Theoretical Foundations for Computational Sampling

计算采样的理论基础

南京大学 尹一通（新基石研究员）
Yitong Yin, Nanjing University

Turing's Proof (1936)

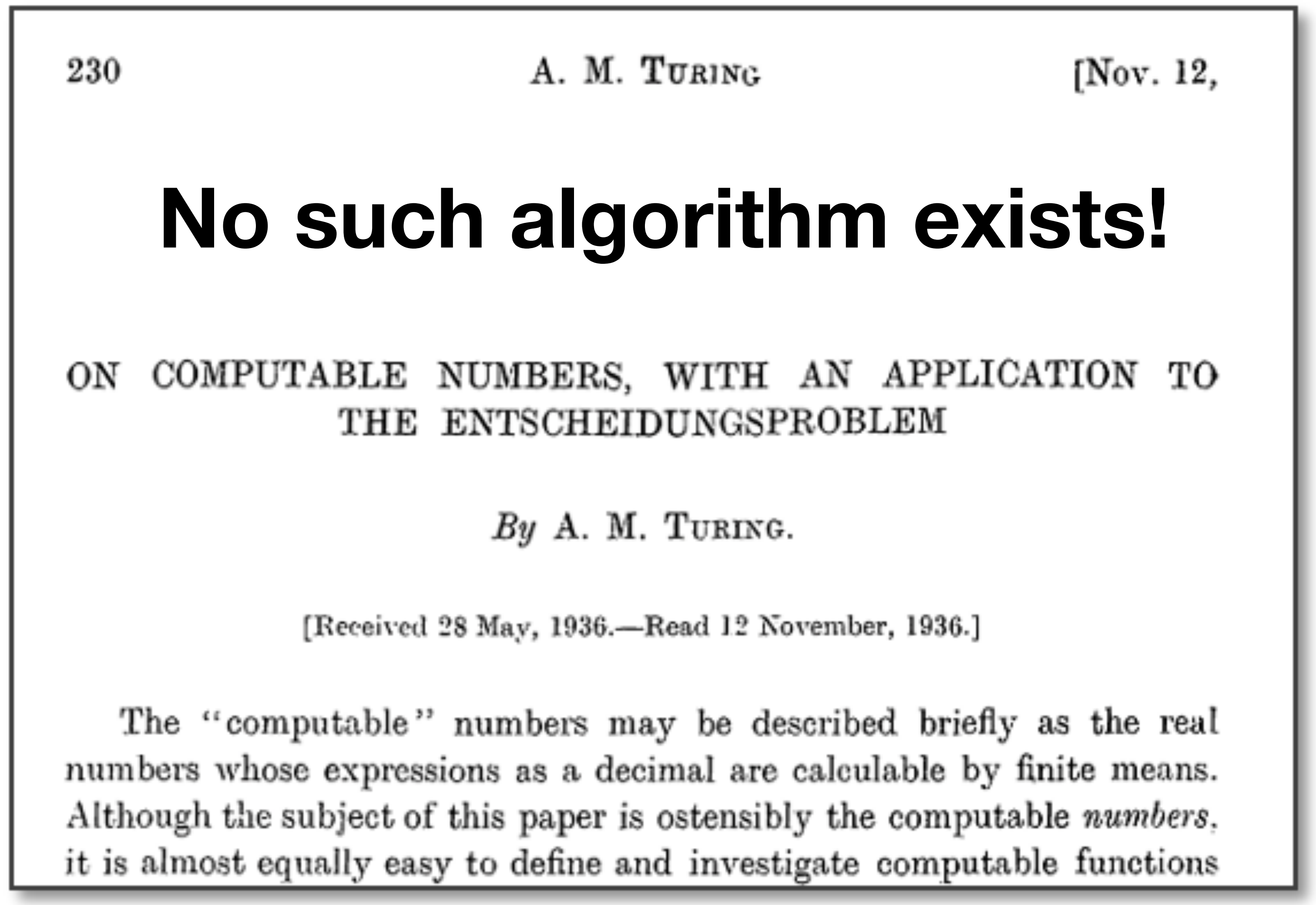
David Hilbert

Entscheidungsproblem (1928):

Give an algorithm which determines the validity of mathematical statements.

Is Mathematics decidable?

Alan Turing



230 A. M. TURING [Nov. 12,

No such algorithm exists!

ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGSPROBLEM

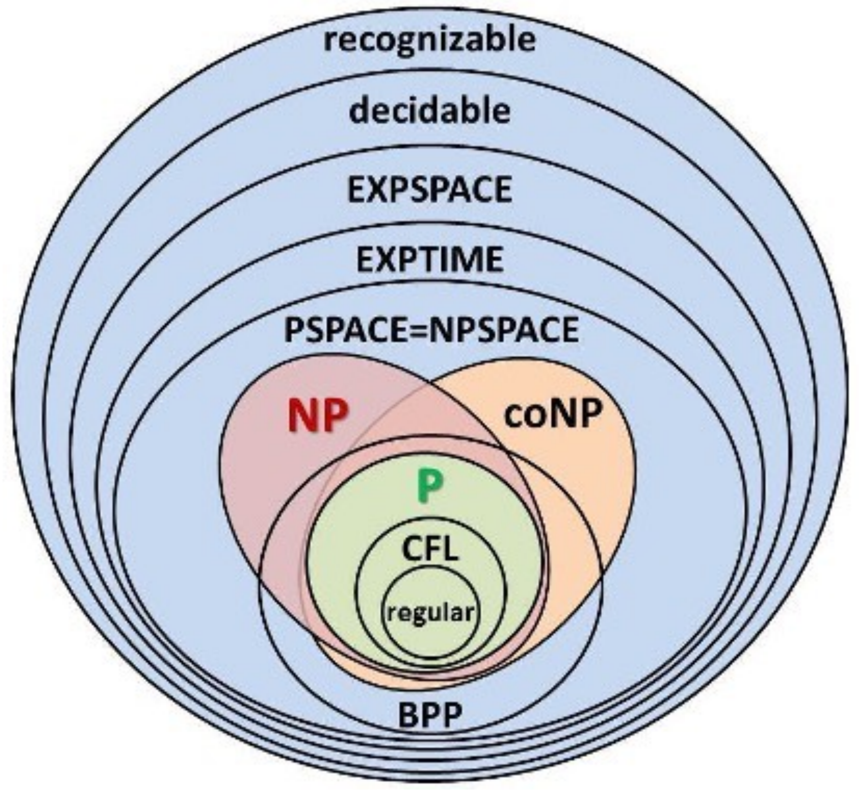
By A. M. TURING.

[Received 28 May, 1936.—Read 12 November, 1936.]

The "computable" numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means. Although the subject of this paper is ostensibly the computable *numbers*, it is almost equally easy to define and investigate computable functions

The Birth of (Theoretical) Computer Science

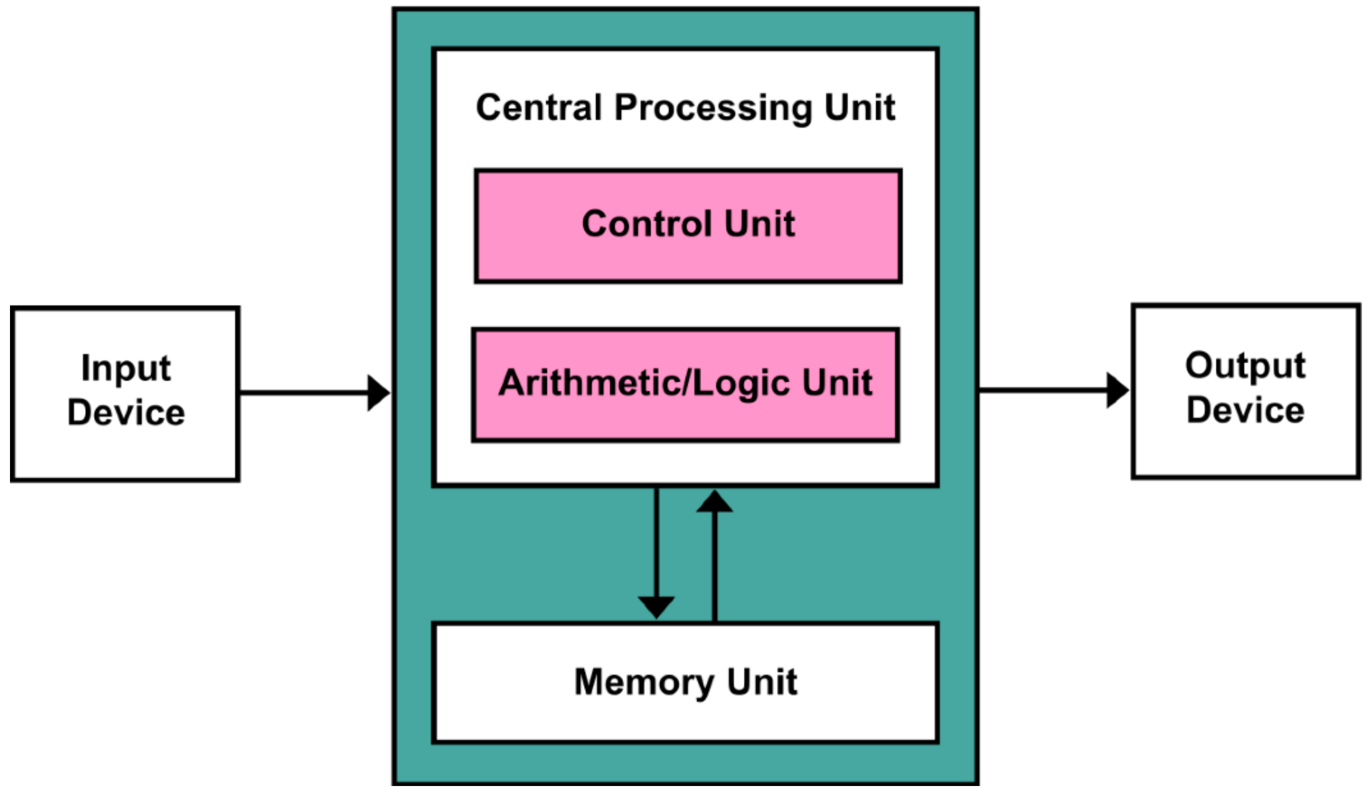
- **Computation is incomplete:** not all problems are computable
- "*What makes a problem easy/hard to resolve by computer?*"

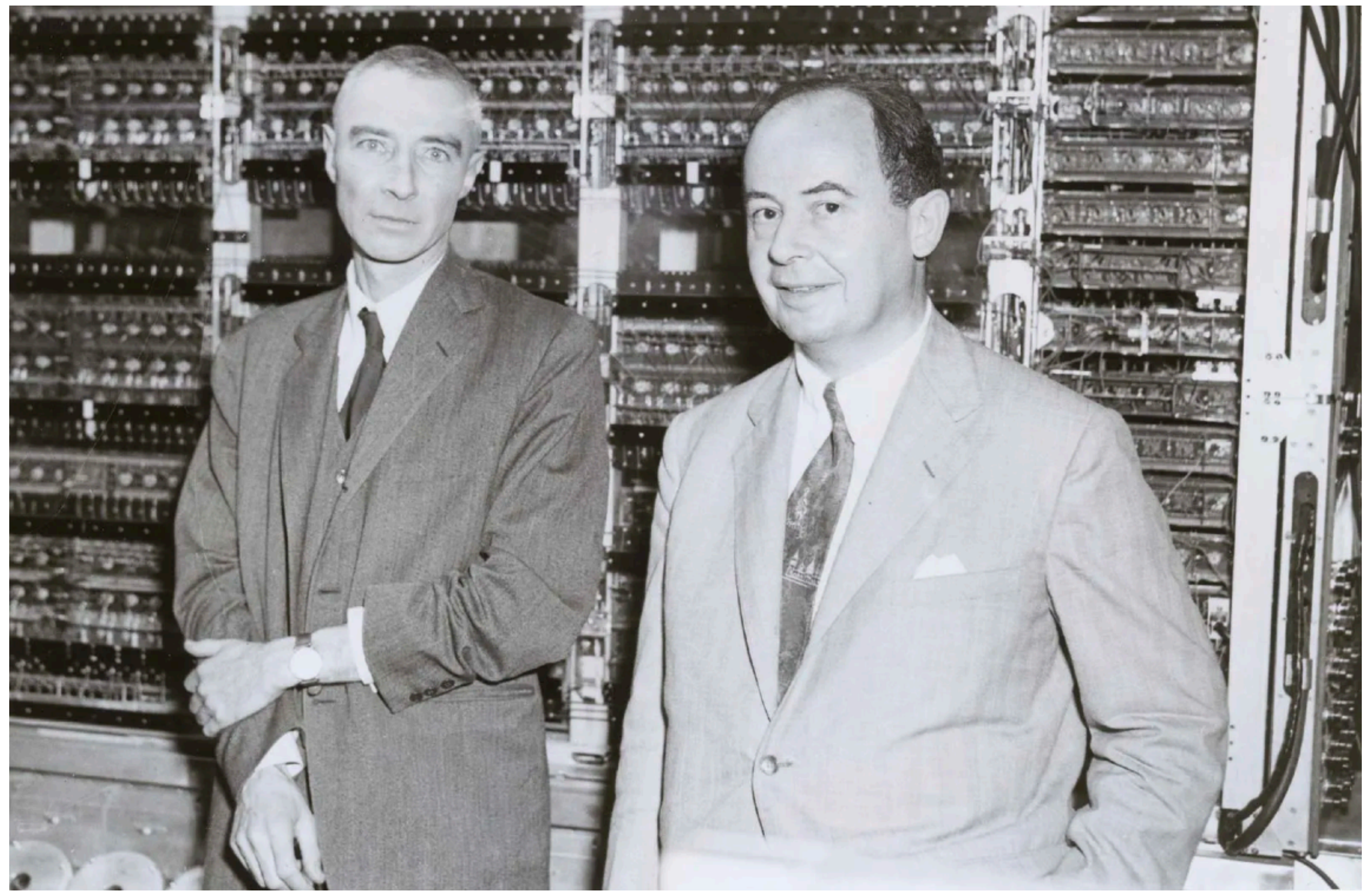


Los Alamos National Lab (1945 ~ 1947)

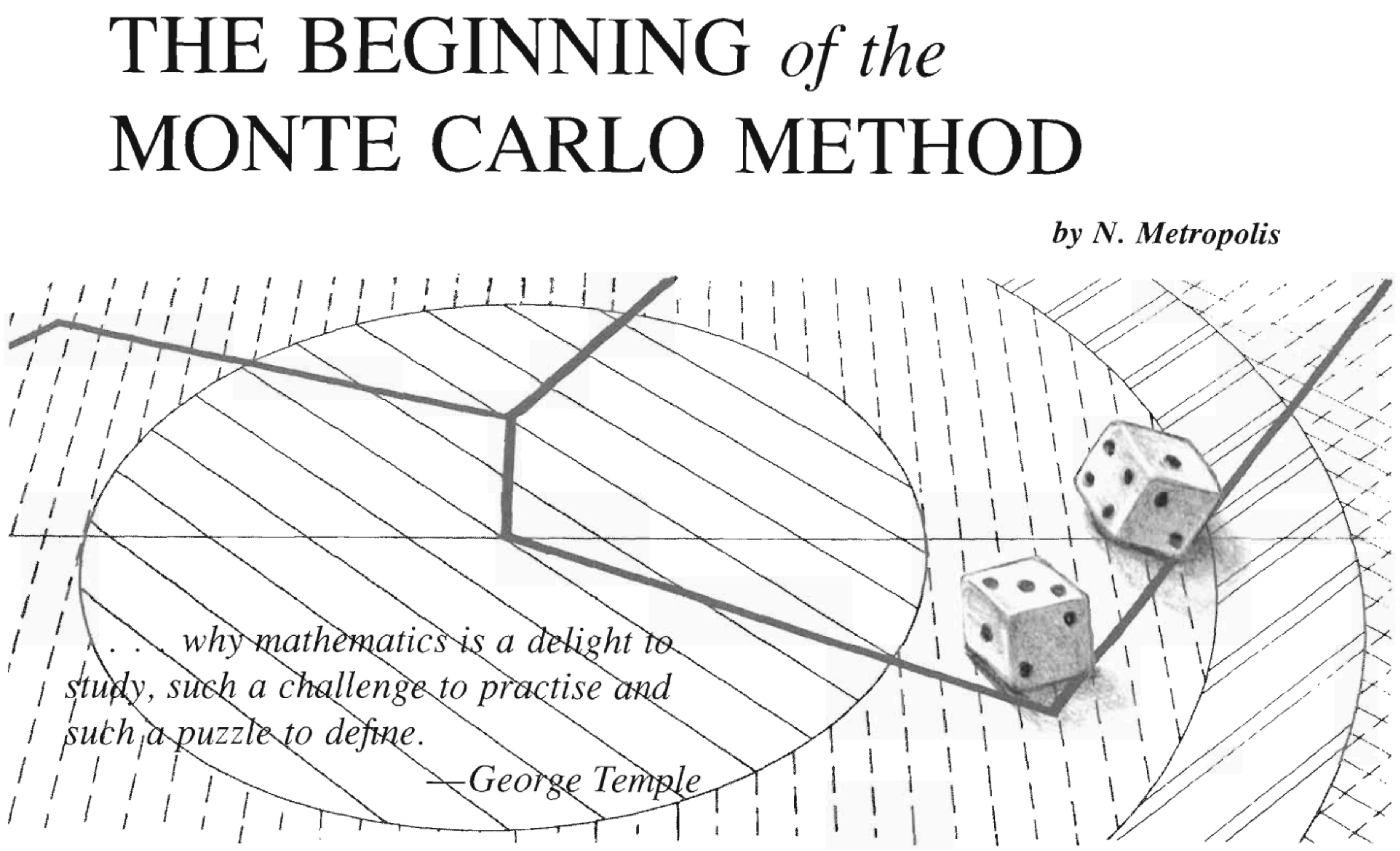
John von Neumann

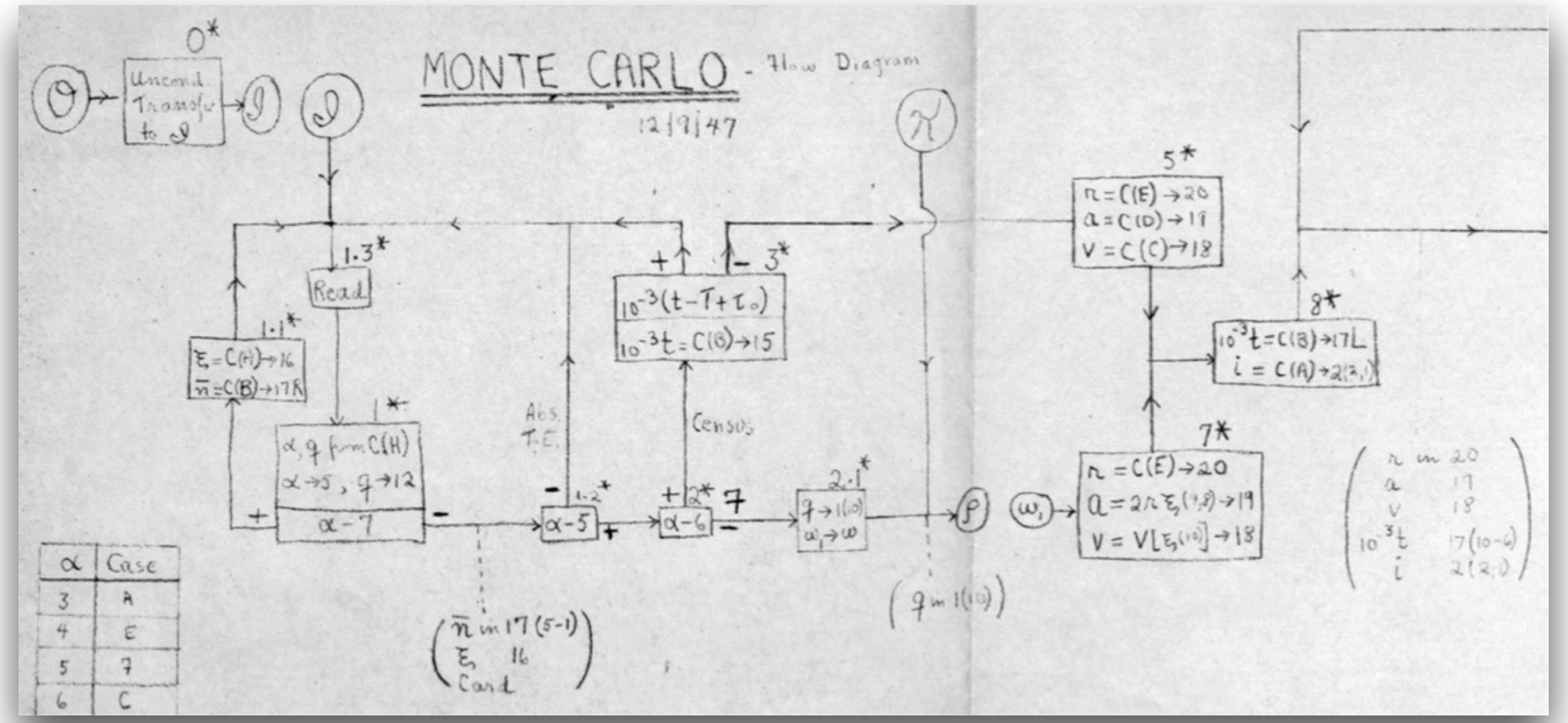
von Neumann Architecture





ENIAC
(Electronic Numerical Integrator and Computer)



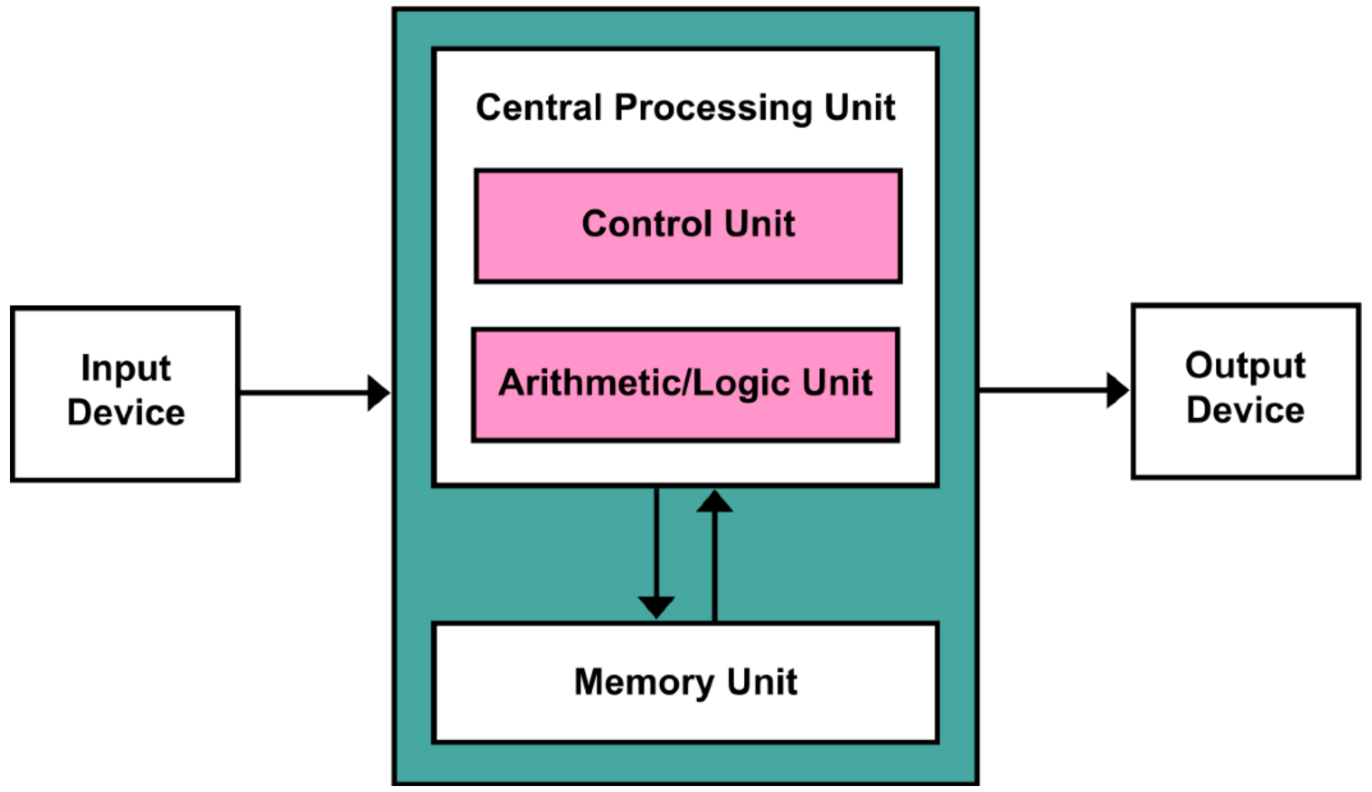


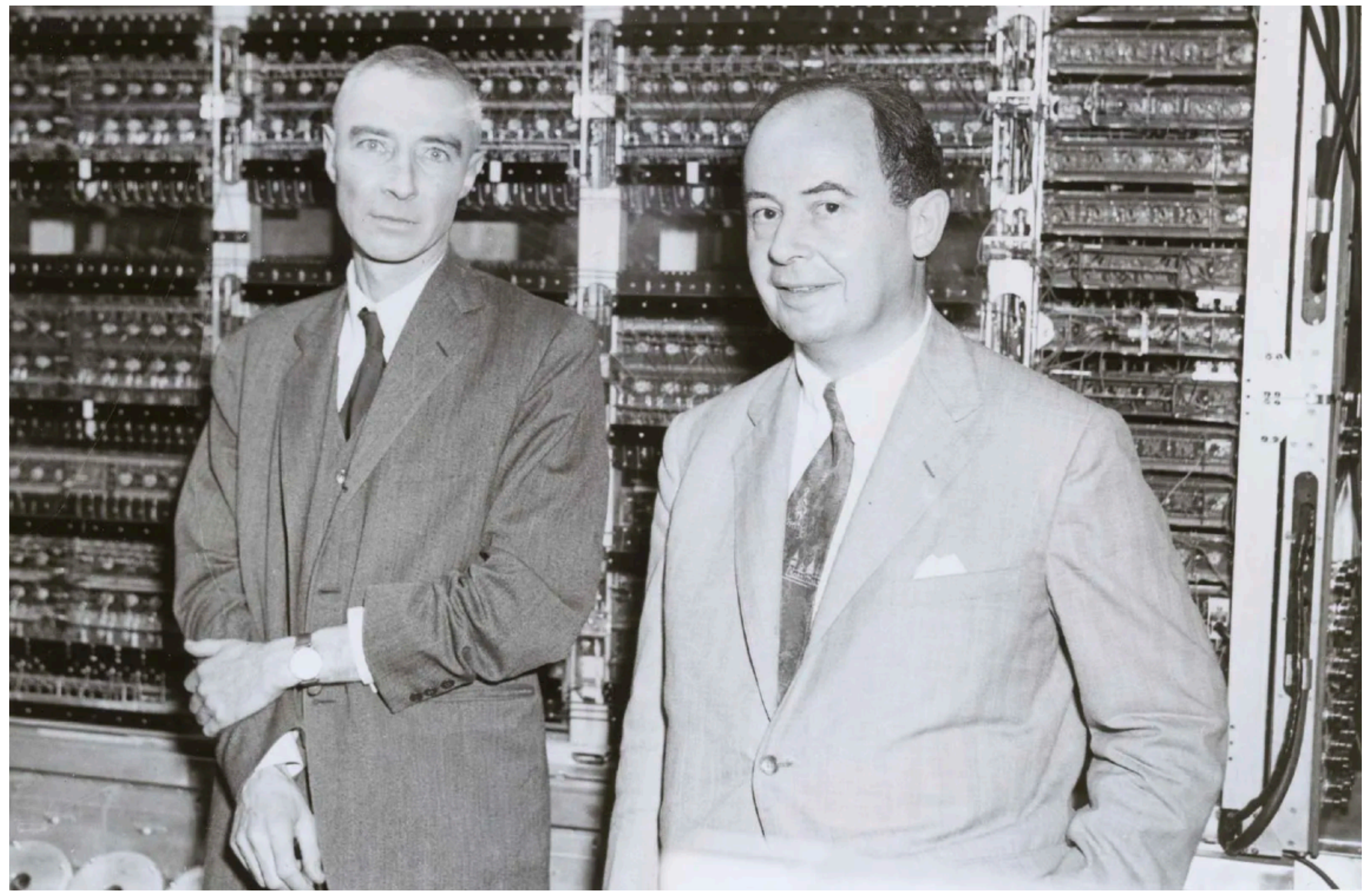
Monte Carlo Method

Los Alamos National Lab (1945 ~ 1947)

John von Neumann

von Neumann Architecture





ENIAC
(Electronic Numerical Integrator and Computer)

THE BEGINNING *of the* MONTE CARLO METHOD

Codename *Monte Carlo*

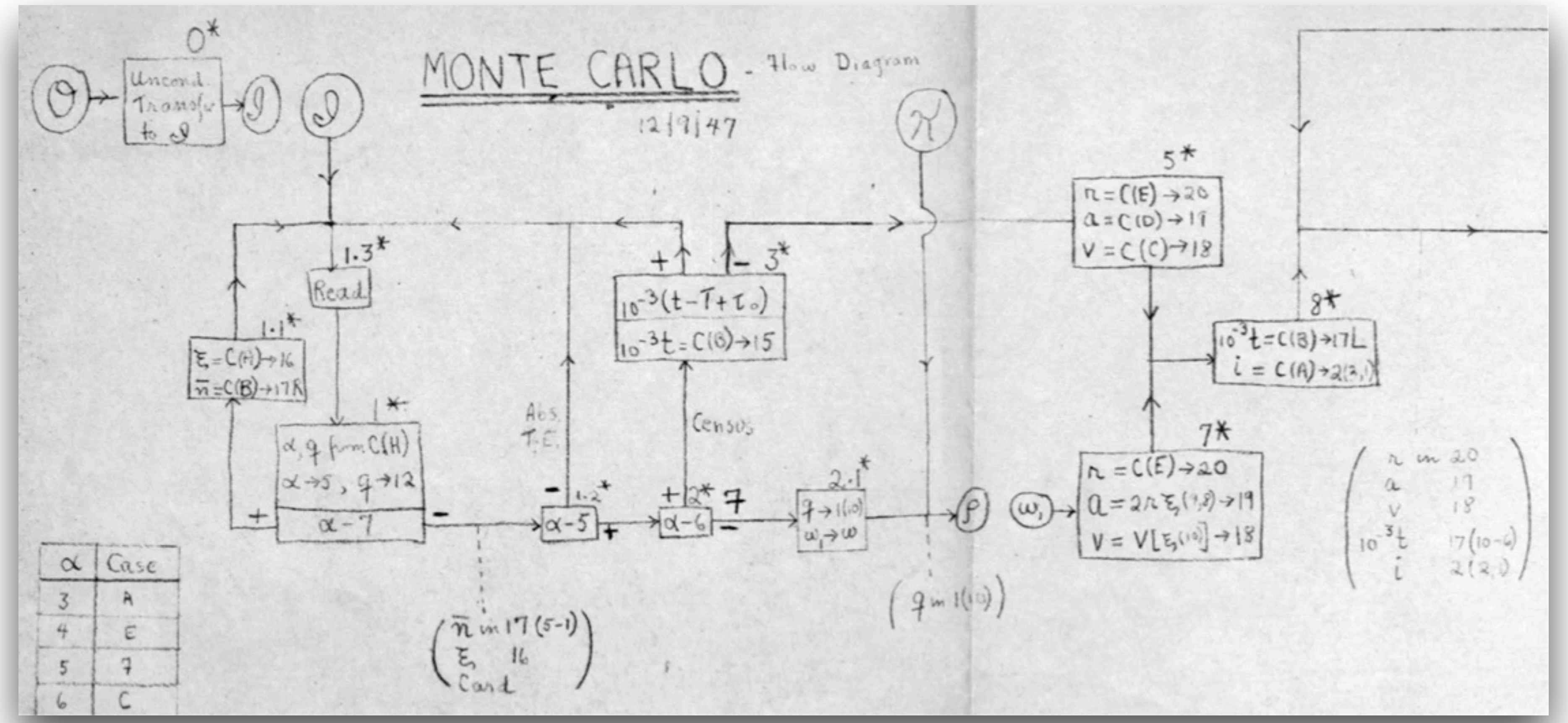
Nicholas Metropolis

Stanislaw Ulam

John von Neumann

Independently

Enrico Fermi



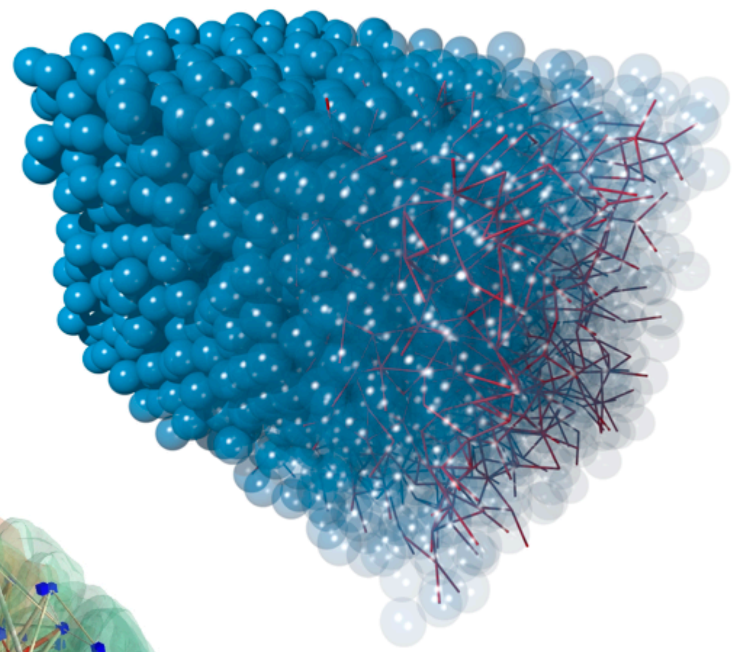
Monte Carlo Method

Computational Sampling

Draw a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ according to distribution μ.

- Can solve the problems [*e.g. neutron diffusion in the core of a nuclear weapon*] that were difficult to solve using conventional, deterministic methods.

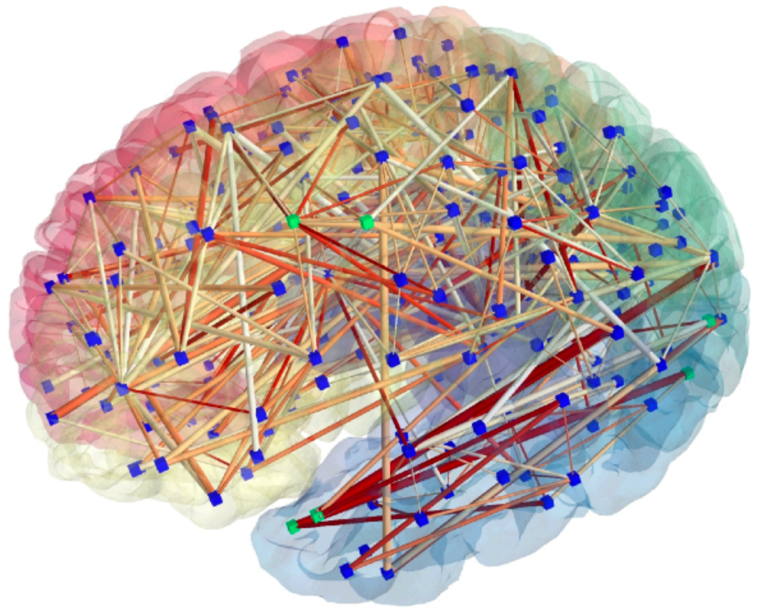
- **Boltzmann distribution** (**Gibbs measure**) in **statistical physics:**



(locally interacting particle states)

$$\mu(\boldsymbol{X}) \propto \exp\left(-\beta \sum_j H_{\Lambda_j}(\boldsymbol{X}_{\Lambda_j})\right)$$

- **Statistical inference/algorithms** in **data science:**



(locally constrained random variables)

$$\mu(\boldsymbol{X}) \propto \prod_j f_j(\boldsymbol{X}_{S_j})$$

- **Integration** in **high dimension** $\int_{\mathbb{R}^n}$, **reliability** of **complex system**, …

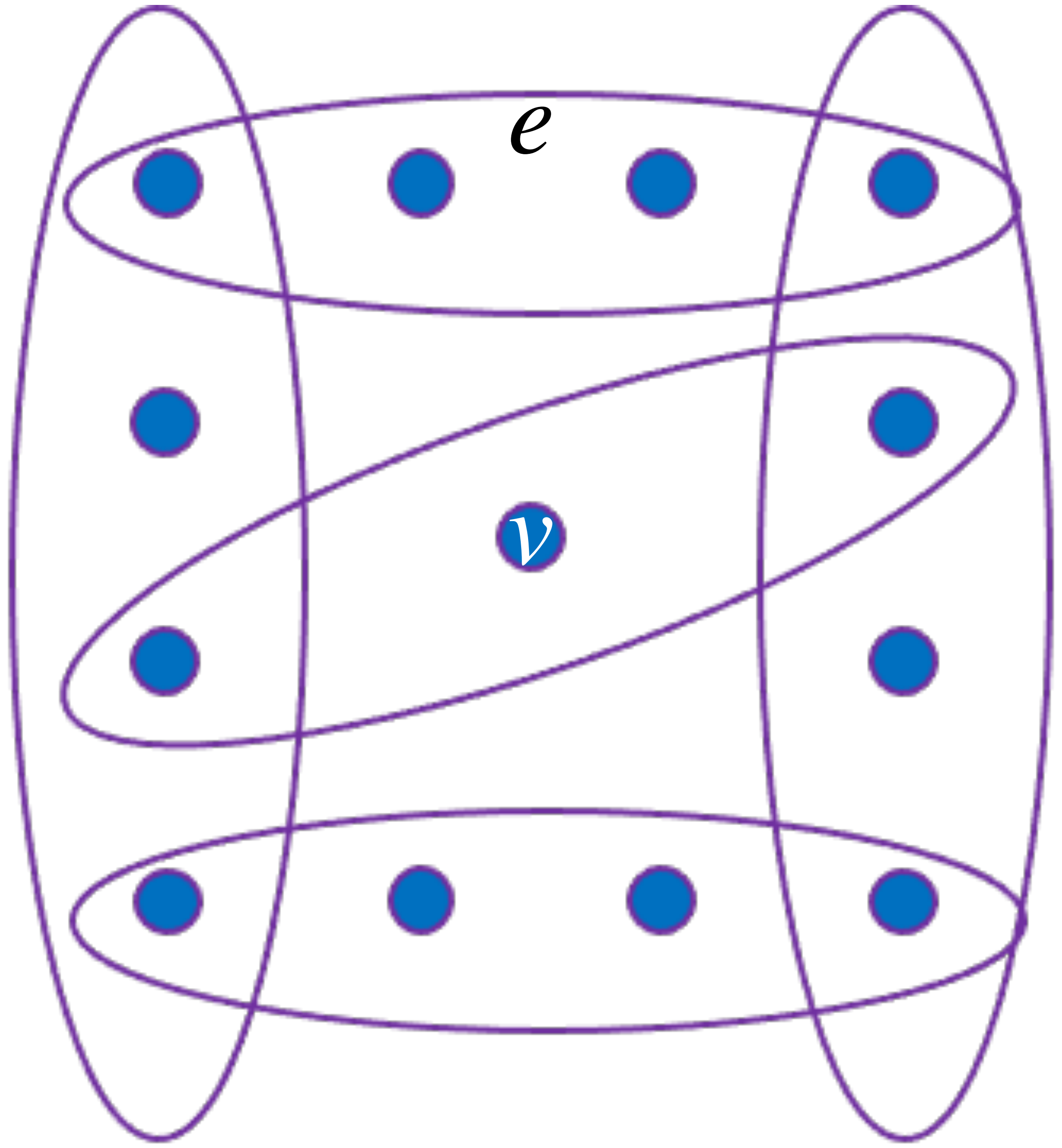
Milestones in Theory of Computing

Draw samples $X \sim \mu \implies$ Approximate $\mu(B) = \int_B \mathrm{d}\mu$, and many more …

- **Computational complexity** of *exact computation*:
  - Leslie Valiant (1979) (**Turing award** 2010): **#P-completeness.**
  - Toda's Theorem (1991) (**Gödel Prize** 1998): $\mathbf{NP}^{\mathbf{NP}^{\cdot^{\cdot^{\cdot}}}} \subseteq \#\mathbf{P}$
  - Bulatov (2013), Dyer-Richerby (2013), Cai-Chen (2017) (**Gödel Prize** 2021): Complexity dichotomy.
- **Monte Carlo method** for *approximate computing*:
  - Dyer-Frieze-Kannan (1991) (**Fulkerson Prize** 1991): **Integration** $\int_B f(x)\,\mathrm{d}x$ of convex f and **volume** $\mathrm{vol}(B)$ of convex body B.
  - Jerrum-Sinclair (1989) (**Gödel Prize** 1996): **Partition function** $Z_G(\beta)$.
  - Jerrum-Sinclair-Vigoda (2004) (**Fulkerson Prize** 2006): **Permanent** $\mathrm{perm}(A)$.

Graphical Model

(Markov random field / factor graph / weighted CSP …)



- Hypergraph $\mathscr{H} = (V, E)$
- vertex $v \in V$ corresponds to a **variable** of domain $[q]$
- hyperedge $e \in E$ (which is a vertex subset $e \subseteq V$) is associated with a **constraint** $f_e : [q]^e \to \mathbb{R}_{\geq 0}$
- **Gibbs distribution** μ over all configurations $\sigma \in [q]^V$:

$$\mu(\sigma) \propto \prod_{e \in E} f_e(\sigma_e)$$

- $\mu(\sigma) = \dfrac{\prod_{e \in E} f_e(\sigma_e)}{Z}$ where $Z := \sum_{\sigma \in [q]^V} \prod_{e \in E} f_e(\sigma_e)$ is called the **partition function**

Markov chain Monte Carlo (MCMC)

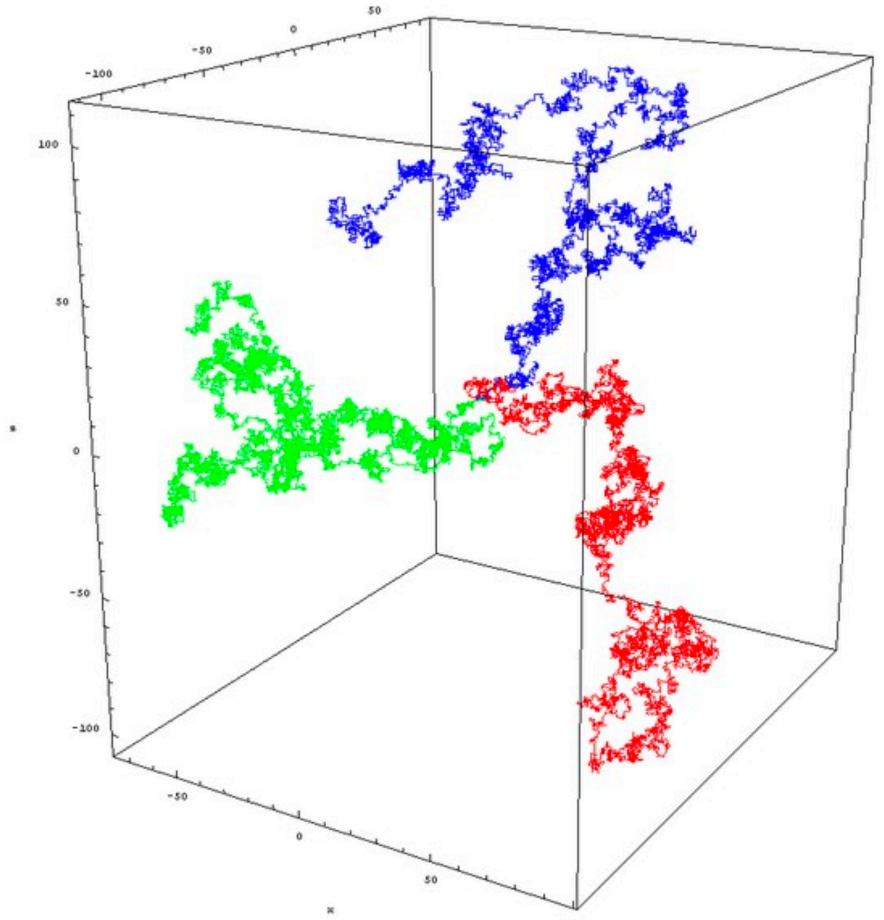
Glauber dynamics [Glauber 1963], *Gibbs sampler* [Geman-Geman 1984]

Draw a random sample $X \in [q]^V$ according to Gibbs distribution μ.

The **Markov chain** maintains an $X \in [q]^V$, at each step:

- pick $v \in V$ uniformly at random;
- update the evaluation of X_v according to its marginal distribution $\mu_v(\cdot \mid X_{N(v)})$.

Random walk in configuration space $[q]^n$



- The Markov chain has **stationary distribution** μ.
- **Mixing time:**
  $$\tau(\epsilon) := \max_{X^{(0)} \in [q]^V} \min\{t \geq 0 \mid d_{\mathrm{TV}}(X^{(t)}, \mu) \leq \epsilon\}$$
- **New sampling algorithms?**

Outline

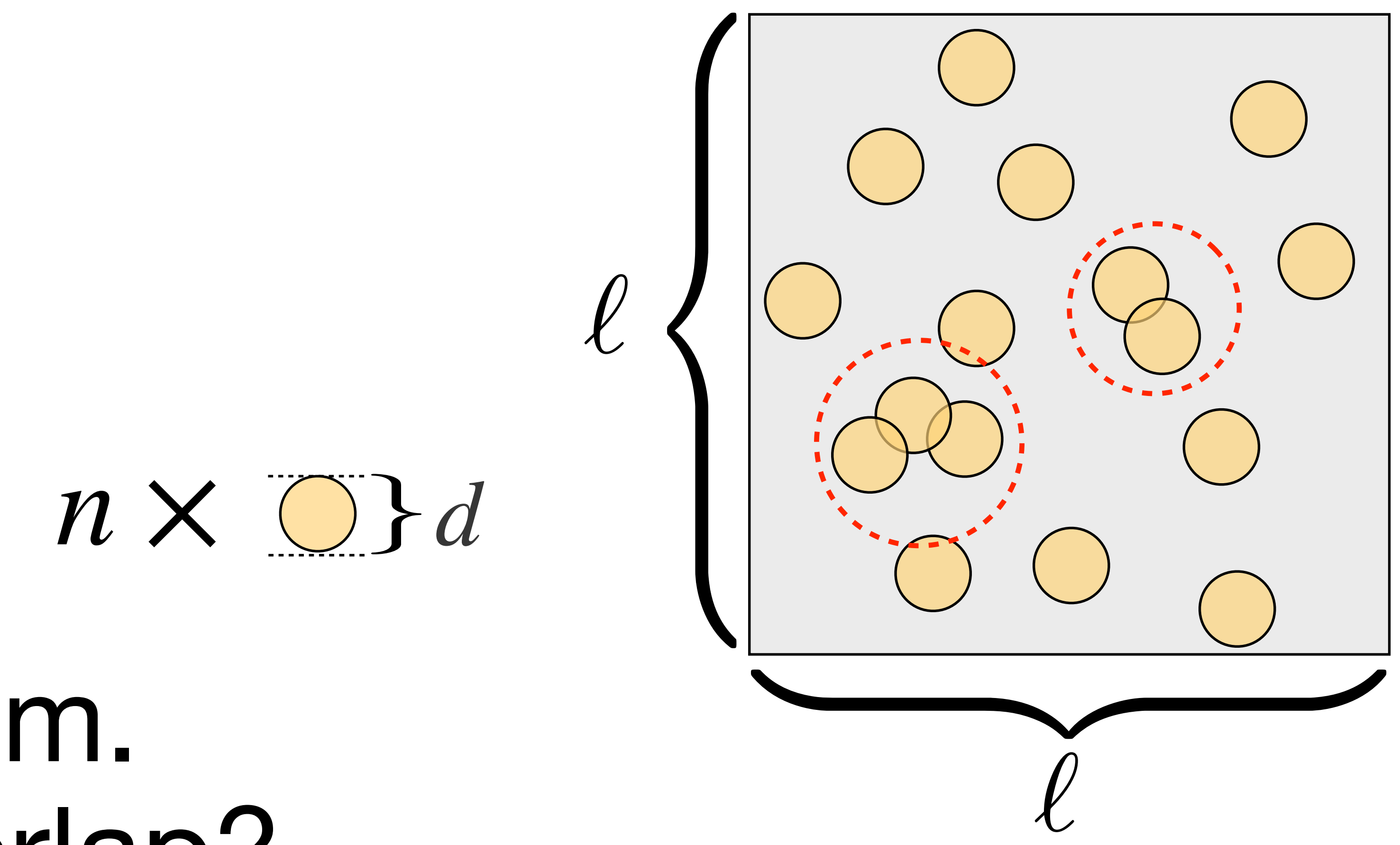
- **Computational Phase Transition of Sampling**
  - Critical phenomenon for sampling *pairwise* interacting variables
  - Computational phase transition for *higher-order* interactions (*sampling Lovász local lemma*)
- **New Paradigm for Computational Sampling**
  - Parallelism of computational sampling
  - Marginal (modular) sampling
  - Dynamic sampling

Outline

- **Computational Phase Transition of Sampling**
  - Critical phenomenon for sampling *pairwise* interacting variables
  - Computational phase transition for *higher-order* interactions (*sampling Lovász local lemma*)
- **New Paradigm for Computational Sampling**
  - Parallelism of computational sampling
  - Marginal (modular) sampling
  - Dynamic sampling

Pennies on a Carpet

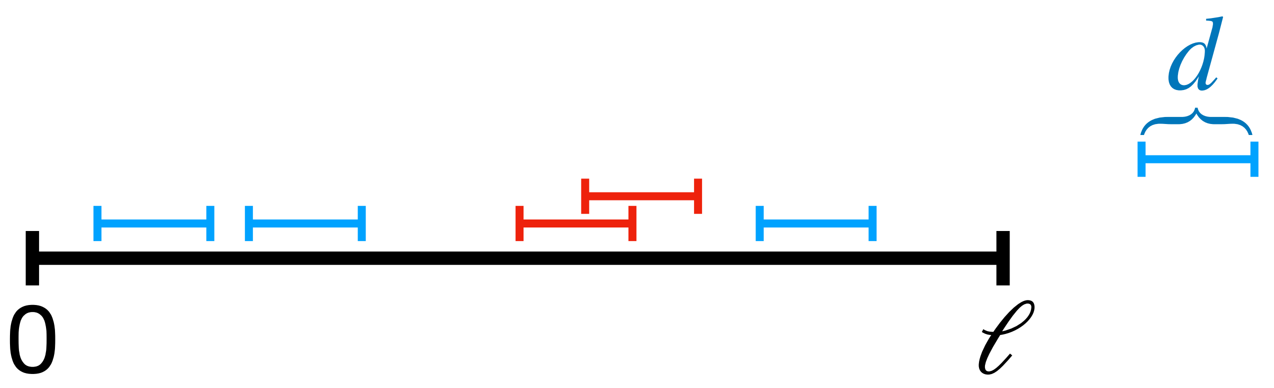
(hard spheres in 2D square)



- Drop n pennies on a square-shape carpet at random. What is the probability that no two pennies will overlap?

- In **1-dimension** (n needles on a line segment):

$$\begin{cases} \left(\dfrac{\ell - nd}{\ell - d}\right)^n & \text{if } \ell \geq nd \\ 0 & \text{otherwise} \end{cases}$$



- In **2-dimension**: Nothing is known about this problem. as of 1979~98.

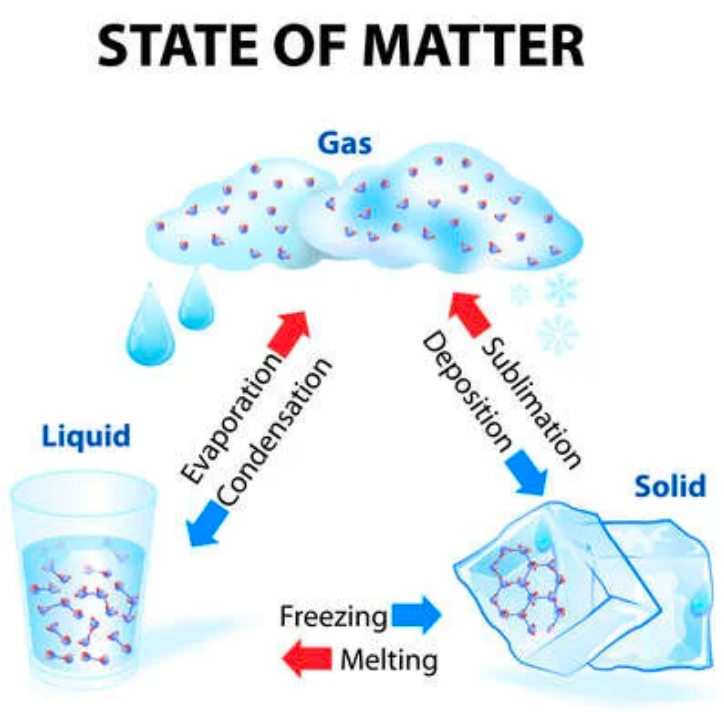
Hard spheres model: This problem is one of the most important problems of statistical mechanics. If we could answer it we would know, for example, why water boils at 100°C, on the basis of purely atomic computations.

Jerrum-Guo (2021): Monte Carlo algorithm for simulating 2D hard spheres

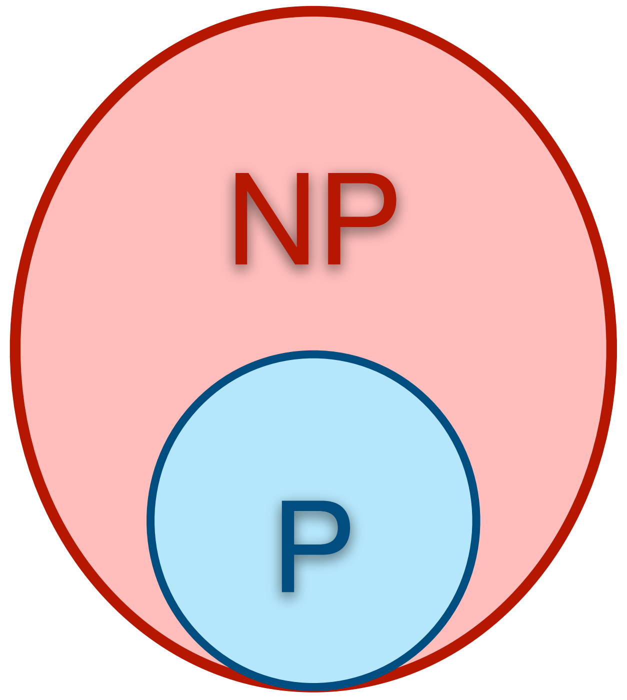
Gian-Carlo Rota
(MIT 18.313)

Computational Phase Transition

Physical
Phase Transition



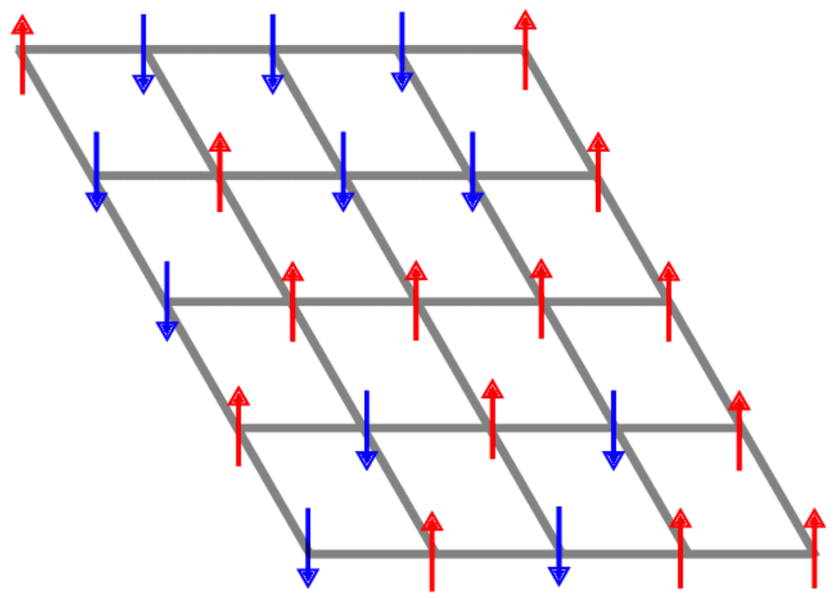
For sampling from
Gibbs measure



Computational
Complexity

- **Boltzmann distribution (Gibbs measure):**

$$\mu(X) \propto \prod_{(f,S)\in\mathscr{C}} f(X_S)$$

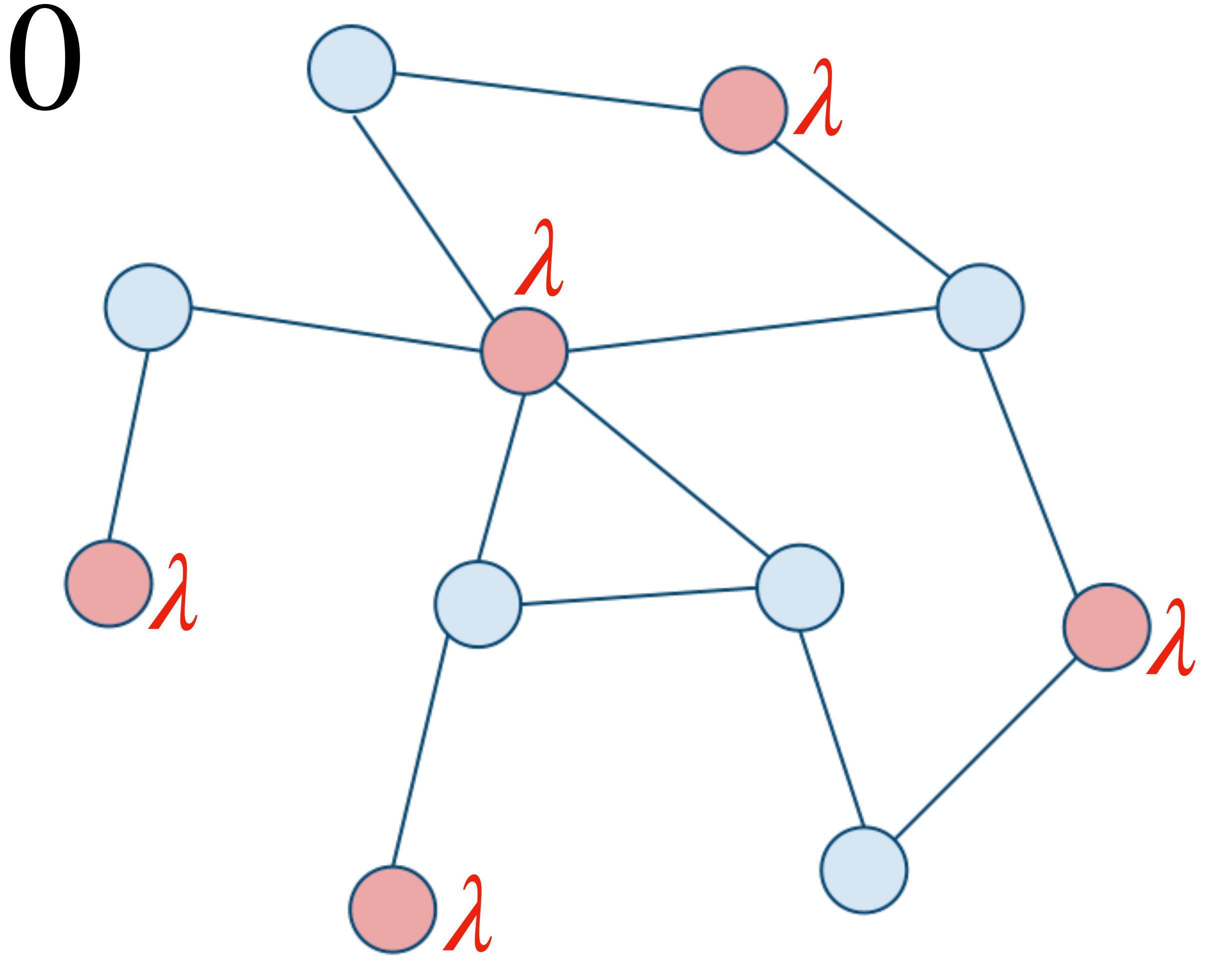


- **locally constrained random variables $\iff$ locally interacting particle states**

- Continuous change of strength of local interaction $\Longrightarrow$ sharp transition of global state (state of matter / computational complexity)

Hardcore Model (*Weighted Independent Set*)

- Sampling graph **independent set** with vertex weight $\lambda > 0$ (**hardcore lattice gas model** with fugacity $\lambda > 0$)



- In graph $G = (V, E)$ of **maximum degree** Δ:

$\mu(I) \propto \lambda^{|I|}$ for independent set I in G

- **Critical threshold** (for phase transition of *uniqueness of Gibbs measure on Δ-degree Bethe lattice*):

$$\lambda_c(\Delta) \triangleq \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{\mathrm{e}}{\Delta-2}$$

- **Computational phase transition** conjecture [Dyer-Frieze-Jerrum, FOCS '99]:

$$\text{Sampling } I \sim \mu \text{ is } \begin{cases} \textbf{NP}\text{-hard} & \text{if } \lambda > \lambda_c(\Delta) \\ \text{poly-time} & \text{if } \lambda < \lambda_c(\Delta) \end{cases}$$

[Sly, FOCS '10 best paper]

Hardcore Sampler

Condition	Time	
$\lambda \le \frac{1}{\Delta - 1}$	$O(n \log n)$	[Bubley-Dyer, FOCS '97]
$\lambda \le \frac{2}{\Delta - 2}$	$O(n \log n)$	[Luby-Vigoda, STOC '97]
$\lambda < \lambda_c(\Delta)$ non-Monte-Carlo	$n^{O(\log \Delta)}$	[Weitz, STOC '06]
$\lambda < \lambda_c(\Delta)$ girth ≥ 7, large Δ	$O(n \log n)$	[Efthymiou-Hayes-Štefankovič-Vigoda-**Y.**, FOCS '16]
$\lambda \le (1-\delta)\lambda_c(\Delta)$	$n^{\exp(O(1/\delta))}$	[Anari-Liu-Oveis Gharan, FOCS '20]
$\lambda \le (1-\delta)\lambda_c(\Delta)$	$n^{O(1/\delta)}$	[Chen-Liu-Vigoda, FOCS '20]
$\lambda < \lambda_c(\Delta)$	$\Delta^{O(\Delta^2)} n \log n$	[Chen-Liu-Vigoda, STOC '21]
$\lambda < \lambda_c(\Delta)$	$O(n^2 \log n)$	[Chen-Feng-**Y.**-Zhang, FOCS '21]
$\lambda < \lambda_c(\Delta)$ *balanced* random walk	$O(n \log n)$	[Anari-Jain-Koehler-Pham-Vuong, STOC '22]
Ind. work { $\lambda < \lambda_c(\Delta)$	$O(n \log n)$	[Chen-Eldan, FOCS '22]
Ind. work { $\lambda < \lambda_c(\Delta)$	$O(n \log n)$	[Chen-Feng-**Y.**-Zhang, FOCS '22]

The Markov chain: (Gibbs sampler)

Starting from $I = \emptyset$, at each step:

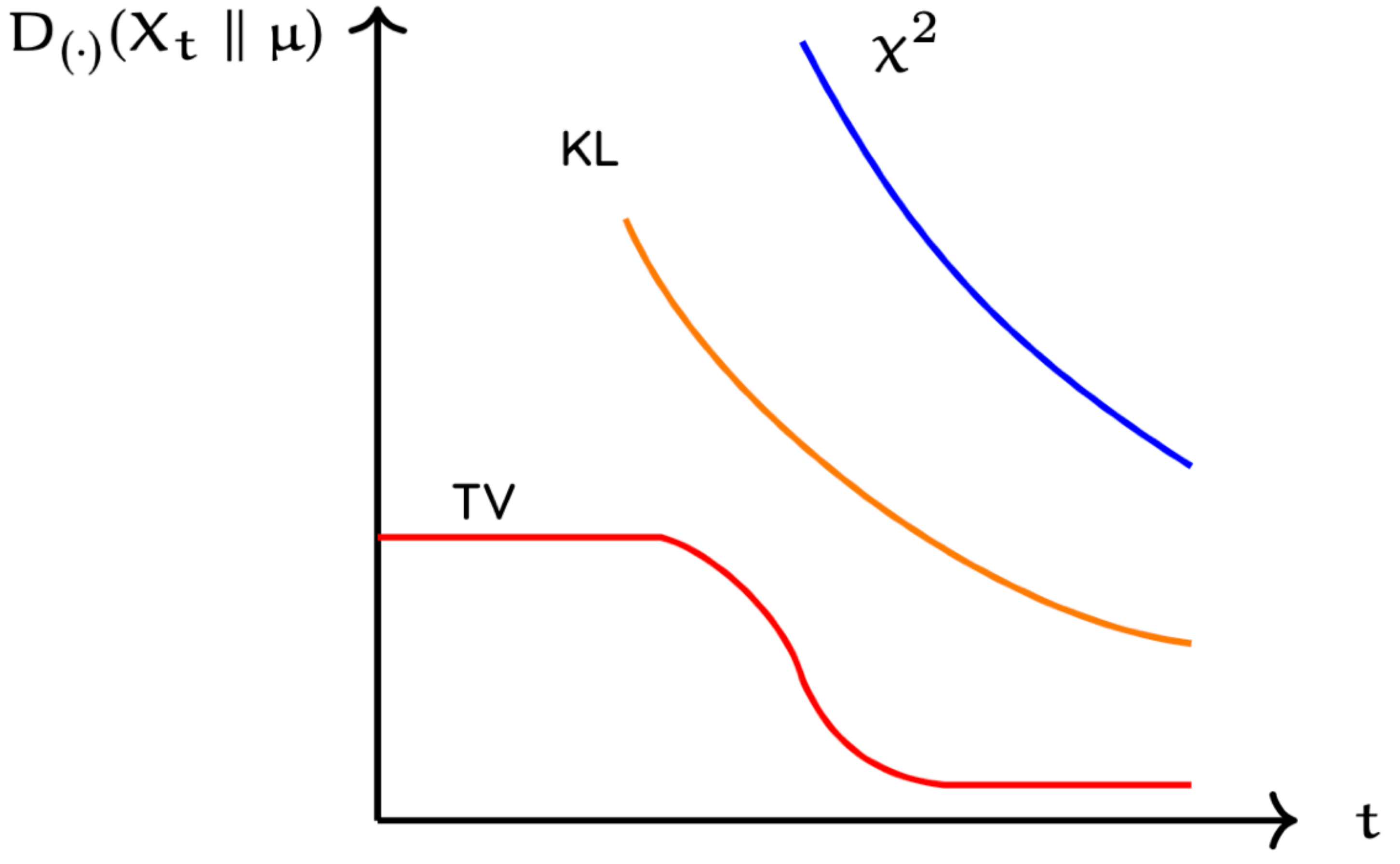
- pick a uniform $v \in V$ at random;
- if $I \cup \{v\}$ is an independent set

$$I \leftarrow \begin{cases} I \cup \{v\} & \text{with prob. } \frac{\lambda}{1+\lambda} \\ I \setminus \{v\} & \text{with prob. } \frac{1}{1+\lambda} \end{cases}$$

strong spatial mixing (SSM)
spectral independence
entropic independence
high-dimensional expander (HDX)
local-to-global argument
modified log-Sobolev inequality
field dynamics
... ...

$$D_f := \mathop{\mathbb{E}}_{\sigma\sim\mu}\left[f\left(\frac{\nu(\sigma)}{\mu(\sigma)}\right)\right]$$

$$D_{\chi^2} : f(x) = (x-1)^2$$
$$D_{\mathrm{KL}} : f(x) = x\log x$$



- **Poincaré inequality:** for Poincaré constant $\kappa_{\mathrm{Poin}} < 1$

$$D_{\chi^2}(X^{(t)} \parallel \mu) \le \kappa_{\mathrm{Poin}} \cdot D_{\chi^2}(X^{(t-1)} \parallel \mu) \implies \tau_{\mathrm{mix}} = O(n^2 \log n)$$

- **Modified log-Sobolev (MLS) inequality:** for MLS constant $\kappa_{\mathrm{MLS}} < 1$

$$D_{\mathrm{KL}}(X^{(t)} \parallel \mu) \le \kappa_{\mathrm{MLS}} \cdot D_{\mathrm{KL}}(X^{(t-1)} \parallel \mu) \implies \tau_{\mathrm{mix}} = O(n \log n)$$

- [Erbar-Henderson-Menz-Tetali '16] proved a *subcritical* MLS inequality via Ricci curvature
- [Weitz '06] proved a decay of correlation property up to critical threshold, which was used in [Anari-Liu-Oveis Gharan '20] [Chen-Liu-Vigoda '20] to imply the mixing of Glauber dynamics by a local-to-global argument in high dimension expanders, which was further refined in [Chen-Liu-Vigoda '21] to prove the fast mixing of $\Theta(n)$-block dynamics
- [Chen-Feng-**Y.**-Zhang '21, '22] invented the field dynamics, which used the mixing of block dynamics to lift the subcritical MLS inequality in [EHMT '16] to the critical threshold

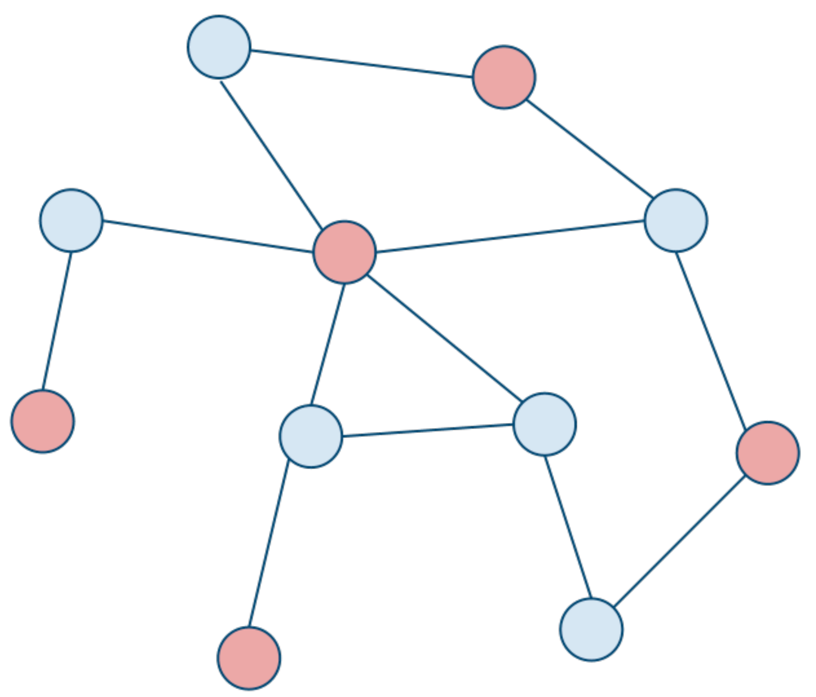
Local-to-Global Argument

Random Walk Algorithm:
(Glauber Dynamics)

Starting from $I = \varnothing$, at each step:

- pick a uniform $v \in V$ at random;
- if $I \cup \{v\}$ is an independent set

$$I \leftarrow \begin{cases} I \cup \{v\} & \text{with prob. } \frac{\lambda}{1+\lambda} \\ I \backslash \{v\} & \text{with prob. } \frac{1}{1+\lambda} \end{cases}$$



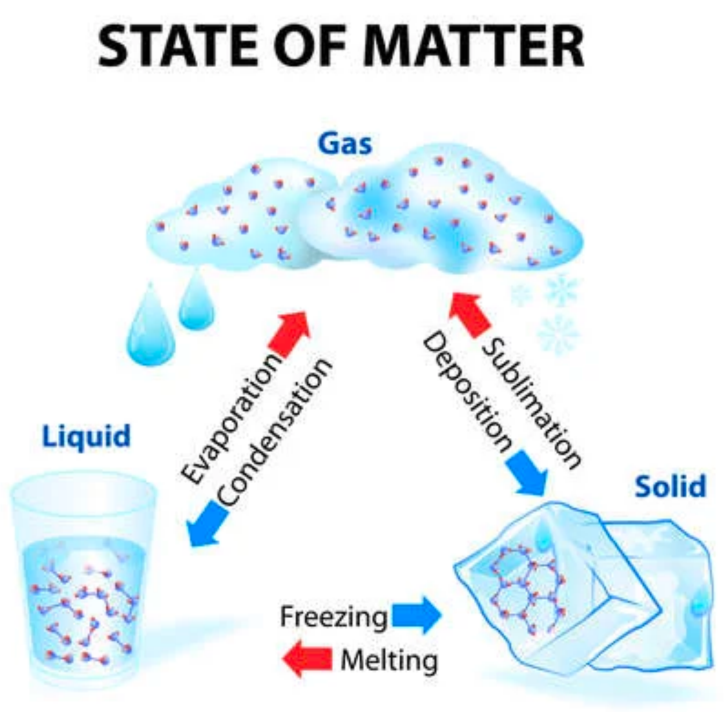
- The transition matrix P for the random walk has size $\exp(\Omega(n)) \times \exp(\Omega(n))$
- The correlation matrix $\mathrm{Corr}(i,j) \triangleq \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}}$ for μ has size $n \times n$
- We prove (essentially) [Chen-Feng-Y.-Zhang, FOCS '21 FOCS '22]:

$$\|\mathrm{Corr}\| = O(1) \Longrightarrow 1 - \kappa_{\{\text{Poin,MLS}\}}(P_\lambda) \geq \Omega\left(1 - \kappa_{\{\text{Poin,MLS}\}}(P_{\lambda/100})\right)$$

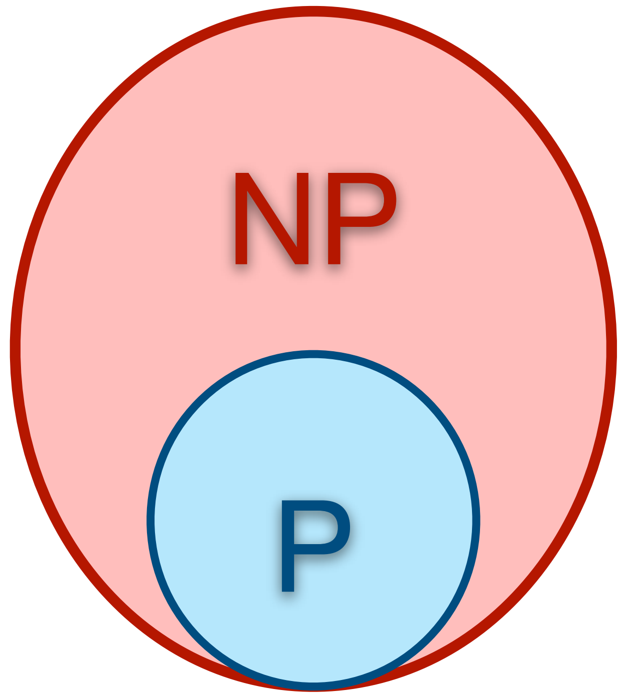
- When $\lambda < \lambda_c(\Delta)$, there is decay of correlation, then $\|\mathrm{Corr}\| = O(1)$, and therefore the Poincare/MLS constant for subcritical case $\lambda/100$ can be lifted to the near critical regime with constant overhead

Computational Phase Transition

Physical
Phase Transition



For sampling from
Gibbs measure



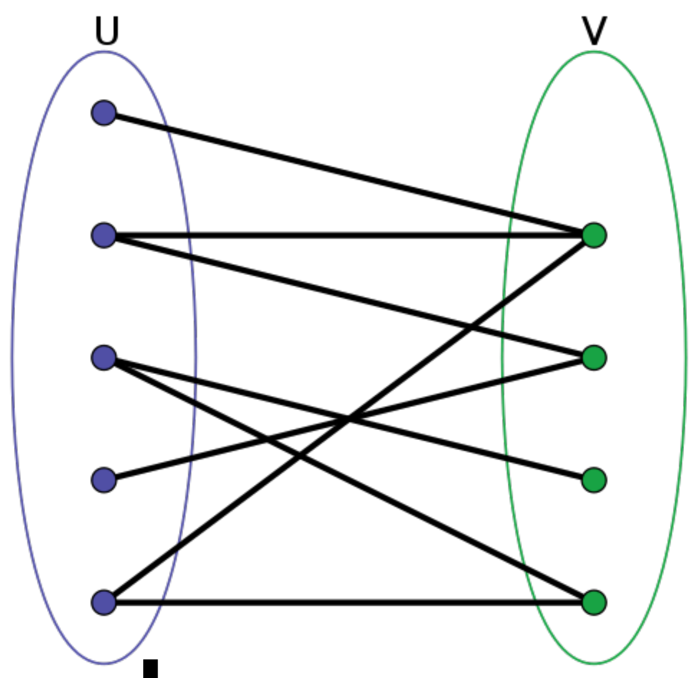
Computational
Complexity

- [Chen-Feng-Y.-Zhang, FOCS '21, '22]:
  For **pairwise** ***negatively*** constrained Boolean variables $X = (X_1, \ldots, X_n) \sim \mu$:
  (*anti-ferromagnetic Ising model / anti-ferromagnetic two-state spin systems*)

$$\text{Sampling } X \sim \mu \text{ is } \begin{cases} \text{poly-time} & \text{within physical phase-transition cond.} \\ \mathbf{NP}\text{-hard} & \text{beyond physical phase-transition cond.} \end{cases}$$

- [Jerrum-Sinclair, '92] (**Gödel Prize** 1996): Sampling **pairwise** ***positively*** constrained Boolean variables (*ferromagnetic Ising model*) in poly-time

Bipartite Hardcore Model (#BIS)



- Sampling **independent set** with vertex weight $\lambda > 0$ in a **bipartite graph**.
- In bipartite graph $G = (U, V, E)$:

$$\mu(I) \propto \lambda^{|I|} \text{ for independent set } I \text{ in } G$$

- **#BIS** (bipartite independent set): sampling independent set in bipartite graph
  - Many sampling/approximate counting problems are #BIS-equivalent: subclasses of #CSP, ferromagnetic spin systems, stable matchings, …
  - The computational complexity of #BIS is still unknown.
- [Chen-Liu-Y., FOCS '23]: For bipartite hardcore model with **one-side** maximum degree Δ, sampling is poly-time tractable if $\lambda < \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$

Higher-Order (k-wise) Interactions

- For k-**wise** interactions, consider **hard constraints** $f : [q]^e \to \{0,1\}$

$$\mu(\sigma) \propto \prod_{e \in E} f_e(\sigma_e)$$

  μ is the uniform distribution over all **constraint satisfaction solutions**

- **Example:** k-CNF (conjunctive normal form)

$$\Phi = (x_1 \vee \neg x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee x_4) \wedge (x_3 \vee \neg x_4 \vee \neg x_5)$$

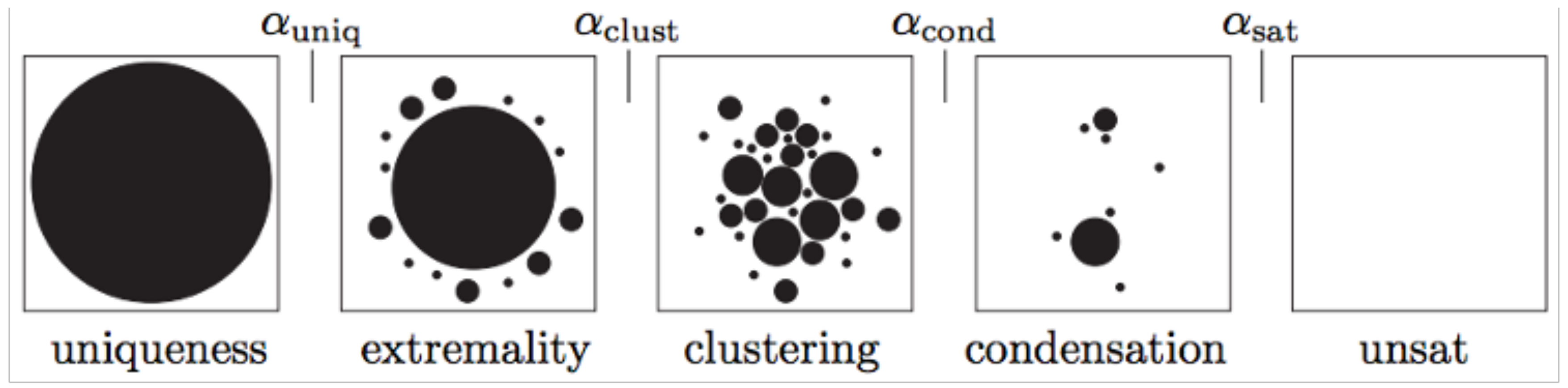
- **SAT** **(Boolean satisfiability)**: determine whether there is a satisfying solution
  - **NP-complete** (Cook–Levin theorem)
  - Sample from μ (**SAT sampler**) $\implies$ **SAT solver**

Solving vs Sampling

- For **Constraint Satisfaction Problem** (CSP):

$$\mu(\sigma) \propto \prod_{e \in E} f_e(\sigma_e)$$

- **Lovász local lemma** (1975): $pD \lesssim 1 \implies$ a SAT solution exists
  - $p = \max_e \Pr_{X \sim [q]^e} [f_e(X) = 0]$: max *violation probability* of any constraint
  - $D = \max_e |\{e' : e \cap e' \neq \varnothing\}|$: max *dependency degree* of any constraint
- **Connectivity of solution space:**
  - The solution space can be highly disconnected



- **Barrier**: MCMC sampling crucially relies on ***connectivity*** of solution space

Overcome the Connectivity Barrier

Projected Markov chain:

Properly construct a subspace $U \subseteq V$;

Sample $X_U \sim \mu_U$ by **simulating** Gibbs sampler on μ_U;

Recover from X_U a satisfying solution $X \sim \mu$;

Idea: project onto lower dimension to improve connectivity

- **Fast sampler in near-linear time** (under *Lovász local lemma like conditions* $pD^{O(1)} \lesssim 1$):
  - SAT [Feng-Guo-Y.-Zhang, STOC 2020, JACM 2021] **projected MCMC**
  - CSP with *atomic* constraints [Feng-He-Y., STOC 2021] **"*compressed*" MCMC**
  - *general* CSP (*constraint satisfaction problem*) [He-Wang-Y., FOCS 2022] ***new algorithm***
- "*Sampling Lovász local lemma*" conjecture: "*sampling is twice-local*"

$$pD^2 \lesssim 1 \implies \text{sampling uniform satisfying solution is poly-time}$$

  - **state-of-the-arts:** $pD^5 \lesssim 1$ *"5-fold local lemma"* [He-Wang-Y., FOCS 2022, SODA 2023]

Outline

- **Computational Phase Transition of Sampling**
  - Critical phenomenon for sampling *pairwise* interacting variables
  - Computational phase transition for *higher-order* interactions (*sampling Lovász local lemma*)
- **New Paradigm for Computational Sampling**
  - Parallelism of computational sampling
  - Marginal (modular) sampling
  - Dynamic sampling

Markov chain Monte Carlo (MCMC)

Gibbs sampler [Geman-Geman 1984] for sampling X from Gibbs distribution μ:

The **Markov chain** maintains an $X \in [q]^V$, at each step:

- pick $v \in V$ uniformly at random;
- update the evaluation of X_v according to its marginal distribution $\mu_v(\cdot \mid X_{N(v)})$.

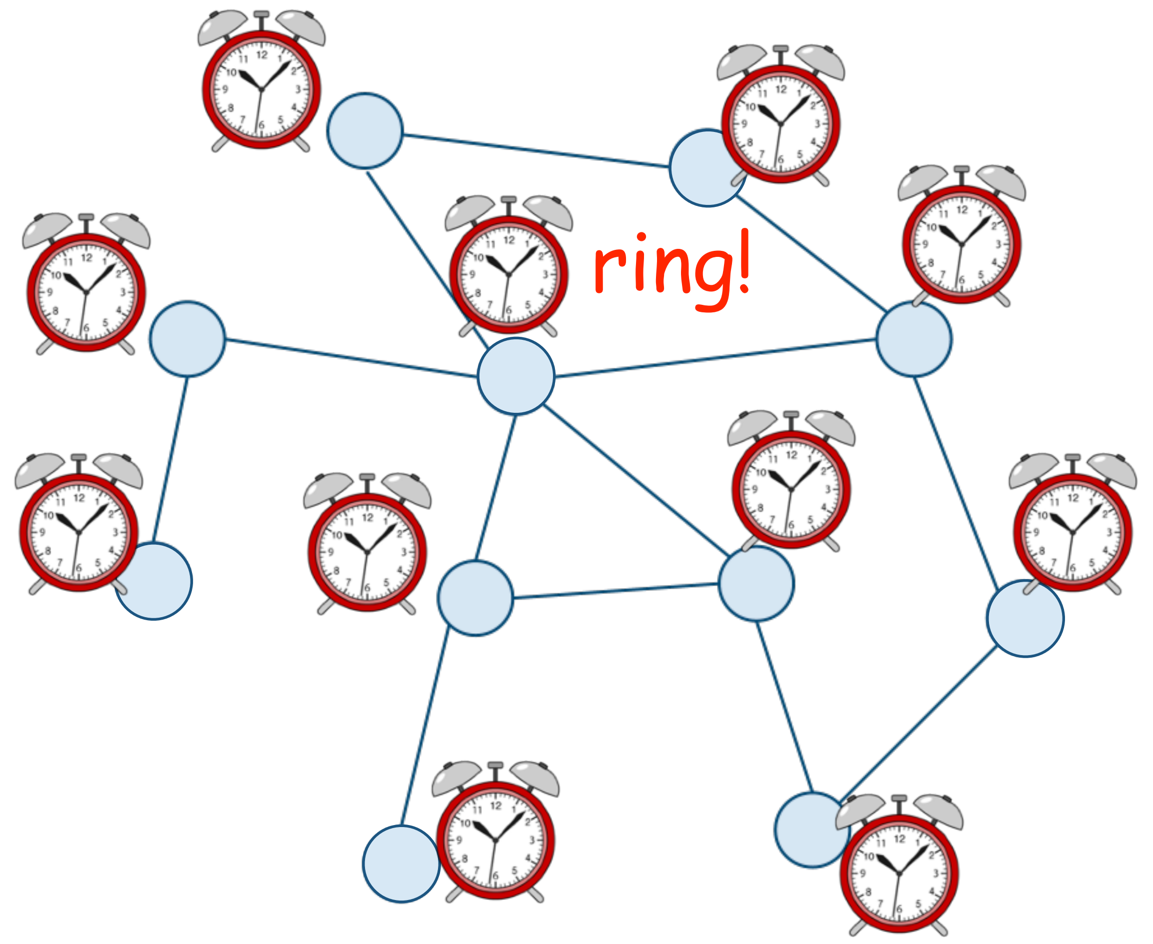
- **Sequential** algorithm that updates a **single-site** at each step.
- Generic lower bound [Hayes-Sinclair '13]:
  - any **single-site dynamics** requires at least $\Omega(n \log n)$ steps to converge

Idealized Parallel Process

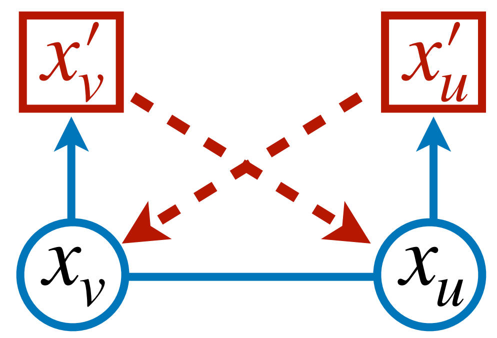
Glauber dynamics [Glauber 1963] for sampling X from Gibbs distribution μ:

The **Markov chain** for $X \in [q]^V$ runs in continuous time:

- Each $v \in V$ holds a rate-1 **Poisson clock**;
- upon v's clock rings: its state X_v is updated atomically according to $\mu_v(\cdot \mid X_{N(v)})$.



- **Idealized** (*continuity* and *atomicity*) **parallel** process for the evolution of real physical world
- **Barrier for concurrency**: updates of adjacent sites
  - $O(\Delta)$ **overhead!**
- **Fully & correctly parallelize MCMC?**



Parallelism of MCMC Sampling

(Fundamental challenge in Theory of Parallel Computing)

- In a seminal paper for parallel computing [Mulmuley-Vazirani-Vazirani, STOC '87]:
  - "*It is possible to sample uniform perfect matching in NC (poly-log rounds)?*"
  - "*It is possible to estimate permanent in NC (parallel counterpart for P)?*"
- Later, [Shang-Hua Teng, 1992]
  - **Proved**: classical MCMC sampling could not be efficiently parallelized "*via standard approaches of parallelization*"
  - **Conjectured**: permanent estimation cannot be done in poly-log rounds
    (perhaps the only problem not known to be P-complete but conjectured intrinsically sequential)

Fully Parallelize MCMC Sampling

Continuous-time Glauber dynamics (1963):

Each $v \in V$ holds a Poisson clock;

upon v's clock rings:

- X_v is updated according to $\mu_v(\cdot \mid x_{N(v)})$;

Algorithm 1: An iterative algorithm for simulating single-site dynamics

Input: initial configuration $X_0 \in Q^V$; update schedule $\mathfrak{T} = (t_i^v)_{v \in V, 0 \le i \le m_v}$;
assignment $\mathfrak{R} = (\mathcal{R}_{(v,i)})_{v \in V, 1 \le i \le m_v}$ of random bits for resolving updates.

initialize $\ell \leftarrow 0$ and $\widehat{X}_v^{(0)}[i] \leftarrow X_0(v)$ for all $v \in V, 0 \le i \le m_v$;
repeat
  $\ell \leftarrow \ell + 1$;
  forall $v \in V$ **in parallel do** $\widehat{X}_v^{(\ell)}[0] \leftarrow X_0(v)$;
  forall *updates* (v,i), *where* $v \in V, 1 \le i \le m_v$, **in parallel do**
    let $\tau \in Q^{N_v^+}$ be constructed as:
      $\forall u \in N_v^+, \tau_u \leftarrow \widehat{X}_u^{(\ell-1)}[j_u]$ for $j_u = \max\{j \ge 0 \mid t_j^u < t_i^v\}$;
    $\widehat{X}_v^{(\ell)}[i] \leftarrow \text{Sample}\left(P_v^\tau, \mathcal{R}_{(v,i)}\right)$;
  end
until $\widehat{X}^{(\ell)} = \widehat{X}^{(\ell-1)}$;

- Suppose that all random choices have been generated:
  - time t_i^v and a *random seed* $R_{(v,i)} \in [0,1]$ for the ith update at $v \in V$
- Dynamical system for the Markov chain:

$$X_t(v) \leftarrow \text{Sample}\left(\mu_v^\tau, R_{(v,i)}\right)$$

where $\tau \in [q]^{N(v)}$ satisfies $\forall u \in N(v)$, $\tau_u = X_{t_j^u}(u)$ for $t_j^u = \max\{t_{j'}^u : t_{j'}^u < t_i^v\}$

Fully Parallelize MCMC Sampling

Continuous-time Glauber dynamics (1963):

Each $v \in V$ holds a Poisson clock;

upon v's clock rings:

- X_v is updated according to $\mu_v(\cdot \mid x_{N(v)})$;

Algorithm 1: An iterative algorithm for simulating single-site dynamics

Input: initial configuration $X_0 \in Q^V$; update schedule $\mathfrak{T} = (t_i^v)_{v \in V, 0 \le i \le m_v}$; assignment $\mathfrak{R} = (\mathcal{R}_{(v,i)})_{v \in V, 1 \le i \le m_v}$ of random bits for resolving updates.

```
initialize ℓ ← 0 and X̂_v^(0)[i] ← X_0(v) for all v ∈ V, 0 ≤ i ≤ m_v;
repeat
    ℓ ← ℓ + 1;
    forall v ∈ V in parallel do X̂_v^(ℓ)[0] ← X_0(v);
    forall updates (v, i), where v ∈ V, 1 ≤ i ≤ m_v, in parallel do
        let τ ∈ Q^{N_v^+} be constructed as:
            ∀u ∈ N_v^+, τ_u ← X̂_u^(ℓ−1)[j_u] for j_u = max{j ≥ 0 | t_j^u < t_i^v};
        X̂_v^(ℓ)[i] ← Sample(P_v^τ, R_(v,i));
    end
until X̂^(ℓ) = X̂^(ℓ−1);
```

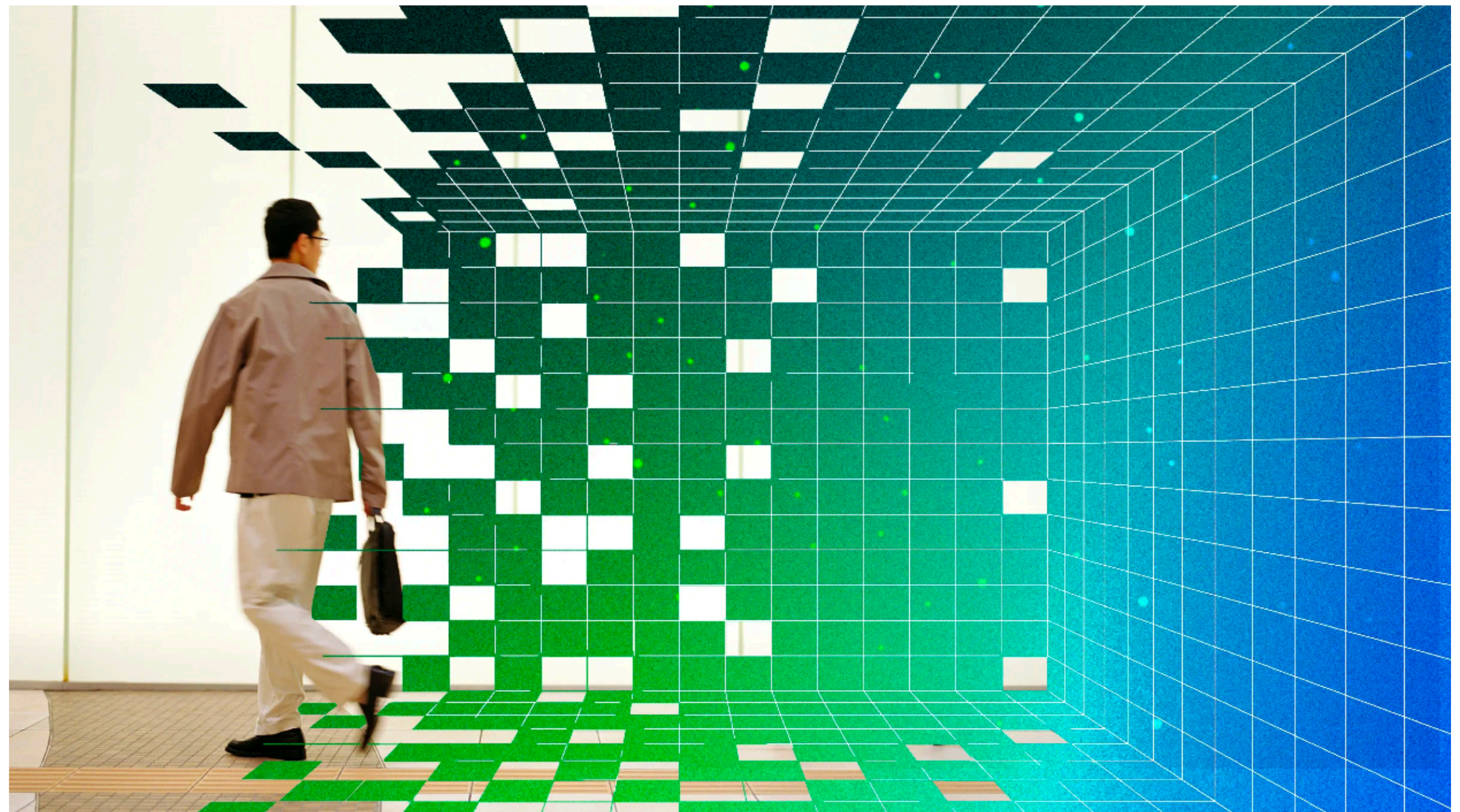
- **Key ideas:**
  - Construct a dynamical system whose fixpoint corresponds to the correct evolution of the chain.
  - Simulate this dynamical system by a locally-iterative message-passing parallel algorithm.
  - A universal coupling of randomness to ensure fast stabilization to the correct fixpoint.
- **Faithful parallel simulation of MCMC with no overhead [Liu-Y., STOC 2022]**
  (when the Dobrushin's influence matrix is O(1)-normed)

Local Evaluating Random Vector

- Evaluate a few X_v's in $\boldsymbol{X} \in [q]^V$ drawn from a Gibbs distribution μ
  (statistical inference / estimation in selected dimensions)

- **Classic MCMC:** have to compute everything even just interested in 1 variable,
  because all variables may be correlated

- **Marginal (modular) sampling:** with some local computational cost
  evaluate X_v in $\boldsymbol{X} \in [q]^V$ drawn from a Gibbs distribution μ



Classic MCMC: *Forward Simulation*

- **MCMC since 1940s:** simulate a dynamical system till it converges to a fixpoint

$$\text{arbitrary initial state } \boldsymbol{X}^{(0)} \xrightarrow{\text{long enough evolution}} \boldsymbol{X}^* \sim \mu \text{ (fixpoint)}$$

The **Markov chain** maintains an $X \in [q]^V$, at each step:

- pick $v \in V$ uniformly at random;
- update the evaluation of X_v according to its marginal distribution $\mu_v(\cdot \mid X_{N(v)})$.

- For **marginal sampler** (which evaluates X_v^*):

It seems necessary to faithfully simulate everything…?

Marginal Sampler via *Backward Deduction*

- Imagine an idealized Glauber dynamics:

$$X^{(-\infty)} \to \cdots \to X^{(-2)} \to X^{(-1)} \to X^{(0)} \sim \mu$$

- Evaluate the $X^{(0)}(v) \Longrightarrow$ sample from μ_v

$\text{pred}_u(t)$: last time u is updated before t

- [Feng-Guo-Wang-Wang-Y., FOCS 2023]: with grand couplings, this terminates within $O(1)$ cost in expectation, and $O(\log n)$ cost with high probability

Dynamic Sampling

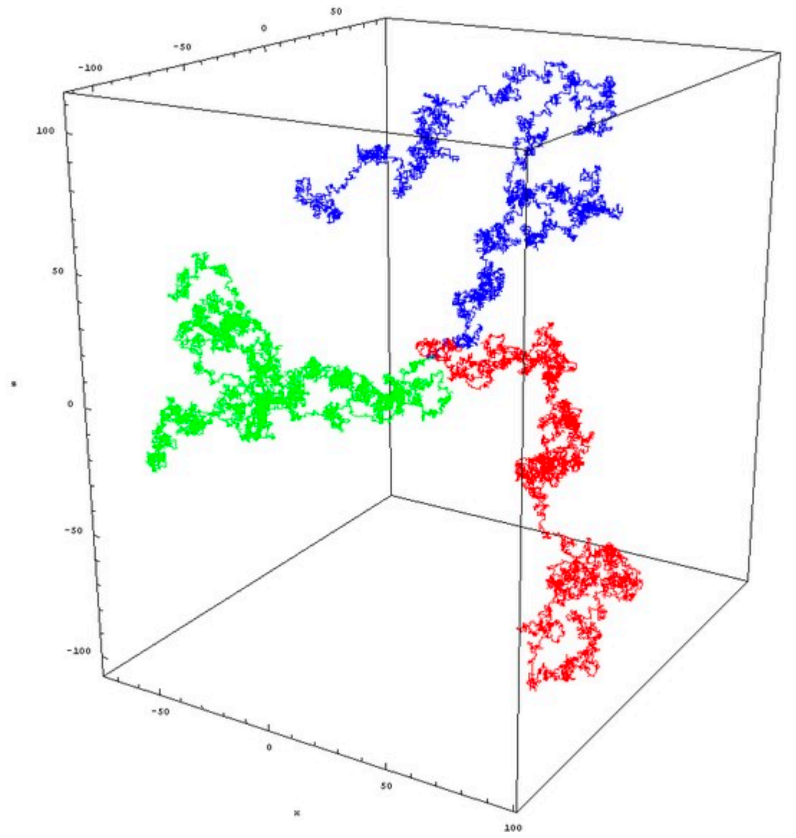
Dynamic Sampling problem: for dynamically changing distributions $\mu \to \mu'$

$X \sim \mu$ → $X' \sim \mu'$

dynamic update

with incremental cost

- Sampling/inference tasks on dynamically changing data:
  - Online data, data streams, network environment, etc.
- Dynamically changing graphical models generated in:
  - Locally-iterative algorithms for learning.
  - Self-reduction procedure in approximate counting.



Classic random walks fail on dynamic data

- **Algorithmic Lipschitz**: transform $X \sim \mu$ to $X' \sim \mu'$ with cost proportional to diff(μ, μ')

Dynamic Sampling

Dynamic Sampling problem: for dynamically changing distributions $\mu \to \mu'$

$\boldsymbol{X} \sim \mu$ —dynamic update / with incremental cost→ $\boldsymbol{X}' \sim \mu'$

Algorithm 1: Dynamic Sampler

Input : a graphical model $\mathcal{I}$ and a random sample $\boldsymbol{X} \sim \mu_{\mathcal{I}}$;
Update: an update (D, Φ_D) which modifies $\mathcal{I}$ to $\mathcal{I}'$;
Output: a random sample $\boldsymbol{X} \sim \mu_{\mathcal{I}'}$;

```
R ← vbl(D);
while R ≠ ∅ do
   (X, R) ← Local-Resample(I′, X, R);
return X;
```

Algorithm 2: Local-Resample($\mathcal{I}$, $\boldsymbol{X}$, $\mathcal{R}$)

Input : a graphical model $\mathcal{I} = (V, E, [q], \Phi)$, a configuration $\boldsymbol{X} \in [q]^V$ and a $\mathcal{R} \subseteq V$;
Output: a new pair $(\boldsymbol{X}', \mathcal{R}')$ of configuration $\boldsymbol{X}' \in [q]^V$ and subset $\mathcal{R}' \subseteq V$;

- for each $e \in E^+(\mathcal{R})$, in parallel, compute $\kappa_e \triangleq \frac{1}{\phi_e(X_e)} \min_{x \in [q]^e : x_{e \cap \mathcal{R}} = X_{e \cap \mathcal{R}}} \phi_e(x)$;
- for each $v \in \mathcal{R}$, in parallel, resample $X_v \in [q]$ independently according to distribution ϕ_v;
- for each $e \in E^+(\mathcal{R})$, in parallel, sample $F_e \in \{0,1\}$ ind. with $\Pr[F_e = 0] = \kappa_e \cdot \phi_e(X_e)$;
- $\boldsymbol{X}' \leftarrow \boldsymbol{X}$ and $\mathcal{R}' \leftarrow \bigcup_{e \in E : F_e = 1} e$;
- **return** $(\boldsymbol{X}', \mathcal{R}')$.

- A **dynamic sampling** algorithm: [Feng-Vishnoi-Y., STOC '19]
  - correct and efficient on *dynamic* data
  - *parallel*, *distributed*, *communication-efficient*
  - *Las Vegas* algorithm for *perfect* sampling

Summary

- **Computational phase transition of sampling**
- **Parallel, marginal (modular), dynamic sampling**
- **Theme of future work:**
  - a unified and critical theory for sampling and solving
  - classify efficient computing through efficient sampling
    - **Turing** (1936): "*What is computation?*"
      by investigating Hilbert's *Entscheidungsproblem*
    - In 2023: "*What is efficient (Monte Carlo) computation?*"
      by classifying: "*What distributions are easy to sample?*"

ENIAC

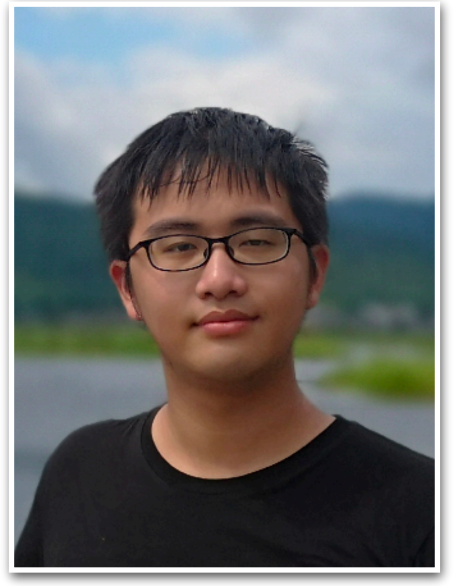
陈小羽
南京大学

凤维明
UC Berkeley

郭珩
Edinburgh

何昆
人民大学

刘景铖
南京大学



刘弘洋
南京大学

王淳扬
南京大学

王家恒
Edinburgh

张驰豪
上海交通大学

张昕渊
南京大学

- [Chen-Liu-**Y.** ’23]: Uniqueness and rapid mixing in the bipartite hardcore model. **FOCS** ’23.
- [Feng-Guo-Wang-Wang-**Y.** ’23]: Towards derandomising Markov chain Monte Carlo. **FOCS** ’23.
- [Chen-Feng-**Y.**-Zhang ’22]: Optimal mixing for two-state anti-ferromagnetic spin systems. **FOCS** ’22.
- [He-Wang-**Y.** ’22]: Sampling Lovász local lemma for general constraint satisfaction solutions in near-linear time. **FOCS** ’22.
- [Liu-**Y.** ’22]: Simple parallel algorithms for single-site dynamics. **STOC** ’22.
- [Chen-Feng-**Y.**, Zhang ’21]: Rapid mixing of Glauber dynamics via spectral independence for all degrees. **FOCS** ’21.
- [Feng-He-**Y.** ’21]: Sampling constraint satisfaction solutions in the local lemma regime. **STOC** ’21.
- [Feng-Guo-**Y.**-Zhang ’20]: Fast sampling and counting *k*-SAT solutions in the local lemma regime. **STOC** ’20. **JACM** ’21.
- [Feng-Vishnoi-**Y.** ’19]: Dynamic sampling from graphical models. **STOC** ’19. **SICOMP** ’21.
- [Efthymiou-Hayes-Štefankovič-Vigoda-**Y.** ’16]: Convergence of MCMC and loopy BP in the tree uniqueness region for the hardcore model. **FOCS** ’16. **SICOMP** ’19.

Thank you!

- [Chen-Liu-**Y.** '23]: Uniqueness and rapid mixing in the bipartite hardcore model. **FOCS** '23.
- [Feng-Guo-Wang-Wang-**Y.** '23]: Towards derandomising Markov chain Monte Carlo. **FOCS** '23.
- [Chen-Feng-**Y.**-Zhang '22]: Optimal mixing for two-state anti-ferromagnetic spin systems. **FOCS** '22.
- [He-Wang-**Y.** '22]: Sampling Lovász local lemma for general constraint satisfaction solutions in near-linear time. **FOCS** '22.
- [Liu-**Y.** '22]: Simple parallel algorithms for single-site dynamics. **STOC** '22.
- [Chen-Feng-**Y.**, Zhang '21]: Rapid mixing of Glauber dynamics via spectral independence for all degrees. **FOCS** '21.
- [Feng-He-**Y.** '21]: Sampling constraint satisfaction solutions in the local lemma regime. **STOC** '21.
- [Feng-Guo-**Y.**-Zhang '20]: Fast sampling and counting *k*-SAT solutions in the local lemma regime. **STOC** '20. **JACM** '21.
- [Feng-Vishnoi-**Y.** '19]: Dynamic sampling from graphical models. **STOC** '19. **SICOMP** '21.
- [Efthymiou-Hayes-Štefankovič-Vigoda-**Y.** '16]: Convergence of MCMC and loopy BP in the tree uniqueness region for the hardcore model. **FOCS** '16. **SICOMP** '19.