2022 ICTP-KIAS School on Statistical Physics for Life Sciences

Life can be thought of as a type of stochastic information processing realized out of thermodynamic equilibrium. Living systems indeed consume energy to sustain order and information flow in fluctuating environments. Recent advances in non-equilibrium statistical physics and information theory offer a new framework to unravel the underlying principles of the dynamical processes in living matter. As an example of the developing interface between biology, information theory and statistical physics, a recently discovered new principle, the thermodynamic uncertainty relation, quantifies the energy-precision trade-off of biological processes and its physical bounds. Another important example of the richness of this interface is offered by machine learning, which couples ideas from neuroscience, computer science and statistical physics.

The new language developed in modern statistical mechanics is then expected to play a key role in the understanding of how order, and eventually life, emerges from disorder. It has been widely demonstrated that statistical mechanics is an essential tool for the development of the above subjects, and great interest in these subjects is currently shared by statistical physics and biophysics communities in Korea. The 2022 ICTP-KIAS School of Statistical Physics for Life Sciences aims to introduce the next generation of students and researchers to these important themes: the quantitative understanding of life sciences, machine learning, and data science from the perspective of modern statistical physics. In addition to the pedagogical lectures, the school includes a small workshop for sharing recent research progresses among scientists at the frontier of this field.

Relevance to scientists in developing countries:

The school can provide unique opportunities for (junior) scientists in Asia Pacific regions to learn modern statistical physics and its applications to quantitative life science, machine learning, and data science. We plan to attract a large contingent of participants from the developing world. The organizers of this school have held the 2016 and 2018 Quantitative Life Science Workshops in Korea. The 2022 ICTP-KIAS school will contribute to the cross-fertilization between physics and biology as well as to long-term scientific collaborations between researchers in Asia and the ICTP scientists in the QLS section.

Venue: Korea Institute for Advanced Study (KIAS, www.kias.re.kr)

Time: Oct 31 - Nov 8, 2022

홈페이지: **https://indico.ictp.it/event/9836/overview**

Organizers:

Changbong Hyeon (KIAS, hyeoncb@kias.re.kr)

Junghyo Jo (Seoul National University, jojunghyo@snu.ac.kr)

Antonio Celani (ICTP, celani@ictp.it)

Scientific committee:

Kunihiko Kaneko (University of Tokyo, kaneko@complex.c.u-tokyo.ac.jp)

Tetsuya Kobayashi (University of Tokyo, tetsuya@sat.t.u-tokyo.ac.jp)

Leihan Tang (Hong Kong Baptist University, lhtang@hkbu.edu.hk)

Hyunggyu Park (KIAS, hgpark@kias.re.kr)

Yongjoo Baek (Seoul National University, y.baek@snu.ac.kr)

Yong Woon Kim (Korea Advanced Institute for Science and Technology, y.w.kim@kaist.ac.kr)

Jae-Hyung Jeon (Pohang University of Science and Technology, jeonjh@postech.ac.kr)

Jae Sung Lee (KIAS, jslee@kias.re.kr)

Deok-Sun Lee (KIAS, deoksunlee@kias.re.kr)

Time table for School Lectures:

	31(Mon)	1(Tue)	2(Wed)	3(Thu)	4(Fri)
9:30-10:45	Registration	Periwal 2	Mehta 1	Periwal 4	Mehta 3
10:45-11:00		Break	Break	Break	Break
11:00-12:15		Lee 2	Periwal 3	Mehta 2	Mehta 4
12:15-2:30	Lunch	Lunch	Lunch	Lunch	Lunch
2:30-3:45	Periwal 1	Discussion	Discussion	Sagawa 1	Sagawa 3
3:45-4:00	Break			Break	Break
4:00-5:15	Lee 1	Lee 3	Lee 4	Sagawa 2	Sagawa 4
	Dinner	Banquet	Dinner	Dinner	Dinner

Data science and statistical physics (Vipul Periwal)
Lecture 1: Comparing the inverse problem in statistical physics to model selection in data science
Lecture 2: Kinetic Ising models for modeling discrete data
Lecture 3: Temperature in data science
Lecture 4: Some new approaches to old problems

Machine learning and statistical physics (Pankaj Mehta)
Lecture 1: Fitting versus Predicting
Lecture 2: Towards Deep Learning: Linear and Logistic Regression
Lecture 3: Introduction to modern Deep Learning
Lecture 4: Why does modern Deep Learning work so well? Rethinking generalization.

Introduction to Stochastic Thermodynamics: from Fluctuation Theorems to Thermodynamic Trade-off relations (Jae Sung Lee)
Lecture 1: Work, Heat, and Stochastic Trajectory in Stochastic Systems
Lecture 2: Entropy Production and Fluctuation Theorems
Lecture 3: Thermodynamic Uncertainty Relations
Lecture 4: Thermodynamic Speed Limit

Thermodynamics of information (Takahiro Sagawa)
Lecture 1: An introduction to information theory
Lecture 2: Thermodynamics of information I
Lecture 3: Thermodynamics of information II
Lecture 4: Autonomous Maxwell's demons

Workshop

	Nov 7 (Monday)	Nov 8 (Tuesday)
9:30 - 10:00	Changbong Hyeon	Antonio Celani
10:10 - 10:40	Hyunggyu Park	Vipul Periwal
10:50 - 11:20	Yongjoo Baek	Deok-Sun Lee
11:30 - 12:00	Jae Kyoung Kim (online)	Junghyo Jo
12:00 - 2:00	Lunch	Lunch
2:00 - 2:30	Tetsuya Kobayashi	
2:40 - 3:10	Jae-Hyung Jeon	
3:20 - 3:50	Yong Woon Kim	
4:00 - 4:30	Kunihiko Kaneko (online)	
4:50 - 6:30	Poster session	

Nov 7 (morning session; chair:)

1. Changbong Hyeon (KIAS, hyeoncb@kias.re.kr)

   Molecular chaperones as nonequilibrium annealing machines

   Proteins and RNA are major constituents of the cell. They fold to specific 3D structures to be able to carry out biological functions. But, not all proteins and RNA fold nicely by themselves. Significant portions of them are prone to misfold, which could potentially be detrimental to the cellular function. To resolve such an issue, biological cells utilize ATP-burning molecular machineries to facilitate the folding of biopolymers. Here, I will address how chaperone activities change the population of functionally competent biopolymers over time, putting a particular emphasis on the nonequilibrium aspect of chaperone-assisted folding. If time allows, I will try to extend our discussion to optimal chaperoning activity for the processing of pre-RNA metabolism.

2. Hyunggyu Park (KIAS, hgpark@kias.re.kr)

   On the origin of the power-efficiency bound of a heat engine

An engine producing a finite power at the ideal (Carnot) efficiency is a dream engine, which is not prohibited by the thermodynamic second law. In 2011, a two-terminal heat engine with asymmetric Onsager coefficients was suggested by Benenti, Saito, and Casati, as a prototypical system to make such a dream come true with non-divergent system parameter values. However, such a system has never been found in spite of many trials. Here, we introduce an exactly solvable two-terminal Brownian heat engine, where asymmetric Onsager coefficients can be realized in the presence of a Lorenz (magnetic) force. Nevertheless, we show that the dream engine regime cannot be accessible even with asymmetric Onsager coefficients, due to an instability keeping the engine from reaching its steady state. This is consistent with recent trade-off relations between the engine power and efficiency, where the (cyclic) steady-state condition is implicitly presumed. We conclude that the inaccessibility to the dream engine is simply originated from the steady-state constraint on the engine.

3. Yongjoo Baek (Seoul National University, y.baek@snu.ac.kr)

A generic mechanism for symmetry-breaking motility in a dilute active fluid

A symmetric object in an active fluid may gain motility due to a ‘negative drag’ which applies in the direction of the object’s velocity. In previous studies, the phenomenon presupposed the presence of polar or nematic order in the active fluid. In this study, we show by a mean-field argument that such symmetry-breaking motility can generally emerge even in dilute and disordered active funds. The phenomenon manifests itself as both continuous and discontinuous transitions associated with the bifurcation of the steady-state velocity of the object. We also numerically show that the critical phenomena accounting for the continuous transition belong to the mean-field Ising universality class regardless of the shape of the object.

4. Jae Kyoung Kim (KAIST/IBS, jaekkim@ibs.re.kr)

Molecular mechanisms for filtering spatio-temporal noise in cells

The circadian clock generates ~24h rhythms everyday via a transcriptional-translational negative feedback loop. Although this involves the daily entry of repressor molecules into the nucleus after random diffusion through a crowded cytoplasm, the period remains extremely consistent. In this talk, I will describe how we used the combination of mathematical modeling and trafficking of clock molecules to identify a key molecular mechanism for such robustness of the circadian clock against spatio-temporal noise. This reveals that

cytoplasmic congestion is the major reason for why obesity, aging, and neurodegenerative disorders cause unstable sleep-wake cycles.

Nov 7 (afternoon session; chair:)

5. Tetsuya Kobayashi (University of Tokyo, tetsuya@sat.t.u-tokyo.ac.jp)

Chemical Thermodynamics of Growing Systems

Self-replication is the determinant feature that distinguishes living systems from non-living systems. It consists of the replication of molecular assemblies by autocatalysis and the growth of the volume that encapsulates them. The search for conditions necessary for the realization of self-replication dates back to von Neumann's theory of self-replicating automata and is deeply related to the question of the origin of life, i.e., how the mechanism of molecular replication emerged in primordial environments that cannot have sophisticated reaction mechanisms. In particular, because real cells are physico-chemical objects, seeking conditions for self-replication that are consistent with chemical thermodynamics and are physically feasible is one of the most important issues for the physical understanding of life.
However, most previous studies have focused on autocatalytic replications alone. On the other hand, cell-like structures and their volume growth, which enclose replication molecules and support the reaction fields, are also essential for the maintenance and evolution of self-replication mechanisms. In particular, if volume growth does not proceed in coordination with internal autocatalytic reactions in a thermodynamically consistent manner, the self-replication reactions can stop or even cycle reversely.
However, little has been known about the environmental conditions under which volume growth coupled with internal autocatalysis is thermodynamically permitted.
Also, there was no theory general enough to address questions, e.g., what characteristics the steady growth state with self-replication has, and how much thermodynamic cost is required to maintain the state.

We developed a new theory that can deal with a wide range of chemical thermodynamics[1,2,3], including those with volume growth. Based on this theory, we have derived the environmental conditions with which the fate of the system is determined: steady growth state, shrinking state, and equilibrium state[4].
We also clarified that the possible steady growth state is thermodynamically constrained, which could characterize the living state.
Moreover, the cost required for maintaining growth was obtained as the thermodynamic dissipation associated with the steady growth state.
The steady growth state is nonequilibrium but different from the conventional nonequilibrium steady state [5] because its origin is the extensivity of thermodynamic systems. If time permits, we also discuss the current limitation of our formulation and possible extensions to have a thorough thermodynamic description of self-replicating systems.

[1] Y.Sughiyama et al, A Hessian Geometric Structure of Chemical Thermodynamic Systems with Stoichiometric Constraints, Phys. Rev. Res (2022)
[2] T. J. Kobayashi et al, Kinetic Derivation of the Hessian Geometric Structure in Chemical Reaction Systems, Phys Rev. Res, (2022)
[3] D. Loutchko et al, Riemannian Geometry of Optimal Driving and Thermodynamic Length and its Application to Chemical Reaction Networks, Phys. Rev. Res to appear, (2022).
[4] Y.Sughiyama et al, Chemical thermodynamics for growing systems, Phys. Rev. Res. , (2022)

[5] T.J. Kobayashi et al, Hessian Geometry of Nonequilibrium Chemical Reaction Networks and Generalized Entropy Production Decompositions, Phys. Rev. Res. (2022)

6. Jae-Hyung Jeon (Pohang University of Science and Technology, jeonjh@postech.ac.kr)

7. Yong Woon Kim (Korea Advanced Institute for Science and Technology, y.w.kim@kaist.ac.kr)

First-passage dynamics of many Brownian particles

First-passage dynamics provides the fundamental understanding of many physical phenomena, ranging from diffusion-limited reactions of molecules to genetic drift in a population of organisms. In recent years, random target search problem received a great deal of attention in the context of first-passage dynamics, which includes abundant examples such as animal foraging, chemical reaction, and search for a lost item. Most previous studies in this field focused on the target search by a single particle. In reality, however, a group of particles is usually involved. In this talk, I will present recent advances related to the first-passage problems by many Brownian particles. For many particles, the problem becomes complicated even in the absence of interactions because the smallest value among first-passage times recorded by respective particles is only meaningful. I will also discuss what is the optimal initial particle distribution to minimize the first-passage time. If time allows, the effects of interactions among search agents are briefly mentioned.

8. Kunihiko Kaneko (University of Tokyo, kaneko@complex.c.u-tokyo.ac.jp)

Dynamical systems theory of robust cell differentiation and reprogramming

In development of multicellular organisms, cells differentiate into several distinct types. To depict the robustness of such differentiation process to noise and environmental perturbations, Waddington introduced the epigenetic landscape. However, how the landscape with branching of valleys (that correspond to cell types) is generated from the interplay between gene expression dynamics and epigenetic modification remains elusive. By introducing a theoretical model with a positive feedback reinforcement between the two, we demonstrate that the robust differentiation process (homeorhesis) is generally resulted, as an interplay between fast oscillatory gene expression and slow epigenetic modification.

By adopting the model, we then discuss “cell reprogramming” : differentiated cells can regain pluripotency, i.e., potential for differentiation, just by overexpressing only four genes. We demonstrated that global attraction to the initial unstable pluripotent-state is indeed possible just by overexpressing only few genes. The generality of this mechanism was confirmed by dynamical-systems analysis, as well as simulations with random and experimentally-extracted gene regulation networks. Robust differentiation and cellular reprogramming are thus universally explained.

References

Kaneko K., Life: An Introduction to Complex Systems Biology, Springer (2006)

Homeorhesis inWaddington's landscape by epigenetic feedback regulation, Y Matsushita, K Kaneko,Physical Review Research 2 (2020), 023083

Dynamical systems theory of cellular reprogramming, Y Matsushita, TS Hatakeyama, K Kaneko, Physical Review Research 4 (2022), L022008

Nov 8 (morning session; chair:)

9. Antonio Celani (ICTP, celani@ictp.it)

The uses of memory in olfactory search

Many organisms, from insects to mammals, have developed exquisite skills in searching for sources of odor from huge distances in turbulent atmospheric conditions. How are these sequences of very sparse detections translated into effective strategies to reach the source in the shortest time? Here we present a Reinforcement Learning algorithm that discovers such strategies with minimal memory requirements.

10. Vipul Periwal (NIH, vipulp@niddk.nih.gov)

Scaling persistent homology to large biological datasets

Manu Aggarwal[1], Vipul Periwal[1]

[1]Laboratory of Biological Modeling, NIDDK/NIH, Bethesda, MD

The structure and relative arrangement of the constituents of any biological system is crucial to its function due to existence of proximity-dependent interactions. Given data measuring the spatial embedding of such constituents, a pattern of interest is a region devoid of the constituents surrounded by a region of high density with constituents close enough to allow interaction, which we colloquially refer to as a hole. Such holes have been shown to have functional significance, for example, chromatin loops in chromosomes enable long range control of gene transcription and three-dimensional voids in protein crystal structures are related to ligand interaction. An algorithm to compute loops and voids is then needed to analyze the deluge of experimental data which often has large experimental uncertainties. Furthermore, identifying voids in a 3D embedding by visual inspection using the human eye is subjective and prone to inconsistencies. An objective mathematically sound method to detect holes and compute their statistical significance is required. Persistent homology (PH) is an approach to topological data analysis (TDA) that can compute the existence of holes in discrete data sets, assigning them a significance based on their robustness to experimental variability in the data set. This information comes at a high computational cost (run time and memory) that has limited applicability of PH to small data sets of a few thousand points. Further, it is commonly restricted to computing only the existence and significance of holes and not their location due to higher computational

costs and a lack of precision in computing their location. We developed Dory, an efficient and scalable algorithm for computing PH along with the location of significant holes with improved precision in large data sets. We used Dory to find protein homologs with significantly different topology by analyzing 180k publicly available crystal structures (PDB) and find chromatin loops in the human genome by analyzing high resolution Hi-C contact maps that result in point clouds with millions of points. In benchmarking different software, Dory was the only one that was able to analyze genome wide Hi-C contact maps. For validation of results, we show that the computed loops in Hi-C data sets and the voids in proteins agree with known biology.

11. Deok-Sun Lee (KIAS, deoksunlee@kias.re.kr)

Evolution of metabolic reactions' heterogeneous popularity

Different environments may create differences in the composition of the cellular metabolism across species. Thousands of bacterial species contain similar numbers of metabolic reactions but the cross-species popularity of reactions is so heterogenous that some reactions are found in all the species while others are in just few species, characterized by a power-law distribution with the exponent one. Introducing an evolutionary model concretizing the stochastic recruitment of chemical reactions into the metabolism of different species at different times and their inheritance to descendants, we show quantitatively how the exponential growth of the number of species containing a reaction and the saturated recruitment rate of brand-new reactions lead to the empirical power-law popularity distribution. The rate of recruiting brand-new reactions first grows exponentially and then saturates as more species are born, giving rise to a crossover in the popularity distribution. The future of the metabolism evolution is discussed within the proposed model.

12. Junghyo Jo (Seoul National University, jojunghyo@snu.ac.kr)

Mirror descent of Hopfield model

Mirror descent is a gradient descent method using a dual space of parametric models. The method has been developed in convex optimization problems, but not yet applied in machine learning. Using the Hopfield model as a prototype of neural networks, we demonstrate that the mirror descent can train the model more effectively than the usual gradient descent. The duality of the mirror descent allows a good parameter initialization, unlike the usual random parameter initialization. Our study can contribute to the way for data-driven parameter initialization of machine learning.