From Crowds to Codes: Minimizing Review Burden in Conference Review Protocols

Xingbo Wang[1], Fang-Yi Yu[2], Yichi Zhang[3]

Peking University[1] George Mason University[2] Rutgers University[3]

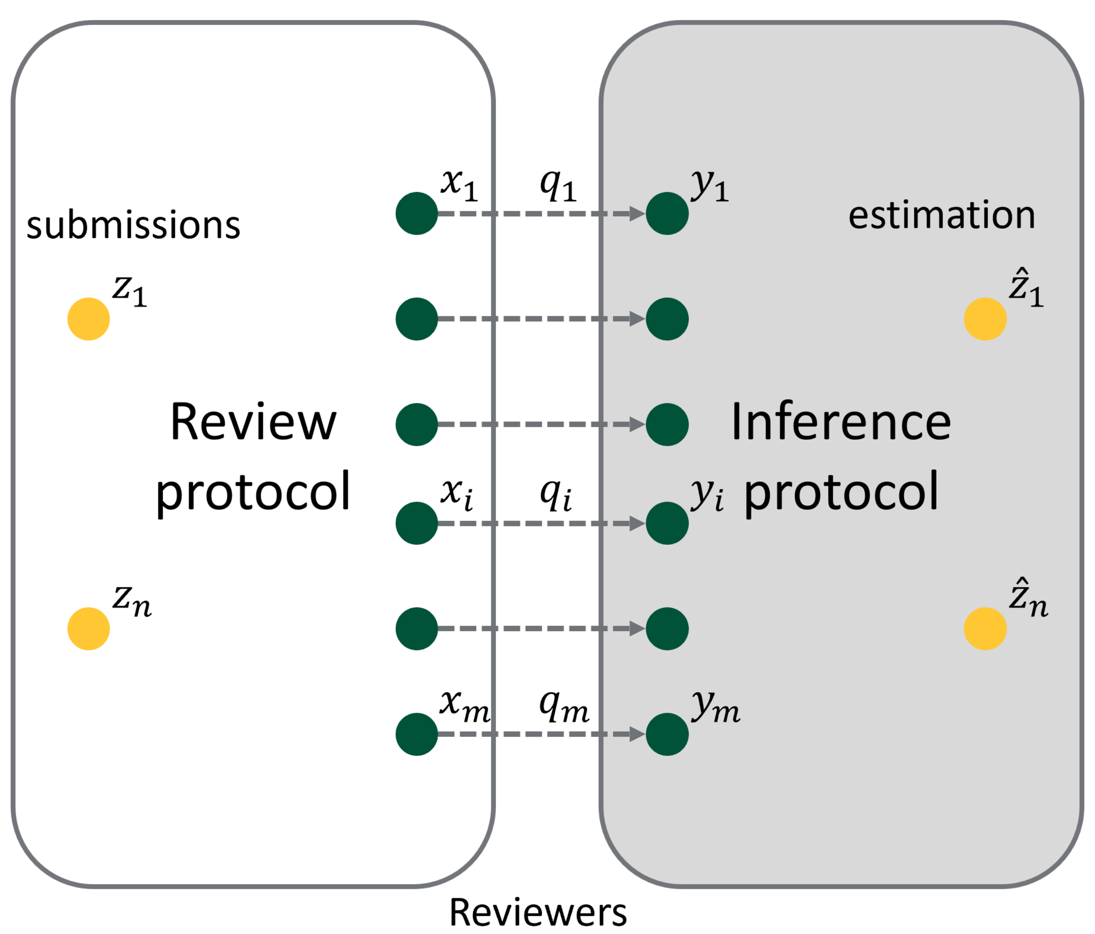
Key Contributions

How can we minimize the *review load ratio*—the number of review tasks per paper—while ensuring error probability at most ϵ?

- Isolated Protocols: Optimal load ratio is $\Theta(\ln n/\epsilon)$ →grows with number of papers n and desired accuracy.
- Joint Protocols: Optimal load ratio is a constant (depends only on reviewer noise) →Joint protocols, e.g., two-phase review, can reduce the review burden.
- Unknown Noise Levels: We analyze how to design efficient protocols without knowing reviewer noise levels.

Conference Review Problem

A conference review problem consists of the number of paper n and the distribution of noise level $\mathcal{D}_{\boldsymbol{q}}$.



- n submissions with quality $z_j \in \{-1, 1\}$ for all $j \in [n]$ uniformly distributed.
- m review tasks each has noise level $q_i \in [0, 1]$, binary question f_i with answer x_i, and a report y_i in $\{-1, 1\}$ so that

$$y_i = \begin{cases} x_i & \text{with probability } 1 - q_i \\ -x_i & \text{otherwise.} \end{cases}$$

- Histogram of noise level $\mathcal{D}_{\boldsymbol{q}}$: quality of PC pool
- Review load ratio $\lambda := \frac{m}{n}$: number of review tasks per paper

Given $(n, \mathcal{D}_{\boldsymbol{q}})$, a conference protocol with load ratio λ consists of

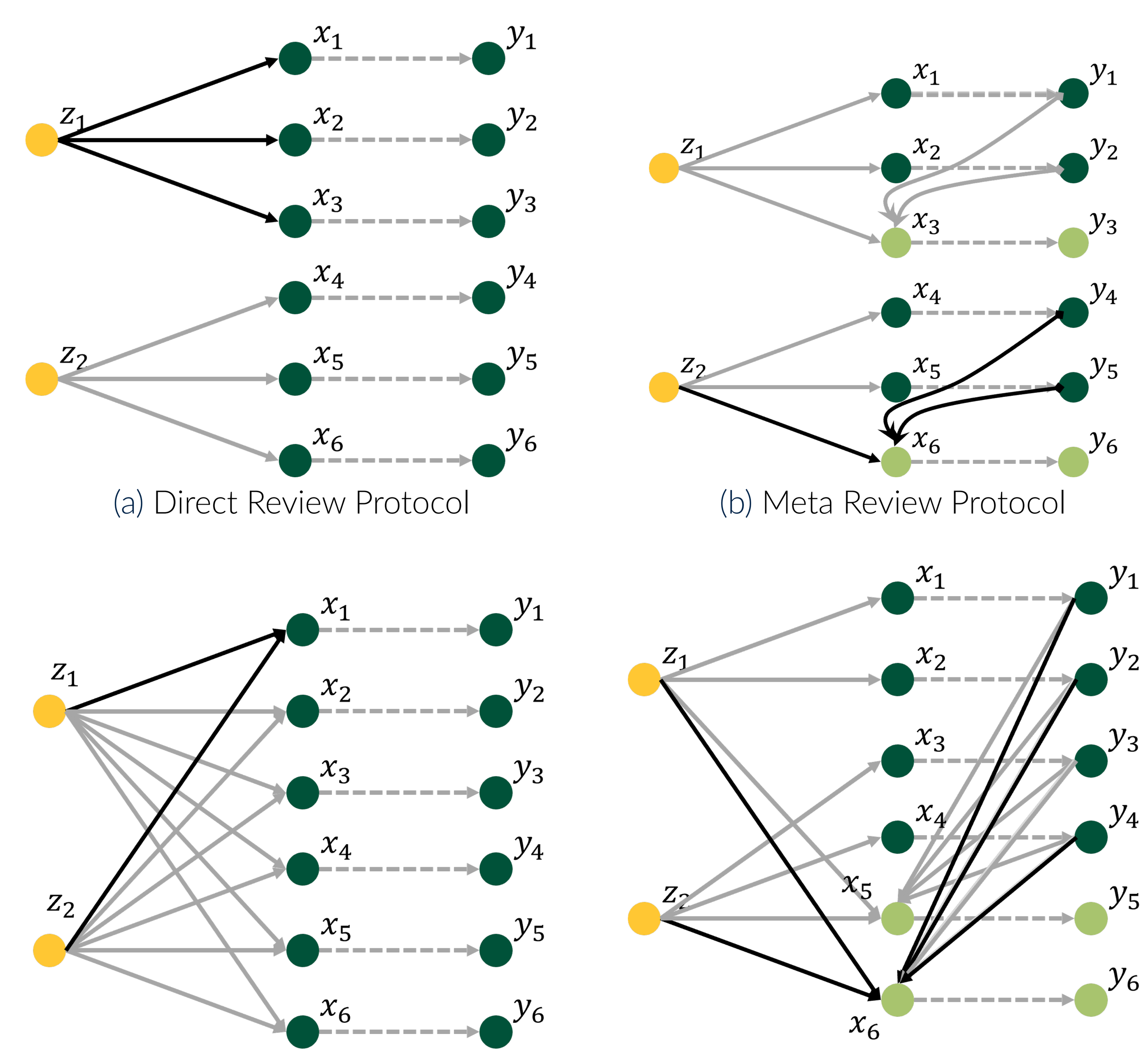
- **Review protocol**: Design $m = \lambda n$ review tasks $\boldsymbol{f} = (f_1, \ldots, f_m)$, and agents submit reports $\boldsymbol{y} = (y_1, \ldots, y_m)$.
- **Inference protocol**: Infer the quality of n papers, $\hat{\boldsymbol{z}}$, using $\boldsymbol{y}$.

Review and Inference Protocols

Review Protocols

We can classify the review protocols based on two key attributes: *isolated* vs *joint* and *parallel* vs *adaptive*.

- A protocol is isolated if each review task pertains to a single paper.
- A protocol is non-adaptive (or parallel) if review tasks are independent of each other and the reports are simultaneously collected.



(a) Direct Review Protocol (b) Meta Review Protocol

(c) Comparison Review Protocol (d) Two-phase Review Protocol

Figure 1. Examples of review protocols and tasks

Examples:

- A task i is a direct review task of paper j, if the report is a noisy estimate of paper j's quality $y_i \overset{q_i}{\leftarrow} f_i(\boldsymbol{z}) = z_j$ and the true answer is $x_i = z_j$
- A task i is second phase review task can be

$$y_i \overset{q_i}{\leftarrow} f_i(\boldsymbol{z}, \boldsymbol{y}_{<i}) = \begin{cases} z_1 \text{ if } y_1 + y_2 \geq y_3 + y_4 \\ z_2 \text{ otherwise.} \end{cases}$$

Inference Protocols

The conference cares about the *error probability*, $\Pr[\hat{\boldsymbol{z}} \neq \boldsymbol{z}]$. Given a review protocol $\boldsymbol{f}$, the optimal inference rule minimizing this error is the maximum a posteriori (MAP) estimator.

A (λ, ϵ)-protocol has the load ratio is upper bounded by λ and error probability is upper bounded by ϵ.

Minimizing the Review Load Ratio

Isolated Protocols

Given $(n, \mathcal{D}_{\boldsymbol{q}})$ and small enough $\epsilon > 0$, an optimal isolated protocol with error probability $P_e = \epsilon$ requires load ratio $\lambda = \Theta\left(\ln \frac{n}{\epsilon}\right)$.

- The reports of direct Review Protocols Blackwell dominates any isolated protocols'.
- Compute the optimal load ratio of direct review protocols.

Joint Protocols

Let $h(q) := q \log_2 \frac{1}{q} + (1-q) \log_2 \frac{1}{1-q}$ and $\lambda^* := \left(1 - \sum_{q \in supp(\mathcal{D}_{\boldsymbol{q}})} \mathcal{D}_{\boldsymbol{q}}(q) h(q)\right)^{-1}$.

For any $\delta, \epsilon > 0$, there is no $(\lambda^* - \delta, \epsilon)$-adaptive protocol, but there exists a $(\lambda^* + \delta, \epsilon)$-nonadaptive conference protocol, when the number of papers n is sufficiently large.

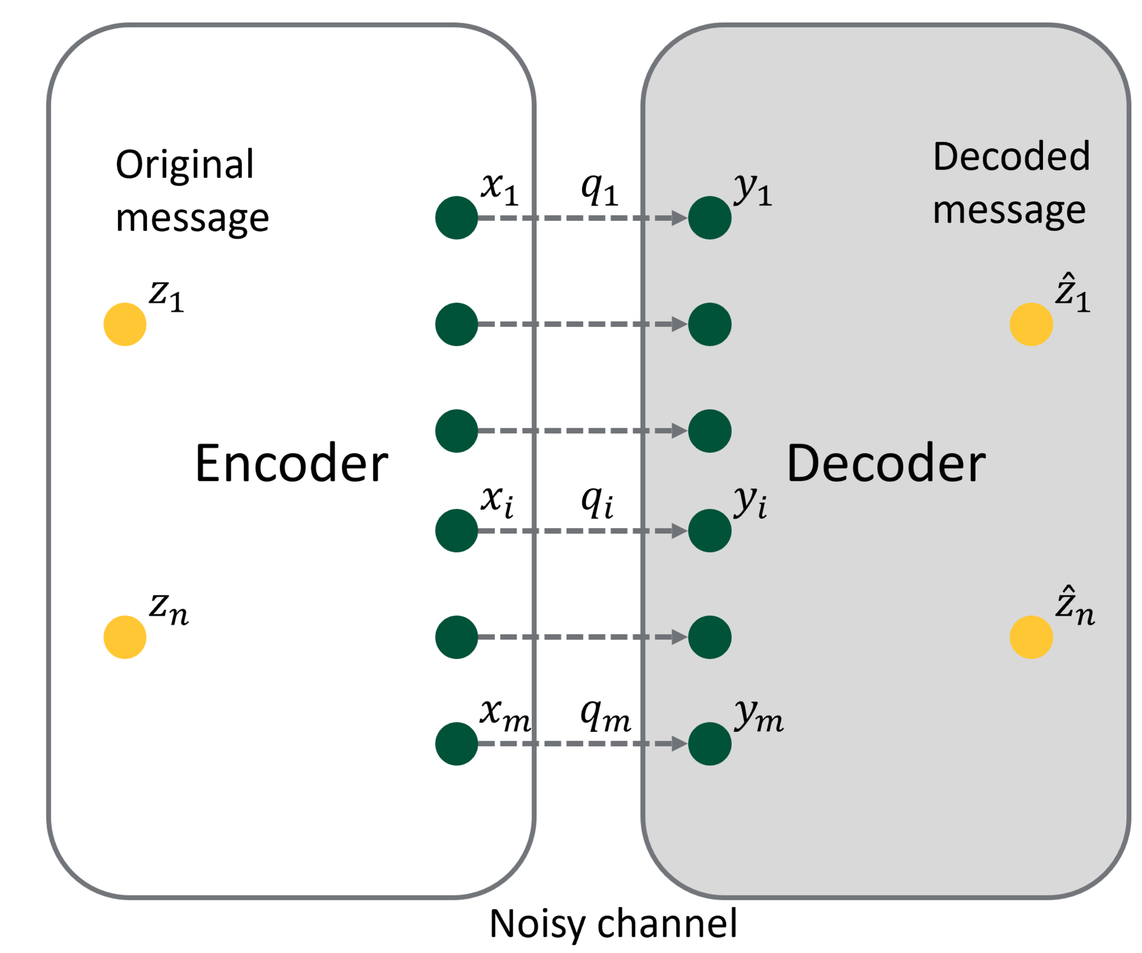
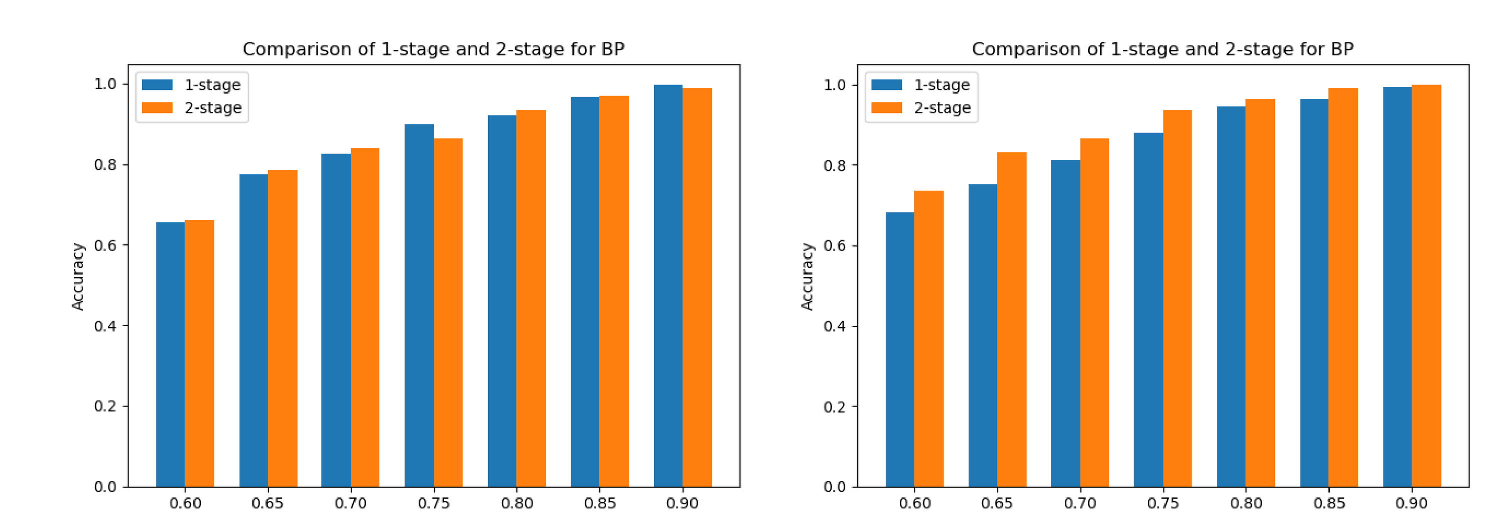


Figure 2. Connection to coding theory

Simulations



(a) Select the bottom 50% most inconsistent papers for the second phase (b) Select the top 75% of papers based on aggregated scores for the second phase

Figure 3. Two-phase review with $1/4$ review tasks in the second phase

The Twenty-Sixth ACM Conference on Economics and Computation