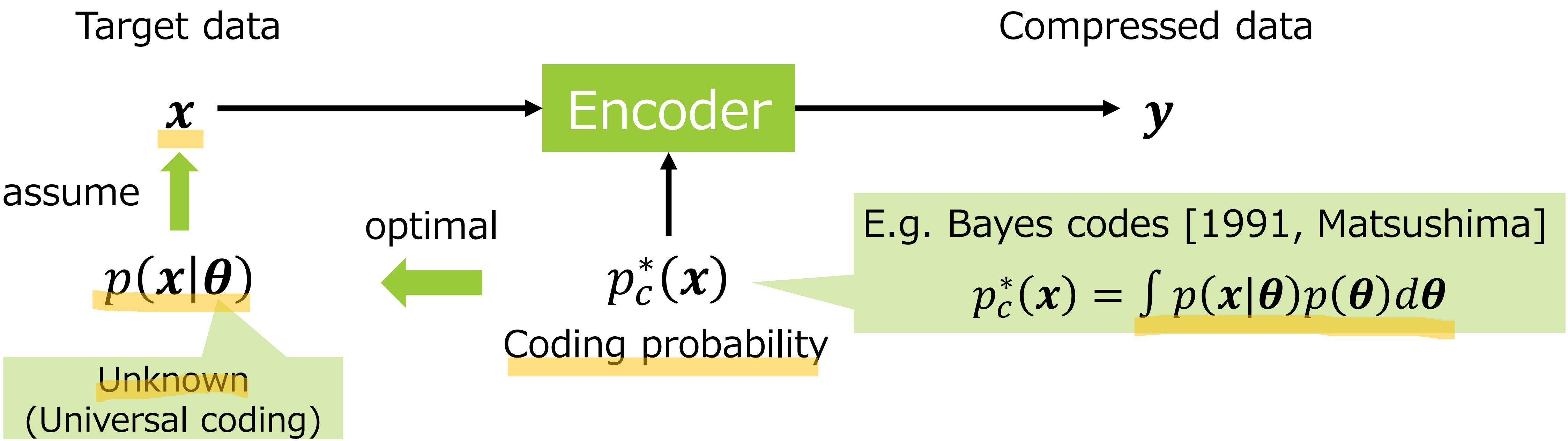
# A Stochastic Model of Block Segmentation Based on the Quadtree and the Bayes Code for It

Waseda University
Center for Data Science Yuta Nakahara
Dept. of Pure and Applied Math. Toshiyasu Matsushima

DCC2020

# Lossless compression of general data

## ■Information theoretical approach

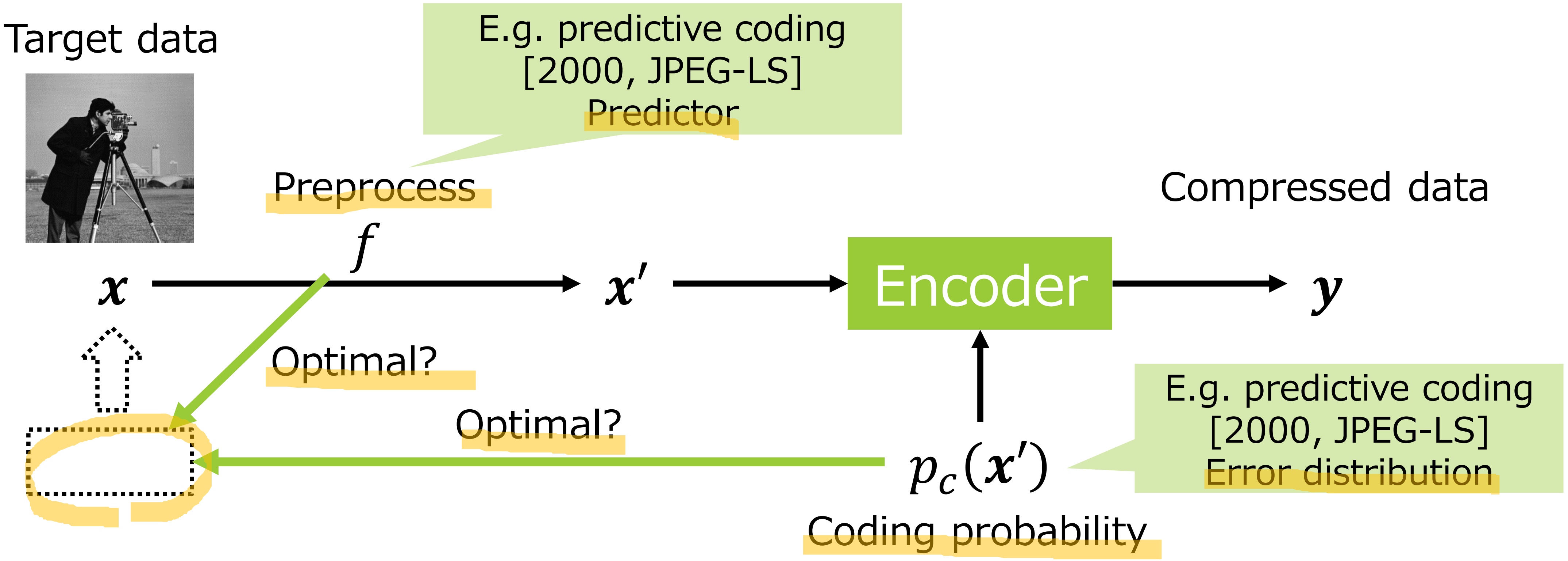



## ■The important factors

- The efficiency of the algorithm to calculate $p_c^*(\boldsymbol{x})$
- The flexibility of the stochastic model $p(\boldsymbol{x}|\boldsymbol{\theta})$

# Lossless image compression

- Approaches in most previous studies



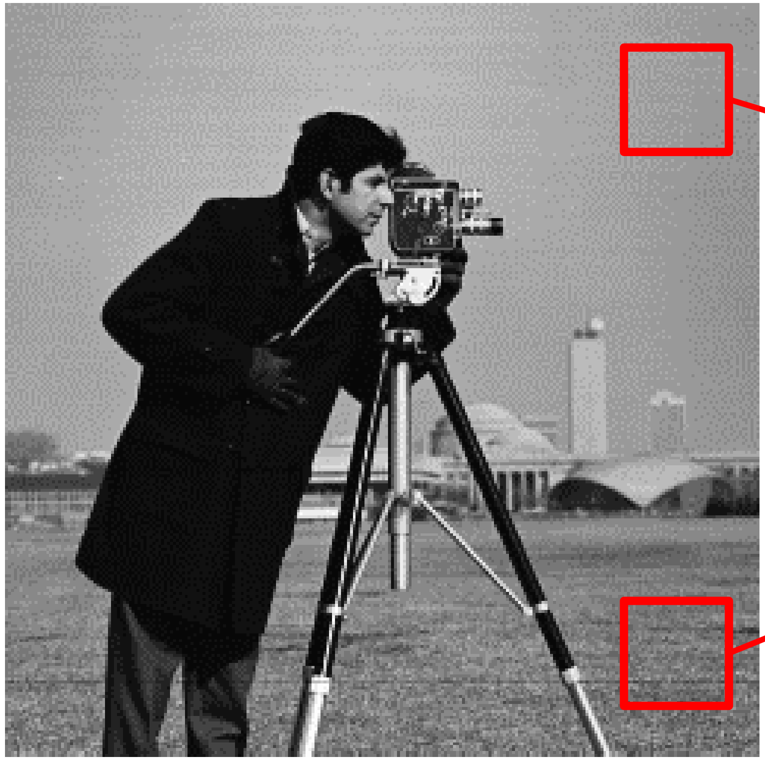

- Difficulty to discuss the optimality of $f$ and $p_c(\boldsymbol{x}')$

# Purposes of this study

1. We propose a stochastic model that effectively represents target images.
2. We derive an optimal coding probability.
3. We derive an efficient algorithm to calculate it without loss of the optimality.

# Target data

- Images with non-stationarity



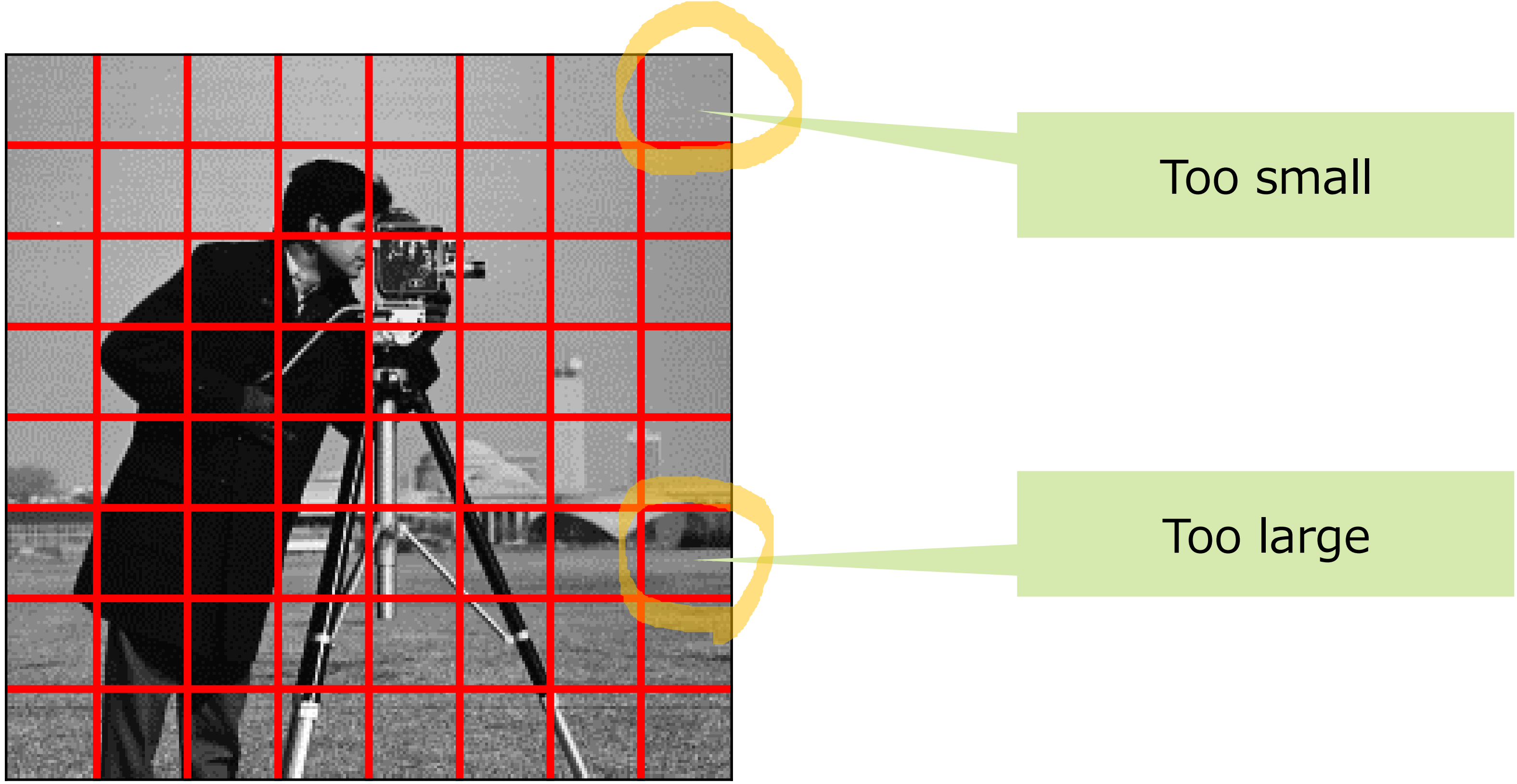
$$p(\boldsymbol{x}|\boldsymbol{\theta})$$

$$\neq$$

$$p(\boldsymbol{x}|\boldsymbol{\theta}')$$

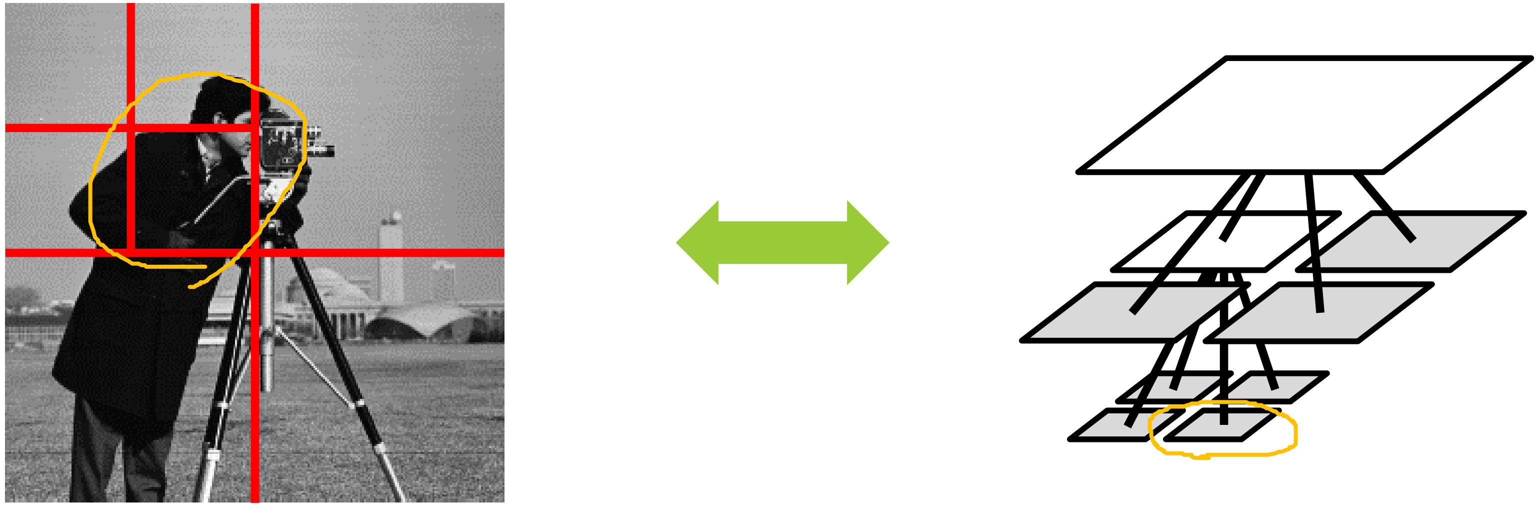
# Trivial way

- Divide the image into fixed size blocks
- Assume different stochastic models to them



# Quadtree

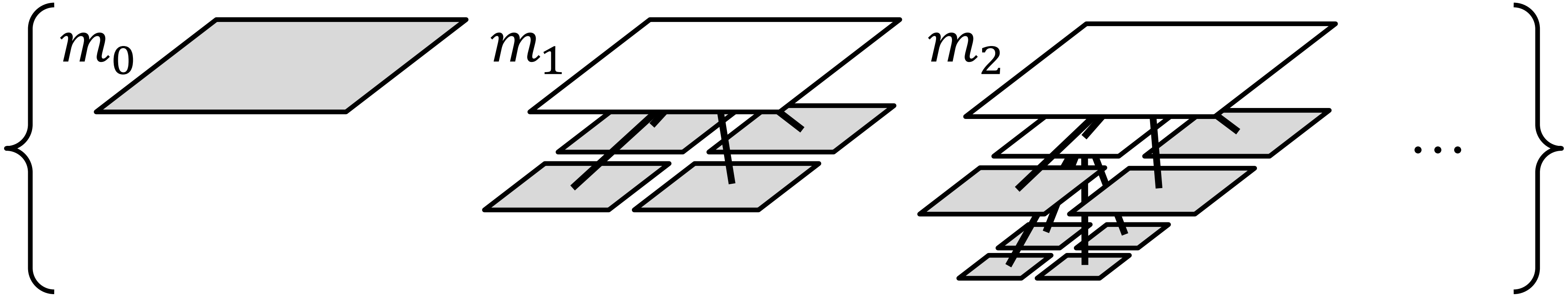
- It effectively represents variable block size segmentation
- 



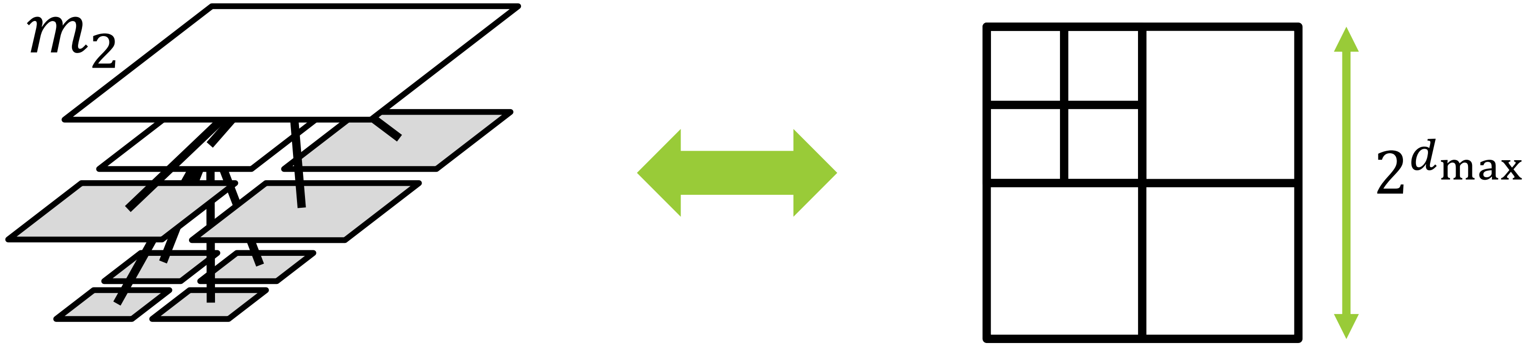
- Previous studies regard it just as a procedure
- In this study, we regard the quadtree as a part of stochastic generative model of the images.

# Proposed stochastic model

- Let both of the width and height are $2^{d_{\max}}$.
- Consider the set of the quadtrees whose depth $\leq d_{\max}$.

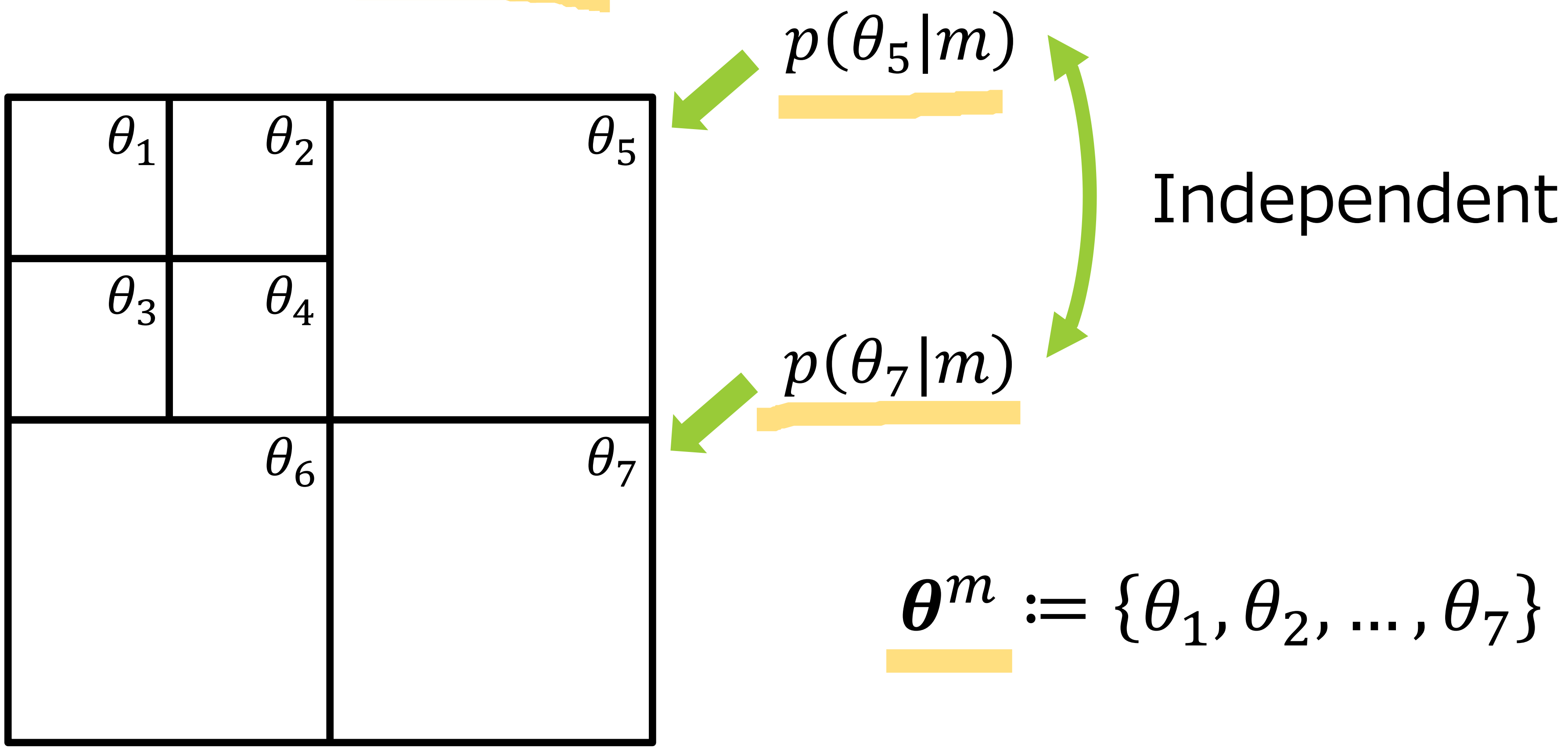
$\mathcal{M} =$ 


- One of them is chosen with probability $p(m)$.



# Proposed stochastic model

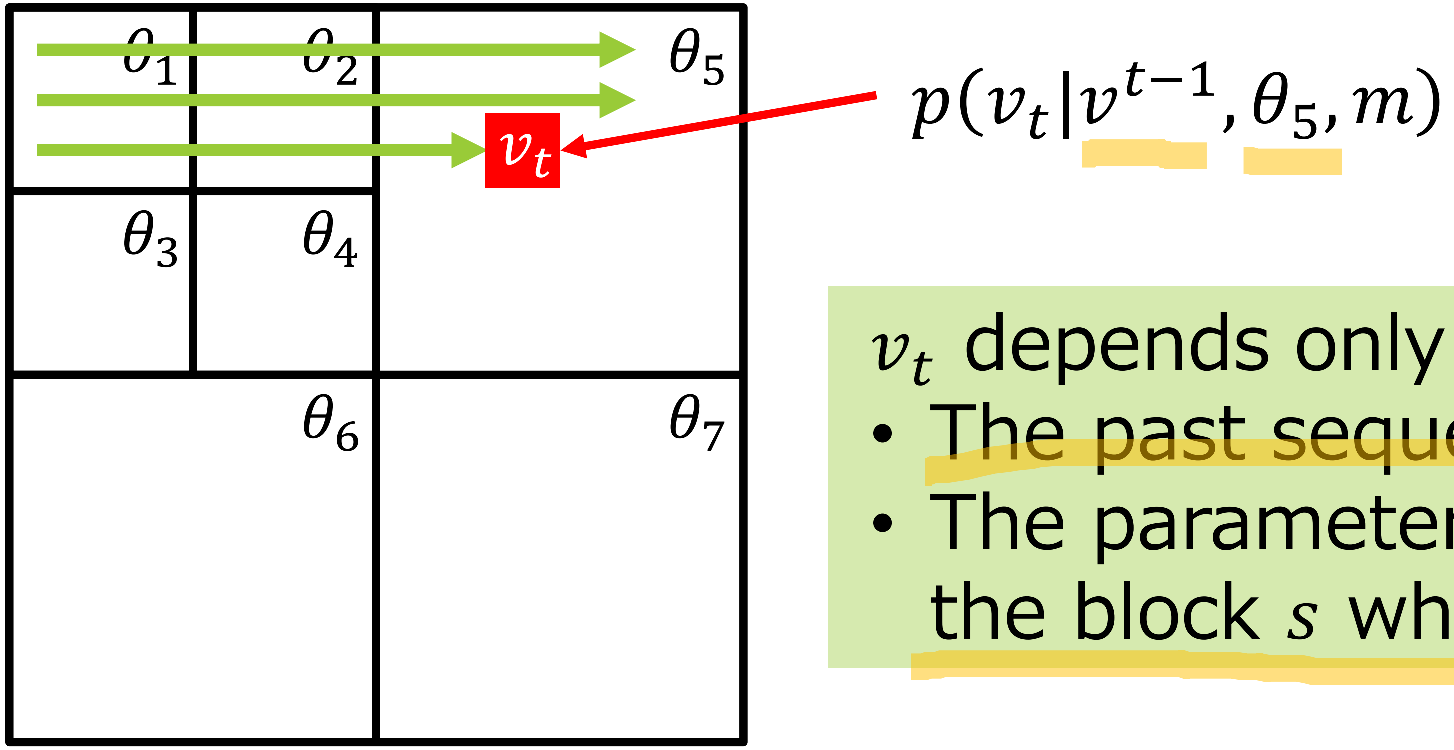
- Parameter $\theta_s$ is independently assigned to each block $s$ with probability $p(\theta_s|m)$.




$$\boldsymbol{\theta}^m := \{\theta_1, \theta_2, \dots, \theta_7\}$$

# Proposed stochastic model

- Pixel value $v_t$ at block $s$ is generated in order of the raster scan with probability $p(v_t|v^{t-1}, \theta_s, m)$.




$v_t$ depends only on
- The past sequence $v^{t-1}$
- The parameter $\theta_s$ of the block $s$ which contains $v_t$

# The Bayes code

- We cannot use $p(v_t|v^{t-1}, \theta^m, m)$ because true $m$ and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner.

- Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

$$p_c^*(v_t|v^{t-1}) = \sum_{m \in \mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \, \mathrm{d}\boldsymbol{\theta}^m$$

- Three computationally hard parts
  - 1. The summation w.r.t. $m$ ← A recursive structure of quadtree
  - 2. The posterior $p(m|v^{t-1})$ ← Special prior (Detailed in the paper)
  - 3. The integral w.r.t. $\boldsymbol{\theta}^m$ ← Conjugate prior

# The Bayes code

- We cannot use $p(v_t|v^{t-1}, \theta^m, m)$ because true $m$ and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner.
- Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

$$p_c^*(v_t|v^{t-1}) = \sum_{m \in \mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \, \mathrm{d}\boldsymbol{\theta}^m$$

- Three computationally hard parts
  - 1. The summation w.r.t. $m$ ← A recursive structure of quadtree
  - 2. The posterior $p(m|v^{t-1})$ ← Special prior (Detailed in the paper)
  - 3. The integral w.r.t. $\boldsymbol{\theta}^m$ ← Conjugate prior

# The Bayes code

- We cannot use $p(v_t|v^{t-1}, \theta^m, m)$ because true $m$ and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner.

- Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

$$p_c^*(v_t|v^{t-1})$$
$$= \sum_{m \in \mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \, \mathrm{d}\boldsymbol{\theta}^m$$

- Three computationally hard parts
  - 1. The summation w.r.t. $m$ ← A recursive structure of quadtree
  - 2. The posterior $p(m|v^{t-1})$ ← Special prior (Detailed in the paper)
  - 3. The integral w.r.t. $\boldsymbol{\theta}^m$ ← Conjugate prior

# The Bayes code

- We cannot use $p(v_t|v^{t-1}, \theta^m, m)$ because true $m$ and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner.

- Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

$$p_c^*(v_t|v^{t-1}) = \sum_{m \in \mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \, \mathrm{d}\boldsymbol{\theta}^m$$

- Three computationally hard parts
  - 1. The summation w.r.t. $m \leftarrow$ A recursive structure of quadtree
  - 2. The posterior $p(m|v^{t-1}) \leftarrow$ Special prior (Detailed in the paper)
  - 3. The integral w.r.t. $\boldsymbol{\theta}^m \leftarrow$ Conjugate prior

# The Bayes code

- We cannot use $p(v_t|v^{t-1}, \theta^m, m)$ because true $m$ and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner.

- Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

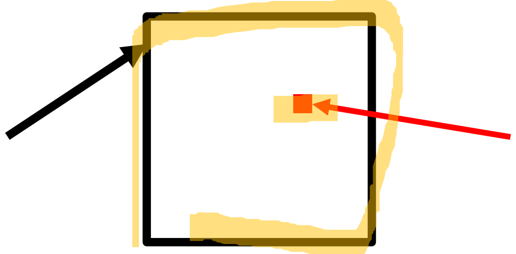
$$p_c^*(v_t|v^{t-1}) = \sum_{m\in\mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m)\, \mathrm{d}\boldsymbol{\theta}^m$$

- Three computationally hard parts
  - 1. The summation w.r.t. $m$ ← A recursive structure of quadtree
  - 2. The posterior $p(m|v^{t-1})$ ← Special prior (Detailed in the paper)
  - 3. The integral w.r.t. $\boldsymbol{\theta}^m$ ← Conjugate prior

# The derived algorithm

- Consider the following case

The whole image

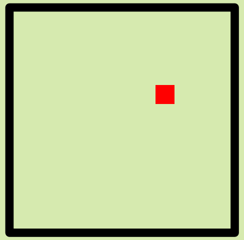


$v_t$: The pixel we want to compress

- $p_c^*(v_t|v^{t-1})$

$$= \underbrace{\left(1 - g_{s_\lambda|t-1}\right)}_{\text{Posterior}} \underbrace{q(v_t|v^{t-1}, s_\lambda)}_{\text{Coding Probability}} + \underbrace{g_{s_\lambda|t-1}}_{\text{Posterior}} \underbrace{p_c(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})}_{\text{Coding Probability}}$$

The model whose minimal block which contains $v_t$ is root block $s_\lambda$



The other models

# The derived algorithm

- $p_c^*(v_t|v^{t-1})$
$$= (1 - g_{s_\lambda|t-1})q(v_t|v^{t-1}, s_\lambda) + g_{s_\lambda|t-1}p_c(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})$$
$$= (1 - g_{s_\lambda|t-1})q(v_t|v^{t-1}, s_\lambda) + \underbrace{g_{s_\lambda|t-1}(1 - g_{s_{\lambda\,\mathrm{child}}|t-1})}_{\text{Posterior}}\underbrace{q(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})}_{\text{Coding Probability}}$$
$$+ \underbrace{g_{s_\lambda|t-1}g_{s_{\lambda\,\mathrm{child}}|t-1}}_{\text{Posterior}}\underbrace{p_c(v_t|v^{t-1}, s_{\lambda\,\mathrm{child\,child}})}_{\text{Coding Probability}}$$

# The derived algorithm

- $p_c^*(v_t|v^{t-1})$

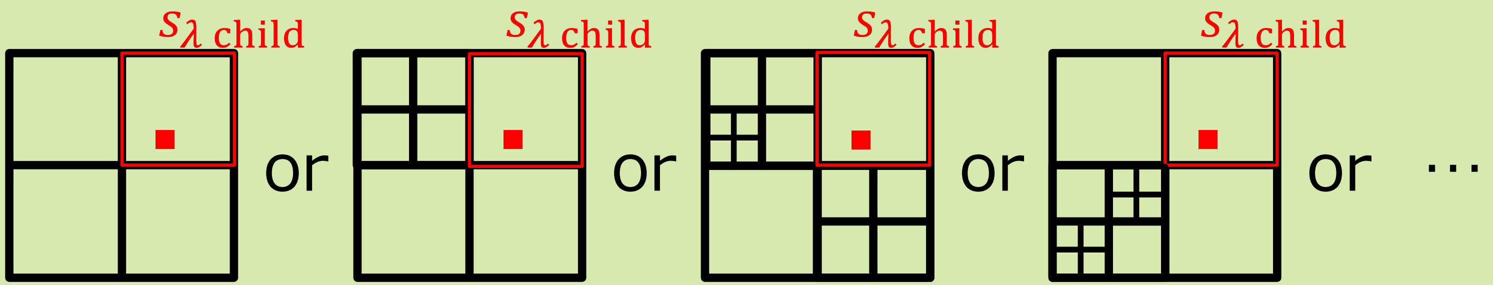
$$= \left(1 - g_{s_\lambda|t-1}\right) q(v_t|v^{t-1}, s_\lambda) + g_{s_\lambda|t-1} p_c(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})$$

$$= \left(1 - g_{s_\lambda|t-1}\right) q(v_t|v^{t-1}, s_\lambda)$$

$$+ g_{s_\lambda|t-1}\left(1 - g_{s_{\lambda\,\mathrm{child}}|t-1}\right) q(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})$$

Posterior Coding Probability

$$+ \ldots_1 g_{s_{\lambda\,\mathrm{child}}|t-1} p_c(v \ldots s_{\lambda\,\mathrm{child\,child}})$$




The models whose minimal block which contains $v_t$ is $s_{\lambda\,\mathrm{child}}$

# The derived algorithm

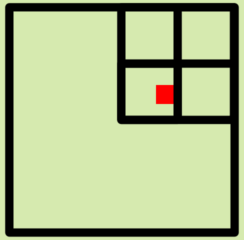
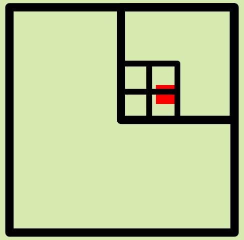
- $p_c^*(v_t|v^{t-1})$

$$= \left(1 - g_{s_\lambda|t-1}\right) q(v_t|v^{t-1}, s_\lambda) + g_{s_\lambda|t-1} \underbrace{p_c(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})}$$

$$= \left(1 - g_{s_\lambda|t-1}\right) q(v_t|v^{t-1}, s_\lambda)$$

$$+ \underbrace{g_{s_\lambda|t-1}\left(1 - g_{s_{\lambda\,\mathrm{child}}|t-1}\right)}_{\text{Posterior}} \underbrace{q(v_t|v^{t-1}, s_{\lambda\,\mathrm{child}})}_{\text{Coding Probability}}$$

$$+ \underbrace{g_{s_\lambda|t-1} g_{s_{\lambda\,\mathrm{child}}|t-1}}_{\text{Posterior}} \underbrace{p_c(v_t|v^{t-1}, s_{\lambda\,\mathrm{child\,child}})}_{\text{Coding Probability}}$$

The other models:  or  or ⋯

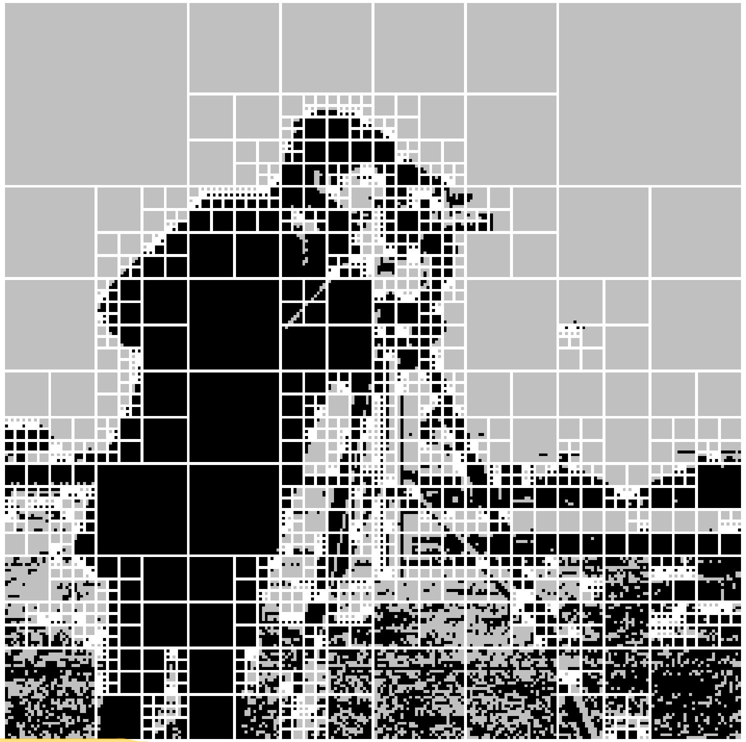
# Experiment 1

- Purpose: To confirmation of the Bayes optimality
- Setting:
  - We generated 1000 binary images (64x64) as follows:
    1. Generate $m$ according to $p(m)$ (detailed in the paper).
    2. Generate $\theta_s$ according to $Beta(\theta_s|\alpha,\beta)$ for each block $s$.
    3. Generate $v_t$ according to $Bern(v_t|\ \theta_s)$ for each block $s$.
  - We compressed them by
    - the proposed method
    - the method with fixed size block segmentation
- Result: Average coding rates (bit/pel)

| Proposed | Fixed size 4 | Fixed size 8 | Fixed size 16 |
|---|---|---|---|
| **0.619** | 0.705 | 0.659 | 0.679 |

# Experiment 2

- Purpose: To demonstrate the model flexibility
- We compressed the binarized image of the cameraman.
- Result: MAP estimated model $m^{\mathrm{MAP}} = \mathrm{argmax}_{m \in \mathcal{M}} p(m|\boldsymbol{v})$.



# Conclusion and future works

- Conclusion
  - We proposed the novel stochastic model with the quadtree.
  - We derived the Bayes optimal coding probability for it.
  - We derived the efficient algorithm to calculate it.

- Future works
  - Our proposed model can be used for other problems
    - Image generation
    - Image inpainting
    - Future extraction