Achieving Secrecy Capacity With Polar Codes and Polar Lattices

Cong Ling
Imperial College London
cling@ieee.org

Sept. 23, 2015
1 Secrecy Coding

2 Polar Codes

3 Polar Lattices
   - Achieving Shannon Capacity
   - Achieving Secrecy Capacity
Wiretap channel

- M confidential message, $M'$ auxiliary message, $X^n$ codeword
- Assuming symmetry and channel degradation, secrecy capacity
  \[ C_s = C(\text{Bob}) - C(\text{Eve}). \]
- Our problem is to construct a (coset) code to achieve $C_s$. 
Security notions

- In information theory, information leakage is measured by mutual information:

**Definition (Strong secrecy)**

A wiretap code is information theoretically secure if $I(M; Z^n) \to 0$.

- In cryptography, it is measured by min-entropy or $l_1$ distance:

**Definition (Semantic security)**

A wiretap code is semantically secure if

$$\sup_{f, p_M} \left( e^{-H_\infty(f(M) | Z^n)} - e^{-H_\infty(f(M))} \right) \to 0.$$
A wiretap code achieves distinguishing security if

$$\max_{m,m'} \mathbb{V}(p_Z^n|M=m, p_Z^n|M=m') \to 0.$$
Techniques for secrecy coding

- **Coding methods**
  - LDPC codes: limited success (Suresh et al.’10).
  - Polar codes: semantically secure (Mahdavifar, Vardy’11).
  - Lattice codes: semantically secure (L., Luzzi, Belfiore, Stehle’12).
  - Resovability (Bloch, Laneman’13).

- **Extractors/hash functions**
  - Invertible extractors (Bellare et al.’12, Cheraghchi et al.’12, Chou et al.’14).
  - Universal hash functions (Hayashi et al.’10, Tyagi et al.’14).
(Binary) Polar codes are **capacity-achieving** for both binary memoryless symmetric (BMS) and binary memoryless asymmetric (BMA) channels.

- Encoding complexity is $O(N \log N)$.
- Using successive decoding, the decoding complexity is $O(N \log N)$.
- Block error probability decays like $2^{-\sqrt{N}}$.
- The construction is **deterministic**.
Among all channels, there are two classes for which it is easy to communicate optimally:

- The **perfect** channels: the output $Y$ completely determines the input $X$.

- The **useless** channels: the output $Y$ is independent of the input $X$.

**Polar coding** is a novel technique to convert any binary-input channel to a mixture of binary-input extremal channels.

- The converting process is **information-preserving**, and of low complexity.

- This technique can also be used for source coding (both lossless and lossy).
How do channels polarize?

Given two copies of a binary input channel \( W : X \rightarrow Y, X \in \mathbb{F}_2 \).
How do channels polarize?

Given two copies of a binary input channel $W : X \rightarrow Y$, $X \in \mathbb{F}_2$.

- Let

\[
X_1 = U_1 + U_2 \\
X_2 = U_2,
\]

i.e.,

\[
[X_1, X_2] = [U_1, U_2] \times \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.
\]
How do channels polarize?

Given two copies of a binary input channel $W : X \rightarrow Y$, $X \in \mathbb{F}_2$.

- Let
  
  \[
  X_1 = U_1 + U_2 \\
  X_2 = U_2,
  \]

  i.e.,
  \[
  [X_1, X_2] = [U_1, U_2] \times \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.
  \]

- This induces two synthetic channels $W^- : U_1 \rightarrow (Y_1, Y_2)$. 

\[ \begin{array}{c}
  \text{Diagram}
  \end{array} \]
How do channels polarize?

Given two copies of a binary input channel $W : X \to Y$, $X \in \mathbb{F}_2$.

- Let
  
  \[
  X_1 = U_1 + U_2 \\
  X_2 = U_2,
  \]
  
  i.e.,
  
  \[
  [X_1, X_2] = [U_1, U_2] \times \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.
  \]

- This induces two synthetic channels $W^- : U_1 \to (Y_1, Y_2)$ and $W^+ : U_2 \to (U_1, Y_1, Y_2)$. 

![Diagram](image.png)
How do channels polarize?

Properties of the process \((W, W) \rightarrow (W^-, W^+):\)

\[ I(W^-) = I(U_1; Y_1 Y_2) \]

\[ I(W^-) = I(U_1; Y_1 Y_2) \leq I(W) \leq I(W^-) \]
How do channels polarize?

Properties of the process \((W, W) \rightarrow (W^-, W^+):\)

\[ I(W^-) = I(U_1; Y_1 Y_2) \]
\[ I(W^+) = I(U_2; U_1 Y_1 Y_2) \]
How do channels polarize?

Properties of the process \((W, W) \rightarrow (W^-, W^+)\):

\[
I(W^-) = I(U_1; Y_1 Y_2)
\]
\[
I(W^+) = I(U_2; U_1 Y_1 Y_2)
\]

\[
I(W^-) + I(W^+) = I(U_1 U_2; Y_1 Y_2)
\]

\[
= I(X_1 X_2; Y_1 Y_2)
\]
How do channels polarize?

Properties of the process \((W, W) \rightarrow (W^-, W^+)\):

\[
I(W^-) = I(U_1; Y_1 Y_2) \\
I(W^+) = I(U_2; U_1 Y_1 Y_2)
\]

\[
I(W^-) + I(W^+) = I(U_1 U_2; Y_1 Y_2) = I(X_1 X_2; Y_1 Y_2)
\]

- Info-preserving:
  \[
  I(W^-) + I(W^+) = 2I(W).
  \]
How do channels polarize?

Properties of the process \((W, W) \rightarrow (W^-, W^+)\):

\[
I(W^-) = I(U_1; Y_1 Y_2)
\]

\[
I(W^+) = I(U_2; U_1 Y_1 Y_2)
\]

\[
I(W^-) + I(W^+) = I(U_1 U_2; Y_1 Y_2)
\]

\[
= I(X_1 X_2; Y_1 Y_2)
\]

- Info-preserving:
  \[
  I(W^-) + I(W^+) = 2I(W).
  \]
- Polarization:
  \[
  I(W^-) \leq I(W) \leq I(W^+).
  \]
Recursive polarization

Given $W : X \rightarrow Y$,

- Duplicate $W$ and obtain $W^-$ and $W^+$.

The transform corresponds to a generator matrix $G_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes m$. 
Recursive polarization

Given $W: X \rightarrow Y$,

- Duplicate $W$ and obtain $W^-$ and $W^+$.  
- Duplicate $W^-$ ($W^+$), and obtain $W^{--}$ and $W^{++}$ ($W^{+-}$ and $W^{+\cdot}$).
Recursive polarization

Given $W : X \rightarrow Y$,

- Duplicate $W$ and obtain $W^-$ and $W^+$.
- Duplicate $W^-$ ($W^+$), and obtain $W^{--}$ and $W^{++}$ ($W^{+-}$ and $W^{+--}$).
- Duplicate $W^{--}$, and obtain $W^{----}$ and $W^{---+}$, ....
Given $W : X \rightarrow Y$,

- Duplicate $W$ and obtain $W^-$ and $W^+$.
- Duplicate $W^-$ ($W^+$), and obtain $W^{--}$ and $W^{++}$ ($W^{+-}$ and $W^{+-}$).
- Duplicate $W^{--}$, and obtain $W^{----}$ and $W^{-----}$, ....
- Keep doing this recursively...

The transform corresponds to a generator matrix $G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes m$. 
Recursive polarization

Given $W : X \rightarrow Y$,

- Duplicate $W$ and obtain $W^-$ and $W^+$.
- Duplicate $W^-$ ($W^+$), and obtain $W^{---}$ and $W^{+++}$ ($W^{+-}$ and $W^{++-}$).
- Duplicate $W^{---}$, and obtain $W^{----}$ and $W^{-----}$, ....
- Keep doing this recursively...
- The transform corresponds to a generator matrix

$$G_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes m.$$
Synthetical channels

For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels $W_s^1 \cdots W_s^m$, $s_i \in \{+, -\}$, which can be also denoted by $W_i^{(i)}$.

$$\sum_i I(U_i; U_{i-1}, Y_{N,1}) = N \cdot I(W).$$

Indicating a successive cancellation (SC) decoding scheme:
- decode $U_1$ from $Y_{N,1}$;
- decode $U_2$ from $U_1, Y_{N,1}$;
- ...;
- decode $U_N$ from $U_{N-1}, Y_{N,1}$.
For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels

$$W^{s_1 \cdots s_m}, s_i \in \{+,-\},$$

which can be also denoted by

$$W^{(i)}_N, U_i \rightarrow U_1^{i-1}, Y_1^N, i = 1, \ldots N.$$
Synthetical channels

- For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels
  \[ W^{s_1 \cdots s_m}, s_i \in \{+, -\} \],
  which can be also denoted by
  \[ W^{(i)}_N, U_i \rightarrow U_1^{i-1}, Y_1^N, i = 1, \ldots, N. \]
  \[ \sum_i I(U_i; U_1^{i-1}, Y_1^N) = N \cdot I(W). \]
Synthesical channels

- For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthesical channels

$$W^{s_1 \cdots s_m}, s_i \in \{+, -\},$$

which can be also denoted by

$$W_N^{(i)}, U_i \rightarrow U_1^{i-1}, Y_1^N, i = 1, \ldots N.$$

$$\sum_i I(U_i; U_1^{i-1}, Y_1^N) = N \cdot I(W).$$

- Indicating a successive cancellation (SC) decoding scheme
Synthetical channels

- For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels $W^{s_1 \cdots s_m}$, $s_i \in \{+, -\}$,

which can be also denoted by

$$W_N^{(i)}, U_i \rightarrow U_1^{i-1}, Y_1^N, i = 1, \ldots, N.$$ 

$$\sum_i I(U_i; U_1^{i-1}, Y_1^N) = N \cdot I(W).$$ 

Indicating a successive cancellation (SC) decoding scheme

- decode $U_1$ from $Y_1^N$;
For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels

$$W^{s_1 \cdots s_m}, s_i \in \{+,-\},$$

which can be also denoted by

$$W^{(i)}_N, U_i \rightarrow U^{i-1}_1, Y^N_1, i = 1, \ldots, N.$$

$$\sum_i I(U_i; U^{i-1}_1, Y^N_1) = N \cdot I(W).$$

Indicating a successive cancellation (SC) decoding scheme
- decode $U_1$ from $Y^N_1$;
- decode $U_2$ from $U_1, Y^N_1$;
Synthetical channels

- For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels
  
  $W^{s_1 \cdots s_m}, s_i \in \{+, -\}$,

  which can be also denoted by

  $W_N^{(i)}, U_i \rightarrow U_1^{i-1}, Y_1^N, i = 1, \ldots, N$.

  $\sum_i I(U_i; U_1^{i-1}, Y_1^N) = N \cdot I(W)$.

  Indicating a successive cancellation (SC) decoding scheme

  - decode $U_1$ from $Y_1^N$;
  - decode $U_2$ from $U_1, Y_1^N$;
  - ...

  $U_1 \rightarrow \cdots \rightarrow Y_1$

  $U_N \rightarrow \cdots \rightarrow Y_N$
Synthetical channels

- For $m$ levels, we transform $N = 2^m$ uses of $W$ to a set of synthetical channels
  \[ W^{s_1 \cdots s_m}, s_i \in \{+, -\}, \]
  which can be also denoted by
  \[ W^{(i)}_N, U_i \rightarrow U_1^{i-1}, Y_1^N, i = 1, \ldots, N. \]
  \[ \sum_i I(U_i; U_1^{i-1}, Y_1^N) = N \cdot I(W). \]
  Indicating a successive cancellation (SC) decoding scheme
  - decode $U_1$ from $Y_1^N$;
  - decode $U_2$ from $U_1, Y_1^N$;
  - ...
  - decode $U_N$ from $U_1^{N-1}, Y_1^N$;
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, $\frac{\#\{i: I(W_N^{(i)}) \in (0.1, 0.9)\}}{N} = 0.0938$. 
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, \[
\frac{\#\{i : I(W_N^{(i)}) \in (0.1, 0.9)\}}{N} = 0.0938.
\]
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, $\frac{\#\{i : I(W_N^{(i)}) \in (0.1, 0.9)\}}{N} = 0.0938$. 
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, 
\[
\frac{\# \{ i : I(W_N^{(i)}) \in (0.1, 0.9) \}}{N} = 0.0938.
\]
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, $\frac{\#\{i: I(W_N^{(i)}) \in (0.1, 0.9)\}}{N} = 0.0938$. 
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$,

$$\frac{\#\{i: l(W_N^{(i)}) \in (0.1, 0.9)\}}{N} = 0.0938.$$
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, \[
\frac{\#\{i : I(W_N^{(i)} \in (0.1,0.9)} \} \quad N = 0.0938.
\]
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, $\frac{\# \{i: I(W^{(i)}_N) \in (0.1, 0.9) \}}{N} = 0.0938$. 
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, $\frac{\# \{ i : l(W^{(i)}_N) \in (0.1, 0.9) \} }{N} = 0.0938$. 
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, \[
\frac{\#\{i: I(W_N^{(i)}) \in (0.1, 0.9)\}}{N} = 0.0938.
\]
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, $\frac{\#\{i: I(\frac{W_i}{N}) \in (0.1,0.9)\}}{N} = 0.0938$. 
An example of polarization

When $W = \text{BEC}(0.5)$ and $N = 2^{12}$, \[
\frac{\#\{i: I(W^{(i)}_N) \in (0.1, 0.9)\}}{N} = 0.0938.
\]
Polarization speed and Bhattacharyya parameter

Definition (Symmetric Bhattacharyya)

For a BMS $W : X \rightarrow Y$, the Bhattacharyya parameter of $W$ is defined by

$$\tilde{Z}(W) = \sum_{y \in Y} \sqrt{W(y|0)W(y|1)}.$$  

- $\tilde{Z}(W) \in [0, 1]$;
- $\tilde{Z}(W) \approx 0$, iff $I(W) = 1$;
- $\tilde{Z}(W) \approx 1$, iff $I(W) = 0$;

Theorem (Polarization speed)

For any $0 < \beta < 0.5$ and $N \rightarrow \infty$,

$$\frac{1}{N} \# \{i : \tilde{Z}(W_N^{(i)}) \leq 2^{-N^\beta} \} = I(W).$$

Moreover, using these good indices $i$ guarantees a block error probability $P_e \leq N2^{-N^\beta}$ under SC decoding.
- Calculate the Bhattacharyya parameters of all synthetical channels.

- Given a rate $R$ and $k = RN$, sort $\{Z(W_N^{(i)})\}$ in an ascending order.

- **Good indexes**: put information bits on set $\Omega$ of the first $k$ elements.

- **Bad indexes**: put frozen bits on $\Omega^c$.

- We can also use $\{I(W_N^{(i)})\}$ to construct a code.
Calculate the Bhattacharyya parameters of all synthetical channels.

Given a rate $R$ and $k = RN$, sort $\{Z(W_{N}^{(i)})\}$ in an ascending order.

**Good indexes:** put information bits on set $\Omega$ of the first $k$ elements.

**Bad indexes:** put frozen bits on $\Omega^c$.

We can also use $\{I(W_{N}^{(i)})\}$ to construct a code.
Calculate the Bhattacharyya parameters of all synthetical channels.

Given a rate $R$ and $k = RN$, sort $\{Z(W_{N}^{(i)})\}$ in an ascending order.

**Good indexes:** put information bits on set $\Omega$ of the first $k$ elements.

**Bad indexes:** put frozen bits on $\Omega^c$.

We can also use $\{I(W_{N}^{(i)})\}$ to construct a code.
- Calculate the Bhattacharyya parameters of all synthetical channels.
- Given a rate $R$ and $k = RN$, sort $\{Z(W_N^{(i)})\}$ in an ascending order.
- **Good indexes:** put information bits on set $\Omega$ of the first $k$ elements.
- **Bad indexes:** put frozen bits on $\Omega^c$.
- We can also use $\{I(W_N^{(i)})\}$ to construct a code.
Calculate the Bhattacharyya parameters of all synthetical channels.

Given a rate $R$ and $k = RN$, sort $\{Z(W_N^{(i)})\}$ in an ascending order.

**Good indexes:** put information bits on set $\Omega$ of the first $k$ elements.

**Bad indexes:** put frozen bits on $\Omega^c$.

We can also use $\{I(W_N^{(i)})\}$ to construct a code.
- Calculate the Bhattacharyya parameters of all synthetical channels.
- Given a rate $R$ and $k = RN$, sort $\{Z(W_N^{(i)})\}$ in an ascending order.
- **Good indexes:** put information bits on set $\Omega$ of the first $k$ elements.
- **Bad indexes:** put frozen bits on $\Omega^c$.
- We can also use $\{I(W_N^{(i)})\}$ to construct a code.
Calculate the Bhattacharyya parameters of all synthetical channels.

Given a rate $R$ and $k = RN$, sort $\{Z(W_N^{(i)})\}$ in an ascending order.

Good indexes: put information bits on set $\Omega$ of the first $k$ elements.

Bad indexes: put frozen bits on $\Omega^c$.

We can also use $\{I(W_N^{(i)})\}$ to construct a code.
Calculate the Bhattacharyya parameters of all synthetical channels.

Given a rate $R$ and $k = RN$, sort $\{Z(W_N^{(i)})\}$ in an ascending order.

**Good indexes:** put information bits on set $\Omega$ of the first $k$ elements.

**Bad indexes:** put frozen bits on $\Omega^c$.

We can also use $\{I(W_N^{(i)})\}$ to construct a code.
BMS wiretap channel with main channel \( V \) and eavesdropping channel \( W \) (Mahdavifar, Vardy’09).

Good indexes: \( G(V) = \{ i : Z(V_N^{(i)}) \leq 2^{-N\beta} \} \).

Bad indexes: \( N(W) = \{ i : Z(W_N^{(i)}) \geq 1 - 2^{-N\beta} \} \Rightarrow I(W_N^{(i)}) \leq 2^{-N\beta'}, \quad 0 < \beta' < \beta < 0.5 \).

Partition
\[
\mathcal{M} = G(V) \cap N(W) \\
\mathcal{R} = G(V) \cap N(W)^c \\
\mathcal{F} = G(V)^c \cap N(W) \\
\mathcal{R} = G(V)^c \cap N(W)^c.
\]
Semantic security

Main result:

\[ I(MF; Z^N) \leq \sum_{i \in \mathcal{R}^c} I(W_N^{(i)}) = \sum_{i \in \mathcal{N}(W)} I(W_N^{(i)}) \leq N2^{-N\beta'}. \]

Semantic security: the induced channel is symmetric, thus uniform distribution maximizes \( I(MF; Z^N) \).

Dealing with ugly bits in \( \mathcal{G}(V)^c \cap \mathcal{N}(W)^c \)

- Correctness of these bits are not guaranteed by SC decoding.
- However, their fraction tends to zero as \( N \to \infty \).
- Can be handled by multi-block transmission (Sasoglu, Vardy’13).
- Open question: Is there an improved single-block decoder achieving reliability?

The technique can be adapted to secret key generation (Chou, Bloch, Abbe’13).
In general, the channel is not necessarily symmetric or degraded.

Secrecy capacity

\[ C_s = \max_{V \rightarrow X \rightarrow Y, Z} I(V; Y) - I(V; Z) \]

- Using universal polar codes (Wei, Ulukus’14)
- Using chaining (multi-block) transmission (Gulcu, Barg’14)
Overview of lattice codes

- We use lattices to construct capacity-achieving codes for the Gaussian channel.
- From the coding point of view, a lattice serves as the interface between Hamming space and Euclidean space.
- For the Gaussian wiretap channel, we need
  - A lattice code achieving the Shannon capacity of Bob’s channel;
  - A secrecy-good lattice for Eve’s channel.
- No efficient schemes exist up until now.
The capacity of the AWGN channel is $\frac{1}{2} \log(1 + SNR)$ (Shannon, 1948).

Lattice codes have been proved to achieve the AWGN channel capacity $\frac{1}{2} \log(1 + SNR)$:

- **Coding**: A lattice $\Lambda$ is **AWGN-good** if it achieves a vanishing error probability as long as its volume-to-noise ratio (VNR) $\geq 2\pi e$.
- **Shaping**: for power constraint
  - Voronoi shaping: The shaping lattice is good for quantization. (Erez, Zamir’04);
  - Gaussian shaping: Applying a discrete Gaussian distribution over an AWGN-good lattice. (Ling, Belfiore’14).

No explicit construction until recently.
Construction of AWGN-good lattices

- Explicit constructions: LDA lattice (di Pietro, Zemor, Boutros’13) and polar lattice (Yan, L., Wu’13).
- Polar lattice is a combination of polar codes and lattice Construction D.
- Standing on the shoulders of giants:
  - Polar codes are capacity achieving for BMCs
  - Sphere-bound-achieving lattices

Arikan

Forney

AWGN good polar lattice
Polar lattice

- Construction D: A lattice $L$ constructed by an $n$-dimensional binary lattice partition chain $\Lambda_1/\cdots/\Lambda_r$ and and $r-1$ binary nested codes $C_1(N, k_1) \subseteq C_2(N, k_2) \cdots \subseteq C_{r-1}(N, k_{r-1})$.

- Let $g_1, \cdots, g_{k_\ell}$ be a basis of $C_\ell$ for $1 \leq \ell \leq r$. Lattice $L$ can be described by

$$L = \left\{ \sum_{\ell=1}^r 2^{\ell-1} \sum_{i=1}^{k_\ell} u_{i\ell}^i g_i + 2^r \mathbb{Z}^N \mid u_{i\ell}^i \in \{0, 1\} \right\}.$$ 

Definition (Polar lattice)

A polar lattice $L$ is constructed by a set of binary nested polar codes $\mathcal{P}_1(N, k_1) \subseteq \mathcal{P}_2(N, k_2) \subseteq \cdots \subseteq \mathcal{P}_r(N, k_r)$ according to Construction D.
Nested polar codes

Using nested Polar codes to construct Polar lattices

\( \tilde{W} \) is a degraded version of \( W \)

\( \tilde{W}^{(i)}_N \) is also degraded to \( W^{(i)}_N \)

\[ Z(\tilde{W}^{(i)}_N) \geq Z(W^{(i)}_N) \]

\[ \{ i : Z(W^{(i)}_N) < 2^{-N^a} \}, \quad \tilde{A} \subseteq A \]

The channel of \( i \)-th level is a degraded version of the channel of \( i + 1 \)-th level.

Conclusion

The polar codes in our multilevel construction are nested. We satisfy the requirement of Construction D.
Simulation results

Block error probabilities of polar lattices with length $N = 1024$ and $N = 8192$ under multistage decoding.
Comparisons

Performance Comparison

Symbol error rate comparison with existing lattices.

<table>
<thead>
<tr>
<th>Lattices</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDPC Lattices</td>
<td>1000</td>
</tr>
<tr>
<td>LDLC</td>
<td>1000</td>
</tr>
<tr>
<td>LDA</td>
<td>1000</td>
</tr>
<tr>
<td>Polar Lattices</td>
<td>1024</td>
</tr>
</tbody>
</table>

Decoding Complexity

- $30 \times r \times O(N \log N), r = 2$
- $200 \times 7 \times 256 \log(256) O(N)$
- $p^2 \times O(N \log N), p = 11$
- $r \times O(N \log N), r = 2$
Gaussian shaping

- We propose **lattice Gaussian coding**: discrete Gaussian distribution over a lattice (L., Belfiore’13).
  - Why Gaussian? Gaussian is capacity-achieving.
  - Why lattice Gaussian? Continuous Gaussian is not implementable; needs structure.
- Only one lattice (AWGN-good) is required.
Discrete Gaussian distribution

- Standard continuous Gaussian distribution

\[ f_{\sigma,c}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|x-c\|^2}{2\sigma^2}} \]

- Discrete Gaussian distribution over lattice \( \Lambda \)

\[ D_{\Lambda,\sigma,c}(\lambda) = \frac{f_{\sigma,c}(\lambda)}{f_{\sigma,c}(\Lambda)}, \quad \forall \lambda \in \Lambda, \]

where

\[ f_{\sigma,c}(\Lambda) \triangleq \sum_{\lambda \in \Lambda} f_{\sigma,c}(\lambda). \]
 Periodic Continuous Gaussian distribution

- \( \Lambda \)-periodic function

\[
f_{\sigma, \Lambda}(x) = \frac{1}{(\sqrt{2\pi\sigma})^n} \sum_{\lambda \in \Lambda} e^{-\frac{\|x - \lambda\|^2}{2\sigma^2}},
\]

- \( f_{\sigma, \Lambda} \) restricted to the quotient \( \mathbb{R}^n / \Lambda \) is a (continuous) probability density.

- It arises, e.g., when Gaussian noise passes through a mod-\( \Lambda \) operator.

- It gets flat as \( \sigma \) increases.
Discrete and continuous versions of lattice Gaussian are the Fourier dual of each other.

Fourier series expansion on the dual lattice $\Lambda^*$

\[
f_{\sigma,\Lambda}(x) = \frac{1}{V(\Lambda)} \sum_{\lambda^* \in \Lambda^*} \hat{f}_\sigma(\lambda^*) e^{j2\pi \langle \lambda^*, x \rangle}
\]

\[
\hat{f}_\sigma(y) = \int f_{\sigma}(x) e^{-j2\pi \langle x, y \rangle} dx = e^{-2\pi^2 \sigma^2 \|y\|^2}
\]
- **Λ-periodic function**

\[ f_{\sigma, \Lambda}(x) = \frac{1}{(\sqrt{2\pi\sigma})^n} \sum_{\lambda \in \Lambda} e^{-\frac{||x-\lambda||^2}{2\sigma^2}}, \]

- \( f_{\sigma, \Lambda} \) restricted to the quotient \( \mathbb{R}^n/\Lambda \) is a (continuous) probability density.
- It arises, e.g., when Gaussian noise passes through a mod-Λ operator.
- It gets flat as \( \sigma \) increases.
**Periodic Continuous Gaussian distribution**

- **Λ-periodic function**

\[
f_{\sigma,\Lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda \in \Lambda} e^{-\frac{\|x-\lambda\|^2}{2\sigma^2}},
\]

- \(f_{\sigma,\Lambda}\) restricted to the quotient \(\mathbb{R}^n/\Lambda\) is a (continuous) probability density.
- It arises, e.g., when Gaussian noise passes through a mod-\(\Lambda\) operator.
- It gets flat as \(\sigma\) increases.
Periodic Continuous Gaussian distribution

- \( \Lambda \)-periodic function

\[
f_{\sigma, \Lambda}(x) = \frac{1}{(\sqrt{2\pi\sigma})^n} \sum_{\lambda \in \Lambda} e^{-\frac{\|x - \lambda\|^2}{2\sigma^2}},
\]

- \( f_{\sigma, \Lambda} \) restricted to the quotient \( \mathbb{R}^n / \Lambda \) is a (continuous) probability density.

- It arises, e.g., when Gaussian noise passes through a mod-\( \Lambda \) operator.

- It gets flat as \( \sigma \) increases.
**Periodic Continuous Gaussian distribution**

- \( \Lambda \)-periodic function

\[
f_{\sigma, \Lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda \in \Lambda} e^{-\frac{\|x-\lambda\|^2}{2\sigma^2}},
\]

- \( f_{\sigma, \Lambda} \) restricted to the quotient \( \mathbb{R}^n / \Lambda \) is a (continuous) probability density.

- It arises, e.g., when Gaussian noise passes through a mod-\( \Lambda \) operator.

- It gets flat as \( \sigma \) increases.
For a lattice $\Lambda$ and for a parameter $\sigma$, the flatness factor is defined by:

$$\epsilon_{\Lambda}(\sigma) \triangleq \max_{x \in \mathcal{R}(\Lambda)} | V(\Lambda)f_{\sigma,\Lambda}(x) - 1 |$$

where $\mathcal{R}(\Lambda)$ is a fundamental region.

$\epsilon_{\Lambda}(\sigma)$ quantifies the maximum variation of $f_{\sigma,\Lambda}(x)$. It is the other facet of the smoothing parameter (Micciancio, Regev’05).
How to do Gaussian shaping?

- Start from a good constellation $X \sim D_{\Lambda, \sigma_s}$.
- If $\epsilon_{\Lambda}(\tilde{\sigma}) \to 0$ where $\tilde{\sigma} = \frac{\sigma_s\sigma}{\sqrt{\sigma_s^2 + \sigma^2}}$

$$I(X; Y) \to \frac{1}{2} \log (1 + \text{SNR}).$$

- Multilevel coding: $X_{1:r} \to D_{\Lambda, \sigma_s}$ where $\Lambda$ is the top lattice.
Input distributions $P_{X_1}, P_{X_2|X_1}, \ldots, P_{X_r|X_1:r-1}$ are generally nonuniform.

By the chain rule of mutual information,

$$I(Y; X_1:r) = \sum_{\ell=1}^{r} I(Y; X_{\ell}|X_1:\ell-1).$$

Let $W_{\ell}$ denote the binary channel corresponds to $I(Y; X_{\ell}|X_1:\ell-1)$. Clearly, $W_{\ell}$ is generally asymmetric.

The afore-mentioned polar coding technique only applies to BMS channels. To achieve capacity $I(Y; X_{\ell}|X_1:\ell-1)$, we need the polar coding technique for asymmetric channels.

The two techniques are closely related.
More general Bhattacharyya parameter

**Definition (Bhattacharyya Parameter for BMA Channels)**

Let $W$ be a BMA channel with input $X$ and output $Y$, and let $P_X$ and $P_{Y|X}$ denote the input distribution and channel transition probability, respectively. The Bhattacharyya parameter channel $W$ is defined as

$$Z(X|Y) = 2 \sum_y \sqrt{P_{X,Y}(0,y)P_{X,Y}(1,y)}.$$ 

- It is the same as the Bhattacharyya parameter $\tilde{Z}(W)$ for BMS channels when $P_X$ is uniform.
- When $P_X$ is nonuniform, $Z(X|Y)$ can be converted to a symmetric Bhattacharyya parameter $\tilde{Z}(\tilde{W})$ for some BMS channel $\tilde{W}$. 
Lemma (Symmetrization)

Let $\tilde{W}$ be a binary input channel with input $\tilde{X} \in \mathcal{X} = \{0, 1\}$ and output $\tilde{Y} \in \{Y, X\}$, built from the asymmetric channel $W$ as in the following figure. Then $\tilde{W}$ is a binary symmetric channel in the sense that

$$P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x}|\tilde{x}) = P_{Y,X}(y, x).$$

Moreover, $Z(W) = \tilde{Z}(\tilde{W})$. 
Lemma (Symmetrization)

Let \( \tilde{W} \) be a binary input channel with input \( \tilde{X} \in X = \{0, 1\} \) and output \( \tilde{Y} \in \{Y, X\} \), built from the asymmetric channel \( W \) as in the following figure. Then \( \tilde{W} \) is a binary symmetric channel in the sense that

\[
P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x}|\tilde{x}) = P_{Y,X}(y, x). 
\]

Moreover, \( Z(W) = \tilde{Z}(\tilde{W}) \).

Example:
We know how to construct a polar code for $\tilde{\mathcal{W}}$. Let $\tilde{U}^{1:N} = \tilde{X}^{1:N} G_N$ and $U^{1:N} = X^{1:N} G_N$. 

\[ \begin{array}{c}
\tilde{F} \\
\tilde{I} \\
\tilde{U}^{1:N}
\end{array} \]
We know how to construct a polar code for $\tilde{\mathcal{W}}$. Let $\tilde{U}^{1:N} = \tilde{X}^{1:N} G_N$ and $U^{1:N} = X^{1:N} G_N$.

- $\tilde{F}$
- $\tilde{I}$

\[ \tilde{F} \]

\[ \tilde{U}^{1:N} \]

\[ \tilde{I} \]

almost random
given $\tilde{U}^{1:i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N}$

almost deterministic
given $\tilde{U}^{1:i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N}$
We know how to construct a polar code for $\tilde{W}$. Let $\tilde{U}^{1:N} = \tilde{X}^{1:N} G_N$ and $U^{1:N} = X^{1:N} G_N$. 
We know how to construct a polar code for $\tilde{\mathcal{W}}$. Let $\tilde{U}^{1:N} = \tilde{X}^{1:N} G_N$ and $U^{1:N} = X^{1:N} G_N$. 

\[ \tilde{X}^{1:N} = X^{1:N} \oplus \tilde{Y}^{1:N} \]

\[ U^{1:N} = U^{1:N} \oplus Y^{1:N} \]

\[ 1 - H(X) \]

Almost random given $\tilde{U}^{i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N}$

Almost deterministic given $U^{i-1}, Y^{1:N}$

Almost deterministic given $U^{i-1}, Y^{1:N}$

Almost random given $U^{i-1}$

Almost deterministic given $U^{i-1}, Y^{1:N}$ and almost random given $U^{i-1}$
We know how to construct a polar code for $\tilde{W}$. Let $\tilde{U}^{1:N} = \tilde{X}^{1:N} G_N$ and $U^{1:N} = X^{1:N} G_N$.

$S$ is a shaping set with proportion $1 - H(X)$. If $P_X$ is uniform, $|S| = 0$ and the two codes are identical.
The shaping for level 1 is done. For the second level, we firstly shape it according to $P_{X_2}$.
The shaping for level 1 is done. For the second level, we firstly shape it according to $P_{X_2}$. Since $X_1$ and $X_2$ are also correlated, we then refine the shaping according to the realization of $X_1$ and make up $P_{X_2|X_1}$. 

\[ I(X_2; X_1) \]

\[ U_{2}^{1:N} \]

\[ U_{2}^{1:i-1} \]

\[ Y_{1:N} \]

\[ X_1^{1:N} \]
The shaping for level 1 is done. For the second level, we firstly shape it according to $P_{X_2}$. Since $X_1$ and $X_2$ are also correlated, we then refine the shaping according to the realization of $X_1$ and make up $P_{X_2|X_1}$. Keep doing this until level $r$. 

\[ I(X_2; X_1) \]
Channel equivalence

Lemma (Equivalence lemma)

Consider a multilevel lattice code constructed from constellation $D_{\Lambda, \sigma_s}$ for a Gaussian channel with noise variance $\sigma^2$. The $\ell$-th symmetric channel $\tilde{W}_\ell$ ($1 \leq \ell \leq r$) which is derived from the asymmetric channel $W_\ell$ is equivalent to the MMSE-scaled $\Lambda_{\ell-1}/\Lambda_\ell$ channel with noise variance $\tilde{\sigma}^2 = \frac{\sigma_s^2 \sigma^2}{\sigma_s^2 + \sigma^2}$.

- The decoding of polar codes for $W_\ell$ can also be converted to that for $\tilde{W}_\ell$.
- The resultant polar lattice is AWGN-good for noise variance $\tilde{\sigma}^2$.
- Consistent with the result that the AWGN capacity is achieved with Gaussian shaping over an AWGN-good lattice and MMSE lattice decoding. (L., Belfiore’13)
Choose a good constellation with negligible flatness factor $\epsilon_{\Lambda}(\tilde{\sigma})$, and construct a multilevel polar code with $r = O(\log \log N)$. Then, for any SNR, the message rate approaches $\frac{1}{2} \log(1 + \text{SNR})$, while the error probability under multistage decoding is bounded by

$$P_e \leq r N 2^{-N\beta'}, \quad 0 < \beta' < 0.5$$

as $N \to \infty$.

- The scheme is fully explicit, with overall complexity $O(N \log N)$.
- One can say that the problem of constructing an efficient code to achieve the capacity of the Gaussian channel is finally settled.
Simulation results

The proportions of the shaping set, information set, and frozen set on each level when $N = 2^{16}$ and SNR = 15 dB.
Simulation results

**Figure:** Lower bounds on the rates achieved by polar lattices with block error probability $5 \times 10^{-5}$ for block lengths $2^{10}$, $2^{12}$, ..., $2^{20}$. 
Wiretap lattice code

- $\Lambda_e \subset \Lambda_b$ chain of lattices in $\mathbb{R}^n$, such that (L., Luzzi, Belfiore, Stehlé’13)
  - $\Lambda_b$ AWGN-good $\Rightarrow$ reliability for Bob
  - $\Lambda_e$ secrecy-good $\Rightarrow$ secrecy against Eve

Nesting ratio: $|\Lambda_b/\Lambda_e| = \lceil e^{Rn} \rceil$, corresponding to secrecy rate $R$.

- Each confidential message $m = 1, \ldots, \lceil e^{nR} \rceil$ is associated to a coset $\Lambda_e + \lambda_m$, $\lambda_m \in [\Lambda_b/\Lambda_e]$
The key idea

- Forget about shaping for now.
- Given message $m$, Alice samples a lattice point uniformly at random from a coset $\Lambda_e + \lambda_m$.
- Due to the channel noise, Eve observes the periodic distribution

$$\frac{1}{(\sqrt{2\pi}\sigma_e)^n} \sum_{\lambda \in \Lambda + \lambda_m} e^{-\frac{\|z-\lambda\|^2}{2\sigma_e^2}}.$$

- If the flatness factor $\epsilon_{\Lambda_e}(\sigma_e)$ is small, it will be close to a uniform distribution, regardless of message $m$.
- Problem: one cannot really sample a point uniformly from a lattice (or its coset).
Gaussian sampling

- Alice actually samples $X^n_m$ from lattice Gaussian distribution

$$X^n_m \sim D_{\Lambda_e + \lambda m, \sigma_s}.$$  

Lemma (Regev’09)

Let $\tilde{\sigma} = \frac{\sigma_s \sigma}{\sqrt{\sigma_s^2 + \sigma^2}}$ and $\sigma'_s = \sqrt{\sigma_s^2 + \sigma^2}$. If $\varepsilon = \epsilon_{\Lambda_e}(\tilde{\sigma}) < \frac{1}{2}$, the continuous distribution resulting from adding Gaussian noise of variance $\sigma^2$ to a discrete Gaussian $D_{\Lambda-c, \sigma_s}$ is uniformly close to $f_{\sigma'_s}(x)$ with $L_\infty$ distance bounded by $4 \varepsilon$.

- It implies that if $\epsilon_{\Lambda_e}(\tilde{\sigma}_e) < \frac{1}{2}$, then:

$$\nabla \left( p_{Z^n|M}(\cdot|m), f_{\sigma'_s} \right) \leq 4 \epsilon_{\Lambda_e}(\tilde{\sigma}_e).$$

- Eve’s received signals converge to the same Gaussian distribution $f_{\sigma'_s}$. This already gives distinguishing security.
Secrecy-good lattice

- Mutual information for Eve is bounded as follows (L., Luzzi, Belfiore, Stehle’13):

\[ I(M; Z^n) \leq 8\varepsilon_n nR - 8\varepsilon_n \log 8\varepsilon_n. \]  

(1)

- A sequence of lattices \( \Lambda^{(n)} \) is secrecy-good if

\[ \epsilon_{\Lambda^{(n)}}(\sigma) = e^{-\Omega(n)}, \quad \forall \gamma_{\Lambda^{(n)}}(\sigma) < 2\pi. \]  

(2)

- Strong secrecy has been obtained. Since we make no a priori assumption on the distribution of \( m \), it also achieves semantic security.

- However, it is hard to construct a lattice with an exponentially vanishing flatness factor.
Construction of secrecy-good lattice

**Definition (Relaxed definition)**

A lattice \( \Lambda_e \) is secrecy-good if it results in vanishing information leakage \( I(\mathcal{M}; Z^N) \).

- Apply the secrecy-coding construction to each level: \( \mathcal{V}_\ell = W(\Lambda_\ell/\Lambda_{\ell+1}, \tilde{\sigma}_b^2) \) and \( W_\ell = W(\Lambda_\ell/\Lambda_{\ell+1}, \tilde{\sigma}_e^2) \), \( I(\mathcal{M}_\ell; Z^{N \mod \Lambda_{\ell+1}}) \leq N2^{-N\beta} \).

- \( \mathcal{A}_\ell = \mathcal{G}(\mathcal{V}_\ell) \cap \mathcal{N}(W_\ell) \)
  \( \mathcal{B}_\ell = \mathcal{G}(\mathcal{V}_\ell) \cap \mathcal{N}(W_\ell)^c \)
  \( \mathcal{C}_\ell = \mathcal{G}(\mathcal{V}_\ell)^c \cap \mathcal{N}(W_\ell) \)
  \( \mathcal{D}_\ell = \mathcal{G}(\mathcal{V}_\ell)^c \cap \mathcal{N}(W_\ell)^c \).

- \( \Lambda_b \):
  \( C_1(N, |\mathcal{A}_1| + |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \cdots \subseteq C_{r-1}(N, |\mathcal{A}_{r-1}| + |\mathcal{B}_{r-1}| + |\mathcal{D}_{r-1}|) \).

- \( \Lambda_e \):
  \( C_1(N, |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \cdots \subseteq C_{r-1}(N, |\mathcal{B}_{r-1}| + |\mathcal{D}_{r-1}|) \).

- \( \Lambda_e \subseteq \Lambda_b \). \( M \rightarrow [\Lambda_b/\Lambda_e] \): \( \lambda_m \in \Lambda_b/\Lambda_e \).
Security proof

- Same polar codes for $W(Z; X_\ell | X_1, \ldots, X_{\ell-1})$ (by equivalence lemma):

  $$I(M_{\ell}; Z^N, X_1^N, \ldots, X_{\ell-1}^N) \leq N2^{-N\beta}.$$  

- \( I(Z^N; M) = \sum_{i=1}^{r} I(Z^N; M_{\ell}| M_1: \ell-1) \)

  $$= \sum_{\ell=1}^{r} h(M_{\ell}) - h(M_{\ell}| Z^N, M_1: \ell-1)$$

  $$= \sum_{\ell=1}^{r} I(M_{\ell}; Z^N, M_1: \ell-1)$$

  $$\leq \sum_{\ell=1}^{r} I(M_{\ell}; Z^N, X_1^N| X_{1: \ell-1}) \leq rN2^{-N\beta'}.$$  

- Therefore strong secrecy is achieved as \( \lim_{N \to \infty} I(M; Z^N) = 0. \)
Secrecy rate

Theorem (Yan, Liu, L.’14)

Consider $\Lambda_b$ and $\Lambda_e$ constructed with the binary lattice partition chain $\Lambda_1/\cdot \cdot \cdot /\Lambda_r$ and $r-1$ binary nested polar codes with block length $N$. By scaling $\Lambda_1$ and $r$ to satisfy the following conditions:

(i) $h(\Lambda_1, \tilde{\sigma}_b^2) \rightarrow \log V(\Lambda_1)$

(ii) $h(\Lambda_r, \tilde{\sigma}_e^2) \rightarrow \frac{1}{2} \log(2\pi e \tilde{\sigma}_e^2)$,

given $\tilde{\sigma}_e^2 > \tilde{\sigma}_b^2$, as $N \rightarrow \infty$, all strong secrecy rates $R$ satisfying

$$R = \sum_{\ell=1}^{r} \lim_{N \rightarrow \infty} \frac{|A_{\ell}|}{N} < \frac{1}{2} \log \frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2} = \frac{1}{2} \log \left( \frac{1 + \text{SNR}_b}{1 + \text{SNR}_e} \right)$$

are achievable for the Gaussian wiretap channel.
Shaping

- Reliable indices $G_\ell \triangleq \{ i \in [N] : Z(U_1^i \mid U_1^{1:i-1}, X_1^N, Y^N) \leq 2^{-N^{1/2}} \}$.
- Secure $N_\ell \triangleq \{ i \in [N] : Z(U_1^i \mid U_1^{1:i-1}, X_1^N, Z_1^N) \geq 1 - 2^{-N^{1/2}} \}$.
- Shaping set $S_\ell \triangleq \{ i \in [N] : Z(U_1^i \mid U_1^{1:i-1}, X_1^N) < 1 - 2^{-N^{1/2}} \}$.

$A^S_\ell = G_\ell \cap N_\ell \cap S_\ell$, $A^{Sc}_\ell = G_\ell \cap N_\ell \cap S^{Sc}_\ell$,
$B^S_\ell = G_\ell \cap N^{Sc}_\ell \cap S_\ell$, $B^{Sc}_\ell = G_\ell \cap N^{Sc}_\ell \cap S^{Sc}_\ell$,
$C^S_\ell = G^{Sc}_\ell \cap N_\ell \cap S_\ell$, $C^{Sc}_\ell = G^{Sc}_\ell \cap N_\ell \cap S^{Sc}_\ell$,
$D^S_\ell = G^{Sc}_\ell \cap N^{Sc}_\ell \cap S_\ell$, $D^{Sc}_\ell = G^{Sc}_\ell \cap N^{Sc}_\ell \cap S^{Sc}_\ell$.

$A^{Sc}_\ell \leftarrow M_\ell$(uniform)
$B^{Sc}_\ell \leftarrow R$(randomly uniform)
$F_\ell$(shared uniform) $\leftarrow$ 
$[N] \setminus A^{Sc}_\ell \cap B^{Sc}_\ell \cap F_\ell(P(U_1^i \mid U_1^{1:i-1}, X_1^{1:N}))$

$A^{Sc}_\ell = A_\ell$, $C^{Sc}_\ell = C_\ell$.

$R = \sum_{\ell=1}^{\ell'} \lim_{N \to \infty} \frac{|A_\ell|}{N} \to \frac{1}{2} \log \left( \frac{1 + SNR_b}{1 + SNR_e} \right)$. 

The induced channel.

**Theorem (Semantic security (Liu, Yan, L.’15))**

For arbitrarily distributed message $M$, the information leakage $I(M; Z^N)$ of the proposed wiretap lattice code is upper-bounded as

$$I(M; Z^N) \leq I(\tilde{M}\tilde{F}; Z^N, \tilde{M}\tilde{F} \oplus MF) \leq rN2^{-N\beta'},$$

where $I(\tilde{M}\tilde{F}; Z^N, \tilde{M}\tilde{F} \oplus MF)$ is the capacity of the symmetrized channel derived from the non-binary channel $MF \rightarrow Z^N$. 
Open questions

- Polarization gives information leakage $2^{-\sqrt{N}}$. How to make it vanish faster?
- How to construct wiretap codes for fading/MIMO channels? We don’t even have an explicit code achieving the Shannon capacity of Rayleigh fading channels.
- Is it possible to use polarization to build new primitives for classic cryptography?
- What about quantum cryptography/coding?