

量子ドット蛍光発光強度時系列データのための隠れマルコフモデル解析

古田達央^A、濱田啓輔^A、小田勝^A、中村和磨^A

九州工業大学大学院工学研究院^A

Hidden Markov model analysis for fluorescent time series of quantum dots

Tatuhiko Furuta^A, Keisuke Hamada^A, Masaru Oda^A, Kazuma Nakamura^A

Graduate School of Engineering, Kyushu Institute of Technology, Kitakyushu^A

We present a hidden Markov model (HMM) analysis for fluorescent time series of quantum dots. In the HMM, the hidden variable specifying the states behind the real fluorescent intensity is introduced, and the time series of the hidden variable are considered. Light-emission duration of the quantum dots are estimated from the obtained hidden variable time-series, and a duration distribution function for the light-emission state, called a blinking plot, is evaluated to analyze the optical process of the quantum dots. In the present study, we describe how to calculate a reliable blinking plot, and focus on the comparison between the results derived with the HMM and conventional approaches. While the conventional approach suffers from noises in the experimental fluorescent data, the hidden-variable time series obtained from the HMM simulation are noise-suppressed, and thus assignments for the light-emission state become stable. The resulting blinking plots is also reliable. To check the quantitative accuracy, we apply the HMM analysis to theoretical benchmark time series which are generated from input distribution functions for the light-emission duration to be analyzed. We found that the HMM well reproduces the original distribution function and its quantitative accuracy is beyond human cognitive ability.

1. はじめに

近年、機械学習を用いて物質科学研究において得られる様々なデータを解析しようという試みが活発化している [1-4]。これまで機械学習は静的物理量の相関解析を対象としてきたが、最近では動的データの解析にも適用が始まっている [5-8]。

本論文では量子ドット蛍光発光強度時系列データに対する機械学習解析について報告する [9]。量子ドットの光学特性は blinking plot (BP) を用いて理解される。量子ドットが光照射に応答して吸収と放射を繰り返している状態を ON 状態と呼び、一方、吸収はするが、発光はしない状態を OFF 状態と呼ぶ。量子ドットの蛍光発光強度時系列データは、この ON 状態と OFF 状態を繰り返してをモニターしたものであるが、この時系列データから ON あるいは OFF 状態の継続時間を評価し、これらの継続時間に関する確率分布を求めたものが BP である。BP を解析することで、量子ドットの光学特性を定量評価できる。

この時系列データは、量子ドットが ON 状態であろうと OFF 状態であろうと、量子ドットを載せた基板からの発光等のノイズを含んでいる。このノイズが量子ドット固有のシグナルに比して大きくなると時系列データの解析自体が人間の認知能力では困難になる。つまり、ON 状態と OFF 状態の判別が不可能になり、各状態の継続時間評価の定量性が損なわれる。結果、得られる BP の信頼性も損なわれる。こうしたノイズを人為的なやり方で除去するのではなく、機械学習を用いて効果的に実施しようという試みは重要である。

本研究では、代表的な時系列データ解析手法である隠れマルコフモデリング (HMM) を用いた時系列データのノイズ除去と定量精度の高い BP の作成手法を提案する。従来までの ON/OFF 判定法と提案手法の比較から提案手法の妥当性の検証を行った。

従来法では、短継続時間イベントの人為的大量発生と長継続時間イベントの顕著な抑制が起こるが、提案手法ではこれらのアーティファクトを改善することができ、信頼性のある BP が得られることが分かった。また提案手法の精度検証のために、答えの分かった BP データを提案手法が正しく再現できるか調べたところ、人間の認知能力では判断不可能なデータでさえ、正確に解析できることが分かった。

2. 手法

HMM は様々な確率変数を用いた確率分布の組み合わせで表現される。本研究において HMM の確率分布は、以下のように表される [10] :

$$p(\mathbf{I}, \mathbf{S}, \Theta, \pi, \mathbf{A}) \\ = p(\mathbf{I}|\mathbf{S}, \Theta)p(\mathbf{S}|\pi, \mathbf{A})p(\Theta)p(\pi)p(\mathbf{A}). \quad (1)$$

\mathbf{I} は観測データであり、本研究においては量子ドット蛍光発光強度時系列データに相当する。本シミュレーションにおいて、時系列データの発光強度の値は標準化された値を用いる。 \mathbf{S} は潜在変数であり、各時刻の観測データの状態を表す。本研究において発光強度の ON, OFF の状態判定を行う変数であり、時系列データの解析結果に相当する。 Θ は観測データ \mathbf{I} の分布形状パラメータであり、その設定は観測データに従う分布に依存する。本研究において発光強度がガウス分布に従うと仮定し、平均 μ 、精度 λ として導入される。 π は初期時刻の潜在変数の確率、 \mathbf{A} は潜在変数の状態遷移確率を表す。

シミュレーションにおいて観測データ \mathbf{I} が得られた下での条件付き分布

$$p(\mathbf{S}, \mu, \lambda, \pi, \mathbf{A}|\mathbf{I}) = \frac{p(\mathbf{I}, \mathbf{S}, \mu, \lambda, \pi, \mathbf{A})}{p(\mathbf{I})} \quad (2)$$

を計算する必要がある。ここで、 $p(\mathbf{I})$ は周辺分布で

ある：

$$p(\mathbf{I}) = \sum_{\mathbf{S}} \iiint p(\mathbf{I}, \mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\pi}, \mathbf{A}) d\boldsymbol{\mu} d\boldsymbol{\lambda} d\boldsymbol{\pi} d\mathbf{A}. \quad (3)$$

本研究において条件付き分布 [式 (2)] を最適化するために、ベイズ推論に基づくブロックギブスサンプリング [11] を用いた。

BP の再現性を調査するため、条件付き分布 [式 (2)] を評価することにより得られた潜在変数の時系列データ \mathbf{S} から算出された ON/OFF の継続時間 $\{\tau_{\alpha}^{\text{ON}}\}$ 及び $\{\tau_{\alpha}^{\text{OFF}}\}$ の各々について確率分布の計算を行った。 $\{\tau_{\alpha}^{\text{ON}}\}$ 及び $\{\tau_{\alpha}^{\text{OFF}}\}$ における各々の確率分布 [12] は

$$P_{\text{ON}}(\tau) = \frac{1}{N_{\text{ON}}} \sum_{\alpha'=1}^{N'_{\text{ON}}} f_{\alpha'}^{\text{ON}} w_{\alpha'}^{\text{ON}} \delta(\tau - \tau_{\alpha'}^{\text{ON}}), \quad (4)$$

$$P_{\text{OFF}}(\tau) = \frac{1}{N_{\text{OFF}}} \sum_{\alpha'=1}^{N'_{\text{OFF}}} f_{\alpha'}^{\text{OFF}} w_{\alpha'}^{\text{OFF}} \delta(\tau - \tau_{\alpha'}^{\text{OFF}}) \quad (5)$$

と表される。ただし N'_{ON} と N'_{OFF} は ON 及び OFF の独立した合計イベント数である。また、 $w_{\alpha'}^{\text{ON}}$ 及び $w_{\alpha'}^{\text{OFF}}$ は重み因子であり、それぞれ

$$w_{\alpha'}^{\text{ON}} = \frac{2\Delta\tau}{\tau_{\alpha'+1}^{\text{ON}} - \tau_{\alpha'-1}^{\text{ON}}}, \quad (6)$$

$$w_{\alpha'}^{\text{OFF}} = \frac{2\Delta\tau}{\tau_{\alpha'+1}^{\text{OFF}} - \tau_{\alpha'-1}^{\text{OFF}}} \quad (7)$$

と表される。ここで $\Delta\tau$ はグリッドの間隔である。最後に式 (4), (5) 内の N_{ON} 及び N_{OFF} は ON 及び OFF の合計イベント数で以下のような和則を満たす：

$$N_{\text{ON}} = \sum_{\alpha'=1}^{N'_{\text{ON}}} f_{\alpha'}^{\text{ON}} \frac{w_{\alpha'}^{\text{ON}}}{S_{\text{ON}}}, \quad (8)$$

$$N_{\text{OFF}} = \sum_{\alpha'=1}^{N'_{\text{OFF}}} f_{\alpha'}^{\text{OFF}} \frac{w_{\alpha'}^{\text{OFF}}}{S_{\text{OFF}}}. \quad (9)$$

ここで S_{ON} 及び S_{OFF} は確率分布 $P_{\text{ON}}(\tau)$ 及び $P_{\text{OFF}}(\tau)$ を規格化するための補正因子である。

3. 結果

対物レンズ (100 \times ; 0.9 NA) 及び avalanche photodetector を備えた走査型レーザー顕微鏡 [13] を用いて単一量子ドットの蛍光発光強度時系列データを測定した。量子ドットを励起するレーザー光 (連続光) の波長及び励起強度はそれぞれ 488 nm, 133 W/cm² とした。実験では、有機金属化合物の熱分解反応を利用して合成した、有機分子 trioctylphosphine oxide と hexadecyl amine により結晶表面を覆った CdSe/ZnS コアシェル量子ドット [14] を使用した。時系列データは、各ステップでの測定時間

(単位時間) を 0.1s として 1200s の長さで測定した。また、この時系列データは真空及び湿潤窒素環境の 2 つの異なる環境下で各々 12 サンプルで測定した。湿潤窒素環境は蒸留水を通して窒素ガスをバブリングすることで用意した。

図 1 及び図 2 にそれぞれ真空及び湿潤窒素環境下における蛍光発光強度 (赤実線) 及び HMM シミュレーションにより得られた潜在変数 \mathbf{S} の時系列データ (緑破線) を示す。時系列データ \mathbf{I} に I_{th} (青点線) を設けて ON 状態 (もしくは OFF 状態) の継続時間を推定する既存の方法を用いるとノイズに起因して短い継続時間が多く算出されてしまう。一方で潜在変数の時系列データにおいてノイズが抑えられているため、閾値 s_{th} (黒実線) を引いて長い継続時間を算出できることが分かる。

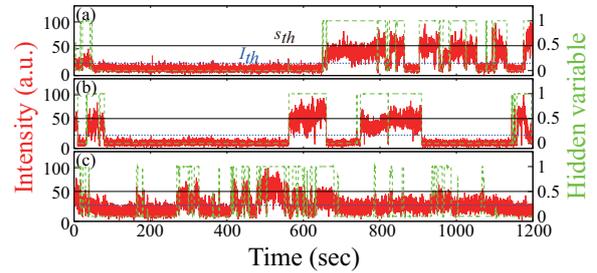


図 1 真空環境下での蛍光発光強度時系列データ (赤実線) 及び HMM シミュレーションで得られた潜在変数 \mathbf{S} の時系列データ (緑破線)。左縦軸が発光強度、右縦軸が潜在変数であり、潜在変数において取り扱いの詳細については [9] に譲るが、1 が ON 状態、0 が OFF 状態を表す。 I_{th} (青点線) は慣習法において時系列データ \mathbf{I} における ON 状態と OFF 状態を区別するための目視で設定された閾値である。 s_{th} (黒実線) は潜在変数時系列データの閾値であり、数値を 0.5 としている。

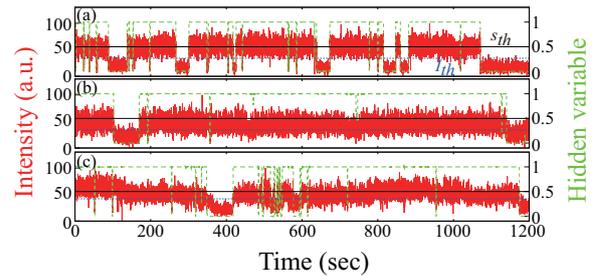


図 2 湿潤窒素環境下での蛍光発光強度時系列データ (赤実線) 及び HMM シミュレーションで得られた潜在変数 \mathbf{S} の時系列データ (緑破線)。

12 サンプルの時系列データの統計平均により、BP を取った結果が図 3 である。各々のデータの傾向を分析するため、以下の関数に従ってフィッティングを行った [15]：

$$f(\tau) = A\tau^{-m}. \quad (10)$$

A は係数、 m は逆指数を表す。各々のプロットの m の値を比較することで、確率分布は慣習法に基づく τ^{-2} に近似され、HMM において τ^{-1} に減衰することが分かった。これは、慣習法ではノイズに起因

する短寿命の ON 状態や OFF 状態が多数生成されてしまうためである。このような短継続時間のイベントは HMM では効果的に排除できるので、結果、正しいべき乗則が得られる。また、HMM に基づいた結果において湿潤窒素環境下における m 値は真空環境下と比較して基本的に小さい値をとることから、単一量子ドットは湿潤窒素環境下において長時間放射と長時間消光をとる傾向があること、一方で、真空環境下では短時間の明滅を繰り返す傾向があることが分かった。

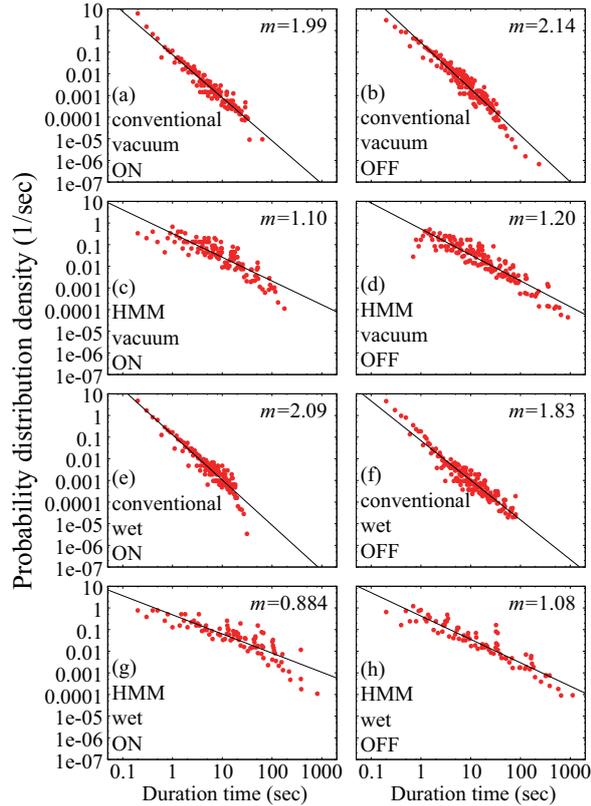


図3 実験蛍光発光強度時系列データにおける ON/OFF 継続時間の確率分布 $P_{ON}(\tau)$ [式 (4)] 及び $P_{OFF}(\tau)$ [式 (5)] (赤点)。黒実線は $f(\tau)$ [式 (10)] であり、逆指数 m は直線の傾きを表す。直線の傾きが小さいほど長継続時間に出現の頻度が高いことを表す。図の左側 [(a), (c), (e), (g)] が ON 継続時間、図の右側 [(b), (d), (f), (h)] が OFF 継続時間の結果である。また、図の上側 4 つ [(a), (b), (c), (d)] は真空環境下、図の下側 4 つ [(e), (f), (g), (h)] は湿潤窒素環境下の結果である。(a), (b), (e), (f) は慣習法、(c), (d), (g), (h) は HMM に基づいた結果を表す。

HMM による時系列データ解析の定量精度を評価するため、量子ドットのプリンキング現象として提案されている以下のモデル関数について考える [16] :

$$p_{ON}(\tau) = A^{ON} \tau^{-q} e^{-\tau/\xi} \theta(\tau - \tau_{min}) \theta(\tau_{max} - \tau), \quad (11)$$

$$p_{OFF}(\tau) = A^{OFF} \tau^{-l} \theta(\tau - \tau_{min}) \theta(\tau_{max} - \tau). \quad (12)$$

ここで A^{ON} 及び A^{OFF} は規格化定数、 $\theta(x)$ はス

テップ関数、 q, ξ, l はモデル関数のパラメータを表す。 τ_{min} 及び τ_{max} は継続時間の下限及び上限カットオフを表す。モデル関数からベンチマーク時系列データの生成を行い、その時系列データを HMM で解析を行い、その解析で得られる時系列データの統計平均がモデル関数を再現できているのかを調べるにより HMM の定量精度の評価を行う。

図4がモデル関数 [式 (11), (12)] により生成された時系列データである。ON 及び OFF 継続時間を各々 20 イベント抽出し、それに基づいて白破線のような蛍光発光強度時系列データを生成する。次に、平均 0、標準偏差 0.5 に従うガウシアンノイズを白破線の時系列データに付与したものが赤実線の時系列データであり、この赤実線の時系列データが HMM シミュレーションの解析対象となる。 Δ は ON と OFF のベースラインの差を表し、このパラメータを 0.7 (TS-I), 0.5 (TS-II), 0.3 (TS-III) と変えた時系列データを用意した。赤実線の時系列データから Δ を小さくする ($\Delta = 0.7 \rightarrow 0.5 \rightarrow 0.3$) につれて ON/OFF 状態を区別するのは困難であることが分かる。 $\Delta = 0.3$ (TS-III) において ON/OFF 状態を目視で区別するのは不可能である。統計平均のため、時系列データは各々 1000 サンプルずつとっている。

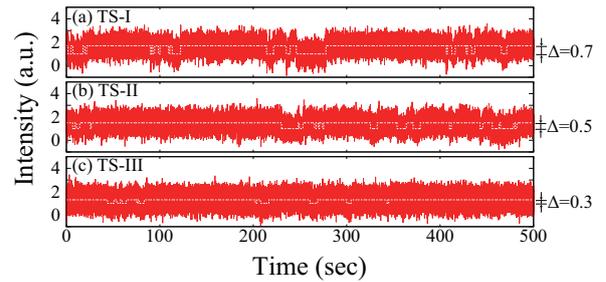


図4 モデル関数 [式 (11), (12)] により生成されたノイズあり (赤実線) 及びノイズなし (白破線) 時系列データ。 Δ はノイズなし時系列データの ON 状態と OFF 状態の発光強度のレベル差であり、(a) $\Delta = 0.7$ (TS-I), (b) $\Delta = 0.5$ (TS-II), (c) $\Delta = 0.3$ (TS-III) のケースに相当する。

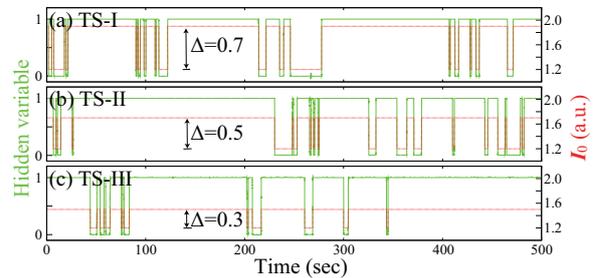


図5 HMM シミュレーションによって得られた潜在変数の時系列データ (緑実線)。(a) が TS-I ($\Delta = 0.7$) [図 4(a)] に対して得られた潜在変数の時系列データを表す。(b), (c) においても同様に各々の TS-II ($\Delta = 0.5$) [図 4(b)] 及び TS-III ($\Delta = 0.3$) [図 4(c)] である。赤点線は図4の白破線を表す。HMM シミュレーションが成功していれば、緑実線は赤点線と一致する。

図4の時系列データに対してHMMシミュレーションを行うことで得られた潜在変数の結果を図5に示す。潜在変数の時系列(緑実線)は赤点線(図4の白破線)をほぼ再現できていることが分かる。 $\Delta = 0.3$ (TS-III)の時系列データにおいてもほぼミスがないことから、HMMの判定正確性が高いことが分かる。

潜在変数の時系列データ1000サンプルの統計平均をとることにより得られたBPの結果を図6に示す。計算により得られたBPはオリジナルの分布関数(黒実線)をよく再現できていることが分かる。図4(c)の時系列データTS-III($\Delta = 0.3$, ノイズ0.5)は人間にはON/OFF状態の識別不可能な時系列データであるが、HMMシミュレーションは正しく状態判定を実施できることが分かる。

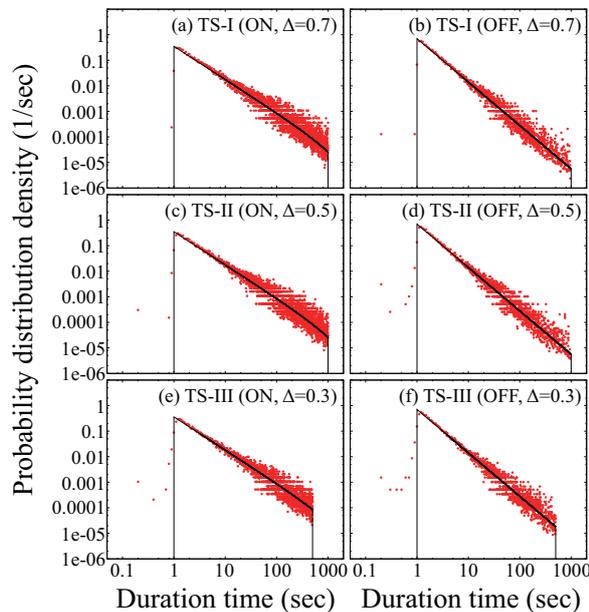


図6 ベンチマーク時系列データにおけるON/OFF継続時間の確率分布 $P_{ON}(\tau)$ [式(4)]及び $P_{OFF}(\tau)$ [式(5)] (赤点)。黒実線がモデル関数の $p_{ON}(\tau)$ [式(11)]及び $p_{OFF}(\tau)$ [式(12)]を表す。図の上部[(a), (b)], 中部[(c), (d)], 下部[(e), (f)]は各々 $\Delta = 0.7, 0.5, 0.3$ の結果に対応する。各々の結果は1000サンプルの統計平均に基づく。また、左側[(a), (c), (e)]はON, 右側[(b), (d), (f)]はOFF継続時間の結果である。

4. まとめ

量子ドットの蛍光発光強度時系列データに対してHMMシミュレーションを行った。量子ドットでは、蛍光発光強度時系列データからON/OFF継続時間を算出し、その統計平均から得られるBPが光学特性の指標となる。慣習法では、蛍光発光強度時系列データに目視で基準値を決めることによりON/OFF継続時間の算出を行っていた。しかしながら、この

手法ではノイズに起因して短い継続時間が不当に多く算出される。本研究ではこのようなアーティファクトを取り除くために時系列データ解析に用いられる機械学習の手法であるHMMを導入した。HMMでは潜在変数と呼ばれる各時刻における観測データの状態を記述する変数が定義され、蛍光発光強度時系列データではなく、潜在変数の時系列データに対して解析を行う。真空・湿潤窒素環境下の実験における単一量子ドットの蛍光発光強度時系列データの解析に適用したところ、潜在変数の時系列データでは慣習法よりもノイズを抑えてON/OFF継続時間を算出できることが分かった。また、定量的理解のために、ON/OFF継続時間のモデル関数に基づいて生成したベンチマーク時系列データに対してHMMシミュレーションを行い、その統計平均により得られたBPがモデル関数を再現できるか検証した。検証によりHMMシミュレーションでは、人間が識別不可能な時系列データに対しても高い解析性能を示すことが分かった。

参考文献

- [1] A. Seko, T. Maekawa, K. Tsuda, and I. Tanaka, Phys. Rev. B 89, 054303 (2014).
- [2] A. Seko, *et al.*, Phys. Rev. Lett. 115, 205901 (2015).
- [3] J. Lee, *et al.*, Phys. Rev. B 93, 115104 (2016).
- [4] Q. Tao, P. Xu, M. Li, and W. Lu, npj Comput. Mater. 7, 23 (2021).
- [5] F. M. Puglisi, P. Pavan, A. Padovani, and L. Larcher, Solid-State Electron. 102, 69 (2014).
- [6] T. Murai, *et al.*, Appl. Phys. Lett. 112, 111903 (2018).
- [7] M. Pirchi, *et al.*, J. Phys. Chem. B 120, 13065 (2016).
- [8] D. H. Nguyen-Le, *et al.*, Eng. Fract. Mech. 235, 107085 (2020).
- [9] T. Furuta, K. Hamada, M. Oda, and K. Nakamura, Phys. Rev. B 106, 104305 (2022).
- [10] L. Rabiner, Proc. IEEE 77, 257 (1989).
- [11] C. S. Jensen, U. Kjærulff, and A. Kong, Int. J. Hum. Comput. Stud. 42, 647 (1995).
- [12] M. Kuno, *et al.*, J. Chem. Phys. 115, 1028 (2001).
- [13] M. Oda, *et al.*, J. Lumin. 127, 198 (2007).
- [14] K. Hashizume, M. Matsubayashi, M. Vacha, and T. Tani, J. Lumin. 98, 49 (2002).
- [15] K. T. Shimizu, *et al.*, Phys. Rev. B 63, 205316 (2001).
- [16] B. Bruhn, F. Qejvanaj, I. Sychugov, and J. Linnros, J. Phys. Chem. C 118, 2202 (2014).