博士論文（要約）

Efficient Analysis of Event Data and Application for Optimal Termination:

Practice in Inelastic Neutron-Scattering Experiments

（イベントデータの効率的な分析と最適停止への応用：

非弾性中性子散乱実験での実践）

武藤　健介

Efficient Analysis of Event Data and Application for Optimal Termination

Practice in Inelastic Neutron-Scattering Experiments

Kensuke Muto

Supervisor
Masato Okada

University of Tokyo, 2020
Department of Complexity Science and Engineering, Graduate School of Frontier Sciences

Abstract

In recent years, a large amount of four-dimensional event data have been obtainable in neutron inelastic scattering experiments conducted by chopper spectrometers at Japan Proton Accelerator Research Complex (J-PARC). As preprocessing, researchers make histograms from obtained event data. At present, the researchers only empirically select bin widths and slice conditions to obtain a two-dimensional histogram, while checking the histogram in a visual approach. The arbitrariness of the process and human cost are significant problems. It is also an essential task to establish an automatic termination strategy of inelastic neutron-scattering measurement to prevent redundancy of the measurement. There is no criterion to assess whether the obtained data is sufficient in event number. By using large-scale data, we can extract the fine features of the measurement target. However, it is not necessary to extract information beyond the resolution of the measurement equipment. In this thesis, we propose methods to resolve these issues.

First, we propose a method that can automatically make a multidimensional histogram from event data. The optimization criterion is based on a cost function representing the tradeoff between the reduction of stochastic fluctuation and extraction of the structure that the data have. In this thesis, we use artificial data to investigate the behavior of our method. The artificial four-dimensional event data were produced, assuming neutron inelastic scattering due to phonons. We applied the proposed method to both sliced two-dimensional event data and the whole four-dimensional event data. Comparing their results, we have found that the optimized bin widths strongly depends on the dimensionality of the data. Moreover, the optimal bin widths are affected by the number of events and the magnitude of the white background noise.

Second, we propose a method to compute the termination criteria and determine whether to continue or terminate the experiment in real time. In the proposed method, researchers compute the optimal bin widths of a histogram for the obtained data. Regarding the termination criterion, the experiment is terminated when the optimal bin widths become smaller than the target resolutions. Since the optimal bin-width calculation can be performed in parallel to the experiment, it is effective as a real-time stopping arrangement. In numerical experiments, we dealt with real inelastic neutron-scattering data of a typical size. As a result of the numerical experiments, the optimal bin widths decrease as the number of events increases. Even the optimal bin widths for data downsampled to 1/5 are comparable with the resolutions limited by the sample size, choppers, and so on. This implies excessive measurement of the inelastic neutron experiments for the moment. Moreover, we show that Bayesian optimization (BO) is useful in searching for the optimal bin widths.

Third, we propose a method for efficient terminating strategy while measuring event data in real time. In general, the computational cost of bin widths optimization grows exponentially with the number of the dimension of the data. Therefore, it is urgent task to reduce the computational cost of the bin widths optimization. We proposed a

method using the prior distribution of BO computed from the information about the cost function obtained in the past. We perform numerical experiments using inelastic neutron-scattering experiment data. As a result of numerical experiments, the proposed method greatly improves the search efficiency of the optimal bin widths. Moreover, it is robust for a hyper parameter.

Acknowledge

This thesis is the summary of my works in the Ph. D course of the Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, the University of Tokyo. Many people have helped me to write this thesis. Firstly, I would like to thank my supervisor Professor Masato Okada for his support. I was greatly influenced by his unique insight. Secondly, Professor Taka-hisa Arima has provided me with a lot of knowledge about condensed matter physics, especially inelastic neutron scattering experiments. In addition, he provided us with the experimental data through his connection with J-PARC. Professor Masato Okada, Professor Taka-hisa Arima, Professor Takeshi Imamura, Associate Professor Ryota Kobayashi, and Lecturer Naoto Yokoya. Their valuable and comprehensive comments helped to improve the draft of this thesis. Thirdly, I am deeply grateful to all of my lab members and alumni. They have made my research life pleasant and fruitful. Without the existence of my peers Mr. Shun Katakami Mr. Tomofumi Nakayama, and Dr. Yuki Yoshida I would not have entered the Ph. D course. Finally, I owe a very important debt of gratitude to my family. They have always trusted my choice and encouraged me.

This thesis was partially supported by JSPS KAKENHI Grant-in-Aid for Scientific Research(A) (No. 18H04106) and JST CREST (JPMJCR1761). The inelastic neutron-scattering data of $Ba_3Fe_2O_5Cl_2$ were obtained by a chopper spectrometer 4SEASONS installed on BL01, J-PARC MLF, Japan, under the proposal No. 2016A0088.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 Data-intensive science

Jim Gray proposed a fourth paradigm, data-intensive science. [3] This is a scientific framework based on "big data" [4, 5, 6] and advanced computational analysis technology. Currently, infrastructures are being actively developed to efficiently apply machine learning to big data.[7, 8, 9] In this section, we introduce the current state of data-intensive science and its challenges. The definition of four paradigms are as follows.

- 1st paradigm: experimental science

This paradigm empirically describes natural phenomena without advanced mathematics and computer.

- 2nd paradigm: theoretical science

Researchers analyze the observed data and reveal the laws behind it. The laws are mainly written in the form of differential equations. Newton's laws of motion, Maxwell's equations and Schrodinger's equation are well known. Then, for many problems, the theoretical models grew too complicated to solve analytically, and researchers had to start simulation. This leads to the third paradigm.

- 3rd paradigm: computational science

Researchers use computers to numerically solve nonlinear equations that cannot be solved analytically The simulation facilitated the analysis of many-body problems. Moreover, these simulations are generating a whole lot of data, along with a huge increase in data from the experimental sciences.

- 4th paradigm: data-intensive science

The fourth paradigm unifies the first to the third paradigms. The world of science has changed. Data-intensive science consists of three basic activities: capture, curation, and analysis. The new model is optimized by using the data to be captured by instruments or generated by simulations. The resulting information and knowledge are to be stored in computers.

Preprocessing and data integration are significant for the data to make it easier to analyze and to reduce the volume. Scientists only get to look at their data fairly late in this pipeline. The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science. Experimental and theoretical sciences have also been data-centric. However, data-intensive science is based on "big data" and more sophisticated analysis. Various elemental technologies; data collection, organization, storage, analysis and visualization, are needed to realize data-intensive science. Therefore, there is an urgent task to develop and integrate these methods. There are several efforts to integrate the science of each discipline with data-intensive science. Materials informatics (MI) is known as a field of study that applies the principles of informatics to materials science and engineering to improve the understanding, use, selection, development, and discovery of materials. In 2011, the Materials Genome Initiative was launched in the United States. [10, 11] It was the start of the research trend on MI all over the world. Big data generated by simulation are analyzed for material development. The organization of these data is an important issue, and several universities and national institutes are building material databases. In recent years, the construction of a real-time data analysis platform has been focused on in inelastic neutron-scattering experiments. We aim to build data-intensive methods for inelastic neutron-scattering experiments in this thesis.

1.1.2 Motivation

In recent years, improvements in simulation and measurement technology have led to the generation of "big data" in various fields. Researchers organize these data to reduce noise and data volume. Currently, most of the preprocessing are conducted empirically by experts. There are the problems of arbitrariness of the process and human cost. Therefore, it is necessary to develop algorithms to automate these processes.

There are many experiments to acquire event data such as neutron scattering experiments, radiation measurements, and spike measurements in neuroscience. We focus on inelastic neutron scattering experiments in this thesis. Researcher investigate the the dynamical structure of materials by using inelastic neutron-scattering experiments. [12]. In recent years, numerous event data have been obtained in inelastic neutron-scattering experiments using high-power accelerator-based neutron sources such as those of ISIS, SNS, J-PARC, and CSNS. A time-of-flight neutron spectrometer designed for inelastic scattering measurements such as MAPS [13], MERLIN [14], HYSPEC [15], ARCS [16], 4SEASONS [17], AMATERAS [18], and HRC [19] produces a large-scale event data. Researchers obtain event data mapped on the four-dimensional (4D) space of transferred energy (E) and momentum ($\mathbf{q}$). When the measurements are independent, the process of observation of the events can be modeled as the Poisson distribution. The data are observed with stochastic fluctuations. At present, researchers create histograms to reduce the fluctuations and analyze them [20]. If the bin widths are set to be wider, stochastic fluctuation of the event data can be reduced by smoothing; however, the structure of the data should be lost.12, 13) Thus, selecting the best bin widths can be considered as an optimization of this trade-off. Currently, researchers empirically select bin widths in a visual approach. Here, the problem is that the widths are chosen for 2D datasliced from 4D data. When cutting a 2D slice, researchers arbitrarily select the slice widthsand position. These arbitrary processes should affect the result of selecting bin widths. The arbitrariness of preprocessing and human costs of trial and error are current problems. Therefore, we propose a method for optimizing the bin widths of histograms for multidimensional event data. Bin widths optimization is formulated as a minimization problem of a cost function in the proposed method.

A practical application of the proposed method is to determine whether to terminate or to continue the experiment. By using a larger amount of data, researchers can extract more detailed features of the target materials. However, it is not necessary to extract information beyond the resolution of the measurement system. At present, there is no criterion to assess whether the obtained data is sufficient in event number, and the measurement usually becomes redundant. We propose a method to compute the termination criteria and determine whether to terminate or continue the experiment in real time. In the proposed method, researchers compute the optimal bin widths of a histogram for the obtained data and terminate the experiment when the optimal bin widths become smaller than the expected resolutions. Here, it is known that the optimal bin widths decrease as the number of events increases.[21, 22] Since the proposed method can be executed in parallel with the experiment, it is helpful for real-time termination decision. Moreover, we conducted numerical experiments using a set of real experimental data and showed that Bayesian optimization (BO)[23, 24] is efficient for searching in the optimal bin widths. Bayesian optimization is known as a method for solving optimization problems for an objective function. Gaussian process (GP)[25] is often used in Bayesian optimization. GP is known as a method for calculating a predictive distribution with input data. For further efficiency, we aim to utilize the information of the cost function computed in the past. We propose a method using the prior distribution of BO computed from the information of the cost function obtained in the past. Regarding the proposed method, we focus on extrapolation of the cost function in the direction of the number of the events. By extrapolating the cost function, we can estimate the cost function when the number of the events increases. We introduce a prior distribution computed by the extrapolated cost function. This prior distribution contains the information of the cost function obtained in the past. In the proposed method, we apply BO to the joint distribution of the prior distribution and the predictive distribution about the cost function with increased data.

1.1.3 Inelastic neutron-scattering experiments

The neutrons have no electric charge, but have a significant magnetic moment. Neutron scattering is caused by the interaction between magnetic moment of electrons or the nuclear force of the atoms. Neutrons have a high transmission rate, and are suitable for identifying the location and structure of light elements such as hydrogen and lithium. We can measure the atomic configuration and dynamics inside a solid simultaneously by using inelastic neutron-scattering experiments. This is applied for elucidating the structure of superconductivity.[26, 27, 28]. We show the outline of the inelastic neutron-scattering experiments below, and the process of the experiment in Fig. 1.2 and Fig. 1.3

The neutron beam is monochromatized by a Fermi chopper [29] and injected into a solid sample. Let the momentum of an incident neutron as $\hbar\mathbf{k}_i$ and, the energy of an incident neutron as $E_i = \frac{\hbar^2|\mathbf{k}_i|^2}{2m_N}$. Here, k, m_N, and $\hbar$ represents the wavenumber, the mass of the neutron, and the conversion Planck's constant, respectively. Let the momentum of an scattered neutron as $\hbar\mathbf{k}_f$ and, the energy of an scattered neutron as $E_f = \frac{\hbar^2|\mathbf{k}_i|^2}{2m_N}$. We can calculate the change of momentum and energy before and after the scattering as $\Delta\mathbf{Q}$ and ΔE. We consider that the incident neutrons are scattered by the nuclei in the solid. Let the momentum of an scattered neutron as $\hbar\mathbf{k}_f$ and, the energy of an scattered neutron as $E_f = \frac{\hbar^2|\mathbf{k}_i|^2}{2m_N}$. Then, we can calculate the change of momentum and energy before and after the scattering as $\Delta\mathbf{Q}$ and ΔE as follows.

$$\begin{aligned} \Delta\mathbf{Q} &= \hbar(\mathbf{k}_i - \mathbf{k}_f + \mathbf{G}) \\ \Delta E &= E_i - E_f \end{aligned} \tag{1.1}$$

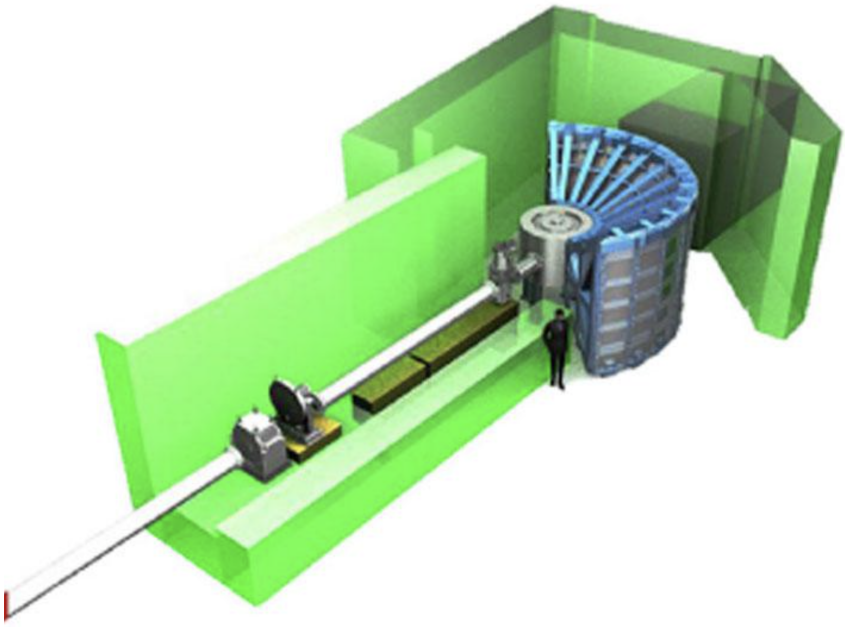


Figure 1.1: A schematic view of 4SEASONS.[1]

Here, $\mathbf{G}$ represents the reciprocal lattice vector. In J-PARC, researchers obtain large amount of event data $(\Delta E, \Delta \mathbf{Q})$ mapped on 4D space. We describe the way to calculate the wavenumber vectors $\mathbf{k}_i$, $\mathbf{k}_f$. From De Broglie's equation, we obtain

$$\mathbf{k} = \frac{m\mathbf{v}}{\hbar}. \tag{1.2}$$

Here, we define m, $\mathbf{v}$, and $\mathbf{k}$ as mass, velocity, and wavenumber vector of a matter, respectively. We can calculate wavenumber vector $\mathbf{k}$ from the velocity of the neutron $\mathbf{v}$ by using Eq. (1.2), Regarding time-of-flight method, researchers measure scattering angles and flighting time to calculate the velocity of the neutron.[30, 31] In previous devices, researchers had to set the location of the detectors for trials. On the other hand, 4SEASONS (Fig. 1.1) has detectors arranged in a spherical shape, and can detect a wide range of scattering angles. In addition, 4SEASONS has succeeded in utilizing multiple incident energies E_i during a single measurement [32, 17]. Therefore, the experimenter can perform measurement with a wider range of momentum space more efficiently than before. The experimenter can detect neutrons up to 300 meV with medium resolution. Since we can perform experiments with high efficiency with 4SEASONS, there are a large amount of inelastic neutron scattering data in J-PARC. However, researchers empirically select bin widths in a visual approach and make a histogram. It is an urgent task to develop a method for optimizing the bin widths for multidimensional event data. In addition, the redundancy of measurements is also a critical issue.

1.1.4 Time-of-flight method

The principle of time-of-flight (TOF) method [33, 34] of a pulsed neutron source is shown in Fig. 1.3. Neutrons fly a certain distance and are selected to be irradiated onto a sample at a speed suitable for the opening interval of a monochromatic chopper. Scattered neutrons are detected by detectors surrounding the sample. Opening of the chopper timing

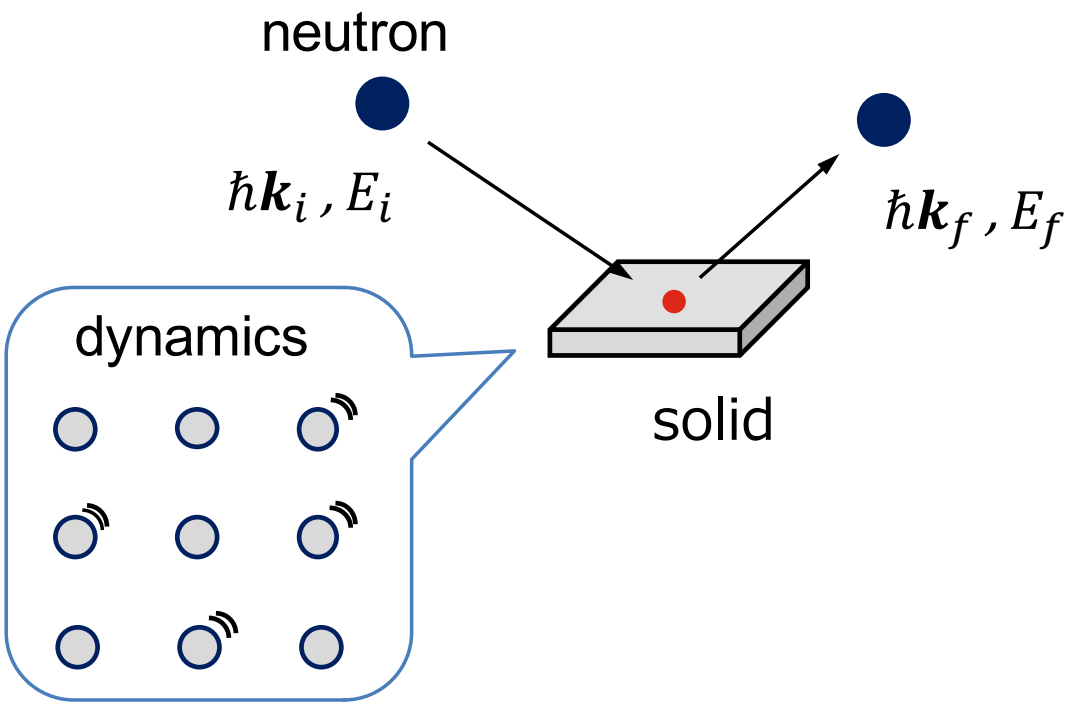


Figure 1.2: The process of inelastic neutron-scattering experiments. Let the momentum of an incident neutron as $\hbar\mathbf{k}_i$ and, the energy of an incident neutron as $E_i = \frac{\hbar^2|\mathbf{k}_i|^2}{2m_N}$. Let the momentum of an scattered neutron as $\hbar\mathbf{k}_f$ and, the energy of an scattered neutron as $E_f = \frac{\hbar^2|\mathbf{k}_i|^2}{2m_N}$. Here, k, m_N, and $\hbar$ represents the wavenumber, the mass of the neutron, and reduced Planck constant, respectively. We can calculate the change of momentum and energy before and after the scattering as $\Delta\mathbf{Q}$ and ΔE in Eq. (1.1).

The energy and momentum transitions can be computed from the total flight time, since the distance of each flight is known. Specifically, the magnitude and energy of the wavenumber vectors can be calculated as follows

$$k_i = \frac{m}{\hbar}\frac{L_1 - L_3}{t_{chm}}, \tag{1.3}$$

$$k_f = \frac{m}{\hbar}\frac{L_2}{t_d - t_s}, \tag{1.4}$$

$$E_i = \frac{1}{2}m\left(\frac{L_1 - L_3}{t_{chm}}\right), \tag{1.5}$$

$$E_f = \frac{1}{2}m\left(\frac{L_2}{t_d - t_s}\right). \tag{1.6}$$

Thus energy transition $\hbar\omega$ is writtens as

$$\hbar\omega = E_i - E_f. \tag{1.7}$$

Here, m_N is the mass of the neutron, L_1, L_2 and L_3 are the distance from the source to the sample, the sample to the detector, and the chopper to the sample, respectively t_{chm}, t_s, and t_d are the flight time of a neutron from the source to the chopper, sample, and detector, respectively. By using TOF method, researchers observe the information of intensity $I(t_d, \phi)$. Currently, the researchers mainly treat the peak information about the intensity for analysis in inelastic neutron-scattering experiments. Peak information is

extracted for the data transformed the form of histogram. The t_{chm} depends on the velocity (energy) of the incident neutron. Since the neutrons fly to the sample at the same speed, t_s is also determined. The equation for the energy resolution is given as follows

$$\left(\frac{\Delta\hbar\omega}{E_i}\right)^2 = \left(\frac{2\Delta t_{chm}}{t_{chm}}\right)\left(1+\frac{L_1}{L_2}\right)^2 + \left(\frac{2\Delta t_m}{t_{chm}}\right)\left(1+\frac{L_3}{L_2}\right)^2. \tag{1.8}$$

Here, Δt_{chm} and Δt_m are opening interval time of the chopper, and the time width of the neutron pulse at the source location. For simplicity, we show the case of $\hbar\omega = 0$. Strictly speaking, error terms of L_2 due to sample size and detector size are also exists, however it is ignored for simplicity. Momentum resolutions depend on the size and arrangement of the detection devices.

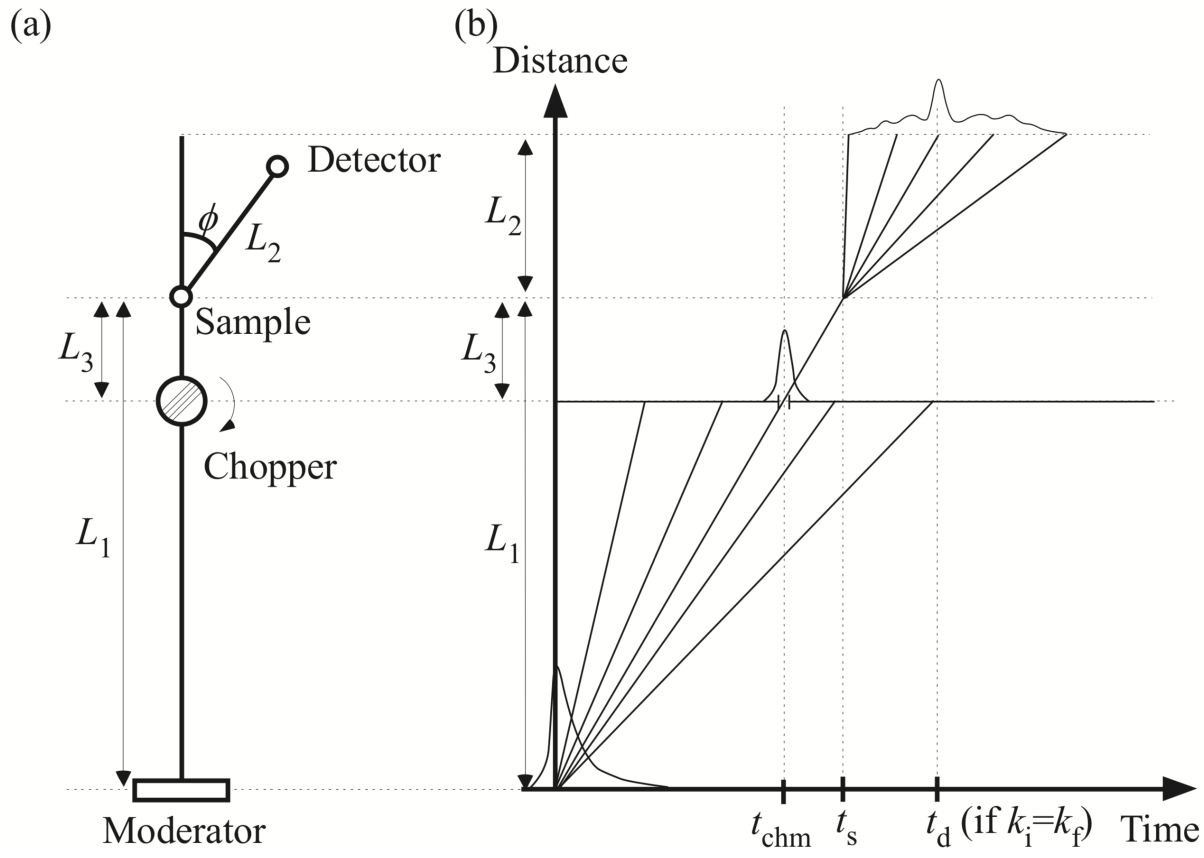


Figure 1.3: (a) A schematic view and (b) a TOF diagram of a conventional chopper spectrometer at a pulsed neutron source.[2] Neutrons selected by a chopper are injected into a solid and scattered neutrons are detected by detectors.

1.1.5 Density estimation

In probability and statistics, density estimation is the construction of an estimate $\hat{\lambda}_t$, based on observed data, of an unknown underlying rate. [35, 36, 37, 38] The data are randomly sampled from the underlying rate λ_t shown in Fig. 1.4. A variety of approaches to density estimation are used, including parametric methods such as Gaussian mixture model and nonparametric methods such as kernel density estimation. The most basic form of density estimation is a rescaled histograms.

A very natural use of density estimates is in the informal investigation of the properties of a given set of data. Density estimates can give valuable indication of such features as skewness and multimodality in the data. An important aspect of density estimation is data visualization. Density estimation is also frequently used in anomaly detection or novelty detection if an observation lies in a very low-density region, it is likely to be an anomaly or a novelty.

In statistics, the mean integrated squared error (MISE) is used in density estimation.[39] The MISE of an estimate of an unknown probability density is given by Among several plausible optimizing principles, such as the Kullback-Leibler divergence[40] or the Hellinger distance[41], we adopt, here, MISE for measuring the goodness-of-fit of an estimate to the unknown underlying rate.

$$\mathrm{MISE} = \int E[(\hat{\lambda}_t - \lambda_t)^2]dt \tag{1.9}$$

E refers to the expectation over different realizations of point events given λ_t. In this thesis, we estimate $\hat{\lambda}_t$ as a histogram, and explore a method to select the bin size that minimizes the MISE. We describes the detail about the method in Chapter. 2.

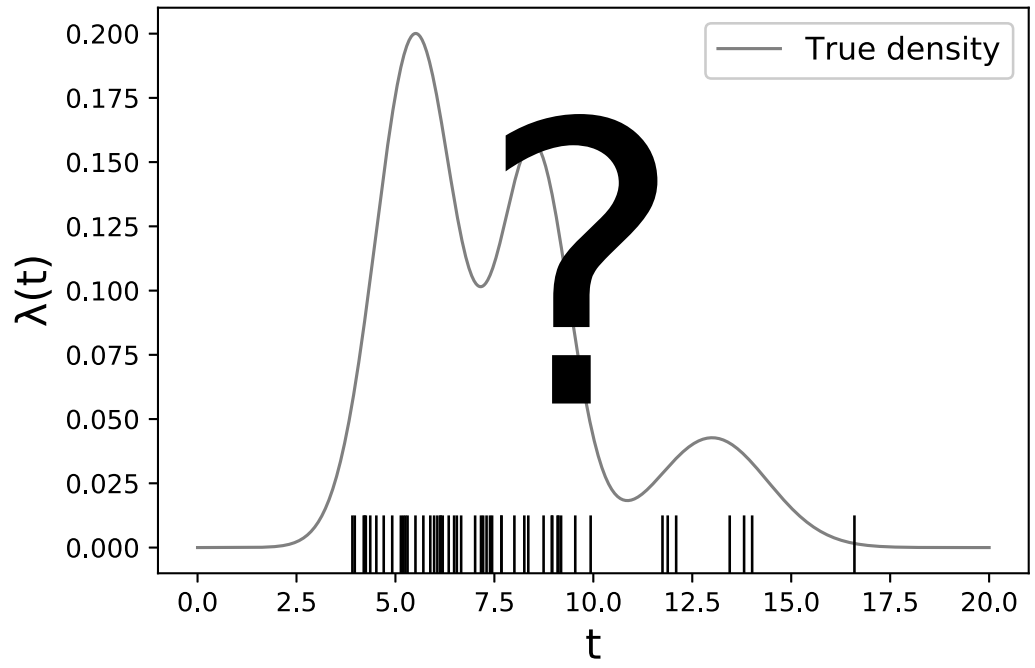


Figure 1.4: Problem setting of density estimation. Researchers estimate the unknown underlying rate $\lambda(t)$ from obtained event data.

We discuss an error in the optimal bin widths search and a fluctuation in bin height. In the proposed method, we conduct grid search. Therefore, an error corresponding to the grid size can exist. In other words, you can improve the accuracy by selecting small grid size. Moreover, we can reduce the computational cost by using summed-area talbes (SAT) [42] when performing a grid search. As increasing the number of events in a bin, we can reduce the fluctuation of the bin height. In current inelastic neutron scattering experiments, since the purpose of the experiment is peak estimation, researchers do not consider the reduction of the fluctuation.

1.2 Summary of contributions

We summarize the contributions of our work.

- Multidimensional Bin-Width Optimization for Histogram and Its Application to Four-Dimensional Neutron Inelastic Scattering Data

We have proposed a method for optimizing bin widths for multidimensional event data on the basis of a histogram bin-width optimization method.[21] Since the multidimensionalization of the bin-width optimization algorithm increases the number of parameters to be optimized, the computational cost increases. We have also proposed a method of reducing the computational cost. We generated event data from a dispersion relation by using Poisson sampling. First, to verify the validity of our method, we applied it to the 2D event data sliced from 4D event data. Second, we applied it to the whole 4D event data and compared the results of 2D bin-width optimization and 4D bin-width optimization. As a result, the optimal bin widths of the whole high-dimensional data and those of the data sliced from the high-dimensional data did not agree with each other. The result implies that the bin-width optimization should be performed for all of the high-dimensional data. In addition, since it is difficult for researchers to visually process 4D data, the proposed method is effective. Third, the optimal bin widths increase as the number of pieces of data increases or the magnitude of the white BG noise decreases. By using our method, we can eliminate arbitrariness in preprocessing and select the optimal bin widths for multidimensional event data. Regarding the contribution to machine learning, we formulate multidimensional cost function for bin-width optimization. Original content can be found in Ref. [22]

- Automatic Termination Strategy of Inelastic Neutron-scattering Measurement Using Bin-width Optimization

To prevent a redundant measurement in inelastic neutron-scattering measurement, we proposed a strategy to determine whether to terminate the measurement. In the proposed method, researchers compute approximate optimal bin widths as the stopping criteria in real time. When the optimal bin widths become smaller than the target resolutions, the experiment can be terminated. In numerical experiments, we dealt with real inelastic neutron-scattering data. In this study, we computed the cost function for real data of an inelastic neutron-scattering experiment for the first time. The optimal bin widths decrease as the number of data pieces increases. Moreover, we showed that Bayesian optimization is effective for searching for the optimal bin widths, especially when the number of data pieces is large. The cost function can be computed in parallel, and the computational resources can be saved by using BO. The estimated computational cost is not too significant to perform the stop-continue decision during the measurement. This work is planned to submit to a journal.

- Efficient Bayesian Bin-width Optimization for On-going Event Measurement

we proposed a strategy to determine efficiently whether to terminate or continue the measurement. As a criterion for determining the termination, we focused on the optimal bin widths of a histogram. For efficient termination judgment, it is necessary to improve the efficiency of bin widths optimization. We proposed a method using the prior distribution of BO computed from the information of the cost function obtained in the past. As a result of numerical experiments, it was found that the proposed method greatly improves the

search efficiency of the optimal bin widths. It was also found that the proposed method is robust for HP. Regarding contributions to machine learning, we formulate a Bayesian optimization method which incorporates information extracted from another domain as a prior distribution in optimal termination problems. This work is planned to submit to a journal.

1.3 Overview of this thesis

With respect to the structure of this thesis, we describe the bin widths optimization of multidimensional histograms in Chapter 2. In Chapter 3 and 4, we describe a proposal of an efficient termination strategy for event measurement. In Chapter 5 and 6, we conclude our contributions and present our future prospects.

Chapter 2

Multidimensional Bin-Width Optimization for Histogram and Its Application to Four-Dimensional Neutron Inelastic Scattering Data

2.1 Introduction

Event data often appear in many experimental methods such as neutron scattering experiments[12, 43, 44, 45, 46], radiation measurements[47, 48, 49, 50], and spike observation[51, 52, 53] in neuroscience. Generally, these data are smoothed by a histogram or a kernel in the process of data analysis.[54, 55, 56] In this paper, we treat smoothing by a histogram as the simplest case of such smoothing. With regard to neutron inelastic scattering experiments, researchers obtain event data mapped on four-dimensional (4D) space of energy and momentum. At Japan Proton Accelerator Research Complex (J-PARC), there are high-intensity Fermi chopper[29] spectrometers such as "4SEASONS"[17], "AMATERAS"[18], and "HRC"[19]. In recent years, numerous pieces of event data have been obtained in neutron inelastic scattering experiments by operating the chopper spectrometers. Researchers use the software called "Utsusemi"[20] to make histograms, while empirically adjusting the bin widths for the reduction of noise and data volume. If the bin widths are set to be wider, stochastic fluctuation of the event data can be reduced by smoothing; however, the structure of the data should be lost.[57, 58] Thus, selecting the best bin widths can be considered as an optimization of this tradeoff. Currently, researchers empirically select bin widths in a visual approach. Here, the problem is that the widths are chosen for 2D data sliced from 4D data. When cutting a 2D slice, researchers arbitrarily select the slice widths and position. These arbitrary processes should affect the result of selecting bin widths. The arbitrariness of preprocessing and human costs of trial and error are current issues.

To solve these problems, we focus on the fact that the neutron count in a specific region follows a Poisson distribution.[59] Several bin-width optimization methods have been proposed in neuroscience, although they have limited applicability for 1D spike sequences.[21, 60] These methods are used for estimating the firing rate, which represents the spike count of neurons per unit time. For the spike time series obtained in many experiment trials, it is known that the spike count in a specific time interval follows a Poisson distribution.[61, 62, 63] Therefore, we tried to develop a method applicable to multidimensional event data such as those obtained from neutron inelastic scattering experiments

on the basis of previous research in the field of neuroscience. In this study, we propose a method for automatically optimizing bin widths for whole multidimensinal data.

The structure of this chapter is as follows. In Sect. 2.2, we formulate a method for optimizing multidimensional bin widths. In Sect. 2.3, we show numerical experiments using artificial data to investigate the behavior of the proposed method. The method is applied to 4D artificial event data obtained by Poisson sampling.[64] The validity of our method is verified in the following steps. First, we apply the proposed method to 2D data sliced from 4D event data. Second, we apply it to the whole 4D event data and compare the results of 2D bin-width optimization and 4D bin-width optimization. We have found that the optimized bin widths strongly depends on the dimensionality of the data. Third, we compute the optimal bin widths by changing the number of data pieces and the magnitude of white background (BG) noise. It is shown that the optimal bin widths increase as the number of data pieces decreases or the magnitude of the white BG noise increases. In Sect. 2.4, we discuss the relationship between the optimal bin widths and the number of pieces of event data as well as the magnitude of white BG noise. The conclusion of this paper is provided in Sect. 2.5.

2.2 Method

2.2.1 Bin-width optimization for one-dimensional event data

First, we introduce the bin-width optimization method for 1D event data.[21] Let us assume that n pieces of event data $t_i \in [0, T]$ $(1 \leq i \leq n)$ are obtained in the observation interval $[0, T]$. Here, let $\lambda(t)$ be a true probability density function that the event data follow. Then, the performance of the estimator $\hat{\lambda}(t)$ for $\lambda(t)$ is evaluated using the mean integrated squared error (MISE) in Eq. (1). In statistics, MISE is used in density estimation[39]. There is Kullback–Leibler (KL) divergence[40] as a criterion for estimation accuracy, but when estimating a probability density with a histogram, KL can be infinite. [54] In this case, we use MISE.

$$\mathrm{MISE} = \frac{1}{T}\int_0^T E\left[\left(\hat{\lambda}(t) - \lambda(t)\right)^2\right] dt \tag{2.1}$$

Here, $E[\cdot]$ represents the expectation over different realizations of event data given $\lambda(t)$. When the observation interval is equally divided into N parts and the estimator $\hat{\lambda}(t)$ is limited to the form of the histogram of the bin width $\Delta = \frac{T}{N}$, Eq. (2.1) can be expressed as

$$\begin{aligned} \mathrm{MISE} &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\Delta}\int_0^{\Delta} E_{Poisson}\left[\left(\hat{\theta}_i - \lambda(t + \Delta(i-1))\right)^2\right] dt, \\ \text{where} \quad E_{Poisson}[\cdot] &:= \sum_{k_i=0}^{\infty} P(k_i|n\Delta\theta_i)[\cdot]. \end{aligned} \tag{2.2}$$

Here, $\hat{\theta}_i$ represents the height of the i th bin. $\hat{\theta}_i$ can be written as $\hat{\theta}_i = \frac{k_i}{n\Delta}$ by using the number of counts k_i in the i th bin. Using the parameter $\theta_i = \frac{1}{\Delta}\int_0^{\Delta}\lambda(t + \Delta(i-1))dt$, which is the average of $\lambda(t)$ in the i th bin, the probability distribution $P(k_i|n\Delta\theta_i)$ can be written as Eq. (2.3). In Eq. (2.3), we assume that the data generation process is a Poisson process.

$$P(k_i|n\Delta\theta_i) = \frac{(n\Delta\theta_i)^{k_i}}{k_i!}\exp(-n\Delta\theta_i) \tag{2.3}$$

Since $E[\hat{\theta}_i] = \theta_i$ holds for any i, $\hat{\theta}_i$ is the unbiased estimator of the parameter θ_i. Hereafter, we denote $\lambda(t + (i-1)\Delta)$ in Eq. (2.2) as an average over an ensemble of segmented probability density functions $\{\lambda(t)\},\ t \in [0, \Delta]$, and $E_{Poisson}[\cdot]$ as $E[\cdot]$.

$$\text{MISE} = \frac{1}{\Delta}\int_0^\Delta \left\langle E\left[(\hat{\theta} - \lambda(t))^2\right]\right\rangle dt \tag{2.4}$$

The quadratic term of Eq. (2.4) can be decomposed as

$$\text{MISE} = \left\langle E\left[(\hat{\theta} - \theta)^2\right]\right\rangle + \frac{1}{\Delta}\int_0^\Delta \left\langle(\lambda(t) - \theta)^2\right\rangle dt. \tag{2.5}$$

The first term is the stochastic fluctuation of the estimator $\hat{\theta}$ around its expected mean value θ $(= E[\hat{\theta}])$, and the second term is the averaged fluctuation of $\lambda(t)$ around its section average θ. We also expand the second term of Eq. (2.5) as

$$\frac{1}{\Delta}\int_0^\Delta \left\langle(\lambda(t) - \theta)^2\right\rangle dt = \frac{1}{\Delta}\int_0^\Delta \left\langle(\lambda(t) - \langle\theta\rangle)^2\right\rangle dt - \left\langle(\theta - \langle\theta\rangle)^2\right\rangle \tag{2.6}$$

$$= \frac{1}{T}\int_0^T (\lambda(t) - \langle\theta\rangle)^2 dt - \left\langle(\theta - \langle\theta\rangle)^2\right\rangle. \tag{2.7}$$

Since the first term in Eq. (2.7) does not depend on the bin width Δ, we eliminate it from MISE and let the cost function $C_n(\Delta)$ be Eq. (2.9). Then, the bin width that minimizes the cost function is taken as the optimal bin width.

$$\frac{1}{n^2}C_n(\Delta) := \text{MISE} - \frac{1}{T}\int_0^T (\lambda(t) - \langle\theta\rangle)^2 dt \tag{2.8}$$

$$= \left\langle E[(\hat{\theta} - \theta)^2]\right\rangle - \left\langle(\theta - \langle\theta\rangle)^2\right\rangle \tag{2.9}$$

The second term in Eq. (2.9) represents the fluctuation of the expected mean value θ around the ensemble average $\langle\theta\rangle$, where $\langle\cdot\rangle$ means the average operation for all bins. Since $\langle\theta\rangle$ is unobservable, it is eliminated by using Eq. (2.10), which is the decomposition rule for an unbiased estimator $\hat{\theta}$.

$$\left\langle E\left[\left(\hat{\theta} - \left\langle E[\hat{\theta}]\right\rangle\right)^2\right]\right\rangle = \left\langle E[(\hat{\theta} - \theta)^2]\right\rangle + \left\langle(\theta - \langle\theta\rangle)^2\right\rangle \tag{2.10}$$

From Eqs. (2.9) and (2.10), the following equation is obtained.

$$\frac{1}{n^2}C_n(\Delta) = 2\left\langle E[(\hat{\theta} - \theta)^2]\right\rangle - \left\langle E\left[\left(\hat{\theta} - \left\langle E[\hat{\theta}]\right\rangle\right)^2\right]\right\rangle \tag{2.11}$$

Since the number of counts in the i th bin follows a Poisson distribution[65] and the estimator of bin height is denoted as $\hat{\theta}_i = \frac{k_i}{n\Delta}$, the next equation holds.

$$E[(\hat{\theta} - \theta)^2] = \frac{1}{n\Delta}E[\hat{\theta}] \tag{2.12}$$

Then, the cost function only has the terms of the estimator $\hat{\theta}$.

$$\frac{1}{n^2}C_n(\Delta) = \frac{2}{n\Delta}\left\langle E[\hat{\theta}]\right\rangle - \left\langle E\left[\left(\hat{\theta} - \left\langle E[\hat{\theta}]\right\rangle\right)^2\right]\right\rangle \tag{2.13}$$

$$C_n(\Delta) = \frac{2\langle E[k]\rangle - \langle E\left[(k - \langle E[k]\rangle)\right]\rangle}{\Delta^2} \tag{2.14}$$

In Eq. (2.14), we apply $\hat{\theta}_i = \frac{k_i}{n\Delta}$. We obtain the estimated cost function $\hat{C}_n(\Delta)$ of the observable value k_i.

$$\hat{C}_n(\Delta) = \frac{2\bar{k} - v}{\Delta^2} \tag{2.15}$$

$$\bar{k} = \frac{n}{N},\ v = \frac{1}{N}\sum_{i=1}^{N}(k_i - \bar{k})^2 \tag{2.16}$$

The steps of the algorithm are as follows.

1. Assume that n pieces of event data are obtained in the observation interval of length T.

2. Set the bin width Δ (number of bins: $N = \frac{T}{\Delta}$). Compute the number of counts k_i in the i th bin ($1 \leq i \leq N$).

Then, compute v by using $\bar{k} = \frac{n}{N}$, as in Eq. (2.17).

$$v = \frac{1}{N}\sum_{i=1}^{N}(k_i - \bar{k})^2 \tag{2.17}$$

3. Compute $\hat{C}_n(\Delta)$ defined by

$$\hat{C}_n(\Delta) = \frac{2\bar{k} - v}{\Delta^2}. \tag{2.18}$$

Repeat steps 2 and 3 above to find the bin width Δ that minimizes $\hat{C}_n(\Delta)$, and let this be the optimal bin width.

2.2.2 Multidimensionalization of bin-width optimization algorithm

Let us consider the case in which n pieces of event data are mapped on $\mathbb{R}^d$ and the observation area is equally divided into N d-dimensional rectangular parallelepipeds just as in the 1D case [bin widths: $(\Delta_1, \Delta_2, ..., \Delta_d)$]. Assuming that the number of counts in the i th bin is k_i, we can compute the estimator of the cost function $\hat{C}_n(\Delta_1, \Delta_2, ..., \Delta_d)$ in the same way as for 1D event data.

$$\hat{C}_n(\Delta_1, \Delta_2, ..., \Delta_d) = \frac{2\bar{k} - v}{(\Delta_1\Delta_2...\Delta_d)^2} \tag{2.19}$$

$$\bar{k} = \frac{n}{N},\ v = \frac{1}{N}\sum_{i=1}^{N}(k_i - \bar{k})^2 \tag{2.20}$$

For multidimensional event data, let the bin widths that minimize $\hat{C}_n(\Delta_1, \Delta_2, ..., \Delta_d)$ by changing bin width $(\Delta_1, \Delta_2, ..., \Delta_d)$ be the optimal bin widths. The estimator of the height of the i th bin is $\hat{\theta}_i = \frac{k_i}{n\Delta_1\Delta_2...\Delta_d}$.

2.2.3 Reducing computational cost by using cumulative sum

Computational cost increases when the algorithm is multidimensionalized, and therefore it is necessary to reduce the cost. Using the summed-area table (SAT) algorithm, we can reduce the cost to compute the number of counts in a bin.[42] We show an outline

of the SAT algorithm in Fig. 2.1. To apply the SAT algorithm, it is necessary to make event data into a sufficiently fine histogram. In the following, we show a concrete method for reducing the computational cost for 1D data. The method below can be expanded to the data of two or more dimensions.

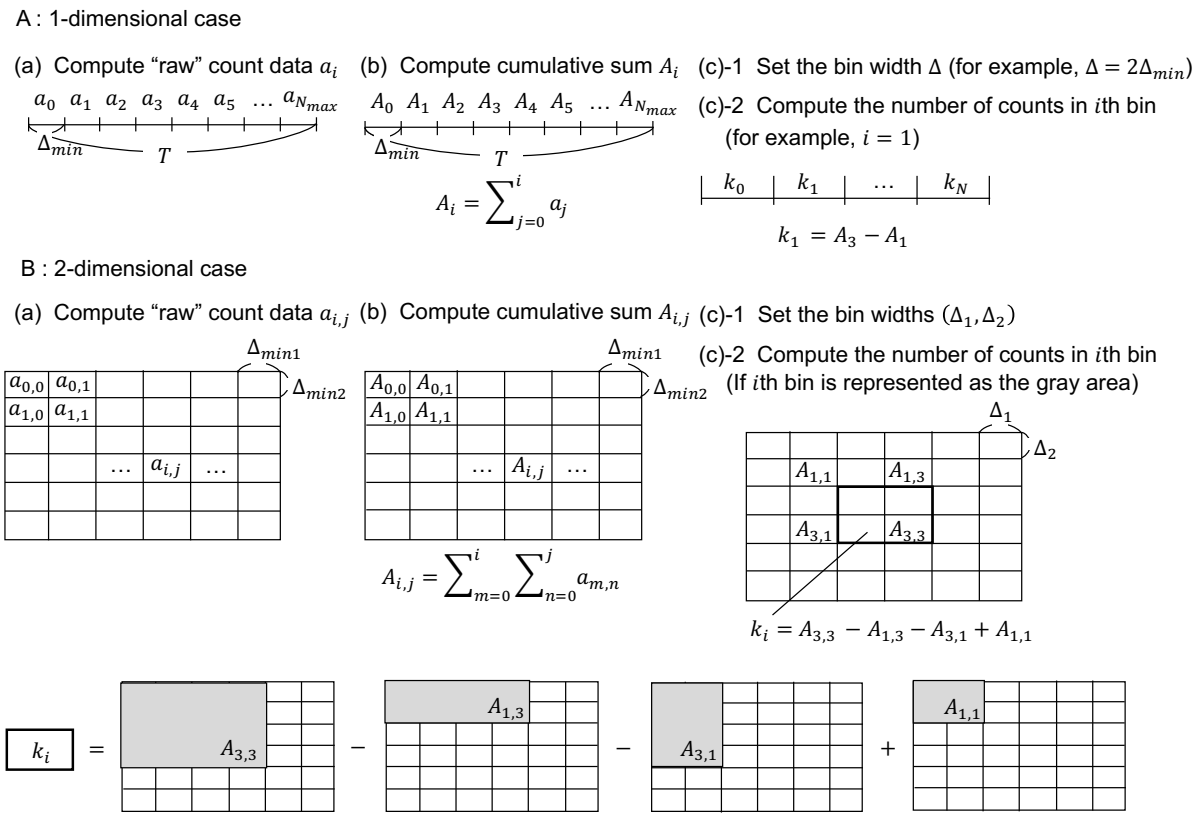


Figure 2.1: Steps of the SAT algorithm. A(a)–A(c) show 1D case. B(a)–B(c) show 2D case. If we do not use the SAT algorithm, we have to check whether each event is in each bin. Since this computational cost is proportional to the number of pieces of event data, the SAT algorithm is effective when the number of pieces of event data is large.

1. Assume that n pieces of event data are obtained in the observation interval of length T.

2. Define the value of the minimum unit of bin width as Δ_{min} and limit the area for searching for the optimal bin width to a positive integer multiple of Δ_{min}. Note that the optimal bin width is limited within the group of $\left\{c\Delta_{min} | 1 \leq c \leq \frac{T}{\Delta_{min}}\right\}$, as described below. We divide the observation interval into $N_{max} = \frac{T}{\Delta_{min}}$ in width Δ_{min}. Let a_i $(1 \leq i \leq N_{max})$ be the number of pieces of event data in the i th bin. Here, a_i represents the number of pieces of "raw" event data obtained from the experiment, and k_i represents the number of pieces of count data after setting a certain bin against the "raw" event data. Here, we define the term "raw" count data as an array a.

3. As a preparation for reducing the computational cost in step 4, compute the cumulative sum A_i as $A_i = \sum_{j=1}^{i} a_j$. The number of counts in an arbitrary section in the observation interval can be computed as the difference in cumulative sums. By using the cumulative sum, the computational cost in step 4 can be reduced.

4. Fix an integer c in the range $1 \leq c \leq \frac{N_{max}}{2}$, set the bin width Δ to $\Delta = c\Delta_{min}$, and equally divide the observation interval into N $(= \frac{T}{\Delta})$. In our approach, we compute

the number k_i of counts in the i th bin as a difference of the elements of cumulative sum A (Fig. 2.1A(c)–2) and obtain v by using $\bar{k} = \frac{n}{N}$ as Eq. (2.20).

$$v = \frac{1}{N}\sum_{i=1}^{N}(k_i - \bar{k})^2 \tag{2.21}$$

5. Compute $\hat{C}_n(\Delta)$ as

$$\hat{C}_n(\Delta) = \frac{2\bar{k} - v}{\Delta^2}. \tag{2.22}$$

Repeat steps 4 and 5 while changing the integer c within $1 \leq c \leq \frac{N_{max}}{2}$, find the bin width Δ that minimizes $\hat{C}_n(\Delta)$, and define this as the optimal bin width.

Basically, we explore all pairs of bin widths $(\Delta_1, \Delta_2, ..., \Delta_d)$ to find the optimal bin widths in the d-dimensional case, as shown in Fig. 2.1B(c)–2.

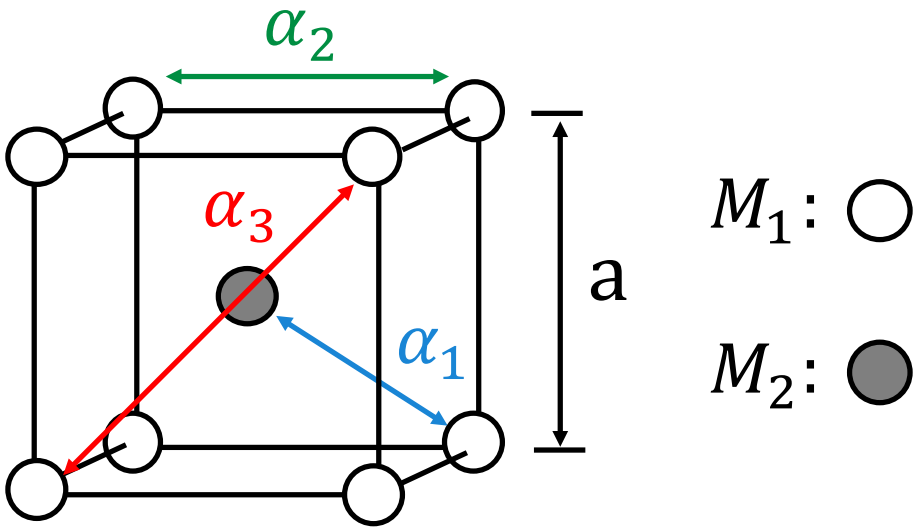


Figure 2.2: (Color online) Assumed diatomic body-centered-cubic lattice model. The mass of a white atom is M_1 and that of a gray atom is M_2. The lattice constant is a, and the spring constants corresponding to the first to third neighboring interactions are α_1–α_3.

2.2.4 Computational cost and memory usage

We discuss about the computational complexity and memory usage of the proposed method. In this paper, we chose naive implementation and nearest neighbor search for comparison The results are shown in table 2.1. In the proposed method, the computational cost and memory usage mostly depend on the maximun division number in each direction. On the other hand, the computational cost depend on the maximun division numbers and the number of events, and the memory usage mostly depends on the number of events. There is a trade-off between computational cost and memory usage. As the number of dimensions increases, SAT uses more memory, however, n_{data} and n_{NN} should be bottleneck of the computational cost.

	computational cost	memory usage
naive	$\mathrm{O}(\mathrm{n}_{data}\prod_{i=1}^{d}N_i\log N_i)$	$\mathrm{O}(\mathrm{n}_{data})$
NN	$\mathrm{O}(\mathrm{n}_{NN}\prod_{i=1}^{d}N_i\log N_i)$	$\mathrm{O}(\mathrm{n}_{data})$
SAT	$\mathrm{O}(\prod_{i=1}^{d}N_i\log N_i)$	$\mathrm{O}(\prod_{i=1}^{d}N_i)$

Table 2.1: The computational complexity and memory usage of the naive implementation (naive), nearest neighbor search (NN), and proposed method (SAT). d, n_{data}, n_{NN}, N_i represents the dimension, number of events, number of the events in the nearest cell, the maximun division number in i-th direction, respectively. We assume that N_i, n_{data}, n_{NN} are sufficiently larger than d.

2.2.5 Error and fluctuation of bin widths optimization

We discuss an error in the optimal bin widths search and a fluctuation in bin height In the proposed method, we set the minimum units of bin widths and conduct grid search. Therefore, an error corresponding to the minimum unit can exist in each direction. The variance of the bin height $\hat{\theta}$ follows the $\frac{\theta}{n\Delta}$. As increasing the number of events in a bin, we can reduce the fluctuation of the bin height. In current experiments, since the purpose of the experiment is peak estimation, researchers do not consider the reduction of the fluctuation. The peaks correspond to the dispersion relation. Researchers perform parameter estimation of the Hamiltonian by using the information of the dispersion relation.

2.3 Results

2.3.1 Applying the proposed method to artificial data

We dealt with the physical model shown in Fig. 2.2, and applied the proposed method to event data generated from the model. We used a diatomic cubic lattice model for which the masses of atoms at the corners and center of each unit cell are different. The first to third neighboring elastic interactions were taken into account. Let the lattice constant be a and the masses of the atoms be M_1 and M_2. We defined the elastic constants corresponding to the first, second, and third neighboring interactions as α_1, α_2, and α_3, respectively. Using this model, we derived the probability density $P(E, q_x, q_y, q_z)$ that the event data follow. The detailed derivation is shown in Appendix. A.1 In this paper, $P(E, q_x, q_y, q_z)$ was calculated with $M_1 = 2 \times 10^{-26}[kg]$, $M_2 = 4 \times 10^{-26}[kg]$, $\alpha_1 = 100[N/m]$, $\alpha_2 = 200[N/m]$, $\alpha_3 = 300[N/m]$, and the event data were generated by Poisson sampling [64] to follow $P(E, q_x, q_y, q_z)$.

First, we set the values of the minimum units of the bin widths to make “raw" count data. The values of minimum units of the bin widths in the energy direction and the momentum directions are set to be $\Delta_{min,E} = \frac{23}{7}, \Delta_{min,q_x} = \Delta_{min,q_y} = \Delta_{min,q_z} = \frac{1}{70}$, respectively. To verify the validity of our method, we applied it to the 2D count data sliced from the “raw" count data. Figures 2.3(a)–2.3(c) are 2D slices of $P(E, q_x, q_y, q_z)$, and Figs. 2.3(d)–2.3(f) are 2D slices of 4D “raw" count data. The total number of pieces of event data, n, is $10,000,000$. Figures 2.4(a)–2.4(c) show the corresponding optimal histograms computed from count data shown in Figs. 2.3(d)–2.3(f), respectively. Here, Figs. 2.3(a)–2.3(c) are sliced from $-\frac{9}{35} \leq q_y, q_z < -\frac{17}{70}$, $0 \leq q_x, q_z < \frac{1}{70}$, and $\frac{1196}{7} \leq E < \frac{1219}{7}$, $\frac{17}{70} \leq q_x < \frac{9}{35}$, respectively. Figures 2.3(d) and 2.4(a) are sliced from $-\frac{2}{7} \leq q_y, q_z < -\frac{17}{70}$. Figures 2.3(e) and 2.4(b) are sliced from $-\frac{2}{70} \leq q_x, q_z < -\frac{1}{70}$. Figures 2.3(f) and 2.4(c) are sliced from $\frac{1173}{7} \leq E < \frac{1242}{7}$, $\frac{8}{35} \leq q_x < \frac{19}{70}$. The optimal bin widths for the sliced count data are $(\Delta_E = \frac{23}{7}, \Delta_{q_x} = \frac{4}{70})$ in the q_x-E plane, $(\Delta_E = \frac{23}{7}, \Delta_{q_y} = \frac{1}{70})$ in the q_y-E plane,

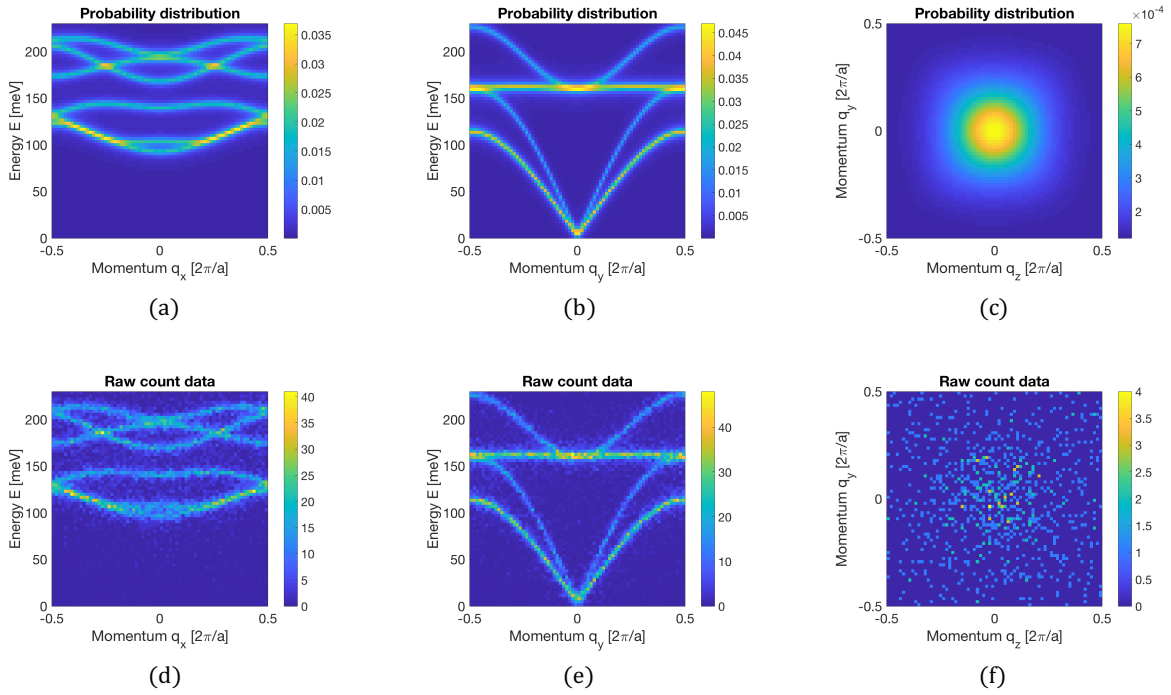


Figure 2.3: (Color online) (a)–(c) 2D slices of the 4D probability distribution $P(E, q_x, q_y, q_z)$ calculated from the dynamical model. (d)–(f) 2D slices of 4D event data generated by the probability distribution $P(E, q_x, q_y, q_z)$. Here, (a) is sliced from $-\frac{9}{35} \leq q_y, q_z < -\frac{17}{70}$. (b) is sliced from $0 \leq q_x, q_z < \frac{1}{70}$. (c) is sliced from $\frac{1196}{7} \leq E < \frac{1219}{7}$, $\frac{17}{70} \leq q_x < \frac{9}{35}$. (d) is sliced from $-\frac{2}{7} \leq q_y, q_z < -\frac{17}{70}$. (e) is sliced from $-\frac{2}{70} \leq q_x, q_z < -\frac{1}{70}$. (f) is sliced from $\frac{1173}{7} \leq E < \frac{1242}{7}$, $\frac{8}{35} \leq q_x < \frac{19}{70}$.

and $(\Delta_{q_y} = \frac{12}{35}, \Delta_{q_z} = \frac{12}{35})$ in the q_z-q_y plane. The optimal bin widths for the slice are much larger in the q_z-q_y plane than in the q_x-E plane and q_y-E plane. Figures 2.5(a)–(c) are 2D cost functions obtained by applying the proposed method to 2D count data (Figs. 2.3(d)–2.3(f)).

Second, we applied the proposed method to all the 4D count data as shown in Figs. 2.4(d)–2.4(f). Figures 2.4(d)–2.4(f) are sliced from $-\frac{2}{7} \leq q_y, q_z < -\frac{17}{70}$, $-\frac{2}{70} \leq q_x, q_z < -\frac{1}{70}$, and $\frac{1173}{7} \leq E < \frac{1242}{7}$, $\frac{8}{35} \leq q_x < \frac{19}{70}$, respectively. Figures 2.5(d)–2.5(f) are 2D slices of a 4D cost function $\hat{C}(\Delta_E, \Delta_{q_x}, \Delta_{q_y}, \Delta_{q_z})$. Figures 2.5(d)–2.5(f) represent $\hat{C}(\Delta_E, \Delta_{q_x}, \frac{3}{70}, \frac{3}{70})$, $\hat{C}(\Delta_E, \frac{3}{70}, \Delta_{q_y}, \frac{3}{70})$, and $\hat{C}(\frac{23}{7}, \frac{3}{70}, \Delta_{q_y}, \Delta_{q_z})$, respectively. Comparison of the results of 2D [Figs. 2.4(a)–2.4(c)] and 4D bin-width optimizations [Figs. 2.4(d)–2.4(f)] shows that the 2D and 4D optimizations are not exactly the same. We conducted an additional numerical experiment to verify the validity of the selected the values of the minimum units of bin widths. Fixing the other bin widths to their optimal values, we computed the cost function more finely in the q_x direction. Here, we set the value of the minimum unit of the bin width in the q_x direction to $\Delta_{min,q_x}/8$ $(= \frac{1}{560})$, and the result is shown in Fig. 2.6. As a result, the optimal bin width computed in this experiment $\Delta_{q_x} = \frac{1}{28}$ is close to $\frac{3}{70}$.

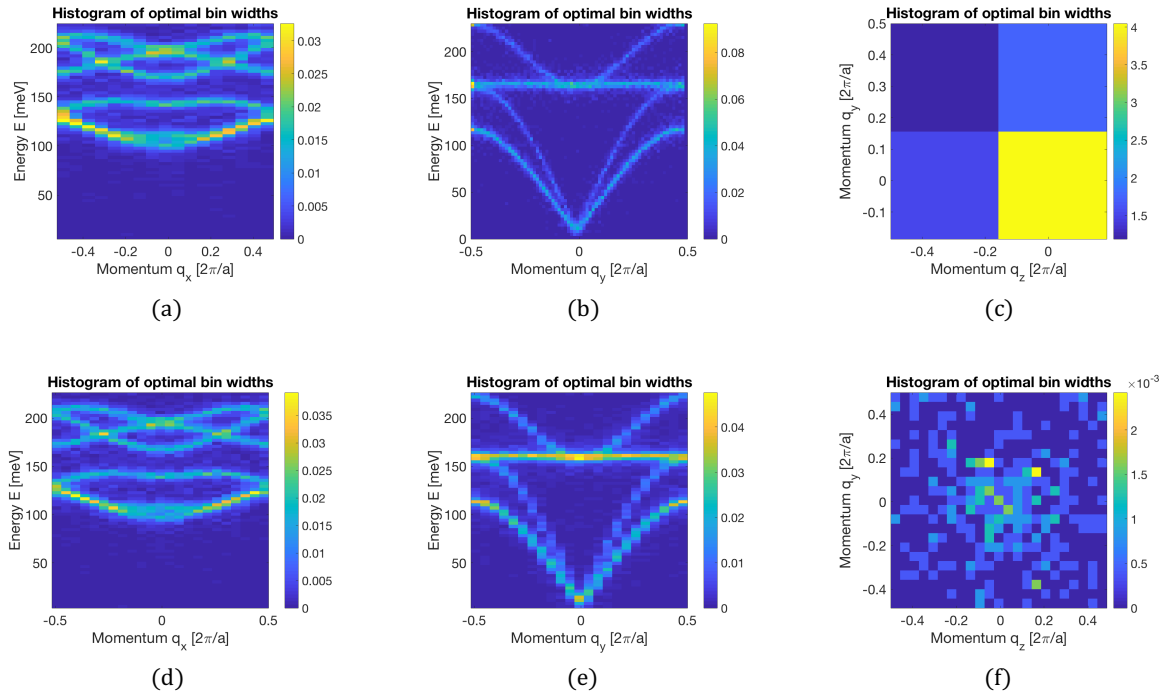


Figure 2.4: (Color online) (a)–(c) 2D histograms made by applying the proposed method to 2D event data in Figs. 2.3(d)–2.3(f), respectively. The optimal bin widths of (a) are $(\Delta_E = \frac{23}{7}, \Delta_{q_x} = \frac{4}{70})$. The optimal bin widths of (b) are $(\Delta_E = \frac{23}{7}, \Delta_{q_y} = \frac{1}{70})$. The optimal bin widths of (c) are $(\Delta_{q_y} = \frac{12}{35}, \Delta_{q_z} = \frac{12}{35})$. (d)–(f) 2D slices of a 4D histogram made by applying the proposed method to 4D count data. (d) is sliced from $-\frac{2}{7} \leq q_y, q_z < -\frac{17}{70}$. (e) is sliced from $-\frac{2}{70} \leq q_x, q_z < -\frac{1}{70}$. (f) is sliced from $\frac{1173}{7} \leq E < \frac{1242}{7}$, $\frac{8}{35} \leq q_x < \frac{19}{70}$.

2.3.2 Investigating properties of proposed method for number of pieces of event data and magnitude of background noise

To investigate the behavior of the proposed method against data, we conducted numerical experiments. In this paper, we focused on the tendency of the obtained optimal bin widths and the robustness of the proposed method against data. While we systematically changed the magnitude of white BG noise and the number of pieces of data, we applied the proposed method to the numerical-experiment data. In the 2D case, we used a probability distribution $P(E, q)$ by adding white BG noise to the intensity function $I(E, q, 0, 0)$ and normalizing it. We defined r_{BG} as the ratio of the sum of the white BG noise to the magnitude of the white BG noise M_{BG} and $P(E, q)$ as follows.

$$r_{BG} = \frac{2\pi E_{max} M_{BG}}{2\pi E_{max} M_{BG} + a \int_0^{E_{max}} \int_{-\frac{\pi}{a}}^{\frac{\pi}{a}} I(E, q, 0, 0) dE dq} \tag{2.23}$$

In this experiment, the values of the minimum units of the bin width in the energy direction and the momentum directions were $\Delta_{min,E} = 1.15$ and $\Delta_{min,q} = 0.005$, respectively. Figure 2.7(a) shows the probability distribution for $r_{BG} = 0$. Figures 2.7(b)–2.7(d) are "raw" count data for $r_{BG} = 0$, $r_{BG} = 0.50$, and $r_{BG} = 0.86$, respectively. We computed the product of optimal bin widths $\Delta_E \Delta_q$ for each data under the condition $r_{BG}, n = 100,000$ as shown in Fig. 2.8. We also show the results under the conditions n,$r_{BG} = 0$ and $r_{BG} = 0.317$ in Figs. 2.9 and 2.10, respectively. We performed the same experiments done to obtain Figs. 2.8–2.10 for the whole 4D event data. The corresponding

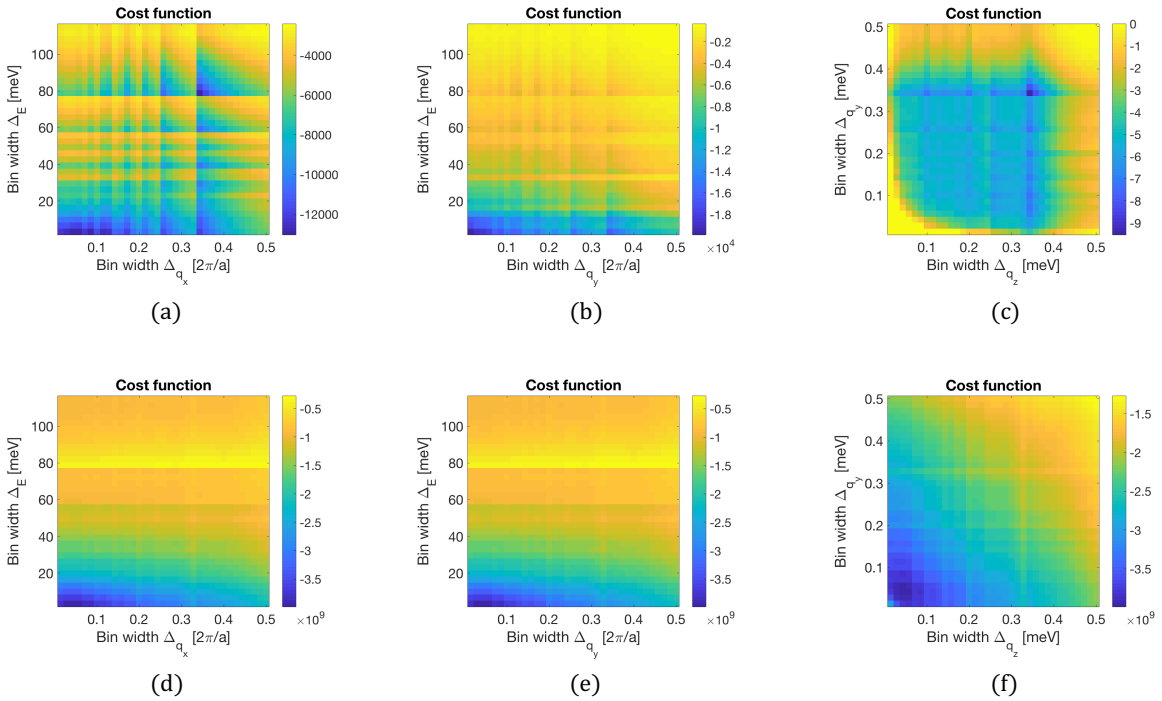


Figure 2.5: (Color online) (a)–(c) 2D cost functions obtained by applying the proposed method to 2D count data in Figs. 2.3(d)–2.3(f), respectively. (d)–(f) 2D slices of a 4D cost function $\hat{C}(\Delta_E, \Delta_{q_x}, \Delta_{q_y}, \Delta_{q_z})$. (d) represents $\hat{C}(\Delta_E, \Delta_{q_x}, \frac{3}{70}, \frac{3}{70})$. (e) represents $\hat{C}(\Delta_E, \frac{3}{70}, \Delta_{q_y}, \frac{3}{70})$. (f) represents $\hat{C}(\frac{23}{7}, \frac{3}{70}, \Delta_{q_y}, \Delta_{q_z})$. In this experiment, the value of the minimum unit of the bin width in the energy direction was $\Delta_{min,E} = \frac{23}{7}$, and those in the momentum directions were $\Delta_{min,q_x} = \Delta_{min,q_y} = \Delta_{min,q_z} = \frac{1}{70}$.

results are shown in Figs. 2.11–2.13. In these experiments, the value of the minimum units of the bin width in the energy direction was $\Delta_{min,E} = 5.75$, and those in the momentum directions were $\Delta_{min,q_x} = \Delta_{min,q_y} = \Delta_{min,q_z} = 0.025$. In the 4D case, r_{BG} is expressed as

$$r_{BG} = \frac{2\pi E_{max} M_{BG}}{2\pi E_{max} M_{BG} + a}. \tag{2.24}$$

As can be seen in Figs. 2.8–2.13, the optimal bin widths tend to increase as the total number of pieces of event data decreases or the magnitude of white BG noise increases. From Figs. 2.8–2.13, we can confirm that the proposed method extracts optimal bin widths almost uniquely in the case of a large amount of event data and a small magnitude of white BG noise.

2.4 Discussion

In this section, we discuss the results of the numerical experiments conducted in the previous section. First, we discuss how the cost function depends on the bin widths by using Eqs. (19) and (20). We obtain

$$\begin{aligned} \hat{C}_n(\Delta_1, \Delta_2, ..., \Delta_d) &\propto -\frac{1}{\Delta}\sum_{i=1}^{N}({k_i}^2 - 2k_i) + \text{const}, \\ \text{and} \quad \Delta &= \Delta_1\Delta_2...\Delta_d. \end{aligned} \tag{2.25}$$

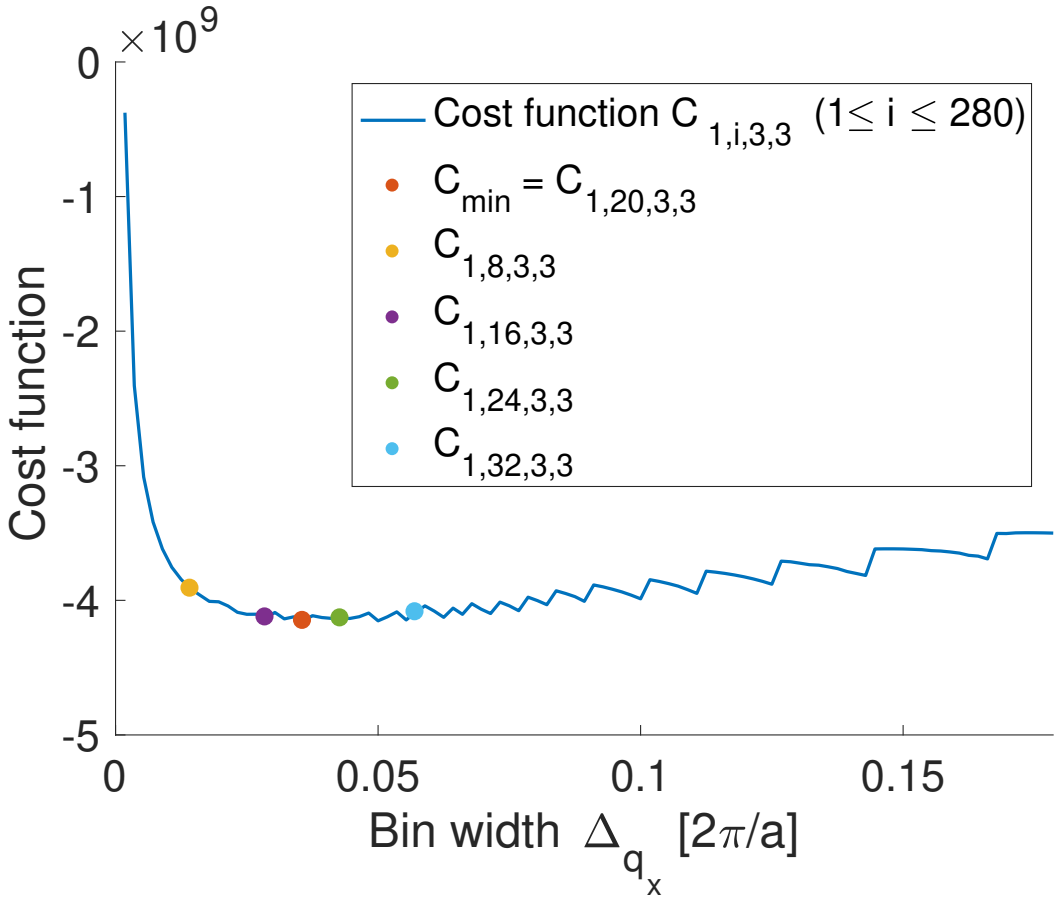


Figure 2.6: (Color online) Line plot representing the cost $C(\Delta_{min,E}, \Delta_{q_x}, 3\Delta_{min,qy}, 3\Delta_{min,qz})$. Δ_{q_x} is limited within the group of $\{i\Delta_{min,q_x}/8 | 1 \leq i \leq 280\}$. Here, Δ_{min,q_x} is equal to $\frac{1}{70}$, as described in Sect. 3. $C_{1,i,3,3}$ represents $C(\Delta_{min,E}, i\Delta_{min,q_x}/8, 3\Delta_{min,qy}, 3\Delta_{min,qz})$, $(1 \leq i \leq 280)$. $C_{1,20,3,3}$ achieves the minimum value of the cost function. The line plot is shown within $1 \leq i \leq 100$.

Since the value of the cost function decreases as the first term of Eq. (25) decreases, we obtained bin widths that increase $\frac{1}{\Delta}\sum_{i=1}^{N}({k_i}^2 - 2k_i)$. There is a tradeoff between an increase in $({k_i}^2 - 2k_i)$, corresponding to increasing bin widths, and an increase in $\frac{1}{\Delta}\sum_{i=1}^{N}$, corresponding to decreasing bin widths. Figures 2.8–2.13 show that the optimal bin widths increase as the total number of event data decreases or the magnitude of white BG noise increases. These results can be interpreted as the total number of event data being small or the magnitude of white BG noise being large and the effect of increasing $({k_i}^2 - 2k_i)$ in the first term of Eq. (25) becoming dominant. In other words, the effect of smoothing becomes dominant.

Second, we discuss discrete patterns in the cost functions. In the cost functions shown in Fig. 2.5, the cost functions abruptly change when the bin widths change. For example, a sudden change can be seen around $(\Delta_E = \frac{552}{7}, \Delta_{q_x} = \frac{12}{35})$ in Fig. 2.5(a). The change in bin widths is continuous, whereas the change in the number of bins is discrete. By changing the bin widths, the number of bins may change. At this time, the value of the cost function greatly changes and results in discrete patterns. Not only the bin widths but also the locations of the slices and the slice widths affect the result. In Fig. 2.4, we compared the result of bin-width optimization for 4D data and 2D data sliced from the 4D data. The results for the two cases were different from each other. When slicing data, researchers

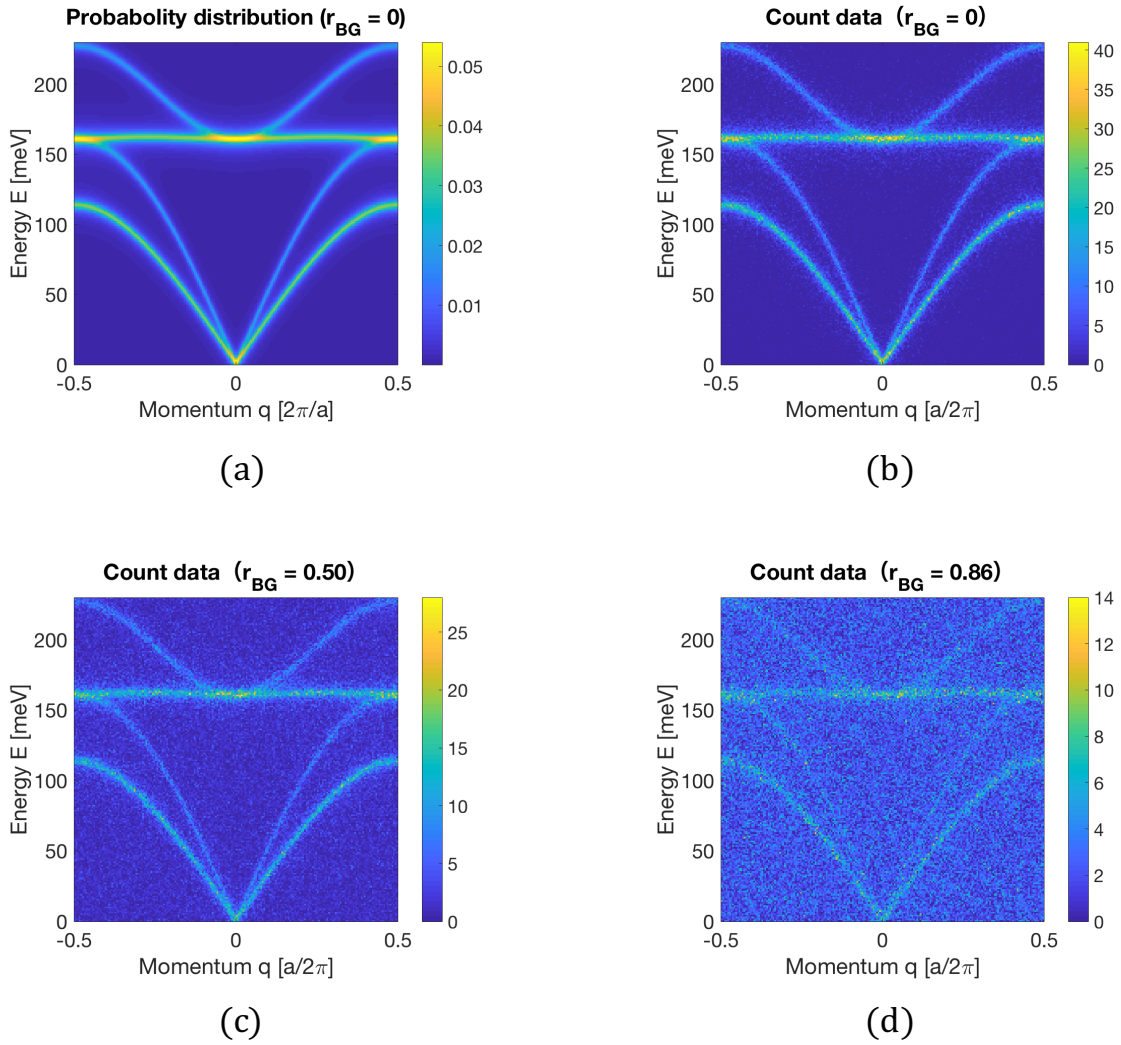


Figure 2.7: (Color online) The dispersion relation and the data. (a) represents the probability distribution $P(E, q)$ for $r_{BG} = 0$. (b), (c), and (d) "raw" count data of $r_{BG} = 0$, $r_{BG} = 0.50$, $r_{BG} = 0.86$, respectively. We set the total number of pieces of event data, n, $100,000$, and the values of the minimum units of the bin width in the energy direction and momentum directions were $\Delta_{min,E} = 1.15$ and $\Delta_{min,q} = 0.005$, respectively.

arbitrarily set the slice widths. Thus, the number of counts in 2D data changes, and the change greatly affects the optimal bin widths. For example, in the histogram for the 2D slice data in Figs. 2.4(a)–2.4(c), the optimal bin widths of Fig. 2.4(c) are extremely large. The total numbers of pieces of event data in Figs. 2.3(d)–2.3(f) are 18380, 18214, and 869, respectively. The large estimated optimal bin widths are large owing to the extremely small number of pieces of event data. Empirical choices of slice and bin widths may sometimes cause such a serious problem.

Third, we discuss the robustness of the proposed method against local minimums and data. According to the results shown in Figs. 2.5(d)–2.5(f) and 2.6, the cost function has several local minimums, although the landscape of the cost function is globally smooth. In this paper, we conducted a grid search. Since experimental data have probabilistic fluctuation and noise, we conducted numerical experiments to investigate the robustness of the proposed method against data. In particular, we focused on the number of pieces of event data and the magnitude of white BG noise. From Figs. 2.8–2.13, we argue that

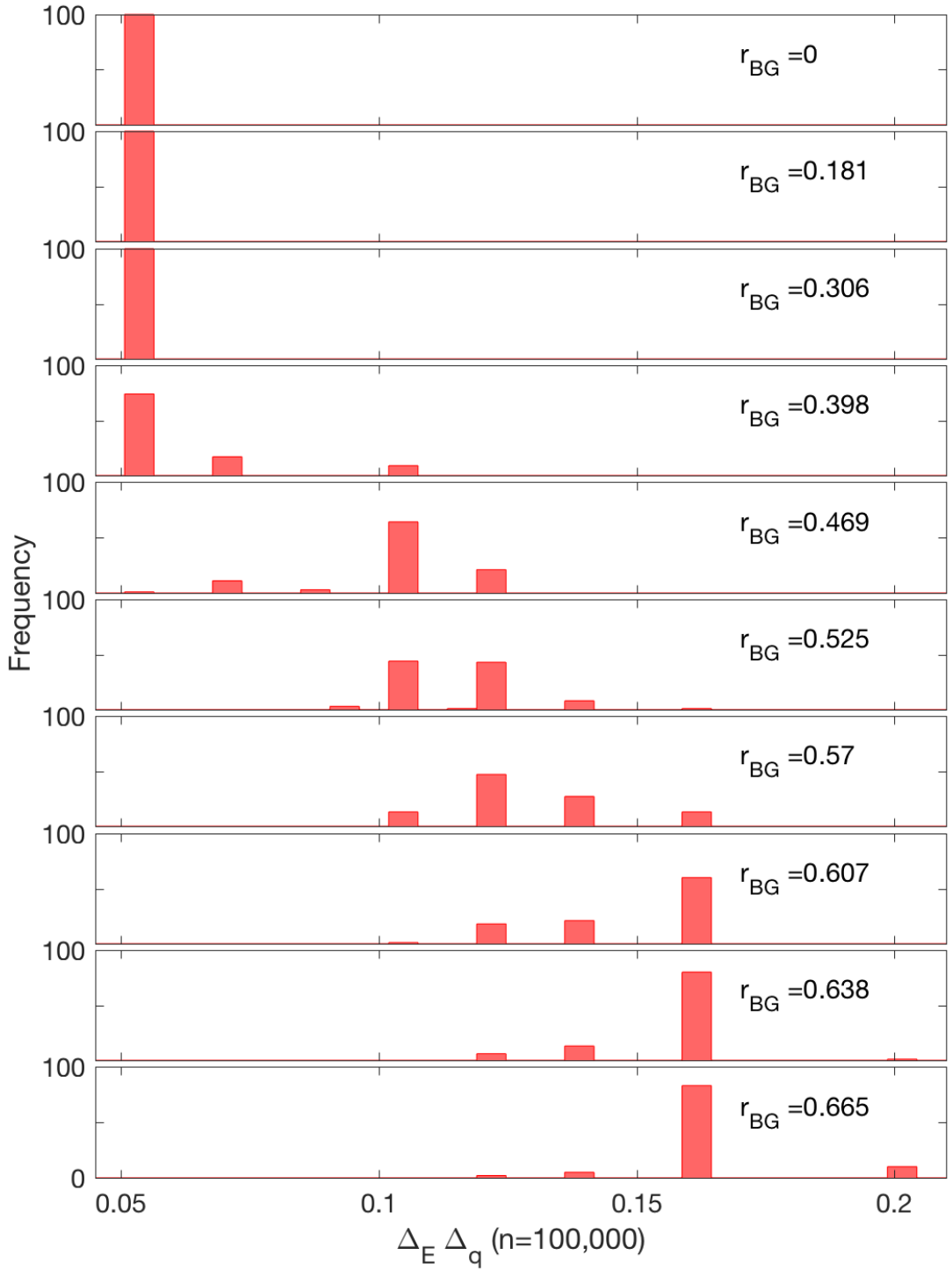


Figure 2.8: (Color online) r_{BG} represents the ratio of white BG noise, defined as Eq. (23). The horizontal label in each histogram shows the product of optimal bin widths $\Delta_E \Delta_q$. The vertical label in each histogram shows the frequency of obtaining $\Delta_E \Delta_q$, and the bin width of each histogram is $\Delta_{min,E} \Delta_{min,q}$. In this experiment, we set the total number of pieces of event data as $n = 100,000$. We set the value of the minimum units of the bin width in the energy direction as $\Delta_{min,E} = 1.15$ and that in the momentum direction as $\Delta_{min,q} = 0.005$.

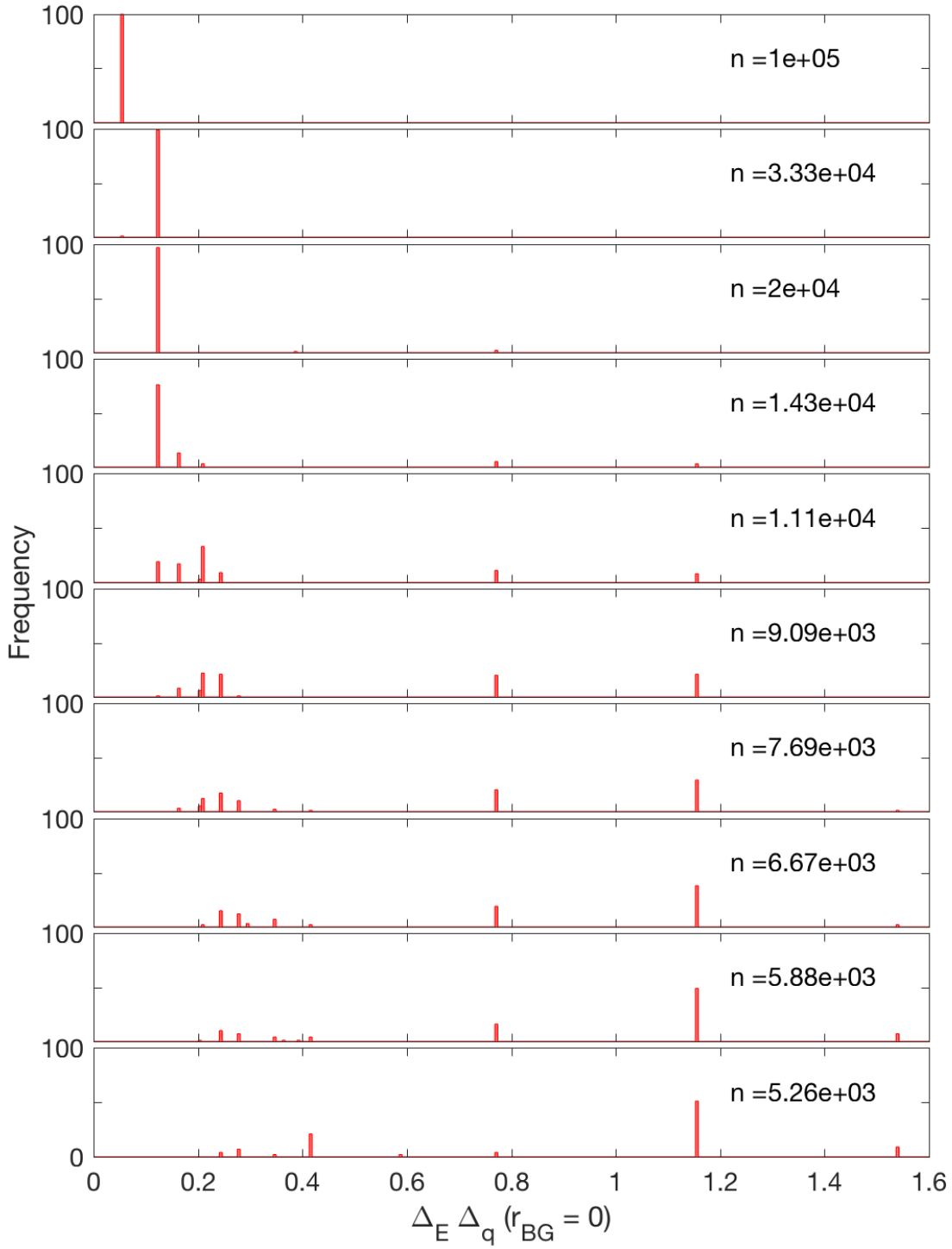


Figure 2.9: (Color online) n represents the total number of pieces of event data. The horizontal label in each histogram shows the product of optimal bin widths $\Delta_E\Delta_q$. The vertical label in each histogram shows the frequency of obtaining $\Delta_E\Delta_q$, and the bin width of each histogram is $\Delta_{min,E}\Delta_{min,q}$. In this experiment, we set $r_{BG} = 0$. We set the value of the minimum units of the bin width in the energy direction as $\Delta_{min,E} = 1.15$ and that in the momentum direction as $\Delta_{min,q} = 0.005$.

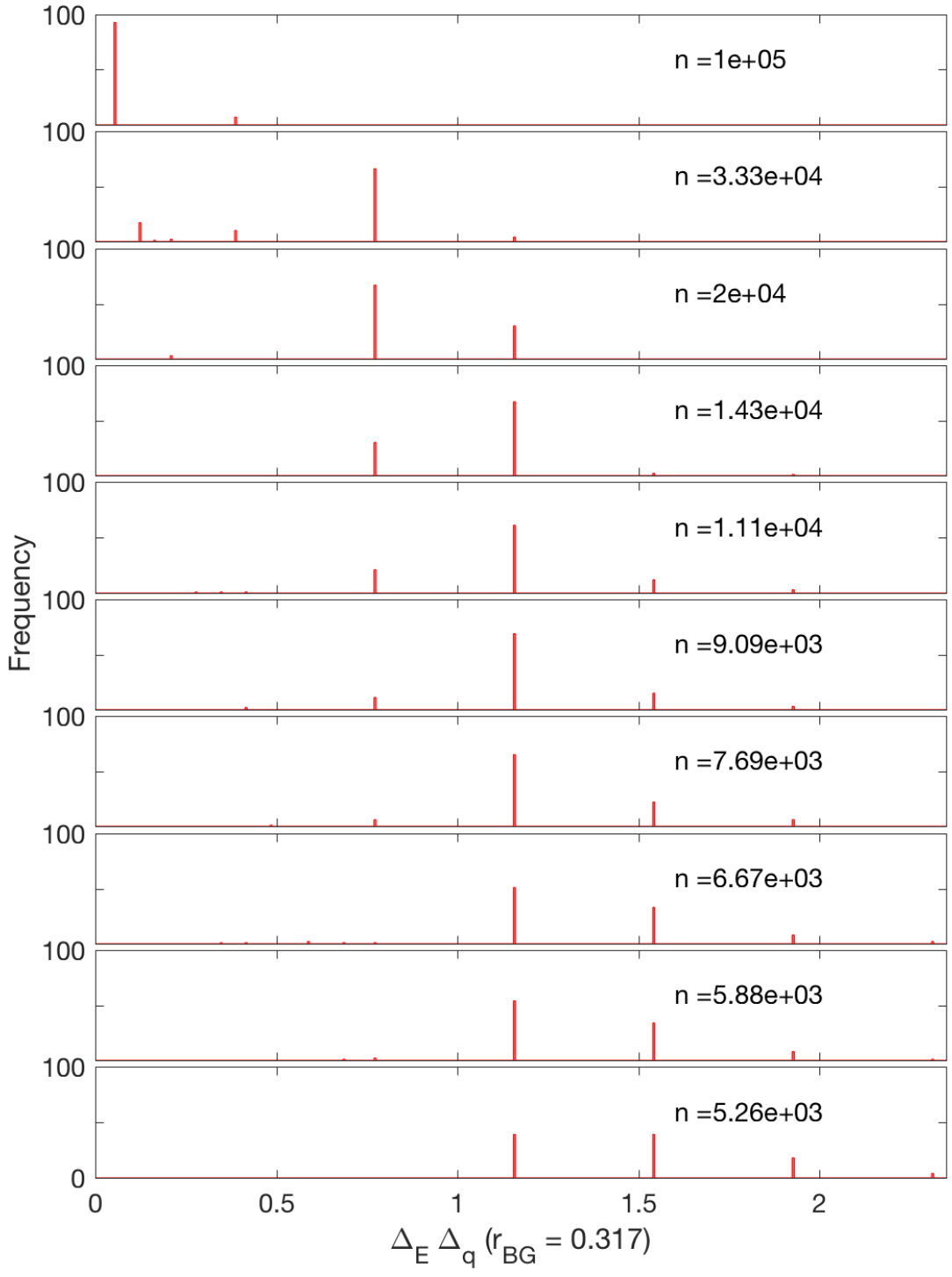


Figure 2.10: (Color online) n represents the total number of pieces of event data. The horizontal label in each histogram shows the product of optimal bin widths $\Delta_E\Delta_q$. The vertical label in each histogram shows the frequency of obtaining $\Delta_E\Delta_q$, and the bin width of each histogram is $\Delta_{min,E}\Delta_{min,q}$. In this experiment, we set $r_{BG} = 0.317$. We set the value of the minimum units of the bin width in the energy direction as $\Delta_{min,E} = 1.15$ and that in the momentum direction as $\Delta_{min,q} = 0.005$.

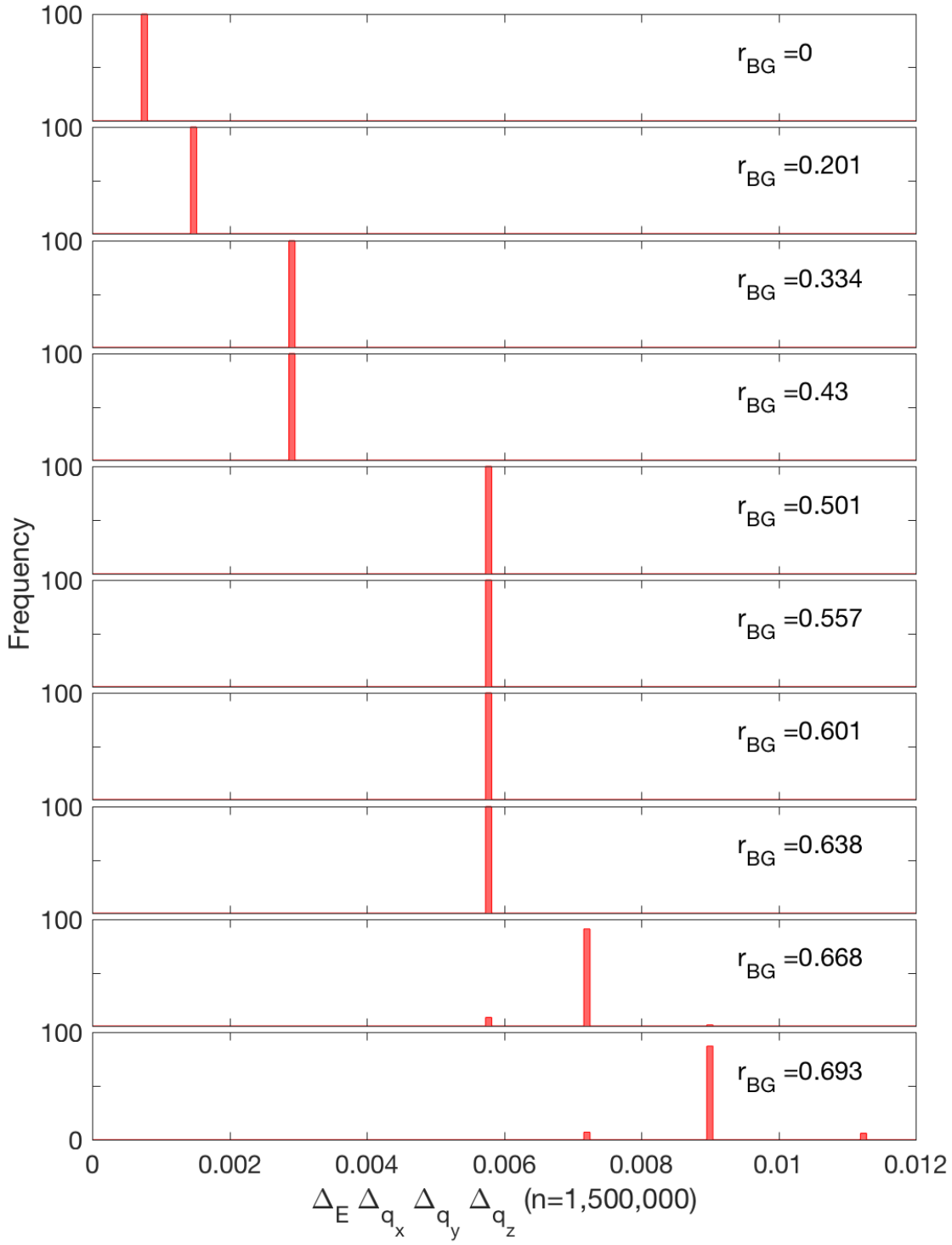


Figure 2.11: (Color online) r_{BG} represents the ratio of white BG noise, defined as Eq. (24). The horizontal label in each histogram shows the product of optimal bin widths $\Delta_E\Delta_{q_x}\Delta_{q_y}\Delta_{q_z}$. The vertical label in each histogram shows the frequency of obtaining $\Delta_E\Delta_{q_x}\Delta_{q_y}\Delta_{q_z}$, and the bin width of each histogram is $\Delta_{min,E}\Delta_{min,q_x}\Delta_{min,q_y}\Delta_{min,q_z}$. In this experiment, we set the total number of pieces of event data as $n = 1,500,000$. We set the value of the minimum units of the bin width in the energy direction as $\Delta_{min,E} = 5.75$ and those in the momentum directions as $\Delta_{min,q_x} = \Delta_{min,q_y} = \Delta_{min,q_z} = 0.025$.

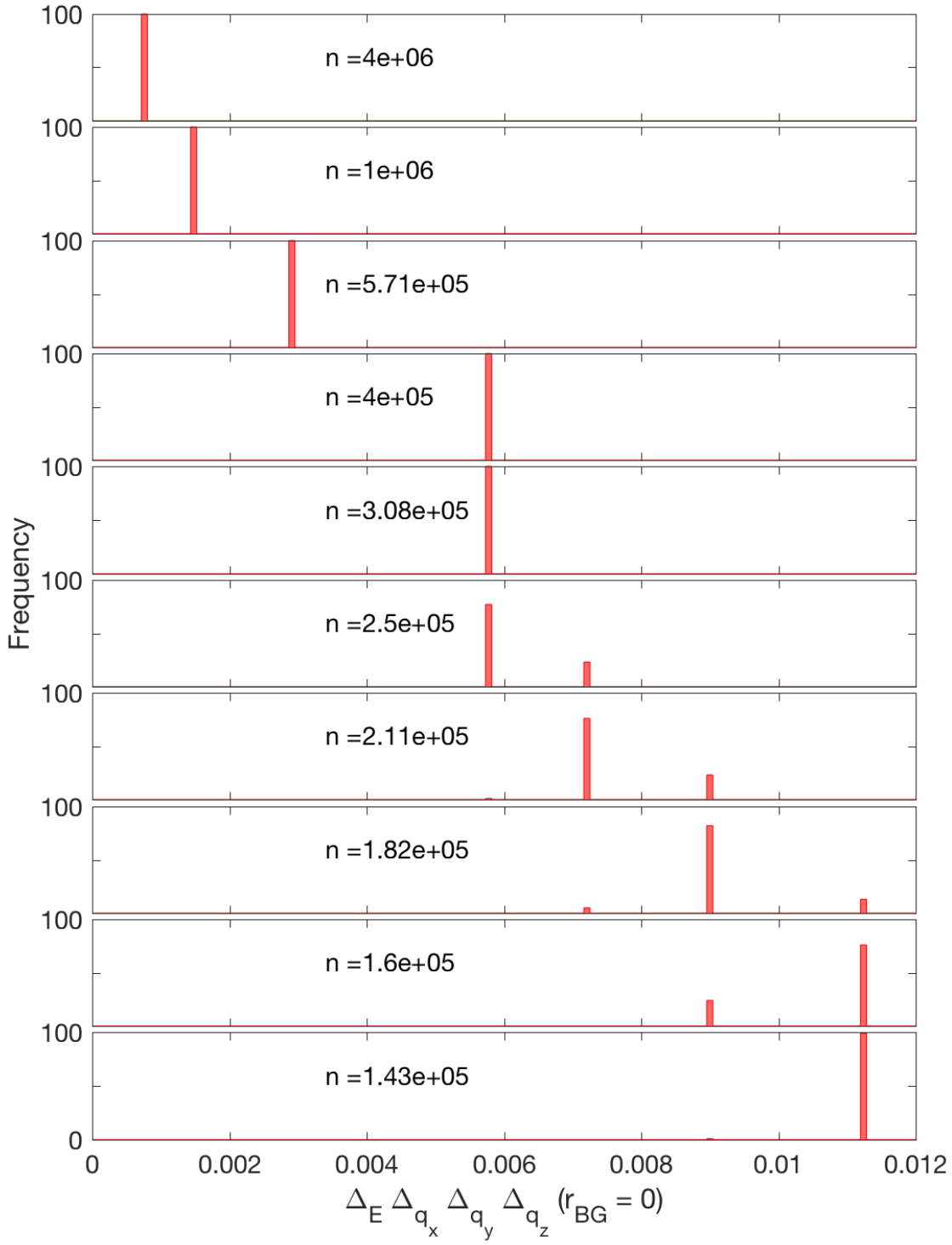


Figure 2.12: (Color online) n represents the total number of pieces of event data. The horizontal label in each histogram shows the product of optimal bin widths $\Delta_E \Delta_{q_x} \Delta_{q_y} \Delta_{q_z}$. The vertical label in each histogram shows the frequency of obtaining $\Delta_E \Delta_{q_x} \Delta_{q_y} \Delta_{q_z}$, and tne bin width of each histogram is $\Delta_{min,E} \Delta_{min,q_x} \Delta_{min,q_y} \Delta_{min,q_z}$. In this experiment, we set $r_{BG} = 0$. We set the value of the minimum units of the bin width in the energy direction as $\Delta_{min,E} = 5.75$ and those in the momentum directions as $\Delta_{min,q_x} = \Delta_{min,q_y} = \Delta_{min,q_z} = 0.025$.

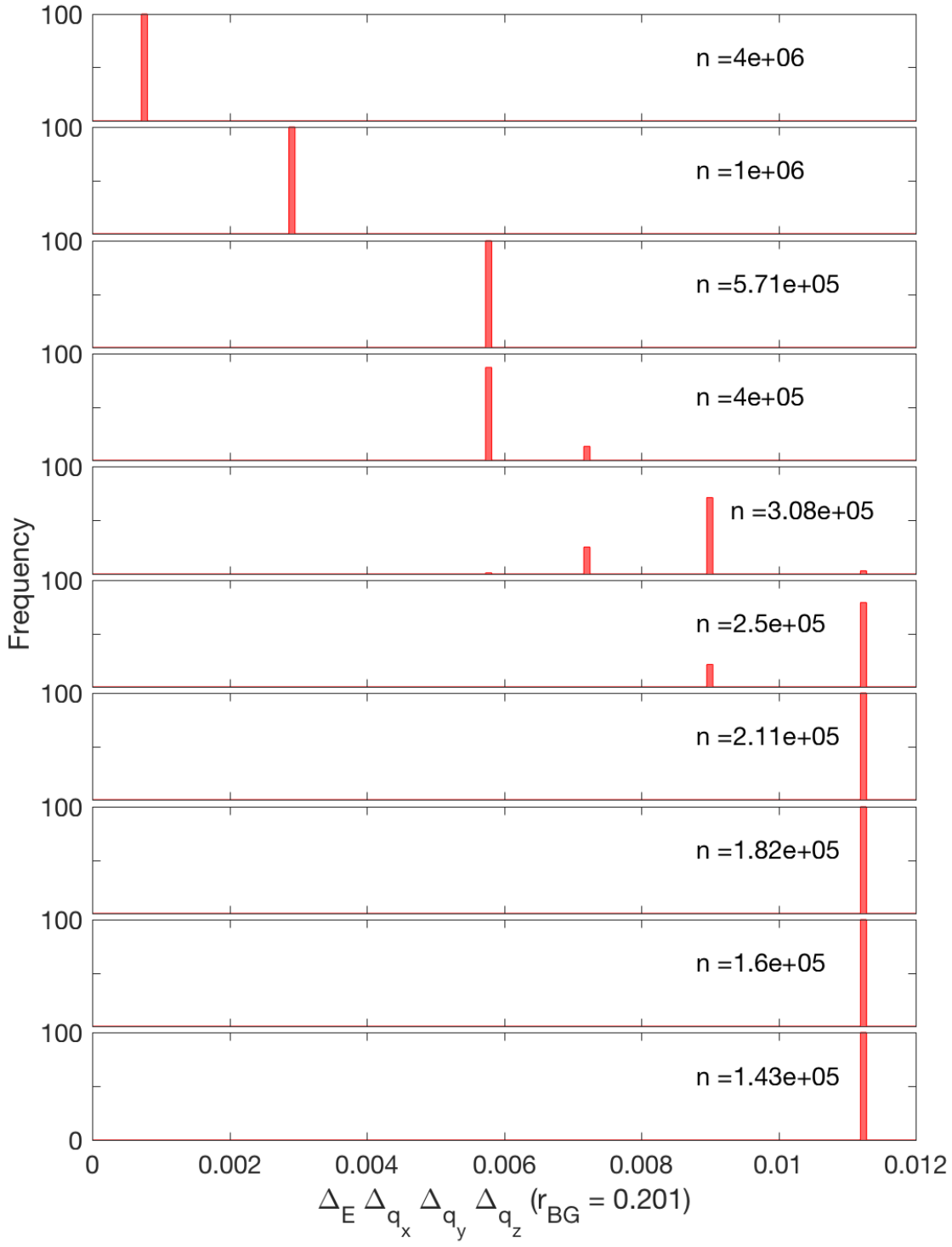


Figure 2.13: (Color online) n represents the total number of pieces of event data. The horizontal label in each histogram shows the product of optimal bin widths $\Delta_E\Delta_{q_x}\Delta_{q_y}\Delta_{q_z}$. The vertical label in each histogram shows the frequency of obtaining $\Delta_E\Delta_{q_x}\Delta_{q_y}\Delta_{q_z}$, and the bin width of each histogram is $\Delta_{min,E}\Delta_{min,q_x}\Delta_{min,q_y}\Delta_{min,q_z}$. In this experiment, we set $r_{BG} = 0.201$. We set the value of the minimum units of the bin width in the energy direction as $\Delta_{min,E} = 5.75$ and those in the momentum directions as $\Delta_{min,q_x} = \Delta_{min,q_y} = \Delta_{min,q_z} = 0.025$.

the proposed method is robust in the case of a large amount of event data and a small magnitude of white BG noise.

2.5 Conclusion

We have proposed a method for optimizing bin widths for multidimensional event data on the basis of a histogram bin-width optimization method.[21] Since the multidimensionalization of the bin-width optimization algorithm increases the number of parameters to be optimized, the computational cost increases. We have also proposed a method of reducing the computational cost.

We generated event data from a dispersion relation by using Poisson sampling. First, to verify the validity of our method, we applied it to the 2D event data sliced from 4D event data. Second, we applied it to the whole 4D event data and compared the results of 2D bin-width optimization and 4D bin-width optimization. As a result, the optimal bin widths of the whole high-dimensional data and those of the data sliced from the high-dimensional data did not agree with each other. The result implies that the bin-width optimization should be performed for all of the high-dimensional data. In addition, since it is difficult for researchers to visually process 4D data, the proposed method is effective. Third, the optimal bin widths increase as the number of pieces of data increases or the magnitude of the white BG noise decreases. By using our method, we can eliminate arbitrariness in preprocessing and select the optimal bin width for multidimensional event data.

Chapter 3

Automatic Termination Strategy of Inelastic Neutron-scattering Measurement Using Bin-width Optimization

第 3 章は雑誌掲載が予定される内容を含むため，インターネット公表できません.

Chapter 4

Efficient Bayesian Bin-width Optimization for On-going Event Measurement

第４章は雑誌掲載が予定される内容を含むため，インターネット公表できません.

Chapter 5

Conclusion

In recent years, a large amount of four-dimensional event data have been obtainable in neutron inelastic scattering experiments conducted by chopper spectrometers. Regarding preprocessing, researchers empirically select bin widths and make histograms from obtained event data. The arbitrariness of the process and human cost are significant issues. It is also an essential task to establish an automatic termination strategy of inelastic neutron-scattering measurement to prevent redundancy of the measurement. There is no criterion to assess whether the obtained data is sufficient in event number. By using large-scale data, we can obtain the fine features of the measurement target. However, it is not necessary to extract information beyond the resolution of the measurement equipment. In this thesis, we proposed methods to resolve these problems.

First, we proposed a method for optimizing bin widths for multidimensional event data on the basis of a histogram bin-width optimization method.[21] Since the multidimensionalization of the bin-width optimization algorithm increases the number of parameters to be optimized, the computational cost increases. We have also proposed a method of reducing the computational cost. As a result of numerical experiments, the optimal bin widths of the whole high-dimensional data and those of the data sliced from the high-dimensional data did not agree with each other. The result implies that the bin-width optimization should be performed for all of the high-dimensional data. In addition, since it is difficult for researchers to visually process 4D data, the proposed method is effective. The optimal bin widths increase as the number of pieces of data increases or the magnitude of the white BG noise decreases. By using our method, we can eliminate arbitrariness in preprocessing and select the optimal bin width for multidimensional event data.

Second, we proposed a strategy to determine whether to terminate the measurement. In the proposed method, researchers compute approximate optimal bin widths as the stopping criteria in real time. When the optimal bin widths become smaller than the target resolutions, the experiment can be terminated. In this study, we computed the cost function for real data of an inelastic neutron-scattering experiment for the first time. The optimal bin widths decrease as the number of data pieces increases. Moreover, we showed that BO is effective for searching for the optimal bin widths, especially when the number of data pieces is large. The cost function can be computed in parallel, and the computational resources can be saved by using BO. The estimated computational cost is not too significant to perform the stop-continue decision during the measurement.

Third, we proposed a strategy to determine efficiently whether to terminate the measurement. As a criterion for determining the termination, we focused on the optimal bin widths of a histogram. For efficient termination judgment, it is necessary to improve the efficiency of bin widths optimization. We proposed a method using the prior distribution of

BO computed from the information of the cost function obtained in the past. As a result of numerical experiments, it was found that the proposed method greatly improves the search efficiency of the optimal bin widths. It was also found that the proposed method is robust for HP.

The proposed methods are applicable not only inelastic neutron-scattering experiments for all general multidimensional event data. In this thesis, we focus on histogram, however, there are several methods that can achieve higher accuracy than histograms. For instance, kernel density estimation (KDE)[66, 67, 56] is suitable for fitting smooth underlying rate. KDE has a problem of computationally complexity, but, GPU acceleration for KDE is being studied.[68, 69, 70] There is a study for applying KDE to 2D small-angle scattering experiment data.[71]

Chapter 6

Future view

6.0.1 Formulation of kernel density estimation

In this chapter, we introduce kernel density estimation (KDE) as a future view. KDE is more computationally expensive than histograms, but high estimation accuracy is expected. We consider the case where N events $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ are independently generated from the true distribution $q(\mathbf{x})$. We attempt to estimate the true distribution $q(\mathbf{x})$ by using the observed data $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$. In KDE, kernel functions are superimposed on data points. We treat Gaussian kernel with covariance matrix Σ.

Let kernel function, density estimator, and likelihood as $k(\mathbf{x}|\boldsymbol{\mu}$, $\hat{p}_{-i}(\mathbf{x}|\Sigma)$, and $l_{\mathrm{LOO}}(\Sigma|\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$ as follows

$$k(\mathbf{x}|\mathbf{x}_i, \Sigma) \qquad \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_i)^T\Sigma^{-1}(\mathbf{x}-\mathbf{x}_i)\right), \quad (6.1)$$

$$\hat{p}_{-i}(\mathbf{x}|\Sigma) \quad := \quad \frac{1}{N-1}\sum_{j\neq i}^{N} k(\mathbf{x}|\mathbf{x}_j, \Sigma), \quad (6.2)$$

$$l_{\mathrm{LOO}}(\Sigma|\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \quad := \quad \prod_{i=1}^{N} \hat{p}_{-i}(\mathbf{x}_i|\Sigma). \quad (6.3)$$

The problem of density estimation comes down to the maximization of the log likelihood $L(\Sigma)$ as follows

$$L(\Sigma) = \log l_{\mathrm{LOO}}(\Sigma|\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \quad (6.4)$$

6.0.2 Sequential update algorithm for the covariance matrix

As the size of the covariance matrix increases, the computational cost increases, and grid search should become difficult. In this section, we consider the selection of a covariance matrix using a sequential update algorithm. [72] We derive a stationary condition by partial

differentiation of the log likelihood with the covariance matrix.

$$\begin{aligned}\frac{\partial L(\Sigma)}{\partial \Sigma^{-1}} &= \sum_{i=i}^{N} \frac{1}{\hat{p}_{-i}(\mathbf{x}_i|\Sigma)} \frac{1}{N-1} \sum_{j\neq i}^{N} \left[k(\mathbf{x}_j|\mathbf{x}_i,\Sigma) \frac{\partial}{\partial \Sigma^{-1}} \log k(\mathbf{x}_j|\mathbf{x}_i,\Sigma) \right] \\ &= \sum_{i=i}^{N} \frac{1}{\hat{p}_{-i}(\mathbf{x}_i|\Sigma)} \frac{1}{N-1} \\ &\quad \times \sum_{j\neq i}^{N} k(\mathbf{x}_j|\mathbf{x}_i,\Sigma) \frac{1}{2} \left[\Sigma - (\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T \right] \end{aligned} \tag{6.5}$$

From Eq. (6.5), assuming $\frac{\partial L(\Sigma)}{\partial \Sigma^{-1}} = 0$, we obtain a self-consistent equation as follows

$$\Sigma = \frac{1}{N(N-1)} \sum_{i=1}^{N} \frac{1}{\hat{p}_{-i}(\mathbf{x}_i|\Sigma)} \sum_{j\neq i}^{N} k(\mathbf{x}_j|\mathbf{x}_i,\Sigma)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T. \tag{6.6}$$

6.0.3 Nearest neighbor search

Each update with the Eq. (6.6) takes computational cost of $O(N^2)$. Therefore, we consider an approximation for the summation for the index j, using only the neighboring data points. Specifically, regarding a real number $r > 0$, we use data points that satisfy the following inequality.

$$(\mathbf{x}_j - \mathbf{x}_i)^T \Sigma^{-1} (\mathbf{x}_j - \mathbf{x}_i) \leq r \tag{6.7}$$

6.1 Numerical experiments

We applied KDE to real data of a neutron scattering experiment. In this experiment, we ignore information about q_y and q_z direction and focus on E–q_x space. The number of total events which are downsampled from the original data is 2768. The data and the density estimation result are shown in Fig. 6.1. Figure. 6.2 shows the results for the exact computation of Eq. (6.6) and the approximation with $r = 15, 20$

The results of numerical experiments show that KDE achieves a smooth density estimation. It is also shown that neighbor search is effective for reducing computational cost. In addition to neighbor search, the computational efficiency is improved by using GPU. In the future, GPU accelerated KDE is expected to play an important role in neutron scattering data.

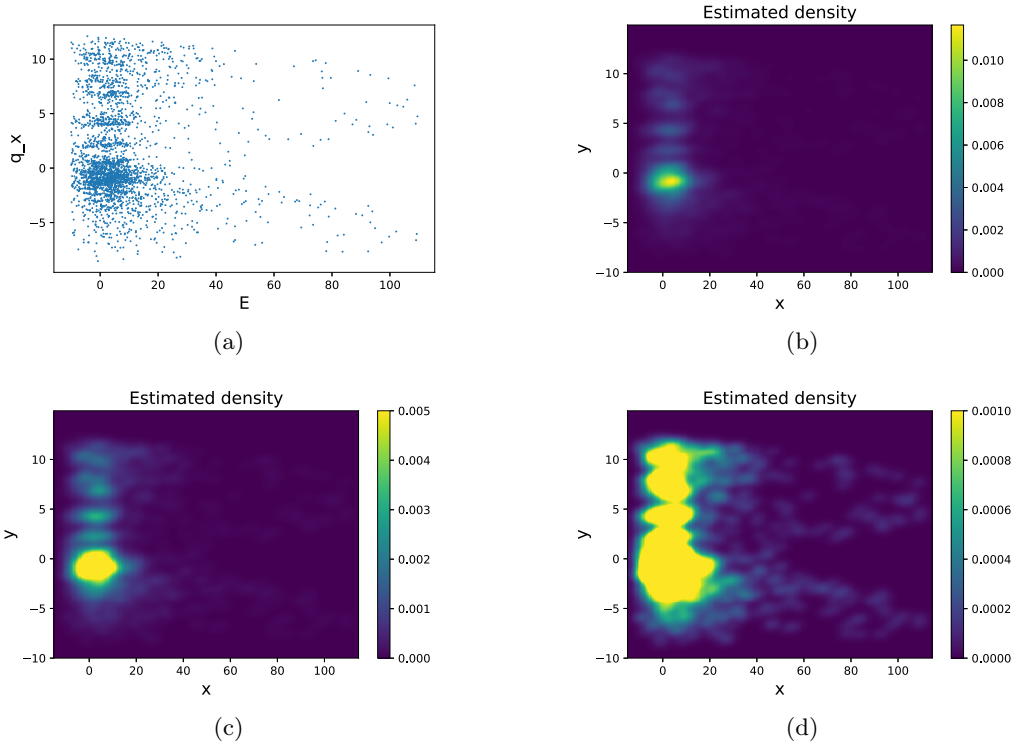
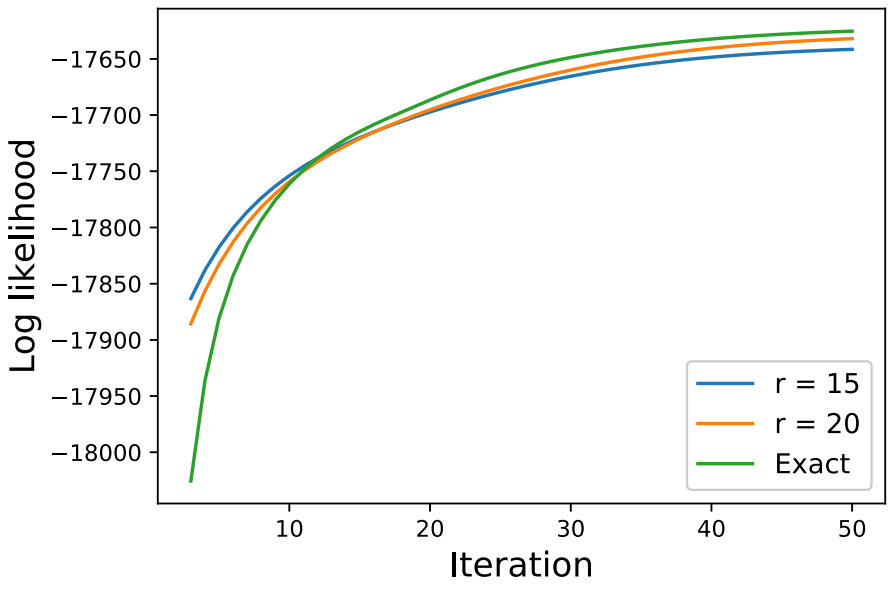


Figure 6.1: (a) is a 2D scatter plot of the inelastic neutron-scattering data. The number of events is 2768. (b)–(d) are the results of kernel density estimation for (a). We change the upper bound of the color bar for (b)–(d).



(a)

Figure 6.2: Log likelihood computed by updating the covariance matrix for each iteration. "Exact" is computed without approximation. $r = 15$ and $r = 20$ represent the result of approximation using Eq. 6.7. The covariance matrices at 50 iterations for "Exact", $r = 20$, and $r = 15$ are $= [[6.5, 0.13], [0.13, 0.18]]$, $[[6.1, 0.096], [0.096, 0.22]]$ and $[[4.8, 0.10], [0.10, 0.24]]$, respectively. The computational cost for $r = 15$ and $r = 20$ are about 30% and 36% of "Exact".

Appendix A

Appendix

A.1 Derivation of the probability density $P(E, q_x, q_y, q_z)$ in Chapter 2

For a triplet of integers (n, m, l), the position of the atom of mass M_1 is $\mathbf{R}_{n,m,l} = na\mathbf{e}_x + ma\mathbf{e}_y + la\mathbf{e}_z$. Here, $\mathbf{e}_j$ is the unit vector in the j-axis direction. Define vectors $\mathbf{h}_j, \mathbf{n}_j, \mathbf{p}_j$, which represent the relative positions of neighboring atoms, as

$$\begin{aligned}
\mathbf{h}_1 &= \frac{a}{2}(\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z),\ \mathbf{h}_2 = \frac{a}{2}(-\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z), \\
\mathbf{h}_3 &= \frac{a}{2}(-\mathbf{e}_x - \mathbf{e}_y + \mathbf{e}_z),\ \mathbf{h}_4 = \frac{a}{2}(\mathbf{e}_x - \mathbf{e}_y + \mathbf{e}_z), \\
\mathbf{h}_j &= \mathbf{h}_{j-4} - a\mathbf{e}_z, (5 \le j \le 8), \\
\mathbf{n}_1 &= a\mathbf{e}_x,\ \mathbf{n}_2 = a\mathbf{e}_y,\ \mathbf{n}_3 = a\mathbf{e}_z, \\
\mathbf{n}_j &= -\mathbf{n}_{j-3}, (4 \le j \le 6), \\
\mathbf{p}_1 &= a(\mathbf{e}_x + \mathbf{e}_z),\ \mathbf{p}_2 = a(\mathbf{e}_y + \mathbf{e}_z),\ \mathbf{p}_3 = a(-\mathbf{e}_x + \mathbf{e}_z), \\
\mathbf{p}_4 &= a(-\mathbf{e}_y + \mathbf{e}_z),\ \mathbf{p}_5 = a(\mathbf{e}_x + \mathbf{e}_y),\ \mathbf{p}_6 = a(-\mathbf{e}_x + \mathbf{e}_y), \\
\mathbf{p}_7 &= -\mathbf{p}_5,\ \mathbf{p}_8 = -\mathbf{p}_6,\ \mathbf{p}_9 = -\mathbf{p}_3, \\
\mathbf{p}_{10} &= -\mathbf{p}_4,\ \mathbf{p}_{11} = -\mathbf{p}_1, \text{ and } \quad \mathbf{p}_{12} = -\mathbf{p}_2.
\end{aligned} \tag{A.1}$$

In Eq. (A.1), let $\hat{\mathbf{h}}_j$, $\hat{\mathbf{p}}_j$, and $\hat{\mathbf{n}}_j$ be the unit vectors in the $\mathbf{h}_j$, $\mathbf{p}_j$, and $\mathbf{n}_j$ directions, respectively. With respect to the atoms of masses M_1 and M_2, Eq. (A.2) is obtained using the equation of motion. Here, $\mathbf{u}_1$ and $\mathbf{u}_2$ are the displacement vectors of atoms whose

masses are M_1 and M_2, respectively.

$$\begin{aligned}
M_1\ddot{\mathbf{u}}_1(\mathbf{r}_1,t) &= \alpha_1 \sum_{j=1}^{8} \left[(\mathbf{u}_2(\mathbf{r}_1+\mathbf{h}_j,t) - \mathbf{u}_1(\mathbf{r}_1,t)) \cdot \hat{\mathbf{h}}_j \right] \hat{\mathbf{h}}_j \\
&+ \alpha_2 \sum_{j=1}^{6} \left[(\mathbf{u}_1(\mathbf{r}_1+\mathbf{n}_j,t) - \mathbf{u}_1(\mathbf{r}_1,t)) \cdot \hat{\mathbf{n}}_j \right] \hat{\mathbf{n}}_j \\
&+ \alpha_3 \sum_{j=1}^{12} \left[(\mathbf{u}_1(\mathbf{r}_1+\mathbf{p}_j,t) - \mathbf{u}_1(\mathbf{r}_1,t)) \cdot \hat{\mathbf{p}}_j \right] \hat{\mathbf{p}}_j \\
(\mathbf{r}_1 &= \mathbf{R}_{n,m,l}) \\
M_2\ddot{\mathbf{u}}_2(\mathbf{r}_2,t) &= \alpha_1 \sum_{j=1}^{8} \left[(\mathbf{u}_1(\mathbf{r}_2+\mathbf{h}_j,t) - \mathbf{u}_2(\mathbf{r}_2,t)) \cdot \hat{\mathbf{h}}_j \right] \hat{\mathbf{h}}_j \\
&+ \alpha_2 \sum_{j=1}^{6} \left[(\mathbf{u}_2(\mathbf{r}_2+\mathbf{n}_j,t) - \mathbf{u}_2(\mathbf{r}_2,t)) \cdot \hat{\mathbf{n}}_j \right] \hat{\mathbf{n}}_j \\
&+ \alpha_3 \sum_{j=1}^{12} \left[(\mathbf{u}_2(\mathbf{r}_2+\mathbf{p}_j,t) - \mathbf{u}_2(\mathbf{r}_2,t)) \cdot \hat{\mathbf{p}}_j \right] \hat{\mathbf{p}}_j \\
(\mathbf{r}_2 &= \mathbf{R}_{n,m,l} + \mathbf{h}_k, (1 \le k \le 8))
\end{aligned} \tag{A.2}$$

Here, we simplify Eq. (A.2) to Eq. (A.3).

$$\begin{aligned}
M_1\ddot{\mathbf{u}}_1(\mathbf{r}_1,t) &= \alpha_1 \sum_{j=1}^{8} \left[(\Delta\mathbf{u}(2,1)(\mathbf{h}_j))) \cdot \hat{\mathbf{h}}_j \right] \hat{\mathbf{h}}_j \\
&+ \alpha_2 \sum_{j=1}^{6} \left[(\Delta\mathbf{u}(1,1)(\mathbf{n}_j)) \cdot \hat{\mathbf{n}}_j \right] \hat{\mathbf{n}}_j \\
&+ \alpha_3 \sum_{j=1}^{12} \left[(\Delta\mathbf{u}(1,1)(\mathbf{p}_j)) \cdot \hat{\mathbf{p}}_j \right] \hat{\mathbf{p}}_j \\
(\mathbf{r}_1 &= \mathbf{R}_{n,m,l}) \\
M_2\ddot{\mathbf{u}}_2(\mathbf{r}_2,t) &= \alpha_1 \sum_{j=1}^{8} \left[\left(\Delta\mathbf{u}_(1,2)(\mathbf{h}_j)\right) \cdot \hat{\mathbf{h}}_j \right] \hat{\mathbf{h}}_j \\
&+ \alpha_2 \sum_{j=1}^{6} \left[(\Delta\mathbf{u}(2,2)(\mathbf{n}_j)) \cdot \hat{\mathbf{n}}_j \right] \hat{\mathbf{n}}_j \\
&+ \alpha_3 \sum_{j=1}^{12} \left[(\Delta\mathbf{u}(2,2)(\mathbf{n}_j)) \cdot \hat{\mathbf{p}}_j \right] \hat{\mathbf{p}}_j \\
(\mathbf{r}_2 &= \mathbf{R}_{n,m,l} + \mathbf{h}_k, (1 \le k \le 8))
\end{aligned} \tag{A.3}$$

We assume the solution of Eq. (A.3) as Eq. (A.4).

$$\begin{pmatrix} \mathbf{u}_1(\mathbf{r}_1,t) \\ \mathbf{u}_2(\mathbf{r}_1+\mathbf{h}_k,t) \end{pmatrix} = \begin{pmatrix} u_{1x}(\mathbf{q}) \\ u_{1y}(\mathbf{q}) \\ u_{1z}(\mathbf{q}) \\ u_{2x}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_k} \\ u_{2y}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_k} \\ u_{2z}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_k} \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_1-\omega t)}$$

$$(1 \le k \le 8) \tag{A.4}$$

Then, $\Delta\mathbf{u}(2,1)(\mathbf{h}_j)$, $\Delta\mathbf{u}(1,1)(\mathbf{n}_j)$, $\Delta\mathbf{u}(1,1)(\mathbf{p}_j)$, $\Delta\mathbf{u}(1,2)(\mathbf{h}_j)$, $\Delta\mathbf{u}(2,2)(\mathbf{n}_j)$, $\Delta\mathbf{p}(2,2)(\mathbf{p}_j)$ in Eq. (A.3) can be denoted as Eq. (A.5).

$$
\begin{aligned}
\Delta\mathbf{u}(2,1)(\mathbf{h}_j) &= \begin{pmatrix} u_{2x}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_j} - u_{1x}(\mathbf{q}) \\ u_{2y}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_j} - u_{1y}(\mathbf{q}) \\ u_{2z}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_j} - u_{1z}(\mathbf{q}) \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_1-\omega t)} \\
\Delta\mathbf{u}(1,1)(\mathbf{n}_j) &= \begin{pmatrix} \left(e^{i\mathbf{q}\cdot\mathbf{n}_j} - 1\right) u_{1x}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{n}_j} - 1\right) u_{1y}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{n}_j} - 1\right) u_{1z}(\mathbf{q}) \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_1-\omega t)} \\
\Delta\mathbf{u}(1,1)(\mathbf{p}_j) &= \begin{pmatrix} \left(e^{i\mathbf{q}\cdot\mathbf{p}_j} - 1\right) u_{1x}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{p}_j} - 1\right) u_{1y}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{p}_j} - 1\right) u_{1z}(\mathbf{q}) \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_1-\omega t)} \\
\Delta\mathbf{u}(1,2)(\mathbf{h}_j) &= \begin{pmatrix} u_{1x}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_j} - u_{2x}(\mathbf{q}) \\ u_{1y}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_j} - u_{2y}(\mathbf{q}) \\ u_{1z}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_j} - u_{2z}(\mathbf{q}) \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_2-\omega t)} \\
\Delta\mathbf{u}(2,2)(\mathbf{n}_j) &= \begin{pmatrix} \left(e^{i\mathbf{q}\cdot\mathbf{n}_j} - 1\right) u_{2x}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{n}_j} - 1\right) u_{2y}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{n}_j} - 1\right) u_{2z}(\mathbf{q}) \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_2-\omega t)} \\
\Delta\mathbf{u}(2,2)(\mathbf{p}_j) &= \begin{pmatrix} \left(e^{i\mathbf{q}\cdot\mathbf{p}_j} - 1\right) u_{2x}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{p}_j} - 1\right) u_{2y}(\mathbf{q}) \\ \left(e^{i\mathbf{q}\cdot\mathbf{p}_j} - 1\right) u_{2z}(\mathbf{q}) \end{pmatrix} e^{i(\mathbf{q}\cdot\mathbf{r}_2-\omega t)}
\end{aligned} \tag{A.5}
$$

We obtain Eq. (A.6) from Eq. (A.3) and Eq. (A.5).

$$
\begin{pmatrix} \mathbf{u}_1(\mathbf{R}_{n,m,l}, t) \\ \mathbf{u}_2(\mathbf{R}_{n,m,l} + \mathbf{h}_k, t) \end{pmatrix} = \begin{pmatrix} u_{1x}(\mathbf{q}) \\ u_{1y}(\mathbf{q}) \\ u_{1z}(\mathbf{q}) \\ u_{2x}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_k} \\ u_{2y}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_k} \\ u_{2z}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{h}_k} \end{pmatrix} e^{i\left(\mathbf{q}\cdot\mathbf{R}_{n,m,l}-\omega t\right)},
$$

$$
(1 \le k \le 8).
$$

$$
\begin{aligned}
\text{Then, we obtain} \quad -\omega^2 M\mathbf{u} &= D\mathbf{u}, \\
M &= diag(M_1, M_1, M_1, M_2, M_2, M_2), \\
\mathbf{u} &= \begin{pmatrix} u_{1x}(\mathbf{q}) \\ u_{1y}(\mathbf{q}) \\ u_{1z}(\mathbf{q}) \\ u_{2x}(\mathbf{q}) \\ u_{2y}(\mathbf{q}) \\ u_{2z}(\mathbf{q}) \end{pmatrix}.
\end{aligned} \tag{A.6}
$$

Here, $\mathbf{q} = (q_x, q_y, q_z) \in \mathbb{R}^3$, $-\frac{\pi}{a} \le q_x, q_y, q_z \le \frac{\pi}{a}$. Using Eq. (A.6), we find solutions that satisfy $\mathbf{u} \neq 0$, $\omega \ge 0$. When calculating the neutron scattering intensity, it is necessary to consider the neutron scattering lengths of the nucleus of each atom, the displacement vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, and the relative angle of the scattering vector $\mathbf{q}$ to the polarization vector. However, because we focus on optimizing bin widths in this paper, these contributions are not taken into consideration. We simply obtain the intensity function $I(E, q_x, q_y, q_z)$ by superimposing the Lorentzian of width 3[meV] centered on $\hbar\omega(\mathbf{q})$ ($\hbar$ is the converted Planck's constant). By normalizing $I(E, q_x, q_y, q_z)$ in the range of $-\frac{\pi}{a} \le q_x, q_y, q_z \le \frac{\pi}{a}$ and $0 \le E \le E_{max}$[meV], we obtain the probability distribution $P(E, q_x, q_y, q_z)$. We generate event data to follow $P(E, q_x, q_y, q_z)$. In this paper, we set $E_{max} = 230$[meV].

A.2 Matrix D in Eq. (A.6)

D in Eq. (A.6) is represented as Eq. (A.7) by using symmetric matrices D^A and D^B.

$$D = \begin{pmatrix} D^A & D^B \\ D^B & D^A \end{pmatrix} \tag{A.7}$$

The elements of D^A and D^B are as follows.

$$\begin{aligned}
D^A_{1,1} &= -\frac{8}{3}\alpha_1 - 4\alpha_2 \sin^2\left(\frac{q_x a}{2}\right) - 2\alpha_3 \left\{(1-\cos(q_z a)\cos(q_x a)) + (1-\cos(q_x a)\cos(q_y a))\right\} \\
D^A_{2,2} &= -\frac{8}{3}\alpha_1 - 4\alpha_2 \sin^2\left(\frac{q_y a}{2}\right) - 2\alpha_3 \left\{(1-\cos(q_x a)\cos(q_y a)) + (1-\cos(q_y a)\cos(q_z a))\right\} \\
D^A_{3,3} &= -\frac{8}{3}\alpha_1 - 4\alpha_2 \sin^2\left(\frac{q_z a}{2}\right) - 2\alpha_3 \left\{(1-\cos(q_y a)\cos(q_z a)) + (1-\cos(q_z a)\cos(q_x a))\right\} \\
D^A_{1,2} &= -2\alpha_3 \sin(q_x a)\sin(q_y a) \\
D^A_{1,3} &= -2\alpha_3 \sin(q_x a)\sin(q_z a) \\
D^A_{2,3} &= -2\alpha_3 \sin(q_y a)\sin(q_z a) \\
D^B_{1,1} &= \frac{8}{3}\alpha_1 \cos\left(\frac{q_x a}{2}\right)\cos\left(\frac{q_y a}{2}\right)\cos\left(\frac{q_z a}{2}\right) \\
D^B_{2,2} &= D^B_{1,1},\ D^B_{3,3} = D^B_{1,1} \\
D^B_{1,2} &= -\frac{8}{3}\alpha_1 \sin\left(\frac{q_x a}{2}\right)\sin\left(\frac{q_y a}{2}\right)\cos\left(\frac{q_z a}{2}\right) \\
D^B_{1,3} &= -\frac{8}{3}\alpha_1 \sin\left(\frac{q_x a}{2}\right)\cos\left(\frac{q_y a}{2}\right)\sin\left(\frac{q_z a}{2}\right) \\
D^B_{2,3} &= -\frac{8}{3}\alpha_1 \cos\left(\frac{q_x a}{2}\right)\sin\left(\frac{q_y a}{2}\right)\sin\left(\frac{q_z a}{2}\right)
\end{aligned} \tag{A.8}$$

Using the symmetry of D^A and D^B, we obtain all the elements of D.

Bibliography

[1] `https://mlfinfo.jp/en/bl01/`.

[2] 中島健次 and 梶本亮. パルス中性子源におけるチョッパー型分光器 (1). **波紋**, 25(1):39–46, 2015.

[3] Tony Hey, Stewart Tansley, Kristin Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA, 2009.

[4] V. Mayer-Schoenberger and K. Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Eamon Dolan, 2014.

[5] G. Ehlers et al. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 2014.

[6] Tasleem Nizam and Syed Imtiyaz Hassan. Big data: A survey paper on big data innovation and its technology. *International Journal of Advanced Research in Computer Science*, 8(5), 2017.

[7] Matteo Migliorini, Riccardo Castellotti, Luca Canali, and Marco Zanetti. Machine learning pipelines with modern big data tools for high energy physics. *Computing and Software for Big Science*, 4(1):1–12, 2020.

[8] Andrii Shelestov, Mykola Lavreniuk, Nataliia Kussul, Alexei Novikov, and Sergii Skakun. Exploring google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *frontiers in Earth Science*, 5:17, 2017.

[9] Raghu Ramakrishnan, Baskar Sridharan, John R Douceur, Pavan Kasturi, Balaji Krishnamachari-Sampath, Karthick Krishnamoorthy, Peng Li, Mitica Manu, Spiro Michaylov, Rogério Ramos, et al. Azure data lake store: a hyperscale distributed file service for big data analytics. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pages 51–63, 2017.

[10] `https://www.mgi.gov/`.

[11] Ankit Agrawal and Alok Choudhary. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. *Apl Materials*, 4(5):053208, 2016.

[12] Stephen W Lovesey. Theory of neutron scattering from condensed matter. 1984.

[13] TG Perring, AD Taylor, R Osbom, D McK Paul, AT Boothroyd, and G Aeppli. Maps: A chopper spectrometer to measure high energy magnetic excitations in single crystals. *Proc. 12th Meeting of the Collaboration on Advanced Neutron Sources (ICANS-XII)*, pages I–60–I–72, 1994.

[14] RI Bewley, RS Eccleston, KA McEwen, SM Hayden, MT Dove, SM Bennington, JR Treadgold, and RLS Coleman. Merlin, a new high count rate spectrometer at isis. *Physica B*, 385–386:1029–1031, 2006.

[15] Barry Winn, Uwe Filges, V Ovidiu Garlea, Melissa Graves-Brook, Mark Hagen, Chenyang Jiang, Michel Kenzelmann, Larry Passell, Stephen M Shapiro, Xin Tong, et al. Recent progress on hyspec, and its polarization analysis capabilities. *EPJ Web of Conferences*, 83:03017, 2015.

[16] Douglas L Abernathy, Matthew B Stone, MJ Loguillo, MS Lucas, O Delaire, Xiaoli Tang, JYY Lin, and B Fultz. Design and operation of the wide angular-range chopper spectrometer arcs at the spallation neutron source. *Rev. Sci. Instruments*, 83:015114, 2012.

[17] Ryoichi Kajimoto, Mitsutaka Nakamura, Yasuhiro Inamura, Fumio Mizuno, Kenji Nakajima, Seiko Ohira-Kawamura, Tetsuya Yokoo, Takeshi Nakatani, Ryuji Maruyama, Kazuhiko Soyama, et al. The fermi chopper spectrometer 4seasons at j-parc. *J. Phys. Soc. Jpn*, 80(Suppl. B):SB025, 2011.

[18] Kenji Nakajima, Seiko Ohira-Kawamura, Tatsuya Kikuchi, Mitsutaka Nakamura, Ryoichi Kajimoto, Yasuhiro Inamura, Nobuaki Takahashi, Kazuya Aizawa, Kentaro Suzuya, Kaoru Shibata, et al. Amateras: a cold-neutron disk chopper spectrometer. *Journal of the Physical Society of Japan*, 80(Suppl. B):SB028, 2011.

[19] Shinichi Itoh, Tetsuya Yokoo, Setsuo Satoh, Shin-ichiro Yano, Daichi Kawana, Junichi Suzuki, and Taku J Sato. High resolution chopper spectrometer (hrc) at j-parc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 631(1):90–97, 2011.

[20] Yasuhiro Inamura, Takeshi Nakatani, Jiro Suzuki, and Toshiya Otomo. Development status of software “utsusemi” for chopper spectrometers at mlf, j-parc. *J. Phys. Soc. Jpn*, 82(Suppl. A):SA031, 2013.

[21] Hideaki Shimazaki and Shigeru Shinomoto. A method for selecting the bin size of a time histogram. *Neural Computation*, 19(6):1503–1527, 2007.

[22] Kensuke Muto, Hirotaka Sakamoto, Keisuke Matsuura, Taka-hisa Arima, and Masato Okada. Multidimensional bin-width optimization for histogram and its application to four-dimensional neutron inelastic scattering data. *J. Phys. Soc. Jpn*, 88(4):044002, 2019.

[23] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25:2951–2959, 2012.

[24] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

[25] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[26] A.D. Christianson et al. Unconventional superconductivity in ba 0.6 k 0.4 fe 2 as 2 from inelastic neutron scattering. *Nature*, 456:930–932, 2008.

[27] S.X. Chi et al. Inelastic neutron-scattering measurements of a three-dimensional spin resonance in the feas-basedbafe1.9ni0.1as2 superconductor. *Phys. Rev. Lett.*, 102, 2009.

[28] J. Rossat-Mignod et al. Investigation of the spin dynamics in yba2cu3o6+x by inelastic neutron scattering. *Physica B: Condensed Matter*, 169, 1991.

[29] G. Ehlers et al. The new cold neutron chopper spectrometer at the spallation neutron source: design and performance. *Review of Scientific Instruments*, 82, 2011.

[30] 遠藤 康夫. **中性子散乱**. 朝倉書店, 2012.

[31] 橋本 竹治. ***X*線・光・中性子散乱の原理と応用**. 講談社, 2017.

[32] M. Nakamura et al. First demonstration of novel method for inelastic neutron scattering measurement utilizing multiple incident energies. *JPSJ*, 78(9), 2009.

[33] John RD Copley and Terrence J Udovic. Neutron time-of-flight spectroscopy. *Journal of research of the National Institute of Standards and Technology*, 98(1):71, 1993.

[34] Alan K Soper. Inelasticity corrections for time-of-flight and fixed wavelength neutron diffraction experiments. *Molecular Physics*, 107(16):1667–1684, 2009.

[35] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[36] David W Scott and Stephan R Sain. Multidimensional density estimation. *Handbook of statistics*, 24:229–261, 2005.

[37] David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[38] Chris Fraley and Adrian E Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631, 2002.

[39] J Steve Marron and Matt P Wand. Exact mean integrated squared error. *The Annals of Statistics*, pages 712–736, 1992.

[40] Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1997.

[41] Rudolf Beran et al. Minimum hellinger distance estimates for parametric models. *The annals of Statistics*, 5(3):445–463, 1977.

[42] Franklin C Crow. Summed-area tables for texture mapping. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 207–212, 1984.

[43] J Rossat-Mignod, LP Regnault, C Vettier, Ph Bourges, P Burlet, J Bossy, JY Henry, and G Lapertot. Neutron scattering study of the yba2cu3o6+ x system. *Physica C: Superconductivity*, 185:86–92, 1991.

[44] J Rossat-Mignod, LP Regnault, C Vettier, P Burlet, JY Henry, and G Lapertot. Investigation of the spin dynamics in yba2cu3o6+ x by inelastic neutron scattering. *Physica B: Condensed Matter*, 169(1-4):58–65, 1991.

[45] BX Yang, TR Thurston, JM Tranquada, and G Shirane. Magnetic neutron scattering study of single-crystal cupric oxide. *Physical Review B*, 39(7):4343, 1989.

[46] JC Smith. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Quarterly reviews of biophysics*, 24(3):227–291, 1991.

[47] AA Edwards, DC Lloyd, and RJ Purrott. Radiation induced chromosome aberrations and the poisson distribution. *Radiation and Environmental Biophysics*, 16(2):89–100, 1979.

[48] FJ Grunthaner, PJ Grunthaner, and J Maserjian. Radiation-induced defects in sio2 as determined with xps. *IEEE Transactions on Nuclear Science*, 29(6):1462–1466, 1982.

[49] K Kobayashi, M Yabashi, Y Takata, T Tokushima, S Shin, K Tamasaku, D Miwa, T Ishikawa, H Nohira, T Hattori, et al. High resolution-high energy x-ray photoelectron spectroscopy using third-generation synchrotron radiation source, and its application to si-high k insulator systems. *Applied physics letters*, 83(5):1005–1007, 2003.

[50] Teruyuki Nakajima and Michael D King. Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. part i: Theory. *Journal of the atmospheric sciences*, 47(15):1878–1893, 1990.

[51] F. Rieke, D. Warland, and W. Bialek R. R. Steveninck. *Spikes: Exploring the Neural Code*. The MIT Press, 1999.

[52] Il Memming Park, Sohan Seth, Antonio RC Paiva, Lin Li, and Jose C Principe. Kernel methods on spike train space for neuroscience: a tutorial. *IEEE Signal Processing Magazine*, 30(4):149–160, 2013.

[53] Asohan Amarasingham, Ting-Li Chen, Stuart Geman, Matthew T Harrison, and David L Sheinberg. Spike count reliability and the poisson hypothesis. *Journal of Neuroscience*, 26(3):801–809, 2006.

[54] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9(2):65–78, 1982.

[55] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Springer, 1994.

[56] H. Shimazaki and S. Shinomoto. Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience*, 29, 2010.

[57] M. P. Wand. Data-based choice of histogram bin width. *The American Statistician*, 51:59–64, 1997.

[58] David W Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.

[59] Walter Hauser and Herman Feshbach. The inelastic scattering of neutrons. *Physical review*, 87(2):366, 1952.

[60] K. Watanabe, H. Tanaka, K. Miura, and M. Okada. Transfer matrix method for instantaneous spike rate estimation. *IEICE TRANSACTIONS on Information and Systems*, E92-D(7):1362–1368, 2009.

[61] Donald Lee Snyder. *Random point processes*. Wiley, 1975.

[62] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[63] Robert E Kass, Valérie Ventura, and Emery N Brown. Statistical issues in the analysis of neuronal data. *Journal of neurophysiology*, 94(1):8–25, 2005.

[64] D.P. Kroese, T. Taimre, and Z.I. Botev. *Handbook of Monte Carlo Methods*. John Wiley and Sons, 2011.

[65] John Frank Charles Kingman. P oisson processes. *Encyclopedia of biostatistics*, 6, 2005.

[66] Philippe Van Kerm. Adaptive kernel density estimation. *The Stata Journal*, 3(2):148–156, 2003.

[67] Travis A O ' Brien, Karthik Kashinath, Nicholas R Cavanaugh, William D Collins, and John P O ' Brien. A fast and objective multidimensional kernel density estimation method: fastkde. *Computational Statistics & Data Analysis*, 101:148–160, 2016.

[68] Panagiotis D Michailidis and Konstantinos G Margaritis. Accelerating kernel density estimation on the gpu using the cuda framework. *Applied Mathematical Sciences*, 7(30):1447–1476, 2013.

[69] Max Heimel, Martin Kiefer, and Volker Markl. Self-tuning, gpu-accelerated kernel density models for multidimensional selectivity estimation. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1477–1492, 2015.

[70] Guiming Zhang, A-Xing Zhu, and Qunying Huang. A gpu-accelerated adaptive kernel density estimation approach for efficient point pattern analysis on spatial big data. *International Journal of Geographical Information Science*, 31(10):2068–2097, 2017.

[71] Kotaro Saito, Masao Yano, Hideitsu Hino, Tetsuya Shoji, Akinori Asahara, Hidekazu Morita, Chiharu Mitsumata, Joachim Kohlbrecher, and Kanta Ono. Accelerating small-angle scattering experiments on anisotropic samples using kernel density estimation. *Scientific reports*, 9(1):1–10, 2019.

[72] José M Leiva-Murillo and Antonio ArtéS-RodríGuez. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. *Pattern Recognition Letters*, 33(13):1717–1724, 2012.