Uniqueness of BP fixed point for Ising models

Yury Polyanskiy

EECS
Massachusetts Institute of Technology

Feb. 25, 2022 Harvard, Cambridge, MA

with: Qian Yu (Princeton) – on the job market!
What is “BP” in the title?

Belief propagation (BP) operator $Q$ is a map of probability measures:

- Fix a probability measure $\mu$ on $\mathbb{R}$ and let $Q\mu$ be the law of

$$R = \sum_{i=1}^{d}(-1)^{X_{i}}F_{\delta}(R_{i}) ,$$

where $R_{i} \overset{iid}{\sim} \mu$ independent of $X_{i} \overset{iid}{\sim} \text{Bern}(\delta)$ and

$$F_{\delta}(x) = 2\text{atanh}((1 - 2\delta)\tanh(x/2))$$
Belief propagation (BP) operator $Q$ is a map of probability measures:

- Fix a probability measure $\mu$ on $\mathbb{R}$ and let $Q\mu$ be the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i),$$

where $R_i \overset{iid}{\sim} \mu$ independent of $X_i \overset{iid}{\sim} \text{Bern}(\delta)$ and

$$F_\delta(x) = 2 \text{atanh}((1 - 2\delta) \tanh(x/2))$$
Belief propagation (BP) operator $Q$ is a map of probability measures:

- Fix a probability measure $\mu$ on $\mathbb{R}$ and let $Q\mu$ be the law of

$$
R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i),
$$

where $R_i \sim iid \mu$ independent of $X_i \sim iid \ Bern(\delta)$ and

$$
F_\delta(x) = 2\text{atanh}((1 - 2\delta)\tanh(\frac{x}{2})).
$$

- Variations:
  - $d$ itself could be random (most often: $d \sim \text{Poi}(\bar{d})$)
  - we can have “side information” or “survey” $S_0 \sim \mu_0$:

$$
R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0
$$
BP-operator: $Q\mu = \text{the law of}$

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_{\delta}(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$Q\mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point
BP-operator: \( Q\mu = \) the law of

\[
R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0
\]

Main Question: Characterize fixed points

\( Q\mu = \mu \)

Note: When \( S_0 = 0 \) choice \( \mu = \delta_0 \) is always a (trivial) fixed point

Theorem (Main result)

There exists at most one non-trivial fixed point \( \mu^* \) and \( Q^k\mu \to \mu^* \) as \( k \to \infty \) for any \( \mu \neq \delta_0 \).
BP-operator: $Q \mu = \text{the law of}$

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_{\delta}(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$Q \mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

Theorem (Main result)

There exists at most one non-trivial fixed point $\mu^*$ and $Q^k \mu \to \mu^*$ as $k \to \infty$ for any $\mu \neq \delta_0$.

Resolves multiple conjectures:

- [Kanade-Mossel-Schramm’2014]: labeled 2-SBM
- [Mossel-Xu’2015]: optimality of local algorithms for 2-SBM
- [Mossel-Neeman-Sly’2016]*: independence to leaf noise in BOT
- [Abbe-Cornacchia-Gu-P.’2021]*: entropy characterization in 2-SBM
BP-operator: $Q\mu = \text{the law of}$

\[
R = \sum_{i=1}^{d}(-1)^{X_i}F_\delta(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0
\]

**Main Question:** Characterize fixed points

\[Q\mu = \mu\]

**Note:** When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

**Theorem (Main result)**

*There exists at most one non-trivial fixed point $\mu^*$ and $Q^k\mu \to \mu^*$ as $k \to \infty$ for any $\mu \neq \delta_0$.***

Resolves multiple conjectures:

- [Kanade-Mossel-Schramm'2014]: labeled 2-SBM
- [Mossel-Xu’2015]: optimality of local algorithms for 2-SBM
- [Mossel-Neeman-Sly'2016]*: independence to leaf noise in BOT
- [Abbe-Cornacchia-Gu-P.'2021]*: entropy characterization in 2-SBM

The last two contain partial result: uniqueness for $(1 - 2\delta)^2d \gg 1$
BP-operator: \( Q \mu = \) the law of

\[
R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, \quad X_i \overset{iid}{\sim} \text{Bern}(\delta), \quad S_0 \sim \mu_0
\]

Main Question: Characterize fixed points

\( Q \mu = \mu \)

Note: When \( S_0 = 0 \) choice \( \mu = \delta_0 \) is always a (trivial) fixed point

Theorem (Main result)

*There exists at most one non-trivial fixed point \( \mu^* \) and \( Q^k \mu \to \mu^* \) as \( k \to \infty \) for any \( \mu \neq \delta_0 \).*

Resolves multiple conjectures:
- [Kanade-Mossel-Schramm'2014]: labeled 2-SBM
- [Mossel-Xu'2015]: optimality of local algorithms for 2-SBM
- [Mossel-Neeman-Sly'2016]*: independence to leaf noise in BOT
- [Abbe-Cornacchia-Gu-P.'2021]*: entropy characterization in 2-SBM

The last two contain partial result: uniqueness for \((1 - 2\delta)^2 d \gg 1\)

This talk: What is this relevant for?
- How information theory helped us prove it?
Three application domains

- Application 1: Statistical Physics
- Application 2: Machine Learning
- Application 3: Information propagation
Model of correlated phenomena: **Ising model** on a graph $G = (V, E)$

$$\mathbb{P}[X = x] = \frac{1}{Z} e^{\beta H(x)}$$

with Hamiltonian

$$H(x) = \sum_{u \sim v} x_u x_v$$

and $x_u \in \{\pm 1\}$ – binary spin variables.

Many graphs are “complex” but “sparse”. So model them as a random $d$-regular graph.
Application 1: Ising Model on Sparse Graphs

- Model of correlated phenomena: **Ising model** on a graph $G = (V, E)$
  
  $\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}$

  with Hamiltonian
  
  $H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$

  and $x_u \in \{\pm 1\} \text{ -- binary spin variables.}$

- Many graphs are “complex” but “sparse”. So model them as a random $d$-regular graph.

- Locally such graphs are just trees.
Application 1: Ising Model on Sparse Graphs

- Model of correlated phenomena: **Ising model** on a graph \( G = (V, E) \)

\[
\mathbb{P}[X = x] = \frac{1}{Z} e^{\beta H(x)}
\]

with Hamiltonian

\[
H(x) = \sum_{u \sim v} x_u x_v
\]

and \( x_u \in \{\pm 1\} \) – binary spin variables.
- Many graphs are “complex” but “sparse”. So model them as a random \( d \)-regular graph.
- Locally such graphs are just trees.
- So let us understand Ising model on trees.
Application 1: Ising Model on trees

Model of correlated phenomena: Ising model on a graph \( G = (V, E) \)

\[
P[X = x] = \frac{1}{Z} e^{\beta H(x)}, \quad H(x) = \sum_{u \sim v} x_u x_v
\]

and \( x_u \in \{\pm 1\} \) – binary spin variables.

Problem: How to extend the above to infinite trees?
Application 1: Ising Model on trees

Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$P[X = x] = \frac{1}{Z} e^{\beta H(x)}, \quad H(x) = \sum_{u \sim v} x_u x_v$$

and $x_u \in \{\pm 1\}$ – binary spin variables.

Problem: How to extend the above to infinite trees?

How to understand Gibbs measure on infinite tree?
Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$P[X = x] = \frac{1}{Z} e^{\beta H(x)}, \quad H(x) = \sum_{u \sim v} x_u x_v$$

and $x_u \in \{\pm 1\}$ – binary spin variables.

Problem: How to extend the above to infinite trees?

How to understand Gibbs measure on infinite tree?

Similar issue on finite locally-tree-like graphs with correlated boundary.
Application 1: Ising Model on infinite trees

For every finite subtree $T$ with boundary $L = \partial T$ we have:

$$\mathbb{P}[X_T = x_T|X_L = y_L] = \frac{1}{Z} e^{\beta H_T(x_T)} 1\{x_L = y_L\}$$

where $H_T$ is the restriction of Hamiltonian to a subtree $T$:

$$H_T(x_T) = \sum_{u \sim v: u \in T, v \in T} x_u x_v$$

As usual there is a phase transition:

- high temp $\beta \leq \frac{1}{\tanh(d)}$: there is a unique Gibbs measure on infinite tree. The choice of boundary condition $y_L$ is irrelevant.
- medium temp $\beta > \frac{1}{\tanh(d)}$: depending on $y_L$ we can get (uncountably many) Gibbs measures on infinite tree.

Problem: How to classify these Gibbs measures?
Application 1: Ising Model on infinite trees

For every finite subtree $T$ with boundary $L = \partial T$ we have:

$$\mathbb{P}[\mathbf{X}_T = \mathbf{x}_T | \mathbf{X}_L = \mathbf{y}_L] = \frac{1}{Z} e^{\beta H_T(\mathbf{x}_T)} 1\{\mathbf{x}_L = \mathbf{y}_L\}$$

where $H_T$ is the restriction of Hamiltonian to a subtree $T$:

$$H_T(\mathbf{x}_T) = \sum_{u \sim v : u \in T, v \in T} x_u x_v$$

As usual there is a phase transition:

- high temp $\beta \leq \frac{1}{\text{atanh}(d)}$: there is a unique Gibbs measure on infinite tree. The choice of boundary condition $\mathbf{y}_L$ is irrelevant.
- medium temp $\beta > \frac{1}{\text{atanh}(d)}$: depending on $\mathbf{y}_L$ we can get (uncountably many) Gibbs measures on infinite tree.
Application 1: Ising Model on infinite trees

For every finite subtree $T$ with boundary $L = \partial T$ we have:

$$
\mathbb{P}[X_T = x_T | X_L = y_L] = \frac{1}{Z} e^{\beta H_T(x_T)} 1\{x_L = y_L\}
$$

where $H_T$ is the restriction of Hamiltonian to a subtree $T$.

**Problem:** How to classify these Gibbs measures?

- **High temp** $\beta \leq \frac{1}{\text{atanh}(d)}$: there is a unique Gibbs measure on infinite tree. The choice of boundary condition $y_L$ is irrelevant.
- **Medium temp** $\beta > \frac{1}{\text{atanh}(d)}$: depending on $y_L$ we can get (uncountably many) Gibbs measures on infinite tree.
Application 1: free boundary Gibbs measure

- For every finite subtree $T$ we define

$$\mathbb{P}_T[X_T = x_T] = \frac{1}{Z} e^{\beta H_T(x_T)}$$

As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ – the free boundary Gibbs measure.

For low temperature $\beta > 1 \tanh(\sqrt{d})$, the measure $\mathbb{P}_\infty$ is not extremal and decomposes as

$$\mathbb{P}_\infty = \sum \alpha \omega_\alpha \mathbb{P}_\alpha$$

**Question:** How to “count” $\mathbb{P}_\alpha$’s?
For every finite subtree $T$ we define

$$\mathbb{P}_T[X_T = x_T] = \frac{1}{Z} e^{\beta H_T(x_T)}$$

As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ – the free boundary Gibbs measure
Application 1: free boundary Gibbs measure

For every finite subtree $T$ we define

$$\mathbb{P}_T[X_T = x_T] = \frac{1}{Z} e^{\beta H_T(x_T)}$$

As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ – the free boundary Gibbs measure

For low temp $\beta > \frac{1}{\text{atanh}(\sqrt{d})}$ measure $\mathbb{P}_\infty$ is not extremal and decomposes as

$$\mathbb{P}_\infty = \sum_{\alpha} w_\alpha \mathbb{P}_\alpha$$
Application 1: free boundary Gibbs measure

For every finite subtree $T$ we define

$$ \mathbb{P}_T[X_T = x_T] = \frac{1}{Z} e^{\beta H_T(x_T)} $$

As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ — the free boundary Gibbs measure

For low temp $\beta > \frac{1}{\text{atanh}(\sqrt{d})}$ measure $\mathbb{P}_\infty$ is not extremal and decomposes as

$$ \mathbb{P}_\infty = \sum_{\alpha} w_\alpha \mathbb{P}_\alpha $$

Question: How to “count” $\mathbb{P}_\alpha$’s?
To understand $P_\infty$ we can sample boundary condition $y_L$ from the distribution $P_T$.
To understand $P_\infty$ we can sample boundary condition $y_L$ from the distribution $P_T$.

... and then compute the finite distribution $P[X_T|X_L = y_L]$. Note: it is a random distribution.

Mezard and Parisi: this random distribution is described by a fixed point $Q\mu = \mu$. 

Mezard-Parisi: Replica-Symmetry Breaking (1RSB)
To understand $P_\infty$ we can sample boundary condition $y_L$ from the distribution $P_T$.

... and then compute the finite distribution $P[X_T|X_L = y_L]$. Note: it is a random distribution.

Mezard and Parisi: this random distribution is described by a fixed point $Q\mu = \mu$.

Our contribution. We prove this ansatz: Iterations $Q \circ Q \cdot \cdot Q\mu_0$ converge to a unique fixed point regardless of $\mu_0$. 
Application 2: Community detection

- Unsupervised clustering problem
- Input: graph
- Want: Label clusters
Application 2: stochastic block model

- A Model for community detection: symmetric $k$-SBM($a, b$), $a, b > 0$
Application 2: stochastic block model

- A Model for community detection: \textit{symmetric $k$-SBM($a, b$), $a, b > 0$}
- $n$ vertices, each assigned a uniformly random color $X_v \sim \text{Unif}[k]$. 
A Model for community detection: symmetric $k$-SBM$(a, b)$, $a, b > 0$
$n$ vertices, each assigned a uniformly random color $X_v \sim \text{Unif}[k]$.
A random graph $G$ with independently selected edges

$$\Pr[(u, v) \in E(G)] = \begin{cases} 
\frac{a}{n}, & \text{if } X_u = X_v \\
\frac{b}{n}, & \text{o/w}
\end{cases}$$
Application 2: stochastic block model

- A Model for community detection: symmetric $k$-SBM$(a, b)$, $a, b > 0$
- $n$ vertices, each assigned a uniformly random color $X_v \sim \text{Unif}[k]$.
- A random graph $G$ with independently selected edges

$$
\mathbb{P}[(u, v) \in E(G)] = \begin{cases} 
\frac{a}{n}, & \text{if } X_u = X_v \\
\frac{b}{n}, & \text{o/w}
\end{cases}
$$

- **Goal 1:** Design algorithms that recover $X$ from $G$.

- **Goal 2:** Characterize fundamental uncertainty $H(X|G)$. 
Application 2: stochastic block model

- A Model for community detection: symmetric $k$-SBM($a, b$), $a, b > 0$
- $n$ vertices, each assigned a uniformly random color $X_v \sim \text{Unif}[k]$.
- A random graph $G$ with independently selected edges

\[
\mathbb{P}[(u, v) \in E(G)] = \begin{cases} 
\frac{a}{n}, & \text{if } X_u = X_v \\
\frac{b}{n}, & \text{o/w}
\end{cases}
\]

- **Goal 1:** Design algorithms that recover $X$ from $G$.
  *Our work:* for $k = 2$ poly-time algo recovering optimal fraction of $X$
- **Goal 2:** Characterize fundamental uncertainty $H(X|G)$.
  *Our work:* closes this as well.
Application 2: stochastic block model

- A Model for community detection: symmetric $k$-SBM($a, b$), $a, b > 0$
- $n$ vertices, each assigned a uniformly random color $X_v \sim \text{Unif}[k]$.
- A random graph $G$ with independently selected edges

$$
P[(u, v) \in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } X_u = X_v \\ \frac{b}{n}, & \text{o/w} \end{cases}
$$

- **Goal 1**: Design algorithms that recover $X$ from $G$.
  **Our work**: for $k = 2$ poly-time algo recovering optimal fraction of $X$
- **Goal 2**: Characterize fundamental uncertainty $H(X|G)$.
  **Our work**: closes this as well.
- **Key fact**: locally the recovery problem looks like BOT (next):

$$
d \triangleq \text{deg}(u) \sim \text{Poi}(a + b), \quad \delta \triangleq P[X_u \neq X_v | u \sim v] = \frac{b}{a + b}
$$
Fix infinite oriented tree $T$ with branching number $\text{br}(T) := d$. 

- Level 0: $X_{0,0}$, $L_0 = 1$
- Level 1: $X_{1,0}$, $X_{1,1}$, $L_1 = 2$
- Level 2: $X_{2,0}$, $X_{2,1}$, $X_{2,2}$, $X_{2,3}$, $L_2 = 2^2$
- ... 
- Level $k$: $X_{k,0}$, $X_{k,1}$, $\ldots$, $X_{k,L_k-2}$, $X_{k,L_k-1}$, $L_k = \text{br}(T)^k$
Application 3: Broadcasting on Trees

- Fix infinite oriented tree $T$ with branching number $\text{br}(T) := d$. 

![Diagram of a tree with levels and branches]

- Level 0: $X_{0,0}$ with $L_0 = 1$
- Level 1: $X_{1,0}$ and $X_{1,1}$ with $L_1 = 2$ and $\text{br}(T) = 2$
- Level 2: Further branching with $L_2 = 2^2$
- Level $k$: $X_{k,0}, X_{k,1}, \ldots, X_{k,L_k-1}$ with $L_k = \text{br}(T)^k$
Application 3: Broadcasting on Trees

- Fix infinite oriented tree $T$ with branching number $\text{br}(T) := d$.
- Root $X_{0,0} \sim \text{Unif}\{\pm 1\}$

```
level 0
  \node (X0) at (0,0) {$X_{0,0}$};
  \node (L0) at (2,0) {$L_0 = 1$};

level 1
  \node (X1) at (-1,-1) {$X_{1,0}$};
  \node (X2) at (1,-1) {$X_{1,1}$};
  \node (L1) at (2,-1) {$L_1 = 2$};
  \node (brT) at (2,-2) {$\text{br}(T) = 2$};

level 2
  \node (X3) at (-2,-2) {$X_{2,0}$};
  \node (X4) at (-1,-2) {$X_{2,1}$};
  \node (X5) at (0,-2) {$X_{2,2}$};
  \node (X6) at (1,-2) {$X_{2,3}$};
  \node (L2) at (2,-2) {$L_2 = 2^2$};

level k
  \node (Xk) at (-k,-k) {$X_{k,0}$};
  \node (Xk-1) at (-k-1,-k-1) {$X_{k,1}$};
  \node (Xk-Lk) at (-k-2,-k-2) {$X_{k,L_k-2}$};
  \node (Xk-Lk-1) at (-k-3,-k-3) {$X_{k,L_k-1}$};
  \node (Lk) at (2,-k) {$L_k = \text{br}(T)^k$};
```
Application 3: Broadcasting on Trees

- Fix infinite oriented tree $T$ with branching number $\text{br}(T) := d$.
- Root $X_{0,0} \sim \text{Unif}\{\pm 1\}$
- For any edge $u \rightarrow v$ set $X_v = \begin{cases} X_u, & \text{w.p. } 1 - \delta \\ -X_u, & \text{w.p. } \delta \end{cases}$

Goal: Reconstruct $X_0, 0$ from $X_k = (X_k, 0, \ldots, X_{k, d_k - 1})$.

Diagram:
- Level 0: $X_{0,0}$ with $L_0 = 1$
- Level 1: $X_{1,0}, X_{1,1}$ with $L_1 = 2$
- Level 2: $X_{2,0}, X_{2,1}, X_{2,2}, X_{2,3}$ with $L_2 = 2^2$
- General: $X_{k,0}, X_{k,1}, \ldots, X_{k,\text{br}(T)^k}$ with $L_k = \text{br}(T)^k$
Application 3: Broadcasting on Trees

- Fix infinite oriented tree $T$ with branching number $\text{br}(T) := d$.
- Root $X_{0,0} \sim \text{Unif}\{\pm 1\}$
- For any edge $u \rightarrow v$ set $X_v = \begin{cases} X_u, & \text{w.p. } 1 - \delta \\ -X_u, & \text{w.p. } \delta \end{cases}$
- Goal: Reconstruct $X_{0,0}$ from $X_k = (X_{k,0}, \ldots, X_{k,d^k-1})$.

\begin{itemize}
  \item \text{level 0: } X_{0,0} \quad L_0 = 1
  \item \text{level 1: } X_{1,0}, X_{1,1} \quad L_1 = 2 \quad \text{br}(T) = 2
  \item \text{level 2: } X_{2,0}, X_{2,1}, X_{2,2}, X_{2,3} \quad L_2 = 2^2
  \item \quad \vdots
  \item \text{level } k: \quad X_{k,0}, \ldots, X_{k,L_k-1} \quad L_k = \text{br}(T)^k
\end{itemize}
Root variable $X_{0,0}$ is the information source.

It spreads along a tree of binary symmetric channels, $BSC_{\delta}$.

**Question:** How to estimate $X_{0,0}$ from a vector of far-away leaves $X_k$?
Root variable $X_{0,0}$ is the information source
It spreads along a tree of binary symmetric channels, $BSC_\delta$.
**Question:** How to estimate $X_{0,0}$ from a vector of far-away leaves $X_k$?
A standard algorithm (belief propagation or BP) computes

$$R^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | X_k]}{\mathbb{P}[X_{0,0} = -1 | X_k]}$$
Broadcasting on Trees and BP fixed point

- Root variable $X_{0,0}$ is the information source
- It spreads along a tree of binary symmetric channels, $BSC_\delta$.
- **Question:** How to estimate $X_{0,0}$ from a vector of far-away leaves $X_k$?
- A standard algorithm (belief propagation or BP) computes

$$R^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | X_k]}{\mathbb{P}[X_{0,0} = -1 | X_k]}$$

- Let $\mu_k$ denote the distribution of $R^{(k)}$ conditioned on $X_k = +1$. Then:

$$R^{(k+1)} = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i), \quad R_i \overset{iid}{\sim} \mu$$

Thus: $R^{(k+1)} \sim \mu_{k+1} \triangleq Q\mu_k$. 

Yury Polyanskiy

Uniqueness of BP fixed point for Ising models
Root variable $X_{0,0}$ is the information source.

It spreads along a tree of binary symmetric channels, $BSC_\delta$.

**Question:** How to estimate $X_{0,0}$ from a vector of far-away leaves $X_k$?

A standard algorithm (belief propagation or BP) computes

$$R^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | X_k]}{\mathbb{P}[X_{0,0} = -1 | X_k]}$$

Let $\mu_k$ denote the distribution of $R^{(k)}$ conditioned on $X_k = +1$. Then:

$$R^{(k+1)} = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i), \quad R_i \overset{iid}{\sim} \mu$$

Thus: $R^{(k+1)} \sim \mu_{k+1} \overset{d}{=} Q\mu_k$.

Can show that $R^{(k)} \xrightarrow{d} \mu^* = \text{BP fixed point.}$
How to estimate $X_{0,0}$ from the \textbf{noisy} observation $Y_k$ of the leaves $X_k$?

\[
\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} 
X_{k,i}, & \text{w.p. } 1 - \tau \\
-X_{k,i}, & \text{w.p. } \tau
\end{cases}
\]
How to estimate $X_{0,0}$ from the noisy observation $Y_k$ of the leaves $X_k$?

$\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1 - \tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$

As before:

$$\tilde{R}^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1|Y_k]}{\mathbb{P}[X_{0,0} = -1|Y_k]}$$

$$\tilde{R}^{(k)} \sim \tilde{\mu}_k \triangleq Q^k \tilde{\mu}_0$$
How to estimate $X_{0,0}$ from the noisy observation $Y_k$ of the leaves $X_k$?

$$\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1 - \tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$$

As before:

$$\tilde{R}^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1|Y_k]}{\mathbb{P}[X_{0,0} = -1|Y_k]}$$

$$\tilde{R}^{(k)} \sim \tilde{\mu}_k \triangleq Q^k \tilde{\mu}_0$$

$$\tilde{\mu}_0 = (1 - \tau)\delta_{-c} + \tau\delta_c, \quad c = \log \frac{\tau}{1 - \tau}.$$
How to estimate $X_{0,0}$ from the noisy observation $Y_k$ of the leaves $X_k$?

$$\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1 - \tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$$

As before:

$$\tilde{R}^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1|Y_k]}{\mathbb{P}[X_{0,0} = -1|Y_k]}$$

$$\tilde{R}^{(k)} \sim \tilde{\mu}_k \triangleq Q^k \tilde{\mu}_0$$

$$\tilde{\mu}_0 = (1 - \tau)\delta_{-c} + \tau \delta_c, \quad c = \log \frac{\tau}{1 - \tau}.$$  

*Our main result:* $Q^k \tilde{\mu}_0 \to \tilde{\mu}_{\infty}$ and $\tilde{\mu}_{\infty}$ does not depend on $\tau$ (!!!)
Proof ideas
Binary symmetric (BMS) channels

Definition (BMS channel)

\[ P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y} \text{ called BMS if there is a bijection } h : \mathcal{Y} \rightarrow \mathcal{Y} \text{ s.t.} \]

\[ P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y \]

- Example 1: BSC\(_{\delta}\) channel is \( Y = (-1)^{\text{Bern}(\delta)}X \)
Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y$$

- Example 1: BSC$_{\delta}$ channel is $Y = (-1)^{\text{Bern}(\delta)}X$
- Example 2: the tree channel $W_k : X_0 \to X_k$
**Binary symmetric (BMS) channels**

**Definition (BMS channel)**

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called **BMS** if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y$$

- Example 1: BSC$_{\delta}$ channel is $Y = (-1)^{\text{Bern}(\delta)}X$
- Example 2: the tree channel $W_k : X_0 \to X_k$
Binary symmetric (BMS) channels

**Definition (BMS channel)**

$P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \rightarrow \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y$$

- Example 1: BSC$_\delta$ channel is $Y = (-1)^{\text{Bern}(\delta)} X$
- Example 2: the tree channel $W_k : X_0 \rightarrow X_k$
- **Blackwell preorder:** We say that $P_{Y|X} \preceq P_{Z|X}$ if there is $P_{Y|Z}$ s.t.

$$P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x)P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y$$
Definition (BMS channel)

\[ P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y} \text{ called BMS if there is a bijection } h : \mathcal{Y} \rightarrow \mathcal{Y} \text{ s.t.} \]

\[ P_{Y|X}(y|x) = P_{Y|X}(h(y)|x) \quad \forall x, y \]

- Example 1: BSC_\delta channel is \( Y = (-1)^{\text{Bern}(\delta)} X \)
- Example 2: the tree channel \( W_k : X_0 \rightarrow X_k \)
- Blackwell preorder: We say that \( P_{Y|X} \preceq P_{Z|X} \) if there is \( P_{Y|Z} \) s.t.

\[ P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x)P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y \]

For example: \( W_{k+1} \preceq W_k \)
Definition (BMS channel)

\( P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y} \) called BMS if there is a bijection \( h : \mathcal{Y} \rightarrow \mathcal{Y} \) s.t.

\[
P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y
\]

- **Question:** How can we compactly represent a BMS?

\[\mu \triangleq \text{Law of } R \text{ under } X = +1.\]

**Tech note:** \( \mu(-dr) = e^{-r}\mu(dr) \) and all \( \mu \)'s in this work are such!
Definition (BMS channel)

\[ P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y} \text{ called BMS if there is a bijection } h : \mathcal{Y} \rightarrow \mathcal{Y} \text{ s.t.} \]

\[ P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y \]

- **Question**: How can we compactly represent a BMS?
- **Sufficient statistic**: Log-likelihood ratio

\[ R = \log \frac{\mathbb{P}[Y|X = +1]}{\mathbb{P}[Y|X = -1]} \]
**Definition (BMS channel)**

$P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y}$ called **BMS** if there is a bijection $h : \mathcal{Y} \rightarrow \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|−x) \quad \forall x, y$$

- **Question:** How can we compactly represent a BMS?
- **Sufficient statistic:** Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X = +1]}{\mathbb{P}[Y|X = -1]}$$

⇒ only need distribution of $R$ under $X = +1$ and under $X = -1$. 
**BMS ↔ µ equivalence**

**Definition (BMS channel)**

\[ P_{Y|X} : \{\pm 1\} \to \mathcal{Y} \] called **BMS** if there is a bijection \( h : \mathcal{Y} \to \mathcal{Y} \) s.t.

\[ P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y \]

- **Question:** How can we compactly represent a BMS?
- **Sufficient statistic:** Log-likelihood ratio

\[ R = \log \frac{\mathbb{P}[Y|X = +1]}{\mathbb{P}[Y|X = -1]} \]

⇒ only need distribution of \( R \) under \( X = +1 \) and under \( X = -1 \).

... But due to symmetry only one is needed!

\[ \mu \triangleq \text{Law of } R \text{ under } X = +1. \]
BMS ↔ $\mu$ equivalence

**Definition (BMS channel)**

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called **BMS** if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y$$

- **Question:** How can we compactly represent a BMS?
- **Sufficient statistic:** Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X = +1]}{\mathbb{P}[Y|X = -1]}$$

$\Rightarrow$ only need distribution of $R$ under $X = +1$ and under $X = -1$.

... But due to symmetry only one is needed!

$$\mu \triangleq \text{Law of } R \text{ under } X = +1.$$  

**Tech note:** $\mu(-dr) = e^{-r}\mu(dr)$ and all $\mu$’s in this work are such!
Definition (BMS channel)

\[ P_{Y|X} : \{\pm 1\} \rightarrow \mathcal{Y} \] called BMS if there is a bijection \( h : \mathcal{Y} \rightarrow \mathcal{Y} \) s.t.

\[ P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \quad \forall x, y \]

- **Question:** How can we compactly represent a BMS?
- **Sufficient statistic:** Log-likelihood ratio

\[ R = \log \frac{P[Y|X = +1]}{P[Y|X = -1]} \]

\( \Rightarrow \) only need distribution of \( R \) under \( X = +1 \) and under \( X = -1 \).

... But due to symmetry only one is needed!

\[ \mu \triangleq \text{Law of } R \text{ under } X = +1. \]

**Tech note:** \( \mu(-dr) = e^{-r}\mu(dr) \) and all \( \mu \)'s in this work are such!

- Denote a BMS corresp. to \( \mu \) as \( \mu \).
Tree channels recursion

In BOT we can build $W_{k+1}$ channel from $W_k$ and $\text{BSC}_\delta$: 
In BOT we can build $W_{k+1}$ channel from $W_k$ and $\text{BSC}_\delta$: 

\[
\begin{align*}
X_0 &\xrightarrow{\text{BSC}_\delta} X_{1,1} \\
&\quad \vdots \\
&\quad \text{BSC}_\delta \\
X_{1,d} &\xrightarrow{\text{BSC}_\delta} X_{1,1} \\
&\quad \vdots \\
&\quad W_k \\
&\quad \vdots \\
&\quad W_k
\end{align*}
\]
Recall: BP-operator $Q_\mu = \text{the law of}$

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$
Recall: **BP-operator** $Q_\mu = \text{the law of}$

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0, \quad R_i \sim \mu, \; X_i \sim \text{Bern}(\delta), \; S_0 \sim \mu_0$$

We can understand the BMS $Q_\mu$ graphically as:
Recall: BP-operator $Q_{\mu} = \text{the law of}$

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0,$$

$R_i \sim \mu, X_i \sim \text{Bern}(\delta), S_0 \sim \mu_0$

We can understand the BMS $Q_{\mu}$ graphically as:
Recall: BP-operator $Q_\mu = \text{the law of}$

\[ R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0, \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0 \]

We can understand the BMS $Q_\mu$ graphically as:

Infinite divisibility: a fixed point channel has property of not changing upon $d$-fold copying of its $\delta$-noisy version.
With these preparations, we can appreciate the following result [EKPS’2000].
With these preparations, we can appreciate the following result \[\text{EKPS'2000}\].

**Theorem (Stringy tree lemma (Evans-Kenyon-Peres-Schulman'2000))**

For all $\mu$ and $\nu$ we have:

\[
\text{BSC}_{\delta} \rightarrow \mu \rightarrow \nu \quad \text{and} \quad \text{BSC}_{\delta} \rightarrow \mu \rightarrow \nu
\]
With these preparations, we can appreciate the following result [EKPS'2000].

**Theorem (Stringy tree lemma (Evans-Kenyon-Peres-Schulman’2000))**

For all $\mu$ and $\nu$ we have:

Copying input bit before adding noise results in a better channel.
With these preparations, we can appreciate the following result [EKPS'2000].

**Theorem (Stringy tree lemma (Evans-Kenyon-Peres-Schulman’2000))**

For all $\mu$ and $\nu$ we have:

- Copying input bit before adding noise results in a better channel.
- Applying iteratively $1 - 2\delta_k = (1 - 2\delta)^k$ [EKPS'2000] bound $W_k$ and showed $Q^k\delta_\infty \rightarrow \delta_0$ if $(1 - 2\delta)^2d < 1$. 
Theorem (Main technical discovery of Yu-P.’2022)

For every $\nu, \delta \in (0, 1)$ and $d \geq 3$ there exists $\epsilon > 0$ such that:

For $d = 2$ this is wrong, but holds for a depth-2 (binary) tree.

Let us see how this result implies unicity of BP-fixed point.
Theorem (Main technical discovery of Yu-P.’2022)

For every $\nu, \delta \in (0, 1)$ and $d \geq 3$ there exists $\epsilon > 0$ such that:

For $d = 2$ this is wrong, but holds for a depth-2 (binary) tree.
For every $\nu, \delta \in (0, 1)$ and $d \geq 3$ there exists $\epsilon > 0$ such that:

For $d = 2$ this is wrong, but holds for a depth-2 (binary) tree. Let us see how this result implies unicity of BP-fixed point.
Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi)\mu(dr) + \phi\mu(-dr)$. Note: this is the law of channel $(\mu \circ BSC_\phi)$. 

Yury Polyanskiy  
Uniqueness of BP fixed point for Ising models
Proof of the main result

- Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi) \mu(dr) + \phi \mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.

- Define degradation index

  $$\phi^*(\mu, \nu) = \inf \{ \phi : \mathcal{B}_\phi \mu \preceq \nu \}.$$ 

  = least amount of noise needed to make $\mu$ worse than $\nu$. 


Proof of the main result I

- Define \( \mathcal{B}_\phi \mu(dr) = (1 - \phi)\mu(dr) + \phi\mu(-dr) \).
  Note: this is the law of channel \( (\mu \circ \text{BSC}_\phi) \).

- Define degradation index

\[
\phi^*(\mu, \nu) = \inf \{ \phi : \mathcal{B}_\phi \mu \preceq \nu \}.
\]

= least amount of noise needed to make \( \mu \) worse than \( \nu \).

- If \( \phi^* > 0 \) then (we will show next) our improved STL implies for some \( \epsilon > 0 \)

\[
\mathcal{B}_{\phi^*} Q \mu \preceq \mathcal{B}_\epsilon Q \nu.
\]
Proof of the main result I

- Define \( \mathcal{B}_\phi \mu(dr) = (1 - \phi) \mu(dr) + \phi \mu(-dr) \).
  Note: this is the law of channel \((\mu \circ \text{BSC}_\phi)\).

- Define degradation index
  \[
  \phi^*(\mu, \nu) = \inf \{ \phi : \mathcal{B}_\phi \mu \preceq \nu \}.
  \]
  = least amount of noise needed to make \( \mu \) worse than \( \nu \).

- If \( \phi^* > 0 \) then (we will show next) our improved STL implies for some \( \epsilon > 0 \)
  \[
  \mathcal{B}_{\phi^*} Q \mu \preceq \mathcal{B}_\epsilon Q \nu.
  \]

- So if \( \mu \) and \( \nu \) are two fixed points, then
  \[
  \mathcal{B}_{\phi^*} \mu \preceq \mathcal{B}_\epsilon \nu,
  \]
  violating minimality of \( \phi^* \).

  Here we also need \( \nu \) – non-trivial, so that \( \phi^* < 1/2 \)
Proof of the main result

- Define $\mathcal{B}_\phi \mu (dr) = (1 - \phi)\mu (dr) + \phi \mu (-dr)$.
  Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.

- Define degradation index

  \[ \phi^* (\mu, \nu) = \inf \{ \phi : \mathcal{B}_\phi \mu \leq \nu \} . \]

  = least amount of noise needed to make $\mu$ worse than $\nu$.

- If $\phi^* > 0$ then (we will show next) our improved STL implies for some $\epsilon > 0$

  \[ \mathcal{B}_{\phi^*} Q \mu \leq \mathcal{B}_\epsilon Q \nu . \]

- So if $\mu$ and $\nu$ are two fixed points, then

  \[ \mathcal{B}_{\phi^*} \mu \leq \mathcal{B}_\epsilon \nu , \]

  violating minimality of $\phi^*$

  Here we also need $\nu$ – non-trivial, so that $\phi^* < 1/2$

- So $\phi^* = 0$, which implies $\mu \leq \nu$. From symmetry $\nu = \mu$ then.
Proof of the main result II

- Define $B_{\phi}\mu(dr) = (1 - \phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.

$\text{Yury Polyanskiy}$

$\text{Uniqueness of BP fixed point for Ising models}$
Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi) \mu(dr) + \phi \mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$
  $$\mathcal{B}_\phi Q\mu \preceq \mathcal{B}_\epsilon Q\nu.$$
Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi) \mu(dr) + \phi \mu(-dr)$. Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$ 
  \[ \mathcal{B}_\phi \mathcal{Q}_\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}_\nu. \]
Define $\mathcal{B}_\phi \mu (dr) = (1 - \phi) \mu (dr) + \phi \mu (-dr)$.

Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.

Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

\[ \mathcal{B}_\phi Q \mu \preceq \mathcal{B}_\epsilon Q \nu. \]
Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi) \mu(dr) + \phi \mu(-dr)$.  
  Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.  
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0 \quad \mathcal{B}_\phi Q \mu \preceq \mathcal{B}_\epsilon Q \nu$.  

\[
\mathcal{B}_\phi Q \mu \preceq \mathcal{B}_\phi Q \mu \preceq \mathcal{B}_\epsilon Q \nu.
\]
Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$ \[ \mathcal{B}_\phi Q\mu \preceq \mathcal{B}_\epsilon Q\nu. \]
Define $\mathbb{B}_\phi \mu(dr) = (1 - \phi)\mu(dr) + \phi\mu(-dr)$. Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.

Remains to show: If $\mathbb{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$ $\mathbb{B}_\phi Q\mu \preceq \mathbb{B}_\epsilon Q\nu$. 
Proof of the main result II

- Define \( \mathcal{B}_\phi \mu(dr) = (1 - \phi) \mu(dr) + \phi \mu(-dr) \).
  Note: this is the law of channel \((\mu \circ \text{BSC}_\phi)\).
- Remains to show: If \( \mathcal{B}_\phi \mu \preceq \nu \) then \( \exists \epsilon > 0 \)
  \[ \mathcal{B}_\phi Q \mu \preceq \mathcal{B}_\epsilon Q \nu. \]
Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1 - \phi)\mu(dr) + \phi\mu(-dr)$. Note: this is the law of channel $(\mu \circ \text{BSC}_\phi)$.
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0 \hspace{1cm} \mathcal{B}_\phi Q\mu \preceq \mathcal{B}_\epsilon Q\nu$.
One word about the proof of improved STL
One word about the proof of improved STL

Blackwell preorder: \( P_{Y|X} \preceq P_{Z|X} \) if there is \( P_{Y|Z} \) s.t.

\[
P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x)P_{Y|Z}(y|z) \iff X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y
\]
One word about the proof of improved STL

Blackwell preorder: \( P_{Y|X} \preceq P_{Z|X} \) if there is \( P_{Y|Z} \) s.t.

\[
P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x) P_{Y|Z}(y|z) \iff X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y
\]

In [EKPS’2000] to prove STL authors constructed the coupling explicitly

Idea 1: Our trick was to avoid the (hard) job of constructing \( P_{Y|Z} \):

\[
\mu \preceq \nu \iff \beta(t; \mu) \leq \beta(t; \nu) \quad \forall t \geq 0,
\]

where \( \beta(t; \mu) = \mathbb{E}_{R \sim \mu} \left[ \tanh \frac{|R|}{2} \vee t \right] \)
One word about the proof of improved STL

**Blackwell preorder:** \( P_Y|X \preceq P_Z|X \) if there is \( P_Y|Z \) s.t.

\[
P_Y|X(y|x) = \sum_z P_Z|X(z|x)P_Y|Z(y|z) \iff X \xrightarrow{P_Y|X} Z \xrightarrow{P_Y|Z} Y
\]

In [EKPS’2000] to prove STL authors constructed the coupling explicitly

**Idea 1:** Our trick was to avoid the (hard) job of constructing \( P_Y|Z \):

\[
\mu \preceq \nu \iff \beta(t; \mu) \leq \beta(t; \nu) \quad \forall t \geq 0,
\]

where \( \beta(t; \mu) = \mathbb{E}_{R \sim \mu}[\tanh \frac{|R|}{2} \lor t] \)

**Idea 2:** Any \( \mu \) is a mixture of elementary ones:

\[
\mathcal{B}_\tau = (1 - \tau)\delta_{-c} + \tau\delta_c, \quad c = \log \frac{\tau}{1 - \tau}.
\]

So key inequalities are checked for \( \text{BSC}_\tau \)'s in place of generic \( \nu \).
Conclusion

\[ F_\delta(x) = \]

\[ Q_\mu \triangleq \text{Law of } \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i), \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta) \]

**Theorem (Main result)**

*There exists at most one non-trivial fixed point \( \mu^* \) of \( Q \).*

- Unusual method based on (new) properties of channel compositions.

---

Yury Polyanskiy

Uniqueness of BP fixed point for Ising models
Conclusion

$$F_\delta(x) =$$

$$Q\mu \triangleq \text{Law of } \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i), \quad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta)$$

Theorem (Main result)

There exists at most one non-trivial fixed point \( \mu^* \) of \( Q \).

- Unusual method based on (new) properties of channel compositions.
- We construct metric \( d(\cdot, \cdot) \) s.t. \( d(Q\mu, Q\nu) < d(\mu, \nu) \) unless both are trivial. This implies \( Q^k \mu_0 \to \mu^* \).
Conclusion

\[ F_\delta(x) = \]

\[ Q\mu \triangleq \text{Law of } \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i), \quad R_i \overset{iid}{\sim} \mu, \quad X_i \overset{iid}{\sim} \text{Bern}(\delta) \]

\textbf{Theorem (Main result)}

\textit{There exists at most one non-trivial fixed point } \mu^* \text{ of } Q.

- Unusual method based on (new) properties of channel compositions.
- We construct metric \(d(\cdot, \cdot)\) s.t. \(d(Q\mu, Q\nu) < d(\mu, \nu)\) unless both are trivial. This implies \(Q^k\mu_0 \rightarrow \mu^*\).
- \textbf{Future work:} Extending to Potts (q-ary) models.
Thank You!

The draft is available here:
https://www.mit.edu/~ypol