

力学系を組み込んだ変分自己符号化器による 顕微鏡動画モデリング：ゼニゴケ精子の 運動を例として

近藤 洋平^{1,2}・南野 尚紀³・上田 貴志^{2,3}

(受付 2023 年 4 月 27 日；改訂 7 月 16 日；採択 7 月 21 日)

要 旨

顕微鏡による観察は、フックによって細胞が発見された17世紀から現代に至るまで、生物学において最も重要な研究方法の一つである。また近年では、光学技術と蛍光タンパク質を用いたツールの発展によって、多くの現象がこれまでになく高い時空間分解能で観測可能になっている。しかし、生み出される顕微鏡動画データは複雑かつ高次元であり、そこから生物学的に興味のあるパラメータを抽出することは必ずしも容易ではない。この問題に対し我々は、教師なし特徴抽出のための変分自己符号化器に力学系を組み込んで顕微鏡動画をモデリングすることで、力学系の状態変数およびパラメータの形で観察対象の生物を特徴づける情報を得る枠組みを提案する。本稿では、植物学において重要な地位を獲得しつつあるモデル生物であるゼニゴケ (*Marchantia polymorpha*) の精子運動様式を対象に、提案する枠組みの有用性を検証した。力学系として位相振動子を用いることで、精子の回転運動の位相とその速度を、暗視野顕微鏡による高速イメージングデータから抽出することができた。

キーワード：力学系、変分自己符号化器、画像解析、植物学、ゼニゴケ。

1. はじめに

生物の形態や行動といった表現型と、遺伝子およびそれから翻訳されるタンパク質の作用とを関連づけることは、現代の生物学において基盤的なアプローチの一つである。例えば、陸上植物の起源の解明において重要な位置を占めているモデル種であるゼニゴケ (*Marchantia polymorpha*) について、その生殖に重要な精子の運動の様式とその分子メカニズムが植物学者の関心を集めている。そこで、精子の運動に変調をきたすよう遺伝子改変された様々な変異体が作出されている。しかし変異体の精子運動を分析する上で課題となるのは、データが顕微鏡動画という複雑かつ高次元の形式をとっている点である。そのような生の高次元データから運動を特徴づけるパラメータ、例えば泳動中の精子回転周期など、を抽出することは容易ではない。この難しさは、最もよく研究されているヒトの精子運動の様式についてさえ顕微鏡動画の分析に関する新たな論文が発表され続けていることからわかる(例えば Saggiorato et al., 2017)。

¹ 自然科学研究機構 生命創成探究センター：〒444-8787 愛知県岡崎市明大寺町東山5-1

² 総合研究大学院大学 先端学術院：〒444-8787 愛知県岡崎市明大寺町東山5-1

³ 基礎生物学研究所：〒444-8585 愛知県岡崎市明大寺町西郷中38

そこで本論文では、精子運動の力学系モデルと変分自己符号化器 (Variational auto-encoder, VAE) (Kingma and Welling, 2014; Rezende et al., 2014) による教師なし特徴抽出の組み合わせによって、顕微鏡動画からの直接的な精子運動周期の抽出を試みる。すなわち、顕微鏡動画を自己符号化器のエンコーダーによって低次元化し、その特徴空間における時間発展を力学系によって制約し、自己符号化器のデコーダーによって未来の時点での顕微鏡画像を予測するという設定の上で、データに基づき力学系と自己符号化を誤差逆伝播法によって同時に学習する。これによって、学習後の力学系の状態変数とパラメータという形で精子回転運動の特徴が抽出できると期待される。

精子回転運動の力学系モデルとして、精子と周囲の流体力学まで考慮した詳細なものも選択肢にあるが、ここでは回転運動を抽象化した単純な振動子を考える。振動子モデル、特に振幅の自由度を無視した位相振動子は、単細胞の緑藻であるクラミドモナスの鞭毛が示す回転同期現象のモデル化などに既に応用がある (Klindt et al., 2017)。また、振動子を採用した状態空間モデルによって時系列から振動成分を抽出する振動子分解と呼ばれる手法も脳イメージングデータを対象に成功を納めており (Matsuda et al., 2022)、生物学分野でのデータ同化の文脈における振動子の有効性を示している。

周辺分野における関連する研究として、ロボティクス等への応用を念頭に、力学系モデルと教師なし学習による特徴抽出器の組み合わせによって動画から直接に力学系モデルのパラメータを推定する試みがなされている。ただし、問題の難しさから、撮影されている物体のセグメンテーションやトラッキングといった追加の情報を利用する手法が多い。そういった追加情報を用いず動画データのみから学習を行う試みとしては Le Guen らの PhyDNet (Le Guen and Thome, 2020) や Jaques らの手法 (Jaques et al., 2020) が挙げられるが、成功しているのは少数の質点の運動といった極めて単純な動態や人工データを扱うケースに限られ、応用までには距離がある。本研究も単純な動画および力学系モデルにおけるものではあるが、実際のデータを用いた生物学的に興味深いパラメータの定量化として、より広い文脈においても意義深い例であると言える。

2. モデルと学習

顕微鏡動画に対して、画像を低次元化するためのエンコーダー・デコーダーと、低次元の中間層における時間発展を記述する位相振動子を考える。モデルの概要を図1に示す。このモデルに入力される長さ T の動画を $X = [x_0, x_1, \dots, x_{T-1}]$, $x_i \in \mathbb{R}^{H \times W}$ とすると、エンコーダーを介して中間層の変数は次のようにサンプリングされる。

$$(2.1) \quad z = \text{Enc}(X)$$

$$(2.2) \quad \phi_0 \sim \text{WrappedNormal}(\phi; \mu_\phi(z), V_\phi)$$

$$(2.3) \quad p \sim \text{Normal}(p; \mu(z), V(z)),$$

ここで $p \in \mathbb{R}^P$ は、例えば画像に写り込む精子以外の夾雑物の有無といった、位相の進行に依らない画像間の変動を記述するための非位相変数である。本研究では p の次元として常に $P = 10$ を用いた。位相変数の初期値 $\phi_0 \in [0, 2\pi)$ は巻き込み正規分布、非位相変数 p は正規分布に従い、それらの分布の母数 $\mu_\phi(z), \mu(z), V(z)$ は潜在変数 z から

$$(2.4) \quad (\mu_\phi(z), \mu(z), V(z))^\top = Bz + C$$

と計算される。 $B \in \mathbb{R}^{3 \times P}$, $C \in \mathbb{R}^3$ はエンコーダーのパラメータと同時に学習によって決まる。一方 V_ϕ は目的関数の評価を簡単にするために定数として扱う。本研究のモデルでは簡単なた

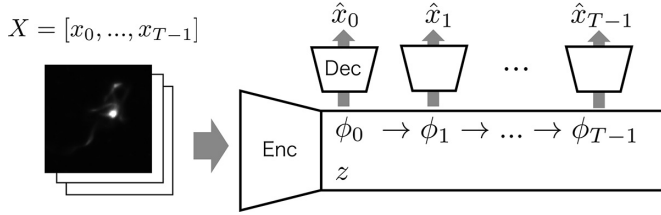


図 1. モデルの概要. 変分自己符号化器の中間層に位相振動子が埋め込まれ, 顕微鏡動画を自己符号化する.

表 1. エンコーダーおよびデコーダーの構造. In: 入力チャンネルサイズ, Out: 出力チャンネルサイズ, Conv: 畳み込みニューラルネットワーク, Trans Conv: 転置畳み込みニューラルネットワーク, SiLU: シグモイド線形ユニット関数 $\text{SiLU}(x) = x/(1 + e^{-x})$. 全ての畳み込み層においてフィルター適用間隔は 2 とした.

	Layer type	In	Out	Filter size	Nonlinear Transformation
Encoder	Conv	T	32	(4, 4)	SiLU
	Conv	32	32	(4, 4)	SiLU
	Conv	32	64	(4, 4)	SiLU
	Conv	64	64	(4, 4)	SiLU
Decoder	Linear	2 + P	$64 \times 6 \times 6$		SiLU
	Trans Conv	64	64	(4, 4)	SiLU
	Trans Conv	64	32	(4, 4)	SiLU
	Trans Conv	32	32	(4, 4)	SiLU
	Trans Conv	32	1	(4, 4)	Sigmoid

めに振幅自由度を無視した位相振動子を採用し, 位相の時間発展を次のように書く.

$$(2.5) \quad \phi_{t+1} = \phi_t + \omega \Delta t.$$

位相速度 ω は初期値を $[0, 0.01]$ の一様分布からサンプルし, ニューラルネットワークのパラメータと同時に学習する.

これら位相変数と非位相変数は, デコーダーによって対応する時点の顕微鏡画像へと復号される.

$$(2.6) \quad \hat{x}_t = \text{Dec}(\cos(\phi_t), \sin(\phi_t), p).$$

そしてこのモデルは元の動画を再構成した $\hat{X} = [\hat{x}_0, \dots, \hat{x}_{T-1}]$ を出力する.

本研究で採用したエンコーダーとデコーダーの構造については表 1 にまとめた. 以上の設定によって, このモデルのパラメータの総数は 318,807 となる.

上で定義したモデルのような確率的自己符号化器を学習するための目的関数として, 変分サイズ法の処方箋に従い, 次のような対数周辺尤度の変分下限が用いられる.

$$(2.7) \quad \log p(X) \geq \mathbf{E}_{q(z|X)}[\log p(X|z)] - D_{\text{KL}}(q(z|X)||p(z)),$$

ここで D_{KL} はカルバック・ライブラー情報量を表す. 提案分布 $q(z|X)$ はエンコーダーによってモデル化される. このように隠れ変数の事後分布を入力から推定するニューラルネットワークを学習するという手法は Amortized inference と呼ばれ深層学習の文脈で多くの例がある. 物理モデルを変分自己符号化器に組み込んだ定式化を扱っている Takeishi and Kalousis, 2021 を

参考にすると、本研究のモデルでは目的関数は

$$(2.8) \quad L(\theta, X) = \mathbf{E}_{q(\phi_0, p|X, \theta)} \log p(X|\phi_0, p, \theta) - D_{\text{KL}}(q(\phi_0, p|X)||p(\phi_0, p))$$

となる。ここでは学習されるモデルパラメータをまとめて θ と書いており、 θ はエンコーダー・デコーダーのパラメータに加えて式(2.4)の B, C と位相速度 ω を含んでいる。右辺第一項は、決定論的なデコーダーについては観測ノイズとして各次元で独立な標準正規分布を考慮して二乗誤差とすることが多く、本研究もそれに従う。右辺第二項については、隠れ変数 (ϕ_0, p) の事前分布がそれぞれ一様分布と標準正規分布とし、位相変数に対する提案分布である巻き込み正規分布に関して分散は学習しないため、学習に影響するのは次のように容易に計算できる部分のみである。

$$(2.9) \quad D_{\text{KL}}(q(\phi_0, p|X)||p(\phi_0, p)) \rightarrow D_{\text{KL}}(\text{Normal}(\mu(X, \theta), V(X, \theta))||\text{Normal}(0, I)).$$

この $\mu(X, \theta), V(X, \theta)$ は式(2.3)に現れる非位相変数の平均 $\mu(X)$ と分散 $V(X)$ と同じものであるが、ここではパラメータ依存性を明示した。

変分下限を目的関数とする確率的自己符号化器すなわち変分自己符号化器の最適化には、計算コストの軽減と局所解の回避を兼ねて、勾配法によるパラメータの更新毎にデータの一部のみを用いる確率的勾配降下法が使用される。本研究でも、全体の長さ $N(> T)$ の動画データを

$$(2.10) \quad [x_0, \dots, x_{N-1}] \rightarrow [x_{0:T-1}, x_{1:T}, \dots, x_{N-T:N-1}]$$

のように分割した上で確率勾配降下法によってモデルを適用した。具体的には、深層学習ライブラリの PyTorch (version 2.0.0) (Paszke et al., 2019) を用いて、本研究で扱う顕微鏡動画それぞれについて学習率 $\eta = 0.0003$ とした Adam オプティマイザーによって 1000 エポック最適化した。本研究では、予測でなくデータ圧縮が主眼であること、目的関数が対数周辺尤度の変分下限であることから、バリデーションデータを利用した Early stopping は実施しなかった。ただし、変分自己符号化器においてもエンコーダーの過学習がおこりうることを示されており (Zhang et al., 2022)、それに対処するための学習アルゴリズムも提案されているが、その実効性は検証途上であると考えられるため、本研究では採用しなかった。

3. データ

先行研究の手順 (Minamino et al., in preparation) に従い、我々はゼニゴケの野生株 Takaragaike-1 から蒸留水によって精子懸濁液を採取し、暗視野顕微鏡によって観察した。そこで得られた 30 ms 間隔の経時観察動画をもとに画像解析ソフトウェア ilastik (Berg et al., 2019) によって個々の精子をトラッキングし、96 px \times 96 px のサイズの動画データを八本作成した。八本の動画のフレーム数は 100, 114, 170, 136, 100, 100, 135, 76 である。

4. 結果と考察

予測時点数を $T = 20$ 、非位相変数の次元を $P = 10$ 、初期位相の分散を $V_\phi = (0.02)^2$ とした学習の結果、位相振動子を埋め込んだ変分自己符号化器は精子の回転運動を再構成するようになった。目的関数値およびその一部である KL 罰則値、すなわち式(2.8)の右辺第二項の値、のエポック履歴を図 2 に示す。本研究で解析した動画間では目的関数値の振る舞いに顕著な違いがみられない。より詳しく調べるために、典型的な再構成例を図 3 に示す。エンコーダー・デコーダーの学習には動画全長を用いているために必ずしも驚くべきことではないが、隠れ変数の推定に用いた入力動画の一部 ($x_{20:19+T}$) よりも先の時点についても、安定した画像再構成が

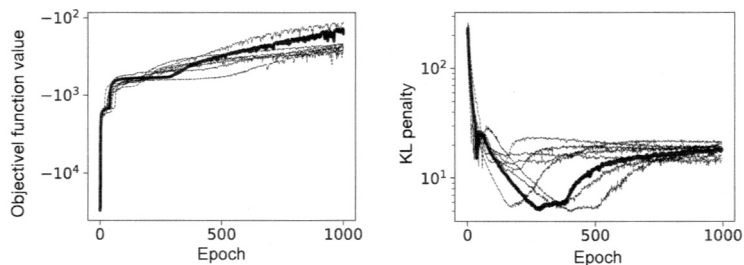


図 2. 八本の精子運動動画データそれぞれについての目的関数値(左)および KL 罰則値(右)のエポック履歴. いずれも確率勾配法におけるミニバッチあたりの値であり, 各ミニバッチは 160 枚の画像を含んでいる. 図 3-5 で用いた動画に関する結果を実線で示した.

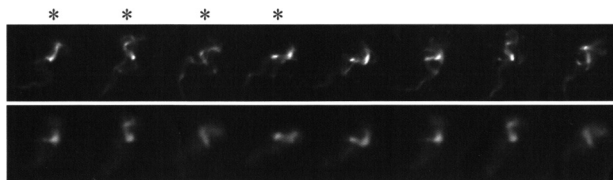


図 3. 一つの精子動画データによって学習したモデルについて, 元の動画の一部(上)と再構成された動画(下). *印は動画の再構成にあたって位相変数と非位相変数の推定に用いられた入力動画の範囲を示す.

なされていることがわかる. ただし再構成画像では鞭毛のような微細な構造は失われ, 明瞭に確認できるのは細胞体のみである.

同じ学習器と動画データを用いて, スライディング・ウィンドウ方式によって動画全長に渡って位相変数 ϕ_0 と非位相変数 p の推定値を計算した結果が図 4 である. 図に示されている値はエンコーダーによって計算された (ϕ, p) の平均値を使い分散を無視している, すなわち式(2.2)および(2.3)における $\mu_\phi(z)$ および $\mu(z)$ を用いている. 言い換えると, 図 4 は隠れ変数の MAP 推定値を示している. また非位相変数 p については煩雑になるのを避けるために (p_0, p_1, p_2) のみ表示している. ここから精子の細胞体の回転運動の周期が安定していることがわかる. 一方で, 位相の進行に依らない画像の変動を吸収するために導入した p については, 期待通り位相変数との明確な相関がない.

そこで我々は, p に符号化されている情報を具体的に調べるために, p の値をエンコーダーによって決められた値から変更し, 再構成画像がいかに変わるかを調べた. 図 5 に各 p_i について $[p_i - 5, p_i - 2, p_i, p_i + 2, p_i + 5]$ と変更した際の再構成画像を示した. p は学習時に標準正規分布に近づくようにカルバック・ライブラー情報量によって制御されるため, 図 4 右からもわかるようにその推定値は概ね $[-2, +2]$ の範囲にとどまる. この範囲の p の変化では再構成画像への影響は視覚的には顕著なものではない. またこの範囲を超えて p を増減させても, 再構成画像にわずかな明暗パターンの変化を引き起こすにとどまることが多く, 容易に解釈が可能な結果ではなかった. 変分自己符号化器の隠れ変数に符号化される画像の特徴が必ずしも解釈可能でない問題は広く認識されており, これを改善するために隠れ変数の分布に様々な制約をかけるなどの手法が提案されているが, 複雑な実データに対する成功例は乏しく, さらなる技術の発展が待たれる.

最後に, 同じモデルと学習方法によって, 八本の動画全てを個別に解析した結果を図 6 に示

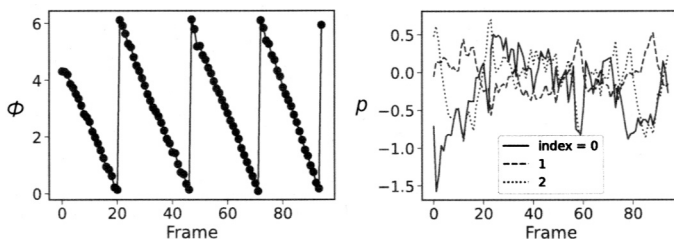


図 4. スライディング・ウィンドウ方式で推定された位相変数(左)と非位相変数(右)の時系列. モデルおよびデータは図 3 で使用したものと同じである.

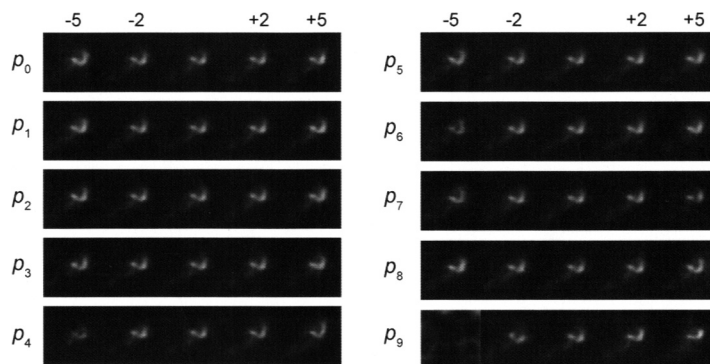


図 5. 各非位相変数 $\{\rho_i\}$ を元の値から変更して再構成した画像. モデルおよびデータは図 3 および図 4 で使用したものと同じである.

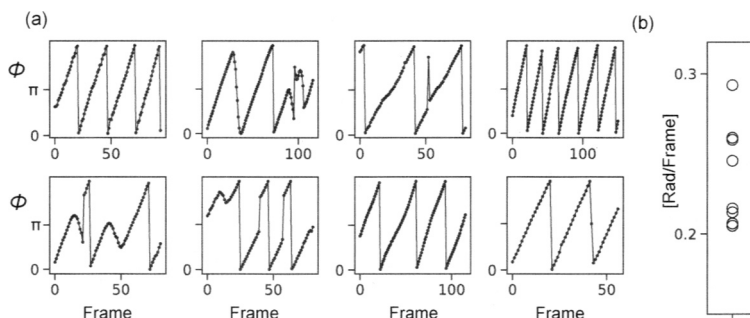


図 6. (a) 八本の精子動画それぞれについて同様の解析を実施して推定された位相時系列. (b) 学習によって得られた位相速度 ω .

す. どの精子も観察時間内では安定した位相速度を持つが, 精子間の差は大きく, 個体差とも呼ぶべきものがあると思われる. また, 一部の精子では位相の跳びが発生している. これはトラッキングされているものではない精子の写り込みなどによって説明できる場合もあるが, 詳細はさらなる研究が必要である. 抽出された回転速度は実時間の単位では 1.3 ± 0.16 Hz であり, これは肉眼による半定量的な評価と一致している. しかし一方で, 精子運動分析専用の観察装置および画像解析ソフトウェアを用いた研究 (Furuichi and Matsuura, 2016) では精子回転を反映していると想定される “lateral head displacement” の頻度について 17–19 Hz というかなり大きな値が報告されている. ここで留意すべきは, ゼニゴケ精子の頭部からは二本の鞭毛が

生えており、さらにそれぞれ異なる運動パターンを示すとの報告もある(宮村 他, 2002)ことから精子頭部の運動パターンは時空間的にかなり複雑になっている可能性があることである。我々の結果と先行研究との差が、注目している運動成分の違いに由来するのか、または精子の状態や周囲の流体の性質の違いに由来するのか、今後明らかにしていくことは精子運動のさらなる理解につながるため重要である。

5. おわりに

本研究では、ゼニゴケ精子の運動様式を特徴づけるパラメータを顕微鏡動画から直接得るという目的に対し、位相振動子を組み込んだ変分自己符号化器を提案し、その有効性を示した。我々の提案モデルは最小限のものであり、その力学系モデル・自己符号化器・目的関数について様々な拡張が考えうるため、以下で簡単に議論したい。

まず力学系モデルの拡張として興味深いのは複数の振動子、さらには振動子間の相互作用を考慮することである。このようなモデルは結合振動子と呼ばれ、理論的な研究によってその挙動が詳細に明らかになっている。単細胞生物の遊泳に関する研究が示すように(Klindt et al., 2017)、結合振動子モデリングによってゼニゴケ精子が持つ二本の鞭毛がいかにして協働し細胞を駆動しているかを明らかにできると期待できる。ただし図3からわかるように、そもそも本研究のモデルでは鞭毛の運動の詳細を捉えられていない。これを解決するためには、より時空間分解能の高いデータを用いることと、より高解像度の画像生成に向けたデコーダーを用いることの両方が必要になるだろう。また本研究で無視した振幅自由度について、ゼニゴケ精子における知見はまだ乏しいものの、遊泳細胞の状態や環境の変化によって鞭毛が振幅変動を示す例は多い。そのため複素ギンツブルグ-ランダウ方程式等のモデルによって振幅自由度を考慮したモデリングも検討に値する。

自己符号化器について、単純により高性能の深層学習器を採用するというを超えて興味深い拡張として、物体中心表現学習(例えばLocatello et al., 2020)のような手法によって、複数の物体が写る画像から単一の精子のみを抽出するというものが考えられる。それによって、精子以外の夾雑物の写り込みに対する頑強性の獲得や、精子間の相互作用の学習が可能になるかもしれない。

変分自己符号化器の目的関数に対しては先行研究で様々な正則化項の追加が提案されており、隠れ変数間の統計的独立性による解釈可能性の改善(Chen et al., 2018)や、入力画像への敵対的攻撃に対する頑強性の向上(Cemgil et al., 2020)が報告されている。ただし、正則化項の追加によって変分下限の近似が緩まり、それが隠れ変数の事後分布の推定におよぼす影響は必ずしも明らかでない。一方で我々が有用と考えているのは、よく使われる変分下限とは異なる周辺対数尤度下限である因子化情報量規準(Hayashi et al., 2015)であり、これはモデルの学習と変数選択を同時に行うことを可能にする(Kondo et al., 2016)。例えば、上で述べたように複数の振動子を考慮する場合には振動子の個数の選択が問題になり、実際松田らは脳活動時系列の振動子による状態空間モデリングにあたって赤池情報量規準を採用しモデル選択している(Matsuda et al., 2022)。我々のモデルは深層学習器を含むためモデル選択を目的として学習を繰り返すことは計算コストが大きいという問題があるが、因子化情報量規準はその解決策になりうる。また、本研究を含め多くのケースで式(2.7)の右辺第一項が表す画像再構成誤差を単純に二乗誤差によって評価しているが、代わりに観察機器の光学的性質を考慮した観測ノイズの分布を用いるなどの工夫の余地はまだまだあるように思われる。

本稿ではゼニゴケ野生株に関する結果のみを示したが、変異体の解析についても現在進行中である。本研究で得られた知見が、未だ謎に包まれている植物精子の運動様式の解明につながる

ることを期待したい。

謝 辞

本研究は MEXT 科研費新学術領域研究 No. 19H05675 の助成を受けたものです。計算機資源に関して基礎生物学研究所・超階層生物学センターの支援に厚く感謝いたします。原稿について富沢瑤子研究員に貴重な助言をいただきました。

参 考 文 献

- Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A. and Kreshuk, A. (2019). ilastik: interactive machine learning for (bio)image analysis, *Nature Methods*, **16**(12), 1226–1232.
- Cemgil, A. T., Ghaisas, S., Dvijotham, K., Gowal, S. and Kohli, P. (2020). The autoencoding variational autoencoder, *Proceedings of the 34th Conference on Neural Information Processing Systems*, 15077–15087.
- Chen, R. T. Q., Li, X., Grosse, R. and Duvenaud, D. (2018). Isolating sources of disentanglement in VAEs, *Proceedings of the 32nd Conference on Neural Information Processing Systems*, 2615–2625.
- Furuichi, T. and Matsuura, K. (2016). Kinetic analysis on the motility of liverwort sperms using a microscopic computer-assisted sperm analyzing system, *Environmental Control in Biology*, **54**(1), 45–49.
- Hayashi, K., Maeda, S. and Fujimaki, R. (2015). Rebuilding factorized information criterion: Asymptotically accurate marginal likelihood, *Proceedings of the 32nd International Conference on Machine Learning*, 1358–1366.
- Jaques, M., Burke, M. and Hospedales, T. M. (2020). Physics-as-inverse-graphics: Unsupervised physical parameter estimation from video, *Proceedings of the 8th International Conference on Learning Representations*, 1–16.
- Kingma, D. and Welling, M. (2014). Auto-encoding variational bases, *2nd International Conference on Learning Representations, Conference Track Proceedings*, <https://arxiv.org/abs/1312.6114>.
- Klindt, G. S., Ruloff, C., Wagner, C. and Friedrich, B. M. (2017). In-phase and anti-phase flagellar synchronization by waveform compliance and basal coupling, *New Journal of Physics*, **19**(11), 113052.
- Kondo, Y., Maeda, S. and Hayashi, K. (2016). Bayesian masking: Sparse Bayesian estimation with weaker shrinkage bias, *Asian Conference on Machine Learning*, 49–64.
- Le Guen, V. and Thome, N. (2020). Disentangling physical dynamics from unknown factors for unsupervised video prediction, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11471–11481.
- Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., Dosovitskiy, A. and Kipf, T. (2020). Object-centric learning with slot attention, *Proceedings of the 34th Conference on Neural Information Processing Systems*, 11525–11538.
- Matsuda, T., Homae, F., Watanabe, H., Taga, G. and Komaki, F. (2022). Oscillator decomposition of infant fNIRS data, *PLoS Computational Biology*, **18**(3), e1009985.
- 宮村新一, 松永 茂, 堀 輝三 (2002). ゼニゴケ精子鞭毛運動の高速度ビデオによる解析, *蘚苔類研究*, **8**(3), 79–83.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A.,

- Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library, *Proceedings of the 33rd Conference on Neural Information Processing Systems*, 8024–8035.
- Rezende, D. J., Mohamed, S. and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models, *Proceedings of the 31st International Conference on Machine Learning*, 1278–1286.
- Saggiorato, G., Alvarez, L., Jikeli, J. F., Kaupp, U. B., Gompper, G. and Elgeti, J. (2017). Human sperm steer with second harmonics of the flagellar beat, *Nature Communications*, **8**(1), 1415.
- Takeishi, N. and Kalousis, A. (2021). Physics-integrated variational autoencoders for robust and interpretable generative modeling, *Proceedings of the 35th Conference on Neural Information Processing Systems*, 14809–14821.
- Zhang, M., Hayes, P. and Barber, D. (2022). Generalization gap in amortized inference, *Proceedings of the 36th Conference on Neural Information Processing Systems*, <https://arxiv.org/abs/2205.11640>.

Dynamical-system-integrated Variational Autoencoder
for Modeling of Time-lapse Microscopy Data:
Application to Liverwort Spermatozoid Motion

Yohei Kondo^{1,2}, Naoki Minamino³ and Takashi Ueda^{2,3}

¹ExCELLS, National Institutes of Natural Sciences

²Graduate Institute for Advanced Studies, SOKENDAI

³National Institute for Basic Biology

Microscopy has been one of the most important research methods in biology from the 17th century, when Hooke discovered the cell, to the present day. In recent years, the development of optical technology and fluorescent protein-based tools has revealed various biological phenomena with unprecedented spatio-temporal resolution. However, such microscopic movie data are complex and high-dimensional, and thus it is not always easy to obtain biologically-relevant parameters from them. To address the issue, we propose a framework for modeling of microscopic movie by integrating a dynamical system into a variational autoencoder, a deep-learning model for unsupervised feature extraction. This allows us to extract biological information as state variables and parameters of the integrated dynamical system. In this article, we have validated our framework by analyzing spermatozoid motion of the liverwort *Marchantia polymorpha*, an emerging model organism for plant biology. Using a phase oscillator, we have succeeded to estimate the phase and speed of spermatozoid rotation from our high-speed dark-field video microscopy.