# Complexity and Second Moment of the Mathematical Theory of Communication

Hsin-Po WANG

Department of Mathematics, University of Illinois at Urbana–Champaign

2021-04-01 PhD Defense Presentation

# Noisy channel

The sender inputs
$X_1^{32} =$ `11001001 00001111 11011010 10100010`.

The channel outputs
$Y_1^{32} =$ `1--01-01 ----1--- -101---0 --0--0-0`.

# Noisy channel

The sender inputs
$X_1^{32} =$ `11001001 00001111 11011010 10100010`.

The channel outputs
$Y_1^{32} =$ `1--01-01 ----1--- -101---0 --0--0-0`.

# Noisy channel

Sender inputs $X_1^{32} \in \mathbb{F}_q^{32}$, where $\mathbb{F}_q$ is called input alphabet. WLoG, we may assume $\mathbb{F}_q$ is a finite field [new idea].

The channel outputs $Y_1^{32}$ according to transition probabilities $P\{Y_j = y \mid X_j = x\} = W(y|x)$ independently for each $j$.

# Noisy-channel coding

The encoder inputs $X_1^{32} \in \mathscr{B} \subsetneq \mathbb{F}_q^{32}$ into a channel.
$\mathscr{B}$ is a *block code* (sometimes a *codebook*) of block length $N = 32$.

The channel outputs $Y_1^{32}$ according to $W(y|x)$.

The decoder, seeing $Y_1^{32} = y_1^{32}$, maximizes the posterior probability
$\hat{X}_1^{32}(y_1^{32}) := \underset{x_1^{32} \in \mathscr{B}}{\operatorname{argmax}} P\{X_1^{32} = x_1^{32} \mid Y_1^{32} = y_1^{32}\}$.

# Noisy-channel coding

The encoder inputs $X_1^{32} \in \mathcal{B} \subsetneq \mathbb{F}_q^{32}$ into a channel.
$\mathcal{B}$ is a *block code* (sometimes a *codebook*) of block length $N = 32$.

The channel outputs $Y_1^{32}$ according to $W(y|x)$.

The decoder, seeing $Y_1^{32} = y_1^{32}$, maximizes the posterior probability
$\hat{X}_1^{32}(y_1^{32}) := \underset{x_1^{32} \in \mathcal{B}}{\operatorname{argmax}} P\{X_1^{32} = x_1^{32} \mid Y_1^{32} = y_1^{32}\}$.

# Noisy-channel coding

The encoder inputs $X_1^{32} \in \mathcal{B} \subsetneq \mathbb{F}_q^{32}$ into a channel.
$\mathcal{B}$ is a *block code* (sometimes a *codebook*) of block length $N = 32$.

The channel outputs $Y_1^{32}$ according to $W(y|x)$.

The decoder, seeing $Y_1^{32} = y_1^{32}$, maximizes the posterior probability
$\hat{X}_1^{32}(y_1^{32}) := \underset{x_1^{32} \in \mathcal{B}}{\operatorname{argmax}} P\{X_1^{32} = x_1^{32} \mid Y_1^{32} = y_1^{32}\}$.

# Noisy-channel coding

The encoder inputs $X_1^{32} \in \mathcal{B} \subsetneq \mathbb{F}_q^{32}$ into a channel.
$\mathcal{B}$ is a *block code* (sometimes a *codebook*) of block length $N = 32$.

The channel outputs $Y_1^{32}$ according to $W(y|x)$.

The decoder, seeing $Y_1^{32} = y_1^{32}$, maximizes the posterior probability
$\hat{X}_1^{32}(y_1^{32}) := \underset{x_1^{32} \in \mathcal{B}}{\text{do-my-best}} P\{X_1^{32} = x_1^{32} \mid Y_1^{32} = y_1^{32}\}$.

# Noisy-channel coding theorem

Channel capacity $C := \sup_{X \sim Q} I(X\,;Y)$ (supremum over input distribution).

Block length is $N$.

Error probability is $P_\mathrm{e} := P\{\hat{X}_1^N \neq X_1^N\}$.

Code rate is $R := \ln|\mathcal{B}| \div \ln|\mathbb{F}_q^N|$. (recall that $\mathcal{B} \subset \mathbb{F}_q^N$)

[Shannon 1948]

*One can find block codes $\mathcal{B}$ such that $P_\mathrm{e} \to 0$ and $R \to C$ as $N \to \infty$.*

*(And $C$ is the greatest number that allows this to happen.)*

# Noisy-channel coding theorem

Channel capacity $C := \sup_{X \sim Q} I(X\,;Y)$ (supremum over input distribution).

Block length is $N$.
Error probability is $P_\mathrm{e} := P\{\hat{X}_1^N \neq X_1^N\}$.
Code rate is $R := \ln|\mathcal{B}| \div \ln|\mathbb{F}_q^N|$. (recall that $\mathcal{B} \subset \mathbb{F}_q^N$)

[Shannon 1948]
*One can find block codes $\mathcal{B}$ such that $P_\mathrm{e} \to 0$ and $R \to C$ as $N \to \infty$.*
*(And C is the greatest number that allows this to happen.)*

# 2nd-order term of coding

How fast do error probability $P_\mathrm{e}$ and code rate $R$ converge to $0$ and $C$ as block length $N \to \infty$? Characterize functions "$P_\mathrm{e}(N)$" and "$R(N)$".

When $R$ is fixed, $P_\mathrm{e} \approx e^{-N}$; or equivalently, $-\ln P_\mathrm{e} \approx N$.
Fano [Fan61], Gallager [Gal65], Shannon–Gallager–Berlekamp [SGB67], [Gal68], [Gal73], Blahut [Bla74], Barg–Forney [BF02], Fàbregas–Land–Martinez [iFLM11], Domb–Zamir–Feder [DZF16].

When $P_\mathrm{e}$ is fixed, $R \approx C - N^{-1/2}$; or equivalently, $(C - R)^{-2} \approx N$.
Wolfowitz [Wol57], Weiss [WEI60], Dobrushin [Dob61], Strassen [Str62], Baron–Khojastepour–Baraniuk [BKB04], Hayashi [Hay09], Polyanskiy–Poor–Verdu [PPV10].

# 2nd-order term of coding

How fast do error probability $P_\mathrm{e}$ and code rate $R$ converge to $0$ and $C$ as block length $N \to \infty$? Characterize functions "$P_\mathrm{e}(N)$" and "$R(N)$".

When $R$ is fixed, $P_\mathrm{e} \approx e^{-N}$; or equivalently, $-\ln P_\mathrm{e} \approx N$.
Fano [Fan61], Gallager [Gal65], Shannon–Gallager–Berlekamp [SGB67], [Gal68], [Gal73], Blahut [Bla74], Barg–Forney [BF02], Fàbregas–Land–Martinez [iFLM11], Domb–Zamir–Feder [DZF16].

When $P_\mathrm{e}$ is fixed, $R \approx C - N^{-1/2}$; or equivalently, $(C-R)^{-2} \approx N$.
Wolfowitz [Wol57], Weiss [WEI60], Dobrushin [Dob61], Strassen [Str62], Baron–Khojastepour–Baraniuk [BKB04], Hayashi [Hay09], Polyanskiy–Poor–Verdu [PPV10].

# 2nd-order term of coding

How fast do error probability $P_\mathrm{e}$ and code rate $R$ converge to $0$ and $C$ as block length $N \to \infty$? Characterize functions "$P_\mathrm{e}(N)$" and "$R(N)$".

When $R$ is fixed, $P_\mathrm{e} \approx e^{-N}$; or equivalently, $-\ln P_\mathrm{e} \approx N$.
Fano [Fan61], Gallager [Gal65], Shannon–Gallager–Berlekamp [SGB67], [Gal68], [Gal73], Blahut [Bla74], Barg–Forney [BF02], Fàbregas–Land–Martinez [iFLM11], Domb–Zamir–Feder [DZF16].

When $P_\mathrm{e}$ is fixed, $R \approx C - N^{-1/2}$; or equivalently, $(C - R)^{-2} \approx N$.
Wolfowitz [Wol57], Weiss [WEI60], Dobrushin [Dob61], Strassen [Str62], Baron–Khojastepour–Baraniuk [BKB04], Hayashi [Hay09], Polyanskiy–Poor–Verdu [PPV10].

# Joint 2nd-order term of coding

When both $R$ and $P_{\mathrm{e}}$ vary, $(P_{\mathrm{e}}, R) \approx (e^{-N^{\pi}}, C - N^{-\rho})$ where $\pi + 2\rho = 1$; or equivalently, $(-\ln P_{\mathrm{e}})(C - R)^{-2} \approx N$.
Altuğ–Wagner [AW10], Polyanskiy–Verdú [PV10], [AW14], Hayashi–Tan [HT15].

This is a two-sided bound:
A code $\mathcal{B}$ exists such that $(-\ln P_{\mathrm{e}})(C - R)^{-2} \approx N$.
No code $\mathcal{B}$ exists such that $(-\ln P_{\mathrm{e}})(C - R)^{-2} \gg N$.

Block length $N$ is your income;
invest in error probability $P_{\mathrm{e}}$ or in code rate $R$ or in both.

# Joint 2nd-order term of coding

When both $R$ and $P_{\mathrm{e}}$ vary, $(P_{\mathrm{e}}, R) \approx (e^{-N^{\pi}}, C - N^{-\rho})$ where $\pi + 2\rho = 1$; or equivalently, $(-\ln P_{\mathrm{e}})(C - R)^{-2} \approx N$.
Altuğ–Wagner [AW10], Polyanskiy–Verdú [PV10], [AW14], Hayashi–Tan [HT15].

This is a two-sided bound:
A code $\mathscr{B}$ exists such that $(-\ln P_{\mathrm{e}})(C - R)^{-2} \approx N$.
No code $\mathscr{B}$ exists such that $(-\ln P_{\mathrm{e}})(C - R)^{-2} \gg N$.

Block length $N$ is your income;
invest in error probability $P_{\mathrm{e}}$ or in code rate $R$ or in both.

# Joint 2nd-order term of coding

When both $R$ and $P_\mathrm{e}$ vary, $(P_\mathrm{e}, R) \approx (e^{-N^\pi}, C - N^{-\rho})$ where $\pi + 2\rho = 1$; or equivalently, $(-\ln P_\mathrm{e})(C - R)^{-2} \approx N$.
Altuğ–Wagner [AW10], Polyanskiy–Verdú [PV10], [AW14], Hayashi–Tan [HT15].

This is a two-sided bound:
A code $\mathscr{B}$ exists such that $(-\ln P_\mathrm{e})(C - R)^{-2} \approx N$.
No code $\mathscr{B}$ exists such that $(-\ln P_\mathrm{e})(C - R)^{-2} \gg N$.

Block length $N$ is your income;
invest in error probability $P_\mathrm{e}$ or in code rate $R$ or in both.

# 2nd-order term analog

| Paradigm | Random variable | Random code |
| --- | --- | --- |
| law of large numbers | $\bar{X} \to \mu$ | $(P_e, R) \to (0, C)$ |
| large deviation principle | $\mathbb{P}\{\bar{X} - \mu > x\} \approx e^{-nI(x)}$ | $P_e \approx e^{-N}$ |
| central limit theorem | $\bar{X} - \mu \sim \mathcal{N}(0, \sigma/\sqrt{n})$ | $C - R \approx N^{-1/2}$ |
| moderate deviation principle | $\frac{-\ln \mathbb{P}\{\bar{X}-\mu>\gamma_n x\}}{\gamma_n^2} \approx nI(x)$ | $\frac{-\ln P_e}{(C-R)^2} \approx N$ |

# 2nd-order term analog

| Paradigm | Random variable | Random code |
|---|---|---|
| law of large numbers | $\bar{X} \to \mu$ | $(P_e, R) \to (0, C)$ |
| large deviation principle | $\mathbb{P}\{\bar{X} - \mu > x\} \approx e^{-nI(x)}$ | $P_e \approx e^{-N}$ |
| central limit theorem | $\bar{X} - \mu \sim \mathcal{N}(0, \sigma/\sqrt{n})$ | $C - R \approx N^{-1/2}$ |
| moderate deviation principle | $\frac{-\ln \mathbb{P}\{\bar{X} - \mu > \gamma_n x\}}{\gamma_n^2} \approx nI(x)$ | $\frac{-\ln P_e}{(C-R)^2} \approx N$ |

# 2nd-order term analog

| Paradigm | Random variable | Random code |
|---|---|---|
| law of large numbers | $\bar{X} \to \mu$ | $(P_\mathrm{e}, R) \to (0, C)$ |
| large deviation principle | $\mathbb{P}\{\bar{X} - \mu > x\} \approx e^{-nI(x)}$ | $P_\mathrm{e} \approx e^{-N}$ |
| central limit theorem | $\bar{X} - \mu \sim \mathcal{N}(0, \sigma/\sqrt{n})$ | $C - R \approx N^{-1/2}$ |
| moderate deviation principle | $\frac{-\ln \mathbb{P}\{\bar{X}-\mu>\gamma_n x\}}{\gamma_n^2} \approx nI(x)$ | $\frac{-\ln P_\mathrm{e}}{(C-R)^2} \approx N$ |

# However...

Achievability via random coding assumes exponential complexity due to the usage of the maximum a posteriori decoder $\operatorname*{argmax}_{x_1^{32} \in \mathcal{B}}$.

Goal: Comparable performance, but with a low-complexity $\operatorname*{do-my-best}_{x_1^{32} \in \mathcal{B}}$.

# However...

Achievability via random coding assumes exponential complexity due to the usage of the maximum a posteriori decoder $\underset{x_1^{32}\in\mathcal{B}}{\mathrm{argmax}}$.

Goal: Comparable performance, but with a low-complexity $\underset{x_1^{32}\in\mathcal{B}}{\mathrm{do\text{-}my\text{-}best}}$.

# 2nd-order term goal

| Paradigm | Random code | Low-complexity code |
|---|---|---|
| law of large numbers | $(P_{\mathrm{e}}, R) \to (0, C)$ | $(P_{\mathrm{e}}, R) \to (0, C)$ |
| large deviation principle | $P_{\mathrm{e}} \approx e^{-N}$ | $P_{\mathrm{e}} \approx e^{-N^{0.99}}$ |
| central limit theorem | $C - R \approx N^{-1/2}$ | $C - R \approx N^{-0.49}$ |
| moderate deviation principle | $\frac{-\ln P_{\mathrm{e}}}{(C-R)^2} \approx N$ | $\frac{-\ln P_{\mathrm{e}}}{(C-R)^2} \approx N^{0.99}$ |

$(\pi, \rho > 0$ and $\pi + 2\rho < 1)$

# 2nd-order term goal

| Paradigm | Random code | Low-complexity code |
|---|---|---|
| law of large numbers | $(P_\mathrm{e}, R) \to (0, C)$ | $(P_\mathrm{e}, R) \to (0, C)$ |
| large deviation principle | $P_\mathrm{e} \approx e^{-N}$ | $P_\mathrm{e} \approx e^{-N^{0.99}}$ |
| central limit theorem | $C - R \approx N^{-1/2}$ | $C - R \approx N^{-0.49}$ |
| moderate deviation principle | $\frac{-\ln P_\mathrm{e}}{(C-R)^2} \approx N$ | $\frac{-\ln P_\mathrm{e}}{(C-R)^2} \approx N^{0.99}$ |

$(\pi, \rho > 0$ and $\pi + 2\rho < 1)$

# Polar coding

| Par | BEC | SBDMC | p-ary | q-ary | finite | BDMC | a-finite |
|---|---|---|---|---|---|---|---|
| LLN | [Ari09] | [Ari09] | [ŞTA09] | [ŞTA09] | [ŞTA09] | [SRDR12] | ?? |
| LDP* | [AT09] | [AT09] | [ŞTA09] | [MT10] | [Sas11] | [HY13] | ?? |
| CLT* | [KMTU10] | [HAU14] | [BGN+18] | ?? | ?? | ?? | ?? |
| MDP* | [GX15] | [GX15] | [BGS18] | ?? | ?? | ?? | ?? |
| LDP | [KSU10] | [KSU10] | ?? | ?? | ?? | ?? | ?? |
| CLT | [FHMV18] | [GRY20] | ?? | ?? | ?? | ?? | ?? |
| MDP | ?? | ?? | ?? | ?? | ?? | ?? | ?? |

# Polar coding

| Par | BEC | SBDMC | p-ary | q-ary | finite | BDMC | a-finite |
|---|---|---|---|---|---|---|---|
| LLN | [Ari09] | [Ari09] | [ŞTA09] | [ŞTA09] | [ŞTA09] | [SRDR12] | ?? |
| LDP$^{\star}$ | [AT09] | [AT09] | [ŞTA09] | [MT10] | [Sas11] | [HY13] | ?? |
| CLT$^{\star}$ | [KMTU10] | [HAU14] | [BGN$^{+}$18] | ?? | ?? | ?? | ?? |
| MDP$^{\star}$ | [GX15] | [GX15] | [BGS18] | ?? | ?? | ?? | ?? |
| LDP | [KSU10] | [KSU10] | ?? | ?? | ?? | ?? | ?? |
| CLT | [FHMV18] | [GRY20] | ?? | ?? | ?? | ?? | ?? |
| MDP | ?? | ?? | ?? | ?? | ?? | ?? | ?? |

# Polar coding

| Par | BEC | SBDMC | p-ary | q-ary | finite | BDMC | a-finite |
|---|---|---|---|---|---|---|---|
| LLN | [Ari09] | [Ari09] | [ŞTA09] | [ŞTA09] | [ŞTA09] | [SRDR12] | ?? |
| LDP$^\star$ | [AT09] | [AT09] | [ŞTA09] | [MT10] | [Sas11] | [HY13] | ?? |
| CLT$^\star$ | [KMTU10] | [HAU14] | [BGN$^+$18] | ?? | ?? | ?? | ?? |
| MDP$^\star$ | [GX15] | [GX15] | [BGS18] | ?? | ?? | ?? | ?? |
| LDP | [KSU10] | [KSU10] | ?? | ?? | ?? | ?? | ?? |
| CLT | [FHMV18] | [GRY20] | ?? | ?? | ?? | ?? | ?? |
| MDP | ?? | ?? | ?? | ?? | ?? | ?? | ?? |

# Polar coding

| Par | BEC | SBDMC | p-ary | q-ary | finite | BDMC | a-finite |
|---|---|---|---|---|---|---|---|
| LLN | [Ari09] | [Ari09] | [ŞTA09] | [ŞTA09] | [ŞTA09] | [SRDR12] | ?? |
| LDP$^\star$ | [AT09] | [AT09] | [ŞTA09] | [MT10] | [Sas11] | [HY13] | ?? |
| CLT$^\star$ | [KMTU10] | [HAU14] | [BGN$^+$18] | ?? | ?? | ?? | ?? |
| MDP$^\star$ | [GX15] | [GX15] | [BGS18] | ?? | ?? | ?? | ?? |
| LDP | [KSU10] | [KSU10] | ?? | ?? | ?? | ?? | ?? |
| CLT | [FHMV18] | [GRY20] | ?? | ?? | ?? | ?? | ?? |
| MDP | ?? | ?? | ?? | ?? | ?? | ?? | ?? |

# Polar coding

| Par | BEC | SBDMC | p-ary | q-ary | finite | BDMC | a-finite |
|---|---|---|---|---|---|---|---|
| LLN | [Ari09] | [Ari09] | [ŞTA09] | [ŞTA09] | [ŞTA09] | [SRDR12] | ?? |
| LDP$^\star$ | [AT09] | [AT09] | [ŞTA09] | [MT10] | [Sas11] | [HY13] | ?? |
| CLT$^\star$ | [KMTU10] | [HAU14] | [BGN$^+$18] | ?? | ?? | ?? | ?? |
| MDP$^\star$ | [GX15] | [GX15] | [BGS18] | ?? | ?? | ?? | ?? |
| LDP | [KSU10] | [KSU10] | ?? | ?? | ?? | ?? | ?? |
| CLT | [FHMV18] | [GRY20] | ?? | ?? | ?? | ?? | ?? |
| MDP | ?? | ?? | ?? | ?? | ?? | ?? | ?? |

# Polar coding road map

Channel transformation manipulates channels.

Channel tree is the result of recursive channel transformation.

Channel parameter measures the channels and keep track of the tree.

Channel process is a syntax candy paraphrasing the tree.

Channel polarization is a phenomenon that channels become extreme.

# Channel transformation

Consider a channel $W = (X \mid Y)$, where input is $X$, output is $Y$.
Make two i.i.d. copies $(X_1 \mid Y_1)$ and $(X_2 \mid Y_2)$.

$W^{(1)} := (X_1 - X_2 \mid Y_1^2)$;
$W^{(2)} := (X_2 \mid (X_1 - X_2) Y_1^2)$. (juxtaposition is tuple concatenation)

# Channel transformation



Consider a channel $W = (X \mid Y)$, where input is $X$, output is $Y$.
Make two i.i.d. copies $(X_1 \mid Y_1)$ and $(X_2 \mid Y_2)$.

$W^{(1)} := (X_1 - X_2 \mid Y_1^2)$;
$W^{(2)} := (X_2 \mid (X_1 - X_2)Y_1^2)$. (juxtaposition is tuple concatenation)

# Channel transformation (other kernel)

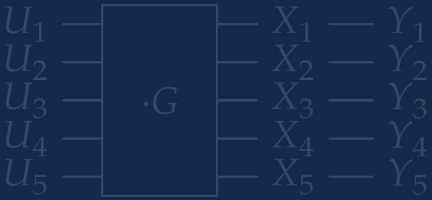
$U_1^2$: two free variables; $G$: a $2 \times 2$ matrix (called kernel);
$X_1^2 := U_1^2 \cdot G$: matrix multiplication; channels generate $Y_1^2$ given $X_1^2$.

$W^{(1)} := (U_1 \mid Y_1^2)$;
$W^{(2)} := (U_2 \mid U_1 Y_1^2)$ (juxtaposition is tuple concatenation).

# Channel transformation (larger kernel)

$U_1^\ell$: many free variables; $G$: an $\ell \times \ell$ matrix kernel;
$X_1^\ell := U_1^\ell \cdot G$; channels generate $Y_1^\ell$ given $X_1^\ell$.

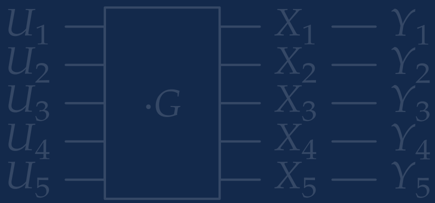
$$
\begin{aligned}
W^{(1)} &:= (U_1 \mid Y_1^\ell);\\
W^{(2)} &:= (U_2 \mid U_1 Y_1^\ell);\\
W^{(3)} &:= (U_3 \mid U_1^2 Y_1^\ell);\\
&\ \vdots\\
W^{(\ell-1)} &:= (U_\ell \mid U_1^{\ell-2} Y_1^\ell);\\
W^{(\ell)} &:= (U_\ell \mid U_1^{\ell-1} Y_1^\ell).
\end{aligned}
$$



# Channel transformation (larger kernel)

$U_1^\ell$: many free variables; $G$: an $\ell \times \ell$ matrix kernel;
$X_1^\ell := U_1^\ell \cdot G$; channels generate $Y_1^\ell$ given $X_1^\ell$.

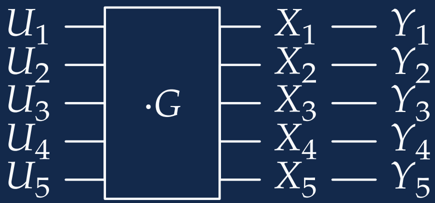
$$
\begin{aligned}
W^{(1)} &:= (U_1 \mid Y_1^\ell);\\
W^{(2)} &:= (U_2 \mid U_1 Y_1^\ell);\\
W^{(3)} &:= (U_3 \mid U_1^2 Y_1^\ell);\\
&\vdots\\
W^{(\ell-1)} &:= (U_\ell \mid U_1^{\ell-2} Y_1^\ell);\\
W^{(\ell)} &:= (U_\ell \mid U_1^{\ell-1} Y_1^\ell).
\end{aligned}
$$



# Channel transformation (larger kernel)

$U_1^\ell$: many free variables; $G$: an $\ell \times \ell$ matrix kernel;
$X_1^\ell := U_1^\ell \cdot G$; channels generate $Y_1^\ell$ given $X_1^\ell$.

$$
\begin{aligned}
W^{(1)} &:= (U_1 \mid Y_1^\ell);\\
W^{(2)} &:= (U_2 \mid U_1 Y_1^\ell);\\
W^{(3)} &:= (U_3 \mid U_1^2 Y_1^\ell);\\
&\vdots\\
W^{(\ell-1)} &:= (U_\ell \mid U_1^{\ell-2} Y_1^\ell);\\
W^{(\ell)} &:= (U_\ell \mid U_1^{\ell-1} Y_1^\ell).
\end{aligned}
$$

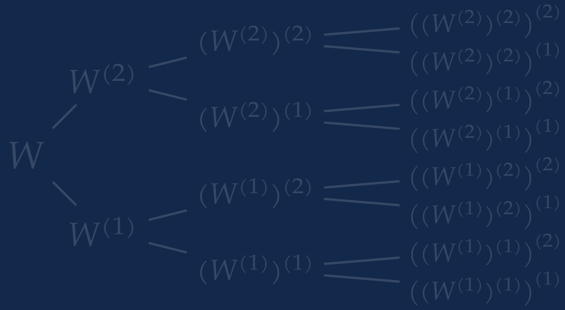


# Channel tree

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$.

For each $j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$.

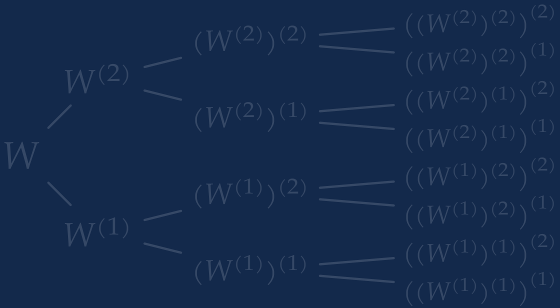


# Channel tree

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$.

For each $j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$.

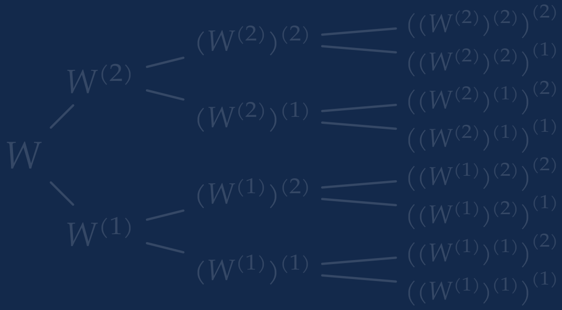


# Channel tree

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$.

For each $j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$.

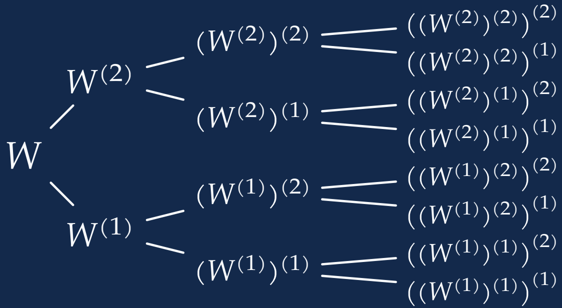


# Channel tree

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$.

For each $j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$.



# Dynamic kernel [new idea*]

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$ using $G$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$ using $G^{(i)}$.

$\forall j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$ using $G^{(ij)}$.

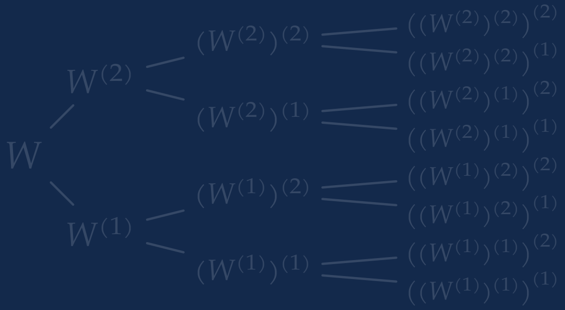
- $W$
  - $W^{(2)}$
    - $(W^{(2)})^{(2)}$
      - $((W^{(2)})^{(2)})^{(2)}$
      - $((W^{(2)})^{(2)})^{(1)}$
    - $(W^{(2)})^{(1)}$
      - $((W^{(2)})^{(1)})^{(2)}$
      - $((W^{(2)})^{(1)})^{(1)}$
  - $W^{(1)}$
    - $(W^{(1)})^{(2)}$
      - $((W^{(1)})^{(2)})^{(2)}$
      - $((W^{(1)})^{(2)})^{(1)}$
    - $(W^{(1)})^{(1)}$
      - $((W^{(1)})^{(1)})^{(2)}$
      - $((W^{(1)})^{(1)})^{(1)}$

# Dynamic kernel [new idea*]

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$ using $G$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$ using $G^{(i)}$.

$\forall j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$ using $G^{(ij)}$.

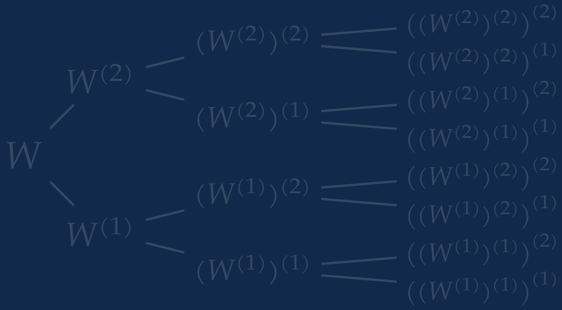


# Dynamic kernel [new idea*]

Channel $W$ grows $W^{(1)}, W^{(2)}, \dots, W^{(\ell)}$ using $G$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \dots, (W^{(i)})^{(\ell)}$ using $G^{(i)}$.

$\forall j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \dots, ((W^{(i)})^{(j)})^{(\ell)}$ using $G^{(ij)}$.

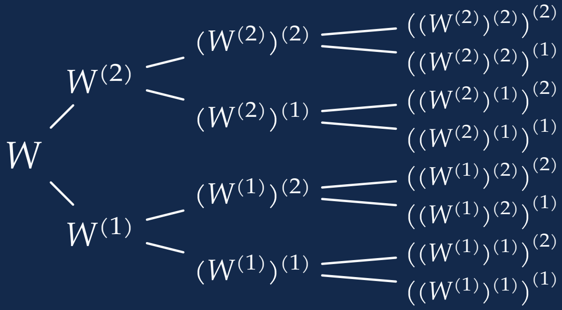


# Dynamic kernel [new idea*]

Channel $W$ grows $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$ using $G$.

For each $i$, channel $W^{(i)}$ grows $(W^{(i)})^{(1)}, \ldots, (W^{(i)})^{(\ell)}$ using $G^{(i)}$.

$\forall j$, channel $(W^{(i)})^{(j)}$ grows $((W^{(i)})^{(j)})^{(1)}, \ldots, ((W^{(i)})^{(j)})^{(\ell)}$ using $G^{(ij)}$.



# Channel parameter ($\ell = 2$ and $n = 3$ exmaple)

Block length $N = \ell^n$; for instance $8 = 2^3$.

Select indices $\mathcal{J} \subseteq \{1, \ldots, \ell\}^n$; for instance $\{122, 212, 221, 222\} \subseteq \{1, 2\}^3$.
Code rate $R = |\mathcal{J}|/N = 4/8$ (nontrivial, due to implementation details).

Error probability $P_e \leq \sum_{ijk \in \mathcal{J}} H\Big(\big((W^{(i)})^{(j)}\big)^{(k)}\Big)$ (nontrivial, due to details);

$H(X \mid Y)$ is conditional entropy with base-$q$ logarithm.

# Channel parameter ($\ell = 2$ and $n = 3$ exmaple)

Block length $N = \ell^n$; for instance $8 = 2^3$.

Select indices $\mathcal{J} \subseteq \{1, \ldots, \ell\}^n$; for instance $\{122, 212, 221, 222\} \subseteq \{1,2\}^3$.
Code rate $R = |\mathcal{J}|/N = 4/8$ (nontrivial, due to implementation details).

Error probability $P_e \le \sum_{ijk \in \mathcal{J}} H\Big(\big((W^{(i)})^{(j)}\big)^{(k)}\Big)$ (nontrivial, due to details);
$H(X \mid Y)$ is conditional entropy with base-$q$ logarithm.

# Channel parameter ($\ell = 2$ and $n = 3$ exmaple)

Block length $N = \ell^n$; for instance $8 = 2^3$.

Select indices $\mathcal{J} \subseteq \{1, \ldots, \ell\}^n$; for instance $\{122, 212, 221, 222\} \subseteq \{1, 2\}^3$.
Code rate $R = |\mathcal{J}|/N = 4/8$ (nontrivial, due to implementation details).

Error probability $P_\mathrm{e} \le \sum_{ijk \in \mathcal{J}} H\Big(\big((W^{(i)})^{(j)}\big)^{(k)}\Big)$ (nontrivial, due to details);
$H(X \mid Y)$ is conditional entropy with base-$q$ logarithm.

# It suffices to understand

$$H(W), H(W^{(i)}), H((W^{(i)})^{(j)}), H\Big(((W^{(i)})^{(j)})^{(k)}\Big), H\Big(\Big(((W^{(i)})^{(j)})^{(k)}\Big)^{(l)}\Big), \ldots$$

Block length $N$ will be $\ell^{\text{where we stop}}$.

Code rate $R$ will be the fraction of small $H$-values.

Block error probability $P_{\mathrm{e}}$ will be $\sum_{\text{those}}$ small $H$-values.

# It suffices to understand

$$H(W), H(W^{(i)}), H((W^{(i)})^{(j)}), H\Big(((W^{(i)})^{(j)})^{(k)}\Big), H\Big(\Big(((W^{(i)})^{(j)})^{(k)}\Big)^{(l)}\Big), \ldots$$

Block length $N$ will be $\ell^{\text{where we stop}}$.

Code rate $R$ will be the fraction of small $H$-values.

Block error probability $P_{\mathrm{e}}$ will be $\sum_{\text{those}}$ small $H$-values.

# It suffices to understand

$$H(W), H(W^{(i)}), H((W^{(i)})^{(j)}), H\Big(((W^{(i)})^{(j)})^{(k)}\Big), H\Big(\Big(((W^{(i)})^{(j)})^{(k)}\Big)^{(l)}\Big), \ldots$$

Block length $N$ will be $\ell^{\text{where we stop}}$.

Code rate $R$ will be the fraction of small $H$-values.

Block error probability $P_\mathrm{e}$ will be $\sum_{\text{those}}$ small $H$-values.

# It suffices to understand

$$H(W), H(W^{(i)}), H((W^{(i)})^{(j)}), H\Big(((W^{(i)})^{(j)})^{(k)}\Big), H\Big(\Big(((W^{(i)})^{(j)})^{(k)}\Big)^{(l)}\Big), \ldots$$

Block length $N$ will be $\ell^{\text{where we stop}}$.

Code rate $R$ will be the fraction of small $H$-values.

Block error probability $P_\mathrm{e}$ will be $\sum_{\text{those}}$ small $H$-values.

# Channel process (a powerful syntax candy)

$W_0 := W$.
$W_{n+1} := W_n^{(J_{n+1})}$, where $J_{n+1} \in \{1, 2, \ldots, \ell\}$ i.i.d. uniform branch chooser.

$H_n := H(W_n)$.

Decide depth $n$, then block length is $N = \ell^n$.
Decide threshold $\theta$, then code rate is $R = P\{H_n < \theta\}$.
Error probability is $P_e < \sum$ small $H_n < \sum \theta = RN\theta \leq N\theta$.

# Channel process (a powerful syntax candy)

$W_0 := W$.
$W_{n+1} := W_n^{(J_{n+1})}$, where $J_{n+1} \in \{1, 2, \ldots, \ell\}$ i.i.d. uniform branch chooser.

$H_n := H(W_n)$.

Decide depth $n$, then block length is $N = \ell^n$.
Decide threshold $\theta$, then code rate is $R = P\{H_n < \theta\}$.
Error probability is $P_e < \sum$ small $H_n < \sum \theta = RN\theta \leq N\theta$.

# Channel process (a powerful syntax candy)

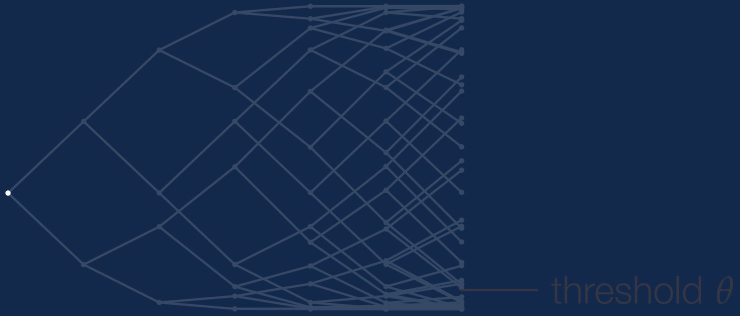
$W_0 := W$.
$W_{n+1} := W_n^{(J_{n+1})}$, where $J_{n+1} \in \{1, 2, \ldots, \ell\}$ i.i.d. uniform branch chooser.

$H_n := H(W_n)$.

Decide depth $n$, then block length is $N = \ell^n$.
Decide threshold $\theta$, then code rate is $R = P\{H_n < \theta\}$.
Error probability is $P_e < \sum$ small $H_n < \sum \theta = RN\theta \leq N\theta$.

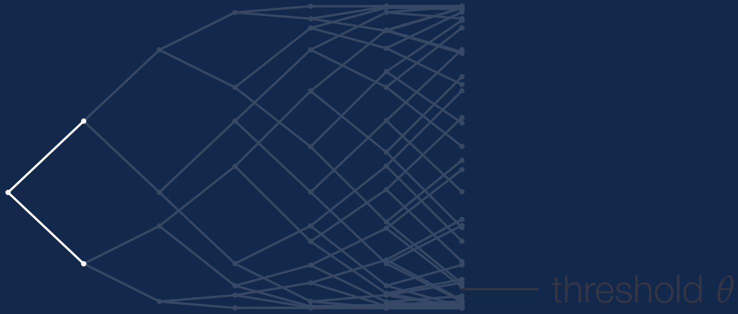
# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.



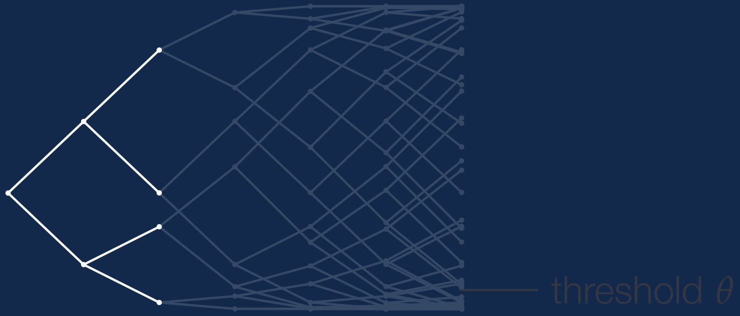
# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.



# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.



# Channel polarization

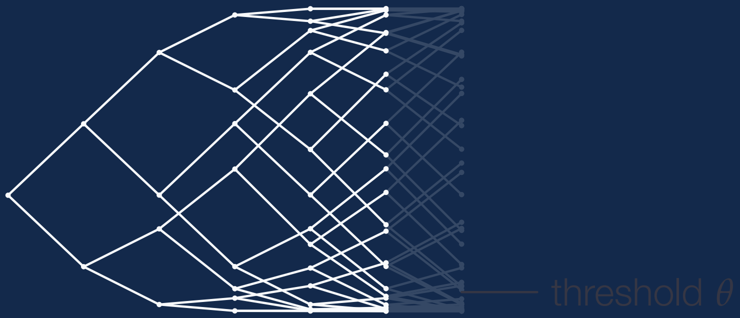
$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.

# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.

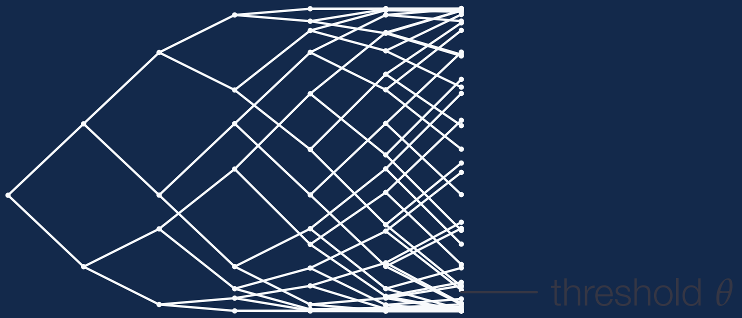
# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.



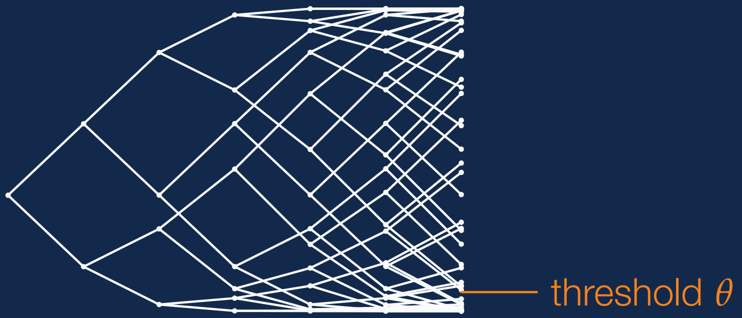
# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.



# Channel polarization

$H_n := H(W_n)$ is a martingale. (Invoke Doob's martingale convergence)
$H_n \to H_\infty$ a.e. as $n \to \infty$; turns out $H_\infty \in \{0, 1\}$ and $P\{H_\infty = 1\} = H_0$.



# It suffices to understand

$$P\{H_n < \text{threshold}\} > C - \text{gap}.$$

Goal: $P\{H_n < e^{-\ell^{\pi n}}\} > C - \ell^{-\rho n}$ for large $n$, where $\pi + 2\rho < 1$.
Then: $N = \ell^n$ and $P_e < Ne^{-N^\pi} \approx e^{-N^\pi}$ and $R > C - N^{-\rho}$.

# It suffices to understand

$$P\{H_n < \text{threshold}\} > C - \text{gap}.$$

Goal: $P\{H_n < e^{-\ell^{\pi n}}\} > C - \ell^{-\rho n}$ for large $n$, where $\pi + 2\rho < 1$.
Then: $N = \ell^n$ and $P_\mathrm{e} < Ne^{-N^\pi} \approx e^{-N^\pi}$ and $R > C - N^{-\rho}$ .

# Proof outline

Local LDP behavior: $Z(W^{(j)}) \leq \ell e^{qZ(W)\ell}(qZ(W))^{\lceil j^2/3\ell \rceil}$,
where $Z$ is a parameter such that $Z \leq q^3\sqrt{H}$ and $H \leq q^3\sqrt{Z}$.

Local CLT behavior: $\sum_{j=1}^{\ell} h(H(W^{(j)})) < 4\ell^{1/2+\alpha}$,
where $\alpha = \ln(\ln \ell)/\ln \ell$ and $h(z) := \min(z, 1-z)^{\alpha}$.

Global MDP behavior: $P\{H_n < e^{-\ell^{\pi n}}\} > C - \ell^{-\rho n}$, where $\pi + 2\rho < 1$,
given local LDP and local CLT behaviors.

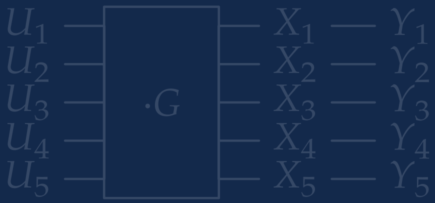
# Local LDP behavior 1/3

Want to prove $Z(W^{(j)}) \le \ell e^{qz\ell}(qz)^{\lceil j^2/3\ell \rceil}$ where $z := Z(W)$.

Fundamental theorem of polar: $Z(W^{(j)}) \le \sum_{u_{j+1}^{\ell} \in \mathbb{F}_q^{\ell-j}} z^{\operatorname{hwt}(0_1^{j-1} 1_j u_{j+1}^{\ell} \cdot G)}$.

RHS is the weight enumerator of a coset code.

$$
\begin{aligned}
W^{(1)} &:= (U_1 \mid Y_1^{\ell});\\
W^{(2)} &:= (U_2 \mid U_1 Y_1^{\ell});\\
&\vdots\\
W^{(\ell)} &:= (U_\ell \mid U_1^{\ell-1} Y_1^{\ell}).
\end{aligned}
$$



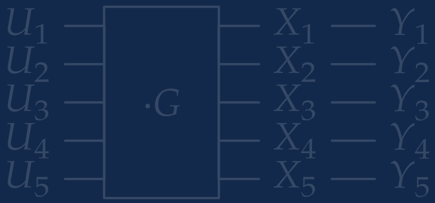
# Local LDP behavior 1/3

Want to prove $Z(W^{(j)}) \le \ell e^{qz\ell}(qz)^{\lceil j^2/3\ell \rceil}$ where $z := Z(W)$.

Fundamental theorem of polar: $Z(W^{(j)}) \le \sum_{u_{j+1}^{\ell} \in \mathbb{F}_q^{\ell-j}} z^{\operatorname{hwt}(0_1^{j-1} 1_j u_{j+1}^{\ell} \cdot G)}$.

RHS is the weight enumerator of a coset code.

$$
\begin{aligned}
W^{(1)} &:= (U_1 \mid Y_1^{\ell}); \\
W^{(2)} &:= (U_2 \mid U_1 Y_1^{\ell}); \\
&\vdots \\
W^{(\ell)} &:= (U_\ell \mid U_1^{\ell-1} Y_1^{\ell}).
\end{aligned}
$$



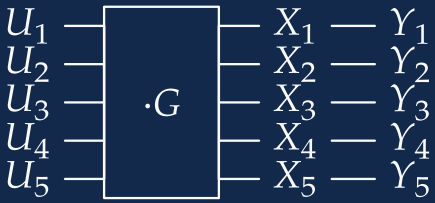
# Local LDP behavior 1/3

Want to prove $Z(W^{(j)}) \le \ell e^{qz\ell}(qz)^{\lceil j^2/3\ell \rceil}$ where $z := Z(W)$.

Fundamental theorem of polar: $Z(W^{(j)}) \le \displaystyle\sum_{u_{j+1}^{\ell} \in \mathbb{F}_q^{\ell-j}} z^{\mathrm{hwt}(0_1^{j-1} 1_j u_{j+1}^{\ell} \cdot G)}$.

RHS is the weight enumerator of a coset code.

$$
\begin{aligned}
W^{(1)} &:= (U_1 \mid Y_1^{\ell}); \\
W^{(2)} &:= (U_2 \mid U_1 Y_1^{\ell}); \\
&\vdots \\
W^{(\ell)} &:= (U_{\ell} \mid U_1^{\ell-1} Y_1^{\ell}).
\end{aligned}
$$



# Local LDP behavior 2/3

Want $\sum_{u^\ell_{j+1}} z^{\mathrm{hwt}(0_1^{j-1} 1_j u^\ell_{j+1} \cdot G)} \leq \ell e^{qz\ell} (qz)^{\lceil j^2/3\ell \rceil}$ for some $G$.

Draw random $\mathbb{G}$ instead; $\mathbb{E}[\mathrm{LHS}] = q^{-j}(1+(q-1)z)^\ell \leq q^{-j}(1+qz)^\ell$.

Compare $(qz)^w$-coefficients: $q^{-j}\binom{\ell}{w}$ vs $\ell \frac{\ell^{w-\lceil j^2/3\ell \rceil}}{(w-\lceil j^2/3\ell \rceil)!}$.

Simplify: $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}\binom{\ell-\lceil j^2/3\ell \rceil}{w-\lceil j^2/3\ell \rceil}$ vs $\ell\binom{\ell}{w-\lceil j^2/3\ell \rceil}$.

# Local LDP behavior 2/3

Want $\sum_{u^\ell_{j+1}} z^{\mathrm{hwt}(0_1^{j-1} 1_j u^\ell_{j+1} \cdot G)} \le \ell e^{qz\ell} (qz)^{\lceil j^2/3\ell \rceil}$ for some $G$.

Draw random $\mathbb{G}$ instead; $\mathbb{E}[\text{LHS}] = q^{-j}(1 + (q-1)z)^\ell \le q^{-j}(1 + qz)^\ell$.

Compare $(qz)^w$-coefficients: $q^{-j}\binom{\ell}{w}$ vs $\ell \frac{\ell^{w-\lceil j^2/3\ell \rceil}}{(w-\lceil j^2/3\ell \rceil)!}$.

Simplify: $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}\binom{\ell-\lceil j^2/3\ell \rceil}{w-\lceil j^2/3\ell \rceil}$ vs $\ell\binom{\ell}{w-\lceil j^2/3\ell \rceil}$.

# Local LDP behavior 2/3

Want $\sum_{u_{j+1}^{\ell}} z^{\mathrm{hwt}(0_1^{j-1} 1_j u_{j+1}^{\ell} \cdot G)} \leq \ell e^{qz\ell} (qz)^{\lceil j^2/3\ell \rceil}$ for some $G$.

Draw random $\mathbb{G}$ instead; $\mathbb{E}[\text{LHS}] = q^{-j}(1 + (q-1)z)^{\ell} \leq q^{-j}(1+qz)^{\ell}$.

Compare $(qz)^w$-coefficients: $q^{-j}\binom{\ell}{w}$ vs $\ell \frac{\ell^{w-\lceil j^2/3\ell \rceil}}{(w-\lceil j^2/3\ell \rceil)!}$.

Simplify: $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}\binom{\ell-\lceil j^2/3\ell \rceil}{w-\lceil j^2/3\ell \rceil}$ vs $\ell\binom{\ell}{w-\lceil j^2/3\ell \rceil}$.

# Local LDP behavior 2/3

Want $\sum_{u_{j+1}^{\ell}} z^{\mathrm{hwt}(0_1^{j-1} 1_j u_{j+1}^{\ell} \cdot G)} \leq \ell e^{qz\ell} (qz)^{\lceil j^2/3\ell \rceil}$ for some $G$.

Draw random $\mathbb{G}$ instead; $\mathbb{E}[\mathrm{LHS}] = q^{-j}(1 + (q-1)z)^{\ell} \leq q^{-j}(1 + qz)^{\ell}$.

Compare $(qz)^w$-coefficients: $q^{-j}\binom{\ell}{w}$ vs $\ell \frac{\ell^{w - \lceil j^2/3\ell \rceil}}{(w - \lceil j^2/3\ell \rceil)!}$.

Simplify: $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}\binom{\ell - \lceil j^2/3\ell \rceil}{w - \lceil j^2/3\ell \rceil}$ vs $\ell\binom{\ell}{w - \lceil j^2/3\ell \rceil}$.

# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)

Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.

# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.

# Local LDP behavior 3/3

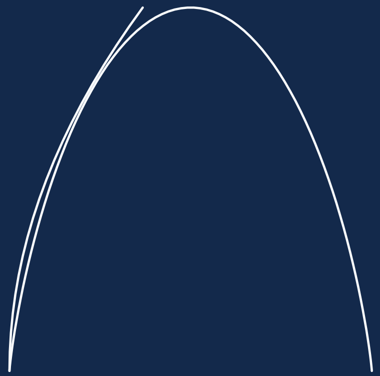
Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.

# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

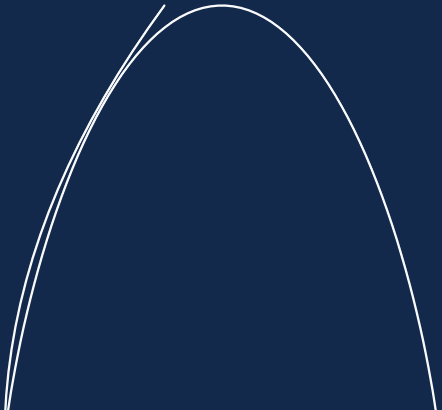
$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

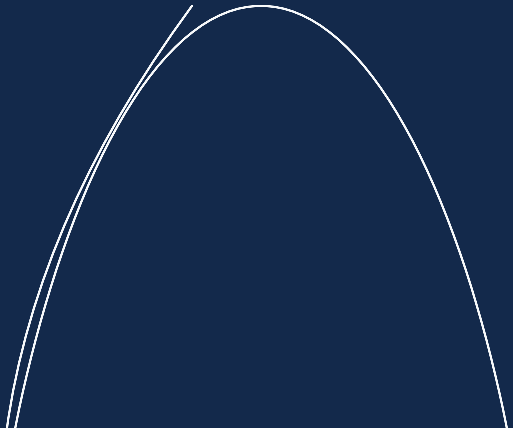
$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.

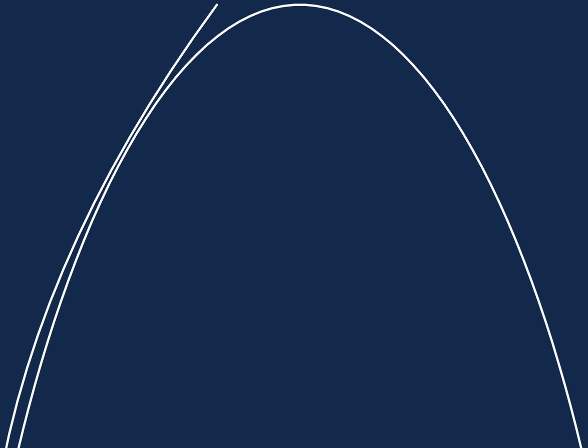


# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)

Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)

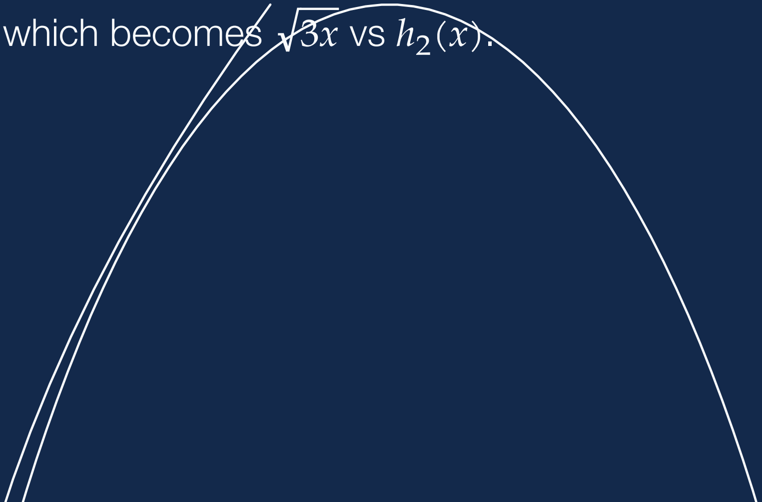
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.

# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)

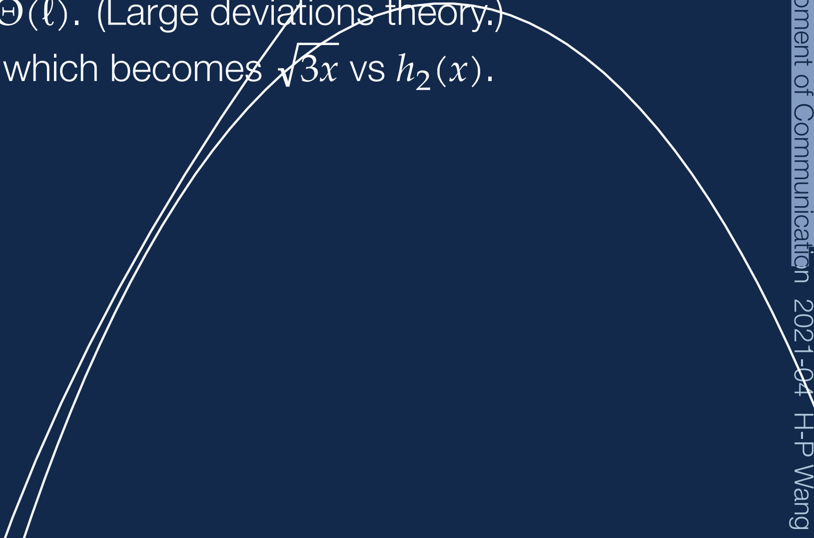
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

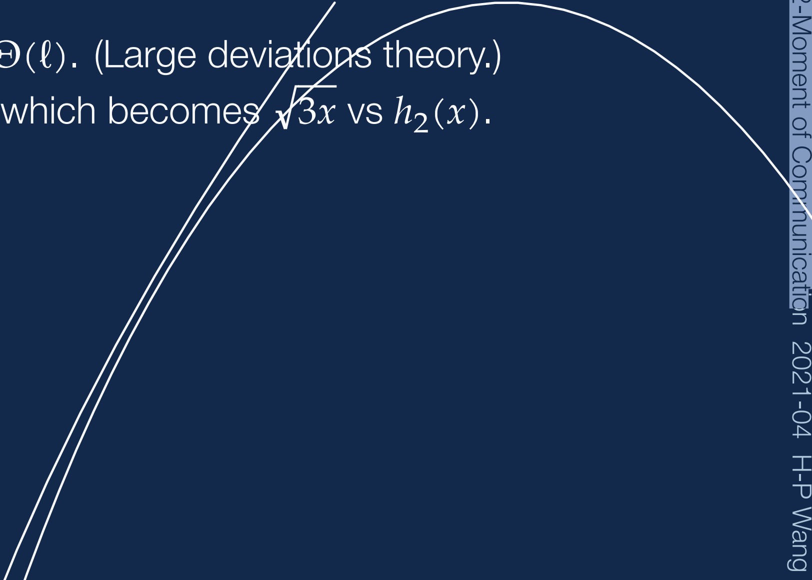
$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

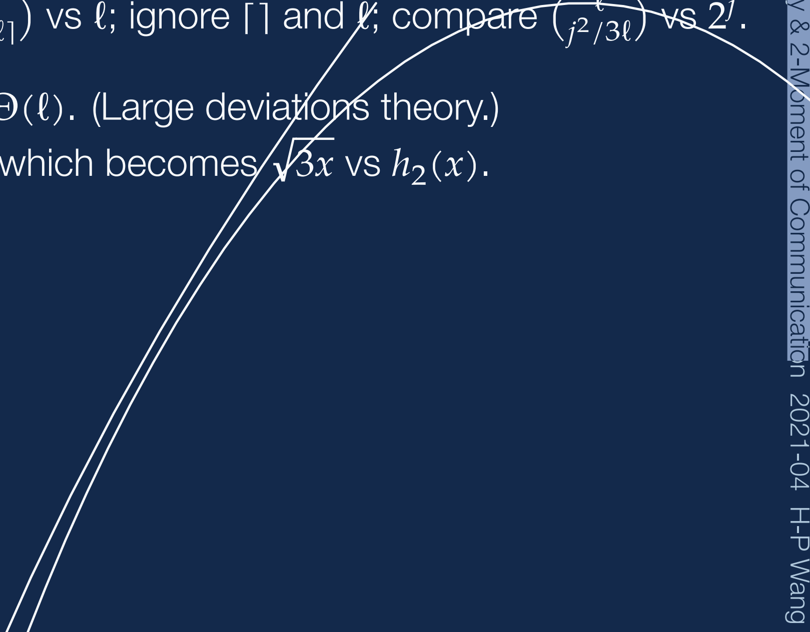
$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell \rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)
Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.

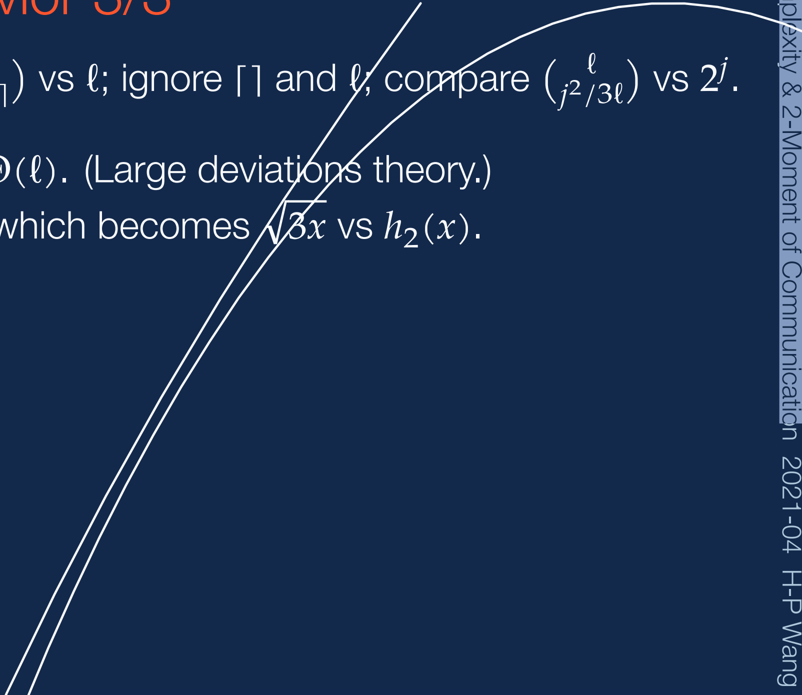


# Local LDP behavior 3/3

Boils down to $2^{-j}\binom{\ell}{\lceil j^2/3\ell\rceil}$ vs $\ell$; ignore $\lceil\,\rceil$ and $\ell$; compare $\binom{\ell}{j^2/3\ell}$ vs $2^j$.

$\binom{\ell}{d} \approx 2^{\ell h_2(d/\ell)}$ for $d = \Theta(\ell)$. (Large deviations theory.)

Hence $h_2(j^2/3\ell^2)$ vs $j$, which becomes $\sqrt{3x}$ vs $h_2(x)$.



# Local CLT behavior 1/4

Want to prove $\sum_{i=1}^{\ell} h(H(W^{(i)})) < 4\ell^{1/2+\alpha}$,
where $\alpha = \ln(\ln \ell) / \ln \ell$ and $h(z) := \min(z, 1-z)^{\alpha}$.

Break into three segments $\begin{cases} \sum_{i=\lceil H(W)+\ell^{-1/2+\alpha}\rceil+1}^{\ell} h(H(W^{(i)})) < \ell^{1/2+\alpha}, \\ \sum_{i=\lfloor H(W)-\ell^{-1/2+\alpha}\rfloor}^{\lceil H(W)+\ell^{-1/2+\alpha}\rceil} h(H(W^{(i)})) < 2\ell^{1/2+\alpha}, \\ \sum_{i=1}^{\lfloor H(W)-\ell^{-1/2+\alpha}\rfloor-1} h(H(W^{(i)})) < \ell^{1/2+\alpha}. \end{cases}$

# Local CLT behavior 1/4

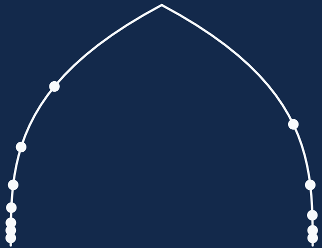
Want to prove $\sum_{i=1}^{\ell} h(H(W^{(i)})) < 4\ell^{1/2+\alpha}$,
where $\alpha = \ln(\ln \ell) / \ln \ell$ and $h(z) := \min(z, 1-z)^{\alpha}$.

Break into three segments $\begin{cases} \sum_{i=\lceil H(W)+\ell^{-1/2+\alpha} \rceil+1}^{\ell} h(H(W^{(i)})) < \ell^{1/2+\alpha}, \\ \sum_{i=\lfloor H(W)-\ell^{-1/2+\alpha} \rfloor}^{\lceil H(W)+\ell^{-1/2+\alpha} \rceil} h(H(W^{(i)})) < 2\ell^{1/2+\alpha}, \\ \sum_{i=1}^{\lfloor H(W)-\ell^{-1/2+\alpha} \rfloor-1} h(H(W^{(i)})) < \ell^{1/2+\alpha}. \end{cases}$

# Local CLT behavior 1/4

Want to prove $\sum_{i=1}^{\ell} h(H(W^{(i)})) < 4\ell^{1/2+\alpha}$,
where $\alpha = \ln(\ln \ell) / \ln \ell$ and $h(z) := \min(z, 1-z)^{\alpha}$.

Break into three segments
$$\begin{cases} \displaystyle\sum_{i=\lceil H(W)+\ell^{-1/2+\alpha}\rceil+1}^{\ell} h(H(W^{(i)})) < \ell^{1/2+\alpha}, \\ \displaystyle\sum_{i=\lfloor H(W)-\ell^{-1/2+\alpha}\rfloor}^{\lceil H(W)+\ell^{-1/2+\alpha}\rceil} h(H(W^{(i)})) < 2\ell^{1/2+\alpha}, \\ \displaystyle\sum_{i=1}^{\lfloor H(W)-\ell^{-1/2+\alpha}\rfloor-1} h(H(W^{(i)})) < \ell^{1/2+\alpha}. \end{cases}$$



# Local CLT behavior 2/4

Want $\sum_{i=j+1}^{\ell} h(H(W^{(i)})) < \ell^{1/2+\alpha}$, where $j := \lceil H(W) + \ell^{-1/2+\alpha} \rceil$.

Jensen LHS; want to show $(\ell - j)h\Big(\frac{1}{\ell - j} \sum_{i=j+1}^{\ell} H(W^{(i)})\Big) < \ell^{1/2+\alpha}$.

$$\begin{aligned} &\vdots \\ W^{(\ell-2)} &:= (U_{\ell-2} \mid U_1^{\ell-3} Y_1^{\ell}), \\ W^{(\ell-1)} &:= (U_{\ell-1} \mid U_1^{\ell-2} Y_1^{\ell}), \\ W^{(\ell)} &:= (U_{\ell} \mid U_1^{\ell-1} Y_1^{\ell}). \end{aligned}$$

$$\sum_{i=j+1}^{\ell} H(W^{(i)}) = H(U_{j+1}^{\ell} \mid U_1^{j} Y_1^{\ell}).$$

# Local CLT behavior 2/4

Want $\sum_{i=j+1}^{\ell} h(H(W^{(i)})) < \ell^{1/2+\alpha}$, where $j := \lceil H(W) + \ell^{-1/2+\alpha} \rceil$.

Jensen LHS; want to show $(\ell - j)h\Big(\frac{1}{\ell-j}\sum_{i=j+1}^{\ell} H(W^{(i)})\Big) < \ell^{1/2+\alpha}$.

$$\begin{aligned} &\vdots \\ W^{(\ell-2)} &:= (U_{\ell-2} \mid U_1^{\ell-3} Y_1^{\ell}), \\ W^{(\ell-1)} &:= (U_{\ell-1} \mid U_1^{\ell-2} Y_1^{\ell}), \\ W^{(\ell)} &:= (U_\ell \mid U_1^{\ell-1} Y_1^{\ell}). \end{aligned}$$

$$\sum_{i=j+1}^{\ell} H(W^{(i)}) = H(U_{j+1}^{\ell} \mid U_1^{j} Y_1^{\ell}).$$

# Local CLT behavior 2/4

Want $\sum_{i=j+1}^{\ell} h(H(W^{(i)})) < \ell^{1/2+\alpha}$, where $j := \lceil H(W) + \ell^{-1/2+\alpha} \rceil$.

Jensen LHS; want to show $(\ell - j)h\Big(\frac{1}{\ell-j}\sum_{i=j+1}^{\ell} H(W^{(i)})\Big) < \ell^{1/2+\alpha}$.

$$\begin{aligned}
&\vdots \\
W^{(\ell-2)} &:= (U_{\ell-2} \mid U_1^{\ell-3} Y_1^{\ell}), \\
W^{(\ell-1)} &:= (U_{\ell-1} \mid U_1^{\ell-2} Y_1^{\ell}), \\
W^{(\ell)} &:= (U_{\ell} \mid U_1^{\ell-1} Y_1^{\ell}).
\end{aligned}$$

$$\sum_{i=j+1}^{\ell} H(W^{(i)}) = H(U_{j+1}^{\ell} \mid U_1^{j} Y_1^{\ell}).$$

# Local CLT behavior 3/4

What is $H(U_{j+1}^{\ell} \mid U_1^j Y_1^{\ell})$? $(j := \lceil H(W) + \ell^{-1/2+\alpha} \rceil)$

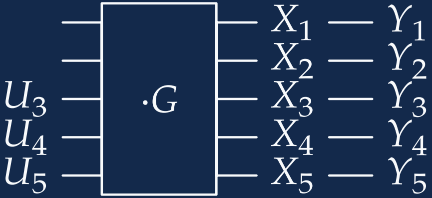
It is the conditional entropy of noisy-channel coding.


Gallager has good bounds.

# Local CLT behavior 3/4

What is $H(U_{j+1}^{\ell} \mid U_1^j Y_1^{\ell})$? $\quad (j := \lceil H(W) + \ell^{-1/2+\alpha} \rceil)$

It is the conditional entropy of noisy-channel coding.

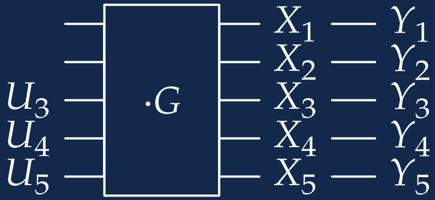



Gallager has good bounds.

# Local CLT behavior 3/4

What is $H(U_{j+1}^{\ell} \mid U_1^j Y_1^{\ell})$?   $(j := \lceil H(W) + \ell^{-1/2+\alpha} \rceil)$

It is the conditional entropy of noisy-channel coding.




Gallager has good bounds.

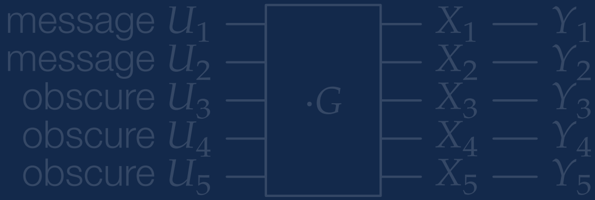
# Local CLT behavior 4/4

The last segment: $\sum_{i=1}^{j} h(H(W^{(i)})) < 4\ell^{1/2+\alpha}$.

Pre-process by Jensen inequality: $jh\left(\frac{1}{j}\sum_{i=1}^{j+1} H(W^{(i)})\right) < 4\ell^{1/2+\alpha}$.

Chain rule: $jh\left(\frac{1}{j}H(U_1^j \mid Y_1^\ell)\right)$, but what is $H(U_1^j \mid Y_1^\ell)$?

Wiretap channel [new idea];
Hayashi has good bounds.



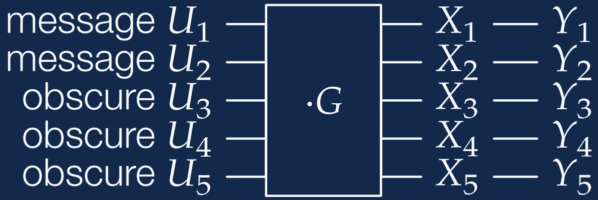
# Local CLT behavior 4/4

The last segment: $\sum_{i=1}^{j} h(H(W^{(i)})) < 4\ell^{1/2+\alpha}$.

Pre-process by Jensen inequality: $jh\Big(\frac{1}{j}\sum_{i=1}^{j+1} H(W^{(i)})\Big) < 4\ell^{1/2+\alpha}$.

Chain rule: $jh\Big(\frac{1}{j}H(U_1^j \mid Y_1^\ell)\Big)$, but what is $H(U_1^j \mid Y_1^\ell)$?

Wiretap channel [new idea];
Hayashi has good bounds.



# A calculus machinery [new idea]

Given local LDP behavior: $Z(W^{(j)}) \le \ell e^{qZ(W)\ell}(qZ(W))^{\lceil j^2/3\ell \rceil}$

and local CLT behavior: $\sum_{j=1}^{\ell} h(H(W^{(j)})) < 4\ell^{1/2+\alpha}$.

eigen: $E[h(H_{n+1}) \mid H_0, \dots, H_n] \le \ell^{-1/2+3\alpha} h(H_n)$.

en23: $P\{Z_n < e^{-n^{2/3}}\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

een13: $P\left\{Z_n < \exp(-e^{n^{1/3}})\right\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

elpin: $P\{Z_n < e^{-\ell^{\pi n}}\} > 1 - H(W) - \ell^{-\rho n}$.

# A calculus machinery [new idea]

Given local LDP behavior: $Z(W^{(j)}) \leq \ell e^{qZ(W)\ell}(qZ(W))^{\lceil j^2/3\ell \rceil}$

and local CLT behavior: $\sum_{j=1}^{\ell} h(H(W^{(j)})) < 4\ell^{1/2+\alpha}$.

eigen: $E[h(H_{n+1}) \mid H_0, \ldots, H_n] \leq \ell^{-1/2+3\alpha} h(H_n)$.

en23: $P\{Z_n < e^{-n^{2/3}}\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

een13: $P\left\{Z_n < \exp(-e^{n^{1/3}})\right\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

elpin: $P\{Z_n < e^{-\ell^{\pi n}}\} > 1 - H(W) - \ell^{-\rho n}$.

# A calculus machinery [new idea]

Given local LDP behavior: $Z(W^{(j)}) \le \ell e^{qZ(W)\ell}(qZ(W))^{\lceil j^2/3\ell \rceil}$

and local CLT behavior: $\sum_{j=1}^{\ell} h(H(W^{(j)})) < 4\ell^{1/2+\alpha}$.

eigen: $E[h(H_{n+1}) \mid H_0, \ldots, H_n] \le \ell^{-1/2+3\alpha} h(H_n)$.

en23: $P\{Z_n < e^{-n^{2/3}}\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

een13: $P\left\{Z_n < \exp(-e^{n^{1/3}})\right\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

elpin: $P\{Z_n < e^{-\ell^{\pi n}}\} > 1 - H(W) - \ell^{-\rho n}$.

# A calculus machinery [new idea]

Given local LDP behavior: $Z(W^{(j)}) \leq \ell e^{qZ(W)\ell}(qZ(W))^{\lceil j^2/3\ell \rceil}$

and local CLT behavior: $\sum_{j=1}^{\ell} h(H(W^{(j)})) < 4\ell^{1/2+\alpha}$.

eigen: $E[h(H_{n+1}) \mid H_0, \ldots, H_n] \leq \ell^{-1/2+3\alpha} h(H_n)$.

en23: $P\{Z_n < e^{-n^{2/3}}\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

een13: $P\left\{Z_n < \exp(-e^{n^{1/3}})\right\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

elpin: $P\{Z_n < e^{-\ell^{\pi n}}\} > 1 - H(W) - \ell^{-\rho n}$.

# A calculus machinery [new idea]

Given local LDP behavior: $Z(W^{(j)}) \leq \ell e^{qZ(W)\ell}(qZ(W))^{\lceil j^2/3\ell \rceil}$

and local CLT behavior: $\sum_{j=1}^{\ell} h(H(W^{(j)})) < 4\ell^{1/2+\alpha}$.

eigen: $E[h(H_{n+1}) \mid H_0, \ldots, H_n] \leq \ell^{-1/2+3\alpha} h(H_n)$.

en23: $P\{Z_n < e^{-n^{2/3}}\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

een13: $P\left\{Z_n < \exp(-e^{n^{1/3}})\right\} > 1 - H(W) - \ell^{(-1/2+4\alpha)n}$.

elpin: $P\{Z_n < e^{-\ell^{\pi n}}\} > 1 - H(W) - \ell^{-\rho n}$.

# Summary of the proof

For local LDP behavior, we investigate the distance of a random matrix.

For local CLT, noisy-channel coding and wiretap-channel coding.

For the global MDP behavior, a calculus machinery is invented/used.

# Summary of the proof

For local LDP behavior, we investigate the distance of a random matrix.

For local CLT, noisy-channel coding and wiretap-channel coding.

For the global MDP behavior, a calculus machinery is invented/used.

# Summary of the proof

For local LDP behavior, we investigate the distance of a random matrix.

For local CLT, noisy-channel coding and wiretap-channel coding.

For the global MDP behavior, a calculus machinery is invented/used.

# Summary of my results so far

For all $\pi + 2\rho < 1$, there exist codes with
error probability $P_\mathrm{e} < e^{-N^\pi}$ and code rate $R > C - N^{-\rho}$.

When only $2 \times 2$ kernels are allowed, at least $\pi, \rho > 0$.

It happens that they have complexity $O(\log N)$ per bit.

Can we reduce the complexity further
(at the expense of worse performance etc)?

# Summary of my results so far

For all $\pi + 2\rho < 1$, there exist codes with error probability $P_{\mathrm{e}} < e^{-N^{\pi}}$ and code rate $R > C - N^{-\rho}$.

When only $2 \times 2$ kernels are allowed, at least $\pi, \rho > 0$.

It happens that they have complexity $O(\log N)$ per bit.

Can we reduce the complexity further
(at the expense of worse performance etc)?

# Summary of my results so far

For all $\pi + 2\rho < 1$, there exist codes with
error probability $P_\mathrm{e} < e^{-N^\pi}$ and code rate $R > C - N^{-\rho}$.

When only $2 \times 2$ kernels are allowed, at least $\pi, \rho > 0$.

It happens that they have complexity $O(\log N)$ per bit.

Can we reduce the complexity further
(at the expense of worse performance etc)?

# Summary of my results so far

For all $\pi + 2\rho < 1$, there exist codes with
error probability $P_{\mathrm{e}} < e^{-N^{\pi}}$ and code rate $R > C - N^{-\rho}$.

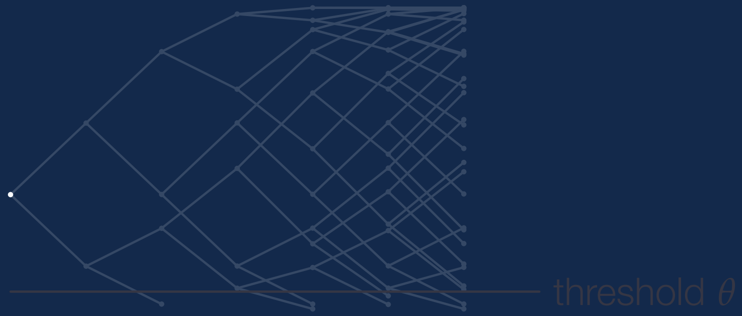
When only $2 \times 2$ kernels are allowed, at least $\pi, \rho > 0$.

It happens that they have complexity $O(\log N)$ per bit.

Can we reduce the complexity further
(at the expense of worse performance etc)?

# Prune the tree for simplicity

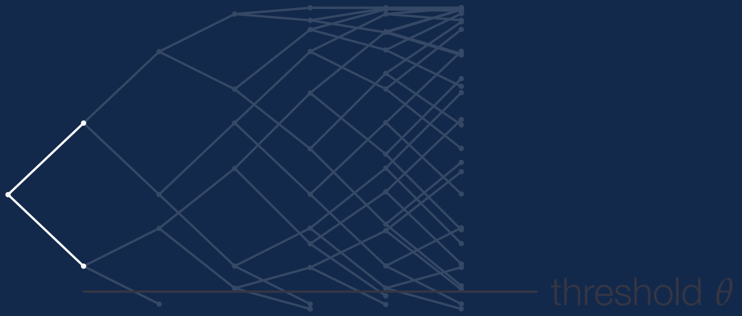
The bottom channel is good enough before we reach our favorite $n$.




Why do we apply transform any further? (Answer: we don't!)

# Prune the tree for simplicity

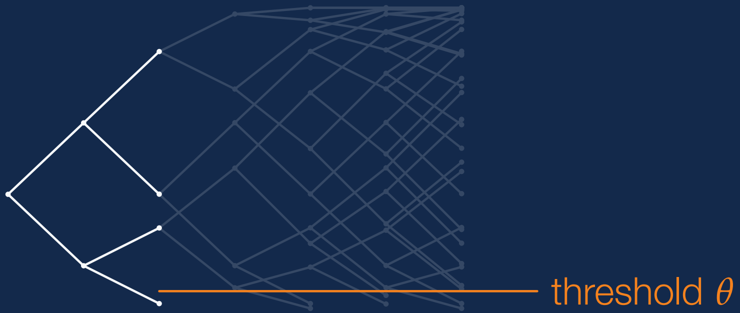
The bottom channel is good enough before we reach our favorite $n$.




Why do we apply transform any further? (Answer: we don't!)

# Prune the tree for simplicity

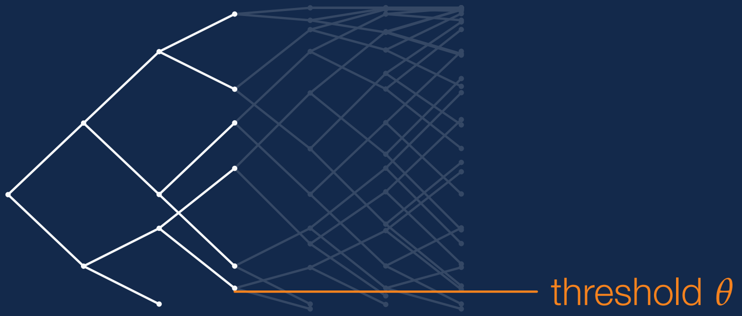
The bottom channel is good enough before we reach our favorite $n$.




Why do we apply transform any further? (Answer: we don't!)

# Prune the tree for simplicity

The bottom channel is good enough before we reach our favorite $n$.




Why do we apply transform any further? (Answer: we don't!)

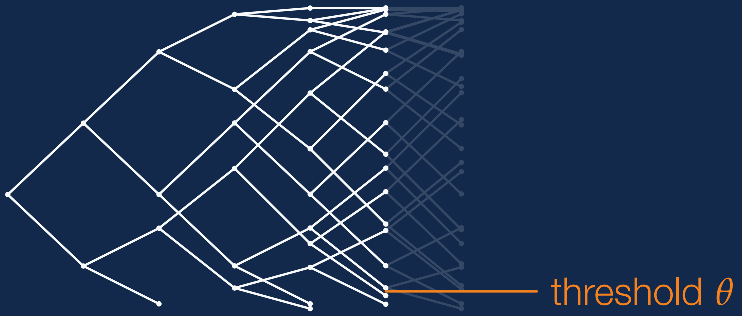
# Prune the tree for simplicity

The bottom channel is good enough before we reach our favorite $n$.


Why do we apply transform any further? (Answer: we don't!)

# Prune the tree for simplicity

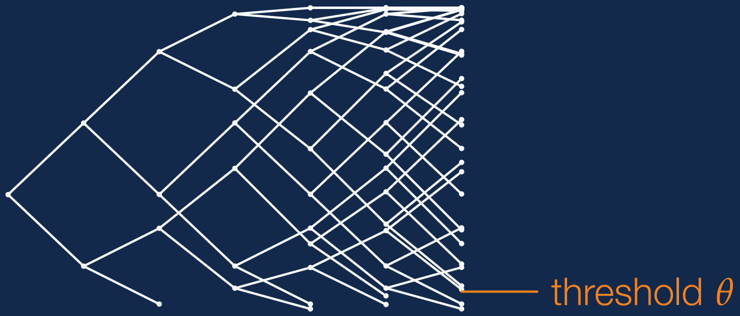
The bottom channel is good enough before we reach our favorite $n$.




Why do we apply transform any further? (Answer: we don't!)

# Prune the tree for simplicity

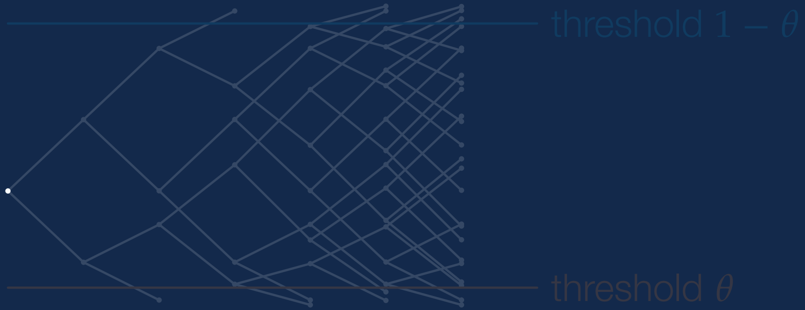
The bottom channel is good enough before we reach our favorite $n$.




Why do we apply transform any further? (Answer: we don't!)

# Prune the other side

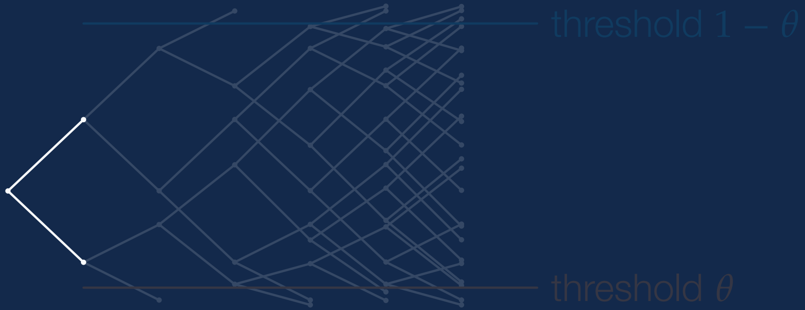
Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Prune the other side

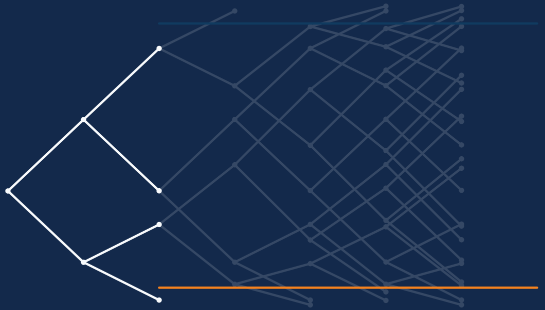
Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Prune the other side

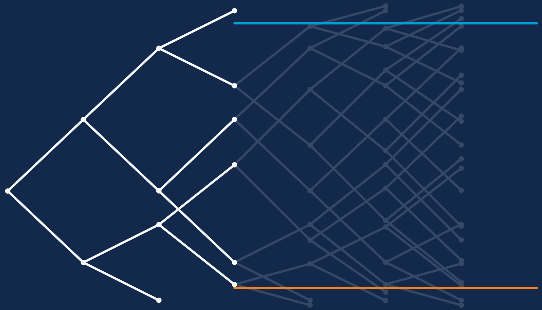
Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Prune the other side

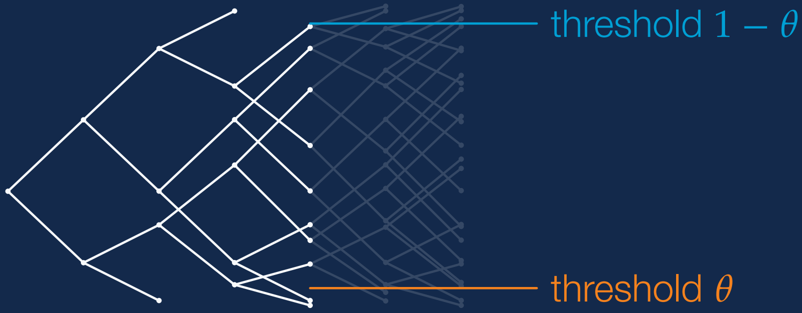
Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Prune the other side

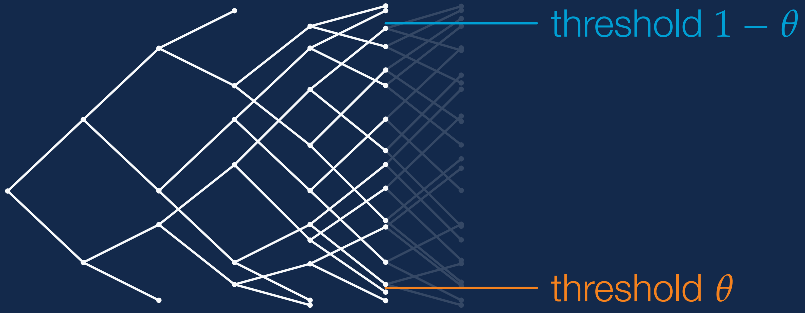
Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Prune the other side

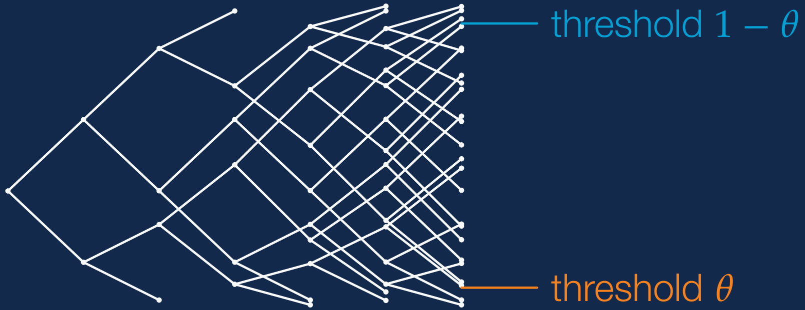
Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Prune the other side

Sometimes, the top channel is too bad.
Do we expect any of its descendants to be good enough?




We don't.

# Stopping time analysis

$W_n$ has children/needs further transformation if $\theta < H_n < 1 - \theta$.

Set $\theta = N^{-10}$; assume $m > O(\log(\log N))$, then $e^{-2^{\pi m}} < \theta$.

Then $P\{H_m < \theta\} > P\{H_m < e^{-2^{\pi m}}\} \geq H(W) - \ell^{-\rho m}$ and
$P\{H_m > 1 - \theta\} > P\{H_m > 1 - e^{-2^{\pi m}}\} \geq 1 - H(W) - \ell^{-\rho m}$

That is to say, $P\{\theta < H_m < 1 - \theta\} \leq 2\ell^{-\rho m}$, hard to stay in the middle.

# Stopping time analysis

$W_n$ has children/needs further transformation if $\theta < H_n < 1 - \theta$.

Set $\theta = N^{-10}$; assume $m > O(\log(\log N))$, then $e^{-2^{\pi m}} < \theta$.

Then $P\{H_m < \theta\} > P\{H_m < e^{-2^{\pi m}}\} \geq H(W) - \ell^{-\rho m}$ and $P\{H_m > 1 - \theta\} > P\{H_m > 1 - e^{-2^{\pi m}}\} \geq 1 - H(W) - \ell^{-\rho m}$

That is to say, $P\{\theta < H_m < 1 - \theta\} \leq 2\ell^{-\rho m}$, hard to stay in the middle.

# Stopping time analysis

$W_n$ has children/needs further transformation if $\theta < H_n < 1 - \theta$.

Set $\theta = N^{-10}$; assume $m > O(\log(\log N))$, then $e^{-2^{\pi m}} < \theta$.

Then $P\{H_m < \theta\} > P\{H_m < e^{-2^{\pi m}}\} \geq H(W) - \ell^{-\rho m}$ and
$P\{H_m > 1 - \theta\} > P\{H_m > 1 - e^{-2^{\pi m}}\} \geq 1 - H(W) - \ell^{-\rho m}$

That is to say, $P\{\theta < H_m < 1 - \theta\} \leq 2\ell^{-\rho m}$, hard to stay in the middle.

# Stopping time analysis

$W_n$ has children/needs further transformation if $\theta < H_n < 1 - \theta$.

Set $\theta = N^{-10}$; assume $m > O(\log(\log N))$, then $e^{-2^{\pi m}} < \theta$.

Then $P\{H_m < \theta\} > P\{H_m < e^{-2^{\pi m}}\} \geq H(W) - \ell^{-\rho m}$ and
$P\{H_m > 1 - \theta\} > P\{H_m > 1 - e^{-2^{\pi m}}\} \geq 1 - H(W) - \ell^{-\rho m}$

That is to say, $P\{\theta < H_m < 1 - \theta\} \leq 2\ell^{-\rho m}$, hard to stay in the middle.

# Geometric complexity

Complexity = #transformations = $\sum_{m=0}^{n} P\{\theta < H_m < 1-\theta\}$.

$$\begin{cases} \sum_{m=O(\log(\log N))}^{n} P\{\theta < H_m < 1-\theta\} \le \sum_{m=O(\log(\log N))}^{n} 2\ell^{-\rho m} = O(1), \\ \sum_{m=0}^{O(\log(\log N))} P\{\theta < H_m < 1-\theta\} \le \sum_{m=0}^{O(\log(\log N))} 1 = O(\log(\log N)). \end{cases}$$

Complexity is $O(\log(\log N))$ per bit, or $O(N\log(\log N))$ per block.

# Geometric complexity

Complexity = #transformations = $\sum_{m=0}^{n} P\{\theta < H_m < 1-\theta\}$.

$$\begin{cases} \displaystyle\sum_{m=O(\log(\log N))}^{n} P\{\theta < H_m < 1-\theta\} \le \sum_{m=O(\log(\log N))}^{n} 2\ell^{-\rho m} = O(1), \\ \displaystyle\sum_{m=0}^{O(\log(\log N))} P\{\theta < H_m < 1-\theta\} \le \sum_{m=0}^{O(\log(\log N))} 1 = O(\log(\log N)). \end{cases}$$

Complexity is $O(\log(\log N))$ per bit, or $O(N\log(\log N))$ per block.

# Geometric complexity

Complexity = #transformations = $\sum_{m=0}^{n} P\{\theta < H_m < 1-\theta\}$.

$$\begin{cases} \sum_{m=O(\log(\log N))}^{n} P\{\theta < H_m < 1-\theta\} \leq \sum_{m=O(\log(\log N))}^{n} 2\ell^{-\rho m} = O(1), \\ \sum_{m=0}^{O(\log(\log N))} P\{\theta < H_m < 1-\theta\} \leq \sum_{m=0}^{O(\log(\log N))} 1 = O(\log(\log N)). \end{cases}$$

Complexity is $O(\log(\log N))$ per bit, or $O(N \log(\log N))$ per block.

# Summary of pruning and whatnot

There exist codes with complexity $O(\log(\log N))$ per bit, error probability $P_{\mathrm{e}} < N^{-9}$, and code rate $R = C - N^{-\rho}$.

(Earlier) we have codes with complexity $O(\log N)$ per bit, error probability $P_{\mathrm{e}} < e^{-N^{\pi}}$, and code rate $R > C - N^{-\rho}$.

Are there codes in between? Yes, continuously.

# Summary of pruning and whatnot

There exist codes with complexity $O(\log(\log N))$ per bit, error probability $P_{\mathrm{e}} < N^{-9}$, and code rate $R = C - N^{-\rho}$.

(Earlier) we have codes with complexity $O(\log N)$ per bit, error probability $P_{\mathrm{e}} < e^{-N^{\pi}}$, and code rate $R > C - N^{-\rho}$.

Are there codes in between? Yes, continuously.

# Summary of pruning and whatnot

There exist codes with complexity $O(\log(\log N))$ per bit, error probability $P_\mathrm{e} < N^{-9}$, and code rate $R = C - N^{-\rho}$.

(Earlier) we have codes with complexity $O(\log N)$ per bit, error probability $P_\mathrm{e} < e^{-N^\pi}$, and code rate $R > C - N^{-\rho}$.

Are there codes in between? Yes, continuously.

# Summary of pruning and whatnot

There exist codes with complexity $O(\log(\log N))$ per bit, error probability $P_{\mathrm{e}} < N^{-9}$, and code rate $R = C - N^{-\rho}$.

(Earlier) we have codes with complexity $O(\log N)$ per bit, error probability $P_{\mathrm{e}} < e^{-N^{\pi}}$, and code rate $R > C - N^{-\rho}$.

Are there codes in between? Yes, continuously.

# Summary

Log-log code taken from (with Duursma)
Log-logarithmic Time Pruned Polar Coding
https://ieeexplore.ieee.org/document/9274497.

MDP code taken from (with Duursma)
Polar Codes' Simplicity, Random Codes' Durability
https://ieeexplore.ieee.org/document/9274521.

# Question?

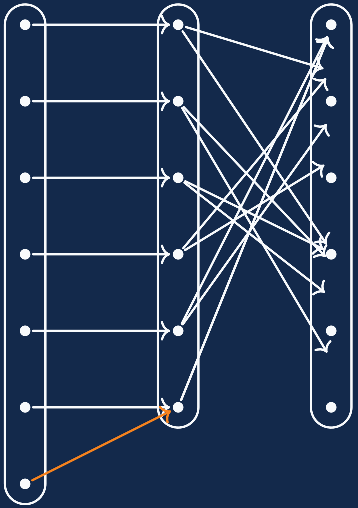
Predefined questions:
What does each chapter in dissertation do?
Why input alphabet is finite field? What is the advantage?
Definition of Bhattacharyya parameter?
References for XYZ?
Your contribution over others?
Future plan?

| Code | Error | Gap | Complexity | Channel |
|---|---|---|---|---|
| random | $e^{-N^\pi}$ | $N^{-\rho}$ | $\exp(N)$ | DMC |
| concatenation | $e^{-N^\pi}$ | $\to 0$ | $\operatorname{poly}(N)$ | DMC |
| RM | $\to 0$ | $\to 0$ | $O(N^2)$ | BEC |
| LDPC | $\to 0$ | $\to 0$ | unclear | SBDMC |
| RA family | $\to 0$ | $\to 0$ | $O(1)$ | BEC |
| old prune | $e^{-N^{1/2}}$ | $O(1)$ | $\Theta(\log N)$ | SBDMC |
| loglog-polar [W.] | $e^{-n^\tau}$ | $N^{-\rho}$ | $O(\log(\log N))$ | DMC |
| MDP-polar [W.] | $e^{-N^\pi}$ | $N^{-\rho}$ | $O(\log N)$ | DMC |

| | Symmetric | | | | | Asymmetric | |
|---|---|---|---|---|---|---|---|
| Par | BEC | SBDMC | p-ary | q-ary | finite | BDMC | a-finite |
| LLN | [Ari09] | [Ari09] | [ŞTA09] | [ŞTA09] | [ŞTA09] | [SRDR12] | [W.] |
| LDP$^\star$ | [AT09] | [AT09] | [ŞTA09] | [MT10] | [Sas11] | [HY13] | [W.] |
| CLT$^\star$ | [KMTU10] | [HAU14] | [BGN$^+$18] | [W.] | [W.] | [W.] | [W.] |
| MDP$^\star$ | [GX15] | [GX15] | [BGS18] | [W.] | [W.] | [W.] | [W.] |
| LDP | [KSU10] | [KSU10] | [W.] | [W.] | [W.] | [W.] | [W.] |
| CLT | [FHMV18] | [GRY20] | [W.] | [W.] | [W.] | [W.] | [W.] |
| MDP | [W.] | [W.] | [W.] | [W.] | [W.] | [W.] | [W.] |

# Input alphabet [new idea]

$$\begin{bmatrix} W(y_1|1) & W(y_2|1) & W(y_3|1) & \cdots \\ W(y_1|2) & W(y_2|2) & W(y_3|2) & \cdots \\ W(y_1|3) & W(y_2|3) & W(y_3|3) & \cdots \\ W(y_1|4) & W(y_2|4) & W(y_3|4) & \cdots \\ W(y_1|5) & W(y_2|5) & W(y_3|5) & \cdots \\ W(y_1|6) & W(y_2|6) & W(y_3|6) & \cdots \\ W(y_1|6) & W(y_2|6) & W(y_3|6) & \cdots \end{bmatrix}$$



# Asymmetric channels [HY13]

Recall $U_j$ is the coordinate as in $X_1^\ell := U_1^\ell \cdot G$.
The difficulty of asymmetric channels is $U_j$ being nonuniform and dependent.

Define synthetic channel $Q^{(i)} := (U_i \mid U_1^{i-1})$.
Define tree $Q^{(i)}, (Q^{(i)})^{(j)}, ((Q^{(i)})^{(j)})^{(k)}, \ldots$; define channel process $\{\boldsymbol{Q}_n\}$.
It polarizes, and at the same pace.

High $H(\boldsymbol{Q}_n)$ low $H(\boldsymbol{W}_n)$ vs both high vs both low.

# Bhattacharyya parameter

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

Non-binary $:= \frac{1}{q-1} \sum_{\substack{x,x' \in \mathbb{F}_q \\ x \neq x'}} \sum_{y \in \mathcal{Y}} \sqrt{W(x,y)W(x',y)}$.

[New idea] $:= \max_{0 \neq d \in \mathbb{F}_q} \sum_{x \in \mathbb{F}_q} \sum_{y \in \mathcal{Y}} \sqrt{W(x,y)W(x+d,y)}$.

# A List of Important Contributions (Chronological)

A combinatorial trick to recover scaling exponent (een13 → elpin).

The pruning technique/stopping time analysis/log-log complexity.

Improved combinatorial trick (en23 → een13 → elpin).

Dynamic kernel, later random dynamic kernel.

Alphabet reduction to finite field (trivial but powerful and last mile).

Improve definition of Bhattacharyya parameter, then and FTPCZ.

Reducing local CLT to noisy-channel and wiretap-channel coding.

For wiretap bound, extend the universal bound via continuation.

A topological argument to show positive $\varrho$ (that is, CLT$^\star$).

[Ari09] E. Arikan.
Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels.
*IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009. doi:10.1109/TIT.2009.2021379.

[AT09] E. Arikan and E. Telatar.
On the rate of channel polarization.
In *2009 IEEE International Symposium on Information Theory*, pages 1493–1495, June 2009. doi:10.1109/ISIT.2009.5205856.

[AW10] Y. Altuğ and A. B. Wagner.
Moderate deviation analysis of channel coding: Discrete memoryless case.
In *2010 IEEE International Symposium on Information Theory*, pages 265–269, June 2010. doi:10.1109/ISIT.2010.5513319.

[AW14] Y. Altuğ and A. B. Wagner.
Moderate deviations in channel coding.
*IEEE Transactions on Information Theory*, 60(8):4417–4426, Aug 2014. doi:10.1109/TIT.2014.2323418.

[BF02] A. Barg and G. D. Forney.
Random codes: minimum distances and error exponents.
*IEEE Transactions on Information Theory*, 48(9):2568–2573, Sep. 2002. doi:10.1109/TIT.2002.800480.

[BGN+18] Jarosław Błasiok, Venkatesan Guruswami, Preetum Nakkiran, Atri Rudra, and Madhu Sudan.
General strong polarization.
In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 485–492, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3188745.3188816.

[BGS18] Jaroslaw Blasiok, Venkatesan Guruswami, and Madhu Sudan.
Polar Codes with Exponentially Small Error at Finite Block Length.
In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:17, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. URL: http://drops.dagstuhl.de/opus/volltexte/2018/9438, doi:10.4230/LIPIcs.APPROX-RANDOM.2018.34.

[BKB04] D. Baron, M. A. Khojastepour, and R. G. Baraniuk.
How quickly can we approach channel capacity?
In *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, volume 1, pages 1096–1100 Vol.1, Nov 2004. doi:10.1109/ACSSC.2004.1399310.

[Bla74] R. Blahut.
Hypothesis testing and information theory.
*IEEE Transactions on Information Theory*, 20(4):405–417, July 1974. doi:10.1109/TIT.1974.1055254.

[Dob61] R. L. Dobrushin.
Mathematical problems in the shannon theory of optimal coding of information.
In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 211–252, Berkeley, Calif., 1961. University of California Press. URL: https://projecteuclid.org/euclid.bsmsp/1200512168.

[DZF16] Y. Domb, R. Zamir, and M. Feder.
The random coding bound is tight for the average linear code or lattice.
*IEEE Transactions on Information Theory*, 62(1):121–130, Jan 2016. doi:10.1109/TIT.2015.2496308.

[Fan61] R.M. Fano.
*Transmission of Information: A Statistical Theory of Communications*.
M.I.T. Press, 1961. URL: https://books.google.com.tw/books?id=VSYIAQAAIAAJ.

[FHMV18] A. Fazeli, H. Hassani, M. Mondelli, and A. Vardy.
Binary linear codes with optimal scaling: Polar codes with large kernels.
In *2018 IEEE Information Theory Workshop (ITW)*, pages 1–5, Nov 2018. doi:10.1109/ITW.2018.8613428.

[Gal65] R. Gallager.
A simple derivation of the coding theorem and some applications.
*IEEE Transactions on Information Theory*, 11(1):3–18, January 1965. doi:10.1109/TIT.1965.1053730.

[Gal68] Robert G. Gallager.
*Information Theory and Reliable Communication*.
John Wiley & Sons, Inc., USA, 1968. doi:10.5555/578869.

[Gal73] R. Gallager.
The random coding bound is tight for the average code (corresp.).
*IEEE Transactions on Information Theory*, 19(2):244–246, March 1973. doi:10.1109/TIT.1973.1054971.

[GRY20] Venkatesan Guruswami, Andrii Riazanov, and Min Ye.
Arikan meets shannon: Polar codes with near-optimal convergence to channel capacity.
In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 552–564, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3357713.3384323.

[GX15] V. Guruswami and P. Xia.
Polar codes: Speed of polarization and polynomial gap to capacity.
*IEEE Transactions on Information Theory*, 61(1):3–16, Jan 2015. doi:10.1109/TIT.2014.2371819.

[HAU14] S. H. Hassani, K. Alishahi, and R. L. Urbanke.
Finite-length scaling for polar codes.
*IEEE Transactions on Information Theory*, 60(10):5875–5898, Oct 2014. doi:10.1109/TIT.2014.2341919.

[Hay09] M. Hayashi.
Information spectrum approach to second-order coding rate in channel coding.
*IEEE Transactions on Information Theory*, 55(11):4947–4966, Nov 2009. doi:10.1109/TIT.2009.2030478.

[HT15] M. Hayashi and V. Y. F. Tan.
Erasure and undetected error probabilities in the moderate deviations regime.
In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1821–1825, June 2015. doi:10.1109/ISIT.2015.7282770.

[HY13] J. Honda and H. Yamamoto.
Polar coding without alphabet extension for asymmetric models.
*IEEE Transactions on Information Theory*, 59(12):7829–7838, Dec 2013. doi:10.1109/TIT.2013.2282305.

[iFLM11] A. G. i. Fàbregas, I. Land, and A. Martinez.
Extremes of random coding error exponents.
In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2896–2898, July 2011. doi:10.1109/ISIT.2011.6034105.

[KMTU10] S. B. Korada, A. Montanari, E. Telatar, and R. Urbanke.
An empirical scaling law for polar codes.
In *2010 IEEE International Symposium on Information Theory*, pages 884–888, June 2010. doi:10.1109/ISIT.2010.5513579.

[KSU10] S. B. Korada, E. Sasoglu, and R. Urbanke.
Polar codes: Characterization of exponent, bounds, and constructions.
*IEEE Transactions on Information Theory*, 56(12):6253–6264, Dec 2010. doi:10.1109/TIT.2010.2080990.

[MT10] R. Mori and T. Tanaka.
Channel polarization on q-ary discrete memoryless channels by arbitrary kernels.
In *2010 IEEE International Symposium on Information Theory*, pages 894–898, June 2010. doi:10.1109/ISIT.2010.5513568.

[PPV10] Y. Polyanskiy, H. V. Poor, and S. Verdu.
Channel coding rate in the finite blocklength regime.
*IEEE Transactions on Information Theory*, 56(5):2307–2359, May 2010. doi:10.1109/TIT.2010.2043769.

[PV10] Y. Polyanskiy and S. Verdú.
Channel dispersion and moderate deviations limits for memoryless channels.
In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1334–1339, Sept 2010. doi:10.1109/ALLERTON.2010.5707068.

[Sas11] Eren Sasoglu.
*Polar Coding Theorems for Discrete Systems*.
PhD thesis, École polytechnique fédérale de Lausanne, Lausanne, 2011. URL: http://infoscience.epfl.ch/record/168993, doi:10.5075/epfl-thesis-5219.

[SGB67] C.E. Shannon, R.G. Gallager, and E.R. Berlekamp.
Lower bounds to error probability for coding on discrete memoryless channels. i.
*Information and Control*, 10(1):65 – 103, 1967. URL: http://www.sciencedirect.com/science/article/pii/S0019995867900526, doi:https://doi.org/10.1016/S0019-9958(67)90052-6.

[SRDR12] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner.
Achieving the capacity of any dmc using only polar codes.
In *2012 IEEE Information Theory Workshop*, pages 114–118, Sep. 2012. doi:10.1109/ITW.2012.6404638.

[ŞTA09] E. Şaşoğlu, E. Telatar, and E. Arikan.
Polarization for arbitrary discrete memoryless channels.
In *2009 IEEE Information Theory Workshop*, pages 144–148, Oct 2009. doi:10.1109/ITW.2009.5351487.

[Str62] V. Strassen.
Asymptotische abschätzungen in shannons informationstheorie.
In *Transactions of the Third Prague Conference on Information Theory*, pages 689–723. Publishing House of the Czechoslovak Academy of Sciences, 1962. URL: https://www.math.cornell.edu/~pmlut/strassen.pdf.

[WEI60] LIONEL WEISS.
On the strong converse of the coding theorem for symmetric channels without memory.
*Quarterly of Applied Mathematics*, 18(3):209–214, 1960. URL: http://www.jstor.org/stable/43636330.

[Wol57] J. Wolfowitz.
The coding of messages subject to chance errors.
*Illinois J. Math.*, 1(4):591–606, 12 1957. doi:10.1215/ijm/1255380682.