Universal Polar Codes for More Capable and Less Noisy Channels and Sources

arXiv: 1312.5990

David Sutter, Joseph M. Renes

Institute for Theoretical Physics

ISIT 2014, Hawaii
## Motivation

### Existence of optimal non-universal codes

Given two DMCs $W : \mathcal{X} \rightarrow \mathcal{Y}$ and $V : \mathcal{X} \rightarrow \mathcal{Z}$ with the same capacity-achieving input distribution and the same capacity. Does there exist a code that achieves the capacity of $W$ but not of $V$, using **optimal decoding**?

- In general, it is desirable to have universal codes
- A non-universal capacity-achieving code could be beneficial for **sending quantum information over a quantum channel** at a rate $> \text{coherent information}$ [Renes et al.'13]
Motivation

**Existence of optimal non-universal codes**

Given two DMCs $W: \mathcal{X} \to \mathcal{Y}$ and $V: \mathcal{X} \to \mathcal{Z}$ with the same capacity-achieving input distribution and the same capacity. Does there exist a code that achieves the capacity of $W$ but not of $V$, using **optimal decoding**?

- In general, it is desirable to have universal codes
- A non-universal capacity-achieving code could be beneficial for **sending quantum information over a quantum channel** at a rate $>\text{coherent information}$ [Renes et al.'13]

Canditate: polar codes
Notation & Definitions

- Given two DMCs $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$
- $X^n$ with $X_i$ i.i.d. Bernoulli($p$), $p \in [0, 1]$;
- $Y^n = W^n X^n$ and $Z^n = V^n X^n$
- $U^n = G_n X^n$ with $G_n := \left( \begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right)^{\log n}$

For $\varepsilon > 0$ consider the two low-entropy sets

- $\mathcal{D}^n_{\varepsilon}(X|Y) := \{ i \in [n] : H(U_i|U^{i-1}, Y^n) \leq \varepsilon \}$
- $\mathcal{D}^n_{\varepsilon}(X|Z) := \{ i \in [n] : H(U_i|U^{i-1}, Z^n) \leq \varepsilon \}$
Notation & Definitions

- Given two DMCs \( W : \mathcal{X} \to \mathcal{Y} \) and \( V : \mathcal{X} \to \mathcal{Z} \)
- \( X^n \) with \( X_i \) i.i.d. Bernoulli\((p)\), \( p \in [0,1] \);
- \( Y^n = W^n X^n \) and \( Z^n = V^n X^n \)
- \( U^n = G_n X^n \) with \( G_n := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{\log n} \)

For \( \varepsilon > 0 \) consider the two low-entropy sets
- \( \mathcal{D}^n_\varepsilon(X|Y) := \{ i \in [n] : H(U_i|U^{i-1}, Y^n) \leq \varepsilon \} \)
- \( \mathcal{D}^n_\varepsilon(X|Z) := \{ i \in [n] : H(U_i|U^{i-1}, Z^n) \leq \varepsilon \} \)

Definition: degraded

- \( V \) is a (stochastically) degraded version of \( W \) if \( \exists \ T : \mathcal{Y} \to \mathcal{Z} \) s.t. \( V(z|x) = \sum_{y \in \mathcal{Y}} W(y|x) T(z|y) \) \( \forall x \in \mathcal{X}, z \in \mathcal{Z} \)
Given two DMCs \( W : \mathcal{X} \to \mathcal{Y} \) and \( V : \mathcal{X} \to \mathcal{Z} \)

- \( X^n \) with \( X_i \) i.i.d. Bernoulli\((p)\), \( p \in [0,1] \);
- \( Y^n = W^n X^n \) and \( Z^n = V^n X^n \)
- \( U^n = G_n X^n \) with \( G_n := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \log n \)

For \( \varepsilon > 0 \) consider the two low-entropy sets

- \( D^n_\varepsilon(X|Y) := \{ i \in [n] : H(U_i|U_i^{-1}, Y^n) \leq \varepsilon \} \)
- \( D^n_\varepsilon(X|Z) := \{ i \in [n] : H(U_i|U_i^{-1}, Z^n) \leq \varepsilon \} \)

**Definition: degraded, less noisy**

- \( V \) is a (stochastically) degraded version of \( W \) if \( \exists T : \mathcal{Y} \to \mathcal{Z} \) s.t. \( V(z|x) = \sum_{y \in \mathcal{Y}} W(y|x) T(z|y) \ \forall x \in \mathcal{X}, z \in \mathcal{Z} \)
- \( W \) is less noisy than \( V \) if \( I(U; Y) \geq I(U; Z) \) \( \forall P_{U,X} \) where \( U \perp X \perp (Y, Z) \)
Notation & Definitions

- Given two DMCs $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$
- $X^n$ with $X_i$ i.i.d. Bernoulli$(p)$, $p \in [0,1]$;
- $Y^n = W^n X^n$ and $Z^n = V^n X^n$
- $U^n = G_n X^n$ with $G_n := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \log n$

For $\varepsilon > 0$ consider the two low-entropy sets

- $D^n_\varepsilon(X|Y) := \{ i \in [n] : H(U_i|U_{i-1}, Y^n) \leq \varepsilon \}$
- $D^n_\varepsilon(X|Z) := \{ i \in [n] : H(U_i|U_{i-1}, Z^n) \leq \varepsilon \}$

Definition: degraded, less noisy, more capable

- $V$ is a (stochastically) degraded version of $W$ if $\exists \ T : \mathcal{Y} \to \mathcal{Z}$ s.t. $V(z|x) = \sum_{y \in \mathcal{Y}} W(y|x) T(z|y) \ \forall x \in \mathcal{X}, z \in \mathcal{Z}$
- $W$ is less noisy than $V$ if $I(U; Y) \geq I(U; Z) \ \forall P_{U,X}$ where $U \rightarrow X \rightarrow (Y, Z)$
- $W$ is more capable than $V$ if $I(X; Y) \geq I(X; Z) \ \forall P_X$
Universality of Polar Codes — History & Contribution

- $W : \mathcal{X} \rightarrow \mathcal{Y}$ and $V : \mathcal{X} \rightarrow \mathcal{Z}$
- $\mathcal{D}_\varepsilon^n(X|Y) := \{ i \in [n] : H(U_i|U^{i-1}, Y^n) \leq \varepsilon \}$
- $\mathcal{D}_\varepsilon^n(X|Z) := \{ i \in [n] : H(U_i|U^{i-1}, Z^n) \leq \varepsilon \}$

Badly understood 😊

Relation between $\mathcal{D}_\varepsilon^n(X|Y)$ and $\mathcal{D}_\varepsilon^n(X|Z)$?

Would be extremely helpful for
- code construction
  - BEC is easy
  - channel up/downgrading [Tal-Vardy’11];
- network coding tasks
  - wiretap channel [Mahdavifar-Vardy’11, Şasoğlu-Vardy’13]
  - broadcast channel [Goela et al.’13]
  - ...
- quantum error correction [Renes et al.’13]
Universality of Polar Codes — History & Contribution

- $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$
- $\mathcal{D}^n_{\varepsilon}(X|Y) := \{i \in [n] : H(U_i|U^{i-1}, Y^n) \leq \varepsilon\}$
- $\mathcal{D}^n_{\varepsilon}(X|Z) := \{i \in [n] : H(U_i|U^{i-1}, Z^n) \leq \varepsilon\}$

Badly understood 😞

Relation between $\mathcal{D}^n_{\varepsilon}(X|Y)$ and $\mathcal{D}^n_{\varepsilon}(X|Z)$?

Would be extremely helpful for
- code construction
  - BEC is easy
  - channel up/downgrading [Tal-Vardy’11];
- network coding tasks
  - wiretap channel [Mahdavifar-Vardy’11, Şasşoğlu-Vardy’13]
  - broadcast channel [Goela et al.’13]
  - ...
- quantum error correction [Renes et al.’13]

Good news 😊

For specific classes of channels a few things are known
Universality of Polar Codes — History & Contribution

- $W: \mathcal{X} \rightarrow \mathcal{Y}$ and $V: \mathcal{X} \rightarrow \mathcal{Z}$
- $\mathcal{D}_{\varepsilon}^n(X|Y) := \{i \in [n]: H(U_i|U^{i-1}, Y^n) \leq \varepsilon\}$
- $\mathcal{D}_{\varepsilon}^n(X|Z) := \{i \in [n]: H(U_i|U^{i-1}, Z^n) \leq \varepsilon\}$

| $V$ degrad. w.r.t. $W$ | $\mathcal{D}_{\varepsilon}^n(X|Z) \subseteq \mathcal{D}_{\varepsilon}^n(X|Y)$ [Arikan’09] |
|----------------------|--------------------------------------------------|
| $W$ less noisy than $V$ |                                                   |
| $W$ more cap. than $V$ |                                                   |
| no relation          |                                                   |
Universality of Polar Codes — History & Contribution

- \( W : \mathcal{X} \rightarrow \mathcal{Y} \) and \( V : \mathcal{X} \rightarrow \mathcal{Z} \)
- \( \mathcal{D}_\varepsilon^n(X|Y) := \{ i \in [n] : H(U_i|U_{i-1}, Y^n) \leq \varepsilon \} \)
- \( \mathcal{D}_\varepsilon^n(X|Z) := \{ i \in [n] : H(U_i|U_{i-1}, Z^n) \leq \varepsilon \} \)

| V degrad. w.r.t. \( W \) | \( \mathcal{D}_\varepsilon^n(X|Z) \subseteq \mathcal{D}_\varepsilon^n(X|Y) \) [Arikan’09] |
|--------------------------|--------------------------------------------------|
| \( W \) less noisy than \( V \) | \( \mathcal{D}_\varepsilon^n(X|Z) \not\subseteq \mathcal{D}_\varepsilon^n(X|Y) \) [Hassani et al.’09] |
| \( W \) more cap. than \( V \) | |
| no relation | |

\( \mathcal{A} \subseteq \mathcal{B} \) means \( |\mathcal{A} \setminus \mathcal{B}| = o(n) \)
Universality of Polar Codes — History & Contribution

- $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$

$\mathcal{A} \subseteq \mathcal{B}$ means $|\mathcal{A}\setminus\mathcal{B}| = o(n)$

$\mathcal{D}_{\varepsilon}^n(X|Y) := \{i \in [n] : H(U_i|U_i^{i-1}, Y^n) \leq \varepsilon\}$

$\mathcal{D}_{\varepsilon}^n(X|Z) := \{i \in [n] : H(U_i|U_i^{i-1}, Z^n) \leq \varepsilon\}$

| $V$ degrad. w.r.t. $W$ | $\mathcal{D}_{\varepsilon}^n(X|Z) \subseteq \mathcal{D}_{\varepsilon}^n(X|Y)$ [Arıkan’09] |
|------------------------|--------------------------------------------------|
| $W$ less noisy than $V$|                                                                                  |
| $W$ more cap. than $V$ |                                                                                  |
| no relation            |                                                                                  |

- $\mathcal{D}_{\varepsilon}^n(X|Z) \nsubseteq \mathcal{D}_{\varepsilon}^n(X|Y)$ [Hassani et al.’09]
- using **optimal decoding** every good code for $V$ is also good for $W$ [Şaşoğlu’11]
### Universality of Polar Codes — History & Contribution

- $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$
- $D_{\delta}^n(X|Y) := \{ i \in [n] : H(U_i|U_i^{i-1}, Y^n) \leq \delta \}$
- $D_{\delta}^n(X|Z) := \{ i \in [n] : H(U_i|U_i^{i-1}, Z^n) \leq \delta \}$

| $V$ degrad. w.r.t. $W$ | $D_{\delta}^n(X|Z) \subseteq D_{\delta}^n(X|Y)$ [Arıkan’09] |
|--------------------------|--------------------------------------------------|
| $W$ less noisy than $V$ | $D_{\delta}^n(X|Z) \not\subseteq D_{\delta}^n(X|Y)$ [Hassani et al.’09] |
| $W$ more cap. than $V$ | using **optimal decoding** every good code for $V$ is also good for $W$ [Şaşoğlu’11] |
| no relation | modified protocols (cf. two previous talks) [Hassani-Urbanke’14], [Şaşoğlu-Wang’14] |
Universality of Polar Codes — History & Contribution

- $W : \mathcal{X} \rightarrow \mathcal{Y}$ and $V : \mathcal{X} \rightarrow \mathcal{Z}$
- $\mathcal{D}_\varepsilon^n(X|Y) := \{i \in [n] : H(U_i|U^{i-1}, Y^n) \leq \varepsilon\}$
- $\mathcal{D}_\varepsilon^n(X|Z) := \{i \in [n] : H(U_i|U^{i-1}, Z^n) \leq \varepsilon\}$

| $V$ degrad. w.r.t. $W$ | $\mathcal{D}_\varepsilon^n(X|Z) \subseteq \mathcal{D}_\varepsilon^n(X|Y)$ [Arıkan’09] |
|------------------------|-------------------------------------------------------------|
| $W$ less noisy than $V$| $\mathcal{D}_\varepsilon^n(X|Z) \subseteq \mathcal{D}_\varepsilon^n(X|Y)$ |
| $W$ more cap. than $V$| • $\mathcal{D}_\varepsilon^n(X|Z) \nsubseteq \mathcal{D}_\varepsilon^n(X|Y)$ [Hassani et al.’09] |
|                       | • using **optimal decoding** every good code for $V$ is also good for $W$ [Şaşoğlu’11] |
|                       | • for $P_X$ that maximizes $I(X; Y) - I(X; Z)$ and $\varepsilon = O(2^{-n^{0.49}})$, $\mathcal{D}_\varepsilon^n(X|Z) \subseteq \mathcal{D}_\varepsilon^n(X|Y)$ |
| no relation           | modified protocols (cf. two previous talks) [Hassani-Urbanke’14], [Şaşoğlu-Wang’14] |
Polar codes are universal for less noisy channels

Theorem: universality for less noisy channels

Let $W: \mathcal{X} \rightarrow \mathcal{Y}$ and $V: \mathcal{X} \rightarrow \mathcal{Z}$ be two DMCs such that $W$ is less noisy than $V$. Then for any $\varepsilon \in (0, 1)$, $n = 2^k$, $k \in \mathbb{N}$ we have $D^n_\varepsilon(X|Z) \subseteq D^n_\varepsilon(X|Y)$.

- Let $V$ and $W$ be symmetric. Every polar code built for $V$ can be used for $W$ with SC decoding
Polar codes are universal for less noisy channels

**Theorem: universality for less noisy channels**

Let $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ be two DMCs such that $W$ is less noisy than $V$. Then for any $\varepsilon \in (0, 1)$, $n = 2^k$, $k \in \mathbb{N}$ we have $D^n_\varepsilon(X|Z) \subseteq D^n_\varepsilon(X|Y)$.

- Let $V$ and $W$ be symmetric. Every polar code built for $V$ can be used for $W$ with SC decoding
- Recall that the class of less noisy channels is strictly larger than the class of degradable channels

**Example: BEC – BSC pair [El Gamal-Kim’11]**

Let $W = \text{BEC}(\alpha)$ for $\alpha \in (0, \frac{1}{2})$ and $V = \text{BSC}(\beta)$. Then

- $0 < \alpha \leq 2\beta$: $V$ is a degraded w.r.t. $W$
- $2\beta < \alpha \leq 4\beta(1 - \beta)$: $W$ is less noisy than $V$
Proof Sketch

**To show:** $D^n_\varepsilon(X|Z) \subseteq D^n_\varepsilon(X|Y)$

**Lemma 1:** [thanks to Chandra Nair]

Let $W$ and $V$ be two DMCs such that $W$ is less noisy than $V$. Then, $W^n$ is less noisy than $V^n$ for all $n \in \mathbb{N}$.

**Lemma 2:** [Csiszár-Körner’78]

Let $W : \mathcal{X} \rightarrow \mathcal{Y}$ and $V : \mathcal{X} \rightarrow \mathcal{Z}$ be two DMCs s.t. $W$ is more capable than $V$. Then $I(X; Y|U) \geq I(X; Z|U) \forall P_{U,X}$, where $U \rightarrow X \rightarrow (Y, Z)$. 
Proof Sketch

To show: \( \mathcal{D}^n_\epsilon(X|Z) \subseteq \mathcal{D}^n_\epsilon(X|Y) \)

Lemma 1: [thanks to Chandra Nair]
Let \( W \) and \( V \) be two DMCs such that \( W \) is less noisy than \( V \). Then, \( W^n \) is less noisy than \( V^n \) for all \( n \in \mathbb{N} \).

Lemma 2: [Csiszár-Körner’78]
Let \( W : \mathcal{X} \to \mathcal{Y} \) and \( V : \mathcal{X} \to \mathcal{Z} \) be two DMCs s.t. \( W \) is more capable than \( V \). Then \( I(X; Y|U) \geq I(X; Z|U) \) \( \forall P_{U,X} \), where \( U \leftarrow X \rightarrow (Y, Z) \).

Recall that
- \( \mathcal{D}^n_\epsilon(X|Y) := \{ i \in [n] : H(U_i|U^{i-1}, Y^n) \leq \epsilon \} \)
- \( \mathcal{D}^n_\epsilon(X|Z) := \{ i \in [n] : H(U_i|U^{i-1}, Z^n) \leq \epsilon \} \)

Lemma 1 implies \( H(U_1|Y^n) \leq H(U_1|Z^n) \)

To show: \( H(U_i|U^{i-1}, Y^n) \leq H(U_i|U^{i-1}, Z^n) \) for \( 2 \leq i \leq n \)
Proof Sketch (con’t)

**To show:** \( H(U_i|U^{i-1}, Y^n) \leq H(U_i|U^{i-1}, Z^n) \) for \( 2 \leq i \leq n \)

**Lemma 1:** [thanks to Chandra Nair]

Let \( W \) and \( V \) be two DMCs such that \( W \) is less noisy than \( V \). Then, \( W^n \) is less noisy than \( V^n \) for all \( n \in \mathbb{N} \).

**Lemma 2:** [Csiszár-Körner’78]

Let \( W : \mathcal{X} \rightarrow \mathcal{Y} \) and \( V : \mathcal{X} \rightarrow \mathcal{Z} \) be two DMCs s.t. \( W \) is more capable than \( V \). Then \( I(X; Y|U) \geq I(X; Z|U) \) \( \forall P_{U,X} \), where \( U \rightarrow X \leftarrow (Y,Z) \).

Consider the Markov chain \( U^{i-1} \rightarrow U^i \rightarrow X^n \rightarrow (Y^n, Z^n) \)

\[
H(U_i|U^{i-1}, Y^n) = H(U_i|U^{i-1}, Y^n) \leq H(U_i|U^{i-1}, Z^n) = H(U_i|U^{i-1}, Z^n)
\]

**Lemma 1 & Lemma 2**
Universality for more capable channels

- Let $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ be two DMCs s.t. $W$ is more capable than $V$
- In general, $D_n^\varepsilon(X|Z) \not\subseteq D_n^\varepsilon(X|Y)$, i.e., a polar code for $V$ cannot be used for $W$ under SC decoding [Hassani et al.’09]
Universality for more capable channels

- Let $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ be two DMCs s.t. $W$ is more capable than $V$.
- In general, $\mathcal{D}_\epsilon^n(X|Z) \not\subseteq \mathcal{D}_\epsilon^n(X|Y)$, i.e., a polar code for $V$ cannot be used for $W$ under SC decoding [Hassani et al.’09].

**Theorem: universality for more capable channels**

Let $P_X$ be such that it maximizes $I(X; Y) - I(X; Z)$. Then for $\epsilon = O(2^{-n\beta})$ with $\beta < \frac{1}{2}$, we have $\mathcal{D}_\epsilon^n(X|Z) \subseteq \mathcal{D}_\epsilon^n(X|Y)$.

Recall: $\mathcal{A} \not\subseteq \mathcal{B}$ means $|\mathcal{A}\setminus\mathcal{B}| = o(n)$

$$ \mathcal{D}_\epsilon^n(X|Y) \cup \mathcal{D}_\epsilon^n(X|Z) $$

$$ \mathcal{G}_\epsilon^n := \mathcal{D}_\epsilon^n(X|Z) \setminus \mathcal{D}_\epsilon^n(X|Y) $$

$$ |\mathcal{G}_\epsilon^n| = o(n) $$
Polar codes are universal for less noisy (symmetric) channels

For a specific input distribution, polar codes are universal for more capable channels

Can this be useful for code construction?

This new insights might be useful for multi-terminal coding tasks
  - wiretap channel [Mahdavifar-Vardy’11, Şaşoğlu-Vardy’13]
  - broadcast channel [Goela et al.’13]
  - ...