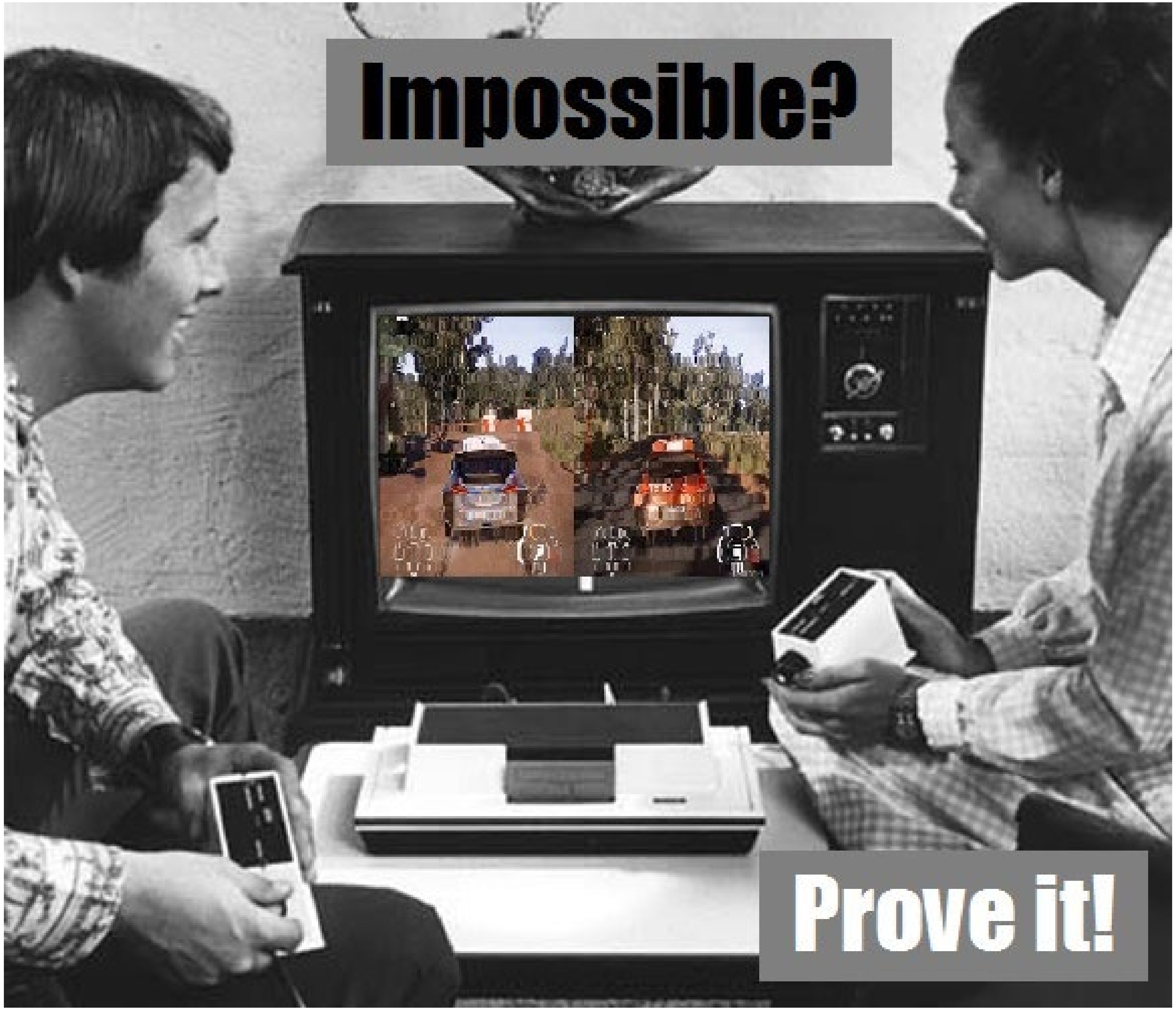
Correlation bounds and all that

Emanuele Viola

Northeastern University

2023 02 16



3

2

1

Announcements

- **Survey:**

  Correlation bounds against polynomials (2008)

  Revised 2022

- **Book:**

  Mathematics of the impossible:

  Computational Complexity

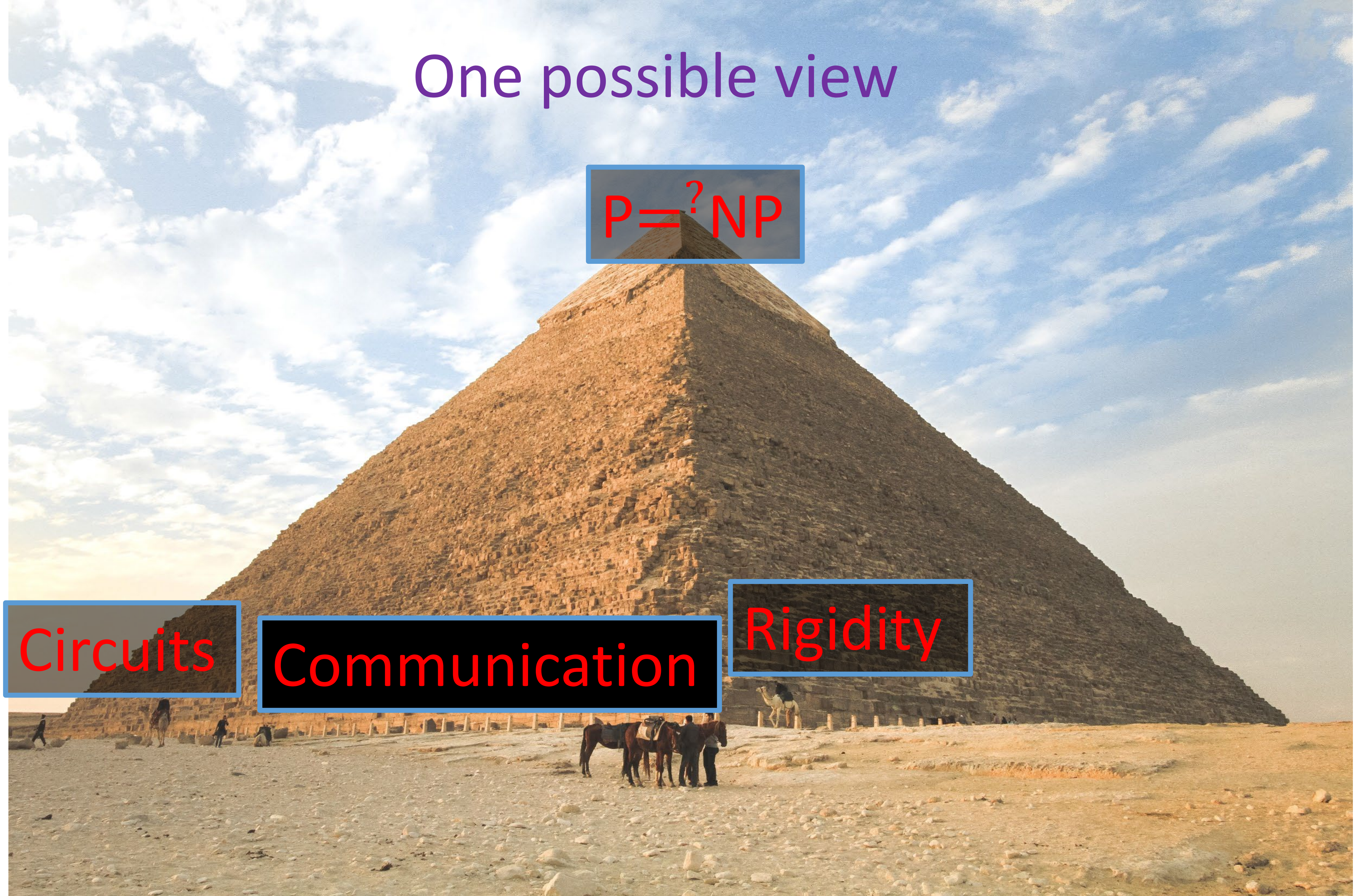
  Being serialized on my blog

One possible view

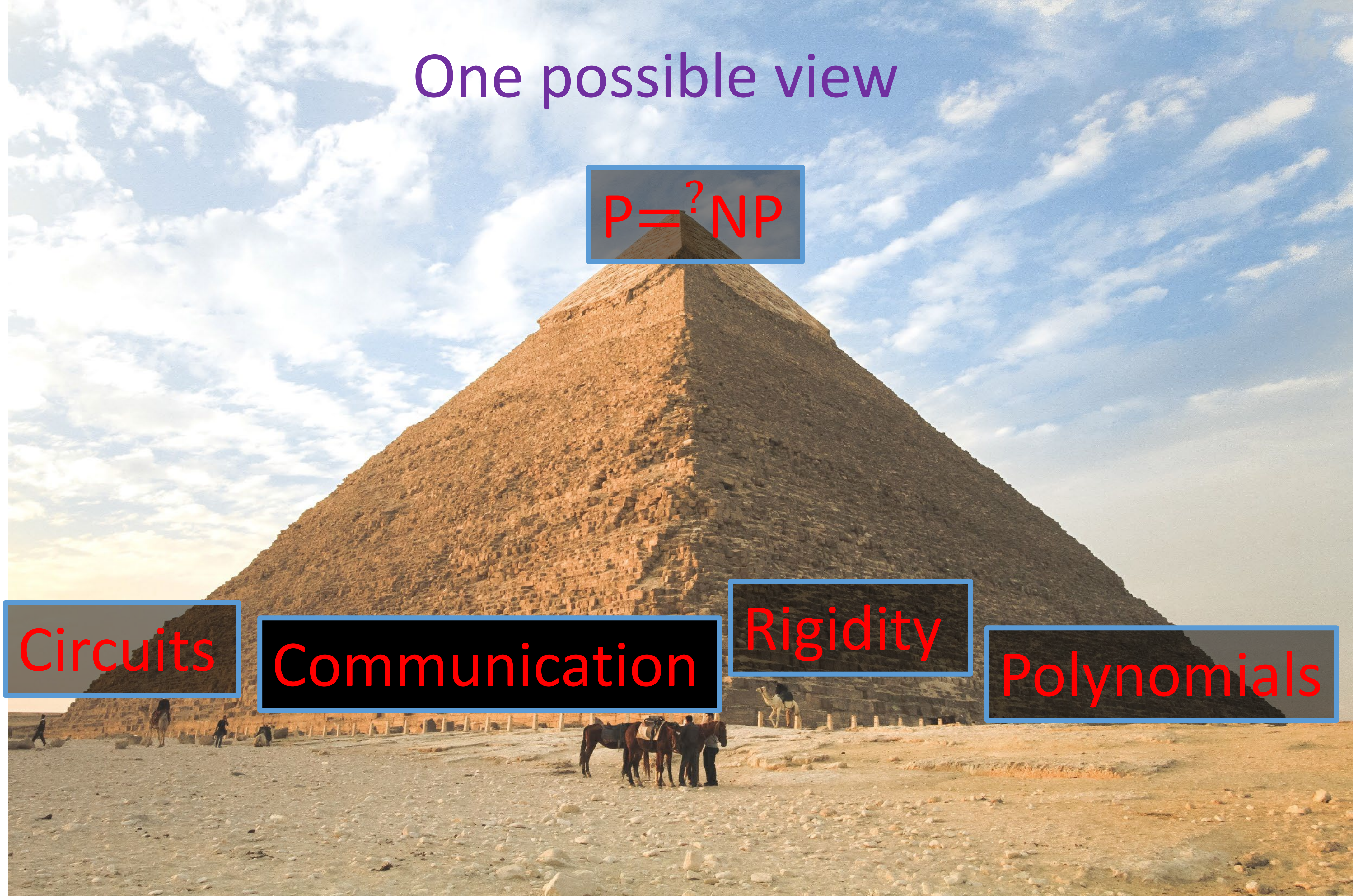
One possible view

One possible view

One possible view



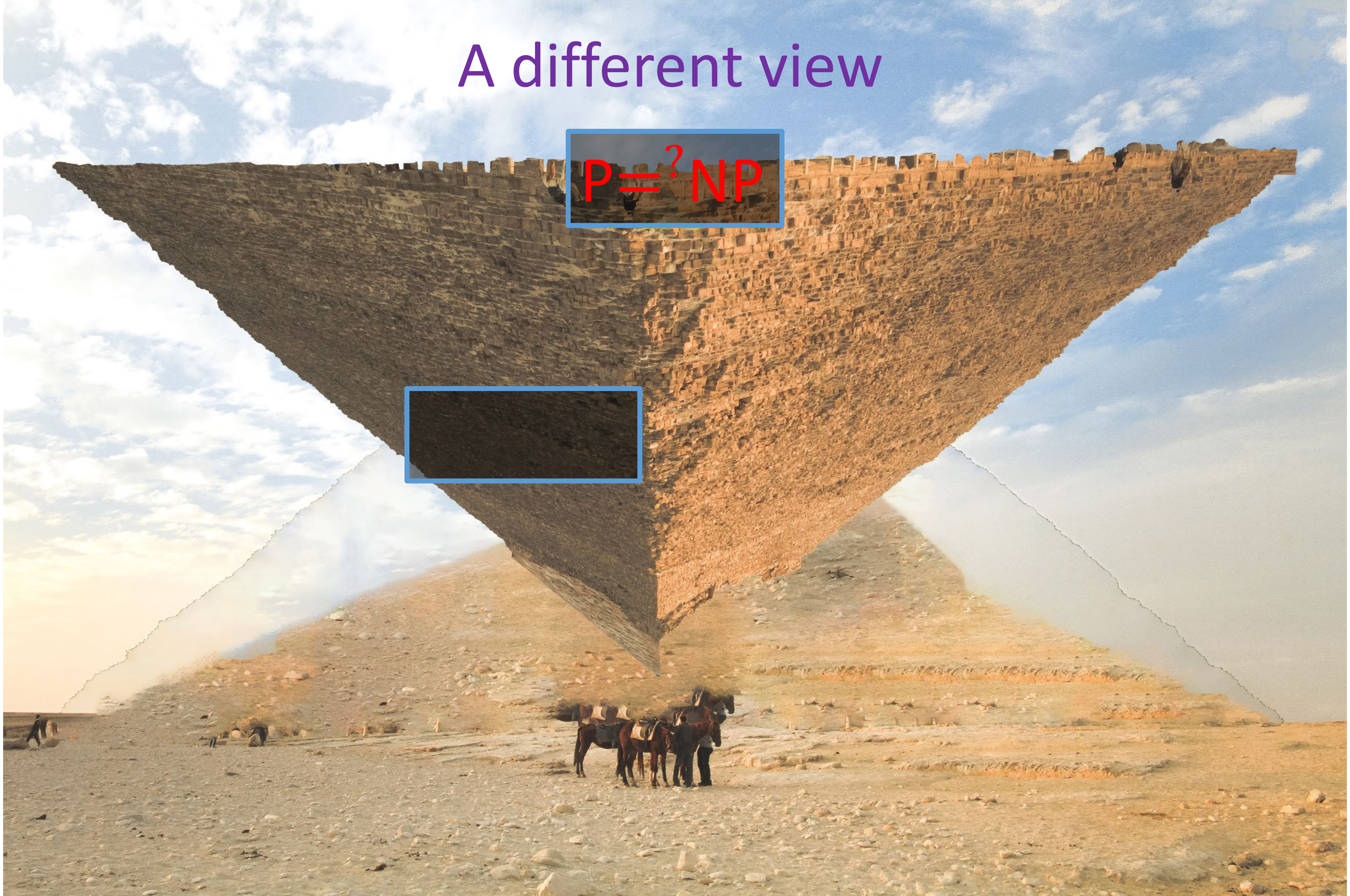
One possible view



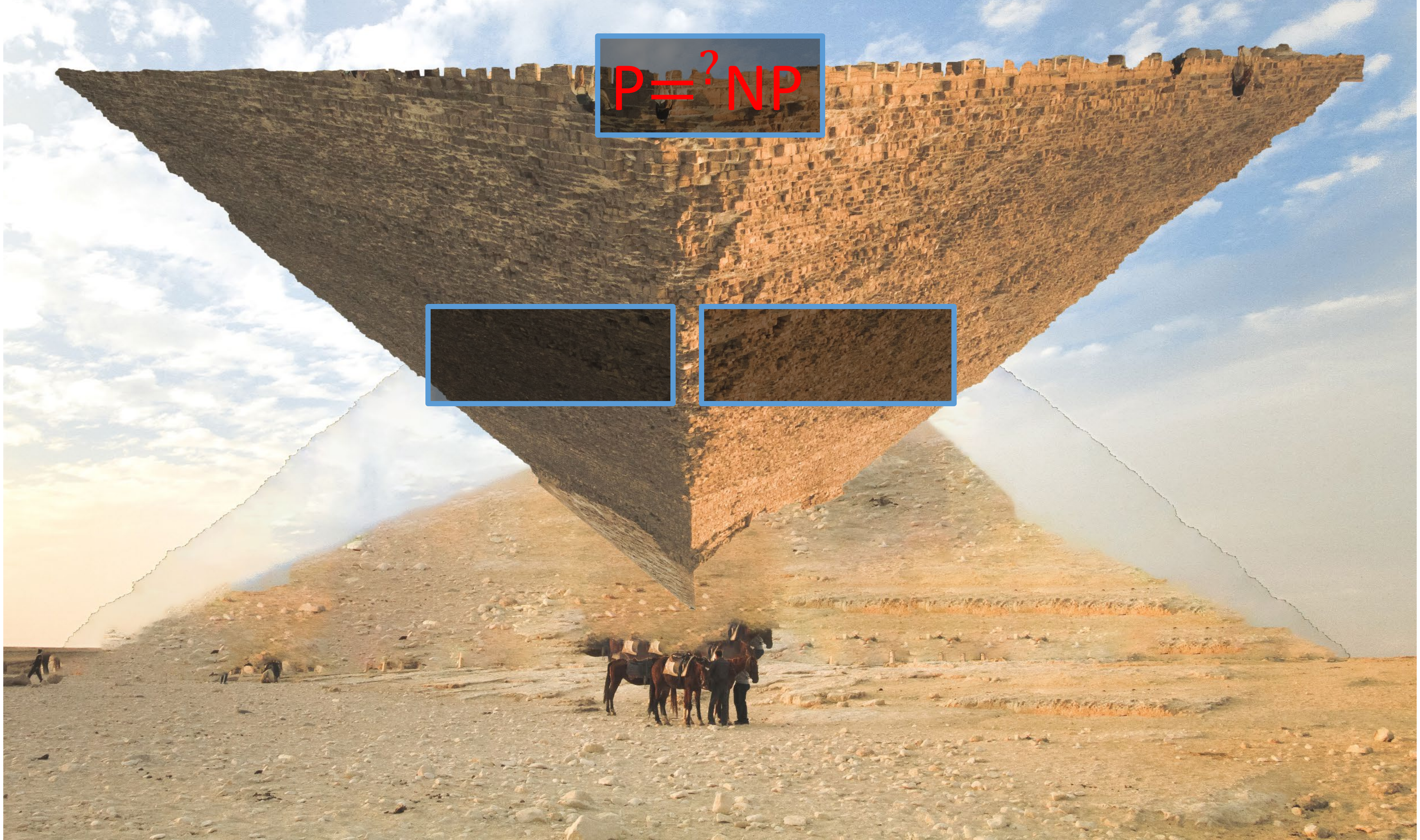
A different view

$P \stackrel{?}{=} NP$

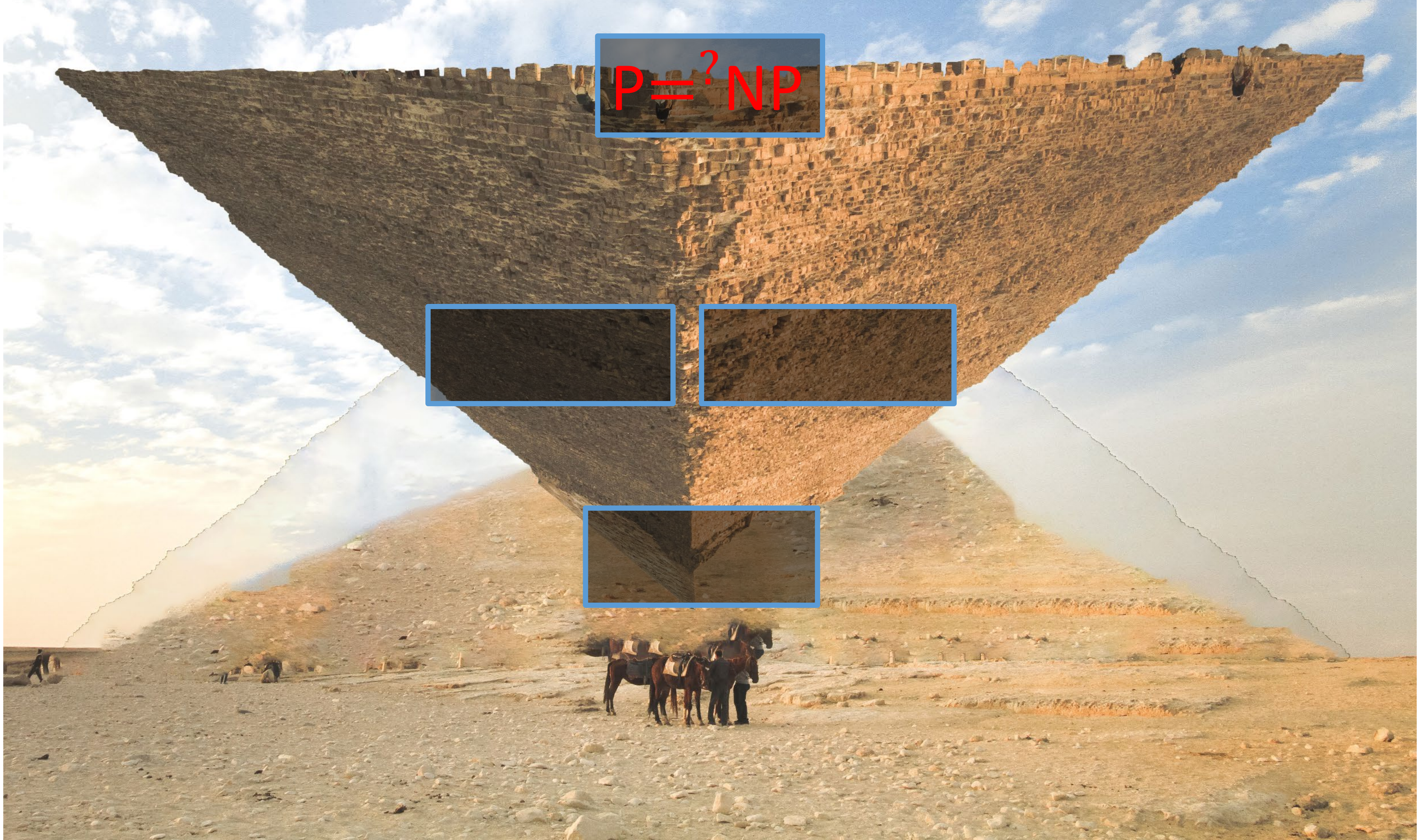
A different view



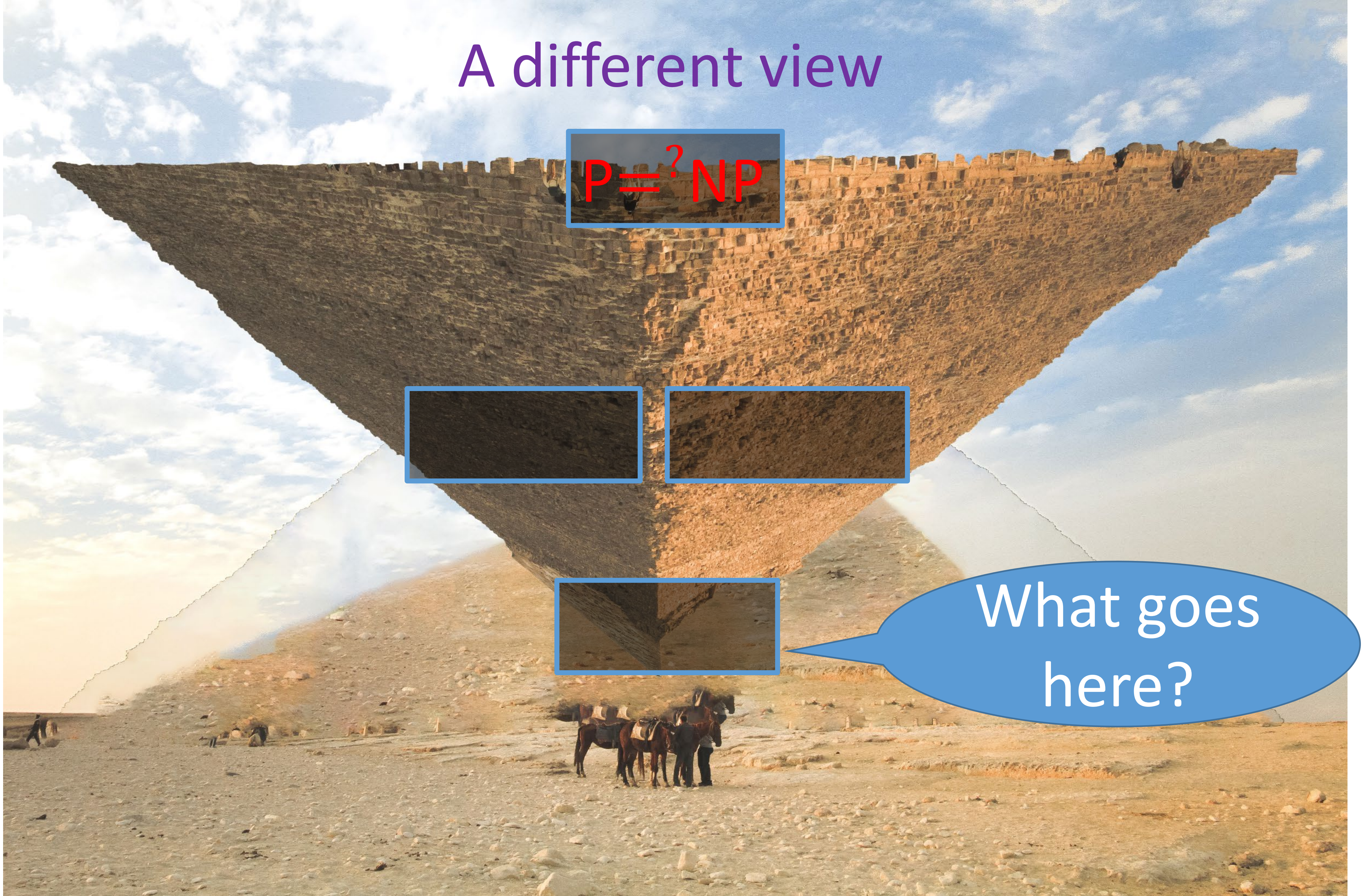
A different view



A different view



A different view



Frontier of P vs. NP

Circuit lower bounds

Frontier of P vs. NP

Circuit lower bounds

Matrix rigidity

Frontier of P vs. NP

Circuit lower bounds

Matrix rigidity

Correlation bounds for polynomials

Frontier of P vs. NP

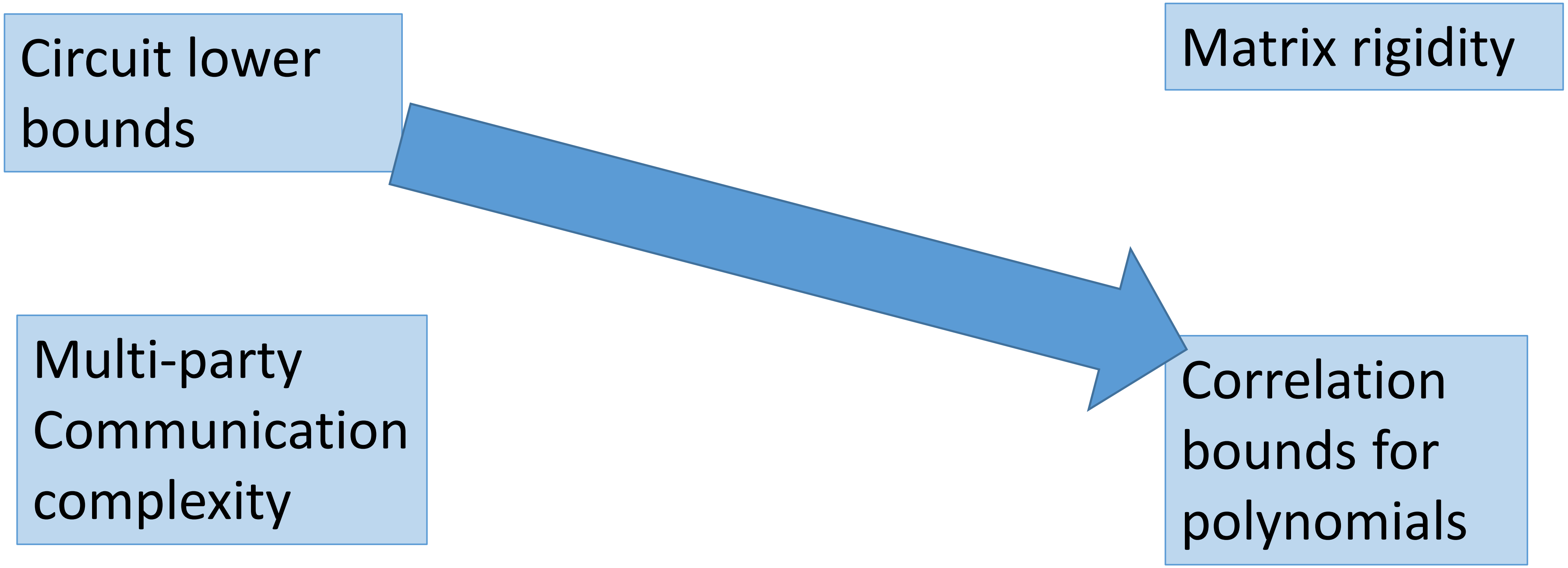
Circuit lower bounds

Matrix rigidity

Multi-party Communication complexity

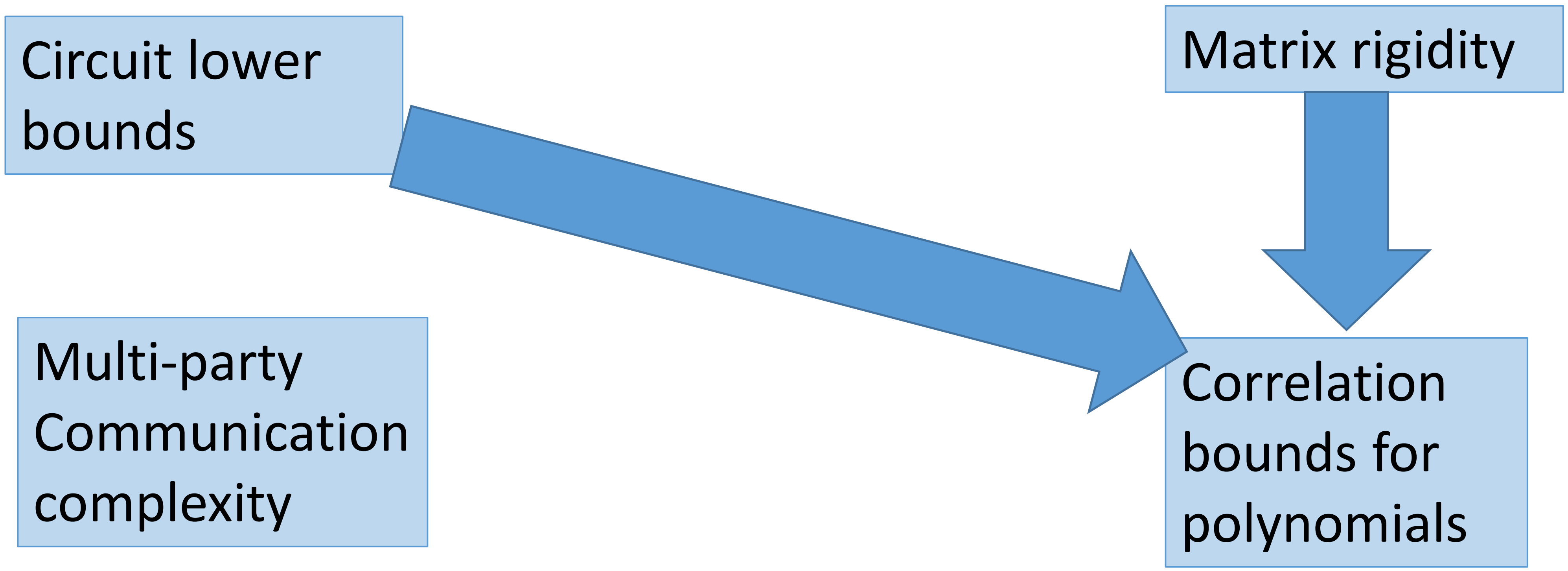
Correlation bounds for polynomials

Frontier of P vs. NP



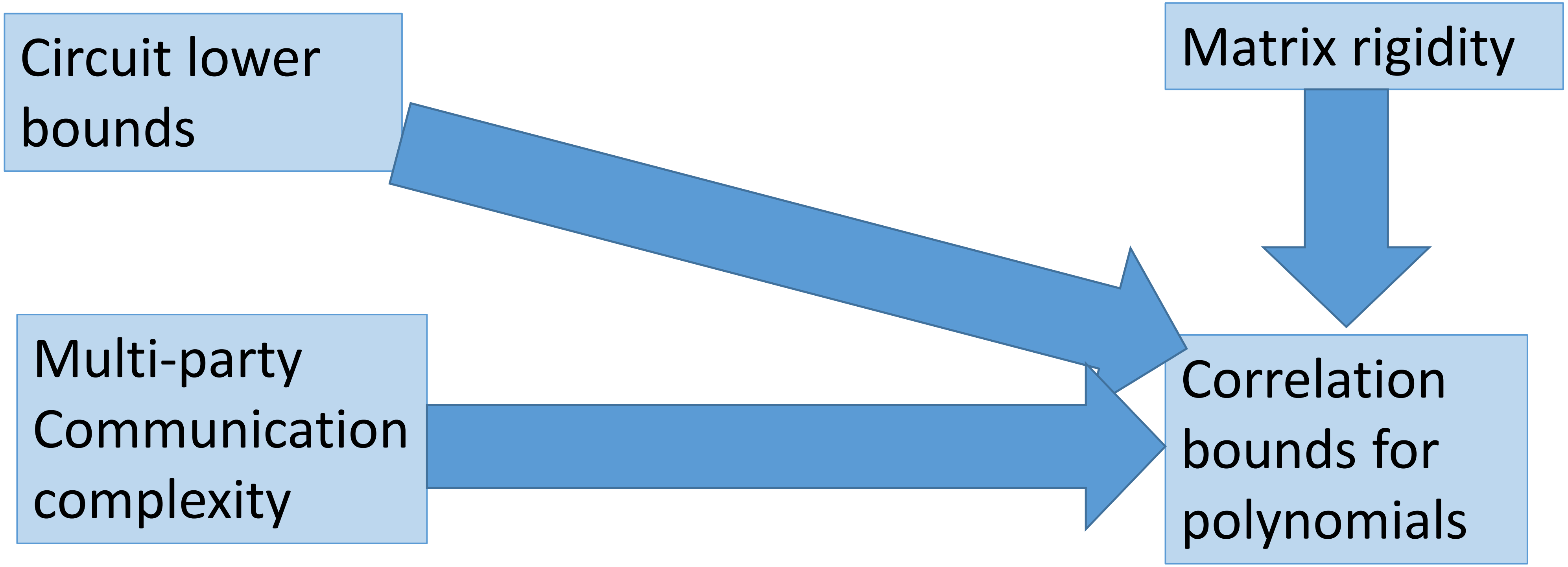
A $\Rightarrow$ B means progress on A requires progress on B

Frontier of P vs. NP



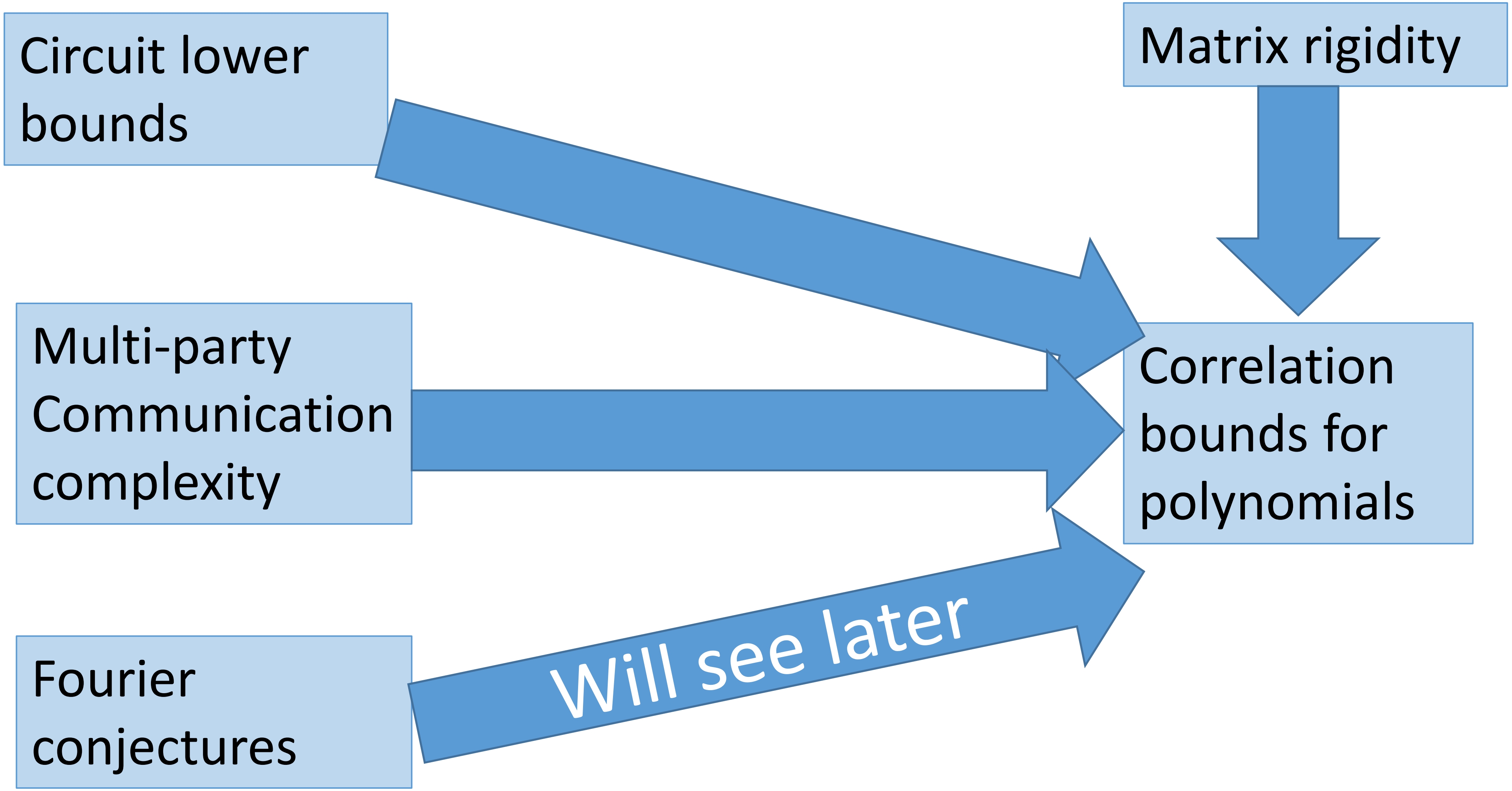
A → B means progress on A requires progress on B

Frontier of P vs. NP



A $\rightarrow$ B means progress on A requires progress on B

Frontier of P vs. NP



My view


$P \stackrel{?}{=} NP$
Circuits
Rigidity
Communication
Polynomials

Correlation bounds for polynomials

- Challenge: Find explicit $f: \{0,1\}^n \to \{0,1\}$ and distribution X such that for every polynomial p of degree d

$$Correlation(f,p) := Pr[f(X) = p(X)] \leq 1/2 + \epsilon$$

- Razborov, Smolenky, 80's: f = Majority, X = uniform, $\epsilon = O\left(\frac{d}{\sqrt{n}}\right)$
- Babai Nisan Szegedy 90's: f = GIP/Mod_3, $\epsilon = 2^{-\Omega(\frac{n}{2^d})}$
- Open: $\epsilon = 1/\sqrt{n}$ for $d = \log(n)$; required to solve any problem on previous slide

Overview

- Introduction
- A couple of recent results on correlation bounds
- Pseudorandom generators, and more recent results

[Chattopadhyay Hatami Hosseini Lovett Zuckerman] STOC 2020

- Def: Local correlation: $\Delta_S(F) := \boldsymbol{E}_{x_{-S}} \left[\boldsymbol{E}_{x_S} [F(x)] - E[F] \right]^2$
- Thm : $\forall\ degree - d\ \ F \quad \exists\, S : |S| \le 2^{poly(d)} :\ \ \Delta_S(F)$ small

  $\Rightarrow$ new correlation bounds for small degrees

- Conjecture : $|S| \le poly(d)$ suffices

  would imply dream correlation bounds for large degrees

[Ivanov Pavlovic V]

- Counterexample to CHHLZ conjecture
- Rules out even weak form, shows what they prove is best possible
- Proof sketch:

  Start with TRIBES DNF

  For any S of size about $n/\log n$: $\boldsymbol{E}_{x-S}$ [TRIBES = 1] $\geq \Omega(1)$

  $$\Rightarrow \left[\ \boldsymbol{E}_{x_S}\,[F(x)]\ - E[F]\right]^2 \text{ large}$$

  Approximate TRIBES by log(n)-degree polynomial F Qed

[Ivanov Pavlovic V]

- Conjecture: Symmetric polynomials maximize correlation with mod 3;
  would imply dream correlation bounds

- Prove the conjecture for degree 2 by “slowly opening directions”

- Prove the conjecture for special classes of degree 3

Overview

- Introduction

- A couple of recent results on correlation bounds

- Pseudorandom generators, and more recent results

Pseudorandom generators

- Explicit, low-entropy distributions that "look random" to polynomials

- Equivalent to correlation bounds for small error

- Case of large error remains unclear

- State-of-the-art [Bogdanov V 2007, Lovett, V]:
  To fool degree-d polynomials sum d independent generators for degree 1

- Can analyze up to $d < 0.01 \log n$. Beyond that is unknown (more later)

Fourier conjectures

- Polarizing random walks: Pseudorandom generators from Fourier bounds [2018 Chattopadhyay Hatami Hosseini Lovett, …]

- To improve generators for polynomials need Fourier conjectures:

$$\sum_{S:|S|=2} |\hat{p}_S| \quad \leq O(d^2)$$ [Chattopadhyay Hatami Lovett Tal]

$$\sum_{S:|S|=k} |\hat{p}_S| \quad \leq 2^{o(dk)}$$ [Chattopadhyay Gaitonde Lee Lovett Shetty]

- Theorem[V]: (Even weaker) conjectures
  $\Rightarrow$ correlation bounds beating Razborov-Smolensky,
  for functions related to majority (e.g., $\sum_{i<j} x_i x_j > 0$)

New correlation bounds

- We prove new correlation bounds which aim to, but don't, resolve conjectures
- Note: Correlation with Majority still open!
- Claim: Smolensky $O(\frac{d}{\sqrt{n}})$ bound for Majority tight under uniform distribution
- Claim: Can do $\Omega\left(\frac{d^2}{n}\right)$ for Majority under every distribution
- Conjecture: This is tight
- Claim: Conjecture holds (thus improving Smolensky) for $d = 1$

Next:

New pseudorandom generators using invariant theory

Pseudorandom generators against polynomials

- Definition:

R : $\{0,1\}^s \to F^n$ fools degree-d polynomials in n variables over finite field F if

$$\text{Statistical-Distance}(\, p\big(R(U)\big),\ p(U)\,) \leq \epsilon$$

for any such polynomial p; U = uniform distribution

Two lines of works

- **Small fields, e.g., {0,1}**

  [Naor Naor '92] Degree 1

  [Bogdanov-Viola '07] Paradigm: To fool degree d, sum d generators for degree 1

  Analysis [BV, Lovett, V '08]: seed length $O(\log n + 2^d)$

  Open problem: Does paradigm work for $d > logn$?

- **Large fields, |F| >> d**

  [Bogdanov '05] Reduces to hitting-set problem

  Optimal hitting sets [Klivans Spielman, B, Lu, Cohen Ta-Shma, Guruswami Xing]

  $\Rightarrow$ seed length $O(d^4 \log n + \log |F|)$, if $|F| > d^6$ Cannot get seed length $< d^2$

- Two lines followed different paradigms

[Derksen V]

- Analyze Bogdanov-Viola paradigm for large degrees over large fields
  $\Rightarrow$ new generators over large fields

- Theorem: Explicit generators against degree-d polynomials with seed length

  (1) Optimal $O(d\log n + \log|F|)$, if $|F| \geq d^4 n^{0.01}$

  (2) Nearly optimal $\tilde{O}(d\log n + log|F|)$, if $|F| \geq d^4 \log^4 n$

  (3) Matching previous best, if $|F| \geq d^4$ (previous work: d^6)
  Smallest possible |F| using Weil's bound

Proof overview

- Definition: Polynomial $g(x_1, x_2, \ldots, x_n)$ over F is decomposable if $g = c\big(h(x_1, x_2, \ldots, x_n)\big)$ for some univariate c of degree ≥ 2

- Lemma: g indecomposable $\Rightarrow$ $g(U)$ close to uniform

- Main Lemma: Construction of polynomials $f_1, f_2, \ldots, f_n$:
  - Few variables, low degree, and
  - preserve indecomposability: $h(f_1, f_2, \ldots, f_n)$ decomposable $\Rightarrow$ h decomposable

- Generator $R(U) := (f_1, f_2, \ldots, f_n)(U)$.

  Proof: Given g, write $g = c(h)$ for max degree c. Note h indecomposable

  $\Rightarrow \boldsymbol{g(U) = c\big(h(U)\big) \approx c(U) \approx c\big(h(f_1, f_2, \ldots, f_n)\big)(U) = g(f_1, f_2, \ldots, f_n)(U)}$

Definition of the f_i

- Let $M_1 . M_2, \ldots$ be all monomials in m variables (of some degree k)
- To fool degree d, take d copies $x^{[1]}, x^{[2]}, \ldots, x^{[d]}$ of the variables
- Define $f_i \coloneqq \sum_{j=1}^{d} M_i^{[j]}$ where $M_i^{[j]}$ is M_i on variables $x^{[j]}$
- "Algebraic" Bogdanov-Viola
  can take any polynomials M_i that fool degree-1 polynomials

Analysis of the f_i

- Assume: $G := g(f_1, f_2, \ldots, f_n)$ decomposable as $c\big(H(x_1, x_2, \ldots, x_n)\big)$.
  Goal: Show $g(x_1, x_2, \ldots, x_n)$ decomposable as $c\big(h(x_1, x_2, \ldots, x_n)\big)$

- G invariant under permuting the copies of the variables (the f_i are)
  $\Rightarrow H$ is invariant

- The f_i are basis for invariant polynomials
  $\Rightarrow H(x_1, x_2, \ldots, x_n) = h(f_1, f_2, \ldots, f_s)$ for some h (possibly $s >> n$)
  $\Rightarrow g(f_1, f_2, \ldots, f_n) = c\big(h(f_1, f_2, \ldots, f_s)\big)$.

- $\Rightarrow g(x_1, x_2, \ldots, x_n) = c\big(h(x_1, x_2, \ldots, x_s)\big)$ and $s = n$. QED

Analysis of the f_i

- We give 3 versions of analysis; different tradeoffs of simplicity and generality

- Can preserve indecomposability over any field, even {0,1}

- For generator, restriction on field size comes only from Weil's bound, used in Lemma: g indecomposable $\Rightarrow$ $g(U)$ close to uniform

A sense of the parameters

- Goal: fool $g(x_1, x_2, \dots, x_n)$ of degree d in n variables

- Pick n distinct monomials of degree k in m variables, need $\binom{m+k}{k} \geq n$

- Previous slides $\Rightarrow$ suffices to fool $g(f_1, f_2, \dots, f_n)$, degree dk in just dm variables

- E.g., set $m = O(\log n)\,, k = O(\log n)\,.$

- Setting uniform values for variables $\Rightarrow$ seed length $O(dm) = O(d\, log\, n \log |F|)$

- Improve to $O(d\, log\, n + \log |F|)$: combine with variant of [Bogdanov '05]
  - Non-standard: degree >> # variables; also better dependence on |F|

Future directions

- Goal: optimal seed length for field size $|F| = O(d^4)$
- May be possible with this approach given suitable extension of Weil's bound (work in progress)

Thanks!