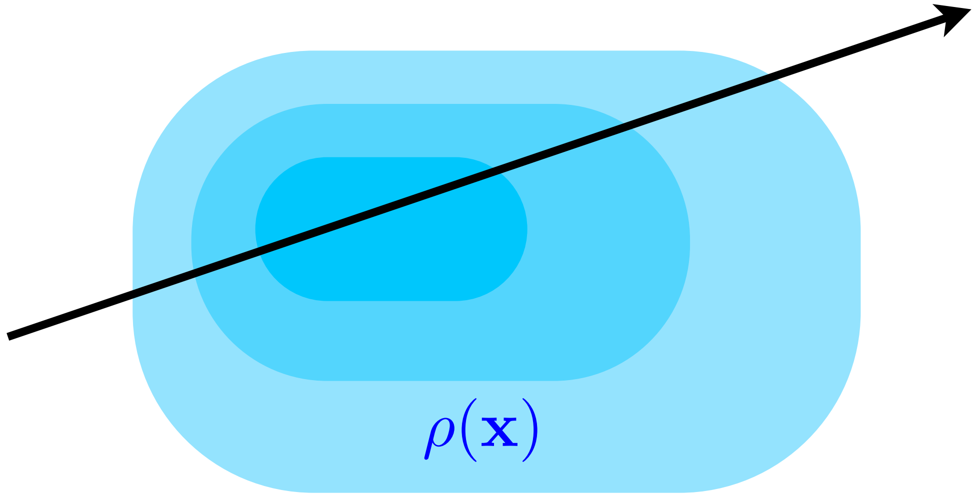
Bayesian Tomography

Tomographie au Laplace

John Skilling (john@skilling.co.uk) Garching, September 2015

An object of density ρ is scanned by line integrals $\int_{\text{line}} \rho(\mathbf{x})\, dx = D$.



Ignore finite width, fan-beam shape, refraction, scattering. Include noise $\pm\sigma$.

Seek to compute $\rho \xleftarrow{\text{infer}} D$.

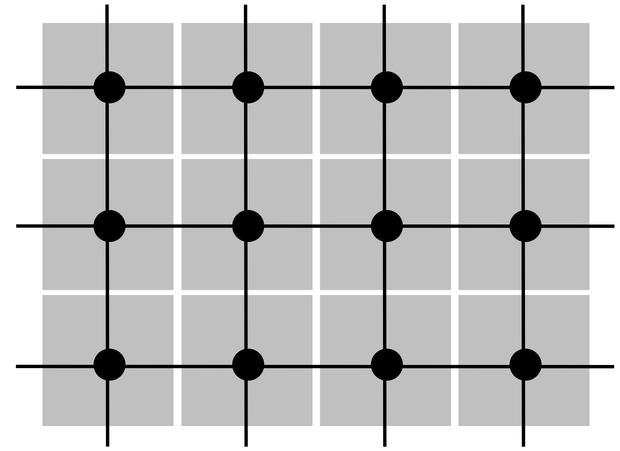
Contents

1. The computing grid (hexagonal)
2. The prior (cartoon pixel colours and bond energies)
3. MCMC requirements (simple moves)
4. The likelihood (chi-not-squared data misfit)
5. Bayes (by nested sampling)
6. Results (display cartoon, posterior samples, mock data)

17 slides, mostly pictures.

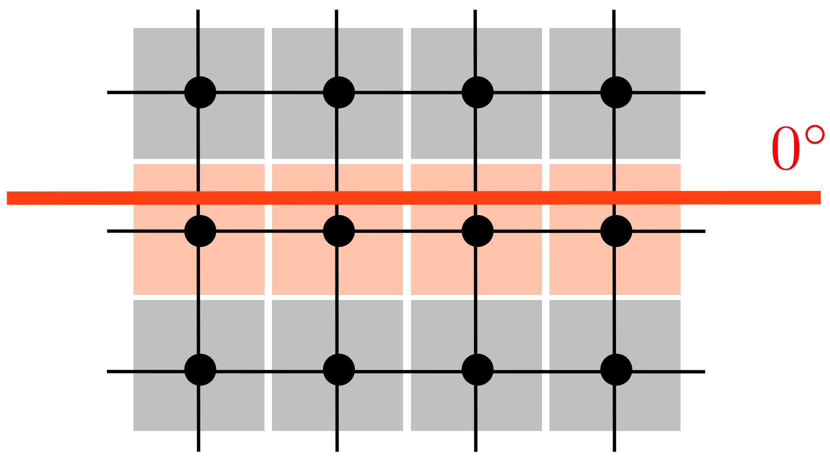
1. The Computing Grid

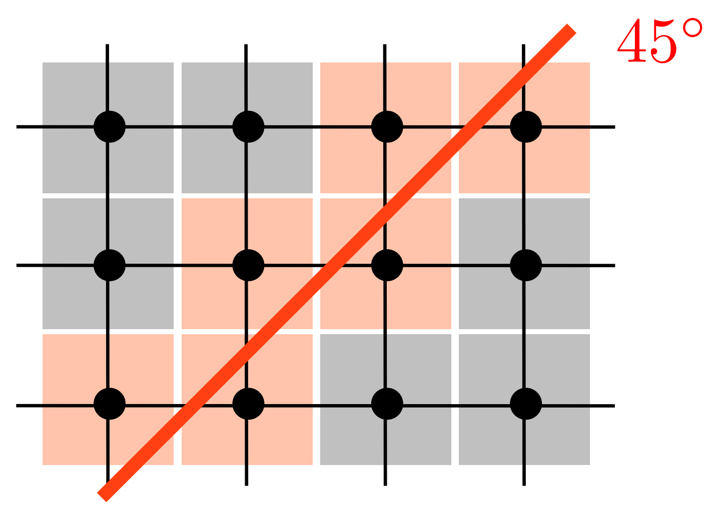
Computation needs finite grid.



Nodes • represent square pixels.
Bonds quantify gradients or discontinuities.

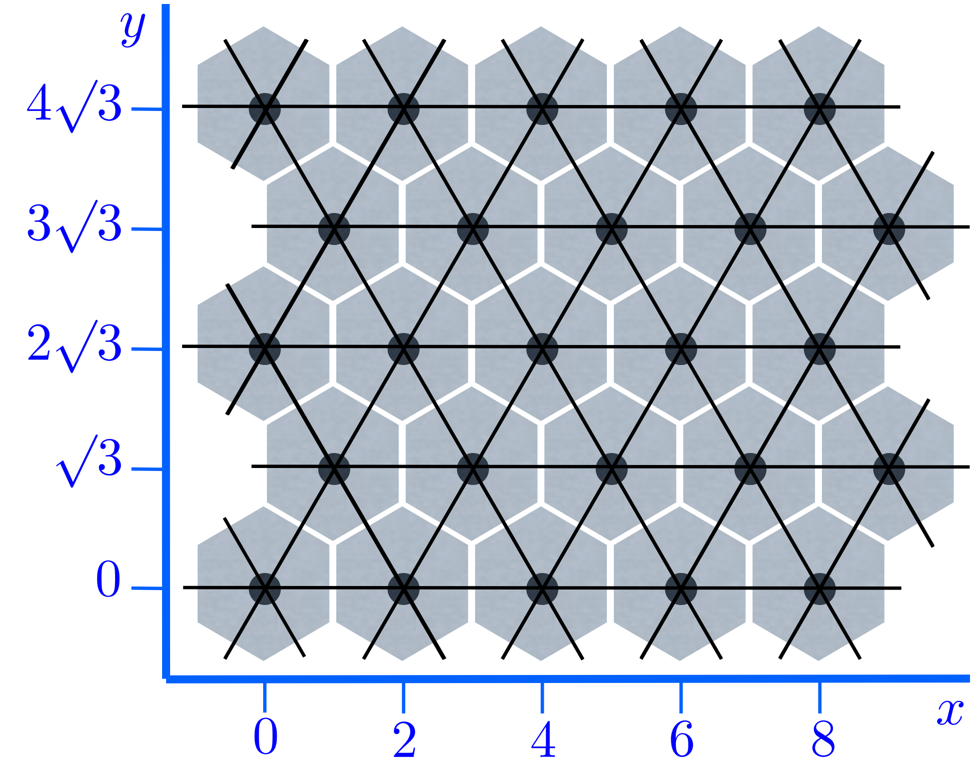
This is badly anisotropic. Pixel density along scan line varies by $\sqrt{2}$.



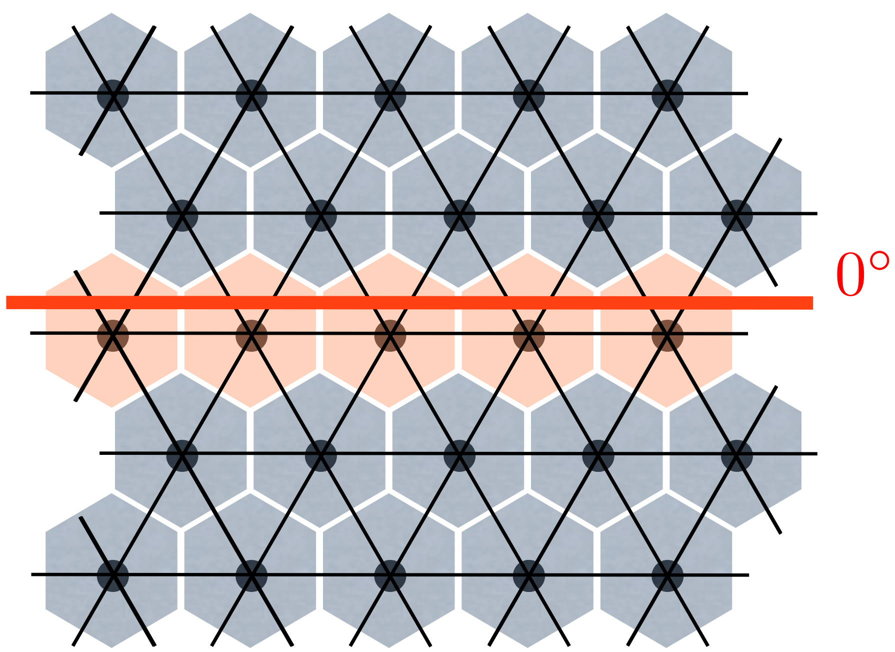


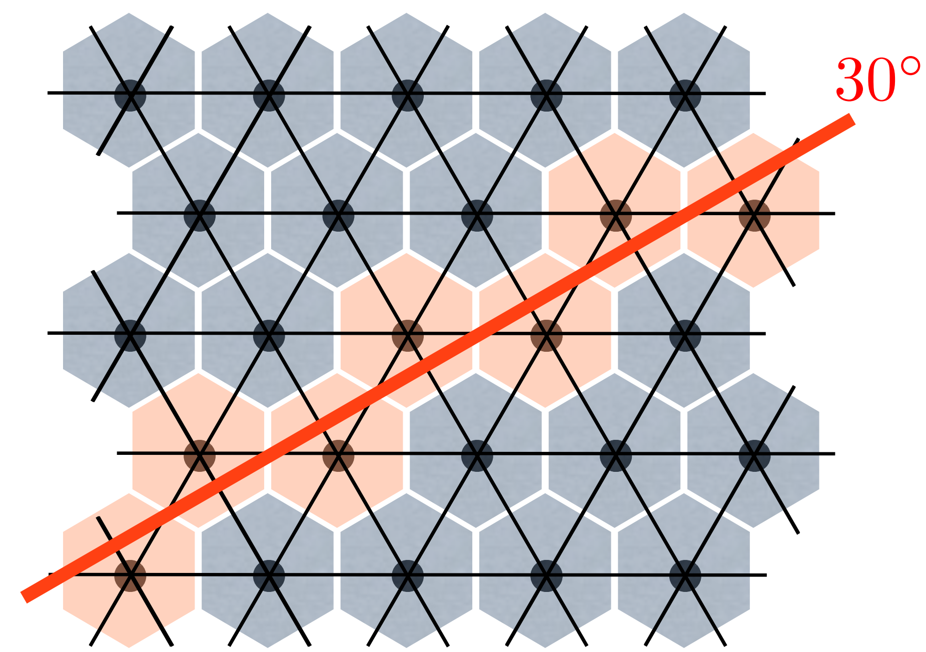
Diagonal edges will be harder to reconstruct. That is bad.

Hexagonal grid (6 neighbours) is better.

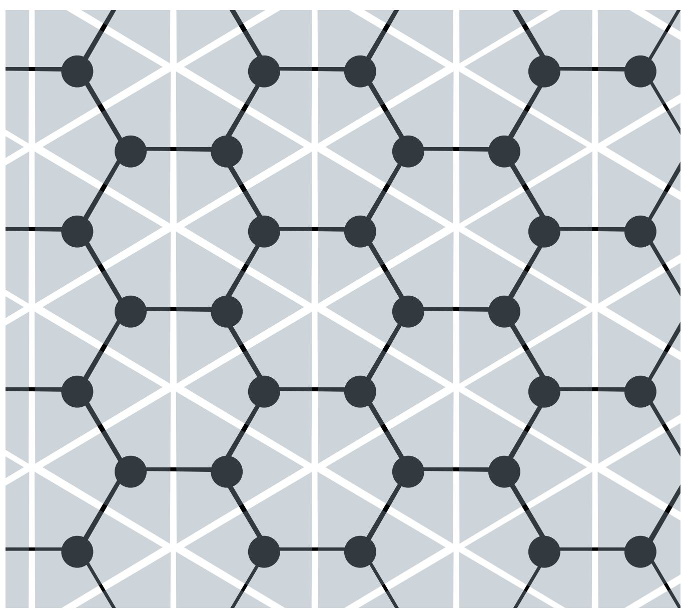


Pixel density along scan lines varies by only $\frac{\sqrt{3}}{2}$ (15% instead of 40%).



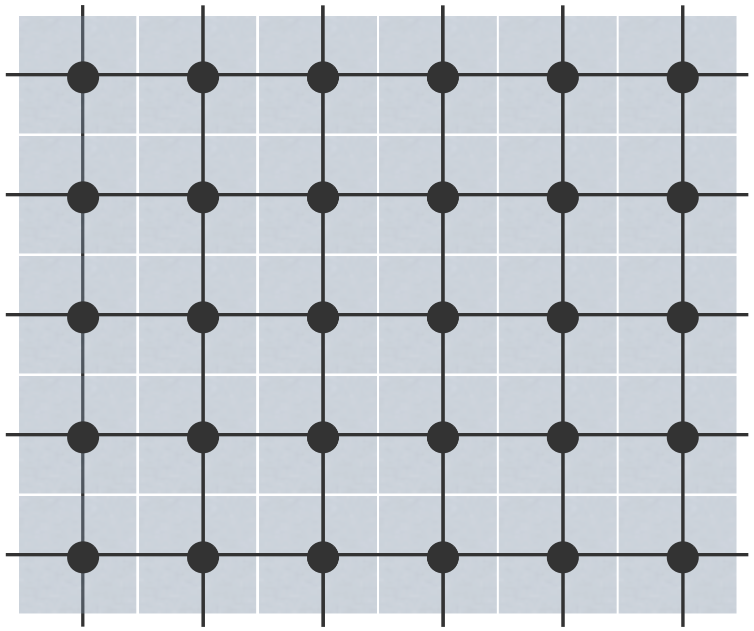


(In 3 dimensions, use 12-neighbour face-centred cubic lattice.)



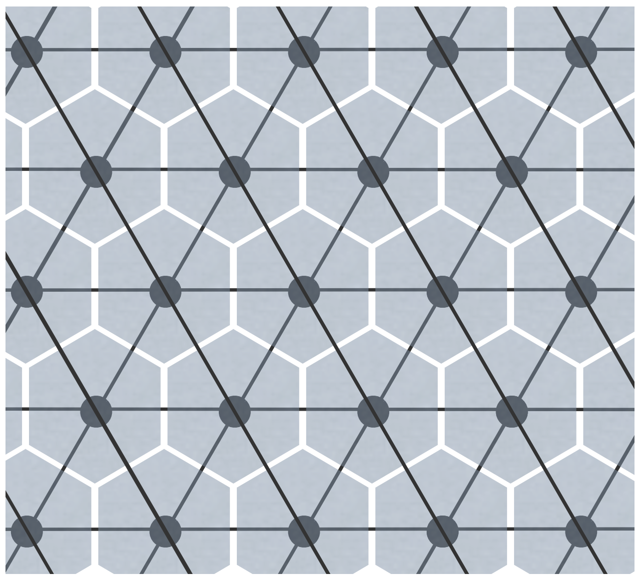
3 neighbours

Bad



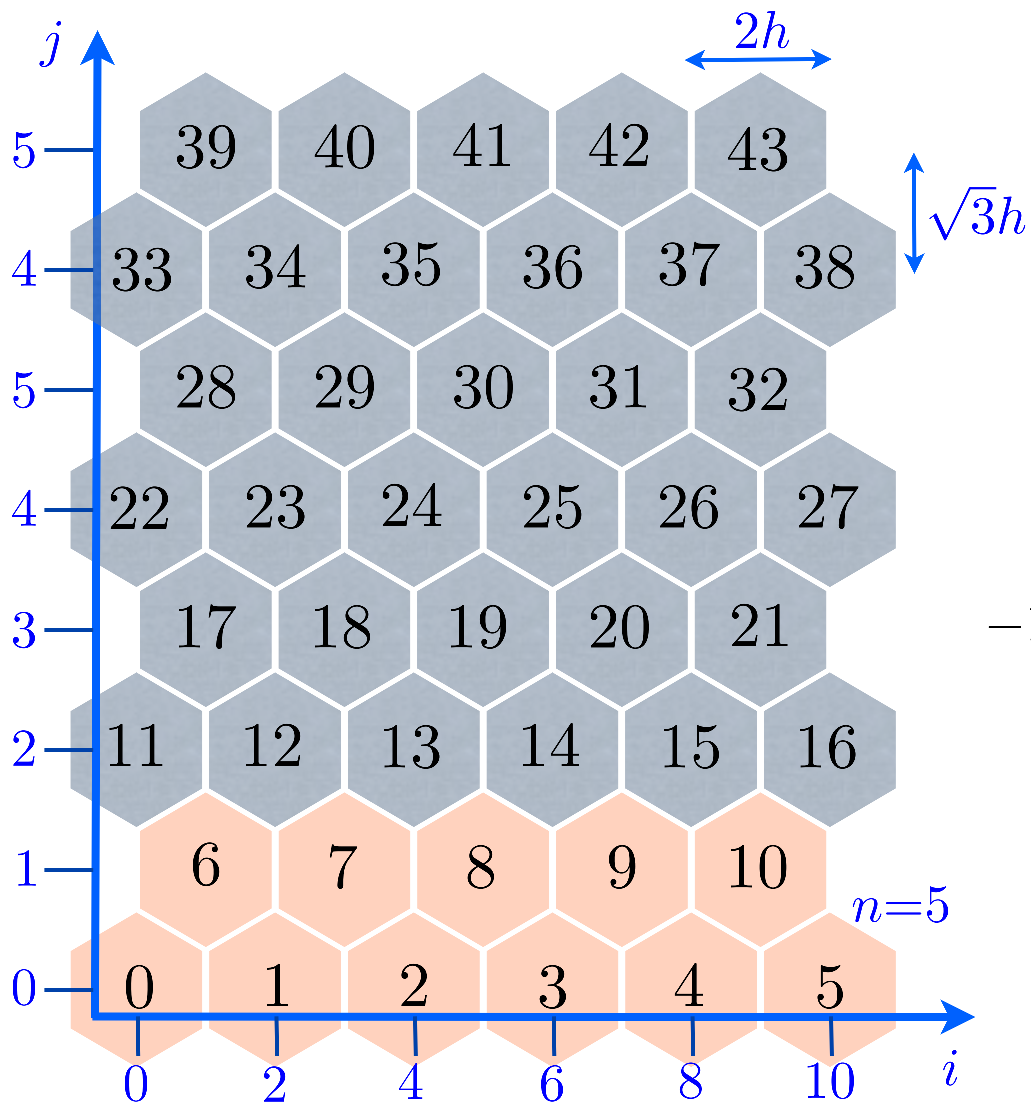
4 neighbours

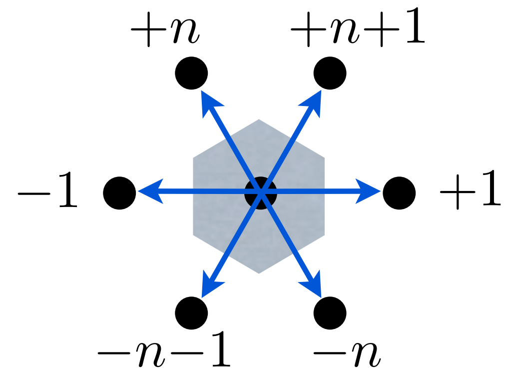
Sad

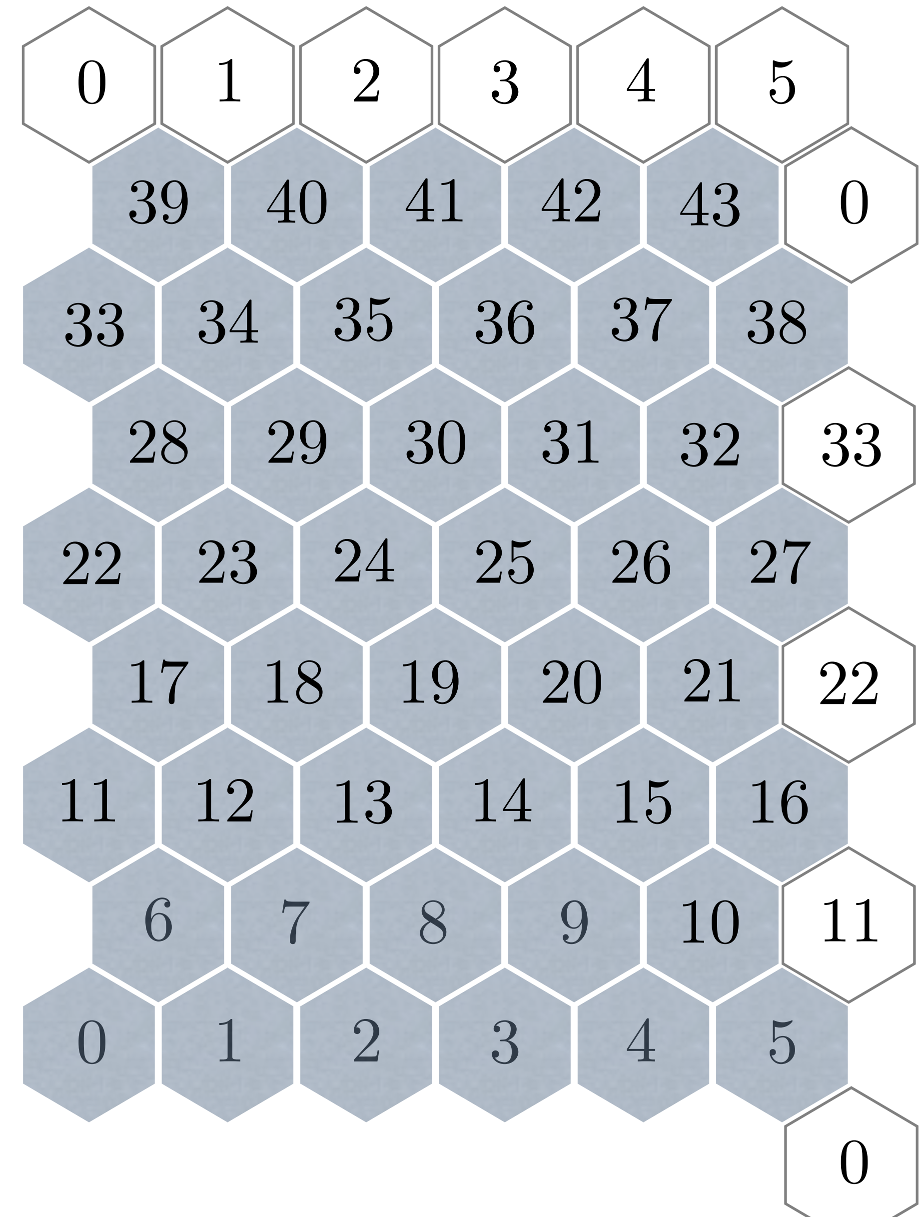


6 neighbours

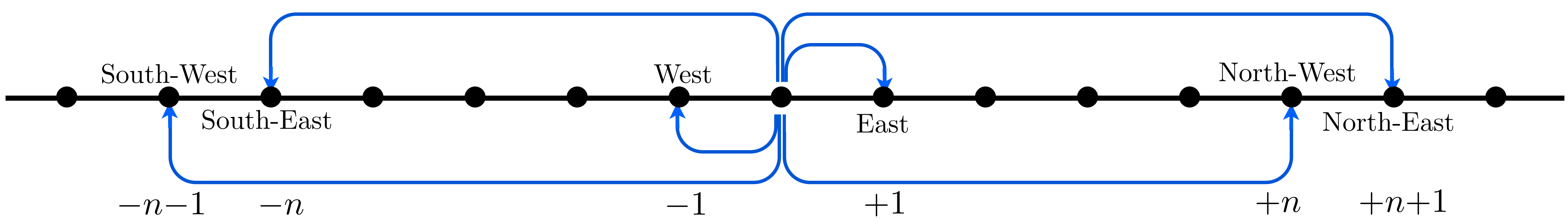
Best



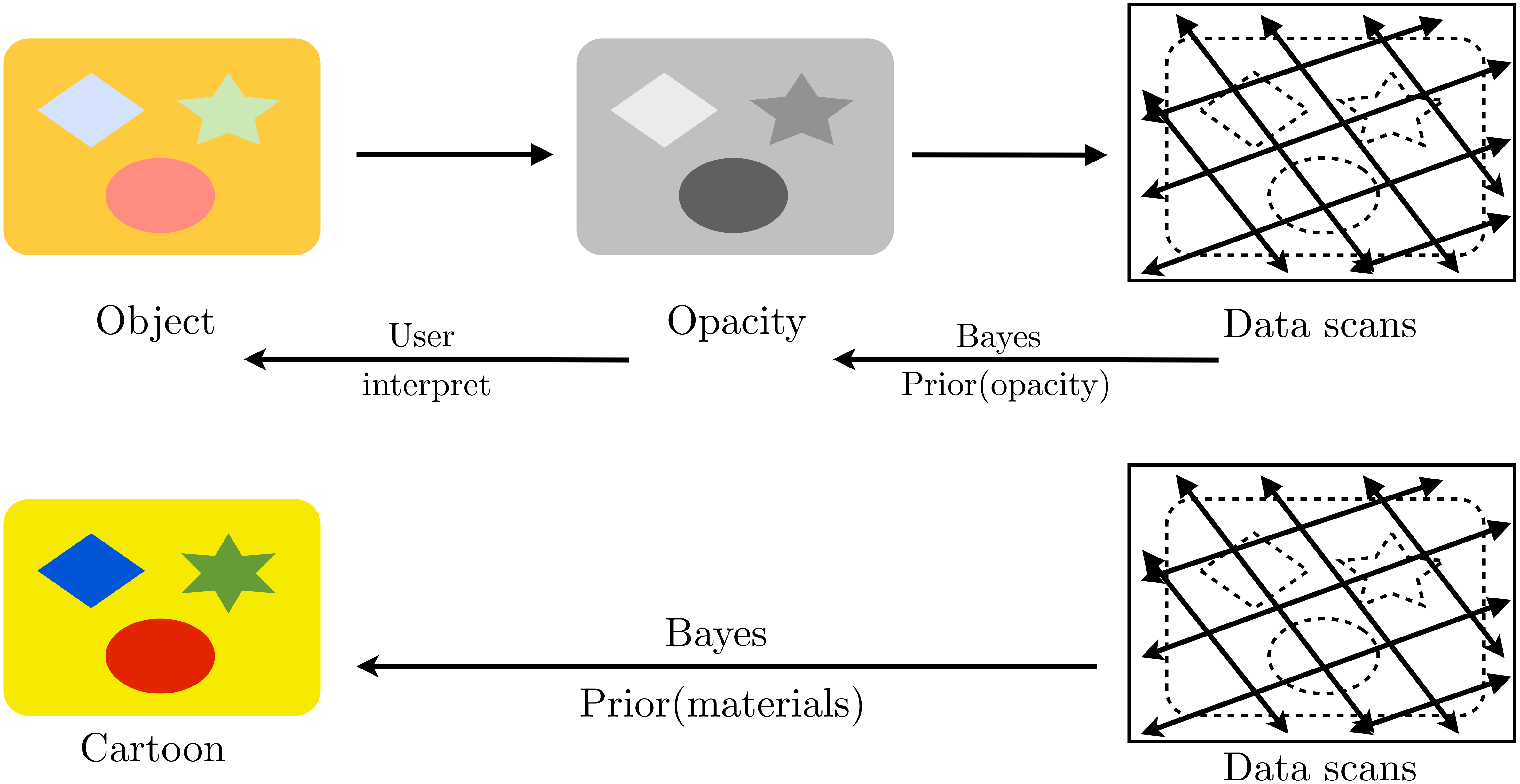




Bond direction is additive constant. Get wraparound periodicity.

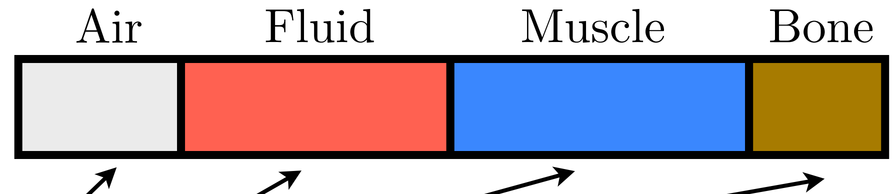


2. The Prior



Cartoons enable **data fusion**, where different properties (X-ray opacity, nuclear spin, sonar reflectivity,. . .) of the same material can be combined into a single multi-technique reconstruction.

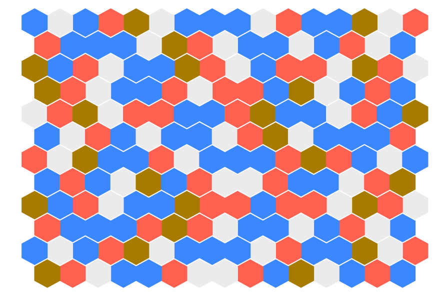
The Prior

Type of material = "colour" 

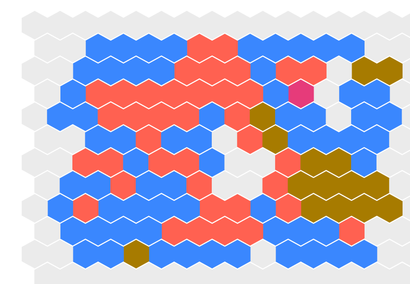
(any number)

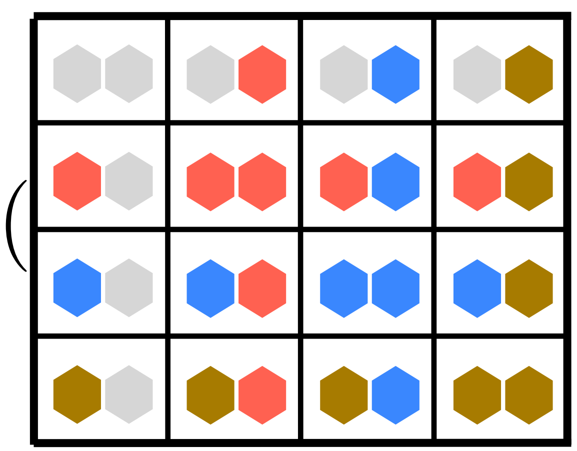
Prob(colour) = expected prior proportions

We do NOT want uncorrelated Prob(colour).
At high resolution, almost surely almost uniform grey.



We DO want coherent patches with reduced boundaries.



Bond Energy () =

		dead		wounded	
	0	9	2	8	Air
	9	0	1	5	Fluid
	2	1	0	1	Muscle
	8	5	1	0	Bone
	Air	Fluid	Muscle	Bone	

$= \dfrac{1}{\text{expected perimeter}}$

$$\text{Prior} = \text{Prob(colours)}\ e^{-\beta\ \text{Energy(bonds)}}$$

(Potts model)

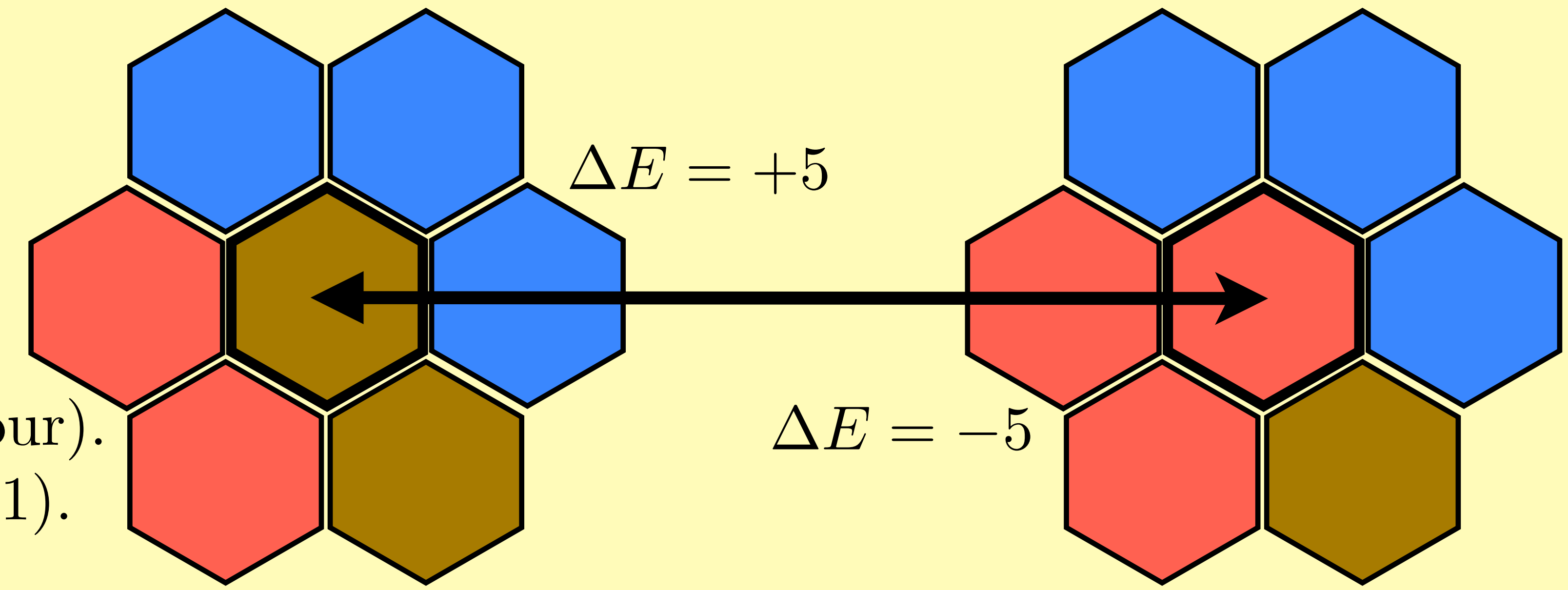
3. MCMC requirements

Start: Sample from prior
Evolve: Move faithfully to prior

Moves will later be modulated by data (rejected by mismatch).
So moves must be **small**, **sympathetic to data**, and **reversible**.

Try individual colour changes.

Select random trial cell.
Select trial colour $\sim$ Prob(colour).
Accept if $e^{-\beta\Delta E} \geq$ Uniform(0,1).



Could also exchange colours.

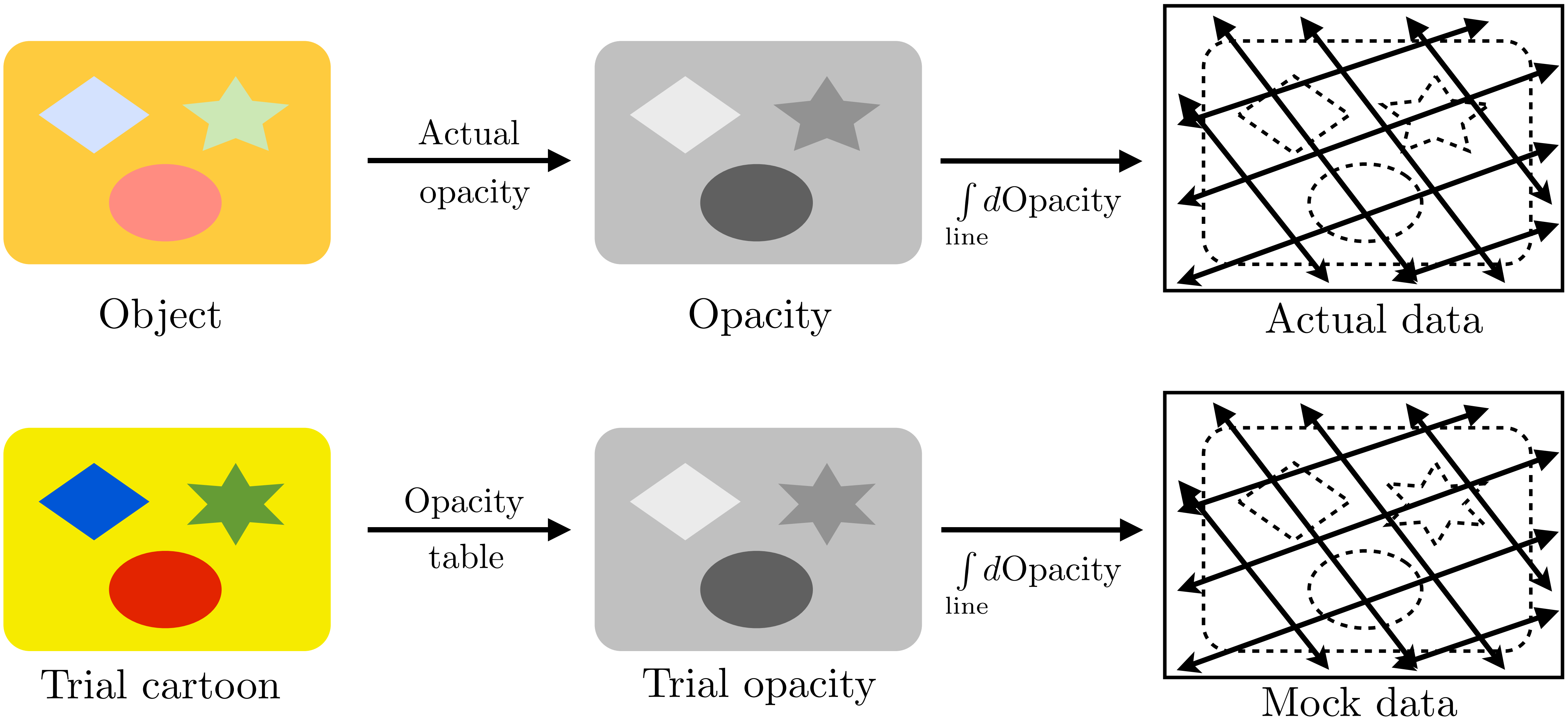
Whatever you choose. . .

Artistry! No clever mathematics. No continuum. No differential geometry. No barycentre.

Simplicity = generality + power

If derivatives **are** wanted, hexagonal neighbours give all of $f, f_x, f_y, f_{xx}, f_{xy}, f_{yy}$.

4. The Likelihood



Prior unknowns θ = cell colours.
Prior assignments = {colour probabilities, bond energies, colour opacities}.
Likelihood is based on

$$\text{Residuals} = \frac{\text{Mock data} - \text{Actual data}}{\text{Noise magnitude}}$$

Commonly,

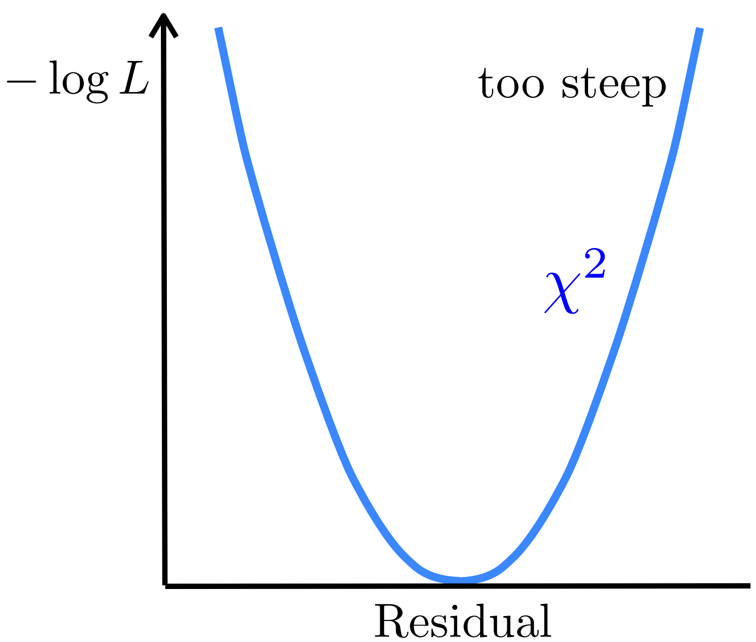
$$\text{Likelihood} = \exp(-\tfrac{1}{2}\chi^2), \qquad \chi^2 = \sum_{k=1}^{N} \text{Residual}_k{}^2$$

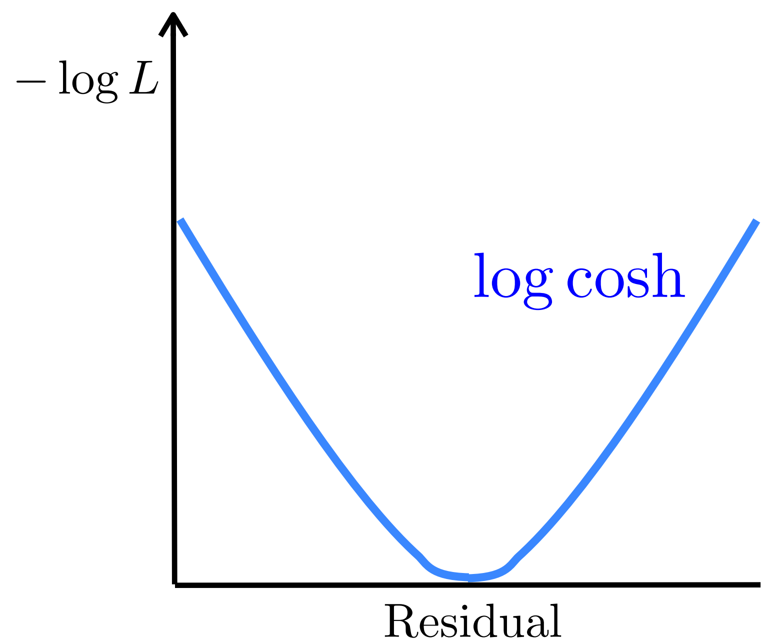
Hope to get $\chi^2 \sim N$ (each residual ~ 1).

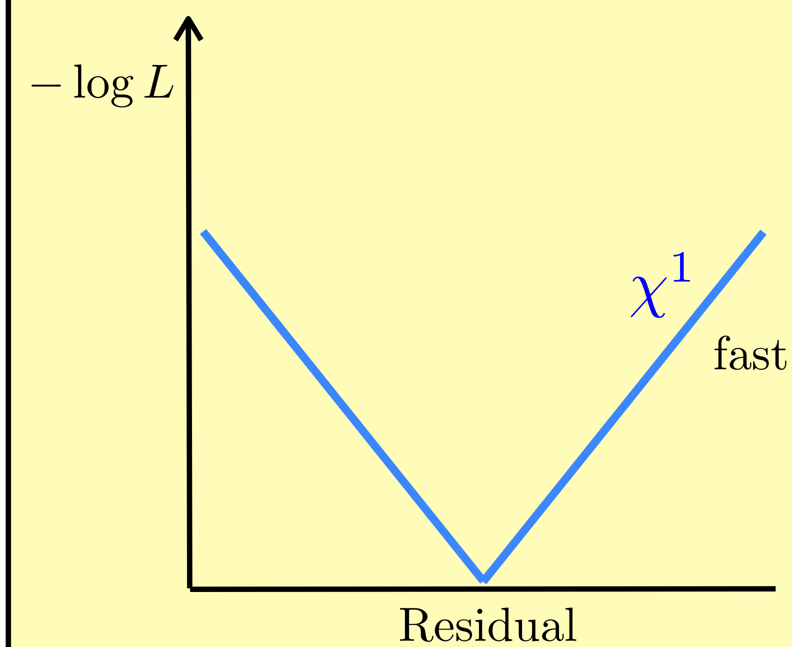
But cartoon is idealisation, **not** completely faithful. We **expect** residuals $\gg 1$.

A 10-sigma residual should **not** be 100 times worse than 1 sigma, and should **not** attract a e^{-50} penalty. The mismatch is systematic, not random noise.

Instead of $\log L = -\frac{1}{2}\sum R_k{}^2$, use $-\sum \log\cosh R_k$ or $-\sum |R_k|$.







$$L = \exp\Big(-\sum_k |\text{Residual}_k|\Big)$$

(Note: Bayes with χ^2 often gives posteriors that are wrongly placed and too definitive, because of imperfect modelling.)

5. Bayes is just sum and product rules for unknown parameters (colours) θ.

$$\underbrace{\Pr(\text{object})}_{\text{Prior } \pi(\theta)} \; \underbrace{\Pr(\text{Data} \mid \text{object})}_{\text{Likelihood } L(\theta)} = \underbrace{\Pr(\text{Data})}_{\text{Evidence } Z} \; \underbrace{\Pr(\text{object} \mid \text{Data})}_{\text{Posterior } P(\theta)}$$

$$Z = \sum_{\theta} L(\theta)\,\pi(\theta) \qquad\qquad P(\theta) = \pi(\theta)\,L(\theta)\,/\,Z$$

Sum and product rules $\Longleftarrow$ $\underbrace{\text{Associativity, Commutativity, Order}}_{\text{basic symmetries}}$

Same unique calculus for truth, myth, cartoons, ... so **we know no truth**.

Prior using limited colours is not true, but is consistent with symmetries.
Chi-not-squared likelihood is not true, but is consistent with symmetries.

Frequentist: "*Let the data speak for themselves!*" ✗

Bayesian: Ask a question, then get the answer. ✓

prior & likelihood $\Longrightarrow$ evidence & posterior

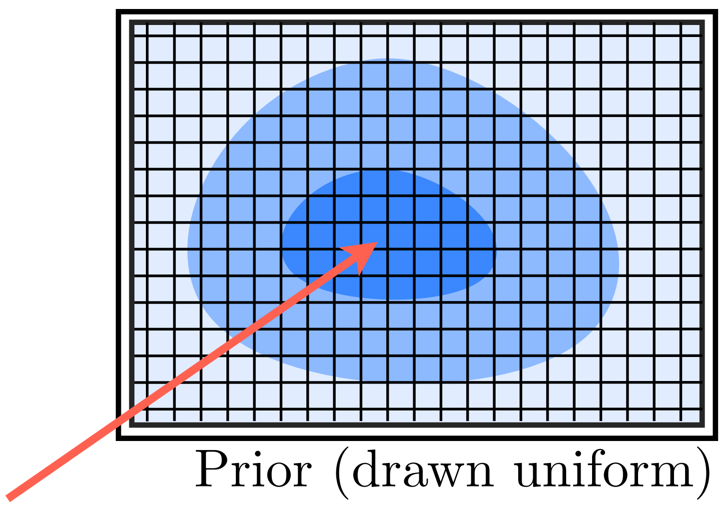
Science relies on epistemic values and judgement.

5. Bayes is just sum and product rules for unknown parameters (colours) θ.

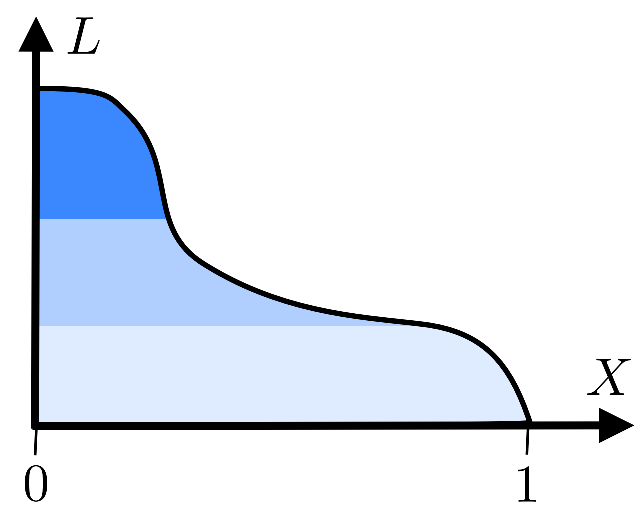
$$\underbrace{\Pr(\text{object})}_{\text{Prior } \pi(\theta)} \; \underbrace{\Pr(\text{Data} \mid \text{object})}_{\text{Likelihood } L(\theta)} = \underbrace{\Pr(\text{Data})}_{\text{Evidence } Z} \; \underbrace{\Pr(\text{object} \mid \text{Data})}_{\text{Posterior } P(\theta)}$$

$$Z = \int L(\theta)\, \pi(\theta) d^n\theta \qquad\qquad P(\theta) = \pi(\theta)\, L(\theta) \,/\, Z$$

Riemann sum $\int L\, dX$ is expensive.



Use Lebesgue $\int X\, dL$ instead.

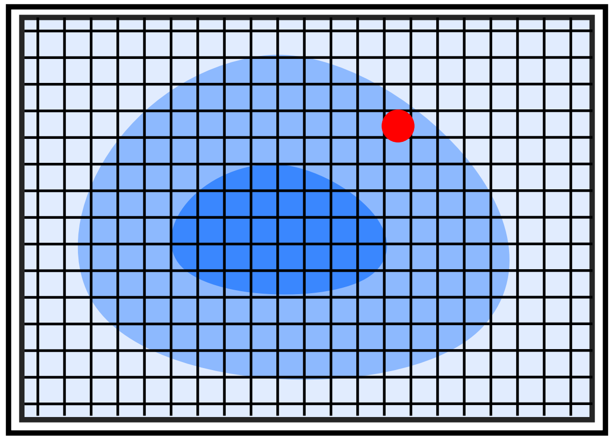


$X(L)$ = prior mass with likelihood $> L$.

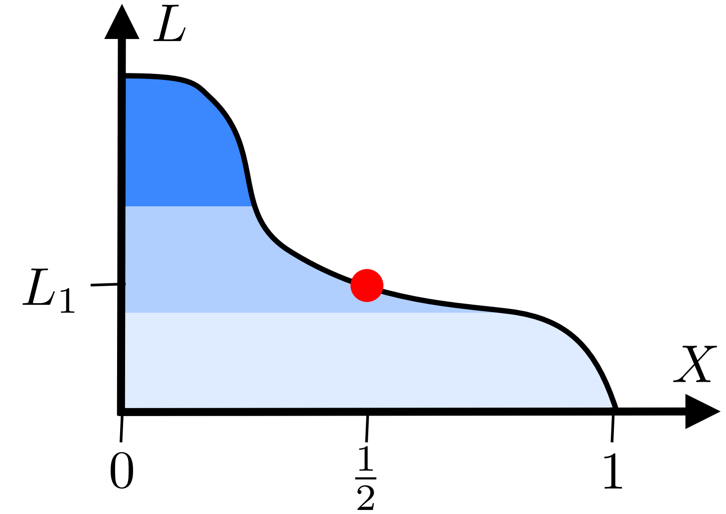
Can not compute X exactly.

Must get X statistically.

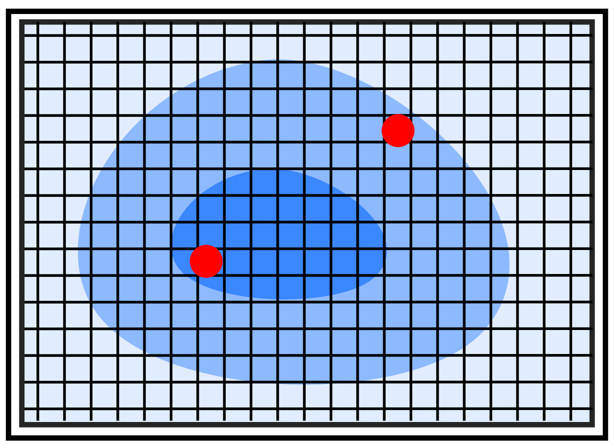
Start with random $\mathbf{x}_1$. Get $L_1 = L(\mathbf{x}_1)$.



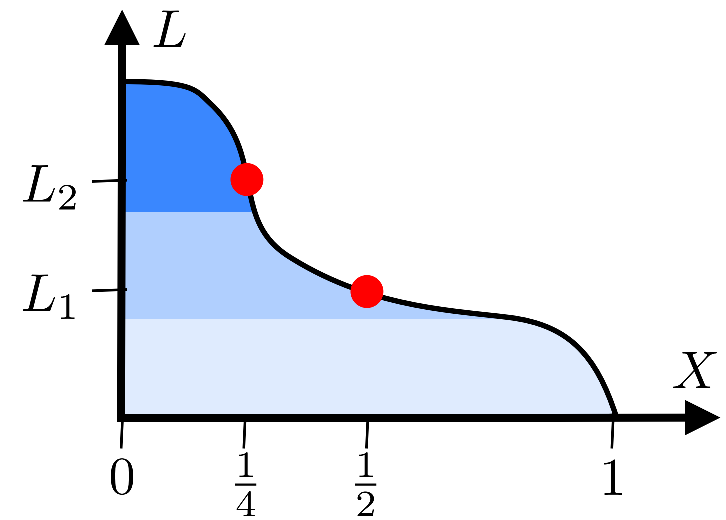
Then $X_1(L_1) \sim \texttt{Uniform}(0, 1) \approx \frac{1}{2}$.



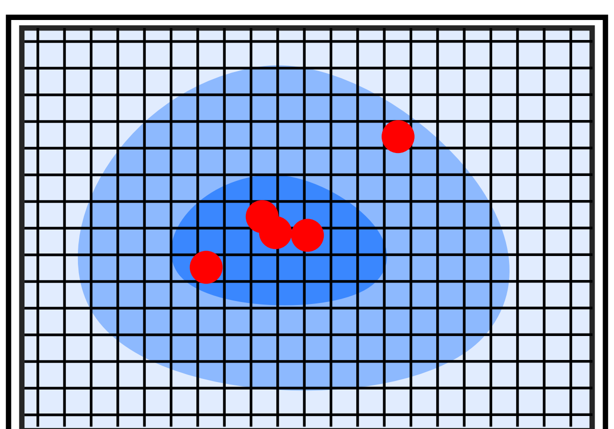
Next, move to $\mathbf{x}_2$ random within $L > L_1$. Get $L_2 = L(\mathbf{x}_2)$.
(! *Oui ou non* !)



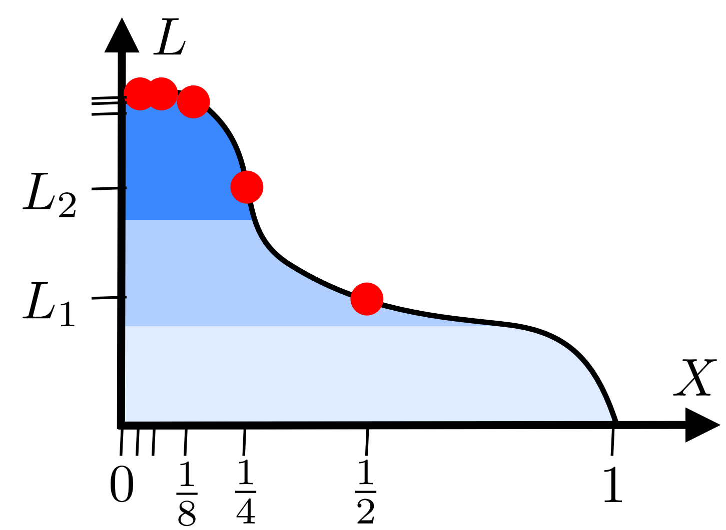
Then $X_2(L_2) \sim \texttt{Uniform}(0, X_1) \approx \frac{1}{4}$.



And so on, with $\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \ldots$.



Compress exponentially to reach the exponentially tiny posterior.



Lebesgue compression is **nested sampling**.

Get sequence $\{$location $\mathbf{x}_r$, likelihood L_r, enclosed prior mass $X_r\}$.

Terminate when information H ("$\sum P \log P$") saturates.
For more accuracy, use more samples.
Guaranteed convergence to truth if $H < \infty$.

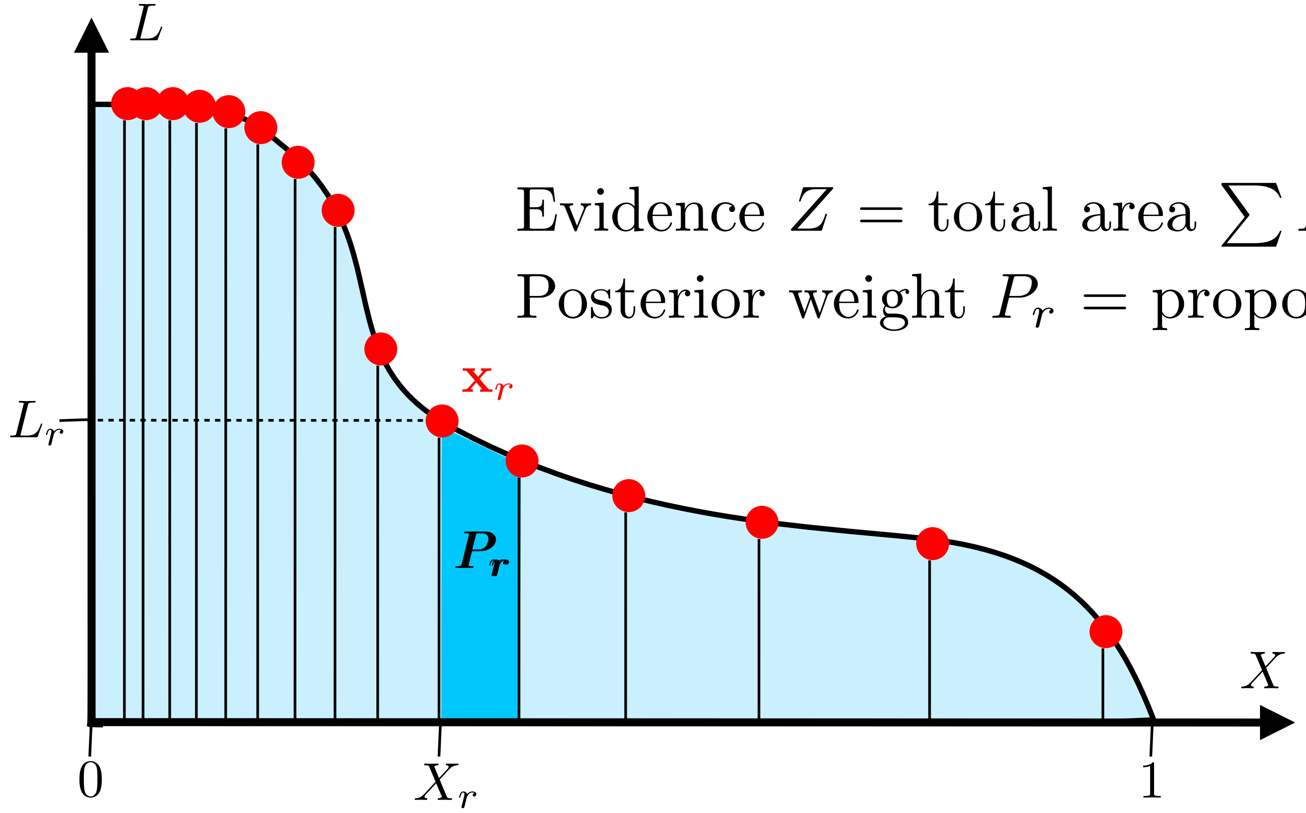
To set β, use relationship $\langle E \rangle = \dfrac{\int E\, n(E) e^{-\beta E}\, dE}{\int n(E) e^{-\beta E}}$ between β and average energy $\langle E \rangle$,
pre-computed $\forall \beta$ from prior model of colours and bonds on assigned grid.*

Start with random colours at $\beta = 0$ (disorder).
Allow β to increase appropriately as the data impose order. ($\beta = \infty$ is uniform colour.)

---

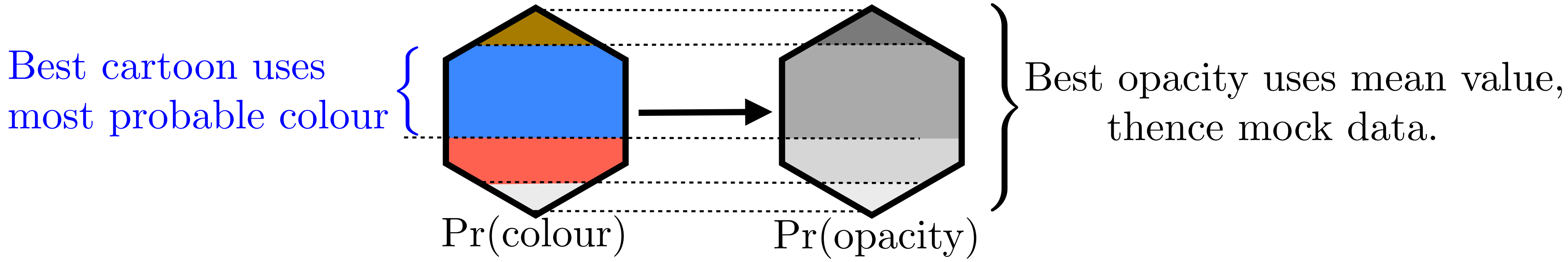
*Best get $n(E)dE$ as nested sampling's prior mass dX in a calibration run, then simulate $\langle E \rangle$).

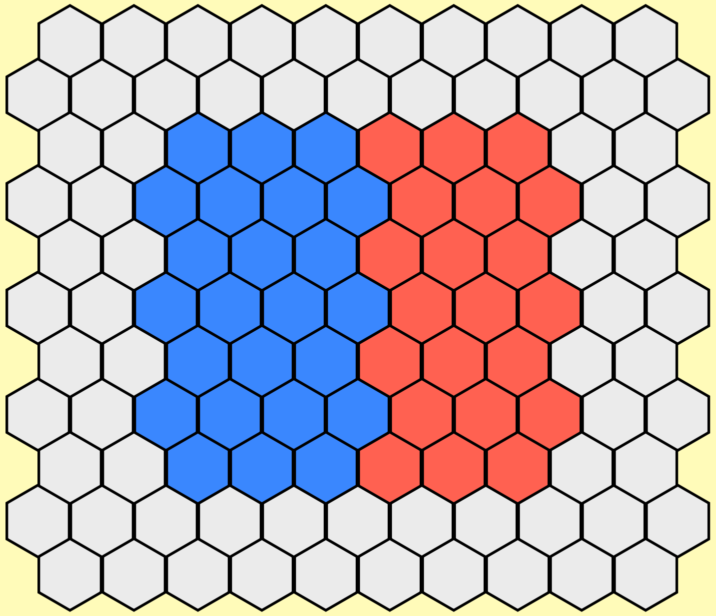
6. Results



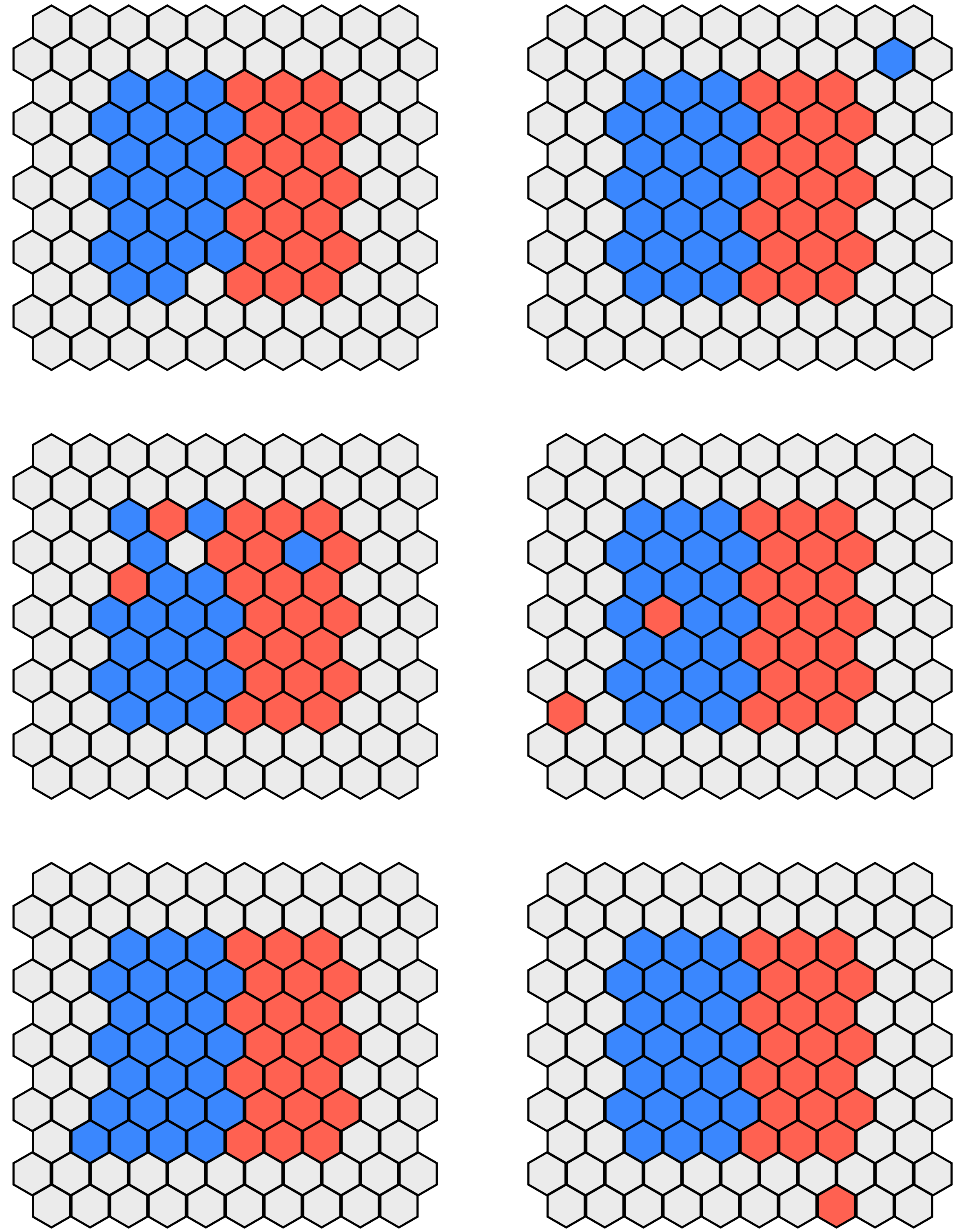
Evidence Z = total area $\sum L_r \, \delta X_r$.
Posterior weight P_r = proportional area.

Each cell accumulates its own posterior colour distribution.





Toy reconstruction

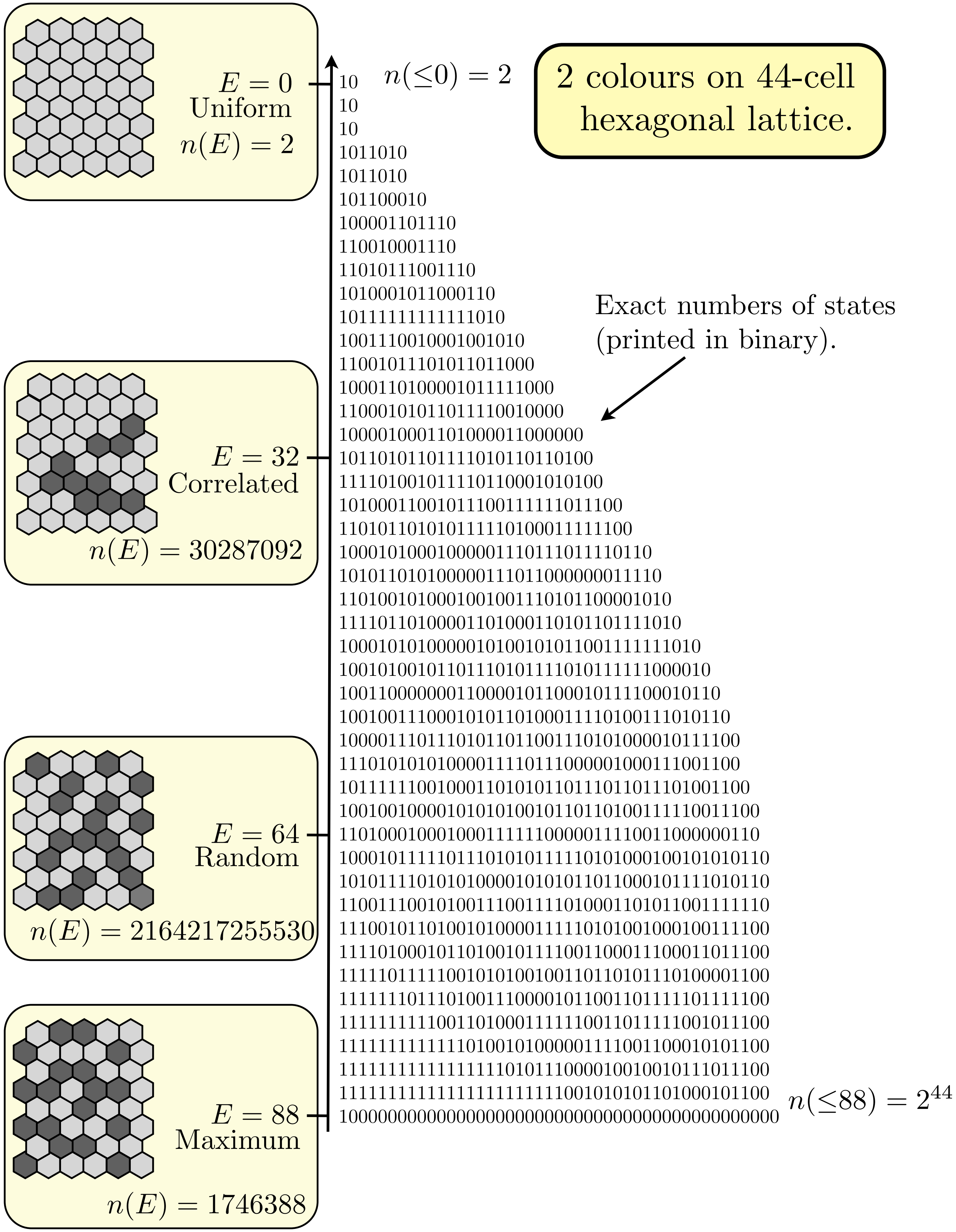


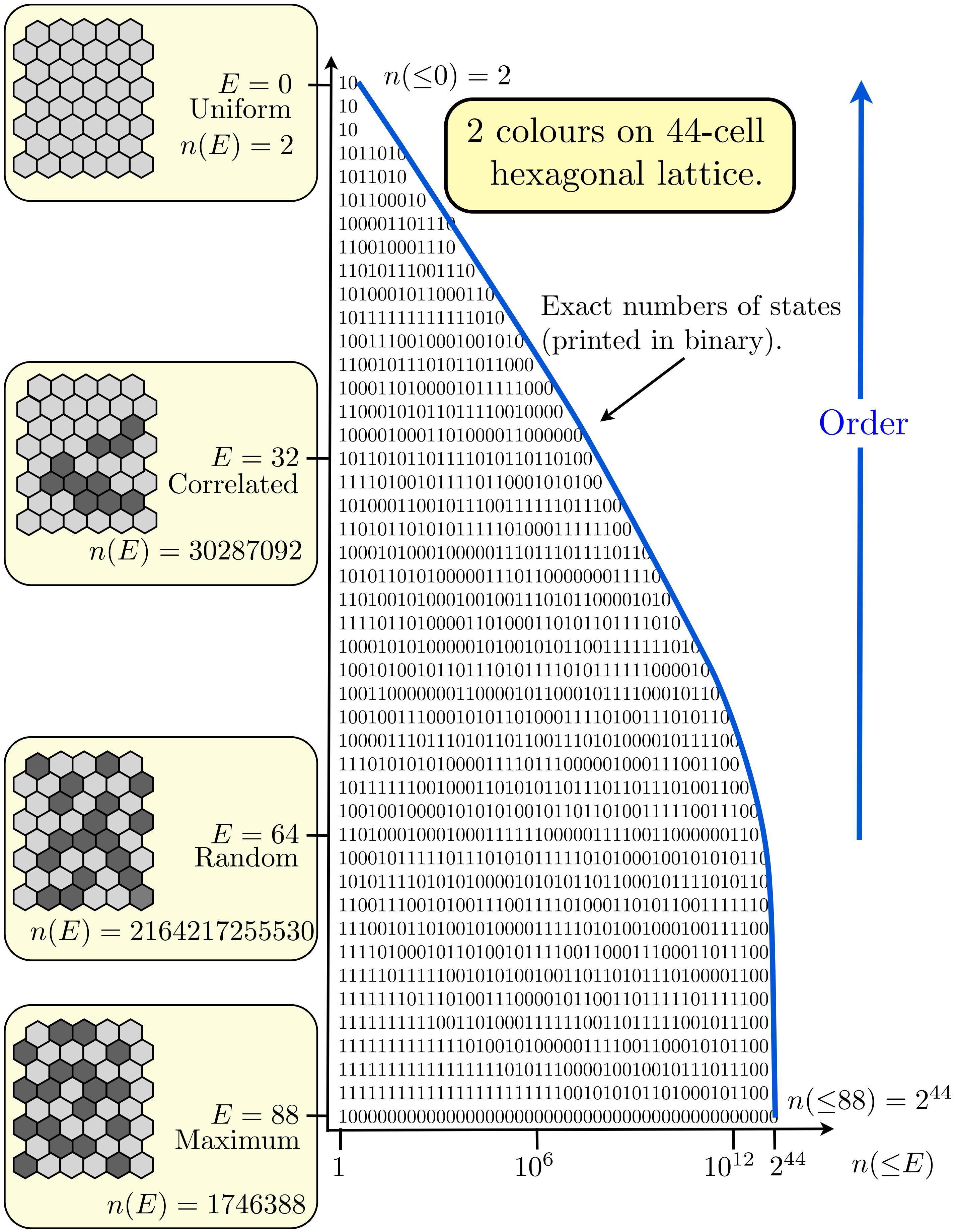
Posterior samples

That's all. It's simple!

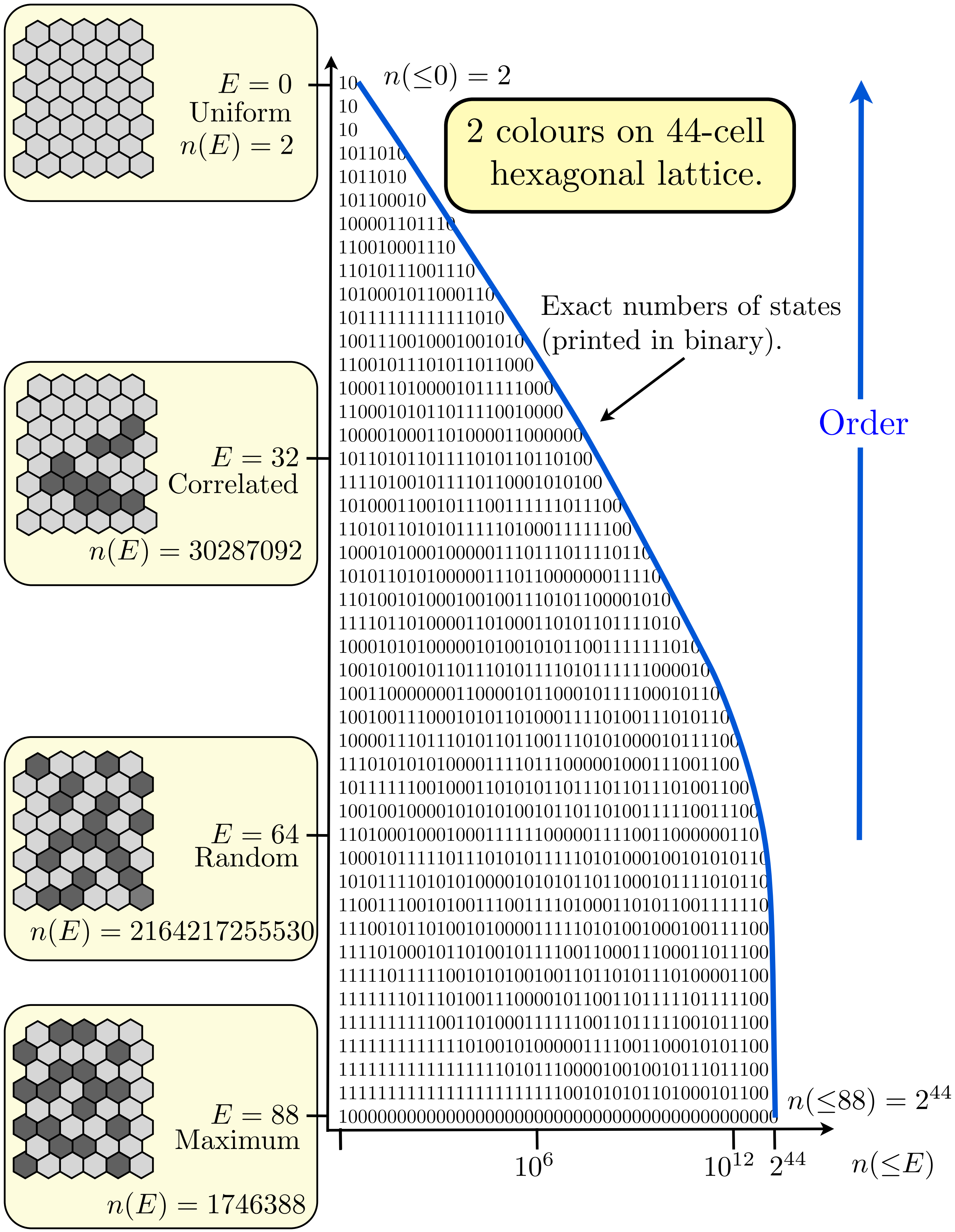
[redacted]

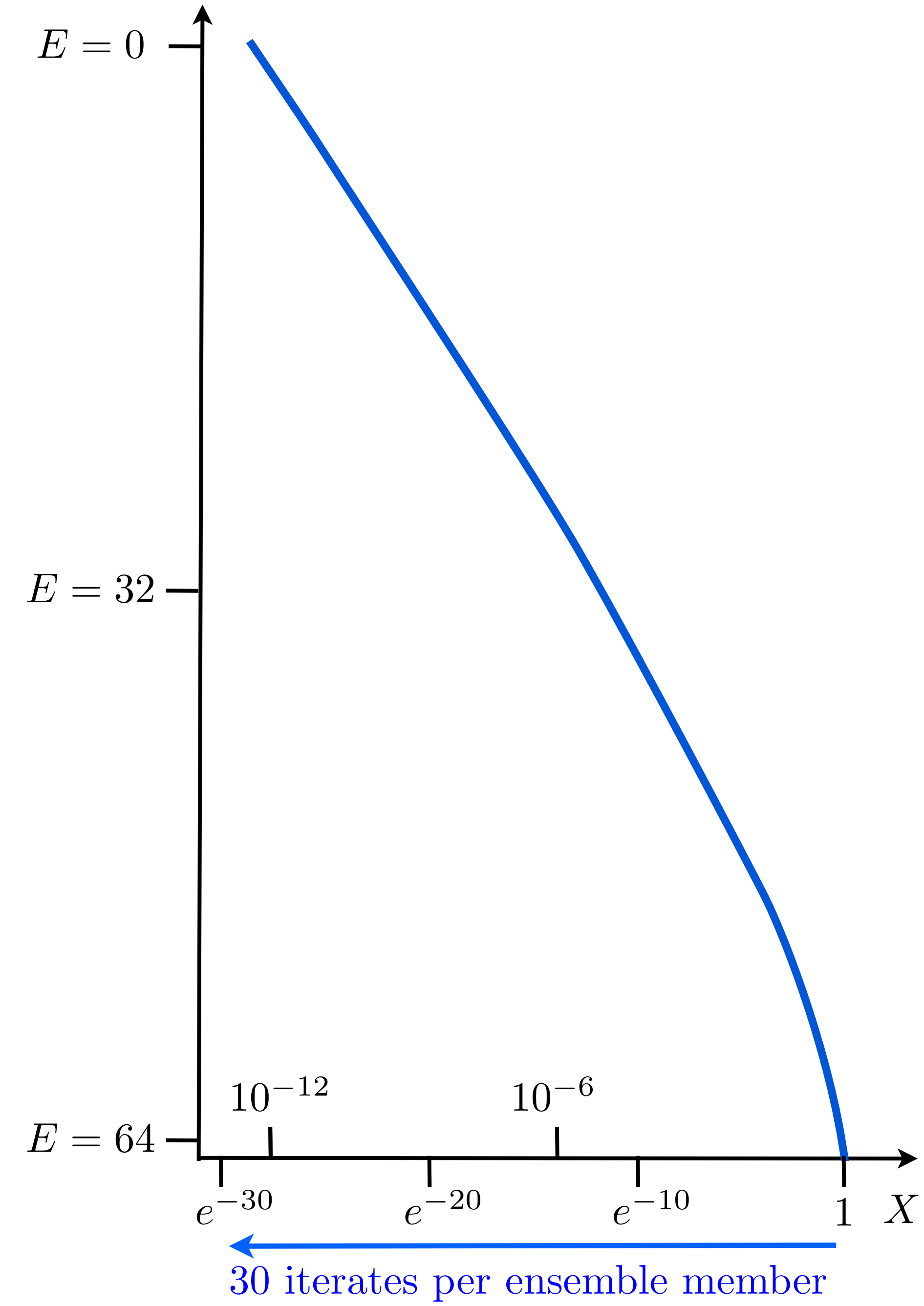
(Ancillary slides follow)


2 colours on 44-cell hexagonal lattice.
E = 0
Uniform
n(E) = 2
E = 32
Correlated
n(E) = 30287092
E = 64
Random
n(E) = 2164217255530
E = 88
Maximum
n(E) = 1746388
n(≤0) = 2
Exact numbers of states (printed in binary).
10
10
10
1011010
1011010
101100010
100001101110
110010001110
11010111001110
1010001011000110
1011111111111010
1001110010001001010
1100101110101011011000
10001101000010111111000
110001010110111110010000
1000010001101000011000000
1011010110111101011011010 0
11110100101111011000101010 0
1010001100101110011111101110 0
1101011010101111101000111111 00
100010100010000011101110111101 10
1010110101000001110110000000111 10
11010010100010010011101011000010 10
111101101000011010001101011011110 10
1000101010000010100101011001111111 010
1001010010110111010111101011111100 0010
1001100000001100001011000101111000 10110
1001001110001010110100011110100111 010110
1000011101110101101100111010100001 0111100
1110101010100001111011100000100011 1001100
1011111110010001101010110111011011 10100 1100
1001001000010101010010110110100111 1100 11100
1101000100010001111111000000111100 1100000110
1000101111101110101011111010100010 0101010110
1010111101010100001010101101100010 1111010110
1100111001010011100111101000110101 1001111110
1110010110100101000011111010100100 0100111100
1111010001011010010111100110001110 0011011100
1111101111100101010010011011010111 0100001100
1111111011101001110000101100110111 1101111100
1111111111001101000111111001101111 1001011100
1111111111111010010100000111100110 0010101100
1111111111111111101011100001001001 0111011100
1111111111111111111111110010101011 0100010110 0
100000000000000000000000000000000000000000000
n(≤88) = 2^{44}



E = 0
Uniform
n(E) = 2
E = 32
Correlated
n(E) = 30287092
E = 64
Random
n(E) = 2164217255530
E = 88
Maximum
n(E) = 1746388
n(≤0) = 2
2 colours on 44-cell hexagonal lattice.
Exact numbers of states (printed in binary).
Order
n(≤88) = 2^44
1
10^6
10^12
2^44
n(≤E)





X is fractional "volume".

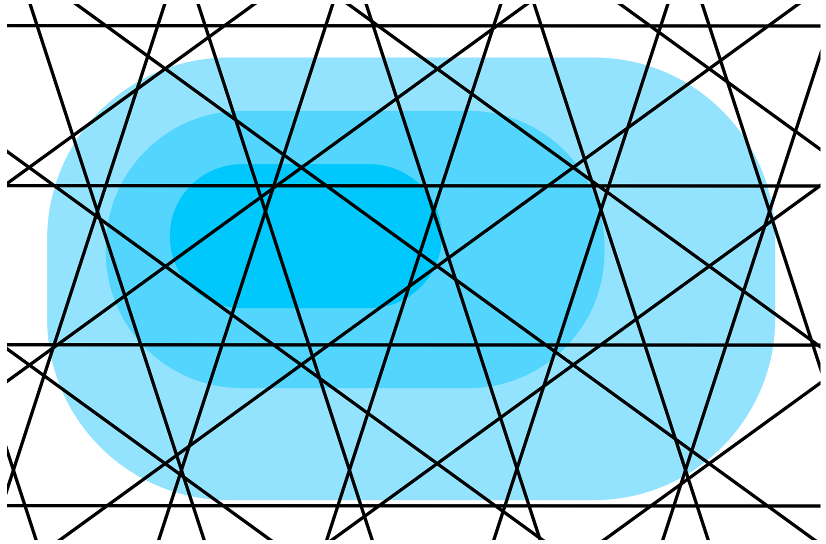
Nested sampling builds the curve uniformly in $\log X$ for *any* energy distribution.

In 1 dimension:- tomography fails because $\int \rho(x)dx$ can not yield detailed ρ.

In 2 dimensions:- direction $\mathbf{n}$ lies on unit circle, and best sampled uniformly; offset $\mathbf{c}$ (conventionally $\perp$ $\mathbf{n}$) is also best sampled uniformly.

I suspect that the number of $\mathbf{n}$ and the number of offsets $\mathbf{c}$ should balance, so that $N \times N$ image needs about N directions and N offsets.

At high resolution, D is Radon transform of ρ, which is invertable, but that ignores (and amplifies) noise.



In 3 dimensions:- direction $\mathbf{n}$ lies on unit sphere, and best sampled uniformly; offset $\mathbf{c}$ (in plane $\perp$ $\mathbf{n}$) is also best sampled uniformly.

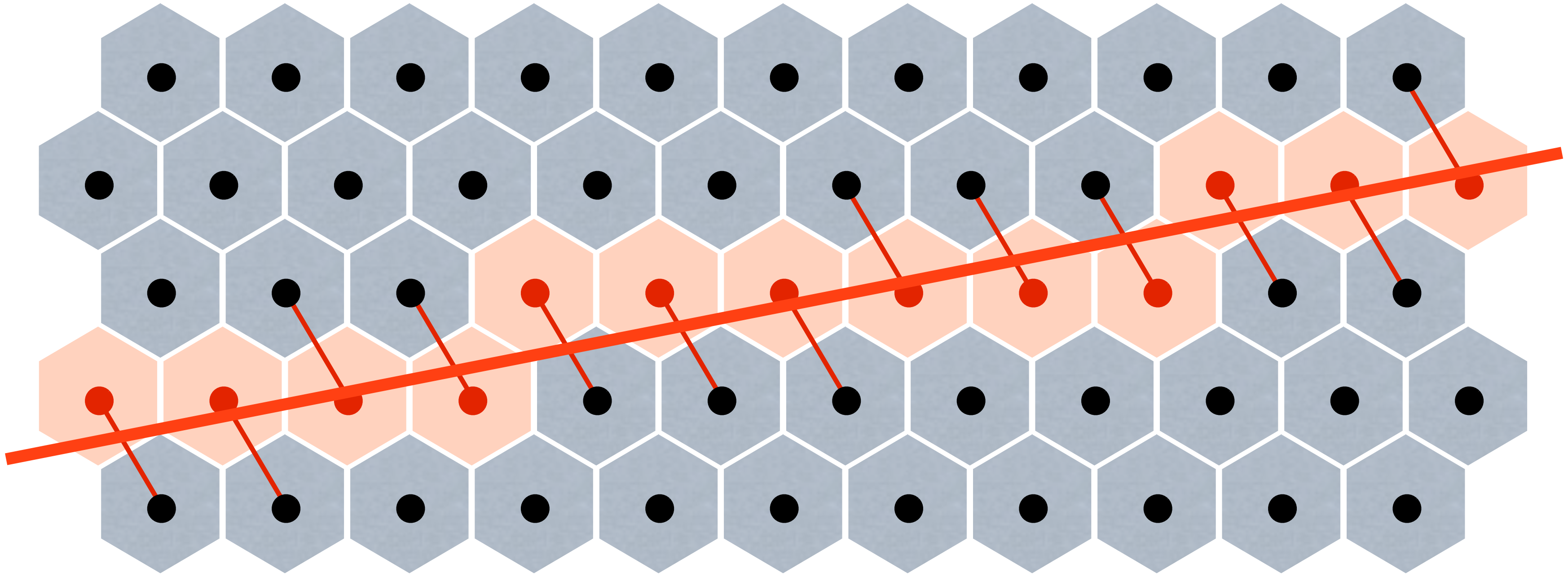
I suspect that the number of $\mathbf{n}$ and the number of offsets $\mathbf{c}$ should balance, so that $N \times N \times N$ image needs about $N^{3/2}$ directions and $N^{3/2}$ offsets.

I know of no analytic inverse (and would not want it).

Use hexagonal grid!

Identify scan nodes as the closest to the scan line at it crosses most-nearly-orthogonal bonds.

Weight nodes by w = length of scan line per node = $\dfrac{\sqrt{3}}{\cos(\theta - 30^\circ)}$



$0 \leq \theta \leq 60^\circ$ scan lines are $\sim$ orthogonal to 120° bonds.