} \\
&= I(W_n) = I_n && \text{Tvrzení 17,}
\end{aligned}$$

  z čehož plyne, že $(I_n)_{n\in\mathbb{N}_0}$ je martingál vzhledem k $(B_n)_{n\in\mathbb{N}_0}$. Z Tvrzení 10 platí $0 \le I(W) \le 1$ pro každý kanál W, tedy $0 \le I_n \le 1$ a z toho i $\sup_n \mathsf{E}|I_n| \le 1 < \infty$. Z Věty 5 pak plyne $I_n \xrightarrow[n\to\infty]{\text{s.j.}} I_\infty$, kde $\mathsf{E}I_\infty = \mathsf{E}I_0 = I(W)$.

- **Rozdělení** I_∞
  Nechť nyní pro spor $I_\infty \in (\delta, 1-\delta)$ s nenulovou pravděpodobností pro nějaké

$\delta > 0$. Z Věty 5 a předchozího bodu máme konvergenci v L_1, která implikuje L_1 Cauchyovskost (Tvrzení 4), tedy pro každé $\varepsilon > 0$ existuje $n_0 \in \mathbb{N}$ takové, že pro všechna $n \geq n_0$ platí

$$\mathsf{E}|I_{n+1} - I_n| < \varepsilon.$$

Ale F je polarizující, tedy z Tvrzení 18 existuje i splňující, že $I_n \in (\delta, 1-\delta)$ právě tehdy, když $|I(W_n^{(i)}) - I(W_n)| > \eta_\delta$ pro nějaké $\eta_\delta > 0$. Protože předpokládáme $I_\infty \in (\delta, 1-\delta)$ s nenulovou pravděpodobností, pak $I_n \in (\delta, 1-\delta)$ s nenulovou pravděpodobností pro nekonečně mnoho n, a tedy pro nekonečně mnoho n platí

$$0 < \frac{1}{\ell}\eta_\delta < \frac{1}{\ell}\mathsf{E}|I(W_n^{(i)}) - I(W_n)| \leq \mathsf{E}|I_{n+1} - I_n| < \varepsilon,$$

kde odhad střední hodnoty plyne z toho, že $I(W_n^{(i)}) = I(W_{n+1})$ s pravděpodobností $\frac{1}{\ell}$. To je ale spor s Cauchyovskostí posloupnosti (volme $\varepsilon < \frac{\eta_\delta}{\ell}$), tedy $I_\infty \in \{0, 1\}$ skoro jistě.

Z předchozího bodu pak máme $\mathsf{E}I_\infty = \mathsf{E}I_0 = I(W)$ a protože $I_\infty \in \{0, 1\}$ skoro jistě, pak i $\mathsf{P}(I_\infty = 1) = \mathsf{E}I_\infty = I(W)$.

- **Konvergence a rozdělení Z_n**
  Plyne přímo z Lemmatu 10 a předchozích bodů.

□

Důsledek. Buď W_0 binární kanál a $F \in \mathbb{F}_2^{\ell\times\ell}$ polarizující. Definujme multimnožiny $\mathcal{W}_n$ $n \in \mathbb{N}_0$ předpisem

$$\mathcal{W}_0 = \{W_0\} \qquad \mathcal{W}_{n+1} = \bigcup_{W\in\mathcal{W}_n} \{W^{(i)} : i \in \{1, \ldots, l\}\},$$

kde $W^{(i)}$ je i-tou F-transformací kanálu W. Z definice má W_n rovnoměrné rozložení na $\mathcal{W}_n$. Pro každé $\delta \in (0, 1)$ pak platí

$$\lim_{n\to\infty} \frac{|\{W \in \mathcal{W}_n : Z(W) < \delta\}|}{l^n} = \lim_{n\to\infty} \mathsf{P}(Z_n < \delta) = I(W_0).$$

Uvažujme nyní multimnožiny

$$\mathcal{V}_n = \operatorname*{arg\,min}_{\substack{\mathcal{W}'\subseteq\mathcal{W}_n \\ |\mathcal{I}|=\lfloor I(W_0)\ell^n\rfloor}} \sum_{W'\in\mathcal{W}'} Z(W').$$

Pak z výše uvedeného máme

$$\lim_{n\to\infty} \sum_{W'\in\mathcal{V}_n} Z(W') = 0.$$

Důsledek výše je vhodné vnímat zejména v kontextu Věty 14, která omezuje pravděpodobnost chyby SC dekódování kosetového kódu právě pomocí součtu Bhattacharyyových parametrů dekódovacích kanálů kódu. Není však stále zjevné, že kanály z $\mathcal{W}_n$ dekódovacími kanály nějakého kosetového kódu jsou. Konstrukce takového kódu je předmětem Kapitoly 4.

3.1 Rychlost polarizace

Výsledek Věty 19 nyní ještě vylepšíme důkazem rychlosti konvergence. Pro něj ale nestačí odhady symetrických kapacit z Tvrzení 18. Následující lemma však poskytuje použitelný odhad pro $Z(W^{(i)})$.

Definice 18. *Buď $A \in \mathbb{F}^{n\times n}$ libovolná matice. Označme i-tý řádkový vektor A jako $\mathbf{a}^i$. Pak pro $i \in \{1,\dots,n\}$ definujeme i-**tou částečnou vzdálenost** A jako číslo $D^{(i)}$ splňující*

$$\begin{aligned} D^{(i)} &= d_H(\mathbf{a}^i, span\{\mathbf{a}^{i+1},\dots,\mathbf{a}^n\}) \quad i = 1,\dots,n-1 \\ D^{(n)} &= w_H(\mathbf{a}^n). \end{aligned}$$

Lemma 20. *Buď W symetrický binární kanál a $F \in \mathbb{F}_2^{\ell\times\ell}$ polarizující. Pak pro každé $i \in \{1,\dots,l\}$ platí*

$$Z(W^{(i)}) \le CZ(W)^{D^{(i)}},$$

kde $W^{(i)}$ je i-tá F-transformace kanálu W a $C \ge 1$ je konstanta.

Důkaz. Viz [5] (*Lemma 13*). □

Podívejme se nyní na důsledky Lemmatu 20 pro náš proces $(Z_n)_{n\in\mathbb{N}_0}$. Pokud by nerovnost platila pro $C = 1$, dostali bychom odhad

$$Z_n \le Z_0^{\prod_{i=1}^n D^{(B_i)}} \Longrightarrow \log_2 Z_n \le (\log_2 Z_0)\ \ell^{n\frac{1}{n}\sum_{i=1}^n \log_\ell D^{(B_i)}}.$$

Ale $(B_n)_{n\in\mathbb{N}}$ jsou rovnoměrně rozdělené i.i.d. náhodné veličiny, tedy ze zákona velkých čísel konverguje výběrový průměr v exponentu na pravé straně k $\mathsf{E}\log_\ell D^{(B_1)}$, tedy asymptoticky budou hodnoty Z_n konvergovat exponenciálně rychle s exponentem daným následující definicí.

Definice 19. *Pro matici $F \in \mathbb{F}_2^{\ell\times\ell}$ nazveme **exponentem** F číslo*

$$E(F) = \frac{1}{\ell}\sum_{i=1}^{\ell} \log_l D^{(i)},$$

kde $D^{(i)}$ je i-tá částečná vzdálenost F.

Obecně však $C > 1$ a analýza výše již nejde přímočaře použít, protože násobení konstantou C může odhad neúměrně zvýšit. Řešení tohoto problému přináší konvergence skoro jistě dokázaná ve Větě 19. S její pomocí nejprve dokážeme asymptotickou lineární konvergenci a pak pro fixní $n \in \mathbb{N}$ najdeme $m \in \mathbb{N}$ takové, že lineární konvergence převáží vliv násobení C mezi Z_m a Z_n. Formálně tento postup ukazuje následující důkaz, který rozšiřuje důkaz rychlosti konvergence z [2] postupem naznačeným v [5].

Tvrzení 21. *Buď $F \in \mathbb{F}_2^{\ell\times\ell}$ polarizující matice a $(W_n)_{n\in\mathbb{N}_0}$, $(Z_n)_{n\in\mathbb{N}_0}$ příslušné procesy z Věty 19. Pak pro každé $\beta < E(F)$ platí*

$$\lim_{n\to\infty} \mathsf{P}[Z_n \le 2^{-\ell^{n\beta}}] = I(W_0).$$

Důkaz. Pro jednoduchost značení definujme $D_n = D^{(B_n)}$. Protože je F polarizující, pak není horní trojúhelníková a tedy $D^{(i)} > 1$ pro nějaké i a tedy $\mathsf{E}D_1 > 1$.

- **Lineární konvergence** Z_n
  Z Lemmatu 20 platí pro $m \in \mathbb{N}$ vztah $\frac{Z_{m+1}}{Z_m} \leq C Z_m^{D_m - 1}$ a tedy

$$\frac{Z_m}{Z_0} \leq C^m \prod_{i=0}^{m-1} Z_i^{D_i - 1}.$$

  Nyní využijeme konvergenci $(Z_m)_{m\in\mathbb{N}}$ skoro jistě a najdeme množinu $A \in \mathcal{F}$ pravděpodobnosti $\mathsf{P}(Z_\infty = 0)$, na které platí $Z_m(\omega) \xrightarrow[m\to\infty]{} 0$. Můžeme očekávat, podobně jako v diskuzi případu $C = 1$ výše, že pro exponenty D_i bude platit zákon velkých čísel, avšak až od bodu, kde mocnění převažuje nad násobením konstantou. Volíme proto $\varepsilon \in (0, \mathsf{E}D_1)$ a najdeme $m_0 = m_0(\omega) \in \mathbb{N}$, od kterého platí $Z_m \leq \frac{C}{4} Z_m^{\mathsf{E}D_1 - \varepsilon}$ – to je splněno pro $Z_m \leq (4C)^{-\frac{1}{\mathsf{E}D_1 - \varepsilon - 1}}$. Potom můžeme pro $m \geq m_0$ psát

$$\begin{aligned}
\frac{Z_m}{Z_0} &\leq C^m \left(\prod_{i=0}^{m_0-1} Z_i^{D_i-1} \right) \left(\prod_{i=m_0}^{m-1} Z_i^{D_i-1} \right) \\
&\leq C^m \prod_{i=m_0}^{m-1} (4C)^{-\frac{D_i-1}{\mathsf{E}D_1-\varepsilon-1}} \\
&\leq C^m \left(\prod_{i=0}^{m_0-1} (4C)^{\frac{l-1}{\mathsf{E}D_1-\varepsilon-1}} \right) \left(\prod_{i=0}^{m-1} (4C)^{-\frac{D_i-1}{\mathsf{E}D_1-\varepsilon-1}} \right) \\
&\leq c(m_0)\, C^m (4C)^{-\frac{\sum_{i=0}^{m-1} D_i - m}{\mathsf{E}D_1-\varepsilon-1}},
\end{aligned}$$

  kde $c(m_0)$ značí obecnou konstantu závislou na m_0. Na jevu $A \cap \left[\frac{1}{m}\sum_{i=0}^{m-1} D_i \geq \mathsf{E}D_1 - \varepsilon\right]$ platí pro dostatečně velká m

$$Z_m \leq c(m_0)\, C^m (4C)^{-m\frac{\mathsf{E}D_1-\varepsilon-1}{\mathsf{E}D_1-\varepsilon-1}} \leq c(m_0) \left(\frac{1}{4}\right)^m \leq \left(\frac{1}{2}\right)^m,$$

  tedy máme

$$\mathsf{P}\left(Z_m \leq \left(\frac{1}{2}\right)^m\right) \geq \mathsf{P}\left(A \cap \left[\frac{1}{m}\sum_{i=0}^{m-1} D_i \geq \mathsf{E}D_1 - \varepsilon\right]\right).$$

  $(D_m)_{m\in\mathbb{N}}$ jsou nezávislé stejně rozdělené náhodné veličiny splňující $\mathsf{E}D_1 < \infty$ a z Tvrzení 6 dostáváme $\mathsf{P}(\frac{1}{m}\sum_{i=0}^{m-1} D_i \geq \mathsf{E}D_1 - \varepsilon) \xrightarrow{m\to\infty} 1$ a protože $\mathsf{P}(A) = \mathsf{P}(Z_\infty = 0)$, pak ze spojitosti pravděpodobnosti dostáváme limitním přechodem v nerovnosti výše

$$\lim_{m\to\infty} \mathsf{P}\left(Z_m \leq \left(\tfrac{1}{2}\right)^m\right) \geq \mathsf{P}(Z_\infty = 0).$$

- **Exponenciální konvergence** Z_n
  Mějme nyní $m_n < n$ přirozená a rozdělme množinu $\{m_n, \ldots, n-1\}$ na k podmnožin $I_1, \ldots, I_k$ po sobě jdoucích prvků velikosti $\sqrt{n}$ (tedy platí $n - m_n = \sqrt{n}k$). Uvažujme jev $B_n = \cap_{j=0}^{k} [\frac{1}{\sqrt{n}} \sum_{i\in I_j} \log_\ell D_i \geq \beta]$. Zde již

nemůžeme přímočaře použít Tvrzení 6, neboť k závisí na n – pro zajištění konvergence potřebujeme získat přesnější odhad. Pro pravděpodobnost jevu B_n platí

$$\begin{aligned}
\mathsf{P}(B) &= 1 - \mathsf{P}(B^c) \\
&\geq 1 - \sum_{j=1}^{k} \mathsf{P}\left(\frac{1}{\sqrt{n}} \sum_{i \in I_j} \log_\ell D_i < \beta\right) \\
&= 1 - \sum_{j=1}^{k} \mathsf{P}\left(\frac{1}{\sqrt{n}} \sum_{i \in I_j} (\log_\ell D_i - \mathsf{E} \log_\ell D_1) < \beta - \mathsf{E} \log_\ell D_1\right) \\
&\geq 1 - k e^{-2\sqrt{n}(\beta - \mathsf{E} \log_\ell D_1)^2} \xrightarrow[n\to\infty]{} 1,
\end{aligned}$$

kde jsme využili Tvrzení 7 pro $0 \leq \log_l D_i \leq 1$.

Pro $p \in \mathbb{N}$ máme p-násobným použitím Lemmatu 20 následující odhad:

$$\begin{aligned}
\log_2 Z_{m_n+p} &\leq (\log_2 C) \sum_{q=0}^{p-1} \prod_{i=1}^{q} D_i + (\log_2 Z_{m_n}) \prod_{i=1}^{p} D_i \\
&\leq (\log_2 C) p \ell^{\sum_{i=1}^{p-1} \log_\ell D_i} + (\log_2 Z_{m_n}) \ell^{\sum_{i=1}^{p} \log_l D_i} \\
&\leq \ell^{\sum_{i=1}^{p} \log_\ell D_i} \left[p \log_2 C + \log_2 Z_{m_n}\right].
\end{aligned}$$

Na jevu B_n z toho odhadu obdržíme pro $p = \sqrt{n}$ následující vztah pro hodnoty Z_n na krajních bodech podmnožin I_j:

$$\log_2 Z_{m_n+(j+1)\sqrt{n}} \leq \ell^{\sqrt{n}\beta} \left[\log_2 Z_{m_n+j\sqrt{n}} + \sqrt{n} \log_2 C\right].$$

Jeho opakovaným použitím získáme

$$\begin{aligned}
\log_2 Z_n &\leq \ell^{(n-m_n)\beta} \log_2 Z_m + \sqrt{n} (\log_2 C) \sum_{j=1}^{k} \ell^{j\sqrt{n}\beta} \\
&\leq \ell^{(n-m_n)\beta} \left[\log_2 Z_{m_n} + \sqrt{n} (\log_2 C)(1 - \ell^{-\sqrt{n}\beta})^{-1}\right],
\end{aligned}$$

kde jsme využili částečných součtů geometrické řady. Ze získaného odhadu je patrné, že na jevu B_n (tedy že na všech podmnožinách I_j je průměrný exponent alespoň β) přispívá v závorce výše násobení konstantou C mezi Z_{m_n} a Z_n asymptoticky pouze řádově $\sqrt{n}$. Zároveň ale víme, že Z_{m_n} konverguje asymptoticky lineárně, tedy $\log_2 Z_{m_n}$ je asymptoticky řádově m_n. Položme proto nyní $m_n = n^{\frac{3}{4}}$. Na jevech $C_n = B_n \cap [Z_{m_n} \leq \left(\frac{1}{2}\right)^{m_n}]$ pak platí

$$\begin{aligned}
\log_2 Z_n &\leq \ell^{(n-m_n)\beta} \left[-m + n^{\frac{1}{2}} (\log_2 C)(1 - \ell^{-\frac{1}{2}\beta})^{-1}\right] \\
&= l^{(n-n^{\frac{3}{4}})\beta} \left[-n^{\frac{3}{4}} + n^{\frac{1}{2}} (\log_2 C)(1 - \ell^{-\frac{1}{2}\beta})^{-1}\right] \\
&\leq -\ell^{n\beta}
\end{aligned}$$

pro dostatečně velká n. Limitním přechodem a spojitostí pravděpodobnosti společně s prvním bodem důkazu dostáváme

$$\begin{aligned}
\lim_{n\to\infty} \mathsf{P}(Z_n \leq 2^{-\ell^{n\beta}}) &\geq \lim_{n\to\infty} P(C_n) \\
&\geq P(Z_\infty = 0) = I(W),
\end{aligned}$$

z čehož plyne dokazované tvrzení, neboť

$$I(W) \leq \mathsf{P}(Z_n \leq 2^{-\ell^{n\beta}}) \leq \mathsf{P}\left(Z_n \leq \frac{1}{2}\right) \xrightarrow[n\to\infty]{} \mathsf{P}(Z_\infty = 0) = I(W).$$

□

Důsledek. Buď $\mathcal{V}_n$ jako v Důsledku pod Tvrzením 19. Pak pro libovolné $\beta < E(F)$

$$\lim_{n\to\infty} \frac{\sum_{W'\in\mathcal{V}_n} Z(W')}{2^{-l^{\beta n}}} = 0.$$

4. Polární kódy

V předchozí kapitole jsme ukázali, že opakovanou aplikací F-transformace dochází k polarizaci výsledných kanálů. Stále jsme ale neukázali, že kanály, které takovou opakovanou aplikací vznikají, jsou dekódovacími kanály nějakého kosetového kódu. Intuitivně bychom mohli z očekávat, že generující matice takového kódu bude mít určitou rekurzivní strukturu. V následující kapitola ukážeme, že stačí za generující matici zvolit Kroneckerovu mocninu matice F. V následujícím textu budeme uvažovat fixní binární kanál W (pokud explicitně neuvedeme jinak).

Definice 20. *Buď $F \in \mathbb{F}^{\ell\times\ell}$. Pak pro každé $n \in \mathbb{N}_0$ a $N = \ell^n$ definujme matici $G_{lN} \in \mathbb{F}_2^{N\times N}$ vztahem*

$$G_1 = I_1 \qquad G_{\ell N} = (I_N \otimes F)R_{\ell N}(I_\ell \otimes G_N),$$

kde R_{lN} je permutační matice odpovídající permutaci na $\{1,\dots,N\ell\}$ dané předpisem

$$Nk + i + 1 \mapsto \ell i + k + 1 \qquad k \in \{0,\dots,\ell-1\},\ i \in \{0,\dots,N-1\}$$

Poznámka. V následující kapitole budeme uvažovat fixní polarizující matici $F \in \mathbb{F}_2^{\ell\times\ell}$ a budeme značit $W_N = W_{G_N}$ a $W_N^{(i)} = W_{G_N}^{(i)}$ pro $N = \ell^n$, $n \in \mathbb{N}$ a $i \in \{1,\dots,N\}$.

Z předpisu v Definici 20 nemusí být zřejmá struktura permutace definující permutační matici R_{lN}. Tu můžeme také zapsat jako

$$p + 1 \mapsto \ell(p \bmod N) + (p \operatorname{div} N) + 1 = \ell(p \bmod \ell^n) + (p \operatorname{div} \ell^n) + 1,$$

kde div (resp. mod) značí celočíselné dělení (resp. zbytek po celočíselném dělení). Buď nyní $p = \sum_{i=0}^{n} p_i\ell^i + 1$, kde $p_i \in \{0,\dots,l-1\}$. Pak působením permutace na p dostáváme

$$\begin{aligned} p + 1 &\mapsto \ell\left(\sum_{i=0}^{n} p_i\ell^i \bmod \ell^n\right) + \left(\sum_{i=0}^{n} p_i\ell^i \operatorname{div} \ell^n\right) + 1 \\ &= \ell\left(\sum_{i=0}^{n-1} p_i\ell^i - 1\right) + p_\ell + 1 \\ &= p_\ell + \sum_{i=1}^{n} p_{i-1}\ell^i + 1 \end{aligned}$$

tedy permutace odpovídá (až na posunutí o jedna z notačních důvodů) rotaci ℓ-árního zápisu čísla p o jednu pozici. Tuto vlastnost lze využít například pro efektivní implementaci (zejména pro $\ell = 2$, kdy operace odpovídá binární rotaci).

Rekurentní vztah pro matici $G_{\ell N}$ lze také přepsat do explicitního tvaru, jak ukazuje následující lemma.

Lemma 22. *Pro $N = \ell^n$, $n \in \mathbb{N}$ platí $G_N = B_N F^{\otimes n}$, kde $B_N = R_N(I_\ell \otimes R_{\frac{N}{\ell}})\dots(I_{\frac{N}{\ell}} \otimes R_\ell)$.*

Důkaz. První ukážeme, že $(I_N \otimes F)R_{\ell N} = R_{\ell N}(F \otimes I_N)$. Pro $\mathbf{u} \in \mathbb{F}_2^{\ell N}$ libovolné a $0 \leq k \leq \ell - 1$, $1 \leq i \leq N$ platí

$$\begin{aligned}[\mathbf{u}(I_N \otimes F)R_{\ell N}]_{Nk+i} &= [\mathbf{u}(I_N \otimes F)]_{\ell(i-1)+(k+1)} = (\mathbf{u}_{(i-1)\ell+1}^{i\ell} F)_{k+1} \\ &= [(\mathbf{u}R_{\ell N})_{i,i+N,\ldots,i+(\ell-1)N} F]_{k+1} = [\mathbf{u}R_{\ell N}(F \otimes I_N)]_{Nk+i}.\end{aligned}$$

Dále postupujeme indukcí. Pro $n = 1$ je tvrzení zřejmé z definice. Pro ℓN dostáváme z Tvrzení 2

$$\begin{aligned}G_{\ell N} &= (I_N \otimes F)R_{\ell N}(I_\ell \otimes G_N) \\ &= (I_N \otimes F)R_{\ell N}(I_\ell \otimes B_N F^{\otimes n}) \\ &= R_{\ell N}(F \otimes I_N)(I_\ell \otimes G_N) \\ &= R_{\ell N}(F \otimes G_N) = R_{\ell N}(F \otimes B_N F^{\otimes n}) \\ &= R_{\ell N}(B_N \otimes I_\ell)(F \otimes F^{\otimes n}) \\ &= B_{\ell N} F^{\otimes (n+1)}.\end{aligned}$$

□

Na základě lemmatu výše se můžeme ptát, zda je permutace $R_{\ell N}$ v definici vůbec nutná, jelikož odpovídá jenom permutaci pozic vstupního slova. Ukazuje se, že tuto permutaci skutečně můžeme vynechat a stále můžeme konstrukci obdobnou té níže provést – takový postup by však byl technicky náročnější. Následující lemma totiž dává do vztahu dekódovací kanály pro různá n a permutace je v jeho důkazu nutná, aby rovnoměrně rozložila pozice vstupního slova mezi jednotlivé kanály (lépe je potřeba této permutace viditelná v důkazu níže). Pokud bychom tuto permutaci vynechali, výsledné kanály by po transformaci nebyly dekódovacími kanály dle naší definice, přesto bychom však mohli upravit Algoritmus 1 tak, aby konstrukce fungovala i s nimi.

Lemma níže je klíčové, byť je jeho důkaz poněkud technický – dává totiž do souvislosti výsledky z Kapitoly 2 a Kapitoly 3.

Lemma 23. *Buď $n \in \mathbb{N}$, $N = \ell^n$ a $k \in \{0, \ldots, N-1\}$. Pak $(W_{\ell N}^{(\ell k+1)}, W_{\ell N}^{(\ell k+2)}, \ldots, W_{\ell N}^{(\ell(k+1)-1)})$ je F-transformací kanálu $W_N^{(k)}$.*

Důkaz. Z Definice 15 máme pro $i \in \{1, \ldots, \ell\}$:

$$\begin{aligned}W_{\ell N}^{(\ell k+i)}(\mathbf{y}, \mathbf{u}_1^{\ell k+i-1} | \mathbf{u}_{\ell k+i}) &= \frac{1}{2^{l(N-k)-i}} \sum_{\mathbf{u}_{\ell k+i+1}^{\ell N}} W_{\ell N}(\mathbf{y}|\mathbf{u}) \\ &= \frac{1}{2^{l(N-k)-i}} \sum_{\mathbf{u}_{\ell k+i+1}^{\ell N}} W^{\ell N}(\mathbf{y}|\mathbf{u}(I_N \otimes F)R_{\ell N}(I_\ell \otimes G_N)) \\ &= \frac{1}{2^{l(N-k)-i}} \sum_{\mathbf{u}_{\ell k+i+1}^{\ell N}} \prod_{j=1}^{\ell} W_N(\mathbf{y}_{(j-1)N+1}^{jN} | (\mathbf{u}(I_N \otimes F)R_{\ell N})_{(j-1)N+1}^{jN}).\end{aligned}$$

Použitím definice Kroneckerova součinu a Definice 20 se snadno nahlédne, že

$$\begin{aligned}(\mathbf{u}(I_N \otimes F)R_{\ell N})_{(j-1)N+1}^{jN} &= (\mathbf{u}(I_N \otimes F))_{j,j+\ell,\ldots,(N-1)\ell+j} \\ &= ((\mathbf{u}_1^\ell F)_j \ (\mathbf{u}_{\ell+1}^{2\ell} F)_j \ \ldots (\mathbf{u}_{(N-1)\ell+1}^{N\ell} F)_j).\end{aligned}$$

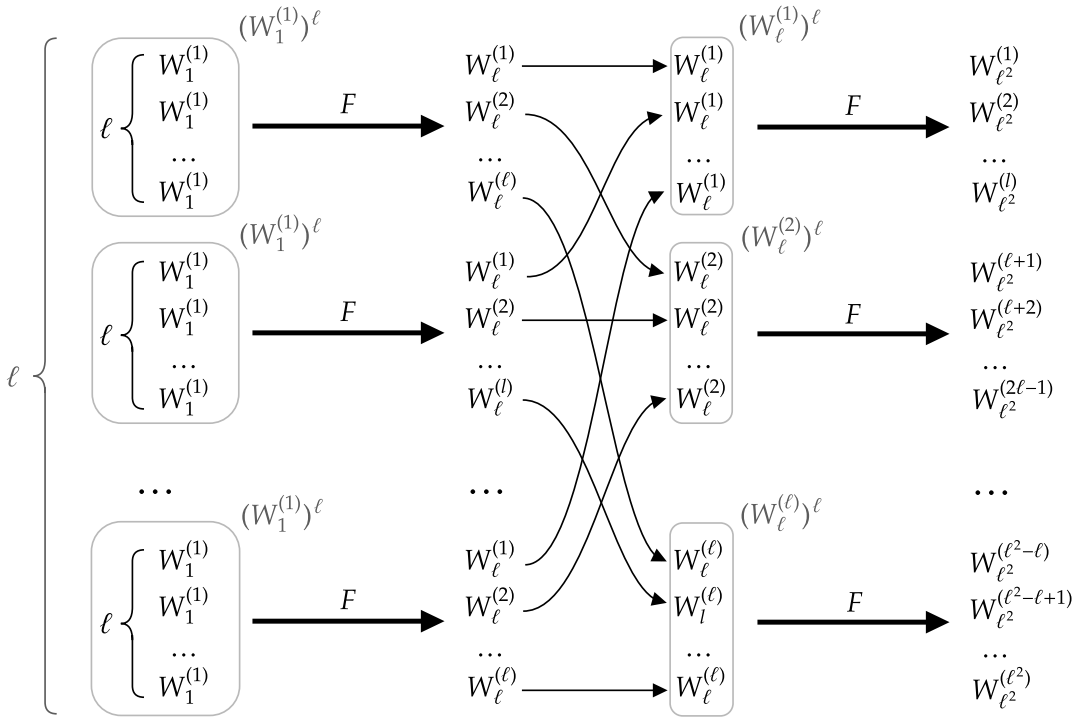
Z regularity F pak vidíme, že sčítáním přes všechny hodnoty $\mathbf{u}_{\ell(k+1)+1}^{lN}$ sčítáme přes všechny hodnoty posledních $N-k-1$ pozic vstupů kanálů W_N výše (odtud také plyne potřeba permutace $R_{\ell N}$, viz diskuzi nad lemmatem). To je ale přesně definice k-té F-transformace kanálu W_N, z čehož seskupením příslušných sum dostáváme

$$\frac{1}{2^{l(N-k)-i}} \sum_{\mathbf{u}_{\ell k+i+1}^{\ell N}} \prod_{j=1}^{\ell} W_N(\mathbf{y}_{(j-1)N+1}^{jN} | (\mathbf{u}(I_N \otimes F)R_{lN})_{(j-1)N+1}^{jN})$$

$$= \frac{1}{2^{l-i}} \sum_{\mathbf{u}_{\ell k+i+1}^{\ell(k+1)}} \prod_{j=1}^{\ell} W_N^{(k)}(\mathbf{y}_{(j-1)N+1}^{jN}, (\mathbf{u}(I_N \otimes F)R_{lN})_{(j-1)N+1}^{(j-1)N+k} | (\mathbf{u}_{\ell k+1}^{\ell(k+1)} F)_j),$$

což je tvar i-té F-transformace kanálu $W_N^{(k)}$ dle Definice 16 (kde v definici identifikujeme vstup F-transformovaného kanálu s $\mathbf{u}_{\ell k+1}^{\ell k+i}$ a výstup s dvojicí $(\mathbf{y}, \mathbf{u}_{\ell(k-1)+1}^{\ell k})$).

□



Obrázek 4.1: Znázornění rekurzivních transformací z Lemmatu 23 pro $n = 2$. Každý kanál v obrázku odpovídá nezávislé realizaci daného kanálu – šedě jsou vyznačeny mocniny kanálů.

Lemma ukazuje, že dekódovací kanály $W_N^{(i)}$ vznikly opakovanou rekurzivní F-transformací kanálu $W = W_1^{(1)}$ (schématicky tento postup ukazuje Obrázek 4.1). Jelikož víme, že takovým rekurzivním procesem dochází k polarizaci $I(W)$-části výsledných kanálů (viz Důsledek pod Větou 19) a pravděpodobnost chyby SC dekódování příslušného kosetového kódu je shora omezená součtem Bhattacharyyových parametrů těchto kanálů (Tvrzení 16), dostáváme tímto postupem konstrukci kosetových kódů asymptoticky dosahujících symetrické kapacity.

Takto popsaná konstrukce dává jednoznačně určené asymptoticky optimální kódy pouze v případě, kdy je kanál W symetrický a jsou tedy splněny předpoklady Tvrzení 16 – v opačném případě postup vede na třídu kosetových kódů s různými hodnotami $\mathbf{u}_{\mathcal{A}^c}$ a obecně různými pravděpodobnostmi chyby dekódování. To zde však nevadí, neboť pro nesymetrické kanály jsou vzniklé kódy neoptimální již v tom smyslu, že asymptoticky dosahují pouze symetrické kapacity kanálu, která je pro nesymetrický kanál obecně různá od jeho kapacity. V následující definici polárních kódů se proto bez značné újmy na obecnosti omezíme pouze na symetrické kanály a nastavíme $\mathbf{u}_{\mathcal{A}^c} = \mathbf{0}_{N-K}$, aby výsledné kódy byly lineární.

Definice 21. *Buď W symetrický binární kanál a $0 < R < I(W)$. Pro $F \in \mathbb{F}_2^{\ell\times\ell}$ polarizující a $N = \ell^n, n \in \mathbb{N}$ rozumíme* ***polárním kódem s parametry*** *(F, N, R, W) kosetový kód s parametry $(G_N, N, \lfloor NR \rfloor, \mathcal{A}, \mathbf{0}_{N-K})$, kde $\mathcal{A}$ je množina splňující*

$$\mathcal{A} = \underset{\substack{\mathcal{I}\subseteq\{1,\dots,N\}\\ |\mathcal{I}|=\lfloor NR\rfloor}}{\arg\min} \sum_{i\in\mathcal{I}} Z(W_N^{(i)}).$$

Rekurzivní struktura polárních kódů přináší efektivní algoritmy jak pro kódování tak pro dekódování (jak již jsme naznačili při zavedení SC dekódování v Kapitole 2). V následujících pozorováních budeme používat běžnou notaci $\mathcal{O}(\cdot)$ pro horní asymptotický odhad, tj. pro funkce $f, g : \mathbb{N} \longrightarrow \mathbb{R}$ píšeme

$$f(n) = \mathcal{O}(g(n)) \iff \exists n_0 \in \mathbb{N}, C > 0\ \forall n \geq n_0\ f(n) \leq Cg(n).$$

Pozorování 24. *(F, N, R, W)-polární kód lze kódovat v čase $\mathcal{O}(N \log N)$.*

Důkaz. Z definice kosetového kódu je kódování vstupního slova $\mathbf{u}_{\mathcal{A}}$ ekvivalentní s násobením $\mathbf{u} \in \mathbb{F}_2^{\ell N}$ s maticí G_{lN}. Z Lemmatu 22 platí $G_{\ell N} = B_{\ell N} F^{\otimes(n+1)}$, kde $B_{\ell N}$ je permutační matice, tedy $\mathbf{u}B_{\ell N}$ lze vyhodnotit v čase $\mathcal{O}(\ell N) = \mathcal{O}(N)$ a získat tím vektor $\mathbf{v} = \mathbf{u}B_{\ell N}$.

Uvažujme nyní složitost vyhodnocení $\mathbf{v}F^{\otimes(n+1)}$. Z Tvrzení 2 platí

$$\begin{aligned}\mathbf{v}F^{\otimes(n+1)} &= \mathbf{v}(F^{\otimes n} \otimes F) = \mathbf{v}(F^{\otimes n} \otimes I_\ell)(I_N \otimes F)\\ &= (\mathbf{v}_1^N F^{\otimes n}\ \mathbf{v}_{N+1}^{2N} F^{\otimes n}\ \dots\ \mathbf{v}_{(\ell-1)N+1}^{\ell N} F^{\otimes n})(I_{\frac{N}{\ell}} \otimes F)\end{aligned}$$

Označme nyní jako $T_{\ell N}$ čas potřebný k výpočtu $\mathbf{v}F^{\otimes(n+1)}$. Protože maticí $(I_{\frac{N}{\ell}} \otimes F)$ dokážeme násobit v čase $\mathcal{O}(\frac{N}{\ell}\ell^2) = \mathcal{O}(N)$ (jedná se o blokově diagonální matici s $\frac{N}{\ell}$ bloky velikosti ℓ), plyne z dokázaného vztahu rekurence

$$T_{\ell N} = \ell T_N + \mathcal{O}(N),$$

tedy $T_N = \mathcal{O}(N \log N)$. Celková časová složitost kódování polárního kódu délky N je tedy $\mathcal{O}(N) + \mathcal{O}(N \log N) = \mathcal{O}(N \log N)$.

□

Pozorování 25. *(F, N, R, W)-polární kód lze dekódovat v čase $\mathcal{O}(N \log N)$.*

Důkaz. Polární kód je kosetovým kódem, budeme tedy uvažovat jeho SC dekódování. Z Algoritmu 1 vidíme, že SC dekodér rozhoduje o svém výstupu na základě hodnot

$$W_N^{(\ell k+i)}(\mathbf{y}, \hat{\mathbf{u}}_1^{\ell k+i-1}|\cdot)$$

pro $i \in \{1, \ldots, N-1\}$ a kde $\cdot$ označuje 0 nebo 1. Z důkazu Lemmatu 23 ale plyne, že tyto hodnoty jsou závislé pouze na hodnotách

$$\{W_N^{(k)}(\mathbf{y}_{(j-1)N+1}^{jN}, (\hat{\mathbf{u}}(I_N \otimes F)R_{\ell N})_{(j-1)N+1}^{(j-1)N+k}|\cdot) : 1 \leq j \leq \ell\}$$

a z nich je lze vypočítat v čase $\mathcal{O}(\ell 2^\ell) = \mathcal{O}(1)$ (provádíme ℓ násobení v $\mathcal{O}(2^\ell)$ členech součtu). Označme nyní čas potřebný pro výpočet všech N hodnot $W_N^{\ell k+i}$ jako T_N. Z výše uvedeného platí následující rekurence:

$$T_{\ell N} = \ell T_N + \mathcal{O}(N).$$

Řešení této rekurence splňuje $T_N = \mathcal{O}(N \log N)$ a protože na základě vypočtených $W_N^{\ell k+i}$ rozhodne SC dekodér v čase $\mathcal{O}(N)$, je celková časová složitost dekódování také $\mathcal{O}(N \log N)$.

□

Výsledky výše byly odvozeny pro obecnou polarizující matici F. Nyní z těchto obecných výsledků obdržíme volbou $F = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ původní konstrukci polárních kódů z [1].

Věta 26. *Buď W symetrický binární kanál. Pak pro libovolné $0 < R < I(W)$ existuje posloupnost jednoznačně určených lineárních $[N, K_N]_2$ kódů takových, že $N \longrightarrow \infty$ a $K_N = \lfloor NR \rfloor$ s dekódovacím algoritmem běžícím v čase $\mathcal{O}(N \log N)$ a pravděpodobností dekódovací chyby $\mathcal{O}(N 2^{-N^\beta})$ pro každé $\beta < \frac{1}{2}$.*

Důkaz. Buď $F = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. Ta je polarizující, neboť je regulární a žádnou permutací sloupců nedostaneme horní trojúhelníkovou matici. Její částečné vzdálenosti jsou z definice $D^{(1)} = 1$ a $D^{(2)} = 2$, její exponent je tedy $E(F) = \frac{1}{2}(\log_2 1 + \log_2 2) = \frac{1}{2}$.

Nyní pro $n \in \mathbb{N}$ definujme $N = 2^n$ a uvažujme (F, N, R, W)-polární kód. Ten je lineární a lze jej z Pozorování 25 dekódovat v čase $\mathcal{O}(N \log N)$. Z Tvrzení 16 je pravděpodobnost chyby dekódování omezena shora pomocí

$$\sum_{i \in \mathcal{A}} Z(W_N^{(i)}),$$

ale z důsledku Tvrzení 21 a volby $\mathcal{A}$ v Definici 21 ve spojení s Lemmatem 23 platí $\sum_{i \in \mathcal{A}} Z(W_N^{(i)}) \in \mathcal{O}(N 2^{-N^\beta})$ pro každé $\beta < E(F) = \frac{1}{2}$.

□

Věta 26 dává pro binární symetrické kanály plně konstruktivní důkaz Shannonovy věty a navíc pro takto zkonstruované kódy poskytuje efektivní dekódovací algoritmus. Jakkoliv je tento důkaz konstruktivní, všimněme si, že naivní konstrukce založená na přímém výpočtu $Z(W_N^{(i)})$ pro všechna i je výpočetně neproveditelná z důvodu velikosti výstupní abecedy. Obecně je otázka efektivní konstrukce polárních kódů netriviální – v [1] je zmíněn možný postup založený na Monte Carlo metodách. Pro $F = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ je v [9] vyložen algoritmus s časovou složitostí $\mathcal{O}(N \log^2 N \log \log N)$ založený na vhodném omezování velikosti výstupní abecedy slučováním výstupních symbolů. Tento výsledek lze vylepšit využitím toho, že na výsledných kanálech existuje částečné uspořádání [6] a dosáhnout tak algoritmu konstrukce, jehož časová složitost je sublineární v N.

Závěr

V práci jsme popsali konstrukci polárních kódů a její základní myšlenky. Vlastním přínosem je zejména obsah Kapitoly 3, kde jsme formulovali Definici 16, doplnili důkazy některých Tvrzení (17, 19, 21 a související Lemma 13) a poskytli vlastní důkaz Tvrzení 18. V Kapitole 2 jsme přiblížili princip SC dekódování na Příkladu 2, v Kapitole 4 jsme pak rozšířili důkazy z [1] na případ obecné polarizační matice (Lemmata 22, 23 a Pozorování 25, 24).

Zde popsané výsledky lze rozšířit několika směry. Nabízí se zejména otázka, jestli nejde popsanou konstrukci zobecnit – například na nebinární kanály nebo použitím obecnější transformace. Dále je možné studovat strukturu polárních kódů při konečných délkách a jejich efektivitou – zde se objevuje podobnost s Reed-Mullerovými kódy (více k této souvislosti lze najít v původním článku [1]). V neposlední řadě je nasnadě otázka existence a možných exponentů polarizujících matic – některé výsledky v tomto směru lze najít v [5].

Seznam použité literatury

[1] Arikan, E. (2009). Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, **55**(7), 3051–3073. doi: 10.1109/TIT.2009.2021379.

[2] Arikan, E. a Telatar, E. (2009). On the rate of channel polarization. In *2009 IEEE International Symposium on Information Theory*, pages 1493–1495. doi: 10.1109/ISIT.2009.5205856.

[3] Cover, T. M. a Thomas, J. A. (2006). *Elements of information theory*. Wiley-Interscience, Hoboken, N.J, 2nd ed edition. ISBN 978-0-471-24195-9. OCLC: ocm59879802.

[4] Klenke, A. (2020). *Probability theory: a comprehensive course*. Universitext. Springer Nature, Cham, Switzerland, third edition edition. ISBN 978-3-030-56401-8. doi: 10.1007/978-3-030-56402-5.

[5] Korada, S. B., Şaşoğlu, E. a Urbanke, R. (2010). Polar codes: Characterization of exponent, bounds, and constructions. *IEEE Transactions on Information Theory*, **56**(12), 6253–6264. doi: 10.1109/TIT.2010.2080990.

[6] Schürch, C. (2016). A partial order for the synthesized channels of a polar code. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 220–224. doi: 10.1109/ISIT.2016.7541293.

[7] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, **27**(3), 379–423. ISSN 0005-8580. Publisher: Nokia Bell Labs.

[8] Sutskover, I., Shamai, S. a Ziv, J. (2005). Extremes of information combining. *IEEE Transactions on Information Theory*, **51**(4), 1313–1325. doi: 10.1109/TIT.2005.844077.

[9] Tal, I. a Vardy, A. (2013). How to construct polar codes. *IEEE Transactions on Information Theory*, **59**(10), 6562–6582. doi: 10.1109/TIT.2013.2272694.

[10] Wyner, A. a Ziv, J. (1973). A theorem on the entropy of certain binary sequences and applications–I. *IEEE Transactions on Information Theory*, **19**(6), 769–772. doi: 10.1109/TIT.1973.1055107.