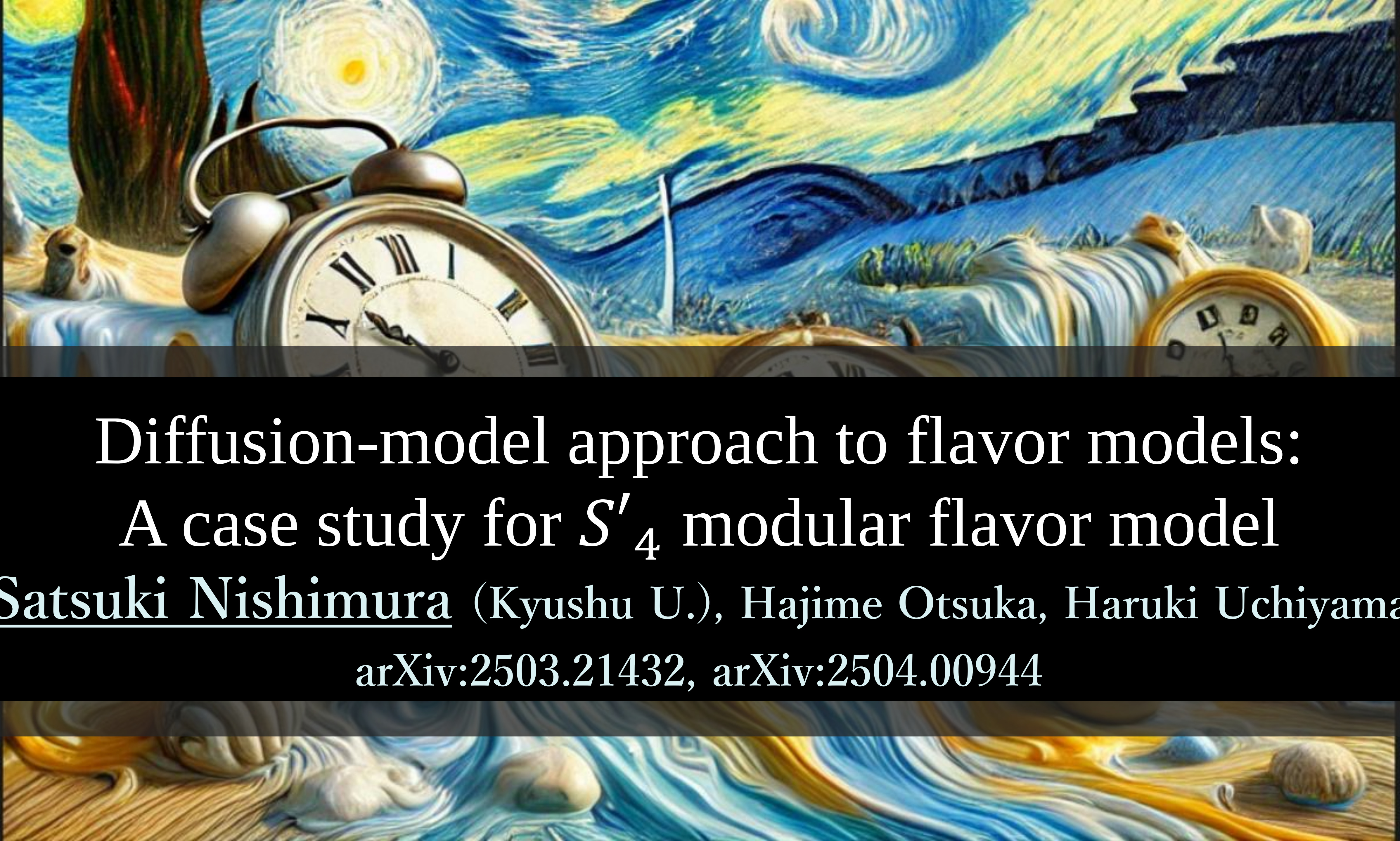


Diffusion-model approach to flavor models: A case study for S'_4 modular flavor model

Satsuki Nishimura (Kyushu U.), Hajime Otsuka, Haruki Uchiyama

arXiv:2503.21432, arXiv:2504.00944

Artificial Intelligence

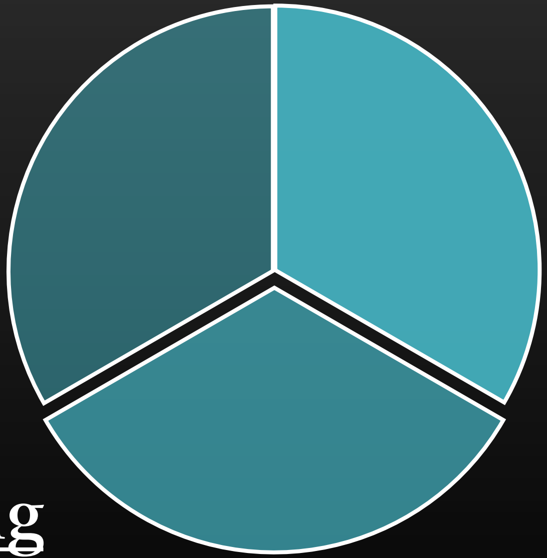
- A generic term for technologies that use computers to perform so-called "intelligent behavior" in place of humans. Ex) Language understanding, Estimation, Problem solving, ⋯

- In particular, the frameworks in which the computers learn itself are called as "machine learning."

Machine Learning ($\subset$ AI)

- A technique in which a computer extracts hidden rules or patterns as it iteratively learns data.

Supervised Learning



Reinforcement Learning

S. Nishimura, C. Miyao, H. Otsuka,
・JHEP12(2023)021 (2304.14176 [hep-ph])
・2409.10023 [hep-ph]

<u>Unsupervised Learning</u>

K. Ishiguro, S. Nishimura, H. Otsuka,
・JHEP08(2024)133 (2312.07181 [hep-th])

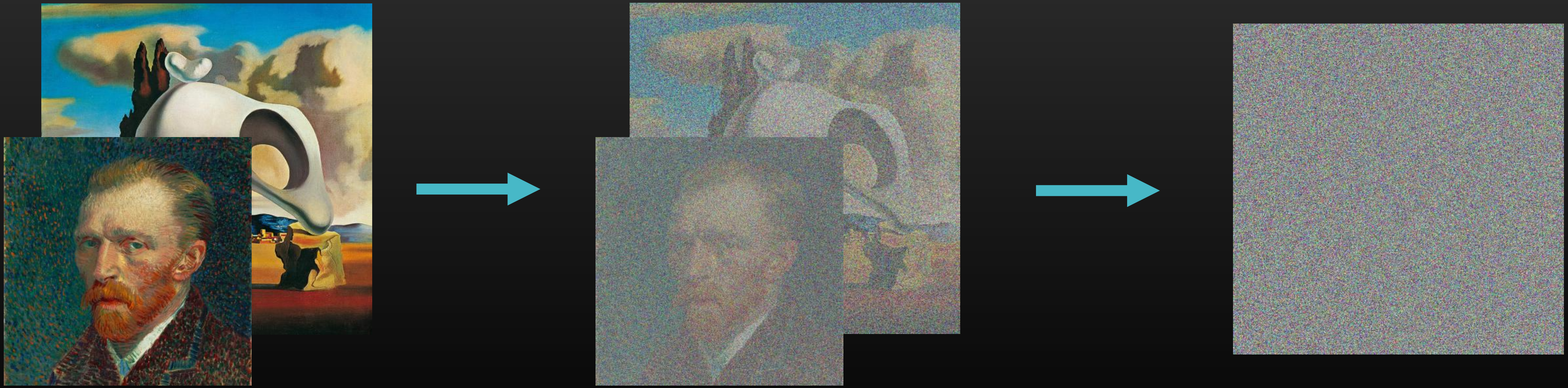
Diffusion Model（DM）

- A type of generative AI, which is used in image generation.
  Ex) AI services “Stable Diffusion” & “DALL-E in ChatGPT”

- Diffusion process ：Add noises to learning data
  Reverse process ：Remove noises from Gaussian noise

Diffusion Model（DM）

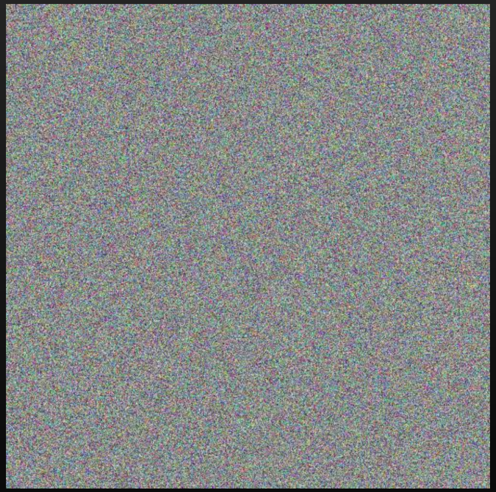
- Diffusion process : Initial pictures G are added noises until they become pure Gaussian noise.



Diffusion Model (DM)

- Reverse process : Pictures G are generated by denoising.

Ex) Make an image with a touch of Salvador Dalí.

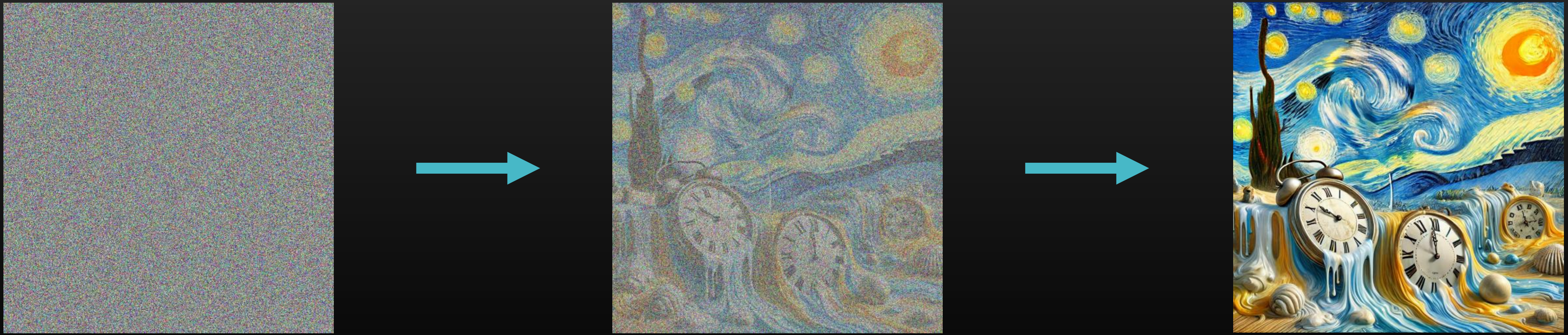


Diffusion Model（DM）

- Conditional DM : Pictures are generated with <u>conditional labels</u> L.

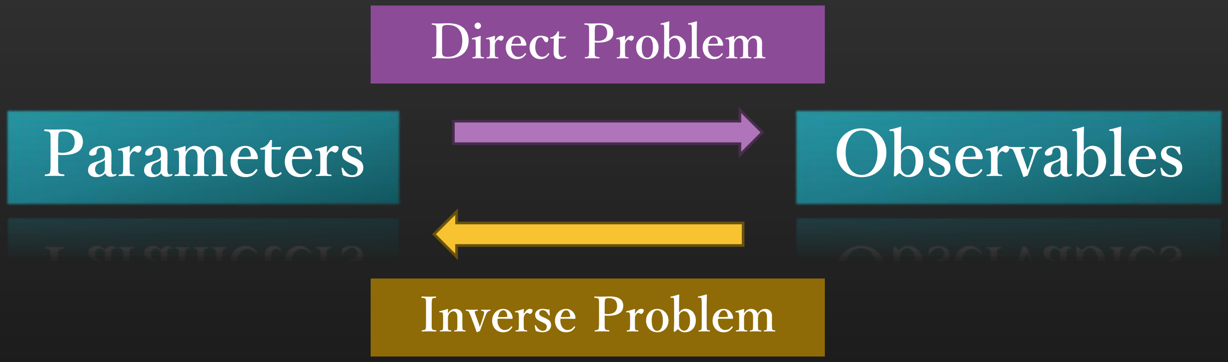
Ex) Make an image of “The Persistence of Memory” by Salvador Dalí,

with a touch of Van Gogh.



Key Point

- Using the diffusion model, we analyze S'_4 modular flavor model.

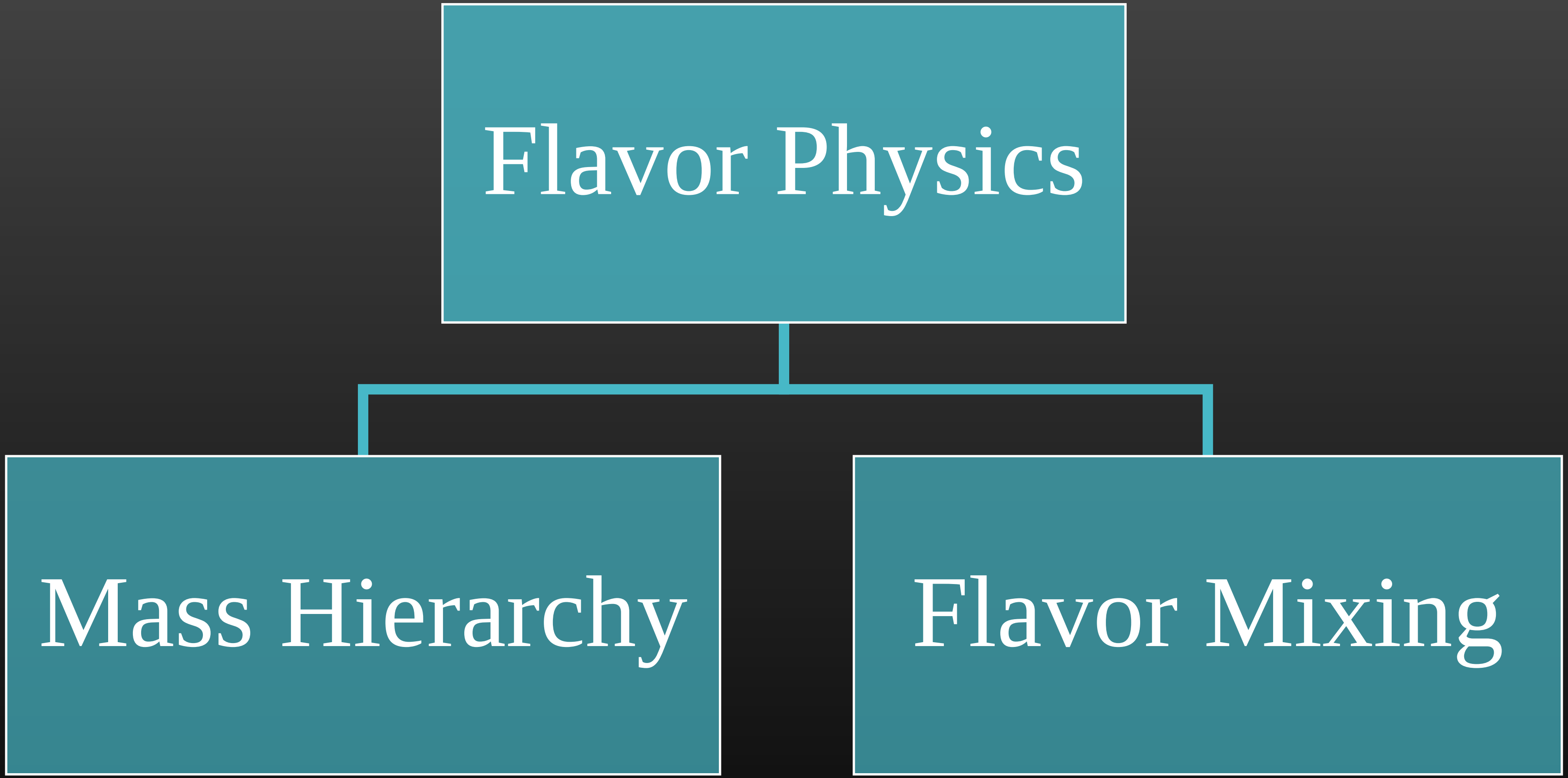


- We found various parameters as an inverse problem, and show the S'_4 model occurs spontaneous CP violation.

The Standard Model (SM)

- SM describes the behavior of elementary particles with a high degree of accurately. It is valid for $\sim 10^{-18}$ m. However, there are many problems. (neutrino masses, generation, ……)

https://higgstan.com/category/phys-item/


Flavor Physics
Mass Hierarchy
Flavor Mixing

Modular Flavor Model（1）

- Modular flavor symmetry can explain
  the hierarchical structure of the quarks & the leptons.
  Ex) $SL(2,Z)$, $\Gamma_2 \simeq S_3$（symmetric group of degree 3）,
  S'_4（double covering of S_4）

- Yukawa couplings are representation of modular symmetry.
  These depend on moduli τ remaining as a degree of freedom.

Modular Flavor Model（2）

Y.Abe, T.Higaki, J.Kawamura, T.Kobayashi
Phys.Lett.B 842 (2023) 137977

- Superpotential of S'_4 modular flavor model

$$W = H_u \left\{ \sum_{a=1}^{2} \alpha_a \left(Q Y_3^{(k_{ua}+k_Q)} \right)_1 u_a^c + \alpha_3 \left(Q Y_{\hat{3}}^{(k_{u3}+k_Q)} u_3^c \right)_1 \right\}$$
$$+ H_d \sum_{i=1}^{3} \left\{ \beta_i \left(Q Y_3^{(k_{di}+k_Q)} d_i^c \right)_1 \right\}$$

- Under modular transformation, modular form with weight k is defined as a function such that $f_i(\gamma\tau) = (c\tau + d)^k \rho_{ij}(\gamma) f_j(\tau)$. Y_3^k means 3 representation with weight k.

Modular Flavor Model（3）

- When representations and weights are fixed,
  Yukawa couplings are calculated using modular forms.

$$Y_u = \begin{pmatrix} \alpha_1 \left[Y_3^{k_{u1}+k_Q}(\tau)\right]_1 & \alpha_2 \left[Y_3^{k_{u2}+k_Q}(\tau)\right]_1 & \alpha_3 \left[Y_{\hat{3}}^{k_{u3}+k_Q}(\tau)\right]_1 \\ \alpha_1 \left[Y_3^{k_{u1}+k_Q}(\tau)\right]_3 & \alpha_2 \left[Y_3^{k_{u2}+k_Q}(\tau)\right]_3 & \alpha_3 \left[Y_{\hat{3}}^{k_{u3}+k_Q}(\tau)\right]_3 \\ \alpha_1 \left[Y_3^{k_{u1}+k_Q}(\tau)\right]_2 & \alpha_2 \left[Y_3^{k_{u2}+k_Q}(\tau)\right]_2 & \alpha_3 \left[Y_{\hat{3}}^{k_{u3}+k_Q}(\tau)\right]_2 \end{pmatrix}$$

Modular Flavor Model（4）

	$\boldsymbol{Q}$	$(\boldsymbol{u_1^c}, \boldsymbol{u_2^c}, \boldsymbol{u_3^c})$	$(\boldsymbol{u_1^c}, \boldsymbol{u_2^c}, \boldsymbol{u_3^c})$	$(\boldsymbol{H_u}, \boldsymbol{H_d})$
S'_4	$\mathbf{3}$	$(\mathbf{1}, \mathbf{1}, \hat{\mathbf{1}}')$	$(\mathbf{1}, \mathbf{1}, \mathbf{1})$	$(\mathbf{1}, \mathbf{1})$
Weight	-4	$(0, -4, -3)$	$(0, -2, -4)$	$(0,0)$

$$Y_u = \begin{pmatrix} \alpha_1 \left[Y_3^{k_{u1}+k_Q}(\tau)\right]_1 & \alpha_2 \left[Y_3^{k_{u2}+k_Q}(\tau)\right]_1 & \alpha_3 \left[Y_{\hat{3}}^{k_{u3}+k_Q}(\tau)\right]_1 \\ \alpha_1 \left[Y_3^{k_{u1}+k_Q}(\tau)\right]_3 & \alpha_2 \left[Y_3^{k_{u2}+k_Q}(\tau)\right]_3 & \alpha_3 \left[Y_{\hat{3}}^{k_{u3}+k_Q}(\tau)\right]_3 \\ \alpha_1 \left[Y_3^{k_{u1}+k_Q}(\tau)\right]_2 & \alpha_2 \left[Y_3^{k_{u2}+k_Q}(\tau)\right]_2 & \alpha_3 \left[Y_{\hat{3}}^{k_{u3}+k_Q}(\tau)\right]_2 \end{pmatrix}$$

- To reproduce masses and mixings of the quarks,
  9 constants α, β and moduli τ should be determined properly.

Motivation

- Diffusion model is known as "Generative AI"
  (generate new data and information based on learning)

- It may be possible to
  - generate unknown parameters that reproduce the flavor structure in an observable-driven way (Inverse problem)
  - consider a wide range of parameters without limiting the search space (in contrast to conventional methods)

Procedure of Learning

- Generating parameters $G = \{\tau, \alpha, \beta\}$ (α, β are real)

  Calculated observables as labels $L = \left\{\frac{m_u}{m_t}, \frac{m_c}{m_t}, \frac{m_d}{m_b}, \frac{m_s}{m_b}, \left|U_{ij}^{\mathrm{CKM}}\right|, J^q\right\}$

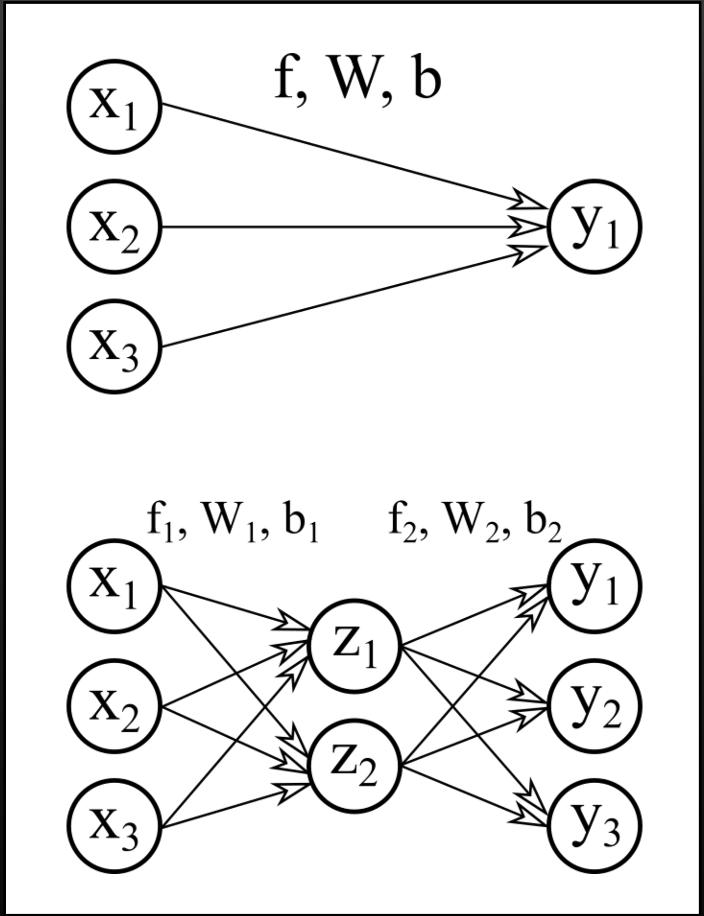
  For various time step $t \in [1, \ldots, T]$ with $T = 1000$,

1）G is prepared randomly, and related label L is calculated

2）A random fluctuation ΔG (noise) is added as $G_t = \frac{T-t}{T} G + \frac{t}{T} \Delta G$

3）A neural network is trained to predict ΔG from G_t & L

Neural Network (1)



- The mathematical model of brains.
  It has input layers (x) & output layers (y).

- Outputs are calculated from the inputs
  using parameters W, b (weight & bias)
  activation function f

$$y_1 = f(W_1 x_1 + W_2 x_2 + W_3 x_3 + b)$$

Neural Network (2)

Input	$\mathbb{R}_G^{10} + \mathbb{R}_{L,t}^{17}$
Hidden 1	$\mathbb{R}^{32} + \mathbb{R}_{L,t}^{17}$
Hidden 2	$\mathbb{R}^{64} + \mathbb{R}_{L,t}^{17}$
Hidden 3	$\mathbb{R}^{128} + \mathbb{R}_{L,t}^{17}$
Hidden 4	$\mathbb{R}^{256} + \mathbb{R}_{L,t}^{17}$
Hidden 5	$\mathbb{R}^{512} + \mathbb{R}_{L,t}^{17}$
Hidden 6	$\mathbb{R}^{512} + \mathbb{R}_{L,t}^{17}$
Hidden 7	$\mathbb{R}^{256} + \mathbb{R}_{L,t}^{17}$
Hidden 8	$\mathbb{R}^{128} + \mathbb{R}_{L,t}^{17}$
Hidden 9	$\mathbb{R}^{64} + \mathbb{R}_{L,t}^{17}$
Hidden 10	$\mathbb{R}^{32} + \mathbb{R}_{L,t}^{17}$
Output	$\mathbb{R}^{10}$

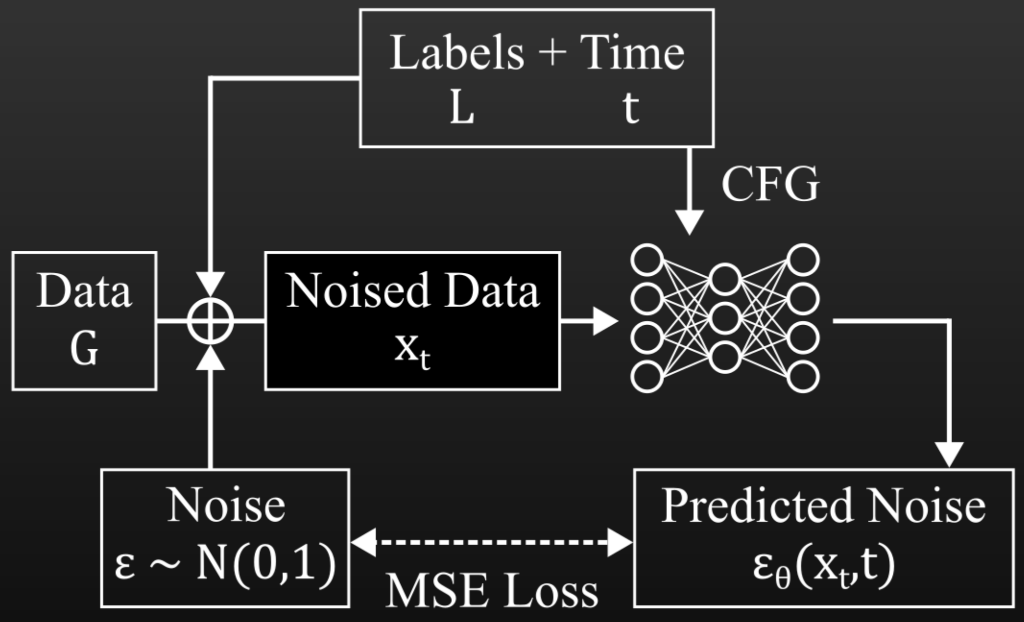
- Loss function is an indicator of the learning status.

- We use mean squared error (MSE)

$$L = (x - \bar{x})^2$$

= difference between added noise and predicted noise.

Neural Network (3)



- In the diffusion model,

  input : noised data & label info

  output : predicted noise

- Weight W & bias b are updated to decrease the difference between actual added noise and predicted noise.

Procedure of Generating

- Generating parameters $G = \{\tau, \alpha, \beta\}$

  Real observables as labels $L_{\mathrm{exp}} = \left\{\frac{m_u}{m_t}, \frac{m_c}{m_t}, \frac{m_d}{m_b}, \frac{m_s}{m_b}, \left|U_{ij}^{\mathrm{CKM}}\right|, J^q\right\}$

1） Pure noise G_T is prepared from Gaussian distribution

2） A predicted fluctuation ΔG is denoised as $G_t = \frac{T-t}{T} G_{t+1} - \frac{t}{T}\Delta G$

3） New data G_0 is generated along experimental label L_{exp}

Calculating Resource

- We used Google Colaboratory with CPU (not GPU)

- The diffusion process takes approximately 1 hour.
  The reverse process requires approximately 4.6 hours
  per generating 10^5 data in a single session.

- While GPUs are strong resource, DM is accessible to everyone.

Progress of Generating（1）

- From the parameters found by DM, 8 observables are calculated.

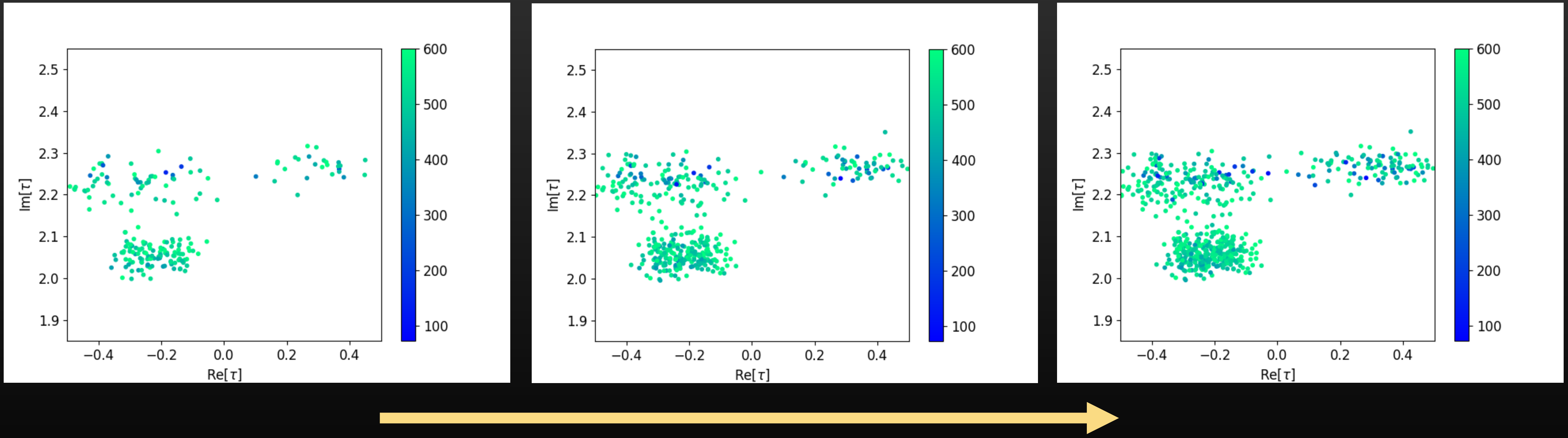
$$\left\{\frac{m_u}{m_t},\frac{m_c}{m_t},\frac{m_d}{m_b},\frac{m_s}{m_b},\theta_{12}^q,\theta_{23}^q,\theta_{13}^q,J^q\right\}$$

- χ^2 is defined from calculated & experimental values $v_{\mathrm{cal}}, v_{\mathrm{exp}}$

$$\chi^2=\sum\chi_i^2=\sum\frac{\left(v_{i,\mathrm{cal}}-v_{i,\mathrm{exp}}\right)^2}{v_{i,\mathrm{error}}^2}$$

Progress of Generating（2）

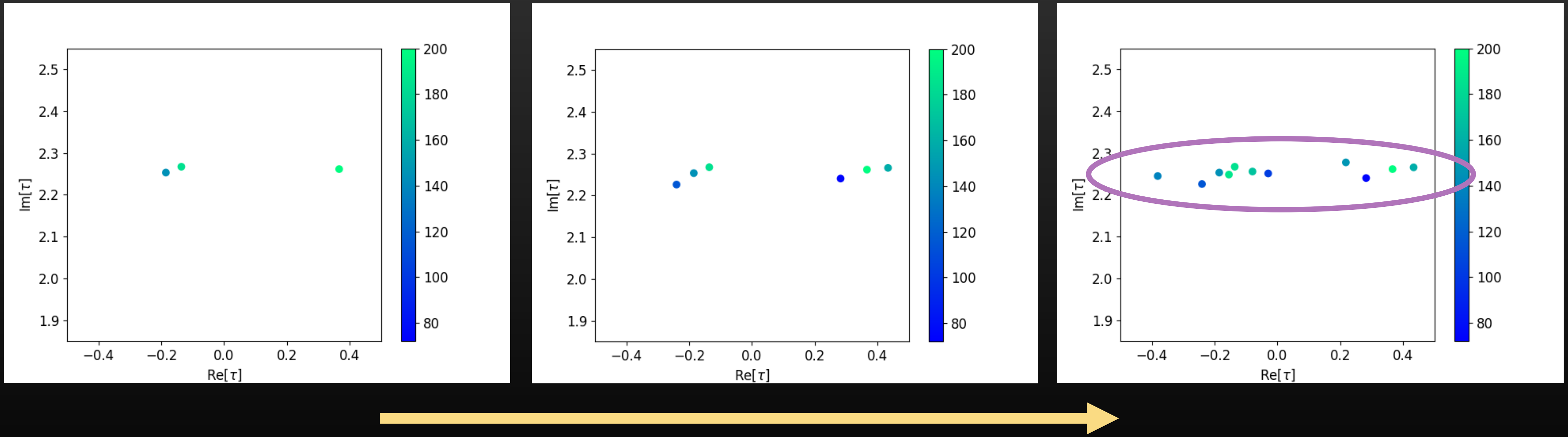
- The distributions of moduli τ with high accuracy are shown.
 (with an increase in the data generated by the DM)



Progress of Generating (3)

- The distributions of moduli τ with high accuracy are shown.

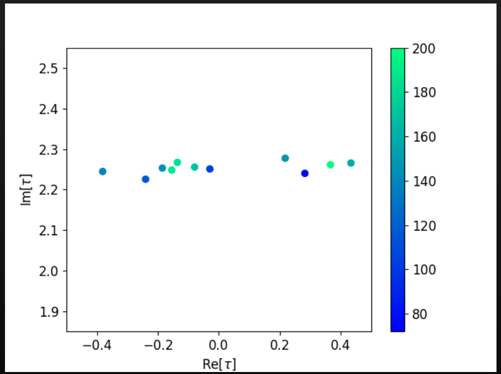
(with an increase in the data generated by the DM)



Result：Various Solutions（1）

- The flavor structure of S'_4 model is sensitive to $\mathrm{Im}\,\tau$.
  It is an arduous task to find good parameter regions.
  Analytically, large $\mathrm{Im}\,\tau$ is preferred. (Ex：$\mathrm{Im}\,\tau = 2.8$)

- We found various solutions with smaller $\mathrm{Im}\,\tau \sim 2.2$
  even without strict limitations for param. region.



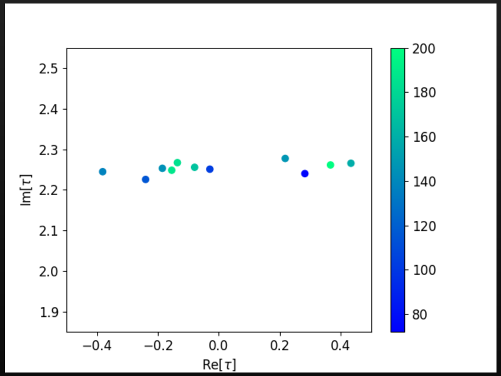
Result：Various Solutions（2）

- In the diffusion process, we adopt following wide region.

$$-\frac{1}{2} \leq \mathrm{Im}\,\tau \leq \frac{1}{2}, \quad \frac{\sqrt{3}}{2} \leq \mathrm{Im}\,\tau \leq 5$$

However, the candidates proposed by AI concentrate $\mathrm{Im}\,\tau \sim 2.2$

- DM automatically suggests promising solutions even in areas that would be difficult to find with human experience alone.

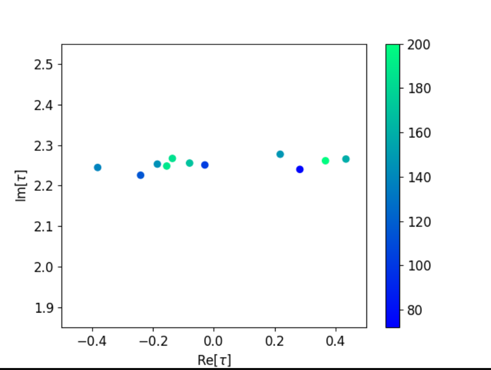


Result : Spontaneous CP Violation

- Coefficients α, β are real. We found $\mathrm{Re}\,\tau$ reproduces Jarlskog invariant, so spontaneous CP violation is derived.

$$J_{\mathrm{DM}} \sim 3.26 \times 10^{-5}, \qquad J_{\mathrm{exp}} = 2.87 \times 10^{-5}$$

- In previous works, complex coefficients are often introduced to reproduce CP violation.
  It is interesting to find the parameters with SCPV.



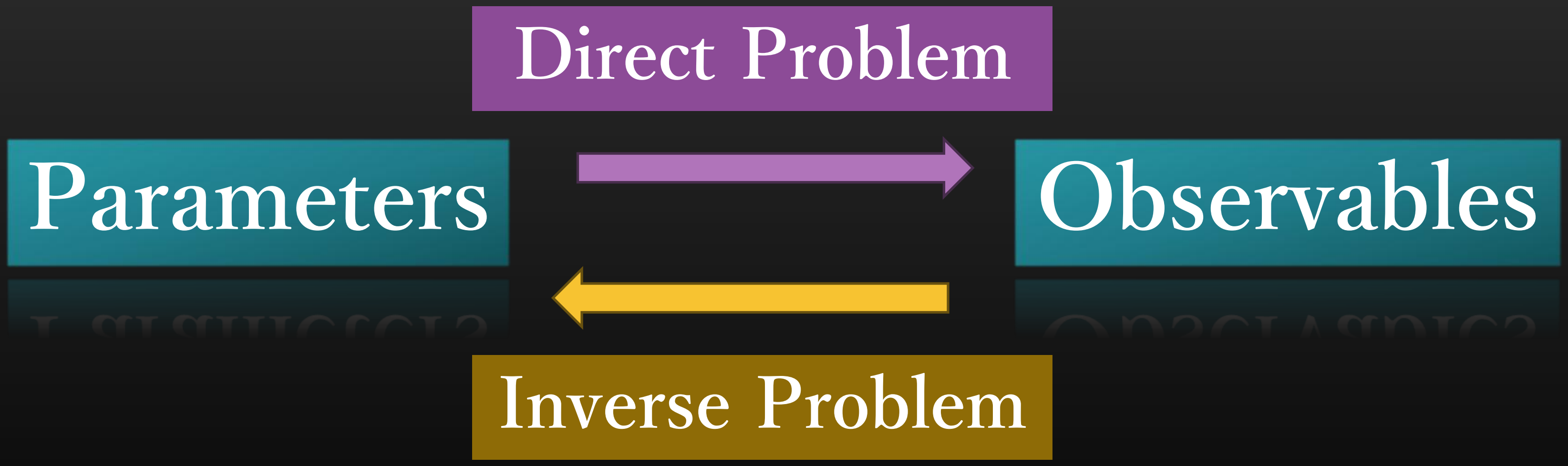
Summary : application of DM for flavor physics

- We focused on the S'_4 modular flavor model, and searched parameters with diffusion model.

- We found various parameters with smaller $\mathrm{Im}\,\tau$ which reproduce the flavor structure of quark sector in an observable-driven way. This approach do not need constraints on searching region.

Summary : application of DM for flavor physics

- Even when coefficients of Yukawa are real,
  the S'_4 model occurs the spontaneous CP violation.



Outlook

- Will spontaneous CP violation be visible with A_4 or other modular symmetries? Will non-trivial parameters be discovered?

- We hope to construct a framework that get predictions from flavor models by simply specifying the experimental values.