Quantitative Steganalysis of LSB Embedding in JPEG Domain

Jan Kodovský, Jessica Fridrich

September 10, 2010 / ACM MM&Sec '10

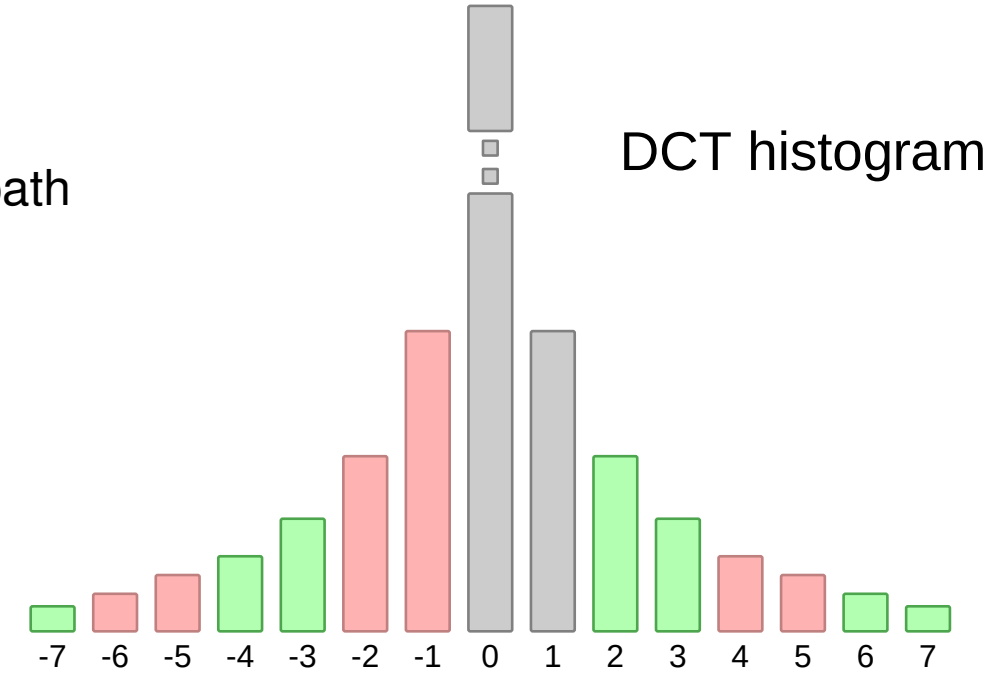
BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Motivation

- Least Significant Bit (LSB) embedding
  - Simplicity, high embedding capacity
  - Used in Jsteg, JP Hide&Seek, and other commercial stego software
- Steganalysis of LSB embedding in spatial domain is mature area
  - [Dumitrescu-2002], [Ker-2008]
- Our focus
  - Transform domain – JPEG format
- Quantitative steganalysis
  - Outputs the estimate of the message length

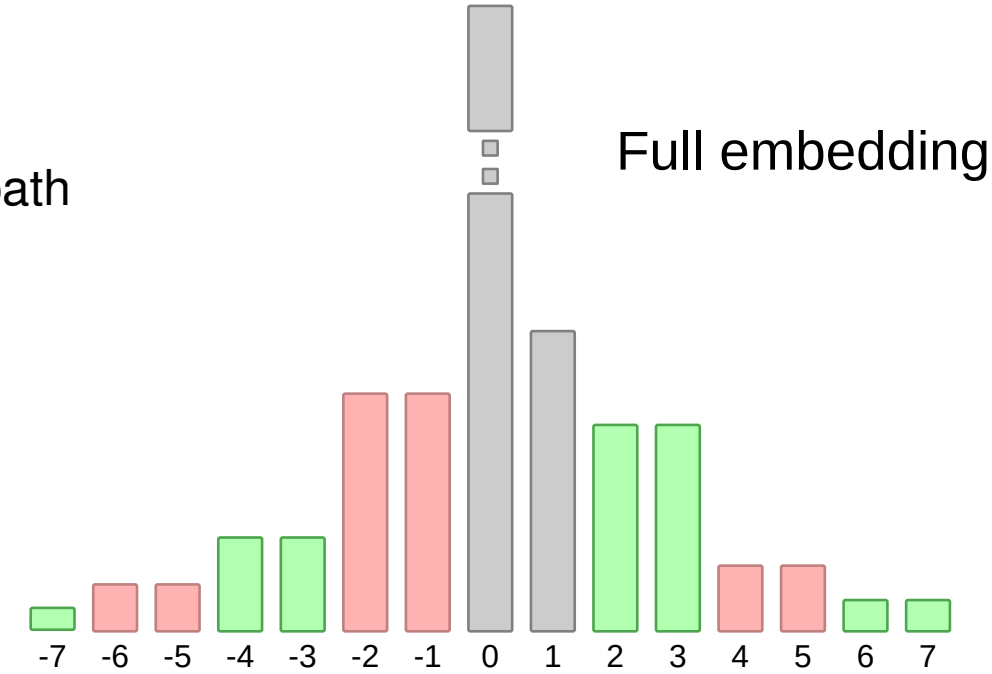
Jsteg

- Jsteg: [Upham-1993]
- LSB replacement
  - Embedding along a pseudo-random path
  - Skipping 0 and 1



Jsteg

- Jsteg: [Upham-1993]
- LSB replacement
  - Embedding along a pseudo-random path
  - Skipping 0 and 1



Embedding violates histogram symmetry

Selected Existing Attacks

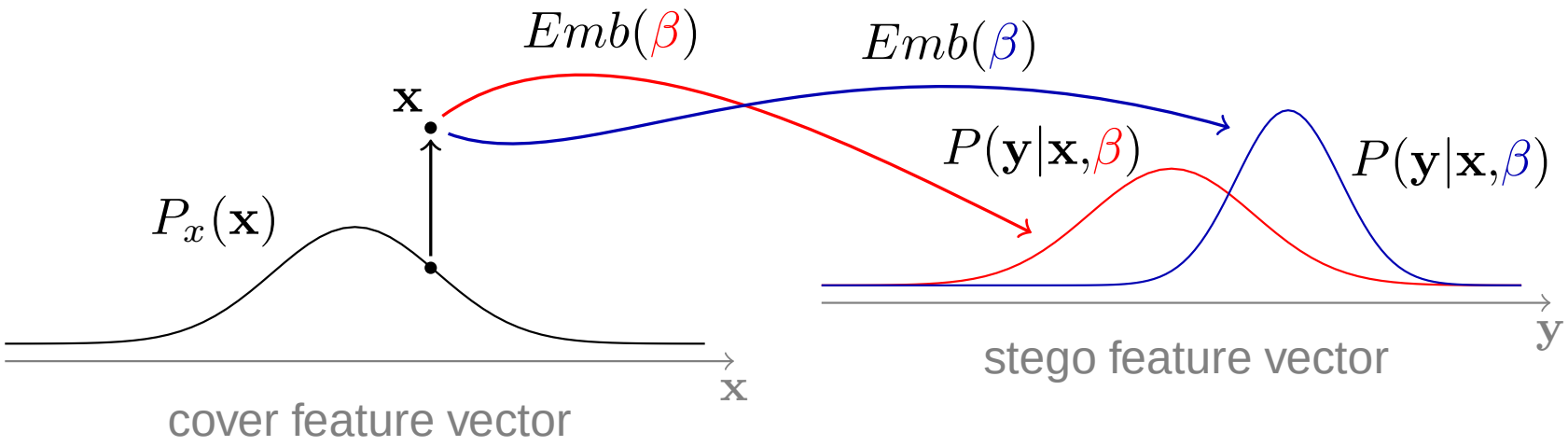
- [Zhang,Ping-2003] – the first quantitative attack
  - Employed violation of histogram symmetry
- [Yu-2004] – histogram-based attack
  - Generalized Cauchy ML fit
  - Chi-square test
- [Lee-2006], [Lee-2007] – Category attack
  - Technically not quantitative
- [Westfeld-2007], [Böhme-2008] – adaptation of spatial domain attacks
- [Pevný-2009] – support vector regression
  - Feature-based non-structural attack
  - Currently the most accurate quantitative attack

Our Goals / Challenges

- Improve the accuracy of existing *quantitative* attacks to Jsteg
- Achieve better performance than the feature-based machine learning approach (SVR)
- Focus on the *structure* of LSB embedding
- Deliver theoretically well-founded modular framework
- Explore the applicability of the proposed attacks to a different LSB embedding paradigms

Maximum Likelihood

β ... change rate



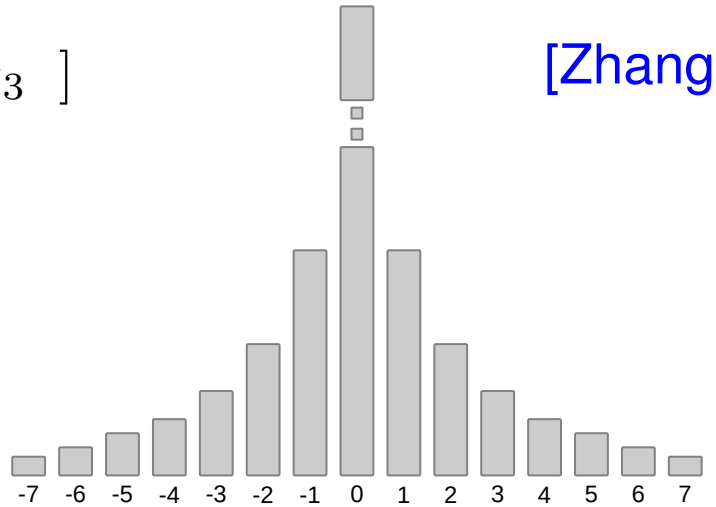
$$P(\mathbf{y},\beta) = \int P(\mathbf{y},\mathbf{x},\beta)\mathrm{d}\mathbf{x} = \int P(\mathbf{y}|\mathbf{x},\beta)P(\mathbf{x},\beta)\mathrm{d}\mathbf{x} = P(\beta)\int P(\mathbf{y}|\mathbf{x},\beta)P_x(\mathbf{x})\mathrm{d}\mathbf{x}$$

$$\hat{\beta} = \arg\max_{\beta\geq 0} P(\mathbf{y}|\beta) = \arg\max_{\beta\geq 0} \int P(\mathbf{y}|\mathbf{x},\beta)P_x(\mathbf{x})\mathrm{d}\mathbf{x}$$

Choice of the feature vector $\mathbf{x}$ is crucial

Features of Zhang & Ping

$$\mathbf{x} = [\ x_1 \quad x_2 \quad x_3\ ]$$



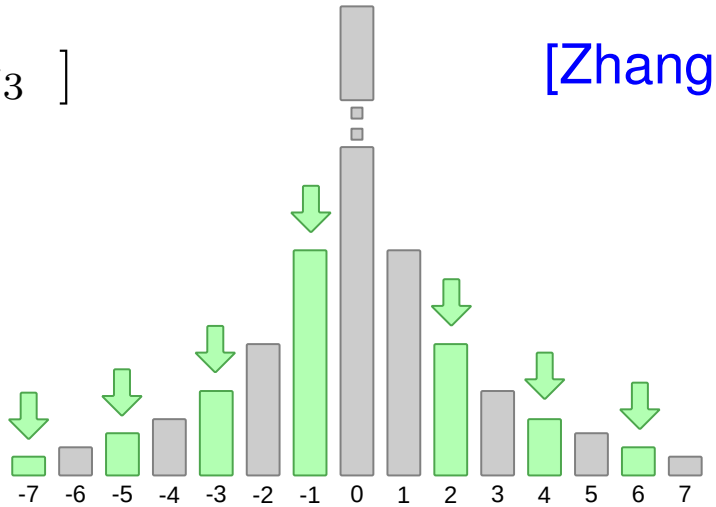
[Zhang,Ping-2003]

- $P(\mathbf{y}|\mathbf{x}, \beta) \longrightarrow$ Binomial distribution $\longrightarrow$ Gaussian approximation
- Embedding invariants: $x_1 + x_2,\ x_3$
- $P_x(\mathbf{x}) \rightarrow$ precover assumption [Ker-2007]

Features of Zhang & Ping

$$\mathbf{x} = [\ x_1 \quad x_2 \quad x_3\ ]$$

[Zhang,Ping-2003]

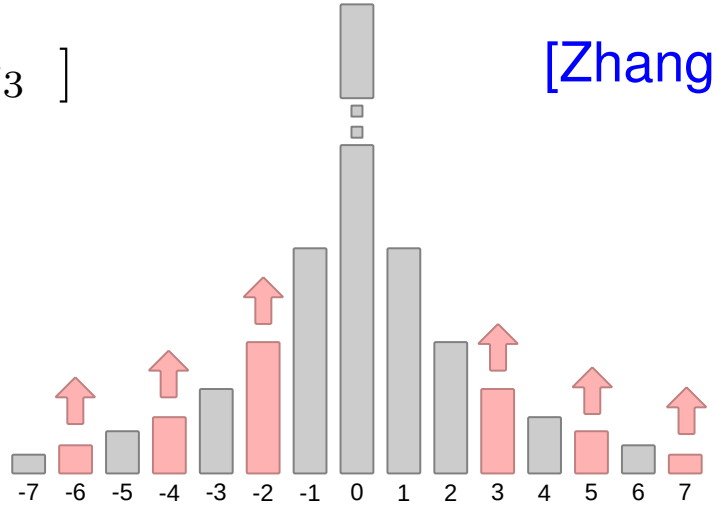


- $P(\mathbf{y}|\mathbf{x},\beta) \longrightarrow$ Binomial distribution $\longrightarrow$ Gaussian approximation
- Embedding invariants: $x_1 + x_2,\ x_3$
- $P_x(\mathbf{x}) \rightarrow$ precover assumption [Ker-2007]

Features of Zhang & Ping

$$\mathbf{x} = [\ x_1 \quad \underbrace{x_2} \quad x_3\ ]$$

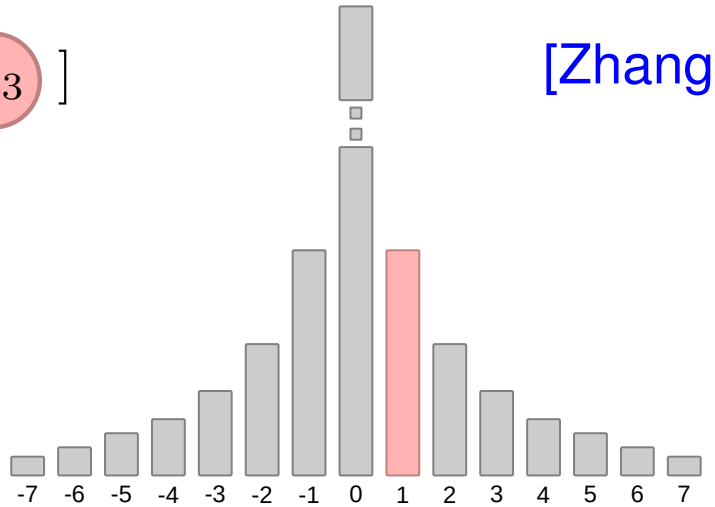
[Zhang,Ping-2003]



- $P(\mathbf{y}|\mathbf{x}, \beta) \longrightarrow$ Binomial distribution $\longrightarrow$ Gaussian approximation
- Embedding invariants: $x_1 + x_2,\ x_3$
- $P_x(\mathbf{x}) \rightarrow$ precover assumption [Ker-2007]

Features of Zhang & Ping

$$\mathbf{x} = [\ x_1 \quad x_2 \quad x_3\ ]$$



[Zhang,Ping-2003]

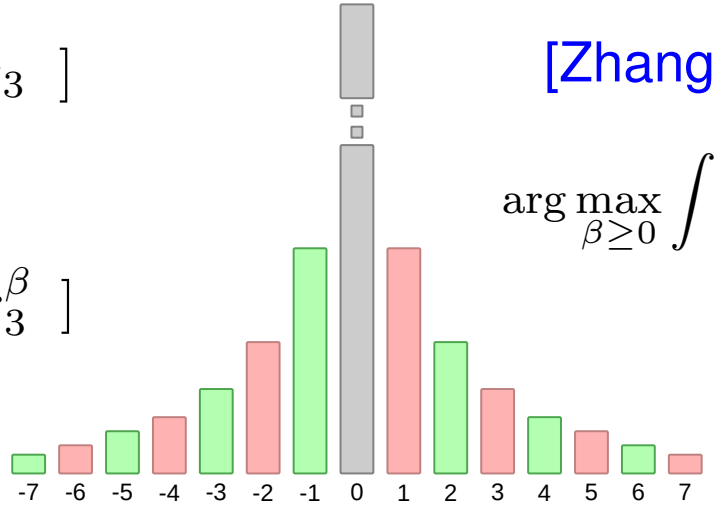
- $P(\mathbf{y}|\mathbf{x}, \beta) \longrightarrow$ Binomial distribution $\longrightarrow$ Gaussian approximation
- Embedding invariants: $x_1 + x_2,\ x_3$
- $P_x(\mathbf{x}) \rightarrow$ precover assumption [Ker-2007]

Features of Zhang & Ping

$$\mathbf{x} = [\ x_1 \quad x_2 \quad x_3\ ]$$

$$Emb(\beta)\ \Big(\ 1-\beta \downarrow \quad \beta \searrow$$

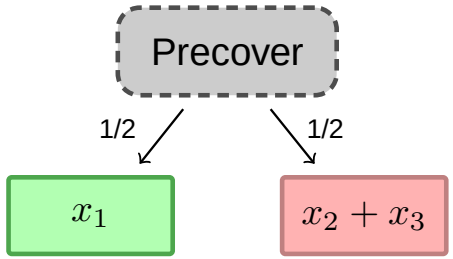
$$\mathbf{y} = [\ x_1^{\beta} \quad x_2^{\beta} \quad x_3^{\beta}\ ]$$



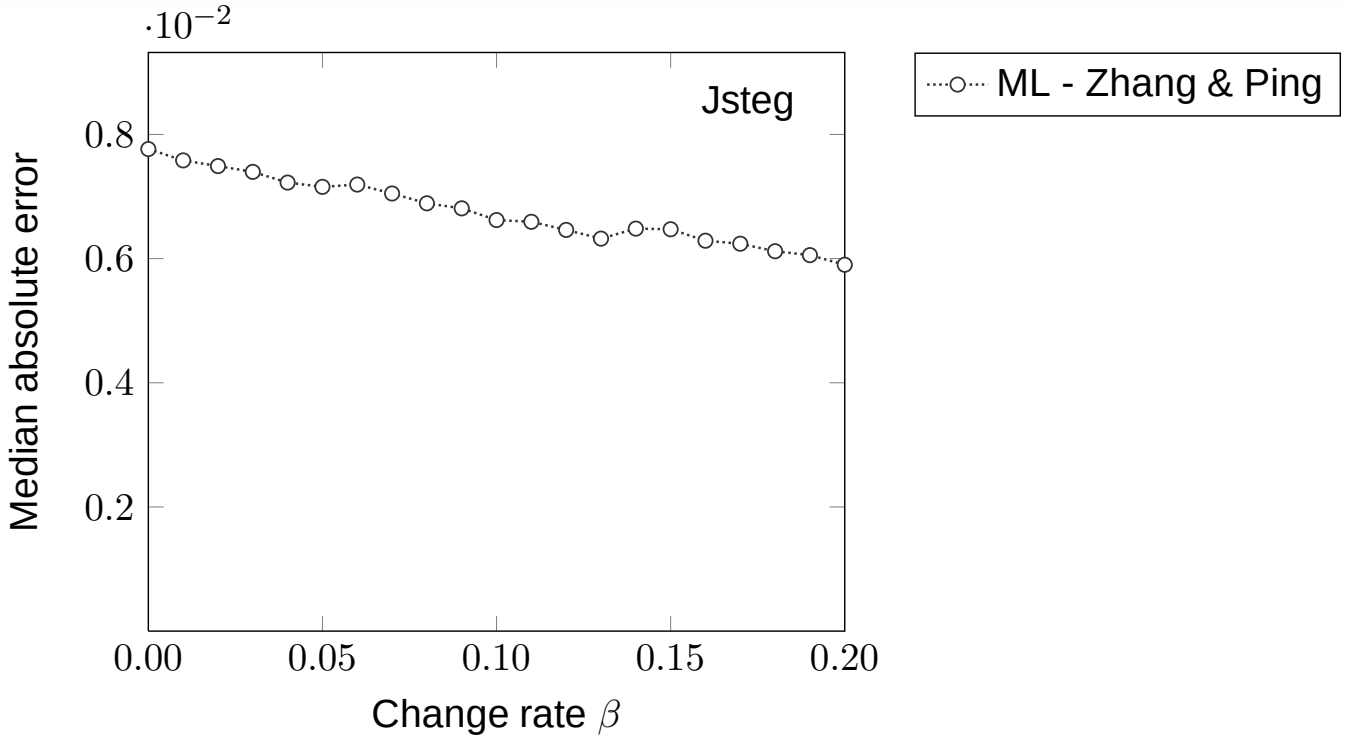
[Zhang,Ping-2003]

$$\arg\max_{\beta \geq 0} \int P(\mathbf{y}|\mathbf{x},\beta) P_x(\mathbf{x}) \mathrm{d}\mathbf{x}$$

- $P(\mathbf{y}|\mathbf{x},\beta) \longrightarrow$ Binomial distribution $\longrightarrow$ Gaussian approximation
- Embedding invariants: $x_1 + x_2,\ x_3$
- $P_x(\mathbf{x}) \rightarrow$ precover assumption [Ker-2007]

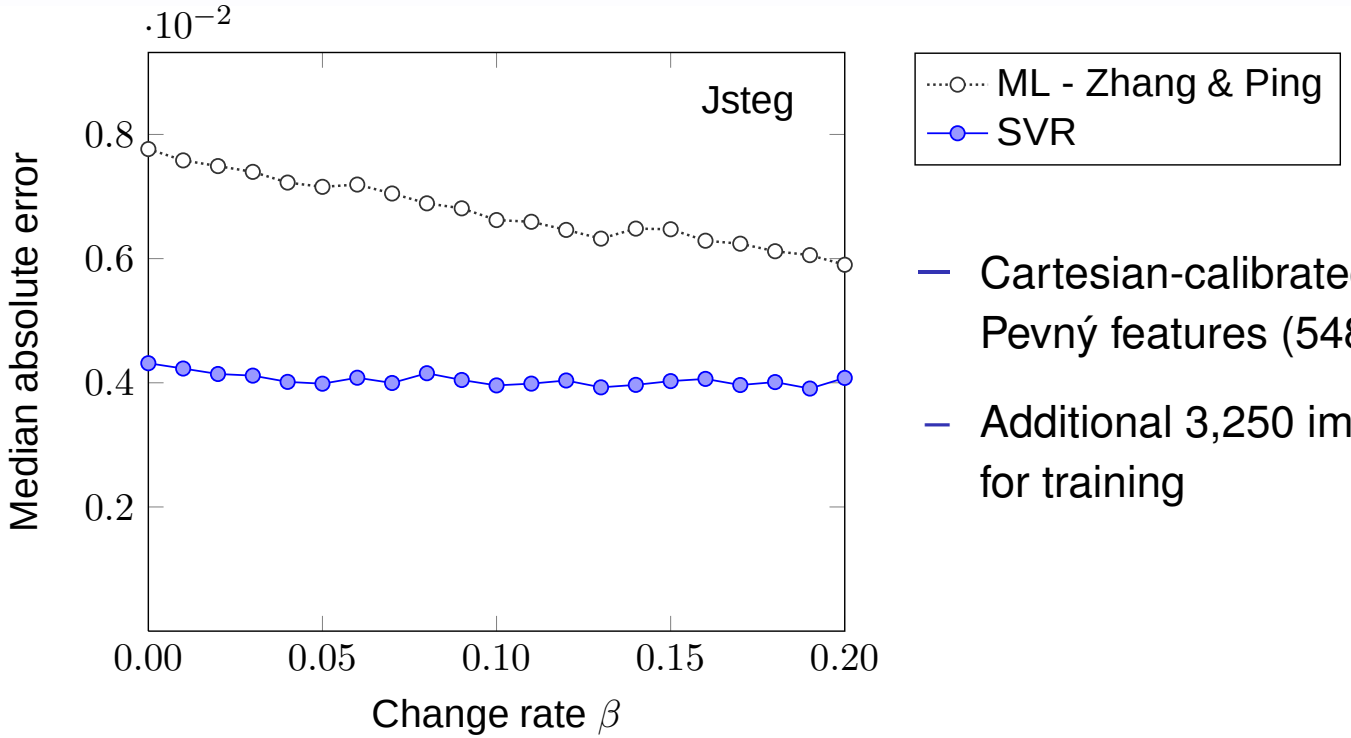


Performance Evaluation



- 3,250 JPEG images – resized and compressed to QF=75
- Performance similar to [Zhang,Ping-2003]
- Assumption $x_1^{\beta} =$ expected value $\Rightarrow$ Zhang & Ping's estimator

Performance Evaluation



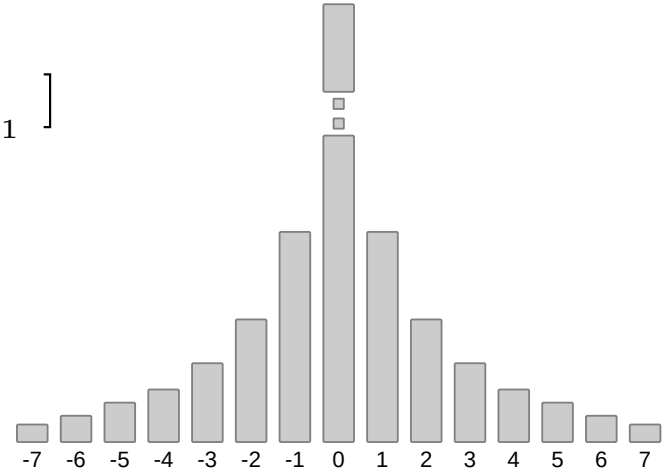
- Cartesian-calibrated Pevný features (548)
- Additional 3,250 images for training

- 3,250 JPEG images – resized and compressed to QF=75
- Performance similar to [Zhang,Ping-2003]
- Assumption $x_1^\beta = \text{expected value} \Rightarrow$ Zhang & Ping's estimator

First-Order Statistics

$$\mathbf{x} = [\, x_{-2L}, \; x_{-2L+1}, \ldots, x_{2R}, \; x_{2R+1} \,]$$

$$\hat{\beta} = \arg\max_{\beta \geq 0} \int P(\mathbf{y}|\mathbf{x}, \beta) P_x(\mathbf{x}) \mathrm{d}\mathbf{x}$$



- Embedding changes in individual LSB pairs are independent

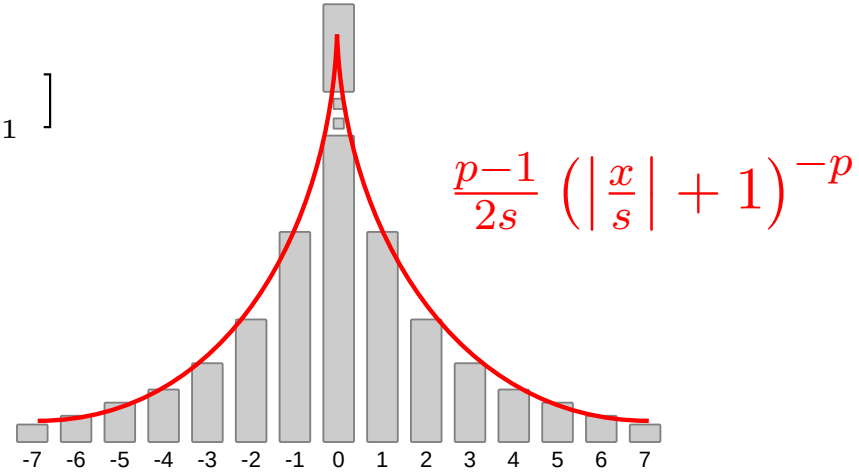
$$P(\mathbf{y}|\mathbf{x}, \beta) = P\left(x_0^\beta | x_0, \beta\right) \cdot P\left(x_1^\beta | x_1, \beta\right) \cdot \prod_k P\left(x_{2k}^\beta, x_{2k+1}^\beta | x_{2k}, x_{2k+1}, \beta\right)$$

- Embedding invariants: $x_0, \; x_1, \; x_{2k} + x_{2k+1}$
- Binomial distribution $\longrightarrow$ Gaussian approximation

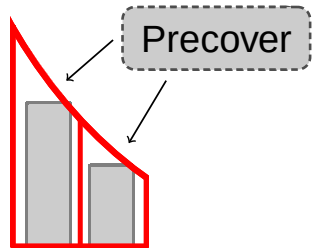
First-Order Statistics

$$\mathbf{x} = [\, x_{-2L}\,,\; x_{-2L+1}\,, \ldots, x_{2R}\,,\; x_{2R+1}\,]$$

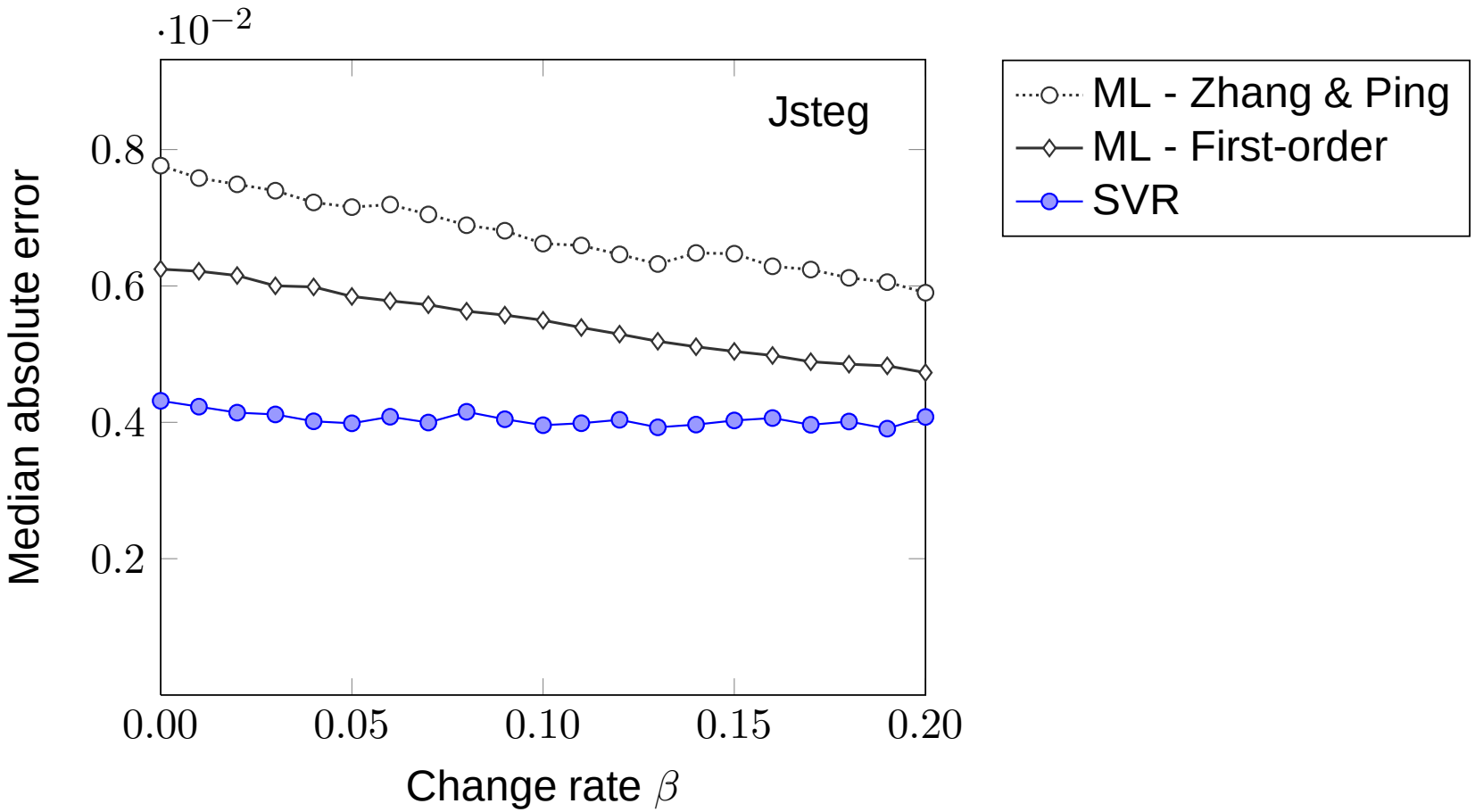
$$\hat{\beta} = \arg\max_{\beta \geq 0} \int P(\mathbf{y}|\mathbf{x}, \beta) P_x(\mathbf{x}) \mathrm{d}\mathbf{x}$$



- DCT coefficients are i.i.d. drawn from generalized Cauchy distribution
- Parameters p and s are ML estimates, given embedding invariants
- Precover assumption for every LSB pair
- Embedding invariants: $x_0,\; x_1,\; x_{2k} + x_{2k+1}$



Performance Evaluation



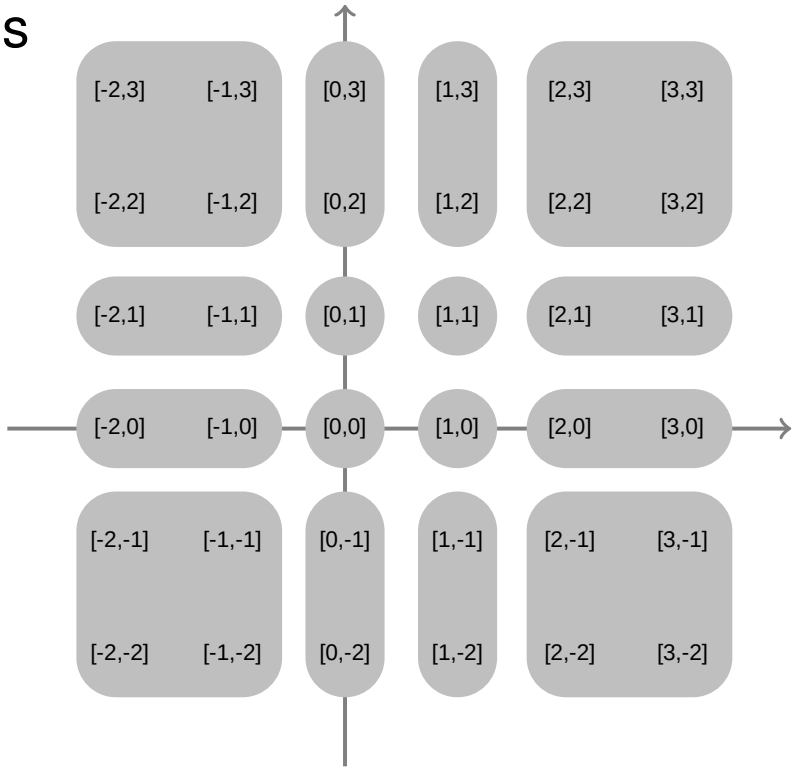
Second-Order Statistics

- DCT coefficients are **not** i.i.d.
- We capture dependencies using adjacency matrix $\mathbf{X}$
- Natural decomposition into k-nodes, $k \in \{1, 2, 4\}$
- Binomial / multinomial distributions

  $\longrightarrow$ Gaussian approximations

$$\arg\max_{\beta \geq 0} \int P(\mathbf{Y}|\mathbf{X}, \beta) P_x(\mathbf{x}) \mathrm{d}\mathbf{x}$$

- Factorization of $P(\mathbf{y}|\mathbf{x}, \beta)$
- Embedding invariants
- Analytic expression

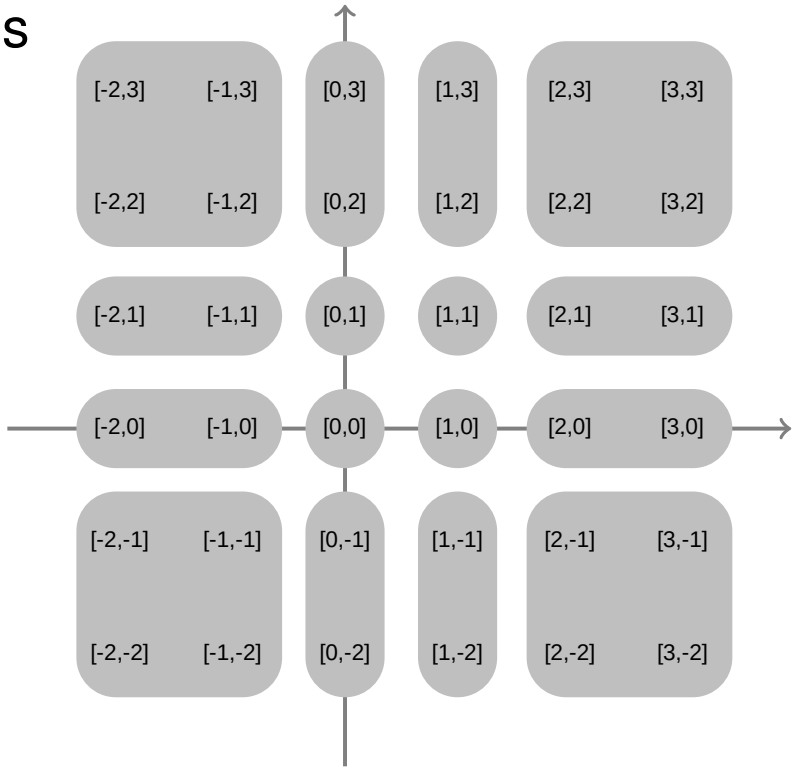


Second-Order Statistics

- DCT coefficients are **not** i.i.d.
- We capture dependencies using adjacency matrix $\mathbf{X}$
- Natural decomposition into k-nodes, $k \in \{1, 2, 4\}$
- Binomial / multinomial distributions

  $\longrightarrow$ Gaussian approximations

$$\arg\max_{\beta \geq 0} \int P(\mathbf{Y}|\mathbf{X}, \beta) P_x(\mathbf{x}) \mathrm{d}\mathbf{x}$$



- Complications arise
- Good parametric model ?
- High complexity

Zero Message Hypothesis (ZMH)

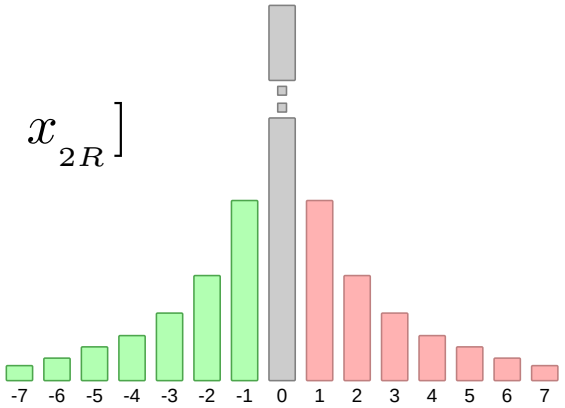
- Alternative heuristic approach
  - Penalty function $z(\mathbf{x}) \geq 0$ satisfying $z(\mathbf{x}^{\beta}) \approx 0$ when $\beta = 0$, $z(\mathbf{x}^{\beta}) > 0$ when $\beta > 0$
  - $z(\mathbf{x})$ should be a quantitative description of a zero message hypothesis capturing a key cover property violated by embedding
  - Assumption: $\mathbf{y} = E[\mathbf{x}^{\beta}] = Emb(\mathbf{x}, \beta)$
  - Assumption: mapping Emb is invertible $\Rightarrow \mathbf{x} = Emb^{-1}(\mathbf{y}, \beta)$

$$\hat{\beta} = \arg\min_{\beta \geq 0} z(Emb^{-1}(\mathbf{y}, \beta))$$

- Comments
  - Low computational complexity – one-dimensional search over β
  - ZMH-based steganalysis is not a new idea! [RS steganalysis,2001]

First-Order Statistics (ZMH)

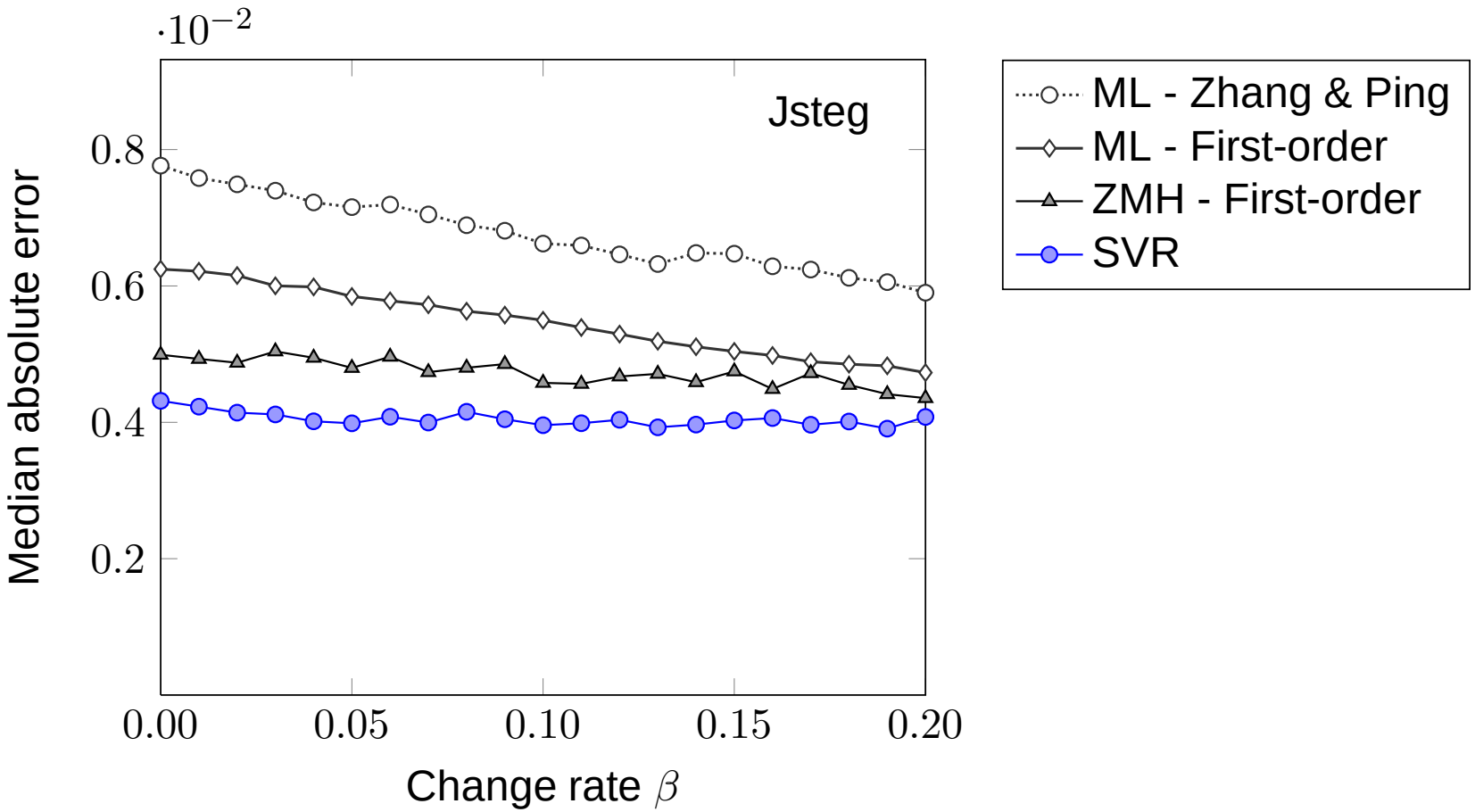
$$\mathbf{x} = [x_{-2L},\ x_{-2L+1}, \ldots, x_{2R-1},\ x_{2R}]$$



- Penalty function $z_{\mathrm{sym}}(\mathbf{x}) = \sum w_k (x_k - x_{-k})^2$
- Weights w_k chosen to minimize the estimator variance

  $\longrightarrow$ least squares steganalysis [Ker-2007]
- Final form of the penalty function: $$z_{\mathrm{sym}}(\mathbf{x}) = \sum_{k>0} \frac{(x_k - x_{-k})^2}{x_k + x_{-k}}$$

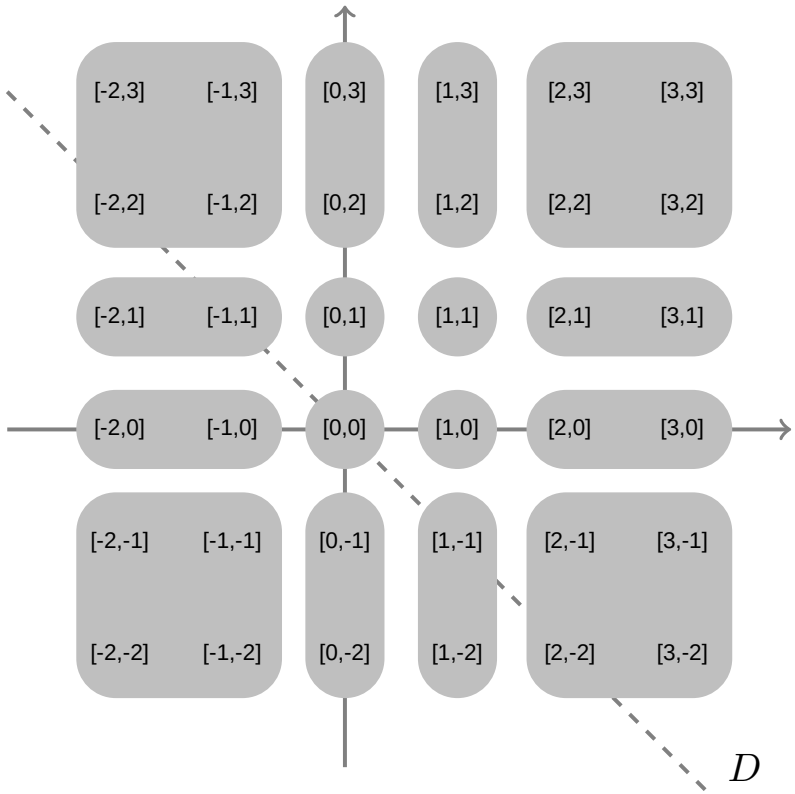
Performance Evaluation



Second-Order Statistics (ZMH)

- Feature vector: adjacency matrix $\mathbf{X}$
- ZMH approach
  - Decomposition into k-nodes
  - Embedding is invertible provided $0 \leq \beta < 1/2$
  - Symmetry about D

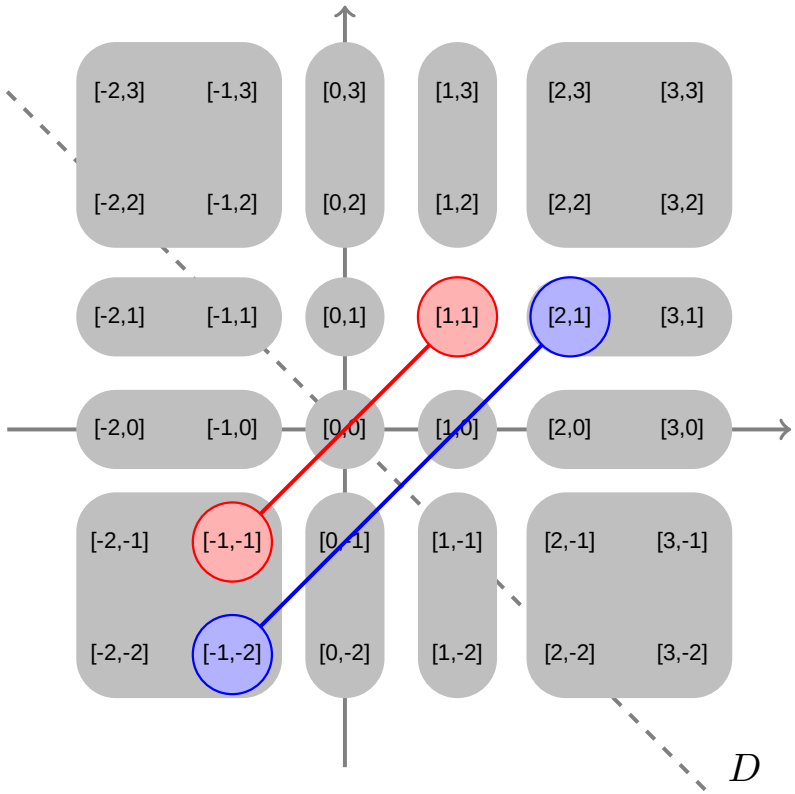
$$z_{\text{adj}}(\mathbf{X}) = \sum_{i,j} \frac{(x_{i,j} - x_{-j,-i})^2}{x_{i,j} + x_{-j,-i}}$$



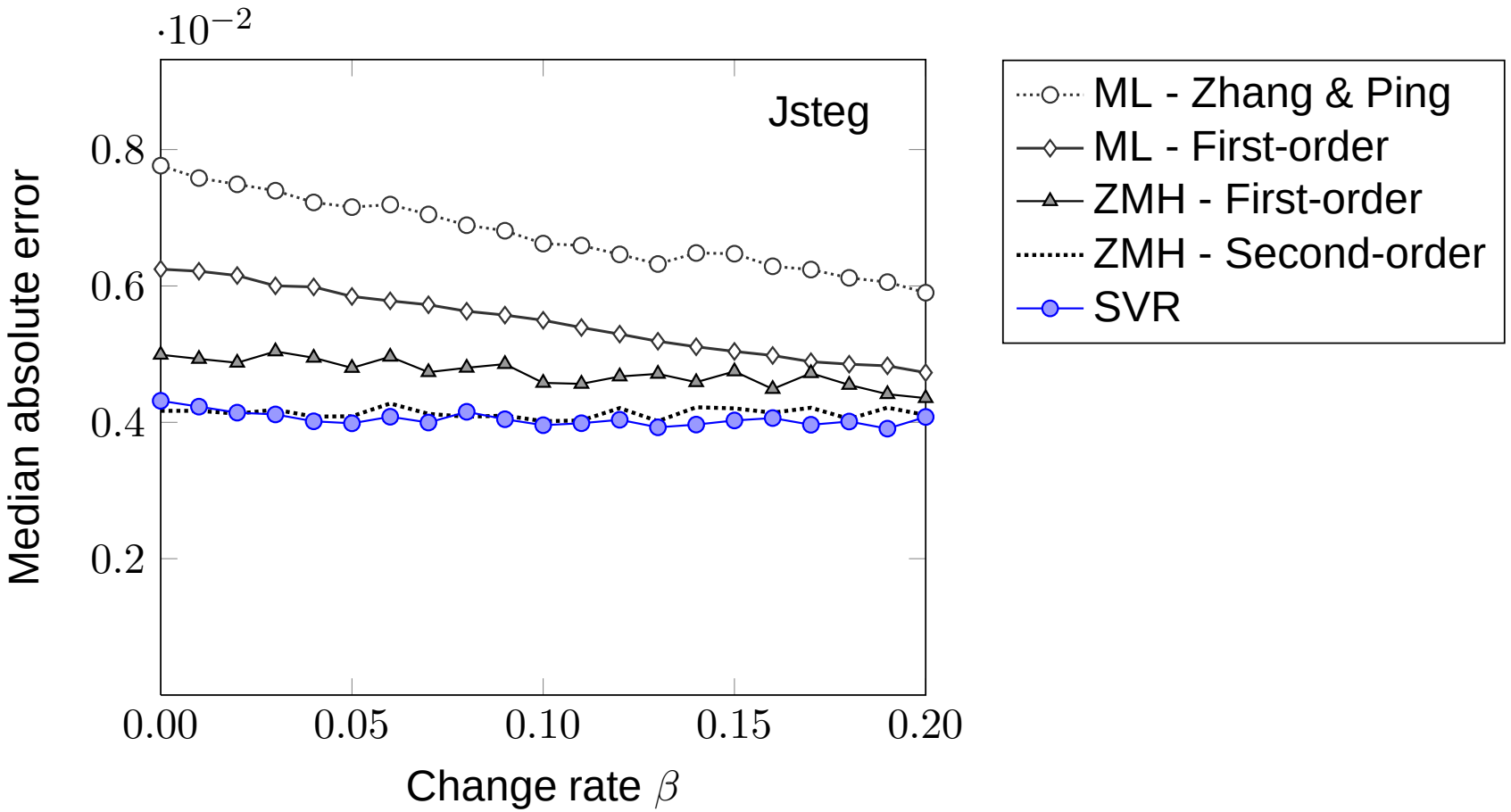
Second-Order Statistics (ZMH)

- Feature vector: adjacency matrix $\mathbf{X}$
- ZMH approach
  - Decomposition into k-nodes
  - Embedding is invertible provided $0 \leq \beta < 1/2$
  - Symmetry about D

$$z_{\text{adj}}(\mathbf{X}) = \sum_{i,j} \frac{(x_{i,j} - x_{-j,-i})^2}{x_{i,j} + x_{-j,-i}}$$



Performance Evaluation



What Else Can You Find in the Paper / Journal Version

- Error analysis of between-image and within-image errors for selected attacks
- Verification of precover assumptions using two different statistical tests
- Discussion & experiments with the symmetrized version of Jsteg
- Conversion of the Category attack [Lee-2006] into a quantitative one through the proposed ZMH framework
- Experiments conducted on two different sources of images
- Results reported in terms of two more security measures: IQR, median bias