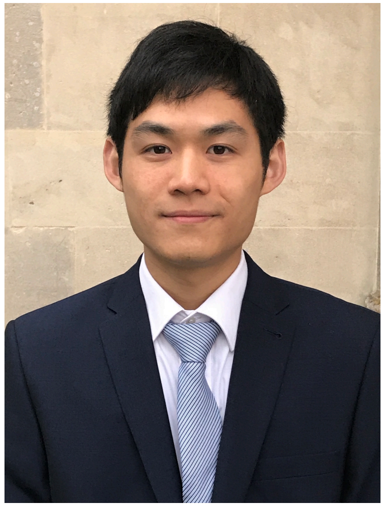
Modulated Sparse Regression Codes

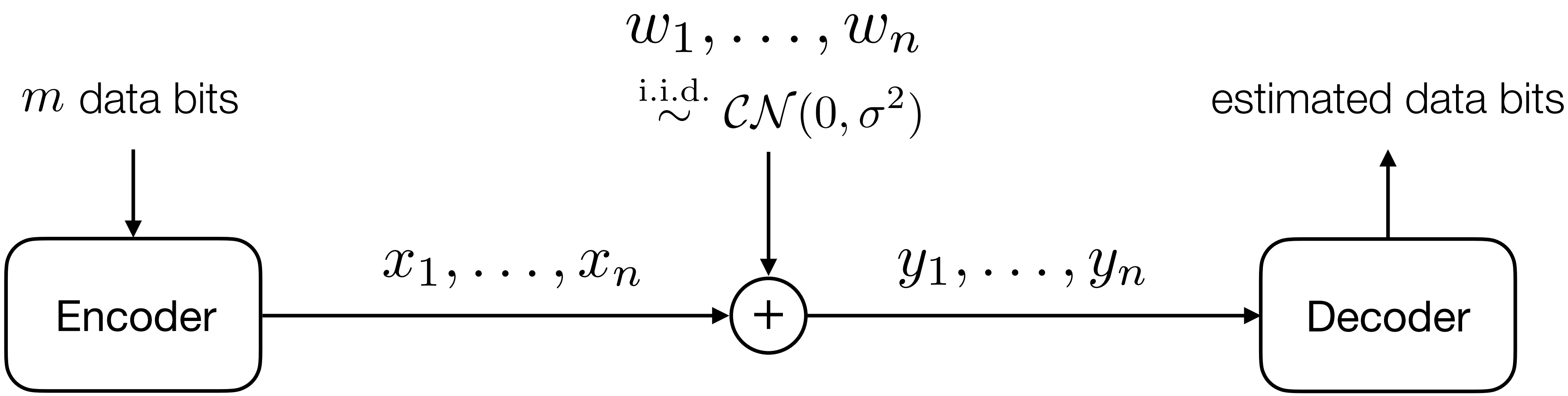
Kuan Hsieh and Ramji Venkataramanan

University of Cambridge, UK

ISIT, June 2020

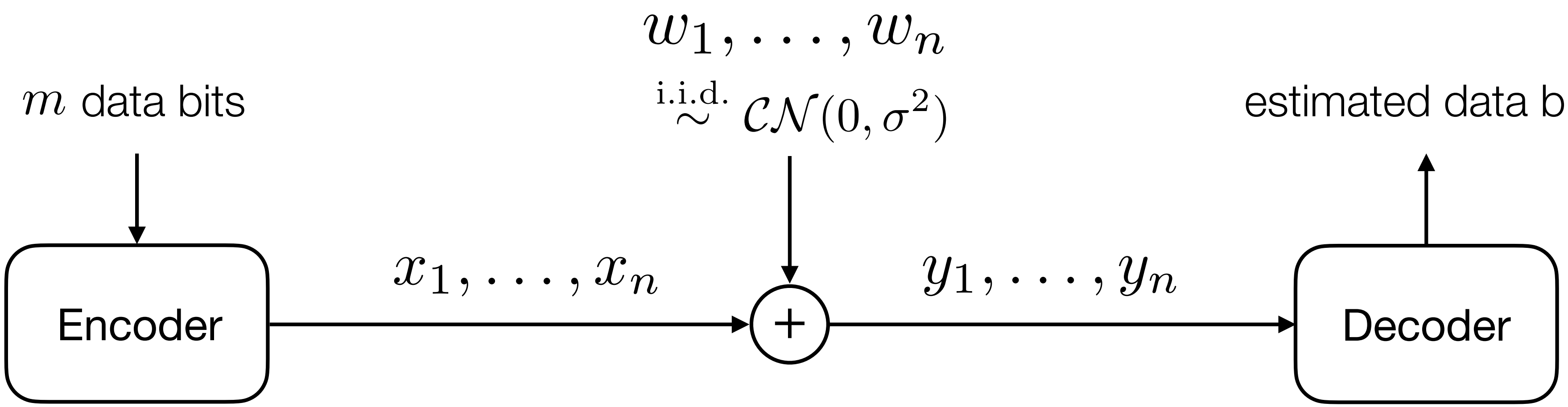


Complex AWGN channel communication



$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}$$

Complex AWGN channel communication



Rate

$$R = \frac{m}{n}$$

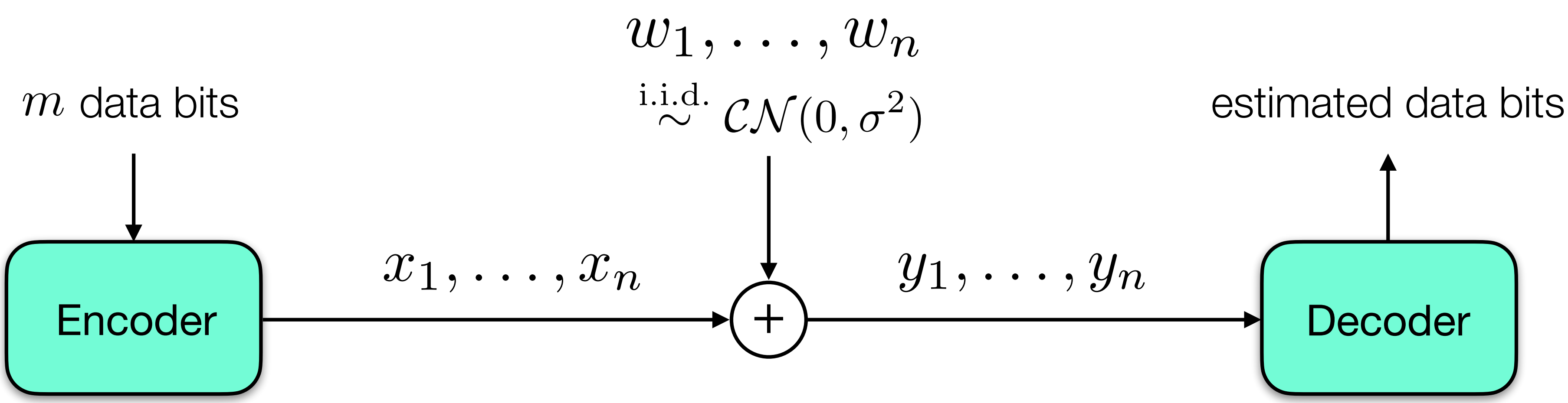
Power constraint

$$\frac{1}{n} \sum_{i=1}^{n} |x_i|^2 \le P$$

Channel capacity

$$\mathcal{C} = \log\left(1 + \frac{P}{\sigma^2}\right)$$

Complex AWGN channel communication



Rate

$$R = \frac{m}{n}$$

Power constraint

$$\frac{1}{n} \sum_{i=1}^{n} |x_i|^2 \leq P$$

Channel capacity

$$\mathcal{C} = \log\left(1 + \frac{P}{\sigma^2}\right)$$

Sparse regression codes (SPARCs)

Encoding

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$$

Codeword $\boldsymbol{x}$

$[x_1, \ldots, x_n]^\top$

Design matrix $\boldsymbol{A}$

ind. Gaussian entries

(sparse)

Message vector $\boldsymbol{\beta}$

encodes data bits

[Joseph and Barron '12]

Sparse regression codes (SPARCs)

Encoding

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$$

Codeword $\boldsymbol{x}$
$[x_1, \ldots, x_n]^\top$

Design matrix $\boldsymbol{A}$
ind. Gaussian entries

(sparse)
Message vector $\boldsymbol{\beta}$
encodes data bits

Decoding

Estimate $\boldsymbol{\beta}$ given $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$

[Joseph and Barron '12]

SPARC encoding

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$$

$$\boldsymbol{\beta} : [\ ...0, 1, 0 \ \vert \ 1, 0, 0, ... \ \vert \ \cdots \ \vert \ 0, 1, 0, ... \ ]^\top$$

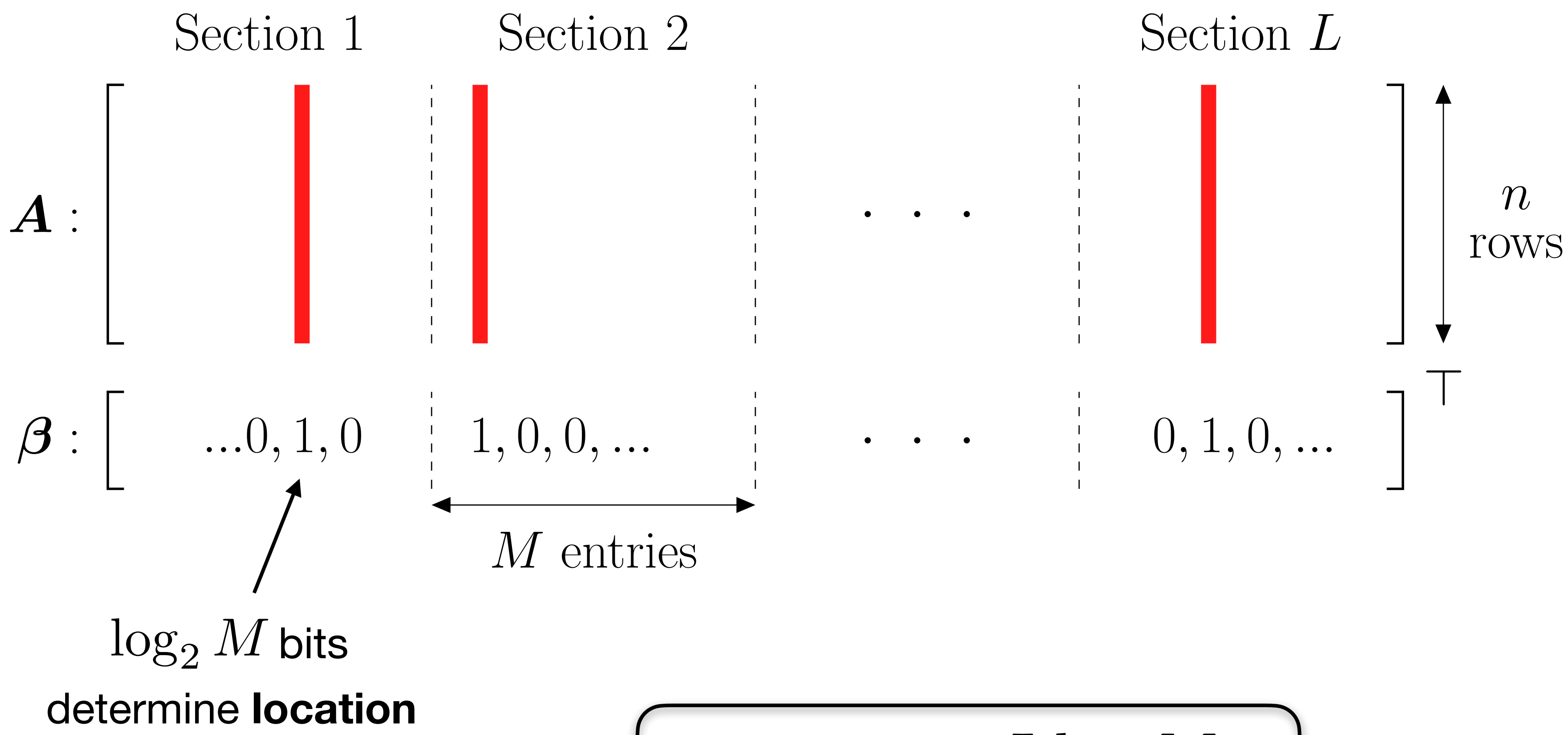
M entries

$\log_2 M$ bits

determine **location**

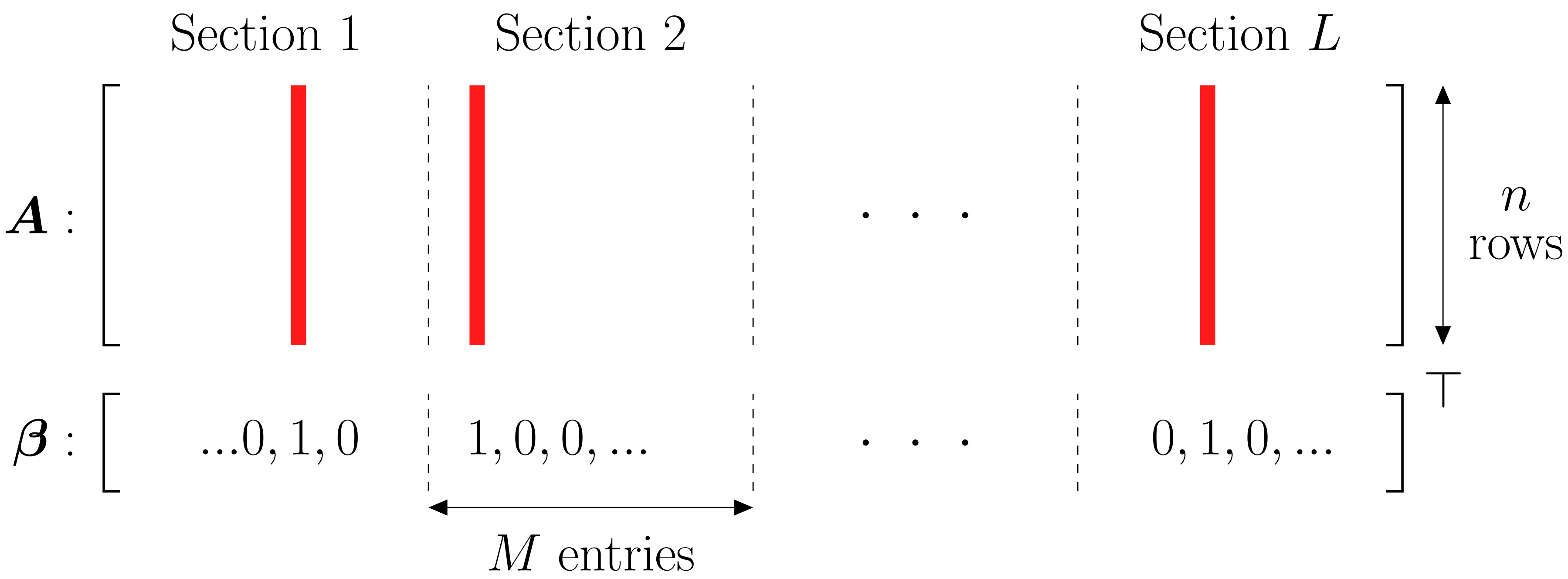
SPARC encoding

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$$



Rate $R = \frac{L \log M}{n}$

SPARC decoding



Estimate $\boldsymbol{\beta}$ given $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$

Section Error Rate: (SER) $$\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}\left\{\widehat{\boldsymbol{\beta}}_{\boldsymbol{\ell}} \neq \boldsymbol{\beta}_{\boldsymbol{\ell}}\right\}$$

Previous results on (unmodulated) SPARCs

Maximum likelihood decoding

[Joseph and Barron '12]

Matrix designs + efficient decoding

Power allocation

Adaptive, Successive Hard-thresholding

[Joseph and Barron '14]

Adaptive, Successive Soft-thresholding

[Cho and Barron '13]

Approximate Message Passing

[Barbier and Krzakala '17]

[Rush, Greig and Venkataramanan '17]

Spatial coupling

Approximate Message Passing

[Barbier et al. '14-'19]

[Rush, Hsieh and Venkataramanan '18, '19, '20]

Previous results on (unmodulated) SPARCs

Maximum likelihood decoding
[Joseph and Barron ’12]

Matrix designs + efficient decoding

Power allocation

Adaptive, Successive Hard-thresholding
[Joseph and Barron ’14]

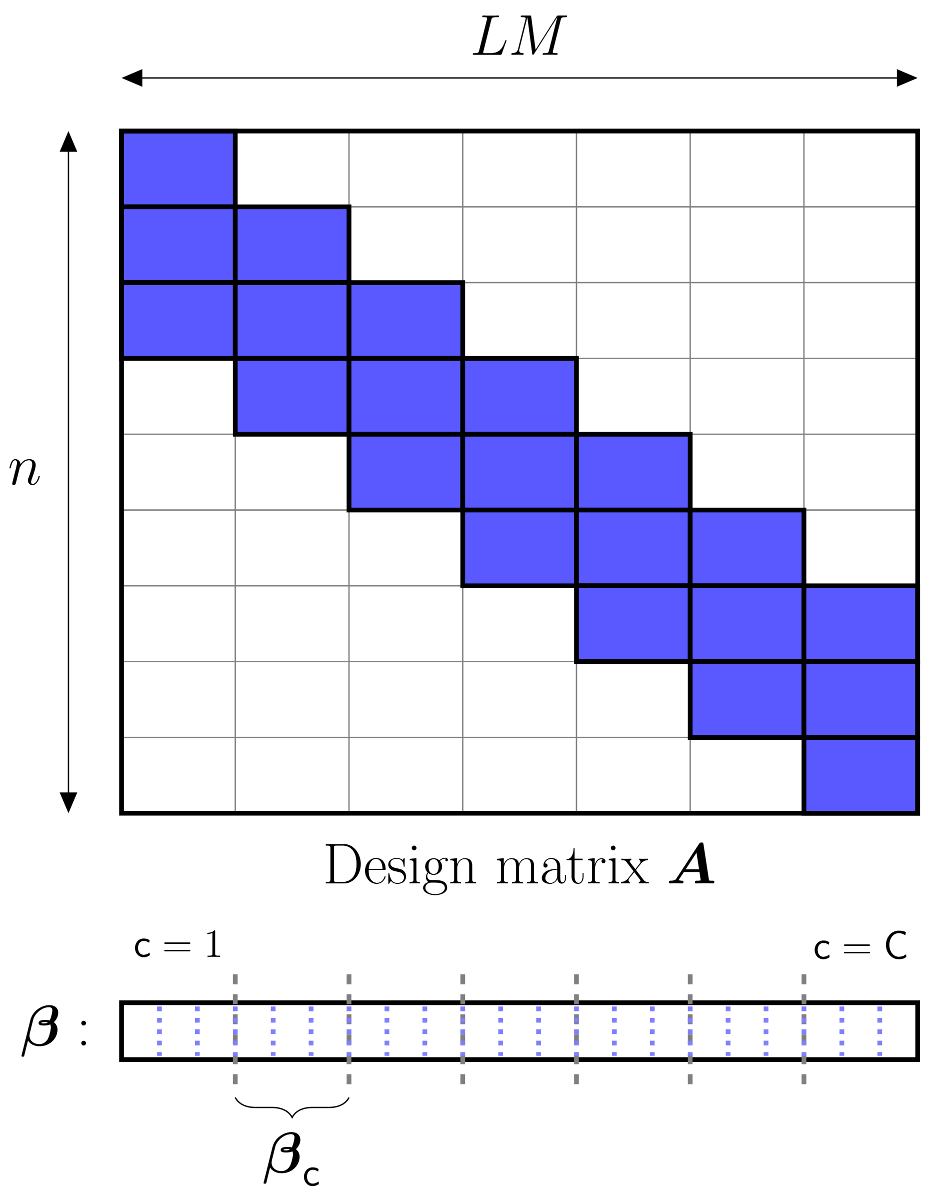
Adaptive, Successive Soft-thresholding
[Cho and Barron ’13]

Approximate Message Passing
[Barbier and Krzakala ’17]
[Rush, Greig and Venkataramanan ’17]

Spatial coupling

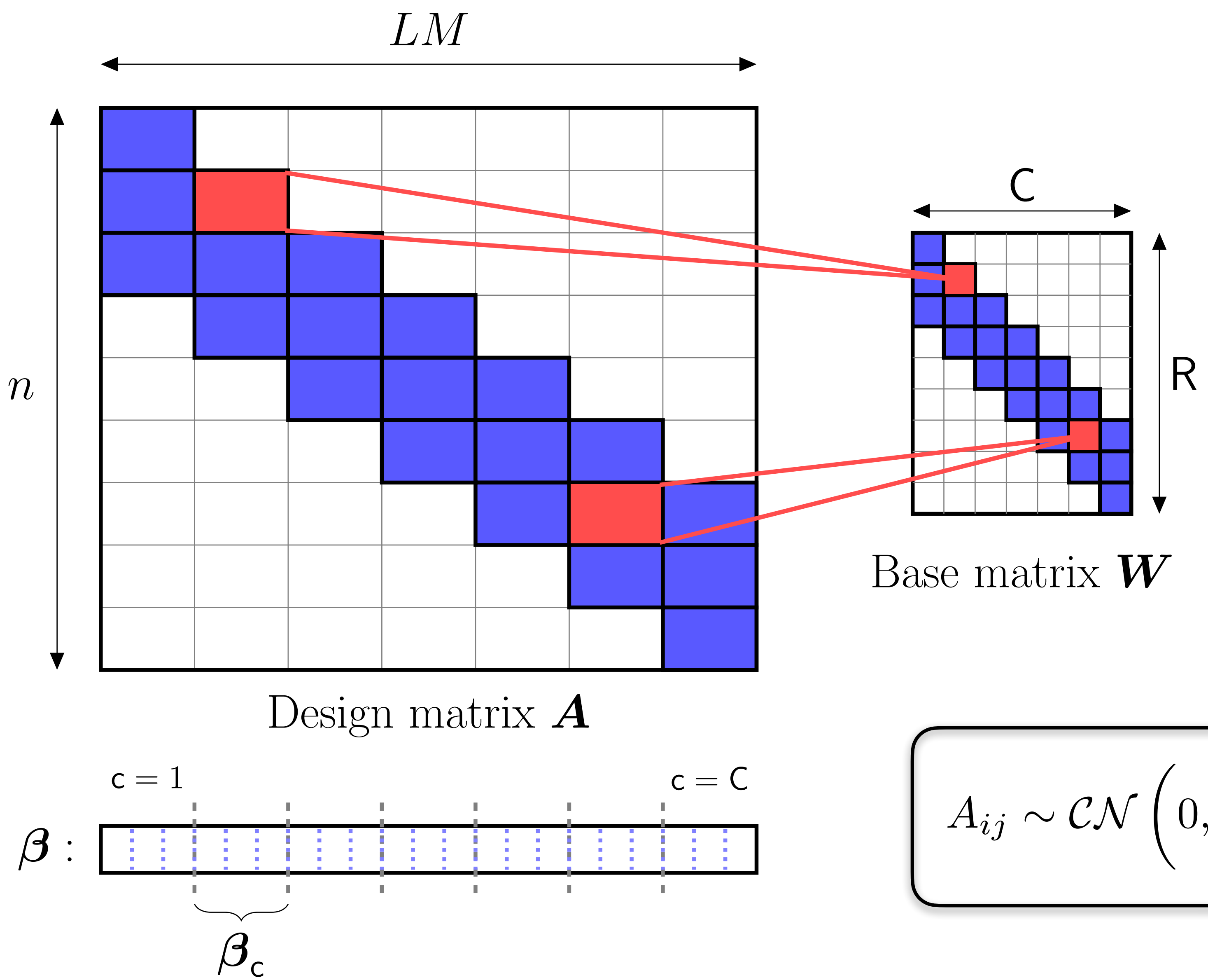
Approximate Message Passing
[Barbier et al. ’14-’19]
[Rush, Hsieh and Venkataramanan ’18, ’19, ’20]

Spatial coupling



[Felstrom and Zigangirov '99] [Kudekar and Pfister '10] [Barbier, Schülke and Krazakala '13, '15] …

Spatial coupling



$$A_{ij} \sim \mathcal{CN}\left(0, \frac{1}{L} W_{\mathsf{r}(i),\mathsf{c}(j)}\right)$$

[Thorpe '03] [Mitchell, Lentmaier, and Costello '15] [Liang, Ma and Ping '17] …

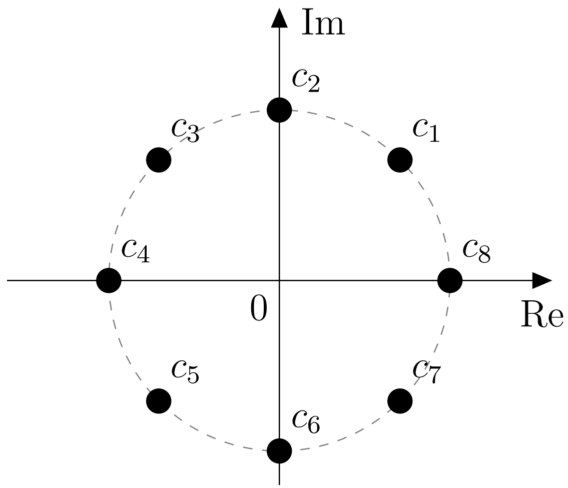
Modulated SPARC encoding $\boldsymbol{x} = \boldsymbol{A\beta}$

$$\boldsymbol{\beta} : \left[\;\ldots 0, \textcolor{red}{a_1}, 0 \;\vdots\; \textcolor{red}{a_2}, 0, 0, \ldots \;\vdots\; \cdots \;\vdots\; 0, \textcolor{red}{a_L}, 0, \ldots \;\right]^\top$$

M entries

$\log_2 M$ bits
determine **location**

$\log_2 K$ bits
determine **value**



E.g. 8-PSK

$$R = \frac{L \log(KM)}{n}$$

K-ary
Phase Shift Keying
(PSK)

AMP decoding $\quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$

Initialise $\widehat{\boldsymbol{\beta}}^0$ to all-zero vector. For $t = 0, 1, 2 \ldots$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}^t + \boldsymbol{v}^t \odot \boldsymbol{z}^{t-1}$$

$$\widehat{\boldsymbol{\beta}}^{t+1} = \eta\left(\widehat{\boldsymbol{\beta}}^t + (\boldsymbol{S}^t \odot \boldsymbol{A})^* \boldsymbol{z}^t \,, \; \boldsymbol{\tau}^t\right)$$

AMP decoding $\quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$

Initialise $\widehat{\boldsymbol{\beta}}^0$ to all-zero vector. For $t = 0, 1, 2 \ldots$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}^t + \boldsymbol{v}^t \odot \boldsymbol{z}^{t-1}$$

$$\widehat{\boldsymbol{\beta}}^{t+1} = \eta\left(\underbrace{\widehat{\boldsymbol{\beta}}^t + (\boldsymbol{S}^t \odot \boldsymbol{A})^* \boldsymbol{z}^t}_{\approx\, \boldsymbol{\beta} \,+\, \text{Gaussian noise}}\,,\ \boldsymbol{\tau}^t\right)$$

Effective noise variance (← $\boldsymbol{\tau}^t$)

AMP decoding $\quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$

Initialise $\widehat{\boldsymbol{\beta}}^0$ to all-zero vector. For $t = 0, 1, 2 \ldots$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}^t + \boldsymbol{v}^t \odot \boldsymbol{z}^{t-1}$$

$$\widehat{\boldsymbol{\beta}}^{t+1} = \eta\left(\underbrace{\widehat{\boldsymbol{\beta}}^t + (\boldsymbol{S}^t \odot \boldsymbol{A})^* \boldsymbol{z}^t}_{\approx\, \boldsymbol{\beta} \,+\, \text{Gaussian noise}}, \ \boldsymbol{\tau}^t\right)$$

Effective noise variance ($\boldsymbol{\tau}^t$)

Bayes-optimal estimator (η)

$$\eta_j(\boldsymbol{s}, \boldsymbol{\tau}) = \mathbb{E}\left[\beta_j \,\middle|\, \boldsymbol{s} = \boldsymbol{\beta} + \sqrt{\boldsymbol{\tau}} \odot \boldsymbol{u}\right]$$

$\boldsymbol{u}$: standard normal random vector

AMP decoding $\quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{w}$

Initialise $\widehat{\boldsymbol{\beta}}^0$ to all-zero vector. For $t = 0, 1, 2 \ldots$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}^t + \boldsymbol{v}^t \odot \boldsymbol{z}^{t-1}$$

Effective noise variance

$$\widehat{\boldsymbol{\beta}}^{t+1} = \eta\left(\underbrace{\widehat{\boldsymbol{\beta}}^t + (\boldsymbol{S}^t \odot \boldsymbol{A})^* \boldsymbol{z}^t}_{\approx\, \boldsymbol{\beta} \,+\, \text{Gaussian noise}},\ \boldsymbol{\tau}^t\right)$$

Bayes-optimal estimator

$$\eta_j(\boldsymbol{s}, \boldsymbol{\tau}) = \mathbb{E}\left[\beta_j \,\middle|\, \boldsymbol{s} = \boldsymbol{\beta} + \sqrt{\boldsymbol{\tau}} \odot \boldsymbol{u}\right]$$

$\boldsymbol{u}$: standard normal random vector

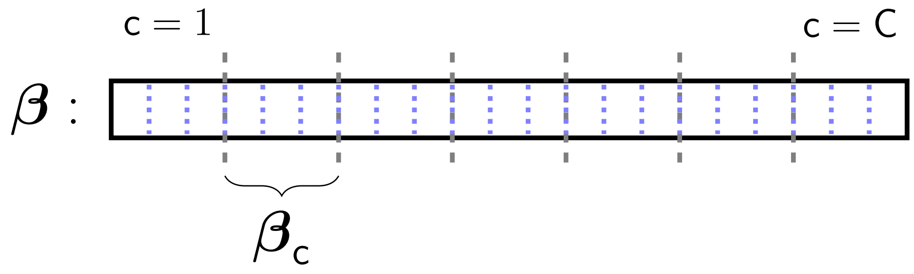
State evolution predicts

$$\|\widehat{\boldsymbol{\beta}}^t - \boldsymbol{\beta}\|^2$$

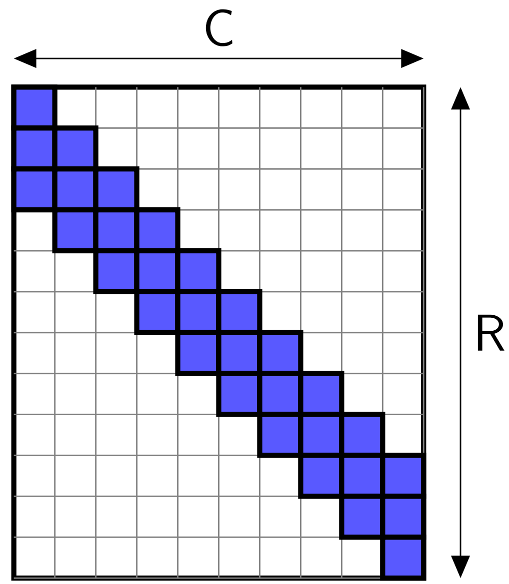
State evolution for K-PSK modulated SPARCs

For large n and L

$$\psi_{\mathsf{c}}^{t} \approx \frac{\|\widehat{\boldsymbol{\beta}}_{\mathsf{c}}^{t} - \boldsymbol{\beta}_{\mathsf{c}}\|^2}{L/\mathsf{C}}$$



E.g.

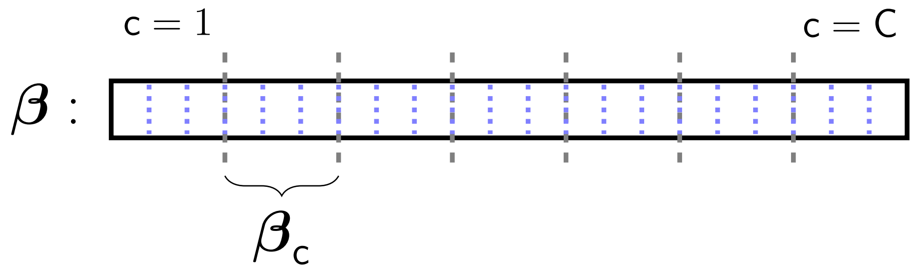


Base matrix $\boldsymbol{W}$

State evolution for K-PSK modulated SPARCs

For large n and L

$$\psi_{\mathsf{c}}^{t} \approx \frac{\|\widehat{\boldsymbol{\beta}}_{\mathsf{c}}^{t} - \boldsymbol{\beta}_{\mathsf{c}}\|^2}{L/\mathsf{C}}$$



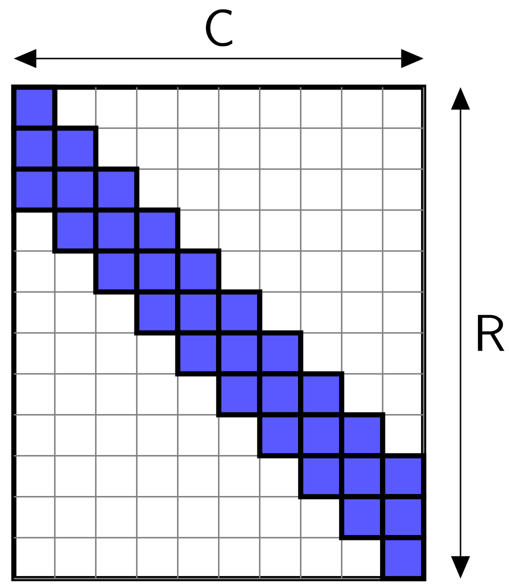
Initialise $\psi_{\mathsf{c}}^{0} = 1$ for $\mathsf{c} = 1, \ldots, \mathsf{C}$. For $t = 0, 1, 2 \ldots$

$$\phi_{\mathsf{r}}^{t} = \sigma^2 + \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \psi_{\mathsf{c}}^{t},$$

$$\tau_{\mathsf{c}}^{t} = \frac{R/2}{\log(KM)} \left[\frac{1}{\mathsf{R}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \frac{W_{\mathsf{rc}}}{\phi_{\mathsf{r}}^{t}} \right]^{-1},$$

$$\psi_{\mathsf{c}}^{t+1} = \mathsf{mmse}_{\beta}\left(\tau_{\mathsf{c}}^{t}\right)$$

E.g.



Base matrix $\boldsymbol{W}$

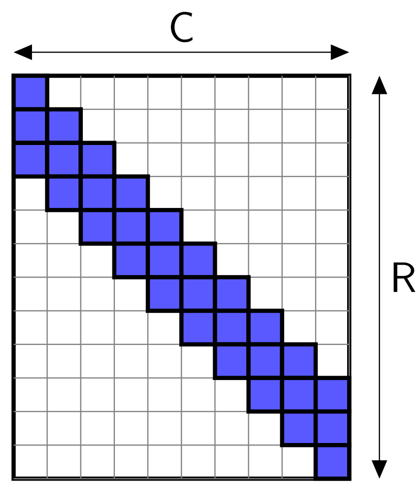
Initialise $\psi_{\mathsf{c}}^0 = 1$ for $\mathsf{c} = 1, \dots, \mathsf{C}$. For $t = 0, 1, 2 \dots$

$$\phi_{\mathsf{r}}^t = \sigma^2 + \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \psi_{\mathsf{c}}^t,$$

$$\tau_{\mathsf{c}}^t = \frac{R/2}{\log(KM)} \left[\frac{1}{\mathsf{R}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \frac{W_{\mathsf{rc}}}{\phi_{\mathsf{r}}^t} \right]^{-1},$$

$$\psi_{\mathsf{c}}^{t+1} = \mathsf{mmse}_{\beta}\left(\tau_{\mathsf{c}}^t\right)$$

E.g.



Base matrix $\boldsymbol{W}$

For $\delta \in (0, \frac{1}{2})$ and $\nu_{\mathsf{c}}^t = \frac{1}{\tau_{\mathsf{c}}^t \log(KM)}$,

Main result

$$\psi_{\mathsf{c}}^{t+1} \leq \begin{cases} \dfrac{(KM)^{-\alpha_1 K \delta^2}}{\delta \sqrt{\log(KM)}} & \text{if } \nu_{\mathsf{c}}^t > 2 + \delta, \\ 1 + \dfrac{(KM)^{-\alpha_2 K \nu_{\mathsf{c}}^t}}{\sqrt{\nu_{\mathsf{c}}^t \log(KM)}} & \text{otherwise.} \end{cases}$$

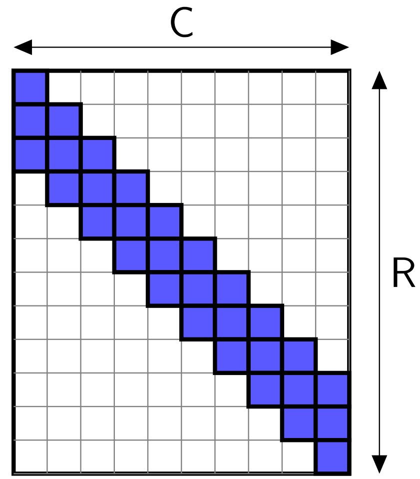
Initialise $\psi_c^0 = 1$ for $c = 1, \ldots, C$. For $t = 0, 1, 2 \ldots$

$$\phi_r^t = \sigma^2 + \frac{1}{C}\sum_{c=1}^{C} W_{rc}\psi_c^t,$$

$$\tau_c^t = \frac{R/2}{\log(KM)}\left[\frac{1}{R}\sum_{r=1}^{R}\frac{W_{rc}}{\phi_r^t}\right]^{-1},$$

$$\psi_c^{t+1} = \mathsf{mmse}_\beta\left(\tau_c^t\right)$$

E.g.



Base matrix $\boldsymbol{W}$

For $\delta \in (0, \frac{1}{2})$ and $\nu_c^t = \frac{1}{\tau_c^t \log(KM)}$,

Main result

$$\psi_c^{t+1} \le \begin{cases} \dfrac{(KM)^{-\alpha_1 K^{\delta^2}}}{\delta\sqrt{\log(KM)}} \\[2em] 1 + \dfrac{(KM)^{-\alpha_2 K^{\nu_c^t}}}{\sqrt{\nu_c^t \log(KM)}} \end{cases} \xrightarrow{\text{fixed } K \text{ and } M\to\infty} \begin{cases} 0 & \text{if } \nu_c^t > 2, \\[2em] 1 & \text{otherwise.} \end{cases}$$

Asymptotic SE for K-PSK modulated SPARCs

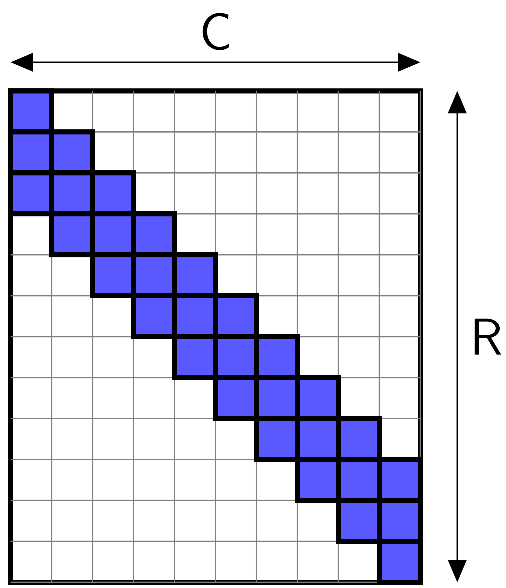
For fixed K, as $M \to \infty$ the state evolution simplifies to:

Initialise $\psi_{\mathsf{c}}^0 = 1$ for $\mathsf{c} = 1, \dots, \mathsf{C}$. For $t = 0, 1, 2 \dots$

$$\phi_{\mathsf{r}}^t = \sigma^2 + \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \psi_{\mathsf{c}}^t,$$

$$\psi_{\mathsf{c}}^{t+1} = \mathbb{1}\left\{ \frac{1}{\mathsf{R}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \frac{W_{\mathsf{rc}}}{\phi_{\mathsf{r}}^t} \le R \right\}.$$

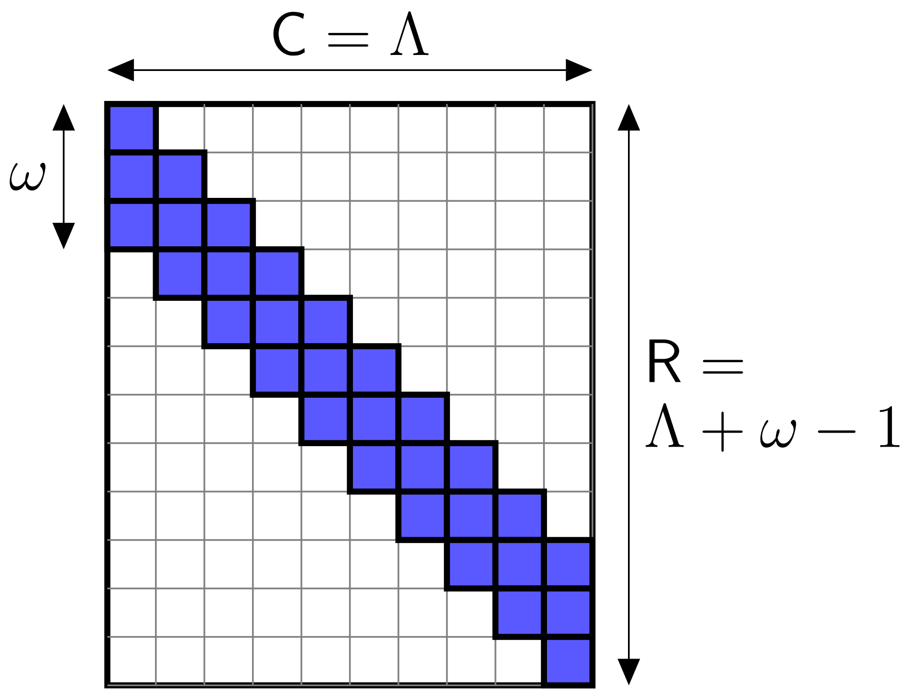
E.g.



Base matrix $\boldsymbol{W}$

Does not depend on K

Theorem for K-PSK modulated SPARCs



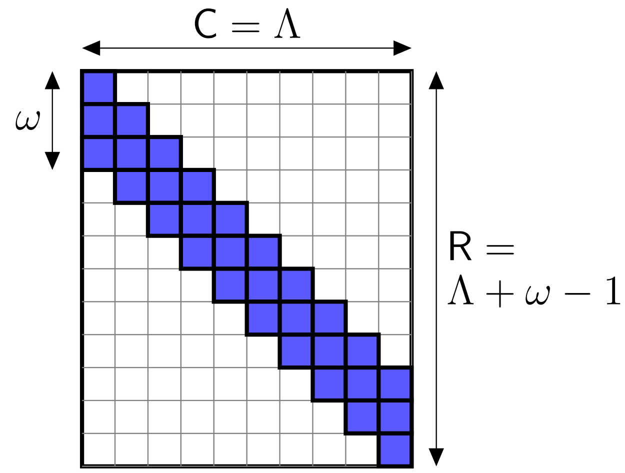
(ω, Λ) base matrix $\boldsymbol{W}$

Theorem for K-PSK modulated SPARCs

Consider a K-PSK modulated complex SPARC constructed with an (ω, Λ) base matrix $\boldsymbol{W}$ with $\omega > \omega^\star$ and rate satisfying $R < \tilde{\mathcal{C}} := \mathcal{C}/(1 + \frac{\omega-1}{\Lambda})$.

As $n \to \infty$, the SER of the AMP decoder after T iterations $= 0$ almost surely, where

$$T \propto \frac{\Lambda}{2\omega(\tilde{\mathcal{C}} - R)}.$$



(ω, Λ) base matrix $\boldsymbol{W}$

Steps of proof

1. Error rate of AMP accurately predicted by state evolution for large code lengths.

   By extending results in [Rush, Hsieh and Venkataramanan '20].

2. For any $R < C$, state evolution predicts vanishing error probability in the large system limit.

   A. Asymptotic state evolution is the same for any K.

      Shown in this work.

   B. Use asymptotic state evolution analysis from unmodulated ($K = 1$) SPARCs.

      Shown in [Rush, Hsieh and Venkataramanan '20].

Steps of proof

1. Error rate of AMP accurately predicted by state evolution for large code lengths.

   By extending results in [Rush, Hsieh and Venkataramanan '20].

2. For any $R < \mathcal{C}$, state evolution predicts vanishing error probability in the large system limit.

   A. Asymptotic state evolution is the same for any K.

   Shown in this work.

   B. Use asymptotic state evolution analysis from unmodulated ($K = 1$) SPARCs.

   Shown in [Rush, Hsieh and Venkataramanan '20].

Steps of proof

1. Error rate of AMP accurately predicted by state evolution for large code lengths.

   By extending results in [Rush, Hsieh and Venkataramanan '20].

2. For any $R < \mathcal{C}$, state evolution predicts vanishing error probability in the large system limit.

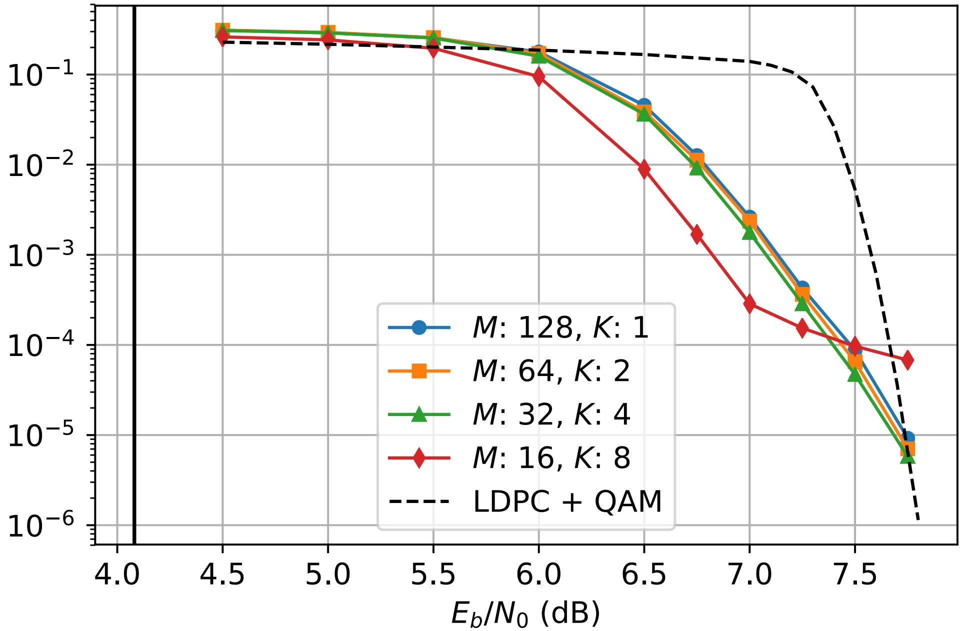
   A. Asymptotic state evolution is the same for any K.

      Shown in this work.

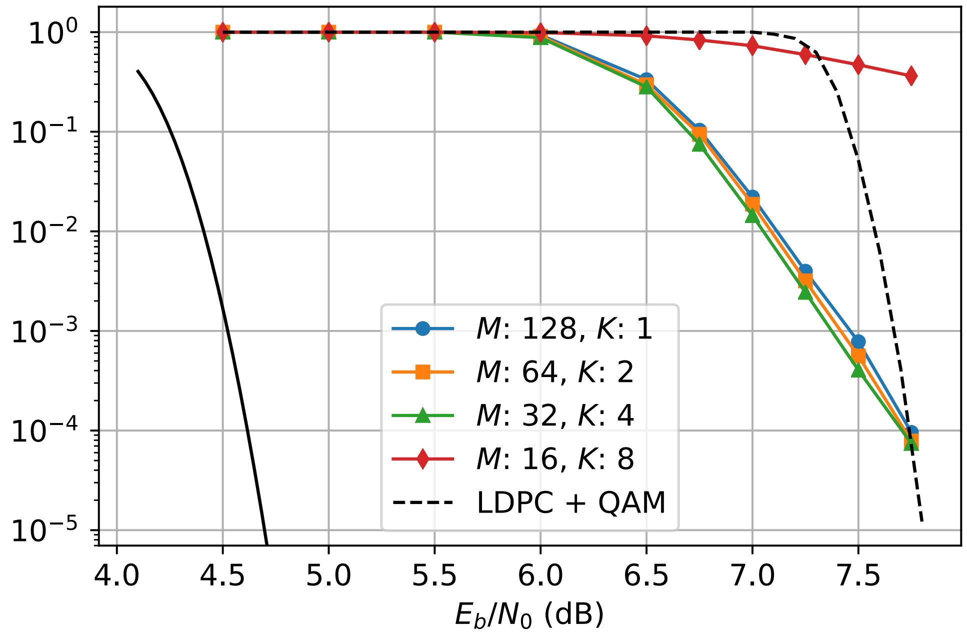
   B. Use asymptotic state evolution analysis from unmodulated ($K = 1$) SPARCs.

      Shown in [Rush, Hsieh and Venkataramanan '18, '20].

Simulation results



Bit error rate



Codeword error rate

$R = 1.6$ bits/dim.

$n \approx 2000$

$L = 960$

$\omega = 6, \Lambda = 32$

$$R = \frac{L \log(KM)}{n}$$

Coded modulation

(6480, 16200) LDPC

DVB-S2 standard

+256 QAM

Computational benefits of modulation

Per iteration complexity AMP decoder (FFT based)

$$O(LM(\log(LM) + K))$$

Let $M_{\text{unmod}} = KM_{\text{mod}}$, then

$$\frac{\text{complexity for unmodulated SPARC}}{\text{complexity for modulated SPARC}} = K \cdot \frac{\log(LM_{\text{unmod}}) + 1}{\log(LM_{\text{unmod}}) + K - \log K}$$

If $K \ll \log(LM_{\text{unmod}})$, approx. K times reduction

(approx. 3.8x in simulation example using $K = 4$)

Background

Sparse regression codes (SPARCs) for the AWGN channel

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\beta}$$

This work

1. SPARCs for the **complex** AWGN channel.
2. Introduce (PSK) **modulation** to SPARC encoding.

Theoretical result

Complex SPARCs with K-PSK modulation
are asymptotically capacity achieving for any fixed K

Numerical result

Modulation can significantly reduce complexity
without sacrificing error performance.