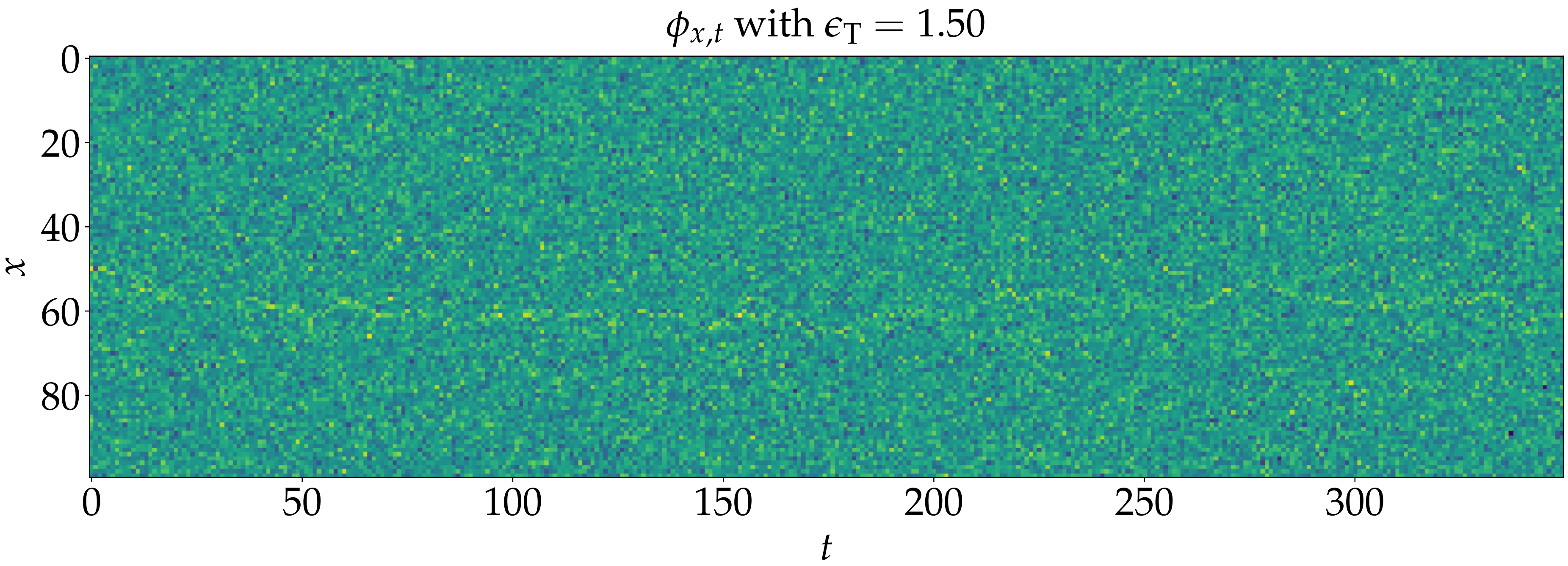
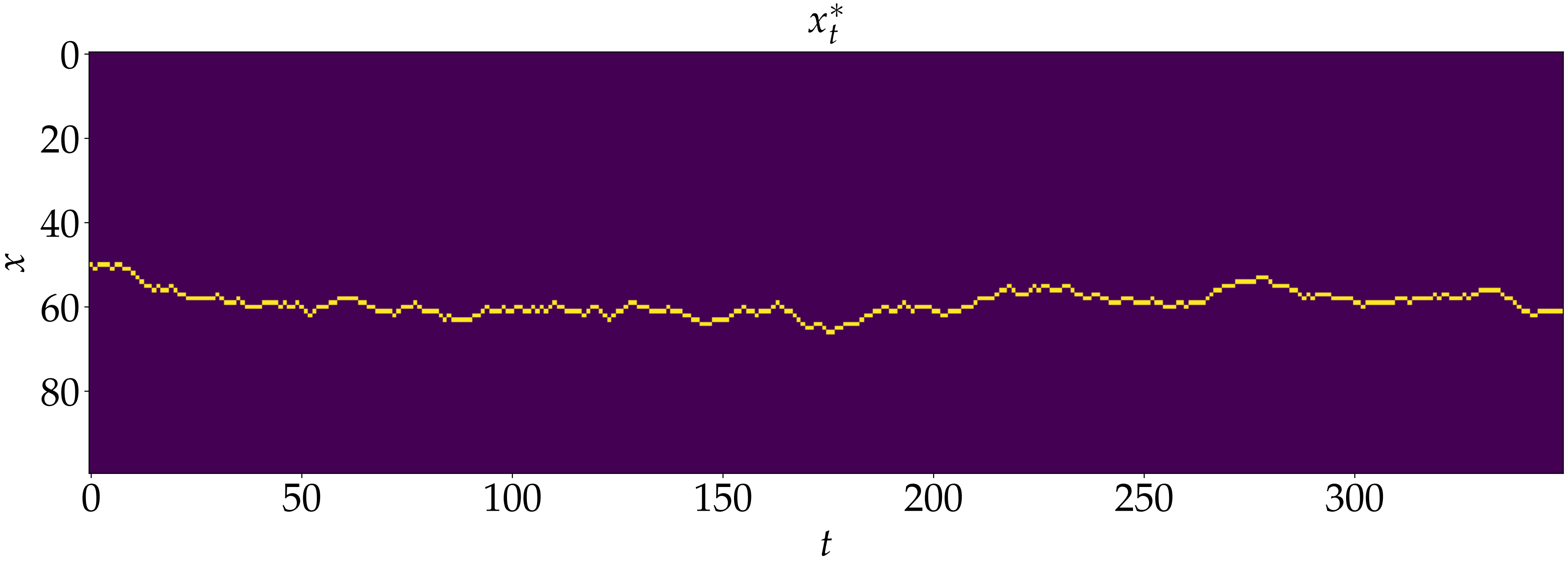
The planted directed polymer:
inferring a random walk from noisy images

Sun Woo P. Kim (King's College London)
Austen Lamacraft (University of Cambridge)
arXiv.2404.07263


$\phi_{x,t}$ with $\epsilon_T = 1.50$
0
20
40
60
80
x
0
50
100
150
200
250
300
t


x_t^*
0
20
40
60
80
x
0
50
100
150
200
250
300
t

Outline

1. Bayesian inference
2. Planted directed polymer
3. Numerical results in 1D
4. Analytic results on a tree
5. Outlook

Bayesian inference

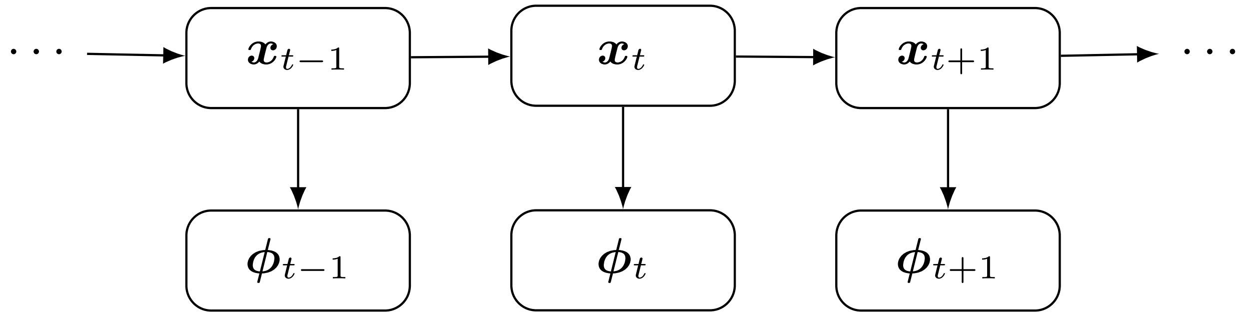
Infer posterior distribution for state X given data Φ

$$p(X|\Phi) = \frac{p(\Phi|X)p(X)}{p(\Phi)}$$

- $p(\Phi|X)$: "likelihood"/measurement model
- $p(X)$: "prior"
- $p(\Phi) = \sum_X p(\Phi|X)p(X)$: "evidence"/normalisation

In general X is high-dimensional

Hidden Markov models



- $X = \boldsymbol{x}_{1:t}$: entire trajectory
- $\Phi = \boldsymbol{\phi}_{1:t}$: measurements over all timesteps

Filtering: posterior for current state given history of measurements $p(\boldsymbol{x}_t \,|\, \boldsymbol{\phi}_{1:t})$

The planted directed polymer

Walker undergoes a random walk $p(x_t \mid x_{t-1}) = \frac{1}{2}\delta_{x_t, x_{t-1}} + \frac{1}{4}\delta_{x_t \pm 1, x_{t-1}}$

Given walker position x_t pixel of image at time t distributed as

$$\phi_{x,t} \sim \mathcal{N}(\epsilon \delta_{x,x_t}, \sigma_{\mathrm{S}}^2)$$

So $p(\boldsymbol{\phi}_t \mid x_t) = \prod_x \left(\frac{e^{-(\phi_{x,t} - \epsilon \delta_{x,x_t})^2 / 2\sigma^2}}{\sqrt{2\pi\sigma^2}} \right) = e^{-\epsilon/2\sigma^2} e^{\epsilon \phi_{x_t,t}/\sigma^2} \boldsymbol{\pi}(\boldsymbol{\phi}_t)$

$\pi(\boldsymbol{\phi}_t)$ is the Gaussian measure

Connection to the directed polymer

Assuming that the kernel only depends on its distance, $p(x_t \mid x_{t-1}) \sim e^{-\frac{1}{2\nu}(x_t - x_{t-1})^2}$

Therefore the posterior is $p(X \mid \Phi) = \frac{1}{Z} q(X \mid \Phi)$, where $Z = \sum_X q(X \mid \Phi)$ and the

unnormalised posterior for entire trajectory is

$$q(X \mid \Phi) = \exp\left[\sum_t \left(-\frac{1}{2\nu}(x_t - x_{t-1})^2 + \beta \phi_{x_t,t}\right)\right] \text{where } \beta = \frac{\epsilon}{\sigma^2}$$

$q(x_t \mid \boldsymbol{\phi}_{1:t}) = \sum_{x_{1:t-1}} q(X \mid \Phi)$ evolves linearly with transfer matrix $T_{x_t,x_{t-1}} = e^{-\frac{1}{2\nu}(x_t - x_{t-1})^2 + \beta\phi_{x_t,t}}$

The directed polymer

- The posterior looks formally identical to Boltzmann probability of the directed polymer
- Competition of elastic potential $\sim 1/\nu$ vs. random environment $\sim \beta$
- Low $\nu\beta$: follows a random walk with $x \sim t^{1/2}$
- High $\nu\beta$: polymer is "pinned" by random potential, $x \sim t^{2/3}$ (superdiffusion)
- In 1D, any finite β results in the low temperature/high $\nu\beta$ phase

Teacher-student scenario (Zdeborová & Krzakala, 2016)

- Teacher generates true state X^* then data Φ with teacher's parameters T
- Student receives only Φ and conducts Bayesian inference assuming student parameters S to generate posterior for inferred state X
- Joint distribution is

$$
\begin{aligned}
p(X, \Phi, X^*) &= p_{\mathrm{S}}(X \mid \Phi) p_{\mathrm{T}}(\Phi \mid X^*) p_{\mathrm{T}}(X^*) \\
&= \frac{p_{\mathrm{S}}(\Phi \mid X) p_{\mathrm{S}}(X) p_{\mathrm{T}}(\Phi \mid X^*) p_{\mathrm{T}}(X^*)}{p_{\mathrm{S}}(\Phi)}
\end{aligned}
$$

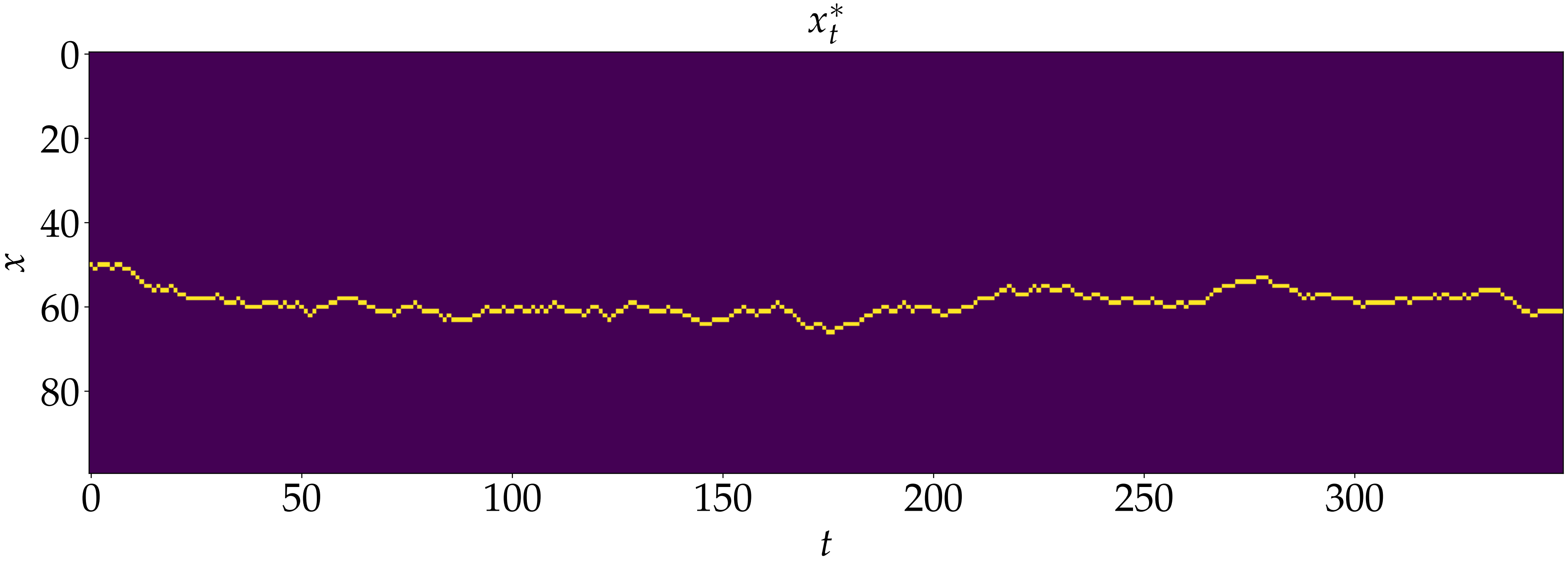
- At Bayes optimality $\mathrm{S} = \mathrm{T}$, X distributed identically to $X^* \Rightarrow x \sim t^{1/2}$
- Note that even with full knowledge of teacher's parameters perfect inference is not possible in general as data is still generated randomly

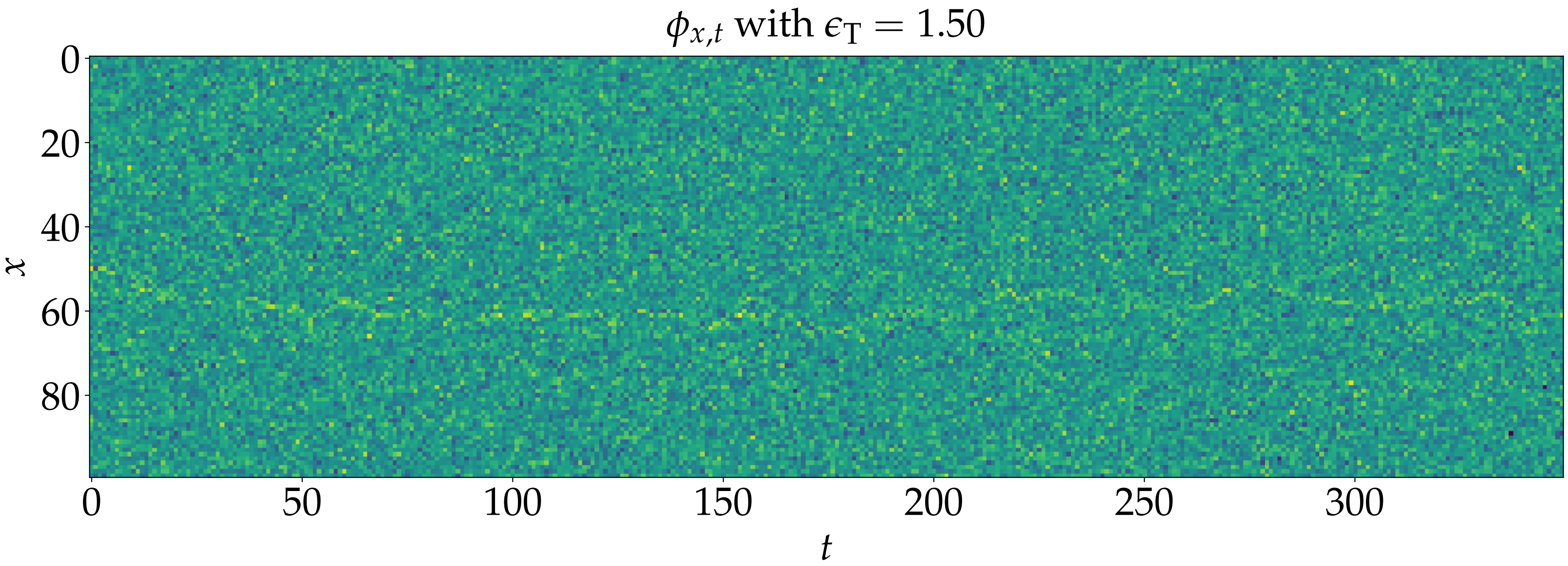
Planting

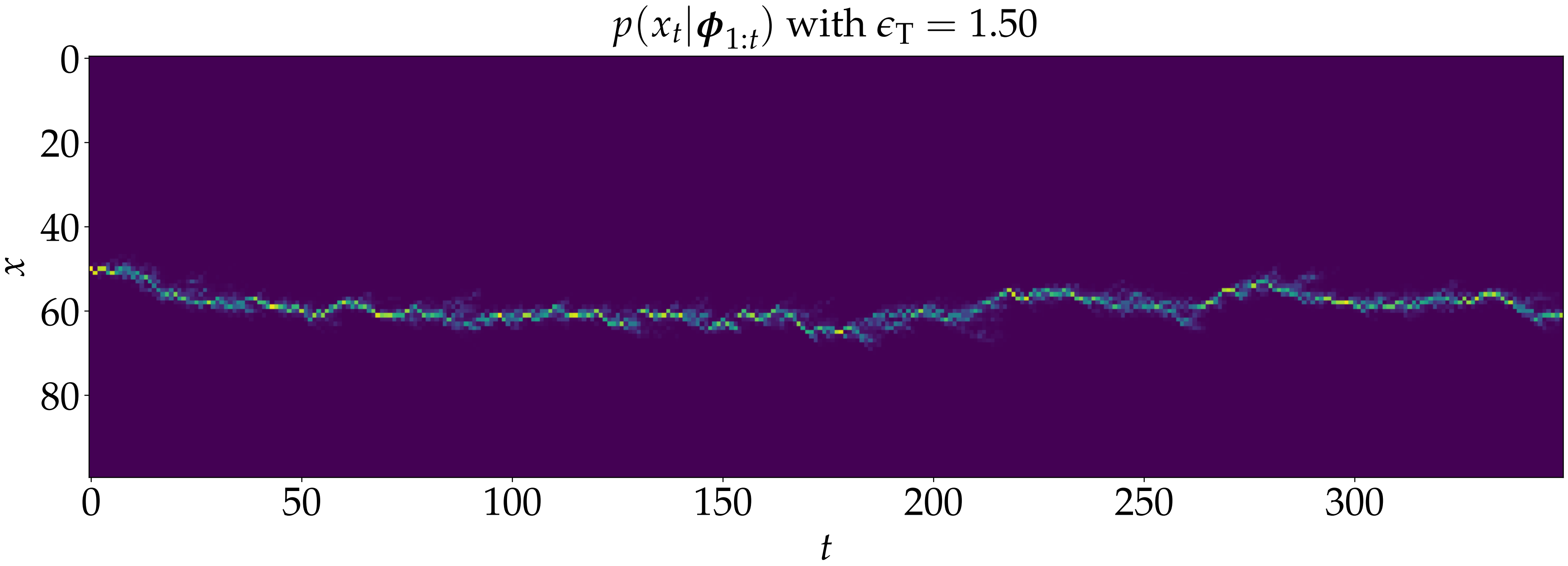
Unlike the original directed polymer the disorder is "planted" by the true path

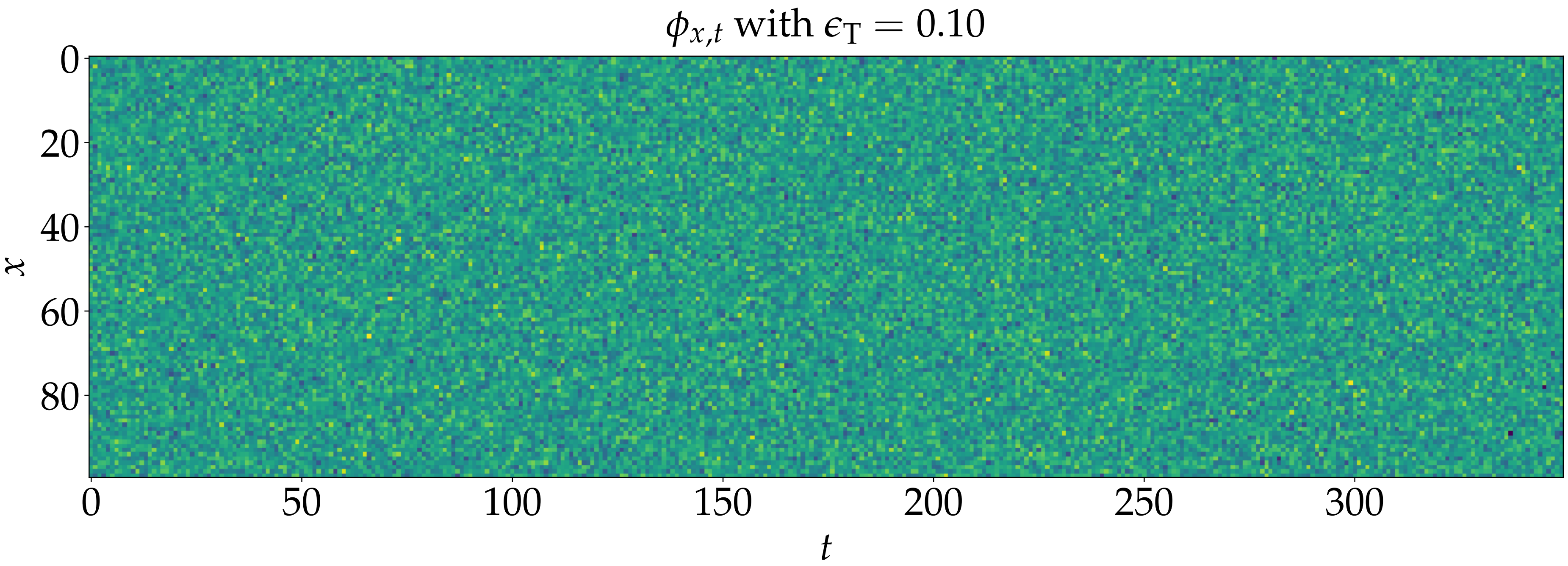
$$
\begin{aligned}
p(\Phi) &= \sum_{X^*} p_{\mathrm{T}}(\Phi \,|\, X^*) p_{\mathrm{T}}(X^*) \\
&\propto \pi_{\mathrm{T}}(\Phi) \sum_{X^*} \exp\left[\sum_t \left(-\frac{1}{2\nu_{\mathrm{T}}}(x_t^* - x_{t-1}^*)^2 + \frac{\epsilon_{\mathrm{T}}}{\sigma_{\mathrm{T}}^2}\phi_{x_t^*,t}\right)\right]
\end{aligned}
$$

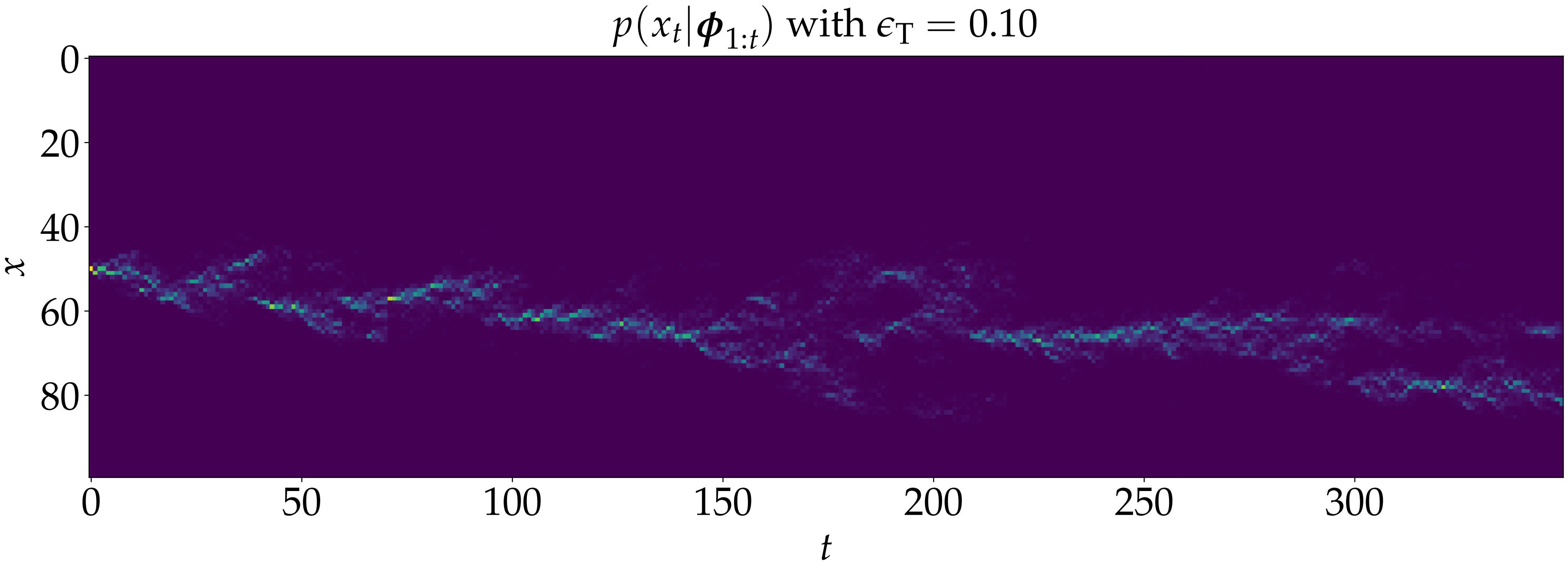
At $\epsilon_{\mathrm{T}} = 0$ coincides with the directed polymer


x_t^*
0
20
40
60
80
x
0
50
100
150
200
250
300
t


$\phi_{x,t}$ with $\epsilon_T = 1.50$
0
20
40
60
80
x
0
50
100
150
200
250
300
t


$p(x_t|\boldsymbol{\phi}_{1:t})$ with $\epsilon_\mathrm{T} = 1.50$
0
20
40
60
80
x
0
50
100
150
200
250
300
t


$\phi_{x,t}$ with $\epsilon_{\mathrm{T}} = 0.10$
0
20
40
60
80
x
0
50
100
150
200
250
300
t

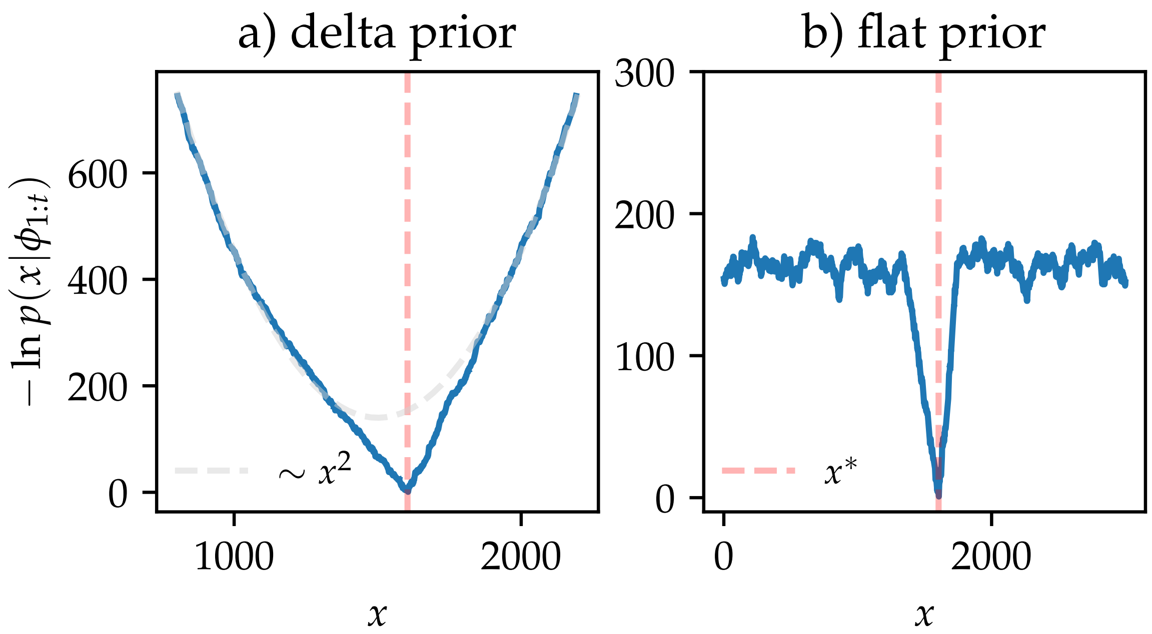

$p(x_t|\boldsymbol{\phi}_{1:t})$ with $\epsilon_{\mathrm{T}} = 0.10$
0
20
40
60
80
x
0
50
100
150
200
250
300
t

Observables

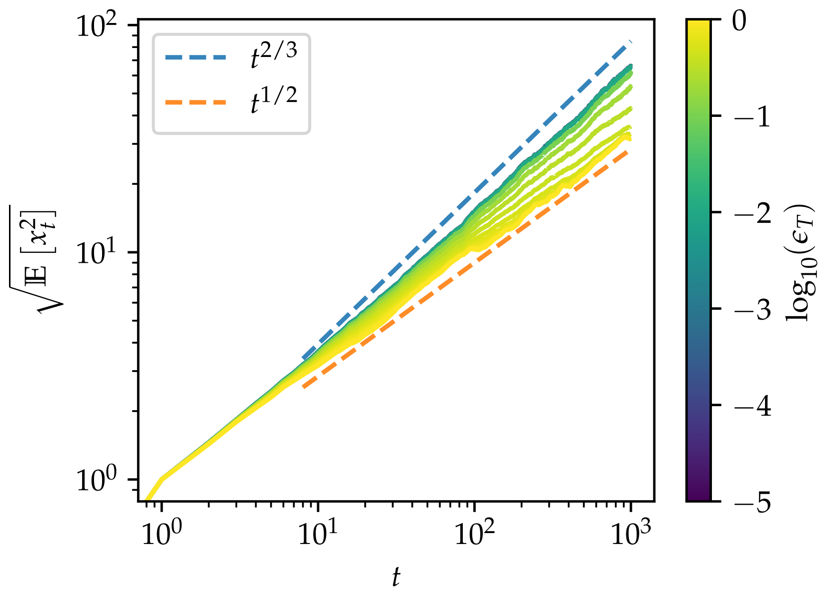
Mean-squared error $\mathrm{MSE}_t = \mathbb{E}\left[(x_t - x_t^*)^2\right]$

Overlap $Y_t = \frac{1}{t}\mathbb{E}\left[\sum_{\tau=1}^{t} \delta_{x_\tau, x_\tau^*}\right]$

"Free energy" profiles

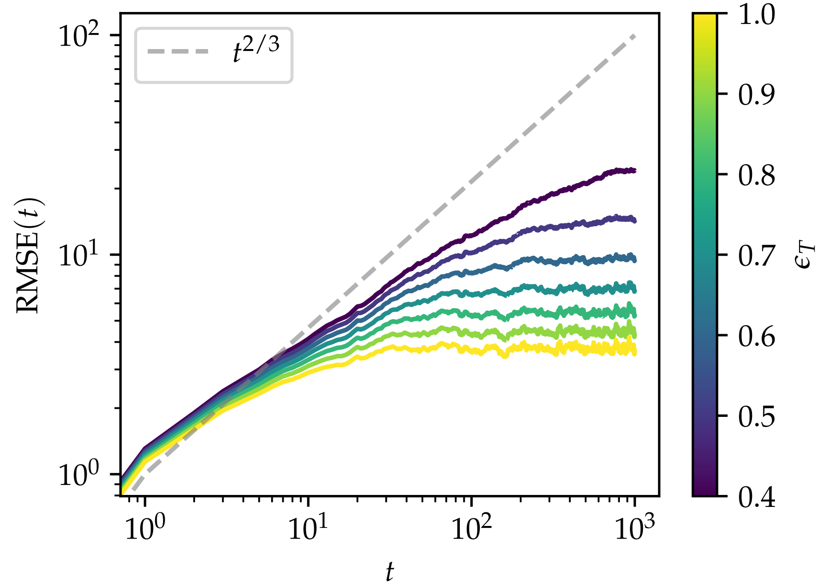


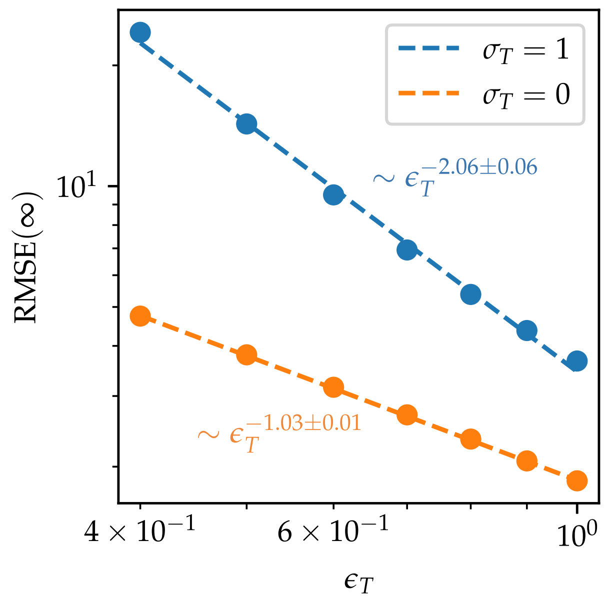
Fluctuations



- Crossover from Gaussian $\sim t^{1/2}$ to KPZ $\sim t^{2/3}$
- No evidence of phase transition

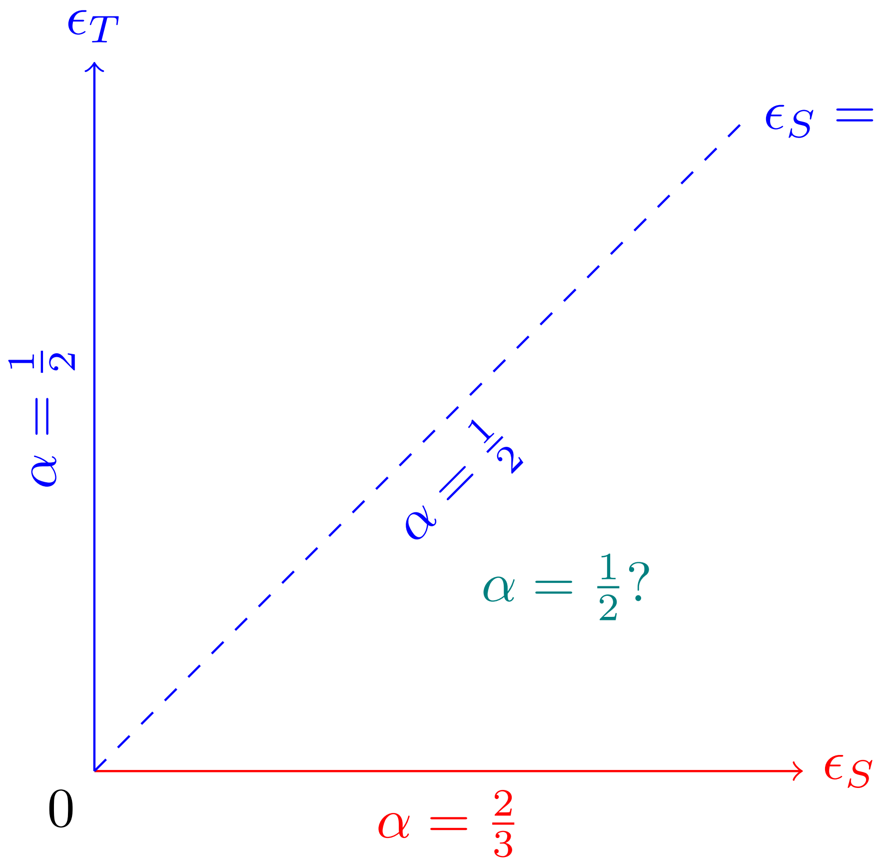
Root-mean-squared error



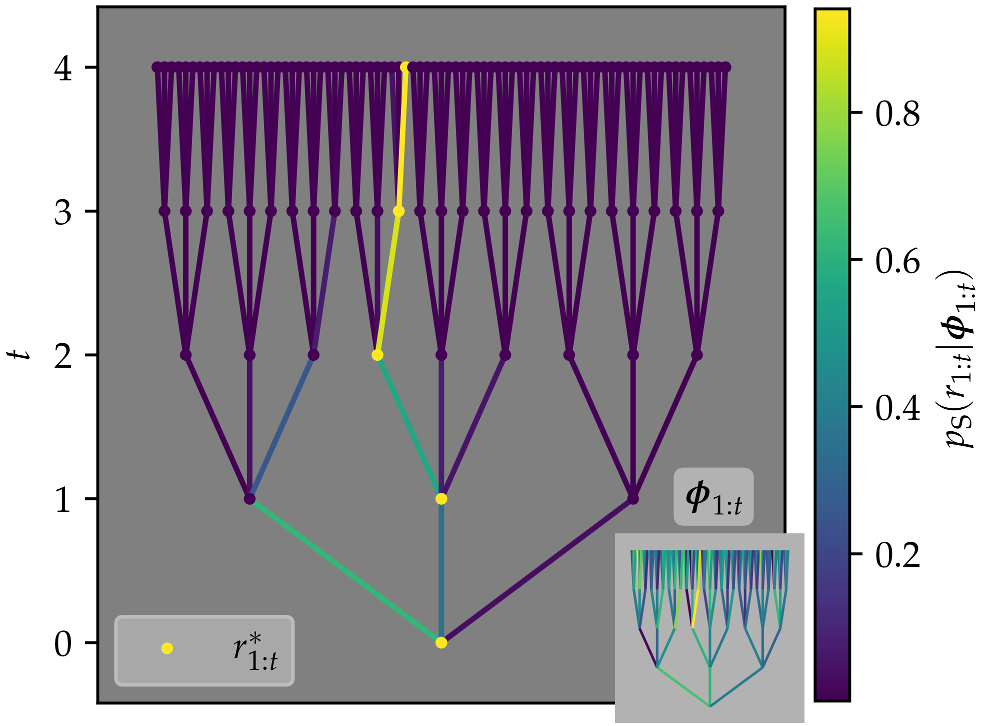


- Numerics points to finite $\mathrm{RMSE}(t)$ with any finite true/teacher signal strength ϵ_T

Conjectured phase diagram



Tree case



- At every timestep, move deeper into the tree with branching number k

Directed polymer on a tree

Directed polymer on a tree studied via generalised random energy model (Derrida and Gardner, 1987) and travelling waves (Derrida and Spohn, 1988)

$$Z = \sum_{r_{1:t}} \exp\left[-\beta \Psi(r_{1:t})\right], \text{ where } \Psi(r_{1:t}) = \psi_{r_1,t} + \cdots + \psi_{r_{1:t},t}$$

GREM: map energy of each spin configuration $\{\sigma\}$ to energy through a branch

Travelling waves approach: use recursive relation of partition function

$$Z(t) = e^{-\beta\phi} \sum_{i=1}^{k} Z^{(i)}(t-1) \text{ and study the generating function}$$

$$G_t(x) := \mathbb{E}\left[\exp\left(-e^{-\beta x} Z(t)\right)\right]$$

Planted directed polymer on a tree

- Due to the structure of the tree, can choose any one to be the true path $r^*_{1:t}$
- Look at overlap with this path: fraction of time the inferred path is equal to the true path

GREM approach

Separate $\phi_{r_{1:\tau},\tau} = \psi_{r_{1:\tau},\tau} + \epsilon_{\mathrm{T}}\delta_{r_{1:\tau},r^*_{1:\tau}}$

Student's posterior $p_{\mathrm{S}}(r_{1:t}\,|\,\Psi, r^*_{1:t}) = q_{\mathrm{S}}(r_{1:t}\,|\,\Psi, r^*_{1:t})/Z(\Psi, r^*_{1:t})$ given by partition function

$$Z(\Psi, r^*_{1:t}) = \sum_{r_{1:t}} \exp\left[\beta\Psi(r_{1:t}) + \beta\epsilon_{\mathrm{T}} t y(r_{1:t}, r^*_{1:t})\right]$$

$\beta = \epsilon_{\mathrm{S}}/\sigma^2_{\mathrm{S}}$, $y(r_{1:t}, r^*_{1:t})$ is the fractional overlap with true path

Each $\psi_{r_{1:\tau},\tau}$ is iid Gaussian distributed with variance σ^2_{T}

Relation to magnetisation on the GREM

Choose true path as the ferromagnetic configuration $\Rightarrow$ overlap is magnetisation!

Write partition function in terms of partial partition function

$$Z(\Psi, r^*_{1:t}) = \sum_y z_y(\Psi) e^{\beta \epsilon_{\mathrm{T}} y t} \text{ where } z_y(\Psi) = \sum_{r_{1:t}|y} e^{\beta \Psi(r_{1:t})} \text{ is another GREM}$$

Use maximum a-posteriori approximation $\lim_{t \to \infty} f = \max_y f_w$

where $f = \frac{1}{t} \mathbb{E}_\Psi \left[\ln Z(\Psi) \right]$ and $f_y = \frac{1}{t} \mathbb{E}_\Psi \left[\ln z_y(\Psi) \right] + \beta y \epsilon_{\mathrm{T}}$

Using results from Derrida, Spohn 1988

Original GREM has phase transition with β

$$\frac{1}{t}\mathbb{E}_{\Psi}\left[\ln z_w\left(\Psi, r^*_{1:t}\right)\right] = (1-w)\beta c_\beta$$

"speed of front" $c_\beta = \begin{cases} c(\beta) & \text{if } \beta \leq \beta_c \\ c\left(\beta_c\right) & \text{if } \beta > \beta_c \end{cases}$

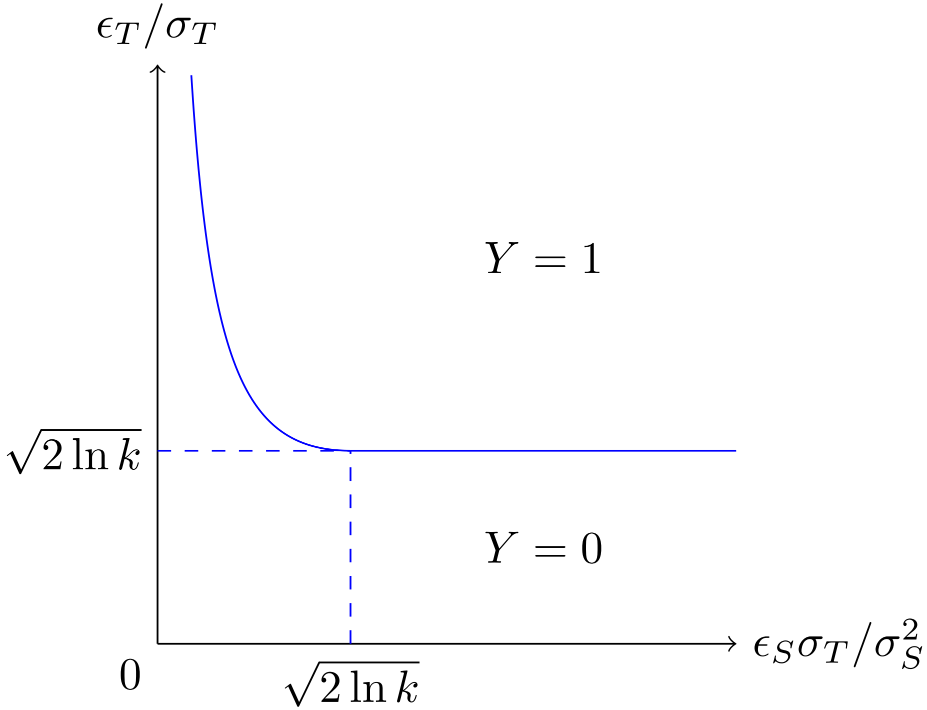
$$c(\beta) = \frac{1}{\beta}\ln\left(k\int d\psi \pi_{\mathrm{T}}(\psi) e^{-\beta\psi}\right) \text{ with } \frac{\partial}{\partial\beta}c(\beta)\bigg|_{\beta=\beta_c} = 0$$

$$f_y = \begin{cases} \left(\beta\epsilon_{\mathrm{T}} - \left[\frac{\ln k}{\beta} + \frac{\beta\sigma_{\mathrm{T}}^2}{2}\right]\right) y + C_1 & \text{if } \beta \leq \beta_c, \\ \left(\beta\epsilon_{\mathrm{T}} - \sigma_{\mathrm{T}}\sqrt{2\ln k}\right) y + C_2 & \text{if } \beta > \beta_c, \end{cases}$$

Resulting in

$$\frac{\epsilon_{\mathrm{T}}}{\sigma_{\mathrm{T}}} = \begin{cases} \ln k \left(\frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2}\right)^{-1} + \frac{1}{2}\frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2} & \text{if } \frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2} \leq \sqrt{2\ln k} \\ \sqrt{2\ln k} & \text{if } \frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2} > \sqrt{2\ln k} \end{cases}$$

Phase diagram on the tree



- Average overlap $Y_{t\to\infty}$ given by y that minimises free energy

Outlook

- Phase transition in higher dimensions? (Offer, 2018)
- Analytic solution using machinery of 1D directed polymer
- Connection to quantum measurement-induced phase transitions (MIPT)

$$\rho_t = M_t U_t \rho_{t-1} U_t^\dagger M_t^\dagger \text{ is like } q_t = T q_{t-1}$$

Traveling waves approach

Recursion for planted case is

$$Z_P\left(\epsilon_{\mathrm{T}}, t+1\right) = e^{\beta\left(\psi - \epsilon_{\mathrm{T}}\right)}\left(Z_P^{(1)}\left(\epsilon_{\mathrm{T}}, t\right) + \sum_{i=2}^{k} Z_P^{(i)}(0,t)\right)$$

Recall generating function $G_{\epsilon_{\mathrm{T}}}(x,t) := \mathbb{E}_{\Psi}\left[\exp\left(-Z_P\left(\epsilon_{\mathrm{T}}, t\right) e^{-\beta x}\right)\right]$

Evolves as

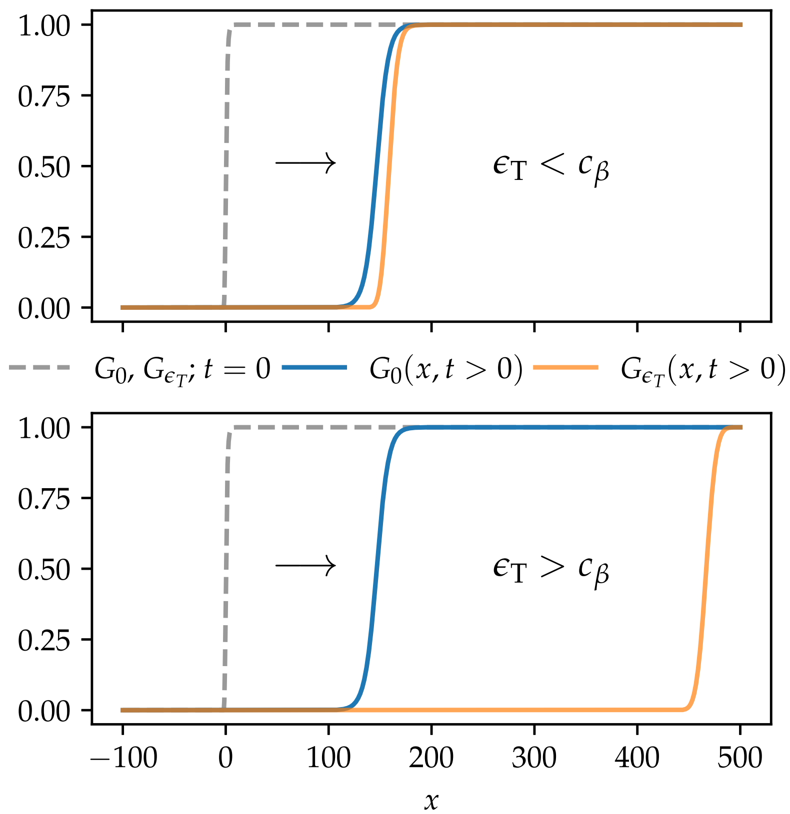
$$G_{\epsilon_{\mathrm{T}}}(x, t+1) = \mathbb{E}_{\psi}\left[G_{\epsilon_{\mathrm{T}}}\left(x+\psi-\epsilon_{\mathrm{T}}, t\right) \times G_0\left(x+\psi-\epsilon_{\mathrm{T}}, t\right)^{k-1}\right]$$

In the continuum limit generating function evolves as

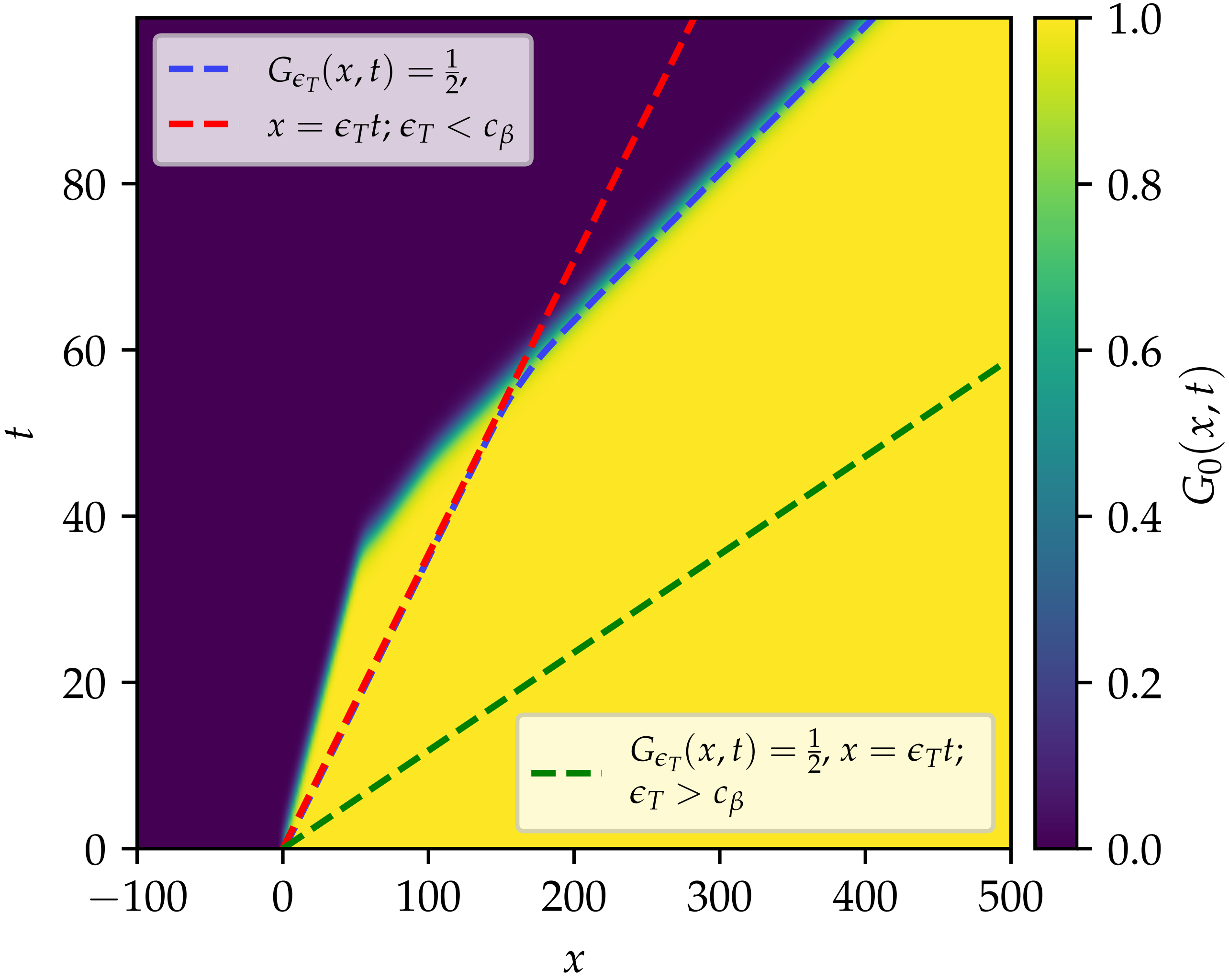
$$\partial_t G_{\epsilon_\mathrm{T}} = D\partial_x^2 G_{\epsilon_\mathrm{T}} - \epsilon_\mathrm{T}\partial_x G_{\epsilon_\mathrm{T}} - \lambda \left(1 - G_0\right) G_{\epsilon_\mathrm{T}}$$

$\epsilon_\mathrm{T} = 0$: FKPP equation, with minimum speed of front

$\epsilon_\mathrm{T} \neq 0$ evolves linearly with characteristic velocity ϵ_T but also with G_0 "carrying" G_{ϵ_T}



Both start with profile $\exp(e^{-\beta x})$



$G_{\epsilon_T}(x,t) = \frac{1}{2}$,
$x = \epsilon_T t; \epsilon_T < c_\beta$
$G_{\epsilon_T}(x,t) = \frac{1}{2}$, $x = \epsilon_T t$;
$\epsilon_T > c_\beta$
80
60
40
20
0
−100
0
100
200
300
400
500
x
t
1.0
0.8
0.6
0.4
0.2
0.0
$G_0(x,t)$

Using results from Derrida, Spohn 1988

G_0 has phase transition with β

$$\text{"speed of front" } c_\beta = \begin{cases} c(\beta) & \text{if } \beta \leq \beta_c \\ c\left(\beta_c\right) & \text{if } \beta > \beta_c \end{cases}$$

$$c(\beta) = \frac{1}{\beta} \ln\left(k \int d\psi \pi_{\mathrm{T}}(\psi) e^{-\beta\psi}\right) \text{ with } \frac{\partial}{\partial\beta} c(\beta)\bigg|_{\beta=\beta_c} = 0$$

So speed of planted generating function is $v(\epsilon_{\mathrm{T}}) = \max(\epsilon_{\mathrm{T}}, c_\beta)$

$G_{\epsilon_T}(x, t)$ switches from 0 to 1 approximately at point $\beta\hat{x} \sim \ln Z_P(\epsilon_T, t)$

At long times $\hat{x}(t) = \max(\epsilon_T, c_\beta)t$

Assuming sublinear fluctuations in $\ln Z_P(\epsilon_T, t)$,

$$\mathbb{E}_\Psi\left[\ln Z_P(\epsilon_T, t)\right] = \beta \max(\epsilon_T, c_\beta)t$$

From definition of overlap

$$Y = \frac{1}{\beta t}\frac{\partial}{\partial \epsilon_{\mathrm{T}}}\mathbb{E}_{\Psi}[\ln Z_P(\epsilon_{\mathrm{T}}, t)] = \frac{\partial}{\partial \epsilon_{\mathrm{T}}} v_{\beta,\epsilon_{\mathrm{T}}}$$

- If $\epsilon_{\mathrm{T}} < c_\beta$, $v(\epsilon_{\mathrm{T}})$ constant with ϵ_{T} and $Y = 0$
- If $\epsilon_{\mathrm{T}} > c_\beta$, $v(\epsilon_{\mathrm{T}}) = \epsilon_{\mathrm{T}}$ constant and $Y = 1$

Resulting in

$$\frac{\epsilon_{\mathrm{T}}}{\sigma_{\mathrm{T}}} = \begin{cases} \ln k \left(\frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2}\right)^{-1} + \frac{1}{2}\frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2} & \text{if } \frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2} \leq \sqrt{2\ln k} \\ \sqrt{2\ln k} & \text{if } \frac{\epsilon_{\mathrm{S}}\sigma_{\mathrm{T}}}{\sigma_{\mathrm{S}}^2} > \sqrt{2\ln k} \end{cases}$$