Asymptotic Mean-Square Optimality of Belief Propagation for Sparse Linear Systems

Dongning Guo
Dept. of EECS,
Northwestern University

and

Chih-Chun Wang
School of ECE,
Purdue University

An Ongoing Research

Asymptotic Mean-Square Optimality of Belief Propagation for Sparse Linear Systems

An Ongoing Research

Asymptotic Mean-Square Optimality of Belief Propagation for Sparse Linear Systems

$$\Downarrow$$

Belief Propagation Is Asymptotically Equivalent to MAP Detection for Sparse Linear Systems

Outline

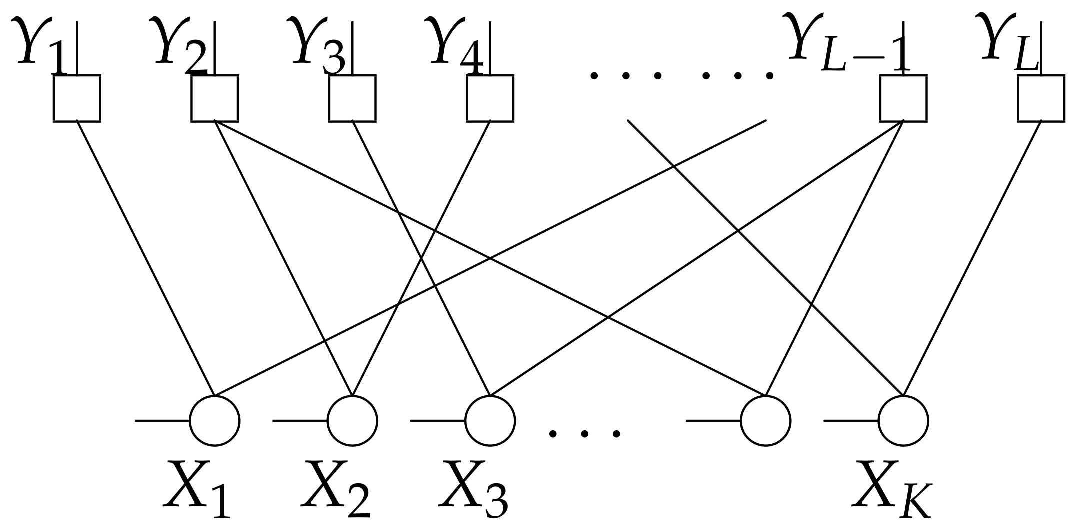
- The system of interest: $\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{N}$.
- Existing results & motivation.
- New results on MAP vs. BP.
- The proofs and the intuition behind.
- Conclusions and future developments.

The Linear System w. AWGN

- $$\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{N} = \sum_{k=1}^{K} \mathbf{S}_k A_k X_k + \mathbf{N}. \quad (1)$$
  - Input: $\mathbf{X} = [X_1, \ldots, X_K]^\top$, where $X_k \sim P_X$ are i.i.d.
  - Amplitudes: $A_k \sim P_A$, i.i.d.
  - Linear system $\mathbf{S}_{L\times K} = [A_1\mathbf{S}_1, \cdots, A_K\mathbf{S}_K]$.
  - Noise $\mathbf{N} \sim \mathcal{N}(0, \mathbf{I})$.
- $$Y_l = \sum_{k=1}^{K} \frac{S_{lk}}{\sqrt{\Lambda_k}} A_k X_k + N_l = W_l + N_l, \qquad l = 1, \ldots, L.$$

  $$\mathbf{S}_k = \frac{1}{\sqrt{\Lambda_k}}[S_{1k}, S_{2k}, \ldots, S_{Lk}]^\top, \mathsf{E}\|\mathbf{S}_k\|^2 = 1.$$
- Applications: MIMO, CDMA, OFDM, statistics, optimization ...

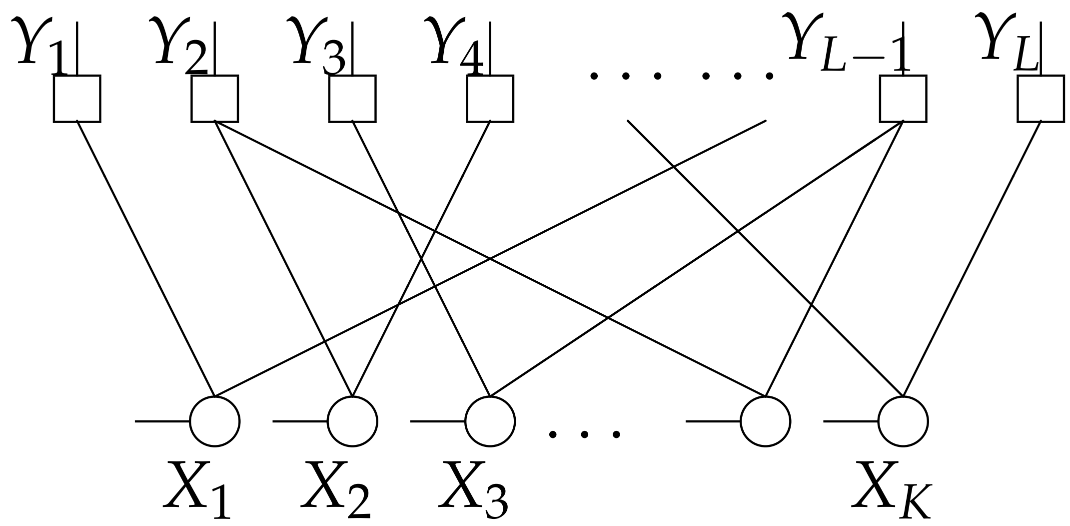
Factor Graph



$$p_{\mathbf{Y}|\mathbf{X},\mathbf{S}}(\mathbf{y}|\mathbf{x},\mathbf{S}) = \prod_{l=1}^{L} \mathcal{N}\left(\sum_{k=1}^{K} \frac{S_{lk}}{\sqrt{\Lambda_k}} A_k x_k, 1\right).$$

- Given $(\mathbf{Y}, \mathbf{S}) = (\mathbf{y}, \mathbf{s})$, compute the maximum *a posteriori* estimate (exact statistical inference on the graph model): $p_{X_k|\mathbf{Y},\mathbf{S}}(\cdot|\mathbf{y}, \mathbf{s})$.

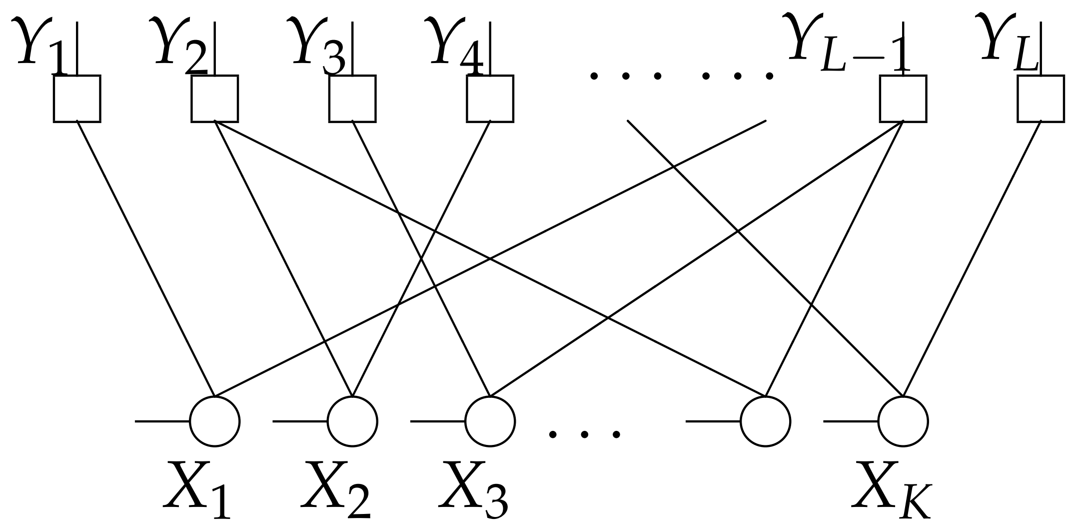
Factor Graph



How about sparse graphs?

- Given $(\mathbf{Y}, \mathbf{S}) = (\mathbf{y}, \mathbf{s})$, compute the maximum *a posteriori* estimate (exact statistical inference on the graph model): $p_{X_k|\mathbf{Y},\mathbf{S}}(\cdot|\mathbf{y}, \mathbf{s})$.

Factor Graph



How about sparse graphs?

$\approx$ Low-Density Generator-Matrix (LDGM) codes

estimate (exact statistical inference on the graph model):

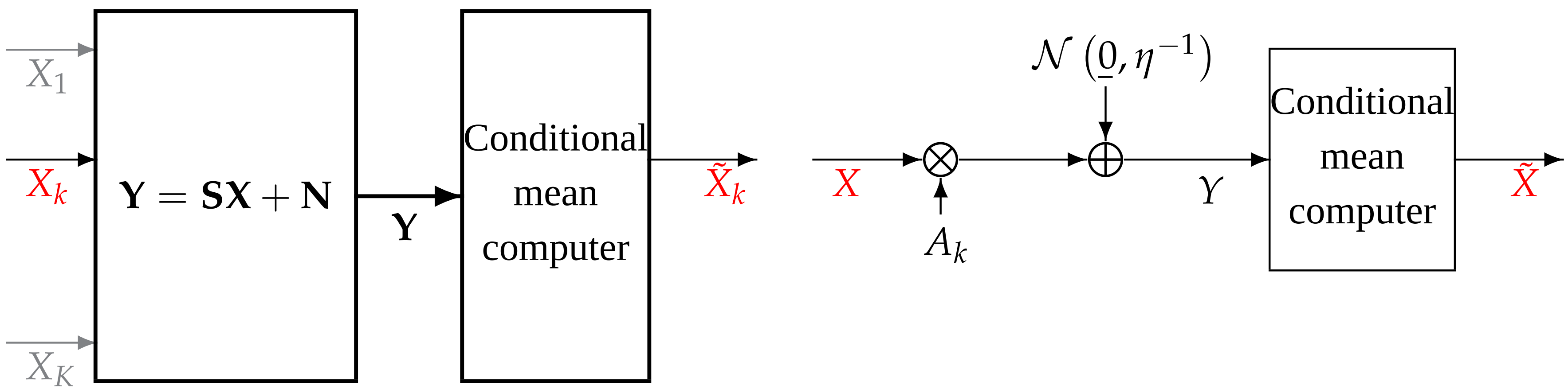
$p_{X_k|\mathbf{Y},\mathbf{S}}(\cdot|\mathbf{y},\mathbf{s})$.

One Important Notation

- The MMSE per dimension of estimating $\mathbf{Z}_{K\times 1}$ given $\mathbf{W}$:

$$\mathcal{E}\left(\mathbf{Z}\,|\mathbf{W}\right) = \frac{1}{K}\mathsf{E}\left\{\|\mathbf{Z} - \mathsf{E}\left\{\mathbf{Z}\,|\,\mathbf{W}\right\}\|^2\right\}$$

Non-rigorous Results Via Statistical Mechanics



Claim 1 (MMSE Estimator) *[Guo & Verdú, IT-05] Suppose $S_{lk} \sim P_S$ are i.i.d. As $K, L \to \infty$ with $K/L \to \beta$,*

$$p_{\tilde{X}_k|X_k}(\cdot|\cdot) \to p_{\tilde{X}|X}(\cdot|\cdot)$$

where

$$\eta^{-1} = 1 + \beta\, \mathcal{E}\left(AX \,|\, \sqrt{\eta}AX + N, A\right)$$

Note: Special case of $X_k = \pm 1$ and $A \equiv 1$ due to Tanaka IT-02.

Previous Results (Cont'd)

Claim 2 (Mutual Information) *[Guo & Verdú, IT '05]*

$$\lim_{K=\beta L\to\infty} \frac{1}{K} I(X_k; \mathbf{Y}|\mathbf{S}) = I(X; \sqrt{\eta}\, AX + N|A)$$

Claim 3 (Mutual Information) *[Guo & Verdú, IT '05]*

$$\lim_{K=\beta L\to\infty} \frac{1}{K} I(\mathbf{X}; \mathbf{Y}|\mathbf{S}) = I(X; \sqrt{\eta}\, AX + N|A) + \frac{\eta - 1 - \log \eta}{2\beta}$$

Previous Results (Cont'd)

Claim 2 (Mutual Information) *[Guo & Verdú, IT '05]*

$$\lim_{K=\beta L\to\infty} \frac{1}{K} I(X_k;\mathbf{Y}|\mathbf{S}) = I(X;\sqrt{\eta}\,AX+N|A)$$

Claim 3 (Mutual Information) *[Guo & Verdú, IT '05]*

$$\lim_{K=\beta L\to\infty} \frac{1}{K} I(\mathbf{X};\mathbf{Y}|\mathbf{S}) = I(X;\sqrt{\eta}\,AX+N|A) + \frac{\eta - 1 - \log\eta}{2\beta}$$

- A good "understanding" of the dense graph case, but no rigorous justification.

The First Justification by [Montanari & Tse]

- Assumptions:
  - Binary input symbols; Equal power ($A_k \equiv 1$);
  - Infinitesimal density, but infinitely spread matrix **S**
    $K = \beta L \to \infty$ first, then average node degree $\bar{\Gamma} \to \infty$.

Theorem 1 (Entropy) *[Montanari and Tse, '06]*

$$\frac{1}{K} H(\mathbf{X}|\mathbf{Y}, \mathbf{S}) \xrightarrow{K=\beta L \to \infty} H(X|\sqrt{\eta} X + N) - \frac{\eta - 1 - \log \eta}{2\beta}$$

The First Justification by [Montanari & Tse]

- Assumptions:
  - Binary input symbols; Equal power ($A_k \equiv 1$);
  - Infinitesimal density, but infinitely spread matrix **S**
    $K = \beta L \to \infty$ first, then average node degree $\bar{\Gamma} \to \infty$.

Theorem 1 (Entropy) *[Montanari and Tse, '06]*

$$\frac{1}{K} H(\mathbf{X}|\mathbf{Y}, \mathbf{S}) \xrightarrow{K=\beta L \to \infty} H(X|\sqrt{\eta}X + N) - \frac{\eta - 1 - \log \eta}{2\beta}$$

Theorem 2 (BER optimality of BP) *[Montanari and Tse, '06]*

$$\lim_{t\to\infty} \lim_{\bar{\Gamma}\to\infty} \limsup_{K=\beta L\to\infty} \left[\mathsf{P}\left(\hat{X}_k^t \neq X_k\right) - \mathsf{P}\left(\hat{X}_k \neq X_k\right)\right] = 0$$

By the area theorem [MeMoRiUr] $\equiv$ the I-MMSE formula [GuShVe].

Overview of New Results

Arbitrary P_X, P_A, P_S

Sandwiching Arguments $\Longrightarrow$ Posterior optimality (this paper)

Equiv. Scalar AWGN Ch. (this paper)

$\Downarrow \quad \uparrow \qquad\qquad \Downarrow$

$X_k = \pm 1$, $S_{lk} = \pm 1$, & $A_k \equiv 1$

I-MMSE Arguments $\Longrightarrow$ Optimality of BP (MontTse06)

Fixed-point formula (MontTse06, GuoWang06)

Sparse Linear Systems

- We consider a general setting: regular/irregular, etc.
- One example:
  - The Doubly-Poisson Ensemble:
    Flip a coin w. $p = \bar{\Gamma}/K$ to decide whether an entry of $\mathbf{S}$ is non-zero.
    Non-zero $S_{lk} \sim P_S$ being symmetrically distributed.
  - The *large-system limit*: $K, L \to \infty$ with $K/L \to \beta$.
  - The *large-sparse-system limit*:
    — asymptotically free of short cycles.

Sparse Linear Systems

- We consider a general setting: regular/irregular, etc.
- One example:
  - The Doubly-Poisson Ensemble:
    Flip a coin w. $p = \bar{\Gamma}/K$ to decide whether an entry of **S** is non-zero.
    Non-zero $S_{lk} \sim P_S$ being symmetrically distributed.
  - The *large-system limit*: $K, L \to \infty$ with $K/L \to \beta$.
  - The *large-sparse-system limit*:
    — asymptotically free of short cycles.
- System load constraint: β is small enough (ex: $\beta < 1.49$) s.t. the solution is unique for

$$\eta^{-1} = 1 + \beta \, \mathcal{E}\left(AX \,|\, \sqrt{\eta}\, AX + N, A\right)$$

Our Main Results

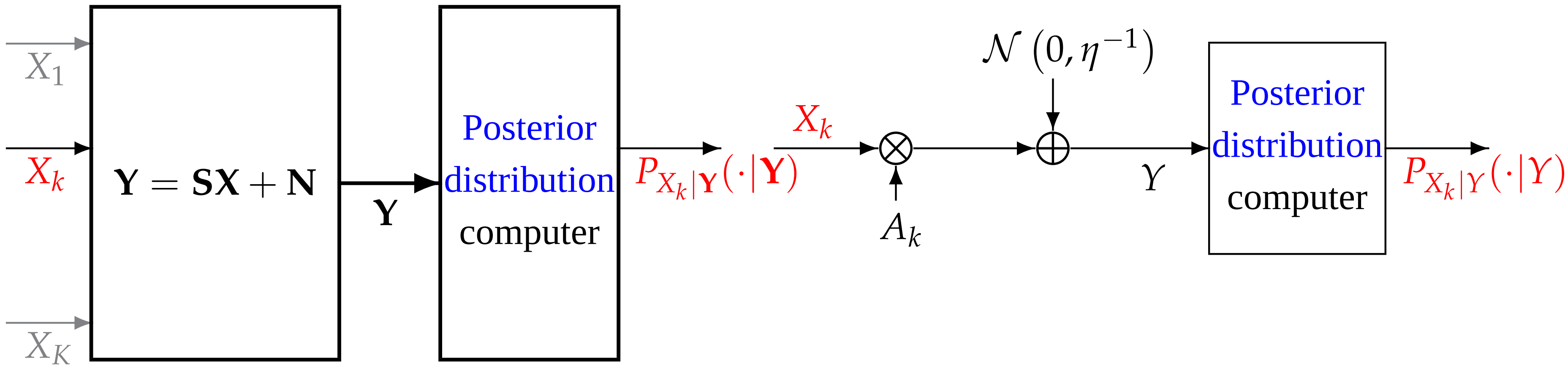
Theorem 3 (Posterior optimality of BP)

$$\lim_{t\to\infty} \limsup_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}^2=o(K)\to\infty}} \left| p^{(t)}_{X_k}(x|\mathbf{Y}^{(t)},\mathbf{S}) - p_{X_k}(x|\mathbf{Y},\mathbf{S}) \right| = 0 \quad i.p., \forall k, x$$

where $\mathbf{Y}^{(t)}$ *consists of all elements of* $\mathbf{Y}$ *that reach* X_k *within* t *iterations.*

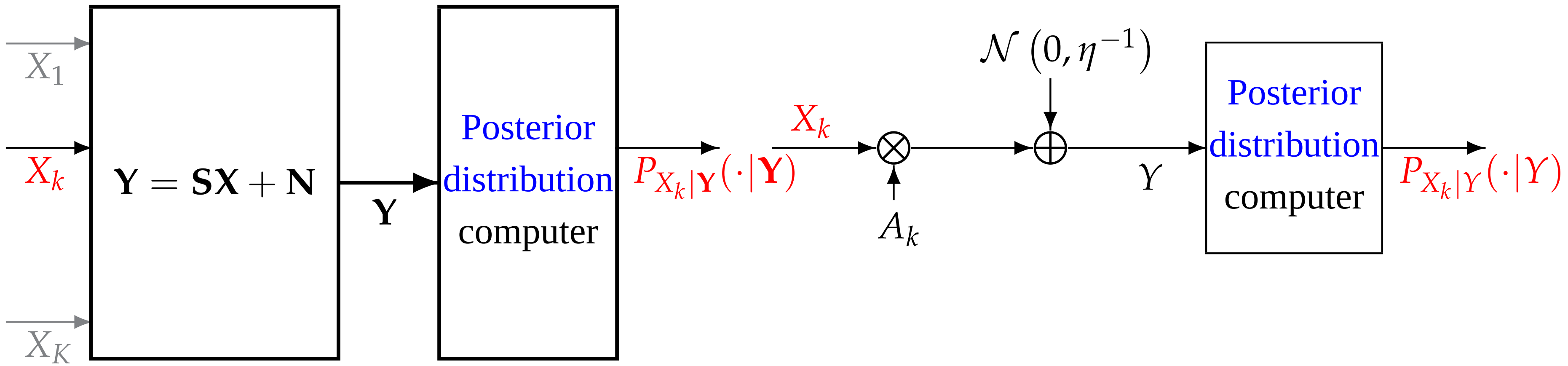
The key message: $\mathbf{Y}^{(t)}$, an infinitesimal portion of the entire observation $\mathbf{Y}$ is sufficient.

Our Main Results (Cont'd)



Theorem 4 (Equivalent Scalar AWGN Channel)
The statistics of the posterior distribution computer are indistinguishable from that of a scalar Gaussian channel.

Our Main Results (Cont'd)

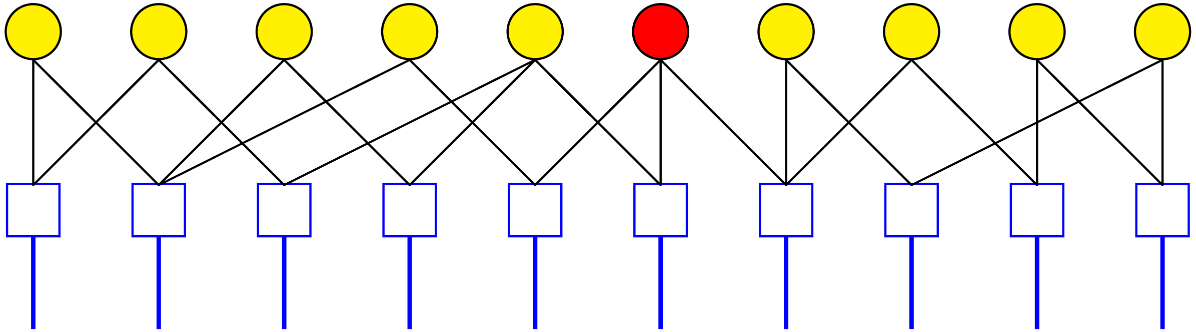


Theorem 4 (Equivalent Scalar AWGN Channel)
The statistics of the posterior distribution computer are indistinguishable from that of a scalar Gaussian channel.

With BP $\equiv$ MAP and the scalar Gaussian channel equivalency, all existing entropy, mutual information, MMSE, BP vs. MAP results are straightforward corollaries.

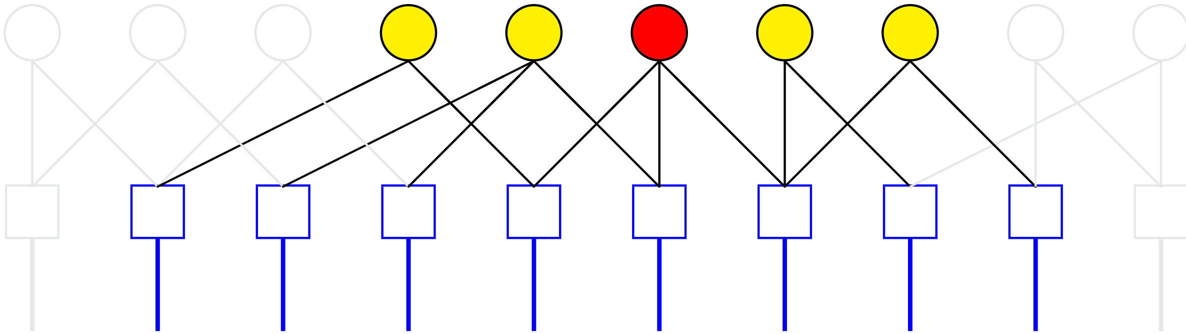
The Proof of BP $\equiv$ MAP

MAP

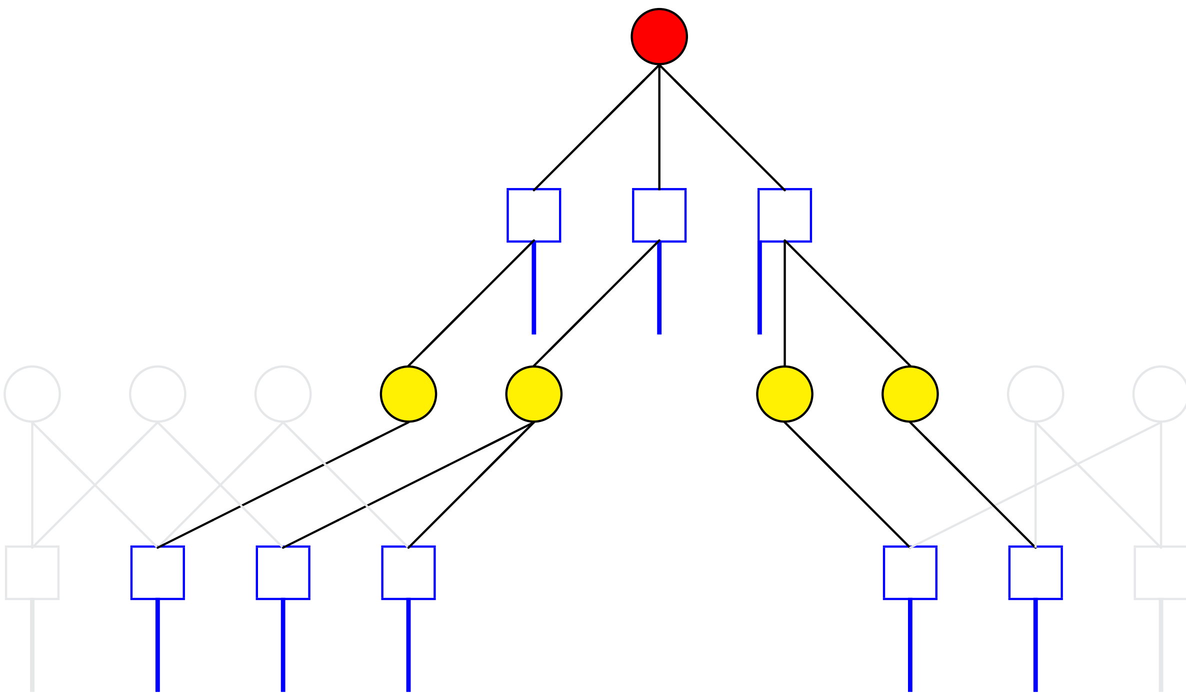


The Proof of BP $\equiv$ MAP

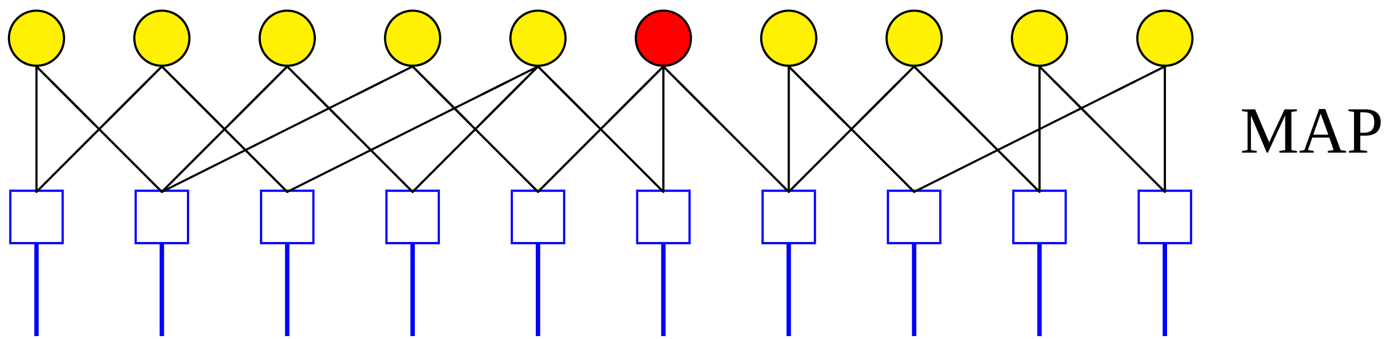
Belief Propagation



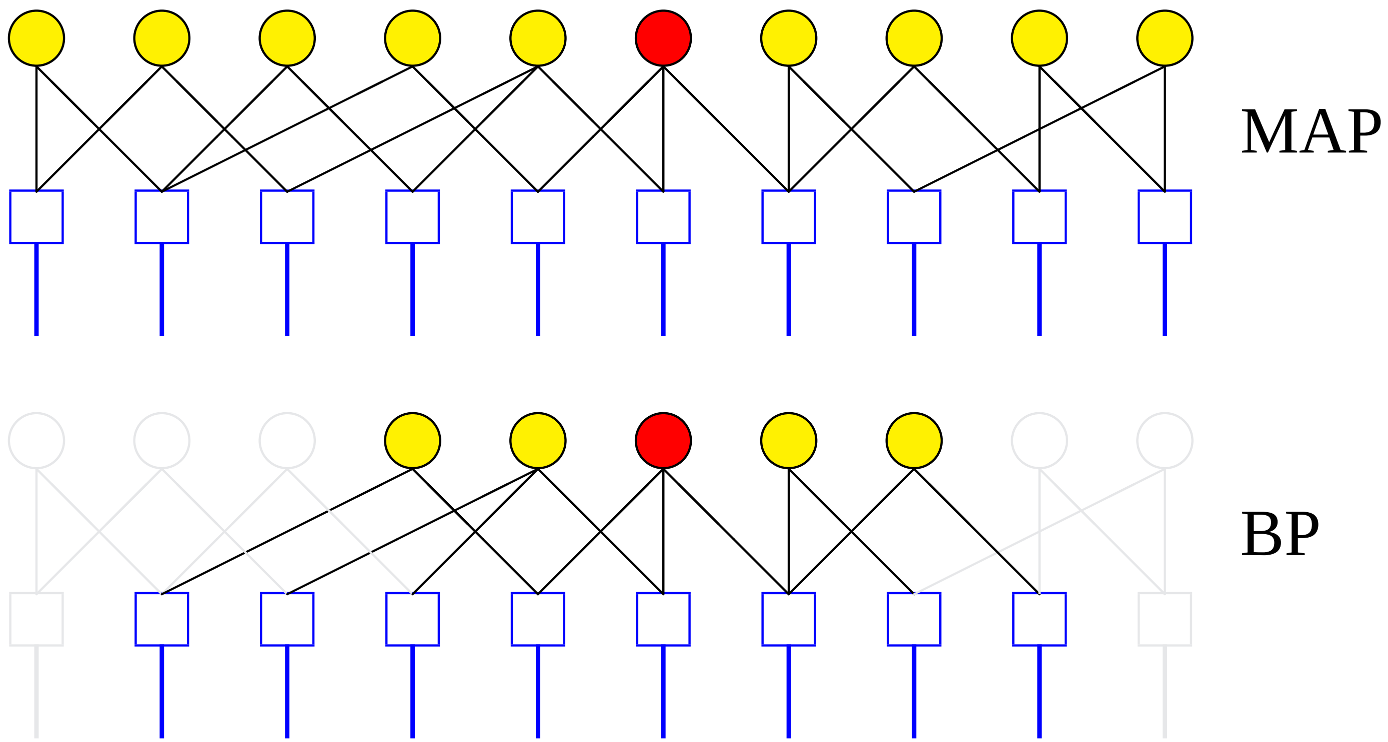
1 iteration



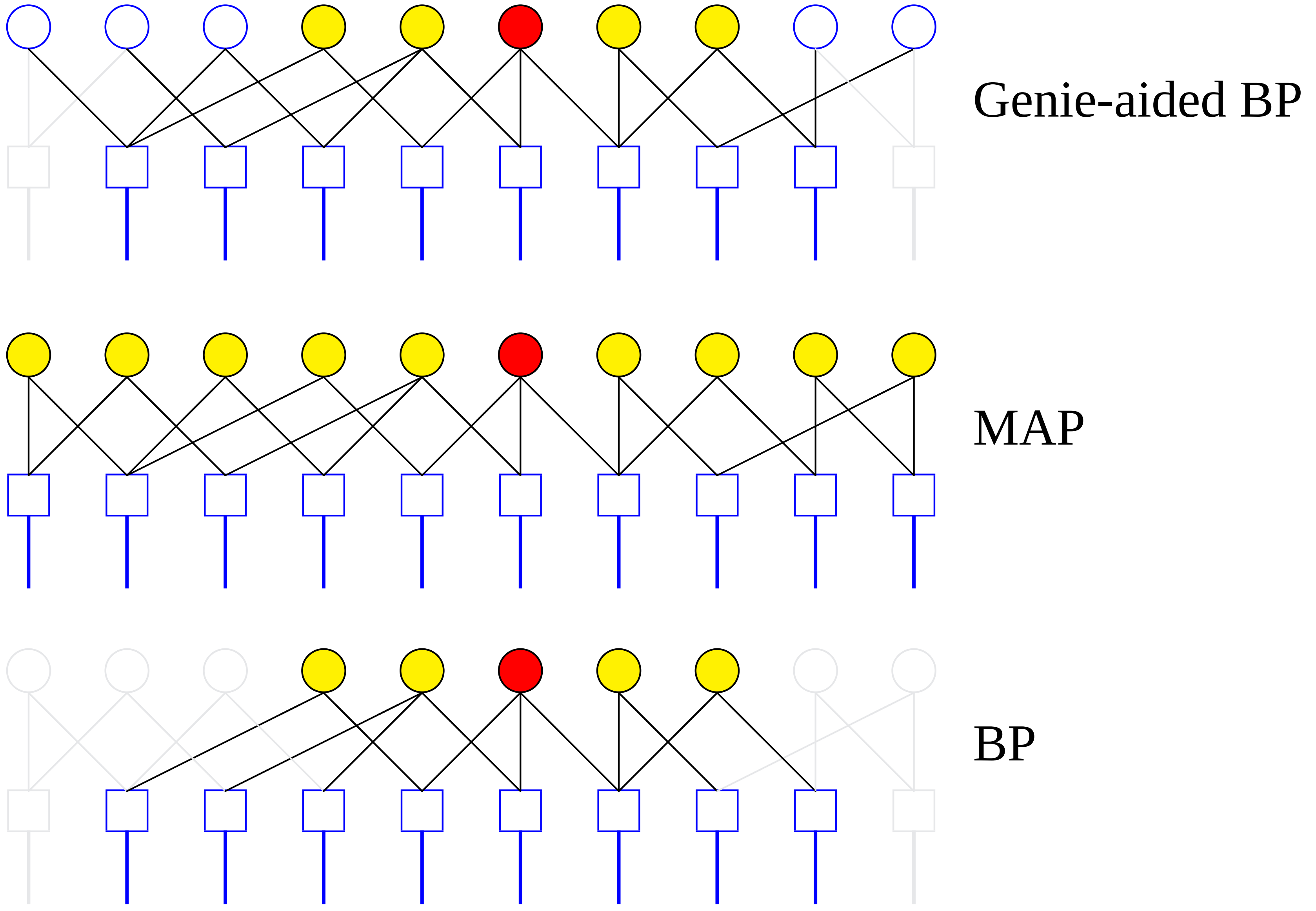
Sandwiching Proof



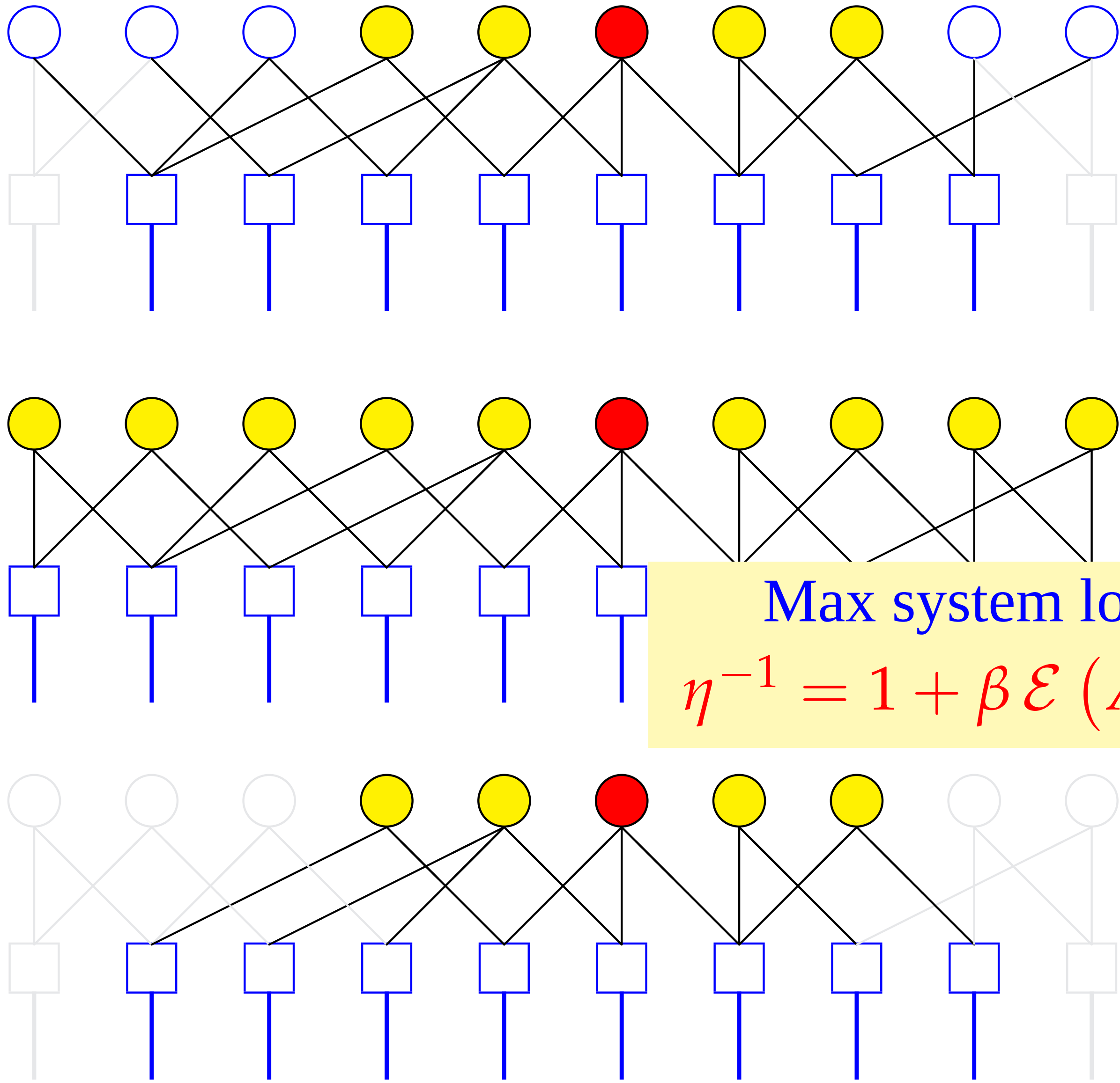
Sandwiching Proof



Sandwiching Proof



Sandwiching Proof



Genie-aided BP

MAP

Max system load constraint on β:

$$\eta^{-1} = 1 + \beta\, \mathcal{E}\left(AX \,\middle|\, \sqrt{\eta}\, AX + N, A\right)$$

BP

Relation to I-MMSE

Theorem 5 (I-MMSE) *[Guo-Shamai-Verdú]*

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(\mathbf{X}; \sqrt{\gamma}\,\mathbf{SX} + \mathbf{N}) = \frac{1}{2}\,\mathcal{E}\left(\mathbf{SX} \,|\, \sqrt{\gamma}\,\mathbf{SX} + \mathbf{N}\right), \quad \gamma \geq 0.$$

Define

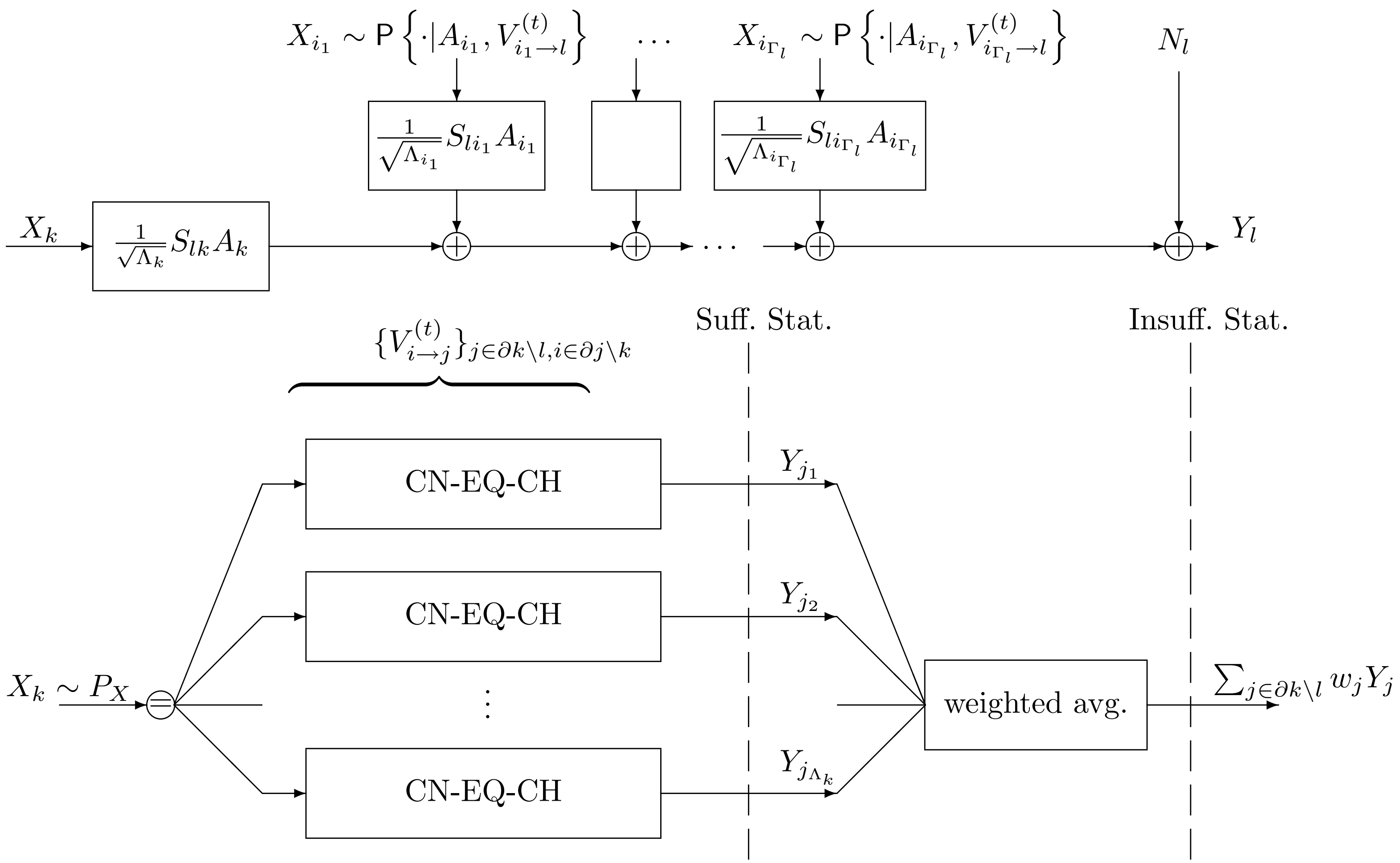
$$C(\gamma) = I\big(X; \sqrt{\gamma\eta(\gamma)}\, AX + N | A\big) + \frac{\eta(\gamma) - 1 - \log \eta(\gamma)}{2\beta}.$$

Theorem 6 *[in this work] For every $\gamma \geq 0$,*

$$C(\gamma) = \frac{1}{2} \int_0^\gamma \eta(\gamma)\, \mathcal{E}_{X,A}\left(\gamma\eta(\gamma)\right)\, \mathrm{d}\gamma.$$

The Proof of the Scalar AWGNC Equivalency

Equivalent Channel Perspective [Wang *et al.*, IT-Jan. 07]



Based on the central limit theorem while addressing the convergence speed competition problem.

Generalization to Broader Ensembles

- Symbol degree Λ_k, chip degree Γ_l.
- Chip-semi-regularity:

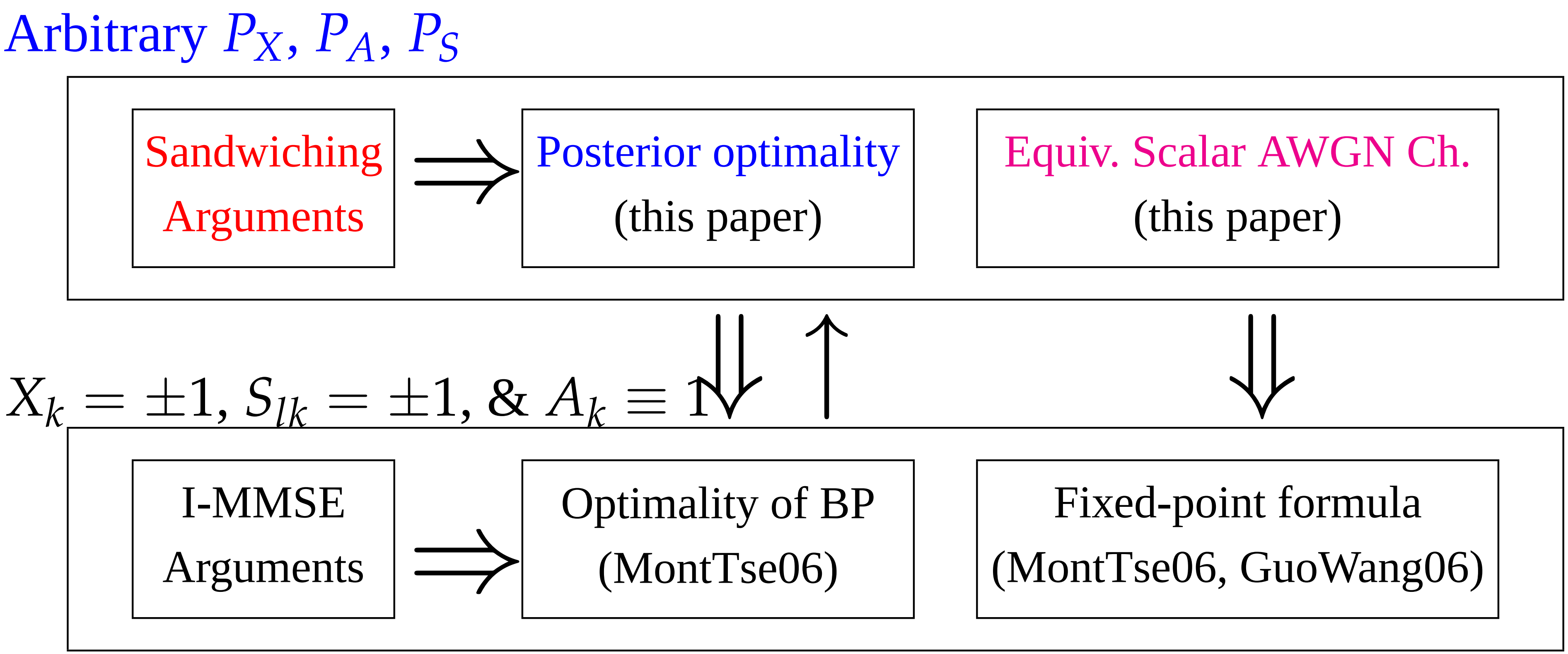
$$\lim_{\substack{K=\beta L \to \infty \\ \bar{\Gamma}^2 = o(K) \to \infty}} \mathsf{P}\left(\left| \Gamma_l - \bar{\Gamma} \right| > \epsilon \bar{\Gamma} \right) = 0, \quad \forall \epsilon > 0, \forall l$$

- Symbol balanced condition: $\exists\, a < \infty$ such that

$$\lim_{K \to \infty} \frac{1}{K} \left| \{ k \mid \Lambda_k > a \bar{\Gamma} \} \right| = 0.$$

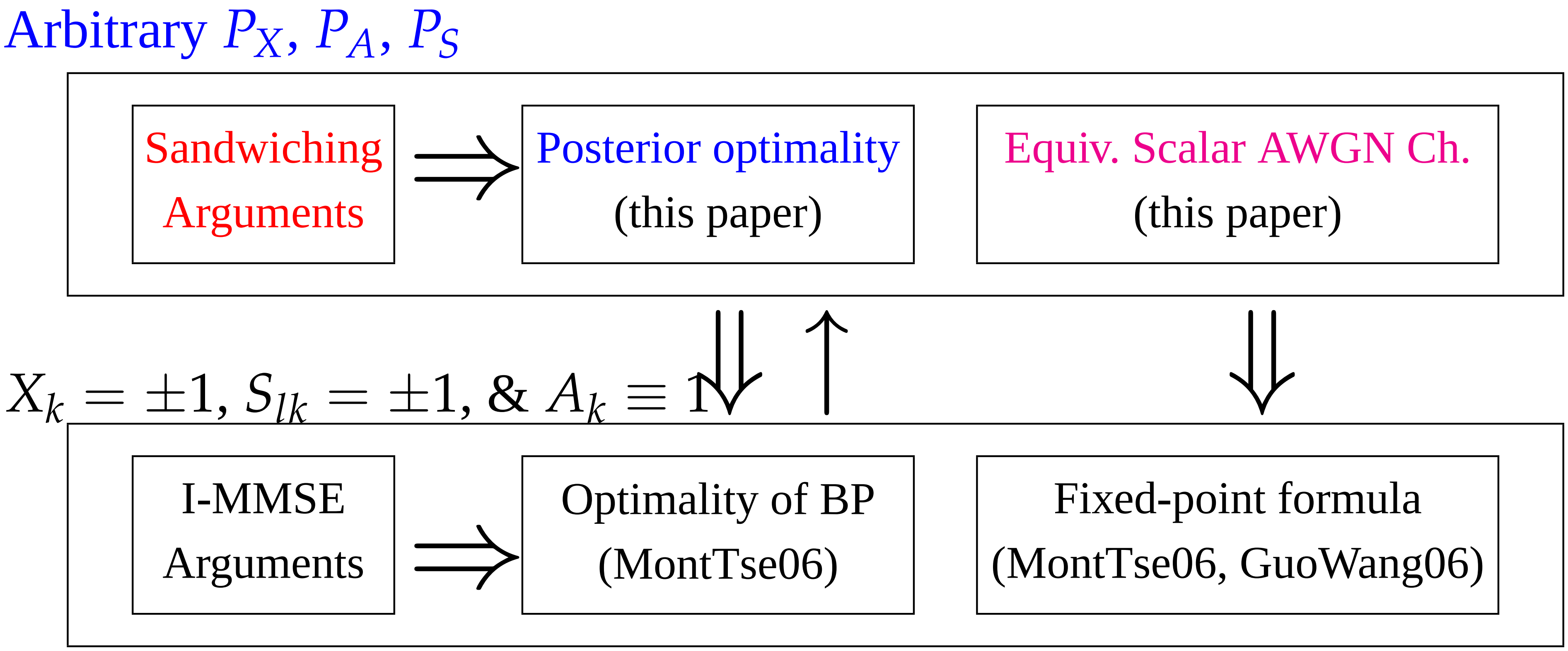
- Examples
  - The graph-based regular/irregular ensembles;
  - The symbol-irregular, chip-Poisson ensembles (Montanari-Tse).

Conclusion & Future Work



- Large sparse systems; small system load (ex: $\beta < 1.49$);
- BP $\equiv$ MAP; Equivalent Scalar AWGN Channels;

Conclusion & Future Work



- Large sparse systems; small system load (ex: $\beta < 1.49$);
- BP $\equiv$ MAP; Equivalent Scalar AWGN Channels;
- Non-Gaussian noise?