Foundations of Distributed Signal Processing

Michael C. Gastpar

EPFL and University of California, Berkeley

September 4, 2012

Outline

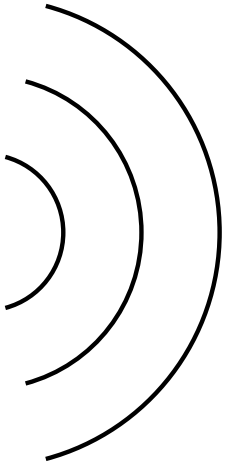
Signal Processing

Distributed Signal Processing

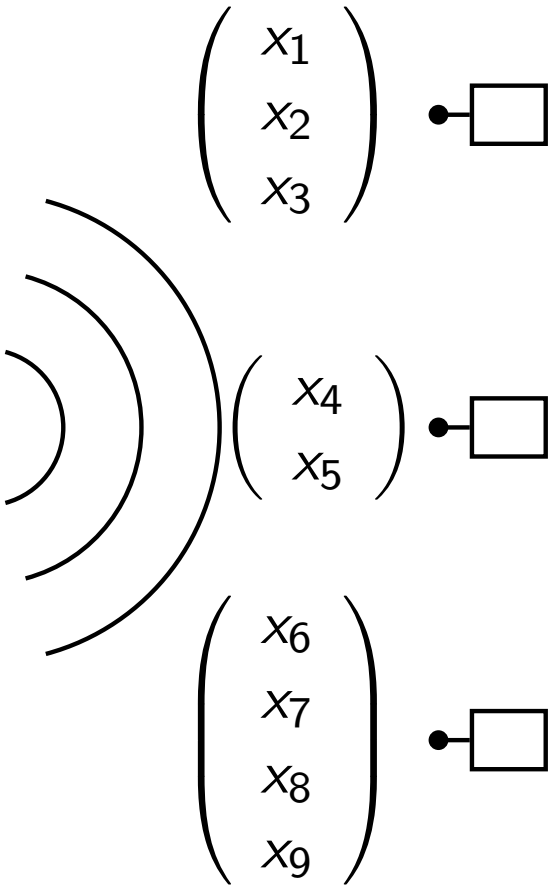
Case Study 1: Transform Coding

Case Study 2: A Wireless Sensor Network Scenario

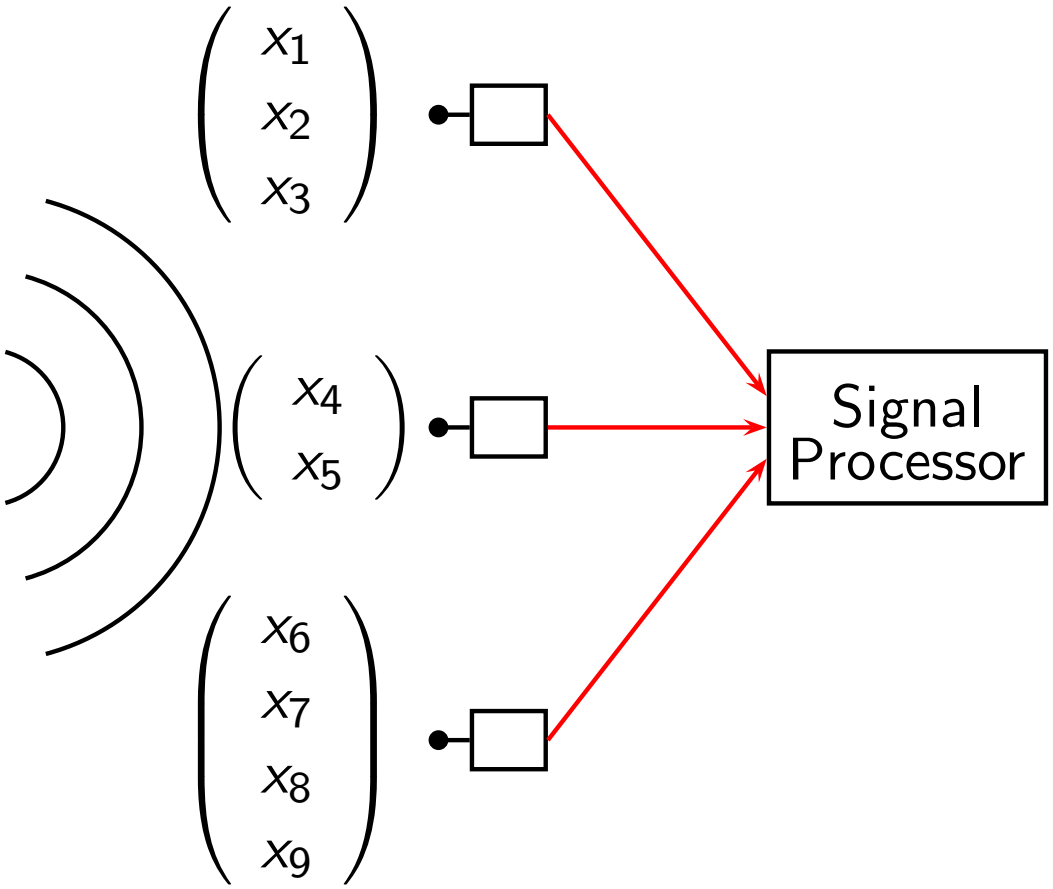
Signal Processing



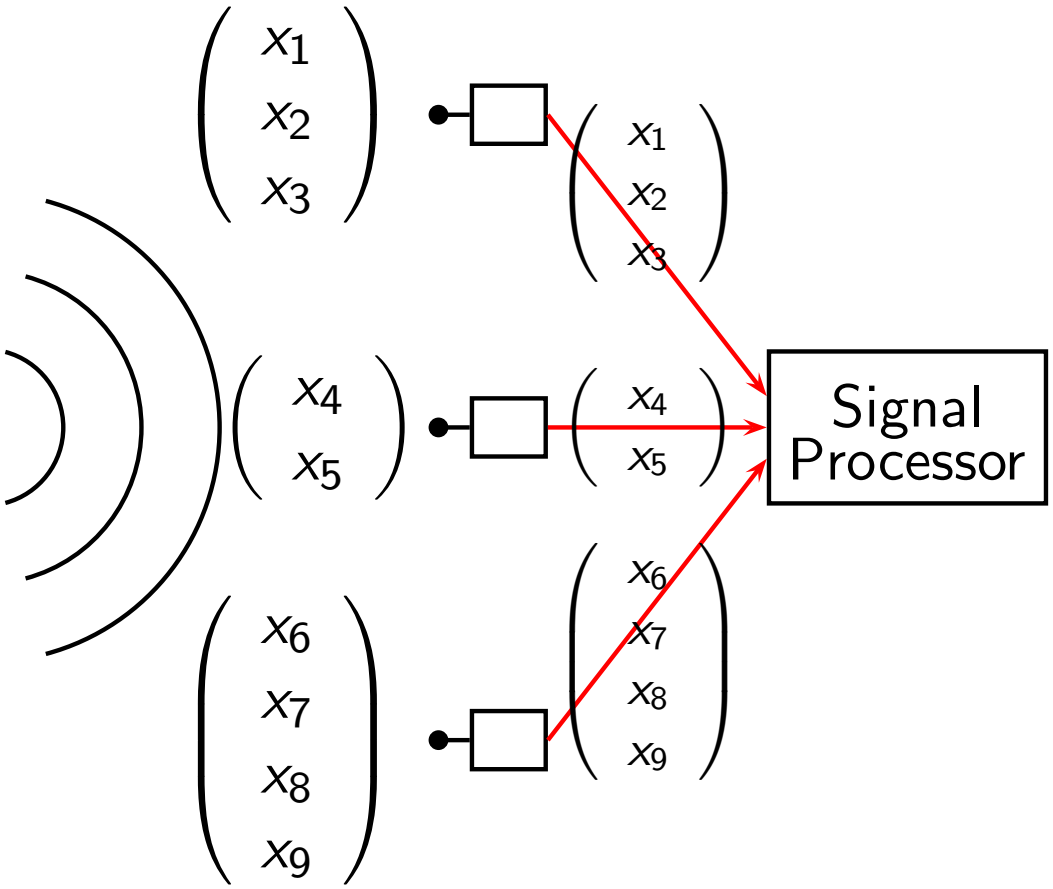
Signal Processing



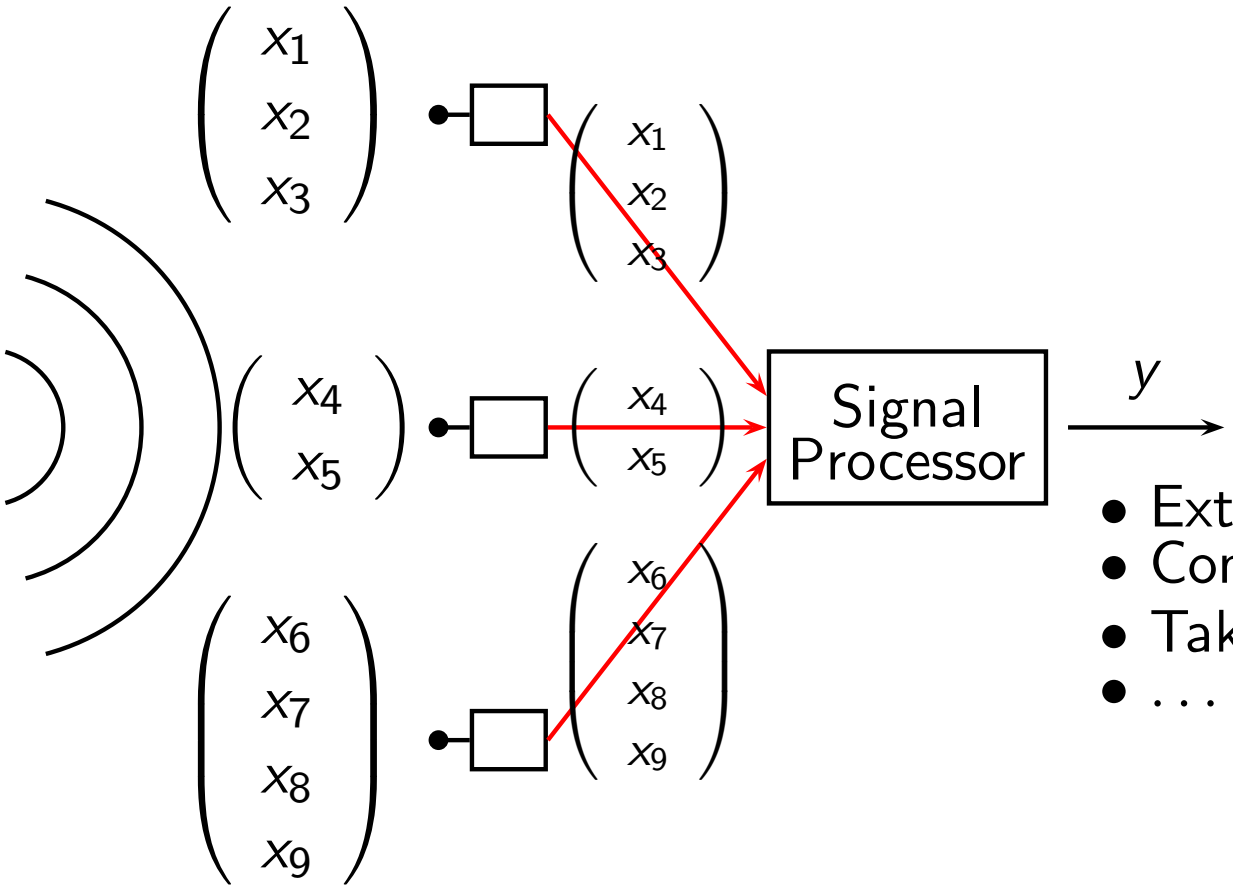
Signal Processing



Signal Processing

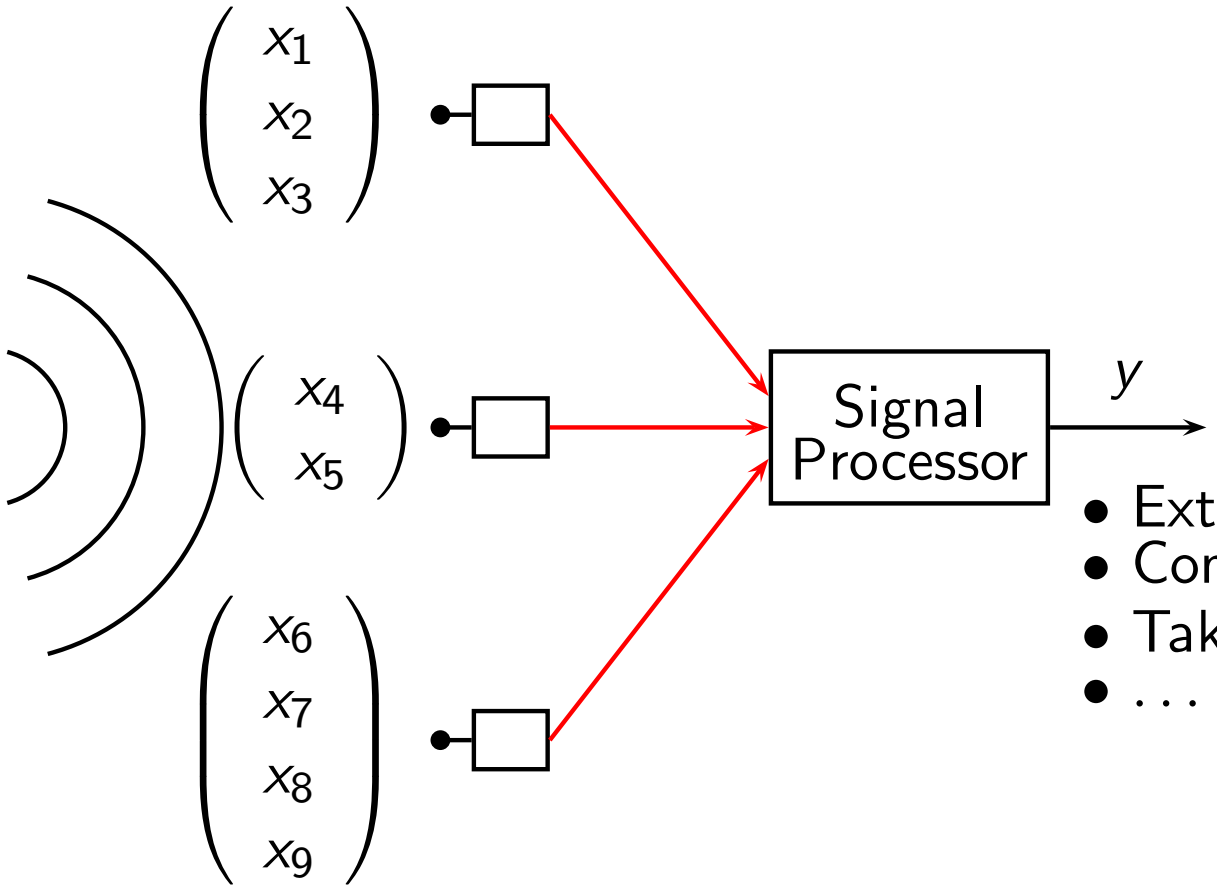


Signal Processing



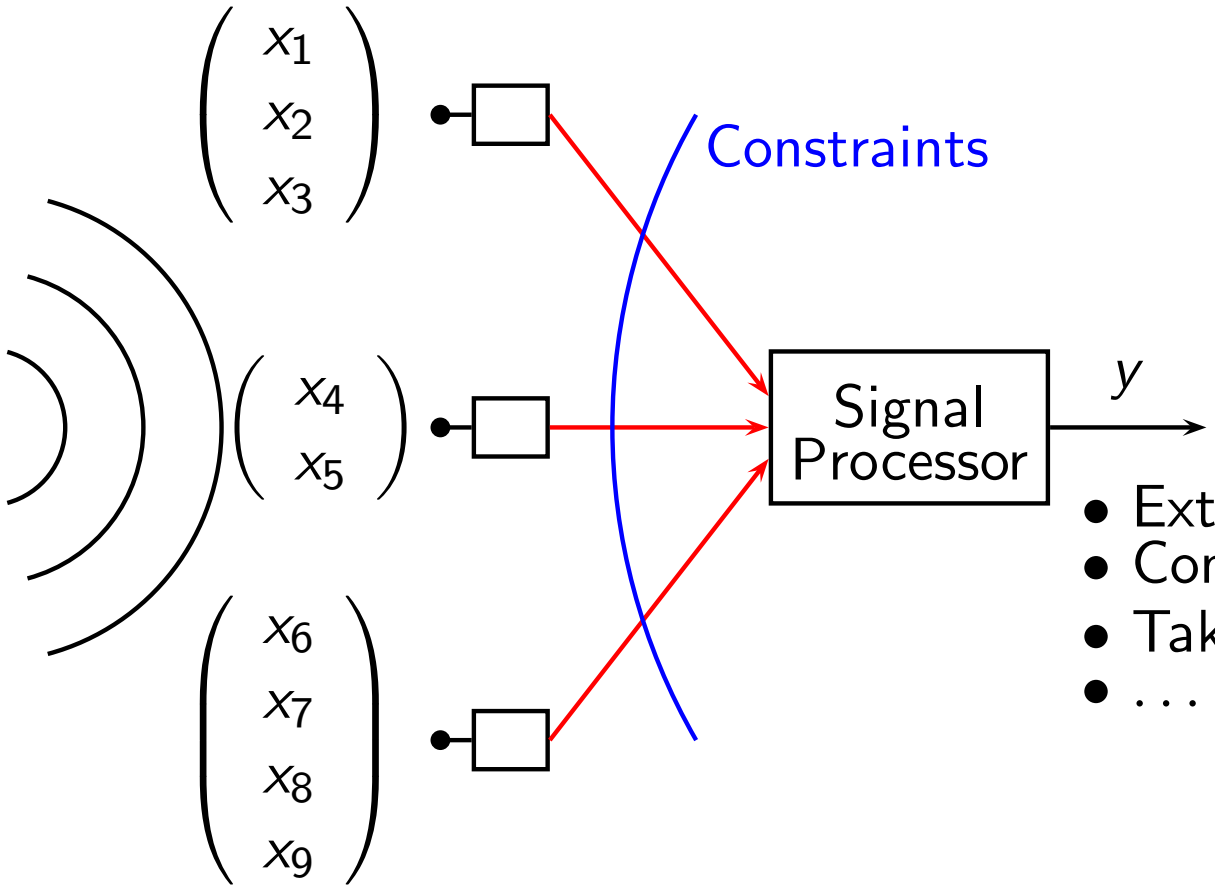
- Extract a feature
- Compress the signal
- Take a decision
- ...

Distributed Signal Processing



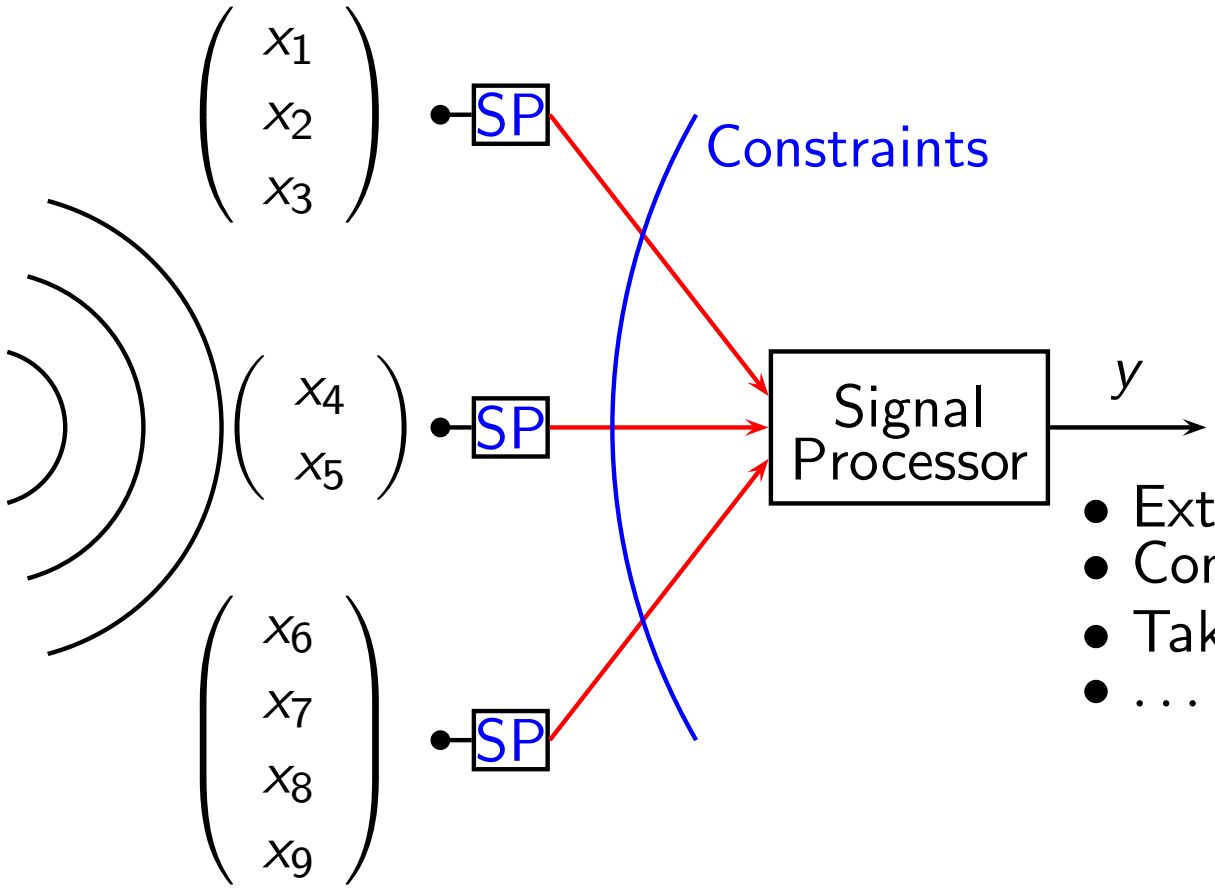
- Extract a feature
- Compress the signal
- Take a decision
- ...

Distributed Signal Processing



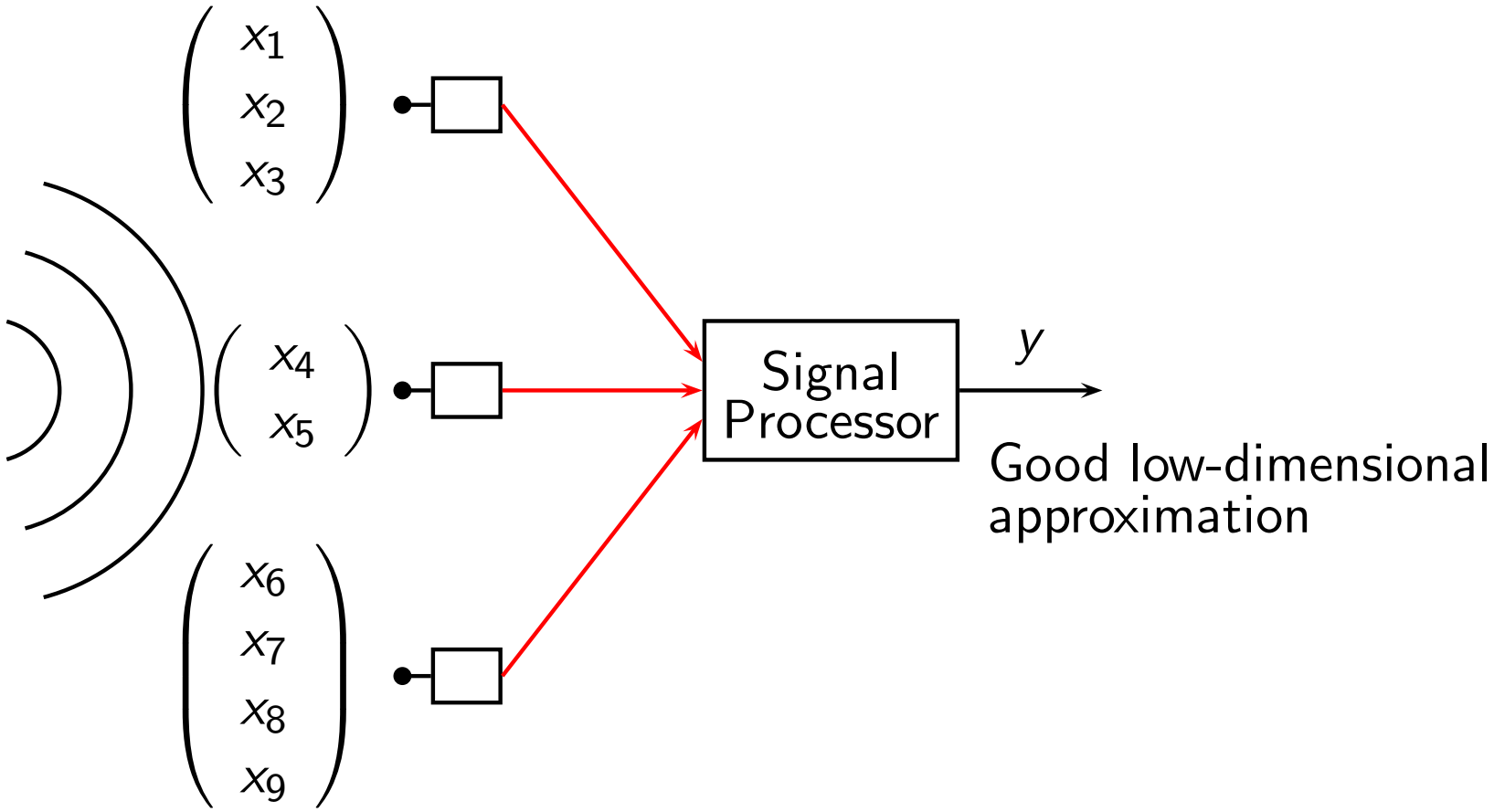
- Extract a feature
- Compress the signal
- Take a decision
- ...

Distributed Signal Processing

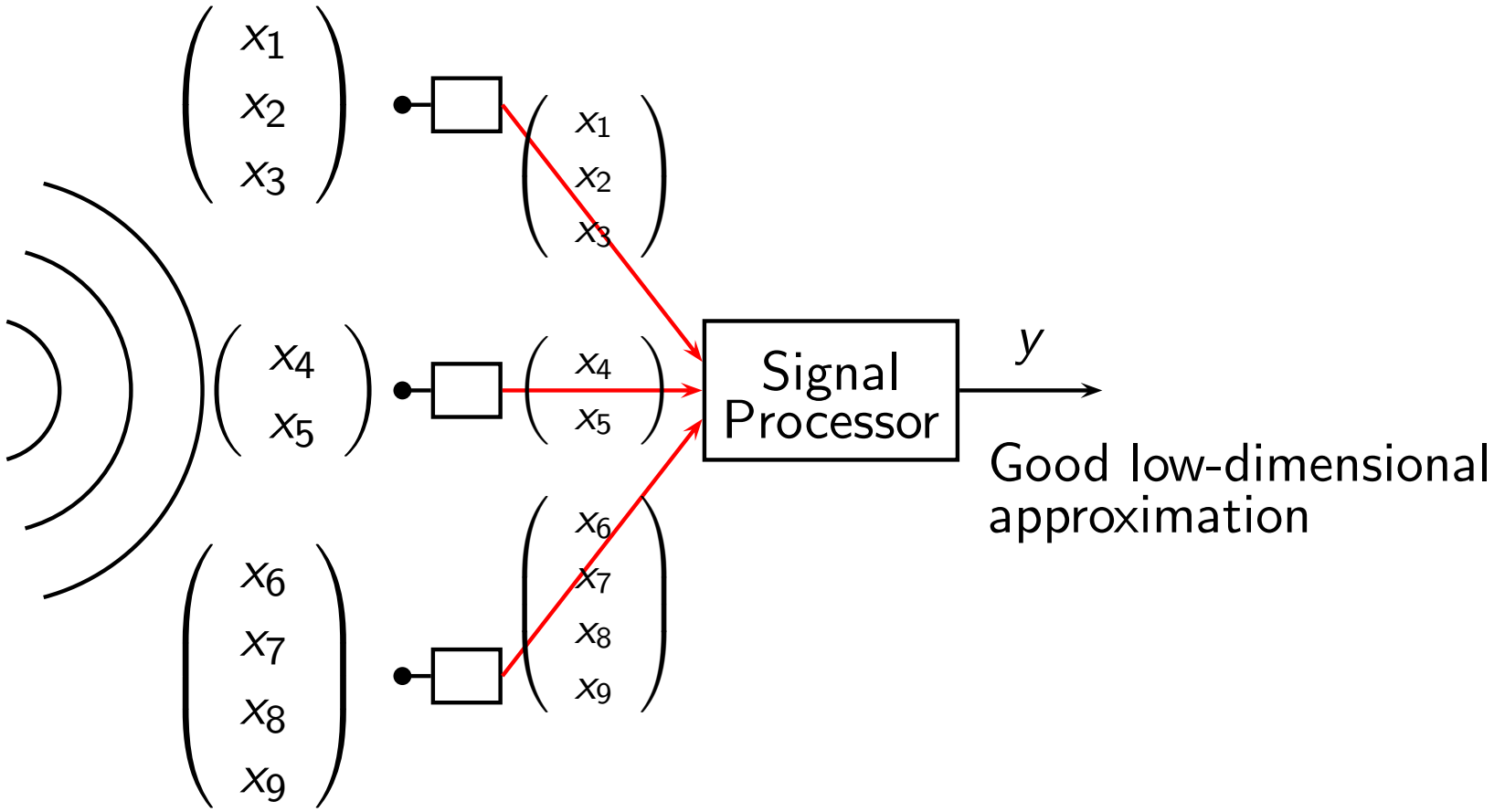


- Extract a feature
- Compress the signal
- Take a decision
- ...

Case Study 1: Transform Coding



Case Study 1: Transform Coding



Case Study 1: Transform Coding

Formal Problem Statement:

- Suppose you only get to keep *linear combinations* of your observations:

$$\begin{aligned} y_1 &= \alpha_{11}x_1 + \cdots + \alpha_{1N}x_N \\ \vdots &\quad \vdots \quad \vdots \\ y_K &= \alpha_{K1}x_1 + \cdots + \alpha_{KN}x_N \end{aligned}$$

  or, equivalently, $\mathbf{Y} = A\mathbf{X}$.

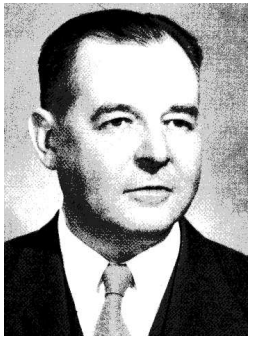
- Want to minimize the *mean-squared error*

$$\mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}(Y_1, \ldots, Y_K)\|^2\right].$$

- How to optimally select the coefficients α_{ij} (i.e., the matrix A)?

Case Study 1: Transform Coding

The solution was found by Hotelling (1933), Karhunen (1947), and Loève (1948).



Hotelling

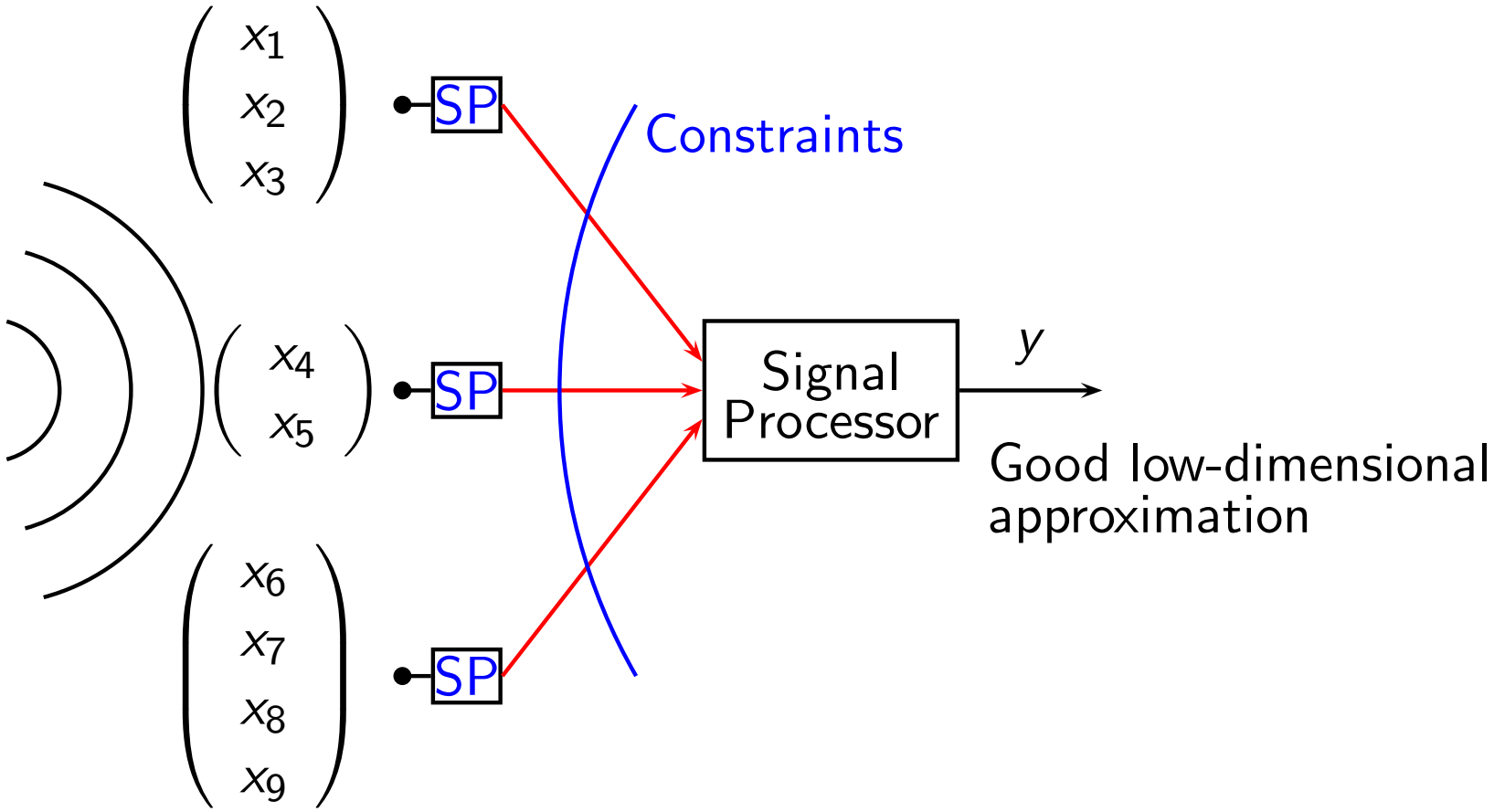
Loève

It takes the shape of an *eigendecomposition.*

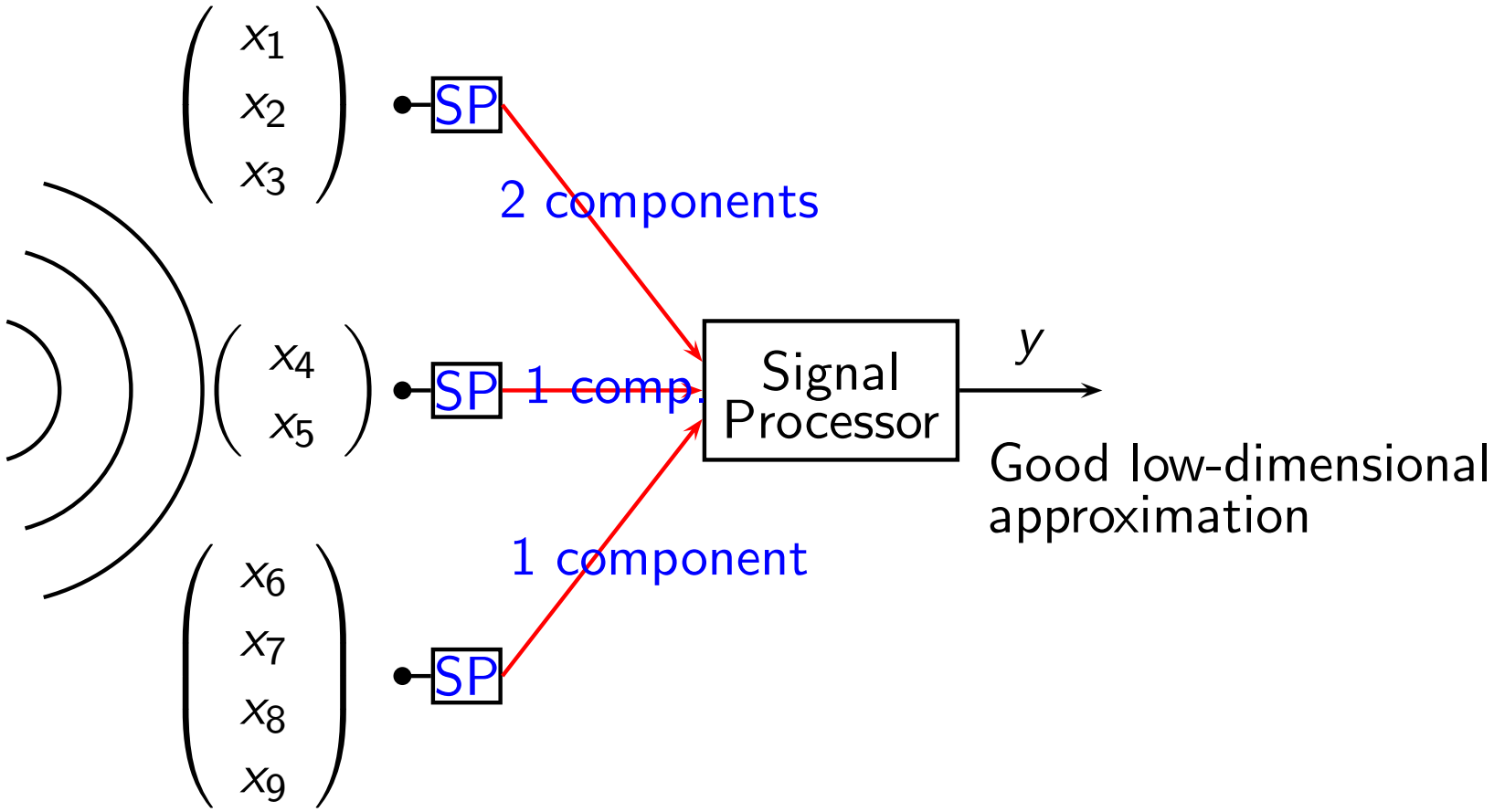
Specifically, let Σ_x be the covariance matrix of the vector $\mathbf{X} = (x_1, x_2, \ldots, x_N)$. Then, the best *linear* transform is given by the eigenvectors of Σ_x :

$$\Sigma_x = Q \Lambda Q^H.$$

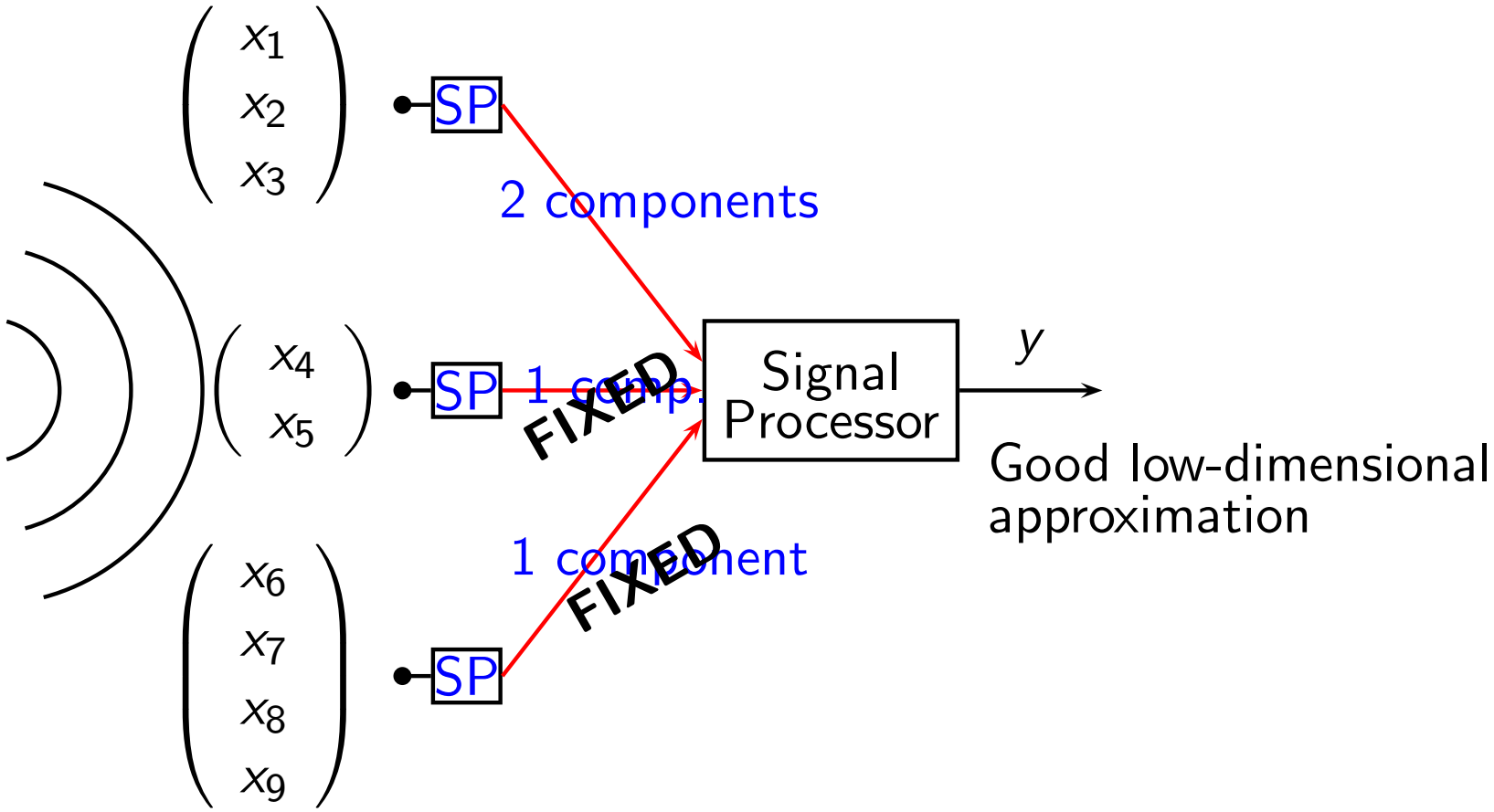
Case Study 1: *Distributed* Transform Coding



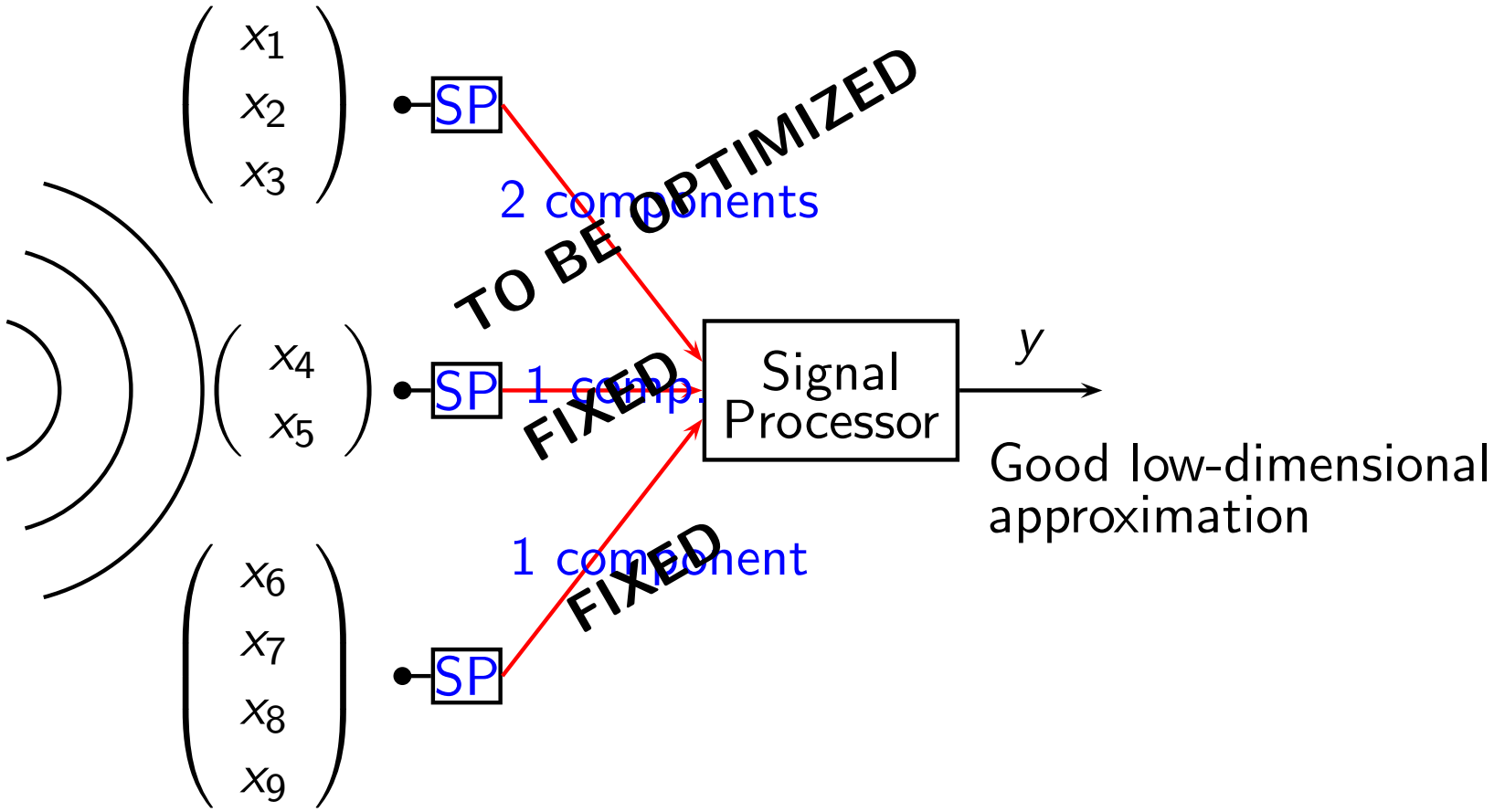
Case Study 1: *Distributed* Transform Coding



Case Study 1: *Distributed* Transform Coding



Case Study 1: *Distributed* Transform Coding



The Distributed Karhunen-Loève Transform

Theorem (Local KLT)

If all else is fixed, the best transform for terminal 1 is given by

$$C_1 = \left(Q_w^{(M)}\right)^T \begin{pmatrix} I_M \\ A_1 \end{pmatrix}.$$

where A_1 consists of the first M columns of

$$A = \begin{pmatrix} \Sigma_{12}^T & \Sigma_2 C_2^T \end{pmatrix} \begin{pmatrix} \Sigma_1 & \Sigma_{12} C_2^T \\ C_2 \Sigma_{12}^T & C_2 \Sigma_2 C_2^T + \Sigma_z \end{pmatrix}^{-1},$$

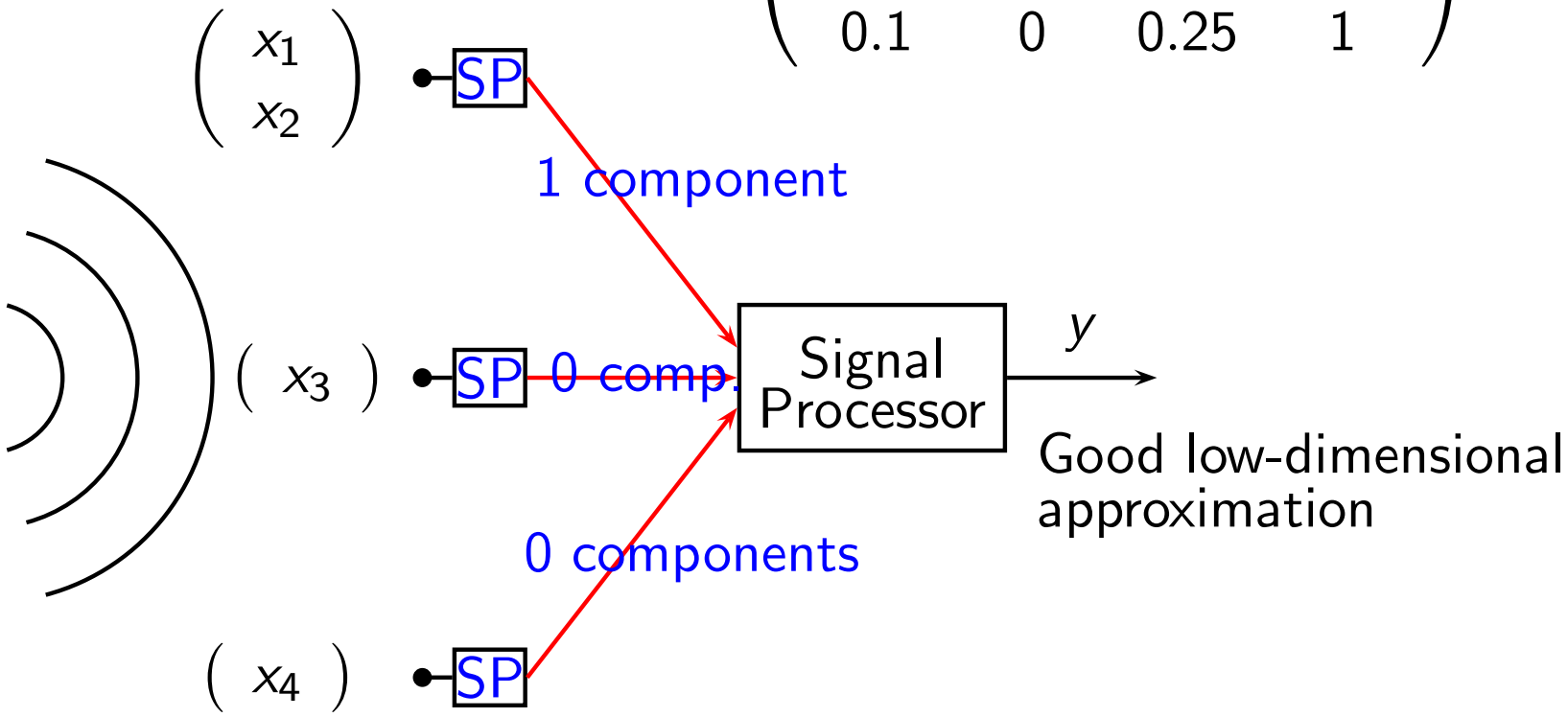
and Q_w is the eigendecomposition of

$$\Sigma_w = \begin{pmatrix} I_M \\ A_1 \end{pmatrix} \left(\Sigma_1 - \Sigma_{12} C_2^T (C_2 \Sigma_2 C_2^T + \Sigma_z)^{-1} C_2 \Sigma_{12}^T\right) \begin{pmatrix} I_M & A_1^T \end{pmatrix}.$$

The Distributed Karhunen-Loève Transform

Example 1 of the Local KLT:

$$\Sigma_x = \begin{pmatrix} 0.11 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}$$



The Distributed Karhunen-Loève Transform

Example 1 of the Local KLT: Let

$$\Sigma_x = \begin{pmatrix} 0.11 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}.$$

According to the standard KLT, Terminal 1 should provide:

$$Y_1 = X_1.$$

However, terminals 2 and 3 provide *nothing*. Therefore, terminal 1's best choice is:

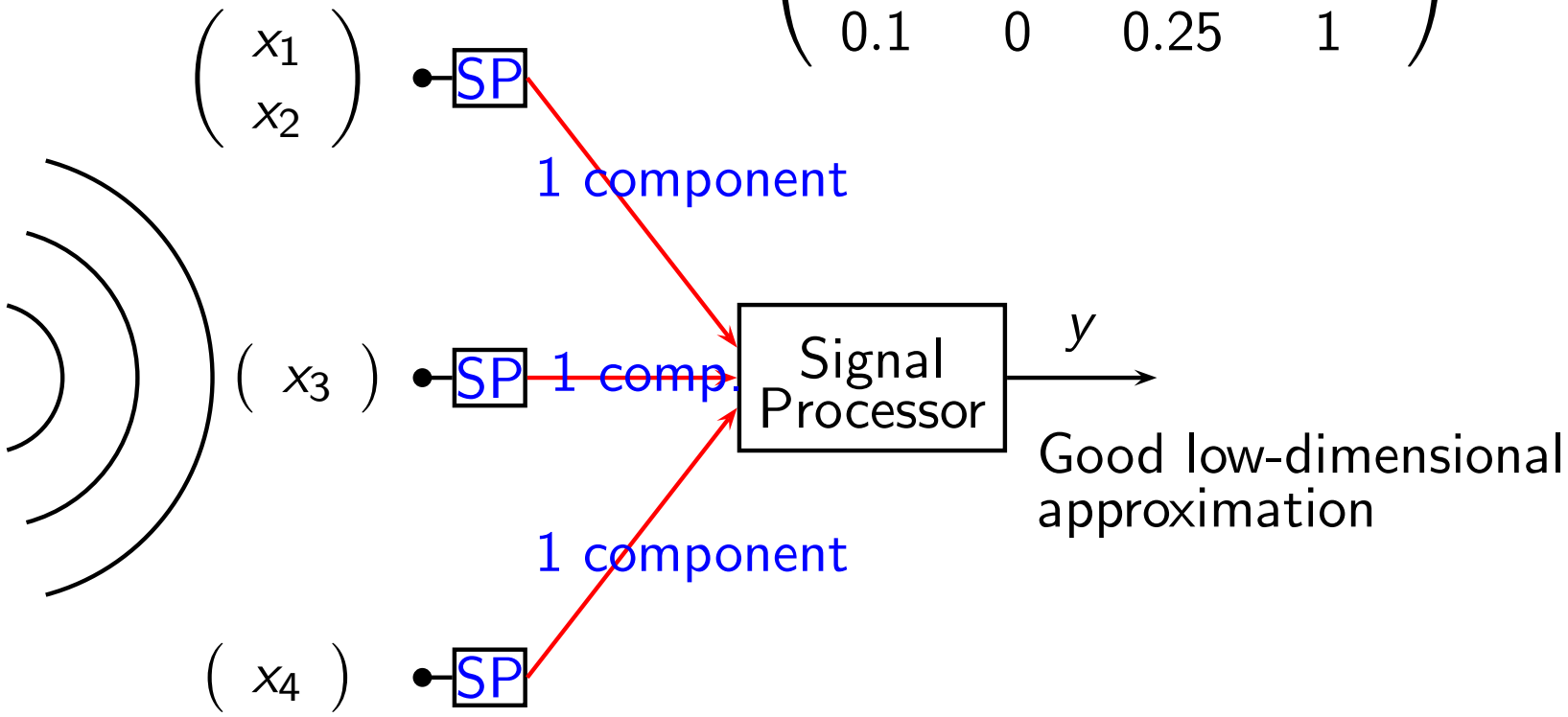
$$Y_1 \approx 1.1119X_1 + 2.6353X_2$$

This leads to a reduction of about 30% in MSE.

The Distributed Karhunen-Loève Transform

Example 2 of the Local KLT:

$$\Sigma_x = \begin{pmatrix} 0.1 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}$$



The Distributed Karhunen-Loève Transform

Example 2 of the Local KLT: Let

$$\Sigma_x = \begin{pmatrix} 0.1 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}.$$

According to the standard KLT, Terminal 1 should provide:

$$Y_1 = X_1.$$

However, terminals 2 and 3 provides *everything*. Therefore, Terminal 1's best choice is:

$$Y_1 \approx -0.9447X_1 + 0.3280X_2.$$

This leads to a reduction of about 65% in MSE.

The Distributed Karhunen-Loève Transform

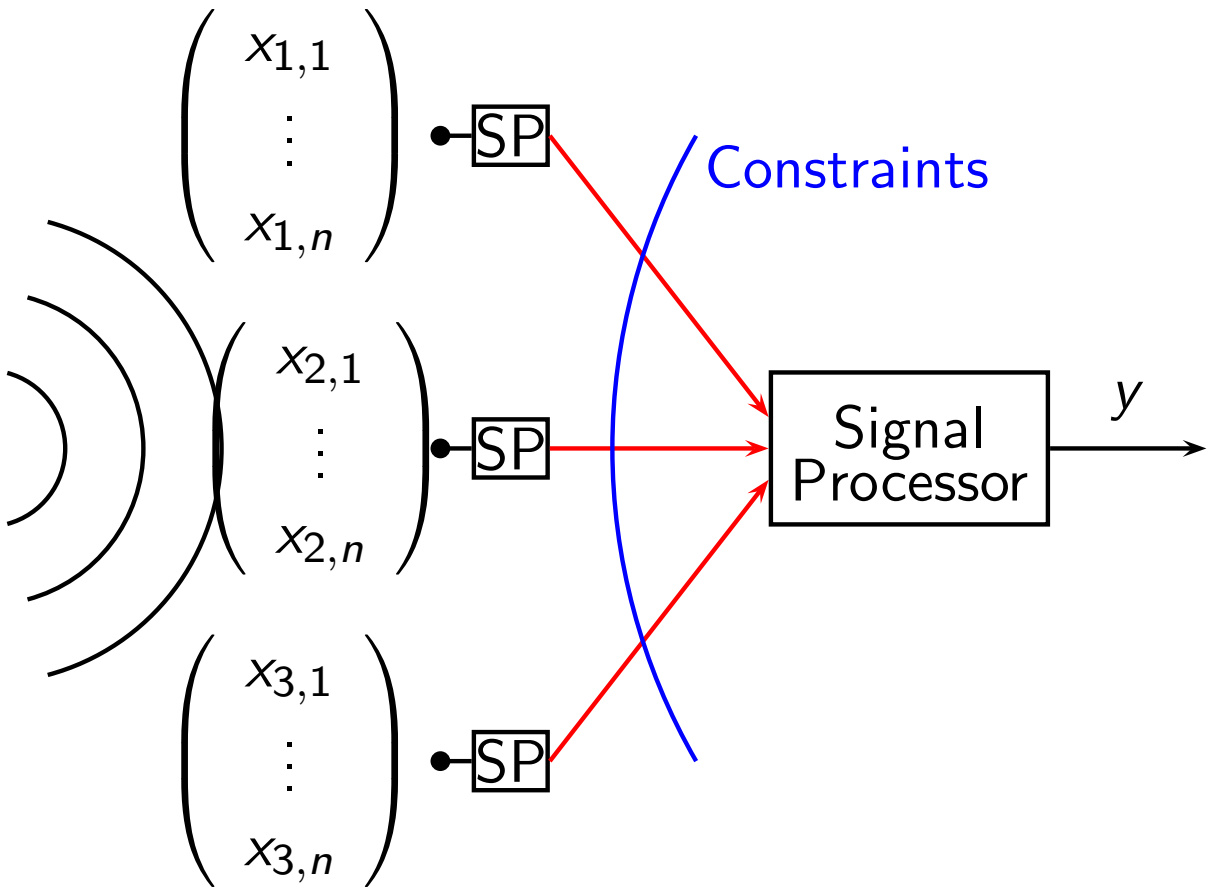
Algorithm (distributed KLT algorithm)

- *Start with arbitrary transform matrices* $C_1, C_2, \ldots, C_K$.
- *In turn, update each matrix according to the local KLT theorem.*
- *Repeat until* $(\epsilon-)$*convergence.*

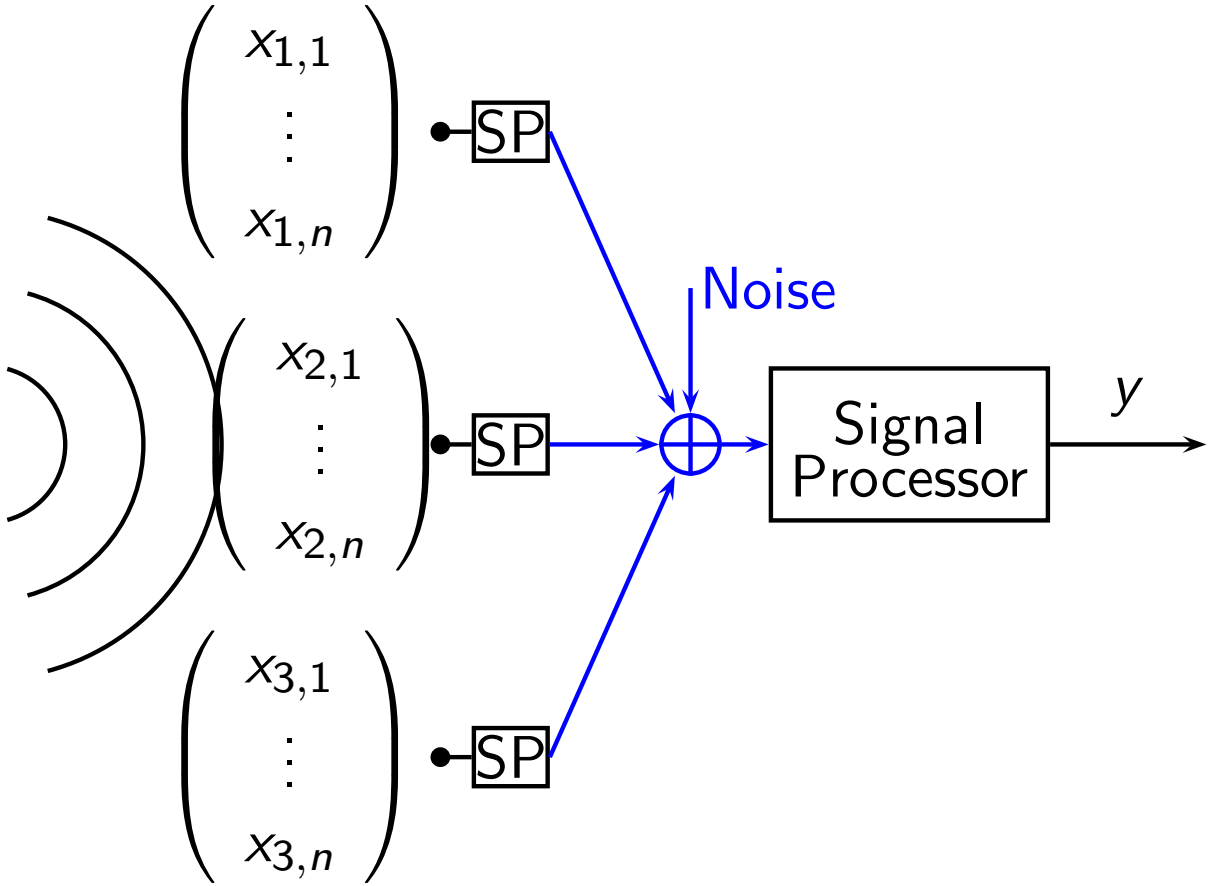
This algorithm is locally convergent, but no global guarantees can be given.

M. Gastpar, P. L. Dragotti, and M. Vetterli, "The Distributed Karhunen-Loève Transform," *IEEE Transactions on Information Theory,* December 2006.

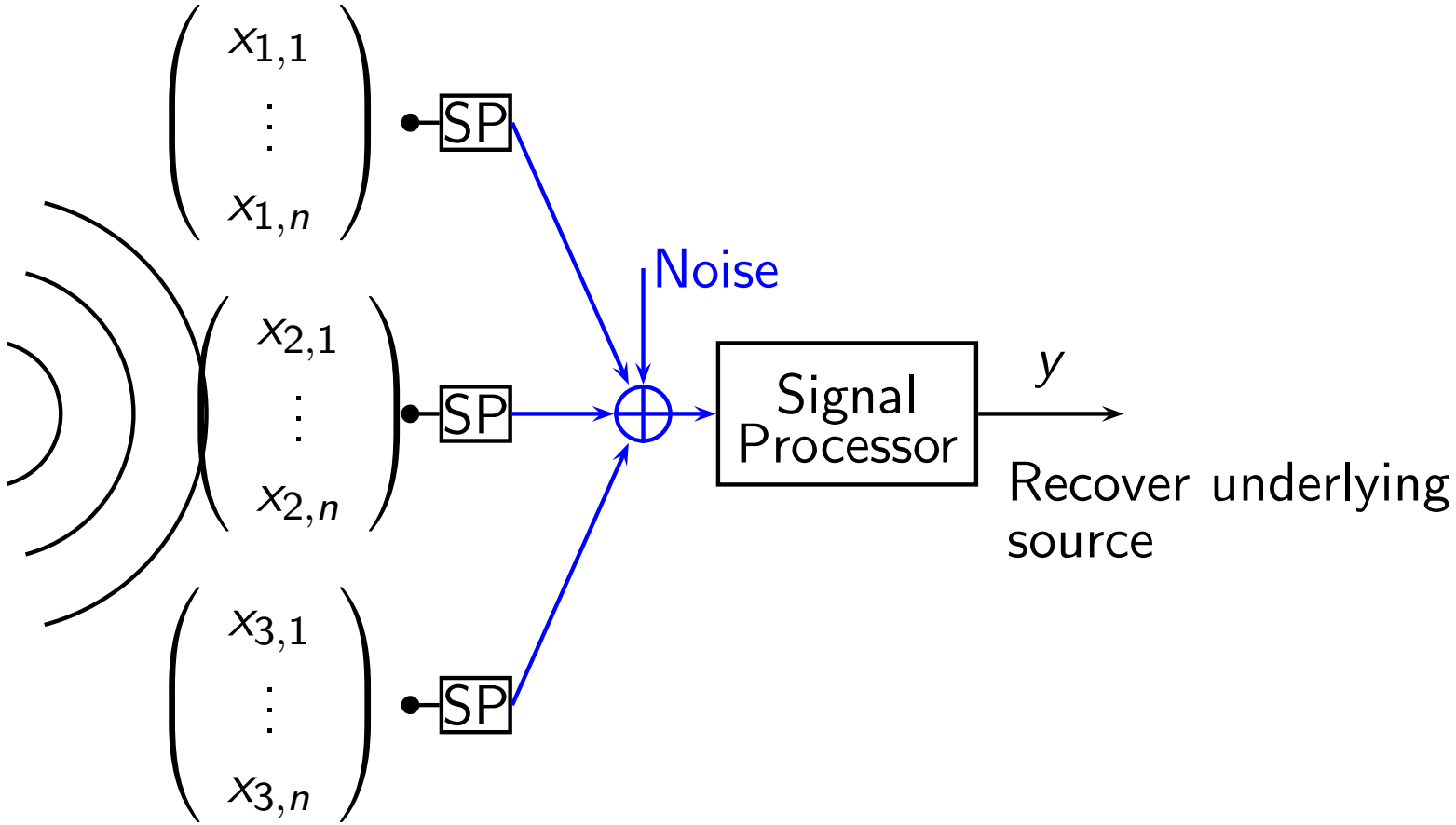
Case Study 2: A Wireless Sensor Network Scenario



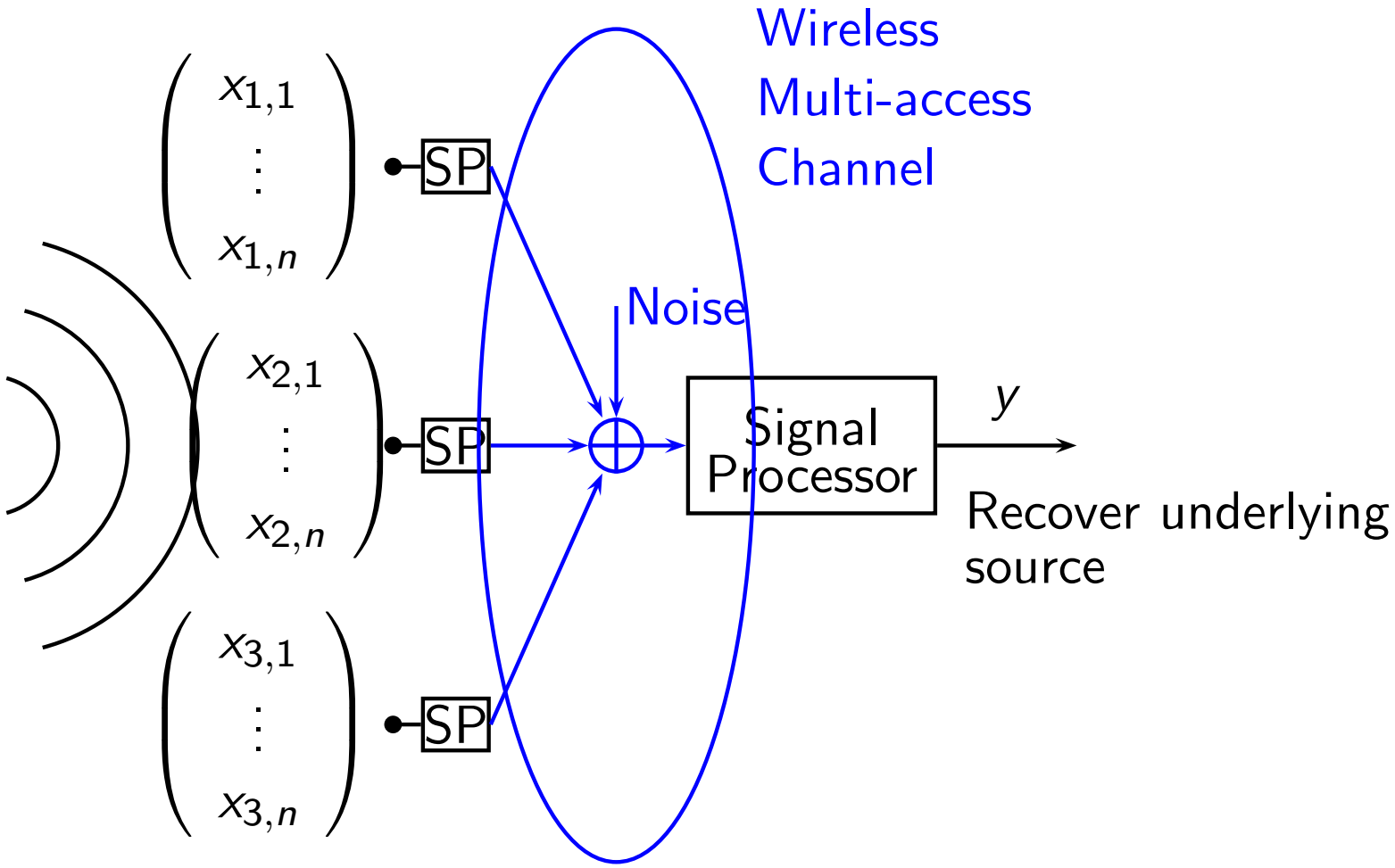
Case Study 2: A Wireless Sensor Network Scenario



Case Study 2: A Wireless Sensor Network Scenario



Case Study 2: A Wireless Sensor Network Scenario



Case Study 2: A Wireless Sensor Network Scenario

- For the wireless multi-access channel with M users and signal-to-noise ratio SNR, the information-theoretic total (sum) capacity is given by

$$C = \frac{1}{2}\log_2(1 + M\ \mathrm{SNR}) \quad \text{Bits.}$$

Case Study 2: A Wireless Sensor Network Scenario

- For the wireless multi-access channel with M users and signal-to-noise ratio SNR, the information-theoretic total (sum) capacity is given by

$$C = \frac{1}{2}\log_2(1 + M\ \mathrm{SNR}) \quad \text{Bits.}$$

- How well does this permit to represent the underlying source?

Case Study 2: A Wireless Sensor Network Scenario

- For the wireless multi-access channel with M users and signal-to-noise ratio SNR, the information-theoretic total (sum) capacity is given by

$$C = \frac{1}{2}\log_2(1 + M\ \mathrm{SNR}) \quad \text{Bits.}$$

- How well does this permit to represent the underlying source?
- This depends on precisely how the underlying source is captured by the sensor observations.

Case Study 2: A Wireless Sensor Network Scenario

- For example, let us suppose that each sensor observes the underlying source corrupted by independent additive noise:

$$
\begin{aligned}
x_1 &= s + w_1 \\
&\vdots \\
x_M &= s + w_M
\end{aligned}
$$

Case Study 2: A Wireless Sensor Network Scenario

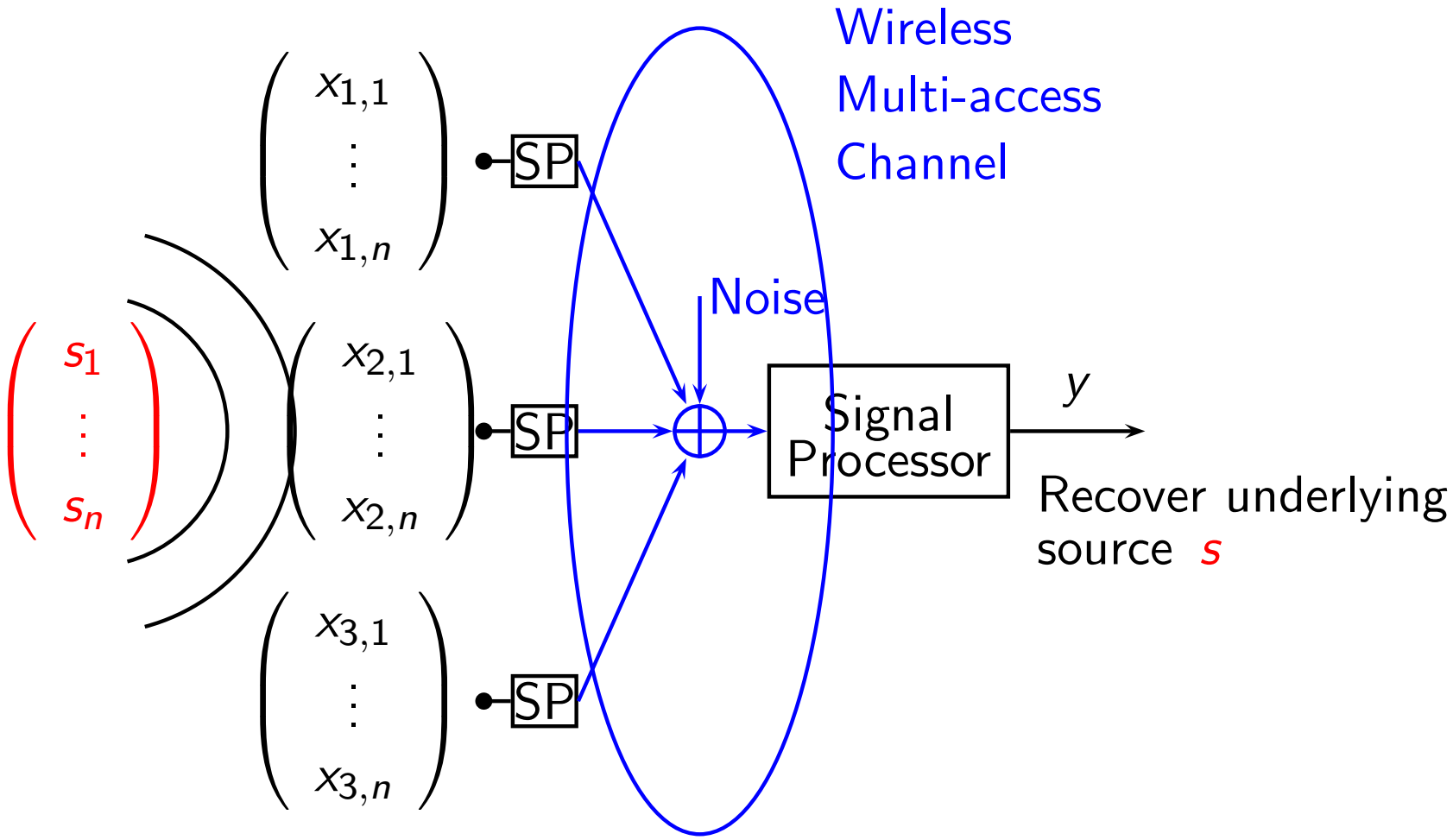
- For example, let us suppose that each sensor observes the underlying source corrupted by independent additive noise:

$$\begin{aligned} x_1 &= s + w_1 \\ &\vdots \\ x_M &= s + w_M \end{aligned}$$

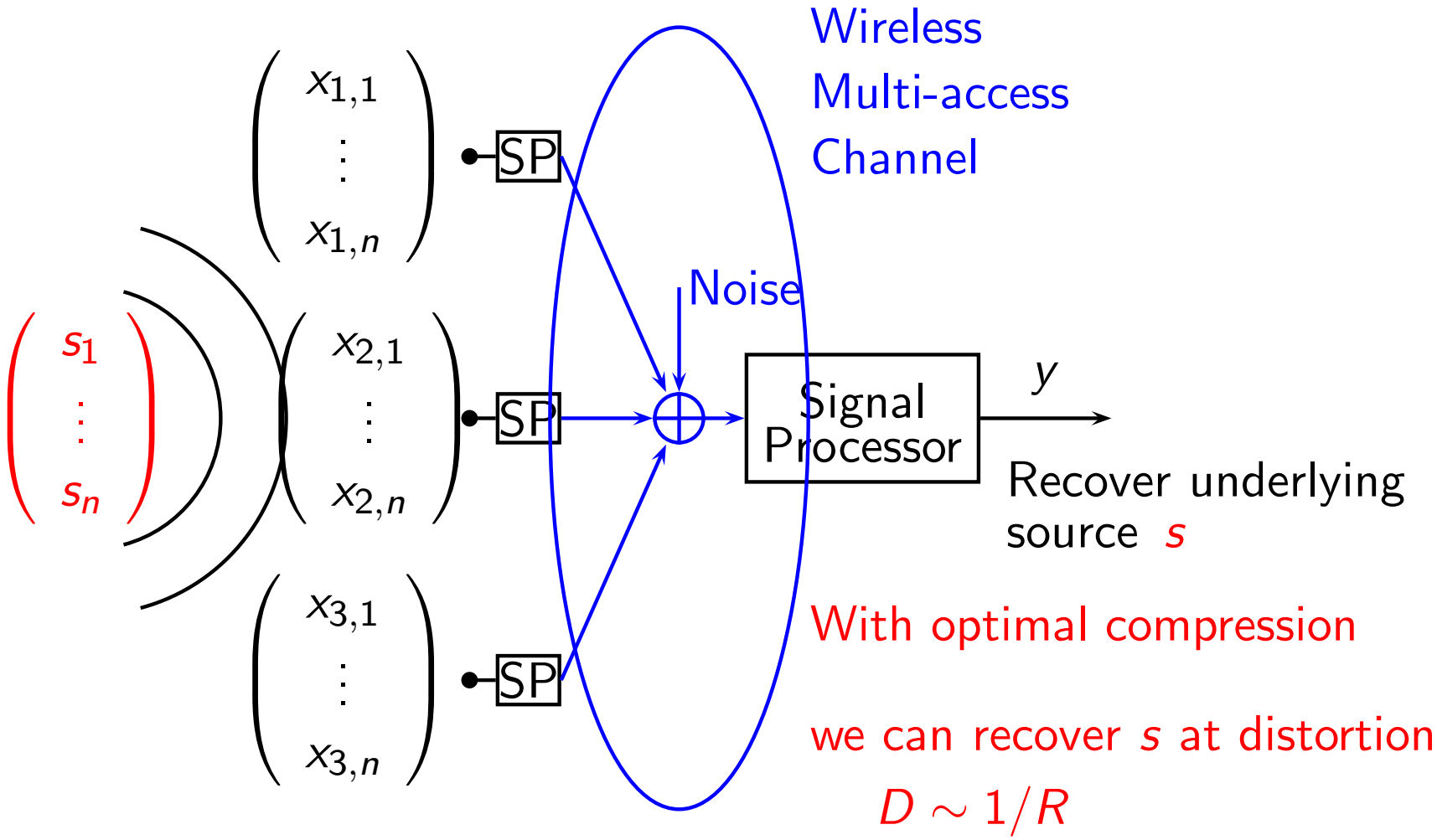
- An intricate information-theoretic argument shows that if the observations are *separately* compressed using a total of R bits, the underlying source S can be recovered with distortion no smaller than

$$D \sim \frac{1}{R}.$$

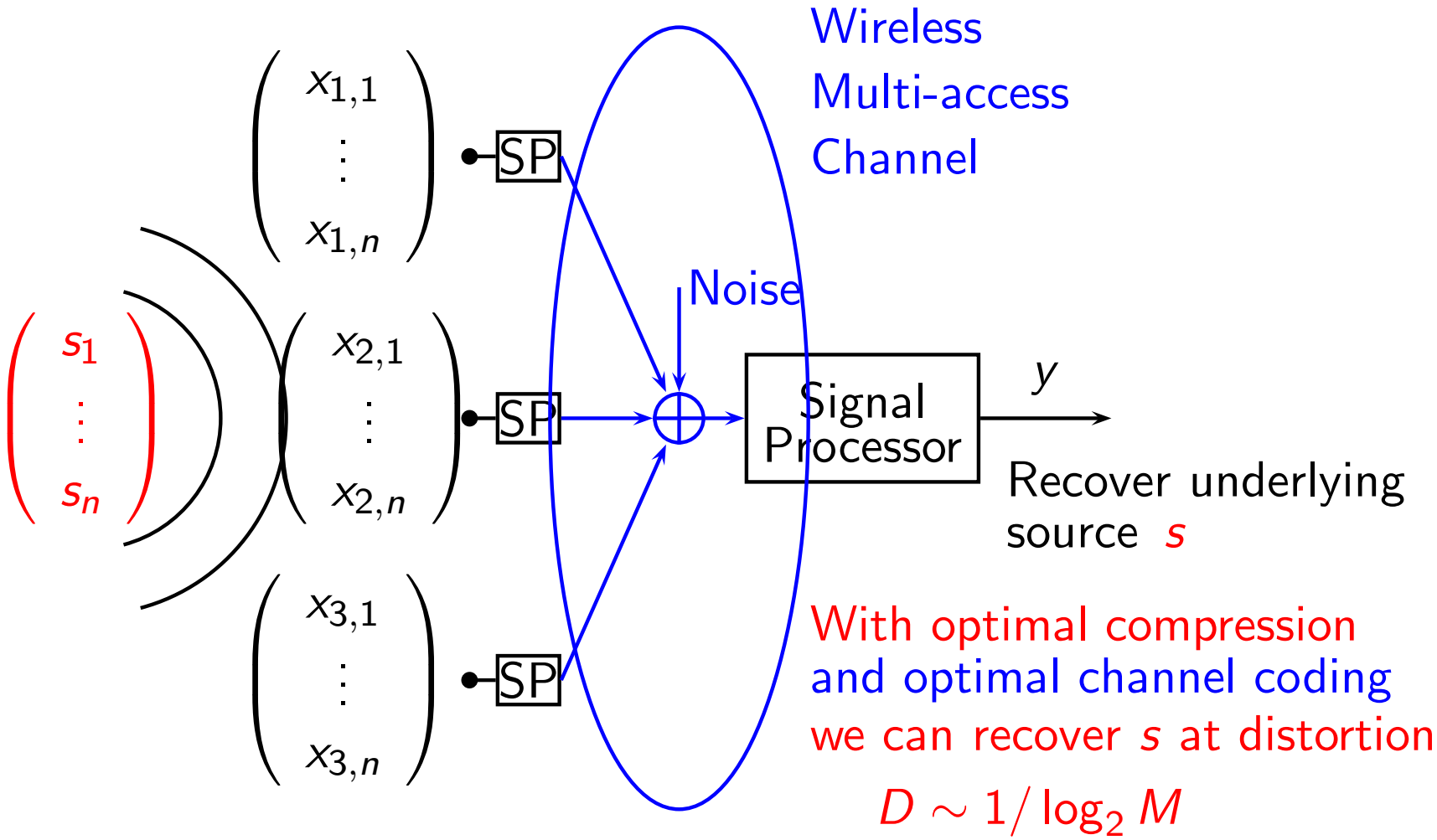
Case Study 2: A Wireless Sensor Network Scenario



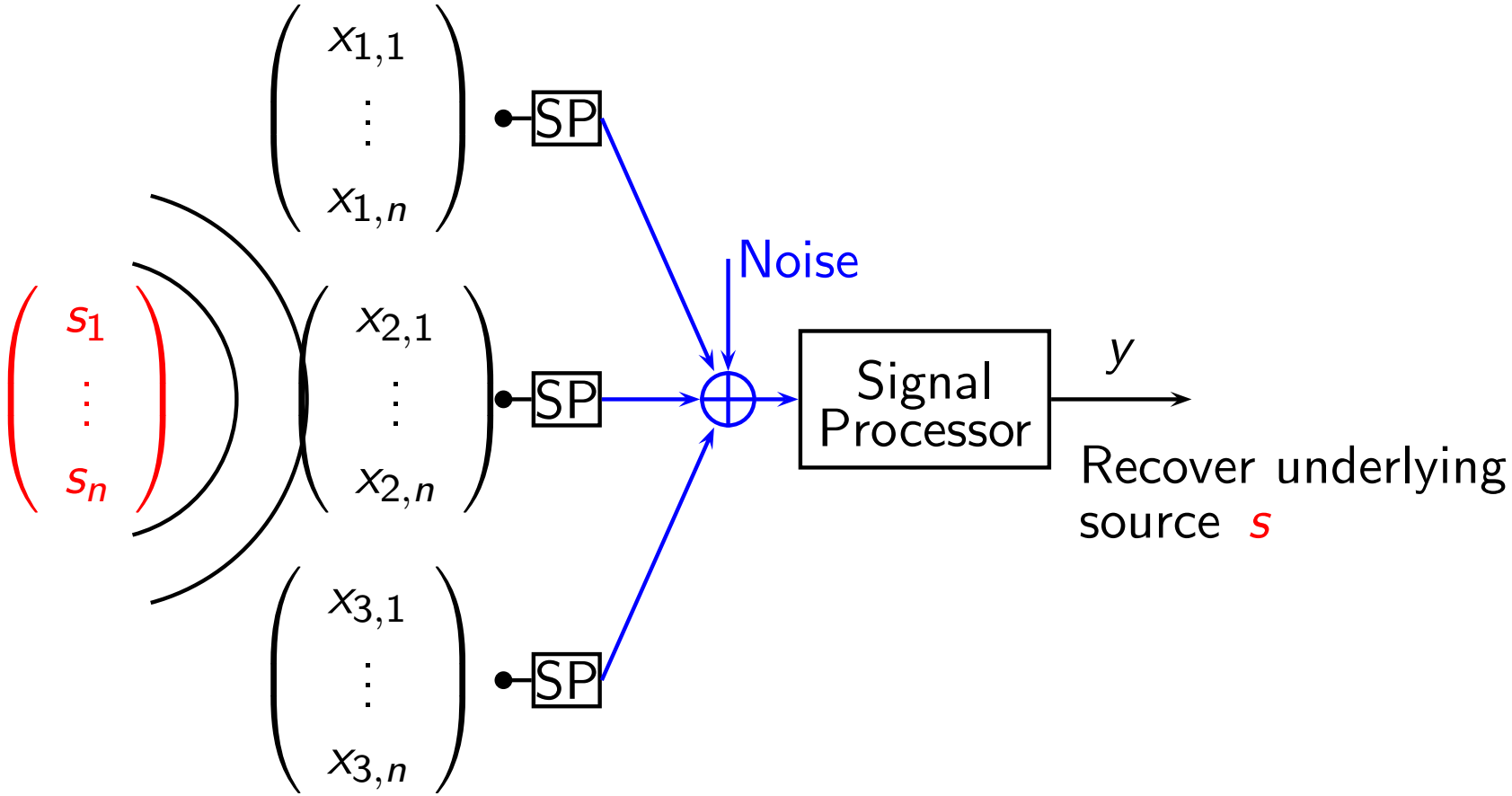
Case Study 2: A Wireless Sensor Network Scenario



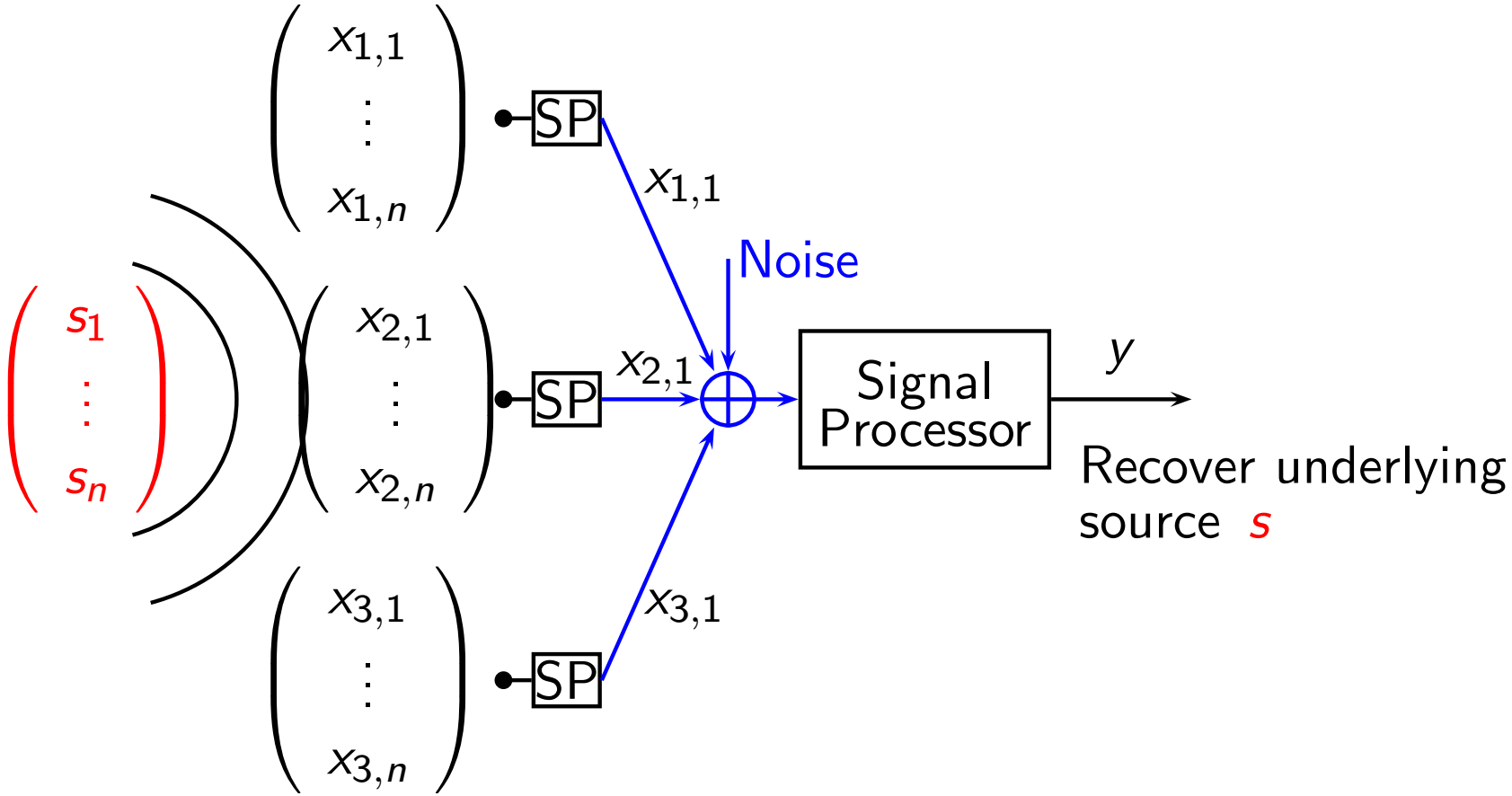
Case Study 2: A Wireless Sensor Network Scenario



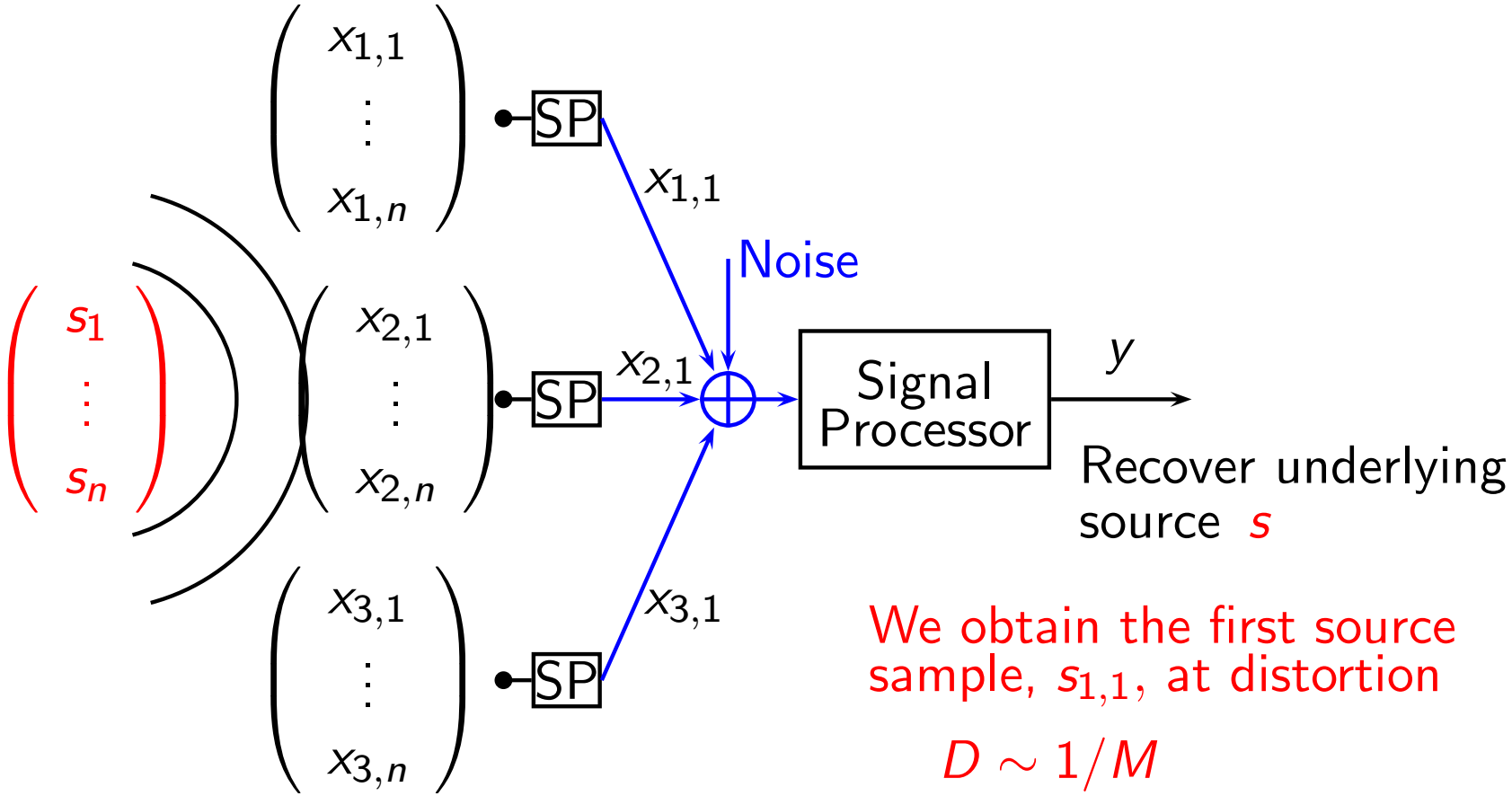
Case Study 2: A Wireless Sensor Network Scenario



Case Study 2: A Wireless Sensor Network Scenario



Case Study 2: A Wireless Sensor Network Scenario



We obtain the first source sample, $s_{1,1}$, at distortion

$$D \sim 1/M$$

Case Study 2: A Wireless Sensor Network Scenario

- Information-theoretically optimal channel coding combined with information-theoretically optimal (distributed) source compression lets us recover the underlying source to within a distortion of

$$D \sim \frac{1}{\log_2 M}.$$

- By contrast, straight uncoded transmission of the sensor observations lets us recover the underlying source to within a distortion of

$$D \sim \frac{1}{M}.$$

M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," *IEEE INFOCOM*,2002.

Case Study 2: A Sensor Network Scenario

What's really going on?

B. Nazer and M. Gastpar, "Computation over Multiple-Access Channels," *IEEE Transactions on Information Theory*, October 2007.

Foundations of Distributed Signal Processing

- "Distributed" Signal Processing means that there are **communication constraints.**

We looked at two versions of imposing communication constraints:

1. **Dimensionality constraints,**
   leading to the *Distributed Karhunen-Loève Transform.*
2. **Wireless "interference-plus-noise" constraints,**
   leading to insights about the insufficiency of legacy network architectures.

Other examples: Pairwise communication only ("gossip"); no centralized scheduling; specific connectivity; noisy linear projections; etc