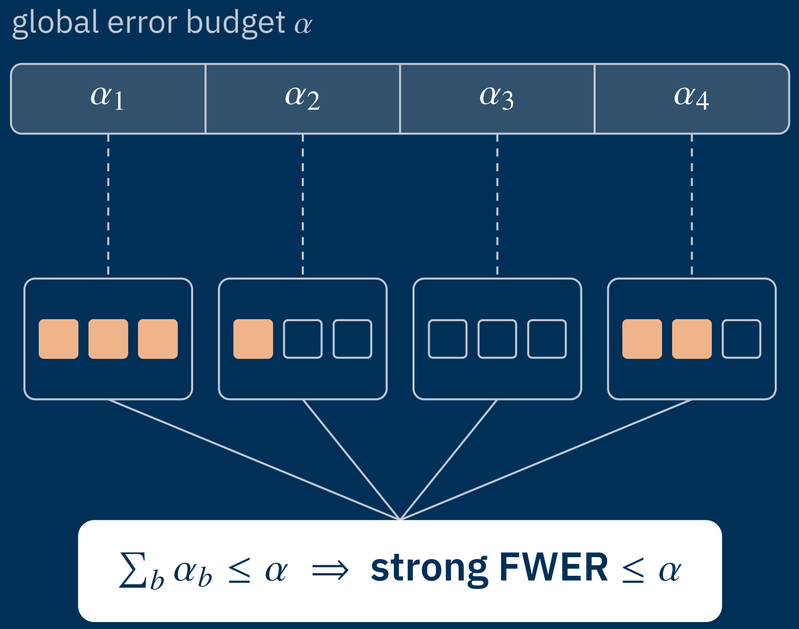
NEURIPS 2026 · MAIN TRACK

BOOST: Power-Optimal Strong-FWER Testing for Block-Structured Multiplicity



Prasanjit Dubey and Xiaoming Huo

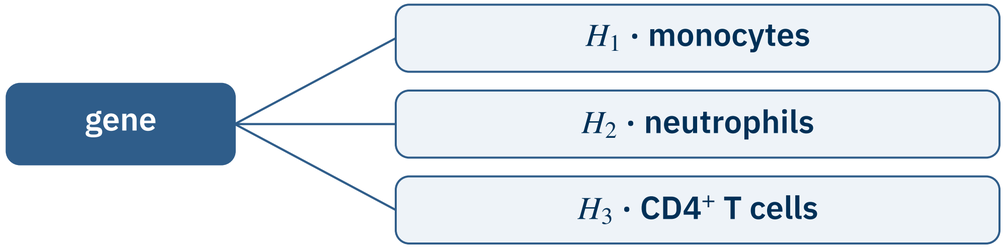
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Confirmatory hypotheses often come in blocks

Fixed by design: each block groups the hypotheses about one scientific unit, before any data are seen

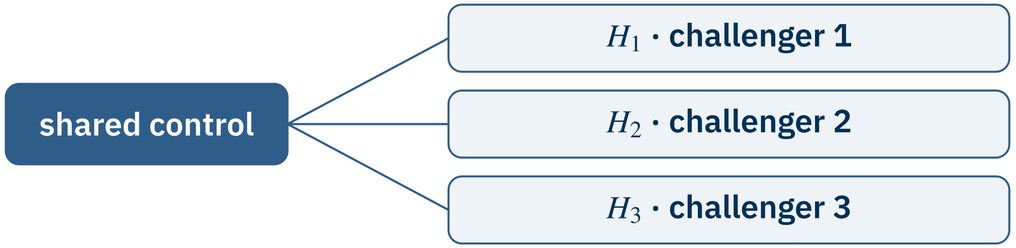
GENOMICS · eQTL MAPPING

One gene, three immune cell types



ONLINE EXPERIMENTS · A/B TESTS

One experiment, three challengers



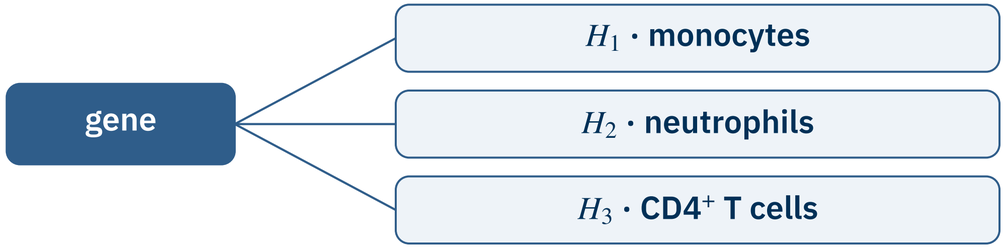
Most powerful tests: Rosset, Heller, Painsky, and Aharoni (2022), JRSS B. Data: Chen et al. (2016); Matias et al. (2021).

Confirmatory hypotheses often come in blocks

Fixed by design: each block groups the hypotheses about one scientific unit, before any data are seen

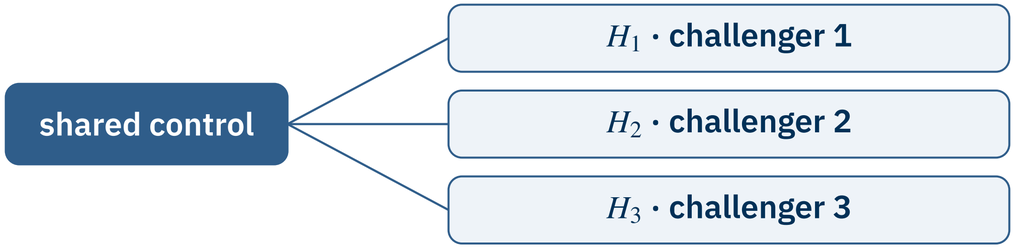
GENOMICS · eQTL MAPPING

One gene, three immune cell types



ONLINE EXPERIMENTS · A/B TESTS

One experiment, three challengers



Strong FWER

$\Pr_{\vec{h}}(\text{any false rejection}) \le \alpha$ for every truth vector $\vec{h}$

One false claim can undo a confirmatory conclusion, so this must hold whichever nulls are true.

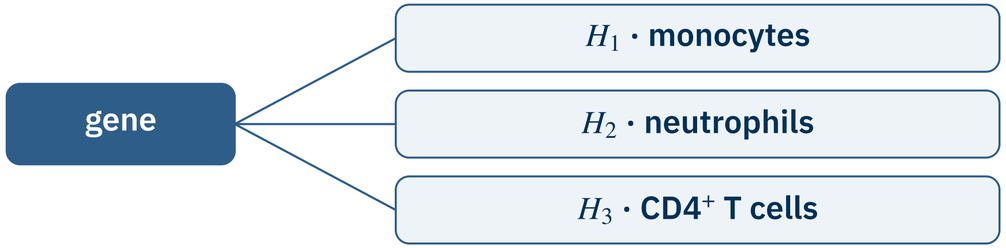
Most powerful tests: Rosset, Heller, Painsky, and Aharoni (2022), JRSS B. Data: Chen et al. (2016); Matias et al. (2021).

Confirmatory hypotheses often come in blocks

Fixed by design: each block groups the hypotheses about one scientific unit, before any data are seen

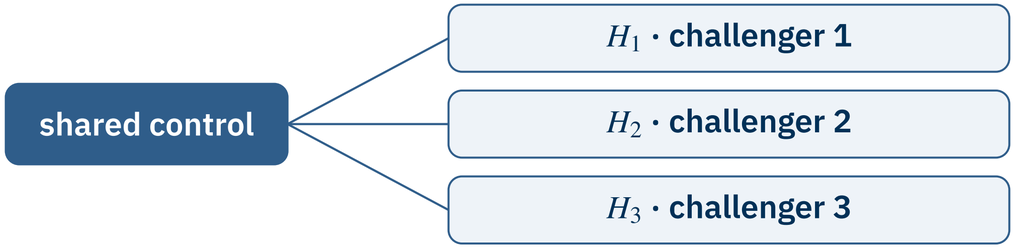
GENOMICS · eQTL MAPPING

One gene, three immune cell types



ONLINE EXPERIMENTS · A/B TESTS

One experiment, three challengers



Strong FWER

$\Pr_{\vec{h}}(\text{any false rejection}) \le \alpha$ for every truth vector $\vec{h}$

One false claim can undo a confirmatory conclusion, so this must hold whichever nulls are true.

Standard FWER procedures

Holm and Hommel use generic thresholds rather than optimizing power for a specified block model.

Most powerful strong-FWER tests

Direct power optimization requires integration in a dimension that grows with the number of hypotheses.

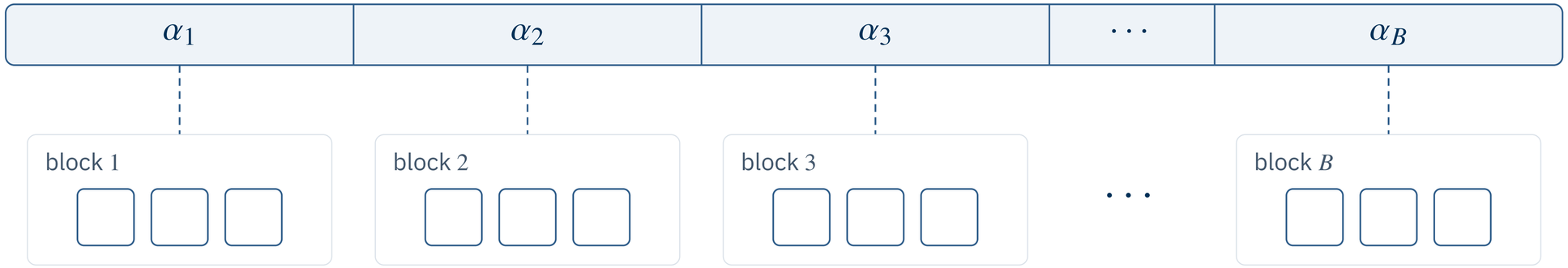
Most powerful tests: Rosset, Heller, Painsky, and Aharoni (2022), JRSS B. Data: Chen et al. (2016); Matias et al. (2021).

BOOST: split the budget, certify each block, compose

Validity first: the outer guarantee accepts any locally certified rule

global budget α

Example:
$\alpha = 5\%$, $B = 10$
$\alpha_b = 0.5\%$ each

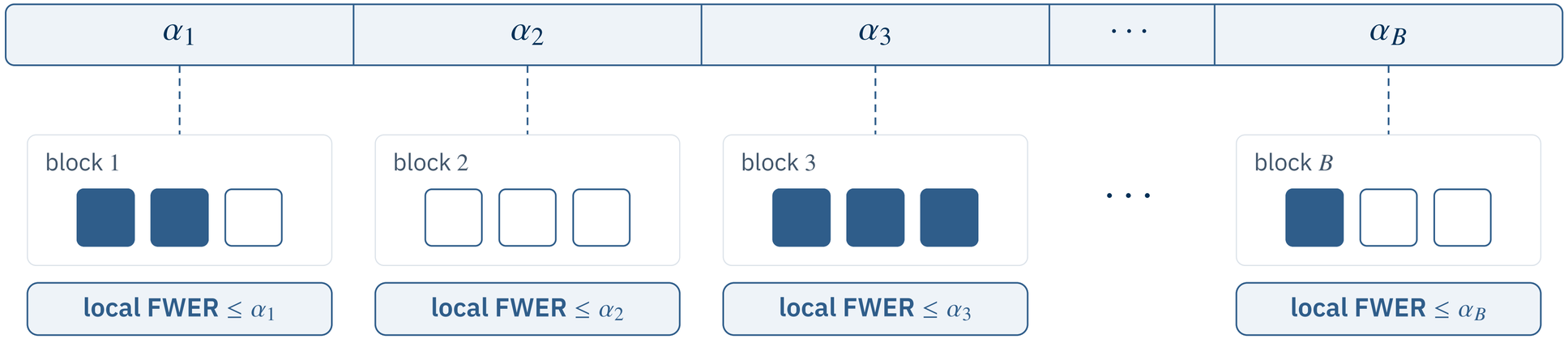


BOOST: split the budget, certify each block, compose

Validity first: the outer guarantee accepts any locally certified rule

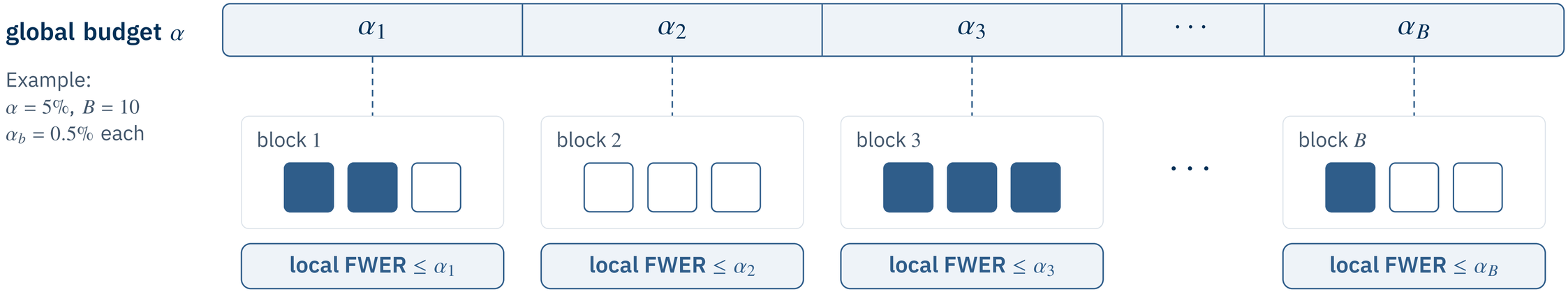
global budget α

Example:
$\alpha = 5\%$, $B = 10$
$\alpha_b = 0.5\%$ each



BOOST: split the budget, certify each block, compose

Validity first: the outer guarantee accepts any locally certified rule



THEOREM 3.3 · LOCAL-TO-GLOBAL COMPOSITION

If each block's rule uses only its p-values and has local strong FWER $\leq \alpha_b$ under the block's actual law, and $\sum_b \alpha_b \leq \alpha$, then for every truth vector $\vec{h}$:

$$\mathrm{FWER}_{\vec{h}} = \Pr_{\vec{h}}\Big(\bigcup_b E_b\Big) \leq \sum_b \Pr_{\vec{h}}(E_b) \leq \sum_b \alpha_b \leq \alpha$$

E_b: block b makes at least one false rejection.

No independence or exchangeability across blocks is assumed.

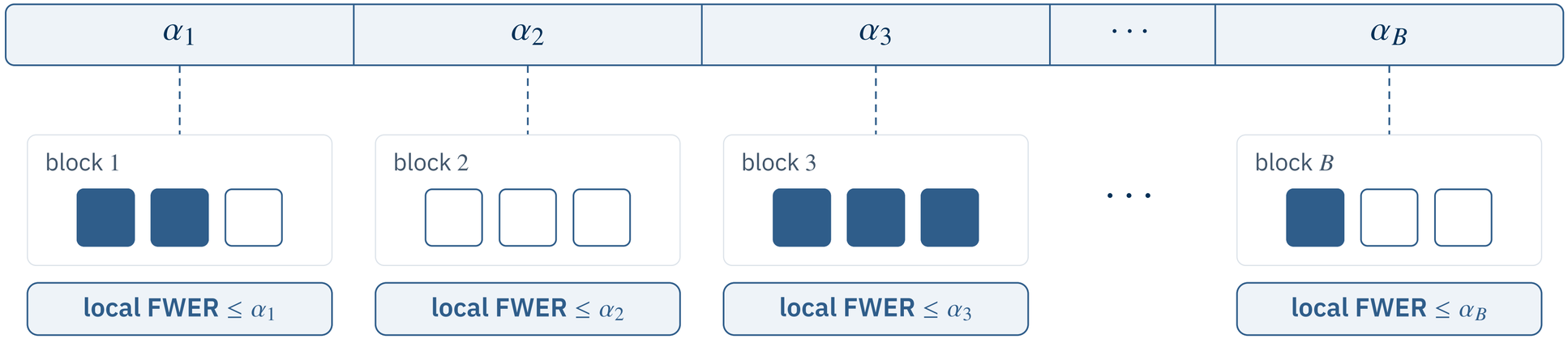
Theorem 3.3 and Table 2 of the paper ($K = 30$, $B = 10$, $\theta = -2$; strongest partial-null configuration; block laws fixed; 5×10^5 runs per ρ).

BOOST: split the budget, certify each block, compose

Validity first: the outer guarantee accepts any locally certified rule

global budget α

Example:
$\alpha = 5\%$, $B = 10$
$\alpha_b = 0.5\%$ each



THEOREM 3.3 · LOCAL-TO-GLOBAL COMPOSITION

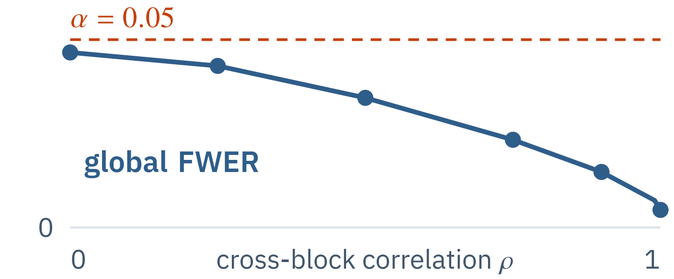
If each block's rule uses only its p-values and has local strong FWER $\le \alpha_b$ under the block's actual law, and $\sum_b \alpha_b \le \alpha$, then for every truth vector $\vec{h}$:

$$\mathrm{FWER}_{\vec{h}} = \Pr_{\vec{h}}\Big(\bigcup_b E_b\Big) \le \sum_b \Pr_{\vec{h}}(E_b) \le \sum_b \alpha_b \le \alpha$$

E_b: block b makes at least one false rejection.

No independence or exchangeability across blocks is assumed.

SIMULATION · CORRELATED BLOCKS



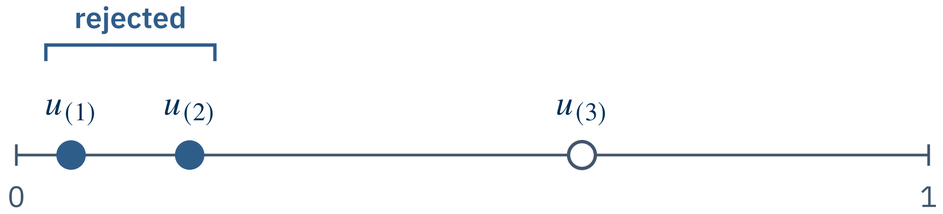
Theorem 3.3 and Table 2 of the paper ($K = 30$, $B = 10$, $\theta = -2$; strongest partial-null configuration; block laws fixed; 5×10^5 runs per ρ).

Power: the best rule per block, the best split

Objective: average power Π_{avg}, the expected fraction of hypotheses rejected under the complete alternative

LOCAL RULE · BLOCK OF THREE

Sort the p-values; reject a prefix



A precomputed rule chooses how many of the smallest p-values to reject, under local strong FWER control.

Product model: within each block, independent p-values, uniform under the null, with a common nonincreasing alternative density.

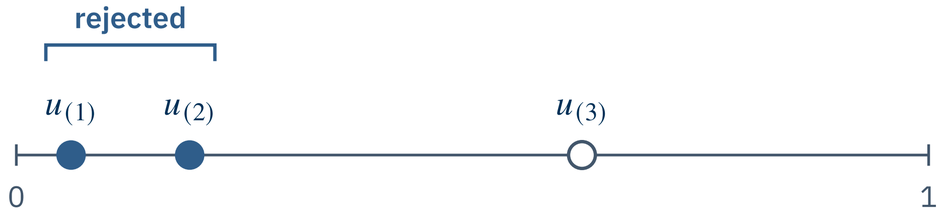
Size-three rule: Dubey and Huo (2025), arXiv:2512.14131. Prefix reduction: Rosset et al. (2022), Theorem 1. Weights $w_b = s_b/K$; Theorem 3.7 uses supergradients.

Power: the best rule per block, the best split

Objective: average power Π_{avg}, the expected fraction of hypotheses rejected under the complete alternative

LOCAL RULE · BLOCK OF THREE

Sort the p-values; reject a prefix

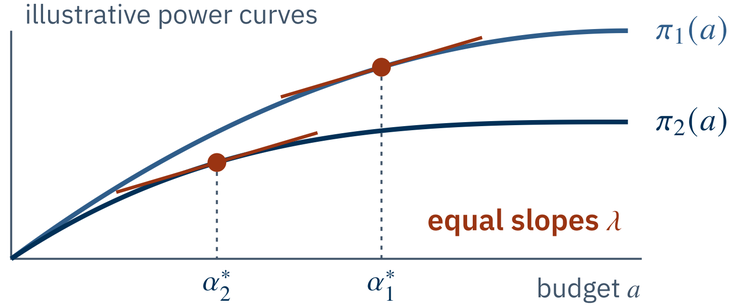


A precomputed rule chooses how many of the smallest p-values to reject, under local strong FWER control.

Product model: within each block, independent p-values, uniform under the null, with a common nonincreasing alternative density.

BUDGET SPLIT · THEOREM 3.7

Equalize the marginal value of budget



$w_b\,\pi_b'(\alpha_b^*) = \lambda$ on every funded block: a one-dimensional search

Identical blocks: α/B each.

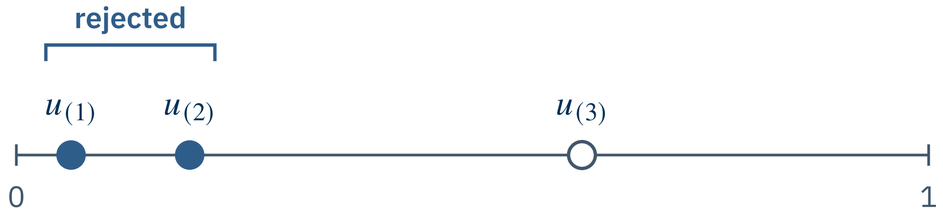
Size-three rule: Dubey and Huo (2025), arXiv:2512.14131. Prefix reduction: Rosset et al. (2022), Theorem 1. Weights $w_b = s_b/K$; Theorem 3.7 uses supergradients.

Power: the best rule per block, the best split

Objective: average power Π_{avg}, the expected fraction of hypotheses rejected under the complete alternative

LOCAL RULE · BLOCK OF THREE

Sort the p-values; reject a prefix

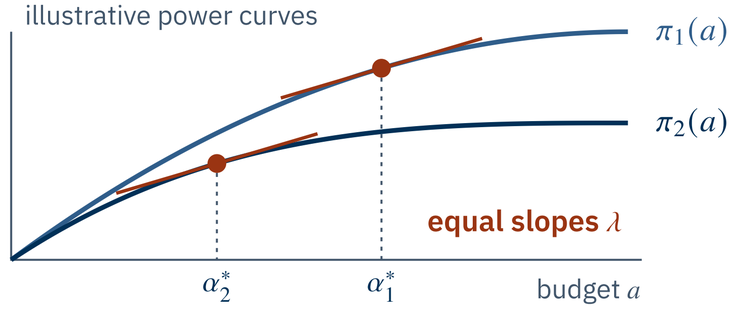


A precomputed rule chooses how many of the smallest p-values to reject, under local strong FWER control.

Product model: within each block, independent p-values, uniform under the null, with a common nonincreasing alternative density.

BUDGET SPLIT · THEOREM 3.7

Equalize the marginal value of budget



$w_b\,\pi_b'(\alpha_b^*) = \lambda$ on every funded block: a one-dimensional search

Identical blocks: α/B each.

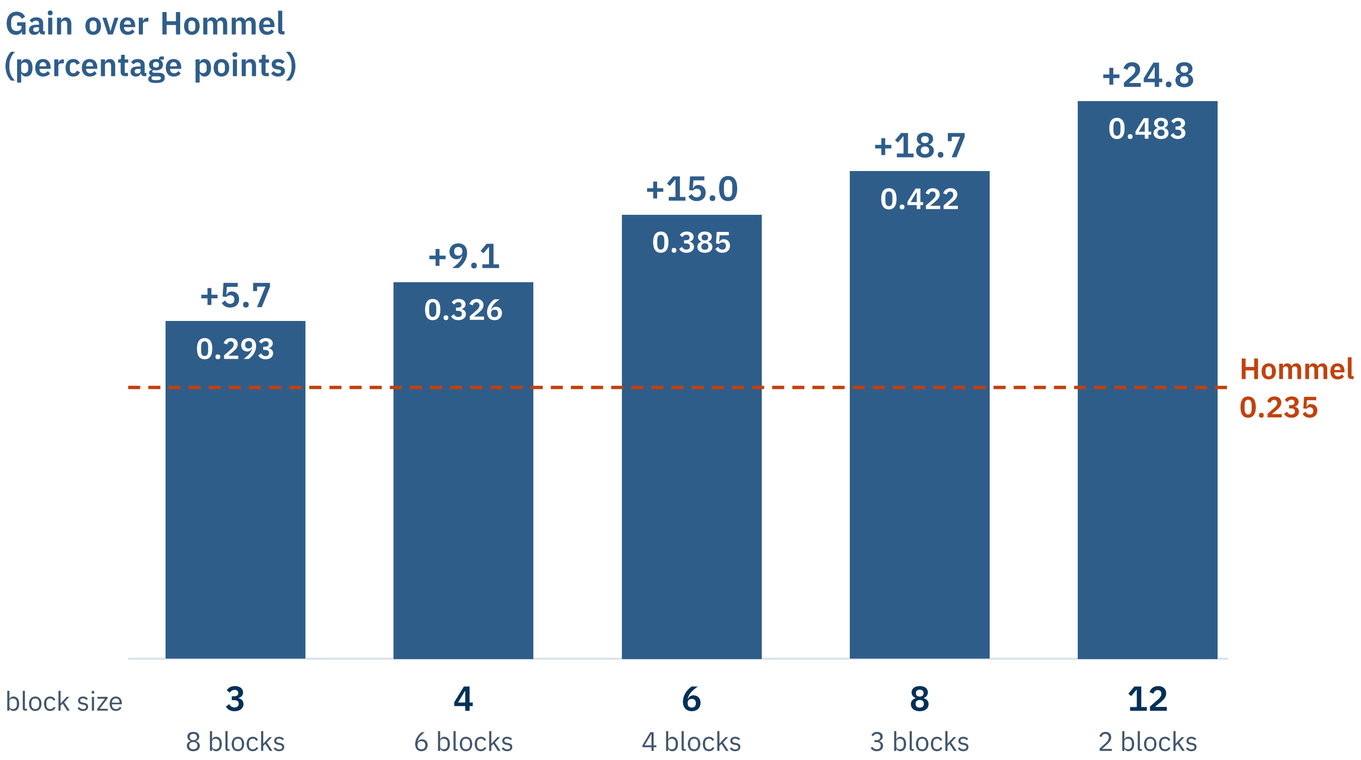
Theorem 3.6: under the product model, exact local optimizers and the optimal split maximize Π_{avg} in the block-separable class.

In practice, rules are numerical approximations validated on held-out draws; once precomputed, BOOST runs in $O(K)$ for bounded block size.

Size-three rule: Dubey and Huo (2025), arXiv:2512.14131. Prefix reduction: Rosset et al. (2022), Theorem 1. Weights $w_b = s_b/K$; Theorem 3.7 uses supergradients.

Larger blocks improve power in this simulation

24 hypotheses in every column, $\alpha = 0.05$, $\theta = -2$: average power of BOOST vs. Hommel, the strongest standard rule here

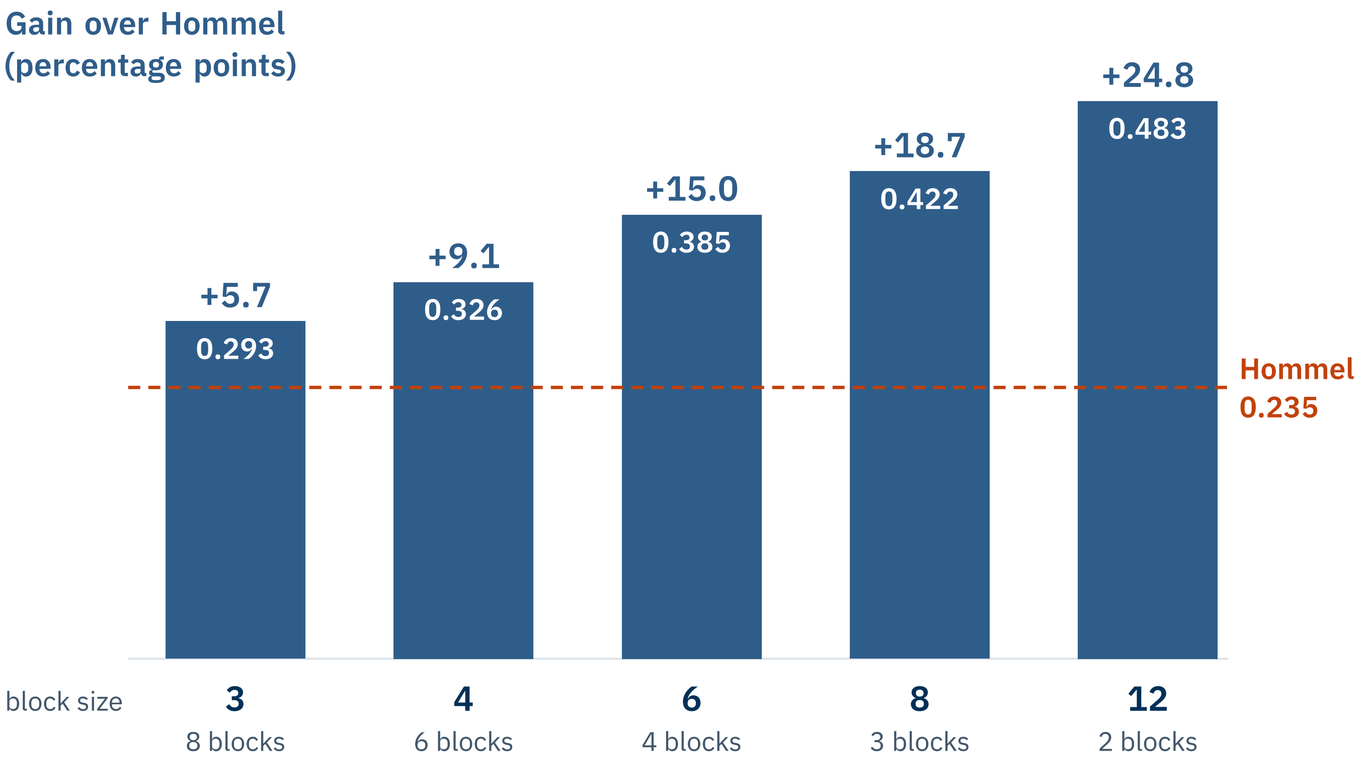


Every FWER upper bound is below 0.05.

Tables 1, 3, and 7 of the paper. Blocks of size 4 to 12 use the local rule of Dubey and Huo (2026, JCGS). FWER bounds are simultaneous 95% confidence bounds.

Larger blocks improve power in this simulation

24 hypotheses in every column, $\alpha = 0.05$, $\theta = -2$: average power of BOOST vs. Hommel, the strongest standard rule here

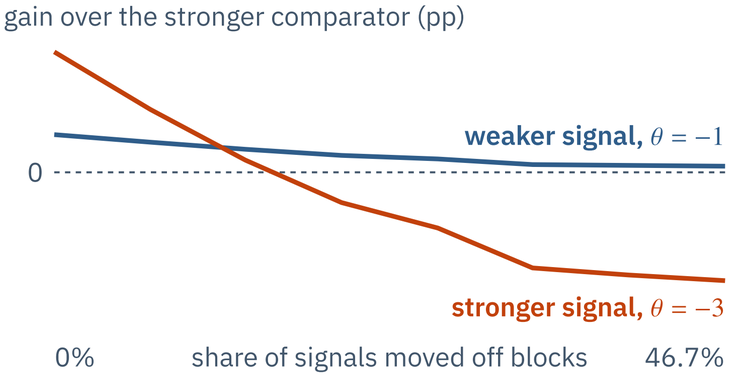


Every FWER upper bound is below 0.05.

ALIGNMENT · K = 30, BLOCKS OF THREE

Gains depend on signal alignment

Aligned signals: +0.7 to +10.1 pp over the stronger of Hommel and Šidák step-down, at every tested signal strength (Table 1).



Weak to moderate signals: the gain stays positive. Strong signals: it can reverse (Table 7).

Tables 1, 3, and 7 of the paper. Blocks of size 4 to 12 use the local rule of Dubey and Huo (2026, JCGS). FWER bounds are simultaneous 95% confidence bounds.

Two applications in two domains

Working-model illustrations: local rules fitted on one fold of blocks, applied to the other

GENOMICS · BLUEPRINT cis-eQTL

1,000 test genes × 3 immune cell types

Individual hypotheses rejected



+3.7% more rejections than Holm

Genes with all three rejected: **103** vs. 92

ONLINE EXPERIMENTS · UPWORTHY ARCHIVE

2,282 test experiments × 3 challengers

Individual hypotheses rejected



+37% more rejections than Holm

Experiments with all three rejected: **7** vs. 3

Table 4 and Appendix E of the paper. Data: Chen et al. (2016) via the eQTL Catalogue (Kerimov et al., 2021); Upworthy Research Archive (Matias et al., 2021).

Two applications in two domains

Working-model illustrations: local rules fitted on one fold of blocks, applied to the other

GENOMICS · BLUEPRINT cis-eQTL

1,000 test genes × 3 immune cell types

Individual hypotheses rejected



+3.7% more rejections than Holm

Genes with all three rejected: **103** vs. 92

ONLINE EXPERIMENTS · UPWORTHY ARCHIVE

2,282 test experiments × 3 challengers

Individual hypotheses rejected



+37% more rejections than Holm

Experiments with all three rejected: **7** vs. 3

Validated under the fitted product laws: composed FWER bounds 0.0455 (BLUEPRINT) and 0.0473 (Upworthy)

Both are below the 0.049 testing budget. A confirmatory analysis would certify each block under its actual within-block joint law; Theorem 3.3 then applies unchanged.

Table 4 and Appendix E of the paper. Data: Chen et al. (2016) via the eQTL Catalogue (Kerimov et al., 2021); Upworthy Research Archive (Matias et al., 2021).

Takeaways

1 **Validity is modular.**
Locally certified blocks compose into strong FWER under any cross-block dependence (Theorem 3.3).

2 **Power is optimized within the block-separable class.**
Under the product model, exact local optimizers and the optimal split maximize average power (Theorems 3.6 and 3.7).

3 **Practical to compute.**
Precompute reusable local rules, then apply BOOST in $O(K)$ time for bounded block size.

Thank you. The paper provides the proofs, further details, and the code.