My research focuses on problems in discrete probability theory, often arising out of questions in statistical physics and theoretical computer science. I am particularly interested in drawing connections with different areas of mathematics like complex analysis, partial differential equations, information theory etc., to solve problems in probability. Some of the areas of my current interest include the study of geometric properties of various polymer models in disordered media; scaling limits and fluctuation exponents of interfaces and phase transitions in various statistical mechanics models; large deviation and counting problems in graph theory and additive combinatorics. Below I will give a high level description of my contributions in the above and other areas of probability theory and describe certain key questions and general research programs that I plan to pursue. To maintain brevity I will refrain from providing precise statements of theorems as they would need more preparation. For more details please refer to my website [https://www.stat.berkeley.edu/~sganguly/](https://www.stat.berkeley.edu/~sganguly/).

1. **Disordered metric geometries on Euclidean space**

The field of disordered media has been the object of intensive mathematical research over the last thirty years. Of much interest are various non-Markovian path measures in such environments commonly known as polymers. Particular examples include models of First and Last Passage Percolation (henceforth to be called FPP and LPP) where, one assigns independent and identically distributed non-negative random weights to each edge or vertex on a lattice and studies the minimum (or maximum) weights of paths between two far away vertices. Formally the metric structure associated to First Passage Percolation (FPP) is defined as follows: on \( \mathbb{Z}^2 \) associate with each edge \( x \in \mathbb{Z}^2 \) an identical independent random variable \( \xi_x \). The polymer weight or the first passage time from \((0,0)\) to \((n,0)\) is defined as

\[
T_n = \min_{\pi} \sum_{x \in \pi} \xi_x,
\]

minimized over all paths \( \pi \in \mathbb{Z}^2 \) from \((0,0)\) to \((n,0)\). The extremal path \( \Gamma_n \) is hence a geodesic in the random metric space induced by the disorder. In directed last passage percolation (LPP) one instead maximizes the weight of all oriented paths between \((0,0)\) and \((n,n)\). A classical growth process known as the Eden Model considers balls of growing radii with respect to the metric induced by FPP (the corresponding process for LPP is known as the Corner Growth process), and one of the key objects of interest in such settings is the dynamics of the interface. The Kardar-Parisi-Zhang (KPZ) universality class includes a wide range of such interface models, in which growth in a direction normal to the surface competes with a surface tension. The law of large numbers i.e., that \( T_n \) converges to some deterministic constant almost surely, follows from standard sub-additive arguments. However the fluctuation theory is well developed only in the exactly solvable cases where the seminal works [Joh00, BDJ99] using powerful tools from probability, algebra and random matrix theory, verify the predictions from KPZ theory. Even though similar statements are believed to be true for FPP and other non-integrable models, the lack of formulae has been a major hindrance to mathematical progress.

**Large deviation of polymers.** Going beyond typical fluctuations, often in many models in statistical mechanics, a lot of information is obtained by a systematic study of rare behavior where one frequently encounters surprising geometric features (examples include behavior of zeros of random polynomials, eigenvalues of random matrices, solid on solid models etc.) This brings us within the realm of the theory of Large deviations which aims at precise understanding of rare events. There has been considerable study on large deviations of polymers in the exactly solvable case but most of the proofs in the literature rely heavily on exact formula and hence do not provide any geometric information about the effect of large deviation on the underlying metric structure or so and the proof techniques do not generalize to non-integrable models. Formally
in the latter setting, for $\delta > 0$ consider the events

$$\mathcal{L}_\delta := \{T_n \leq (1 - \delta)E(T_n)\} \text{ (lower tail)} \quad \text{and} \quad \mathcal{U}_\delta := \{T_n \geq (1 + \delta)E(T_n)\} \text{ (upper tail)}.$$  

It goes back to the work of Kesten [Kes86] in 1986 whose arguments imply that $\mathbb{P}(\mathcal{L}_\delta) = e^{-\Theta(n)}$ whereas $\mathbb{P}(\mathcal{U}_\delta) = e^{-\Theta(n^2)}$ under certain assumptions.  To establish a large deviation principle (LDP), one asks if the log-probabilities (rate functions) have a limit, i.e., whether the quantities $\log \mathbb{P}(\mathcal{L}_\delta)$ and $\log \mathbb{P}(\mathcal{U}_\delta)$ converge as $n$ grows to infinity. In many statistical mechanics models the rate function typically involves a variational problem which contains vital information about the underlying dynamics. In the above setting the lower tail problem can be attacked using a standard sub-additive argument, and even though the convergence of both in the context of the exactly solvable LPP was proven by Johansson [Joh00] crucially using tools from integrable probability, the above question about FPP had remained open since Kesten’s work.

- In joint work with Basu and Sly [BGS17b] using purely geometric arguments, we resolve this, establishing a LDP for a general class of FPP and non integrable LPP models.

A natural geometric object of interest is the polymer $\Gamma_n$ conditioned on these rare events. Deuschel and Zeitouni in 1998 [DZ99] studied such questions in the context of Poissonian LPP (another canonical exactly solvable LPP model with connections to maximal increasing subsequences of random permutations) and proved relying on analysis of Young tableaux, that conditioned on the upper tail event the polymer still stays localized near a straight line. However the situation for the lower tail event has been open since.

- With Basu and Sly in [BG17a] we resolve this question showing that in this case the polymer is delocalized and is macroscopically (linear in $n$) away from the straightline joining the endpoints. (see Figure 1)

In both the above works our arguments do not use any tools from integrable probability and is the first result of its kind for a general class of non-integrable models including FPP and high dimensional variants.

**Future Directions:** The above works open several directions for further research.

- An intriguing geometric question is whether the FPP metric conditioned on the upper tail event converges to some deterministic metric? The argument in [BG17b] essentially relies on sub-sequential limits which exists as a consequence of tightness. A natural approach towards answering the above, involves analyzing the rate function and showing that different sub-sequential limits cannot be optimal from the LDP perspective. Can one using the above strategy or otherwise deduce interesting geometric properties of this metric? A long term program is to understand behavior of Brownian motion via stochastic homogenization theory (see [Bis11]) on random metric spaces to deduce their conformal structure.

- Another question of considerable interest is to prove a large deviation principle for the Parabolic-Anderson model which is another canonical example of diffusion in a random potential. Cranston, Grathier and Mountford had found the order of decay for the large deviation events in the latter analogous to Kesten’s work for FPP, relating it to the tail behavior of the noise variables but a large deviation principle is still missing. We believe that the techniques we have developed have the potential to be useful in such settings.

**Extremal Isoperimetry and phase separation in disordered media:** Continuing with the topic of interfaces in random media under constraints, we end this section with a discussion about fluctuation of interfaces in a model related to LPP, constrained to exhibit global curvature. This causes a naturally occurring competition between the latter and local fluctuations appearing in formation of crystals and droplets formed in spin models. When the local fluctuation is of Brownian nature, this reflects an important aspect of the KPZ universality class.

- To understand fluctuations of such sets, with Basu and Hammond in [BG17a] we study the polymer in exactly solvable LPP under the additional requirement of enclosing an atypically large area thereby forcing the path to have a quadratic curvature and lose its integrable properties. We prove that the maximum length of the facets of the constrained path’s convex hull and the maximum local roughness or inward deviation are governed by exponents $3/4$, and $1/2$ respectively.

**Conjecture:** The motivation for our study comes from random walk on the unique infinite cluster $C_\infty$ in super-critical Bernoulli percolation on $\mathbb{Z}^2$; the behavior of which is intimately connected to geometric properties of $C_\infty$.

---

1. In the context of LPP, reversing the perspective of FPP, the probability of the upper tail event decays exponentially in $n$ whereas the lower tail event decays exponentially in $n^2$, instead.
such as the Anchored isoperimetric profile, (boundary length of isoperimetrically extremal sets \( B(r) \) with fixed volume \( r \), see [BLPR12]). The boundary of the latter can be interpreted as an FPP analogue to the interface that we study (see [BGH17a]) and hence alluding to universality, we conjecture that the corresponding exponents for \( B(r) \) are \( 3/4 \), and \( 1/2 \) as well. This complements the Wulff crystal shape theorem proved in [BLPR12].

2. LAPLACIAN GROWTH MODELS AND SELF-ORGANIZED CRITICALITY

Aside from the models in the KPZ universality class described above, the class of Laplacian growth models where the growth rate is determined by the harmonic measure of random walk, have also generated much interest. In these and other related models, simple local, stochastic or deterministic rules have global effects which is an important aspect of models of Self-organized criticality. I now describe various results including the resolution of various conjectures about such models, which provide a window into the tools of discrete potential theory: harmonic functions, martingales, etc.

2.1. Interface dynamics in Laplacian growth models. The so called Internal Diffusion limited aggregation (IDLA) is a canonical example, where on a graph, particles are emitted from a vertex and perform a random walk till they find an unvisited site which they occupy and stop. Lawler Bramson and Griffeath in their seminal work proved a limiting shape theorem for the growth cluster on \( \mathbb{Z}^2 \) and also bounds on the fluctuation. Study of interfaces in such context appear naturally in multi-species variants of IDLA and so on. Motivated by the above, Propp in 2003, introduced a competing version of IDLA and called it Competitive erosion. In this model, given an initial coloring of the vertices on a graph \( G \) with red and blue colors, red and blue particles are alternately emitted from vertices \( u \) and \( v \) respectively, which perform random walk on \( G \) and on hitting a particle of opposite color, removes the latter and occupies the site. When the given graph \( G \) is a discretization of a simply connected domain \( U \) in the complex plane, and the sources \( u, v \) are points on the boundary of \( U \), based on conformally invariant nature of Reflecting Brownian motion, Propp conjectured that as the mesh size goes to zero, at stationarity, the territories of the two colors would be separated by a geodesic with respect to the hyperbolic metric.

- With Peres in [GP15], via developing a convergence theory of solutions of discrete PDE’s to their continuous counterparts under Neumann (reflecting) boundary conditions, we confirm Propp’s conjecture on regular enough domains.

In another related case, when the graph \( G \) is an initially empty, infinite lattice \( \mathbb{Z}^d \) and red and blue particles start from the origin alternately and occupy the first site (excluding the origin) which is either empty or of the opposite color; although a priori not obvious, simulations indicate surprisingly coherent red and blue territories which are random unlike the above example.

- With Levine and Sarkar in [GLS17] we study the process on the line and show that the growth cluster has diameter \( O(n^{1/4}) \) after \( n \) rounds of the process and find the scaling limit of the random red and blue territories in terms of extremal functionals of Brownian motion.

Future Directions: The interface dynamics in all of these models, are governed by various curvature flows and it would be rather interesting to explore whether one can produce nice representation of the dynamics using results about such flows from the theory of PDE’s. Two particularly interesting cases, where experiments produce intriguing outcomes, are competitive erosion on the planar lattice \( \mathbb{Z}^2 \) or on a smooth compact domain with more than two colors and hence we have the following natural open problem for further work.

Question: What is the diameter of the growth cluster in planar competitive erosion at time \( n \)?

2.2. Absorbing state phase transition in models of self organized criticality. Other than shape theorems and fluctuations of interfaces, another widely studied aspect of particle systems, specially in models of non-equilibrium statistical mechanics, is whether it exhibits a phase transition as some underlying parameter, like the number of particles, is tuned. We consider a class of such
models of self organized criticality known as the Activated Random Walk (ARW(\(\mu, \lambda\))), where one starts with active particles at every site on the line with density \(\mu\). A particle can be either active or sleepy. Each active particle does a continuous time nearest neighbor symmetric random walk on \(\mathbb{Z}\) at rate one and falls asleep at rate \(\lambda > 0\). A sleepy particles does not move but becomes active when an active particle steps on it. For a fixed sleep rate \(\lambda\), as the particle density \(\mu\) increases, it is expected that the system shows a transition from almost sure local fixation to staying active forever almost surely. The long range correlation in such models makes them extremely challenging to analyze mathematically. One of the first rigorous results about ARW was established in [RS12] where it is shown that for every \(\lambda > 0\), there is a critical particle density \(\mu_c \in \left[ \frac{\lambda}{\lambda + 1}, 1 \right]\) such that ARW(\(\mu, \lambda\)) locally fixates almost surely when \(\mu < \mu_c\) and stays active almost surely when \(\mu > \mu_c\). However a well known conjecture in this area is that for any \(\lambda\), the critical density \(\mu_c\) is strictly less than 1.

- With Basu and Hoffman ([BGH17a]), we positively resolve the above conjecture for small \(\lambda\) and in fact show that indeed \(\mu_c\) goes to zero as \(\lambda\) tends to zero.
- Subsequently, jointly with the above coauthors, along with Richey in [BGHR17] we prove a parallel quantitative phase transition in terms of the number of steps taken by the process on the finite cycle \(\mathbb{Z}/n\mathbb{Z}\).

**Future Directions:** As already mentioned, there is a rich family of conjectures in the physics literature while only a handful of them have been verified rigorously. The following is a question analogous to the main result in [BGH17a] about the process on the plane that I plan to work on:

**Question:** Show that the critical density is strictly less than one for some value of \(\lambda\), for ARW on the plane. There has been some progress on understanding ARW on transient lattices but the planar version remains the most challenging case.

### 3. Large Deviation and Counting Problems in Sparse Settings

Recall that the polymer weight in (1.1) is an infimum of exponentially many linear functions of independent variables and hence is a highly non-linear function. At first glance, polynomials, seem much simpler non-linear functions. However the theory of large deviations for the latter is already highly non trivial specially when the underlying noise space is in some sense sparse. A canonical example of the above problems, concerns the count of triangles in a random graph which corresponds to a polynomial of degree 3 of independent variables. Formally, let \(G_{n,p}\) be the Erdős-Rényi random graph on \(n\) vertices where every edge is independently included with probability \(p\), and let \(X_H\) be the number of copies of a fixed graph \(H\) in it. Thus \(X_H\) can be thought of as a polynomial of independent Bernoulli variables. The upper tail problem for \(X_H\) asks to estimate the large deviation rate function given by

\[
R_H(n,p,\delta) := -\log \mathbb{P}(X_H \geq (1+\delta)\mathbb{E}[X_H])
\]

for fixed \(\delta > 0\). For \(H = K_3\) after a series of bounds, the correct order of the rate function was settled fairly recently by Chatterjee, and independently by DeMarco and Kahn. Subsequently much progress has been made in that front, propelled by the seminal work of Chatterjee and Varadhan who proved a large deviation principle in the dense regime \((0 < p < 1\) fixed) relying on Szemerédi’s regularity lemma via the theory of graph limits. In the sparse regime \((p \to 0)\), in the absence of graph limit tools, the understanding of large deviations for a fixed graph \(H\) (be it even a triangle) remained very limited until a breakthrough paper [CD] that reduced it to a natural entropic variational problem in a certain range of \(p\).

- With Bhattacharya, Lubetzky and Zhao [BGLZ17], we solve the variational problem for any fixed graph \(H\) in the sparse case when \(p\) goes to zero. Relying on various entropy inequalities and delicate combinatorial results about matchings, qualitatively, we show that asymptotically a sparse Erdős-Rényi random graph with more copies of a graph \(H\) than typical, looks like a typical Erdős-Rényi along with a planted clique or planted anti-clique to get the required boost in the \(H\) count\(]^2\).

A parallel story has been developed in the world of additive combinatorics where the object of interest is the count of arithmetic progressions in random subsets of \(\mathbb{Z}/n\mathbb{Z}\). Using their machinery, Chatterjee and Dembo also established a large deviation for the number of arithmetic progressions of length 3 in random subsets of

\[^2\text{An anti-clique is the graph obtained by taking a set of vertices and connecting them to all the vertices of the graph.}\]
relying crucially on the well known connections between classical Fourier analysis and progressions of length 3. However for higher order progressions this connection breaks down and the corresponding LDP was left open.

The beautiful theory of higher order Fourier analysis developed by Gowers, can be used to prove an LDP in such settings (see [BGSZ16]) but the bounds are not strong enough to let \( p \) decay to zero polynomially fast. Subsequently Eldan in [Eld16] using the theory of Gaussian processes, changed the condition of the gradient having a small covering number to the condition of having a small Gaussian width.

- By analyzing the associated Gaussian processes, and relying on other recently developed additive combinatorics tools, with Bhattacharya, Shao and Zhao in [BGSZ16], we establish an LDP for arithmetic progressions of arbitrary length in the polynomially sparse regime and solve the associated variational problems by reducing them to extremal problems in additive combinatorics.

**Future Directions:** In spite of an explosion of activity in this area in the last decade or so, several important questions still remain open. The above works establish a connection between large deviation problems and extremal problems in graph theory about maximizing or minimizing subgraph counts in a graph given the number of edges. One of my plans is to investigate this further, specially in the context of the lower tail problem which is largely unexplored, and drawing connections to the deep theory of flag algebra which was developed by Razborov, to answer Turan type problems in extremal graph theory.

**Question:** A concrete question in this direction is to find the phase diagram for large deviations for general graphs in the upper tail problem and to understand completely the case of triangles in the lower tail problem. As a follow up of the work in [BGLZ17], I plan to analyze the random graph measure conditioned on the upper tail event, in the dense case, i.e., when \( p \) does not go to zero with \( n \). Conjectures in statistical physics relate these conditioned measures to the so called **Stochastic Block Model**, a canonical example of a random graph exhibiting clustering.

### 4. Other interests

I summarize in this section some other areas of probability that I have worked on.

**Random walks, relaxation to equilibrium:** One of the fundamental quantities of interest about a random walk on a finite graph is its mixing time i.e., the time taken to converge to equilibrium. The last couple of decades have witnessed intense activity pioneered by Diaconis and Aldous to prove sharp mixing timing results also known as the Cutoff phenomenon in Markov chains which refer to an abrupt convergence to stationarity. Recently with Martinelli in [GM16], we considered random walk on the group of uni-upper triangular matrices with entries in \( \mathbb{F}_2 \) which forms an important example of a nilpotent group. We prove a cutoff result for the mixing of finitely many columns in the upper triangular matrix walk at the same location as the East process (a spin system that was introduced in the physics literature by Jäckle and Eisinger in 1991 to model the behavior of cooled liquids near the glass transition point) of the same dimension. Our proof relies on ingredients from our work jointly with Lubetzky in [GLM15] on cutoff results for the East process. An important related setting that I plan to work on in the future, is the random walk on the group \( \text{SL}_n(\mathbb{Z}_2) \).

**Growth rate of harmonic functions and speed of random walk:** Another much studied aspect of random walk is the relation between harmonic functions on graphs and speed of random walks that has generated a lot of activity in probability, geometric group theory and other related areas. It is a classical fact that the standard random walk \( \{X_n\} \) on \( \mathbb{Z}^d \) exhibits **diffusive behavior**. Recently jointly with Lee and Peres in [GLP17], solving an open problem from [BDCK+15] we show that there do not exist sub-linear harmonic functions on stationary graphs of polynomial growth by proving sub-diffusive behavior of random walk infinitely often, on stationary graphs of polynomial volume growth relying on tools from embedding theory. A natural direction that I want to pursue is to explore if such methods can be used to understand speed of random walk on critical percolation clusters.

**Gibbs measure on Random matrices and lattice gauge theory:** Lattice Gauge theories have been studied in the physics literature as discrete approximations to quantum Yang-Mills theory for a long time. Primary statistics of interest in these models are expectations of the so called “Wilson loop variables”. Jointly with Basu in [BGI17], we continue the program...
initiated by Chatterjee [Cha15] to understand Wilson loop expectations in Lattice Gauge theories in a certain limit through gauge-string duality. The objective in our work is to better understand the underlying combinatorics in the strong coupling regime, by giving a more geometric picture of string trajectories involving correspondence to objects such as decorated trees and non-crossing partitions. Using connections with Free Probability theory, we provide an elaborate description of loop expectations in the planar setting, which provides certain insights about structures of higher dimensional trajectories as well. A natural direction I want to pursue is to explore the weak coupling regime where the results in the above papers do not hold.

REFERENCES


