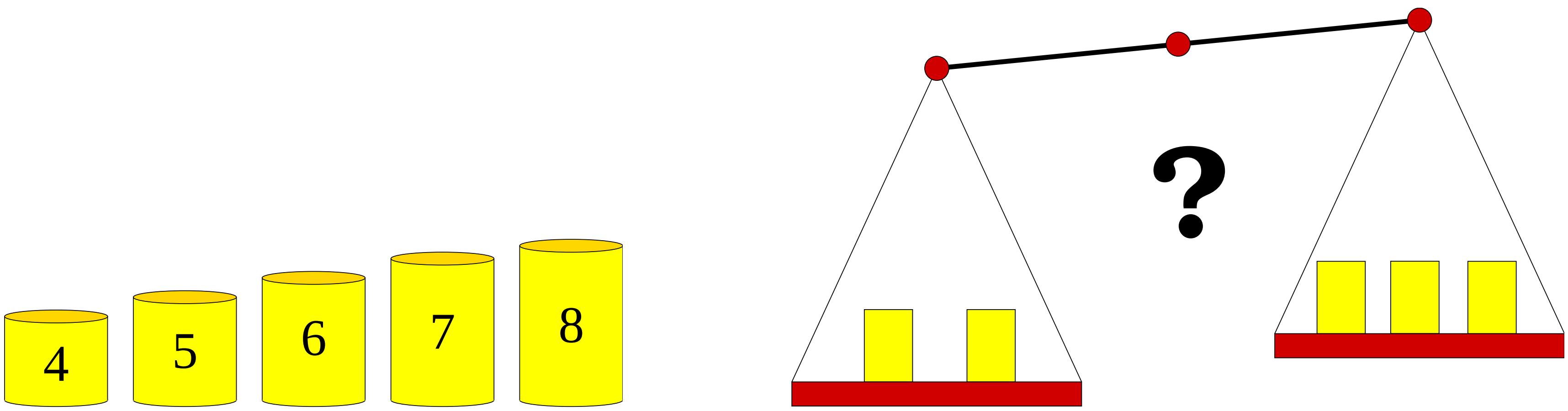
The Easiest Hard Problem

Phasetransitions in Integer Partitioning

Stephan Mertens, Magdeburg University

Christian Borgs, Microsoft
Jennifer Chayes, Microsoft
Boris Pittel, Ohio State University

Balancing Numbers



Number Partitioning Problem (NPP): Given n positive integers $\{X_1, X_2, \ldots, X_n\}$. Find a partition $\sigma_j = \pm 1$ that minimizes the discrepancy

$$d = |\sum_{j=1}^{n} X_j \sigma_j|.$$

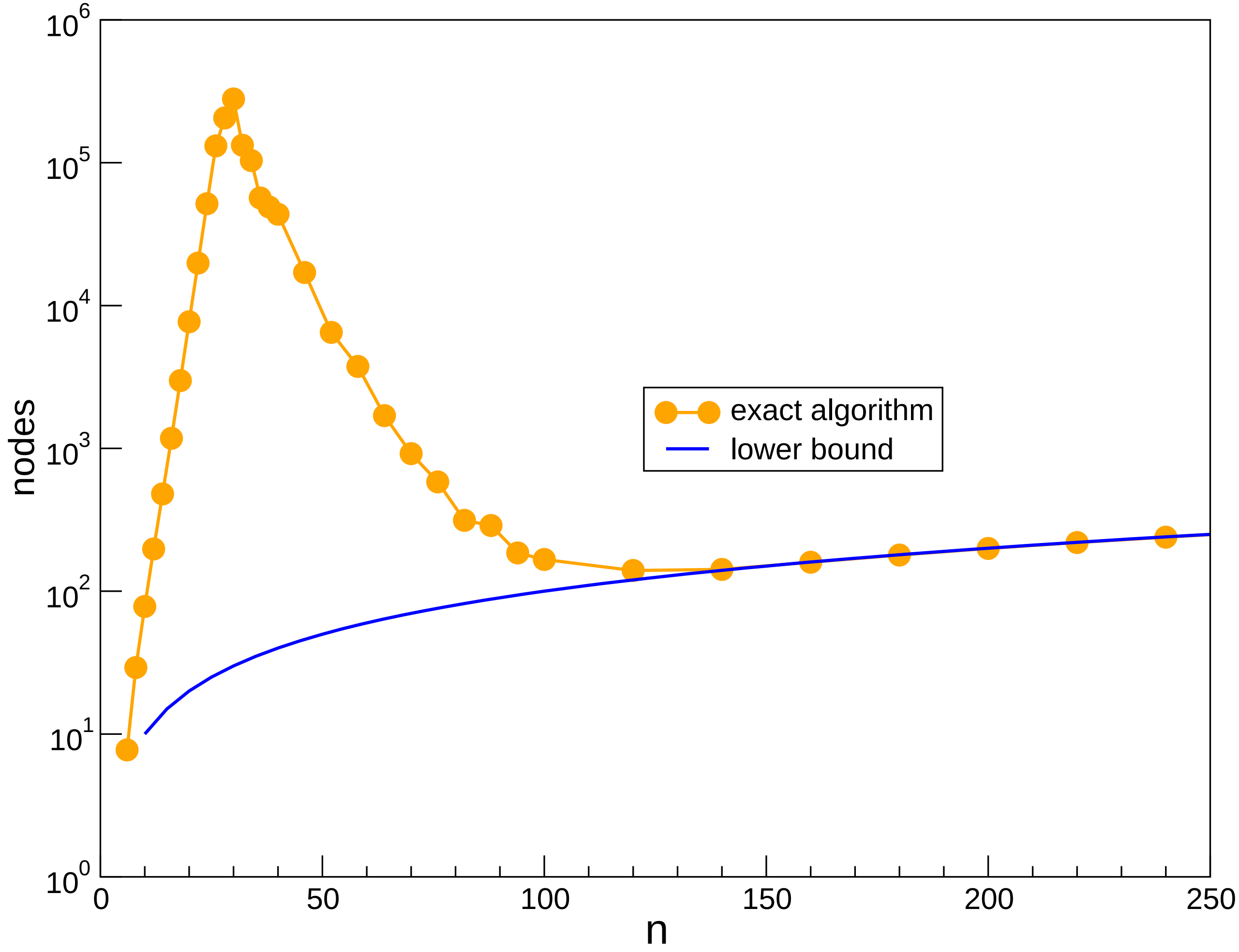
The easiest hard problem[1]

Integer partioning is

- hard because it's NP-hard.
- easy because there is a hard-easy phasetransition.
- easiest because exact and rigorous results are known on the phasetransition.

[1] Brian Hayes *American Scientist* **90**, March-April 2002, 113-117

Example

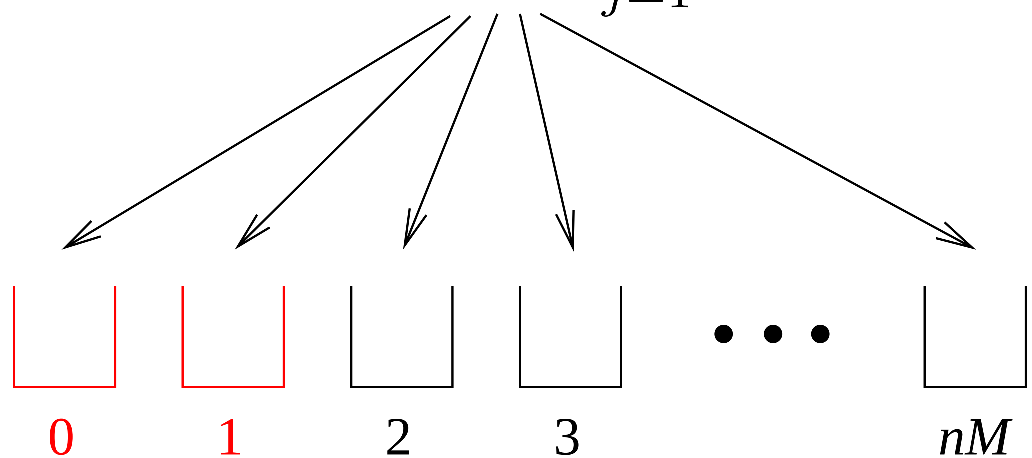


Finding the optimum partition of random 25-bit integers.

Pigeonhole Principle

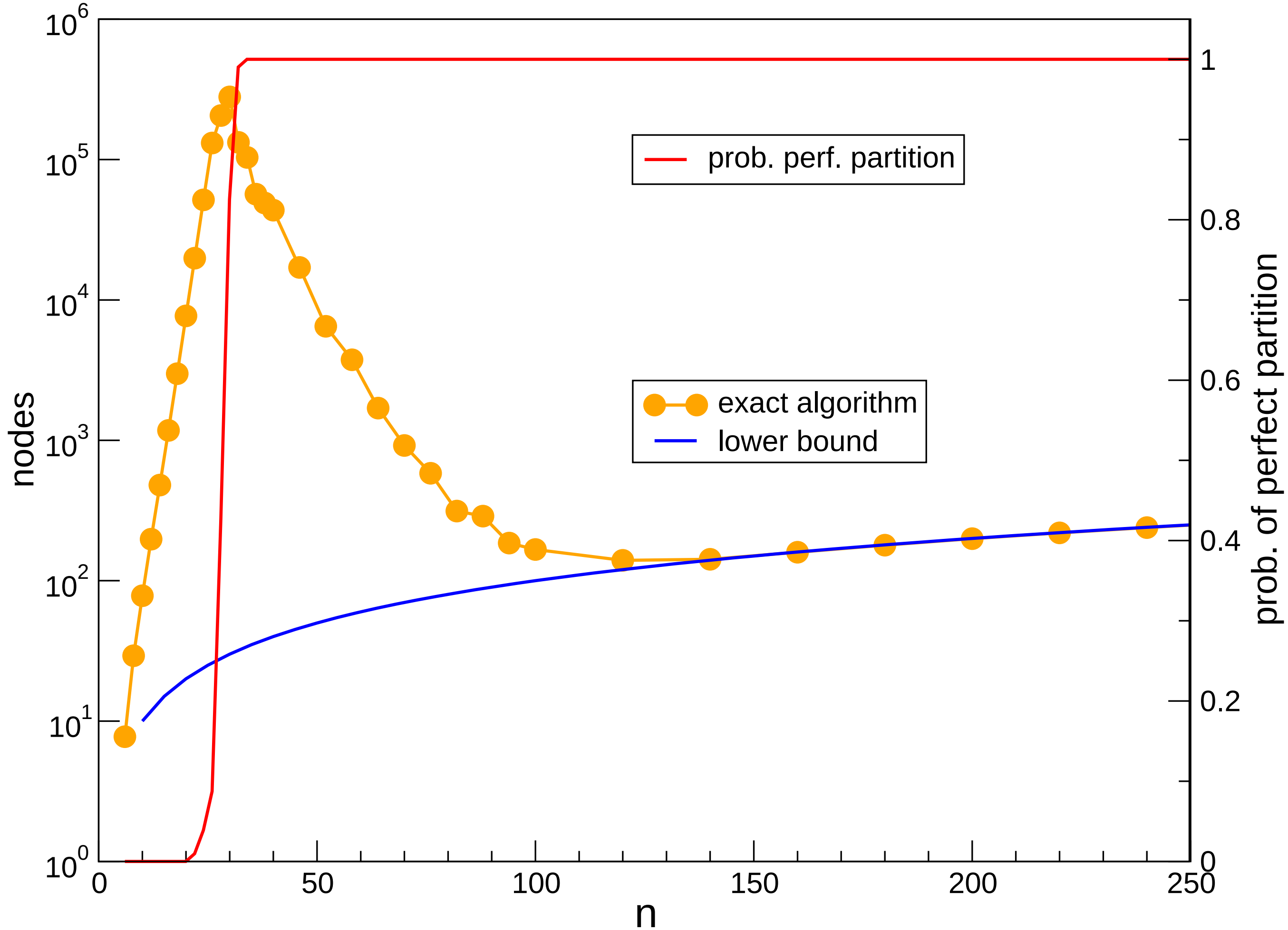
$$M := \max X_j$$

$$2^n \text{ values } d = \left| \sum_{j=1}^{n} X_j \sigma_j \right|$$



$$M = 2^{\kappa n} \Rightarrow \quad \# \text{ of perfect partitions } \approx \frac{1}{n} 2^{(1-\kappa)n}$$

Perfect Partitions



Perfect partition: $d = |\sum_{j=1}^{n} X_j \sigma_j| \in \{0, 1\}$

Phasetransition

- $\kappa < \kappa_c$: exponentially many perfect partitions
  $\kappa > \kappa_c$: no perfect partitions

- I. Gent, T. Walsh (1996): numerics $\Rightarrow \kappa_c(n \to \infty) \approx 0.96$

- S.M. (1998): non rigorous entropy calculation

$$\kappa_c = 1 - \frac{1}{n} \log_2 \sqrt{\frac{\pi}{6} n} \tag{1}$$

- C. Borgs, J. Chayes, B. Pittel (2001): proof of (1) and many other rigorous results.

Constrained Number Partitioning

Given n positive integers $\{X_1, X_2, \ldots, X_n\}$ and an integer $B \in \{0, 1, \ldots, n\}$. Find a partition $\sigma_j = \pm 1$, $j = 1, \ldots, n$ that minimizes the discrepancy

$$d = |\sum_{j=1}^{n} X_j \sigma_j|.$$

under the constraint $\sum_{j=1}^{n} \sigma_j = B =: bn$

Overconstrained Instances

Find optimum partition with 8 vs. 2 numbers (i.e. $b = 0.6$) of

$$2 \quad 10 \quad 3 \quad 8 \quad 5 \quad 7 \quad 9 \quad 5 \quad 3 \quad 2$$

$$\underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad} \qquad \underline{\quad}\ \underline{\quad}$$

Sorted partition: Put the largest numbers in the smaller set,

$$\underline{2}\ \underline{2}\ \underline{3}\ \underline{3}\ \underline{5}\ \underline{5}\ \underline{7}\ \underline{8} \qquad \underline{9}\ \underline{10}$$

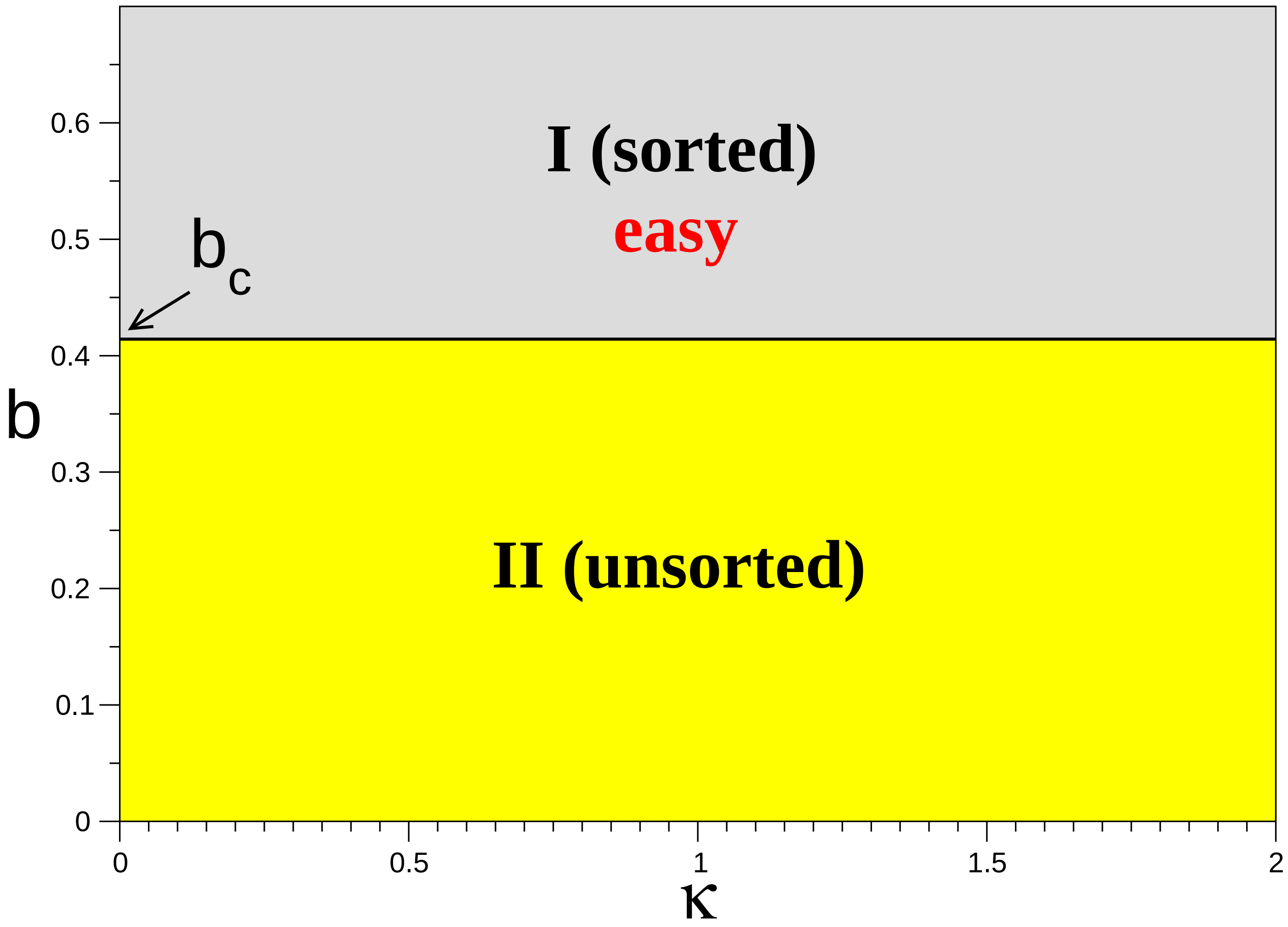
$$\Sigma = 35 \qquad\qquad \Sigma = 19$$

Critical Disbalance

$$
\begin{aligned}
\frac{\langle d_s \rangle}{nM} &= \int_0^{\frac{1+b}{2}} x\,dx - \int_{\frac{1+b}{2}}^{1} x\,dx \\
&= \left(\frac{1+b}{2}\right)^2 - \frac{1}{2} \\
&\geq 0
\end{aligned}
$$

$$b > b_c = \sqrt{2} - 1 = 0.4142\ldots$$

Sorted-Unsorted Phasetransition



Counting perfect partitions

$$\begin{aligned}
Z_n(\kappa, b) &= \sum_{\sigma} \mathbb{I}\Big(\sum_j X_j \sigma_j = 0, \sum_j \sigma_j = bn\Big) \\
&= \frac{2^n}{\pi^2} \int_{-\pi/2}^{\pi/2} dm \int_{-\pi/2}^{\pi/2} dy\, e^{-inbm} \prod_j \cos(X_j y + m) \\
&= \frac{2^n}{\pi^2} \int_{-\pi/2}^{\pi/2} dm \int_{-\pi/2}^{\pi/2} dy\, e^{nG(y,m)}
\end{aligned}$$

with

$$G(y, m) = -ibm + \frac{1}{n} \sum_{j=1}^{n} \log\cos(X_j y + m). \tag{2}$$

Saddlepoint Method

Entropy $s_n(\kappa, b) := \frac{1}{n} \log Z_n(\kappa, b)$

$$s_n(\kappa, b) = (1-\kappa) \log 2 - b\tilde{m} + \langle \log \cosh(x\tilde{y} + \tilde{m}) \rangle - \frac{1}{n} \log \left(\frac{\pi n}{2} \sqrt{|g|} \right)$$

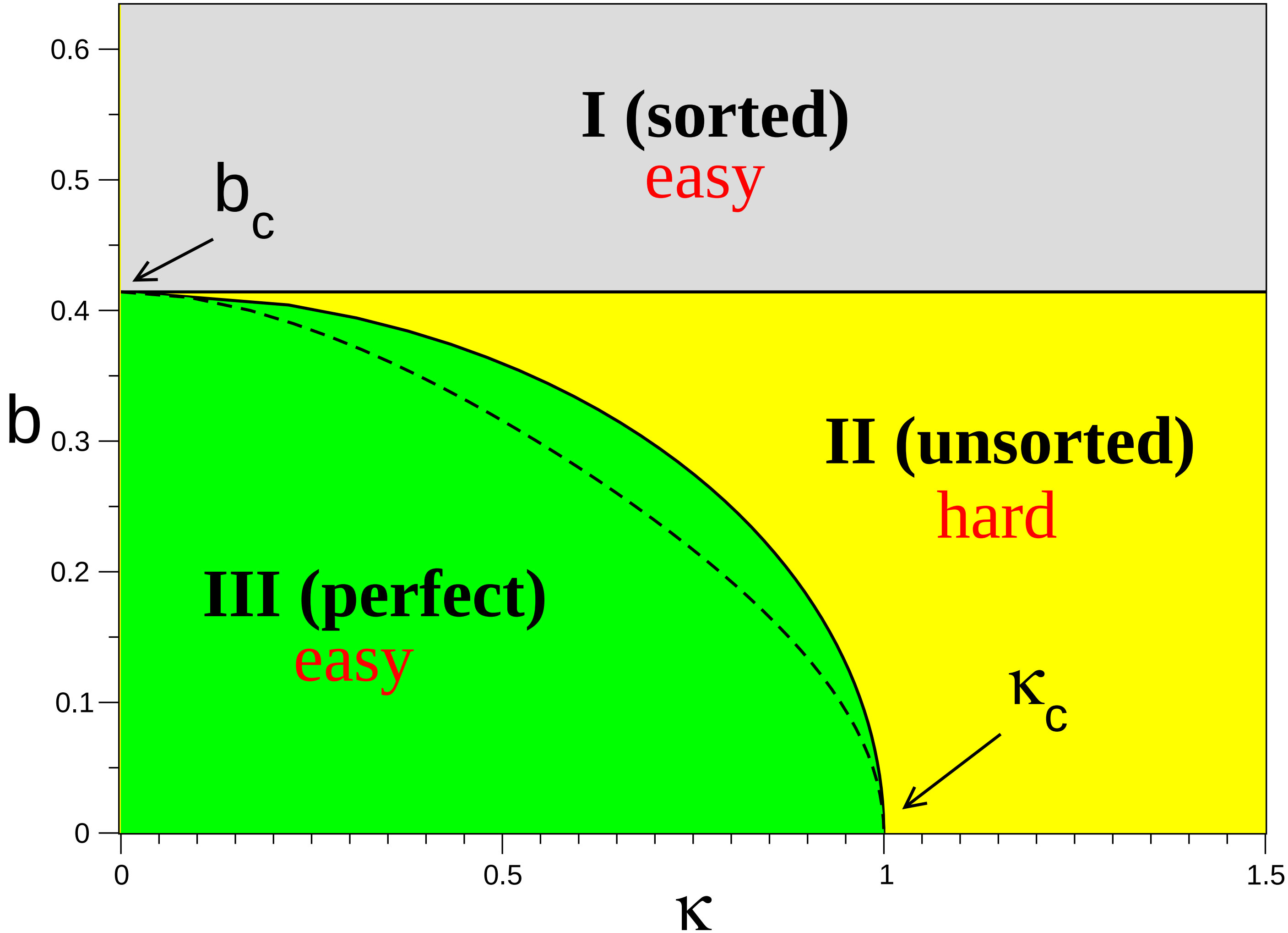
where $\tilde{m}$ and $\tilde{y}$ are the solutions of the saddlepoint equations

$$\langle \tanh(x\tilde{y} + \tilde{m}) \rangle = b \qquad \langle x \tanh(x\tilde{y} + \tilde{m}) \rangle = 0$$

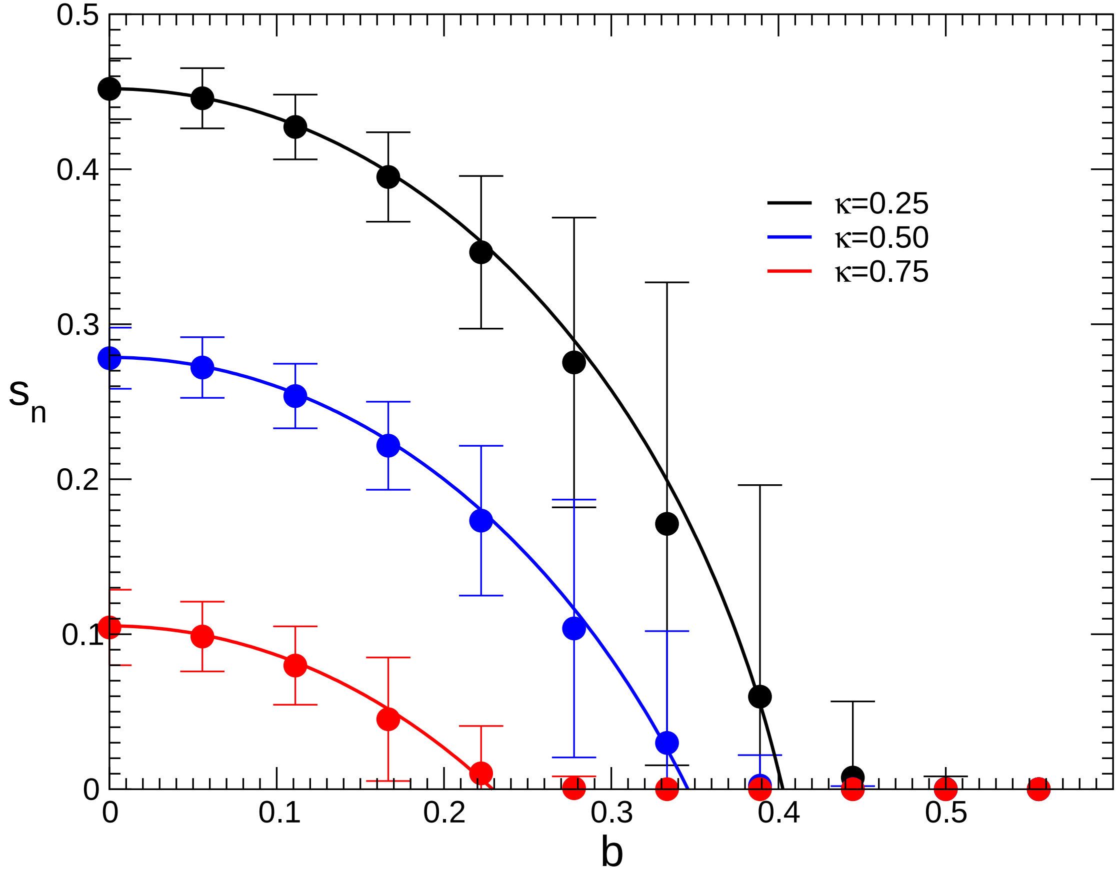
and

$$g = \left\langle \frac{x^2}{\cosh^2(x\tilde{y} + \tilde{m})} \right\rangle \left\langle \frac{1}{\cosh^2(x\tilde{y} + \tilde{m})} \right\rangle - \left\langle \frac{x}{\cosh^2(x\tilde{y} + \tilde{m})} \right\rangle^2$$

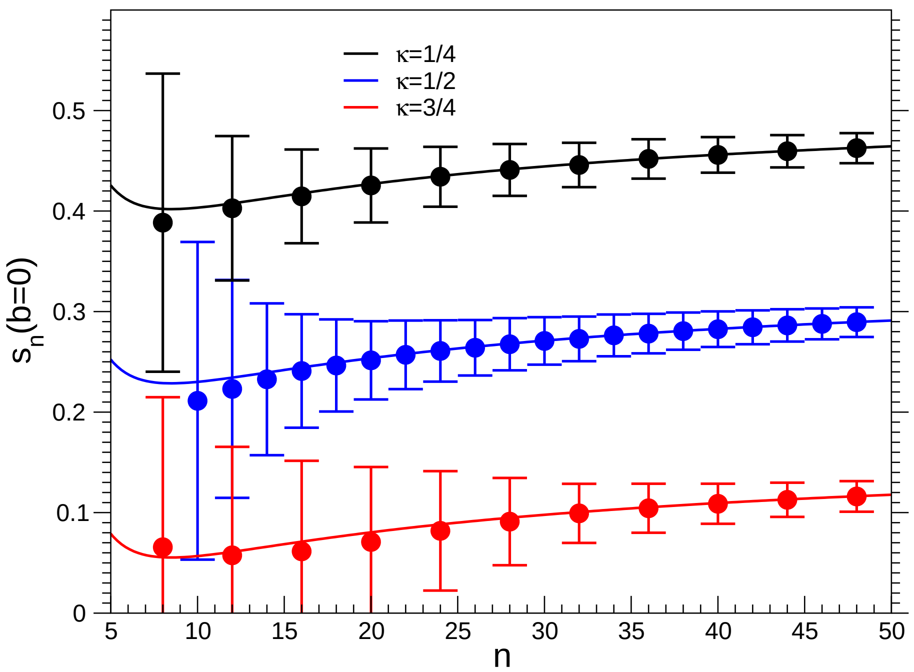
Phasediagram

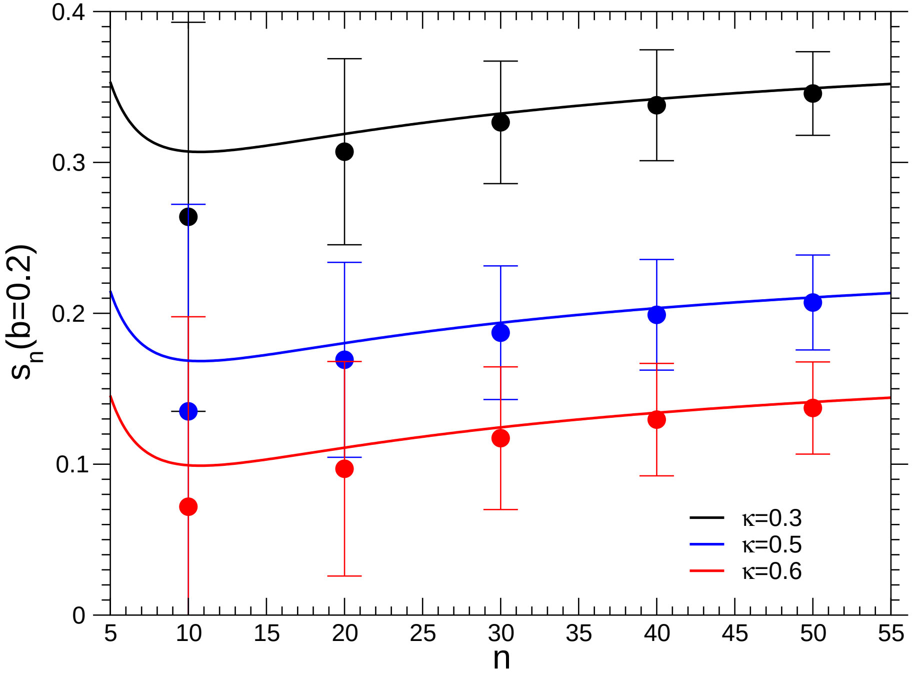


From Phase III to Phase II

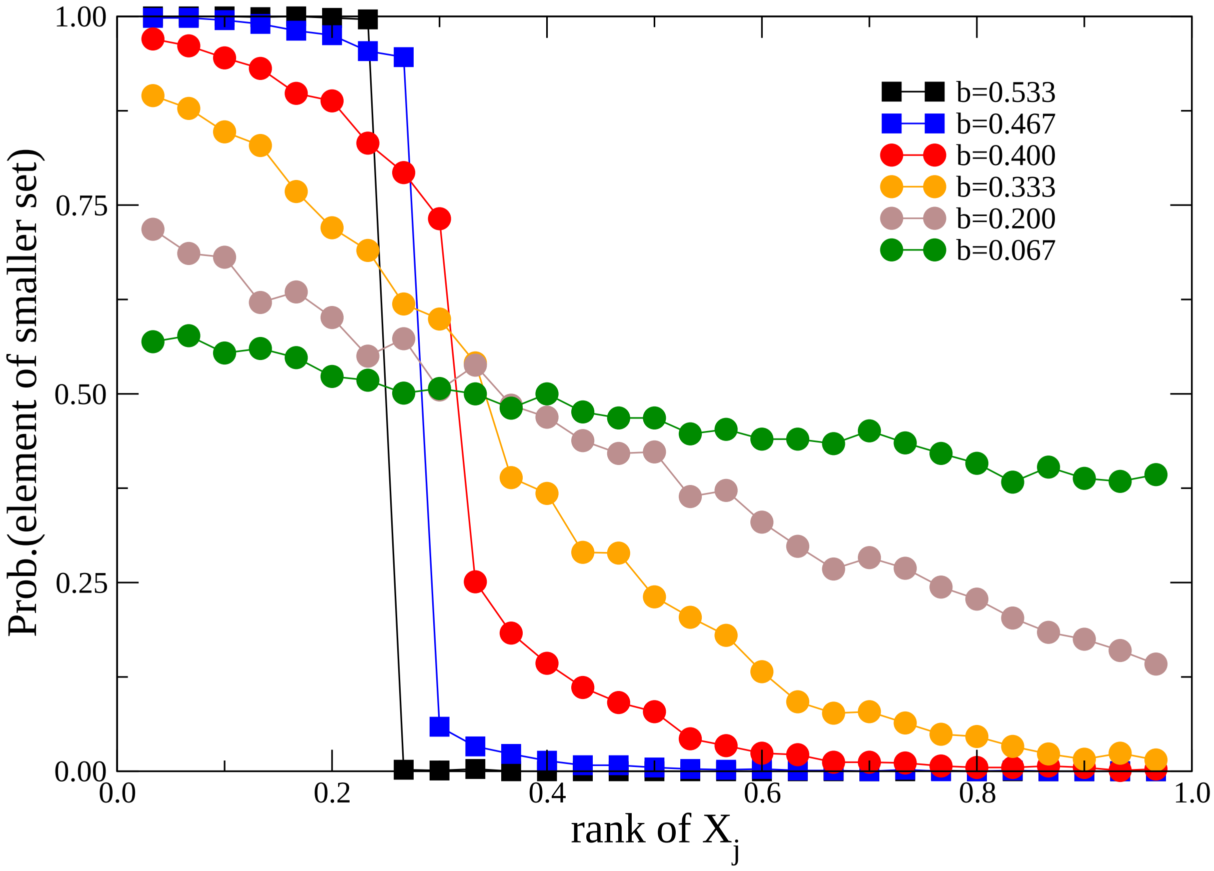


Concentration for $n \to \infty$





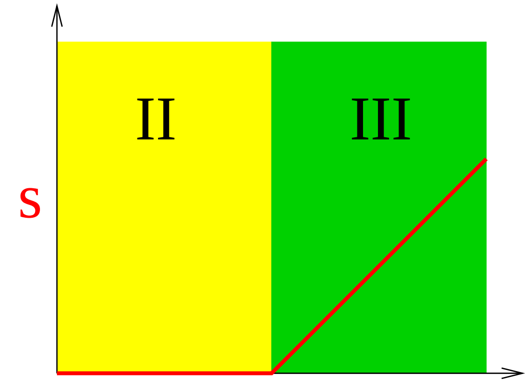
From Phase I to Phase II



Order Parameters

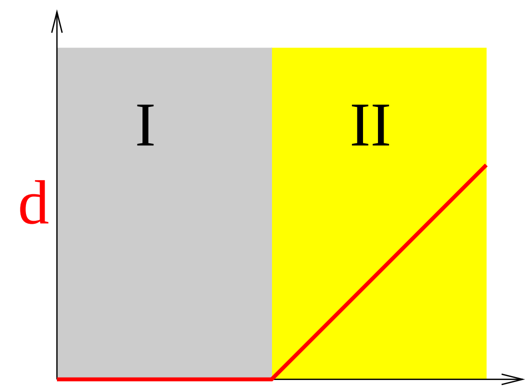
Entropy of perfect partitions

$$s_n(\kappa, b) = \frac{1}{n} \log Z$$



Hamming distance to sorted partition

$$d_n(b) = \frac{1}{2n} \sum_{j=1}^{n} (1 - \sigma_j^{\mathrm{opt}} \sigma_j^{\mathrm{sorted}})$$

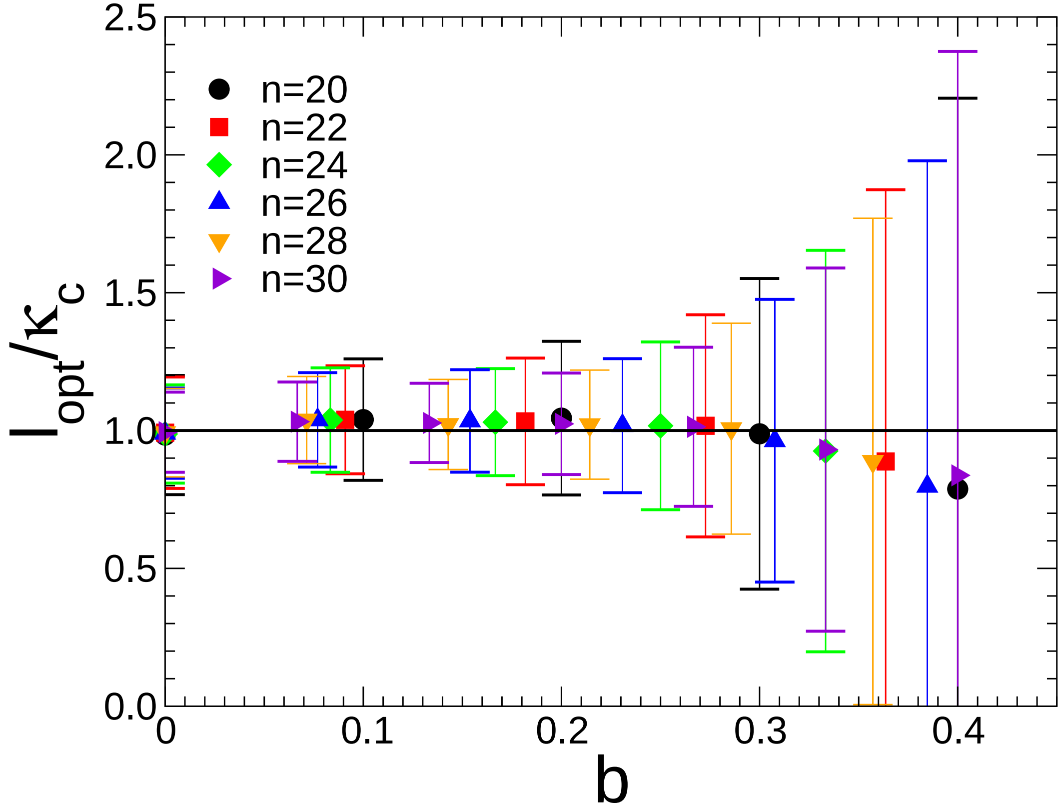


Optimal Discrepancy: Heuristic

- X_j's are κn-bit integers
- $d = |\sum_j \sigma_j X_j|$ is a κn-bit integer
- Phases II, III: at most $\kappa_c n$ bits of these can be 0
- choose the $\kappa_c n$ most significant bits:

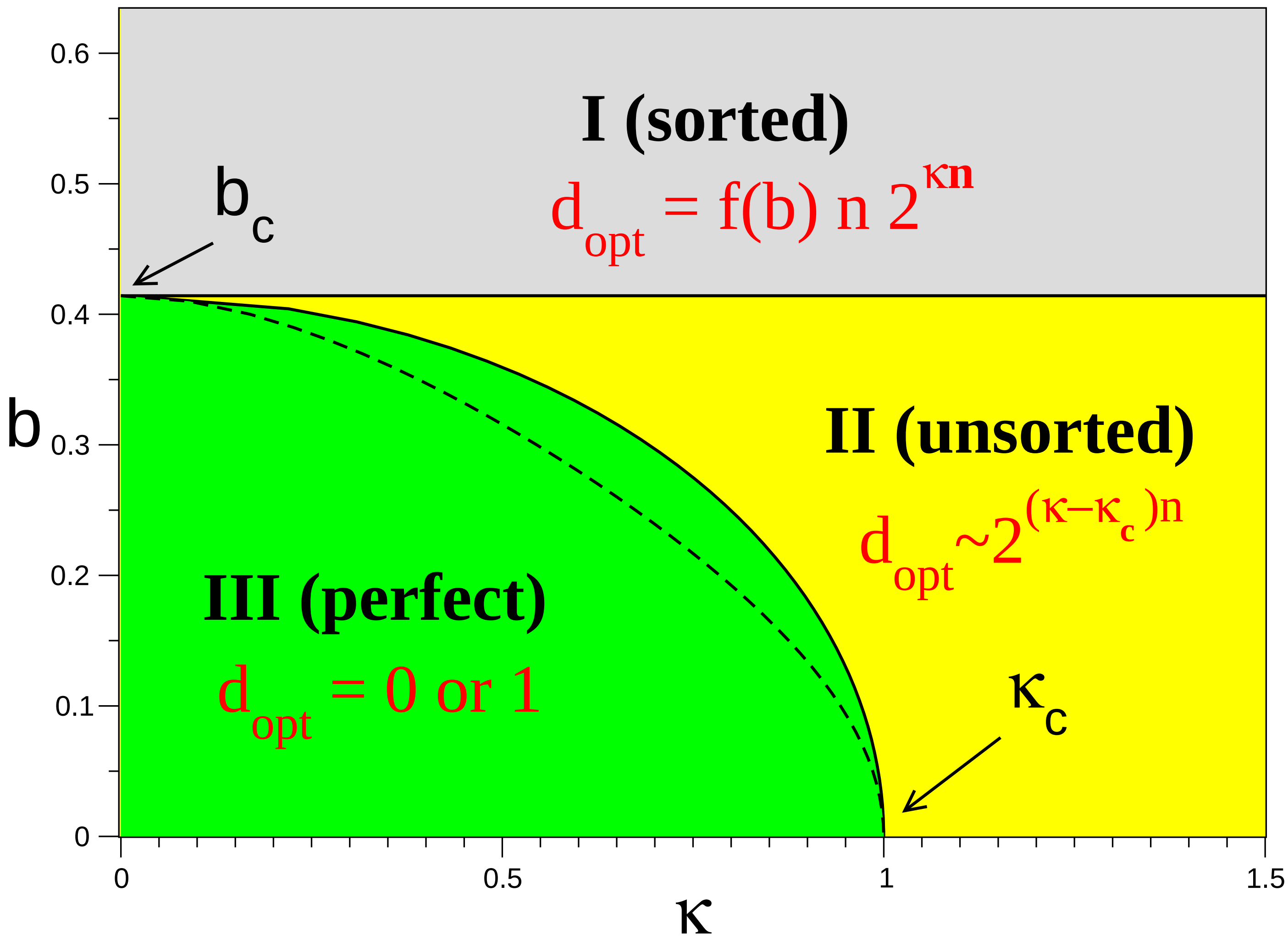
$$d_{\text{opt}}(\kappa, b) \approx 2^{[\kappa - \kappa_c]n}$$

Optimal Discrepancy in Phase II



$$l_{\text{opt}}(n) := \frac{1}{n}\log_2\left(\frac{2^{\kappa n}}{d_{\text{opt}}}\right)$$

Optimal Discrepancies



Conclusions

Constrained Integer Partitioning

- has two easy-hard phasetransitions

- control parameters: numerical resolution κ
  and imposed disbalance b

- order parameters: entropy of perfect partitions
  distance to sorted partition

- simple considerations yield reasonable results

- sharp rigoros results are in preparation