Template Attacks on Different Devices

COSADE 2014

Omar Choudary and Markus G. Kuhn

Paris, 15 April 2014

UNIVERSITY OF CAMBRIDGE

Outline

- Template Attacks [Chari et al., CHES ’02]

Outline

- Template Attacks [Chari et al., CHES ’02]
- Problems when using different devices

Outline

- Template Attacks [Chari et al., CHES ’02]
- Problems when using different devices
- Extensive evaluation of TA on different devices
  - 4 devices and 5 acquisition campaigns
  - several compression methods
  - several methods to improve attack

Outline

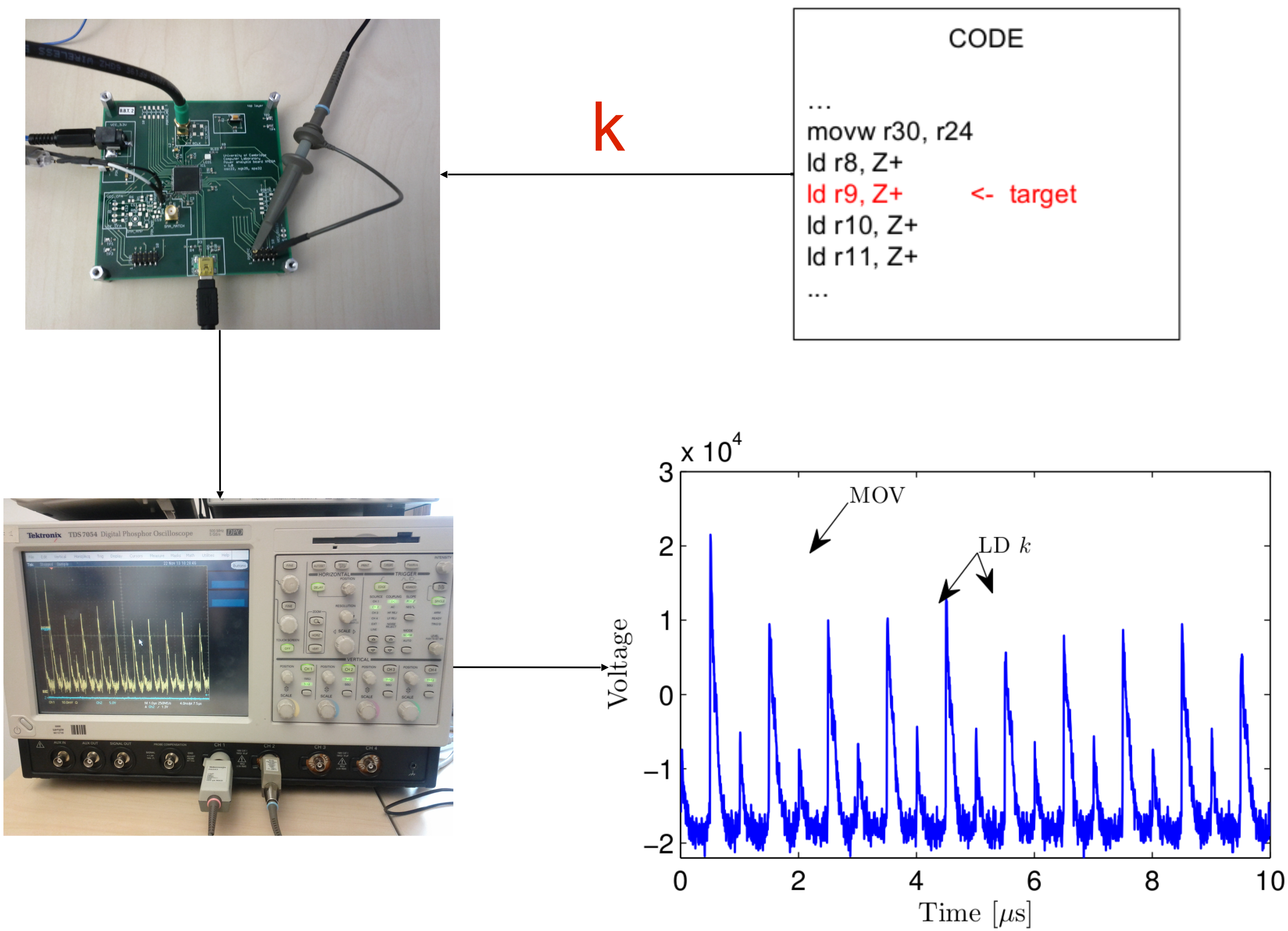
- Template Attacks [Chari et al., CHES ’02]
- Problems when using different devices
- Extensive evaluation of TA on different devices
  - 4 devices and 5 acquisition campaigns
  - several compression methods
  - several methods to improve attack
- PCA and LDA
  - Guideline for PCA/LDA to make it efficient
  - Method for improving PCA

Template Attacks on DPA contest v4

Participant	Submission date	Key found	Max PGE < 10	Key found (stable)	Max PGE stable < 10	Time/Trace (ms)	Attack type
Liran Lerman Université Libre de Bruxelles, Belgium	19/09/2013	22	13	22	13	24 ms	Profiling
Amir Moradi RUB, Germany	02/10/2013	174	148	174	148	305 ms	Non Profiling
Tang Ming Wuhan University, China	03/11/2013	763	465	990	482	271 ms	Non Profiling
Frank Schuhmacher Segrids, Germany	26/02/2014	1	1	1	1	5 ms	Profiling
Hideo Shimizu Toshiba Corporation Corporate Research & Development Center, Japan	28/02/2014	1	1	1	1	30 ms	Profiling
Xavier Bodart, Liran Lerman Université Libre de Bruxelles, Belgique	06/03/2014	21	17	21	17	400 ms	Profiling

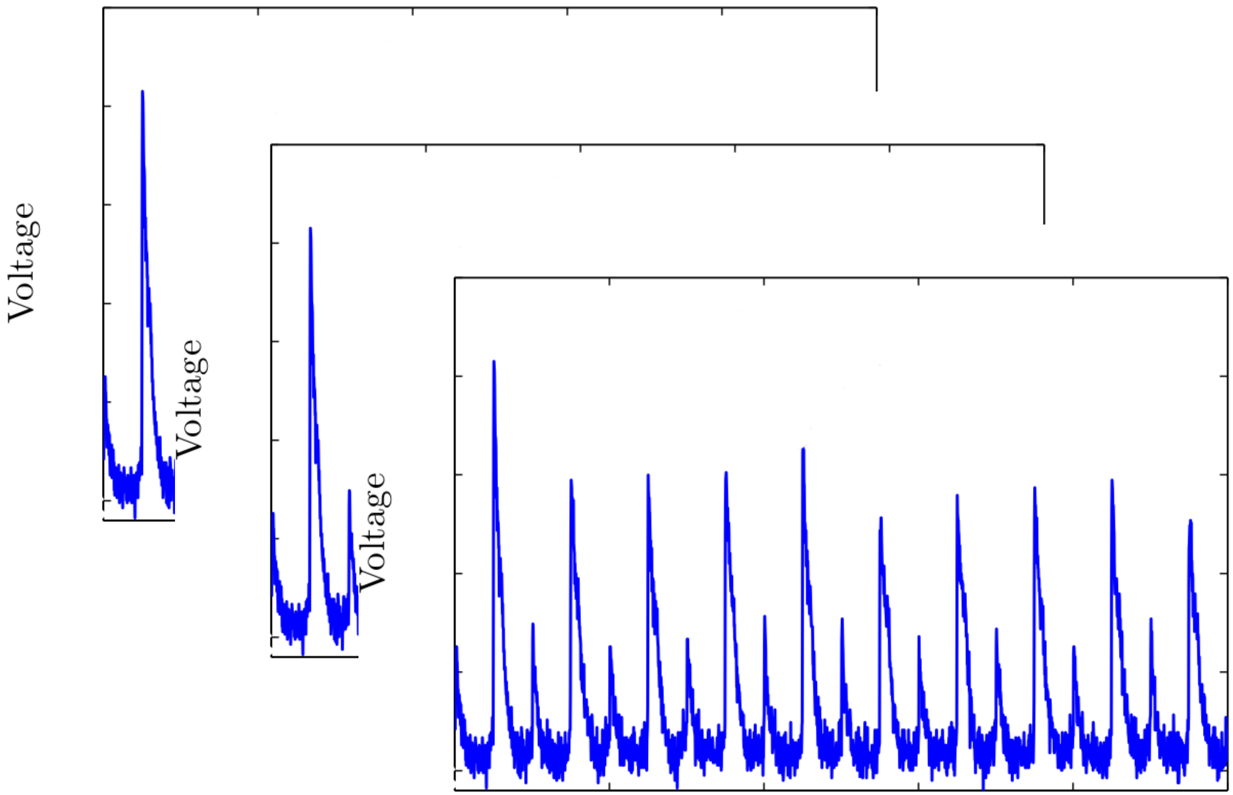
- **Key found**: Number of traces needed to find the correct key
- **Max PGE < 10**: Number of traces for the maximum Partial Guessing Entropy to be below 10
- **Key found (stable)**: Number of traces needed to find the correct key for good
- **Max PGE stable < 10**: Number of traces for the maximum Partial Guessing Entropy to be stable below 10
- **Time/Trace**: Mean time per trace

Template Attacks – Setup



Template Attacks on Different Devices

Template Attacks – Profiling

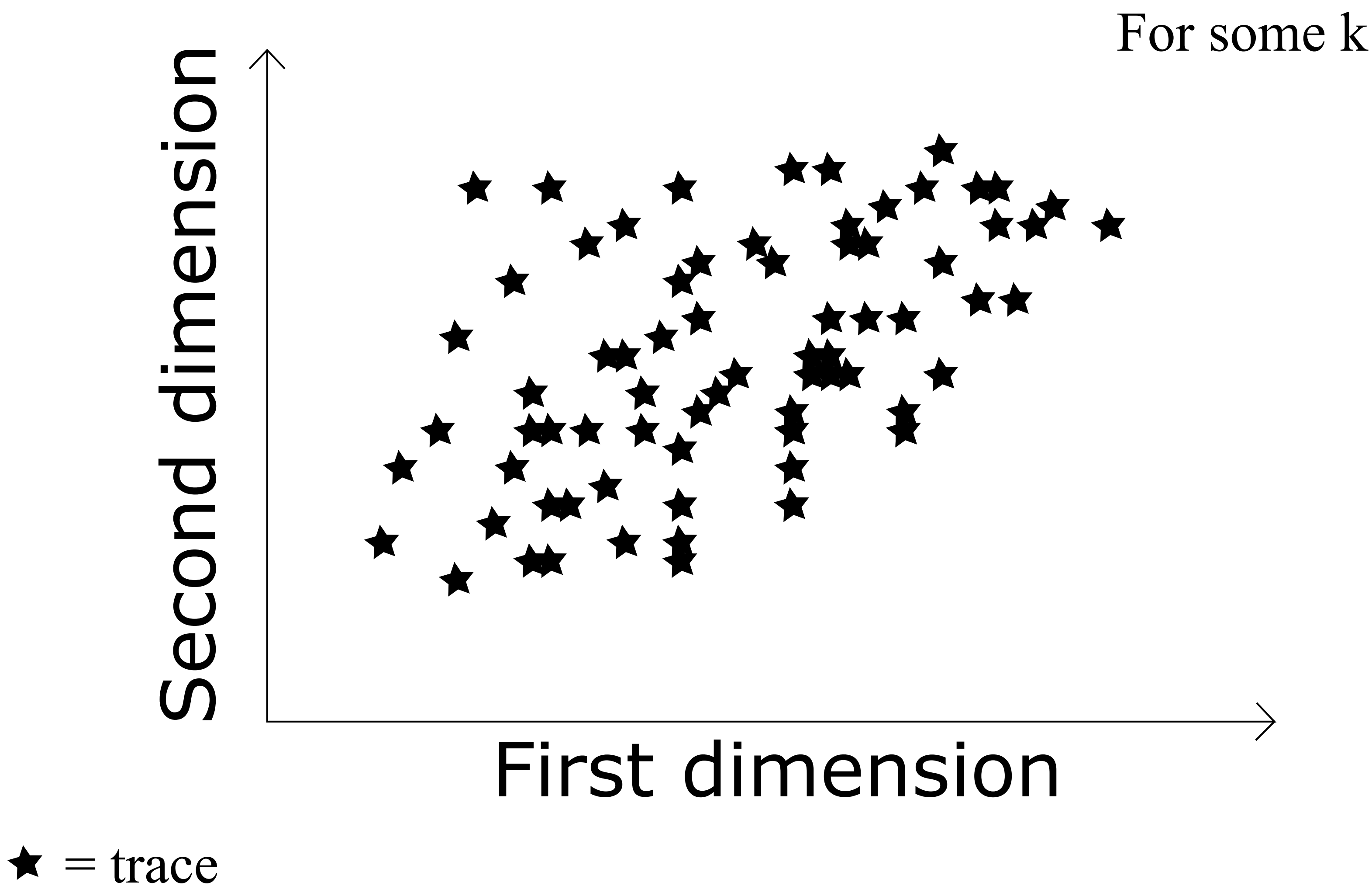


k = 0, 1, 2, …, 255

$n_{\mathrm{p}} = 1000$ profiling traces per k

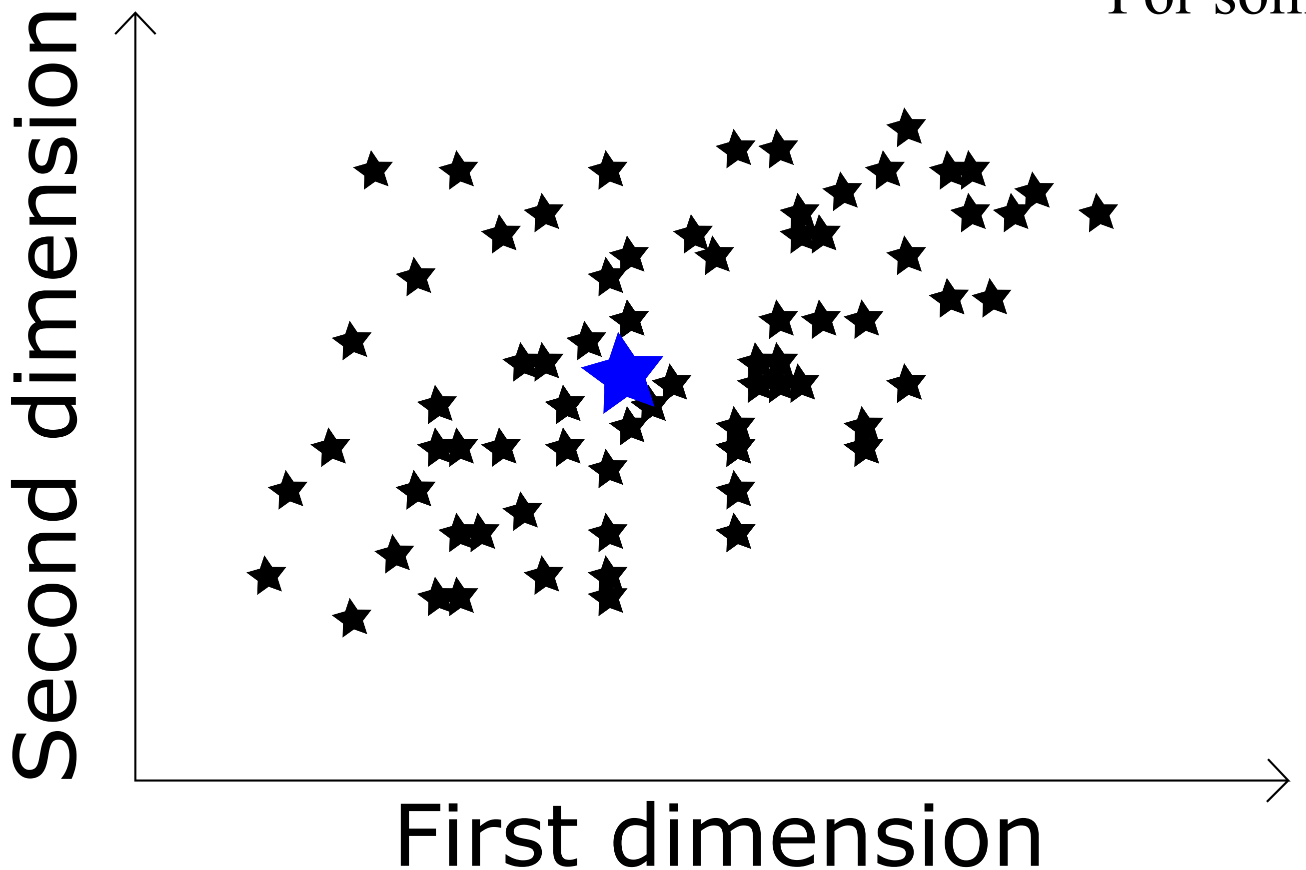
$m = 2500$ samples per trace

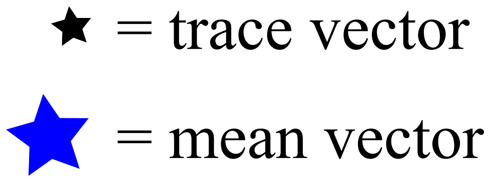
Data space – cloud of traces



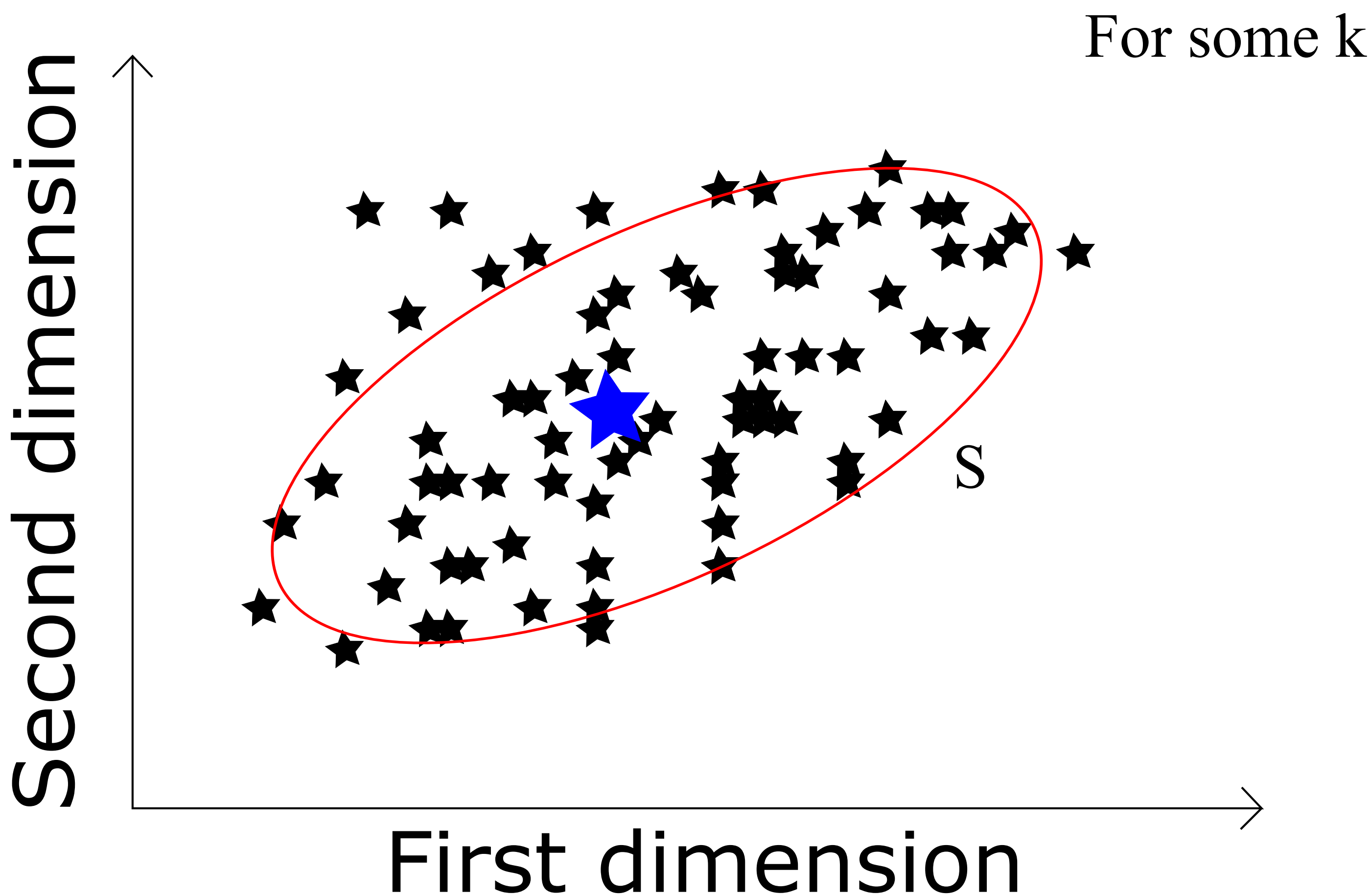
Data space – mean vector

For some k





Data space – covariance matrix

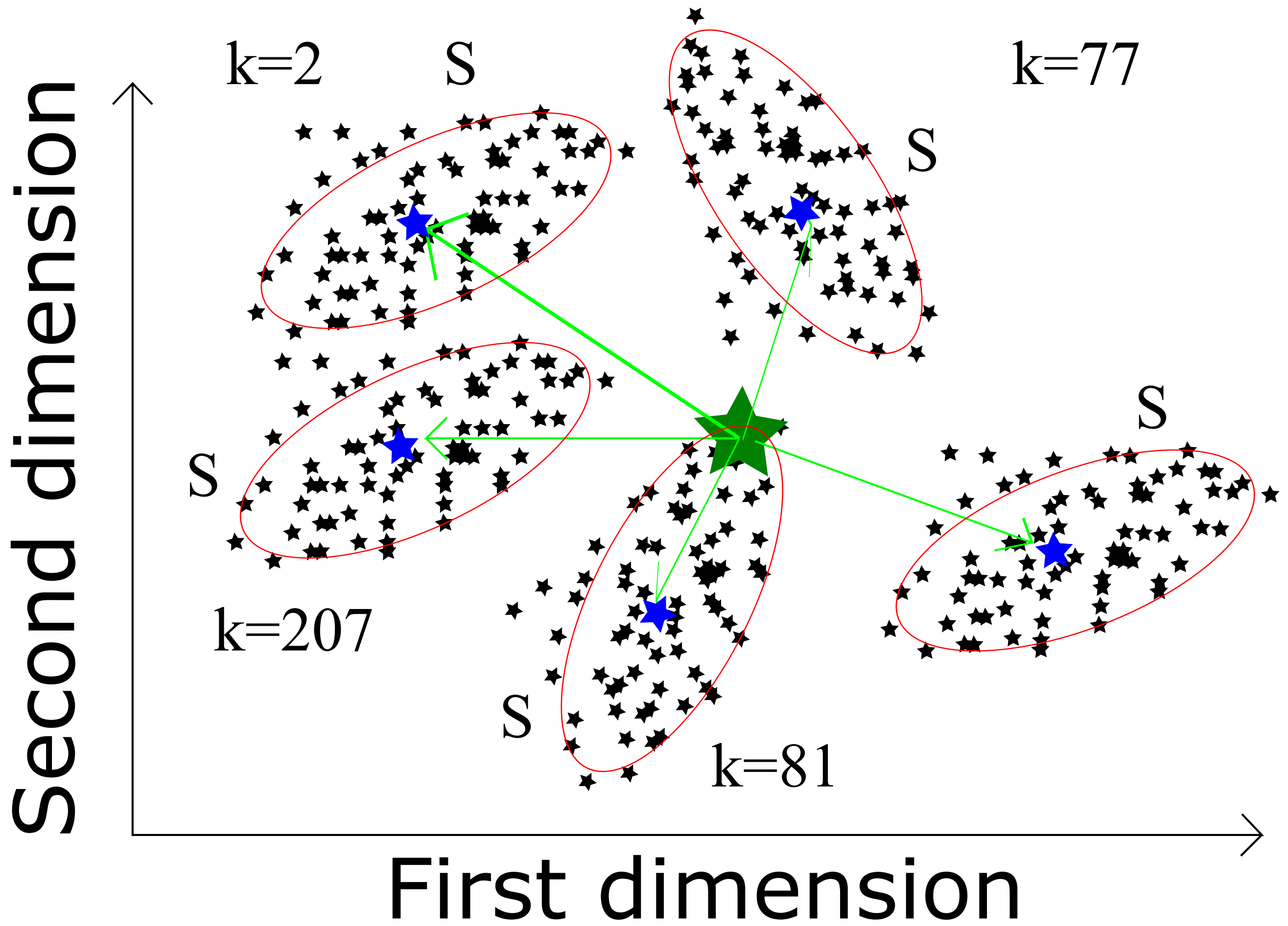


★ = trace vector

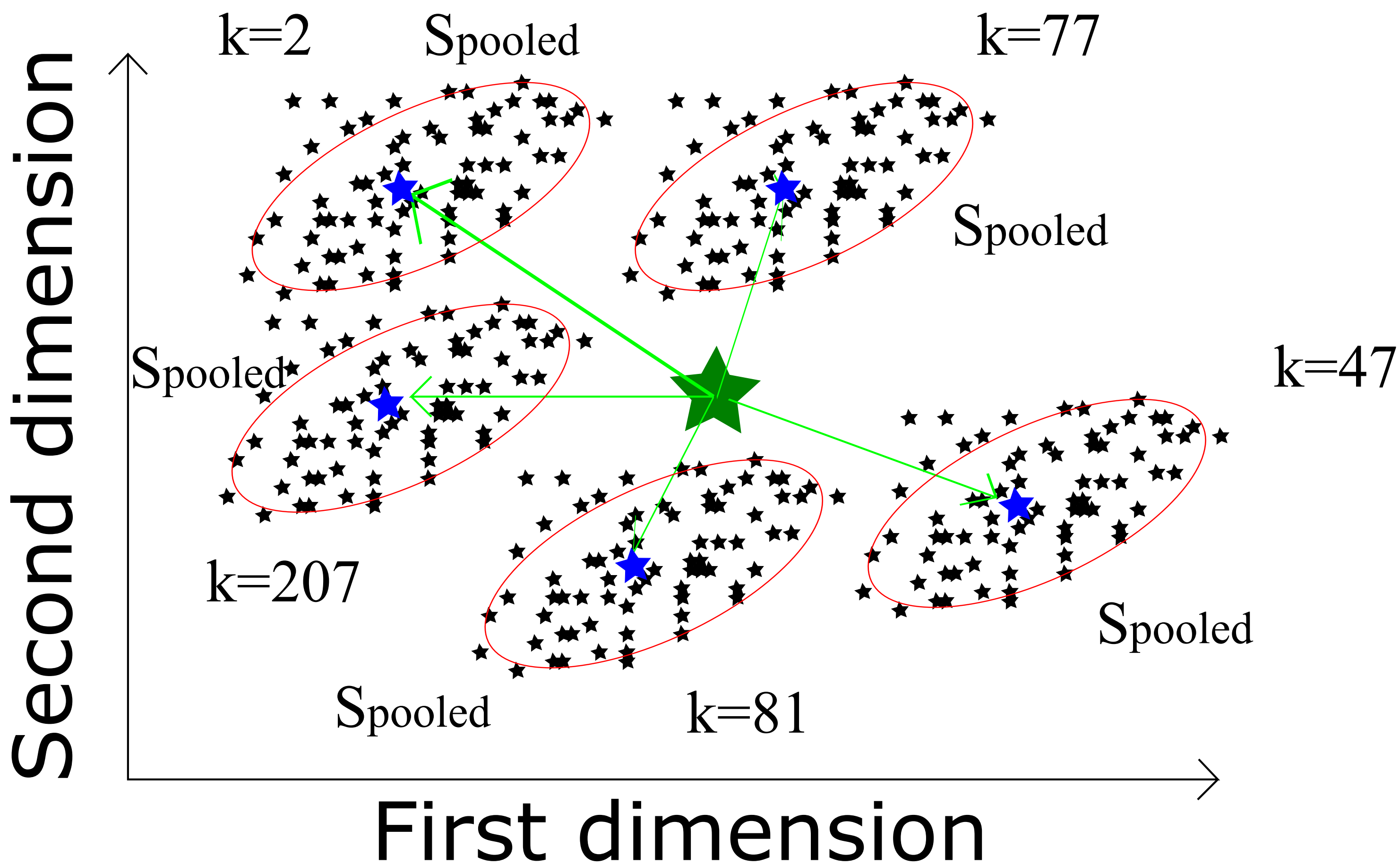
= covariance

★ = mean vector

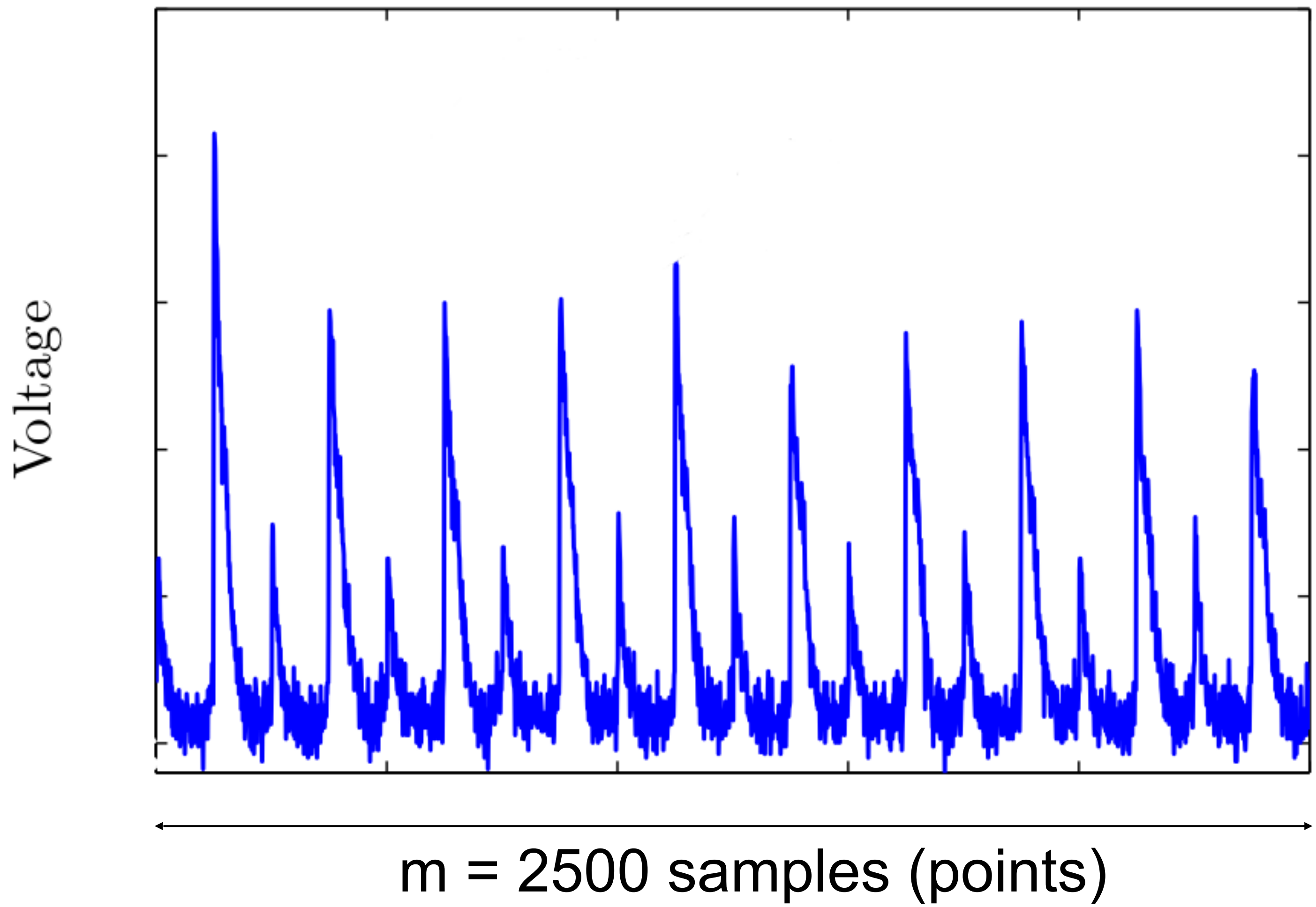
Data space – individual covariances



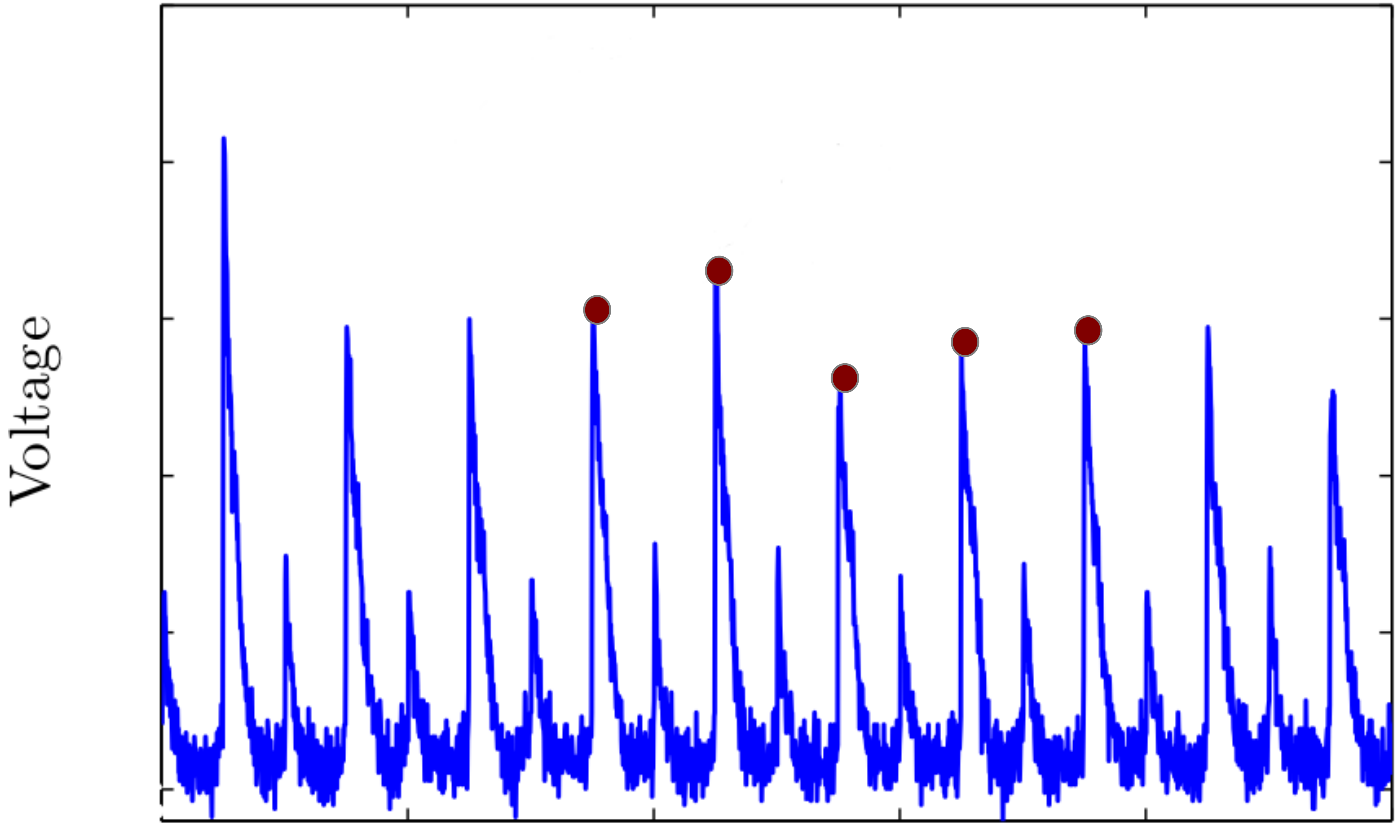
Data space – pooled covariance



Template Attacks – Compression

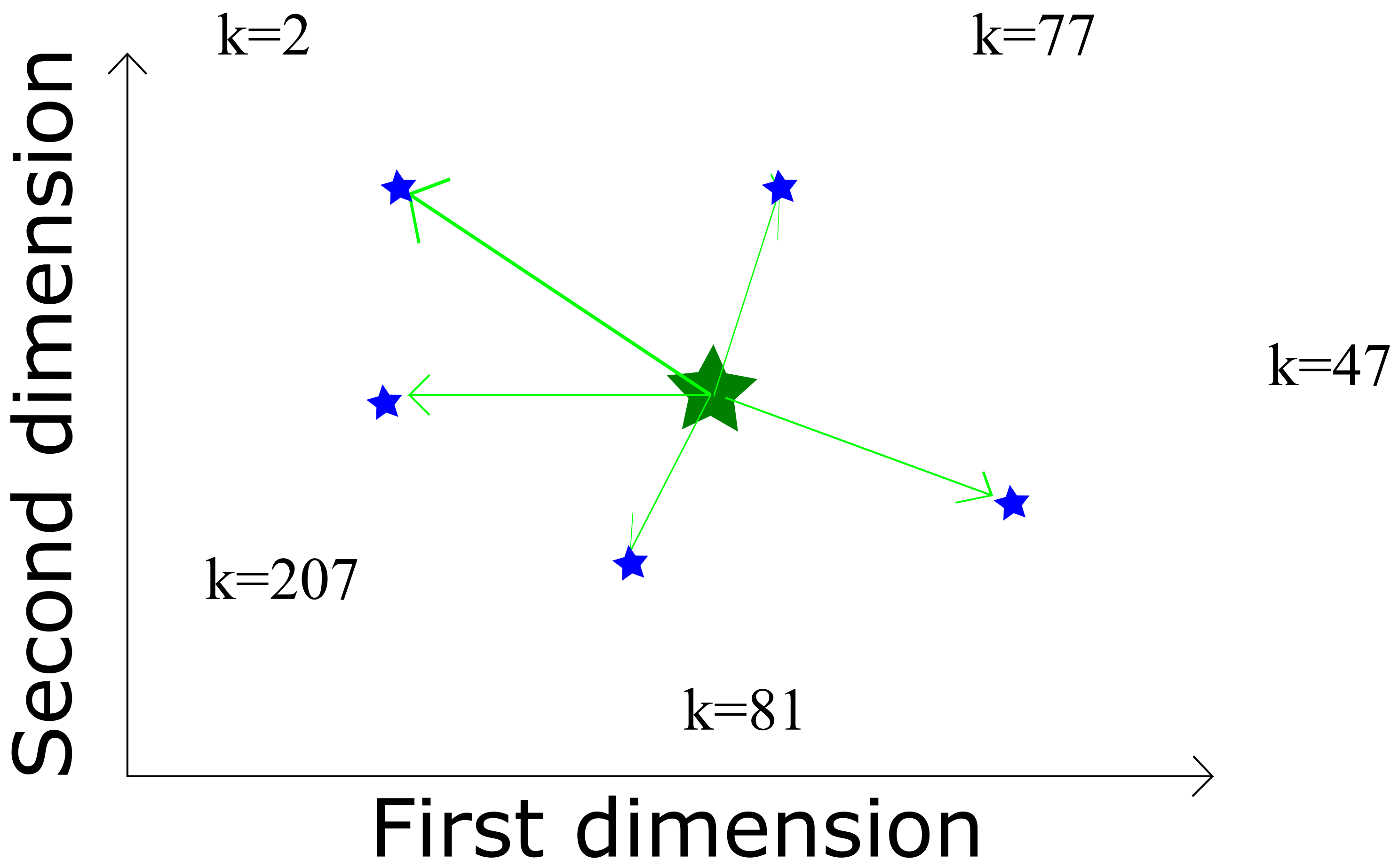


Select samples

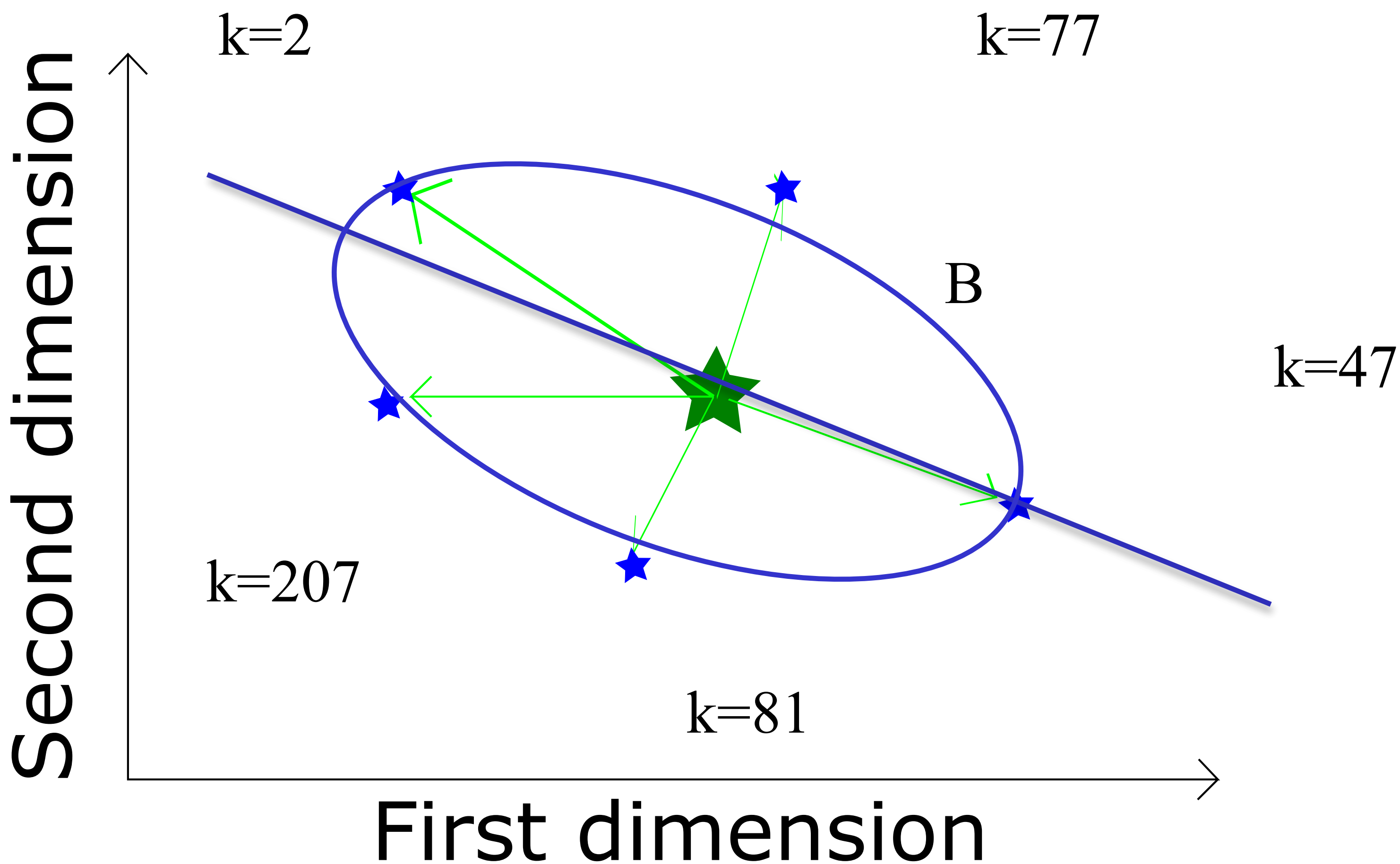


$m = 5$

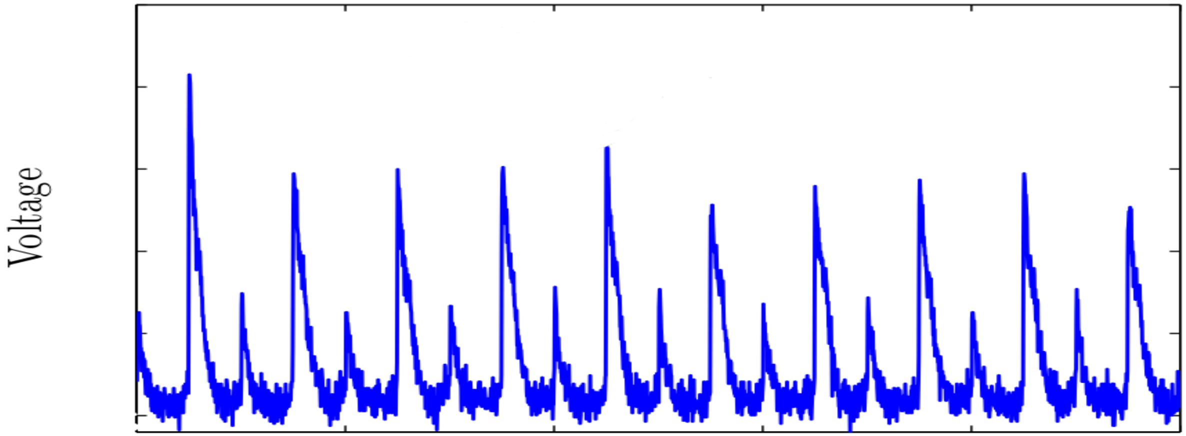
Principal Component Analysis (PCA)



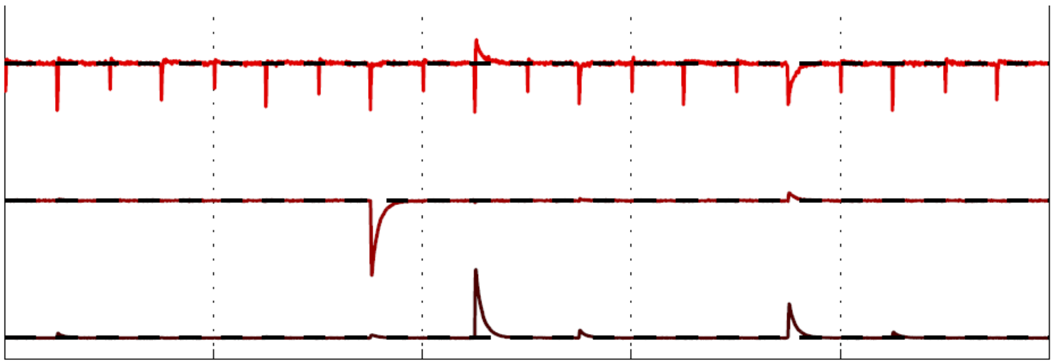
Principal Component Analysis (PCA)



Principal Component Analysis (PCA)

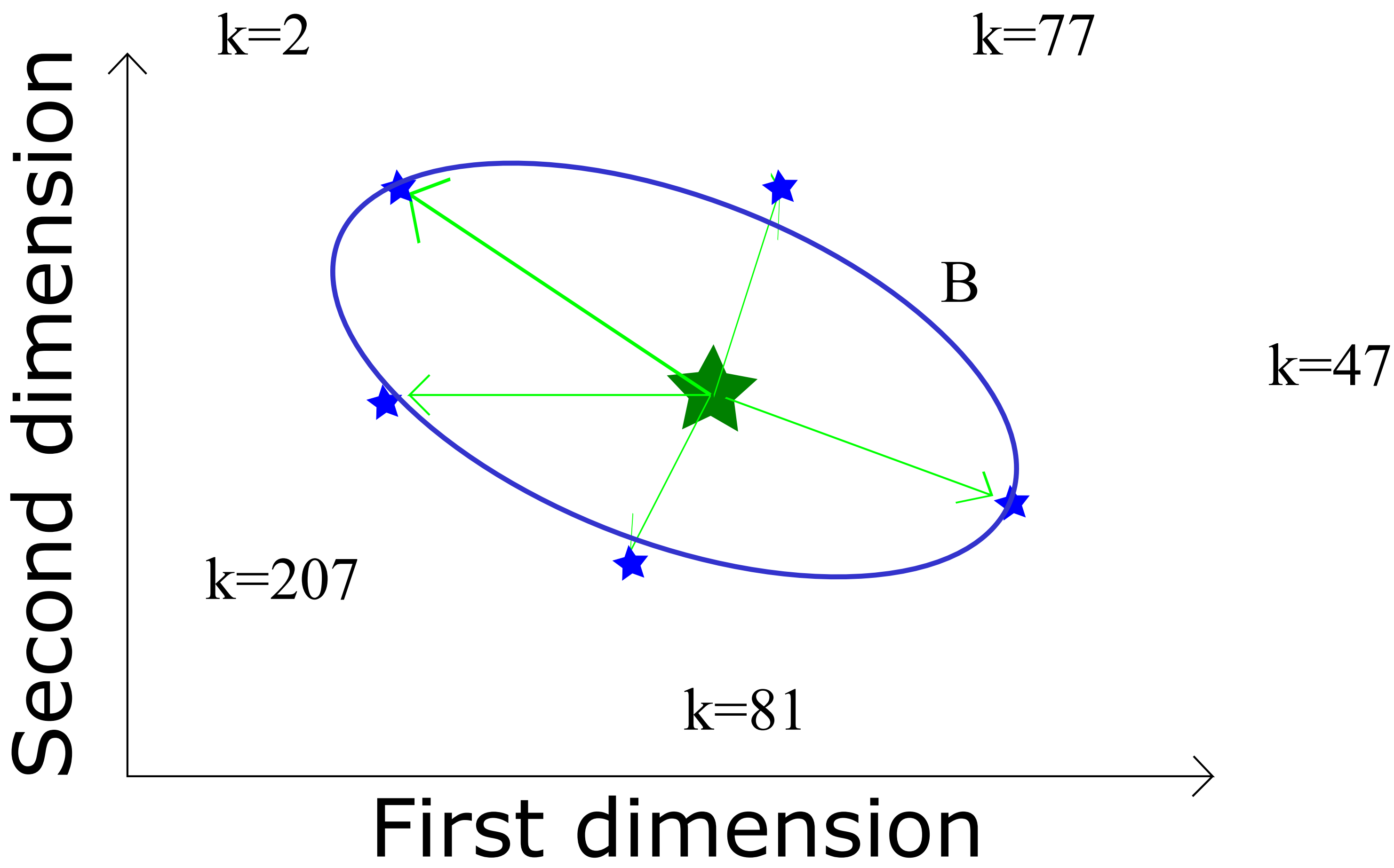


$m = 3$

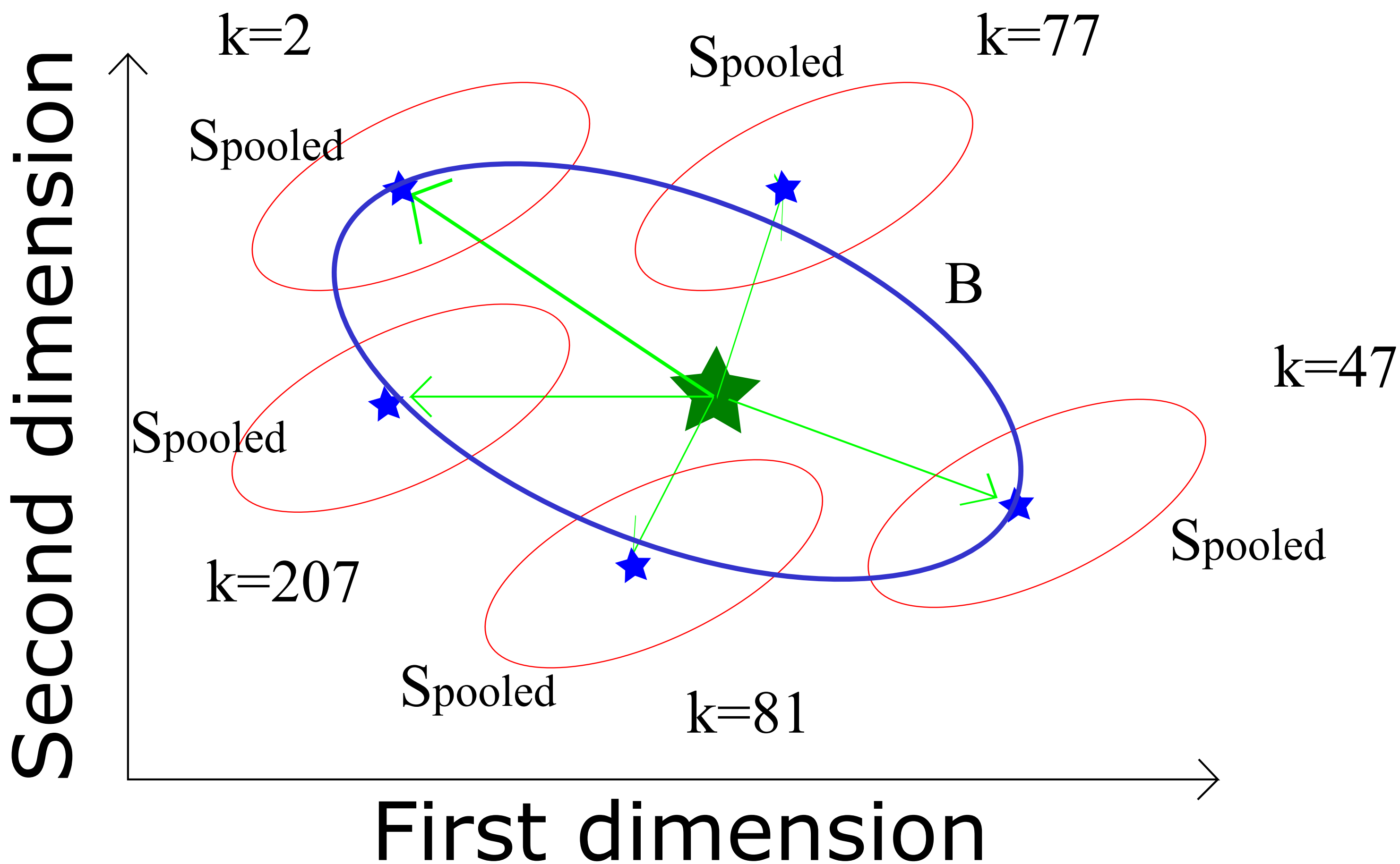


U=SVD(B)

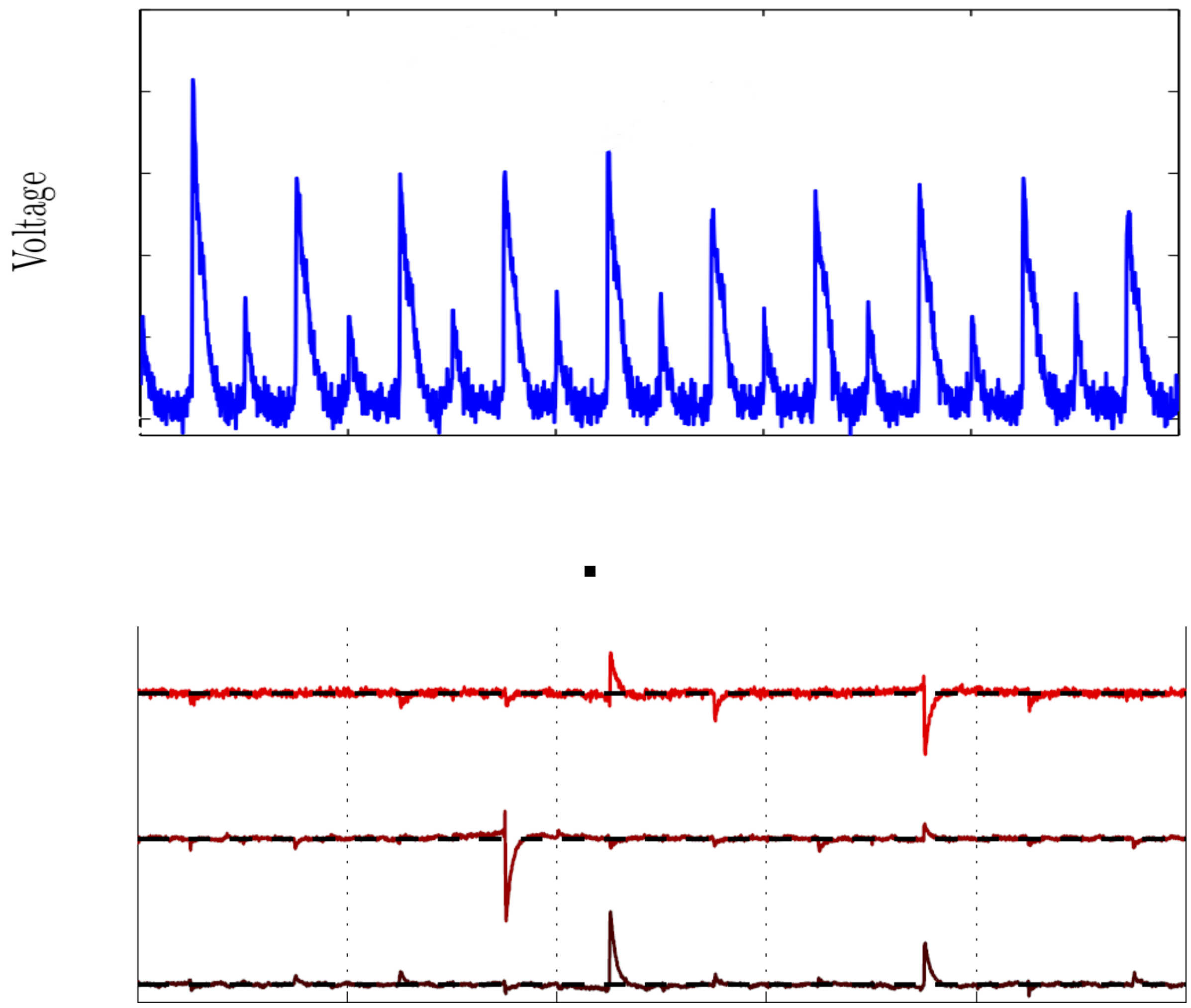
Linear Discriminant Analysis (LDA)



Linear Discriminant Analysis (LDA)



Linear Discriminant Analysis (LDA)

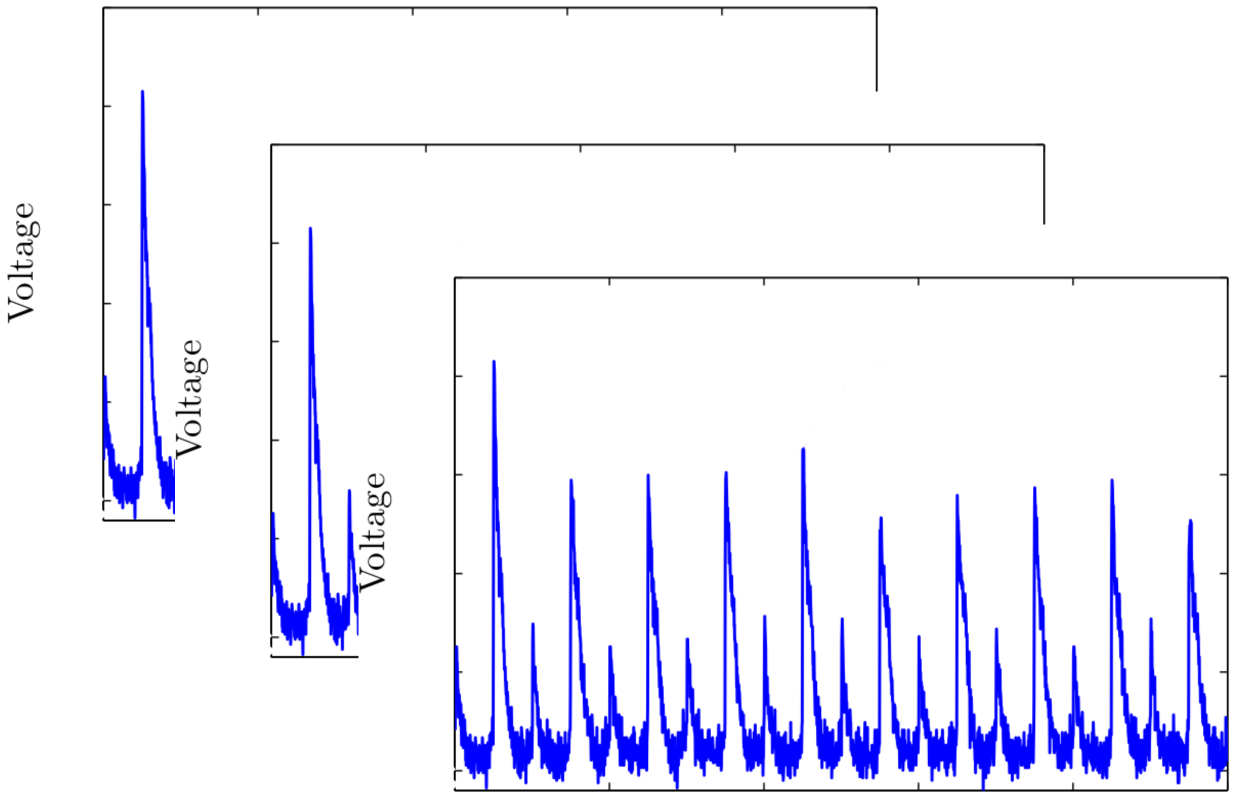


$m = 3$

U=SVD(B/S)

Template Attacks on Different Devices

Template Attacks – Attack



k = ???

$1 \leqslant n_a \leqslant 1000$

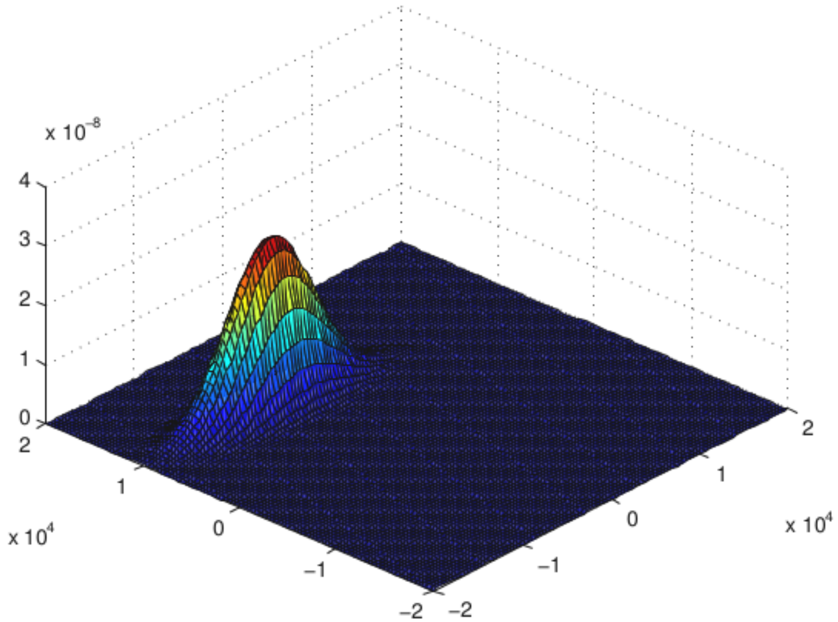
Template Attacks – Attack

k = 0, 1, 2, …, 255

Option 1: Multivariate Gaussian Distribution
[Chari et al., CHES '02]

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_\mathrm{a}}\}$$

$$\mathrm{d}(k \mid \mathbf{X}) = \prod_{\mathbf{x} \in \mathbf{X}} \frac{1}{\sqrt{(2\pi)^m |\mathbf{S}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_k)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_k)\right)$$



$$k^\star = \arg\max_k \mathrm{d}(k \mid \mathbf{X})$$

Template Attacks – Attack

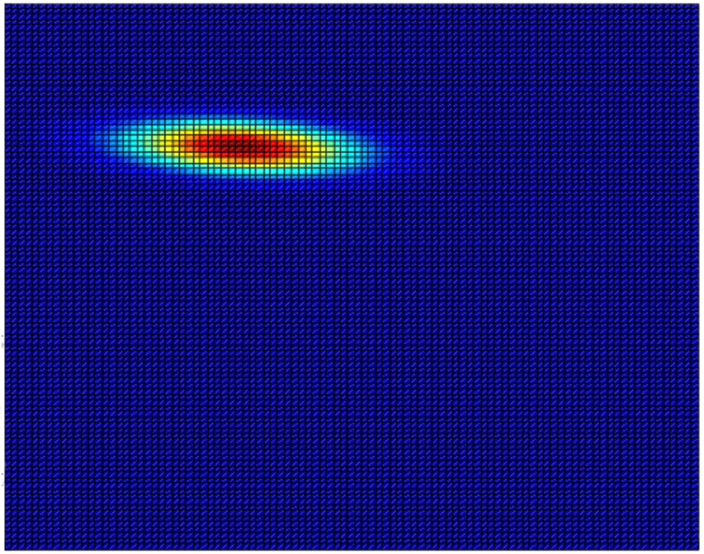
k = 0, 1, 2, …, 255

Option 2: Mahalanobis Distance or Linear Discriminant [Choudary and Kuhn, CARDIS '13]

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_\mathrm{a}}\}$$

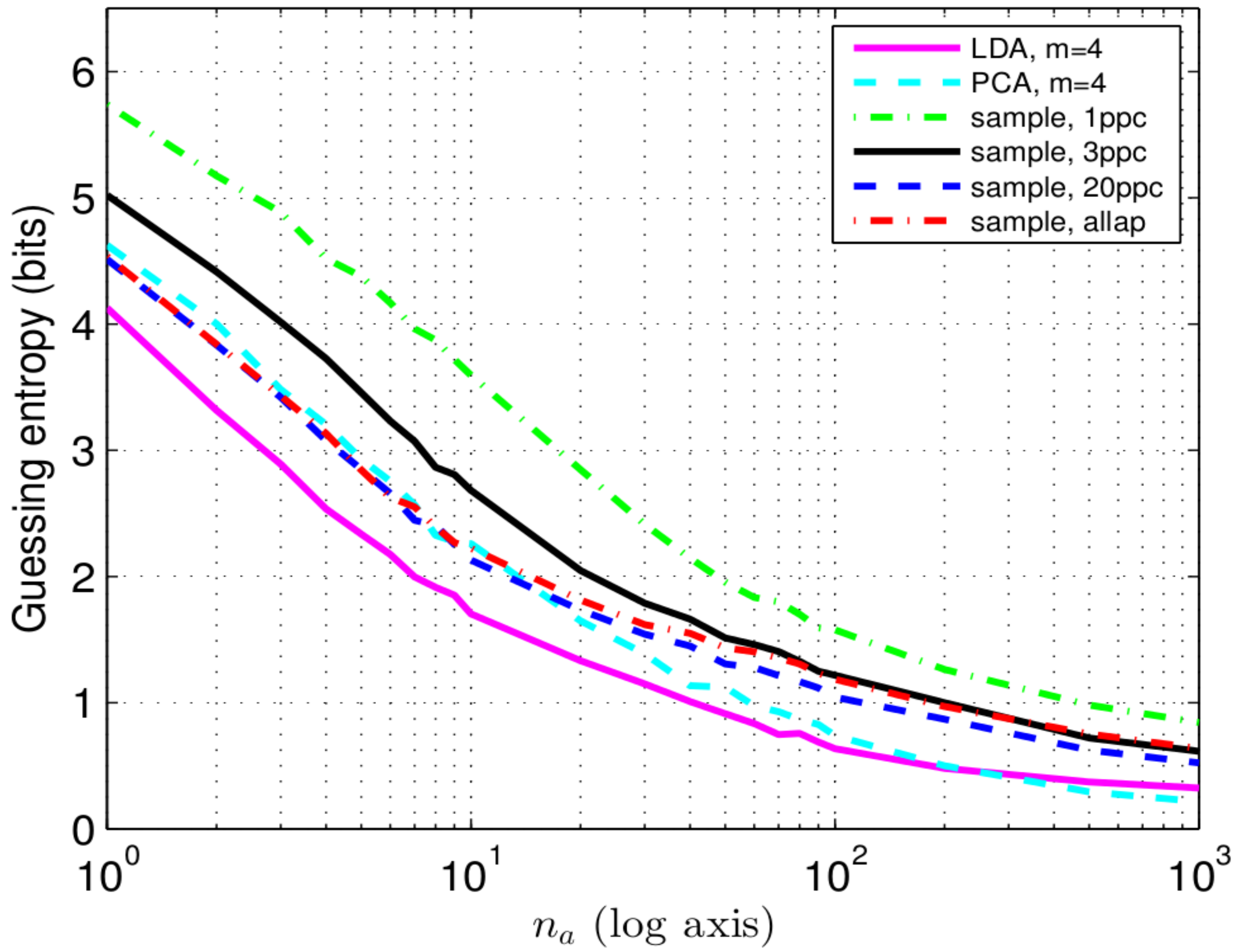
$$\mathrm{d}_{\mathrm{MD}}(k \mid \mathbf{X}) = -\frac{1}{2} \sum_{\mathbf{x} \in \mathbf{X}} (\mathbf{x} - \bar{\mathbf{x}}_k)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_k)$$

$$\mathrm{d}_{\mathrm{Linear}}(k \mid \mathbf{X}) = \bar{\mathbf{x}}_k' \mathbf{S}^{-1} \left(\sum_{\mathbf{x} \in \mathbf{X}_{k\star}} \mathbf{x} \right) - \frac{n_\mathrm{a}}{2} \bar{\mathbf{x}}_k' \mathbf{S}^{-1} \bar{\mathbf{x}}_k$$



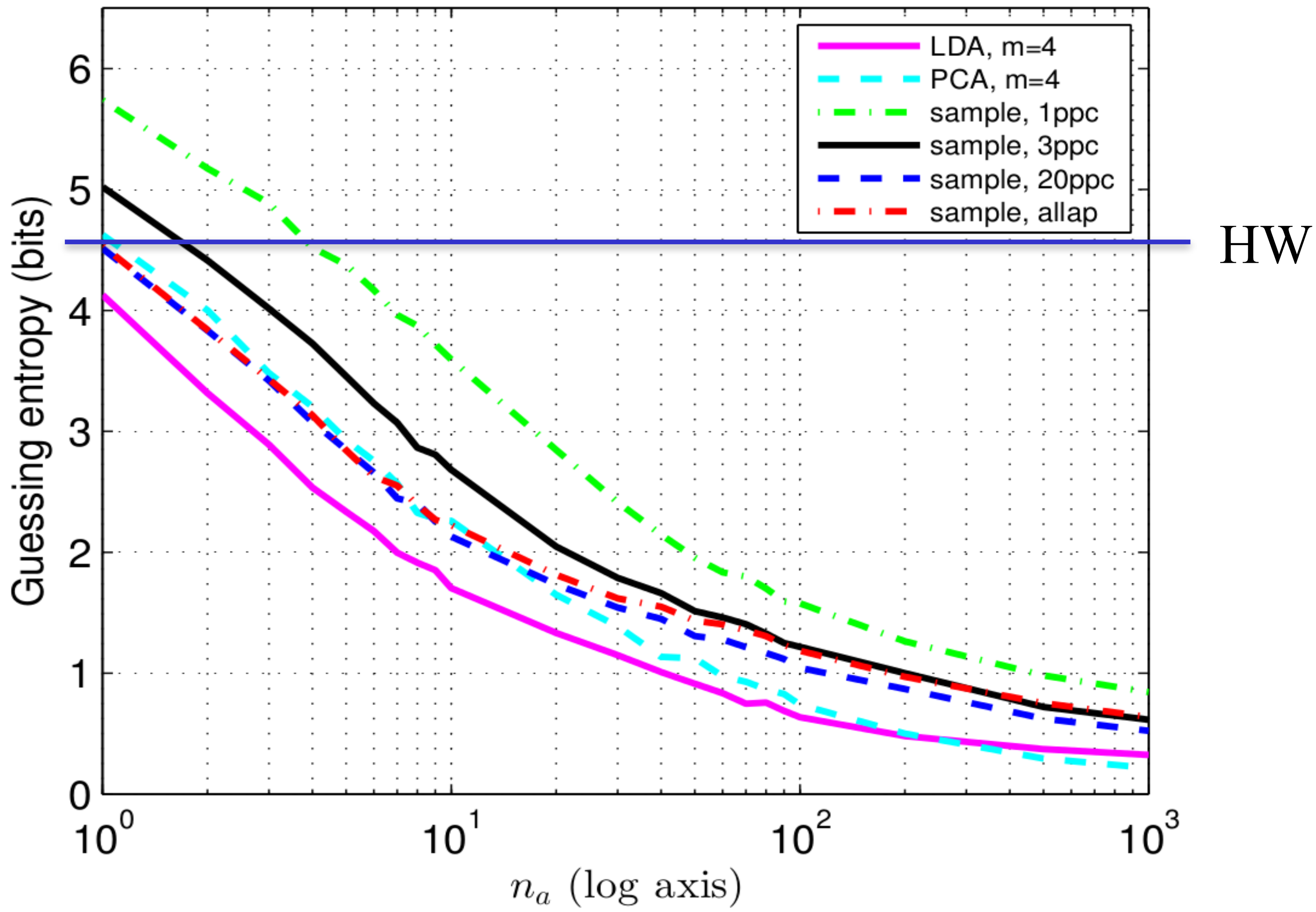
$$k^\star = \arg\max_k \mathrm{d}(k \mid \mathbf{X})$$

TA on same campaign [CARDIS '13]



Template Attacks on Different Devices

TA on same campaign [CARDIS '13]



Template Attacks on Different Devices

Using different devices in template attack during profiling versus attack phase

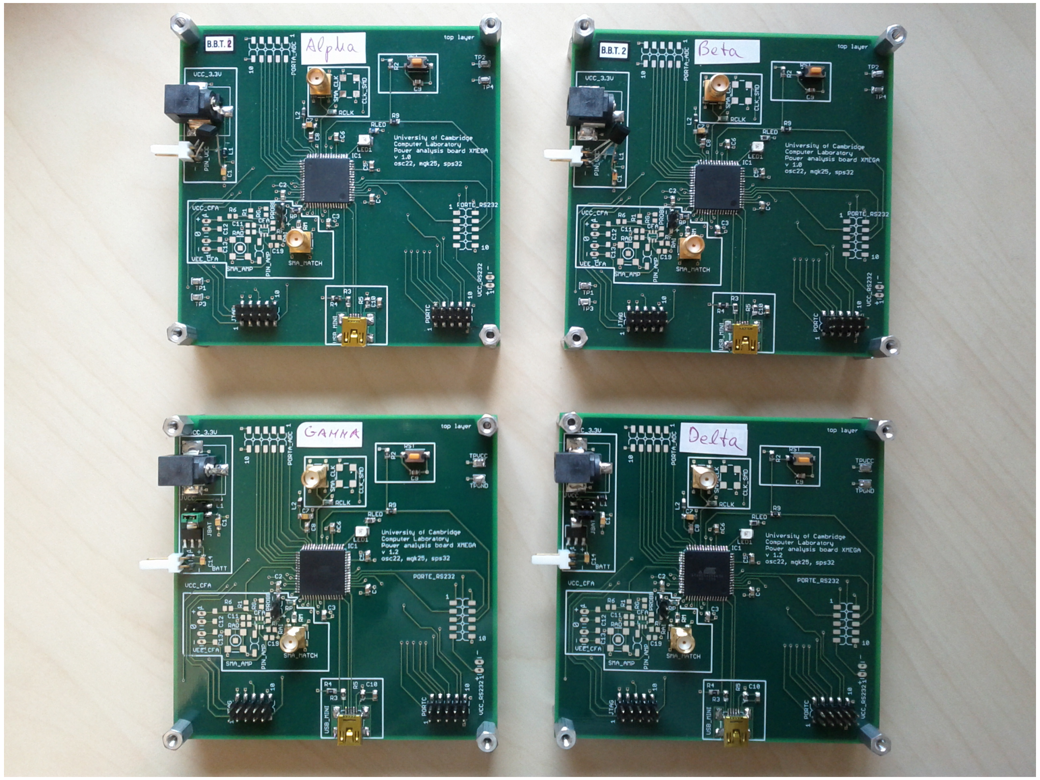
- [Renauld et al., Eurocrypt ’11]
  - Bad results across different ASIC devices
  - Used 20 different devices
  - Sample selection with 1 to 3 samples
- [Elaabid et al., Journal Crypto Engineering ’12]
  - Bad results on same device but different campaigns
  - PCA with 1 principal component

Our evaluation

- 4 different devices (Atmel XMEGA 8-bit µC)

Alpha

Beta



Gamma

Delta

Our evaluation

- 4 different devices (Atmel XMEGA 8-bit µC)
- Code same as our CARDIS ’13 scenario

CODE

```
…
movw r30, r24
ld r8, Z+
ld r9, Z+        <-  target
ld r10, Z+
ld r11, Z+
…
```

Our evaluation

- 4 different devices (Atmel XMEGA 8-bit µC)
- Code same as our CARDIS ’13 scenario
- 5 acquisition campaigns
  - 1 per device
  - 1 additional campaign on one device

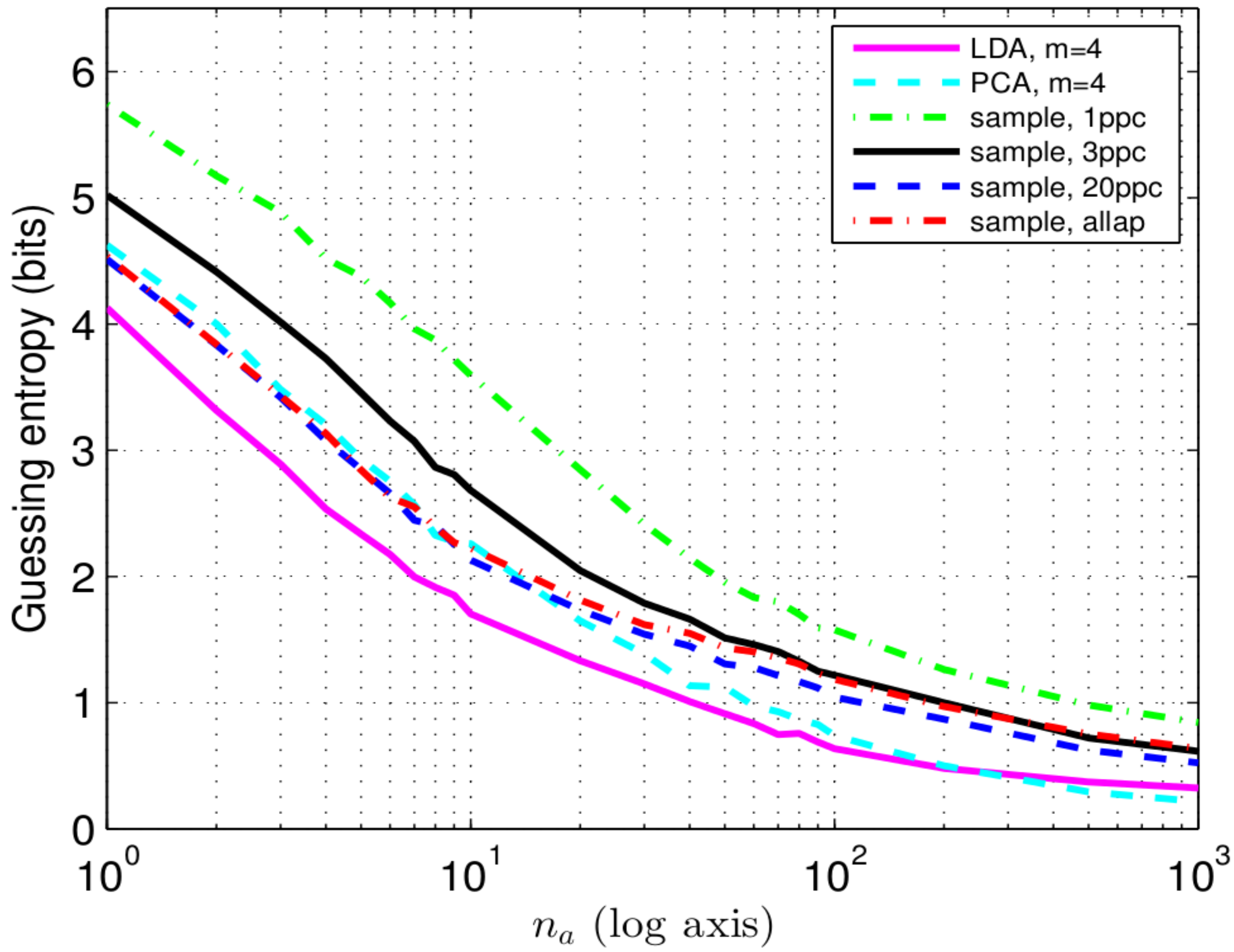
Our evaluation

- 4 different devices (Atmel XMEGA 8-bit µC)
- Code same as our CARDIS ’13 scenario
- 5 acquisition campaigns
  - 1 per device
  - 1 additional campaign on one device
- Several compressions with different params

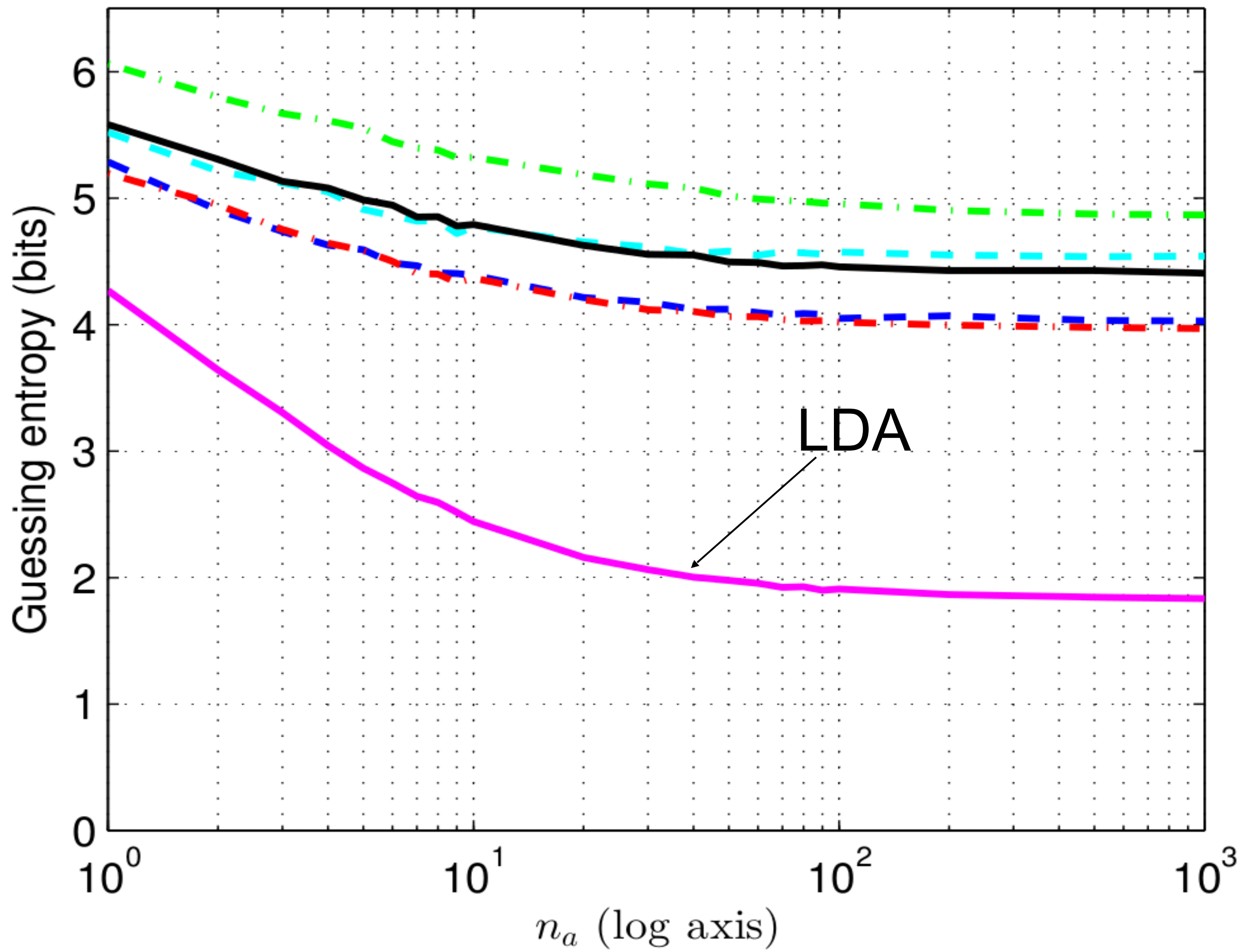
Our evaluation

- 4 different devices (Atmel XMEGA 8-bit µC)
- Code same as our CARDIS ’13 scenario
- 5 acquisition campaigns
  - 1 per device
  - 1 additional campaign on one device
- Several compressions with different params
- Several methods to improve TA

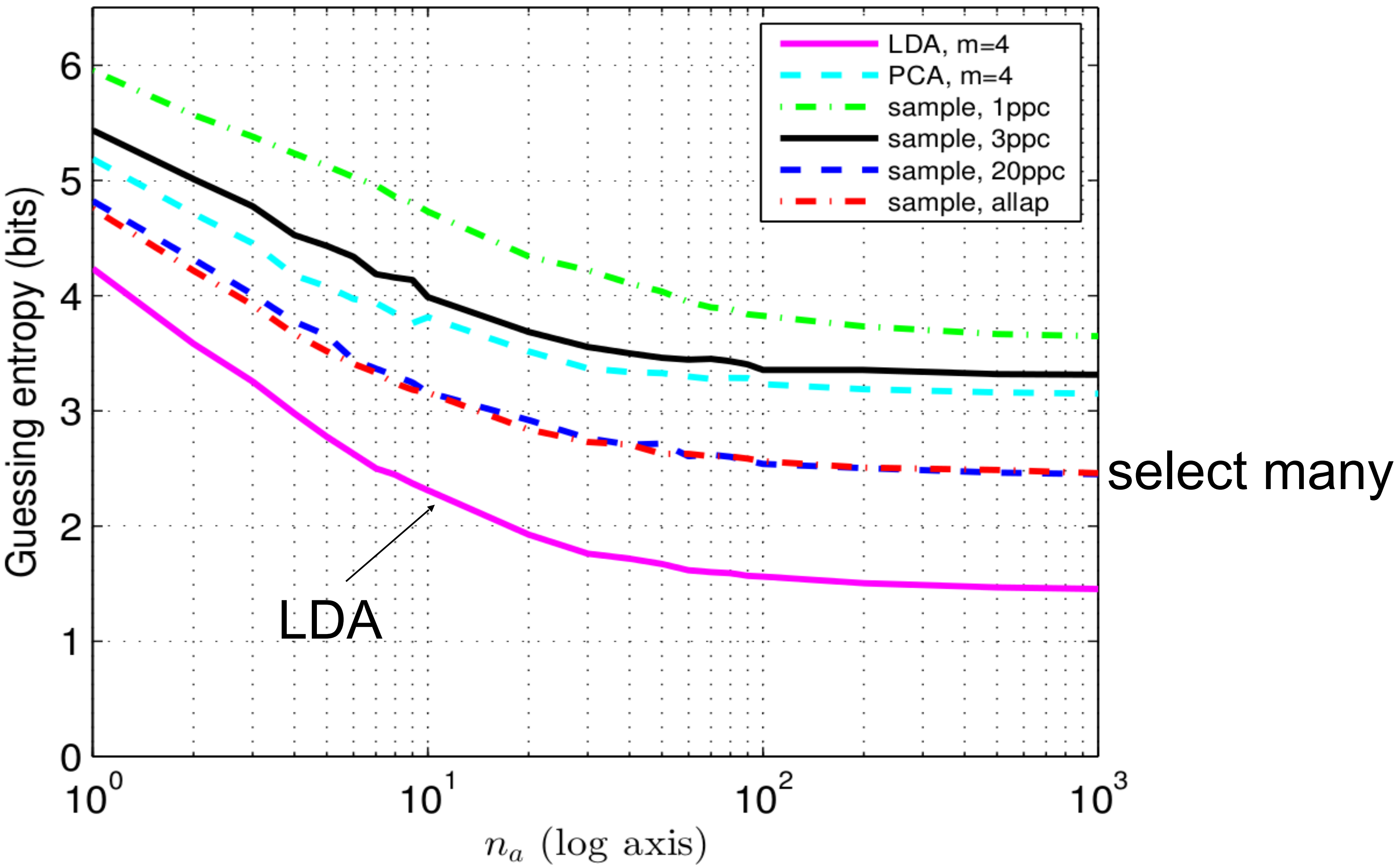
Standard TA (Meth. 1) same device



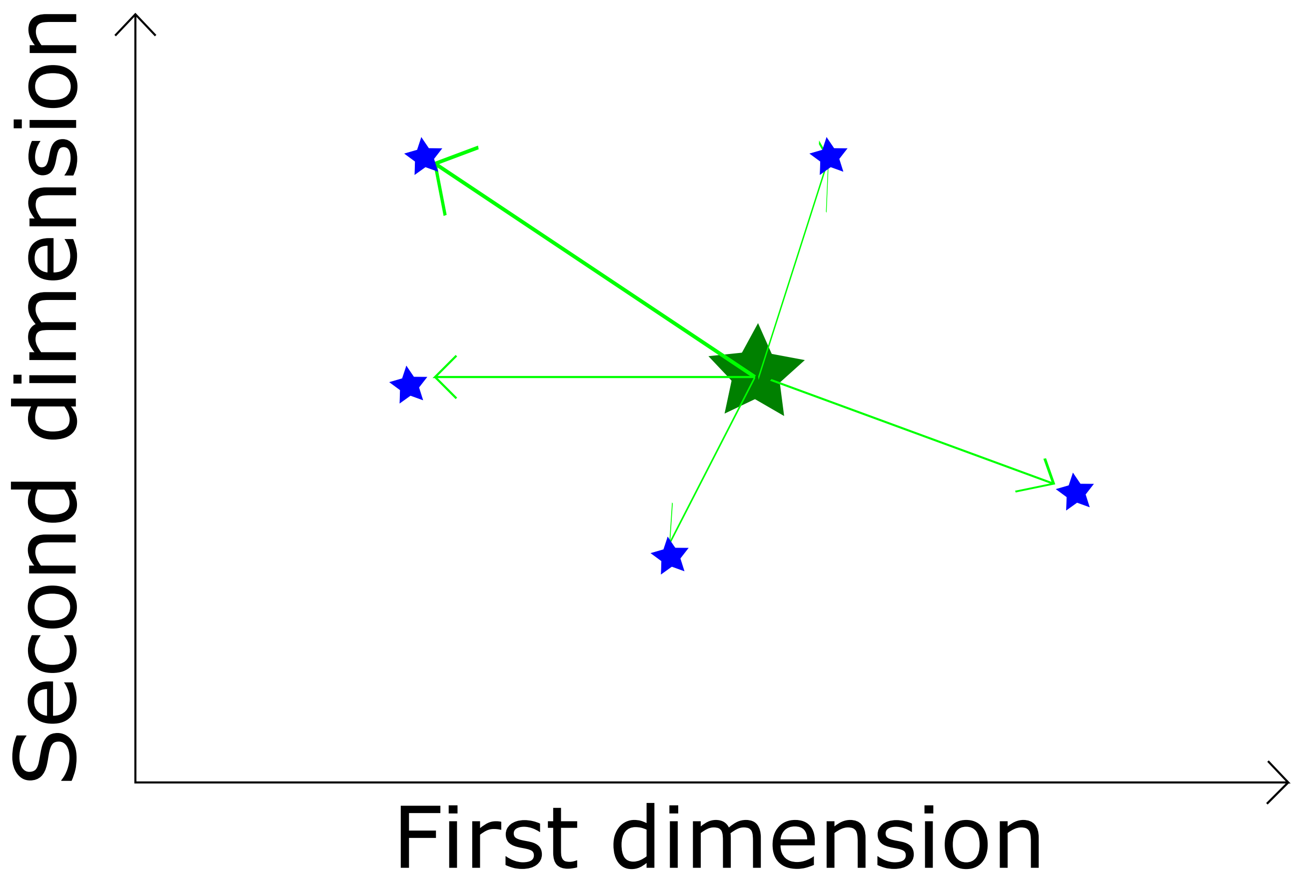
Standard TA (Meth. 1) different devices



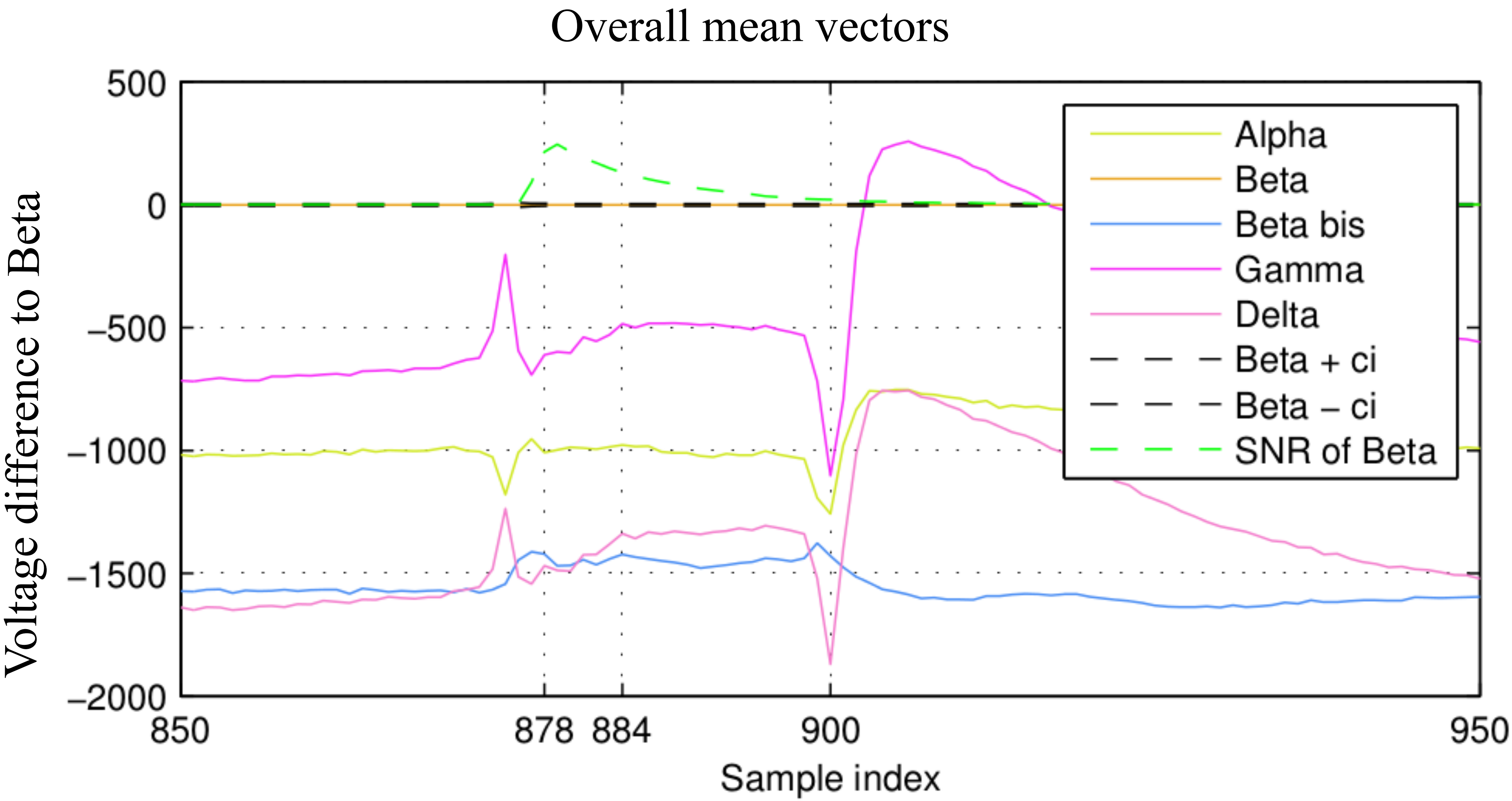
Profiling on 3 devices (Meth. 2)



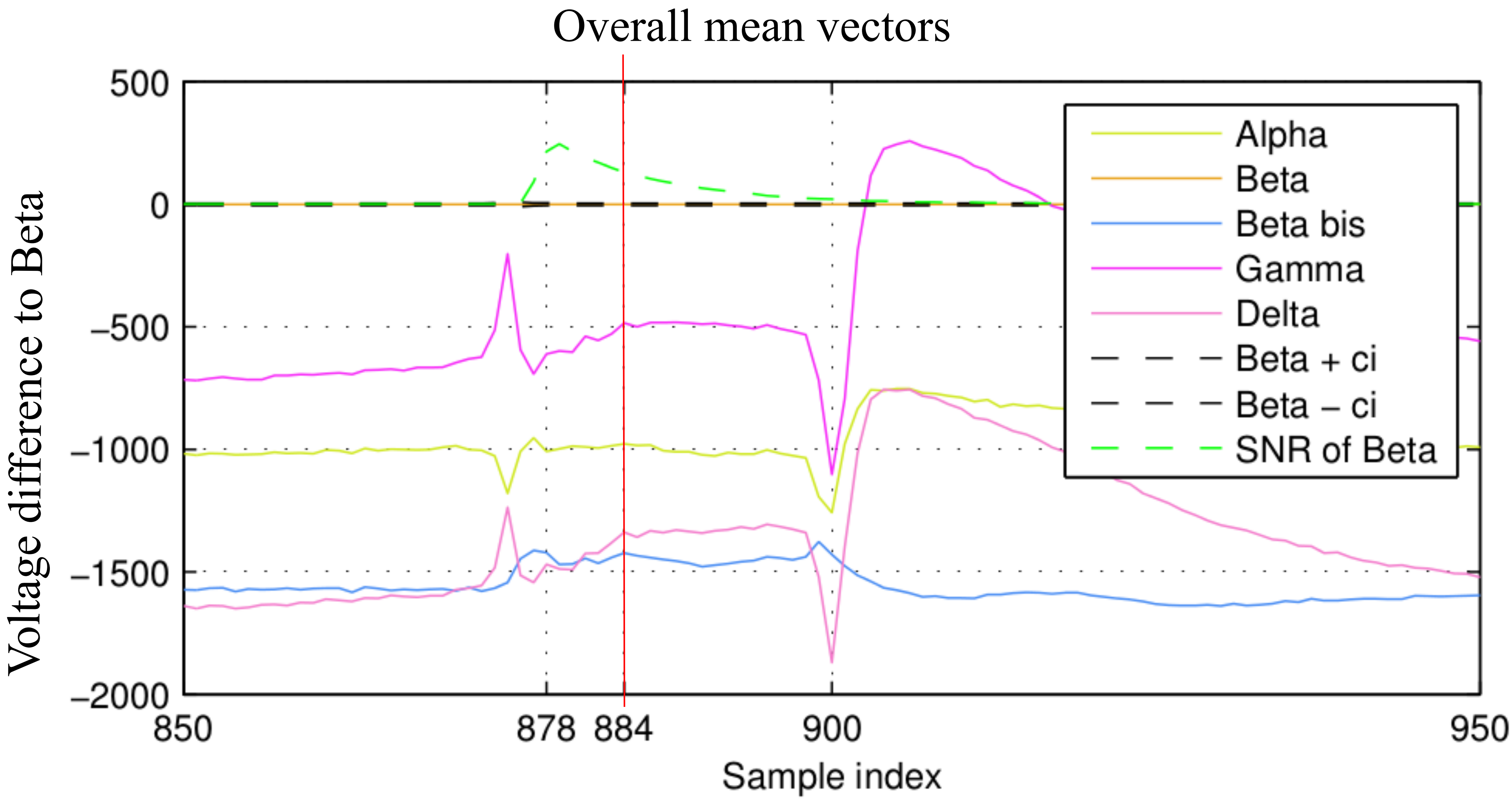
Analysis of overall mean vectors



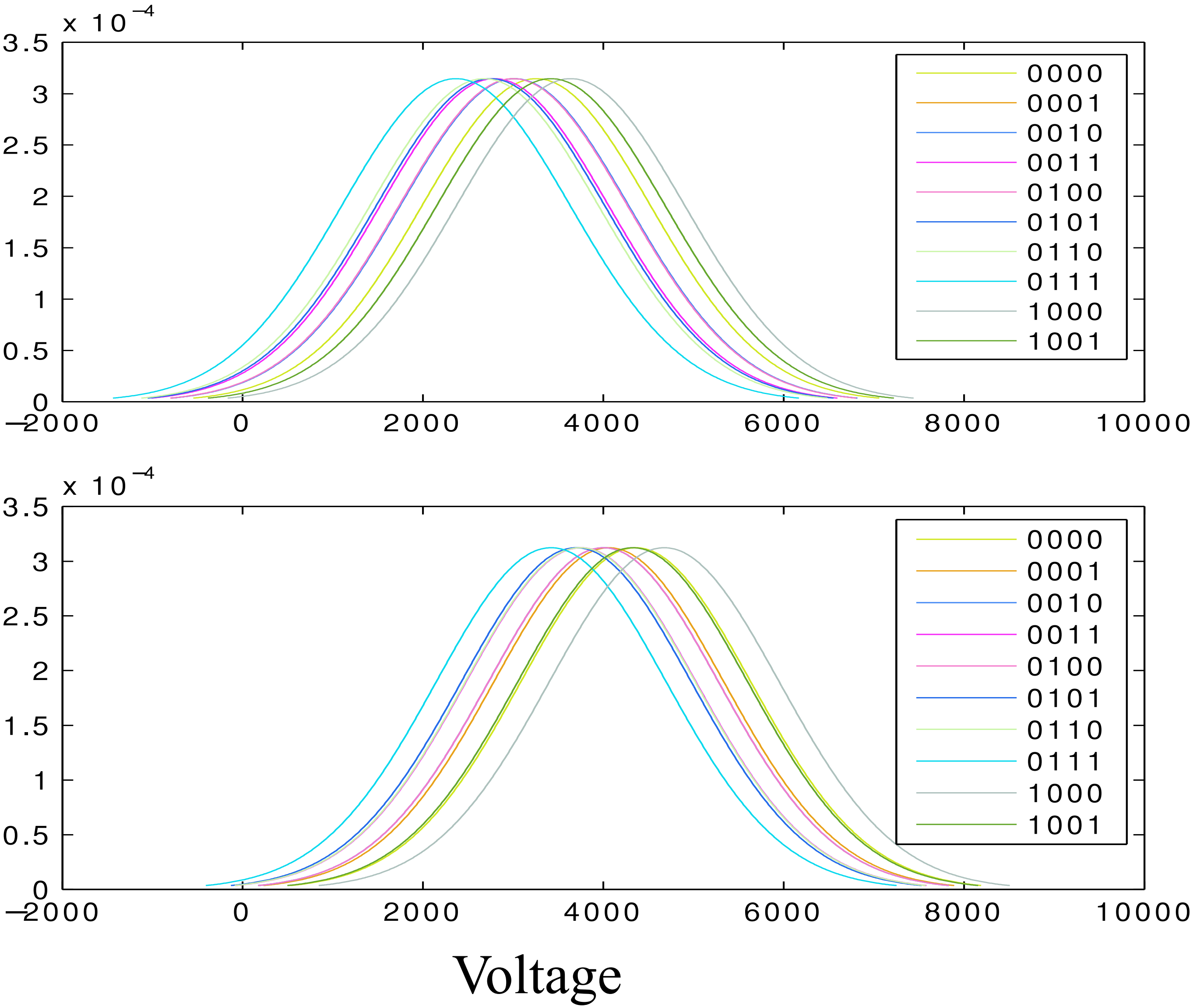
Major problem: low-frequency offset



Major problem: low-frequency offset



Major problem: low-frequency offset



k = 0, 1, …, 9

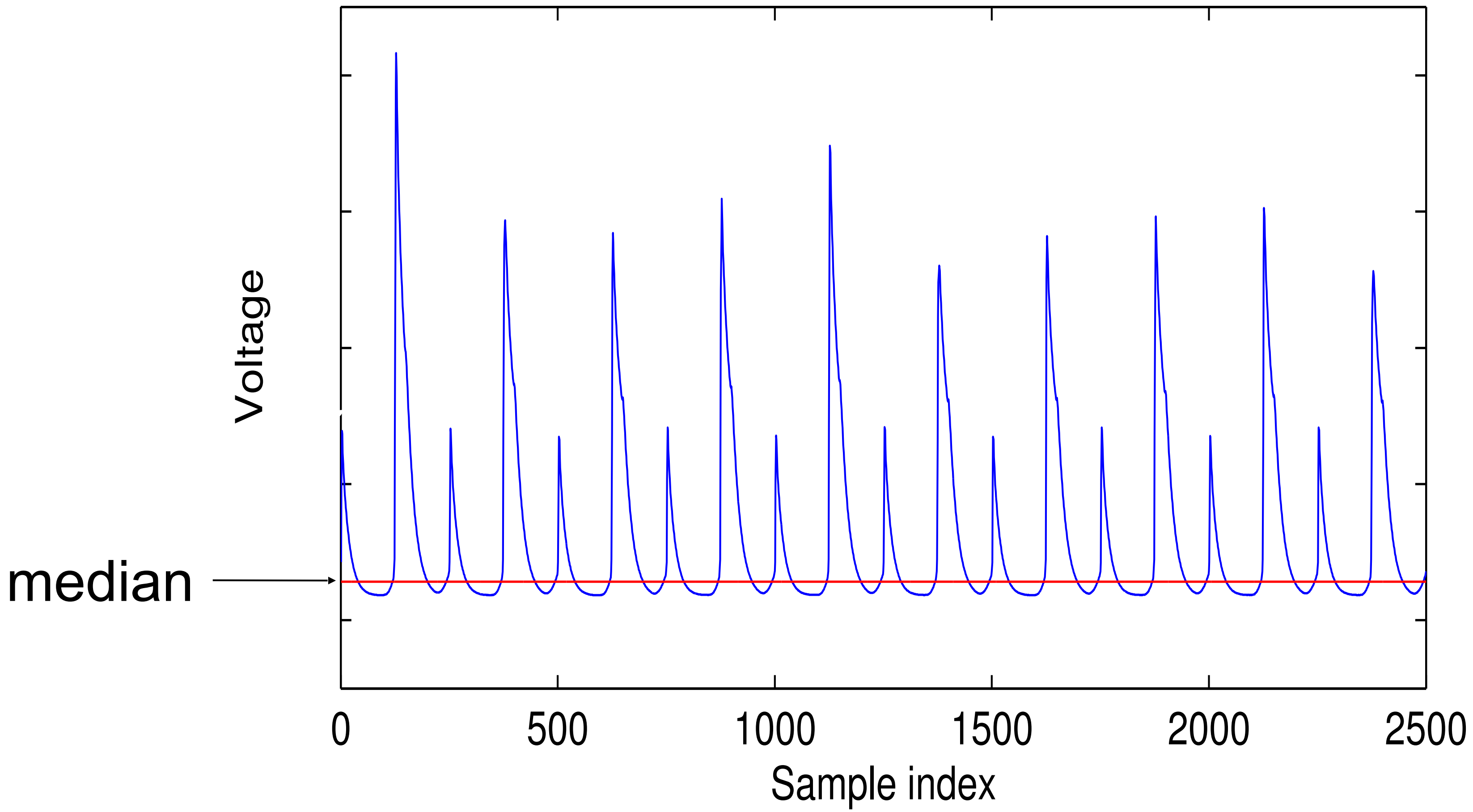
Alpha

Sample j = 884

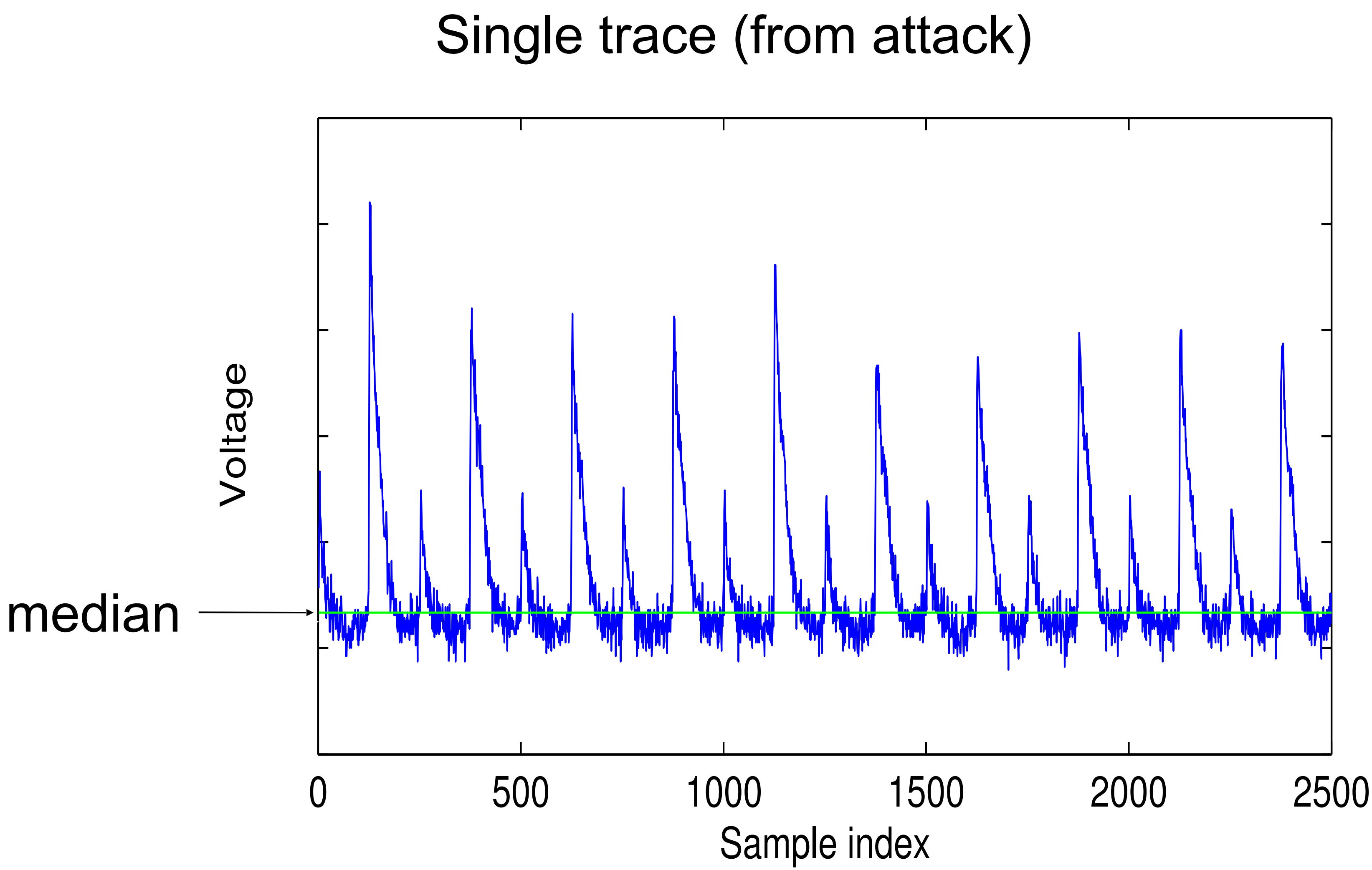
Beta

Adapt for the offset (Meth. 3)

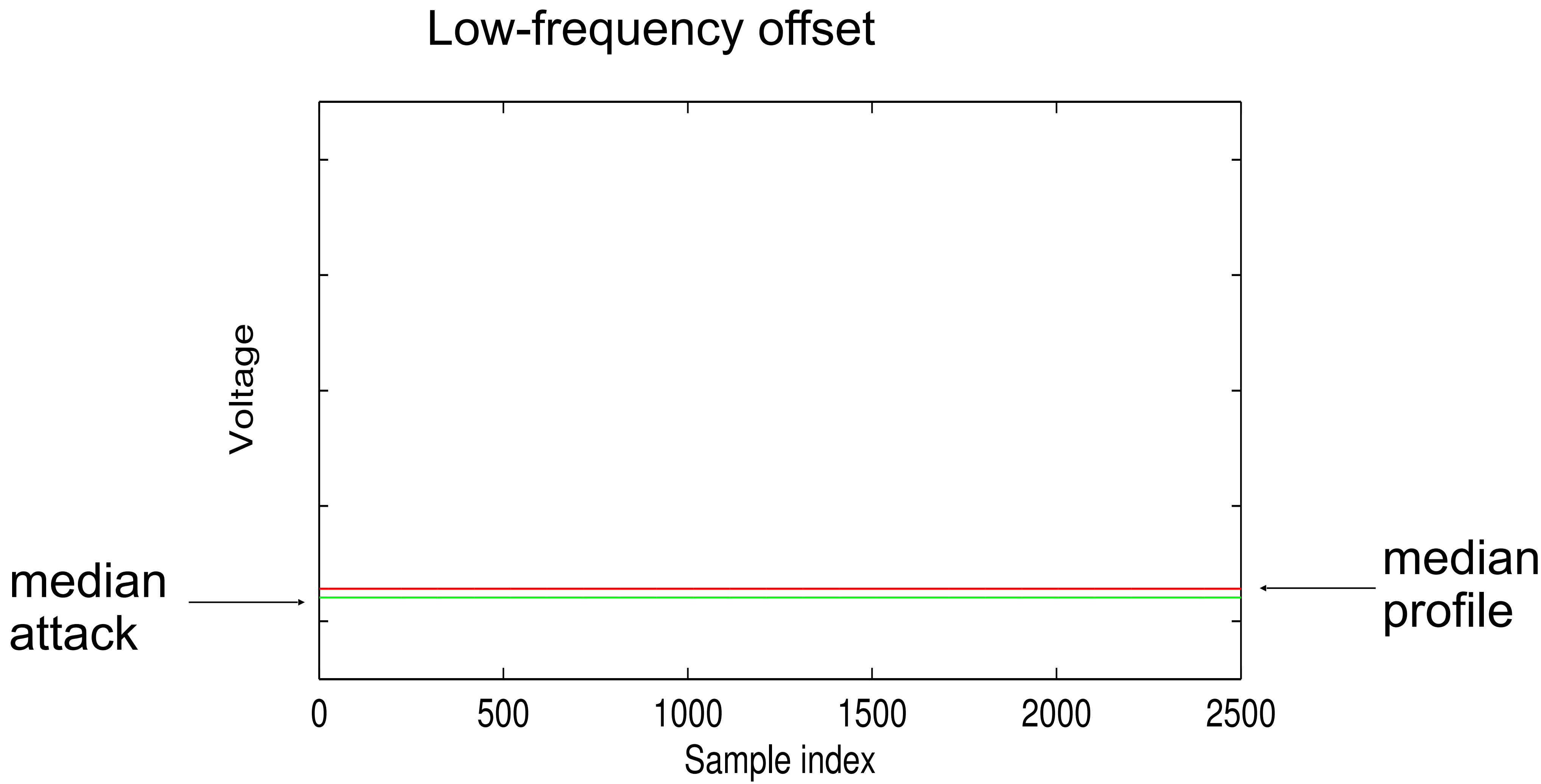
Overall mean trace (from profiling)



Adapt for the offset (Meth. 3)

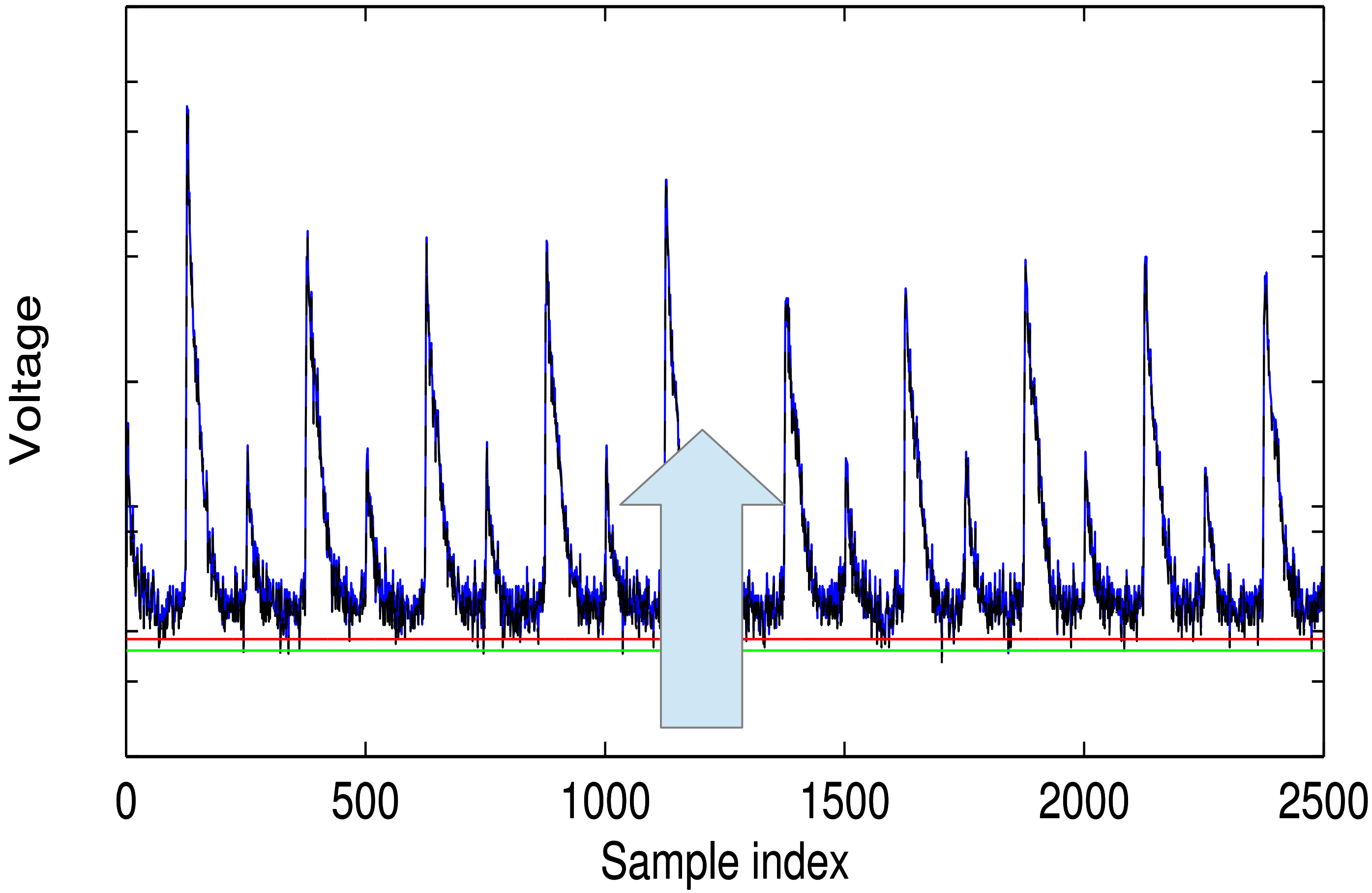


Adapt for the offset (Meth. 3)

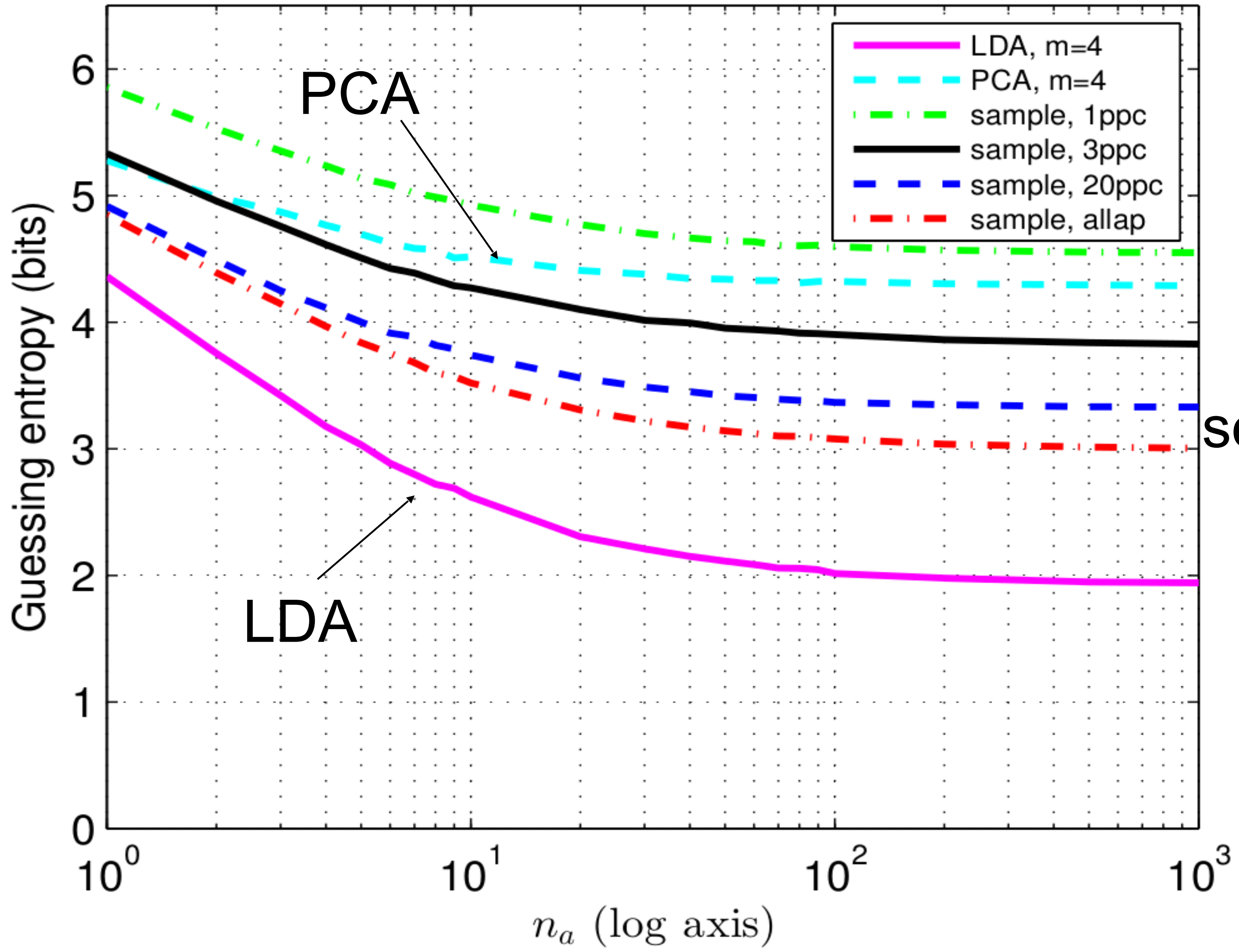


Adapt for the offset (Meth. 3)

Shift attack trace with offset

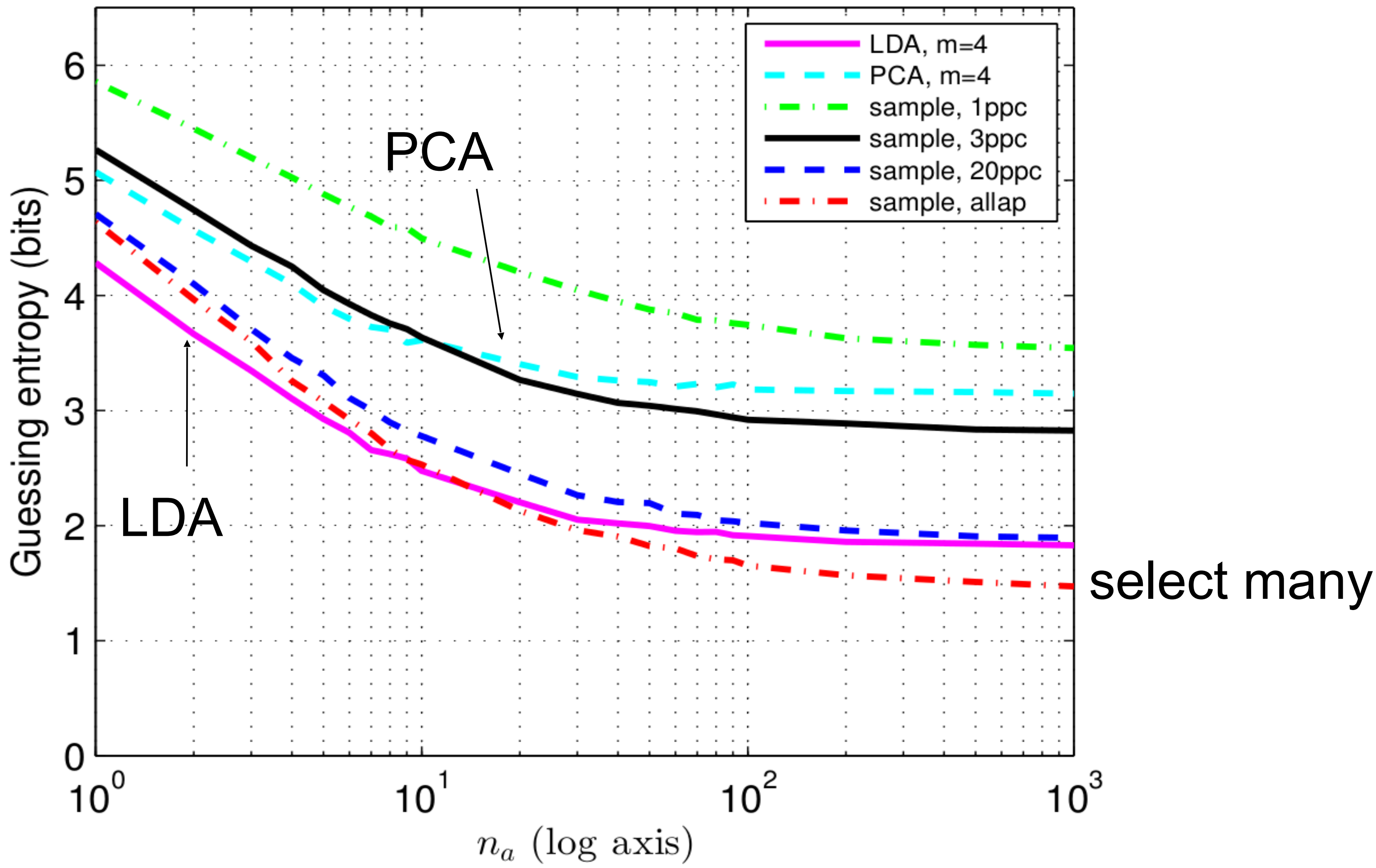


Adapt for the offset (Meth. 3)

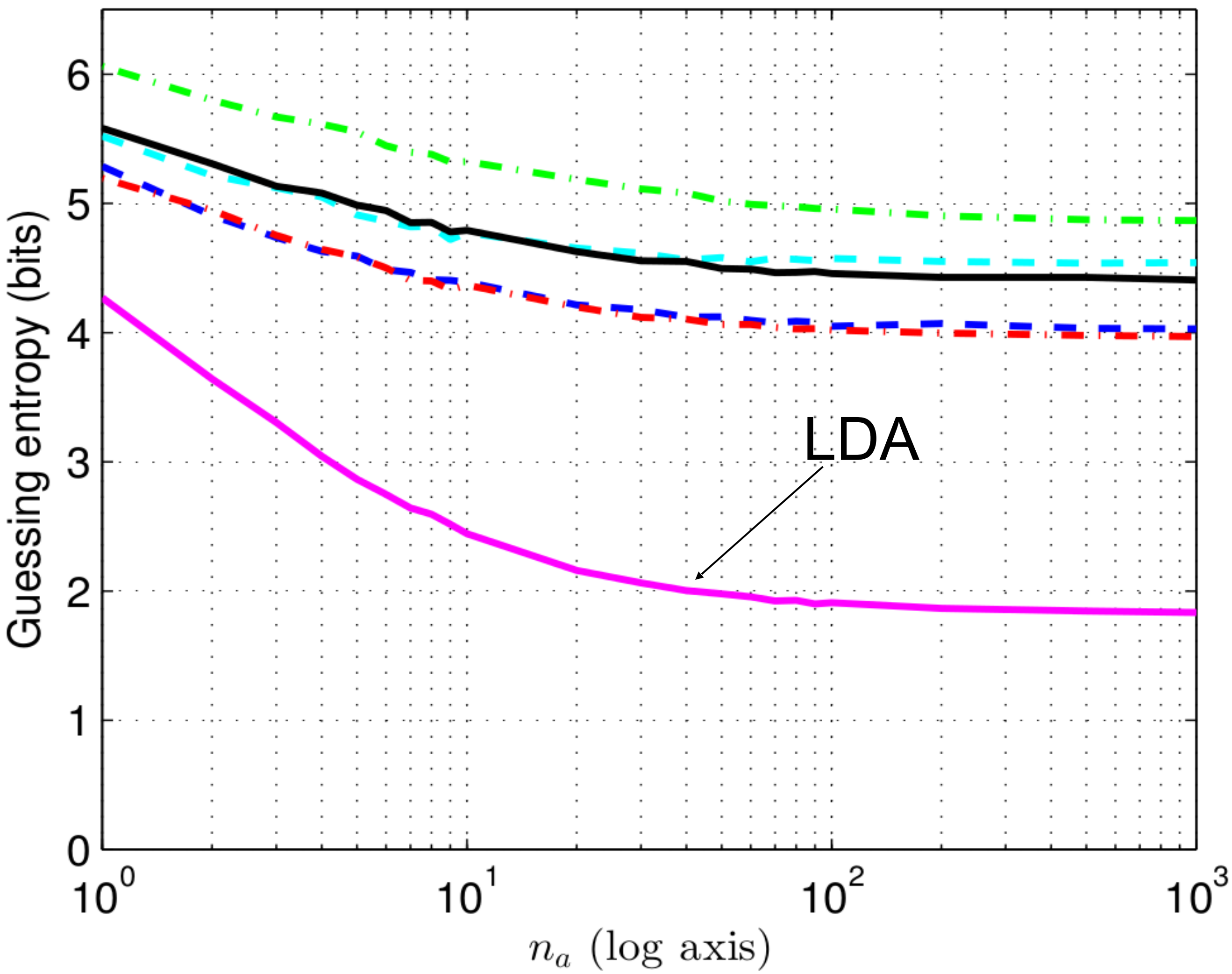


select many

Profile on 3 devices & adapt offset (Meth. 4)



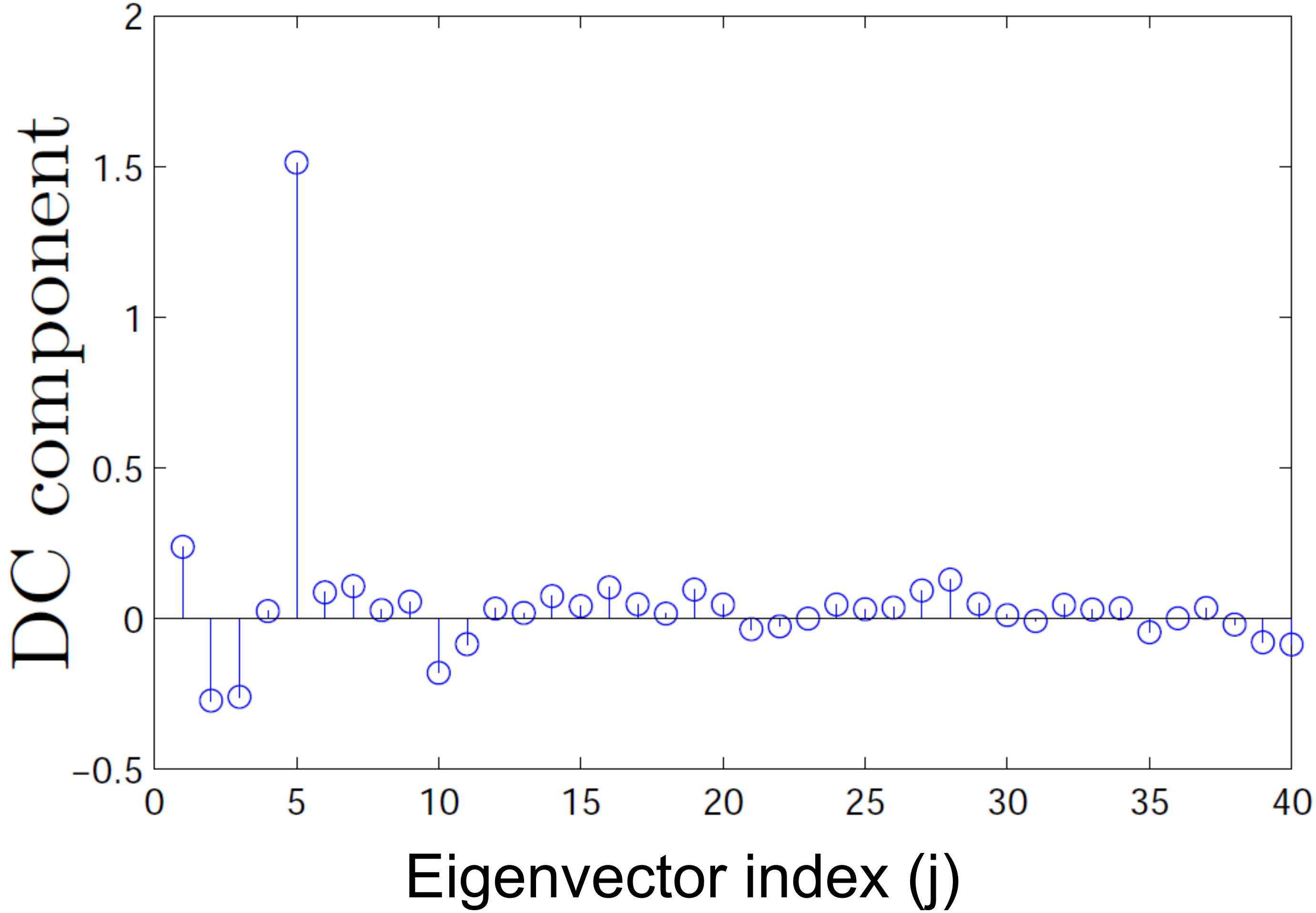
Standard TA works well with LDA



Standard TA works well with LDA

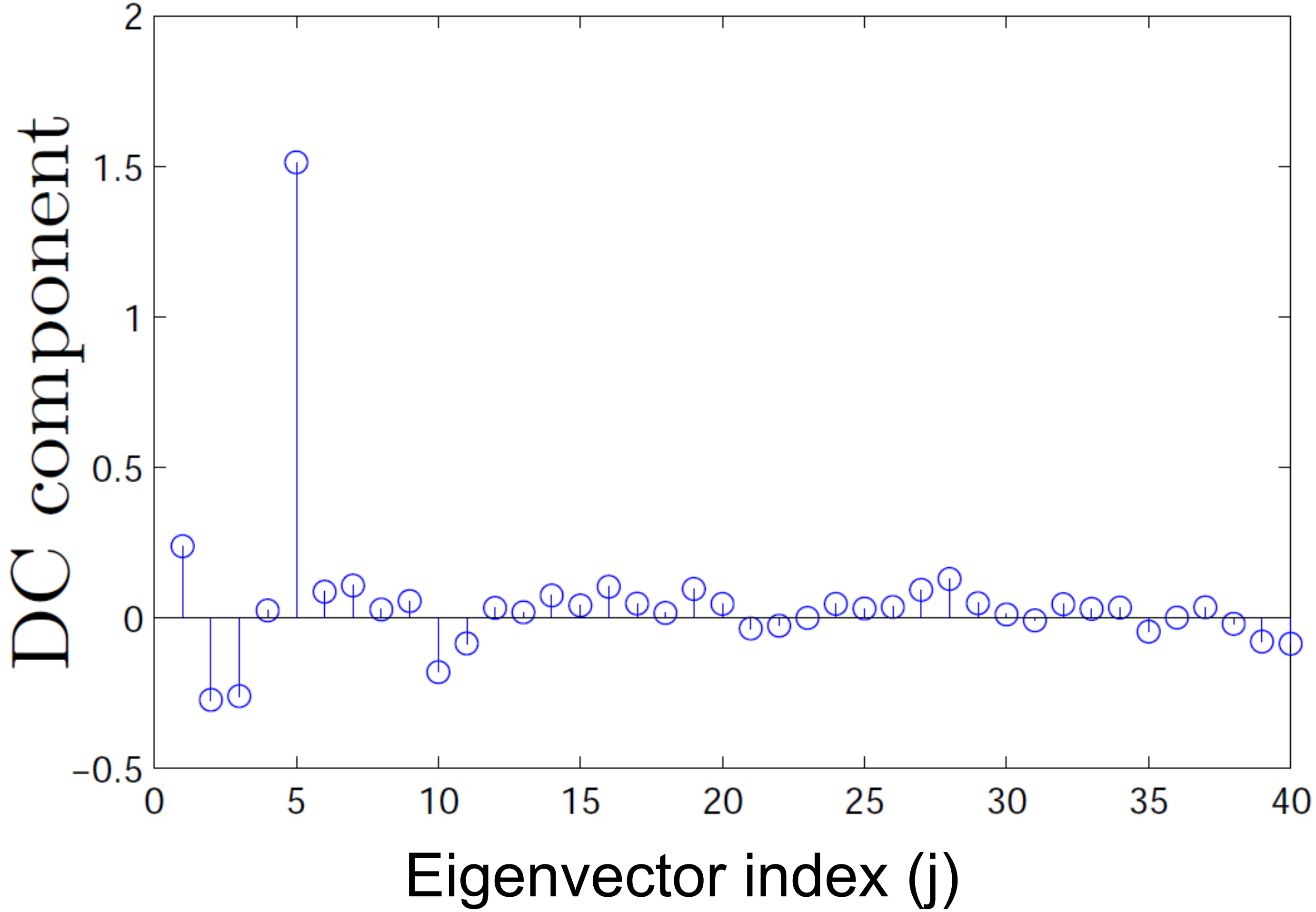
- LDA uses common covariance matrix $\mathrm{S}_{\mathrm{pooled}}$ in computation of eigenvectors
- $\mathrm{S}_{\mathrm{pooled}}$ captures noise factors, such as temperature variations
  - Our acquisition campaigns took several hours to complete
- If variation due to noise is similar across campaigns then LDA can be useful

How to select LDA eigenvectors (1)



$$DC(\boldsymbol{u}_j) = u_j^1 + \ldots + u_j^m$$

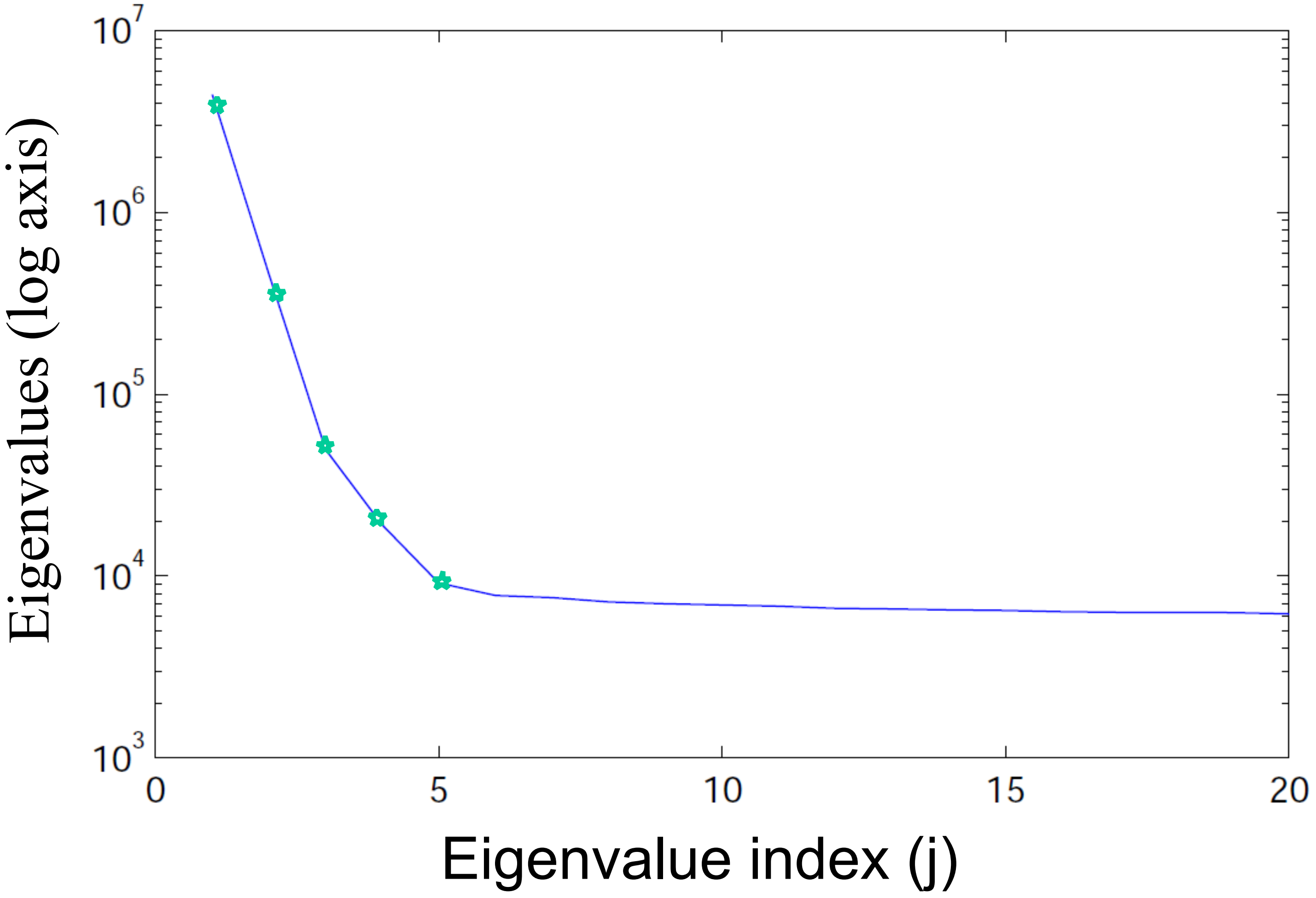
How to select LDA eigenvectors (1)



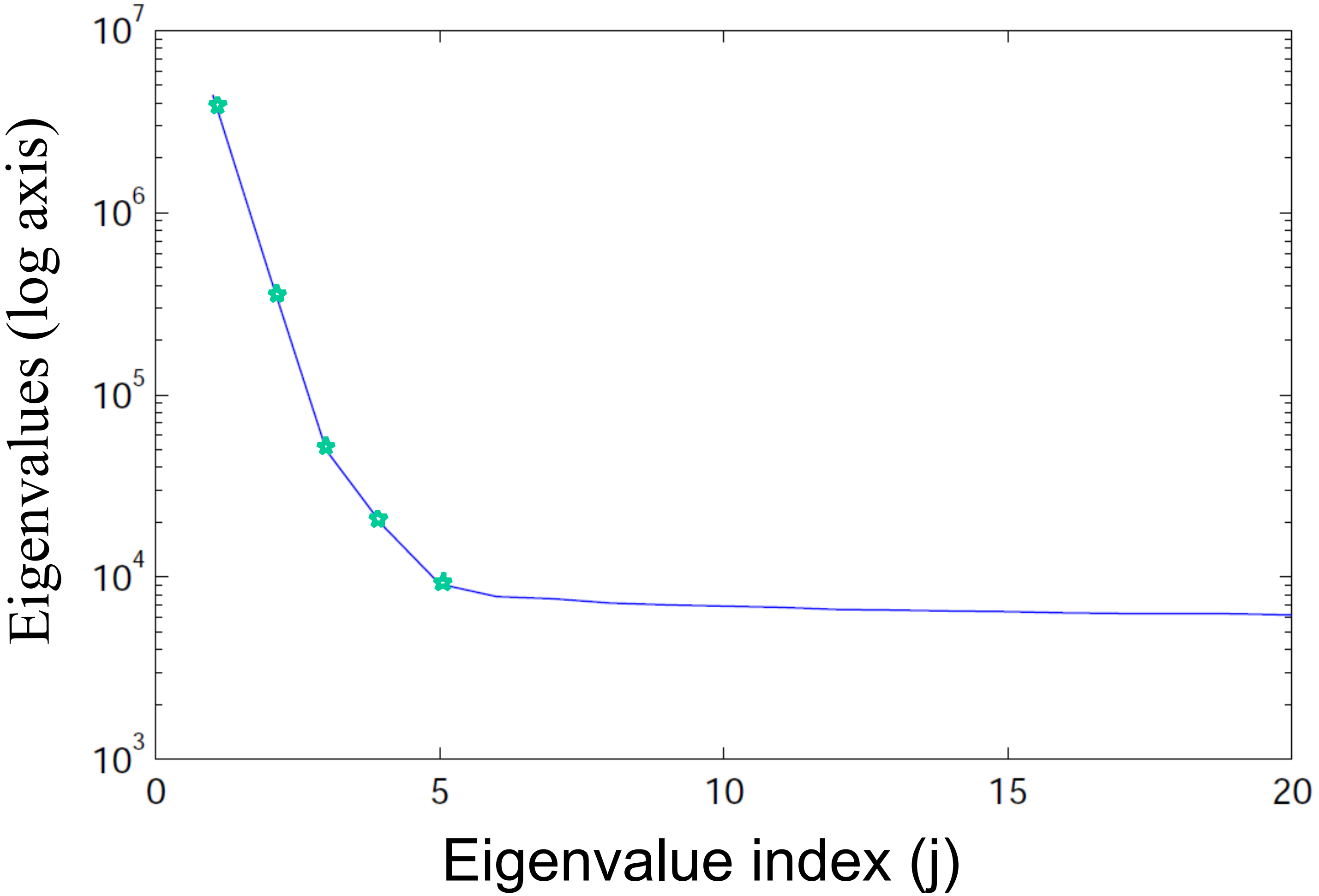
$$m = 4$$

$$DC(\boldsymbol{u}_j) = u_j^1 + \ldots + u_j^m$$

How to select LDA eigenvectors (2)

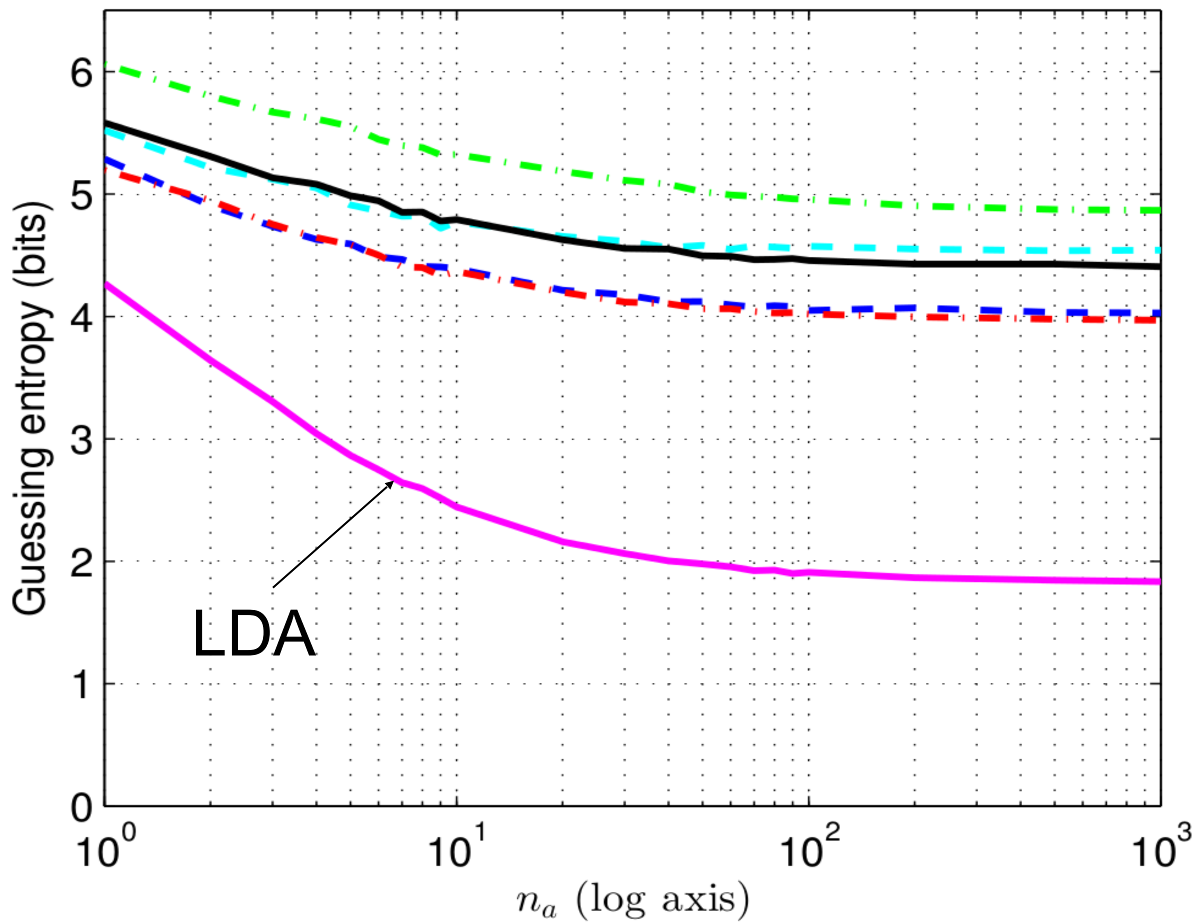


How to select LDA eigenvectors (2)



$m = 4$

How to select LDA eigenvectors

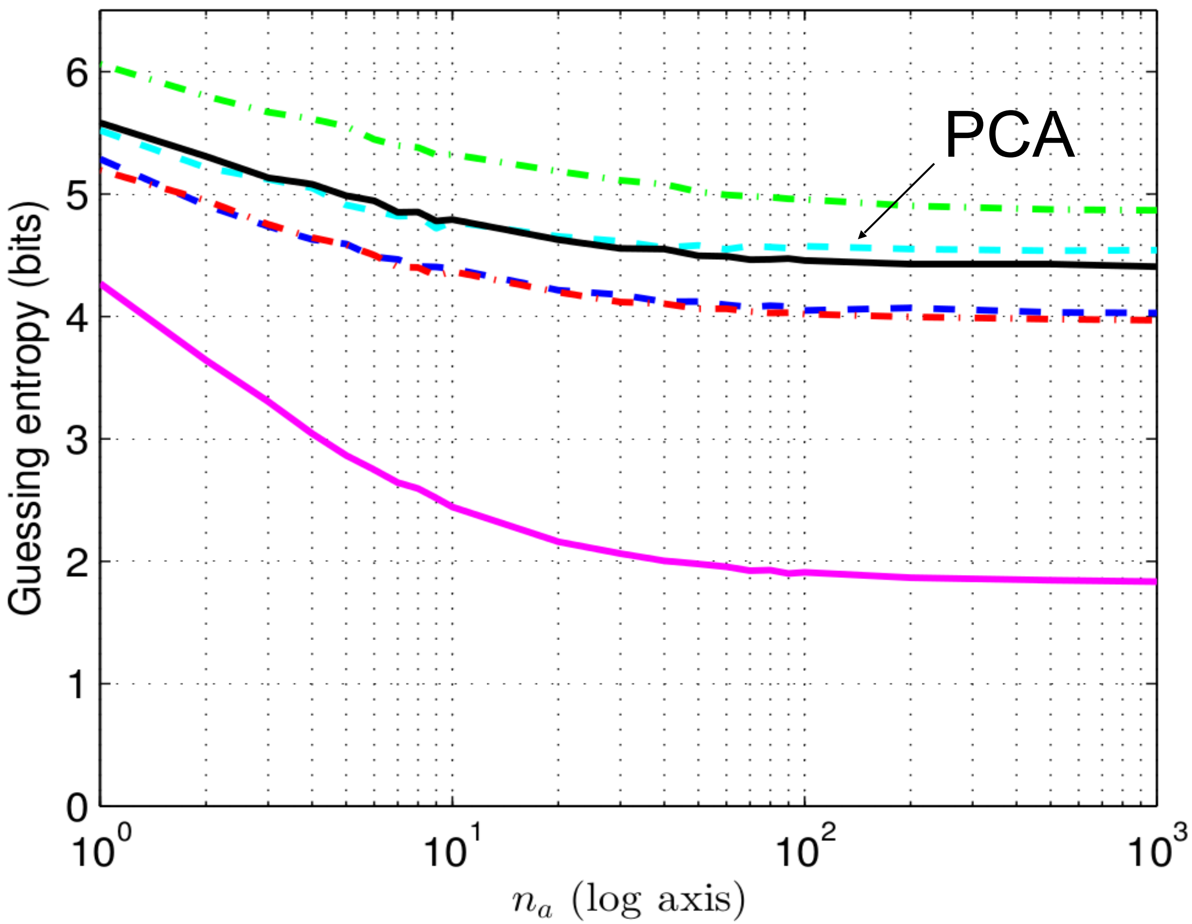


$m = 4$

Good selection of m was only by chance!

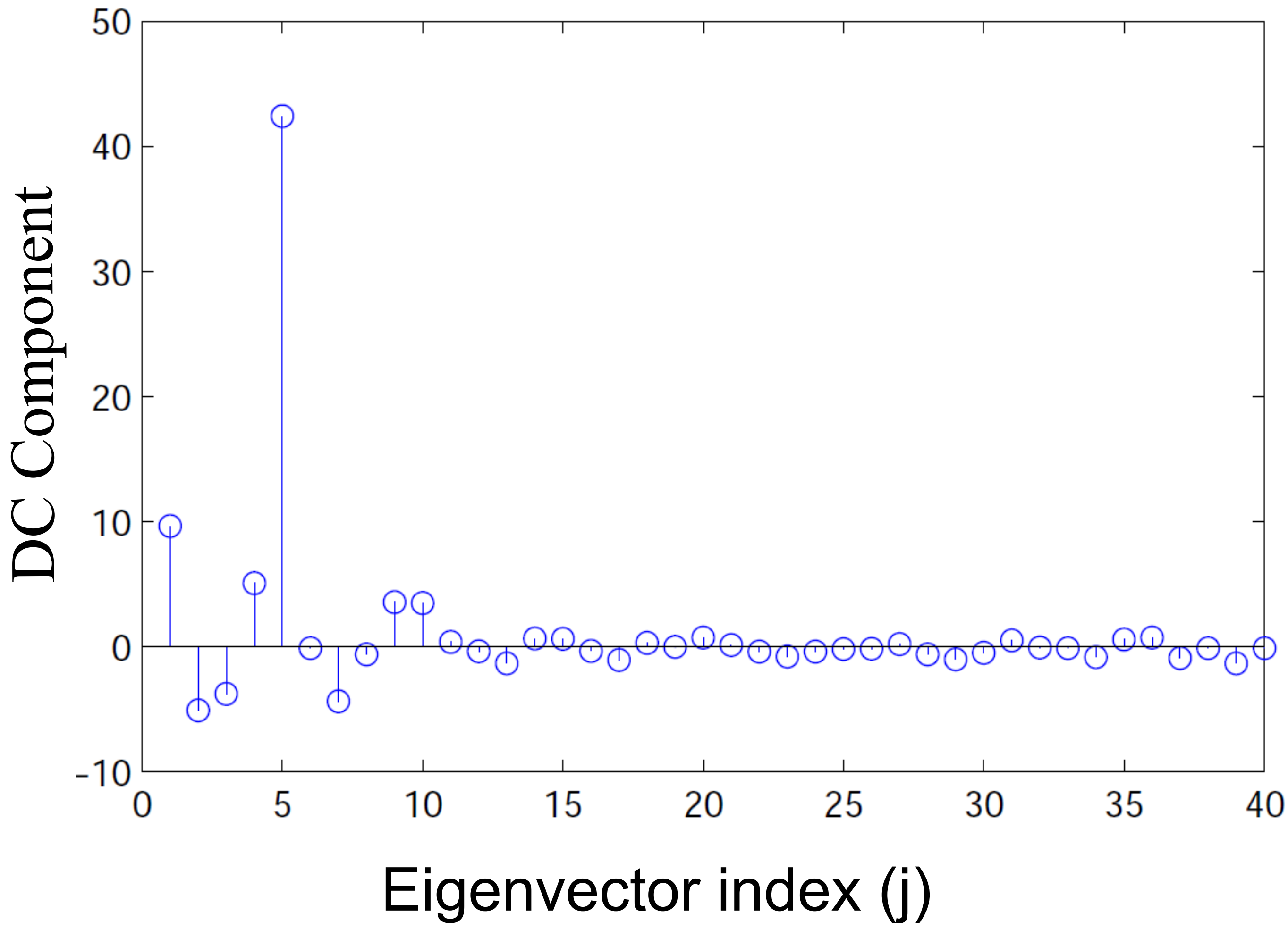
We should look at DC component of eigenvectors

Can we improve PCA?

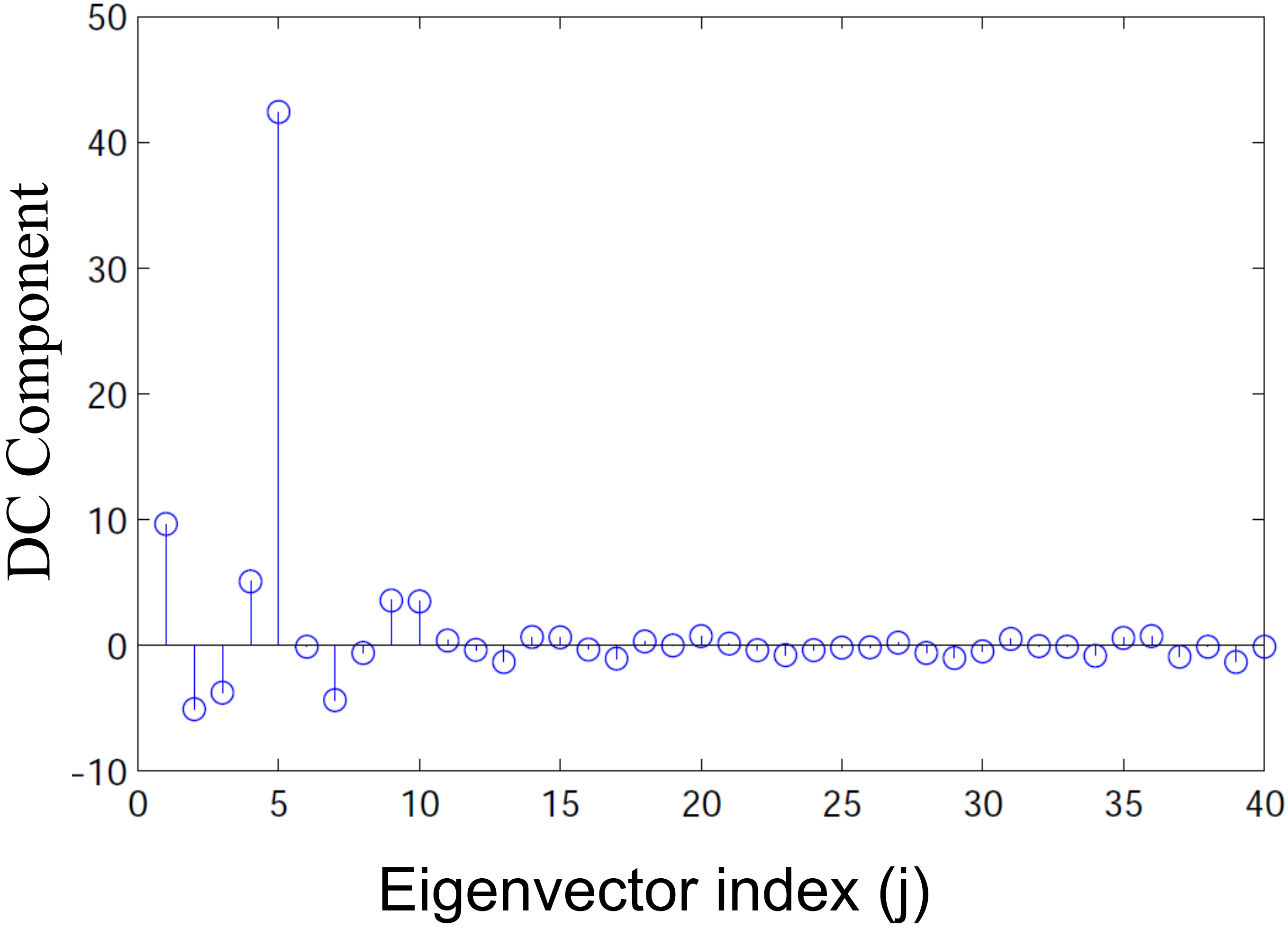


$m = 4$

Can we improve PCA?

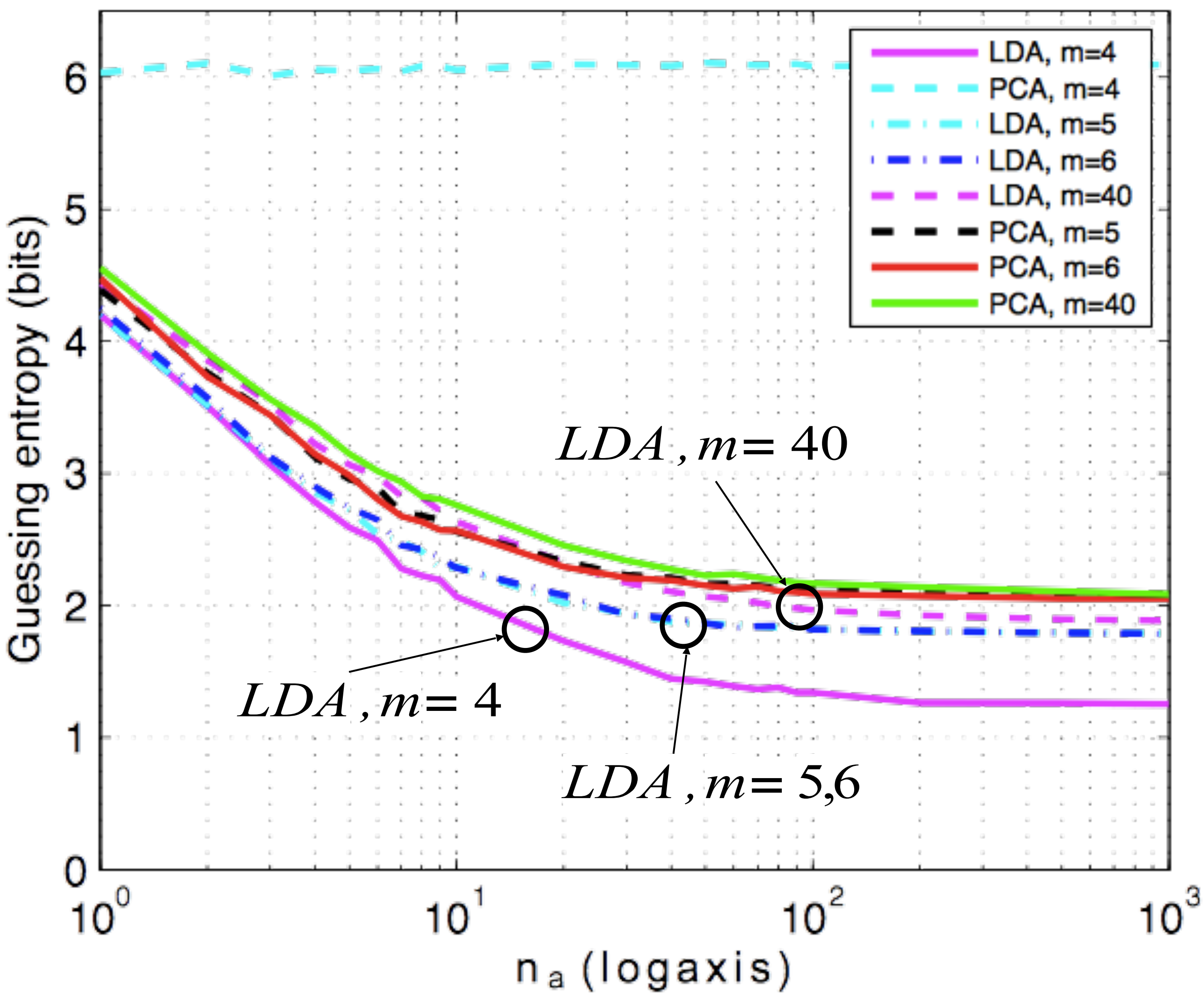


Can we improve PCA?



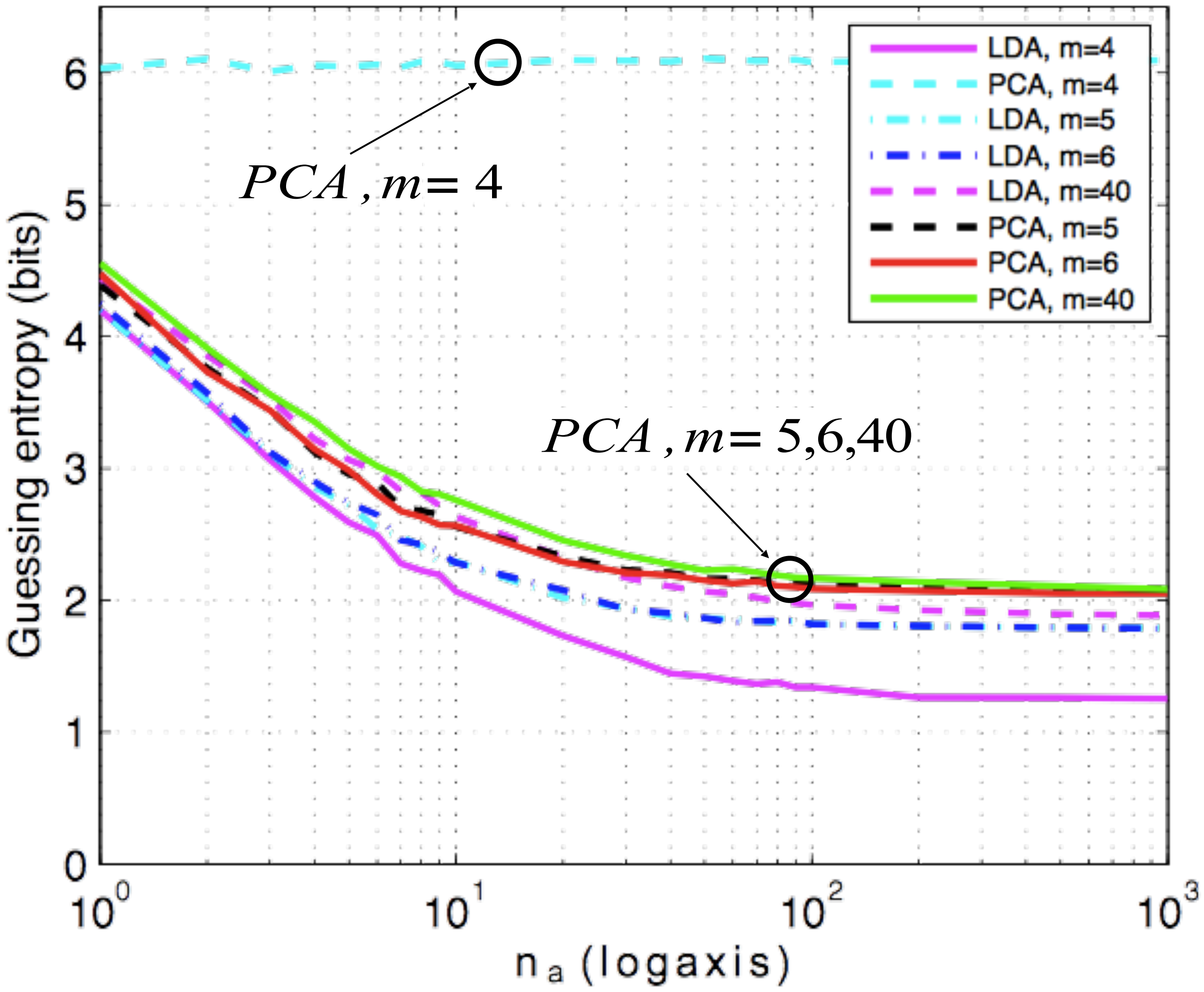
$m \geqslant 5$

Standard TA with PCA and LDA



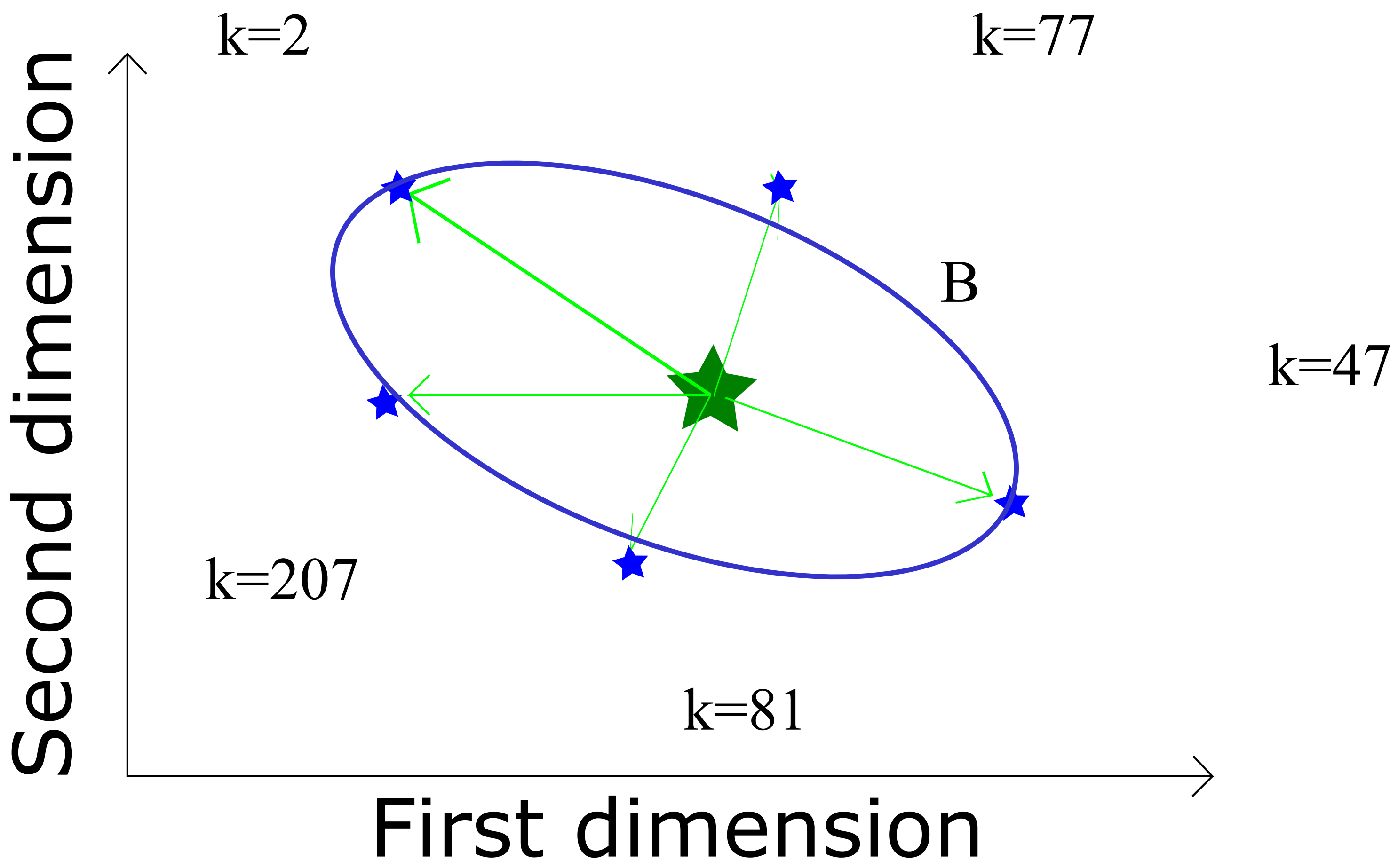
Template Attacks on Different Devices

Standard TA with PCA and LDA

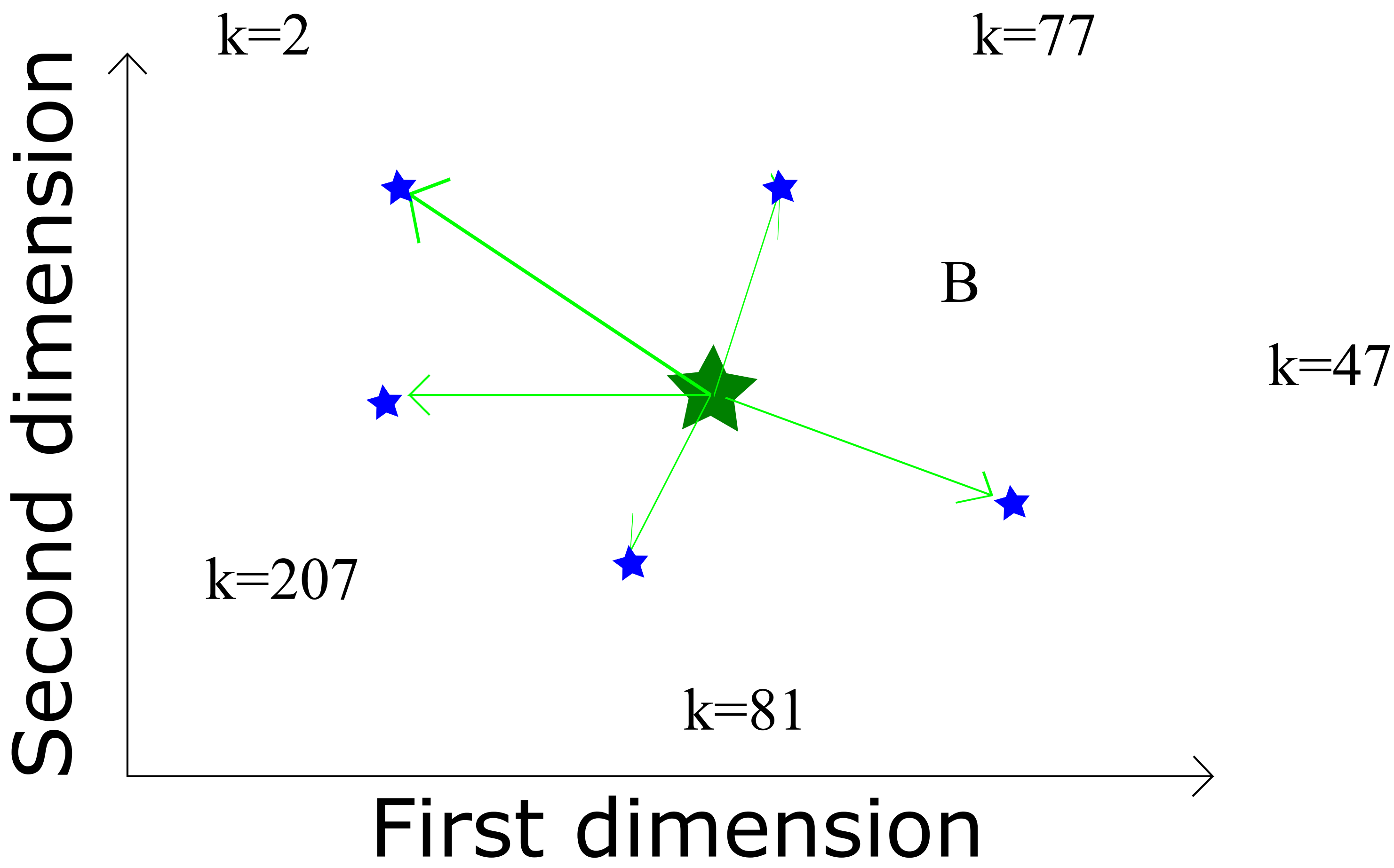


Template Attacks on Different Devices

Method 5: improving PCA



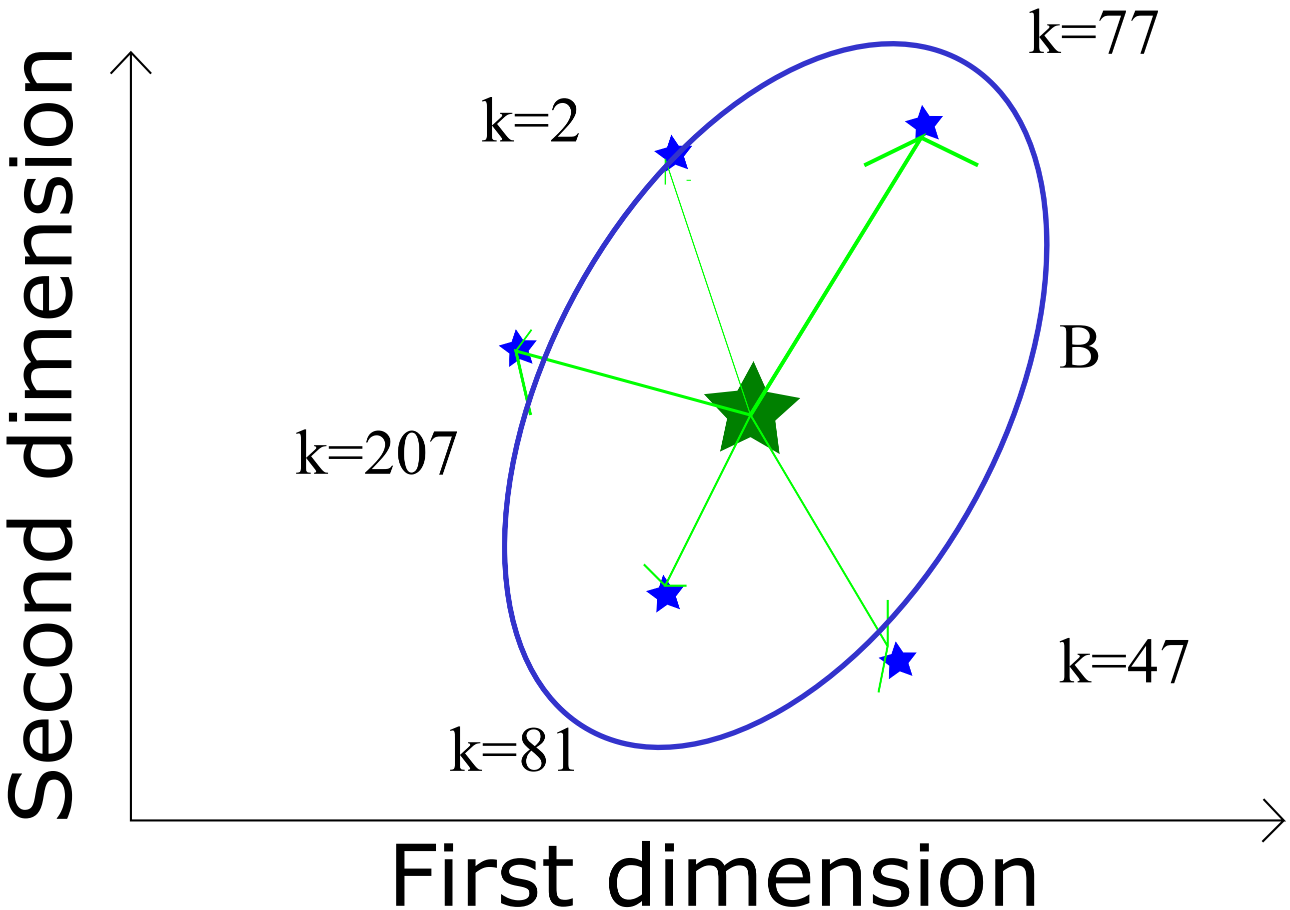
Method 5: improving PCA



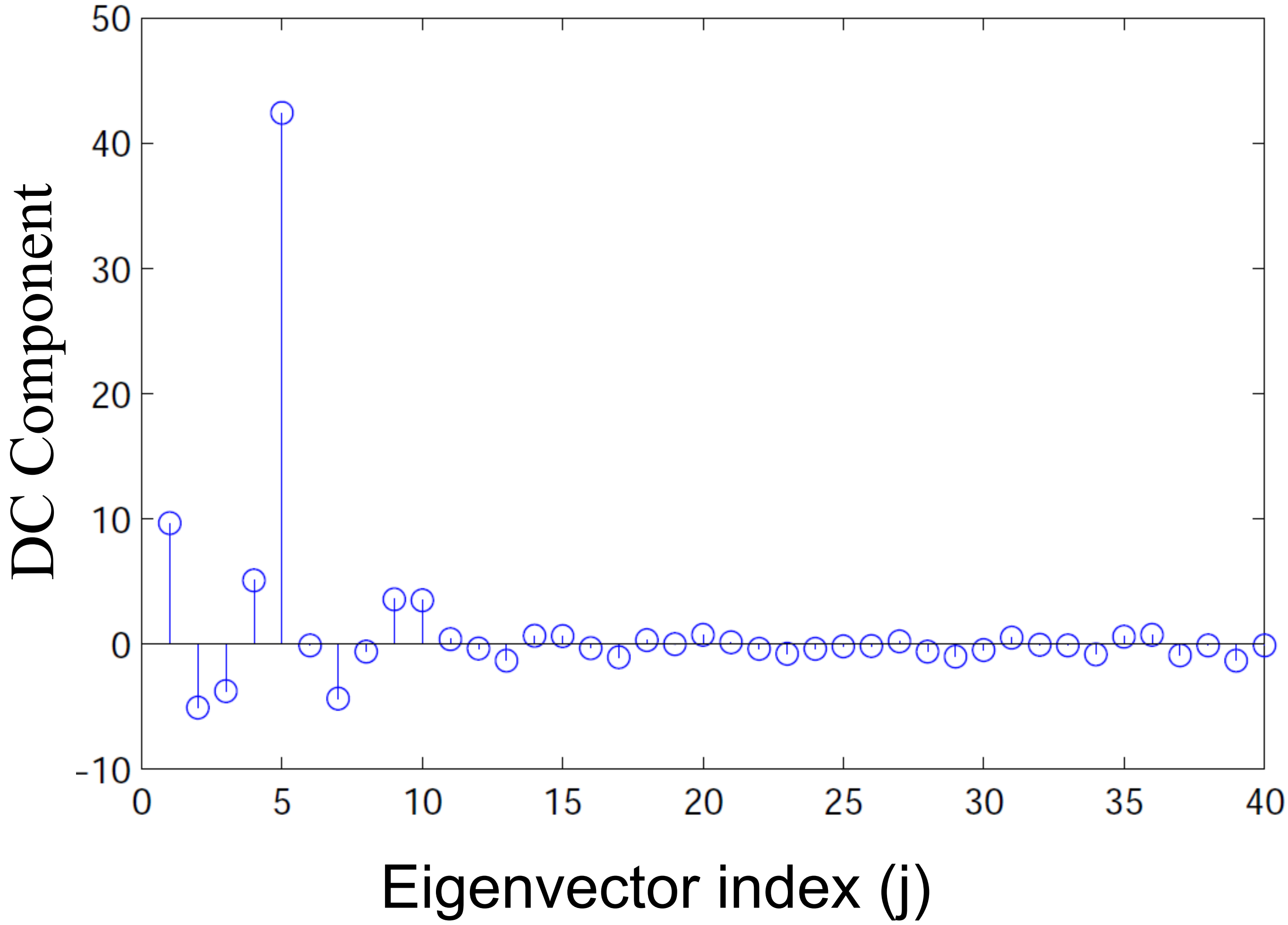
Method 5: improving PCA

- We add random offsets to mean vectors
- This forces DC offset in first eigenvector
  - which should remove DC offset from other eigenvectors, due to orthogonality of eigenvectors

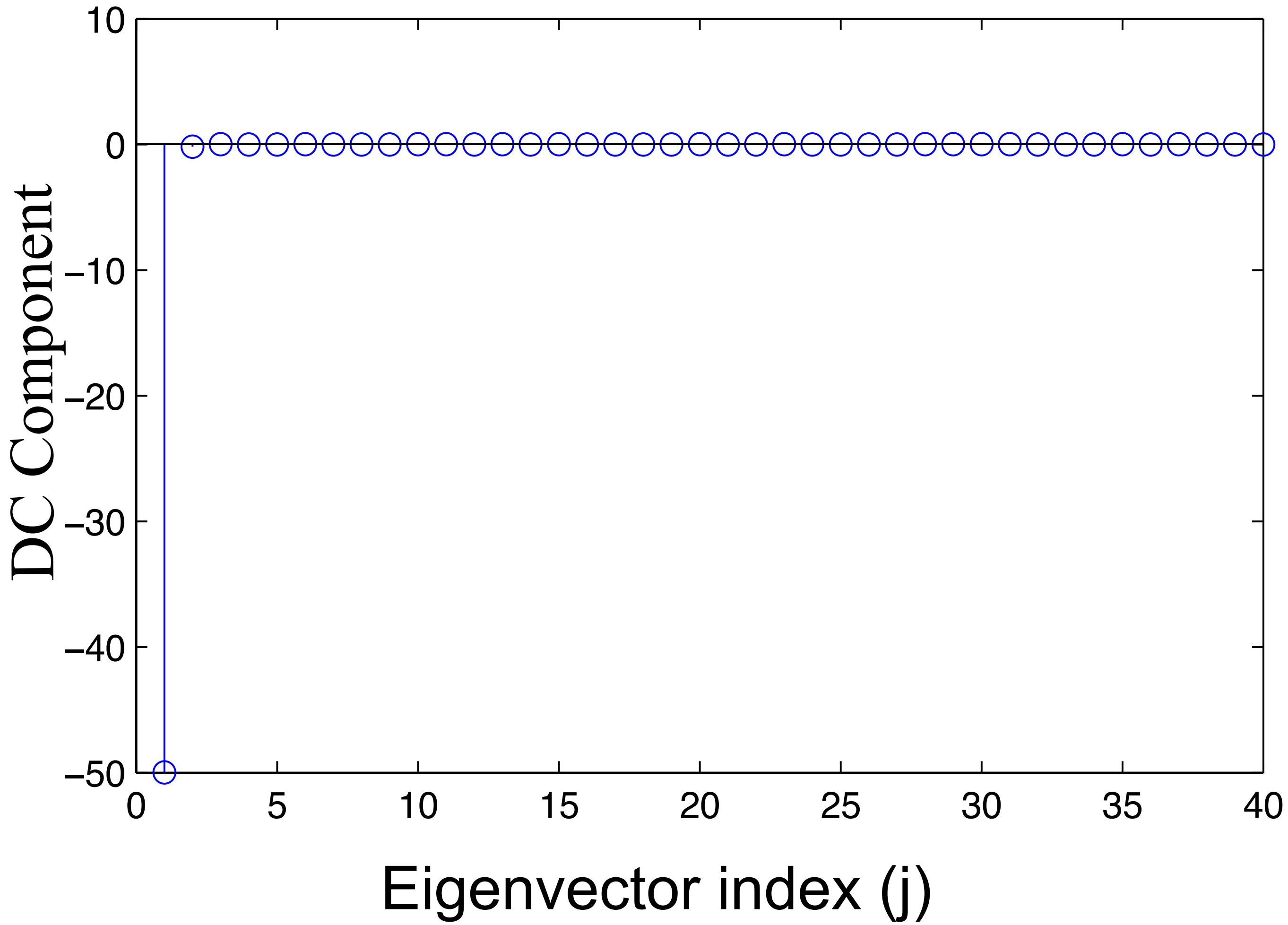
Method 5: improving PCA



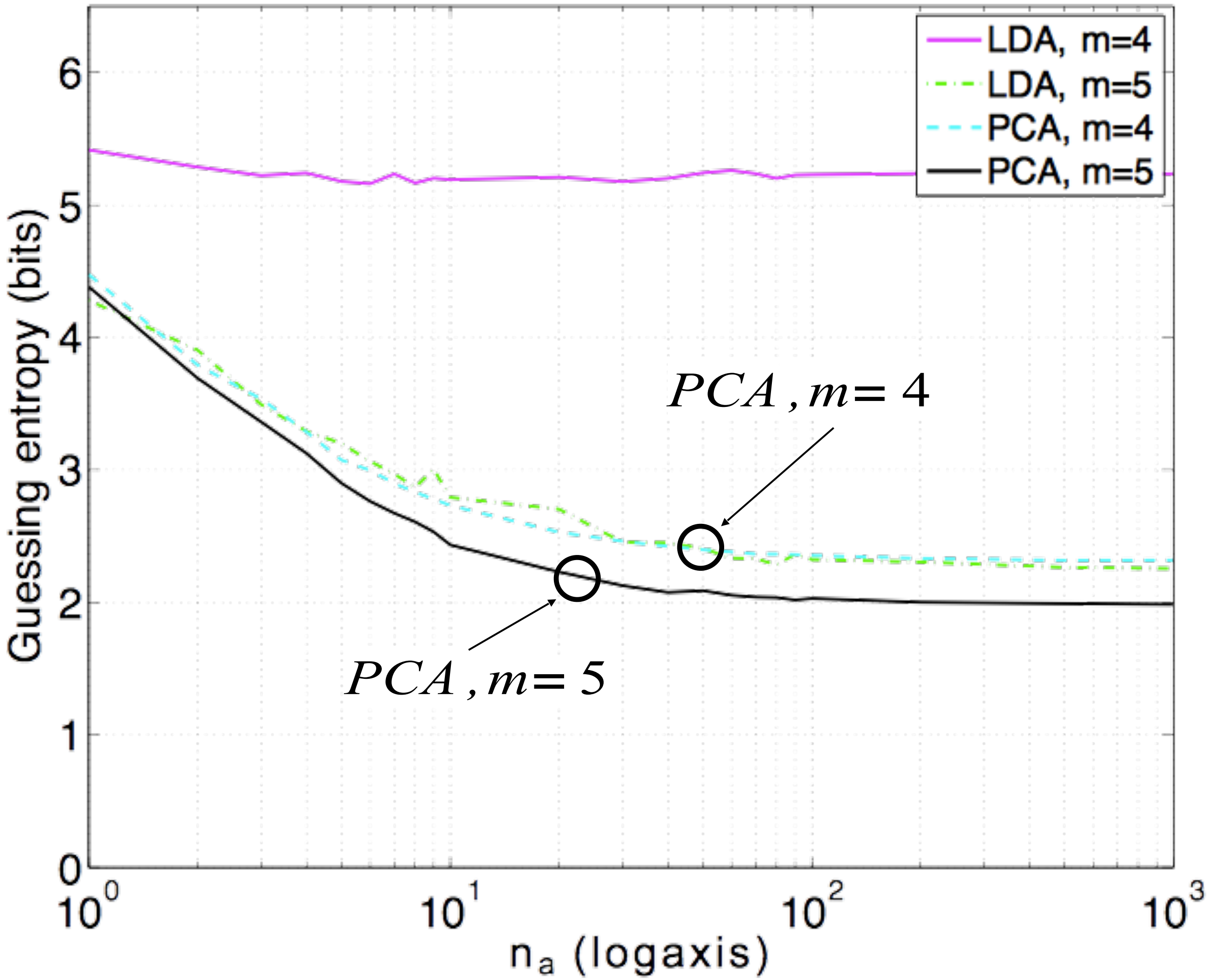
DC offset of PCA eigenvectors: before Method 5



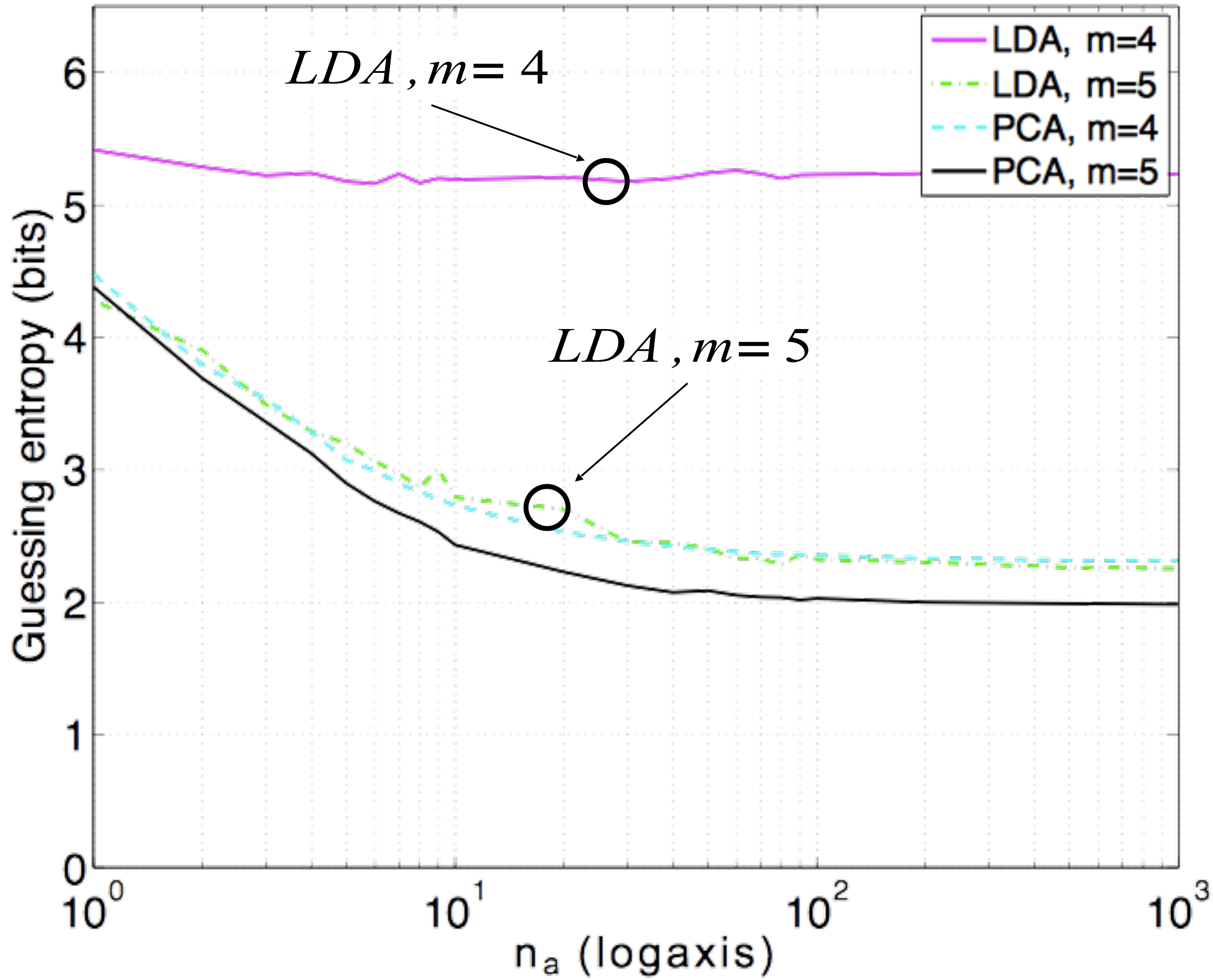
DC offset of PCA eigenvectors: after Method 5



Method 5: improving PCA



Method 5: improving PCA



Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Conclusions

- Extensive evaluation of TA on different devices
  - 4 devices, 5 campaigns
  - Tested compression methods: LDA, PCA, 1/3/20/5%-ile sample selection
  - 5 methods to improve TA
- Inter-device differences similar to inter-campaign differences
- Mostly low frequency offset
- Profiling on multiple devices and manipulation of DC offset can help
- But PCA and LDA can work with standard TA
  - Need to look at DC component
- Improved PCA by forcing in a DC eigenvector
- **Take away message**: compression method matters very much in this case
  - Previous studies may have missed this fact

Questions

Speaker: Omar Choudary
omar.choudary@cl.cam.ac.uk

Co-author: Markus Kuhn
markus.kuhn@cl.cam.ac.uk

Security Group
Computer Laboratory, University of Cambridge