

---

---

X-RAY SPECTRAL STUDY OF SUPERNOVA REMNANTS  
USING UNSUPERVISED DEEP LEARNING

---

---

HIROYOSHI IWASAKI

*Department of Physics  
Graduate School of Science  
Rikkyo University*

*3-34-1 Nishi Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan*



A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL OF SCIENCE, RIKKYO UNIVERSITY  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

JANUARY, 2020



# Abstract

Recent rapid development of deep learning algorithms, which can implicitly capture structures in high-dimensional data, opens a new chapter in astronomical data analysis. We report here a new implementation of deep learning techniques for X-ray analyses.

We apply variational autoencoder (VAE) using a deep neural network for spatio-spectral analysis of data obtained by *Chandra* X-ray Observatory from three type-Ia supernova remnants (SNRs); *Tycho's* SNR, *Kepler's* SNR, and SN 1006. VAE, which is one of the best-known generative models, is capable of extracting features from complex data. We established an unsupervised learning method combining VAE and Gaussian mixture model (GMM), where the dimensions of the observed spectral data are reduced by VAE, and clustering in the feature space is performed by GMM.

For X-ray data, Poisson statistics is appropriate because each bin in an X-ray spectrum represents a number of photons. As Ichinohe & Yamada (2019) shows for an ideal case, Poisson statistics is important for VAE training with such X-ray spectral dataset. We also newly implemented the data processes at the input and output of VAE in order to apply Poisson reconstruction loss for the training, and applied the model to observational X-ray spectral data of SNRs.

We found that some characteristic spatial structures can be automatically recognized by this method, which uses only spectral properties. In *Tycho's* SNR, this method automatically found the synchrotron dominated filaments, the iron knot on the eastern rim, and the layered structure of the north-western ejecta rim. In *Kepler's* SNR, which is interacting with asymmetric dense circumstellar medium (CSM), the regions where the shocked CSM emitting thermal X-ray spectra were automatically recognized by our method. This method also unveiled that the synchrotron dominated forward shock and the layered structure in the northern part of *Kepler's* SNR. In SN 1006, our method also recognized the synchrotron dominated regions in the north-eastern and south-western of the SNR, and ejecta dominated regions inside the SNR.

We also show the relation between the VAE latent space and the original data space, using the decoder to generate spectra from given latent parameters. For *Kepler's* SNR, the VAE has extracted the latent axes corresponding to the relations of Fe and intermediate-mass element (IME) line blends, continuum emission, the N and O blends, Fe L blend peak energy reflecting the electron temperature or plasma ionization state. For SN 1006, we also found that VAE has captured the feature of synchrotron emission and line emissions of O, Ne, Mg, and IMEs in the latent space. VAE is capable of unveiling the meaning of the latent axes, and help us to understand the dimensionality reduction result.

These results show that unsupervised machine learning can be useful for extracting characteristic spatial structures from spectral information in observational data (without detailed spectral analysis), which would reduce human-intensive preprocessing costs for understanding fine structures in diffuse astronomical objects, e.g., SNRs or galaxy clusters. Our method is also applicable to temporally variable data, i.e., light curves, because the training uses only spectral information. Furthermore, our method can also be applied to other energy bands. Such data-driven analysis techniques can be used to select regions to extract spectra for detailed analysis and help us make the best use of the large amount of spectral data currently available and arriving in the coming decades.



# Contents

Chapter 1	Introduction	1
Chapter 2	Machine Learning and Their Applications in Astronomy	3
2.1	Machine Learning and Deep Learning . . . . .	3
2.1.1	Supervised and Unsupervised Learning . . . . .	5
2.1.2	Generative and Discriminative Models . . . . .	6
2.2	Machine Learning Algorithms . . . . .	7
2.2.1	Supervised Machine Learning Algorithms . . . . .	7
2.2.2	Dimensionality Reduction . . . . .	7
2.2.3	Unsupervised Clustering . . . . .	8
2.3	Machine Learning Applications in Astronomy and Astrophysics . . . . .	9
Chapter 3	Supernova Remnants and the Observations	17
3.1	Supernovae . . . . .	17
3.1.1	Thermonuclear Supernovae . . . . .	18
3.2	Supernova Remnants . . . . .	19
3.3	X-Ray Radiative Processes . . . . .	20
3.3.1	Line Emission . . . . .	20
3.3.2	Bremsstrahlung . . . . .	21
3.4	<i>Chandra</i> X-Ray Observatory . . . . .	22
3.4.1	ACIS . . . . .	23
3.4.2	Spatial Resolution . . . . .	24
3.4.3	Effective Area . . . . .	25
3.4.4	Energy Resolution . . . . .	27
Chapter 4	New Method Using Deep Learning	29
4.1	Our Method Applied for X-Ray Data of SNRs . . . . .	29
4.1.1	Variational Autoencoder . . . . .	31
4.1.2	VAE Architecture Using Poisson Reconstruction Loss . . . . .	34
4.1.3	Gaussian Mixture Model . . . . .	35
4.1.4	t-SNE . . . . .	35
4.1.5	Hierarchical Clustering . . . . .	36

4.2	Demonstration of the Method with MNIST . . . . .	36
Chapter 5	Demonstration with Tycho’s SNR	39
5.1	Introduction for Tycho’s SNR . . . . .	39
5.2	<i>Chandra</i> ACIS-I Data Set . . . . .	40
5.3	Unsupervised Dimensionality Reduction and Clustering . . . . .	43
5.3.1	VAE Dimensionality Reduction . . . . .	43
5.3.2	GMM Clustering . . . . .	45
5.4	Detailed Results of Clustering . . . . .	48
5.5	Detailed Analyses of the Regions Suggested by Machine Learning . . . . .	54
5.5.1	Spectral Analysis of the Fe Knot . . . . .	54
5.5.2	Spectral Analysis of NW Ejecta . . . . .	58
5.6	Discussions . . . . .	61
Chapter 6	Demonstration with Kepler’s SNR	65
6.1	Introduction for Kepler’s SNR . . . . .	65
6.2	<i>Chandra</i> ACIS-S Data Set . . . . .	66
6.3	Results of Unsupervised Machine Learning . . . . .	69
6.3.1	The VAE Architecture . . . . .	69
6.3.2	The Results with the Best VAE Model . . . . .	72
6.3.3	Reconstruction of Spectra . . . . .	73
6.3.4	Spectra Generated by the Decoder . . . . .	74
6.3.5	Combined Spectra for Each Latent Axis . . . . .	76
6.3.6	Results of GMM . . . . .	79
6.4	Discussions . . . . .	88
Chapter 7	Demonstration with SN 1006	91
7.1	Introduction for SN 1006 . . . . .	91
7.2	<i>Chandra</i> ACIS Data Set . . . . .	92
7.3	Dimensionality Reduction by VAE . . . . .	94
7.3.1	The VAE Architecture . . . . .	94
7.3.2	The Latent Variables . . . . .	95
7.3.3	Reconstruction of Spectra . . . . .	96
7.3.4	Spectra Generated by the Decoder . . . . .	97
7.3.5	Averaged Spectra for Each Latent Axis . . . . .	98
7.3.6	Results of GMM . . . . .	102
7.4	Application of t-SNE . . . . .	109
7.4.1	Dimensionality Reduction by t-SNE . . . . .	109
7.4.2	Hierarchical Clustering . . . . .	109
7.5	Discussions . . . . .	111
7.5.1	Dimensionality Reduction by VAE . . . . .	111

7.5.2	Dimensionality Reduction by t-SNE . . . . .	113
7.5.3	Hierarchical Clustering . . . . .	113
Chapter 8	Conclusions	115
	Acknowledgements	119
	Bibliography	120



# Chapter 1

## Introduction

In the past decade, machine learning, especially deep learning using a deep neural network (DNN), has occupied an important position in data science because of its rapid development and high versatility. It is capable to capture features in big and high-dimensional data, and is applied in many areas. It has the potential to assist in analysis of astronomical data and to extract important information from rich astronomical data in a less biased way as reviewed in Section 2.3.

Astronomical observations produce complex multidimensional data that include spatial, temporal, and spectral information. Especially X-ray data obtained from a single observation of a diffuse source, e.g., a supernova remnant (SNR), a star forming region, a galaxy, or a galaxy cluster, may contain all these types of information. It tends to be difficult for human to capture features in such complex data.

In the near future, a dramatic improvement in the energy resolution of X-ray observations is expected; e.g., *XRISM* will have an energy resolution of several electron volts (Tashiro et al., 2018), and *Athena* will have a spectral resolution of 2.5 eV up to 7 keV at a spatial resolution of  $\sim 5$  arcsec with  $\sim 4000$  pixels (Barret et al., 2018). In their spectra, emission lines, which are blended due to the limited spectral resolutions of present detectors (e.g.; charge coupled devices, CCDs), will be resolved each other. The observation of a diffuse source, e.g., SNR, will produce high-dimensional big data containing a large amount of highly-resolved images and spectra. These upcoming data will allow us to perform more detailed plasma diagnoses and to reveal the plasma kinematics, thermodynamics, ionization, and elemental compositions. Detailed analyses of such data are quite important to exploit information of physical processes, but may require excessive human resources. Thus, automatic methods to discover features and pre-analyse the data are required to exploit the full potential of upcoming instruments.

An SNR is a diffuse source that may have a different radiative process due to a different physical state and process in each spatial structure. Spectroscopy of SNRs can mainly study two key issues; nucleosynthesis of SN explosion, and cosmic-ray acceleration at shock waves. X-ray spectra of an SNR contains thermal components emitted from thermal plasma and nonthermal components emitted from accelerated electrons. Spatially-resolved spectral analyses of thermal emission from SNRs have revealed the spatial variations of temperature, ionization, densities or elemental abundances due to asymmetric nucleosynthesis in an SN (e.g.; Yamaguchi et al., 2017), and also SNR evolutions, for example, shock heating of electrons (e.g.; Yamaguchi et al., 2014), or the interaction with ambient media (e.g.; Reynolds et al., 2007). Meanwhile, the non-thermal spectra from SNRs have reveal acceleration of electrons (e.g.;

Uchiyama et al., 2007) and protons (e.g.; Warren et al., 2005).

Plasma motions of young SNRs on the line of sight have been measured using the Doppler shifts of X-ray emission lines (e.g.; Hayato et al., 2010; Sato & Hughes, 2017a,b). Further, comparison between some observations operated in some years allow us to measure proper motions in some SNRs (e.g.; Vink, 2008; Katsuda et al., 2008, 2010, 2013). Combining the Doppler shifts and proper motions in an SNR, the three dimensional plasma motion, which provides constraints to establish realistic SN explosion models, can be reconstructed (e.g.; Williams et al., 2017; Millard et al., 2019).

Machine learning applications have been studied in some fields of astronomy, which have already entered the big data era. Especially supervised classification tasks of supernovae (SNe) or galaxy shape in optical observations have been studied well (e.g.; Charnock & Moss, 2017; Dieleman et al., 2015; Schaefer et al., 2018). On the other hand, machine learning applications to capture characteristic features from astronomical data, especially for spatially-resolved spectral analyses in diffuse sources, have not been explored well. Such analyses in diffuse sources are important to reveal key topics of astrophysical phenomena, although the observations of diffuse sources obtain complex multidimensional data and the analyses tend to be difficult.

In this thesis, we propose a new automatic method using DNN to capture features in such observational data for the spatially-resolved spectral analyses, and applied it to diffuse astronomical sources for the first time. We selected an appropriate DNN model, a variational autoencoder (VAE; Kingma & Welling, 2013), which is one of well-known DNN generative models, to automatically exploit essential features from observational data, because the DNN model is capable to capture nonlinear features and the features in the data may be represented nonlinearly. We explore the method combining nonlinear dimensionality reduction by a VAE and clustering by a Gaussian mixture model (GMM) in order to practically realize unsupervised classification. This method allows us to capture spatial structures owing to classification of spectra in individual spatial bins. We also newly developed the input and output of VAE to apply Poisson reconstruction loss for VAE training in order to be adapted to observational X-ray spectral data.

Our demonstrations of the method applied to SNRs show that the method successfully extracted some characteristic spatial structures in SNRs. The demonstrations show that our method can be useful for extracting characteristic spatial structures from spectral information in observational data and can help us for understanding the fine structures in diffuse objects. In addition, our method is not limited to SNRs in X-ray and is expected to widely apply to other classes of astronomical sources and to other energy bands.

This thesis is organized as follows. In Chapter 2, we briefly review machine learning methods and the previous applications for astronomy. Chapter 3 introduces supernovae and their remnants, and X-ray observations with the instruments on board *Chandra*. In Chapter 4, we describe our machine learning method. Chapter 5, Chapter 6, and Chapter 7 present the individual results and discussions for *Tycho*'s SNR, *Kepler*'s SNR, and SN 1006, respectively. Chapter 8 concludes this thesis.

## Chapter 2

# Machine Learning and Their Applications in Astronomy

### 2.1 Machine Learning and Deep Learning

Machine learning (for details, see e.g. Bishop, 2006) is a method to train a computer program for intelligent tasks through experiences. ‘Learning’ is a process based on data to make a machine learning model enable to accurately behave with unknown data, and not to memorize all data, i.e., ‘generalization’. In other words, machine learning aims at capturing universal empirical knowledge. Overfitting, where machine learning does not generalize and fit too closely to a particular dataset, tends to occur for a complex model with a small size dataset. Although generalization is not an easy problem for machine learning, deep learning, which is a category of machine learning and is recently rapidly developed, has highly performed generalization in many tasks.

Deep learning (for details, see e.g. Bishop, 2006; Goodfellow et al., 2016) is one of the most successful machine learning algorithms, and is applied in many areas. Recent computational advances have made it possible to train deep neural networks at a reasonable time and cost, and such techniques have become very popular in many areas including physics and astronomy.

An artificial neural network (ANN), which consists of several layers of multiple formal neurons, mimics the functioning of animal brains (e.g., LeCun et al., 2015) and can implicitly capture features embedded in high-dimensional data. A schematic diagram of an ANN is shown in Figure 2.1.1. Each neuron transforms its input  $x$  as  $f(\sum \mathbf{W}x + \mathbf{b})$ , where  $\mathbf{W}$  and  $\mathbf{b}$  are tunable weight parameters and biases, and  $f(x)$  is a nonlinear function called ‘activation function’. When each neuron in one layer is connected to all neurons in the next layer, the layer is called a fully connected (FC) layer. By contrast, in a convolutional layer, each neuron is connected to only a part of neurons, meanwhile the weight matrix corresponds to the kernel of convolution. A network using convolutional layers is specifically called a convolutional neural network (CNN). The processing layers learn representations of data with multiple levels of abstraction. Multilayer ANN architectures, such as the multilayer perceptron, can reveal complex, nonlinear relations.

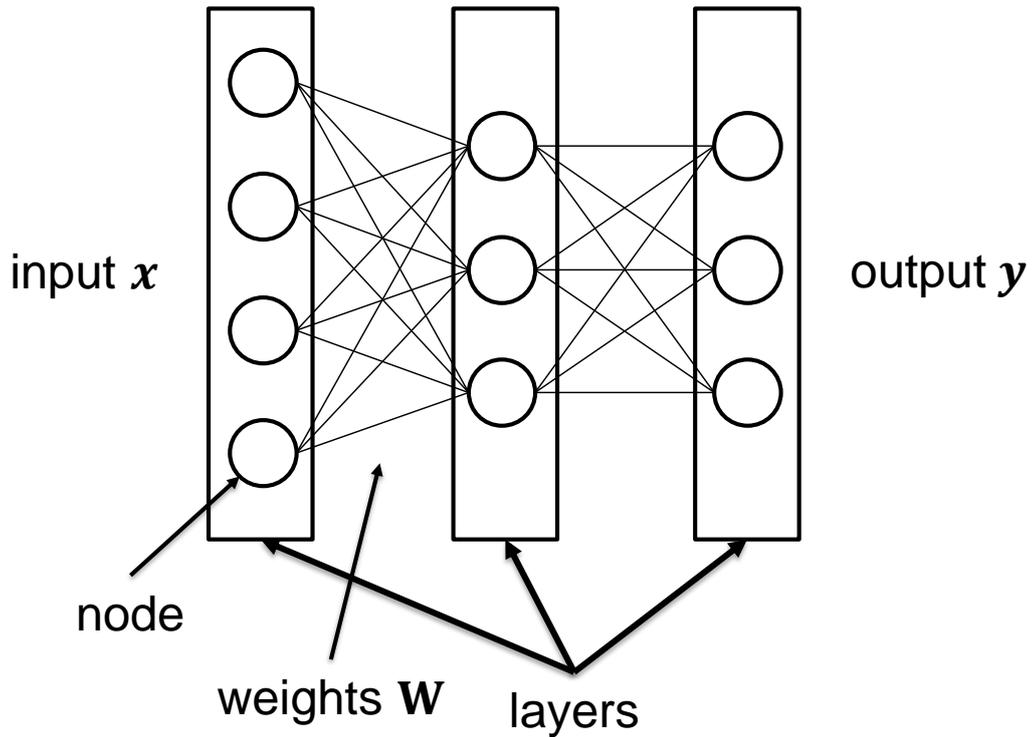


Figure 2.1.1: Schematic diagram of an ANN.

In the case of a simple sequential model, ANN has a layered structure like

$$\mathbf{y} = f^{(N)}\left(\sum \mathbf{W}^{(N)} f^{(N-1)}\left(\dots f^{(2)}\left(\sum \mathbf{W}^{(2)} f^{(1)}\left(\sum \mathbf{W}^{(1)} \mathbf{x}\right)\right)\dots\right)\right), \quad (2.1.1)$$

where  $N$ ,  $\mathbf{x}$ , and  $\mathbf{y}$  are number of layers, input, and output vectors, respectively (e.g.; Bishop, 2006; Goodfellow et al., 2016). Each component of  $f^{(n)}(\sum \mathbf{W}^{(n)} \mathbf{z}^{(n)})$  represents the  $n$ th layer, where  $\mathbf{z}^{(n)}$  is the output of the  $(n - 1)$ th activation function. For  $N > n \geq 1$ , the layers are called intermediate layers, or hidden layers, which learn representations of data. The last layer  $f^{(N)}(\sum \mathbf{W}^{(N)} \mathbf{z}^{(N)})$  is the output layer, which determines the output shape and the variable scale.

An ANN constructed with only one layer cannot be trained for non-linearly-separable problems, in which two classes cannot be separated by one hyperplane. However, an ANN architecture constructed with two or more layers is capable of approximating any functions. While multiple-layer ANNs are difficult to train, the back propagation algorithm made them trainable at a reasonable time. A deep multiple-layer ANN is also called a deep neural network (DNN). As the number of layers increase, training tends to become more difficult because of some troubles, e.g., vanishing gradient. Therefore various techniques have been developed.

Gradient descent algorithms are used for deep learning training. ‘Learning’ of an ANN is finding the weights contained in the layers that optimise the cost function. In the most simple way, using the gradient of the cost function  $E(\mathbf{W})$ ,

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla E(\mathbf{W}^{(t)}) \quad (2.1.2)$$

where  $\eta$  is a learning rate determining the change  $\Delta \mathbf{W}^{(l)}$  in a step (e.g.; Bishop, 2006). The learning rate is usually a hyperparameter, i.e., a parameter predefined by a human. The cost function is generally a complex non-convex function, which has not only the global minimum but also huge amount of local minima. ANNs, however, can perform well on local minima having enough good cost function values, even if they have not found the global minimum yet. It is one of mysteries of deep learning. There are also other problems for training, e.g., saddle points and plateaux, which have 0 gradients, and oscillations in narrow valleys. Thus, optimizer algorithms have been improved to avoid such the problems, e.g., Adam (Kingma & Ba, 2014).

The gradient descent algorithms calculate a derivative of the cost function by the parameters, thus the activation functions need to be differentiable. We, here, introduce some popular activation functions and show them in Figure 2.1.2. Recent reviews and comparisons of activation functions are available in Nwankpa et al. (2018); Eger et al. (2018). Sigmoid and tanh have shapes like a step function but smooth, and take a value  $0 < f_{\text{sigmoid}} < 1$  and  $-1 < \tanh < 1$ , respectively, thus they are often used in binary classification tasks. A sigmoid function for input  $\mathbf{u}$  can be written as

$$f(\mathbf{u}) = \frac{1}{1 + e^{-\mathbf{u}}}. \quad (2.1.3)$$

Softmax is an extension of sigmoid for categorical classification tasks. It takes a value between 0 and 1, and the sum over  $d$  is always 1, where  $d$  is a variable number of vector  $\mathbf{u}$ . A softmax for the  $d$ th variable of  $D$ -dimensional vector  $\mathbf{u}$  can be written as

$$f(u_d) = \frac{e^{u_d}}{\sum_{d=1}^D e^{u_d}}. \quad (2.1.4)$$

They are usually used in output layers and not in hidden layers. Rectified linear unit (ReLU; Nair & Hinton, 2010) function works better in hidden layers for smooth training. A ReLU function can be written as

$$f(\mathbf{u}) = \max\{0, \mathbf{u}\}. \quad (2.1.5)$$

ReLU is not smooth at the threshold (usually 0), which has a bend, and so globally smooth functions approximating ReLU have been developed, e.g., softplus (Dugas et al., 2001) written as

$$f(\mathbf{u}) = \log(1 + e^{\mathbf{u}}). \quad (2.1.6)$$

### 2.1.1 Supervised and Unsupervised Learning

Machine learning algorithms are mainly classified into three classes; supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms take a dataset containing data and labels for training to predict the data label. By contrast, unsupervised learning algorithms use a dataset that is not labelled, and find structures in the data. Reinforcement learning algorithms find the way of agents to choose actions in an environment so as to maximize the rewards.

Supervised learning algorithms have two classes; classification, which classifies data into some classes, and regression, which predicts continuous values. Some algorithms that simultaneously perform object

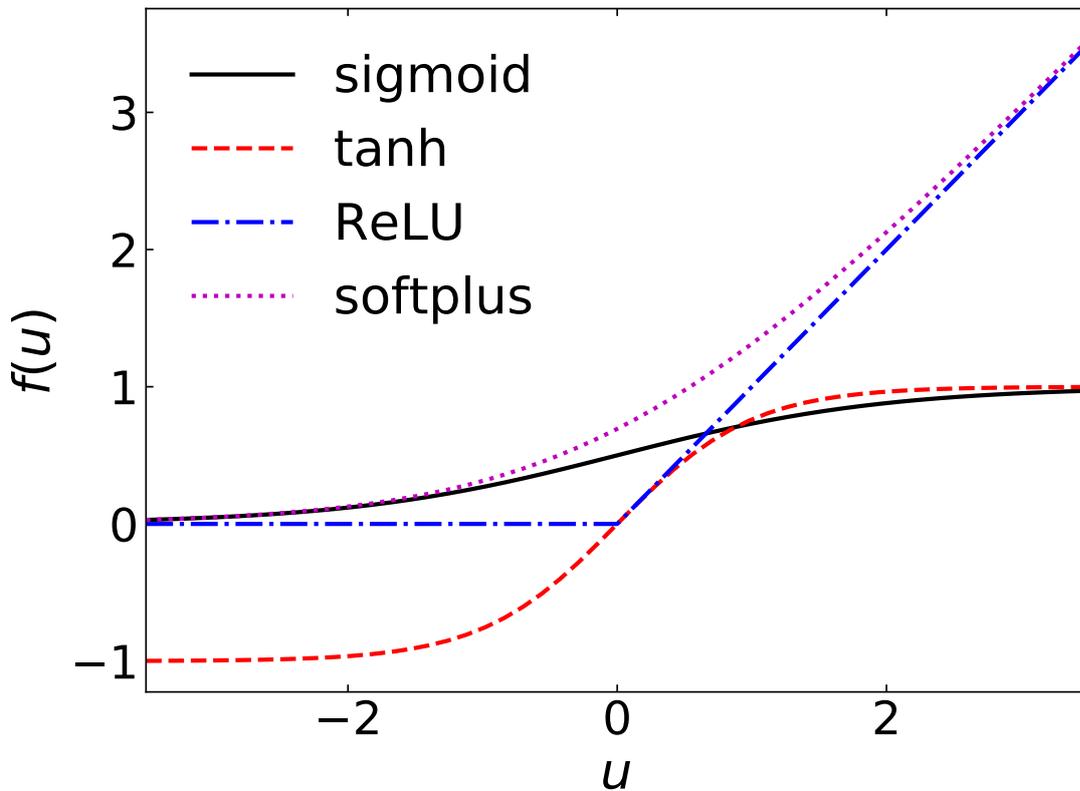


Figure 2.1.2: Activation functions.

detection and classification of detected objects, have been developed e.g., Region-based CNN (R-CNN) and the derivations (He et al., 2017), You Only Look Once (YOLO; Redmon et al., 2015; Redmon & Farhadi, 2018), and Single Shot MultiBox Detector (SSD; Liu et al., 2015). Segmentation algorithms allow us to perform pixel-by-pixel classification in images, e.g. U-net (Ronneberger et al., 2015).

Unsupervised learning algorithms include methods of dimensionality reduction, feature extraction, manifold learning, and clustering. Unsupervised learning algorithms are useful to capture the features lying in data. Unsupervised learning is applicable to problems in which what to output cannot be defined before training. It is a merit of unsupervised learning that lower cost is spent for dataset preparation than supervised learning because no labelling is required.

### 2.1.2 Generative and Discriminative Models

Machine learning methods can also be categorised differently; generative models and discriminative models. Generative models directly optimize joint probability distributions  $P(x, y)$ , where  $x$ , and  $y$  are explanation variable (input), and target variable (output), respectively, which are the mechanisms to generate the data.

Generative adversarial network (GAN; Goodfellow et al., 2014), variational autoencoder (VAE; Kingma & Welling, 2013) and their derivations are well-known generative models of DNN. VAE is

an architecture of autoencoders (AEs), which are neural network models connecting an encoder and a decoder. In Section 4.1.1, we introduce VAE in detail.

GAN, which is capable to generate samples just like real data, is one of the most successful neural network models especially in the field of image data. However, the training of GAN is practically difficult and one needs to select carefully the architecture and the hyperparameters, which are predefined by a human. GAN has an architecture combining a generator and a discriminator. During GAN training, the generator tries to produce fake samples, meanwhile the discriminator tries to distinguish samples from training dataset or from the generator.

By contrast, usual classification or regression methods like we have introduced above are discriminative models, which model conditional distributions  $P(y|x)$ .

## 2.2 Machine Learning Algorithms

Many machine learning techniques have been developed. In this section, we briefly introduce some representative methods.

### 2.2.1 Supervised Machine Learning Algorithms

Random Forest (RF; Breiman, 2001) is an ensemble machine learning algorithm, which makes an ensemble of multiple simple models to obtain better prediction performance, using decision trees. RF can measure the importance of each feature. RF is a nonparametric algorithm because the decision tree is nonparametric. RF is applicable to classification and regression.

Support vector machine (SVM; Burges, 1998) divides data with hyperplanes, which maximize the margins. Using the ‘kernel trick’, which projects dataset to a linearly-separable space with a kernel (e.g., radial basis function<sup>1)</sup>, polynomial), SVM is capable to divide nonlinear features in dataset. SVMs are applicable to classification, regression and clustering tasks.

### 2.2.2 Dimensionality Reduction

Principal component analysis (PCA) is one of the best-known classical dimensionality reduction methods. PCA attempts to find directions (namely principal components), which have large variance in dataset, and transforms data axes to the new axes orthogonally. Therefore, the transformed axes have linear relations to the original data axes. The principal components have a sequence such that the first principal component corresponds to the largest variance in the dataset meanwhile the last principal component has the lowest variance.

Manifold learning, e.g., t-distributed Stochastic Neighbour Embedding (t-SNE; van der Maaten & Hinton, 2008) and isomap (Tenenbaum et al., 2000), is an approach to nonlinear dimensionality reduc-

---

<sup>1)</sup> radial basis function (RBF) kernel, i.e., Gaussian kernel

Table 2.2.1: Machine learning methods

Method	Property and Usage
Machine Learning Models	
RF	ensemble method with decision trees for classification/regression
SVM	linear or nonlinear (kernel) classification/regression
PCA	orthogonal transformation for dimensionality reduction
t-SNE	nonlinear manifold learning dimensionality reduction
isomap	nonlinear manifold learning dimensionality reduction
GMM	each sample is assigned to all categories with different weights (soft clustering)
k-means	each sample is assigned to one category (hard clustering)
ANN Models	
DNN	classification or regression on nonlinear features
CNN	classification or regression on image data
RNN	classification or regression on time-series data
R-CNN	object detection model combining region proposal and classifier/regressor
YOLO	fast object detection using one sequential network
SSD	fast object detection using one sequential network
U-net	segmentation model using only CNN
AE	an ANN generative model combining encoder and decoder for dimensionality reduction
VAE	an AE with multidimensional-Gaussian latent variables
GAN	an ANN generative model using generator and discriminator

tion, assuming that a low-dimensional data are embedded in a higher-dimensional space. We describe t-SNE in Section 4.1.4.

### 2.2.3 Unsupervised Clustering

There are two major types of clustering methods; (1) hard clustering methods, in which each data point is assigned to only one category (e.g., k-means, SVM), and (2) soft clustering methods, which assign each point to all the categories with different weights (e.g., mixture model).

k-means is one of the best-known clustering algorithms. Each data sample is assigned to the cluster whose centre is the nearest to the sample, thus this method is a hard clustering. This method attempts to minimize the square distance between each data sample and the assigned cluster centre, which is called a distortion measure.

Gaussian mixture model (GMM) is a well-known soft clustering method, and we describe it in Section 4.1.3 in detail.

Hierarchical clustering can categorise data points hierarchically, and we describe it in Section 4.1.5.

## 2.3 Machine Learning Applications in Astronomy and Astrophysics

Recent and upcoming large scale surveys or instrumental improvements make astronomical observations big-data in the point of quantity and quality. This trend accelerates applications of machine learning or data scientific methods to exploit the full potential of the rich data. Recently, rapid development of deep learning accelerates to apply the algorithms.

In modern astronomy, the universe and the objects are observed with multimessengers, i.e. multiwavelength electromagnetic waves, neutrinos, and gravitational waves. Observational data of each messenger have quite different characteristics, and therefore have different problems of observations or analyses from other messengers.

In optical or infrared, large scale survey projects produce huge amount of observational data, which are beyond the capacity of analysis by human experts. Time domain observations require fast and accurate detection of time variability, e.g., discriminating supernovae (SNe) from data containing non-astronomical signals (e.g., noises, and artificial objects), whose numbers are sometimes some orders of magnitude more than that of true SNe signals.

Classification methods of machine learning have been applied to such large amount of data, especially wide-field-of-view surveys. In the image classification terms, CNNs, which are the most successful methods in the deep learning and are state of the art, have classified observational image data of galaxy structures (e.g.; Dieleman et al., 2015; Ribli et al., 2019) and gravitational lensing (e.g.; Schaefer et al., 2018)

Some ANN and non-ANN classification algorithms have been applied to SN type classification, e.g., recurrent neural networks (RNN; Charnock & Moss, 2017), ensemble learning of DNN, RF, and AUC<sup>2)</sup> boosting (Morii et al., 2016). Automatic SN type classifications enable real-time analysis and triggering for follow-up observations for rare type SNe or type Ia SNe, which are available as standard candles of the universe.

In addition, for optical survey images, segmentation algorithms have been applied to detect astronomical objects (Mask R-CNN; Burke et al., 2019), and classify the images in pixel level (U-net; Hausen & Robertson, 2019).

Regression methods were also applied to direct prediction of physical parameters, e.g, gravitational lenses (Hezaveh et al., 2017), and stellar effective temperature from optical observational data by *GAIA* using RF (Bai et al., 2019).

In radio interferometry, incomplete sampling of the spatial frequency space makes imaging from observed visibilities an under-determined problem. Recently, sparse modelling algorithms made it possible to reconstruct high resolution images from radio interferometry (Honma et al., 2014; Akiyama et al., 2017), and supported the first imaging of the black hole shadow of M87 (Event Horizon Telescope Collaboration et al., 2019a) (Figure 2.3.1). In addition, radio observations obtain multidimensional data,

---

<sup>2)</sup> Area under the Receiver operating characteristic Curve (AUC)

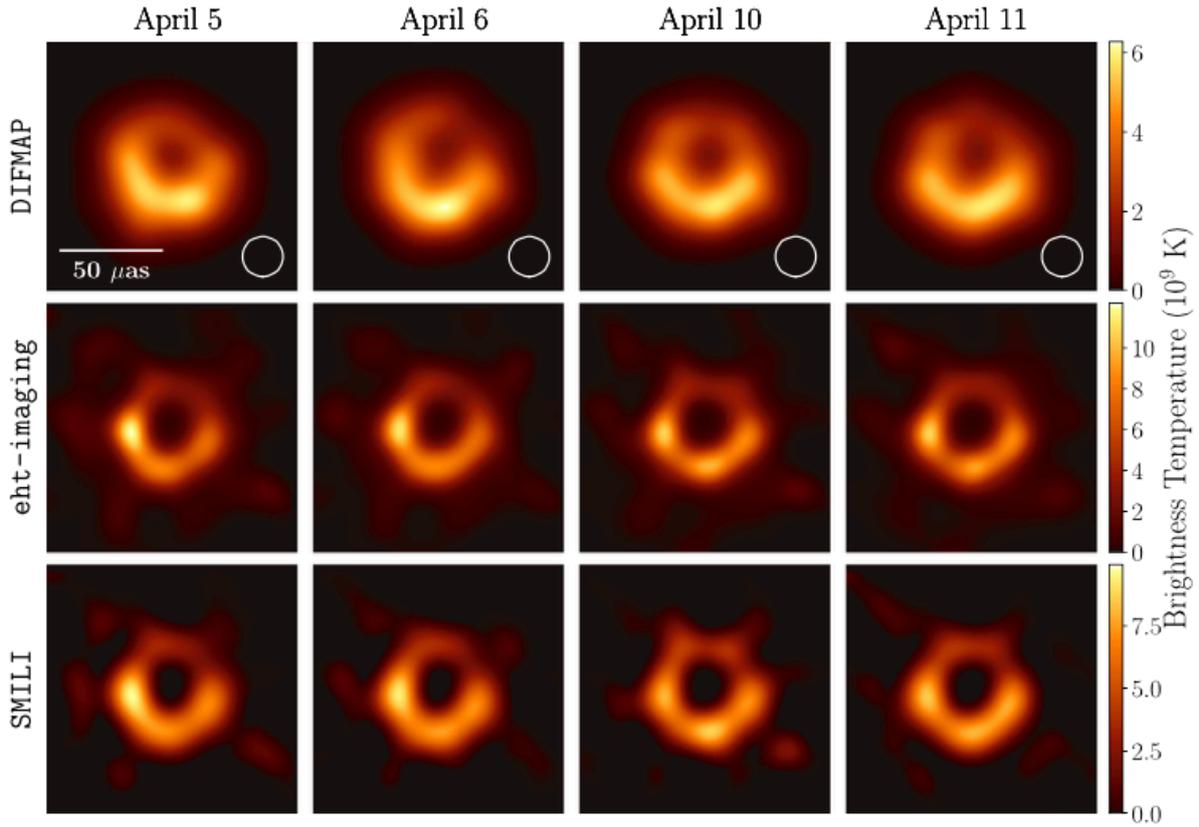


Figure 2.3.1: Fiducial images of M87 on all four observed days from each of the three imaging methods. The top row shows the images by an ordinary method, CLEAN from DIFMAP, while the middle and bottom rows are produced by sparse modelling methods, eht-imaging and SMILI, respectively. The sparse modelling methods reconstruct finer images than the ordinary one. This figure is Figure 11 of (Event Horizon Telescope Collaboration et al., 2019b).

which contain spatial, temporal, and velocity information.

In X-ray and  $\gamma$ -ray, multidimensional data containing spatial, temporal, and spectral information are obtained in an observation. In the next decade, the energy resolution of X-ray observation will be improved (several tens or hundreds of electron volts by CCD to several electron volts by microcalorimeter) by *XRISM* (Tashiro et al., 2018) or *Athena* (Barret et al., 2018). Automatic spectral analysis methods are studied for incoming such high resolution spectra, e.g., a DNN regression algorithm was applied to directly predict the physical parameters (electron temperatures, emission measures, and redshifts) of X-ray plasma spectra (Ichinohe et al., 2018). Figure 2.3.2 shows the spectrum of Perseus cluster observed by *Hitomi* and the predicted spectra by the DNN model.

In observations of  $\gamma$  ray, neutrino, and gravitational wave, astronomical signals are some orders of magnitude less than background signals, thus effective event selection are required. Machine learning classifications have been applied to event selection of such detectors. In high-energy  $\gamma$ -ray astronomy, some boosted algorithms have been implemented, RF for *MAGIC* (Albert et al., 2008), and boosted decision trees (BDT) for H.E.S.S. (Ohm et al., 2009; Becherini et al., 2011) and *VERITAS* (Krause et al., 2017). CNNs are going to be applied for H.E.S.S. (Shilon et al., 2019), *VERITAS* (Feng et al., 2017), and

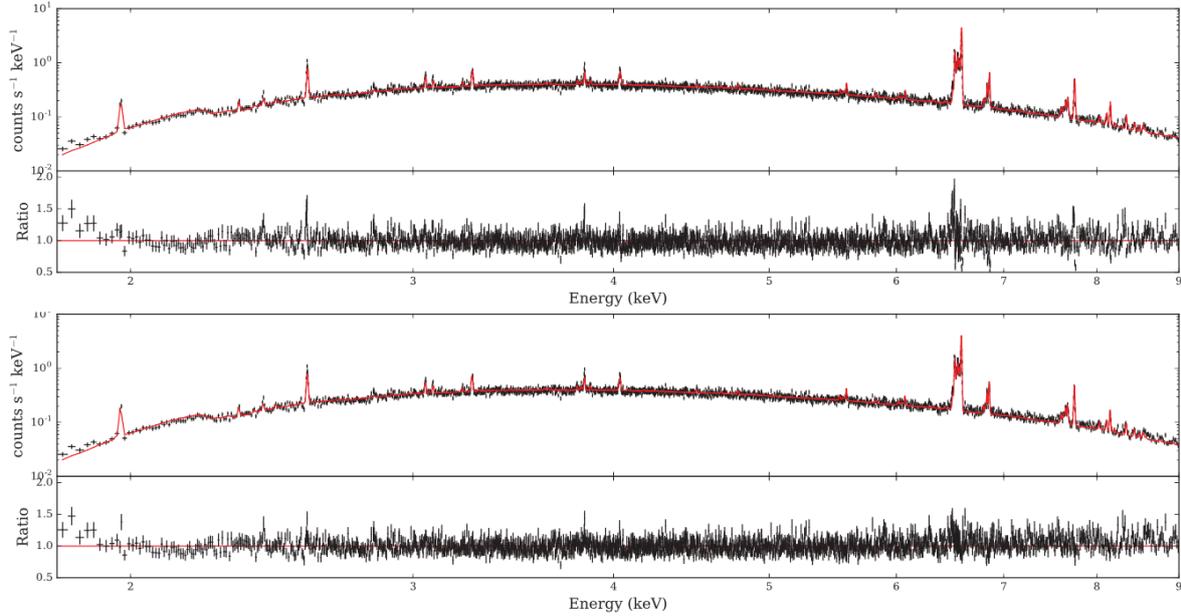


Figure 2.3.2: The top panel shows the spectrum (black points) and the plasma model (red line) based on the parameters predicted by the DNN. In the bottom panel, the model is the best-fitting one used the DNN prediction as the initial parameters. This figure is taken from Ichinohe et al. (2018).

*CTA* (Nieto Castaño et al., 2017). The *IceCube* Collaboration is also preparing to use DNN for the event type classification (Kronmueller & Theo, 2019).

The Sun is the nearest star from the earth, therefore a large amount of high-resolution observations are obtained from radio to  $\gamma$  ray. CNNs have successfully classified solar images (e.g.; Armstrong & Fletcher, 2019). Some machine learning methods and DNN also successfully predicted solar flares using physical parameters extracted from observational data based on the solar science knowledge (e.g.; Nishizuka et al., 2017, 2018).

GANs or VAEs, which are well-known generative models of DNN, have also been applied to some astronomical analyses. GAN has been applied for recovering features from optical observations that are low signal to noise and low angular resolution (Schawinski et al., 2017). GANs have also been applied for fast simulator generating images, e.g. cosmic web (Rodríguez et al., 2018), the cosmic microwave background (Mishra et al., 2019). VAE has been applied for anomaly detection of X-ray plasma spectra (Ichinohe & Yamada, 2019). Generative models, whose outputs are robust to noises, are also available for denoising of observational data (e.g. Díaz Baso et al., 2019).

The supervised classification tasks of astronomical objects or event selection in observations have been studied well. The supervised methods accurately work on data, which are classifiable by the predefined classes of the model, although they cannot work on data including unknown features, e.g., a classifier of SN types cannot classify an event whose type has not been discovered. On the other hand, unsupervised machine learning methods, which extract characteristic features from observational data, have not been explored enough, especially for spatially resolved analysis in diffuse sources, e.g., SNRs, galaxies, and galaxy clusters. Such analysis in diffuse sources tend to be difficult because of the complexity of high-

dimensional data, although the analysis can reveal a lot of important physics, e.g., chemical evolution of the Universe. Thus, this thesis propose a new method for the spatially-resolved spectral analysis in diffuse sources.

X-ray observations of SNRs, which we focused on, produce complex multidimensional data that include spatial, temporal, and spectral information. Because of this complexity, conventional analysis can be prone to human prejudice and oversight. SNRs have been observed by X-ray telescopes e.g. *Chandra*, *XMM-Newton*, and *Suzaku*, many times, and the huge amount of rich data are archived and available. As we mentioned above, microcalorimeter on e.g., *XRISM* or *Athena* will obtain new highly resolved spectroscopic data in the near future. Detailed analysis of such data may require excessive human resources. Thus, automatic and less biased methods to discover features and pre-analyse the data are required to exploit the full potential of upcoming instruments.

To take advantage of the rich data contained in SNR observations and to extract essential information without human bias, some classical machine learning techniques have been explored. Warren et al. (2005) and Warren (2006) demonstrated the separation of mostly featureless and line-dominated emission from *Tycho*'s SNR using a linear dimensionality reduction method, PCA. They extracted 12 new axes from 12 broad spectral channels and found that the image of the first principal axis corresponds to the contrast between Si- and Fe-rich emission and the hard continuum emission as shown in the top right-hand panel of Figure 2.3.3. Sato & Hughes (2017a) also performed PCA of the narrow-band spectra of *Tycho*'s SNR, separating the Si He  $\alpha$  band into 18 bins, and found that the first three principal components correspond to the line equivalent width, line energy centroid, and line energy width, respectively. In Figure 2.3.4, the principal component corresponding to the line centroid is shown. Burkey et al. (2013) demonstrated clustering of four-band line fluxes extracted from 5000 spatial regions of Kepler's SNR using a GMM and identified the shocked circumstellar-medium (CSM) region (Figure 2.3.5).

Most previous applications of machine learning techniques to analysis of SNR data have been limited to linear methods. However, the value of each spectral bin depends nonlinearly on the underlying physical parameters; e.g., the bremsstrahlung continuum emission is exponentially affected by the plasma temperature. In other words, the data space of X-ray spectra is not flat. Although PCA, which linearly transforms the data into another orthogonal expression, might provide approximate results in some cases, it is reasonable to choose a model capable of expressing nonlinear relations when the problem exhibits such features. DNNs are likely to obtain more effective expressions from data spaces than linear methods because of their ability to handle nonlinear relations. In this research, we examine the potential of DNNs to extract features embedded in nonlinear relations.

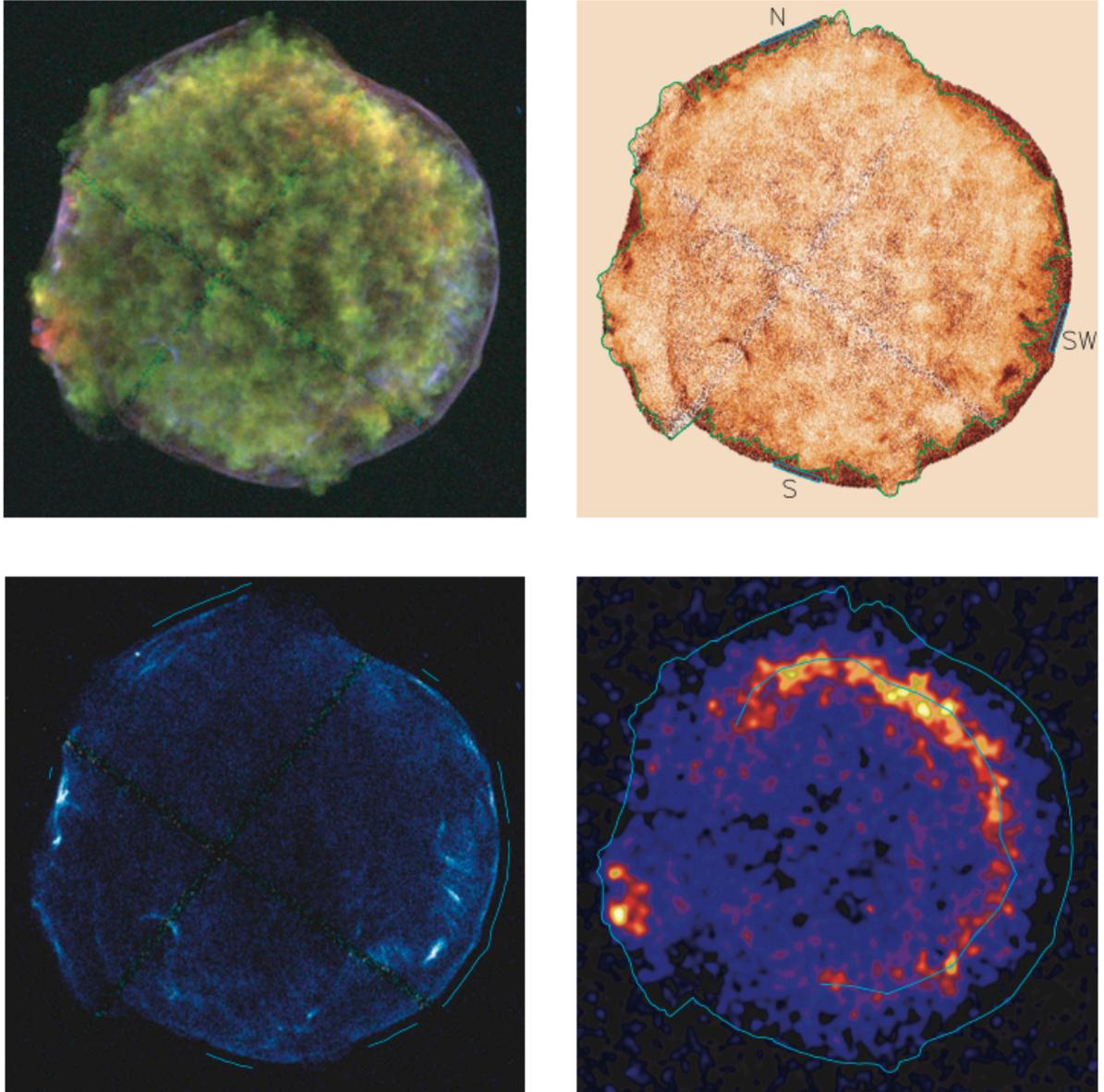


Figure 2.3.3: The top left-hand panel shows a three-colour image of *Tycho*'s SNR; red, 0.95–1.26 keV; green, 1.63–2.26 keV; blue, 4.1–6.1 keV. The top right-hand panel shows the first principal component that separates line-rich emission from continuum emission. The green contour indicates the contact discontinuity location. The bottom left-hand panel shows the continuum image (4–6 keV). The bottom right-hand panel shows the Fe K  $\alpha$  image from which the continuum (4–6 keV) is subtracted. The outer contour notes the forward-shock location, meanwhile the inner contour indicates the reverse-shock location. This figure is Figure 1 of Warren et al. (2005).

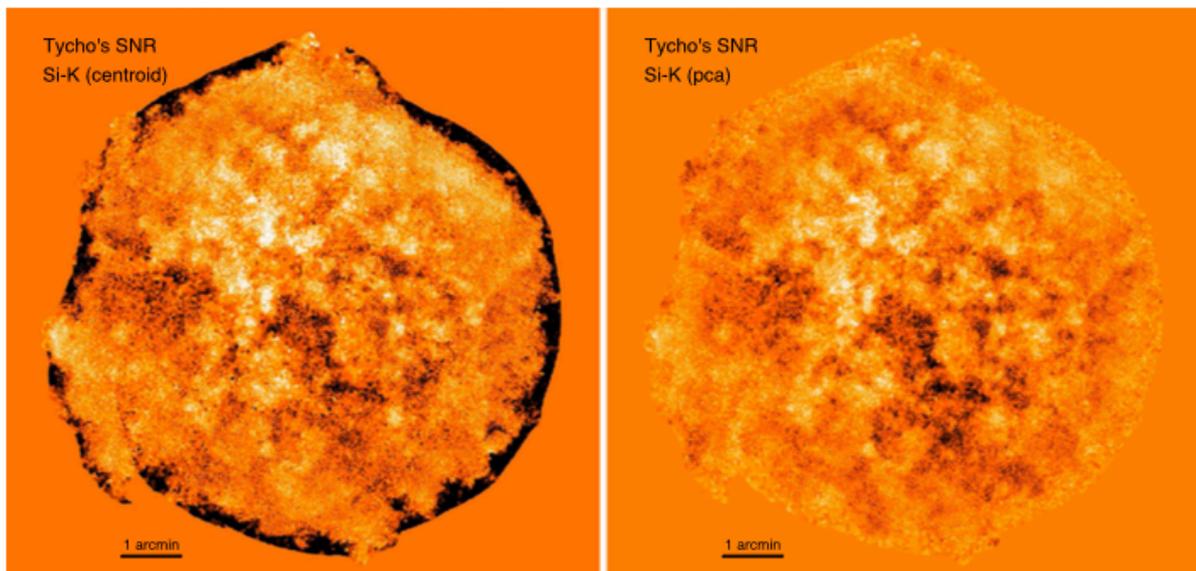


Figure 2.3.4: The left-hand panel shows a mean photon energy map in the Si He  $\alpha$  band (1.6-2.1 keV). The right-hand panel shows a principal component that separates redshift and blueshift of Si He  $\alpha$  line. This figure is Figure 4 of Sato & Hughes (2017a).

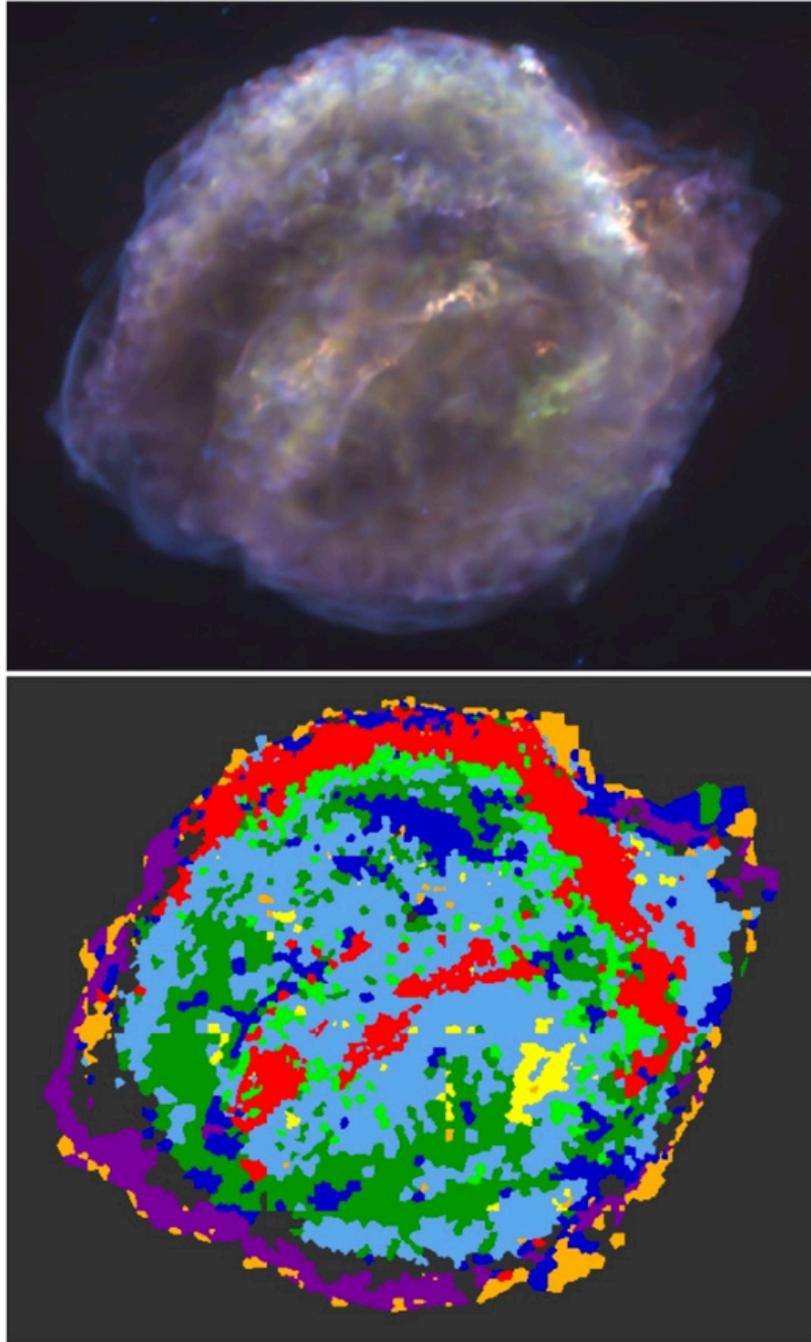


Figure 2.3.5: The top panel shows a three-colour image of *Kepler*'s SNR; red, 0.3–0.72 keV; green, 0.72–1.7 keV; blue, 1.7–8 keV (Reynolds et al., 2007). The bottom panel shows the division of *Kepler*'s SNR into clusters. This figure is Figure 1 of Burkey et al. (2013).



## Chapter 3

# Supernova Remnants and the Observations

Supernovae (SNe), which are explosions of massive stars or white dwarfs, have been believed as the supplier of heavy elements. Unfortunately, we cannot directly observe spatial structures or distributions of SNe, because SNe are just point sources from the earth. However, the supernova remnants (SNRs), which are explosively ejected elements synthesised in the progenitors, expand in the interstellar medium (ISM) for tens or hundreds of thousands years. The SNRs in our Galaxy, or the neighbour galaxies (e.g., Large or Small Magellanic Clouds) are observed as diffuse sources, and allow us to investigate the spatially-resolved elemental distributions, compositions and motions. Such studies reveal the SN explosion mechanisms and the elemental synthesis in detail.

The synthesised elements are ionized by shock heating, thus the thermal plasma emit line emissions in X-ray band. Using the line emissions, plasma diagnoses indicate such important information of SNe.

In addition, galactic SNRs have also been believed to supply cosmic rays. In their blast waves, charged particles are accelerated to relativistic energies by diffusive shock acceleration. In X-ray band, accelerated electrons emit synchrotron emission mainly around the blast waves by interaction with magnetic fields. Therefore, the studies of shock waves and accelerated particles in SNRs allow us to reveal the mechanisms of cosmic-ray acceleration.

### 3.1 Supernovae

SNe are divided into two major categories based on the explosion processes; thermonuclear explosion of a white dwarf, and core collapse explosion of a massive star.

In an spectroscopic view, thermonuclear SNe correspond to type Ia SNe, which have no hydrogen line but silicon absorption lines in their spectra. Type Ia SNe have much less variation of the peak brightness than that of core collapse SNe. Thus the progenitors of type Ia SNe seem to be similar objects. Type Ia SNe are used as ‘standard candles’ of the Universe to measure the distances, and supplied the evidence of accelerating expansion of the Universe.

On the other hand, core collapse SNe correspond to the rest of spectroscopic classes, i.e., type Ib, Ic, and II. Core collapse SNe have rich variation in both spectra and lightcurves, reflecting variation of explosion details and the environments around them.

We describe the thermonuclear SNe below because we analysed only them in this thesis.

### 3.1.1 Thermonuclear Supernovae

Thermonuclear explosion occurs in a C-O white dwarf, which contains carbon and oxygen synthesised by helium burning in a progenitor star. White dwarfs are supported by their self gravities by degenerate pressures of electrons. The limit of white dwarf mass, which the degenerate pressure can support, is Chandrasekhar mass;

$$M_{\text{Ch}} \approx 1.454M_{\odot}, \quad (3.1.1)$$

where  $M_{\odot}$  denotes the unit of mass equal to the solar mass (Chandrasekhar, 1931).

The thermonuclear explosion needs a mechanism, which a C-O white dwarf reaches the Chandrasekhar mass, but the progenitor system has been under debates. The popular scenarios are single degenerate (SD) scenario and double degenerate (DD) scenario. The SD scenario assumes a binary system of a white dwarf and a star, and accretion of materials from the donor star grows the white dwarf. By contrast, the DD scenario assumes a binary system of two white dwarfs, and a merge of the white dwarfs causes an explosion. In addition, the core degenerate scenario assumes a binary system of a white dwarf and a massive evolved star like the SD scenario, but the explosion occurs after the white dwarf merges with the hot core of companion.

In thermonuclear SNe, more Fe group elements ( $\sim 0.6 M_{\odot}$ ) are synthesised than in core collapse SNe.

The nucleosynthesis models of thermonuclear explosion are divided into three broad categories; detonation, deflagration, and delayed detonation (DDT) models. The DDT models are currently the most popular ones.

Pure detonation models cause explosive nucleosynthesis due to compress and heat the material by a supersonic shock wave. In these models, almost all of white dwarf elements burn iron group elements. These models were ruled out by the optical spectroscopic observations of type Ia SNe, which contain significant amount of intermediate-mass elements (IMEs).

In pure deflagration models (e.g.; Nomoto et al., 1984; Thielemann et al., 1986), the burning front proceeds subsonic. At the burning front, convections caused by Rayleigh-Taylor instability maintain the nuclear fusion due to mix un-burnt material into the hot burning zone. When the burning front velocity becomes comparable to the material velocity, the burning front is quenched, and a narrow region containing rich IMEs is formed at the quenching front. However, the model produces no high-velocity IMEs, and hence the velocity range of IMEs is narrower than the optical spectra of type Ia SNe. In addition, the ratio  $^{54}\text{Fe}$  to  $^{56}\text{Fe}$  is predicted too high and disagrees with the observations of type Ia SNe.

In DDT model (e.g.; Khokhlov, 1991; Seitenzahl et al., 2013), propagation of burning front changes from deflagration to detonation while the explosion. The explosion starts as deflagration. The burning front propagating into outer and lower densities makes the sonic speed lower, and is finally quenched. And then, the shock wave is accelerated to supersonic, and the detonation burns the rest of white dwarf.

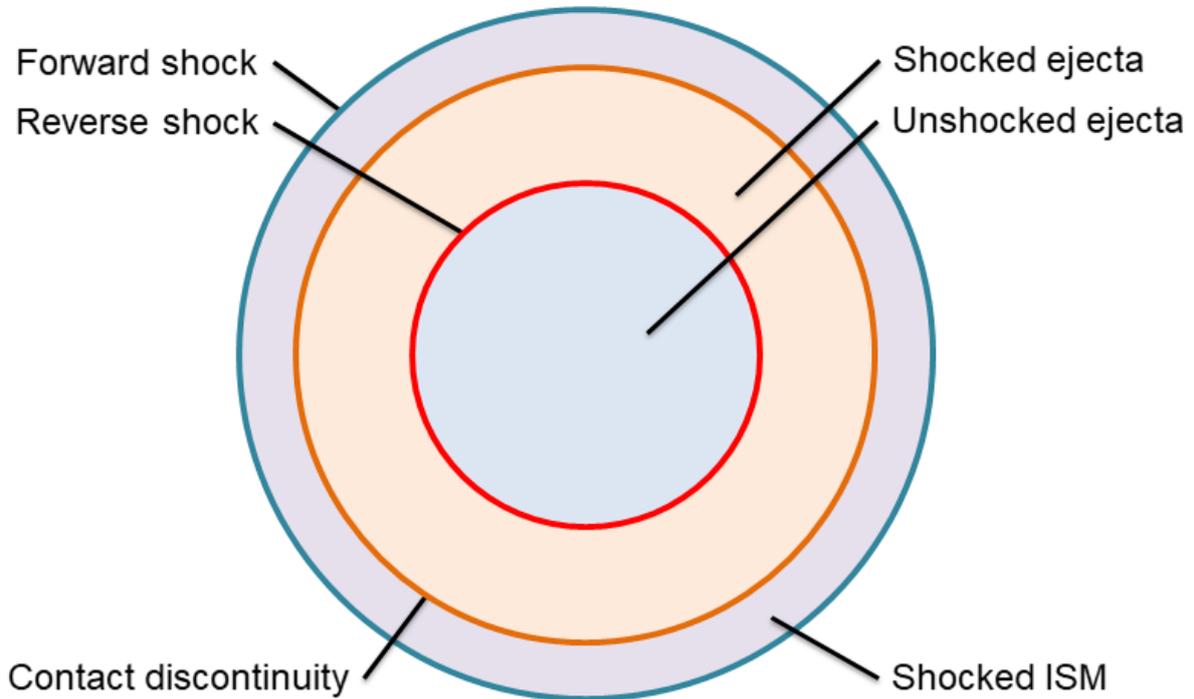


Figure 3.2.1: An schematic view of an SNR in the transition phase.

## 3.2 Supernova Remnants

SNRs have various morphologies each other, reflecting variation of the SN explosions and surrounding environments. Each SNR has rich structure inside it, which may have different radiative process due to different physical state each other.

Figure 3.2.1 shows a simple schematic view of an SNR in the transition phase. Global structure of a young SNR is following. In an SNR, a forward shock (FS) expands outermost of the SNR, and sweeps up ISM. At the FS, charged particles are accelerated, and make non-thermal emissions, e.g., synchrotron emission. The swept-up ISM is located between the FS and ejecta, which is usually at the inner side of FS. A boundary of ejecta and shocked ISM is called a contact discontinuity (CD), and is a outermost surface of ejecta.

The ejecta is heated from outermost toward the centre of the SNR by reverse shock propagating inward. The shocked ejecta is ionized and begins to emit thermal emissions, while unshocked ejecta, which a RS has not reached yet, makes no emission. After the shock-heating, the temperature of ejecta become non-equilibrium, i.e., temperatures of electron and all species of ion are different. In a young SNR, the RS has not reach the centre, and shock-heating is on going. Such SNR shows an electron heating at the RS

(e.g.; Yamaguchi et al., 2014), or a gradation of ionization state of ejecta, which reflects the elapsed time since the ejecta is shock-heated (e.g.; Sato et al., 2018). In middle-aged or old SNRs, all species of ions and electron become ionization equilibration if the ionization time scale reaches  $n_e t \gtrsim 10^{12} \text{ cm}^{-3} \text{ s}$ , where  $n_e$  and  $t$  are an electron density and elapsed time since the ejecta was shock-heated, respectively.

In addition, elemental abundance tends to be different in each position, reflecting nucleosynthesis in an SN explosion. A spherical explosion is expected to form stratified structure (i.e., heavy iron group elements are located innermost, and IMEs are at the outer side) that originated in difference of burning regimes. Some observations suggest the stratified interior to the outer ejecta (e.g.; Hwang & Gotthelf, 1997). An asymmetric explosion may make asymmetric structures, e.g., a distorted distribution, and a knot. Some type Ia SNRs show the asymmetric ejecta distributions, and are suggested that the origins are caused by the explosions (e.g.; Yamaguchi et al., 2012; Uchida et al., 2013)

Each SNR is expected to maintain structures imposed during an explosion. Therefore, detailed studies of SNRs can reveal the explosion mechanisms of SNe. A recent 3D simulation connecting SN and SNR reveals that the imprint of explosion maintains in the SNR morphology for hundreds years (Ferrand et al., 2019). Sato et al. (2019) found that clumpy morphology of *Tycho's* SNR is originated from a explosion through comparisons of *Chandra* observations and 3D simulations.

On the other hand, an interaction of an SNR and the environment surrounding it can cause the expansion deceleration, morphology deformation, and plasma evolution. Hydrodynamic instabilities affect morphology of an SNR during the evolution. Some SNRs have complex structures due to interact ambient media, which have been formed by the progenitors (e.g.; Burkey et al., 2013).

### 3.3 X-Ray Radiative Processes

An X-ray spectrum from an SNR is sum of emissions from charged particles and also multiplied absorptions. The thermal components, which are thermal bremsstrahlung and line emissions, are emitted by a thermal plasma. On the other hand, the non-thermal components, which are synchrotron, inverse Compton, and non-thermal bremsstrahlung, are emitted by accelerated charged particles, i.e., cosmic rays, having non-thermal energy distributions. In this section, we briefly describe the general properties of thermal radiation (for details, see e.g. Rybicki & Lightman, 2008; Kaastra et al., 2008; Longair, 2011).

#### 3.3.1 Line Emission

When an ion is excited by an interaction of another particle, usually an electron, then it returns to the ground state by emitting a photon of the appropriate energy. The emitted photons have discrete energies and make line spectra because the energy levels of bounded electrons are quantized. Hence, emissions caused by collisional excitations are called as line emissions or bound-bound emissions.

A line energy tells us a charge state and an atomic number of emitting ions. A line intensity reflects the ion and electron densities, which determine the probability of collisions between the ions and electrons,

although it also depends on the electron temperature reflecting the kinetic energy distribution. Therefore, we can obtain the plasma ionisation state from the line intensity ratio of different charge state ions, and also plasma motion from the redshift or blueshift of line energies.

For a hydrogen-like (H-like) ion that has only one bounded electron, the transition energy,  $E_{nm}$ , between the principle quantum number  $n$  and  $m$  is

$$E_{nm} \sim Z^2 R_y \left( \frac{1}{n^2} - \frac{1}{m^2} \right), \quad (3.3.1)$$

where  $Z$  is the atomic number of ion,  $R_y$  is the Rydberg constant ( $R_y = \frac{\alpha^2}{2} m_e c^2 \simeq 13.6$  eV),  $\alpha$  is the fine structure constant  $\alpha = \frac{e^2}{\hbar c} \simeq \frac{1}{137}$ ,  $c$  is the speed of light,  $m_e$  is the rest mass of electron,  $e$  is the unit of electron charge,  $\hbar = h/2\pi$  and  $h$  is the Planck constant. The H-like transitions are called Lyman series. Ly  $\alpha$ , Ly  $\beta$ , and Ly  $\gamma$  corresponds to  $2p \rightarrow 1s$ ,  $3p \rightarrow 1s$ , and  $4p \rightarrow 1s$ , respectively, where  $s$  and  $p$  mean orbital angular momenta  $l = 0, 1$ , respectively.

If an ion has two or more bounded electrons, the transition processes become more complicated. The fine structure, which determined by the spins and orbital angular momenta of electrons, makes a lot of fine energy levels and transition rules. For a helium-like (He-like) ion that has two bounded electrons, there are three types of strong emission lines in thermal plasmas of astronomical objects: a resonance line ( $1s2p \ ^1P_1 \rightarrow (1s)^2 \ ^1S_0$ ), inter combination lines ( $1s2p \ ^3P_{2,1} \rightarrow (1s)^2 \ ^1S_0$ ), and a forbidden line ( $1s2p \ ^3S_1 \rightarrow (1s)^2 \ ^1S_0$ ).

In X-ray observations except for lower energy band of grating highly-energy-resolved observations, lines from fine structures of an element cannot be resolved and appear as a line blended them because of the energy resolution of detector. The transitions of the principle quantum number  $n = 2 \rightarrow 1$ ,  $3 \rightarrow 1$ ,  $4 \rightarrow 1$  are called as K  $\alpha$ , K  $\beta$ , and K  $\gamma$ , respectively. For He-like ions, K  $\alpha$ , K  $\beta$ , and K  $\gamma$  are especially expressed as He  $\alpha$ , He  $\beta$ , and He  $\gamma$ , respectively.

### 3.3.2 Bremsstrahlung

In plasma, when an electron and an ion collide, the electron will mainly be deflected from its path since  $m_e \ll m_{\text{ion}}$  where  $m_e$  and  $m_{\text{ion}}$  are the mass of electron and ion, respectively. The accelerated electron motion causes an emission of photons, which is called bremsstrahlung or free-free emission.

We consider the collisions of electrons and ions whose charge number is  $Z$  in a plasma with the electron and ion densities  $n_e$  and  $n_{\text{ion}}$ . The total power of bremsstrahlung emission from single-speed electrons whose speed is represented with  $v$ , per unit volume  $V$ , angular frequency  $\omega$  is

$$\frac{dW(v, \omega)}{d\omega dV dt} = \frac{2^4 \pi e^6}{3 \sqrt{3} m_e^2 c^3} n_e n_{\text{ion}} Z^2 \frac{1}{v} g_{\text{ff}}(v, \omega), \quad (3.3.2)$$

where  $c$ ,  $e$ , and  $t$  are the light speed, the unit of electron charge, and time, respectively, and  $g_{\text{ff}}$  is known as a Gaunt factor.

The bremsstrahlung spectrum is determined by the electron velocity distribution of plasma. Bremsstrahlung from thermal plasma is called thermal bremsstrahlung. For thermal equilibrium plasma, the

electron distribution takes Maxwell-Boltzmann distribution, which is given as

$$P(v) = \left( \frac{m_e}{2\pi k_B T_e} \right)^{3/2} \exp\left(-\frac{m_e v^2}{2k_B T_e}\right), \quad (3.3.3)$$

where  $T_e$  and  $k_B$  are the electron temperature and the Boltzmann constant, respectively.

We next estimate the emission integrated over the thermal distribution. A moving electron must have an kinetic energy more than emitted photon energy  $h\nu$ , and thus, at least, the incident velocity must be

$$\frac{1}{2}m_e v^2 \geq h\nu, \quad (3.3.4)$$

where  $\nu$  is the emitted photon frequency. Therefore, the lower limit of velocity integration is set as

$$v_{\min} = \sqrt{\frac{2h\nu}{m_e}}. \quad (3.3.5)$$

This cut-off is called a photon discreteness effect.

We obtain the bremsstrahlung spectrum averaged over the electron velocities as

$$\frac{dW(T_e, \omega)}{d\omega dV dt} = \frac{\int_{v_{\min}}^{\infty} \frac{dW(T_e, \omega)}{d\omega dV dt} dP}{\int_{v_{\min}}^{\infty} dP} = \frac{\int_{v_{\min}}^{\infty} \frac{dW(T_e, \omega)}{d\omega dV dt} v^2 \exp(-m_e v^2 / 2k_B T_e) dv}{\int_{v_{\min}}^{\infty} v^2 \exp(-m_e v^2 / 2k_B T_e) dv}. \quad (3.3.6)$$

The Gaunt factor depend on electron velocities, thus the integration can be replaced to the velocity-averaged Gaunt factor  $\tilde{g}_{\text{ff}}(\nu)$ . Using  $d\omega = 2\pi d\nu$ , the emissivity of thermal Bremsstrahlung can be derived as

$$\frac{dW(T_e, \nu)}{d\nu dV dt} = \frac{2^5 \pi e^6}{3m_e c^3} \sqrt{\frac{2\pi}{3k_B m_e}} n_e n_{\text{ion}} Z^2 T_e^{-1/2} e^{-\frac{h\nu}{k_B T_e}} \tilde{g}_{\text{ff}}(\nu). \quad (3.3.7)$$

The emissivity can be written in CGS units,

$$\frac{dW(T_e, \nu)}{d\nu dV dt} = 6.8 \times 10^{-38} Z^2 n_e n_{\text{ion}} T_e^{-1/2} e^{-\frac{h\nu}{k_B T_e}} \tilde{g}_{\text{ff}} \quad (3.3.8)$$

The thermal bremsstrahlung spectrum has characteristic shape  $\propto T_e^{-1/2} e^{-\frac{h\nu}{k_B T_e}}$ , which includes a exponential cut off determined by the electron temperature. Since the shape of  $T_e^{-1/2}$ , the emission from a plasma whose electron temperature is between  $\sim 0.1$  keV and  $\sim 1$  keV contributes to soft X-ray bands as a continuum component. The intensity is proportional to  $T_e^{-1/2}$  that reflects the electron velocity and  $n_e n_{\text{ion}}$  due to the two-body collision of electron and ion. As a result, we can obtain  $k_B T_e$  and  $n_e n_{\text{ion}}$  in a plasma if we determine the shape and intensity of thermal bremsstrahlung spectrum in an X-ray observation.

### 3.4 Chandra X-Ray Observatory

In this section, we summarize the instruments whose observation data are used in the subsequent chapters. The information is based on ‘‘The Chandra Proposers’ Observatory Guide’’ (Chandra IPI team, 2018).

The *Chandra* X-ray Observatory was launched on 23rd July 1999 as one of NASA's four Great Observatories, and has been still operated observations. *Chandra* has performed the highest spatial resolution ( $<0'.5$ ) in X-ray telescopes. As Figure 3.4.1 shows, *Chandra* combines an remarkable high-resolution X-ray telescope, High Resolution Mirror Assembly (HRMA), and two types of imaging detectors, a charge coupled device (CCD) array, Advanced CCD Imaging Spectrometer (ACIS) and a microchannel plate array, High Resolution Camera (HRC). These detectors can perform high spatial resolution with wide energy band of 0.1–10 keV, and ACIS can simultaneously have moderate spectral resolution. We can observe in higher energy resolution by grating with High Energy Transmission Grating (HETG) or Low Energy Transmission Grating (LETG).

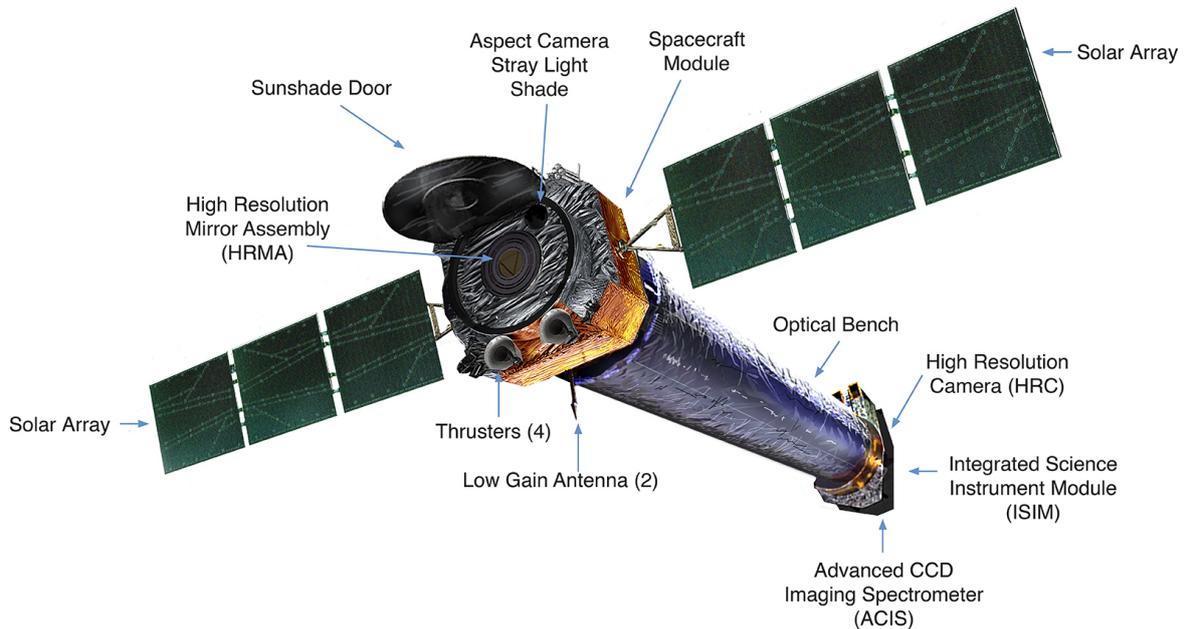


Figure 3.4.1: An illustration of entire *Chandra* spacecraft. (© NASA/CXC & J. Vaughan)

The HRMA, shown schematically in Figure 3.4.2, contains the nested Wolter Type-I mirrors, i.e., 4 pairs of parabolic mirrors and hyperboloid mirrors.

### 3.4.1 ACIS

ACIS (Garmire et al., 2003), which is a CCD-array detector located in the HRMA focal plane, can simultaneously acquire the incident position, time, and energy of each X-ray photon. Thus, an ACIS observation obtains both a high-resolution image and moderate resolution spectra. As shown in Figure 3.4.3, ACIS is comprised of two CCD arrays; a square array with  $2 \times 2$  chips, ACIS-I, and a sequential array with  $1 \times 6$  chips, ACIS-S. ACIS-I is mainly used in an imaging observation. By contrast, ACIS-S is used in both a imaging observation and a grating observation with LETG or HETG to obtain higher resolution spectrum. S1 and S3 chips in ACIS-S are backside-illuminated (BI) CCDs, and the other ACIS-S chips and all ACIS-I chips are frontside-illuminated (FI) CCDs.

Each CCD chip has  $8'.3 \times 8'.3$  field of view (FoV), and  $1024 \times 1024$  pixel format. Thus, ACIS-I and

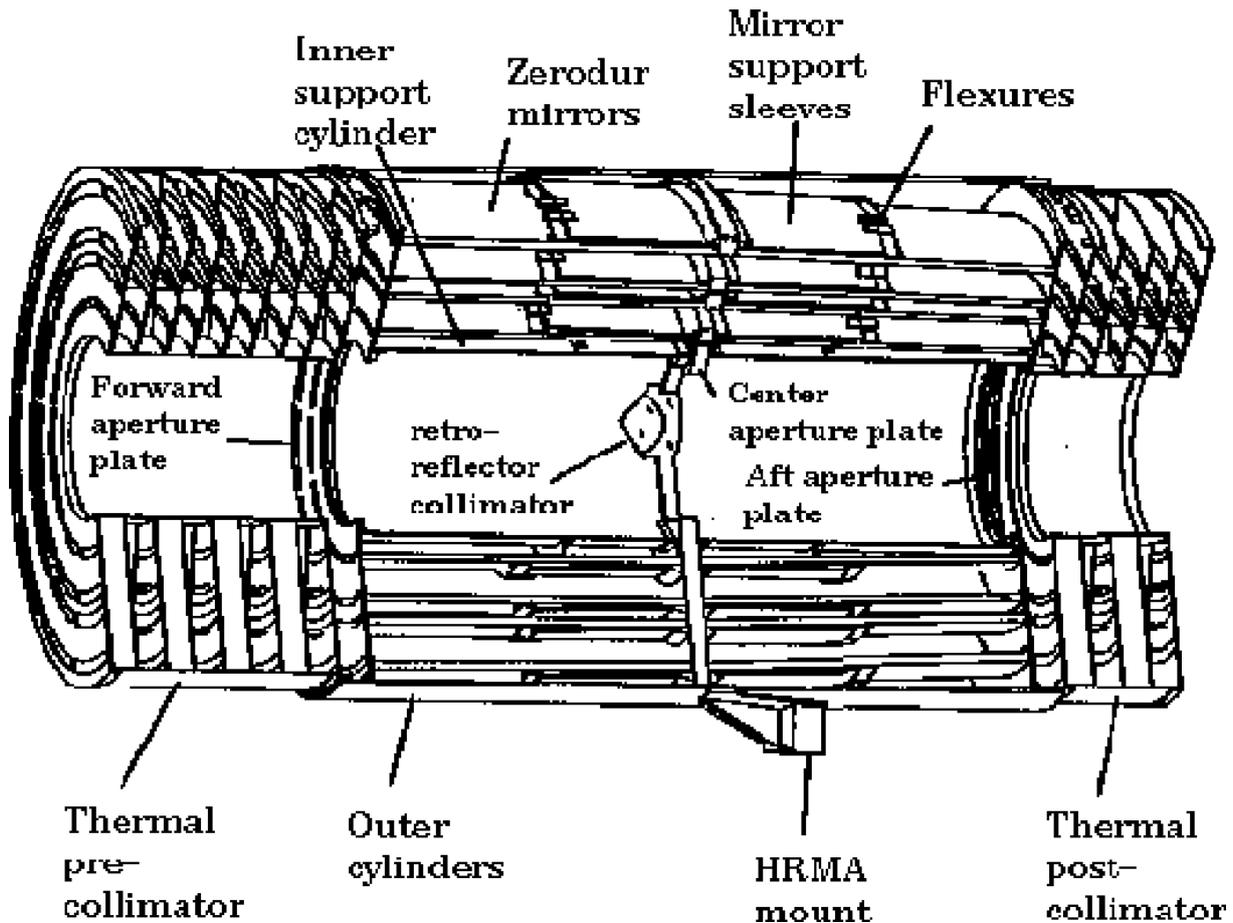


Figure 3.4.2: A schematic illustration of HRMA. This image is taken from Chandra IPI team (2018).

ACIS-S have FoV of approximately  $16' \times 16'$ , and  $8' \times 50'$ , respectively. There are approximately  $11''$  gaps between CCD chips, however the gaps are filled some exposure because of the dithering. Unless specially requested, the spacecraft dithered in a Lissajous pattern during an observation because of the purpose described above and smoothing out pixel-to-pixel variation in the response.

### 3.4.2 Spatial Resolution

The HRMA point-spread function (PSF) is a function of radius and X-ray energy. As shown in the left panel of Figure 3.4.4, the 50% encircled energy fraction of on-axis PSF is below  $0.5''$  at the almost energy bands although the encircled energy radius increases with the X-ray energy. For high energies, the PSFs are broadened at small radii because focus of high energies does not coincide with HRMA common focus. Each of four right panels in Figure 3.4.4 shows the off-axis encircled energy radius as a function of off-axis angle on each ACIS-I chip. The off-axis encircled energy radius on each ACIS-I chip is axially asymmetric with respect to the HRMA optical axis, because the HRMA aim point is located near the inner corner of ACIS-I3 chip. In each chip, the encircled energy radius increases, i.e., the PSF broadens, as the off-axis angle increases. Nevertheless, the off-axis 90% encircled energy radii in high-energy X ray (6.40 keV) are below  $15''$ .

The ACIS pixel size corresponds to  $0.4920''$ . The spatial resolution for on-axis is limited by the CCD

# ACIS FLIGHT FOCAL PLANE

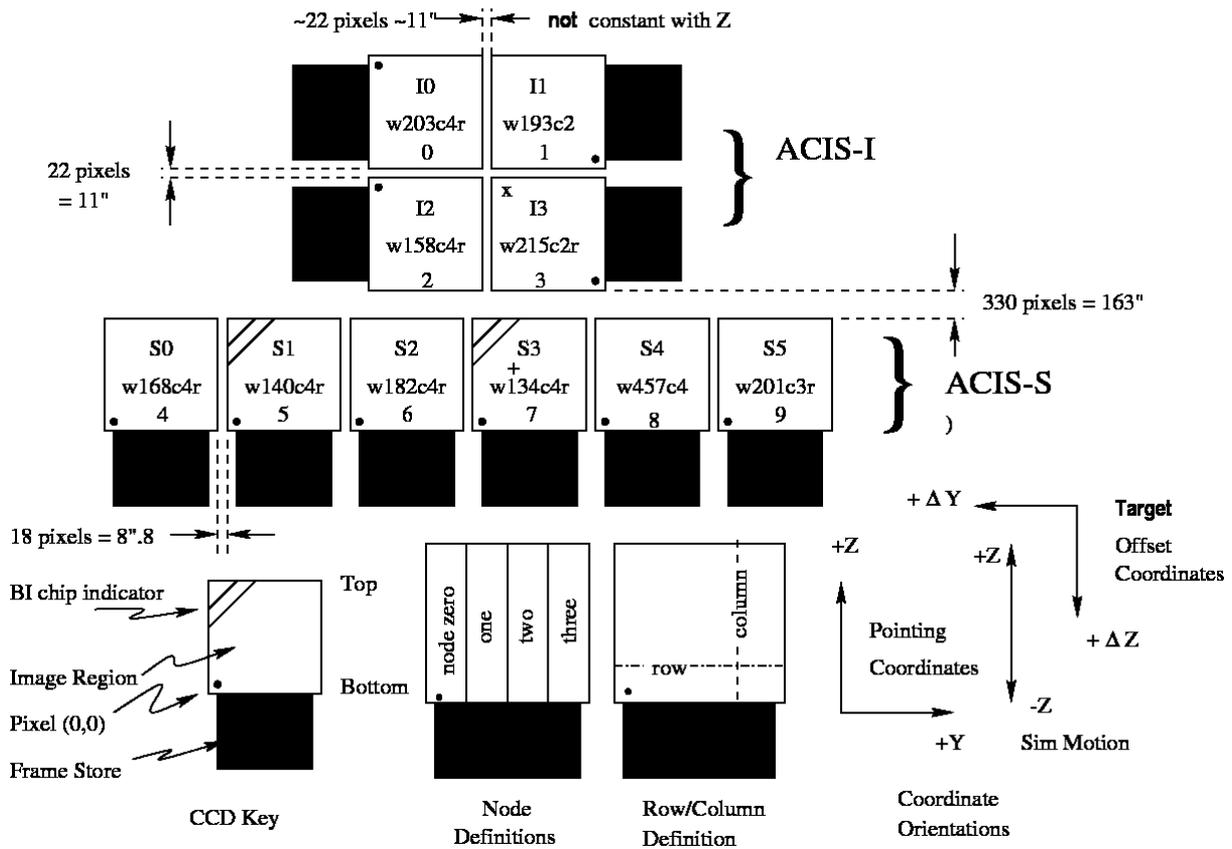


Figure 3.4.3: A schematic drawing of the ACIS focal plane, taken from Chandra IPI team (2018).

pixel size. On the other hand, the HRMA PSF increases with off-axis angle and becomes dominating factor for off-axis.

In the both analyses in Chapter 5 and Chapter 6, we defined the spatial bin sizes of  $3''.94 \times 3''.94$ . As shown in the left panel in Figure 3.4.4, more than 90% photons are expected to remain in the on-axis spatial bin in 6.4 keV or less energy bands. As shown in the four right panels in Figure 3.4.4, the 90% encircled energy radii are approximately  $2''$  and  $3''$  in 1.49 keV and 6.40 keV, respectively, at the off-axis angle of  $4'$  approximately corresponding to the apparent radius of *Tycho*'s SNR. Therefore, more than 90% of photons are expected to maintain in each spatial bin over the entire *Tycho*'s SNR in the low-energy band. By contrast, for *Kepler*'s SNR, the 90% encircled energy radii are approximately  $2''$  in 6.40 keV at the off-axis angle of  $2'$  approximately corresponding to the apparent radius. Thus, the photon leakage from or to the neighbour spatial bins are negligible over the entire *Kepler*'s SNR, even in high-energy bands.

### 3.4.3 Effective Area

The HRMA unobscured geometric aperture is  $1145 \text{ cm}^2$ . The support structures obstruct less than 10% of the HRMA aperture. The HRMA effective area is derived from the ray-trace simulation along

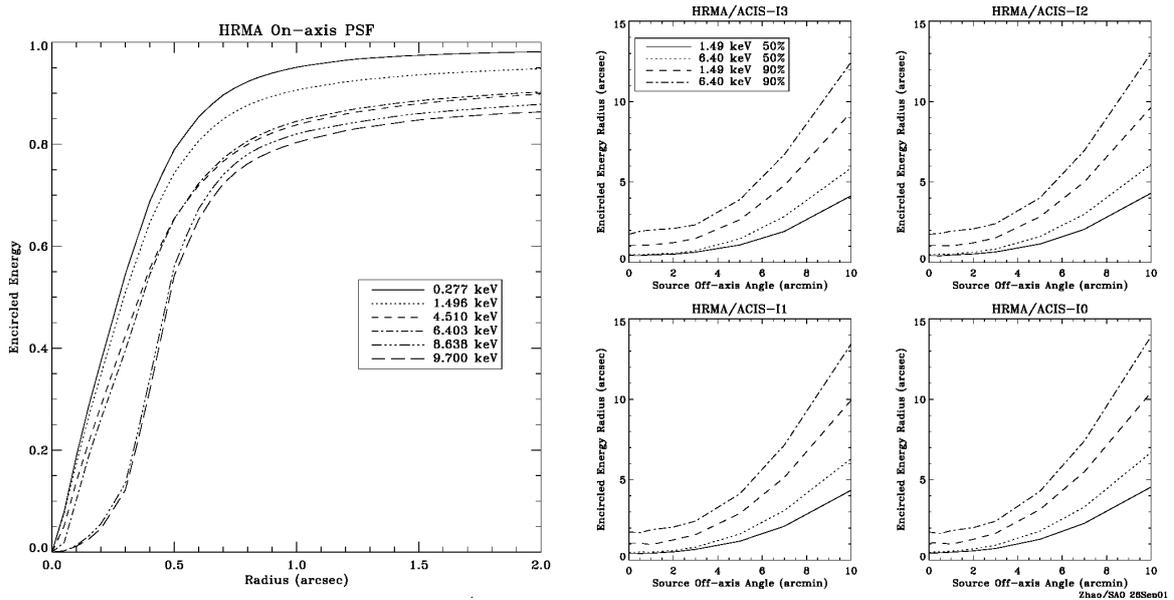


Figure 3.4.4: The *left panel* shows the encircled energy fraction as a function of angular radius for an on-axis point source and for different X-ray energies. The *right panel* shows the dependence of encircled energy radii for circles enclosing 50% and 90% at 1.49 keV and 6.40 keV on off-axis angle on the four ACIS-I chips. This image is taken from Chandra IPI team (2018).

with empirical corrections based on the ground calibration data.

The left panel of Figure 3.4.5 shows the on-axis HRMA effective area and the products of the effective area and quantum efficiency (QE) of ACIS or HRC. The effective area of HRMA/ACIS are above  $50 \text{ cm}^2$  in the energy range of 1.0–9.0 keV. The difference of effective areas between HRMA/ACIS-S3 and HRMA/ACIS-I3 is caused by the difference of QEs between BI and FI; the QE of BI is higher than FI for low energy, while QE of FI is higher than BI for high energy.

The right panel of Figure 3.4.5 shows the off-axis vignetting effect on the HRMA effective area. The effective area maintains more than 85% of on-axis area at off-axis angle  $4'$ , which approximately corresponds to the angle from the aim point to edge in ACIS-S3 chip, or more than 60% at  $8'$ , which approximately corresponds to the angle from the aim point to chip edge of ACIS-I, for the X-ray energy above 8 keV.

In the FI chips, the QEs vary with position and decrease by 5–15% furtherest from the read-out at energies above  $\sim 4$  keV, because of CTI. By contrast, the QE variation with row number is much smaller in the BI chips.

Optical blocking filter (OBF), which is a polyimide plate put between two thin aluminium layers is placed between each ACIS array and HRMA in order to shut optical photons out. The out-gassed molecular build on the cold ACIS OBFs, thus the ACIS effective area below 2 keV has continuously decreased. The optical depth of the molecular contaminant of the ACIS OBFs is shown in Figure 3.4.6.

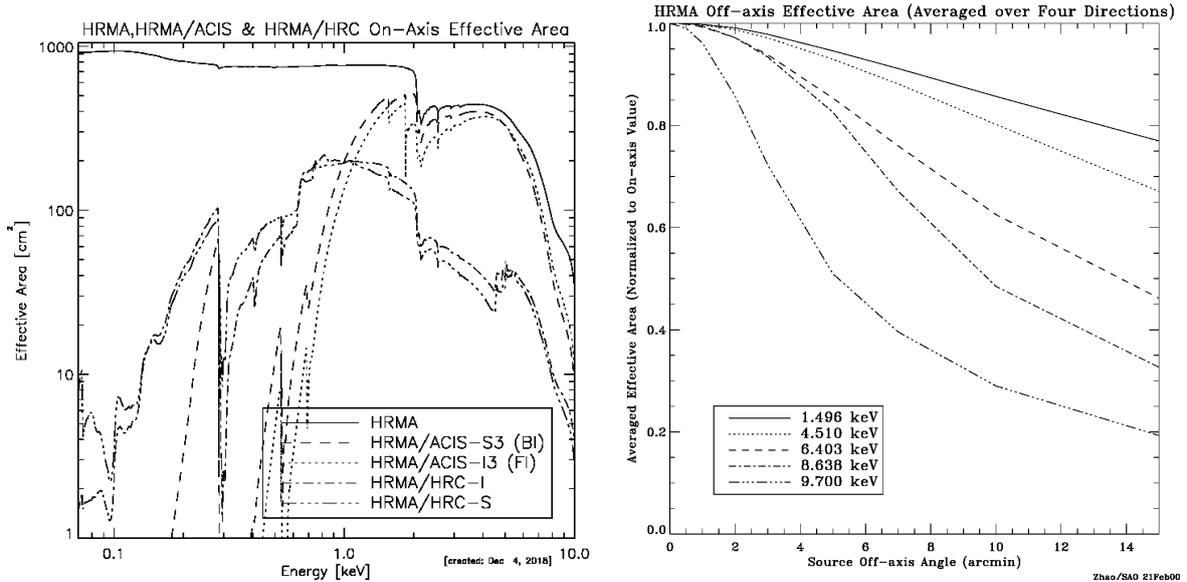


Figure 3.4.5: The *left panel* shows the on-axis HRMA/ACIS effective areas as functions of X-ray energy. The HRMA effective area is calculated by the ray-trace simulation and scaled by the ground-based calibration data. The effective areas are products of HRMA effective area and ACIS or HRC quantum efficiencies. The structure near 2 keV is due to the iridium M-edge. The *right panel* shows the HRMA off-axis effective area as a function of off-axis angle for selected energy. The effective area is averaged over four azimuthal directions and is normalized to the on-axis area. These images are taken from Chandra IPI team (2018).

### 3.4.4 Energy Resolution

The ACIS energy resolution varies roughly as the square root of the X-ray energy. As shown in the left panel of Figure 3.4.7, the ACIS FI CCDs originally approached the theoretical limit for the energy resolution at almost all energy band. After the launch, the CCD chips were encountered by low-energy protons ( $\sim 100$  keV) reflected through the X-ray telescope onto the focal plane during radiation belt passages. As the result of energy deposits on the buried channels at the HRMA side of the FI chips by low-energy protons, the energy resolution of FI chips has decrease and become a function of the row number due to increased charge transfer inefficiency (CTI). By contrast, BI chips remain nearly at pre-launch values, because the buried channels and gates face in the direction opposite to HRMA and approximately  $40 \mu\text{m}$  silicon protected them from low-energy protons. After the degradation, the operation was changed to move ACIS to a sheltered position (i.e., no detector remains in the focal position) during radiation belt passages thus no further degradation has been encountered. The CTI has been managed to cool the focal plane temperature  $-120^\circ\text{C}$ . As shown in the middle and right panels of Figure 3.4.7, much of the lost energy resolution of FI chips has been recovered due to a CTI correction algorithm, which was developed by the ACIS instrumental team and is supplied in the CIAO tool.

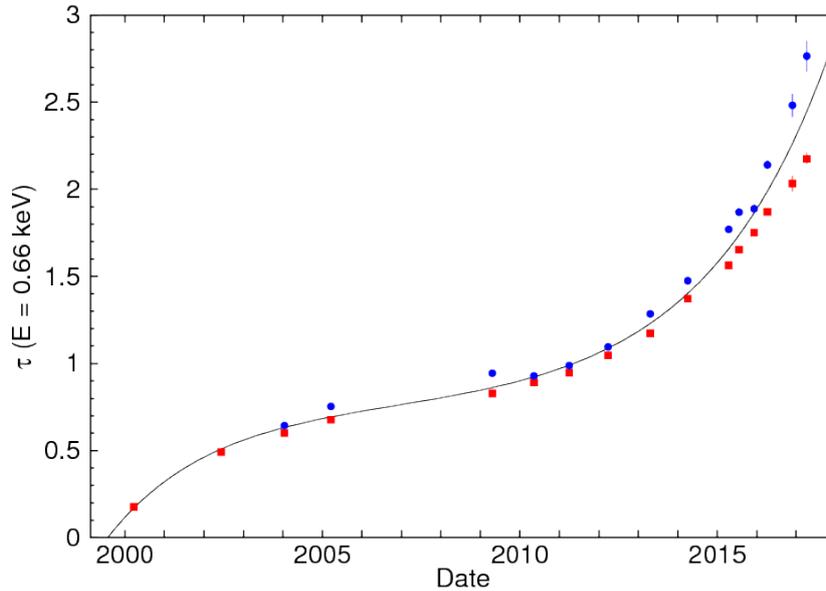


Figure 3.4.6: Optical depth of the contaminant of the ACIS OBFs measured at 0.66 keV from observations of the galaxy cluster A 1795, taken from Chandra IPI team (2018). Red squares and blue circles denote observations with the target on ACIS-S and ACIS-I, respectively. The solid curve represents the best-fit relation to the data.

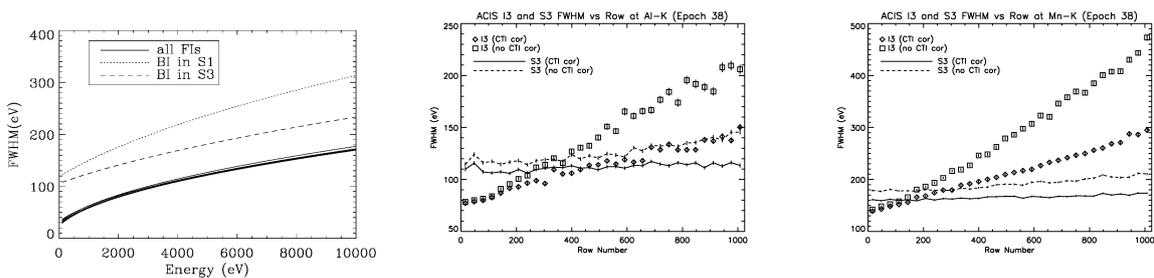


Figure 3.4.7: The *left panel* shows the pre-launch energy resolution (FWHM) of all the ACIS FI and BI chips as a function of energy. The *middle and right panel* show the energy resolution (FWHM) of all the ACIS FI and BI chips as a function of row number (CHIPY), taken from Chandra IPI team (2018). The *middle panel* shows Al K  $\alpha$  (1.49 keV), while the *four right panels* show Mn K  $\alpha$  (5.9 keV). The data points represent I3 chip with or without CTI correction, while the lines represent S3 chip. The data were taken from 2009 May through 2009 July, on I3 node 3, and S3 node 0, in which the aimpoints are.

## Chapter 4

# New Method Using Deep Learning

### 4.1 Our Method Applied for X-Ray Data of SNRs

In this work, we aim to categorise individual spatial bins of an X-ray image using only the spectral properties. We expect to automatically extract unknown features by a machine learning method. In this case, supervised machine learning methods are not applicable. We developed a method that combines two unsupervised learning methods, the variational autoencoder (VAE; Kingma & Welling, 2013) and the Gaussian mixture model (GMM), for spatially resolved spectroscopy of the X-ray data obtained by *Chandra* from supernova remnants (SNRs) (Iwasaki et al., 2019a,b).

Fine energy bins can retain rich information in spectra, while broader energy bins, e.g., bins corresponding line blends of individual element (e.g.; Warren et al., 2005; Burkey et al., 2013), lose information such as ionization states and Doppler shifts. In the near future, high-energy-resolution spectra obtained by a microcalorimeter require finer energy binning to maintain the rich information. However, clustering of the raw dataset, which has large number of energy bins, is difficult.

Although an energy spectrum has large number of energy bins, plasma models can generate a spectrum with smaller number of parameters. For example, APEC, which is an emission model for plasma in collisional ionization equilibrium, has only parameters of electron temperature, emission measure, abundances of individual elements, and redshift. In addition NEI, which is a model for non-equilibrium ionization plasma, has the same parameters of APEC and also only one more parameter, ionization time scale. Therefore, dimensionality reduction to extract essential parameters from data is a reasonable strategy.

Energy bins in a spectrum emitted from thermal plasma have nonlinear relations to each other. Therefore, it is also reasonable to select a dimensionality reduction method that is capable to capture nonlinear features.

A deep neural network (DNNs) summarized in Chapter 2 is one of models that are able to represent nonlinear features. VAE whose architecture is constructed with DNN is a viable technique for dimensionality reduction. As a merit of using VAE, the decoder of VAE can artificially generate samples, and thus we can interpret the latent parameters, which have lower dimension than original dataset, using the decoder. Most of manifold learning algorithms (e.g., t-SNE, isomap) cannot generate samples, and do

not allow interpretation in such a way.

Similar methods in which dimensionality reduction and unsupervised clustering are integrated have already been applied in several astronomical analyses, e.g., classification of galaxy spectra (principal component analysis (PCA) & GMM; Hurley et al., 2012), classification of supernova (autoencoder (AE), isomap & k-means; Ishida et al., 2017) or separation of galactic and extragalactic objects (AE & support vector machine (SVM); Khramtsov & Akhmetov, 2018). Karmakar et al. (2018) applied the method using GMM with VAE for stellar cluster detection, and the method determined the stellar cluster regions as accurately as other state-of-the-art results. They used monochrome image data of an infrared Galactic plane survey. They reduced the dimension of flattened patches in a monochrome observational image whose shape is a vector of  $1 \times l^2$  for an image patch of  $l \times l$  by some VAEs constructed with only fully-connected (FC) layers, and then classified the patches in the latent space into two classes (i.e., stellar cluster region and background region) by GMM. Although they used the same machine learning models as we used, the method is different from our method. They focused on the binary classification, i.e., stellar cluster or the others, while we are interested in the spectral variation in an astronomical object. The data type is also quite different, where their method focused only on the local spatial information, while our method uses spectral information in each spatial points.

In this thesis, we explore the method of combining nonlinear dimensionality reduction by a VAE and clustering by a GMM for automatic investigation of the spatial structures of a diffuse object for the first time. The brief description of our method is following. (1) We train VAE using training dataset, and check the performance using validation dataset. (2) We obtain latent variables, whose dimensions are reduced from the original data using VAE by inputting a post-training dataset. (3) We apply GMM to the latent variables, and obtain a clustering map. After this method, we can perform spectral analyses in ordinary way using regions appearing in the clustering map, and can obtain information of physical processes.

For VAE training, we use a dataset randomly divided into training set and validation set, which are pre-processed, e.g., normalization. In order to select the best VAE architecture, we train some VAE models having different architectures from each other, i.e., number of latent axes, number of layers, and number of nodes in each layer. In each epoch of training, VAE weights are updated using a cost function for the training data, and then the cost function for the validation data is calculated in order to check the overfitting. After the training, we select the best VAE architecture using the validation cost of each trained VAE model. Finally, we input post-training data into the best VAE model, and obtain the latent expression as a result of dimensionality reduction by VAE.

After the VAE processing, we apply GMM to the latent variables extracted from the post-training dataset by the VAE. The number of categories is optimized by checking Bayesian information criterion (BIC).

We describe the detail of each machine learning model in our method below.

### 4.1.1 Variational Autoencoder

Although an X-ray spectrum of, e.g., an SNR, may have a large number of energy bins, they are not completely independent of each other because of the finite instrumental energy resolution and underlying emission process. The number of essential parameters generating the spectrum might be small as described above. Therefore, before classifying the spectrum, we reduced the dimensionality by compressing the information in the raw input spectra. To capture nonlinear relationships between the essential parameters, we employed an unsupervised DNN architecture, namely, a VAE.

An AE (Baldi & Hornik, 1989; Hinton, 1990; Hinton et al., 2011; Hinton & Salakhutdinov, 2006) is a DNN architecture connecting an encoder and a decoder. The encoder is trained to encode the input data as latent variables  $z$  by reducing the input dimensions. At the same time, the decoder is trained to reproduce the original input from the latent variables  $z$ . The dimension of the latent space is smaller than that of the original space. One can thus obtain a compressed latent expression of the original data if the AE is successfully trained. The latent space is expected to capture nonlinear relationships in the input data because of the capability of a DNN.

The VAE is a variant of the AE. In the VAE, a multidimensional Gaussian distribution is assumed for the latent variables. Unlike a normal AE, which computes the latent variables directly, the encoder of a VAE computes the means  $\mu$  and variances  $\sigma$ . A set of latent variables  $z$  is sampled from a multidimensional Gaussian whose means  $\mu$  and variances  $\sigma$  are calculated by the encoder. The decoder decompresses the set of latent variables  $z$ . We will describe the theoretical view later.

VAE models have some advantages over normal AEs, e.g., more stable training and a better latent manifold structure (Tolstikhin et al., 2017). Before the study, we compared normal AEs to the VAE and found that training of the normal AE is sometimes unstable and that in many cases, some latent variables do not represent any features. Therefore, we chose the VAE for this study.

Generative models including VAEs try to reproduce the generative process of the dataset, which contains some latent parameters  $z$ , although the true process and parameters cannot be observed. The generative models like VAEs and GANs try to minimize discrepancy between the data and model distributions. However the most standard divergences are intractable, especially when the data distribution is unknown and the model is parametrized by DNNs. The previous research provides some tricks to address the issue. VAE models try to maximize the marginal log-likelihood  $\ln p_{\theta}(\mathbf{x})$ . In this case, the variational lower bounds can be used as described below.

VAE realizes the process  $p_{\theta}(z)p_{\theta}(\mathbf{x}|z)$  where values  $z$  are generated from some prior  $p_{\theta}(z)$  and then data samples  $\mathbf{x}$  are generated from some conditional distribution  $p_{\theta}(\mathbf{x}|z)$ . However, the marginal likelihood  $p_{\theta}(\mathbf{x}) = p_{\theta}(z)p_{\theta}(\mathbf{x}|z)/p_{\theta}(z|\mathbf{x})$  is intractable. In order to solve it, an approximate posterior  $q_{\phi}(z|\mathbf{x})$  is introduced as an approximation to the intractable true posterior  $p_{\theta}(z|\mathbf{x})$ . In an AE architecture, the encoder acts as a recognition model  $q_{\phi}(z|\mathbf{x})$  to compute  $z$  from given data samples  $\mathbf{x}$ , and then the decoder  $p_{\theta}(\mathbf{x}|z)$  generates  $\mathbf{x}$  from the codes  $z$ . Through training, both recognition model (encoder) parameters  $\phi$  and generative model (decoder) parameter  $\theta$  will be learned.

For VAE latent variables, the prior takes on the isotropic multidimensional Gaussian  $p_{\theta}(z) =$

$\mathcal{N}(z; \mathbf{0}, \mathbf{I})$ , where  $\mathbf{I}$  is an identity matrix. As an approximation of the intractable true posterior  $p_{\theta}(z|x^{(i)})$ , a multidimensional Gaussian form with a diagonal covariance is assumed:

$$q_{\phi}(z|x^{(i)}) = \ln \mathcal{N}(z; \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{(i)2}\mathbf{I}), \quad (4.1.1)$$

where  $\boldsymbol{\mu}^{(i)}$  and  $\boldsymbol{\sigma}^{(i)}$  are the mean and standard deviation of the approximate posterior for the  $i$ th data sample (Kingma & Welling, 2013).

In the implementation of VAE, the encoder branches after the FC layers and connects to a layer of  $\boldsymbol{\mu}$  and  $\boldsymbol{\sigma}$ , which have the same number of nodes as the dimension of  $z$ . The latent variables  $z$  are sampled as  $z = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$  using the *reparametrization trick* (Kingma & Welling, 2013; Jimenez Rezende et al., 2014). With  $\odot$  we signify an operator of element-wise product, which is called Hadamard product. Here  $\boldsymbol{\epsilon}$  is a vector consisting of random numbers sampled from a Gaussian distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ .

Kingma & Welling (2013) describes the marginal likelihood of the  $i$ th data point can be written as

$$\ln p_{\theta}(x^{(i)}) = D_{\text{KL}}(q_{\phi}(z|x^{(i)})||p_{\theta}(z|x^{(i)})) + \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; x^{(i)}). \quad (4.1.2)$$

The first term is the KL divergence between the approximate and true posteriors. The KL divergence always takes a non-negative value, thus  $\ln p_{\theta}(x^{(i)}) \geq \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; x^{(i)})$  and the second term  $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; x^{(i)})$  is called the variational lower bound, or the evidence lower bound (ELBO), on the marginal likelihood of the  $i$ th data point. As a result, the problem to maximize the log-likelihood  $\ln p_{\theta}(x^{(i)})$  can be changed to the problem to maximize the lower bound.

The lower bound can be written as (Kingma & Welling, 2013)

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; x^{(i)}) = -D_{\text{KL}}(q_{\phi}(z|x^{(i)})||p_{\theta}(z)) + \mathbb{E}_{q_{\phi}(z|x^{(i)})}[\ln p_{\theta}(x^{(i)}|z)]. \quad (4.1.3)$$

VAE models try to find the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$  to optimize the variational lower bound. The optimization problem results in that the decoder tries to accurately reconstruct the data samples with the penalty of the second term. Meanwhile, the encoder tries to simultaneously obtain two conflicting issues; (1) it tries to match the encoded distribution  $q_{\phi}(z|x^{(i)})$  to the prior  $p_{\theta}(z)$  as measured by the KL divergence of the first term, (2) it makes sure that the latent codes provided to the decoder are informative enough to reconstruct the data samples as shown in the second term.

For VAE whose both the prior  $p_{\theta}(z)$  and the  $q_{\phi}(z|x^{(i)})$  are Gaussian distribution, the KL divergence in Eq. 4.1.3 can be calculated as

$$D_{\text{KL}}(q_{\phi}(z|x^{(i)})||p_{\theta}(z)) = -\frac{1}{2} \sum_{j=1}^J (1 + \ln(\sigma_j^{(i)})^2 - (\mu_j^{(i)})^2 - (\sigma_j^{(i)})^2), \quad (4.1.4)$$

where  $\mu_j^{(i)}$  or  $\sigma_j^{(i)}$  is the  $j$ th element of  $\boldsymbol{\mu}^{(i)}$  or  $\boldsymbol{\sigma}^{(i)}$  (Kingma & Welling, 2013).

The second term of Eq. 4.1.3 is the expectation of log-likelihood  $\ln p_{\theta}(x^{(i)}|z)$  based on the latent code calculated by  $q_{\phi}(z|x^{(i)})$  for the  $i$ th data point. This is, in other word, the reconstruction loss of the decoder based on the probability distribution of the latent code  $z$  predicted by the encoder. Depending on the type of data, the reconstruction loss can be modelled using Gaussian, Poisson, or Bernoulli distributions for real-valued data, positive- and discrete-valued data, or binary data, respectively; we describe this in detail below.

If we assume that the data are generated from a Gaussian random process, the data point  $\mathbf{x}$  is sampled from a Gaussian distribution whose mean takes the predicted variable  $\mathbf{y}$  and standard deviation is  $\sigma$ . In this case, the reconstruction loss can be written as

$$\begin{aligned}\mathcal{L}_{\text{Gauss}} &= \ln \mathcal{N}(\mathbf{x}^{(i)}; \mathbf{y}^{(i)}, \sigma^{(i)2}\mathbf{I}) \\ &= \ln \frac{1}{\sqrt{2\pi}|\sigma^{(i)}|} \exp\left(-\frac{(\mathbf{y}^{(i)} - \mathbf{x}^{(i)})^2}{2\sigma^{(i)2}\mathbf{I}}\right) \\ &= \sum_{d=1}^D \frac{(y_d^{(i)} - x_d^{(i)})^2}{2\sigma_d^{(i)2}} + \text{const.}\end{aligned}\quad (4.1.5)$$

where  $y_d^{(i)}$  is the  $d$ th value of the decoder output  $\mathbf{y}^{(i)}$  corresponding to the  $i$ th input  $\mathbf{x}^{(i)}$  (Bishop, 2006). If we assume that each data point is sampled with a Gaussian noise whose mean is 0 and standard deviation is 1, the reconstruction loss is simplified by ignoring the standard deviation as

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}^{(i)} - \mathbf{x}^{(i)})^2 \quad (4.1.6)$$

where  $N$  is the number of data samples (Bishop, 2006). This log likelihood is the mean square error (MSE), which is used the least squared method and also many machine learning tasks.

If the data is positive and discrete values, which are sampled by Poisson random process, the reconstruction loss is written as

$$\begin{aligned}\mathcal{L}_{\text{Poisson}} &= \sum_{d=1}^D \ln \frac{e^{-y_d^{(i)}} y_d^{(i)x_d^{(i)}}}{x_d^{(i)}!} \\ &= \sum_{d=1}^D [-y_d^{(i)} + x_d^{(i)} \ln y_d^{(i)} - \ln(x_d^{(i)}!)] \\ &\simeq \sum_{d=1}^D (-y_d^{(i)} + x_d^{(i)} \ln y_d^{(i)} - x_d^{(i)} \ln x_d^{(i)} + x_d^{(i)} - \frac{1}{2} \ln x_d^{(i)})\end{aligned}\quad (4.1.7)$$

where  $y_d^{(i)}$  is the  $d$ th value of the decoder output  $\mathbf{y}^{(i)}$  corresponding to the  $i$ th input  $\mathbf{x}^{(i)}$  (Cash, 1979; Ichinohe & Yamada, 2019). Each bin in an X-ray spectrum represents the number of photons, which are Poissonian variables. Each bin of X-ray spectra finely resolved in energy, space, or time contains a small number of photons, and thus, Poisson log likelihood is appropriate in such a case. A Poisson distribution asymptotes to a Gaussian distribution for sufficiently large values. Therefore, the Gaussian log likelihood or MSE can be an alternative of the Poisson loss if the data consists only of large numbers.

For the binary-valued data sampled from a Bernoulli distribution, the reconstruction loss can be written as

$$\begin{aligned}\mathcal{L}_{\text{Bernoulli}} &= \ln \prod_{i=1}^N p(c_1|\mathbf{x}^{(i)})^{y^{(i)}} (1 - p(c_1|\mathbf{x}^{(i)}))^{1-y^{(i)}} \\ &= \sum_{i=1}^N y^{(i)} \ln p(c_1|\mathbf{x}^{(i)}) + (1 - y^{(i)}) \ln(1 - p(c_1|\mathbf{x}^{(i)}))\end{aligned}\quad (4.1.8)$$

where  $p(c_1|\mathbf{x}^{(i)})$  is the probability to predict the category  $c_1$  for an input  $\mathbf{x}^{(i)}$  (Bishop, 2006). This is a popular loss function for classification tasks of machine learning, namely a cross entropy between a data distribution and a model distribution.

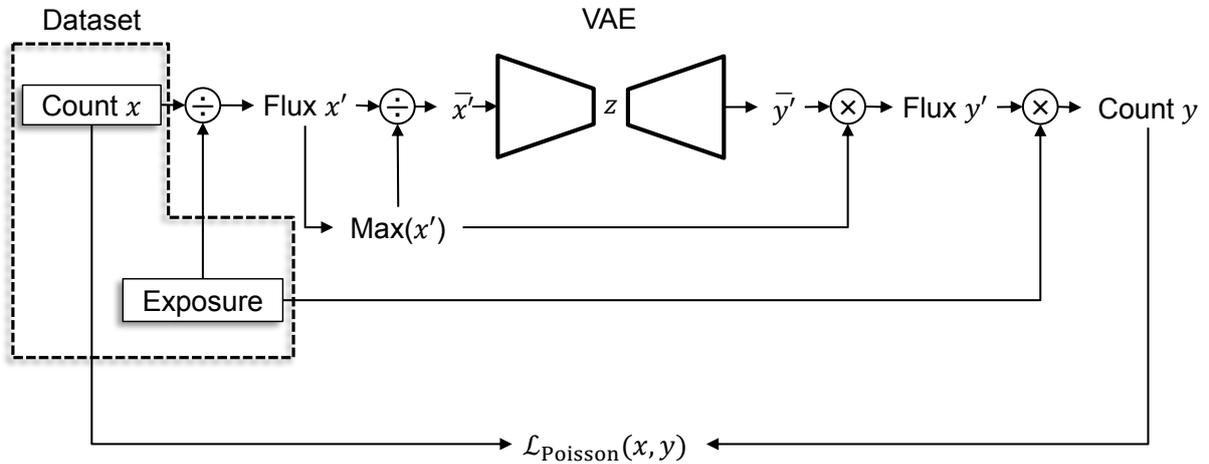


Figure 4.1.1: Diagram of the pre- and post-processing for the VAE. See the main text for the details of the network.

#### 4.1.2 VAE Architecture Using Poisson Reconstruction Loss

Each bin of an X-ray spectrum represents a number of photons, thus Poisson log likelihood is appropriate to the reconstruction loss in this case. Ichinohe & Yamada (2019) implemented the Poisson log likelihood for the VAE reconstruction loss to reconstruct simulated X-ray spectra whose spectral bins contain numbers of photons. However, in real observational data, the count in each spatial and spectral bin depends not only on the photon flux from the SNR but on the exposure time, the effective area, and spatial- and spectral-bin sizes. We want the VAE to learn only the information of the SNR emission. Further, the variables should be scaled to  $\sim 1$  for efficient calculation with a DNN.

In order to apply the VAE using a Poisson reconstruction loss of Eq. 4.1.7 to the observational data, we newly modified the input and output of VAE as shown in Figure 4.1.1. We prepared a dataset consisting of count maps and exposure maps, which consist images of the effective area at each sky point. Each data sample is converted into a photon flux spectrum by dividing the count spectrum by the exposures, and then is fed to the input of the VAE. The output is re-converted into the count spectrum by multiplying the output by the exposures. This manipulation makes the VAE accept and reconstruct photon flux variables, which are independent of exposures, and simultaneously enables the Poisson log likelihood to be calculated using the input and output as count variables. Furthermore, each input as photon flux

is scaled by arbitrary values, e.g., the maximum of each sample, and then the output is rescaled by the same value.

### 4.1.3 Gaussian Mixture Model

After the spectral information was compressed into a several-dimensional latent expression using the VAE as explained in the previous section, we classified it using the GMM.

As described in Section 2.1, clustering algorithms are divided into two classes; hard and soft clustering. A soft clustering method is appropriate for this study for the following reasons. (1) In the latent variable coded by the VAE, the characteristic distribution is not necessarily separate, and components with apparently different trends overlap considerably, particularly around the value 0. It is generally difficult to draw clear boundaries when multiple components overlap. (2) The physical conditions change continuously throughout the SNR.

The GMM is a well-known soft clustering method that has become popular for astrophysical data analysis (e.g., Davoodi et al., 2006; Hurley et al., 2012; Burkey et al., 2013). It describes the data distribution as multiple multidimensional Gaussians; each Gaussian represents a clustering category. Every data point is represented by a weighted superposition of all the categories. The probability that a data point belongs to a certain category, which is also referred to as the *responsibility*, is represented by the ratio of the value of the Gaussian corresponding to the category to the sum of the values of all the Gaussians for this data point. We used the GMM GaussianMixture in scikit-learn 0.19.0 (Pedregosa et al., 2011), which is a Python library providing a machine learning framework.

### 4.1.4 t-SNE

There are some manifold learning algorithms, which can be used for dimensionality reduction without neural networks. If the data dimension is not large and can be compressed into a few dimension space, such techniques can be alternatives of VAEs.

t-distributed Stochastic Neighbour Embedding (t-SNE) (van der Maaten & Hinton, 2008) is a nonparametric dimensionality reduction method to learn a manifold in the data space. It visualizes similarity data that retaining the local structure of the data while also revealing some important global structure, such as clusters at multiple scales in high-dimensional data. It converts similarities between data points to joint probabilities using Gaussian distribution and Student t-distribution as a heavy-tailed distribution in the high-dimensional space and the low-dimensional map, respectively. It minimizes the KL divergence between the joint probabilities  $p_{ij}$  in the high-dimensional space and the joint probabilities  $q_{ij}$  in the low-dimensional space.

t-SNE has a cost function that is not convex depending on choice of several optimization parameters, and hence may construct a solution differently for each run with different initialization.

The standard t-SNE is a nonparametric algorithm, thus the trained model cannot be reused for new data points or other dataset. By contrast, parametric t-SNE is also presented (Van Der Maaten, 2009). In their experiments, parametric t-SNE outperformed other parametric dimensionality reduction techniques, PCA, neighbourhood components analysis (NCA; Goldberger et al., 2005), and AEs.

### 4.1.5 Hierarchical Clustering

Hierarchical clustering, which build a hierarchy of clusters, is one category of hard clustering methods. Hierarchical clustering algorithms have two classes; agglomerative and divisive. In agglomerative algorithms, each data point begins in the individual cluster, and the nearest pair of clusters are merged into one until all data points are contained in one cluster. In divisive algorithms, one cluster initially contains all data points, and continues to be splits into two clusters until each cluster contains only one data point.

Some agglomerative algorithms linking clusters have been developed, e.g., single linkage, complete linkage, group average, and Ward’s method.

The Ward’s method minimizes the variance of data points in each cluster (Ward Jr, 1963). In each linkage step, the method tries to minimize the difference between the variance of merged cluster and sum of variance of individual clusters before they are merged. When two clusters  $s$  and  $t$  are merged into a new cluster  $u$ , the distance between the new cluster and the another cluster  $v$  is written as <sup>1)</sup>:

$$\begin{aligned} d(u, v) &= \text{Var}(u \cup v) - \text{Var}(v) - \text{Var}(u) \\ &= \sqrt{\frac{|v| + |s|}{T} d(v, s)^2 + \frac{|v| + |t|}{T} d(v, t)^2 - \frac{|v|}{T} d(s, t)^2} \end{aligned} \quad (4.1.9)$$

where  $d$  is distance between two clusters,  $T = |v| + |s| + |t|$ , and  $|*|$  is the cardinality, i.d., the number of data points (Ward Jr, 1963). Euclidean distances are usually used for  $d(v, s)$ ,  $d(v, t)$ , and  $d(s, t)$ .

The Ward’s method is robust to outliers. The cluster tends to have a spherical shape because the variance is based on an isotropic Gaussian.

## 4.2 Demonstration of the Method with MNIST

We tested out unsupervised machine learning method combining the VAE and GMM with a handwritten digits dataset in order to demonstrate that the method can accurately classify a dataset, before we apply the method to astronomical data.

The MNIST database is the best-known dataset of handwritten digits in the machine learning filed. The database has fixed-size monochrome images of handwritten digits from 0 to 9. Each sample contains a  $28 \times 28$  image whose pixel value is 0–255, and a label (i.e., a correct answer of each image), which is an integer from 0 to 9. The training set has 60000 samples, and the test set has 10000 samples. We use the labels of the dataset only to measure the accuracy after the models are trained, because this is an unsupervised learning method. In this demonstration, the dataset in Keras (Chollet et al., 2015) was used.

**Data preprocessing:** The images were normalized with the maximum value 255 in order to limit the

---

<sup>1)</sup> The description of SciPy implementation is available in <https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html#scipy.cluster.hierarchy.linkage>.

values within a range of 0–1. Each image of 28×28 matrix was flattened to a 784 vector.

**VAE:** The VAE architecture was below: The encoder and decoder were constructed from two FC layers with 256 nodes per layer and the ReLU activation function, respectively. The shape of input and output layers are 784-d vectors. The output layer has a sigmoid activation function. We prepared models with latent dimensions ranging from 1 to 30 and trained them for 200 epochs. We selected the latent dimension by checking the corresponding loss value for the test data. There was little difference between cases with the latent dimension more than 12, thus we used the latent dimension 13 for the clustering.

We used the deep learning framework Keras 2.0.7 (Chollet et al., 2015) with the TensorFlow 1.3.0 (Abadi et al., 2015) backend. The binary cross-entropy was used as the loss function. Nesterov-accelerated adaptive moment estimation (Nadam; Dozat, 2016) (which, according to our tests, provides the fastest convergence among the optimizers) was used for optimization. The training was performed for 200 epochs with a batch size of 100.

**GMM:** We then classified the latent parameters  $\mu$  for test data using GMM. The number of category was fixed to the number of class contained in the dataset, ten (i.e., the number from 0 to 9). We used the GaussianMixture in scikit-learn 0.19.0 (Pedregosa et al., 2011), which is a Python library providing a machine learning framework.

The GMM category IDs, which have no a priori meaning, were sorted in the order of the data labels.

In Figure 4.2.1, the first four axes of the latent  $\mu$  are shown, where the colour of each point corresponds to the GMM category with the highest responsibility for each sample of the test data.

**Results:** We evaluate the clustering result with the *accuracy*, i.e., (the number of correct predictions)/(the total number of predictions). The accuracy is 90.87% for all test data when the VAE model with latent dimension 13 was used. The confusion matrix of GMM is shown in Figure 4.2.2.

The accuracies for the individual classes 0, 2, 4, and 6 were more than 95%, and for 1, and 9 more than 90%, respectively. On the other hand, the accuracies for 5, 7, and 8 were lower than the others, approximately 80%. The 19.6% of 5 was classified to 3; the 14.7%, 4.5% of 7 were classified to 2, and 9, respectively; the 3.4%, 8.7%, and 2.1% of 8 were classified to 2, 3, and 5, respectively.

By contrast, when we classified the MNIST dataset using GMM clustering without a dimensional reduction by a VAE, the accuracies were 50% for the test data and training data. The comparison clearly shows that dimensional reduction by VAE is effective for GMM clustering of high-dimensional data.

In this demonstration, our method classified handwritten digits with the accuracy of more than 90%. Because of this result, our method is expected to be also applicable to astronomical data.

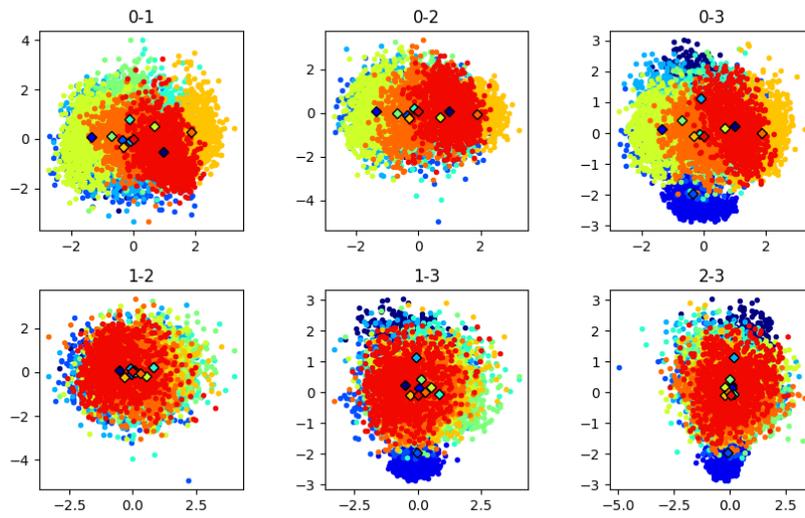


Figure 4.2.1: Scatter plot showing the  $\mu$  of VAE latent parameter obtained by inputting each digits image, where two of the thirteen axes are chosen. Each point was colour-coded for each category classified by GMM clustering. The centroid of each category is shown as an open black diamond.

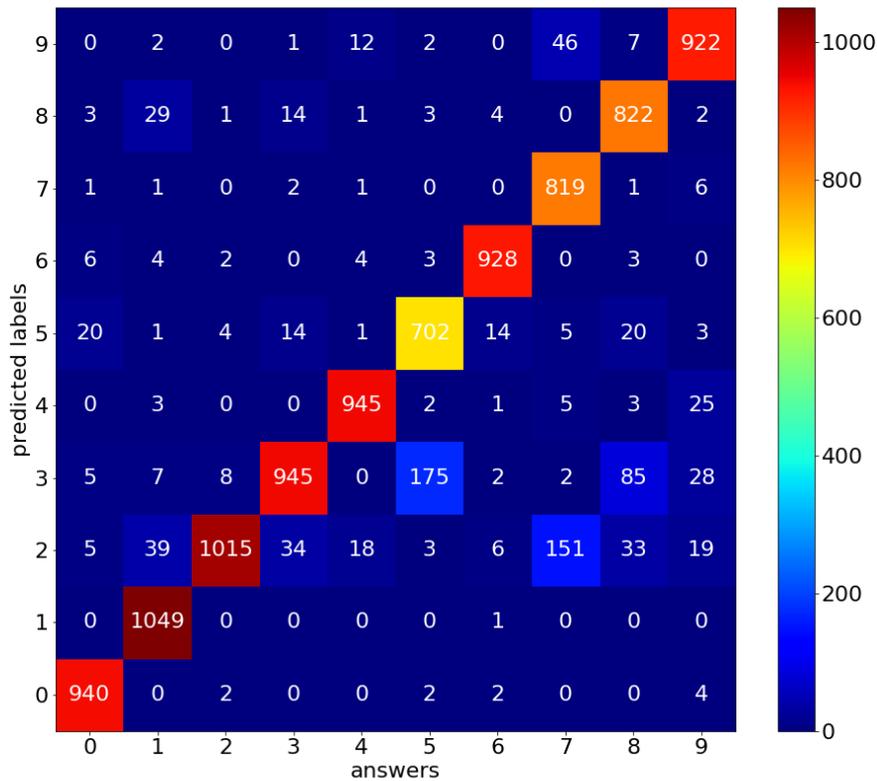


Figure 4.2.2: Confusion matrix of GMM for each answer label vs predicted class. The colour scale shows the number of samples corresponding category. The numbers of samples are also shown on the corresponding bins as texts.

## Chapter 5

# Demonstration with Tycho's SNR

Most of the contents in this chapter are taken from Iwasaki et al. (2019a).

### 5.1 Introduction for Tycho's SNR

X-ray spectra from supernova remnants (SNRs) are known to contain rich multidimensional information. We chose Tycho's SNR as a target for application of our machine learning method because this is one of the best studied SNRs.

Supernovae (SNe) explosively eject elements synthesised in the progenitor materials (so-called ejecta), forming blast waves. The resulting bright X-ray-emitting structures are called SNRs. As we review in Chapter 3, X-ray observations of SNRs allow us to investigate both the chemical evolution of the Universe and the mechanisms of cosmic-ray acceleration.

SNe have long been assumed to supply heavy elements synthesised during the explosion. The ejecta, that is, the X-ray-emitting hot plasma in SNRs, reveals the nuclear burning regimes of elemental synthesis.

*Tycho's* SNR is the remnant of SN 1572, which is known to be a type Ia explosion from the light-echo spectrum (Krause et al., 2008). In X-ray spectra of *Tycho's* SNR, line emission from intermediate-mass elements (IMEs; e.g., Si, S, Ar, and Ca) and Fe synthesised during the supernova explosion are clearly seen. In addition, secondary Fe-peak elements (e.g., Cr, Mn, and Ni, which are synthesised together with Fe) have been detected (e.g., Tamagawa et al., 2009). The global morphology of *Tycho's* SNR features radial gradation of the plasma ionization state and the electron or ion temperature, which are caused by reverse shock (RS) heating. The gradation features appear as differences in the peak radii of the emission lines and are seen especially clearly in the north-western (NW) projected ejecta limb.

X-ray imaging using *ASCA* showed that the Fe K emission clearly peaks at a smaller radius than the Fe L and IME line emission and that the Fe-K-emitting plasma was hotter and less ionized (Hwang & Gotthelf, 1997; Hwang et al., 1998). Warren et al. (2005) measured the averaged forward shock (FS) and RS radius as 251 arcsec and 183 arcsec using the feature-less emissions and the Fe K  $\alpha$  lines, respectively, from a *Chandra* observation. Yamaguchi et al. (2014) showed electron heating at the RS on the NW limb

using the Fe K  $\alpha$  and K  $\beta$  lines. They also measured the RS radius as 158 arcsec using the Fe K  $\beta$  lines of immediate postshock, low-ionization ejecta, which peak at a smaller radius than Fe K  $\alpha$  emission from a relatively highly ionized component.

Lu et al. (2015) used *Chandra* observations to show a systematic increase in the S-to-Si line flux ratio with increasing radius resulting from RS propagation in the ejecta and reported the elapsed ionization time since the ejecta was shock-heated. Sato & Hughes (2017a) also found a gradual increase in the line centroids of Fe K  $\alpha$  from the inner side of the SNR to the outer side beyond the radius of the peak intensity and interpreted it as a difference in the elapsed ionization time.

By contrast, the eastern region exhibits an unusual morphological structure called the Fe knot, where several clumps outrun the FS. Detailed analysis of *Suzaku* and *Chandra* data suggests that the Fe knot did not originate in the deep, dense core of the progenitor white dwarf but was instead synthesised under incomplete Si burning or the  $\alpha$ -rich freeze-out regime (Yamaguchi et al., 2017).

In addition, galactic SNRs are widely believed to supply cosmic rays up to the ‘knee’ energy of the cosmic-ray spectrum at  $10^{15}$  eV, accelerating particles to relativistic energies in their blast waves by diffusive shock acceleration. The accelerated electrons emit the non-thermal X-ray synchrotron emission observed from the limbs of young SNRs (e.g., Koyama et al., 1995; Eriksen et al., 2011).

The X-ray synchrotron emission from electrons accelerated at the FS was observed from the limb of *Tycho*'s SNR (Hwang et al., 2002). Cosmic-ray proton acceleration at the FS was also reported (Warren et al., 2005). Eriksen et al. (2011) found non-thermal stripes, which are synchrotron-dominated filamentary structures running perpendicular to the FS in the projected interior of the remnants, and interpreted them as evidence for particle acceleration to the ‘knee’ energy in regions of enhanced magnetic turbulence.

*Tycho*'s SNR, which is one of the brightest SNRs in the X-ray band and has various interior structures, is one of the best benchmark objects for testing a new analysis method. High spatial–spectral-resolution data from *Tycho*'s SNR were obtained by *Chandra*. In this research, we apply our method to the X-ray data from *Tycho*'s SNR to investigate the morphological structures without human bias by automatic classification of each spatial point based only on the physical features reflected in the spectrum.

## 5.2 *Chandra* ACIS-I Data Set

*Tycho*'s SNR was observed by the ACIS-I of *Chandra* for 145.6 ks, 142.1 ks (two obsIDs), 734.1 ks (nine obsIDs), and 146.98 ks in 2003, 2007, 2009, and 2015, respectively, as summarised in Table 5.2.1. In 2007 and 2009, there were one and five observations, respectively, with exposure times exceeding 80 ks. The observation centres are  $00^{\text{h}}25^{\text{m}}19^{\text{s}}00$ ,  $+64^{\circ}08'10''00$  in 2003, 2007, 2009; and  $00^{\text{h}}25^{\text{m}}19^{\text{s}}30$ ,  $+64^{\circ}07'55''60$  in 2015, respectively. Their FoV is approximately  $16' \times 16'$  and covers the entire SNR.

*Tycho*'s SNR was also observed by ACIS-S with HETG for an effective exposure time of 442.80 ks in 2017. Grating observations are quite different from non-grating ones, thus we excluded the grating observation from the datasets used for variational autoencoder (VAE) trainings. *Tycho*'s SNR was also

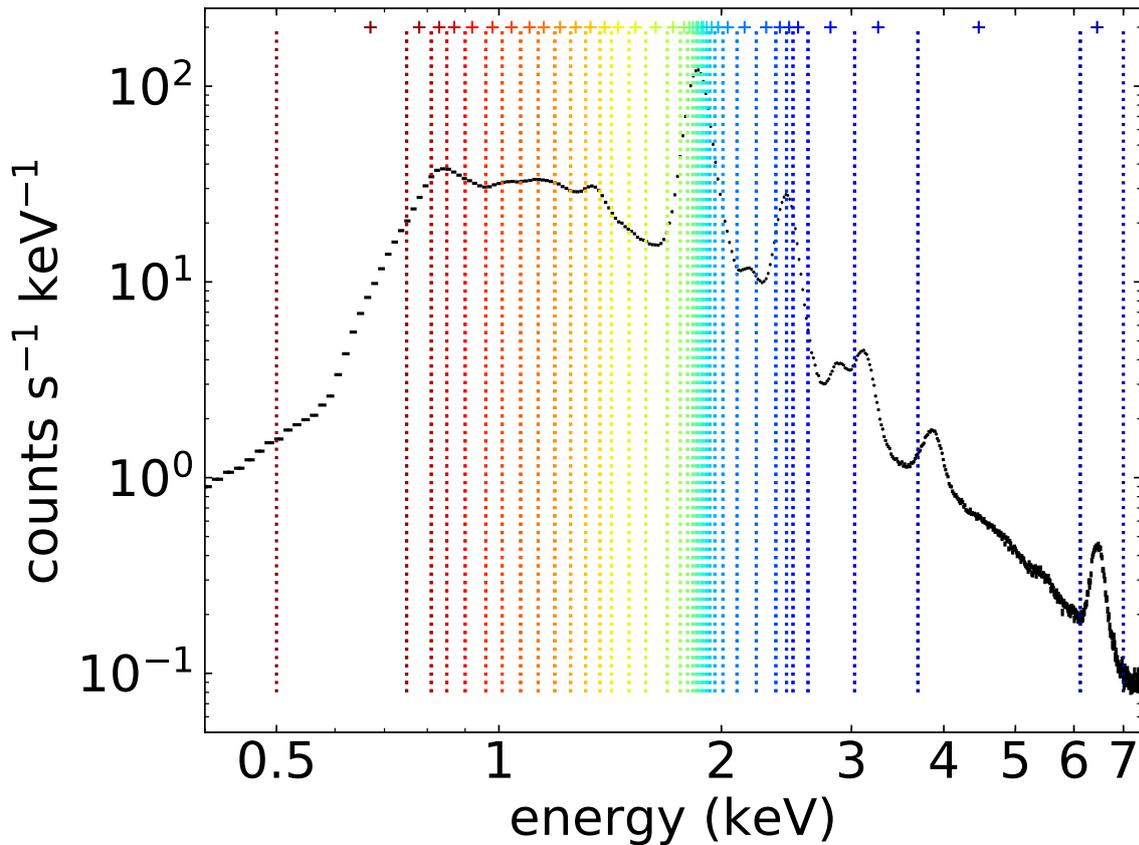


Figure 5.2.1: Spectrum of entire *Tycho*'s SNR. The vertical dotted lines are boundaries of the narrow energy bands. The cross represents the centre of gravity of each energy band.

observed by a backside-illuminated CCD chip of ACIS-S for an effective exposure time of 48.9 ks in 2000. The backside-illuminated chip has a different instrumental response and NXB spectrum from frontside-illuminated chips of ACIS-I. The exposure time of ASIC-S observation in 2000 is much shorter than the total exposure time of ACIS-I observations, thus we decided to use only datasets observed by the ACIS-I. We performed X-ray analysis using CIAO (version 4.9; Fruscione et al., 2006) and CalDB (version 4.7.6) provided by the *Chandra* X-ray Center<sup>1)</sup>.

We determined the spatial-bin size of 3.94 arcsec in order to maintain fine spatial information, i.e., more than 90% of photons are expected to remain in an spatial bin at an off-axis position as described in Section 3.4.2. As a result, an image size becomes  $146 \times 143$  spatial bins to omit regions outside of the SNR.

Finer spectral binning is expected to preserve more information such as the line width, line-centroid shift, and composition of weak lines. However, finer binning results in lower counts in each spectral bin. We employed an objective method of spectral binning to achieve fine binning and adequate photon statistics in each bin at the same time. As shown in Figure 5.2.1, the spectrum of the entire *Tycho*'s SNR was created in the 0.5–7 keV band using only an observation set (ObsID is 10095) and was divided into

<sup>1)</sup> Available at <http://cxc.harvard.edu>

Table 5.2.1: Observations of *Tycho*'s SNR by *Chandra*

ObsID	Exposure (ks)	Date	Grating	Dataset
115	48.91	2000 Sep. 20	–	No usage
3837	145.6	2003 Apr. 29	–	training/validation <sup>a</sup>
7639	108.87	2007 Apr. 23	–	training/validation <sup>a,b</sup>
8551	33.27	2007 Apr. 26	–	training/validation data with 7639 <sup>b</sup>
10093	118.35	2009 Apr. 13	–	training/validation <sup>a</sup> post-training <sup>c</sup>
10094	89.97	2009 Apr. 18	–	training/validation <sup>a</sup> post-training <sup>c</sup>
10095	173.37	2009 Apr. 23	–	training/validation <sup>a</sup> post-training <sup>c</sup>
10096	105.72	2009 Apr. 27	–	training/validation <sup>a</sup> post-training <sup>c</sup>
10097	107.43	2009 Apr. 11	–	training/validation <sup>a</sup> post-training <sup>c</sup>
10902	39.53	2009 Apr. 15	–	post-training <sup>c</sup>
10903	23.92	2009 Apr. 17	–	post-training <sup>c</sup>
10904	34.7	2009 Apr. 13	–	post-training <sup>c</sup>
10906	41.12	2009 May 3	–	post-training <sup>c</sup>
15998	146.98	2015 Apr. 28	–	training/validation <sup>a</sup>
19291	39.98	2017 Oct. 30	HETG	No usage
19292	19.83	2017 Oct. 26	HETG	No usage
19293	49.99	2017 Oct. 17	HETG	No usage
20799	22.17	2017 Nov. 17	HETG	No usage
20813	47.76	2017 Oct. 21	HETG	No usage
20819	44.50	2017 Oct. 29	HETG	No usage
20820	30.50	2017 Oct. 27	HETG	No usage
20821	25.63	2017 Oct. 19	HETG	No usage
20822	13.91	2017 Oct. 23	HETG	No usage
20832	50.42	2017 Nov. 01	HETG	No usage
20833	34.63	2017 Nov. 03	HETG	No usage
20834	35.89	2017 Nov. 04	HETG	No usage
20835	27.59	2017 Nov. 06	HETG	No usage

<sup>a</sup> individual training/validation data

<sup>b</sup> merged training/validation data observed in 2007

<sup>c</sup> merged post-training data

37 narrower energy bins such that each energy bin excluding the last bin (6.12–7.00 keV) had a count averaged over spatial bins inside the SNR of more than 100 counts (spatial-bin)<sup>-1</sup> for the total exposure time observed in 2009, including the background. The low-energy side (0.5–2.6 keV), with a high count rate, is divided into rather narrow bands (bandwidth/band centroid energy < 8%). For example, the Si He  $\alpha$  emission line (1.69–2.01 keV) is divided into 11 bins. On the other hand, the high-energy part of the spectrum (2.6–7.0 keV) is divided into four wider bands because the statistics are not as good as at lower energies. In individual spatial bins, the flux values in the 37 narrow energy bands are combined, and the resulting 37-dimensional vectors are used as the input dataset.

The flux image of each band was created with the coordinate ranges set to omit regions outside of the SNR. We did not subtract the backgrounds from the images because most of *Tycho*'s SNR is sufficiently an order of magnitude brighter than the background that we can safely ignore the contributions from the non-X-ray background and cosmic X-ray background between 0.5 keV and 7.0 keV.

The averaged expansion velocity of *Tycho*'s SNR is approximately 0.3 arcsec yr<sup>-1</sup> (Katsuda et al., 2010), which is significant for the entire dataset. Thus, in our analysis we do not mix observations from different years. Eight individual observations<sup>2)</sup> with exposure times exceeding 80 ks in 2003, 2007, 2009, and 2015 were used for training. In addition, we also used the shorter observation taken in 2007 by co-adding it with the longer one taken in the same year. Eighty per cent of the spatial bins in each flux image were chosen randomly and used as training data, and the rest were used as validation data. The actual size of the training and validation datasets were 150,781 and 36,808, respectively, excluding the spatial bins with zero flux in all the narrow energy bands (i.e., a 37-dimensional zero vector). All the observations from 2009 (a year which has the longest total exposure) were summed and used for the post-training analysis.

## 5.3 Unsupervised Dimensionality Reduction and Clustering

### 5.3.1 VAE Dimensionality Reduction

Figure 5.3.1 shows a diagram of the VAE architecture. The encoder and decoder were constructed from two fully-connected (FC) neural network layers with 256 nodes per layer and rectified linear unit (ReLU; Nair & Hinton, 2010) activation. The output layer has same dimension of the input layer and sigmoid activation function. We determined the latent dimension as follows. We prepared models with latent dimensions ranging from 2 to 10 and trained them for 100 epochs. We selected the latent dimension whose corresponding loss value for the validation data was the lowest.

We used the deep learning framework Keras 2.0.7 (Chollet et al., 2015) with the Tensorflow 1.3.0 (Abadi et al., 2015) backend. Nesterov-accelerated adaptive moment estimation (Nadam; Dozat, 2016) (which, according to our tests, provides the fastest convergence of the optimisers) was used for optimisation. The training was performed for 100 epochs with a batch size of 100.

---

<sup>2)</sup> ObsID 3837, 7839, 10093, 10094, 10095, 10096, 10097, 15998

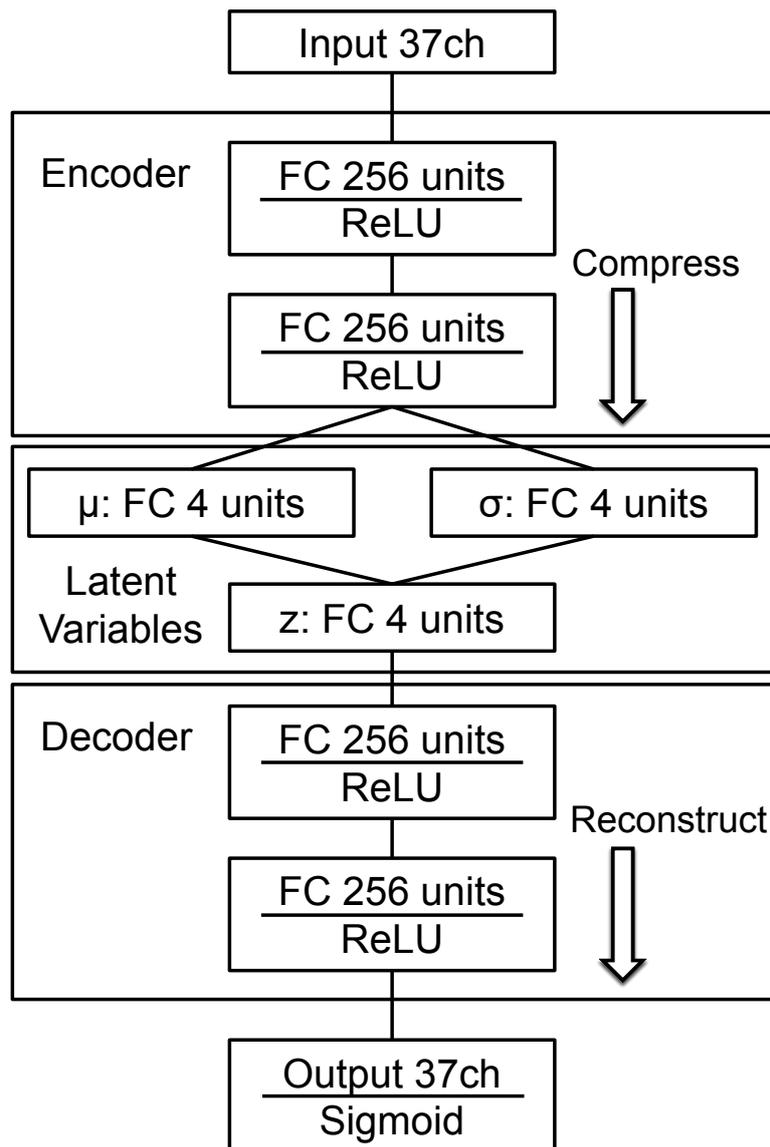


Figure 5.3.1: Diagram of the VAE architecture. See the main text for the details of the network.

Training typically takes 16–20 min on a workstation with an eight-core Intel Xeon E5 CPU. We also tested the training of a model using GPU computations on an NVIDIA GeForce GTX 1080Ti graphics card. Training on the GPU with a batch size of 4096 typically ran for 40 s and required approximately 300 MiB of GPU memory.

We extracted the latent expressions  $\mu$  from the data of *Tycho*'s SNR observed in 2009 using the encoder of the trained VAE. Each panel in Figure 5.3.2 shows the four-dimensional coordinates of the latent parameters  $\mu$  that are obtained by the VAE from a merged set of all the observations from 2009. The images for each axis of  $\mu$  have the same colour scale.

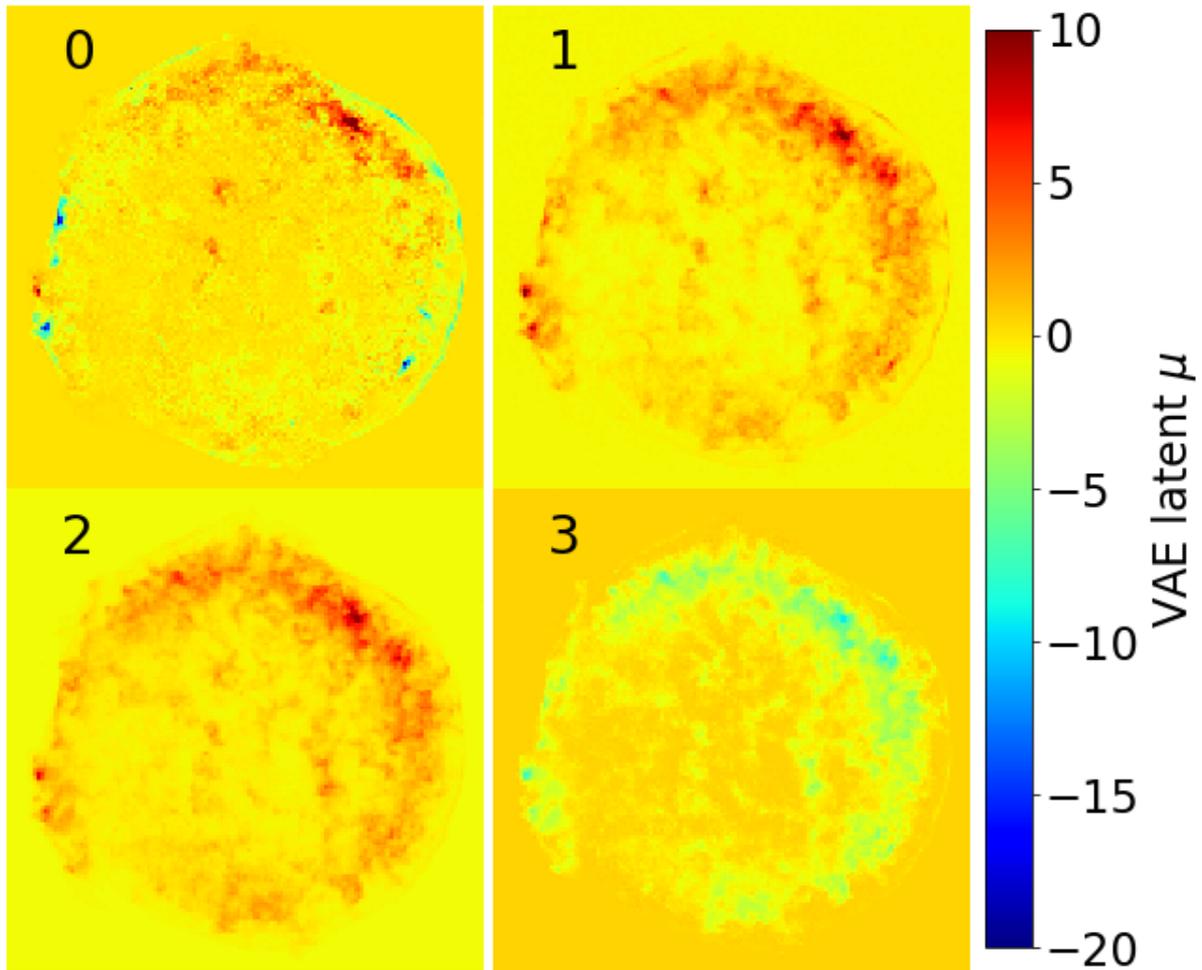


Figure 5.3.2: Images showing the values of the VAE latent parameters  $\mu$  as an image for each axis when the test data (a 37-colour image observed in 2009) are input. The images, which are standardised to improve the appearance, share a colour scale.

### 5.3.2 GMM Clustering

We applied Gaussian mixture model (GMM) soft clustering to the obtained latent expressions. The optimal number was determined by checking the Bayesian information criterion (BIC) for three to nine of the used clusters. There was little difference between cases with seven to nine clusters; thus, for the analysis we used eight clusters. To visualise the features extracted by the VAE, we plot clusters in different colours in the scatter plot shown in Figure 5.3.3.

The scatter plots in Figure 5.3.3 show the distributions of the latent variables, which are colour-coded according to the category assigned by GMM clustering, projected onto all six ( $=_4C_2$ , where 4 is the latent dimension) different two-dimensional planes passing through the origin. In the latent space, the compressed expressions form a radial distribution consisting of several branches around 0. The main

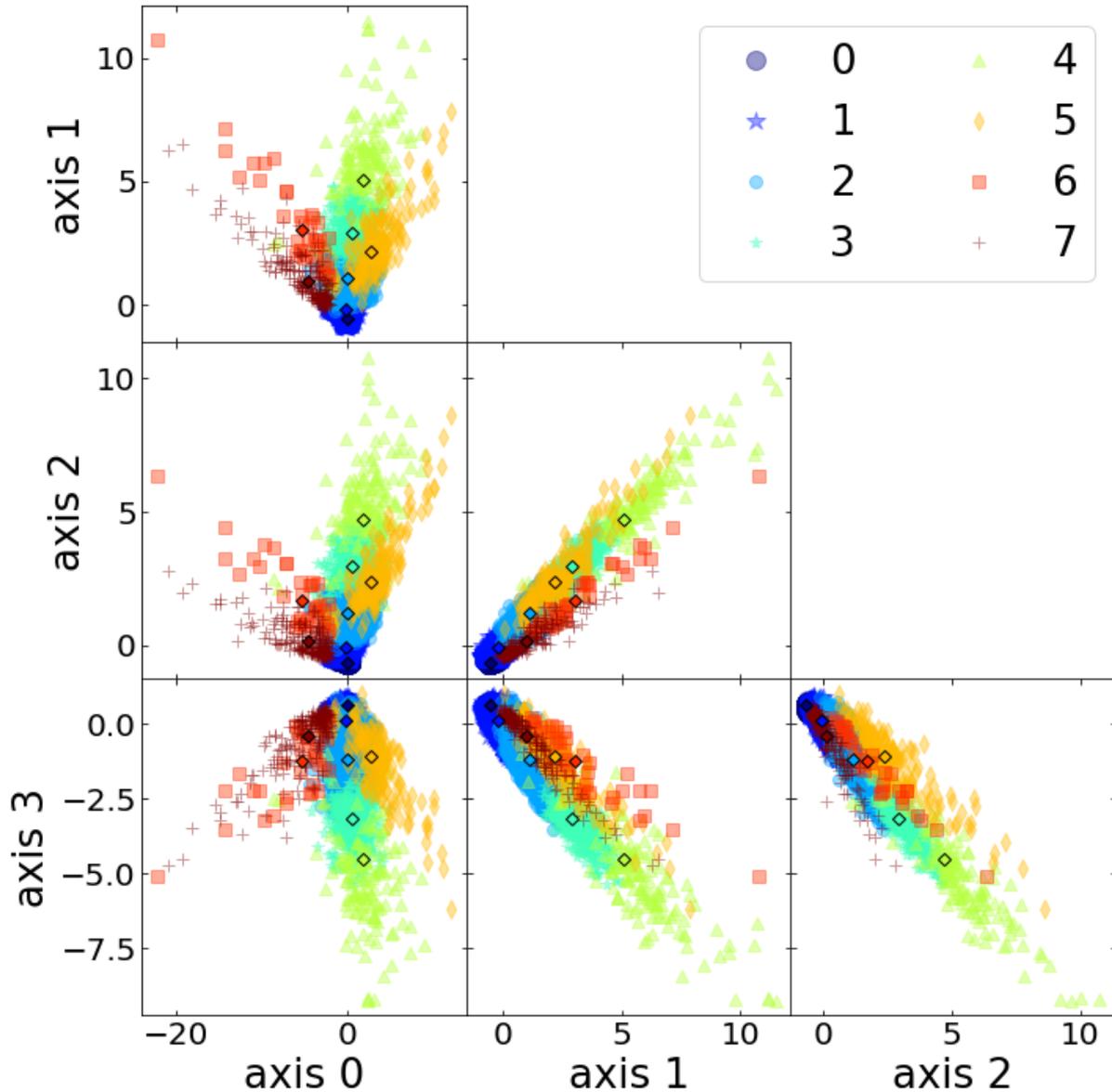


Figure 5.3.3: Scatter plot showing the  $\mu$  of VAE latent parameter obtained by inputting each spatial bin of the test data observed in 2009, where two of the four axes are chosen. Each point was colour-coded for each category classified by GMM clustering. The centroid of each category is shown as an open black diamond.

thick branch is classified as categories 1–4, whereas the three branches extending in different directions are classified as categories 5, 6, and 7, respectively. The data points around 0 are classified as category 0.

Each panel of Figure 5.3.4 shows the responsibility of each GMM category. The middle panel of Figure 5.3.5 shows the division of *Tycho*'s SNR into GMM categories. The colours of the spatial bins correspond to the highest responsibility category, as obtained by the method when the merged data from 2009 were used. The spatial bins of each category have a spatially coherent distribution.

The right panel of Figure 5.3.5 shows the same image as the middle panel of Figure 5.3.5, but the spatial bins whose assigned category has a responsibility of  $\leq 90\%$  are masked. Thus, the spatial bins

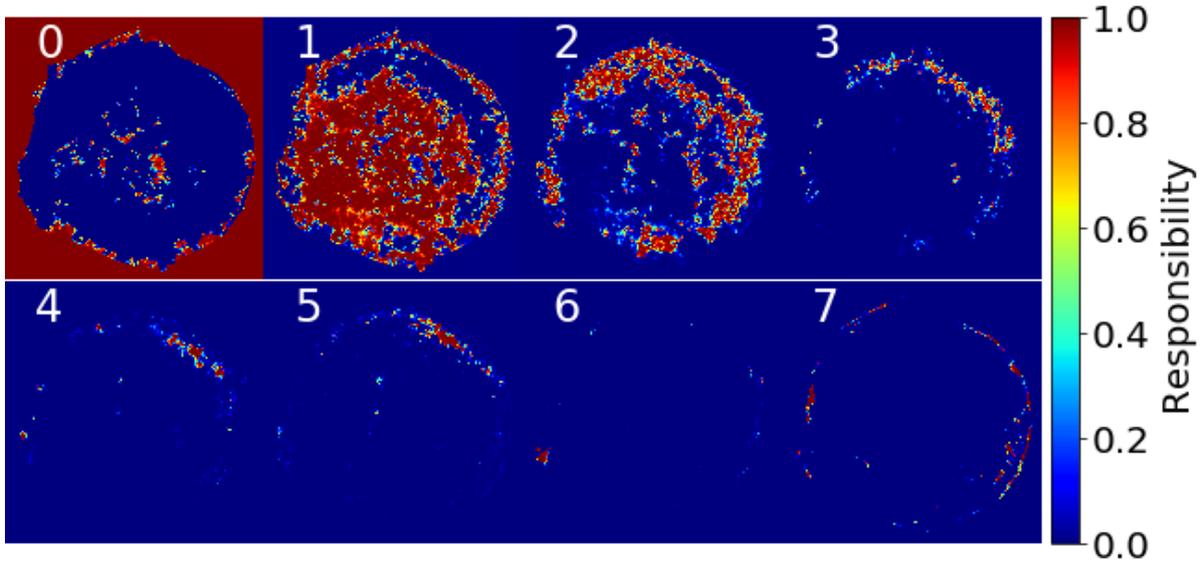


Figure 5.3.4: GMM responsibility of each spatial bin of test dataset observed in 2009 for each category. The responsibility is between 0 and 1; blue represents a responsibility of 0, and red represents a responsibility of 1.

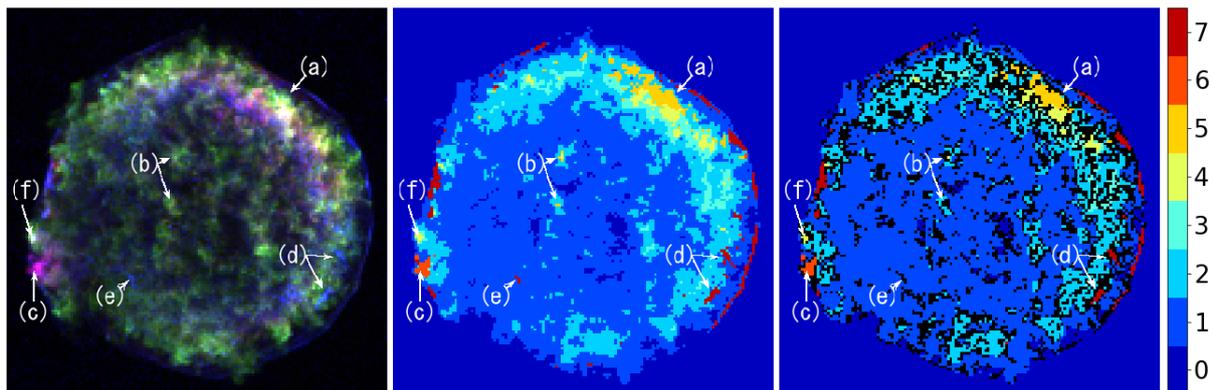


Figure 5.3.5: The *left panel* shows a three-colour image (red: Fe L blend band, 0.7–0.95 keV; green: Si He  $\alpha$  band, 1.75–1.95 keV; blue: Fe K  $\alpha$  band, 6.2–6.9 keV). In the *middle panel*, for each spatial bin of the test data, the GMM categories with the highest responsibility are assigned and colour-coded. The *right panel* shows only spatial bins with a category of more than 90% responsibility, which are assigned the colours representing the category. The other spatial bins, which have responsibilities below 90%, appear black.

that remain coloured in the right panel of Figure 5.3.5 are robustly assigned to some category, and thus are expected to have some spectral features distinct from those of the other categories.

For comparison, a traditional three-colour image is shown in the left panel of Figure 5.3.5. Some regions that appear similar in the three-colour image are assigned to different categories. For example, the blob on the eastern rim (region marked ‘c’ in Figure 5.3.5) and the annular layer to the north-west (inner layer region marked ‘a’ in Figure 5.3.5) both appear reddish in the three-colour image (Figure 5.3.5, left

Table 5.4.1: GMM categories.

Category No.	Location	Feature
0	outside of SNR	very dark, background
1	inside of SNR	dark region, between CD and FS
2–4	rim of ejecta	bright ejecta
5	NW rim	weak Fe emission
6	blobs in the Fe knot	strong Fe line emission
7	rim, filaments	power-law radiation dominant

panel), although they are assigned to different categories in the GMM image (categories 3 and 6; see the middle panel of Figure 5.3.5).

We also note that the clusters corresponding to the layered structure in the NW part of the SNR (for detailed analysis, see Section 5.5.1) are revealed by clustering for any number of Gaussians between seven and nine. On the other hand, the regions dominated by featureless emission and the Fe knot described in Section 5.5.2 are separated into two clusters only when eight or nine categories are assumed.

## 5.4 Detailed Results of Clustering

On the basis of the GMM classification, we extracted the representative spectra of each category by combining all the spatial bins assigned to a certain category with responsibilities above 90%. The combined spectra are shown in Figure 5.4.1. The background was extracted from an annular region surrounding the SNR and subtracted from the spectra.

Table 5.4.1 summarises the physical interpretation of each category. Category 0 is localised mainly outside of *Tycho's* SNR and also contains some dark regions inside the SNR. Categories 1–5, the main regions of which coincide spatially with the ejecta, form a layered structure, and the assigned category numbers, 1–5, change from the inner side of the SNR to the outer side. Category 1 represents faint regions inside of the SNR, which mainly include the unshocked ejecta in projection and the swept-up interstellar medium (ISM)/CSM between the FS and the contact discontinuity (CD). Most of categories 2–5 appears as layered structure, which is most clearly seen in the northern part of the SNR (region marked ‘a’ in Figure 5.3.5), especially category 5, which has a responsibility above 90% and is located only in the NW region of the SNR. Two blobs of categories 4 and 5 close to the SNR centre (marked ‘b’ in Figure 5.3.5) coincide with regions with a measured blue shift (Sato & Hughes, 2017a). Thus, categories 4 and 5 are interpretable as ejecta limbs.

Category 6 is localised at the edge in the eastern part of the SNR (region ‘c’ in Figure 5.3.5), which is associated with the reddish region in the left panel of Figure 5.3.5. This region is a substructure in the Fe knot analysed by Yamaguchi et al. (2017) in detail. The spectrum of category 6 has strong Fe line emission, although the IME emission is weaker in Figure 5.4.1. As shown in the bottom panel of

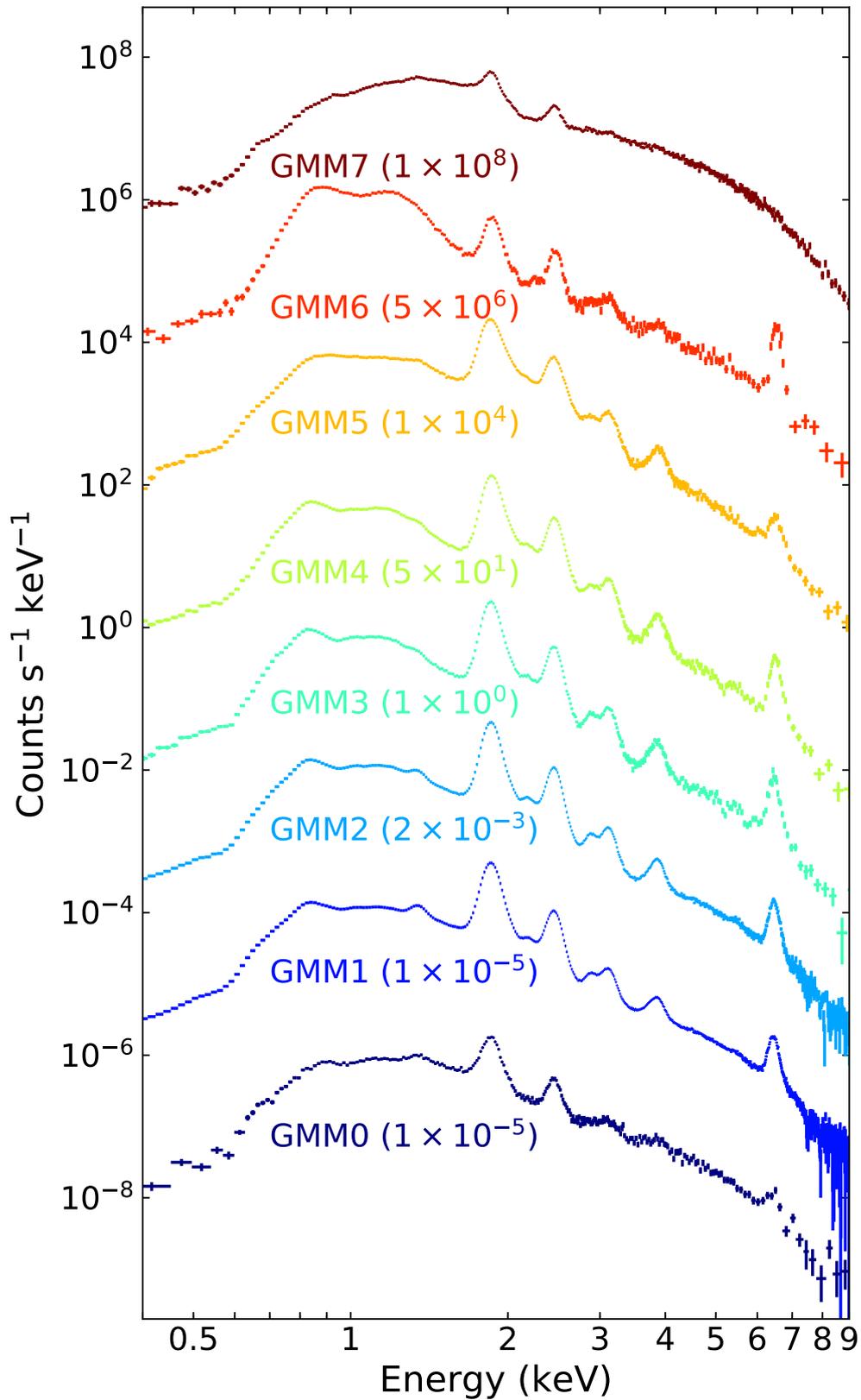


Figure 5.4.1: Spectrum (background-subtracted) of the region inside the SNR for which each GMM category has a responsibility above 90%. The normalisation of each spectrum is adjusted by multiplying by the factor given in the panel.

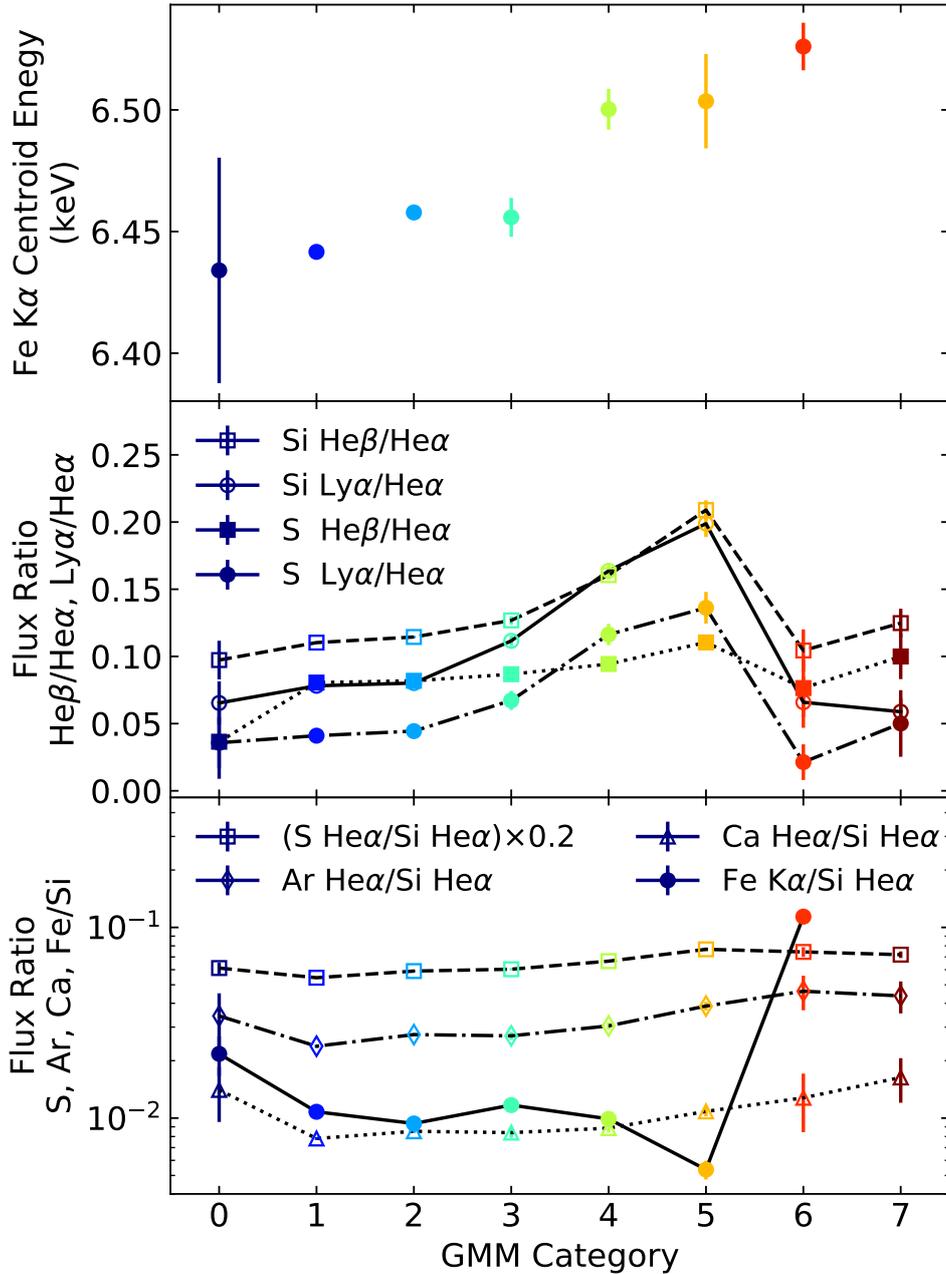


Figure 5.4.2: Model fitting results of *Tycho*'s SNR spectra (best-fitting Gaussian parameters of responsibility of each GMM category above 90%). The *top panel* shows the Fe K  $\alpha$  centroid energies. The *middle panel* shows the flux ratios of different transition lines of Si or S (Ly  $\alpha$  or He  $\beta$  over He  $\alpha$ ). The open circles and solid line correspond to Si Ly  $\alpha$ /He  $\alpha$ , open squares and dashed line to Si He  $\beta$ /He  $\alpha$ , closed circles and dot-dashed line to S Ly  $\alpha$ /He  $\alpha$ , and closed squares and dotted line to S He  $\beta$ /He  $\alpha$ . The *bottom panel* shows the K  $\alpha$  line flux ratios of different elements (S, Ar, Ca, and Fe over Si). The open squares and dashed line correspond to S He  $\alpha$ /Si He  $\alpha$ , open diamonds and dot-dashed line to Ar He  $\alpha$ /Si He  $\alpha$ , open triangles and dotted line to Ca He  $\alpha$ /Si He  $\alpha$ , and closed circles and solid line to Fe K  $\alpha$ /Si He  $\alpha$ . The errors are at 90% C.L.

Figure 5.4.2, category 6 covers a portion of the ejecta having the strongest Fe K  $\alpha$  line in the SNR.

Category 7 corresponds spatially to the FS at the edge of the SNR, the filament and stripe structure at 'd' inside the SNR on the western side, and the bright arc at 'e' in the SNR on the south-east in Figure 5.3.5. These structures are associated with bluish regions in the left panel of Figure 5.3.5. In the spectrum of category 7 in Figure 5.4.1, continuum emission is dominant, although weak contamination by line emission (Si, S, and so on) appears.

Table 5.4.2: Best-fitting parameters for the spectrum of each GMM category with responsibility  $\geq 0.9$ 

	Region							
	0 <sup>a</sup>	1	2	3	4	5	6	7
Fe K $\alpha$ lines								
Centroid (keV)	$6.434^{+0.044}_{-0.046}$	$6.442 \pm 0.002$	$6.458 \pm 0.003$	$6.456 \pm 0.008$	$6.500^{+0.007}_{-0.008}$	$6.504^{+0.020}_{-0.019}$	$6.526^{+0.009}_{-0.010}$	–
Width (keV)	$0.131^{+0.060}_{-0.054}$	$0.088 \pm 0.003$	$0.079 \pm 0.0043$	$0.094 \pm 0.012$	$0.097^{+0.010}_{-0.011}$	$0.117 \pm 0.032$	$0.060^{+0.014}_{-0.016}$	–
Normalisation <sup>b</sup>	$1.80^{+0.45}_{-0.42}$	$317 \pm 5$	$126 \pm 3$	$15.1 \pm 0.8$	$14.8 \pm 0.8$	$7.15^{+0.83}_{-0.80}$	$6.30 \pm 0.45$	–
Line Flux Ratio (%)								
Si He $\beta$ /Si He $\alpha$	$9.7^{+1.5}_{-1.4}$	$11.0 \pm 0.1$	$11.5 \pm 0.1$	$12.7 \pm 0.3$	$16.1 \pm 0.3$	$20.9^{+0.9}_{-0.7}$	$10.4 \pm 1.6$	$12.5^{+1.0}_{-1.1}$
Si Ly $\alpha$ /Si He $\alpha$	$6.5 \pm 1.6$	$7.8 \pm 0.1$	$8.0 \pm 0.1$	$11.2 \pm 0.4$	$16.4 \pm 0.4$	$19.9^{+0.8}_{-1.0}$	$6.6^{+2.0}_{-1.9}$	$5.9 \pm 1.0$
S He $\beta$ /S He $\alpha$	$3.7^{+2.4}_{-2.8}$	$8.1 \pm 0.1$	$8.2 \pm 0.1$	$8.7 \pm 0.4$	$9.4 \pm 0.3$	$11.0 \pm 0.5$	$7.7 \pm 2.1$	$10.0 \pm 1.7$
S Ly $\alpha$ /S He $\alpha$	$3.6^{+5.3}_{-1.9}$	$4.1 \pm 0.2$	$4.4 \pm 0.2$	$6.7 \pm 0.7$	$11.6 \pm 0.8$	$13.6 \pm 1.2$	$2.1^{+2.9}_{-1.3}$	$5.0^{+2.3}_{-2.5}$
K $\alpha$ Line Flux Ratio (%)								
S He $\alpha$ /Si He $\alpha$	$30.6^{+1.9}_{-2.1}$	$27.2 \pm 0.1$	$29.5 \pm 0.1$	$30.2 \pm 0.4$	$33.3^{+0.4}_{-0.5}$	$38.4^{+0.8}_{-0.9}$	$37.2 \pm 2.2$	$36.0 \pm 1.4$
Ar He $\alpha$ /Si He $\alpha$	$3.4^{+1.2}_{-1.1}$	$2.38^{+0.04}_{-0.05}$	$2.74^{+0.04}_{-0.05}$	$2.70^{+0.12}_{-0.13}$	$3.04 \pm 0.13$	$3.87^{+0.19}_{-0.20}$	$4.63^{+1.10}_{-0.95}$	$4.37^{+0.88}_{-0.83}$
Ca He $\alpha$ /Si He $\alpha$	$1.40^{+0.52}_{-0.45}$	$0.78 \pm 0.01$	$0.85 \pm 0.02$	$0.84 \pm 0.05$	$0.89 \pm 0.05$	$1.08 \pm 0.06$	$1.28^{+0.52}_{-0.43}$	$1.63^{+0.44}_{-0.43}$
Fe K $\alpha$ /Si He $\alpha$	$2.17^{+0.54}_{-0.51}$	$1.08 \pm 0.02$	$0.94 \pm 0.02$	$1.17 \pm 0.06$	$0.99 \pm 0.05$	$0.54 \pm 0.06$	$11.4 \pm 0.9$	–

The errors are at 90% C.L.

<sup>a</sup> Here, we used the spectrum extracted only from regions inside the SNR for category 0.

<sup>b</sup> Photon flux at 1 keV in units of  $10^{-6}$  ph cm $^{-2}$  s $^{-1}$ .

The medium- and high-energy parts of the spectra in Figure 5.4.1 have clear line emission features. To investigate the spectral properties quantitatively, we fitted the spectra with a model of an absorbed power law for the continuum emission plus Gaussians for the emission lines of He  $\alpha$ , Ly  $\alpha$ , He  $\beta$ , He  $\gamma$ , and Ly  $\beta$  of Si, S, Ar, and Ca (excluding Ca Ly  $\beta$ ) and Fe K  $\alpha$ .

The weak lines such as Lyman lines and He  $\gamma$  of each element shared the fitting parameters of line centroids, widths, and intensities with a strong line (such as He  $\alpha$  or He  $\beta$ ) of the corresponding element following Hayato et al. (2010). We assumed a hydrogen column density of  $7 \times 10^{21} \text{ cm}^{-2}$  (Cassam-Chenai et al., 2007) and standard ISM abundances (Wilms et al., 2000). The results of the model fitting are shown in Table 5.4.2 and Figure 5.4.2. In category 7, continuum emission is dominant; thus, the Fe K  $\alpha$  line cannot be detected.

In the top and middle panels of Figure 5.4.2, the centroid energy of Fe K  $\alpha$  is higher, and the Ly  $\alpha$ /He  $\alpha$ , He  $\beta$ /He  $\alpha$  line flux ratios of Si and S are higher, except for category 6. These trends correspond spatially to higher ionization and temperature in the outer side of the SNR. Especially in the NW region of the SNR, in which categories 1–5 form a layered structure, the physical parameters of ionization and the temperature appear to change from inner to outer the SNR. We analyse the NW region in detail in Section 5.5.2. On the other hand, category 6, which is located on the rim in the eastern region of the SNR, has different characteristics. For category 6, the centroid energy of Fe K  $\alpha$  is highest among all the categories, but the Ly  $\alpha$ /He  $\alpha$ , He  $\beta$ /He  $\alpha$  line flux ratios of Si and S are lower than the others. We analyse the Fe knot region, including category 6, in detail in Section 5.5.1.

In the bottom panel of Figure 5.4.2, the line flux ratios of He  $\alpha$  of S, Ar, or Ca to Si increase gradually from the inner side toward the outer side of the SNR. On the other hand, the Fe K  $\alpha$ /Si He  $\alpha$  line flux ratios show a different trend. These ratios decrease toward the outer side of the main SNR shell and are lowest in the region of category 5, although the ratio for category 6, which is located the outer edge of the SNR, is only an order of magnitude higher than the others. Category 6 clearly has different spectral features from the other regions.

Although it is difficult to fully understand how each feature in the raw data (i.e., spectral structure in this case) affects the latent expressions, we think the low-energy sides of the spectra also contribute significantly to the clustering because the low-energy part is divided into finer bins than the high-energy part is. For example, whereas the Fe K  $\alpha$  blend consists of only 1 energy bin in our binning method, the energy band of the Fe L blend between 0.75 and 1.31 keV is divided into 10 bins.

The L-shell blend of Fe<sub>XVII</sub>, Fe<sub>XVIII</sub>, and Fe<sub>XIX</sub> ( $n = 3 \rightarrow 2$ ) structure at  $\sim 0.83$  keV is divided into 4 bins and appears strongly in the spectra of categories 1–4. In addition, the Mg He  $\alpha$  line at  $\sim 1.35$  keV appears strongly only in categories 1 and 2, and the O Ly  $\alpha$  line structure at  $\sim 0.65$  keV appears strongly in the spectra of categories 0, 1, and 7. We think that these structures contribute to distinguishing these categories from others.

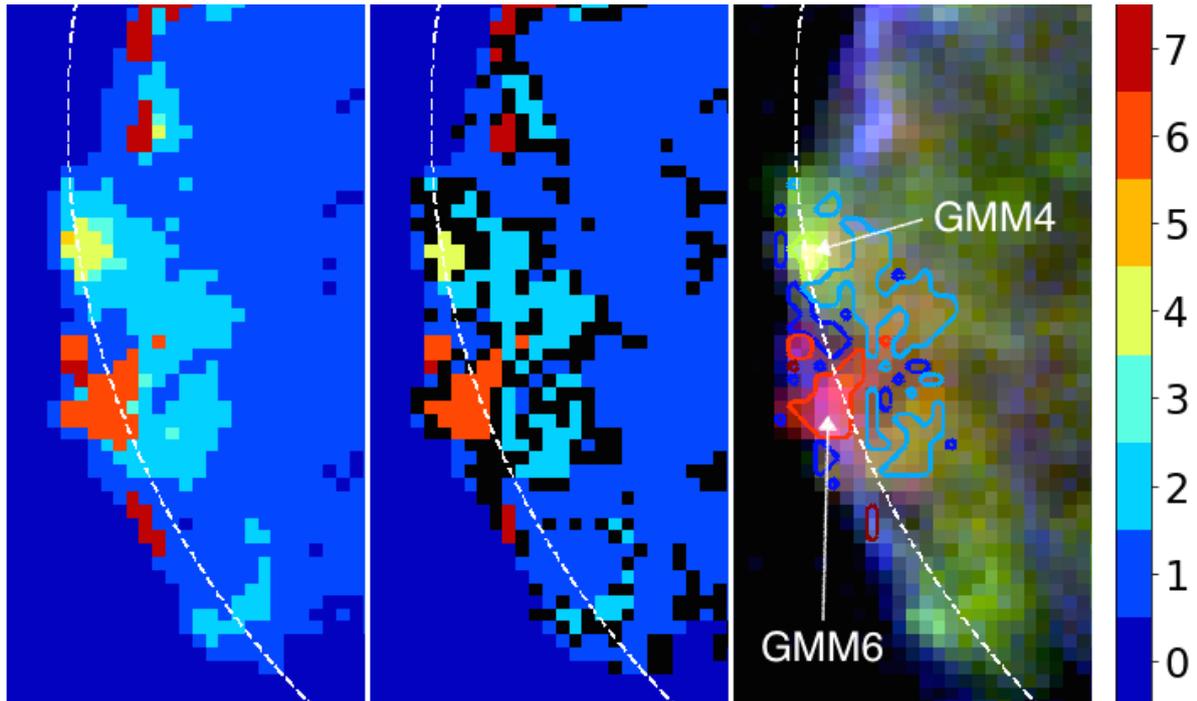


Figure 5.5.1: Substructure of the Fe knot on the eastern rim of *Tycho*'s SNR. In the *left-hand panel*, for each spatial bin of the test data, categories with the highest responsibility are assigned and colour-coded. In the *middle panel*, only spatial bins with a category having a responsibility above 90% are selected and assigned the colours representing the categories. The other spatial bins, which have responsibilities below 90%, appear black. The *right-hand panel* shows a three-colour image (red, Fe L blend band, 0.7–0.95 keV; green, Si He  $\alpha$  band, 1.75–1.95 keV; blue, continuum band, 4.6–5.1 keV). The dashed line corresponds to the azimuthal-averaged radius of the FS (251''; Warren et al., 2005).

## 5.5 Detailed Analyses of the Regions Suggested by Machine Learning

As shown in the previous section, the unsupervised machine learning method can discover spatial structures. In this section, we choose two regions of the revealed structure and analyse them in detail.

### 5.5.1 Spectral Analysis of the Fe Knot

The Fe knot located along the eastern rim of *Tycho*'s SNR represents unusual morphological features in which several iron-rich clumps outrun the FS. The Fe knot can be divided into substructures, and Yamaguchi et al. (2017) analysed in detail these fine regions. The regions defined by Yamaguchi et al. (2017) (see Figure 5.5.2) have the following counterparts in our analysis (Figure 5.5.1): 'A' and 'E', category 6; 'B', category 4; 'C' and 'D', category 2; 'X' and 'Y', category 7.

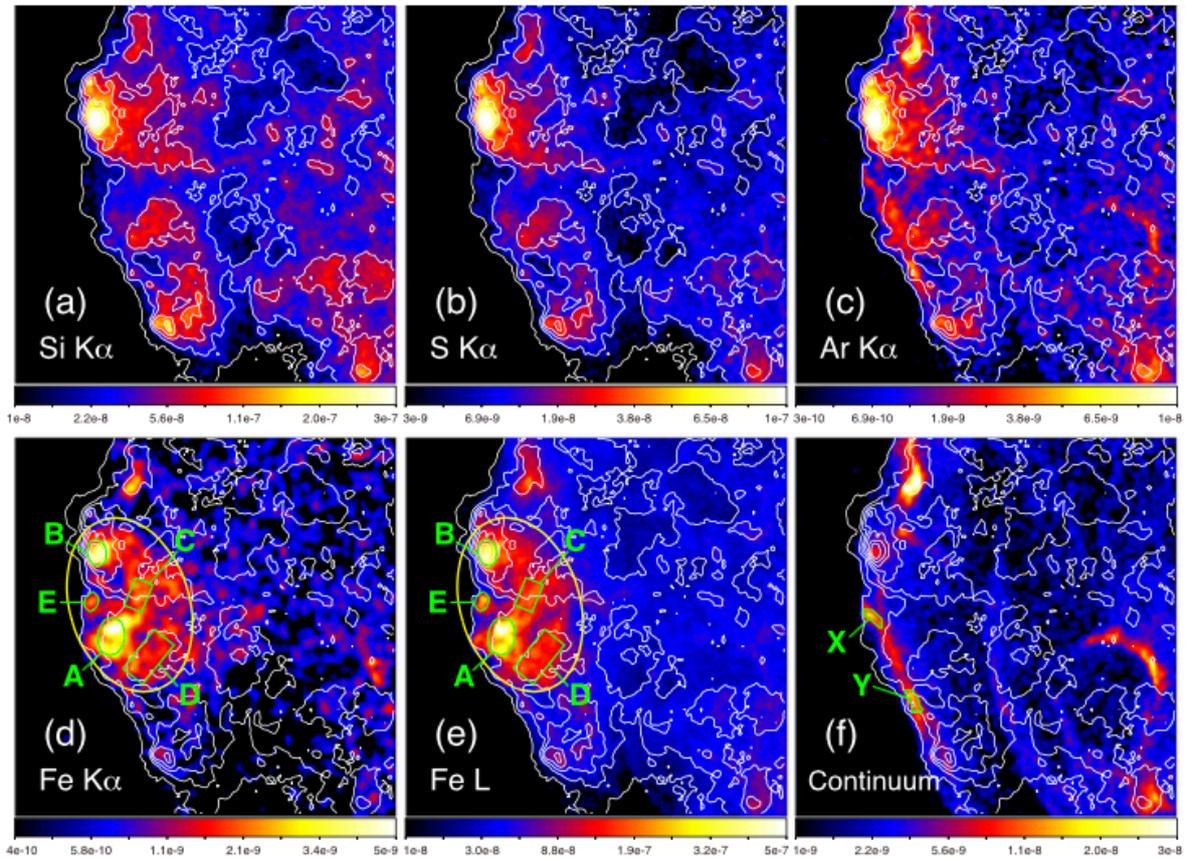


Figure 5.5.2: Eastern region of *Tycho's* SNR observed by *Chandra/ACIS*; (a) 1.8–1.92 keV, (b) 2.4–2.52 keV, (c) 3.07–3.18 keV, (d) 6.35–6.6 keV, (e) 0.8–1.25 keV, and (f) 4.2–6.0 keV. The contour is 1.8–1.92 keV (Si K  $\alpha$ ). The yellow ellipse encloses the entire Fe knot. This figure is taken from Yamaguchi et al. (2017).

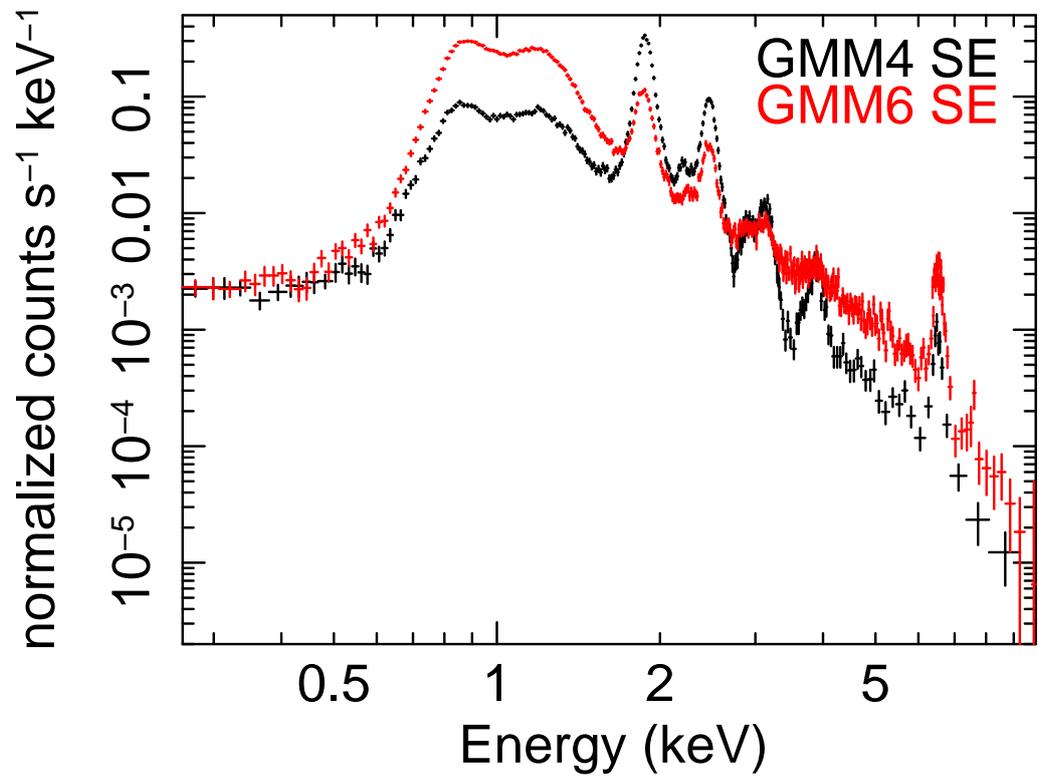


Figure 5.5.3: Spectra of GMM categories 4 and 6 with GMM responsibilities above 90% in the Fe knot, as shown in Figure 5.5.1. The black spectrum is GMM category 4, and the red one is category 6.

Table 5.5.1: Best-fitting parameters for the spectrum of each GMM category in the Fe knot with responsibility  $\geq 0.9$ 

	Region						
	1	2	4	6	7		
Power Law							
Photon Index	$2.97^{+0.13}_{-0.14}$	$2.51^{+0.27}_{-0.26}$	$1.92^{+0.60}_{-0.56}$	$2.85^{+0.14}_{-0.16}$	$2.66^{+0.15}_{-0.21}$		
Surface Brightness <sup>a</sup>	$141^{+27}_{-25}$	$92^{+42}_{-28}$	$94^{+99}_{-38}$	$303^{+70}_{-64}$	$413^{+99}_{-120}$		
Fe K $\alpha$ Lines							
Centroid (keV)	$6.504^{+0.024}_{-0.021}$	$6.509^{+0.009}_{-0.008}$	$6.501 \pm 0.018$	$6.526^{+0.009}_{-0.010}$	–		
Width (keV)	$0.043^{+0.039}_{-0.047}$	$0.076 \pm 0.012$	$0.043^{+0.030}_{-0.056}$	$0.060^{+0.014}_{-0.016}$	–		
Surface Brightness <sup>a</sup>	$0.86 \pm 0.16$	$2.66 \pm 0.17$	$5.19^{+0.80}_{-0.79}$	$5.43 \pm 0.39$	–		
K $\alpha$ Line Flux Ratio (%)							
S He $\alpha$ /Si He $\alpha$	$37.3 \pm 1.9$	$31.3 \pm 0.5$	$34.7 \pm 0.84$	$37.2 \pm 2.2$	$27.9^{+6.5}_{-11}$		
Ar He $\alpha$ /Si He $\alpha$	$3.44^{+0.72}_{-0.67}$	$2.97^{+0.20}_{-0.21}$	$3.60^{+0.27}_{-0.28}$	$4.63^{+1.10}_{-0.95}$	$6.0^{+2.4}_{-2.3}$		
Ca He $\alpha$ /Si He $\alpha$	$1.06^{+0.39}_{-0.32}$	$1.02 \pm 0.08$	$1.03 \pm 0.12$	$1.28^{+0.52}_{-0.43}$	$2.1 \pm 1.5$		
Fe K $\alpha$ /Si He $\alpha$	$2.79 \pm 0.51$	$1.77 \pm 0.11$	$0.87 \pm 0.13$	$11.4 \pm 0.9$	–		

The errors are at 90% C.L.

<sup>a</sup> Surface brightness at 1 keV in units of  $10^{-5}$  ph cm $^{-2}$  s $^{-1}$  arcsec $^{-1}$ .

We extracted spectra from the substructure representative of each category with responsibilities above 90%. The results of fitting by the model described in the previous section are summarised in Table 5.5.1.

Figure 5.5.3 shows the spectra of the regions of categories 4 and 6 in the Fe knot, which correspond to the regions marked ‘GMM4’ and ‘GMM6’ in Figure 5.5.1, respectively, with responsibilities above 90%. The spectrum of category 4 has enhanced IME line radiation, although one of the category 6 features has weaker IME lines and stronger Fe K  $\alpha$  emission. The regions of the Fe knot corresponding to categories 4 and 6 are characterised by the lowest and highest ratios of the Fe K  $\alpha$  and Si He  $\alpha$  line fluxes, respectively. The Fe/Si flux ratio in category 6 is also quite high, as shown in the bottom panel of Figure 5.4.2; thus, the region has a spectral feature unique to not only the entire SNR but also the Fe knot. The machine learning extracted a characteristic structure in which the Fe emission is much stronger.

The He  $\alpha$  line flux ratios of S, Ar, or Ca to Si in categories 1, 2, 4, and 6 are approximately constant in the Fe knot. This implies a constant electron temperature in the knot, in agreement with Yamaguchi et al. (2017).

The S He  $\alpha$ /Si He  $\alpha$  line flux ratios in categories 1 and 2 in the Fe knot are higher than those in the entire SNR (Table 5.4.2). This reflects a higher ionization state in the Fe knot, which is located on the SNR rim (i.e., at a large radius), than that typical of the shocked and unshocked ejecta in the inner part of the SNR, where categories 1 and 2 are most common.

The Fe K  $\alpha$ /Si He  $\alpha$  line flux ratios in categories 1 and 2 in the Fe knot are higher than those in the entire SNR (Table 5.4.2). However, this ratio for category 4 in the Fe knot is comparable to that in the entire SNR. As shown in the left image of Figure 5.5.1, the clump emitting Fe coincides with the blob emitting IME in the category 4 region.

## 5.5.2 Spectral Analysis of NW Ejecta

In the NW region of *Tycho*'s SNR, categories 1–5 and 7 are layered. We extracted a spectrum from each GMM category with responsibilities above 90% in the NW of the SNR. In addition, we divided the NW part of the SNR into annular regions along the layered features representing the categories, as shown in Figure 5.5.4. The centre of the annuli was determined to be R.A. = 00<sup>h</sup>25<sup>m</sup>17<sup>s</sup>.754, +64°08′06″.549 so that those annuli align with the layered structure. The regions were labelled NW1 to NW10 from the inner to the outer side. We extracted the spectra from these regions, adopting a background spectrum extracted from a box region outside of the SNR. We performed model fitting in the 1.7–9 keV band using the same models as in Section 5.4 to investigate the line emission of IMEs and Fe. The best fitting parameters are shown in Figure 5.5.5, Figure 5.5.6, and Figure 5.5.7, and each point and each bin of radius correspond to the centre and the range between the inner and outer edges of each region, respectively.

The surface brightness peak of the Fe K  $\alpha$  line forms the innermost layer, as compared to the He  $\alpha$  lines of Si, S, Ar, and Ca, in Figure 5.5.5. This trend was previously seen in *ASCA* observations (Hwang & Gotthelf, 1997) and *XMM-Newton* observations (Decourchelle et al., 2001).

In Figure 5.5.6, the line flux ratios of Ly  $\alpha$ /He  $\alpha$  and He  $\beta$ /He  $\alpha$  of Si and S increase from the inner region NW5 to the outer one NW10 (156.7–227.3 arcsec) corresponding to the GMM categories 3, 4,

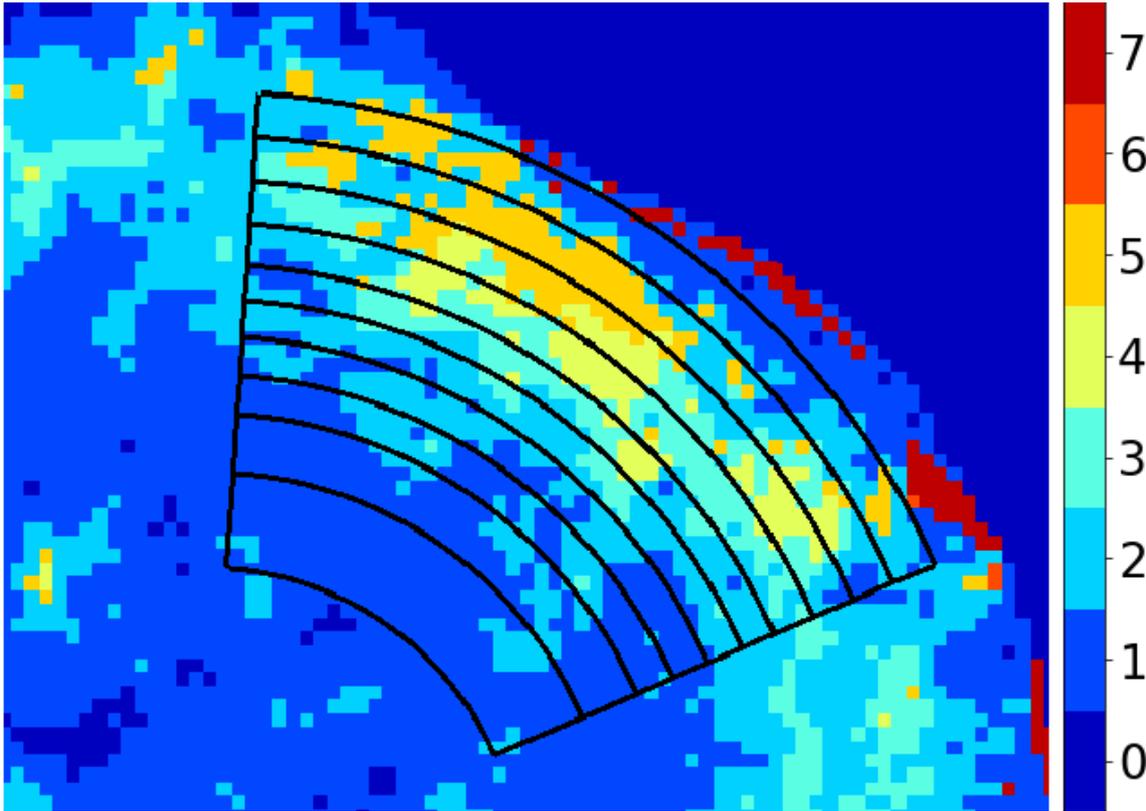


Figure 5.5.4: Annular regions in the NW region of *Tycho*'s SNR.

and 5. These trends correspond to a radial gradient of temperature or ionization, which is higher on the outer side of the SNR. A temperature gradient in the ejecta is suggested, as it is hottest near the RS and cooler near the CD, because the Fe K  $\alpha$  emission peak is located interior to those of Fe L and Si (Hwang & Gotthelf, 1997; Decourchelle et al., 2001). Furthermore, Yamaguchi et al. (2014) showed electron heating near the RS of *Tycho*'s SNR. If we attribute our results to a temperature gradient, our findings imply an opposite trend to that inferred in these works. On the other hand, a variation in the Fe ionization state near the RS (Yamaguchi et al., 2014) was reported. Moreover, a radial gradient of the ionization age was suggested by the Si He  $\alpha$ /S He  $\alpha$  flux ratio (Lu et al., 2015) and Fe K  $\alpha$  centroids (Sato & Hughes, 2017a). Thus, our results are consistent with these works if we interpret the radial dependence of the line flux as arising from the ionization age gradient induced by RS propagation.

Then we fit the spectra with models in the 4.2–10 keV band to investigate the K  $\alpha$  and K  $\beta$  lines of Fe and the K  $\alpha$  lines of secondary Fe-peak elements (Cr and Mn). The Gaussian widths of the Fe K  $\beta$ , Cr K  $\alpha$ , and Mn K  $\alpha$  lines are linked to those of Fe K  $\alpha$ .

The centroid energies of the Fe K  $\alpha$  lines shown in Figure 5.5.7 are flat between NW1 and NW5 (90.0–167.0 arcsec) corresponding to GMM category 1, and begin to increase at NW5 around 162 arcsec corresponding to GMM categories 2 and 3. By contrast, the line width of Fe K  $\alpha$  (except in the outermost region, NW10) and the centroid energies of Fe K  $\beta$  do not change significantly. The peak of the surface brightness of Fe K  $\alpha$  is in the NW6 region between 177.3 arcsec and 189.4 arcsec, and the peak of Fe K  $\beta$  is in the inner region NW5 between 156.7 arcsec and 167.0 arcsec, which is consistent with the results

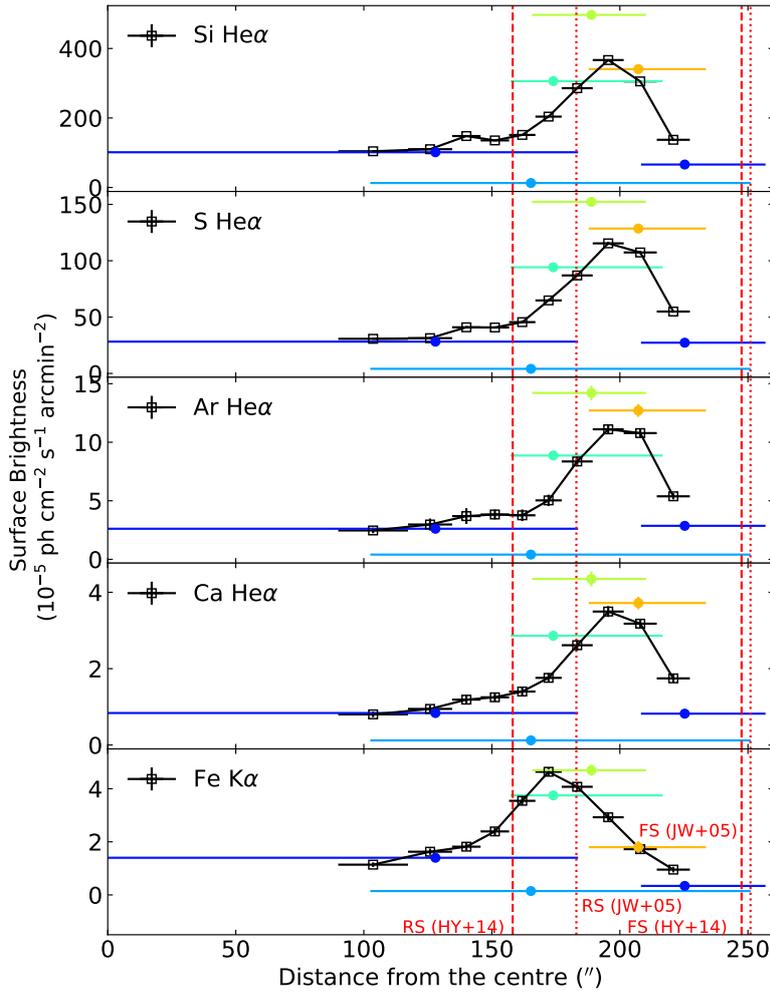


Figure 5.5.5: Radial dependence of the surface brightness of the lines of Si He  $\alpha$ , S He  $\alpha$ , Ar He  $\alpha$ , Ca He  $\alpha$ , and Fe K  $\alpha$  in the NW regions. The closed circles correspond to the spectra of GMM categories with GMM responsibilities above 90% in the NW of the SNR. The open squares and dotted lines correspond to the spectra extracted from the annuli in Figure 5.5.4. The errors are at 90% C.L. The FS and RS positions are determined by Yamaguchi et al. (2014, vertical dashed line) and Warren et al. (2005, dotted line).

of Yamaguchi et al. (2014).

Warren et al. (2005) estimated the averaged RS radius to be 183 arcsec. The RS radius is located in region NW7, which coincides with the turning point of the surface brightness of IME He  $\alpha$  and Fe K  $\alpha$  in Figure 5.5.5. By contrast, Yamaguchi et al. (2014) estimated the RS radius as 158 arcsec in the NW quadrant, which is interior to and more realistic than the former one. The RS radius is located near the boundary of regions NW4 and NW5 corresponding to the inner edge of GMM category 3 and near the centroid of GMM category 2. It coincides with the turning point of the centroid of Fe K  $\alpha$  in Figure 5.5.7. Thus, it seems that categories 3–5 are ejecta shocked by the RS, and most of categories 1 and 2 are located inside the RS in projection.

The transitions of the flux ratios of line emission (Figure 5.5.6) or the line centroid energy (Fig-

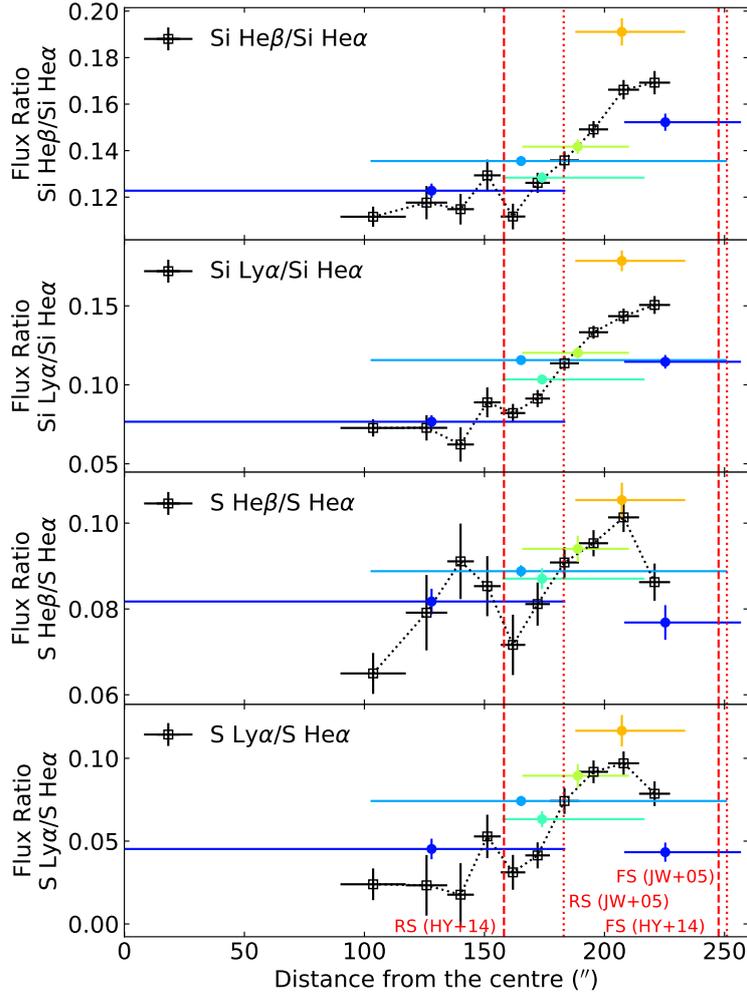


Figure 5.5.6: Radial dependence of the line ratio of  $\text{Ly } \alpha/\text{He } \alpha$  and  $\text{He } \beta/\text{He } \alpha$  for Si and S in the NW regions. The closed circles correspond to the spectra of GMM categories with GMM responsibilities above 90% in the NW of the SNR. The open squares and dotted lines correspond to the spectra extracted from the annuli in Figure 5.5.4. The errors are at 90% C.L. The FS and RS positions are determined by Yamaguchi et al. (2014, vertical dashed line) and Warren et al. (2005, dotted line).

ure 5.5.7) described above appear at the RS position. Thus, the coincidences suggest that the features reflect the plasma state of the ejecta caused by RS heating.

We also note the detection of  $\text{Fe K } \beta$  in regions NW3–NW8 excluding NW7,  $\text{Cr K } \alpha$  in regions NW6 and NW8, and  $\text{Mn K } \alpha$  in region NW8 with  $3\sigma$  or more.

## 5.6 Discussions

We implemented an unsupervised machine learning method combining the VAE and GMM, where the dimensions of the observed data are reduced by the VAE, and clustering in feature space is done by the

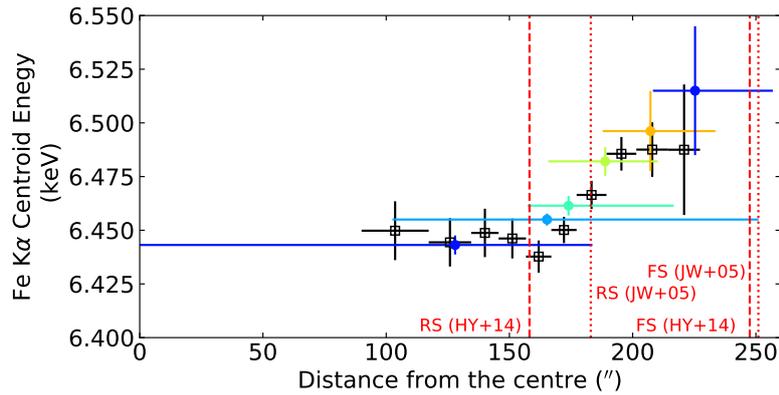


Figure 5.5.7: Radial dependence of the centroid energies of Fe K  $\alpha$  lines in the NW regions. The closed circles correspond to the spectra of GMM categories with GMM responsibilities above 90% in the NW of the SNR. The open squares and dotted lines correspond to the spectra extracted from the annuli in Figure 5.5.4. The errors are at 90% C.L. The FS and RS positions are determined by Yamaguchi et al. (2014, vertical dashed line) and Warren et al. (2005, dotted line).

GMM, and applied the method to *Tycho*'s SNR, one of the best-known SNRs.

Our unsupervised machine learning method automatically revealed spatial structures which have been discussed in the literature (see, e.g., Yamaguchi et al., 2017). This demonstration shows that the method is a powerful tool for data analyses that makes it possible to exploit the rich information contained in data obtained by X-ray observations of SNRs. It may be possible to discover SNR physics by post-training analyses using the results of machine learning.

It is also worth noting that the method discovered the spatial structures automatically, although no spatial information was used in the model. This means that the method can extract physical feature based only on the spectral information.

As demonstrated in Sections 5.3, 5.4, and 5.5, the VAE extracts features using the relative intensities of lines as well as the properties of the continuum spectrum. These characteristics of thermal X-ray spectra reflect the plasma conditions (e.g., temperature, ionization, elemental abundances, and electron or ion densities). When the data distribution in feature space is categorized by the GMM, the entire region is divided into a small number of clusters. As shown by our analysis, clustering can reveal both sharp, knot-like features and continuous changes in physical parameters. Sharp structures are classified as a single category. For example, the Fe-rich blob in the Fe knot on the eastern rim of the SNR, shown in Section 5.5.1, is assigned to category 6. Meanwhile, the synchrotron dominated FS or inner filaments, which have featureless spectrum, is assigned to category 7. By contrast, if physical parameters change gradually, clustering may result in a layered spatial structure like that seen in the NW regions of the SNR (see Section 5.5.2 for details).

The reason that each individual spectrum is classified in a certain category is not yet clear from the network outputs but needs to be investigated and interpreted by human experts, as we mentioned in Section 5.4. It would be useful if the network itself provided the reason, e.g., by highlighting the spectral features that cause the spectrum to be assigned to a particular category. Unveiling the reasoning process

of the network is a significant problem (see, e.g., Smilkov et al., 2017, for recent reviews).

Developing such methods is important for making the best use of the currently available data and for addressing the growing quality and quantity of future observational data. Model fitting of spectra is generally time-intensive; thus, the difficulty of spectral analysis is expected to increase steeply with successful implementation of an X-ray microcalorimeter (such as *Athena*; Barret et al., 2018). The suggested unsupervised method can reveal characteristic features directly from raw observational data without spectral model fitting. It can be an efficient tool to define regions for spectral extraction.

Our method implemented in this work is not limited to SNRs and can be applied to other classes of sources such as galaxy clusters. The method is equally applicable to temporally and spatially variable data, because the training uses only spectral information. Furthermore, our method can also be applied to other energy bands; e.g., it is expected to have good applicability to radio observational data, which contain spatial, temporal, and velocity information (i.e., they have the same dimensions as X-ray data: spatial, temporal, and spectral information).

The deep learning architecture can be improved. In this method, it is a problem that the VAE used to reduce the dimensions of the data tends to form a single peak distribution around 0 in feature space; thus, the boundaries of the extracted data distribution are not clear. Using the architecture of the Wasserstein autoencoder (WAE; Tolstikhin et al., 2017) or Gaussian Mixture VAE (GMMVAE; Dilokthanakul et al., 2016) may improve the structure of the latent manifold. A model using convolutional layers, e.g., a convolutional VAE, can be applied to use the spatial information in a dataset.



## Chapter 6

# Demonstration with Kepler's SNR

### 6.1 Introduction for Kepler's SNR

*Kepler's* SNR (G4.5+6.8), which is the remnant of SN 1604, is one of the most well-studied supernova remnants (SNRs) in the Galaxy. *Kepler's* SNR is generally believed as a remnant of type Ia supernova (SN), even without light echo spectrum. This is mainly because X-ray observations revealed strong emission of silicon, sulphur, and iron but less oxygen from the shocked ejecta (e.g.; Reynolds et al., 2007). Its location is  $590(d/5 \text{ kpc})$  pc away from the Galactic plane, where  $d$  is the distance from the earth.

*Kepler's* SNR has an almost circular shape with some asymmetric features. The most well-known asymmetric features are the two protrusions located in the east and west side of the SNR, often referred to as 'Ears' (Cassam-Chenaï et al., 2004; Tsebrenko & Soker, 2013). Tsebrenko & Soker (2013) proposed that the 'ears' could have been formed by pre-explosion jets in the core-degenerate scenario. Sun & Chen (2019) showed that the thermal plasma in 'ears' mainly consists of Si- and S-rich ejecta, supporting the pre-SN jets scenario for the origin of the 'ears'. By contrast, the interpretation by Burkey et al. (2013) was that they represent an equatorial plane of the pre-SN mass loss of the donor star.

*Kepler's* SNR is interacting with dense (a few particles per  $\text{cm}^3$ ), asymmetric, and nitrogen-rich ambient medium. Douvion et al. (2001) found infrared emission from warm dust and indicated that it was emitted from the dust in the shocked circumstellar medium (CSM). Blair et al. (2007) estimated the total shocked CSM mass to be  $\sim 1 M_{\odot}$ . Williams et al. (2012) showed the CSM dust consists of two components; lower-temperature dust behind fast shocks penetrating into ambient medium, and higher-temperature dust behind slower shocks penetrating into dense material coincides with the optical dense knot. Katsuda et al. (2015) found a nitrogen-rich shocked CSM component from the X-ray spectrum of the entire SNR, and the southern part has higher N abundance. The prominent X-ray emission from oxygen, neon, and magnesium with nearly solar O/Fe abundance ratios has a spatial association with the infrared emission from warm CSM dust (e.g.; Reynolds et al., 2007; Burkey et al., 2013). Sun & Chen (2019) revealed that the X-ray emitting plasma divided into two; hot and low- $n_e t$  plasma corresponding to the newly shock-heated CSM, and cold and high- $n_e t$  plasma corresponding mainly to the ejecta material. The hot plasma is spatially associated with the *Spitzer*  $24 \mu\text{m}$  image tracing the warm CSM

dusts.

The dense CSM environment requires the single-degenerate (SD) scenario. However, no surviving companion star (red giant, asymptotic giant branch (AGB), or post AGB donor star) has been discovered in the central region of the SNR (e.g.; Kerzendorf et al., 2014; Sato & Hughes, 2017b; Ruiz-Lapuente et al., 2018). X-ray observations and hydrodynamic simulations (e.g. Burkey et al., 2013) suggested that the companion is most likely an AGB star with the initial mass of  $\sim 4 M_{\odot}$  (Sun & Chen, 2019), which is an evolved star with a slow and massive wind. In infrared observations by *Spitzer*, Williams et al. (2012) found strong silicate dust suggesting the wind from an O-rich AGB star.

Vink (2008) measured the expansion of *Kepler's* SNR, and determined the shock velocity of  $4200 \text{ km s}^{-1}$  in the eastern part of the SNR for the distance to the SNR of 4 kpc. Katsuda et al. (2008) shows that the northern half of the SNR expands more slowly than the southern part, suggesting an uneven ejecta distribution between the northern and southern shells. Although the density environments are different, some ejecta knots are expanding nearly freely (Sato & Hughes, 2017b; Millard et al., 2019). Sato & Hughes (2017b) measured the kinematic centre at  $\alpha_{J2000}=17^{\text{h}}30^{\text{m}}41^{\text{s}}.189\pm 3'.6$  and  $\delta_{J2000}=-21^{\circ}29'24''.63\pm 3''.5$  based on the expansion velocities of knots. Kasuga et al. (2018) measured the Fe-ejecta velocities in the line-of-sight, and shows their asymmetric motions.

The distance to *Kepler's* SNR is still uncertain with ranging from  $\sim 4.0$  kpc to  $> 7$  kpc. Recent works often use the distance of 5 kpc or 5.1 kpc (e.g. Sato & Hughes, 2017b; Sun & Chen, 2019), which was measured using  $\text{H I}$  absorption (Reynoso & Goss, 1999) and recently optical proper motion of Balmer shocks (Sankrit et al., 2016). Millard et al. (2019) estimated the distance to be  $\sim 4.8$  to 8.2 kpc based on radial velocity measurement using X-ray grating observations by *Chandra*.

## 6.2 *Chandra* ACIS-S Data Set

*Kepler's* SNR was observed by the ACIS-S of *Chandra* for 48.82 ks, 46.17 ks, 741.04 ks (six obsIDs), and 139.11 ks (two obsIDs) in 2000, 2004, 2006, and 2014, respectively, as summarised in Table 6.2.1. In 2007 and 2009, there were five and one observations, respectively, with exposure times exceeding 80 ks. The observation centres are  $17^{\text{h}}30^{\text{m}}41^{\text{s}}.00$ ,  $-21^{\circ}29'17''.02$  in 2000, 2004;  $17^{\text{h}}30^{\text{m}}42^{\text{s}}.00$ ,  $-21^{\circ}29'00''.00$  for two obsIDs in 2006 (6714 and 6716);  $17^{\text{h}}30^{\text{m}}41^{\text{s}}.24$ ,  $-21^{\circ}29'31''.45$  for the others in 2006; and  $17^{\text{h}}30^{\text{m}}41^{\text{s}}.20$ ,  $-21^{\circ}29'31''.40$  in 2014, respectively.

*Kepler's* SNR was observed by ACIS-S and was centred on the ACIS-S3 chip, which is a backside-illuminated CCD chip. The field of view (FoV) of the ACIS-S3 chip is approximately  $8' \times 8'$  and covers the entire SNR. *Kepler's* SNR was also observed by ACIS-S with HETG for an effective exposure time of 147.59 ks in 2016. Grating observations are quite different from non-grating ones, thus we decided to exclude the grating observation from the datasets used for variational autoencoder (VAE) trainings.

We performed X-ray analysis using CIAO (version 4.11; Fruscione et al., 2006) and CalDB (version 4.8.3) provided by the *Chandra* X-ray Center<sup>1)</sup>.

<sup>1)</sup> Available at <http://cxc.harvard.edu>

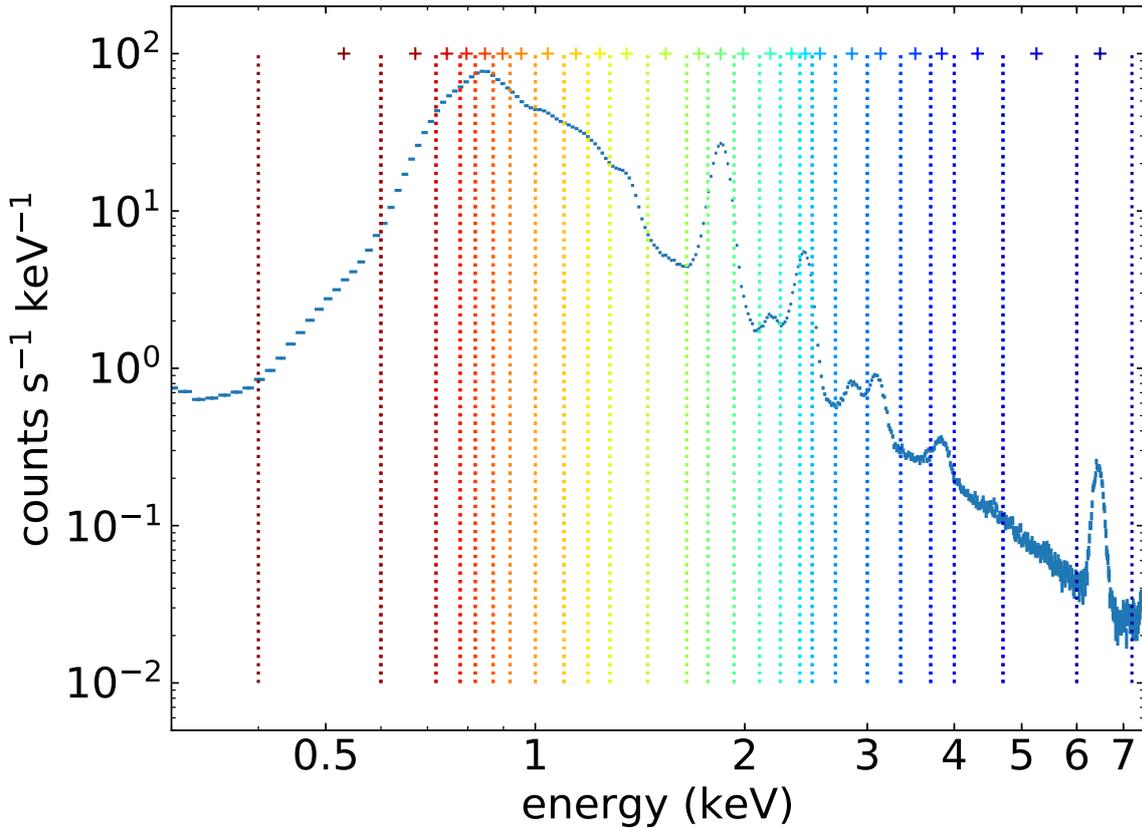


Figure 6.2.1: Spectrum of entire *Kepler's* SNR. The vertical dotted lines are boundaries of the narrow energy bands. The cross represents the centre of gravity of each energy band.

As shown in Figure 6.2.1, the spectrum of the entire *Kepler's* SNR was created in the 0.4–7.2 keV band (using only the data of ObsID 6715, which has the longest exposure) and was divided into 26 narrower energy bins such that each clearly resolved emission line is divided into an independent energy bin. The low-energy side (0.72–1.45 keV), with a high count rate, is divided into rather narrow bands (band-width/band-centroid-energy < 10%). The medium-energy band (1.45–4.0 keV) has a lower count rate than the lower energy side, but is also divided into rather narrow bands because intermediate-mass element (IME) emission line blends are clearly resolved. Each strong line blend of Si He  $\alpha$  or S He  $\alpha$  was divided into three energy bins, the peak and both wings. Their line blend tails also contain other transition lines (e.g., Ly  $\alpha$ ) whose intensities are quite weaker than He  $\alpha$ . The lower- and higher-energy part of the spectrum (0.4–0.72 and 4.0–7.2 keV) is divided into two and three wider bands respectively, because the statistics are not as good as at the low- or medium-energy sides and the emission line blends have broader widths. The continuum dominated band (4.0–6.0 keV) was divided into two bins. The flux values in the 26 narrow energy bands corresponding to a single spatial bin are combined, and the resulting 26-dimensional vectors are used as the input dataset.

The flux image of each band was created with the coordinate ranges set to omit regions outside of the SNR. We prepared two spatial bin sizes, 3.94 arcsec, and 1.97 arcsec, resulting in image sizes of

Table 6.2.1: Observations of *Kepler*'s SNR by *Chandra*

ObsID	Exposure (ks)	Date	Grating
116	48.82	2000 Jun. 30	–
4650	46.17	2004 Oct. 26	–
6714	157.82	2006 Apr. 27	–
6715	159.13	2006 Aug. 03	–
6716	158.02	2006 May 05	–
6717	106.81	2006 Jul. 13	–
6718	107.80	2006 Jul. 21	–
7366	51.46	2006 Jul. 16	–
16004	102.72	2014 May 13	–
16614	36.39	2014 May 16	–
17901 <sup>a</sup>	147.59	2016 Jul. 20	HETG

<sup>a</sup> It is excluded from the dataset because it is a grating observation.

$59 \times 71$ , and  $118 \times 138$  spatial bins, respectively. We did not subtract the backgrounds from the images because most of *Kepler*'s SNR is sufficiently an order of magnitude brighter than the background that we can safely ignore the contributions from the non-X-ray background and cosmic X-ray background in the corresponding energy range.

The flux images whose spatial bin size is 3.94 arcsec of all individual observations and the merged images of them were used for training. In addition, we also used the merged images whose spatial bin size is 1.97 arcsec to increase the training data size. We removed the spatial bins whose 0.4 and 7.2 keV photon fluxes were less than a given threshold in order to exclude the samples outside of the SNR. We defined the threshold such that  $F_{\text{thresh}} = |\mathbf{F}_{\text{bkg}}| + 5\sigma(\mathbf{F}_{\text{bkg}})$ , where  $|\mathbf{F}|$  is the mean of flux,  $\sigma(\mathbf{F})$  is the standard deviation of flux, and  $\mathbf{F}_{\text{bkg}}$  is the photon fluxes in the background region. The threshold of photon flux were  $2.24 \times 10^{-7}$ , and  $1.34 \times 10^{-6}$  photons  $\text{s}^{-1} \text{cm}^{-2}$  for the spatial bin size 3.94 arcsec, and 1.97 arcsec, respectively.

Twenty percent of the spatial bins in each flux image were chosen randomly and used as validation data, and the rest were used as training data. We exclude the spatial bins with zero flux in all the narrow energy bands (i.e., a 26-dimensional zero vector) from both the training and validation datasets. All the observations with the spatial bin size of 3.94 arcsec were summed and used for the post-training analysis. We call it the merged data.

## 6.3 Results of Unsupervised Machine Learning

### 6.3.1 The VAE Architecture

We examined some VAE models having different number of dimensions in the latent variables or hidden layers to find the better models. We prepared VAE models, which have 4 or 8 dimensional latent variables  $z$  and each hidden layer containing 128 or 256 nodes, and trained them with Poisson log likelihood reconstruction loss (Eq. 4.1.7) through 100 epochs. For these training, each data sample was normalized by the maximum value of each sample in order to focus only on the spectral shape.

In all the cases of model with 8 dimensional latent variables, 4 in 8 latent axes of  $z$  seem to have means, although the 4 rest axes of  $z$  seem to be random values and are not committed to informative dimensionality reductions as shown in Figure 6.3.1. Histograms of the latent variables in Figure 6.3.1 show that the meaningless four axes of  $\mu$  are almost consistent with 0 although the meaningful four axes of  $\mu$  have extended distributions. We conclude that 4 axes of latent variables are sufficient for the present dataset, and decided to use the model with 4 dimension latent variables. For the models with 4 dimensional  $z$ , we decided to use the model with 128 nodes in each hidden layer, which had the better loss, while the loss values are almost same if we changed the number of nodes in each hidden layer.

For comparison, we trained the models also with the dataset scaled by the entire maximum value, which maintain the flux information in each spatial bin. The five in eight latent axes of  $z$  seem to have means, although the rest three axes of  $z$  seem to be random values as shown in Figure 6.3.2. An axis (assigned as  $z_1$ ), which newly appeared, seems to represent the flux intensity information. In Figure 6.3.2 the histograms of latent variables show that the source and background bins are clearly separated in  $z_1$  corresponding to brightness information although they cannot be separated in the other axes of  $z$ . In each axis of  $\mu$  in Figure 6.3.2, the distribution of background bins had a sharp peak centring on 0 with a smaller variance than source bins. However the distributions  $\log(\sigma^2)$  of background bins, which peaked in larger values than source bins, made the distribution  $z$  extended wider than  $\mu$  for background bins. It means the spectra of background bins tend to have larger random noise than those of source bins since the lower photon count makes the Poissonian fluctuation larger in each spectral bin. By contrast, the histograms in Figure 6.3.1 show that the source and background bins are not separated so clearly but the peaks of distribution  $\mu$  and  $z$  are shifted. Each distribution of  $\log(\sigma^2)$  for the background bins had a peak in larger value than source bins. In this point, the results using data normalized sample by sample had a same trend as the results using data scaled by the entire maximum.

The comparison shows that the flux information could be veiled by the spatial-bin-by-bin normalization, but the information of photon counts still appeared because it affected the quality of spectral shape and also had been used in the Poisson reconstruction loss function through training.

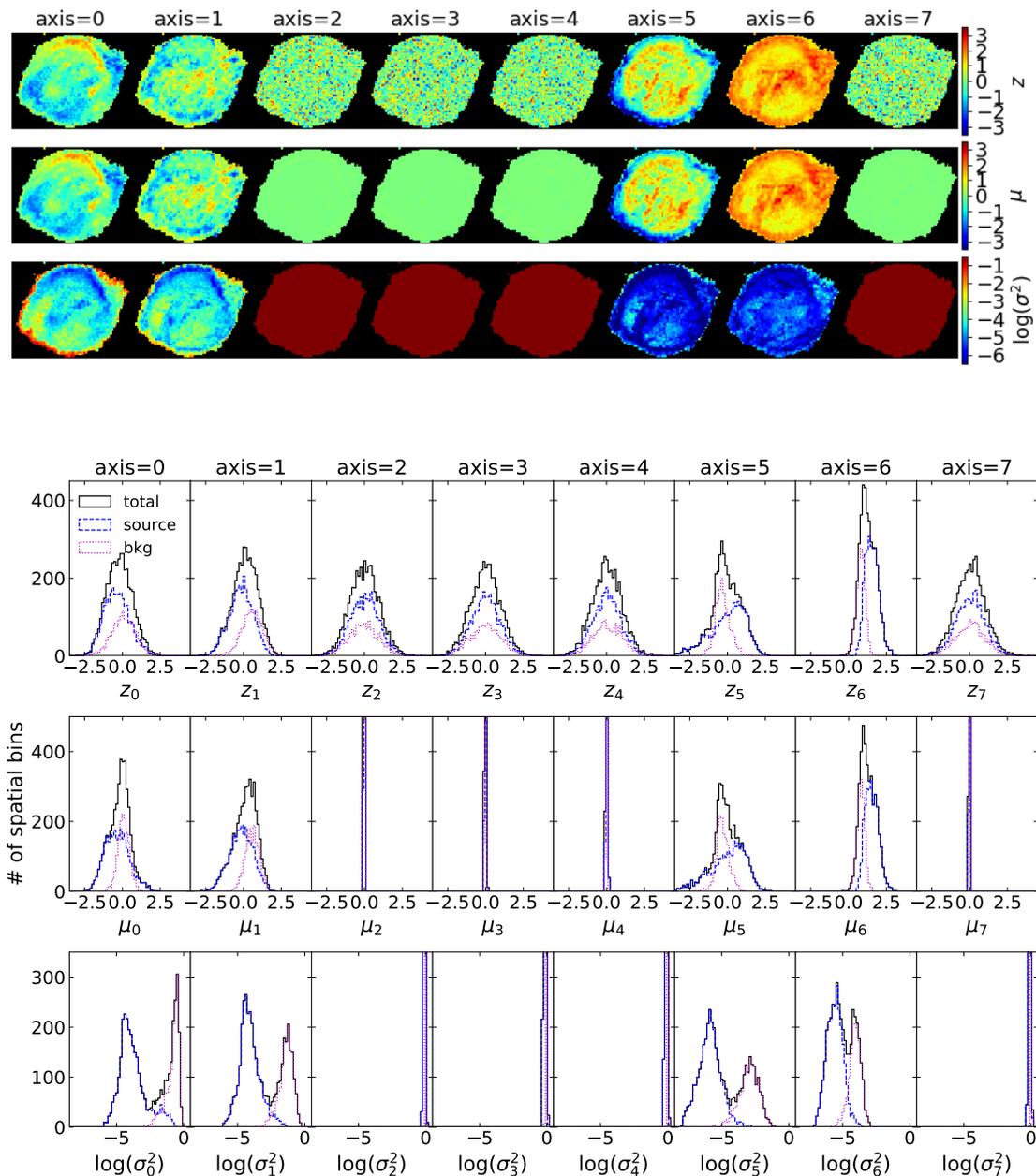


Figure 6.3.1: The 8-d latent parameters of the VAE having 256-d neurons in each FC layer after 100 epochs training. The upper and lower panels show images and histograms of the latent parameters  $z$ ,  $\mu$ , and  $\log(\sigma^2)$  respectively for each axis when the merged data are input. Each data sample was normalized by the maximum value of each sample to focus only on the spectral shape.

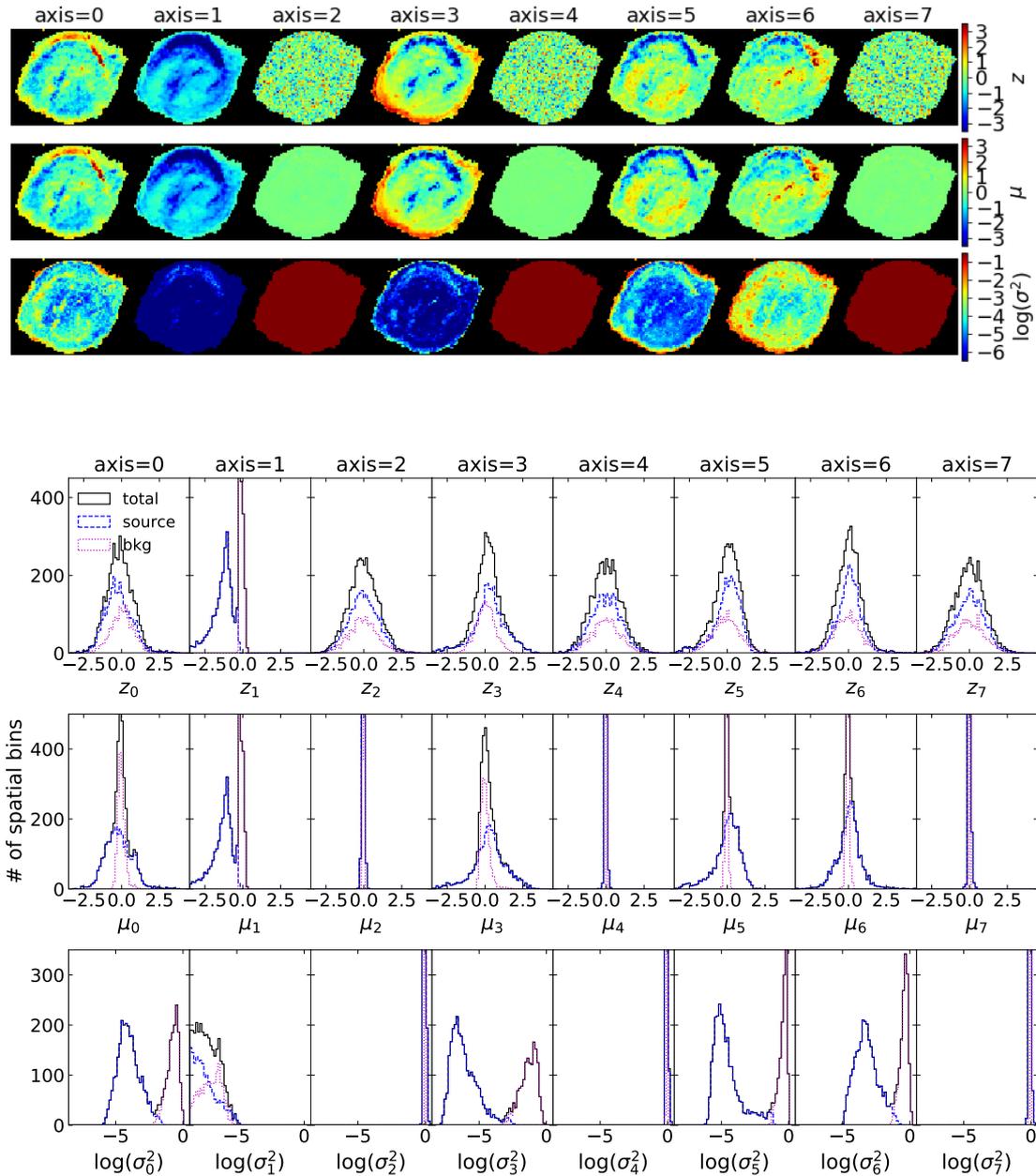


Figure 6.3.2: The 8-d latent parameters of the VAE having 256-d neurons in each FC layer after 100 epochs training. The upper and lower panels show images and histograms of the latent parameters  $z$ ,  $\mu$ , and  $\log(\sigma^2)$  respectively for each axis when the merged data are input. The dataset was scaled by the entire maximum value to keep information of intensities.

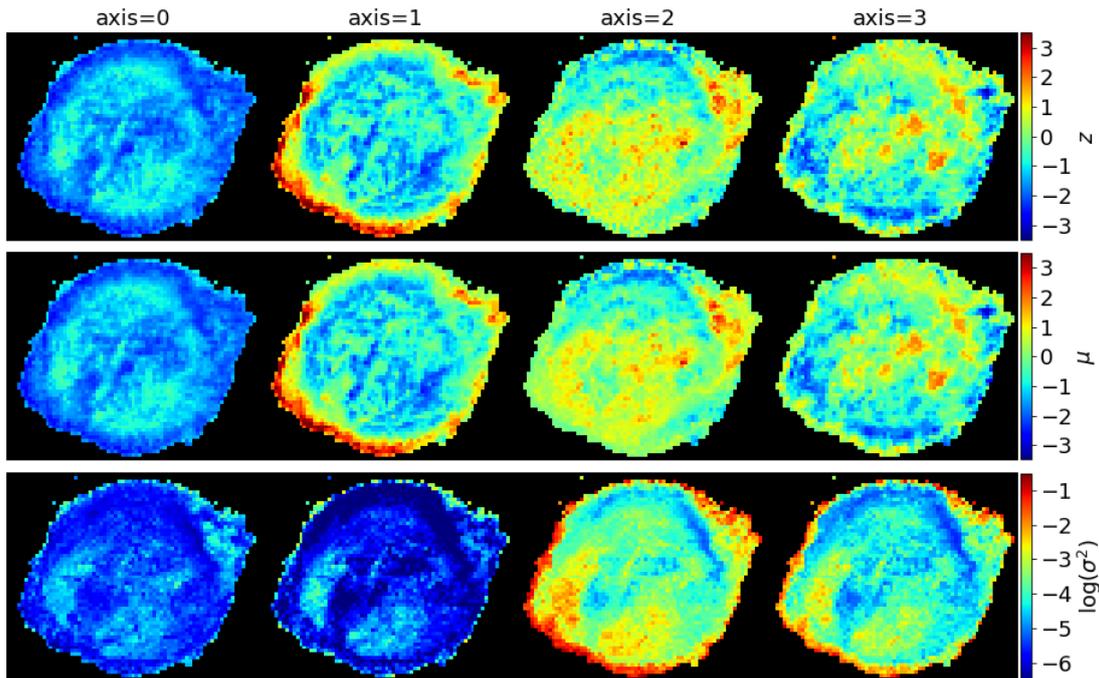


Figure 6.3.3: Images showing the values of the VAE latent parameters  $z$ ,  $\mu$ , and  $\log(\sigma^2)$  in the top, middle, and bottom row as an image for each axis when the merged data (a 26-colour image) are input into a model having 4-d latent variables and 128-d neurons in each FC layer after 100 epochs training. Each data sample was normalized a maximum value of each sample when it was input to the VAE. The images in each row share a colour scale.

### 6.3.2 The Results with the Best VAE Model

As the result of the examination we showed above, we decided to use the VAE architecture with the 4 dimensional latent variables and 2 fully-connected (FC) layers containing 128 nodes. We show the latent variables of  $z$ ,  $\mu$ , and  $\log(\sigma^2)$  for each axis when the merged data normalized spatial bin by bin are input in Figure 6.3.3. We also show the latent variables as scatter plots (for only  $\mu$ ) in Figure 6.3.4 and histograms in Figure 6.3.5 dividing source bins having larger fluxes than threshold and the rest, i.e. background bins. In the scatter plots of Figure 6.3.4, the latent parameters  $\mu$  are projected onto all six ( $=_4C_2$ , where 4 is the latent dimension) different two-dimensional planes passing through the origin. The source and background bins seem to be approximately separable in the space  $\mu_0$  vs  $\mu_1$ . In Figure 6.3.5, the distributions of  $z$  and  $\mu$  for the source and background bins have different peaks. For each axis of  $\log(\sigma^2)$ , the background bins whose peak is larger than that for the source bins, made the corresponding axis of  $z$  extended more widely.

We modelled the latent variable distribution for the source bins using Gaussian as shown in Figure 6.3.6. Although the distributions were distorted, we decided to use the parameters of Gaussian as the approximate centre and standard deviation of each data distribution. The positions corresponding

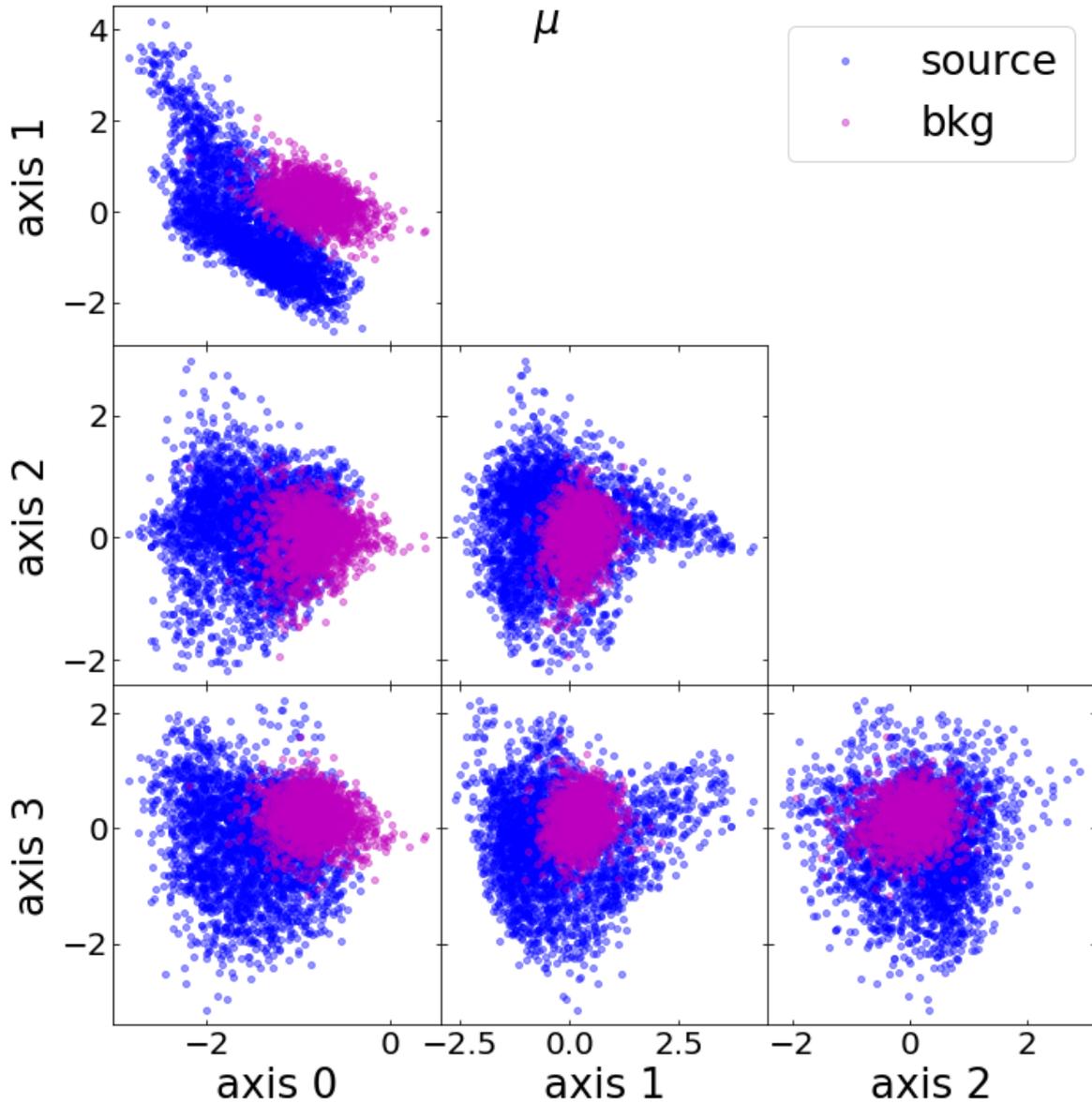


Figure 6.3.4: Scatter plot showing the  $\mu$  of VAE latent parameter obtained by inputting each spatial bin of merged data, where two of the four axes are chosen. The blue and magenta points represent the spatial bins in the source and background region, respectively.

to  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  of the best fitting Gaussian were shown as vertical dotted lines in Figure 6.3.6. The axes 0, 2, and 3 of  $\mu$  and  $z$  had approximately symmetric distributions although the axis 1 had a long tail over the  $3\sigma$  position on the positive side. The trends can be seen in the scatter plots of  $\mu$  in Figure 6.3.4, especially the tail of positive  $\mu_1$  side protruded from the main cluster.

### 6.3.3 Reconstruction of Spectra

We showed some spectra from samples of the merged data reconstructed by the VAE in Figure 6.3.7. The black squares and solid line are the original input spectra, on the other hand the red closed circle

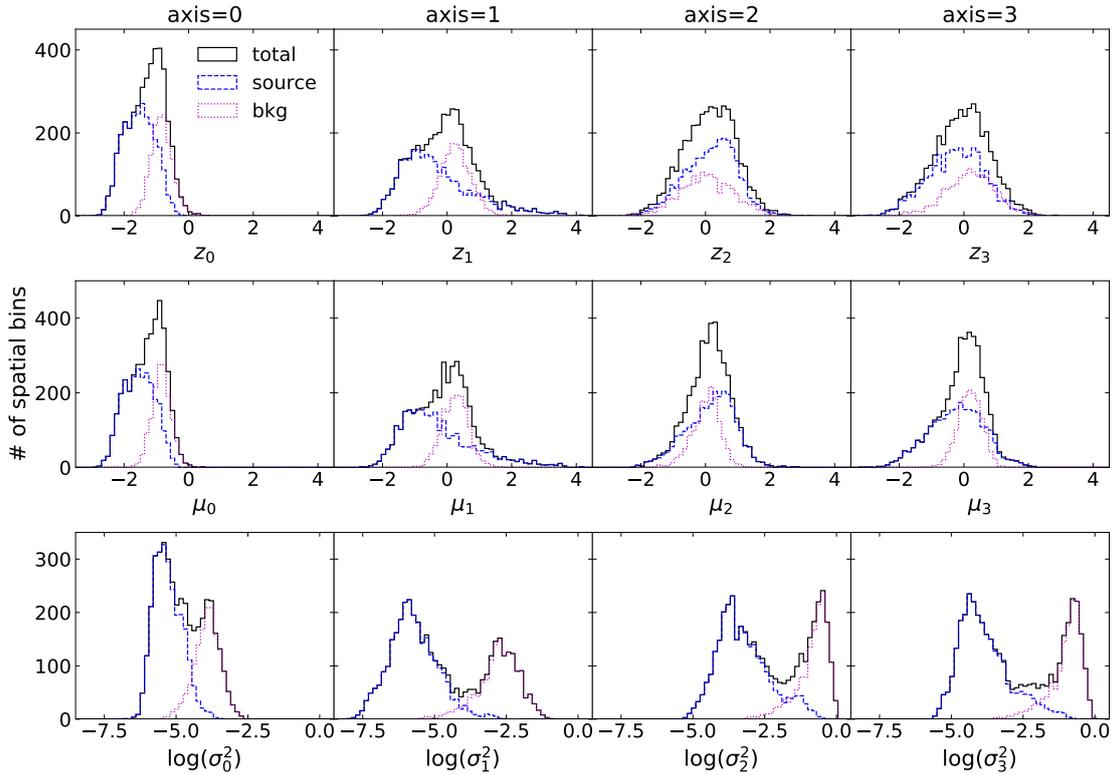


Figure 6.3.5: Histogram showing each axis of  $z$ ,  $\mu$ , and  $\log(\sigma^2)$  of VAE latent parameter obtained by inputting each spatial bin of merged data. The black solid, blue dashed, and magenta dotted lines represent spatial bins in the entire dataset, the source region, and the background region, respectively.

and dashed line are the reconstructed spectra. The box regions marked with ‘A’ to ‘D’ were shown on each image in Figure 6.3.8. The region ‘A’ corresponds to the synchrotron dominated spectrum from the forward shock (FS) in the SE of the SNR. The region ‘B’ has a spectrum with strong iron emission line blends. The ‘C’ and ‘D’ correspond to regions with strong infrared emissions from shocked CSM. The region ‘E’ is chosen outside of the SNR as a background spectrum for comparison. The generated spectrum for background is smooth because the fluctuations caused by low statistics are removed by the VAE.

#### 6.3.4 Spectra Generated by the Decoder

We studied which features are represented by each axis of  $z$  to generate spectra by the decoder. A decoder is capable of generating spectra using the latent parameters beyond real dataset, thus it can emphasize the features committed by each latent parameter. It can be an advantage of generative model. In Figure 6.3.9, Figure 6.3.10, and Figure 6.3.11 we showed the spectra generated by the decoder for given latent parameter sets.

In Figure 6.3.9 the generated spectra were shown with  $z_1$  changing from -2.5 to 4.0, while  $z_0$  was fixed at the best fitting Gaussian mean, -1.5, and  $z_2$  and  $z_3$  were changed between  $\pm 2.5$  for each column and

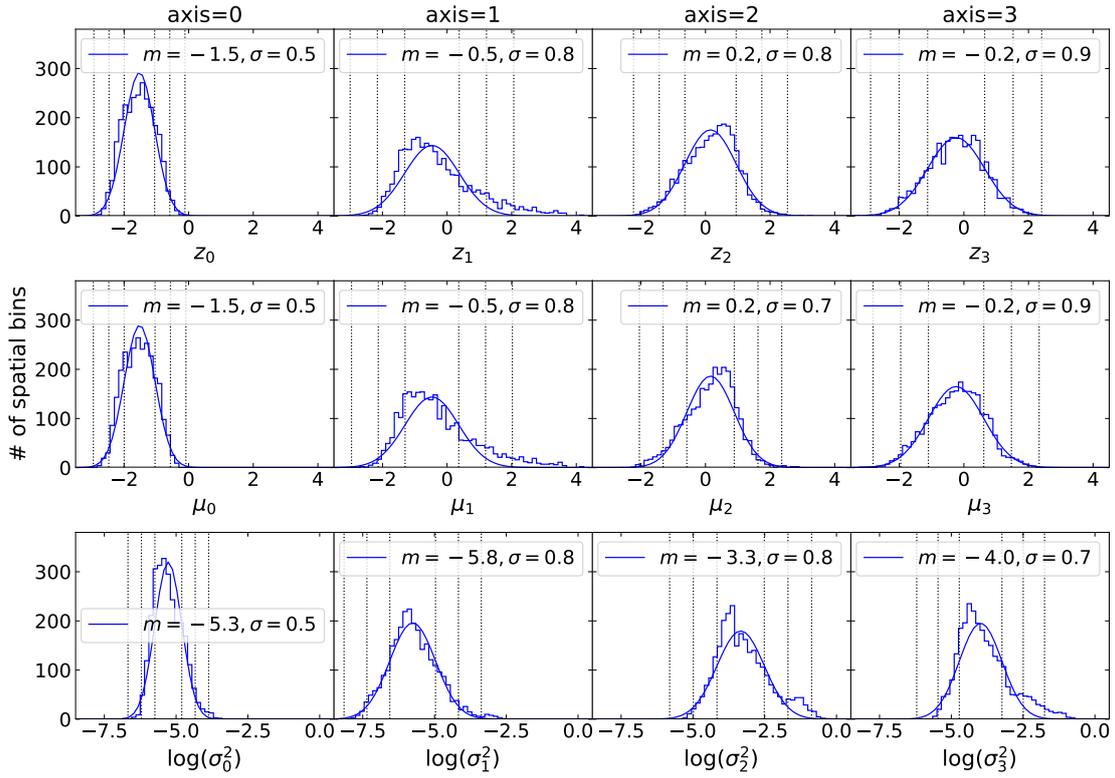


Figure 6.3.6: Histogram showing each axis of  $\mathbf{z}$ ,  $\boldsymbol{\mu}$ , and  $\log(\sigma^2)$  of VAE latent parameter obtained by inputting each spatial bin of merged data in the source region. The curve in each panel is Gaussian fitted on each histogram, the best fitting parameters of Gaussian mean  $m$  and standard deviation  $\sigma$  are shown in each legend. The vertical lines correspond to the  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  positions of each Gaussian.

row, respectively. These spectra for any  $z_2$  and  $z_3$  shows that the continuum dominates the spectra with large  $z_1$ , while the lines clearly appeared with small  $z_1$ . Therefore,  $z_1$  seems to represent the relation of continuum and line intensity. The high  $z_2$  value made the value high for the lowest energy bin (0.4–0.6 keV) including N Ly  $\alpha$  line blend, the O Ly  $\alpha$  line blend bin (0.6–0.72 keV), and the S K  $\alpha$  line blend (2.4–2.5 keV). When the  $z_3$  value becomes low, the IME lines became strong, meanwhile the peak of Fe L blend shifted to low energy side between 0.72–0.92 keV. Thus  $z_3$  seems to represent the IME line intensities and ionization state of Fe.

Figure 6.3.10 shows the spectra generated when the inputs for  $z_0$  were -1.5, -0.5, and -2.5 corresponding to the Gaussian mean and the shift of  $\pm 2\sigma$ . The  $z_1$  was fixed at the Gaussian mean, -0.5, while  $z_2$  and  $z_3$  were changed between  $\pm 2.5$  for each column and row, respectively. In this case, the  $z_0$  changed only among negative values because of the  $z_0$  distribution for the SNR dataset. When  $z_0$  is large, the generated spectrum became hard. The trends for  $z_2$  and  $z_3$  are same as the case changing  $z_1$  with  $z_0$  fixed.

Figure 6.3.11 shows the spectra generated when  $z_0$  changed between  $\pm 2.5$  for each row, which was beyond the range obtained by inputting of merged data. The  $z_1$  was changed between -2.5–4.0 in each panel, while  $z_2$  was fixed at the Gaussian mean, 0.2, and  $z_3$  changed between  $\pm 2.5$  for each column.

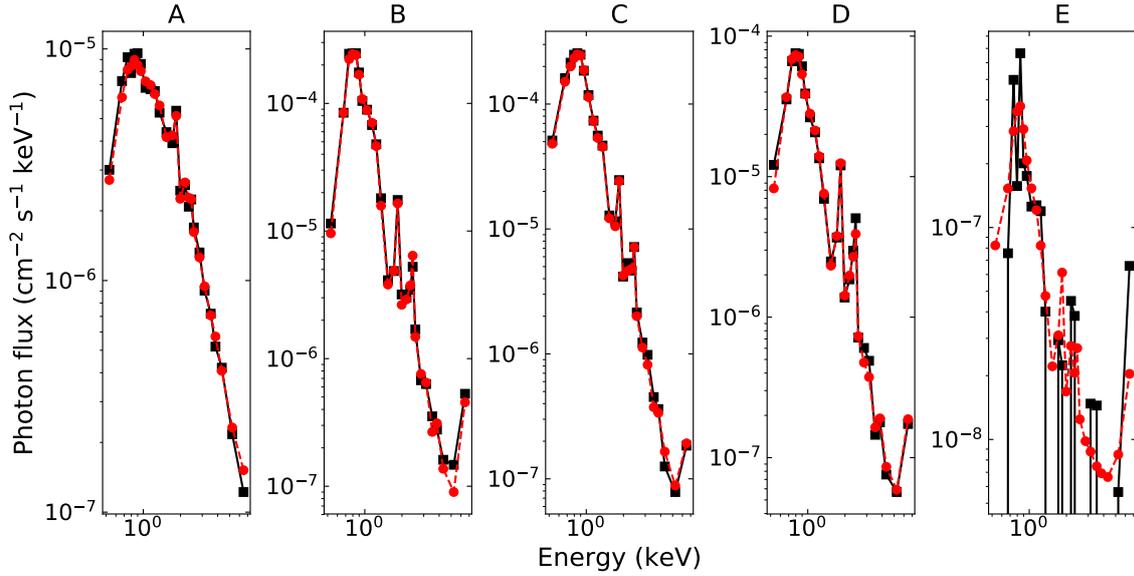


Figure 6.3.7: Sample plot comparing original and reconstructed spectra by the trained VAE as black square and red circle markers, respectively. The spectra marked with ‘A’ to ‘D’ correspond to spatial bins in the source region shown as white box regions in Figure 6.3.8, while the spectrum ‘E’ is the background region.

The small  $z_0$  values made the spectral shape curved upward. The middle of spectrum was pushed up by the soft continuum, stronger IME line blends, and line blends of Fe L, Ne, and Mg within 0.9–1.4 keV. On the other hand, the high  $z_0$  values made the spectra curved downward. The strong Fe L blend of 0.7–0.9 keV pushed up the low-energy side meanwhile hard continuum and Fe K  $\alpha$  blend of 6–7 keV pushed up the high-energy side, although the weaker IME lines of 1.6–4 keV suppressed the middle of spectrum. Therefore, possible  $z_0$  represents the information of global spectral shape, which reflects the relation of Fe and IME line emissions, and also continuum. In addition, when  $z_0$  is low and also  $z_1$  is sufficiently high, the generated spectrum is highly dominated by continuum component.

### 6.3.5 Combined Spectra for Each Latent Axis

We show as contours the regions corresponding to more than  $+1\sigma$  or less than  $-1\sigma$  on each axis of  $z$  in Figure 6.3.12. The contours corresponding to  $+3\sigma$  for  $\mu_1$  and  $z_1$  are also shown in Figure 6.3.13.

We extracted spectra from the regions corresponding to more than  $+1\sigma$  or less than  $-1\sigma$  for each axis of  $z$ . In addition, for  $z_1$ , a spectrum was also extracted from regions corresponding to more than  $+3\sigma$ . The background-subtracted spectra for each axis were shown in each panel of Figure 6.3.14. The annulus surrounding the SNR was determined as the background region. Point-like sources were removed from the annulus, as shown in Figure 6.3.13.

The spectrum from regions of  $z_1 \geq 1\sigma$  had strong continuum and furthermore the spectrum from more than regions of  $z_1 \geq 3\sigma$  was almost continuum dominated, whereas the spectrum from regions of  $z_1 \leq -1\sigma$  had strong emission lines. It supports the interpretation of  $z_1$  mentioned in Section 6.3.4.

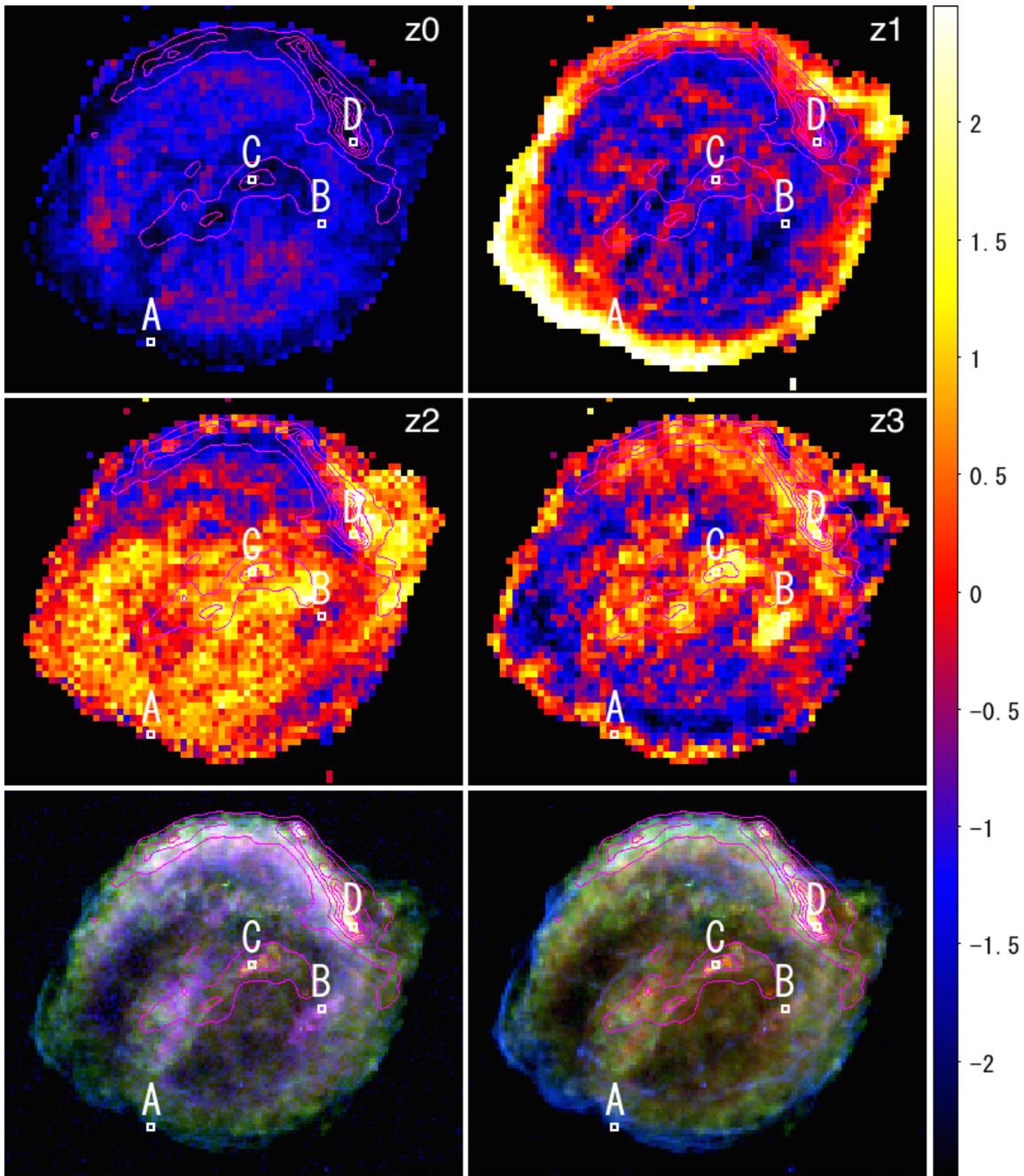


Figure 6.3.8: In the upper four panels, images showing the  $z$  of VAE latent parameters for each axis when the merged data are input. The images of the  $z$  share a colour scale. In the bottom left panel, the three-colour image showing the flux images of Fe L band, 0.72–0.95 keV (red), Si K  $\alpha$  band, 1.65–2.06 keV (green), and Fe K  $\alpha$  band, 6.0–7.0 keV (blue). In the bottom right panel, the three-colour image showing the flux images of O band, 0.4–0.72 keV (red), Si K  $\alpha$  band, 1.65–2.06 keV (green), and continuum band, 4.2–6.0 keV (blue). The *Spitzer* MIPS 24  $\mu$ m observation in 2004 is shown on each panel as magenta contours tracing warm dust of shocked CSM.

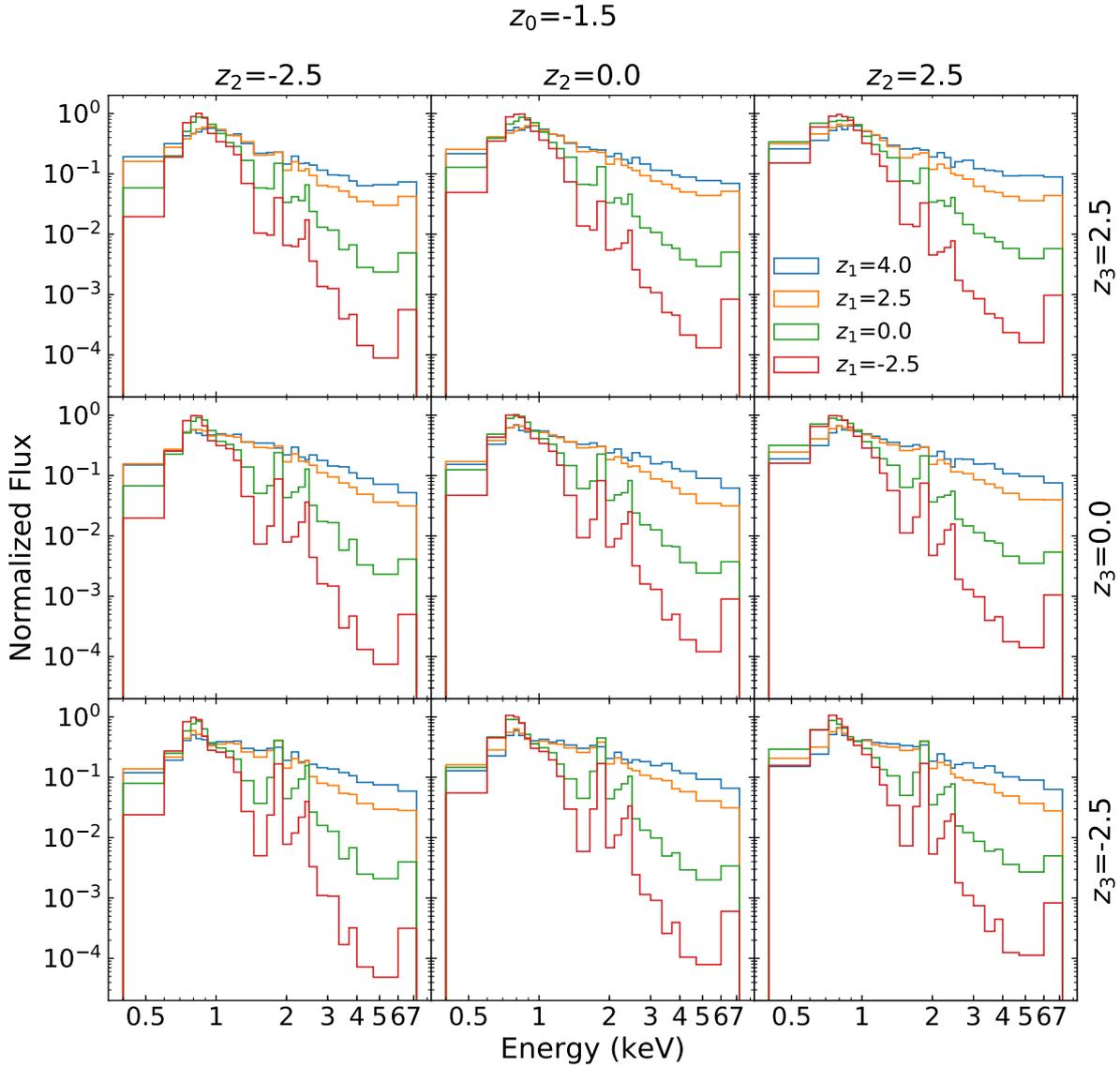


Figure 6.3.9: Generated spectra by the decoder after 100 epochs training. The variation of  $z_1$  shows in each panel, when  $z_0$  is fixed -1.5.

The spectrum from regions of  $z_0 \leq -1\sigma$  had stronger O and Mg line features and a softer continuum than the spectrum from regions of  $z_0 \geq 1\sigma$ , which had strong Fe L emission lines. The spectrum from regions of  $z_2 \geq 1\sigma$  also had stronger O and Mg line features than the spectrum from regions of  $z_2 \leq -1\sigma$ , which had strong Fe L emission lines. It is consistent with the results in Section 6.3.4. The regions of  $z_0 \leq -1\sigma$  are spatially located on the edges and also in the central small parts of the SNR. The regions of  $z_2 \geq 1\sigma$  are spatially located at the west and the centre of the SNR and overlapped with infrared contours. By contrast, the regions of  $z_0 \geq 1\sigma$  are located at inner parts of the SNR. The regions of  $z_2 \leq 1\sigma$  were mainly at the northern rim and southern parts of the SNR. It seems that  $z_0$  and  $z_2$  represent thermal plasma features, e.g., discriminating the shocked CSM and ejecta, and supports the interpretations in Section 6.3.4.

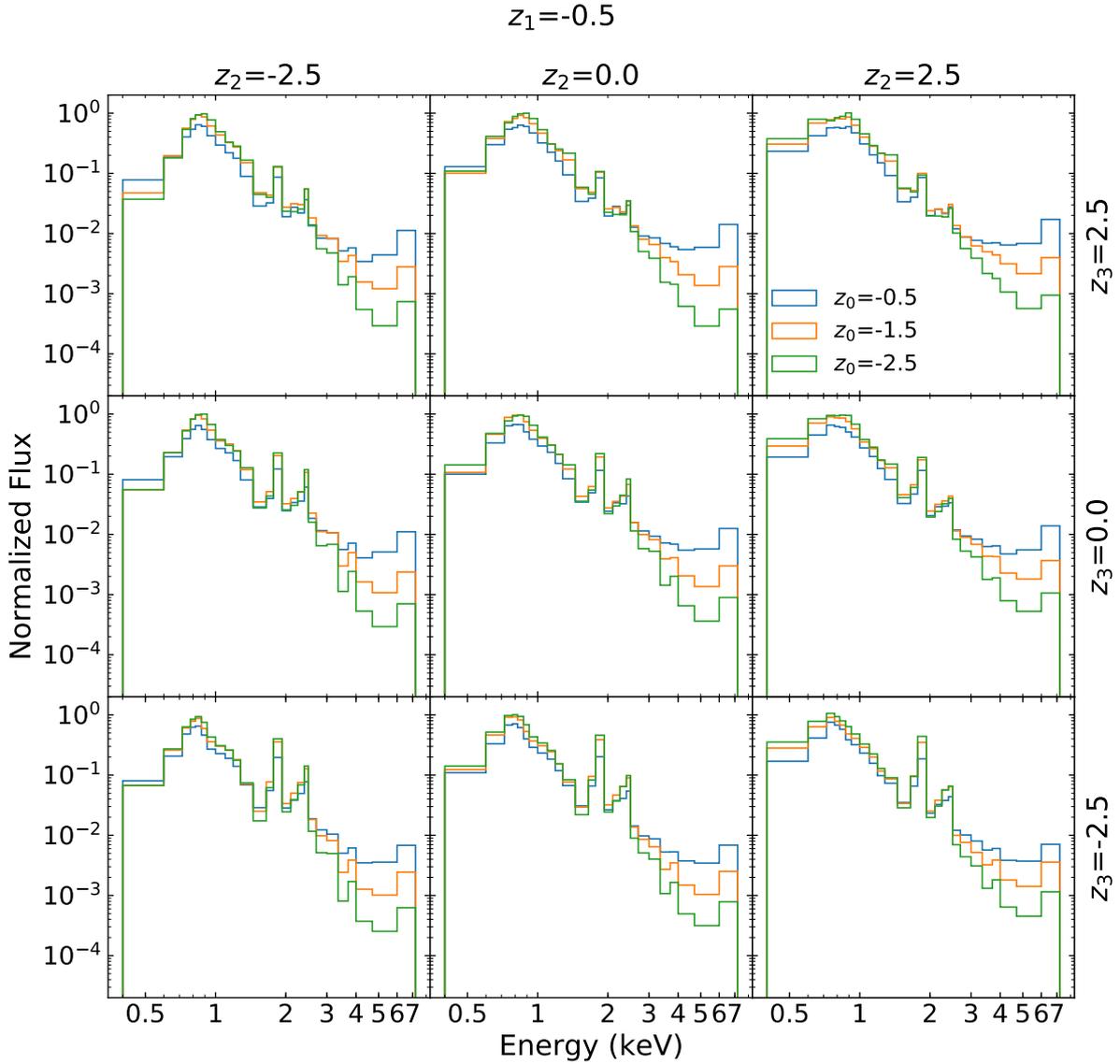


Figure 6.3.10: Generated spectra by the decoder after 100 epochs training. The variation of  $z_0$  shows in each panel, when  $z_1$  is fixed -0.5.

The spectrum from regions of  $z_3 \leq -1\sigma$  had strong IME lines and Fe L peaks in lower energy than that from regions of  $z_3 \geq 1\sigma$ . It supports the interpretation of  $z_3$  mentioned in Section 6.3.4. The regions of  $z_3 \leq 1\sigma$  were mainly at the inner side of the rim, while the regions of  $z_3 \geq 1\sigma$  were at north-western (NW) inner rim and central parts.

### 6.3.6 Results of GMM

We applied Gaussian mixture model (GMM) clustering to the latent expressions obtained by inputting the source bins of the merged data. The optimal number of clusters was determined to be 12 by checking the Bayesian information criterion (BIC) by changing the number of clusters from 3 to 20.

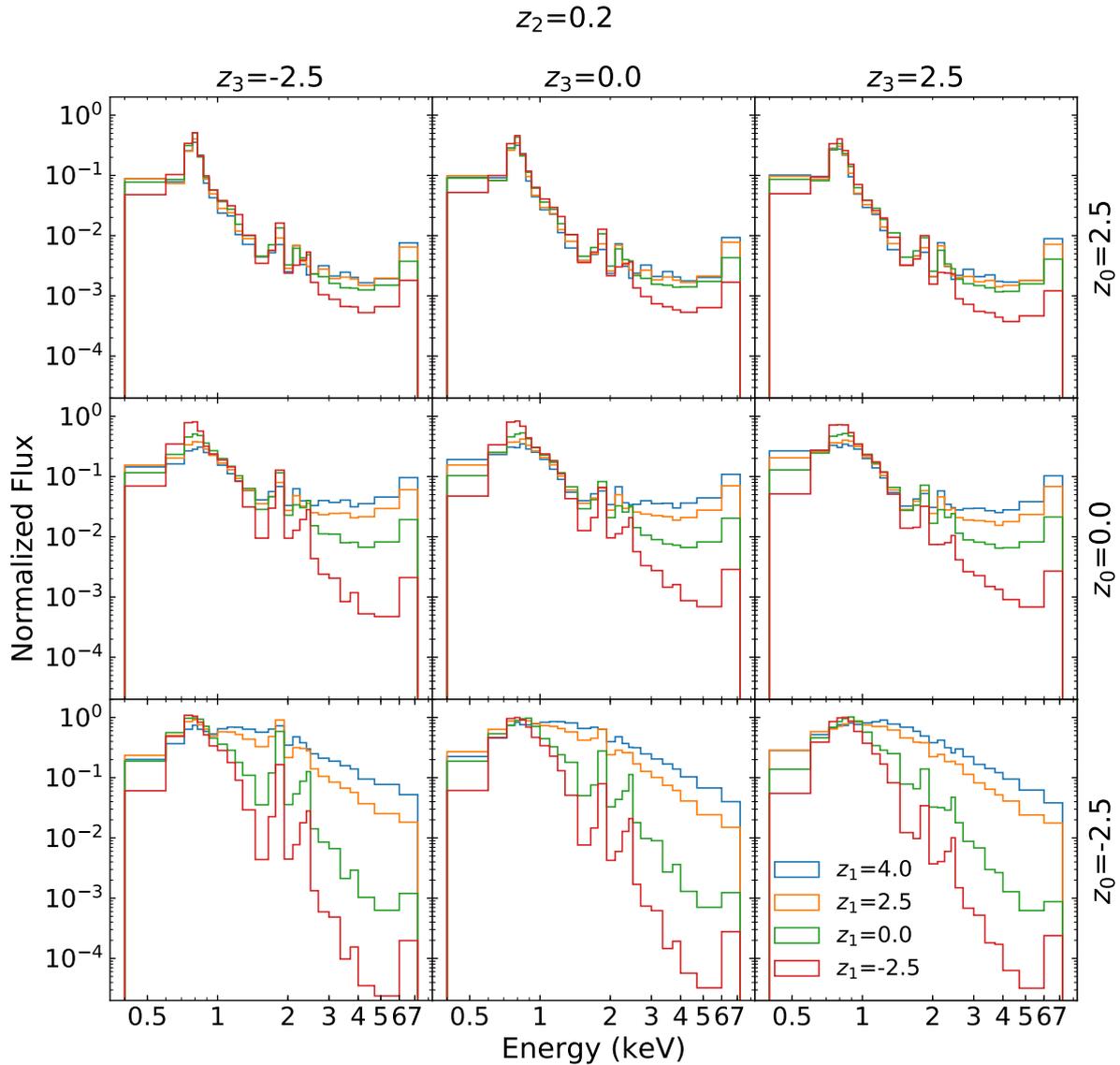


Figure 6.3.11: Generated spectra by the decoder after 100 epochs training. The variation of  $z_1$  shows in each panel, when  $z_2$  is fixed 0.2.

The scatter plots in Figure 6.3.15 show the distributions of the latent variables  $\mu$  same as the source bins in Figure 6.3.4, which are colour-coded according to the category assigned by GMM clustering.

Each panel of Figure 6.3.16 shows the responsibility of each GMM category. We show the division of *Kepler's* SNR into GMM categories in the top left-hand panel of Figure 6.3.17. The top right-hand panel of Figure 6.3.17 shows the same image as the left-hand of it, but the spatial bins whose assigned category has a responsibility of  $\leq 80$  per cent are masked. Hence, the spatial bins that remain coloured in the top right-hand panel of Figure 6.3.17 are robustly assigned to the category. It also means that those bins are expected to have some spectral features distinct from those of the other categories. We describes the individual spectral features later (summarized in Table 6.3.1).

The middle left-hand panel of Figure 6.3.17 shows the  $H\alpha$  emission observed by WFC3 on *HST*

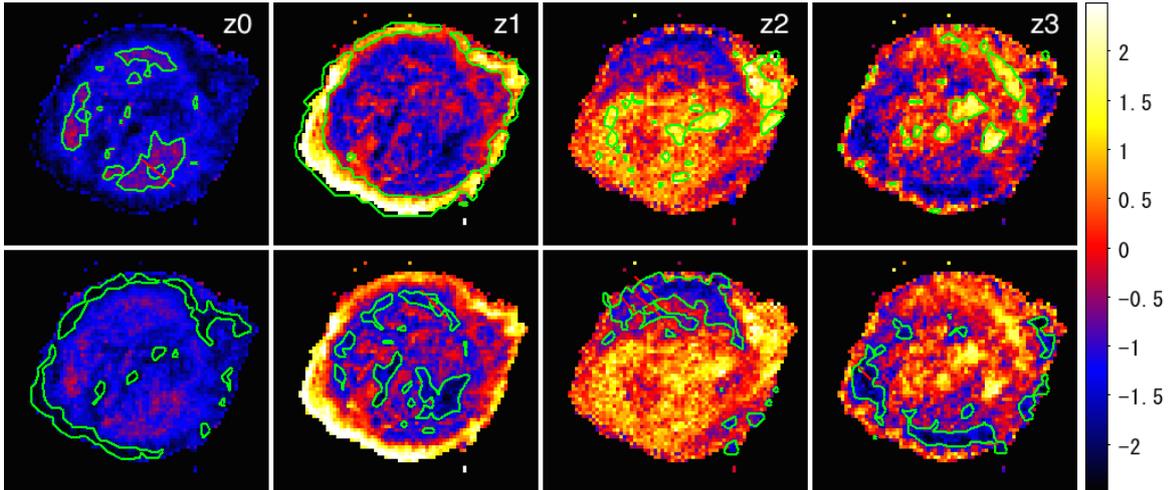


Figure 6.3.12: Images of  $z$  for each axis obtained by inputting of merged data. In the upper or lower eight panels, the contours correspond to the  $+1\sigma$  and  $-1\sigma$  level for each latent axis, respectively. The Background region is masked in each panel.

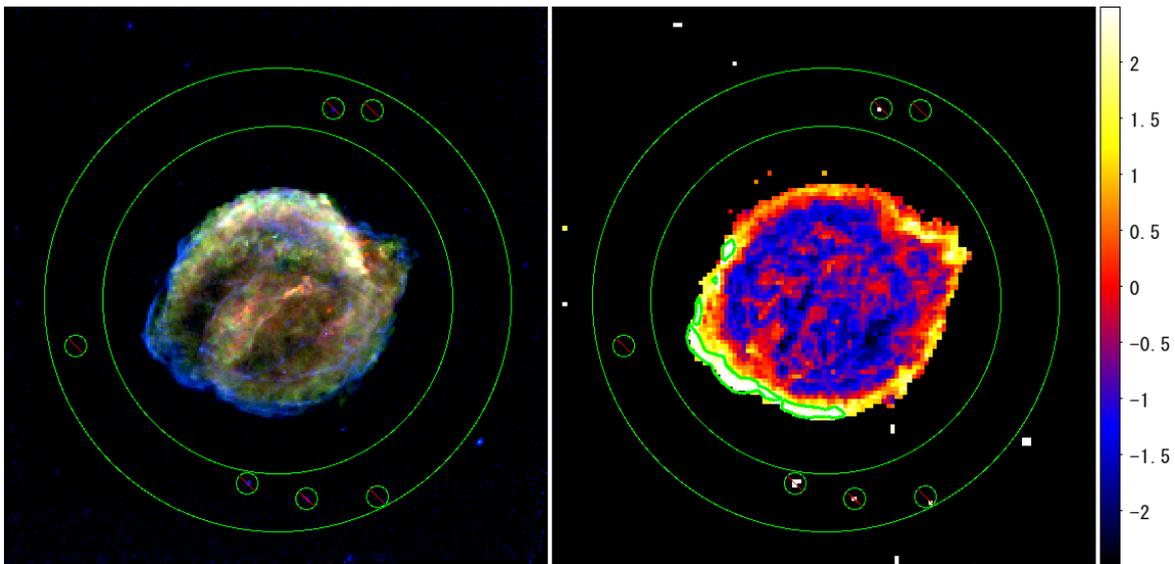


Figure 6.3.13: The left-hand panel is the three-colour image, which is same as the bottom right panel of Figure 6.3.8. The right-hand panel show  $z_1$  obtained by inputting of merged data, and masked the background region. The contours correspond to the  $+3\sigma$  level. In each panel, the green line annulus shows the region for background subtraction.

in 2013. The middle right-hand panel of Figure 6.3.17 shows the  $24\ \mu\text{m}$  infrared image observed by MIPS on *Spitzer* in 2004 tracing warm dust. On each panel of Figure 6.3.17, the warm dust emission is overlapped as contours. The warm dust emission traces the dense CSM heated by shocks (Blair et al., 2007; Williams et al., 2012).

For comparison, traditional three-coloured images are shown in the bottom row of Figure 6.3.17. In the left-hand panel, Fe L blend band,  $0.72\text{--}0.95\ \text{keV}$  is shown as red; Si K  $\alpha$  blend,  $1.65\text{--}2.06\ \text{keV}$  as green; and Fe K  $\alpha$  blend,  $6.0\text{--}7.0\ \text{keV}$  as blue. In the right-hand panel, O and Fe L band,  $0.4\text{--}0.7\ \text{keV}$

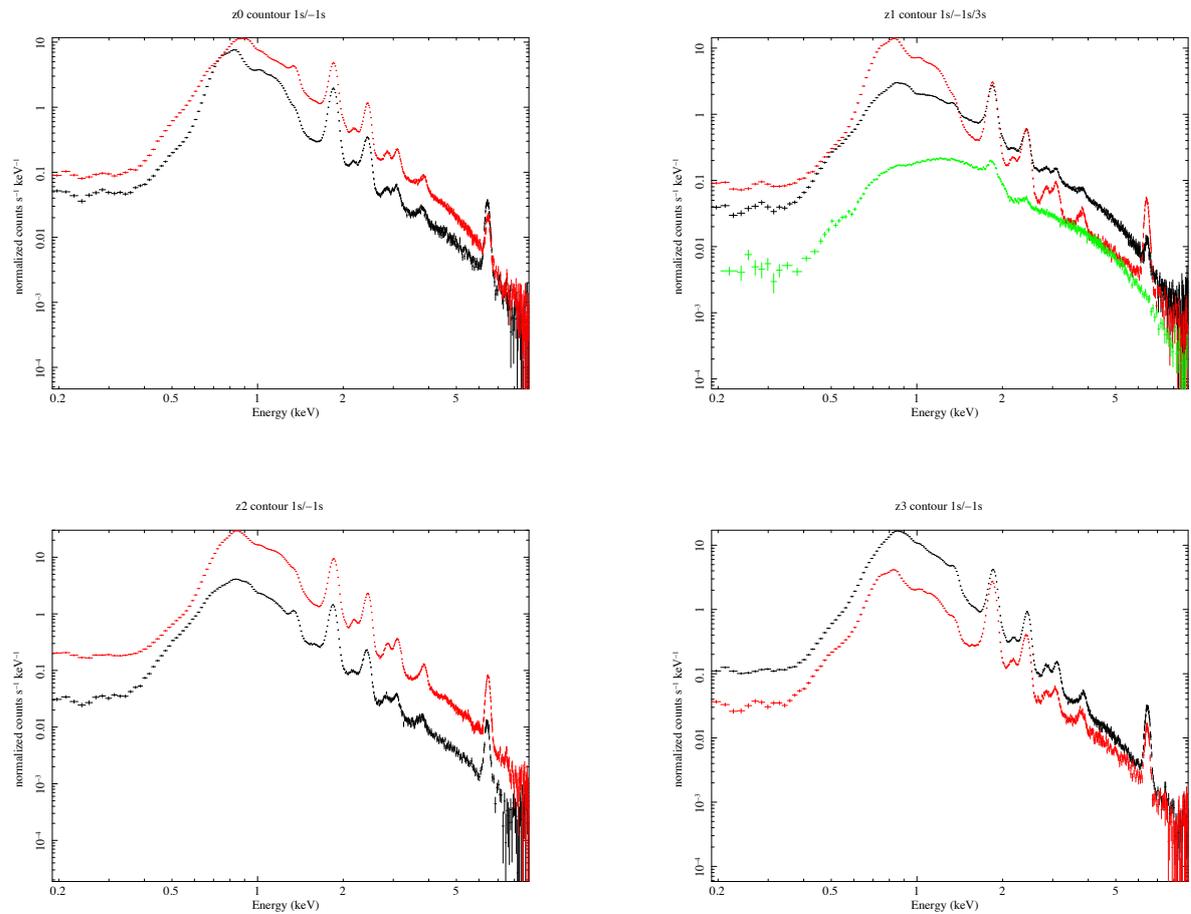


Figure 6.3.14: The black and red spectra were extracted from regions more than  $+1\sigma$  and less than  $-1\sigma$ , respectively, in each VAE  $z$  axis. The green spectrum was extracted from regions more than  $+3\sigma$  for VAE  $z_1$  axis.

is shown as red; Si K  $\alpha$  blend, 1.65–2.06 keV as green (same as the left-hand panel); and continuum dominated band, 4.2–6.0 keV as blue.

We calculated the mean of VAE  $z$  values over spatial bins whose responsibilities were  $\geq 80$  per cent for each GMM category, and also input it to the decoder for generating a spectrum. Meanwhile, a spectrum was also generated by inputting each GMM cluster centroid to  $z$  of the decoder. As shown in Figure 6.3.18, these two spectra were almost same for each category. We describe the interpretation of each category later.

We show the combined spectrum extracted from the regions whose responsibilities were  $\geq 80$  per cent for each GMM category in Figure 6.3.19. The background was extracted from the annular region defined as Figure 6.3.13 and subtracted from the spectra.

We interpreted the feature of each category and summarized in Table 6.3.1.

As shown in Figure 6.3.17, the bins assigned to category 0 are spatially located in the outer edge of northern part of the SNR, especially outside of the category 4 or 6. The spectrum has lower counts than the other categories because the region is smaller and has a lower flux.

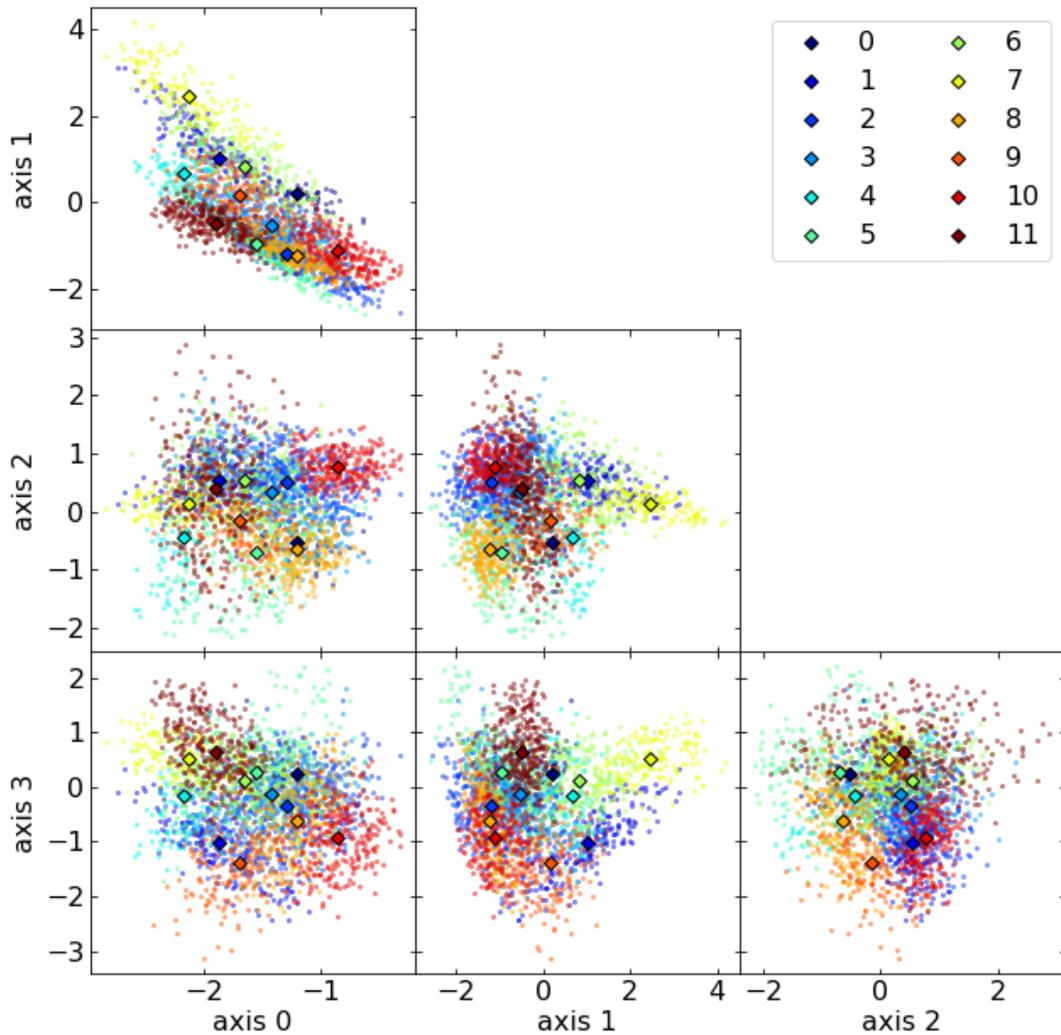


Figure 6.3.15: Scatter plot showing the  $\mu$  of VAE latent parameter obtained by inputting each spatial bin of the merged data, where two of the four axes are chosen. Each point was colour-coded for each category classified by GMM clustering. The centroid of each category is shown as an open black diamond.

The bins assigned to category 1 is mainly located the inner side of the front shock of the eastern and south-eastern (SE) part of the SNR. The category 1 is also located in the outer edges of the western and NW part of the SNR.

As shown in Figure 6.3.16, the category 2 extends inside the SNR with lower responsibility than some other categories excluding the NW part of the SNR.

The category 3 is spatially located inside of the SNR. As shown in Figure 6.3.17, this category is inside of the categories 7, 8, and 9 in the south-western (SW) part of the SNR. Meanwhile the bins are surrounded by the regions of category 8 in the northern part of the SNR. The category has relatively high continuum emissions.

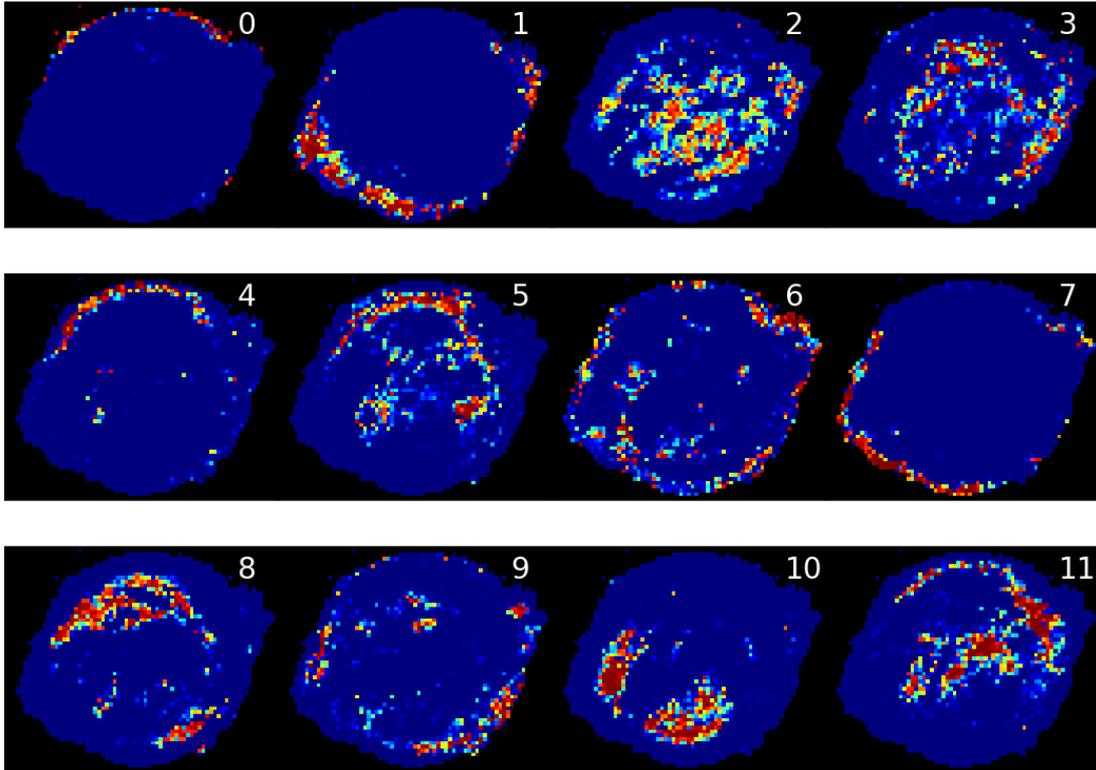


Figure 6.3.16: GMM responsibility for each spatial bin of merged dataset for each category. The responsibility is between 0 and 1; blue and red represent a responsibilities of 0 and 1, respectively.

Table 6.3.1: GMM categories.

Category No.	Location	Feature
0	outer edges of SNR	faint emission
1	inner side of FS	dark region, weak Fe lines
2	inner side of SNR	weak emission
3	rim of SNR	strong continuum
4	northern rim	strong Si lines
5	northern rim; clumps inside SNR	strong Fe L lines
6	outer edges; blobs inside of SNR	strong continuum and weaker emission lines
7	FS	synchrotron dominant
8	rim of SNR	strong Fe K $\alpha$ and weaker continuum
9	outer side of ejecta rim	weaker Fe lines
10	SW part inside SNR	Shocked ejecta
11	NW rim and central clumps	shocked dense CSM

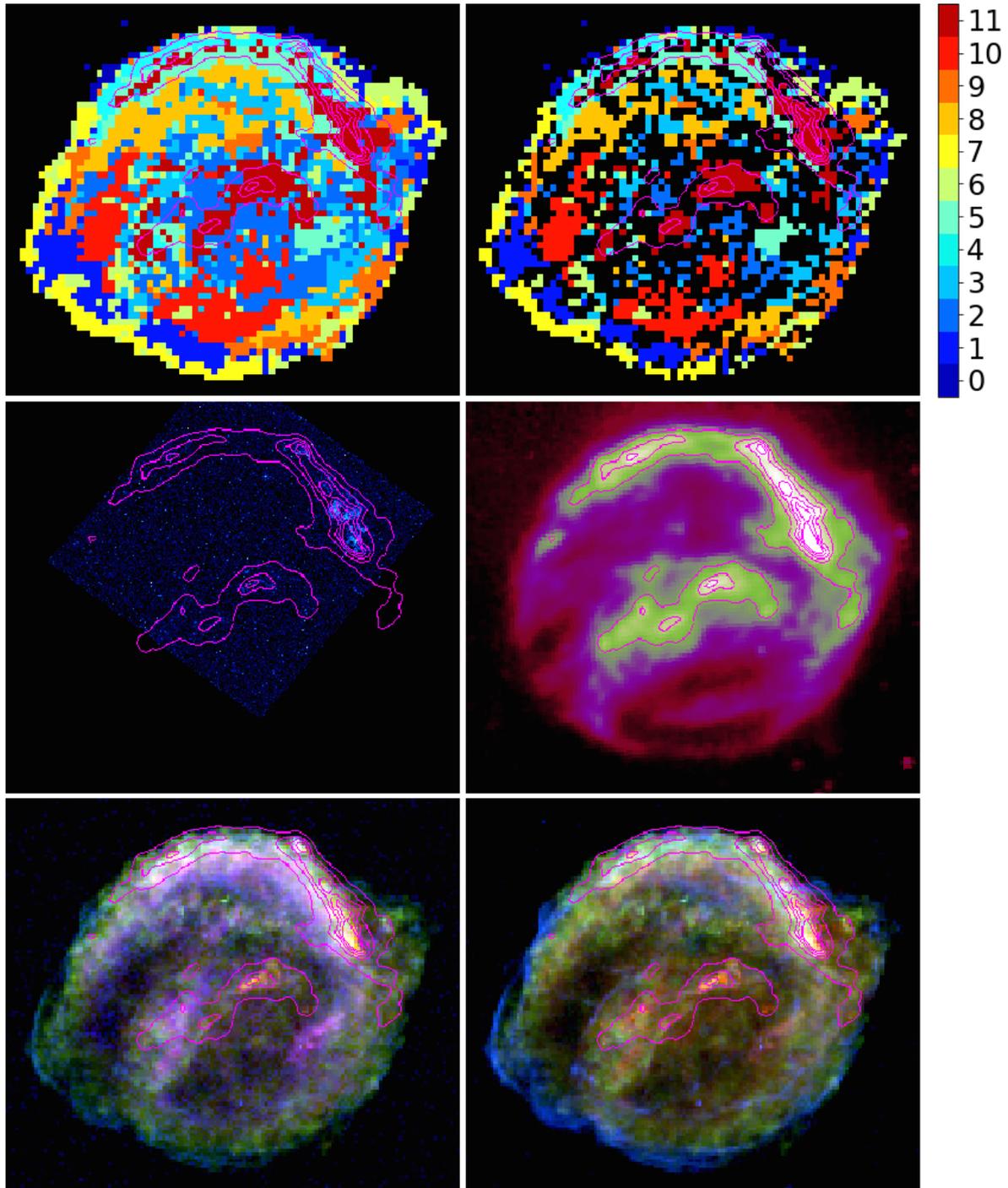


Figure 6.3.17: The top left-hand panel shows the GMM category with the highest responsibility for each spatial bin of the merged data. The top right-hand panel shows only spatial bins with a category of more than 80 per cent responsibility, which are assigned the colours representing the category. The other spatial bins, which have responsibilities less than 80 per cent, appear black. The middle left-hand panel shows the H  $\alpha$  image observed by WFC3 on *HST* in 2013. The middle right-hand panel shows the warm dust emissions, 24  $\mu$ m infrared observed by MIPS on *Spitzer* in 2004. The bottom panels are same as the bottom panels in Figure 6.3.8. The bottom panels are same as the three-colour images in Figure 6.3.8. On each panel, the *Spitzer* MIPS 24  $\mu$ m observation in 2004 is shown as contours tracing warm dust of shocked CSM.

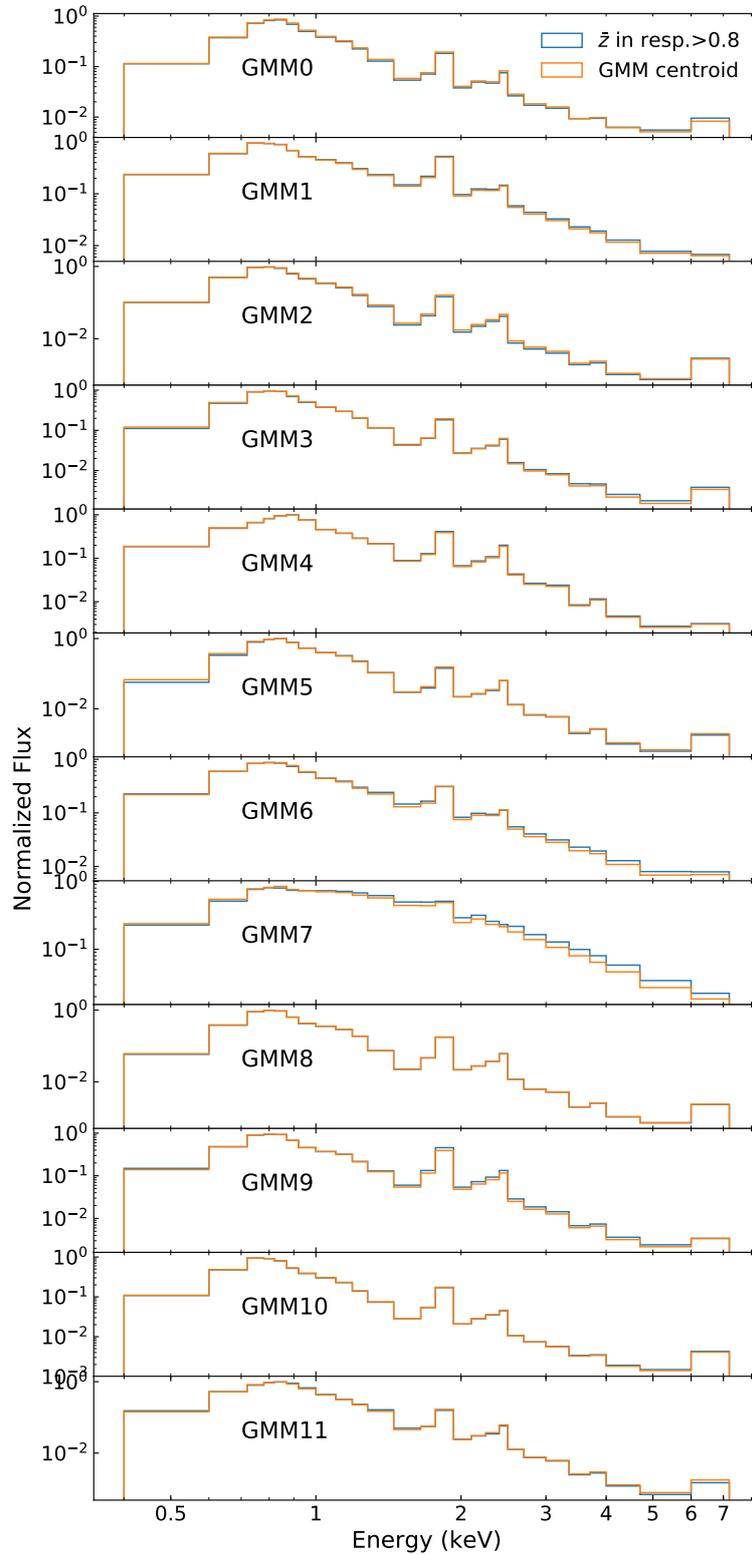


Figure 6.3.18: Spectra generated by the decoder from given  $z$  values, which are the latent vector  $\bar{z}$  averaged over the spatial bins having the responsibility more than 80 per cent and the category centroid for each GMM category. The vertical axes have arbitrary unit because the bins were divided the maximum of each spectrum for the VAE training.

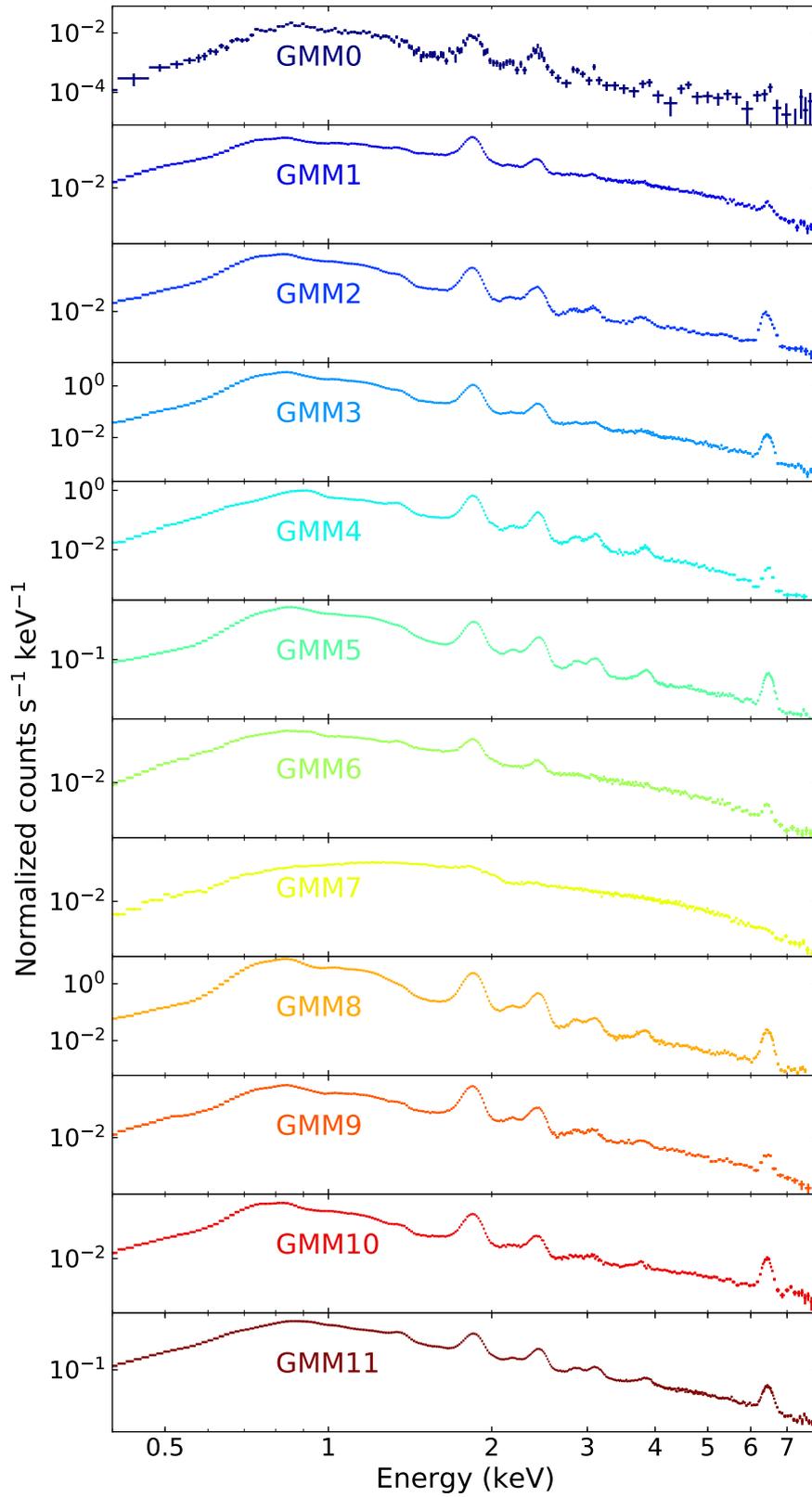


Figure 6.3.19: Spectrum (background-subtracted) of the region inside the SNR for which each GMM category has a responsibility more than 80 per cent.

In the northern rim of the SNR, the categories 4, 11, 5, 8, and 3 appear as a layered structure from the outer side of the SNR to the inner side as shown in Figure 6.3.17.

The category 4 is spatially located in the outer edge of the northern rim. The bins of the northern part are outside of the brightest Fe L layer.

The category 5 is spatially located inside the categories 4 and 11 in the northern layered structure and also the SW and SE bright spots. As shown in Figure 6.3.19, the spectrum has strong Fe emission lines.

The category 6 coincides with the outer edges of the SNR, which have strong continuum emissions e.g., a part of the western ‘ear’ as shown in Figure 6.3.17. This category also coincides with the blobs with strong continuum emission inside the SNR. The bins of category 6 tends to be adjacent to the category 7.

The bins assigned to the category 7 are spatially located in the edge of the SNR, mainly in the eastern half. The eastern half of category 7 approximately corresponds to the regions more than  $+3\sigma$  for VAE  $z_1$  axis in Figure 6.3.13. In our case, we also found a consistent result that the spectrum of category 7 is dominated by synchrotron emission as shown in Figure 6.3.19.

The category 8 is spatially adjacent inner side of the category 5 in the northern layered structure, and also inside of the category 9 in the southern rim of the SNR. The bins of the northern part spatially coincide with the layer with strong Fe K  $\alpha$  emissions inside of the brightest Fe L layer represented by the category 5. Furthermore, the continuum emission is weaker than the bins surrounding the category 8.

The category 9 is adjacent to the outer side of the categories 3 and 8 on the outer rim of the SNR, while some bins are at the inner side of the category 8 in the northern part of the SNR. Further, a clump of category 9 is in the western ‘ear’. These region have strong Si K lines but weaker Fe emission.

As shown in Figure 6.3.16 the category 10 is located almost only in the eastern and the SE regions inside the SNR. The category 10 spatially coincides with the region, which has the weak dust emission or the weak X-ray continuum as shown in Figure 6.3.17. These regions seem to be less associated to dense CSM. Therefore the spectra mainly arise from the shocked ejecta.

The regions of category 11 spatially coincide with the dust emission as shown in Figure 6.3.17. Thus the category 11 seems to be classified as the spectra sufficiently containing the emission from shocked CSM. The category 11 also corresponds to the part of category R (red) in Burkey et al. (2013), which contains the bulk of the CSM.

## 6.4 Discussions

In this work, we used the Poisson log likelihood for the reconstruction loss to train the VAEs. As shown in Figure 6.3.7, the VAE has successfully generated the input spectra for both the high count cases and the low count case. In the low count spectrum, the VAE removed the fluctuations caused by low statistics and generate a smooth spectrum. It shows that the Poisson loss can efficiently lead the VAE training with count data.

We implemented the preprocessing method to normalize each sample by the maximum value of each

sample and compared the training results with or without this spatial-bin-by-bin normalization. We found that the the distribution of the data samples in the latent space is based on only spectral shape. For SNR spectra, this preprocessing is effective because the shape, such as line intensity ratios, ratios of lines and the continuum, the indices and cut-off energies of continuum, is important. On the other hand, in other cases, e.g., a galaxy cluster, the X-ray emissivities corresponding to the plasma densities are essential. The data preprocessing needs to be selected according to the class of source and the focusing physical features.

We demonstrated the interpretation of the features represented by the latent axes using the spectra generated from the latent parameters by the decoder. We found that the  $z_0$  determined the global spectral shape with Fe L, Fe K  $\alpha$ , and IME line blends.

The  $z_1$  represented the ratio of the emission lines and the continuum component, and the featureless spectra as the data sample distribution made a long tail.

We found that the  $z_2$  makes N Ly  $\alpha$  and O Ly  $\alpha$  blends stronger, while the axis makes S K  $\alpha$  blend weaker in the southern half of the the SNR. It is consistent with the previous result (Katsuda et al., 2015). Thus, we concluded that the  $z_2$  represent the line ratio of shocked CSM and shocked ejecta.

The  $z_3$  determined the intensities of IME lines and the peak energy of Fe L blend, which may be originated by the electron temperature or ionization state.

As summarized in Section 6.3.6, we obtain the result of unsupervised classification by the method combining the VAE and the GMM. Our method revealed synchrotron dominated FS, the layered structure in the northern rim, and the region interacting dense CSM.



## Chapter 7

# Demonstration with SN 1006

### 7.1 Introduction for SN 1006

SN 1006 is one of historical supernova remnants (SNRs), and is known to be a type Ia explosion. The distance to SN 1006 is  $\sim 2.2$  kpc (Winkler et al., 2003), and the apparent diameter is approximately 30 arcmin. Therefore, the observation can reveal spatial structures in detail.

The absorbing column density towards SN 1006 is low ( $N_{\text{H}} = 6.8 \times 10^{20} \text{ cm}^{-2}$ ; Dubner et al., 2002) among the Galactic SNRs because of the high latitude of  $14^\circ 6'$ . Thus, the emission lines from low-atomic-number elements e.g., oxygen, are clearly observed in the soft X-ray band.

The surroundings of SN 1006 is homogeneous and relatively low density (averaged atomic density  $\sim 0.3 \text{ cm}^{-3}$ ; Dubner et al., 2002). Thus, the evolutionary state of the SNR may still be the earliest of the young Galactic SNRs (like *Tycho's* SNR and *Kepler's* SNR), although a thousand years has been passed since the explosion. In the SNR, most of oxygen atoms have not been completely ionized and are emitting strong X-ray lines.

In SN 1006, the shock heating by reverse shock (RS) is on going, and the RS has just recently been reached the outer part of Fe rich ejecta in the SNR interior (e.g.; Yamaguchi et al., 2008; Uchida et al., 2013; Li et al., 2016). Yamaguchi et al. (2008) detected Fe K  $\alpha$  line emission for the first time in the SNR, and found that the Fe ionization state is much lower than the other thermal component from *Suzaku* observations. This suggests that Fe is located inner side of the SNR and has been heated by the RS more recently than the other elements.

Uchida et al. (2013) found that the ejecta distribution of O-burning and incomplete Si-burning elements (i.e., Si, S, and Ar) is asymmetric, while the distribution of C-burning elements, i.e., O, Ne, and Mg, is relatively uniform in the SNR interior, and hence suggested an asymmetric SN explosion. *XMM-Newton* observations also confirmed the asymmetry, which the elemental abundances of heavy elements e.g., Si, and Fe higher in the SE of the SNR than the north-western (NW) (Li et al., 2016, 2015).

Li et al. (2016, 2015) also found the ejecta of 'dark belt' in the SE of the SNR interior has lower ionization state than the surrounding regions.

SN 1006 has a bi-polar global structure, i.e., the two prominent shells located in the north-eastern (NE) and south-western (SW) quarter of the SNR. Koyama et al. (1995) found that the X-ray spectrum of the

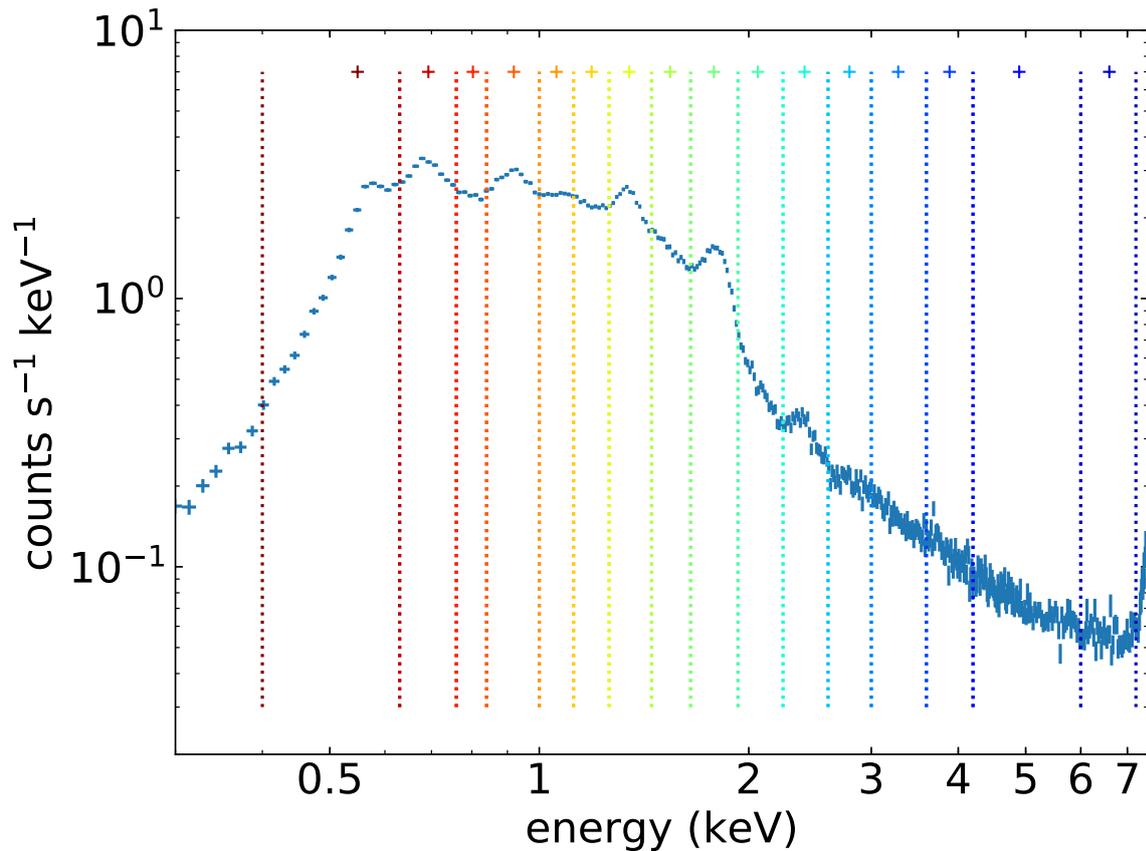


Figure 7.2.1: Spectrum of entire SN 1006. The vertical dotted lines are boundaries of the narrow energy bands. The cross represents the centre of gravity of each energy band.

two bright limbs is synchrotron dominated featureless one. This result provided the first clear evidence for diffusive shock acceleration of charged particles, that is the evidence for cosmic-ray acceleration in SNR shocks. Using radio polarization observations, Reynoso et al. (2013) found that the magnetic field surrounding the SNR is aligned with a direction from NE to SW, and the two polar limbs are highly polarized.

## 7.2 *Chandra* ACIS Data Set

SN 1006 was observed by the ACIS of *Chandra* for 68.09 ks, 88.98 ks, 217.56 ks (eleven obsIDs), 68.87 ks, and 669.85 ks (ten obsIDs) in 2000, 2001, 2003, 2008, and 2012, respectively, as summarised in Table 7.2.1.

SN 1006 has approximately 30' diameter, which is larger than the FoV of ACIS (approximately 16'×16' for ACIS-I, 8'×50' for ACIS-S). In 2003 and 2012, the entire SNR was covered by approximately ten FoVs of ACIS-I. The rims of SNR were also observed by ACIS-S in 2000 and 2008 for the NE rim, and in 2001 and 2012 for the NW rim.

The frontside-illuminated chips of ACIS-I have covered the entire SNR, although the backside-illuminated chips of ACIS-S have observed only some parts of NE and NW rims. The backside-illuminated chip has a different instrumental response and non-X-ray background (NXB) spectrum from frontside-illuminated chips of ACIS-I, thus we decided to use only the datasets observed by the frontside-illuminated chips.

We performed X-ray analysis using CIAO (version 4.11; Fruscione et al., 2006) and CalDB (version 4.8.3) provided by the *Chandra* X-ray Center<sup>1)</sup>.

Table 7.2.1: Observations of SN 1006 by *Chandra*

ObsID	Exposure (ks)	Date	Instruments	Positions
732	68.09	2000 Jul 10	ACIS-S1–4, I2,3	NE
1959	88.98	2001 Apr. 26	ACIS-S1–4, I2,3	NW
3838	20.13	2003 Apr. 08	ACIS-I0–3, S3	NE
4385	19.79	2003 Apr. 08	ACIS-I0–3, S3	SE
4386	19.79	2003 Apr. 08	ACIS-I0–3, S3	S
4387	19.79	2003 Apr. 09	ACIS-I0–3, S3	SW
4388	19.79	2003 Apr. 09	ACIS-I0–3, (S3)	W
4389	19.79	2003 Apr. 09	ACIS-I0–3, (S3)	NW
4390	19.8	2003 Apr. 09	ACIS-I0–3, (S3)	N
4391	19.91	2003 Apr. 10	ACIS-I0–3, (S3)	N
4392	19.59	2003 Apr. 10	ACIS-I0–3, S3	N inner
4393	19.59	2003 Apr. 11	ACIS-I0–3, S3	Central
4394	19.59	2003 Apr. 11	ACIS-I0–3, S3	Central S
9107	68.87	2008 Jun. 24	ACIS-S1–4, I2,3	NE
13737	87.09	2012 Apr. 20	ACIS-S1–4	NW
13738	73.47	2012 Apr. 23	ACIS-I0–3	W
14424	25.39	2012 Apr. 27	ACIS-I0–3	W
13739	100.07	2012 May 04	ACIS-I0–3	SW
13740	50.41	2012 Jun. 10	ACIS-I0–3	Central-W
13741	98.48	2012 Apr. 25	ACIS-I0–3	SE
13742	79.04	2012 Jun. 15	ACIS-I0–3	S
14423	25.02	2012 Apr. 25	ACIS-I0–3	Central
13743	92.56	2012 Apr. 28	ACIS-I0–3	N
14435	38.32	2012 Jun. 08	ACIS-I0–3	Central-E

As shown in Figure 7.2.1, the spectrum of the entire SN 1006 was created in the 0.4–7.2 keV band and was divided into 16 narrower energy bins such that each clearly resolved emission line is divided into an independent energy bin.

<sup>1)</sup> Available at <http://cxc.harvard.edu>

Because SN 1006 is darker and larger than *Tycho*'s SNR or *Kepler*'s SNR, we decided to divide both the energy and space into wider bins than what are done in the previous chapters. In the low-energy side (0.4–2.6 keV) with a relatively high count rate, we divided the energy band so that each band corresponds to each emission line structure clearly appearing in the spectrum. By contrast, the spectrum in the high-energy side has no clear emission line features. The energy band above 7.2 keV is dominated by NXB, thus we did not use the band.

The flux values in the 16 narrow energy bands corresponding to a single spatial bin are combined, and the resulting 16-dimensional vectors are used as the input dataset.

The flux image of each band was created with the coordinate ranges set to omit regions outside of the SNR. We prepared three spatial bin sizes, 31.5 arcsec, 15.7 arcsec, and 7.87 arcsec, resulting in image sizes of  $67 \times 66$ , and  $131 \times 131$  spatial bins,  $261 \times 261$  spatial bins, respectively.

We did not subtract the backgrounds from the images following the case of *Tycho*'s SNR and *Kepler*'s SNR.

The flux images whose spatial bin size is 15.7 arcsec of all individual observations and the merged images of them were used for training. In addition, we also used the merged images whose spatial bin size is 7.87 arcsec and the images of each individual obsID whose spatial bin size is 31.5 arcsec to increase the training data size. We removed the spatial bins whose 0.4 keV and 7.2 keV photon fluxes were less than a given threshold in order to exclude the samples outside of the SNR. The threshold of photon flux were  $6.7 \times 10^{-6}$ ,  $2.0 \times 10^{-6}$ , and  $6.0 \times 10^{-7}$  photons  $\text{s}^{-1} \text{cm}^{-2}$  for the spatial bin size 31.5 arcsec, 15.7 arcsec, and 7.87 arcsec, respectively.

Twenty per cent of the spatial bins in each flux image were chosen randomly and used as validation data, and the rest were used as training data. The actual size of the training and validation datasets were 236304 and 59077, respectively, excluding the spatial bins with zero flux in all the narrow energy bands (i.e., a 16-dimensional zero vector). All the observations with the spatial bin size of 15.7 arcsec were summed and used for the post-training analysis.

## 7.3 Dimensionality Reduction by VAE

### 7.3.1 The VAE Architecture

We applied variational autoencoder (VAE) to dimensionality reduction of the SN 1006 dataset. At first, we examined some VAE models having different number of fully-connected (FC) layers of encoder or decoder (2 or 3), or different number of dimensions in the latent variables (4 or 8 dimensions) or in the hidden layers (64, 128 or 256 nodes per layer). We trained each models through 100 epochs, using Poisson log likelihood for the reconstruction losses as described in Section 4.1.2. For these training, each data sample was normalized by the maximum value of each sample in order to focus only on the spectral shape as described in Section 4.1.2.

For the models whose encoders and decoders have 2 FC layers containing 256 nodes in each layer, three latent axes of  $z$  seem to have means, although the rest axes of  $z$  seem to be random values and

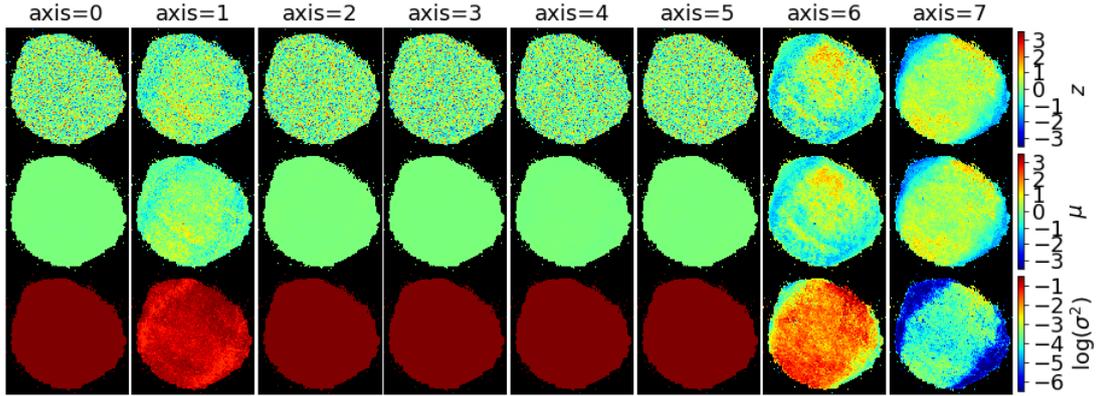


Figure 7.3.1: Images showing the values of the VAE latent parameters  $\mathbf{z}$ ,  $\boldsymbol{\mu}$ , and  $\log(\sigma^2)$  in the top, middle, and bottom row as an image for each axis when the merged data (a 16-colour image) are input into a model having 8-d latent variables and 256-d neurons in each FC layer after 100 epochs training. Each data sample was normalized a maximum value of each spatial bin when it was input to the VAE. The images in each row share a colour scale.

are not committed to informative dimensionality reductions. In Figure 7.3.1, we show an example of the latent parameters of  $\mathbf{z}$ ,  $\boldsymbol{\mu}$ , and  $\log(\sigma^2)$  for each axis when the spatial-bin-by-bin normalized merged data are input. The axes 1, 6 and 7 correspond to meaningful images of latent variables, while each spatial bin for the other axes of  $\mathbf{z}$  seems to be random values.

For the other models, only two latent axes of  $\mathbf{z}$ , which are similar to the axes 6 and 7 in Figure 7.3.1, are meaningful, although the rest of the axes of  $\mathbf{z}$  seem to be meaningless. Hence, we want to examine what the extra axis represents.

For the models with 3 meaningful latent axes, the model with 8 dimensional  $\mathbf{z}$  have better loss function. We analyse the results obtained by this architecture.

### 7.3.2 The Latent Variables

In Figure 7.3.2, we show the latent parameters of  $\mathbf{z}$ ,  $\boldsymbol{\mu}$ , and  $\log(\sigma^2)$ , which are same parameters as shown in Figure 7.3.1. The distributions of the source and background bins for axes 6 and 7 of  $\mathbf{z}$  and  $\boldsymbol{\mu}$  have different peaks. The background bins have not been used for training. For the axis 1,  $\mu_1$  has a narrow distribution like the axes 6 and 7, meanwhile the peak of  $\log(\sigma_1^2)$  are larger than these of axes 6 and 7. Thus  $z_1$  has broader distribution than  $\mu_1$ .

We modelled the latent variable distribution of axes 1, 6 and 7 for the source bins using Gaussian as shown in Figure 7.3.3. Although the distributions were distorted, we decided to use the parameters of Gaussian as the approximate centre and standard deviation of each data distribution. The positions corresponding to  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  of the best fitting Gaussian are shown as the vertical dotted lines in each panel of Figure 7.3.3. The axes 1 and 6 of  $\boldsymbol{\mu}$  and  $\mathbf{z}$  had approximately symmetric distributions although the axis 7 had a long tail beyond  $-3\sigma$  level on the negative side. The negative tail corresponds to the NE and SW rims of the SNR as shown in the panels of axis 7 of Figure 7.3.1. The rims have synchrotron

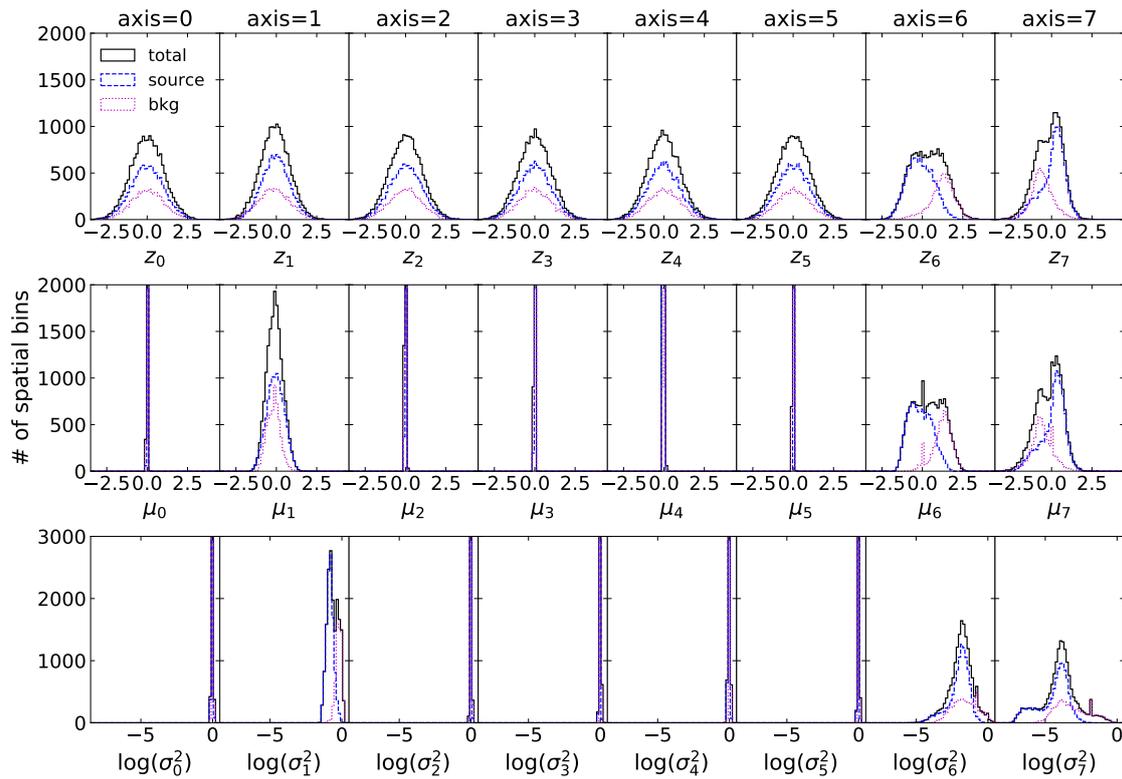


Figure 7.3.2: Histogram showing each axis of  $\mathbf{z}$ ,  $\boldsymbol{\mu}$ , and  $\log(\sigma^2)$  of VAE latent parameter obtained by inputting each spatial bin of merged data. The black solid, blue dashed, and magenta dotted lines represent spatial bins in the entire dataset, the source region, and the background region, respectively.

dominated spectra (e.g.; Li et al., 2018), and thus the featureless spectra seems to be expressed as a tail in the latent parameter distribution. This result is same as the case of *Tycho*'s SNR and *Kepler*'s SNR (see Chapter 5 and Chapter 6).

Furthermore, a filament-like structure inside the southern part of the SNR was extracted in the latent axis 6 as positive values and also in the axis 1 as negative values. Li et al. (2015) showed that the structure, which they named the ‘dark belt’, has high Ne abundance and lower  $n_e t$  than the surrounding regions.

### 7.3.3 Reconstruction of Spectra

We showed some spectra from samples of the merged data reconstructed by the VAE in Figure 7.3.4. The black squares and solid lines are the original input spectra, and the red closed circle and dashed lines are the reconstructed spectra. The box regions marked with ‘A’ to ‘F’ are shown on each image in Figure 7.3.5. The region ‘A’ and ‘F’ correspond to the synchrotron-dominated emission from the forward shock (FS) in the SW and NE of the SNR, respectively. The region from ‘B’ to ‘E’ have spectra with emission lines from thermal plasma.

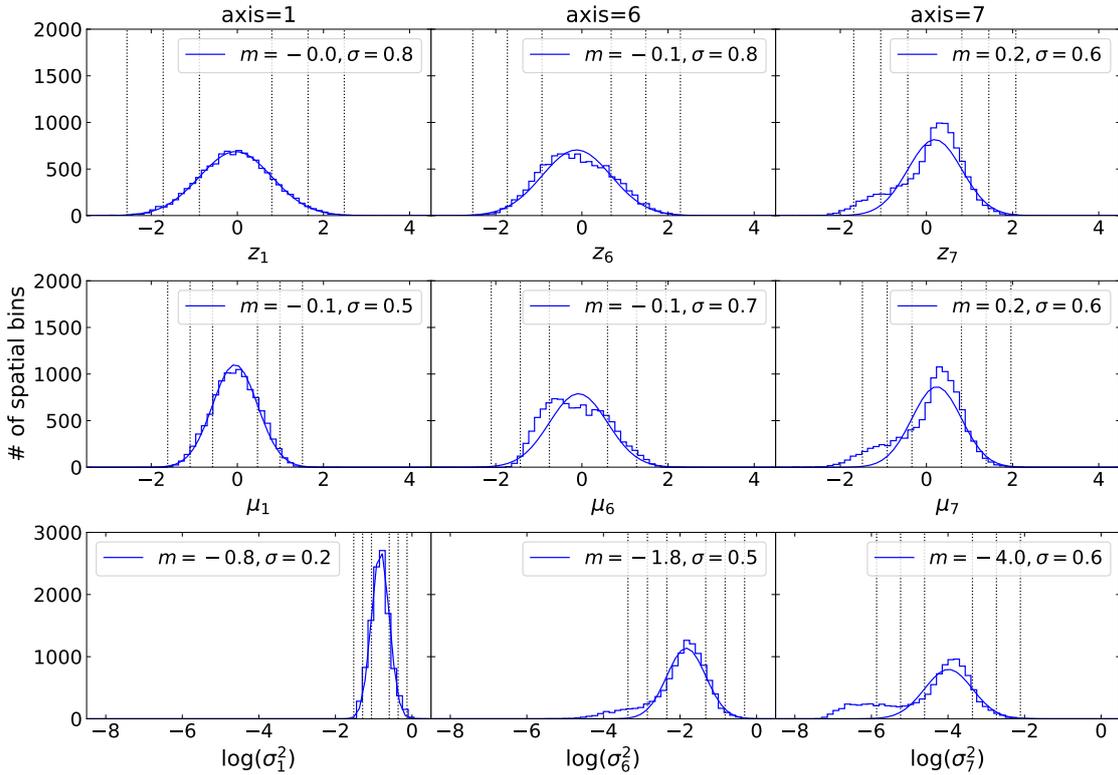


Figure 7.3.3: Histograms showing the meaningful axes of  $z_i$ ,  $\mu_i$ , and  $\log(\sigma_i^2)$ , where  $i = 1, 6, 7$ , of the latent parameters obtained by inputting each spatial bin of merged data in the source region. The curve in each panel is Gaussian fitted on each histogram, the best fitting parameters of Gaussian mean  $m$  and standard deviation  $\sigma$  are shown in each legend. The vertical lines correspond to the  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  positions of each Gaussian.

### 7.3.4 Spectra Generated by the Decoder

We studied which features are represented by each axis of  $z$  for the decoder to generate spectra.

Figure 7.3.6 shows the spectra generated with  $z_1$  and  $z_6$  changing between  $\pm 2.4$  corresponding to  $\pm 3\sigma$  for each column and row, respectively, while  $z_7$  changes between  $-2.8$  and  $2.0$  corresponding to  $-5\sigma$  and  $+3\sigma$  because the distribution of this axis has a long tail beyond  $-3\sigma$  as described above (Figure 7.3.3). The other axes of  $z$ , which seem to be meaningless, are fixed to 0.

The continuum dominated spectra are generated for  $z_6 \leq 0$  and  $z_7 < 0$ . When  $z_6$  becomes larger, the spectrum becomes harder. Thus, the negative pair of  $z_6$  and  $z_7$  represents a featureless spectrum, and simultaneously  $z_6$  determines the spectral hardness. For  $z_7 > 0$  the intermediate-mass element (IME) line features, i.e. Si and S, appear in the spectra. Therefore,  $z_7$  determines the ratio of continuum and line components.

When  $z_1$  decreases, the values of O Ly  $\alpha$  and Fe L blend band (0.63–0.76 keV), and Mg K  $\alpha$  blend band (1.26–1.45 keV) decrease, while the value of 1.45–1.65 keV bin corresponding to the continuum

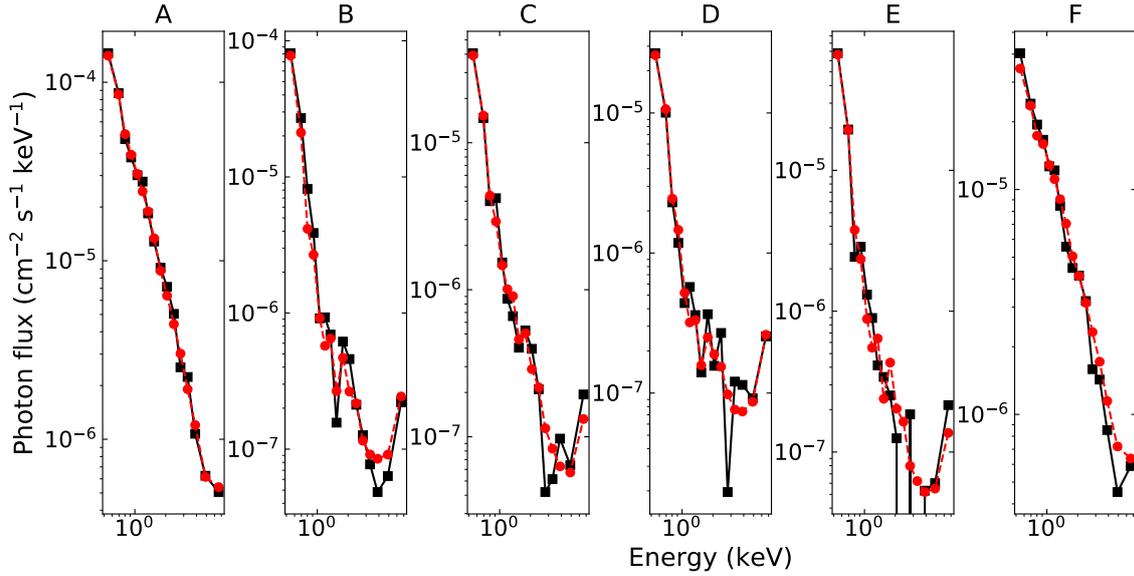


Figure 7.3.4: Sample plot comparing original and reconstructed spectra by the trained VAE as black square and red circle markers, respectively. The spectra marked with ‘A’ to ‘F’ correspond to spatial bins in the source region shown as white box regions in Figure 7.3.5.

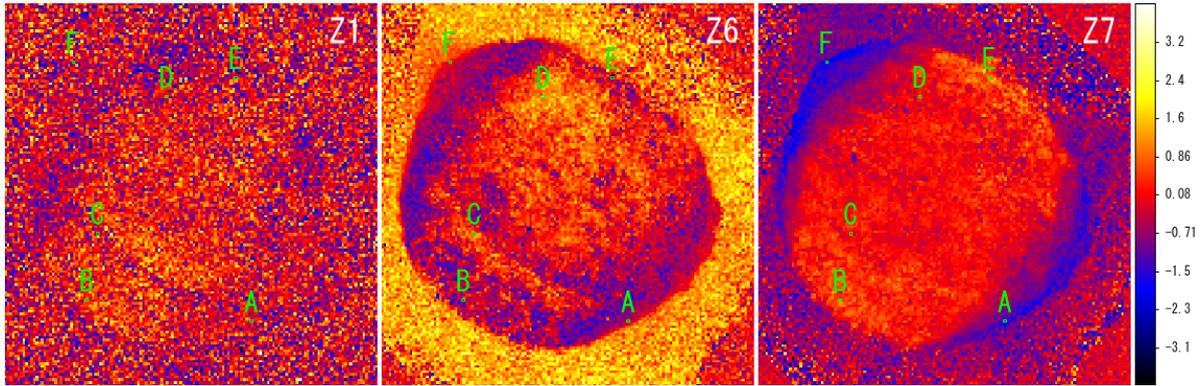


Figure 7.3.5: Images showing the  $z$  of VAE latent parameters for axes 1, 6, and 7 when the merged data observed by the FI chips are input. The images of the  $z$  share a colour scale.

band between Mg K  $\alpha$  and Si K  $\alpha$ , which may contains weak Mg or Al lines, increases. Thus,  $z_1$  seems to determine the line emissions from O, Ne, and Mg and continuum in the low energy side.

### 7.3.5 Averaged Spectra for Each Latent Axis

In order to check the relation between each latent axis and the original observational data, we show the averaged spectra over the spatial bins, which have more than  $+1\sigma$ ,  $+2\sigma$ , or less than  $-1\sigma$ ,  $-2\sigma$ , for each axis 1, 6, and 7 of the latent variable  $z$  in Figure 7.3.7. In addition, for  $z_7$ , the averaged spectrum over the spatial bins corresponding to  $z_7 < -3\sigma$  is also shown. Figure 7.3.8 shows the ratios of the spectra in Figure 7.3.7 to the spectrum averaged over the source region normalized by the values of energy bin dominated by the continuum component (4.2–6.0 keV).

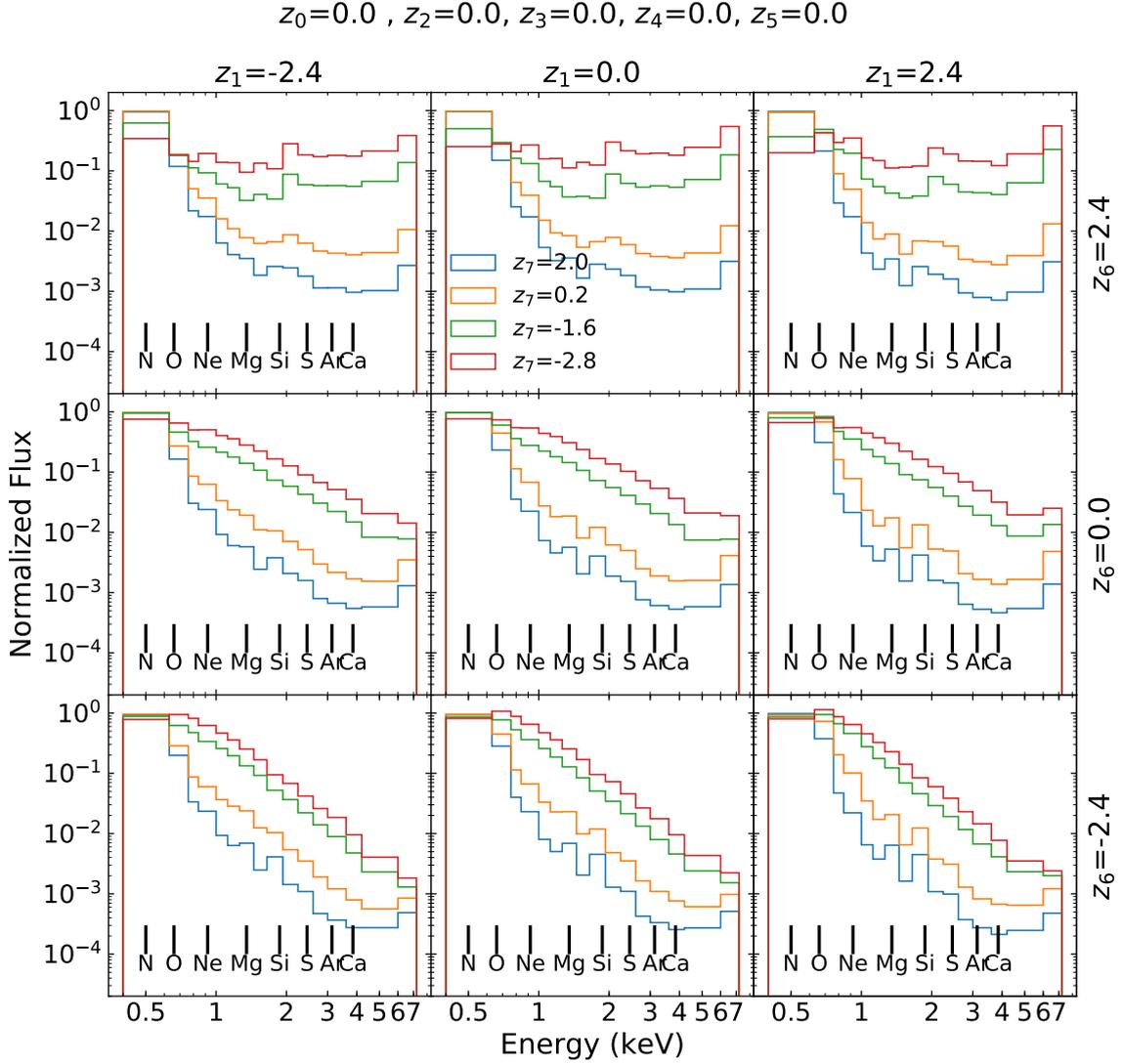


Figure 7.3.6: Generated spectra by the decoder after 100 epochs training. The variation of  $z_7$  is shown in each panel, when meaningless axes are fixed 0.

For  $z_1 < -1\sigma$ , the spectral values from O Ly  $\alpha$  to Ne He  $\alpha$  bands, 0.63–1.0 keV, are lower than the mean of source region, meanwhile those of  $z_1 \geq -1\sigma$  are higher than the mean of source region. In addition, the continuum dominated band between 1.45–1.65 keV has lower value for  $z_1 < -1\sigma$  than the mean of source region. These support the interpretation of  $z_1$  in the previous section.

Each spectrum of  $z_6 \geq 1\sigma$  has a clear Si K  $\alpha$  line feature with a weak continuum component. Each spectrum of  $z_7 \geq 1\sigma$  has a strong Si K  $\alpha$  line feature with weaker continuum. On the other hand, each spectrum of  $z_6 < -1\sigma$  is stronger and softer featureless one than those of  $z_6 \geq 1\sigma$ . Each spectrum of  $z_7 < -1\sigma$  is stronger but harder featureless one than those of  $z_7 \geq 1\sigma$ .

These support the interpretation that the negative sides of  $z_6$  and  $z_7$  determine the continuum component, while their positive sides determine the behaviour of IME line bands as we described in the previous section.

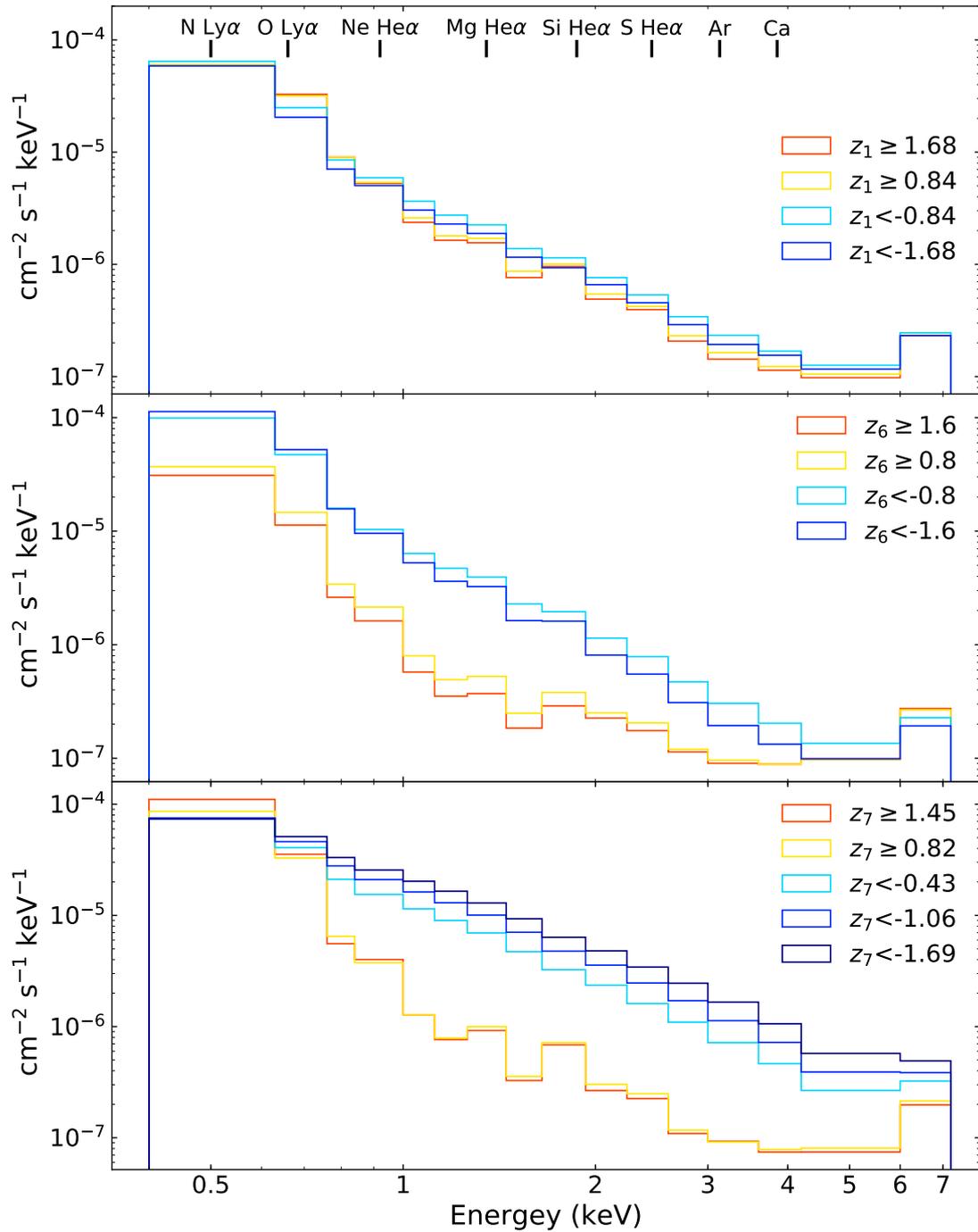


Figure 7.3.7: Spectrum averaged over the spatial bins, which have more than  $+1\sigma$ ,  $+2\sigma$ , or less than  $-1\sigma$ ,  $-2\sigma$ , ( $-3\sigma$  only for  $z_7$ ), respectively, for each latent axes 1, 6, and 7 of  $z$ .

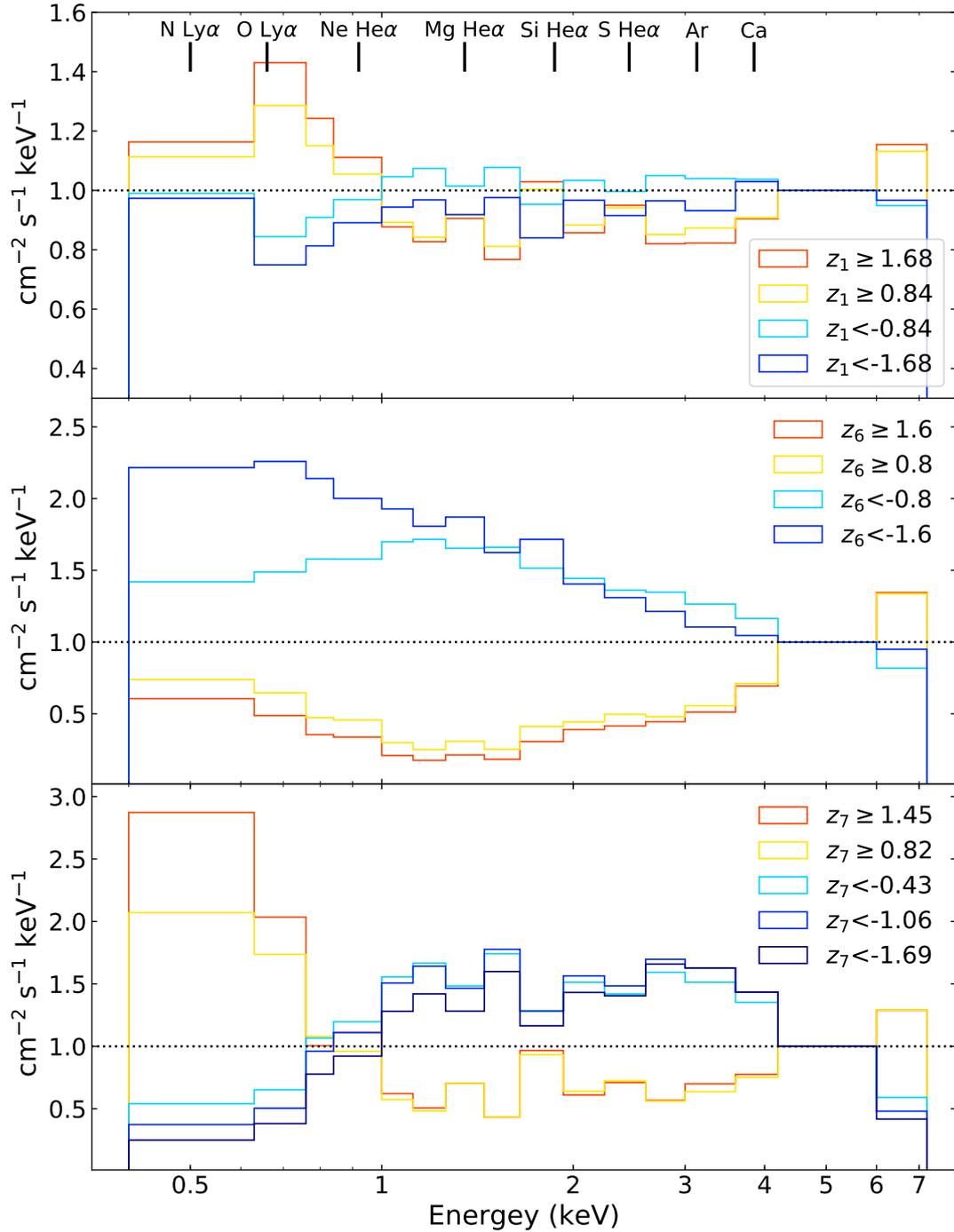


Figure 7.3.8: Same spectra in Figure 7.3.7 but the value of each energy bin is normalized by the average of the source region. The horizontal dotted lines correspond to the ratio of 1.0.

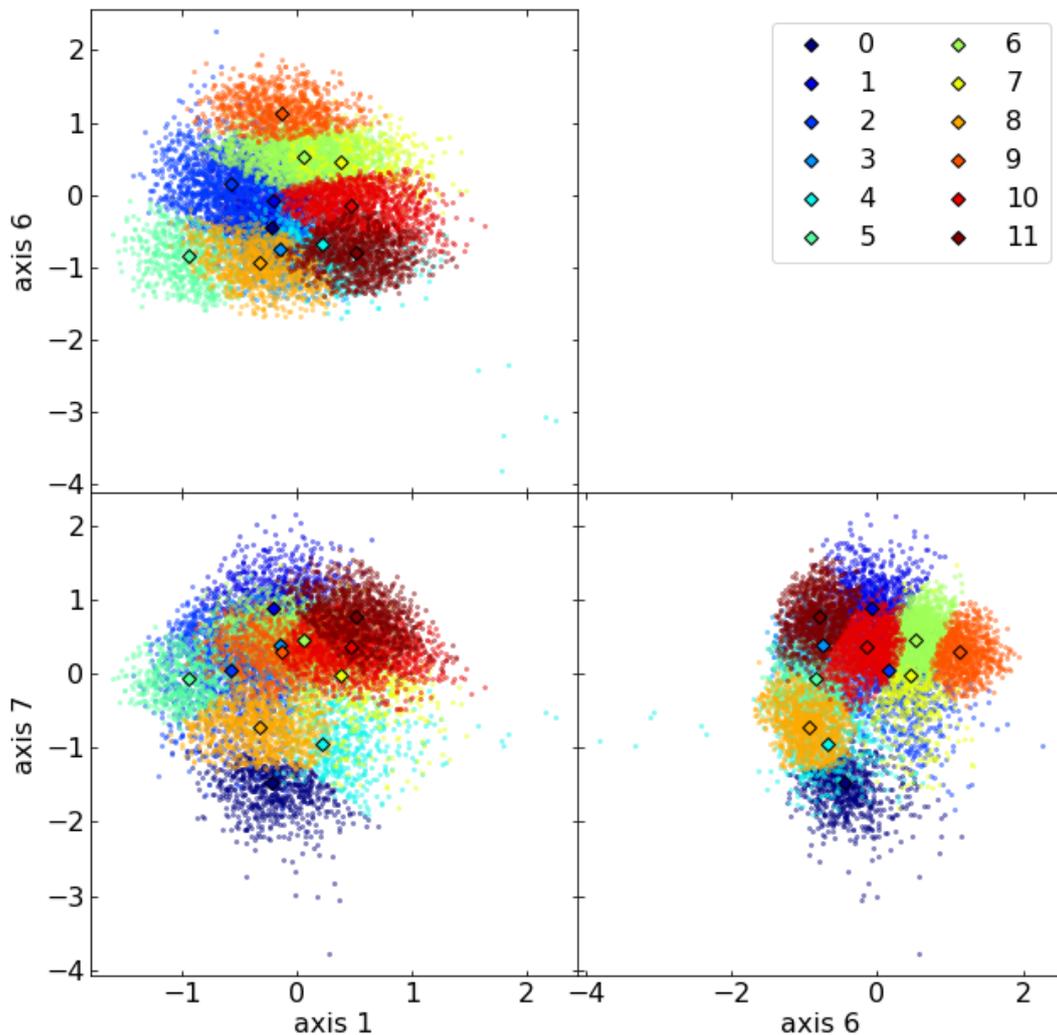


Figure 7.3.9: Scatter plot showing the  $\mu$  of VAE latent parameter obtained by inputting each spatial bin of the merged data, where two of the three axes are chosen. Each point was colour-coded for each category classified by GMM clustering. The centroid of each category is shown as an open black diamond.

### 7.3.6 Results of GMM

We applied Gaussian mixture model (GMM) clustering to the latent expressions obtained by inputting the source bins of the merged data. We only used the three meaningful latent axes (1, 6, and 7) for clustering. The optimal number of clusters was determined to be 12 by checking the Bayesian information criterion (BIC) by changing the number of clusters from 2 to 30.

The scatter plots in Figure 7.3.9 show the distributions of the latent variables  $\mu$ , which are colour-coded according to the category assigned by GMM clustering.

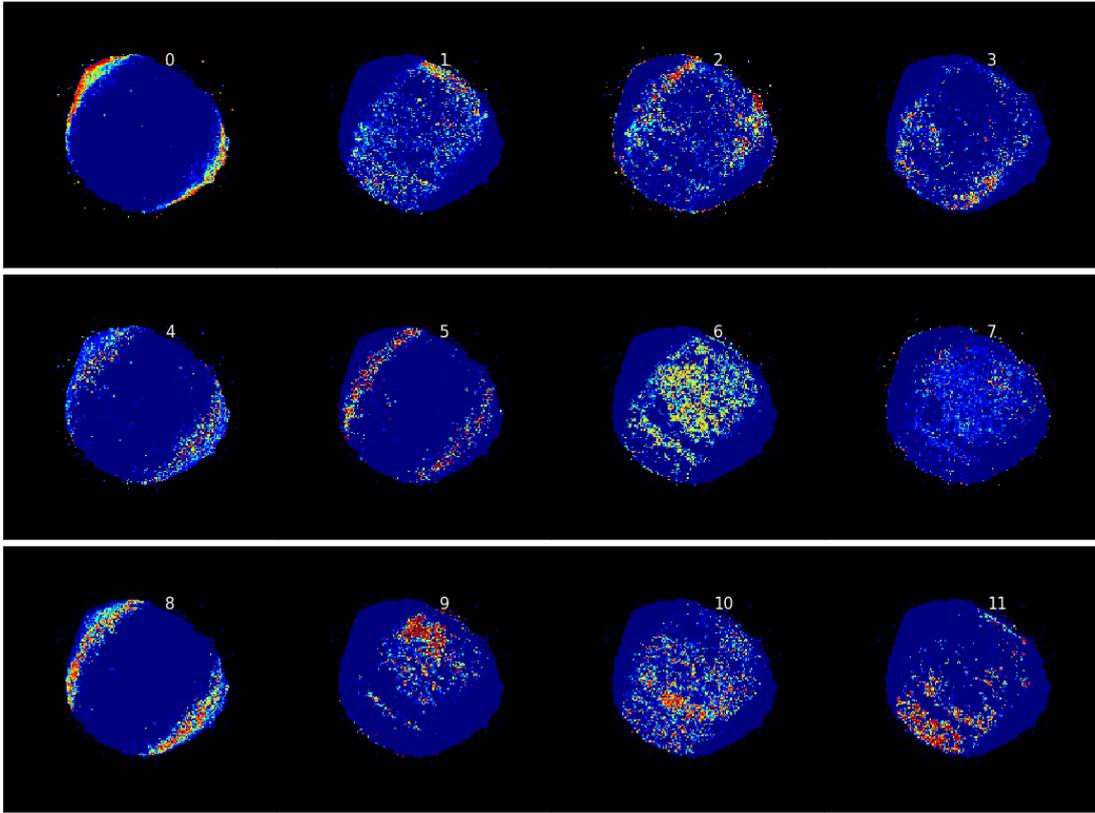


Figure 7.3.10: GMM responsibility for each spatial bin of merged dataset for each category. The responsibility is between 0 and 1; blue and red represent a responsibilities of 0 and 1, respectively.

Each panel of Figure 7.3.10 shows the responsibility of each GMM category. We show the division of SN 1006 into GMM categories in the left-hand panel of Figure 7.3.11. The middle panel of Figure 7.3.11 shows the same image as the left-hand of it, but the spatial bins whose assigned category has a responsibility of  $\leq 80$  per cent are masked. Hence, the spatial bins that remain coloured in the middle panel of Figure 7.3.11 are robustly assigned to the category. It also means that those bins are expected to have some spectral features distinct from those of the other categories. For comparison, traditional three-coloured images are shown in the right-hand panel of Figure 7.3.11.

In Figure 7.3.12, we show the VAE latent values corresponding to the centroid vector of each GMM category. We can discuss the spectral features of each category using the latent values, where we have interpreted the features represented by the latent axes in Section 7.3.4.

We also show spectra generated by inputting each GMM cluster centroid to  $z$  of the decoder in Figure 7.3.13. Figure 7.3.14 shows the ratios of the spectra in Figure 7.3.13 to the spectrum generated by inputting a zero vector to the decoder, where each spectrum is normalized by the values of energy bin dominated by continuum component (4.2–6.0 keV).

We interpreted the feature of each category and summarized in Table 7.3.1. As shown in Figure 7.3.10, the SNR is divided to two parts; the rim of NE and SW, and the central part of the SNR. The categories 0,

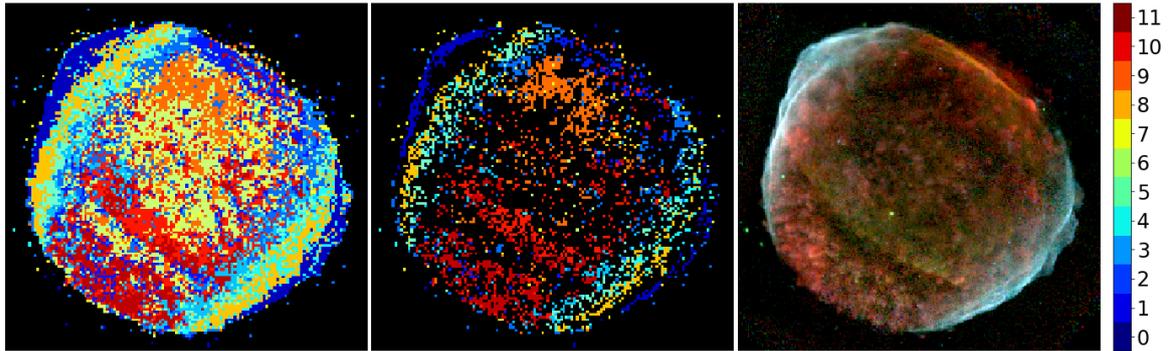


Figure 7.3.11: The left-hand panel shows the GMM category with the highest responsibility for each spatial bin of the merged data. The middle panel shows only spatial bins with a category of more than 80 per cent responsibility, which are assigned the colours representing the category. The other spatial bins, which have responsibilities less than 80 per cent, appear black. The right-hand panel shows a three-colour image of N and O band, 0.4–0.6 keV (red), Ne band, 0.8–1.0 keV (green), and Mg band, 1.27–1.45 keV (blue).

Table 7.3.1: GMM categories.

Category No.	Location	Feature
0	NE and SW edges of SNR	synchrotron dominant
1	Northern rim	O, Ne, Mg, and IME lines
2	Northern part inside the FS	hard continuum and weaker lines
3	Southern part inside the FS	soft continuum and strong lines
4	inside FS, Point sources	synchrotron dominant
5	inside FS	synchrotron dominant
6	dark belt, central part	hard continuum and line emissions
7	blobs	hard continuum and line emissions
8	inside FS	soft synchrotron dominant
9	NE	hard continuum and weak lines
10	Southern inner rim	strong O, Ne, Mg, Si, and S lines
11	Southern rim	strong O, Ne, Mg, Si, and S lines

4, 5, and 8 correspond to the NE and SW rims, while the categories 1, 6, 7, 9, 10, and 11 correspond to the central part. The categories 2 and 3 are the middle of the two classes.

In Figure 7.3.12, the cluster centroids of categories 0, 4, and 8 have larger negative values of  $z_7$  than that of  $-1\sigma$ , which represents continuum dominated spectra as described in Section 7.3.4, and thus the generated spectra of the categories are featureless ones in Figure 7.3.13. The centroid of category 5 has a small negative value of  $z_7$ , and the minimum value of  $z_1$  ( $< -1\sigma$ ), which represents the strong continuum component in the low energy band (Section 7.3.4), and thus the category is also expected to have a soft continuum dominated spectrum. The centroid of category 0 has the minimum value of  $z_7$  (approximately  $-3\sigma$ ) in Figure 7.3.12, and the category spatially corresponds to the outer edges of the

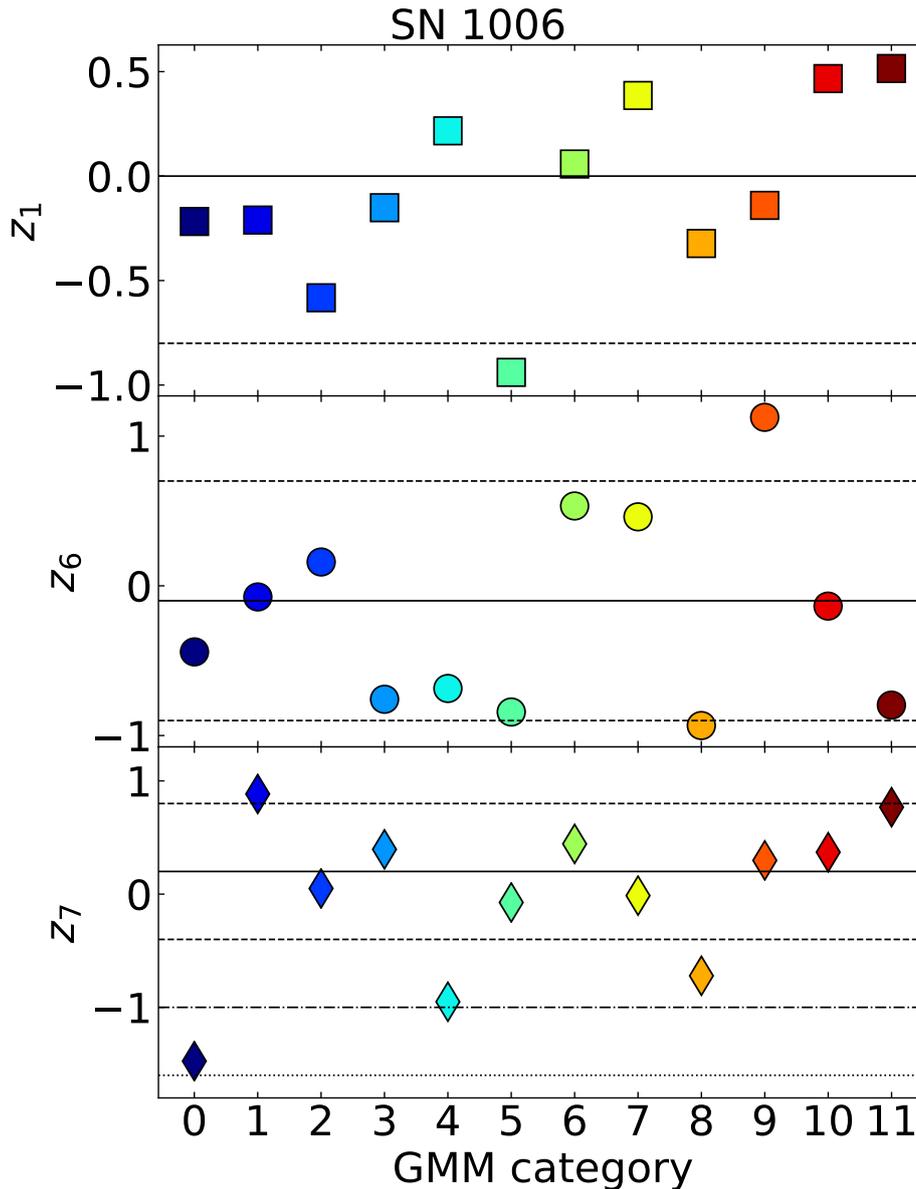


Figure 7.3.12: VAE latent values corresponding to the GMM centroids. The colour in each marker correspond to the assigned GMM category. The horizontal solid, dashed, dash-dotted, and dotted lines correspond the mean,  $\pm 1\sigma$ ,  $\pm 2\sigma$ , and  $\pm 3\sigma$  positions shown in Figure 7.3.3

NE and SW rims, especially in the polar cap geometries in Figure 7.3.10. It suggests that the category 0 is purely synchrotron dominated region at the FS. The centroid of category 8 has the minimum value of  $z_6$  ( $< -1\sigma$ ) in Figure 7.3.12, where the negative side of the latent axis represents a soft spectrum (Section 7.3.4). Therefore, the category 8 is expected to have the softest featureless spectrum. The categories 4 and 5 whose regions are located inside the category 8 in the NE and SW rims, also have large negative values of  $z_6$ , and are expected to have soft spectra. By contrast, the centroid of category 0 has smaller negative value of  $z_6$  than ones of the categories 4, 5, and 8, and thus the result suggests that the spectrum of category 0 is harder than ones of the inner regions.

The categories 2 and 3 are located in northern half and southern half, respectively, of the inner side of the NE and SW rims (Figure 7.3.10). In Figure 7.3.12, the centroid of category 2 has a value more than the mean of  $z_7$ , while that of category 3 is positive but less than the mean of  $z_7$ . In addition, the centroid of category 2 has a positive value of  $z_6$ , while category 3 has a negative one. As a result, in Figure 7.3.13 and Figure 7.3.14, the generated spectrum of category 2 is a softer one containing stronger emission lines than that of category 3. It suggests that the southern half of the rims has softer spectrum containing stronger line emissions than the northern half.

The centroids of categories 1, 3, 6, 9, 10, and 11 have values of more than the mean of  $z_7$  in Figure 7.3.12, and thus the spectra are expected to have clear line emissions. The centroid of category 1, which is mainly located in the outer edge of the northern part of the SNR (Figure 7.3.10), has the maximum value of  $z_7$  ( $> 1\sigma$ ). The centroid of category 1 has a negative value of  $z_1$ , and approximately the mean value of  $z_6$  (Figure 7.3.12). As a result, in Figure 7.3.13, the generated spectrum of category 1 is slightly hard but has also a strong continuum component in the low energy band.

The category 6 is located in the central part of the SNR and the dark belt with lower responsibility than some other categories (Figure 7.3.10). The category 7 is also located in the central part of the SNR, but most of the spatial bins have lower responsibility of the category than ones of some other categories. Both centroids of categories 6 and 7 have similar positive values of  $z_6$  in Figure 7.3.12, and thus their spectra are expected to be hard as shown in Figure 7.3.14. The centroid of category 7 has larger value of  $z_1$  than most of other category, and a small negative value of  $z_7$ . Therefore, the category 7 have weak line emission from especially O, Ne, Mg as shown in Figure 7.3.14.

The category 9 is mainly located in the northern part of the SNR, which is darker than surrounding regions, and one is also located in the part of dark belt (Figure 7.3.11). The centroid of category 9 has the maximum value of  $z_6$  ( $> 1\sigma$ ), a positive value of  $z_7$ , and negative value of  $z_1$  (Figure 7.3.12). It expects that the category 9 has a hard spectrum containing line emissions, and the lines emitted from O, Ne, Mg may be weak as shown in Figure 7.3.14.

The category 10 is located in the southern inner part of the SNR, especially inside of the dark belt, while the category 11 is located the southern part of the SNR, excluding the dark belt (Figure 7.3.10). The centroids of categories 10 and 11 have positive values of  $z_7$ , negative values of  $z_6$ , and similar positive values of  $z_1$  (Figure 7.3.12).  $z_1$  of the category 11 centroid is the maximum value among ones of the all categories. It suggests that the categories have soft spectra containing line emissions from IMEs, and also O, Ne, and Mg.  $z_6$  of the category 11 centroid is lower than that of category 10, which is approximately the mean of  $z_6$ , and thus the category 11 is expected to have a softer spectrum than that of the category 10. The generated spectra in Figure 7.3.14 also show the features described above.

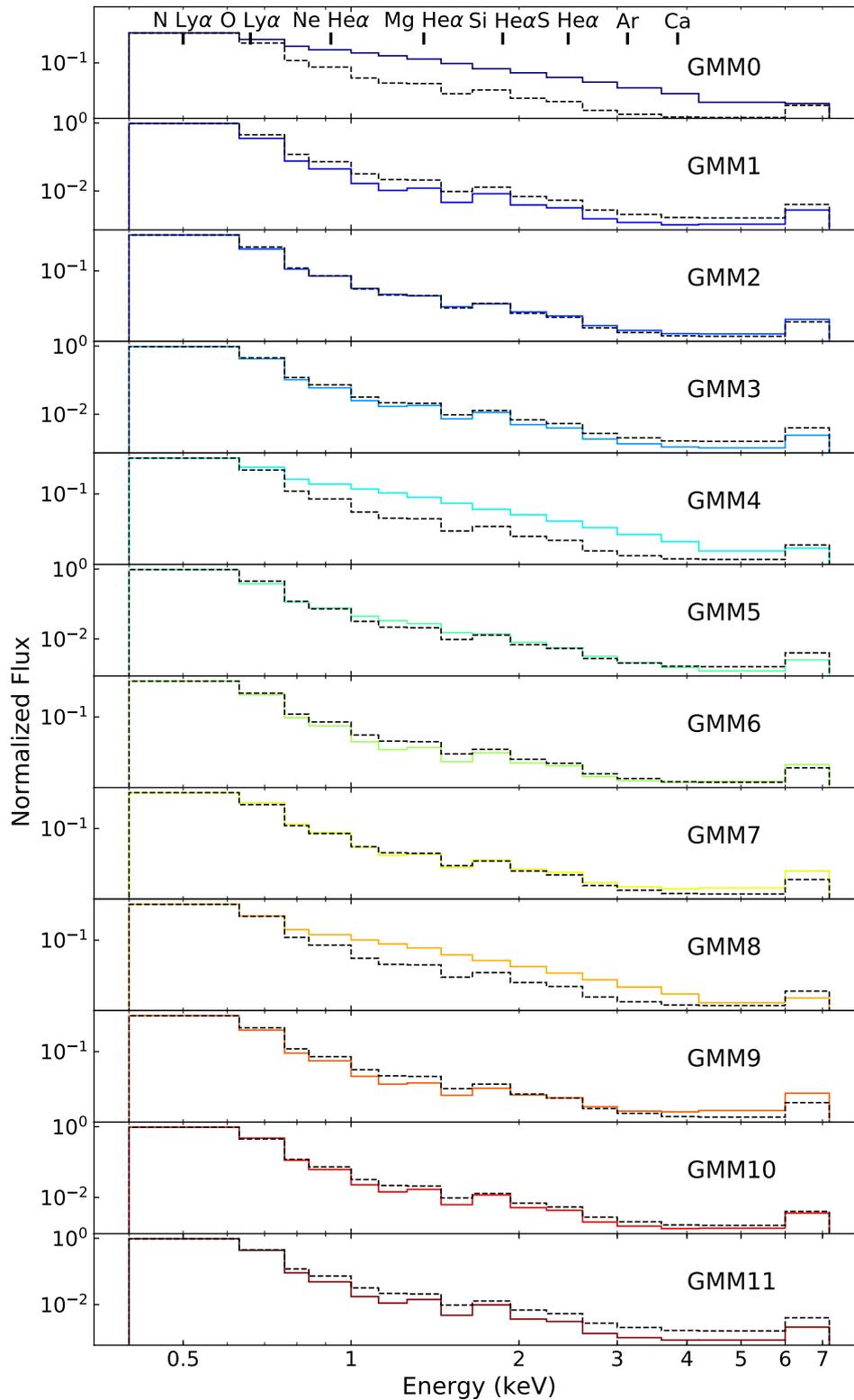


Figure 7.3.13: Spectra generated by the decoder from given  $z$  values, which are the latent vector  $\bar{z}$  of the category centroid for each GMM category. The dashed black line is spectrum generated by inputting a latent vector averaged over the source region to the decoder. The vertical axes have arbitrary unit because the bins were divided the maximum of each spectrum for the VAE training.

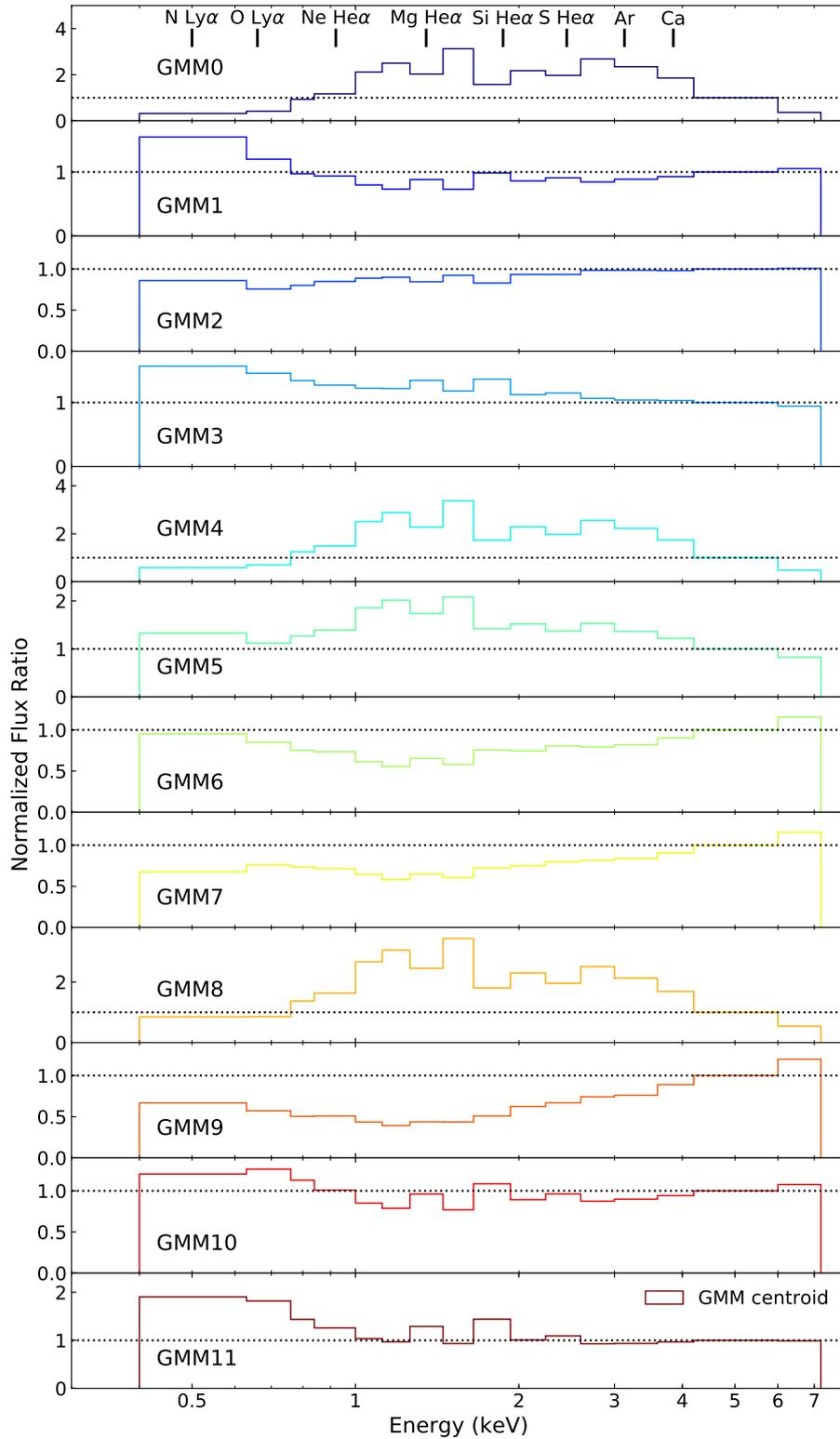


Figure 7.3.14: Same spectra in Figure 7.3.13 but the value of each energy bin is normalized by the spectrum generated by inputting a latent vector averaged over the source region to the decoder. The horizontal dotted lines correspond to the ratio of 1.0.

## 7.4 Application of t-SNE

### 7.4.1 Dimensionality Reduction by t-SNE

If the data can be reduced to a few dimensions (e.g., 2 dimensions), and also the data sample size and dimension are not so large, there are alternative methods for the dimensionality reduction, e.g. t-SNE (van der Maaten & Hinton, 2008). The t-SNE algorithm can transform data into a small dimensional (usually 2-dimensional) space, although the trained model cannot be reused for new data points or other dataset because it is a nonparametric algorithm.

We examined t-SNE to reduce the dimensions of the SN 1006 dataset for comparison with VAE. For the dataset having 10908 samples, i.e., the number of spatial bins in the source region, and 16 feature dimensions, i.e., the number of energy bins, t-SNE spend only 5.6 min to transform the data into 2-dimensional space on a desktop computer with a quad-core Intel Core i5 CPU. We used TSNE in scikit-learn 0.19.0 (Pedregosa et al., 2011).

We embedded the merged data, i.e., a  $10908 \times 16$  matrix, into a 2-dimensional map as shown in Figure 7.4.1. The values of each axis are shown in each panel of Figure 7.4.2.

The axes 0 and 1 are similar to the VAE latent axes 6 and 7 (see also Figure 7.3.1), although the t-SNE axes are opposite in sign to the VAE ones, where the sign is not an intrinsic feature. For the axis 0, the inner part of the NE and SE rims have positive values, while the northern part of the SNR have negative values. Furthermore, the filament-like structure inside the southern part of the SNR have negative values, while the bins of the inner side have large positive values. For the axis 1, the NE and SE rims of the SNR have large positive values, while the SE and NW rims have large negative values.

### 7.4.2 Hierarchical Clustering

We applied the hierarchical clustering using Ward’s linkage criterion with Euclidean metric to the data points in the 2-dimensional space embedded by t-SNE. This clustering method makes a hierarchical tree of clusters where the pairwise distances of data points are measured. We used ‘ward’ in SciPy 0.19.1 for the linkage.

We show some clustering results cutting the tree at the level for cluster distances changing from 4000 to 500 in Figure 7.4.3. For this clustering, we need not determine the number of clusters but the cutting level.

At first, the data is divided into two large clusters, corresponding to the NE and SW rims (ID 0) and inner part of the SNR (ID 1), respectively, for the distance level of 4000. In the 2-dimensional map embedded by t-SNE, the rim cluster (ID 0) coincides the ring like distribution in the positive side of axis 1, while the inner-part cluster (ID 1) coincides the distribution in the negative side of axis 0.

For the distance level of 2500, the inner-part cluster (ID 1) was divided into two more clusters; (new ID 1) the southern part of the SNR, and (new ID 2) the northern part and the filament like structure inside

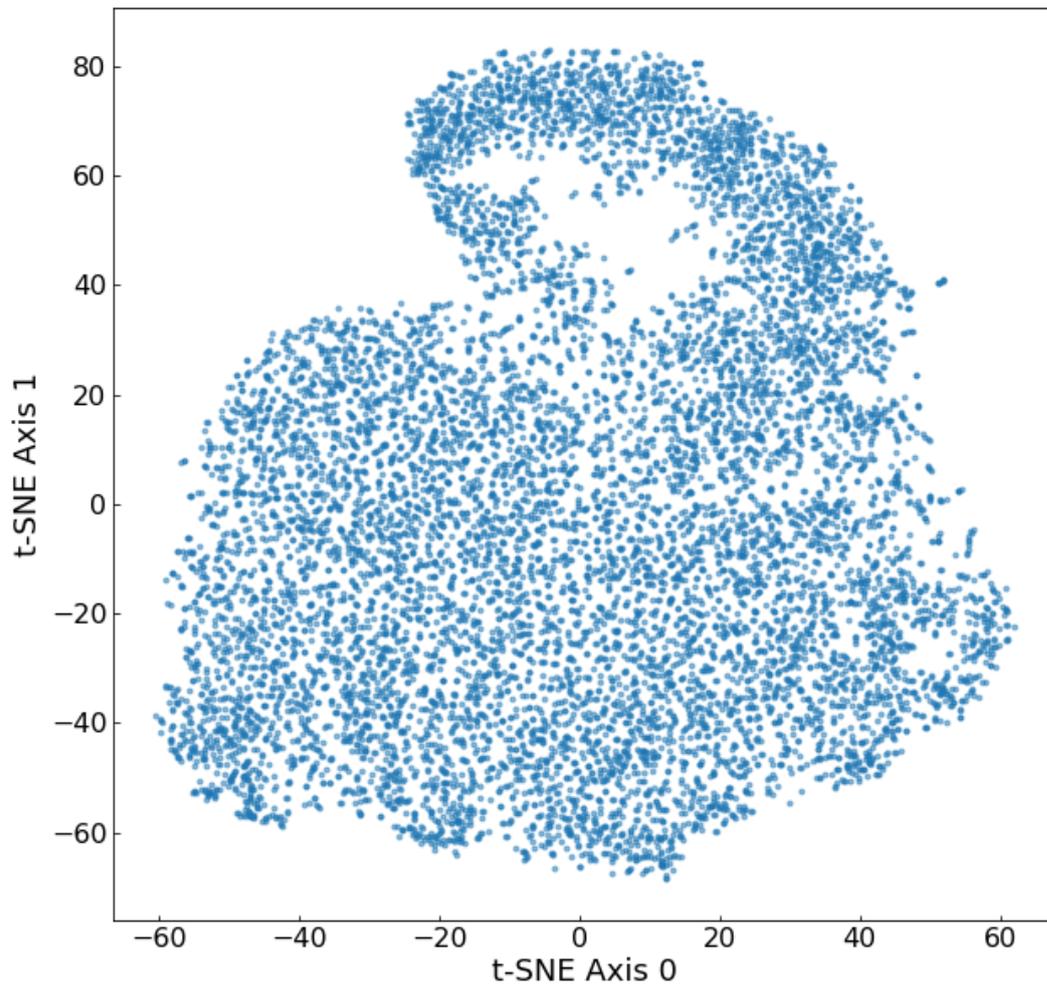


Figure 7.4.1: Scatter plot showing the 2-dimensional map embedded from the merged data observed by ACIS-FI chips (a 16-colour image). Each data sample was normalized a maximum value of each sample when it was input to the t-SNE.

of the southern SNR.

For the threshold of 1600, the previous cluster 0 was divided into 2 clusters; (new ID 0) outer side of the rim, and (new ID 1) inner side of the rim. Meanwhile, the previous cluster 2 was also divided into 2 clusters; (new ID 3) the northern rim and the part of filament, and (new ID 4) the inner side of the SNR.

For the threshold of 1000, the previous cluster 2 divided into 2 clusters, however we could not find the difference between them from only the spatial distribution. Lastly, for the threshold of 500, the data points are divided into 18 clusters.

From the hierarchical relations of clusters, we can understand that the SE and NW rims are sufficiently different from the inner side of the SNR. In the SE and NW rims, layered structures, which may actually

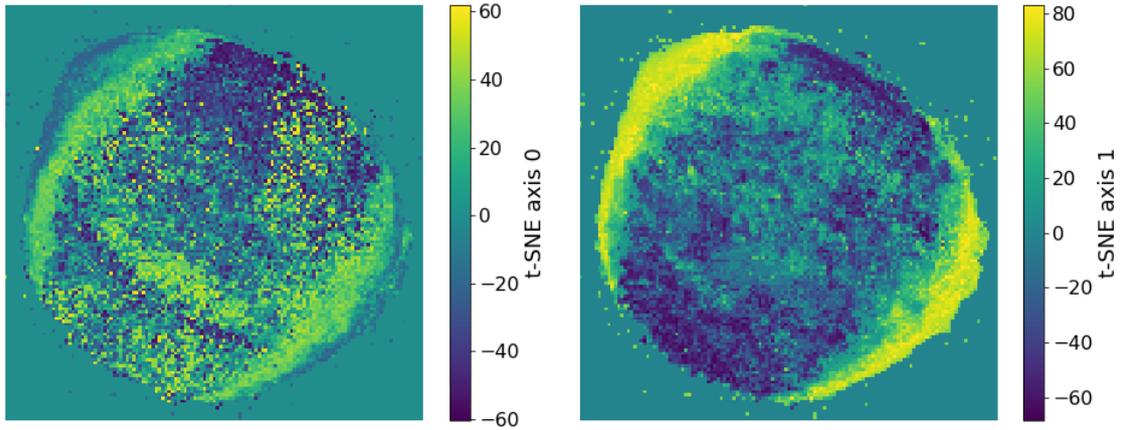


Figure 7.4.2: Images showing the values for each axis of the 2-dimensional space embedded from the merged data observed by ACIS-FI chips (a 16-colour image).

be continuous changes of spectra have appeared. Furthermore, the outer edge of SW rim extends to the NW, and the NW part is assigned to an independent category for the threshold of 500. On the other hand, the inner side of the SNR is roughly divided into four part; the southern rim, the NE part, the ‘dark belt’, and outer side of the northern rim.

## 7.5 Discussions

### 7.5.1 Dimensionality Reduction by VAE

We interpreted the features represented by each latent axis using the spectra generated by the decoder.

We found that the pair of axes  $z_6$  and  $z_7$  represented the continuum component and the IME lines. The axis  $z_7$  determines the ratio of the continuum component and the line emissions. Meanwhile, the axis  $z_6$  determines the spectral hardness. Furthermore, the axis  $z_1$  represents the soft band features, i.e., the ratio of the continuum component and the emission lines of O, Ne, and Mg.

Both the synchrotron emission and the lines in the low-energy band are important topics, on which several previous works have been focused in SN 1006 (e.g.; Koyama et al., 1995; Uchida et al., 2013). We conclude that the VAE successfully extracted such important physical features from the spectral data.

The ‘dark belt’, which is darker than the surrounding regions, also appeared in the latent variables, although the only spectral shapes were used for the training. It shows that the ‘dark belt’ is not only dark but also has different spectral shape from the surrounding regions. It is consistent with the previous result (Li et al., 2015).

Unveiling the reasoning process of the DNN is a significant problem. However, we can discuss the meaning of the latent variables using the decoder. We conclude that this is a merit of a generative model, and thus VAEs can be recommended to use such tasks.

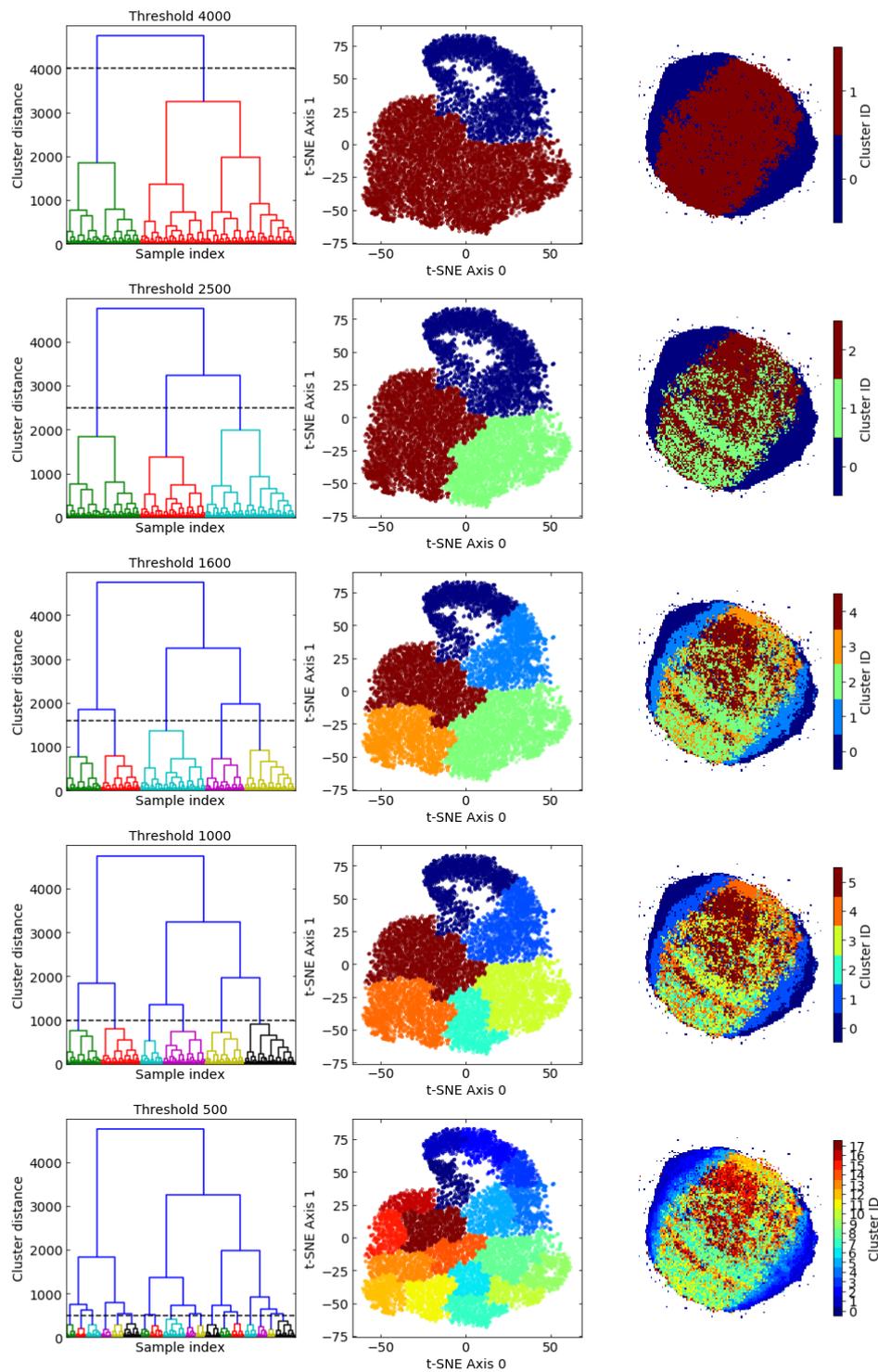


Figure 7.4.3: Each row showing hierarchical clustering result for each threshold of clustering distances. In the left-hand column, the Ward clustering dendrograms show the relations of every spatial bins and cluster distances. The middle column shows 2-dimensional maps embedded from the merged data by t-SNE. Each point was colour-coded for each category classified by Ward clustering. In the right-hand column, images show the Ward clustering category for each spatial bin of the merged data.

In the distribution of latent space parameters, the samples that have featureless spectra made a tail distribution in the latent axis 7. This result is same as the cases of *Tycho's* SNR and *Kepler's* SNR.

The VAE models, which we trained, have only 2 or 3 meaningful latent axes. This is less than the *Kepler's* SNR case. In this case, the number of spectral bins is 16, which is less than the case of *Kepler's* SNR (26 bins) or *Tycho's* SNR (37 bins). Thus, the information included in the dataset may be less than the other cases. The finer spectral binning, which maintains more information, may allow the VAE to extract more features.

In the current case, we divided the O, Ne, and Mg lines into independent spectral bins. It allowed the VAE to capture the line features on the latent space.

### 7.5.2 Dimensionality Reduction by t-SNE

Nonparametric algorithms tend to consume a lot of memory. However, in this case, the sample size is not large, and thus t-SNE has been trained within a reasonable time.

This method has successfully embedded the dataset into 2 dimensional space. We found that the result of t-SNE was similar to the results of VAEs, and thus the t-SNE performed the dimensionality reduction as well as the VAE in this case. Therefore, t-SNE can be an alternative dimensionality reduction method as a VAE in such a case.

However, t-SNE is not capable of generating the data samples like the decoder of VAE. Therefore, t-SNE cannot directly show the meaning of the embedded space axes. It is a weak point of t-SNE.

### 7.5.3 Hierarchical Clustering

The hierarchical clustering can find the relations of clusters appearing in the dendrogram. Further, a short pairwise distance of clusters makes the cluster IDs close to each other, while a longer pairwise distance makes the difference of IDs wider for the hierarchical clustering. Thus the images showing the division into categories are visible and easy to understand the relations of clusters, i.e., the similarly coloured bins have similar spectra.

On the other hand, the clustering cannot show the responsibilities of categories for each data points because of hard clustering. Therefore, this method is unavailable to determine the region to extract a spectrum for spectral analyses so that the region includes only the spatial bins robustly assigned to the category. It is a weak point of hard clustering.



## Chapter 8

# Conclusions

We newly implemented an unsupervised machine learning method combining the variational autoencoder (VAE) and Gaussian mixture model (GMM), where the dimensions of the observed data are reduced by the VAE, and clustering in feature space is done by the GMM, and applied the method to *Tycho's* SNR, *Kepler's* SNR, and SN 1006, ones of the best-known SNRs. It is the first application of this method to diffuse astronomical objects.

As described in Section 2.3, most previous applications of machine learning techniques to X-ray analysis of SNR data have been limited to linear methods, e.g., principal component analysis (PCA) or GMM. However, the value of each spectral bin depends nonlinearly on the underlying physical parameters. Therefore, DNNs, which are capable of handling nonlinear relations, are appropriate for SNR X-ray spectra. In our method using VAE, the ability to extract nonlinear relations from data space is improved.

For X-ray data, Poisson statistics is appropriate because each bin in an X-ray spectrum represents a number of photons. As Ichinohe & Yamada (2019) shows for an ideal case, Poisson statistics is important for VAE training with such X-ray spectral dataset. We also newly implemented the data processes at the input and output of VAE in order to apply Poisson reconstruction loss for the training as described in Section 4.1.2, and applied the model to observational X-ray spectral data.

As demonstrated with *Tycho's* SNR in Chapter 5, the VAE extracts features using the relative intensities of lines as well as the properties of the continuum spectrum. We found that our method successfully reveals the characteristic spatial structures, e.g., the Fe knot in the south-east of the SNR, the layered structure in the north-western ejecta rim, and the synchrotron dominated filaments.

In Chapter 6, the VAE extracts features using only spectral shape of *Kepler's* SNR. We found that our method revealed the characteristic spatial structures, such as the synchrotron dominated forward shock, the layered structure in the northern rim, and the region interacting dense circumstellar medium (CSM).

We also applied our method to SN 1006 in Chapter 7. We found that the VAE have successfully captured some important physical features; the intensity and hardness of synchrotron emission, the line ratio of Si K  $\alpha$  and S K  $\alpha$ , and the emission lines of O, Ne, Mg. Furthermore, the 'dark belt', which is darker than the surrounding regions, was also represented in the latent variables using only spectral shape information.

We also show the relation between the VAE latent space and the original data space, using the decoder to generate spectra from given latent parameters. As shown in Chapter 6, we found that the VAE has

extracted the latent axes corresponding to the relations of Fe and intermediate-mass element (IME) line blends, continuum emission, the N and O blends, and the Fe L blend peak energy reflecting the electron temperature or plasma ionization state. It shows that VAE successfully captured the features of plasma spectra. We emphasize that VAE is capable of unveiling what the latent axes represent, and help us to understand the dimensionality reduction result, although parameters in a deep learning model are hardly interpreted and thus deep learning is often called a ‘black box’. The interpretability of latent variables is a merit of our method using the generative model, VAE. Other dimensionality reduction methods of manifold learning, e.g., t-SNE and isomap, cannot generate data, thus are not capable to directly interpret the embedded spaces.

We also applied the method combining t-SNE and hierarchical clustering to SN 1006. We found that t-SNE converted the spectral data into lower dimensional space than that of dataset as well as VAEs. However, as described above, t-SNE cannot directly show the meaning of the embedded space axes, because it cannot generate the data samples like a decoder of VAE.

Our unsupervised machine learning method automatically revealed spatial structures which have been discussed in the literature (see, e.g., Yamaguchi et al., 2017). This demonstration shows that the method is a powerful tool for data analyses that makes it possible to automatically exploit the rich information contained in data obtained by X-ray observations of SNRs. It may be possible to discover SNR physics (e.g., plasma evolution, interaction with ambient media, or cosmic-ray acceleration), and supernova explosion mechanism (e.g., nucleosynthesis, asymmetric explosion, or progenitor type), by post-training analyses using the results of machine learning.

We conclude that our unsupervised method is an efficient tool to define regions for spectral extraction, and can be important for making the best use of the currently available data and future observational data obtained by upcoming instruments. Model fitting of spectra generally spends many hours; thus, the difficulty of spectral analysis is expected to increase steeply by upcoming an X-ray microcalorimeter (such as *Athena*; Barret et al., 2018). As demonstrated in Chapter 5, Chapter 6, and Chapter 7, our method spends much less time than spatially resolved spectral model fitting. For example, the VAE training on GPUs ran less than 1 min with the large data set of *Tycho*’s SNR. Thus, our method is expected to reveal characteristic features directly from raw observational data for such high-resolution spectroscopy without spectral model fitting, and help us to understand the astronomical plasma phenomena in more detail.

In order to focus only an information of spectral shape, we applied the spatial-bin-by-bin normalization for the VAE training on *Kepler*’s SNR, and SN 1006. We found that this method is effective for analyses of SNRs, in which spectral shapes are important. The data preprocessing is important, and thus it is better to carefully select information to maintain in data set for the type of source and the focusing physical feature.

It is also worth noting that the method discovered the spatial structures automatically, although no spatial information was used in the model as described in Section 5.6. This means that the method can extract physical feature based only on the spectral information.

Thus, the method is equally applicable to temporally and spatially variable data, because the training uses only spectral information. In addition, our method implemented in this work is not limited to SNRs in X-ray and can be applied to other classes of sources and to other energy bands. Our method

is expected to apply to a lot of astronomical objects observed in several energy bands. Furthermore, in the multimessenger astronomy, our method is capable of applying to datasets combined observations of multiwavelength or/and some other messenger.



# Acknowledgements

First of all, I wish to express my best gratitude to my supervisor Prof. Yasunobu Uchiyama for guiding and encouraging me throughout six years of my academic life at Rikkyo university. I would like to thank Prof. Hiroya Yamaguchi for giving me an idea of prototype of this work and precise advices. I am deeply grateful to Dr. Yuto Ichinohe for kindly encouraging me and giving a lot of suggestive comments on this thesis. I also thank Dr. Dmitry Khangulyan for giving many suggestive comments for my paper. I thank Dr. Masato Taki, and Dr. Mikio Morii for useful information and discussions about machine learning, and also Dr. Toshiki Sato for beneficial discussions. I am grateful to Dr. Shigehiro Nagataki, Dr. Gilles Ferrand and all the member of RIKEN ABBL for kindly welcoming me in their meeting and lunch, and discussing about SNR simulations and analyses. I really appreciate constructive advice on this thesis by Prof. Shunji Kitamoto and Prof. Jiro Murata. I deeply thank Dr. Akio Hoshino for his kind help in Maryland when I visited for a conference. I am grateful to Prof. Shunji Kitamoto, Prof. Makoto Taguchi, and Prof. Shingo Kameda for many helpful comments in the Astro and Solar-Terrestrial Physics Laboratory of Rikkyo university. I want to offer my special thanks to Naomi Tsuji, Masanori Arakawa, and all the previous and current colleagues, who made my research life fruitful. Finally, I really thank my family for their support and encouragement.



# Bibliography

- Abadi M., et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/>
- Akiyama K., et al., 2017, ApJ, 838, 1
- Albert J., et al., 2008, Nuclear Instruments and Methods in Physics Research A, 588, 424
- Armstrong J. A., Fletcher L., 2019, Sol. Phys., 294, 80
- Bai Y., Liu J., Bai Z., Wang S., Fan D., 2019, AJ, 158, 93
- Baldi P., Hornik K., 1989, Neural networks, 2, 53
- Barret D., et al., 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 106991G (arXiv:1807.06092), doi:10.1117/12.2312409
- Becherini Y., Djannati-Ataï A., Marandon V., Punch M., Pita S., 2011, Astroparticle Physics, 34, 858
- Bishop C. M., 2006, Pattern recognition and machine learning. springer
- Blair W. P., Ghavamian P., Long K. S., Williams B. J., Borkowski K. J., Reynolds S. P., Sankrit R., 2007, ApJ, 662, 998
- Breiman L., 2001, Machine learning, 45, 5
- Burges C. J., 1998, Data mining and knowledge discovery, 2, 121
- Burke C. J., Aleo P. D., Chen Y.-C., Liu X., Peterson J. R., Sembroski G. H., Yao-Yu Lin J., 2019, arXiv e-prints, p. arXiv:1908.02748
- Burkey M. T., Reynolds S. P., Borkowski K. J., Blondin J. M., 2013, ApJ, 764, 63
- Cash W., 1979, ApJ, 228, 939
- Cassam-Chenaï G., Decourchelle A., Ballet J., Hwang U., Hughes J. P., Petre R., et al. 2004, A&A, 414, 545
- Cassam-Chenaï G., Hughes J. P., Ballet J., Decourchelle A., 2007, ApJ, 665, 315
- Chandra IPI team ., 2018, The Chandra Proposars' Observatory Guige, Version 21.0
- Chandrasekhar S., 1931, ApJ, 74, 81
- Charnock T., Moss A., 2017, ApJ, 837, L28
- Chollet F., et al., 2015, Keras, <https://github.com/keras-team/keras>
- Davoodi P., et al., 2006, AJ, 132, 1818
- Decourchelle A., et al., 2001, A&A, 365, L218
- Díaz Baso C. J., de la Cruz Rodríguez J., Danilovic S., 2019, arXiv e-prints, p. arXiv:1908.02815
- Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
- Dilokthanakul N., Mediano P. A. M., Garnelo M., Lee M. C. H., Salimbeni H., Arulkumaran K., Shanthan M., 2016, preprint, (arXiv:1611.02648)

- Douvion T., Lagage P. O., Cesarsky C. J., Dwek E., 2001, *A&A*, 373, 281
- Dozat T., 2016, ICLR workshop paper
- Dubner G. M., Giacani E. B., Goss W. M., Green A. J., Nyman L. Å., 2002, *A&A*, 387, 1047
- Dugas C., Bengio Y., Bélisle F., Nadeau C., Garcia R., 2001, in Leen T. K., Dietterich T. G., Tresp V., eds, *Advances in Neural Information Processing Systems 13*. MIT Press, pp 472–478, <https://papers.nips.cc/paper/1920-incorporating-second-order-functional-knowledge-for-better-option-pricing>
- Eger S., Youssef P., Gurevych I., 2018, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, pp 4415–4424, doi:10.18653/v1/D18-1472, <https://www.aclweb.org/anthology/D18-1472>
- Eriksen K. A., et al., 2011, *ApJ*, 728, L28
- Event Horizon Telescope Collaboration et al., 2019a, *The Astrophysical Journal*, 875, L1
- Event Horizon Telescope Collaboration et al., 2019b, *ApJ*, 875, L4
- Feng Q., Lin T. T. Y., VERITAS Collaboration 2017, in Brescia M., Djorgovski S. G., Feigelson E. D., Longo G., Cavuoti S., eds, *IAU Symposium Vol. 325, Astroinformatics*. pp 173–179 (arXiv:1611.09832), doi:10.1017/S1743921316012734
- Ferrand G., Warren D. C., Ono M., Nagataki S., Röpke F. K., Seitzzahl I. R., 2019, *ApJ*, 877, 136
- Fruscione A., et al., 2006, in *Proc. SPIE*. p. 62701V, doi:10.1117/12.671760
- Garmire G. P., Bautz M. W., Ford P. G., Nousek J. A., Ricker Jr. G. R., 2003, in Truemper J. E., Tananbaum H. D., eds, *Proc. SPIE Vol. 4851, X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy*. pp 28–44, doi:10.1117/12.461599
- Goldberger J., Hinton G. E., Roweis S. T., Salakhutdinov R. R., 2005, in *Advances in neural information processing systems*. pp 513–520
- Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, arXiv e-prints, p. arXiv:1406.2661
- Goodfellow I., Bengio Y., Courville A., 2016, *Deep Learning*. MIT Press
- Hausen R., Robertson B., 2019, arXiv e-prints, p. arXiv:1906.11248
- Hayato A., et al., 2010, *ApJ*, 725, 894
- He K., Gkioxari G., Dollár P., Girshick R., 2017, arXiv e-prints, p. arXiv:1703.06870
- Hezaveh Y. D., Levasseur L. P., Marshall P. J., 2017, *Nature*, 548, 555
- Hinton G. E., 1990, in , *Machine learning*. Elsevier, pp 555–610
- Hinton G. E., Salakhutdinov R. R., 2006, *Science*, 313, 504
- Hinton G. E., Krizhevsky A., Wang S. D., 2011, in *International Conference on Artificial Neural Networks*. pp 44–51
- Honma M., Akiyama K., Uemura M., Ikeda S., 2014, *PASJ*, 66, 95
- Hurley P. D., Oliver S., Farrah D., Wang L., Efstathiou A., 2012, *MNRAS*, 424, 2069
- Hwang U., Gotthelf E. V., 1997, *ApJ*, 475, 665
- Hwang U., Hughes J. P., Petre R., 1998, *ApJ*, 497, 833
- Hwang U., Decourchelle A., Holt S. S., Petre R., 2002, *ApJ*, 581, 1101
- Ichinohe Y., Yamada S., 2019, *MNRAS*, 487, 2874

- Ichinohe Y., Yamada S., Miyazaki N., Saito S., 2018, *MNRAS*, 475, 4739
- Ishida E. E. O., et al., 2017, in *Astroinformatics*. Cambridge University Press, pp 247–252, doi:10.1017/S174392131601293X
- Iwasaki H., Ichinohe Y., Uchiyama Y., 2019a, *MNRAS*, 488, 4106
- Iwasaki H., Ichinohe Y., Uchiyama Y., Yamaguchi H., 2019b, in Teuben P. J., Pound M. W., Thomas B. A., Warner E. M., eds, *Astronomical Society of the Pacific Conference Series Vol. 521, Astronomical Data Analysis Software and Systems XXVII* (Nov. 2018). p. 79
- Jimenez Rezende D., Mohamed S., Wierstra D., 2014, arXiv e-prints, p. arXiv:1401.4082
- Kaastra J. S., Paerels F. B. S., Durret F., Schindler S., Richter P., 2008, *Space Sci. Rev.*, 134, 155
- Karmakar A., Mishra D., Tej A., 2018, in *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*. pp 122–126
- Kasuga T., Sato T., Mori K., Yamaguchi H., Bamba A., 2018, *PASJ*, 70, 88
- Katsuda S., Tsunemi H., Uchida H., Kimura M., 2008, *ApJ*, 689, 225
- Katsuda S., Petre R., Hughes J. P., Hwang U., Yamaguchi H., Hayato A., Mori K., Tsunemi H., 2010, *ApJ*, 709, 1387
- Katsuda S., Long K. S., Petre R., Reynolds S. P., Williams B. J., Winkler P. F., 2013, *ApJ*, 763, 85
- Katsuda S., et al., 2015, *ApJ*, 808, 49
- Kerzendorf W. E., Childress M., Scharwächter J., Do T., Schmidt B. P., 2014, *ApJ*, 782, 27
- Khokhlov A. M., 1991, *A&A*, 245, 114
- Khrantsov V., Akhmetov V., 2018, preprint, (arXiv:1805.08160)
- Kingma D. P., Ba J., 2014, arXiv e-prints, p. arXiv:1412.6980
- Kingma D. P., Welling M., 2013, preprint, (arXiv:1312.6114)
- Koyama K., Petre R., Gotthelf E. V., Hwang U., Matsuura M., Ozaki M., Holt S. S., 1995, *Nature*, 378, 255
- Krause O., Tanaka M., Usuda T., Hattori T., Goto M., Birkmann S., Nomoto K., 2008, *Nature*, 456, 617
- Krause M., Pueschel E., Maier G., 2017, *Astroparticle Physics*, 89, 1
- Kronmueller M., Theo G., 2019, in *36th International Cosmic Ray Conference (ICRC2019)*. p. 937 (arXiv:1908.08763)
- LeCun Y., Bengio Y., Hinton G., 2015, *Nature*, 521, 436
- Li J.-T., Decourchelle A., Miceli M., Vink J., Bocchino F., 2015, *MNRAS*, 453, 3953
- Li J.-T., Decourchelle A., Miceli M., Vink J., Bocchino F., 2016, *MNRAS*, 462, 158
- Li J.-T., et al., 2018, *ApJ*, 864, 85
- Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. C., 2015, arXiv e-prints, p. arXiv:1512.02325
- Longair M. S., 2011, *High energy astrophysics*. Cambridge university press
- Lu F. J., Ge M. Y., Zheng S. J., Zhang S. N., Long X., Aschenbach B., 2015, *ApJ*, 805, 142
- Millard M. J., et al., 2019, arXiv e-prints, p. arXiv:1905.04475
- Mishra A., Reddy P., Nigam R., 2019, arXiv e-prints, p. arXiv:1908.04682
- Morii M., et al., 2016, *PASJ*, 68, 104
- Nair V., Hinton G. E., 2010, in *Proceedings of the 27th international conference on machine learning*

- (ICML-10). pp 807–814
- Nieto Castaño D., Brill A., Kim B., Humensky T. B., Consortium C., 2017, in 35th International Cosmic Ray Conference (ICRC2017). p. 809 (arXiv:1709.05889)
- Nishizuka N., Sugiura K., Kubo Y., Den M., Watari S., Ishii M., 2017, *ApJ*, 835, 156
- Nishizuka N., Sugiura K., Kubo Y., Den M., Ishii M., 2018, *ApJ*, 858, 113
- Nomoto K., Thielemann F. K., Yokoi K., 1984, *ApJ*, 286, 644
- Nwankpa C., Ijomah W., Gachagan A., Marshall S., 2018, arXiv preprint arXiv:1811.03378
- Ohm S., van Eldik C., Egberts K., 2009, *Astroparticle Physics*, 31, 383
- Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, 12, 2825
- Redmon J., Farhadi A., 2018, arXiv e-prints, p. arXiv:1804.02767
- Redmon J., Divvala S., Girshick R., Farhadi A., 2015, arXiv e-prints, p. arXiv:1506.02640
- Reynolds S. P., Borkowski K. J., Hwang U., Hughes J. P., Badenes C., Laming J. M., Blondin J. M., 2007, *ApJ*, 668, L135
- Reynoso E. M., Goss W. M., 1999, *AJ*, 118, 926
- Reynoso E. M., Hughes J. P., Moffett D. A., 2013, *AJ*, 145, 104
- Ribli D., Dobos L., Csabai I., 2019, arXiv e-prints, p. arXiv:1902.08161
- Rodríguez A. C., Kacprzak T., Lucchi A., Amara A., Sgier R., Fluri J., Hofmann T., Réfrégier A., 2018, *Computational Astrophysics and Cosmology*, 5, 4
- Ronneberger O., Fischer P., Brox T., 2015, arXiv e-prints, p. arXiv:1505.04597
- Ruiz-Lapuente P., Damiani F., Bedin L., González Hernández J. I., Galbany L., Pritchard J., Canal R., Méndez J., 2018, *ApJ*, 862, 124
- Rybicki G. B., Lightman A. P., 2008, *Radiative processes in astrophysics*. John Wiley & Sons
- Sankrit R., Raymond J. C., Blair W. P., Long K. S., Williams B. J., Borkowski K. J., Patnaude D. J., Reynolds S. P., 2016, *ApJ*, 817, 36
- Sato T., Hughes J. P., 2017a, *ApJ*, 840, 112
- Sato T., Hughes J. P., 2017b, *ApJ*, 845, 167
- Sato T., Katsuda S., Morii M., Bamba A., Hughes J. P., Maeda Y., Ishida M., Frascchetti F., 2018, *ApJ*, 853, 46
- Sato T., Hughes J. P., Williams B. J., Morii M., 2019, *ApJ*, 879, 64
- Schaefer C., Geiger M., Kuntzer T., Kneib J. P., 2018, *A&A*, 611, A2
- Schawinski K., Zhang C., Zhang H., Fowler L., Santhanam G. K., 2017, *MNRAS*, 467, L110
- Seitzzahl I. R., et al., 2013, *MNRAS*, 429, 1156
- Shilon I., et al., 2019, *Astroparticle Physics*, 105, 44
- Smilkov D., Thorat N., Kim B., Viégas F., Wattenberg M., 2017, arXiv e-prints, p. arXiv:1706.03825
- Sun L., Chen Y., 2019, *ApJ*, 872, 45
- Tamagawa T., et al., 2009, *PASJ*, 61, S167
- Tashiro M., et al., 2018, in *Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray*. p. 1069922, doi:10.1117/12.2309455
- Tenenbaum J. B., De Silva V., Langford J. C., 2000, *science*, 290, 2319
- Thielemann F. K., Nomoto K., Yokoi K., 1986, *A&A*, 158, 17

- 
- Tolstikhin I., Bousquet O., Gelly S., Schoelkopf B., 2017, preprint, (arXiv:1711.01558)
- Tsebrenko D., Soker N., 2013, MNRAS, 435, 320
- Uchida H., Yamaguchi H., Koyama K., 2013, ApJ, 771, 56
- Uchiyama Y., Aharonian F. A., Tanaka T., Takahashi T., Maeda Y., 2007, Nature, 449, 576
- Van Der Maaten L., 2009, in Artificial Intelligence and Statistics. pp 384–391
- Vink J., 2008, ApJ, 689, 231
- Ward Jr J. H., 1963, Journal of the American statistical association, 58, 236
- Warren J. S., 2006, PhD thesis, Rutgers The State University of New Jersey - New Brunswick
- Warren J. S., et al., 2005, ApJ, 634, 376
- Williams B. J., Borkowski K. J., Reynolds S. P., Ghavamian P., Blair W. P., Long K. S., Sankrit R., 2012, ApJ, 755, 3
- Williams B. J., et al., 2017, ApJ, 842, 28
- Wilms J., Allen A., McCray R., 2000, ApJ, 542, 914
- Winkler P. F., Gupta G., Long K. S., 2003, ApJ, 585, 324
- Yamaguchi H., et al., 2008, PASJ, 60, S141
- Yamaguchi H., Tanaka M., Maeda K., Slane P. O., Foster A., Smith R. K., Katsuda S., Yoshii R., 2012, ApJ, 749, 137
- Yamaguchi H., et al., 2014, ApJ, 780, 136
- Yamaguchi H., Hughes J. P., Badenes C., Bravo E., Seitzzahl I. R., Martínez-Rodríguez H., Park S., Petre R., 2017, ApJ, 834, 124
- van der Maaten L., Hinton G., 2008, Journal of machine learning research, 9, 2579