# Block delay under random linear combinations on a random access erasure collision channel

Nan Xie and Steven Weber
Department of Electrical and Computer Engineering
Drexel University, Philadelphia, PA 19104

***Abstract*—We consider the delay associated with transmissions of random linear combinations (RLC) of a block of packets, each held at a collection of independent transmitters and sent over an erasure and collision multiple access channel to a receiver. Our figure of merit is the expected time until the receiver recovers the block of packets. The transmitters make transmission decisions independently in space and time, and these contention probabilities are the design variable in the problem. The erasure collision channel consists of independent and identically distributed erasure channels between each transmitter and the receiver, with the property that multiple messages arriving at the receiver collide. We study the block delay under both RLC and for the case where each active transmitter selects a packet uniformly at random. Our main result identifies the contention probability vector on the erasure collision channel that maximizes the probability of message reception at the receiver in a time slot, and thus minimizes the expected block delay per packet.**

***Index Terms*—random linear combinations/coding; block delay; random access; collision channel; erasure channel.**

## I. Introduction

The primary focus of this paper is to study the impact of network coding on the delay incurred when multiple transmitters contend for a wireless channel in a random-access manner to send packets over an erasure channel, and multiple packets arriving at a receiver collide, resulting in packet loss. Network coding takes the form of random linear combinations (RLC) of the source packets held at each transmitter, and is compared against a scheme where each contending transmitter randomly selects a packet (RSP) in each time slot. Delay is measured as the time until the receiver recovers the packet block.

Multiple packets are required for there to be a block delay improvement of RLC over RSP since, as is well-understood, this improvement arises from the fewer redundant receptions of RLC compared with RSP. Our prior work [1] (and others) has studied the delay in the related problem consisting of a *single* transmitter and *multiple* receivers, whereas in this work we consider the scenario of *multiple* transmitters and a *single* receiver. Although modeling cooperative transmitters is of natural interest, in this work we suppose the transmitters to be operating *independently* (but nonetheless are assumed to hold in common the packets to be sent to the receivers). Non-cooperative transmissions of shared information to a common receiver arises in settings where the transmitters are unwilling or unable to cooperate with one another (e.g., when owned by separate carriers) but the information comes from a common source (e.g., multicast from the source to the transmitters).

This research has been supported by the National Science Foundation under award #CCF-1421828.

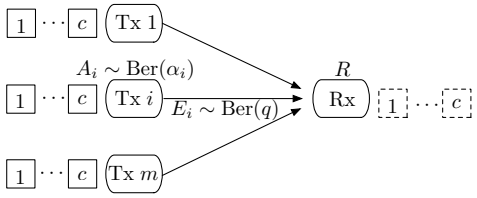


Fig. 1: The system model: $m$ transmitters, each holding a common block of $c$ packets, make transmission attempt decisions in each time slot (denoted by the independent RVs $\mathbf{A} = (A_1, \ldots, A_m)$, with $A_i \sim \mathrm{Ber}(\alpha_i)$), and send messages over an *erasure collision channel* (with IID erasure RVs $\mathbf{E} = (E_i \sim \mathrm{Ber}(q), i \in [m])$ from transmitter $i$), to a receiver desiring the block of packets. Transmitters use either random linear combinations (*RLC*) of the packets or randomly selected packets (*RSP*) for transmission in each slot. Reception is indicated by the RV $R$, where multiple packet arrivals at the receiver result in a collision, and thus no reception. Transmissions end as soon as the receiver recovers the block.

There are two key aspects of real-world wireless channels we wish to capture in our model: erasures and collisions. The erasure aspect captures the inherent uncertainty and unreliability of a wireless channel, while the collision aspect captures the inability of the receiver to obtain information from simultaneous transmissions. The combined erasure collision channel provides an interesting tension in selecting the contention probabilities for the transmitters: erasures encourage more frequent transmission to avoid packet loss in the channel, while collisions encourage less frequent transmission to avoid packet loss at the receiver. We optimize this tradeoff in Thm. 1.

### A. Related work

Refer to [1] for references on delay performance of network coding (in particular [4][5][8][9] therein). The general problem of computation over multiple access channels has been studied from an information theoretic perspective [2]. So called network-coded multiple access (NCMA) [3] [4] exhibits throughput gains by applying physical layer network coding (PNC) and multiuser detection (MUD) to wireless multiple-access. For additional references on applying PNC in a random

access setting, see e.g., [5] [6]. It should be noted much work on random access is about *collision resolution*. In [7], which builds upon ZigZag decoding (see [4] therein), collisions are viewed as linear combinations of the original packets and improvements of delay performance have been shown. Finally, the paradigm of *coded random access* [8] is based on the idea of successive cancellation and leverages the tool of *codes on graphs*: thus the design space of random access protocols has been expanded and performance improvements shown.

### B. Contributions

Our contribution is to optimize the expected block delay in the context of the multi-transmitter to receiver erasure collision channel, using both random linear combinations and randomly selected packets. Our main result, Thm. 1, gives the optimal contention probability vector for maximizing the probability of successful reception at the receiver, which is a tractable but nontrivial nonlinear optimization problem.

## II. Model

The model is illustrated in Fig. 1. Define $[k] \equiv \{1, \ldots, k\}$ for $k \in \mathbb{N}$. $\mathbf{e}_i$ denotes a unit vector with 1 in position $i$. RV and IID stand for "random variable" and "independent & identically distributed" respectively.

We consider a slotted-time erasure collision channel serving $m$ transmitters (e.g., base stations) and a receiver (e.g., a mobile user). There is a block of $c$ identically sized packets, labeled $[c]$, and we suppose each transmitter is able to transmit exactly one packet, or one linear combination of packets, in each time slot. Each of the $m$ transmitters *holds* all $c$ packets, and the receiver *wants* all $c$ packets.

**Delay.** The RV $T$ denotes the *block delay*: the number of time slots until the receiver holds all $c$ packets, or (equivalently for our model) holds $c$ (linearly independent) RLCs.

**Erasure channel.** The channel is a slotted-time erasure collision channel with $q \in (0, 1)$ the nonerasure probability from each transmitter $i$ to the receiver. In each time slot $t$ there is an independent (in both time and space) realization of an erasure process, captured by an $m$ vector of Bernoulli (denoted by Ber) RVs $\mathbf{E} = (E_1, \ldots, E_m)$ where $E_i \sim \mathrm{Ber}(q)$, and $E_i = 1$ (0) represents a nonerasure (erasure) respectively.

**Random access transmissions.** The transmitters employ random access in that in each time slot each of the $m$ transmitters makes an independent random decision $A_i \sim \mathrm{Ber}(\alpha_i)$ of whether or not to transmit (contend), with $\alpha_i \in [0, 1]$ the contention probability. Observe the contention decision RVs $\mathbf{A} = (A_1, \ldots, A_m)$ are independent in time and space, but not identically distributed. The contention probabilities $\boldsymbol{\alpha} \equiv (\alpha_i, i \in [m])$ is the control in the system to be designed.

**Collision receptions.** The transmission decisions $\mathbf{A}$ and erasure pattern $\mathbf{E}$ in a time slot determine which transmissions, if any, reach the receiver. The collision aspect of the channel asserts successful receptions occur when exactly one message is received at the receiver, and these successes are denoted by the reception RV $R$. Note that multiple transmissions imply neither collision nor successful reception, since any or all transmissions may be erased.

**RLC vs. RSP.** The transmission paragraph above clarified *when* transmitters contend, but not *which* packets are sent. We consider two different packet selection models, named RLC, e.g., network coding, and RSP. Under RLC, each transmitter that elects to contend selects $c$ coefficients uniformly at random from a large[1] field and uses these to form a random linear combination of the $c$ packets, which it then transmits. Under RSP, each transmitter that elects to contend selects one packet from the available block of $c$ packets uniformly at random, each with probability $1/c$. *The key questions of interest in this paper are $i$) how to select the contention probabilities $\boldsymbol{\alpha}$ as a function of $(c, m, q)$ for both RLC and RSP, and $ii$) how the optimized RLC and RSP delays differ and scale in $(c, m, q)$.*

**Feedback.** Each transmitter repeatedly contends as discussed above until the receiver recovers the block, i.e., either receives $c$ (linearly independent) RLCs, or receives each of the $c$ packets under RSP. When the receiver recovers the block it immediately notifies the transmitters of this fact with a single broadcast bit. Aside from this, we assume the receiver does not otherwise respond with instantaneous feedback. We assume this in order to minimize the required feedback bandwidth[2].

## III. Block delay analysis under RLC and RSP

We derive expressions for the expected block delay under RLC in §III-A and under RSP in §III-B, then characterize the optimal contention probabilities in §III-C.

### A. Delay under RLC

As we assume the field size is sufficiently large so that the possibility of dependent combinations may be ignored, it follows that the relevant state of the receiver under RLC is simply the number of successfully received combinations. The state space is $\mathcal{X} = \{0, \ldots, c\}$, with transient states $\mathcal{T} = \{0, \ldots, c-1\}$ and absorbing (final) state $\mathcal{A} = \{c\}$. The state advances from $x$ to $x+1$ in each time slot in which there is a reception. As the probability of reception is independent of the state and therefore constant in time, it follows that the random block delay $T_{\mathrm{RLC}}$ is a negative binomial RV, counting the number of time slots required to acquire $c$ successes where each attempt is successful with some (to be determined) reception probability $f \in (0, 1)$, i.e., $T_{\mathrm{RLC}} \sim \mathrm{NegBin}(c, f)$, with $\mathbb{P}(T_{\mathrm{RLC}} = n) = \binom{n-1}{c-1}(1-f)^{n-c} f^c$ for $n \in \{c, c+1, \ldots\}$, and expected block delay per packet $\frac{\mathbb{E}[T_{\mathrm{RLC}}]}{c} = \frac{1}{f}$. The RLC expected block delay per packet, $\mathbb{E}[T_{\mathrm{RLC}}]/c$, is minimized by maximizing the reception probability $f = f(\boldsymbol{\alpha})$. The following result gives the reception probability.

[1] The field size is assumed to be suitably large to justify ignoring the possibility of $i$) selecting the all zero vector and $ii$) two random linear combinations being not linearly independent. Analysis of these phenomena is possible, but is omitted here.

[2] Consider instead a RSP model where the receiver broadcasts the packet index it successfully received that slot, if any. In this case each transmitter would know the state of the receiver, and it is natural to then adapt the contention probability and packet selection to this state via dynamic programming. Although interesting, this is outside the scope of this paper.

**Proposition 1.** *The RLC message reception probability is, with* $\mathbf{p} = (p_i, i \in [m])$ *and* $p_i = \alpha_i q$*, given by:*

$$f(\mathbf{p}) = \mathbb{E}[R] = \mathbb{P}(R = 1) = \sum_{i \in [m]} p_i \prod_{j \neq i} (1 - p_j). \quad (1)$$

*Proof:* In a given time slot suppose $\mathbf{A} = (A_i, i \in [m])$ are the random transmission attempt decisions, with $\mathbf{A}$ independent and $A_i \sim \mathrm{Ber}(\alpha_i)$, and let $\mathbf{E} = (E_i, i \in [m])$ be the IID erasure realizations, with $E_i \sim \mathrm{Ber}(q)$. Form the independent RVs $\mathbf{S} = (S_i, i \in [m])$ with $S_i = A_i E_i$ indicating reception from transmitter $i$. Clearly, $S_i \sim \mathrm{Ber}(p_i)$, with $p_i = \alpha_i q \in [0, q]$ and $\mathbf{p} = (p_i, i \in [m])$. Observe $S_1 + \cdots + S_m$ is the random number of packets arriving at the receiver, and as such the collision channel model requires the indicator RV of a successful reception equal $R = \mathbf{1}\{S_1 + \cdots + S_m = 1\}$. The probability of a successful reception is therefore (1). ∎

### B. Randomly selected packet (RSP)

As the packet selection process is uniformly random, it follows that the probability of the receiver obtaining a new packet in a time slot depends upon the set of currently obtained packets only through the number of such packets. As such, the state space is again $\mathcal{X} = \{0, \ldots, c\}$, with absorbing state $\mathcal{A} = \{c\}$ and transient states $\mathcal{T} = \{0, \ldots, c-1\}$, with state transitions from $x$ to $x + 1$ in each time slot in which the receiver obtains a *new* packet (unlike the case of RLC, where the state advances for *any* reception). As the transition probability is now state dependent, it follows that the random block delay $T_{\mathrm{RSP}}$ may be expressed as $T_{\mathrm{RSP}} = T_1 + \cdots + T_c$, where $T_k \sim \mathrm{Geo}(f_k)$ is a geometric RV with parameter $f_k$ (defined below) representing the number of time slots between the reception of new packet $k-1$ and new packet $k$. Note the $(T_1, \ldots, T_c)$ are independent by the governing model assumptions. It follows that $\mathbb{E}[T_{\mathrm{RSP}}] = \sum_{k=1}^{c} f_k^{-1}$. The success probability is given in the following result.

**Proposition 2.** *Having received* $k-1$ *of the* $c$ *packets, the probability of receiving a* new *message under RSP is, with* $\mathbf{p} = (p_i, i \in [m])$ *and* $p_i = \alpha_i q$*, given by (with* $f(\mathbf{p})$ *in (1)):*

$$f_k = f(\mathbf{p}) \left(1 - \frac{k-1}{c}\right), \; k \in [c]. \quad (2)$$

*Proof:* Suppose the receiver holds $k-1$ of the $c$ packets. The probability $f_k$ of a new reception in a time slot is (2) since $i)$ $f(\mathbf{p})$ is the probability of receiving a (single) packet, and $ii)$ $1 - (k-1)/c$ is the probability the received packet is *new*, i.e., not one of the $k-1$ packets already held. ∎

Therefore the RSP expected block delay per packet is

$$\frac{\mathbb{E}[T_{\mathrm{RSP}}]}{c} = \frac{1}{c} \sum_{k=1}^{c} f_k^{-1} = \frac{1}{c f(\mathbf{p})} \sum_{k=1}^{c} \left(1 - \frac{k-1}{c}\right)^{-1} = \frac{H(c)}{f(\mathbf{p})}. \quad (3)$$

Here, $H(c) \equiv \sum_{k=1}^{c} \frac{1}{k}$ is the $c$th harmonic number, with $H(c) \approx \log c + \gamma$, for $\gamma \approx 0.577$ the Euler-Mascheroni constant. In particular, (3) ensures the expected RSP block delay per packet is minimized by maximizing the reception probability $f(\mathbf{p})$. Furthermore, for any choice of $\mathbf{p}$, the ratio of the expected delays per packet of RSP over RLC equals

$$\frac{\mathbb{E}[T_{\mathrm{RSP}}]/c}{\mathbb{E}[T_{\mathrm{RLC}}]/c} = H(c). \quad (4)$$

That is, the performance ratio $i)$ is independent of the choice of $\boldsymbol{\alpha}$ and in fact independent of the number of transmitters $m$, and $ii)$ grows logarithmically in the blocklength $c$.

### C. Optimal contention for maximum reception probability

The main result of the paper, Thm. 1, gives the optimal contention vector $\boldsymbol{\alpha}^* \in [0, 1]^m$ to maximize the probability of a successful reception over the erasure collision channel.

Consider maximizing $f(\mathbf{p})$ for $q \in (0, 1)$ and $m \in \mathbb{N}$:

$$\max f(\mathbf{p}) \;\; \text{s.t.} \;\; \mathbf{0} \leq \mathbf{p} \leq q\mathbf{1}. \quad (5)$$

The following theorem, the proof of which is given in the Appendix, characterizes $\mathbf{p}^*$, and by extension, the optimal contention probabilities $\boldsymbol{\alpha}^*$ (with $\alpha_i^* = p_i^*/q$ for $i \in [m]$).

**Theorem 1.** *The solution to (5) is (up to permutation):*

1) $\mathbf{p}^* = q\mathbf{1}$*, when* $q \in (0, 1/m]$*;*
2) $\mathbf{p}^* = \sum_{i=1}^{t} q\mathbf{e}_i$*, when* $q \in (1/(t+1), 1/t)$ *for* $t \in [m]$*;*
3) $\mathbf{p}^* = \sum_{i=1}^{t-1} q\mathbf{e}_i + p_s \mathbf{e}_t$*, for any* $p_s \in [0, q]$*, when* $q = 1/t$ *for* $t \in \{2, \ldots, m\}$.

*Here* $t = t^*(q)$ *may be interpreted as the optimal number of contending/active transmitters.*

Several points bear mention:

1) The cases are not fully disjoint (e.g, when $t = m$); in some cases the solution is not unique.
2) The objective function $f(\mathbf{p})$ is the throughput of an equivalent (erasure-free) collision channel under the finite-user slotted Aloha protocol, with contention probabilities $(p_i \equiv \alpha_i q, i \in [m])$. The key difference of (5) is the restriction of the contention probability $p_i$ to $[0, q]$, instead of the usual domain of $[0, 1]$.
3) As will be seen in the proof, this restriction of the domain makes this optimization non-trivial. Further, our current techniques seems to also critically hinge on the assumption of homogeneous erasure probabilities $\mathbf{q} = q\mathbf{1}$.
4) The solution shows the impact of the nonerasure parameter $q$ on the optimal contention probability. For $q$ small ($\leq 1/m$) the risk of erasures outweighs that of collision, and as such it is optimal for *all* transmitters to contend. For $q$ large ($\geq 1/2$) the reverse is true, and it is optimal for a *single* transmitter to contend. The $t^*(q)$ falls from $m$ to 1 in a piecewise constant manner for $q \in [1/m, 1/2]$.
5) The solution demonstrates that, regardless of $q$, each contention probability obeys $\alpha_i^* \in \{0, 1\}$, meaning each transmitter either always or never transmits.

## IV. Conclusion

We presented the optimization problem of minimizing the expected block delay per packet over the erasure collision channel, using both RLC and RSP transmissions. Our main result (Thm. 1) is an explicit characterization of the optimal

contention probability vectors $\boldsymbol{\alpha}$ to maximize the reception probability, and thereby minimize both block delays. Future work will consider: $i$) transmitter-specific erasure probabilities and $ii$) $n$ receivers, and select the contention probabilities to optimize the *anycast* and *broadcast* block delays.

## REFERENCES


[1] N. Xie and S. Weber, "Delay on broadcast erasure channels under random linear combinations," *IEEE Trans. Inf. Theory*, accepted for publication in November 2016, https://arxiv.org/abs/1310.4412.

[2] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, October 2007.

[3] L. Lu, L. You, , and S. C. Liew, "Network-coded multiple access," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2853–2869, December 2014.

[4] G. Cocco, S. Pfletschinger, and M. Navarro, "Seek and decode: Random access with physical–layer network coding and multiuser detection," *Trans Emerging Tel Tech*, 2016.

[5] J. Goseling, M. Gastpar, and J. H. Weber, "Random access with physical–layer network coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3670–3681, July 2015.

[6] G. Cocco, C. Ibars, D. Gündüz, and O. del Rio Herrero, "Collision resolution in multiple access networks with physical-layer network coding and distributed fountain coding," in *Proc. ICASSP*, May 2011.

[7] A. ParandehGheibi, J. K. Sundararajan, and M. Médard, "Collision helps–algebraic collision recovery for wireless erasure networks," in *WiNC*, 2010.

[8] E. Paolini, Č. Stefanović, G. Liva, and P. Popovski, "Coded random access: applying codes on graphs to design random access protocols," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 144–150, June 2015.


## APPENDIX A
## PROOF OF THM. 1.

The following inequalities hold (both tight only at $z = 0$):

$$\frac{z}{1+z} \leq \log(1+z) \leq z, \ \forall z > -1. \tag{6}$$

The following lemma will be used in the proof of Thm. 1.

**Lemma 1.** *For $t \in \mathbb{N}$ and $q \in \left(\frac{1}{t+1}, \frac{1}{t}\right)$ it holds that $g(t;q) \equiv tq(1-q)^{t-\frac{1}{q}} - 1 > 0$.*

*Proof:* **First**, let $t \geq 2$. View $g(t;q)$ as a function of $t$, parameterized by $q$, and defined on $t \in (1/q - 1, 1/q)$, with

$$\frac{\mathrm{d}^2 g}{\mathrm{d}t^2} = q(1-q)^{t-\frac{1}{q}} \log(1-q)(2 + t\log(1-q)). \tag{7}$$

Applying (6):

$$2 + t\log(1-q) \geq 2 + t\frac{-q}{1-q} > 2 - \frac{1}{1-q} > 0, \tag{8}$$

where the second inequality follows from $t < 1/q$, and the last inequality holds if $q < 1/2$ (which is true as $q \in (1/(t+1), 1/t)$ and $t \geq 2$). This implies $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(t;q) < 0$ meaning $g(t;q)$ is strictly concave in $t \in (1/q - 1, 1/q)$. Thus the infimum of $g(t;q)$ occurs at its extreme point(s). Since $g(1/q - 1; q) = 0 = g(1/q; q)$, it follows that $g(t;q) > 0$ for $t \in (1/q - 1, 1/q)$. **Second**, let $t = 1$. We wish to show $g(1;q) \equiv q(1-q)^{1-\frac{1}{q}} - 1 > 0$ for $q \in (1/2, 1)$. The first derivative of $g(1;q)$ is $\frac{1}{q}(1-q)^{1-\frac{1}{q}}(2q + \log(1-q))$. We can verify $2q + \log(1-q)$ is decreasing in $q$ and has a single root on $q \in (1/2, 1)$. Therefore on $(1/2, 1)$, $g(1;q)$ is first increasing and then decreasing, with infimum $g(1;1/2) = 0 = \lim_{q \to 1} g(1;q)$, implying $g(1;q)$ is strictly positive for $q \in (1/2, 1)$. ∎

*Proof of Thm. 1:* We will prove case 1 first. Cases 2 and 3 will then be proved together. Define the following notation

$$x_i = x_i(\mathbf{p}) \equiv p_i \prod_{j \neq i} (1 - p_j), \quad f = f(\mathbf{p}) \equiv \sum_{i=1}^{m} x_i$$
$$\pi = \pi(\mathbf{p}) \equiv \prod_{j} (1 - p_j), \quad \pi_i = \pi_i(\mathbf{p}) \equiv \frac{\pi}{1 - p_i}. \tag{9}$$

Algebra yields the following partial derivatives:

$$\frac{\partial x_j(\mathbf{p})}{\partial p_i} = \begin{cases} \frac{\pi}{1-p_j}, & i = j \\ -\frac{\pi p_j}{(1-p_i)(1-p_j)}, & i \neq j \end{cases}, \quad \frac{\partial f(\mathbf{p})}{\partial p_i} = \frac{\pi_i - f}{1 - p_i}. \tag{10}$$

**Case** 1: $q \in (0, 1/m]$. From (10):

$$\frac{\partial f(\mathbf{p})}{\partial p_i} = \pi_i \left(1 - \sum_{j \neq i} \frac{p_j}{1 - p_j}\right).$$

This shows $\frac{\partial f}{\partial p_i} \geq 0$ iff $\sum_{j \neq i} \frac{p_j}{1-p_j} \leq 1$, which is equivalent to $\sum_{j \neq i} \frac{1}{1-p_j} \leq m$. As each $p_i \in [0, q]$, we have $\frac{1}{1-p_i} \leq \frac{1}{1-q}$, and thus it follows that

$$\sum_{j \neq i} \frac{1}{1 - p_j} \leq \sum_{j \neq i} \frac{1}{1 - q} = \frac{m-1}{1-q} \leq m, \tag{11}$$

where the last inequality follows from $q \in (0, 1/m]$. This means when $q \in (0, 1/m]$, *all* the partial derivatives $\frac{\partial f}{\partial p_i}$ are nonnegative, and thus the maximizer in this case is $\mathbf{p}^* = q\mathbf{1}$.

**Cases** 2 **and** 3: $q \in \bigcup_{t=1}^{m} (1/(t+1), 1/t] \cap (0, 1)$. The Lagrangian is:

$$\mathcal{L}(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{p}) + \sum_{i=1}^{m} \lambda_i (-p_i) + \sum_{i=1}^{m} \nu_i (p_i - q), \tag{12}$$

with Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_i, i \in [m])$ for $-\mathbf{p} \leq \mathbf{0}$, and $\boldsymbol{\nu} = (\nu_i, i \in [m])$ for $\mathbf{p} - q\mathbf{1} \leq \mathbf{0}$. The first-order Karush-Kuhn-Tucker (KKT) necessary conditions for a local *maximizer* are, for each $i \in [m]$: $i$) stationarity, $\frac{\partial \mathcal{L}}{\partial p_i} = 0$; $ii$) primal feasibility, $-p_i \leq 0$ and $p_i - q \leq 0$; $iii$) dual feasibility, $\lambda_i \leq 0$ and $\nu_i \leq 0$; and $iv$) complementary slackness, $\lambda_i(-p_i) = 0$ and $\nu_i(p_i - q) = 0$. Applying (10):

$$\frac{\partial f}{\partial p_i} - \lambda_i + \nu_i = \frac{\pi_i - f}{1 - p_i} - \lambda_i + \nu_i = 0. \tag{13}$$

Applying complementary slackness to the above expression allows us to conclude that if a maximizer $\mathbf{p}$ has any nonzero component(s) strictly less than $q$, then all these nonzero components must equal each other. This follows from observing $\pi_i = \pi_j$ iff $p_i = p_j$ for $\mathbf{0} \leq \mathbf{p} \leq q\mathbf{1}$ with $q \in (0, 1)$. This motivates us to partition the feasible set $[0, q]^m$ according to whether a point $\mathbf{p}$ has any nonzero value strictly less than $q$ among all its $m$ components $(p_i, i \in [m])$. In the following, we will find the best candidate(s) from each of the two categories and then choose the best between them.

First, consider the case when $\mathbf{p}$ does not have any nonzero value strictly less than $q$. These points may also be called "quasi-uniform" (QU) points. Let $m'$ $(\leq m)$ be the total number of indices taking the unique nonzero value $q$. Such

a point $\mathbf{p}$ can be expressed as $\mathbf{p} = \sum_{j=1}^{m'} q\mathbf{e}_j$, and the corresponding objective function $f(\mathbf{p}) = \frac{q}{1-q}h(m';q)$ where the function $h(m';q) \equiv m'(1-q)^{m'}$. We can verify the logarithm of $h$ is concave, and its unique stationary point (maximizer) is $\bar{m'} = 1/\phi(q)$ where $\phi(q) \equiv -\log(1-q)$. We can verify (applying (6))

$$\frac{1}{q} - 1 < \frac{1}{\phi(q)} < \frac{1}{q}, \ \forall q \in (0,1). \tag{14}$$

Note that for $q \in (1/(t+1), 1/t]$, we have

$$t - 1 \leq \frac{1}{q} - 1 < \bar{m'} < \frac{1}{q} < t + 1, \tag{15}$$

from which we can see that the maximizer $m'^*$ (as it has to be an integer) of $f$ comes from $\{t-1, t, t+1\}$. Toward this, we evaluate $h$ at these points and find

$$h(m';q)|_{m'=t-1} \leq h(m';q)|_{m'=t} > h(m';q)|_{m'=t+1}, \tag{16}$$

where the left inequality holds with equality when $q = 1/t$ but is strict when $q \in (1/(t+1), 1/t)$. Therefore the maximizer in this category, when $q \in (1/(t+1), 1/t]$, is $m'^* = t$ (together with $m'^* = t-1$ if $q = 1/t$ for $t \geq 2$), and the corresponding objective function is

$$f|_{m'^*=t} = \frac{q}{1-q}h(m';q)\bigg|_{m'^*=t} = tq(1-q)^{t-1}. \tag{17}$$

Second, consider the case when $\mathbf{p}$ has some nonzero value strictly less than $q$. Recall there can be only one[3] such distinct nonzero value. As such, we parameterize the points in this category as $\mathbf{p}(p_s, k, m')$ where $p_s$ ($s$ for "small", not an index) denotes this common nonzero value, $k$ is the number of indices taking $p_s$, and $m'$ represents the total number of indices taking nonzero values. Naturally, $p_s \in (0, q)$, $k \in \{1, \ldots, m'\}$ (when $k = m'$ it means it does not have any nonzero component value $q$), and $k \leq m' \in \{1, \ldots, m\}$.

Under the $\mathbf{p}(p_s, k, m')$ parameterization, $f(\mathbf{p}(p_s, k, m')) =$

$$\pi(\mathbf{p}(p_s, k, m'))\left(\frac{p_s}{1-p_s}k + \frac{q}{1-q}(m'-k)\right). \tag{18}$$

Combining (13) (applied to $p_s$, with complementary slackness: $\lambda_s = \nu_s = 0$) and (18) gives:

$$m' = \frac{k(q-p_s) + 1 - q}{(1-p_s)q}. \tag{19}$$

As such, the objective $f(\mathbf{p})$ may be equivalently expressed as

$$f(\mathbf{p}(p_s, k, m')) = (1-q)^{m'-1}\left(\frac{1-p_s}{1-q}\right)^{k-1}. \tag{20}$$

Note $(p_s, k, m')$ are mutually constrained as in (19). Here we *enlarge the feasible set* so that the domain of $(p_s, k, m')$ is $[0, q] \times [1, m'] \times [1, \infty)$. In particular, although (19) still has to hold, $m'$ need not be an integer, nor does $k$. Consequently, the best candidate from this enlarged feasible set will yield an upper bound of the maximum that are actually attainable by some point(s) in this category.

Viewing $m'$ as an exogenous parameter, and substituting (19) into (20), the partial derivatives of the bivariate function $f = f(p_s, k)$ w.r.t. $k$ and $p_s$ may be computed as

$$\frac{\partial f}{\partial k} = \frac{(1-q)^{\frac{k(q-p_s)+1-q}{q(1-p_s)}}\left(\frac{1-p_s}{1-q}\right)^k}{q(1-p_s)^2}g_k(p_s;q),$$
$$\frac{\partial f}{\partial p_s} = \frac{(k-1)(1-q)^{\frac{k(q-p_s)+1-q}{q(1-p_s)}}\left(\frac{1-p_s}{1-q}\right)^k}{q(1-p_s)^3}g_{p_s}(p_s;q), \tag{21}$$

where

$$g_k(p_s;q) \equiv (q-p_s)\log(1-q) + q(1-p_s)\log\left(\frac{1-p_s}{1-q}\right),$$
$$g_{p_s}(p_s;q) \equiv (p_s-1)q + (q-1)\log(1-q). \tag{22}$$

We can verify $\frac{\mathrm{d}^2 g_k}{\mathrm{d}p_s^2} = \frac{q}{1-p_s} > 0$, meaning the function $g_k$ is convex in $p_s \in [0, q]$ and hence it attains the maximum value at its boundary (extreme) point(s). We then check $g_k(0;q) = 0 = g_k(q;q)$. This says $g_k \leq 0$ and hence $\frac{\partial f}{\partial k} \leq 0$. Regarding the sign of $\frac{\partial f}{\partial p_s}$, note when $k = 1$, $\frac{\partial f}{\partial p_s} = 0$ for all $p_s \in [0, q]$; when $k > 1$, the sign is the same as that of $g_{p_s}$. Since $g_{p_s}$ is linearly increasing in $p_s$ and we can verify (using (6) to show $g_{p_s}(0;q) < 0 < g_{p_s}(q;q)$), via the intermediate value theorem, that there exists a unique root $\bar{p}_s \in (0, q)$ of $g_{p_s} = 0$, it follows that $\frac{\partial f}{\partial p_s} < 0$ ($> 0$) when $p_s < \bar{p}_s$ ($> \bar{p}_s$). This says for any given $k$, the maximum of $f(p_s, k)$ is attained at either $p_s = 0$ or $q$ (or both). We can compute that $f(0, k) = f(q, k)$. Recall for any fixed $p_s$, $f$ is decreasing in $k$, these together mean the maximum of $f$ is attained when $k$ is further set to 1. In fact we can verify the maximum is given by

$$f(p_s, 1) = (1-q)^{\frac{1}{q}-1}, \ \forall p_s \in [0, q], \tag{23}$$

and occurs (over the enlarged feasible set) when $m' = 1/q$.

To summarize (thus far): when $q \in (1/(t+1), 1/t]$ the best maximizer(s) from the first category (QU points) yields ((17)) the objective function $tq(1-q)^{t-1}$, and the candidates from the other category (points with nonzero component(s) strictly less than $q$) have the objective function *upper bounded* by ((23)) $(1-q)^{\frac{1}{q}-1}$. For the latter, note when $q = 1/t$ for $t \in \{2, \ldots, m\}$, this upper bound is tight and the optimal (from (19)) $m'^* = 1/q = t$ meaning the maximizer in this case is $\sum_{i=1}^{t-1} q\mathbf{e}_i + p_s\mathbf{e}_t$ for all $p_s \in (0, q)$.

Finally, the global maximizer will be found if we can order the maxima from the two categories. Toward this, we'll show

$$tq(1-q)^{t-1} \geq (1-q)^{\frac{1}{q}-1}, \forall q \in (1/(t+1), 1/t], t \in [m], \tag{24}$$

and it is tight when $q = 1/t$ for $t \in \{2, \ldots, m\}$.

When $q = 1/t$, that (24) holds with equality can be verified easily and the maximizers (from the two categories) can be written in a unified manner, namely $\mathbf{p}^* = \sum_{i=1}^{t-1} q\mathbf{e}_i + p_s\mathbf{e}_t$. When $q$ lies in $(1/(t+1), 1/t)$, (24) follows from Lem. 1. This means the maximizer from the QU category is strictly superior, and hence the global maximizer in this case is given by $\mathbf{p}^* = \sum_{i=1}^{t} q\mathbf{e}_i$. ■

[3]This is the number of distinct *values* in the nonzero components of $\mathbf{p}$ (not across different $\mathbf{p}$'s, and not the number of *indices* taking nonzero values).