Efficient size estimation and impossibility of termination in uniform dense population protocols

David Doty[1]
Department of Computer Science, University of California, Davis
doty@ucdavis.edu

Mahsa Eftekhari[2]
Department of Computer Science, University of California, Davis
mhseftekhari@ucdavis.edu

Abstract

We study *uniform* population protocols: networks of anonymous agents whose pairwise interactions are chosen at random, where each agent uses an *identical* transition algorithm that does not depend on the population size n. Many existing polylog(n) time protocols for leader election and majority computation are nonuniform: to operate correctly, they require all agents to be initialized with an approximate estimate of n (specifically, the exact value $\lfloor \log n \rfloor$).

Our first main result is a uniform protocol for calculating $\log(n) \pm O(1)$ with high probability in $O(\log^2 n)$ time and $O(\log^7 n \log\log n)$ states ($O(\log\log n)$ bits of memory). The protocol is converging but not *terminating*: it does not signal when the estimate is close to the true value of $\log n$. If it could be made terminating, this would allow composition with protocols, such as those for leader election or majority, that require a size estimate initially, to make them uniform.

However, our second main result implies that the protocol *cannot* be made terminating. We show that a uniform protocol for *any* task requiring more than constant time cannot be terminating even with probability bounded above 0, if infinitely many initial configurations are *dense*: any state present initially is the state of $\Omega(n)$ agents. (In particular no leader is allowed.) The result holds no matter the memory or time permitted.

Finally, we show that *with* an initial leader, our size-estimation protocol can be made terminating with high probability, with the same asymptotic time and space bounds.

1 Introduction

Population protocols [5] are networks that consist of computational entities called *agents* with no control over the schedule of interactions with other agents. In a population of n agents, repeatedly a random pair of agents is chosen to interact, each observing the state of the other agent before updating its own state.[3] They are an appropriate model for electronic computing scenarios such as sensor networks and for "fast-mixing" physical systems such as animal populations [33], gene regulatory networks [13], and chemical reactions [31], the latter increasingly regarded as an implementable "programming language" for molecular engineering, due to recent experimental breakthroughs in DNA nanotechnology [17, 32].

All problems computable with zero error probability by a constant-state population protocol are computable in $O(n)$ time [6, 21]; the benchmark for "efficient" computation is thus sublinear time, ideally polylog(n). For example, the transition $x, q \to y, y$ (starting with at least as many q as the "input" state x) computes $f(x) = 2x$ in expected time $O(\log n)$, whereas $x, x \to y, q$ computes $f(x) = \lfloor x/2 \rfloor$ exponentially slower: expected time $O(n)$ [16].

[1] Supported by NSF grant CCF-1619343.
[2] Supported by NSF grant CCF-1619343.
[3] Using message-passing terminology, each agent sends its entire state of memory as the message.

Although the original model [5] assumed a set of states and transitions that is constant with respect to n, for important distributed computing problems such as leader election [22], majority computation [1], and computation of other functions and predicates [10] no constant-state protocol can stabilize in sublinear time with probability 1.[4] This motivated the study of protocols in which the set of states and transitions grows with n (essentially adding a non-constant *memory* to each agent). Such protocols achieve leader election and majority computation using $O(\text{polylog}(n))$ time, while keeping the number of states "small": typically $O(\text{polylog}(n))$ [1–3, 11, 12], although $O(\log \log n)$ states suffice for leader election [24].

Unfortunately, most of these sublinear-time protocols [1–3, 11, 12] are *nonuniform*: the set of states and transitions are allowed to depend arbitrarily on n. This capability is used to initialize each agent with an approximate estimate of n (the exact value $\lfloor \log n \rfloor$); without this estimate, the protocols do not work at all. A representative example portion of such a protocol is shown in Fig 1: each agent has an internal "counter", which increments upon each encounter with an x. When the counter reaches $\log n$, the protocol terminates (or moves to a different "stage"). In all of these cases, it is not obvious how to remove this encoding of $\log n$ while preserving correctness.[5]

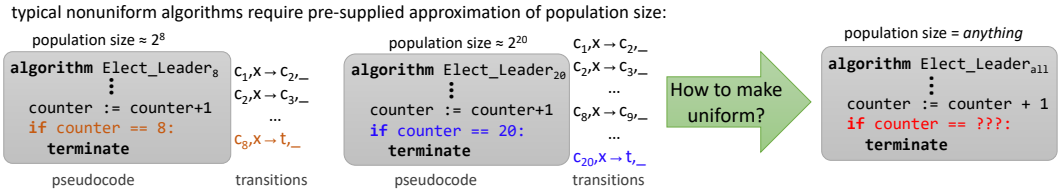

Figure 1 Many population protocols with $\omega(1)$ states use nonuniform algorithms: the value $\log_2 n$ is "hardcoded" into the reactions. Above, "`terminate`" could mean terminate the whole algorithm, or it could mean "move to the next stage of the algorithm".

More desirable would be a *uniform* protocol in which each agent's local algorithm for computing the outputs, given the inputs, has no knowledge of n. Such an algorithm may produce outputs longer than its inputs, retaining the ability to use a number of states that grows with the population size. A uniform protocol can be deployed into *any* population without knowing in advance the size, or even a rough estimate thereof.

Although the original model of population protocols, stipulating a constant set of states, is uniform, when the number of states vary with the population size in this manner, little is known about the computational abilities and limitations of uniform protocols (with a few interesting exceptions [14, 15, 27, 28]).

[4] A protocol *stabilizes* when it becomes unable to change the output. A protocol *converges* in a given random execution when the output stops changing, though it could take longer to subsequently stabilize. Known time lower bounds [1, 10, 22] are on stabilization, not convergence. Recently Kosowski and Uznanski [26] achieved a breakthrough result, showing constant-state protocols for leader election and all decision problems computable by population protocols (the *semilinear* predicates), which converge with high probability in polylog(n) time, and for any $\epsilon > 0$, probability 1 protocols for the same problems that converge in $O(n^\epsilon)$ expected time. The latter protocols require $\Omega(n)$ time to stabilize, as would any constant-state protocol due to the cited time lower bounds.

[5] As we prove in Section 4, uniform protocols with sufficiently "dense" initial configurations *cannot* properly terminate.

1.1 Contribution

Nonuniform protocols in the literature [1–3,11,12] initialize each agent with the value $\lfloor \log n \rfloor$. Hence we study the problem of computing an approximate estimate of $\log n$.

Our first main result, Theorem 3.1, is a uniform protocol, starting from a configuration where all n agents are in an identical state, that with high probability computes $\log n \pm O(1)$ (storing the value in every agent), using $O(\log^2 n)$ time and $O(\log^7 n \log\log n)$ states.[6]

One might hope to use this protocol as a subroutine to "uniformize" existing nonuniform protocols for leader election and majority [1–3, 11, 12].[7] Suppose the size-estimating protocol could be made terminating, eventually producing a termination "signal" that with high probability does not appear until the size estimate has converged. This would allow composition with other protocols requiring the size estimate. It has been known since the very beginning of the population protocol model [5] that termination cannot be guaranteed with probability 1. However, some leader-driven protocols can be made terminating with high probability, including simulation of register machines [6] or exact (but slow) population size counting [27].

Our second main result, Theorem 4.1, shows that this is impossible to do with our leaderless size-estimation protocol and a very wide range of others. The production of such a terminating signal cannot be delayed, even with probability bounded above 0, by more than $O(1)$ time in any uniform protocol where infinitely many valid initial configurations are α-*dense* for some $\alpha > 0$, meaning that all states present are the state of at least αn agents. Since virtually all non-trivial computation with population protocols requires $\Omega(\log n)$ time[8] (including leader election, and computation of predicates and functions such as majority and $f(x) = 2x$), this implies that no uniform terminating protocol can solve these problems from dense initial configurations. The hypothesis of density is crucial: with a *leader*, high-probability termination is possible in a uniform protocol [6]. The hypothesis of uniformity is also crucial; if each agent can *initially* store a value $f(n) = \Omega(n)$ (e.g., $f(n) = \sqrt{n}$ or n), then a termination signal can be delayed until some agent experiences $f(n)$ interactions, an event whose expected time grows unboundedly with n if f grows sufficiently fast.

The first main result answers affirmatively open question 5 of [20], and the second main result answers negatively open questions 1-3 of [20].

Finally, Theorem 3.11 shows that *with* an initial leader, our size estimation can be made terminating with high probability, with the same asymptotic time and space bounds.

1.2 Related work

The work of this paper was inspired by recent work on nonuniform polylog time leader election/majority [1–4, 11, 12, 24]; the fact that those protocols require an approximate size estimate is the direct motivation for seeking a protocol that can compute such an estimate (though unfortunately due to Theorem 4.1, composition of our protocol with these, if possible, will not be straightforward).

Self-stabilizing leader election and exact size counting. Cai, Izumi, and Wada [14] (using different terminology) show an impossibility result for uniform population protocols,

[6] It appears difficult to compute $\lfloor \log n \rfloor$ exactly, rather than within a positive additive constant, since for all k, such a protocol could distinguish between the exact population sizes $2^k - 1$ and 2^k.

[7] Some protocols for leader election [24,25] are uniform, but other protocols [1,3,11,12] have the benefit of simplicity and may possibly be easier to reason about and compose with other protocols.

[8] One way to see that $\Omega(\log n)$ is a lower bound on most interesting computation is that, by a standard coupon collector argument, this is the time required for each agent to have at least one interaction.

that no protocol electing a leader can be uniform if it is also required to be *self-stabilizing*: correct with probability 1 from *any* initial configuration. In fact, it must be nonuniform in a very strong way: the *exact* population size must be encoded into each agent. Self-stabilizing exact size computing has also been shown to be possible with a leader [9] in $O(n \log n)$ time and $O(n)$ states for the leader and 2 states for the other agents, all asymptotically optimal parameters in the self-stabilizing setting [7].

Exact size counting. In the less restrictive setting in which all agents start from a predetermined state, Michail [27] proposed a uniform terminating protocol (where the agents know when they have converged on the correct output) in which a pre-elected leader computes the exact population size n in $O(n \log n)$ time with high probability. Going from the terminating to the less restrictive *converging* criterion (where agents eventually converge on the correct size, but do not know when this occurs), exact size counting is possible in $O(\log n \log\log n)$ time and $O(n^{60})$ states [20], *without* a leader (all agents start in the same state).

Approximate size estimation. Alistarh, Aspnes, Eisenstat, Gelashvili, and Rivest [1] have shown a uniform protocol that in $O(\log n)$ expected time and states converges to an approximation n' of the true size n, computing an integer k such that with high probability $\frac{1}{2}\log n \leq k \leq 9 \log n$, i.e., $\sqrt{n} \leq 2^k \leq n^9$. We improve this from a constant multiplicative error in approximating $\log n$ to a constant *additive* error, or in other words we estimate the population size to within a constant multiplicative factor (instead of a polynomial factor as in [1]), but use $O(\log^2 n)$ time and $O(\log^7 n \log\log n)$ states.

2 Model

To formally define *uniform* computation in population protocols, the agents' transition algorithm is modeled as a multitape Turing machine. (Our protocol describes a constant number of integer fields comprising each agent's state, which could be implemented as one field per tape.) When two agents interact they are both in an initial configuration of the Turing machine, their input is the tape content in the halting configuration of the last time each of them had an interaction. (We assume the Turing machine state and tape head positions are not transmitted, since they can be assumed WLOG identical in every halting configuration.) The space usage (in bits) s is defined as normal for Turing machines: the maximum number of tape cells that are written during the computation. The number of states is then 2^s, where s is the maximum space usage of any agent during an execution of the protocol. For ease of understanding, we will use standard population protocol terminology and not refer explicitly to details of the Turing machine definition except where needed. Therefore a *state* $s \in \Lambda$ always refers to the TM tape content of an agent (leaving out TM state and tape head positions since these are identical in all initial configurations), where Λ is the set of all states, a *configuration* $\vec{c} \in \mathbb{N}^\Lambda$ is a vector indexed by a state, where $\vec{c}(s)$ is the *count* of state s in the population. We set the output of our protocol the value stored in a special field labeled "output". This definition is not indisputable since the output of a protocol can also be defined as a function of the fields stored in an agent's memory. Therefore, agents do not necessarily carry the output in their memory. Though the latter definition does not reduce our protocol's space usage it might benefit other algorithms to save memory. We leave the discussion that which definition is more suited to the readers.

The traditional definition of population protocols assumes a deterministic transition function. Several papers [1,11] indicate how to use the randomness built into the interaction scheduler to provide nearly uniform random bits to the agents, using various *synthetic coin*

techniques, showing that the deterministic model can effectively simulate a model with a randomized transition. For brevity and simplicity of presentation, we will simply assume in the model that each agent has access to a source of uniformly random bits. So instead of a transition function, we have a transition *relation* $\Delta \subseteq \Lambda^4$. We assume that each agent has access to independent uniformly random bits, pre-written on a read-only tape, allowing the Turing machine to be deterministic even though it is computing a nondeterministic relation.

Throughout this paper, n denotes the number of agents in the population. Repeatedly, a pair of agents is selected uniformly at random to interact, where they run the transition algorithm on the pair of states they were in prior to the interaction, and storing the output states for their next interactions. The *time* until some event is measured as the number of interactions until the event occurs, divided by n, also known as parallel time. This represents a natural model of time complexity in which we expect each agent to have $O(1)$ interactions per unit of time, hence across the whole population, $\Theta(n)$ total interactions occur per unit time. All references to "time" in this paper refer to parallel time. An *execution* is a sequence of configurations $\vec{c}_0, \vec{c}_1, \ldots$ such that for all i, applying a transition to $\vec{c}_i$ results in $\vec{c}_{i+1}$. $\log n$ is the base-2 logarithm of n, and $\ln n$ is the base-e logarithm of n.

3 Fast protocol for estimating $\log n$ within $O(1)$ additive error

In this section we describe a uniform protocol for computing the value of $\log n$ with an additive error, i.e., estimating the population size to within a constant multiplicative factor. We say a population protocol is *leaderless* if all agents start in the same state, and we say a protocol *has an initial leader* if one agent starts in a special state ℓ, and all other $n-1$ agents start in a different state $f \neq \ell$.

▶ **Theorem 3.1.** *There is a uniform leaderless population protocol that, with probability at least $1 - 12/n$, computes and stores in each agent an integer k such that $|k - \log n| \leq 4.7$, taking $O(\log^2 n)$ time and $O(\log^7 n \log\log n)$ states.*

Interpreted as an estimate of the population size n, the protocol estimates n within multiplicative factor $2^{4.7} < 26$.

We note that the protocol is not *stabilizing*: it has a positive probability of error. It is an open question to create a protocol using expected polylog(n) time and states that computes $\log(n) \pm O(1)$ with probability 1.

The protocol is described and its time and state complexity analyzed in Subsection 3.2. Most of the analysis of the approximation closeness amounts to proving a bound on the moment-generating function of a maximum of geometric random variables, so that the Chernoff technique can be applied to sums of such variables. This is shown in Section C.

3.1 Intuition

Alistarh et al. [1] describe a protocol for estimating $\log n$ within a constant multiplicative factor. A $\frac{1}{2}$-geometric random variable is the number of flips needed to get one head when flipping a fair coin. In their protocol, each agent generates an independent geometric random variable $\mathbf{G}_i$, then propagates the maximum $\mathbf{M} = \max_{1 \leq i \leq n} \mathbf{G}_i$ by *epidemic*: transitions of the form $i, j \to j, j$ for $i \leq j$, which in $O(\log n)$ time "infect" all agents with the maximum. It is known that $\mathrm{E}[\mathbf{M}] \approx \log n$ [23], and $1/2 \log n < X < 9 \log n$ with probability $\geq 1 - O(1)/n^3$ [1].

We take the obvious extension of this approach: do this K times in parallel and take an average. The estimated average is within $O(1)$ of $\log n$ so long as $K = \Omega(\log n)$ (Co-

rollary C.10). One problem to solve first is how to calculate K; after all, $K = \Theta(\log n)$ scales with n, so with a uniform protocol it cannot be encoded into the agents at the start. The agents estimate it using the protocol of [1]. Since that protocol is converging but not terminating (provably it cannot be made terminating by Theorem 4.1), each time an agent updates its value of K, it reinitializes the remainder of its state.

However, a trickier problem remains: a naïve approach to implementing "averaging of K numbers" requires storing $K = \Theta(\log n)$ numbers in each agent, each having value $\Theta(\log n)$, implying the number of states is $\Theta((\log n)^{\log n}) = \Theta(n^{\log \log n})$. This is even more than the $O(n^{60})$ sufficient to quickly compute *exactly* n [20]. We require a subtler approach.

Each agent will eventually generate K different geometrically distributed random numbers (in addition to the first one that is used to calculate K). However, only *two* of these numbers will ever be stored in the agent at a time. Each agent randomly guesses an index $1 \leq i \leq K$; we say that the agent is *responsible* for index i. It then generates a geometric random variable, propagating the maximum *among all those agents also responsible for* i. Furthermore, an agent responsible for i iterates through each index $j \in \{1, \ldots, K\} \setminus \{i\}$. It waits to encounter an agent responsible for j, generates another geometric random variable $\mathbf{G}$, propagating $\mathbf{G}$ to that agent, before moving on to index $j + 1$. Therefore, the K maxima have the same distribution as in the above memory-intensive protocol: each agent generates exactly K independent geometric random variables, with the maximum of each being propagated among the agents responsible for its index.

After generating K different geometric random variables, to compute the average of the K population-wide maxima, each agent repeatedly cycles through the indices $j \in \{1, \ldots, K\}$, at each index j waiting to encounter an agent responsible for j, adding the latter agent's stored maximum to a running sum. Upon adding the K'th number, the agent divides the sum by K and stores the result in its output field. It resets j to 1 and repeats this cycle forever. Once all K maxima have been propagated to the appropriate agents, each agent will complete a full cycle and calculate the correct average, finally converging on an output.

The time is $O(\log^2 n)$ by the following rough analysis (details follow). Each index $i \in \{1, \ldots, K\}$ has approximately $\frac{n}{K} = \Theta(\frac{n}{\log n})$ agents responsible for it. Therefore, an agent responsible for i waits $O(\log n)$ interactions on average in between each interaction with another agent responsible for i. This slows down the standard $O(\log n)$ time epidemic by factor $O(\log n)$, so each epidemic completes in time $O(\log^2 n)$. By the union bound it is unlikely for any of the K epidemics to take longer to finish. Also, each agent must cycle through the K indices, but moving from index j to $j + 1$ also requires $\Theta(\log n)$ interactions for each agent, so it takes time $\Theta(\log^2 n)$ time for an agent to complete each cycle. Again the union bound shows low probability for any agent to take too long to complete a cycle.

3.2 Formal specification of protocol

In this protocol, agents start by generating one geometric random variable (called $\mathtt{G}_0$) and continue by propagating the maximum among the population. By Lemma C.4 and Corollary C.7 the propagated maximum is in the interval of $[\log n - \log \ln n, 2 \log n + 3/2]$ with probability $\geq 1 - O(1)/n$. Thus, it could be used to estimate K, which is the number of independent additional geometric random variables each agent will generate, with the K maxima propagated across the population.

Each time an agent figures out it was not storing the maximum value of $\mathtt{G}_0$, it restarts its PropagateMaxGRV protocol. Agents generate a new random variable from 1 to `K` as their `ResponsibleIndex` and propagate the maximum for the newly generated index.

The process of restarting and updating the `ResponsibleIndex` will continue until each

Protocol 1 LogSizeEstimation(rec, sen)

▷ initial state of agent:
  $\mathtt{G}_0 \leftarrow \frac{1}{2}$-geometric random variable (used for estimating $\mathtt{K}$)
  $\mathtt{K} \leftarrow$ Smallest power of two $\geq c_{\text{init}} \cdot (1/2 \log \mathtt{G}_0 + \mathtt{G}_0)$ ▷ ($c_{\text{init}} = 4$ for our main analysis)
  `ResponsibleIndex` $\leftarrow$ number i selected uniformly at random from $\{1, \ldots, \mathtt{K}\}$ ▷ agent is responsible for propagating the maximum of the i'th of the $\mathtt{K}$ values of `GeometricRV` across the population
  `GeometricRV` $\leftarrow \frac{1}{2}$-geometric random variable
  `GeneratedIndex` $\leftarrow 1$
  `TemporaryGeometricRV` $\leftarrow 0$
  `AveragingIndex` $\leftarrow 0$
  `sum` $\leftarrow 0$ ▷ Sum of the $\mathtt{K}$ maxima of `GeometricRV`
  `ave` $\leftarrow 0$ ▷ Stores `sum`/$\mathtt{K}$; this is the output field
PropagateMaxInitialEstimate(rec, sen)
PropagateMaxGRV(rec, sen)
if rec.`GeneratedIndex` $> \mathtt{K}$ **then**
  Averaging(rec, sen)

Subprotocol 2 PropagateMaxInitialEstimate(rec, sen)

▷ Maximum generated geometric variable for $\mathtt{G}_0$ will be propagated.
if rec.$\mathtt{G}_0 <$ sen.$\mathtt{G}_0$ **then**
  rec.$\mathtt{G}_0 \leftarrow$ sen.$\mathtt{G}_0$
  Restart(rec)

agent stores the same value in their $\mathtt{G}_0$. In the Restart protocol, all random variables are generated from the beginning. Therefore, we can ignore what have been stored or propagated before the final restart.

Although propagating the random variables is distributed among the population, each agent performs the averaging locally. Agents have a field `sum` for accumulating the maximum variable associated to each index in $\{1, \ldots, \mathtt{K}\}$.

▶ **Lemma 3.2.** *The value $\mathtt{K}$ in* LogSizeEstimation *is in the interval* $[4 \log n, 24 \log n]$ *with probability* $\geq 1 - 1.32/n$.

Proof. Corollary C.7 gives us a very tight bound for the expected value of the maximum of n independent geometric random variables with high probability. We use this corollary to bound $\mathrm{E}[\mathtt{G}_0]$. By Lemma C.4, $\log n + 1 < \mathrm{E}[\mathtt{G}_0] < \log n + 3/2$. Then we can write:

$$\begin{aligned} \Pr[\mathtt{G}_0 \leq \mathrm{E}[\mathtt{G}_0] - 1/2 \log \mathtt{G}_0] &< \Pr[\mathtt{G}_0 \leq \mathrm{E}[\mathtt{G}_0] - (1 + 1/2 \log\log n)] \\ &< \Pr[\mathtt{G}_0 \leq \mathrm{E}[\mathtt{G}_0] - (1 + \log \ln n)] \\ &\leq n^{-1} \text{ by Corollary C.7} \end{aligned}$$

Therefore, setting $\mathtt{K}$ to be the smallest power of two $\geq 4 \cdot (1/2 \log \mathtt{G}_0 + \mathtt{G}_0)$ guarantees $\mathtt{K}$ to be greater then or equal $4 \log n$ with probability $\geq 1 - 1/n$.

Also, by Lemma C.4 and Corollary C.7, $\mathtt{K} \leq 2 \cdot 4 \cdot (2 \log n + 1/2 \log \mathtt{G}_0 + 3/2) \leq 8 \cdot (3 \log n)$ for high values of n with probability $\geq 1 - 0.32/n$. ◀

▶ **Lemma 3.3.** LogSizeEstimation *uses $O(\log^7 n)$ states with probability* $\geq 1 - 2.32/n$.

Subprotocol 3 RESTART(rec)

▷ state: $\mathtt{K} \leftarrow$ Smallest power of two $\geq c_{\text{init}} \cdot (1/2 \log \mathtt{G}_0 + \mathtt{G}_0)$
$\mathtt{ResponsibleIndex} \leftarrow$ number selected uniformly at random from $\{1, \ldots, \mathtt{K}\}$
$\mathtt{GeometricRV} \leftarrow \frac{1}{2}$-geometric random variable
$\mathtt{GeneratedIndex} \leftarrow 1$
$\mathtt{TemporaryGeometricRV} \leftarrow 0$
$\mathtt{sum} \leftarrow 0$
$\mathtt{ave} \leftarrow 0$

Subprotocol 4 PROPAGATEMAXGRV(rec, sen)

if rec.ResponsibleIndex = sen.ResponsibleIndex and rec.GeometricRV < sen.GeometricRV **then**
▷ propagate maximum r.v. between agents responsible for same index
rec.GeometricRV ← sen.GeometricRV
if rec.GeneratedIndex = rec.ResponsibleIndex **then**
▷ geometric r.v. for agent's responsible index was already generated, so skip
rec.GeneratedIndex ← rec.GeneratedIndex + 1
else if rec.GeneratedIndex = sen.ResponsibleIndex **then**
▷ generate geometric r.v. for this index, propagate it, then forget it
rec.GeneratedIndex ← rec.GeneratedIndex + 1
rec.TemporaryGeometricRV ← $\frac{1}{2}$-geometric random variable
if sen.GeometricRV < rec.TemporaryGeometricRV **then**
sen.GeometricRV ← rec.TemporaryGeometricRV

Proof. By setting the number of i.i.d geometric random variables $= nK$ and $\lambda = 2 \ln nK$ in Corollary C.6 we bound all geometric random variables to be $\leq 3 \log nK \leq 15 + 3 \log n + 3 \log \log n \leq 4 \log n$ with probability $\geq 1 - 3.5/(nK) \geq 1 - \frac{1}{n}$ for sufficiently large n. Here, all fields carried by each agent in protocol LOGSIZEESTIMATION are listed:

$\mathtt{G}_0$	$[\log n - \log \ln n - 1/2, 2 \log n + 3/2]$	by Corollary C.7
K	$[4 \log n, 24 \log n]$	by Lemma 3.2
ResponsibleIndex	$[4 \log n, 24 \log n]$	by Lemma 3.2
GeneratedIndex	$[4 \log n, 24 \log n]$	by Lemma 3.2
GeometricRV	$[1, 4 \log n]$	by Lemma 3.3
TemporaryGeometricRV	$[1, 4 \log n]$	by Lemma 3.3
sum	$[1, 96 \log^2 n]$	by Lemmas 3.2, 3.3
AveragingIndex	$[4 \log n, 12 \log n]$	by Lemma 3.2
ave	$[1, 4 \log n]$	by Lemma 3.3

Noting that agents never use both of GeneratedIndex and AveragingIndex at the same time in LOGSIZEESTIMATION protocol; hence the same space can be shared between the variables. Also, once $\mathtt{G}_0$ is generated, it is only needed to compute K, so the space for K subsumes that for $\mathtt{G}_0$, which is not needed after K is generated. Since, K is a power of two storing only the power suffices. Thus we have one variable that is $O(\log^2 n)$, one with $O(\log \log n)$ space, and after the described optimizations, 5 additional variables that are $O(\log n)$. Therefore, each agent needs $O(\log^7 n \log \log n)$ states with probability $\geq 1 - (1.32 + 1)/n$. ◀

Subprotocol 5 AVERAGING(rec, sen)

```
if rec.AveragingIndex = rec.ResponsibleIndex then
    rec.AveragingIndex ← rec.AveragingIndex + 1
    rec.sum ← rec.sum + rec.GeometricRV
else if rec.AveragingIndex = sen.ResponsibleIndex then
    rec.AveragingIndex ← rec.AveragingIndex + 1
    rec.sum ← rec.sum + sen.GeometricRV
if rec.AveragingIndex > K then
    rec.ave ← ⌊rec.sum/K⌋
    rec.AveragingIndex ← 1
    rec.sum ← 0
```

In this protocol, we have each agent choosing the `ResponsibleIndex` uniformly random from 1 to `K`, which has the same distribution as throwing n balls into `K` bins. For a fast convergence of the protocol, there should be enough agents (balls) assigned to each index (bins). We want to show that the time for each epidemic is $O(\log^2 n)$ with high probability.

The next lemma, proven in Section A, bounds the number of agents selecting each index.

▶ **Lemma 3.4.** *With probability at least* $1 - \left(\frac{2}{e}\right)^{\sqrt{n}}$*, for each index* $i \in \{1, \ldots, K\}$*, at least* $\frac{n}{2K}$ *and at most* $\frac{2n}{K}$ *agents are responsible for* i*.*

Each agent repeatedly cycles a counter from 1 to K, incrementing when encountering an agent responsible for the index equal to its counter. (This is used both to generate the geometric random variables for each index, and later to repeatedly compute the sum of their maxima.) The following lemma, proven in Section A, bounds the time required to complete one of these cycles.

▶ **Lemma 3.5.** *With probability* $\geq 1 - 1/n$*, it takes* $O(\log^2 n)$ *time for all agents, starting with a counter at index* 1*, to have all their counters reach* K *and reset back to* 1*.*

▶ **Corollary 3.6.** AVERAGING *protocol takes* $\leq 2304 \log^2 n$ *time with probability* $\geq 1 - 1/n$*.*

▶ **Corollary 3.7.** *Generating all* $K \cdot n$ *random variables takes* $\leq 2304 \log^2 n$ *time with probability at least* $1 - 1/n$*.*

To analyze the time complexity of our protocol, we require the time bounds for completing an epidemic from the paper [6]. The current form is taken from [20].

▶ **Lemma 3.8** ([6])**.** *Let* T *denote the time to complete an epidemic. Then* $\mathrm{E}[T] = \frac{n-1}{n} H_{n-1}$*,* $\Pr\left[T < \frac{1}{4} \ln n\right] < 2e^{-\sqrt{n}}$*, and for any* $\alpha_u > 0$*,* $\Pr\left[T > \alpha_u \ln n\right] < 4n^{-\alpha_u/4+1}$*.*

▶ **Lemma 3.9.** *With probability at least* $1 - exp(-n/24 \log n)$*, it takes* $\leq 3 \cdot 2048 \log^2 n$ *time for all agents to propagate all* K *independent maximum geometric random variables.*

Proof. Fix the i^{th} maximum geometric random variable, we bound the time for completing an epidemic to propagate it within the subpopulation that are responsible for the i^{th} maximum geometric random variable. Lemma 3.4 claims w.h.p at least $\frac{n}{2K}$ and at most $\frac{2n}{K}$ agents are propagating each individual epidemic. So, the probability that the next interaction is among the responsible subpopulation for the i^{th} maxima is $\geq \frac{\binom{\frac{n}{2K}}{2}}{\binom{n}{2}} \geq \frac{1}{8K^2}$. Also, by setting $\alpha_u = 8$ in Lemma 3.8 the probability that an epidemic takes longer than $8 \ln n_0$ is less than $4/n_0$ for any population size n_0. Thus, we can model the completion of an epidemic in the

responsible subpopulation by counting the number of trials of a biased coin until $8 \ln n_0$ heads appears where the probability of seeing a head equals $p = 1/8\mathtt{K}^2$.

With $m = 128\mathtt{K}^2 n_0 \ln n_0$ trials, probability of success $p = 1/8\mathtt{K}^2$, and number of successes $8 \ln n_0$, we have that $\Pr[\mathbf{B}(m,p) \leq 8n_0 \ln n_0] = \Pr[\mathbf{NB}(8n_0 \ln n_0, p) \geq m]$. We use a Chernoff bound for binomial random variables, which states that, for $0 < \delta \leq 1$, setting $\mu = \mathrm{E}[\mathbf{B}(m,p) = mp]$, $\Pr[\mathbf{B}(m,p) \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$. Setting $\delta = \frac{1}{2}$ and $\mu = (128\mathtt{K}^2 n_0 \ln n_0)p = 16 n_0 \ln n_0$,

$$\begin{aligned}
\Pr\left[\mathbf{NB}(8n_0 \ln n_0, p) \geq 128\mathtt{K}^2 n_0 \ln n_0\right] &= \Pr\left[\mathbf{B}(128\mathtt{K}^2 n_0 \ln n_0, p) \leq 8n_0 \ln n_0\right] \\
&= \Pr\left[\mathbf{B}(128\mathtt{K}^2 n_0 \ln n_0, p) \leq (1-\delta)\mu\right] \\
&\leq e^{-\mu\delta^2/2} \\
&= e^{-2n_0 \ln n_0} \\
&< e^{-n/\mathtt{K}} \qquad \text{by Lemma 3.4} \\
&< e^{-n/24\log n} \qquad \text{by Lemma 3.2}
\end{aligned}$$

. The above inequality shows that an epidemic takes more than $3 \cdot 2048 n \log^2 n \geq 128\mathtt{K}^2 n_0 \ln n_0$ interactions or equivalently $3 \cdot 2048 \log^2 n$ time with probability at most $\exp(-n/24 \log n)$. ◀

▶ **Lemma 3.10.** LogSizeEstimation *protocol takes $O(\log^2 n)$ time with probability at least $1 - 7/n$ for all agents to store the average of K independent maximas.*

Proof. We assume in the worst case that individual sub-protocols run consecutively rather than in parallel, and add their required time to upper bound their parallel time. By Lemmas 3.8, 3.9 and Corollaries 3.6, 3.7 we have: $8 \ln n + 24^3 \log^2 n + 2304 \log^2 n + 2304 \log^2 n = O(\log^2 n)$.

By the union bound, the probability that LogSizeEstimation takes $> O(\log^2 n)$ is $\leq \frac{4}{n} + \frac{1}{n} + \frac{1}{n} + \frac{1}{n} = 7/n$ for sufficiently large n. ◀

Finally, we combine these results to prove the main result of this section.

Proof of Theorem 3.1. By Lemma 3.3, LogSizeEstimation protocol uses $O(\log^7 n \log\log n)$ states with probability at least $1 - 2.32/n$. Lemma 3.10 proves that this protocol takes $O(\log^2 n)$ time with probability at least $1 - 7/n$. Finally, Corollary C.10 guaranties the average of k obtained maxima are with an additive error $O(1)$ of $\log n$ with probability at least $1 - 2/n$. By the union bound, after $O(\log^2 n)$ time agents will store the estimation of $\log n$ with an additive error $O(1)$ with probability at least $1 - 12/n$. ◀

3.3 Terminating size estimation with a leader

It is possible to make the size-estimation protocol terminating if we start with an initial leader.

▶ **Theorem 3.11.** *There is a uniform terminating population protocol with an initial leader that, with probability $\geq 1 - O(1)/n$, computes and stores in each agent an integer k such that $|k - \log n| \leq 4.7$, taking $O(\log^2 n)$ time and $O(\log^7 n \log\log n)$ states.*

Proof. By Theorem 4.1, a leader (or a $o(n)$-size junta of leaders) is required for termination to work with positive probability. In the presence of a leader, the population can simulate a phase clock as described in Angluin et al. [6]. By [6, Corollary 1], there is a constant

$k_1 = \max(8c, \frac{8d}{c})$ for the number of phases that it takes at least $d \ln n$ to reach phase k_1 with probability at least $1 - \frac{1}{n^c}$.

We change the algorithm to have two main stages. In stage one, the maximum $\mathtt{G}_0$ gets propagated. If we set the number of phases in a phase clock greater than 64, then reaching the maximum phase takes at least $8 \ln n$ time with probability at least $1 - \frac{1}{n}$. By Lemma 3.8, $8 \ln n$ time is sufficient for propagating the maximum of $\mathtt{G}_0$ among the population with high probability. By Lemma C.4 and Corollary C.7 w.h.p. the maximum of $\mathtt{G}_0$ among n agents is in interval of $[\log n - \log \ln n, 2 \log n + 3/2]$. So, after the phase clock reaches k_1 the leader can terminate stage one and move to stage two; where agents compute the sum of $\mathtt{K}$ maximas of n geometric random variable. By setting the number of phases equal to $k_2 \cdot 2\mathtt{G}_0$, we can set a timer to count up until $\frac{k_2}{8} \ln n \log n$ time with probability at least $1 - \frac{1}{n}$ for some "big" k_2 [6, Corollary 1]. When the phase clock reaches $k_2 \cdot 2\mathtt{G}_0$, leader terminates stage and report the $\mathtt{ave}$ value it computed. ◀

4 Termination

Though the concept of termination has been referenced and studied in population protocols [8, 27, 28], to our knowledge no formal definition exists. We give an abstract definition that captures the behavior of most existing protocols that "perform a computational task".

Let P be a population protocol with a set I of "valid" initial configurations for P, where each agent's memory has a Boolean field $\mathtt{terminated}$ that is FALSE in every configuration in I.[9] We say that a configuration $\vec{c}$ of P is *terminated* if at least one agent in $\vec{c}$ has $\mathtt{terminated} = \text{TRUE}$. Let $\kappa > 0$ and $t : \mathbb{N} \to \mathbb{N}$. We say that P is *κ-t-terminating* if, for all $\vec{i} \in I$, with probability $\geq \kappa$, P reaches from $\vec{i}$ to a terminated configuration $\vec{c}$, but takes time $\geq t(n)$ to do so.

This definition leaves totally abstract which particular task (e.g., leader election) is assumed to have terminated. The idea is that if the task will not be complete before time $t(n)$ with high probability, then no agent should set $\mathtt{terminated} = \text{TRUE}$ until time $\geq t(n)$ with high probability. So proving an upper bound on $t(n)$ in the definition of terminating implies that no protocol can be terminating if it requires time $> t(n)$ to converge.

The definition is applicable beyond the narrow goal of terminating a population protocol. It says more generally that a "signal" is produced after some amount of time. This signal may be used to terminate a protocol, move it from one "stage" to another, or it may be some specific Boolean value relevant to a specific protocol, where in any case the value will start FALSE for all agents and eventually be set to TRUE for at least one agent.

Let $\alpha > 0$. We say a configuration $\vec{c}$ is *α-dense* if, for all $s \in \Lambda$, $\vec{c}(s) > 0 \implies \vec{c}(s) \geq \alpha n$. (Recall $n = \|\vec{c}\|$.) In other words, every state present occupies at least fraction α of the population. We say protocol P with valid initial configuration set I is *i.o.-dense* if there exists $\alpha > 0$ such that infinitely many $\vec{i} \in I$ are α-dense. In particular, an i.o.-dense protocol does not always have an initial *leader*: a state present in count 1 in every $\vec{i} \in I$.

The next theorem, our second main result, shows that termination is impossible for uniform i.o.-dense protocols that require more than constant time.

▶ **Theorem 4.1.** *Let $\kappa > 0$ and $t : \mathbb{N} \to \mathbb{N}$, and let P be a uniform i.o.-dense population protocol. If P is κ-t-terminating, then $t(n) = O(1)$.*

[9] In the language of states, we partition the state set Λ into disjoint subsets Λ_T and Λ_N such that $\Lambda_T \cup \Lambda_N = \Lambda$ and Λ_T are precisely the states with $\mathtt{terminated} = \text{TRUE}$.

Let Λ be the (possibly infinite) set of all states of a population protocol. Since the protocol may be randomized, we consider a transition *relation* $\Delta \subseteq \Lambda^4$, writing $a, b \to c, d$ to denote that $(a, b, c, d) \in \Delta$ (i.e., if agents in states a and b interact, then one of the possible random outcomes is to change to states c and d). For $\rho \in (0, 1]$, we write $a, b \xrightarrow{\rho} c, d$ to denote that when states a and b interact, with probability ρ they transition to c and d. Say that ρ is the *rate constant* of transition $a, b \xrightarrow{\rho} c, d$. If there exist $a, b \in \Lambda$ and $\rho' \geq \rho$ such that $a, b \xrightarrow{\rho'} c, d$, we write $c \in \text{PROD}_\rho(a, b)$ and $d \in \text{PROD}_\rho(a, b)$. (In other words, $c \in \text{PROD}_\rho(a, b)$ if c is produced with probability at least ρ whenever a and b interact). For any $\Gamma \subseteq \Lambda$ and $\rho \in [0, 1]$, define $\text{PROD}_\rho(\Gamma) = \{s \in \Lambda \mid (\exists a, b \in \Gamma)\ s \in \text{PROD}_\rho(a, b)\}$.[10]

Let $\Lambda^0 \subseteq \Lambda$. For all $i \in \mathbb{N}^+$, define $\Lambda^i_\rho = \Lambda^{i-1}_\rho \cup \text{PROD}_\rho(\Lambda^{i-1}_\rho)$. For any $m \in \mathbb{N}$, if $s \in \Lambda^m_\rho$, we say s is *m-ρ-producible from Λ^0*.[11] For a configuration $\vec{c}$, we say s is *m-ρ-producible from $\vec{c}$* if s is m-ρ-producible from $\Lambda^0 = \{s \in \Lambda \mid \vec{c}(s) > 0\}$, the set of states present in $\vec{c}$.[12]

Our main technical tool is the following lemma, a variant of the "timer/density lemma" of [19] (see also [1]). The original lemma states that in a protocol with $O(1)$ states, from any sufficiently large α-dense configuration, in $O(1)$ time all states appear with δ-density (for some $0 < \delta < \alpha$). The proof is similar to that of [19], but is re-tooled to apply to protocols with a non-constant set of states (also to use the discrete-time model of population protocols, instead of the continuous-time model of chemical reaction networks).[13] The key new idea is that, even if a protocol has infinitely many states (of which only finitely many can be produced in finite time), for any subset of states Λ^m_ρ "producible via only m transitions, each having rate constant at least ρ", all states in Λ^m_ρ are produced in constant time with high probability from sufficiently large configurations.

▶ **Lemma 4.2.** *Let $\alpha > 0$, $m \in \mathbb{N}^+$, $\rho \in (0, 1]$, and P be a population protocol. Then there are constants $\epsilon, \delta, n_0 > 0$ such that, for all $n \geq n_0$, for all α-dense configurations $\vec{c}$ of P with $n = ||\vec{c}||$, the following holds. Let Λ^m_ρ be the set of states m-ρ-producible from $\vec{c}$. For $s \in \Lambda$ and $t > 0$, let $\mathbf{C}_{t,s}$ be the random variable denoting the count of s at time t, assuming at time 0 the configuration is $\vec{c}$. Then* $\Pr\left[(\forall s \in \Lambda^m_\rho)\ \mathbf{C}_{1,s} \geq \delta n\right] \geq 1 - 2^{-\epsilon n}$.

A self-contained proof is in Section D. We now use Lemma 4.2 to prove Theorem 4.1.

Proof of Theorem 4.1. Assume P is κ-t-terminating; we will show $t(n) = O(1)$. Let $(\vec{c}_i)_{i=1}^\infty$ be an infinite sequence of α-dense initial configurations in I. Dickson's Lemma [18] states that every infinite sequence in $\mathbb{N}^k$ has an infinite nondecreasing subsequence, so assume without loss of generality that $\vec{c}_i \leq \vec{c}_{i+1}$ for all $i \in \mathbb{N}$. Let $\Lambda^0 = \{s \in \Lambda \mid \vec{c}_0(s) > 0\}$ be the set of states present in $\vec{c}_0$.

---

[10] In other words, $\text{PROD}_\rho(\Gamma)$ is the set of states producible by a single transition, assuming that only states in Γ are present, and that the only transitions used are those that have probability at least ρ of occurring when their input states interact.

[11] If s is m-ρ-producible from Λ^0, then in other words, s is producible from any sufficiently large configuration that contains only states in Λ^0, using at most m different *types* of transitions, each of which has probability at least ρ. More than one instance of each transition, however, may be necessary. For instance, with transitions $x_i, x_i \xrightarrow{\rho} x_{i+1}, q$ for all $i \in \mathbb{N}^+$, x_m is m-ρ-producible from $\Lambda_0 = \{x_1\}$, but 2^m transitions of type $x_1, x_1 \to x_2, q$ must be executed, followed by 2^{m-1} of type $x_2, x_2 \to x_3, q$, etc.

[12] Note that s may be m-ρ-producible from $\vec{c}$, but not actually producible from $\vec{c}$, if the counts in $\vec{c}$ are too small for the requisite transitions to produce s.

[13] Alistarh et al. [1] also prove a variant applying to protocols with $\omega(1)$ states, but for a different purpose: to show that all states in Λ appear as long as $|\Lambda| \leq \frac{1}{2} \log \log n$. However, beyond that bound, the lemma does not hold [24]. In our case, we are not trying to show that all states in Λ appear, only those in some constant size subset of states, all of which are m-ρ-producible from the initial configuration.

By hypothesis $\Pr[P \text{ terminates from } \vec{c}_0] \geq \kappa > 0$. Thus there is at least one finite execution $\mathcal{E}$ starting with $\vec{c}_0$ and ending in a terminated configuration. Let $m = |\mathcal{E}|$ be the length of this execution. Let ρ be the minimum rate constant of any transition in $\mathcal{E}$. Then every state appearing in configurations in $\mathcal{E}$ is m-ρ-producible from $\vec{c}_0$, i.e., is in Λ_ρ^m where $\Lambda^0 = \{s \in \Lambda \mid \vec{c}_0(s) > 0\}$ is the set of states present in $\vec{c}_0$.

For any $\ell \geq 1$, since $\vec{c}_0 \leq \vec{c}_\ell$, all states in Λ_ρ^m are m-ρ-producible from $\vec{c}_\ell$ as well. By Lemma 4.2, there are constants $\epsilon, \delta, n_0 > 0$ such that, for all $\ell \in \mathbb{N}$ such that $n = \|\vec{c}_\ell\| \geq n_0$, letting $\mathbf{C}_{1,s}$ be the random variable denoting the count of s at time 1, assuming at time 0 the configuration is $\vec{c}_\ell$, $\Pr\left[\left(\forall s \in \Lambda_\rho^m\right)\ \mathbf{C}_{1,s} \geq \delta n\right] \geq 1 - 2^{-\epsilon n}$.

However, Λ_ρ^m contains terminated states, so for all $\vec{c}_\ell$ with $\|\vec{c}_\ell\| \geq n_0$, with probability $\geq 1 - 2^{-\epsilon n}$, P terminates within time 1. Since $2^{-\epsilon n} < \kappa$ for sufficiently large n, this implies that if P is κ-t-terminating, then $t(n) \leq 1$ for sufficiently large n. Thus $t(n) = O(1)$. ◀

Observe how the assumption of uniformity is used in the proof: we take a set of transitions used on the population $\vec{c}_0$ and apply it to a larger population $\vec{c}_\ell$. In a nonuniform protocol, the transitions may not be legal to apply to $\vec{c}_\ell$. As a concrete example, in a nonuniform protocol such as that sketched in Fig. 1, an agent increments a counter using transitions such as $c_7, x \to c_8, y$ until the counter exceeds $\log n$, then produces a termination signal t via a transition $c_8, x \to t, y$. The transition $c_8, x \to t, y$ producing this signal is not legal in a population larger than twice n, since the value $\log n$ is at least 1 larger in such a protocol. In this example, the transition of the larger protocol with the same input states simply increments the counter without producing a termination signal: $c_8, x \to c_9, y$.

Acknowledgements.

The authors are grateful to Eric Severson for helpful comments on earlier drafts.

References

1 Dan Alistarh, James Aspnes, David Eisenstat, Rati Gelashvili, and Ronald L Rivest. Time-space trade-offs in population protocols. In *SODA 2017: Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2560–2579. SIAM, 2017.

2 Dan Alistarh, James Aspnes, and Rati Gelashvili. Space-optimal majority in population protocols. In *SODA 2018: Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2221–2239. SIAM, 2018.

3 Dan Alistarh and Rati Gelashvili. Polylogarithmic-time leader election in population protocols. In *42nd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 9135 of *Lecture Notes in Computer Science*, pages 479 – 491. Springer, Berlin, Heidelberg, 2015.

4 Dan Alistarh, Rati Gelashvili, and Milan Vojnović. Fast and exact majority in population protocols. In *PODC 2015: Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing*, pages 47–56. ACM, 2015.

5 Dana Angluin, James Aspnes, Zoë Diamadi, Michael J. Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18(4):235–253, March 2006.

6 Dana Angluin, James Aspnes, and David Eisenstat. Fast computation by population protocols with a leader. *Distributed Computing*, 21(3):183–199, September 2008.

7 James Aspnes, Joffroy Beauquier, Janna Burman, and Devan Sohier. Time and Space Optimal Counting in Population Protocols. In *20th International Conference on Principles of Distributed Systems (OPODIS 2016)*, volume 70, pages 13:1–13:17, 2017.

8 James Aspnes and Eric Ruppert. An introduction to population protocols. *Bulletin of the European Association for Theoretical Computer Science*, 93:98–117, October 2007.

9 Joffroy Beauquier, Janna Burman, Simon Claviere, and Devan Sohier. Space-optimal counting in population protocols. In *DISC 2015: International Symposium on Distributed Computing*, pages 631–646. Springer, 2015.

10 Amanda Belleville, David Doty, and David Soloveichik. Hardness of computing and approximating predicates and functions with leaderless population protocols. In *ICALP 2017: 44th International Colloquium on Automata, Languages, and Programming*, volume 80 of *LIPIcs*, pages 141:1–141:14, 2017.

11 Petra Berenbrink, Dominik Kaaser, Peter Kling, and Lena Otterbach. Simple and Efficient Leader Election. In *1st Symposium on Simplicity in Algorithms (SOSA 2018)*, volume 61, pages 9:1–9:11, 2018.

12 Andreas Bilke, Colin Cooper, Robert Elsässer, and Tomasz Radzik. Brief announcement: Population protocols for leader election and exact majority with $O(\log^2 n)$ states and $O(\log^2 n)$ convergence time. In *PODC 2017: Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 451–453. ACM, 2017.

13 James M Bower and Hamid Bolouri. *Computational modeling of genetic and biochemical networks*. MIT press, 2004.

14 Shukai Cai, Taisuke Izumi, and Koichi Wada. Space complexity of self-stabilizing leader election in passively-mobile anonymous agents. In Shay Kutten and Janez Žerovnik, editors, *Structural Information and Communication Complexity*, pages 113–125, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

15 Ioannis Chatzigiannakis, Othon Michail, Stavros Nikolaou, Andreas Pavlogiannis, and Paul G. Spirakis. Passively mobile communicating machines that use restricted space. *Theoretical Computer Science*, 412(46):6469–6483, October 2011.

16 Ho-Lin Chen, David Doty, and David Soloveichik. Deterministic function computation with chemical reaction networks. *Natural Computing*, 13(4):517–534, 2013. Special issue of invited papers from DNA 2012.

17 Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, 8(10):755–762, 2013.

18 Leonard E. Dickson. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *American Journal of Mathematics*, 35:413–422, 1913.

19 David Doty. Timing in chemical reaction networks. In *SODA 2014: Proc. of the 25th Annual ACM-SIAM Symp. on Discrete Algorithms*, pages 772–784, 2014.

20 David Doty, Mahsa Eftekhari, Othon Michail, Paul G. Spirakis, and Michail Theofilatos. Brief announcement: Exact size counting in uniform population protocols in nearly logarithmic time. In *DISC 2018: 32nd International Symposium on Distributed Computing*, 2018.

21 David Doty and Monir Hajiaghayi. Leaderless deterministic chemical reaction networks. *Natural Computing*, 14(2):213–223, 2015. Preliminary version appeared in DNA 2013.

22 David Doty and David Soloveichik. Stable leader election in population protocols requires linear time. *Distributed Computing*, 31(4):257–271, 2018. Special issue of invited papers from DISC 2015.

23 Bennett Eisenberg. On the expectation of the maximum of iid geometric random variables. *Statistics & Probability Letters*, 78(2):135 – 143, 2008.

24 Leszek Gasieniec and Grzegorz Stachowiak. Fast space optimal leader election in population protocols. In *SODA 2018: ACM-SIAM Symposium on Discrete Algorithms*, 2018. to appear.

25 Leszek GKasieniec, Grzegorz Stachowiak, and Przemysław Uznański. Almost logarithmic-time space optimal leader election in population protocols. Technical Report 1802.06867, arXiv, 2018.

26 Adrian Kosowski and Przemyslaw Uznanski. Population protocols are fast. *CoRR*, abs/1802.06872, 2018.

27 Othon Michail. Terminating distributed construction of shapes and patterns in a fair solution of automata. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing*, pages 37–46, 2015. Also in *Distributed Computing*, 2017.

28 Othon Michail, Ioannis Chatzigiannakis, and Paul G Spirakis. Terminating population protocols via some minimal global knowledge assumptions. In *Stabilization, Safety, and Security of Distributed Systems (SSS)*, pages 77–89. Springer, 2012.

29 Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge university press, 2005.

30 Philippe Rigollet. Lecture notes for MIT course 18.s997: High dimensional statistics, 2015. URL: `https://ocw.mit.edu/courses/mathematics/18-s997-high-dimensional-statistics-spring-2015/lecture-notes/`.

31 David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008.

32 Niranjan Srinivas, James Parkin, Georg Seelig, Erik Winfree, and David Soloveichik. Enzyme-free nucleic acid dynamical systems. *Science*, 358(6369):eaal2052, 2017.

33 Vito Volterra. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma*, 2:31–113, 1926.

A Proofs for correctness of size estimation protocol

This section contains proofs of lemmas required to analyze the correctness and time/space complexity of the size estimation protocol of Theorem 3.1.

We use a Chernoff bound for Poisson random variables [29].

▶ **Theorem A.1** ([29]). *Letting* $\mathbf{P}(\lambda)$ *be a Poisson random variable with rate* λ*, for all* $l \in \mathbb{N}$*:*

For $\lambda > l$, $\Pr[\mathbf{P}(\lambda) \leq l] \leq e^{-\lambda}\left(\frac{e\lambda}{l}\right)^n$.

For $\lambda < l$, $\Pr[\mathbf{P}(\lambda) \geq l] \leq e^{-\lambda}\left(\frac{e\lambda}{l}\right)^n$.

We now prove some lemmas used in the main text.

Lemma 3.4. *With probability at least* $1 - \left(\frac{2}{e}\right)^{\sqrt{n}}$*, for each index* $i \in \{1, \ldots, K\}$*, at least* $\frac{n}{2K}$ *and at most* $\frac{2n}{K}$ *agents are responsible for* i*.*

Proof. Agents choose a random index from the set of $\{1, \ldots, K\}$ and stay responsible for the chosen index. We can model this event by the problem of throwing n balls in to K bins. It is well known [29] that if we approximate the distribution of tossing m balls into n bins by a distribution where each bin gets a number of balls given by a Poisson distribution with rate m/n (the so-called "Poisson heuristic"), then any event whose probability is monotonicially increasing or decreasing with m, and which occurs in the Poisson case with probability at most p, occurs in the exact case with probability at most $2p$. Setting $l = \lambda/2$ in the theorem A.1 results in

$$\begin{aligned}
\Pr[\mathbf{P}(\lambda) \leq \lambda/2] &\leq e^{-\lambda}\left(\frac{e\lambda}{\lambda/2}\right)^{\lambda/2} \\
&= e^{-\lambda}(2e)^{\lambda/2} \\
&= \left(\frac{2}{e}\right)^{\lambda/2} \\
&\leq \left(\frac{2}{e}\right)^{\frac{n}{24\log n}} \\
&< \left(\frac{2}{e}\right)^{\sqrt{n}}.
\end{aligned}$$

The last inequality is correct if we set $\lambda = n/\mathtt{K}$ and from Lemma 3.2 we know that $4\log n \leq \mathtt{K} \leq 24\log n$. Also by setting $l = 2\lambda$ in the theorem A.1 results in

$$\begin{aligned}
\Pr[\mathbf{P}(\lambda) \geq 2\lambda] &\leq e^{-\lambda}\left(\frac{e\lambda}{2\lambda}\right)^{2\lambda} \\
&= e^{-\lambda}\left(\frac{e}{2}\right)^{2\lambda} \\
&= \left(\frac{e}{4}\right)^{\lambda} \\
&\leq \left(\frac{e}{4}\right)^{\frac{n}{24\log n}} \\
&\leq \left(\frac{2}{e}\right)^{\frac{n}{24\log n}} \\
&< \left(\frac{2}{e}\right)^{\sqrt{n}}.
\end{aligned}$$

The lemma follows by using the Poisson approximation bound. ◀

Lemma 3.5. *With probability $\geq 1 - 1/n$, it takes $O(\log^2 n)$ time for all agents, starting with a counter at index 1, to have all their counters reach K and reset back to 1.*

Proof. Fix a single agent. We set $\mathbf{X}_i$ to be random variable representing the number of interactions that takes for this agent to increase counter from $i-1$ to i. Once all agents' `GeneratedIndex` reach $\mathtt{K}$, by Lemma 3.4 at least $\frac{n}{2\mathtt{K}}$ agents are responsible for each index with high probability. Therefore, selecting the current agent as the sender and an agent who is responsible for index i as the receiver gets stochastically dominated by a geometric random variable with success probability of $p = \frac{1}{n} \cdot \frac{1}{2\mathtt{K}} = \frac{1}{2n\mathtt{K}}$. Thus, $\mathrm{E}[\mathbf{X_i}] = 2n\mathtt{K}$ where $\mathtt{K}$, the number of total indices, is in interval $[4\log n, 12\log n]$ with high probability (Lemma 3.2). We desire to bound the sum of $\mathtt{K}$ geometric random variables, each with success probability p, otherwise known as a negative binomial random variable $\mathbf{NB}(\mathtt{K}, p)$.

With m trials, probability of success p, and number of successes k, we have that $\Pr[\mathbf{B}(m,p) \leq \mathtt{K}] = \Pr[\mathbf{NB}(\mathtt{K},p) \geq m]$. We use a Chernoff bound for binomial random variables, which states that, for $0 < \delta \leq 1$, setting $\mu = \mathrm{E}[\mathbf{B}(m,p) = mp]$, $\Pr[\mathbf{B}(m,p) \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$. Setting $\delta = \frac{1}{2}$ and $\mu = 8n\mathtt{K}^2 p = 4\mathtt{K}$,

$$\begin{aligned}
\Pr\left[\mathbf{NB}(\mathtt{K},p) \geq 8n\mathtt{K}^2\right] &= \Pr\left[\mathbf{B}(8n\mathtt{K}^2,p) \leq \mathtt{K}\right] \\
&= \Pr\left[\mathbf{B}(8n\mathtt{K}^2,p) \leq (1-\delta)\mu\right] \\
&\leq e^{-\mu\delta^2/2} \\
&= e^{-4\mathtt{K}(1/4)/2} \\
&= e^{-\mathtt{K}/2} \\
&< e^{-2\log n} \quad \text{by Lemma 3.2} \\
&< \frac{1}{n^2}.
\end{aligned}$$

.

By the union bound over all n agents, the probability that any agent takes more than $8n\mathtt{K}^2 \leq 2304n\log^2 n$ (since $\mathtt{K} \leq 24\log n$ by Lemma 3.2) interactions (equally $2304\log^2 n$ time) is less than $\frac{1}{n}$. ◀

B Simulation

Simulation results are shown in Fig. 2.

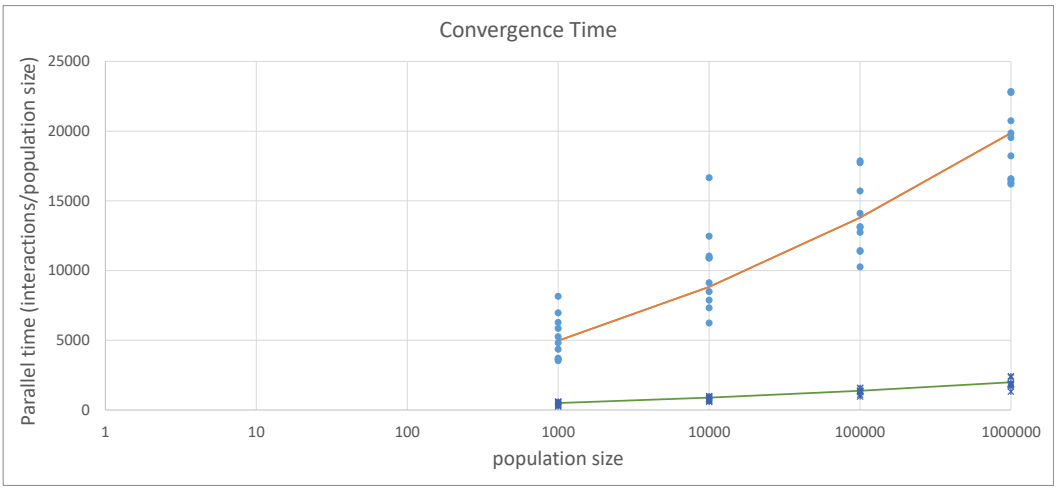

Figure 2 Simulated convergence time of the protocol. Although the proofs give only that the estimate of $\log n$ is likely to get within additive error of 5, in practice the estimate is always within 2, so this is how we define convergence in the experiment. The dots indicate the convergence time of individual experiments. The population size axis is logarithmic (i.e., exactly $O(c \log_{10} n)$ time complexity would correspond to a straight line with slope c). Recall setting c_{init} lower increases the probability that the initial estimate $\mathtt{G}_0$ is too low, i.e., $c_{\text{init}} \cdot \mathtt{G}_0 \leq \log n$. The circular dots in the plot (larger values) are 10 experiments at each value of $n \in \{10^3, 10^4, 10^5, 10^6\}$ with $c_{\text{init}} = 4$, the value used in our main analysis. The line fitted to them goes through the value $50 \log n$ at each value of n. The crosses in the plot (smaller values) are 10 experiments at each value of $n \in \{10^3, 10^4, 10^5, 10^6\}$ with $c_{\text{init}} = 1$, which appears to work in practice (maintaining that the approximation error remains within 2 of $\log n$) and faster by a significant constant. The line fitted to them goes through the value $5 \log n$ at each value of n. For $n \leq 100$ the protocol has large probability of failure to converge due to some index $i \in \{1, \ldots, K\}$ having none of the agents choose to be responsible for i.

C Chernoff bound on sums of maxima of geometric random variables

C.1 Sub-exponential random variables

▶ **Definition C.1.** Let $\alpha, \beta > 0$ and let $\mathbf{X}$ be a random variable. We say $\mathbf{X}$ is *α-β-sub-exponential* if, for all $\lambda > 0$, $\Pr[|\mathbf{X} - \mathrm{E}[\mathbf{X}]| \geq \lambda] \leq \alpha e^{-\lambda/\beta}$.

The following lemma is well-known; we prove it explicitly since the exact form is convenient for our purposes but is more general than typically expressed. It shows that exponential tail bounds for $\Pr[|\mathbf{X} - \mathrm{E}[\mathbf{X}]| > \lambda]$ give bounds on the moment-generating functions of the random variables $\mathbf{X} - \mathrm{E}[\mathbf{X}]$ and $\mathrm{E}[\mathbf{X}] - \mathbf{X}$. The proof is modeled on Rigollet's proof of the analogous lemma for sub-gaussian random variables proven in [30].

▶ **Lemma C.2** ([30]). *Let $\mathbf{X}$ be a α-β-sub-exponential random variable. Then for all $s \in \left[-\frac{1}{2\beta}, \frac{1}{2\beta}\right]$, we have $\mathrm{E}\left[e^{s(\mathbf{X}-\mathrm{E}[\mathbf{X}])}\right], \mathrm{E}\left[e^{s(\mathrm{E}[\mathbf{X}]-\mathbf{X})}\right] \leq 1 + 2\alpha\beta^2 s^2$.*

Proof. Let $k \in \mathbb{N}^+$. Then

$$\begin{aligned}
\mathrm{E}\left[|\mathbf{X} - \mathrm{E}[\mathbf{X}]|^k\right] &= \int_0^\infty \Pr\left[|\mathbf{X} - \mathrm{E}[\mathbf{X}]|^k \geq \lambda\right] d\lambda = \int_0^\infty \Pr\left[|\mathbf{X} - \mathrm{E}[\mathbf{X}]| \geq \lambda^{1/k}\right] d\lambda \\
&\leq \int_0^\infty \alpha e^{-\lambda^{1/k}/\beta} d\lambda = \alpha\beta^k k \int_0^\infty e^{-u} u^{k-1} du \qquad \text{substituting } u = \beta\lambda^{1/k} \\
&= \alpha\beta^k k\Gamma(k) = \alpha\beta^k k!,
\end{aligned}$$

where $\Gamma(k) = \int_0^\infty e^{-u}u^{k-1}du$ is the *gamma function*, known to equal $(k-1)!$ for $k \in \mathbb{N}^+$. Then for all $s \in \left[-\frac{1}{2\beta}, \frac{1}{2\beta}\right]$,

$$\begin{aligned}
\mathrm{E}\left[e^{s(\mathbf{X}-\mathrm{E}[\mathbf{X}])}\right] &= \mathrm{E}\left[\sum_{k=0}^\infty \frac{(s(\mathbf{X} - \mathrm{E}[\mathbf{X}]))^k}{k!}\right] \qquad \text{Taylor expansion of the exponential function} \\
&= \sum_{k=0}^\infty \frac{s^k \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])^k\right]}{k!} \qquad \text{dominated convergence theorem} \\
&= 1 + s\underbrace{\mathrm{E}[\mathbf{X} - \mathrm{E}[\mathbf{X}]]}_{=0} + \sum_{k=2}^\infty \frac{s^k \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])^k\right]}{k!} \\
&\leq 1 + \sum_{k=2}^\infty \frac{|s|^k \mathrm{E}\left[|\mathbf{X} - \mathrm{E}[\mathbf{X}]|^k\right]}{k!} \qquad \text{odd terms can only get larger} \\
&\leq 1 + \sum_{k=2}^\infty \frac{|s|^k \alpha\beta^k k!}{k!} = 1 + \alpha\sum_{k=2}^\infty (|s|\beta)^k = 1 + \alpha s^2\beta^2 \sum_{k=0}^\infty (|s|\beta)^k \\
&\leq 1 + \alpha\beta^2 s^2 \sum_{k=0}^\infty \frac{1}{2^k} \qquad \text{since } |s| \leq \frac{1}{2\beta} \\
&= 1 + 2\alpha\beta^2 s^2.
\end{aligned}$$

The bound for $\mathrm{E}\left[e^{s(\mathrm{E}[\mathbf{X}]-\mathbf{X})}\right]$ is derived by a similar argument. ◀

The following Chernoff bound is well-known, but stated in a more convenient form for our purposes.

▶ **Lemma C.3.** *Let $\alpha, \beta > 0$ and $K \in \mathbb{N}^+$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_K$ be i.i.d. α-β-sub-exponential random variables. Define $\mathbf{S} = \sum_{i=1}^{K} \mathbf{X}_i$. Then for all $t \geq 0$,*

$$\Pr\left[|\mathbf{S} - \mathrm{E}\left[\mathbf{S}\right]| \geq t\right] \leq 2\frac{\left(1+\alpha/2\right)^K}{e^{t/(2\beta)}}.$$

Proof. Then for all $s, t > 0$,

$$\begin{aligned}
\Pr\left[\mathbf{S} - \mathrm{E}\left[\mathbf{S}\right] > t\right] &= \Pr\left[e^{s(\mathbf{S}-\mathrm{E}[\mathbf{S}])} > e^{st}\right] \\
&\leq \frac{\mathrm{E}\left[e^{s(\mathbf{S}-\mathrm{E}[\mathbf{S}])}\right]}{e^{st}} && \text{Markov's inequality} \\
&= e^{-st}\mathrm{E}\left[e^{s\left(\left(\sum_{i=1}^{K}\mathbf{X}_i\right)-\mathrm{E}[\mathbf{S}]\right)}\right] \\
&= e^{-st}\mathrm{E}\left[e^{s\sum_{i=1}^{K}(\mathbf{X}_i-\mathrm{E}[\mathbf{X}_i])}\right] && \text{linearity of expectation} \\
&= e^{-st}\mathrm{E}\left[\prod_{i=1}^{K} e^{s(\mathbf{X}_i-\mathrm{E}[\mathbf{X}_i])}\right] \\
&= e^{-st}\prod_{i=1}^{K}\mathrm{E}\left[e^{s(\mathbf{X}_i-\mathrm{E}[\mathbf{X}_i])}\right]. && \text{independence of the } \mathbf{X}_i\text{'s}
\end{aligned}$$

By Lemma C.2, for all $|s| \leq \frac{1}{2\beta}$, $\mathrm{E}\left[e^{s(\mathbf{X}_i-\mathrm{E}[\mathbf{X}_i])}\right] \leq 1 + 2\alpha\beta^2 s^2$, so letting $s = \frac{1}{2\beta}$,

$$\Pr\left[\mathbf{S} - \mathrm{E}\left[\mathbf{S}\right] > t\right] \leq e^{-st}\left(1 + 2\alpha\beta^2 s^2\right)^K = e^{-t/(2\beta)}\left(1+\alpha/2\right)^K.$$

The proof that $\Pr\left[\mathrm{E}\left[\mathbf{S}\right] - \mathbf{S} \geq t\right] < e^{-t/(2\beta)}\left(1+\alpha/2\right)^K$ is symmetric. By the union bound, $\Pr\left[|\mathbf{S} - \mathrm{E}\left[\mathbf{S}\right]| \geq t\right] < 2 \cdot e^{-t/(2\beta)}\left(1+\alpha/2\right)^K$. ◀

C.2 Geometric random variables and their maximum

We say $\mathbf{G}$ is a *p-geometric random variable* if it is the number of consecutive flips until the first H (including the H), when flipping a coin with $\Pr\left[H\right] = p$. Thus $\mathrm{E}\left[\mathbf{G}\right] = \frac{1}{p}$; in particular $\mathrm{E}\left[\mathbf{G}\right] = 2$ if $p = \frac{1}{2}$.

Defining $\mathbf{M} = \max_{1 \leq i \leq n} \mathbf{G}_i$, where each $\mathbf{G}_i$ is an i.i.d. $\frac{1}{2}$-geometric random variable, it is known [23] that $\mathrm{E}\left[\mathbf{M}\right] \approx \log n$. Lemma C.5 shows a tail bound on $\mathbf{M}$ for general p-geometric random variables, which we will later apply to the case $p = \frac{1}{2}$.

We first require a technical lemma relating $\mathrm{E}\left[\mathbf{M}\right]$ and $\log n$ more precisely, and more generally for p-geometric random variables for $p \neq \frac{1}{2}$. Let $H_n = \sum_{i=1}^{n} \frac{1}{n}$ be the n'th harmonic number. Let $\gamma = \lim_{n\to\infty}\left(H_n - \ln n\right) \approx 0.577$ be the Euler-Mascheroni constant; for all $n \geq 50$ we have $H_n - \ln n - \gamma \leq 0.01$.

▶ **Lemma C.4.** *Let $\mathbf{G}_1, \ldots, \mathbf{G}_n$ be i.i.d. p-geometric random variables with $q = 1 - p \geq \frac{1}{e}$, $n \geq 50$, and let $\mathbf{M} = \max_{1 \leq i \leq n} \mathbf{G}_i$. Let $\epsilon_1 = 0.01$ and $\epsilon_2 = 0.0006$. Then for all $\lambda > 0$, $\frac{\ln n + \gamma}{\ln 1/q} + 1/2 - \epsilon_2 < \mathrm{E}\left[\mathbf{M}\right] < \frac{\ln n + \gamma + \epsilon_1}{\ln 1/q} + 1/2 + \epsilon_2$; particularly for $q = p = 1/2$, we have: $\log n + 1 < \mathrm{E}\left[\mathbf{M}\right] < \log n + 3/2$.*

Proof. Eisenberg [23] showed that if $q \geq \frac{1}{e}$, then $\frac{1}{\lambda}H_n - 0.0006 \leq \mathrm{E}\left[\mathbf{M}\right] - 1/2 < \frac{1}{\lambda}H_n + 0.0006$, where $q = e^{-\lambda}$, i.e. $\lambda = \ln(1/q)$. Thus $\frac{1}{\lambda}H_n + 1/2 - \epsilon_2 \leq \mathrm{E}\left[\mathbf{M}\right] < \frac{1}{\lambda}H_n + 1/2 + \epsilon_2$ i.e., $\frac{\ln n + \gamma}{\ln 1/q} + 1/2 - \epsilon_2 < \mathrm{E}\left[\mathbf{M}\right] < \frac{\ln n + \gamma + \epsilon_1}{\ln 1/q} + 1/2 + \epsilon_2$. ◀

▶ **Lemma C.5.** *Let $\mathbf{G}_1, \ldots, \mathbf{G}_n$ be i.i.d. p-geometric random variables with $q = 1 - p \geq \frac{1}{e}$, $n \geq 50$, and let $\mathbf{M} = \max_{1 \leq i \leq n} \mathbf{G}_i$. Let $\epsilon_1 = 0.01$ and $\epsilon_2 = 0.0006$. Then for all $\lambda > 0$, $\Pr[\mathrm{E}[\mathbf{M}] - \mathbf{M} \geq \lambda] \leq \exp\left(-q^{1/2+\epsilon_2-(\gamma+\epsilon_1)\ln q-\lambda}\right)$ and $\Pr[\mathbf{M} - \mathrm{E}[\mathbf{M}] \geq \lambda] \leq q^{\lambda-1/2-\epsilon_2-\gamma\ln q} + q^{2\lambda-1-2\epsilon_2-2\gamma\ln q}$.*

Proof. For each $t \in \mathbb{N}$, $\Pr[\mathbf{G}_i \geq t] = q^{t-1}$, so $\Pr[\mathbf{G}_i \leq t] = 1 - \Pr[\mathbf{G}_i \geq t+1] = 1 - q^t$.

Since the $\mathbf{G}_i$'s are independent, $\Pr[\mathbf{M} \leq t] = \prod_{i=1}^{n}(1-q^t) = (1-q^t)^n$.

Below we use Lemma C.4 and the inequalities $e^x\left(1 - \frac{x^2}{n}\right) \leq \left(1 + \frac{x}{n}\right)^n \leq e^x$ for $n > 1, |x| < n$.

Setting $t = \mathrm{E}[\mathbf{M}] - \lambda$, we have

$$\begin{aligned}
\Pr[\mathbf{M} \leq \mathrm{E}[\mathbf{M}] - \lambda] &= \left(1-q^t\right)^n \\
&= \left(1 - q^{(\mathrm{E}[\mathbf{M}]-\lambda)}\right)^n \\
&< \left(1 - q^{\log_{1/q} n-(\gamma+\epsilon_1)\ln q+1/2+\epsilon_2-\lambda}\right)^n \\
&= \left(1 - q^{\log_{1/q} n} q^{1/2+\epsilon_2-(\gamma+\epsilon_1)\ln q-\lambda}\right)^n \\
&= \left(1 - \frac{q^{1/2+\epsilon_2-(\gamma+\epsilon_1)\ln q-\lambda}}{n}\right)^n \\
&< \exp\left(-q^{1/2+\epsilon_2-(\gamma+\epsilon_1)\ln q-\lambda}\right) \quad \text{since } \left(1+\frac{x}{n}\right)^n \leq e^x
\end{aligned}$$

Similarly, letting $t = \mathrm{E}[\mathbf{M}] + \lambda - 1$, we have

$$\begin{aligned}
&\Pr[\mathbf{M} \geq \mathrm{E}[\mathbf{M}] + \lambda] \\
=\ & 1 - \Pr[\mathbf{M} \leq \mathrm{E}[\mathbf{M}] + \lambda - 1] \\
=\ & 1 - \left(1-q^t\right)^n \\
=\ & 1 - \left(1 - q^{\mathrm{E}[\mathbf{M}]+\lambda-1}\right)^n \\
<\ & 1 - \left(1 - q^{\log_{1/q} n+1/2-\epsilon_2-\gamma\ln q+\lambda-1}\right)^n \\
=\ & 1 - \left(1 - q^{\log_{1/q} n} q^{\lambda-1/2-\epsilon_2-\gamma\ln q}\right)^n \\
=\ & 1 - \left(1 - \frac{q^{\lambda-1/2-\epsilon_2-\gamma\ln q}}{n}\right)^n \\
<\ & 1 - \exp\left(-q^{\lambda-1/2-\epsilon_2-\gamma\ln q}\right)\left(1 - \frac{q^{2(\lambda-1/2-\epsilon_2-\gamma\ln q)}}{n}\right) \quad \text{since } e^x\left(1-\frac{x^2}{n}\right) \leq \left(1+\frac{x}{n}\right)^n \\
=\ & 1 - \exp\left(-q^{\lambda-1/2-\epsilon_2-\gamma\ln q} + \ln\left(1 - \frac{q^{2\lambda-1-2\epsilon_2-2\gamma\ln q}}{n}\right)\right) \\
\leq\ & q^{\lambda-1/2-\epsilon_2-\gamma\ln q} - \ln\left(1 - \frac{q^{2\lambda-1-2\epsilon_2-2\gamma\ln q}}{n}\right) \quad \text{since } 1 - e^x \leq -x \\
\leq\ & q^{\lambda-1/2-\epsilon_2-\gamma\ln q} + \frac{2q^{2\lambda-1-2\epsilon_2-2\gamma\ln q}}{n} \quad \text{since } \ln(1-x) \geq -2x \text{ if } x < 0.7 \\
\leq\ & q^{\lambda-1/2-\epsilon_2-\gamma\ln q} + q^{2\lambda-1-2\epsilon_2-2\gamma\ln q} \quad \text{since } n \geq 2. \quad \blacktriangleleft
\end{aligned}$$

The following corollary for the special case of $p = \frac{1}{2}$ is used for our main result, showing that a maximum of $\frac{1}{2}$-geometric random variables is α-β-sub-exponential for $\alpha = 3.31, \beta = 2$.

▶ **Corollary C.6.** *Let $\mathbf{G}_1, \ldots, \mathbf{G}_n$ be i.i.d. $\frac{1}{2}$-geometric random variables, $n \geq 50$, and let $\mathbf{M} = \max_{1 \leq i \leq n} \mathbf{G}_i$. Then for all $\lambda > 0$, $\Pr[|\mathbf{M} - \mathrm{E}[\mathbf{M}]| \geq \lambda] < 3.31e^{-\lambda/2}$.*

Proof. By Lemma C.5 and the union bound,

$$\begin{aligned}\Pr\left[|\mathbf{M} - \mathrm{E}\left[\mathbf{M}\right]| \geq \lambda\right] &< \exp\left(-q^{1/2+\epsilon_2-(\gamma+\epsilon_1)\ln q-\lambda}\right) + q^{\lambda-1/2-\epsilon_2-\gamma\ln q} + q^{2\lambda-1-2\epsilon_2-2\gamma\ln q} \\ &= \exp\left(-2^{\lambda-(\gamma+\epsilon_1)\ln 2-1/2-\epsilon_2}\right) + 2^{1/2+\epsilon_2-\lambda-\gamma\ln 2} + 2^{1+2\epsilon_2-2\lambda-2\gamma\ln 2} \\ &< 3.31e^{-\lambda/2}. \quad \text{justified below}\end{aligned}$$

To see the final inequality, note that

$$\begin{aligned}\exp\left(-2^{\lambda-(\gamma+\epsilon_1)\ln 2-1/2-\epsilon_2}\right) &< \exp\left(-2^{\lambda-1}\right) \\ &= \exp\left(-2^{\lambda}/2\right) \\ &\leq \exp\left(-\lambda/2\right)\end{aligned}$$

$$\begin{aligned}2^{1/2+\epsilon_2-\lambda-\gamma\ln 2} &= 2^{1/2+\epsilon_2-\gamma\ln 2}\cdot 2^{-\lambda} \\ &= 2^{1/2+\epsilon_2-\gamma\ln 2}\cdot 4^{-\lambda/2} \\ &< 2^{1/2+\epsilon_2-\gamma\ln 2}\cdot e^{-\lambda/2} \\ &< 1.1\cdot e^{-\lambda/2}.\end{aligned}$$

$$\begin{aligned}2^{1+2\epsilon_2-2\lambda-2\gamma\ln 2} &= 2^{1+2\epsilon_2-2\gamma\ln 2}\cdot 2^{-2\lambda} \\ &= 2^{1+2\epsilon_2-2\gamma\ln 2}\cdot 16^{-\lambda/2} \\ &< (1.1)^2\cdot e^{-\lambda/2}.\end{aligned}$$

So, their sum is less than $3.31e^{-\lambda/2}$. ◀

The following is a stronger corollary of Lemma C.5 for $p = \frac{1}{2}$, which does not fit the definition of α-β-sub-exponential, but which will be useful elsewhere in our analysis.

▶ **Corollary C.7.** *Let* $\mathbf{G}_1, \ldots, \mathbf{G}_n$ *be i.i.d.* $\frac{1}{2}$*-geometric random variables,* $n \geq 50$*, and* $\mathbf{M} = \max_{1\leq i\leq n} \mathbf{G}_i$*. Then* $\Pr\left[\mathbf{M} \geq \mathrm{E}\left[\mathbf{M}\right] + \log n\right] < 0.32\cdot n^{-1}$ *and* $\Pr\left[\mathbf{M} \leq \mathrm{E}\left[\mathbf{M}\right] - (1 + \log\ln n)\right] < n^{-1}$*.*

Proof. Note that for all $\lambda > 0$, setting $q = \frac{1}{2}$,

$$\begin{aligned}\Pr\left[\mathbf{M} - \mathrm{E}\left[\mathbf{M}\right] \geq \lambda\right] &< q^{\lambda-1/2-\epsilon_2-\gamma\ln q} + q^{2\lambda-1-2\epsilon_2-2\gamma\ln q} \\ &= 2^{1/2+\epsilon_2-\lambda-\gamma\ln 2} + 2^{1+2\epsilon_2-2\lambda-2\gamma\ln 2}.\end{aligned}$$

Setting $\lambda = \log n$,

$$\begin{aligned}2^{1/2+\epsilon_2-\lambda-\gamma\ln 2} &= 2^{1/2+\epsilon_2-\log n-\gamma\ln 2} \\ &= 2^{1/2+\epsilon_2-\gamma\ln 2}\cdot 2^{-\log n} \\ &< 0.11\cdot n^{-1}.\end{aligned}$$

$$\begin{aligned} 2^{1+2\epsilon_2-2\lambda-2\gamma\ln 2} &= 2^{1+2\epsilon_2-2\log n-2\gamma\ln 2} \\ &= 2^{1+2\epsilon_2-2\gamma\ln 2}\cdot 2^{-2\log n} \\ &< 0.21\cdot n^{-2} \\ &< 0.21\cdot n^{-1}. \end{aligned}$$

Combining these, we obtain the first inequality $\Pr[\mathbf{M}-\mathrm{E}[\mathbf{M}]\geq\log n]<0.32\cdot n^{-1}$. To see the second, note that setting $q=\frac{1}{2}$, and letting $\lambda=1+\log\ln n$

$$\begin{aligned} \Pr[\mathrm{E}[\mathbf{M}]-\mathbf{M}\geq\lambda] &< \exp\left(-q^{1/2+\epsilon_2-(\gamma+\epsilon_1)\ln q-\lambda}\right) \\ &= \exp\left(-2^{\lambda-(\gamma+\epsilon_1)\ln 2-1/2-\epsilon_2}\right) \\ &< \exp\left(-2^{\lambda-1}\right) \\ &= \exp\left(-2^{1+\log\ln n-1}\right) \\ &= \exp\left(-\ln n\right) \\ &= n^{-1}. \quad \blacktriangleleft \end{aligned}$$

▶ **Lemma C.8.** *Let $n,K\in\mathbb{N}^+$, $n\geq 50$. Let $\mathbf{M}_1,\ldots,\mathbf{M}_K$ be i.i.d. random variables, each of which is the maximum of n i.i.d. $\frac{1}{2}$-geometric random variables. Define $\mathbf{S}=\sum_{i=1}^{K}\mathbf{M}_i$. Then for all $t\geq 0$, $\Pr[|\mathbf{S}-\mathrm{E}[\mathbf{S}]|\geq t]\leq 2\cdot e^{K-t/4}$.*

Proof. By Corollary C.6 and Lemma C.3, for $\alpha=3.31<2e-2$ and $\beta=2$, we have

$$\Pr[|\mathbf{S}-\mathrm{E}[\mathbf{S}]|\geq t] < 2\frac{(1+\alpha/2)^K}{e^{t/(2\beta)}} < 2\frac{\left(1+\frac{2e-2}{2}\right)^K}{e^{t/4}} = 2\cdot(e)^K\cdot e^{-t/4} = 2\cdot e^{K-t/4}. \quad \blacktriangleleft$$

▶ **Corollary C.9.** *Let $a>4$, $n\in\mathbb{N}^+$, $n>50$, $K\geq\frac{\ln n}{\frac{a}{4}-1}$, and $\delta_0=1/2+\gamma/\ln 2-\epsilon_2$. Let $\mathbf{M}_1,\ldots,\mathbf{M}_K$ be i.i.d. random variables, each of which is the maximum of n i.i.d. $\frac{1}{2}$-geometric random variables. Define $\mathbf{S}=\sum_{i=1}^{K}\mathbf{M}_i$. Then*

$$\Pr\left[\left|\frac{\mathbf{S}}{K}-\log n-\delta_0\right|\geq a\right]\leq\frac{2}{n}.$$

Proof. We first manipulate the expression in the conclusion of the corollary to put it in a form where we can apply Lemma C.8.

$$\begin{aligned} &\Pr\left[\frac{\mathbf{S}}{K}-\log n-\delta_0\geq a\right] \\ =\ &\Pr[\mathbf{S}-K(\log n+1/2+\gamma/\ln 2-\epsilon_2)\geq aK] \\ <\ &\Pr[\mathbf{S}-\mathrm{E}[\mathbf{S}]\geq aK]. \qquad \text{Since } K(\log n+1/2+\gamma/\ln 2-\epsilon_2)\leq E[\mathbf{S}] \end{aligned}$$

$$\begin{aligned} &\Pr\left[\log n+\delta_0-\frac{\mathbf{S}}{K}\geq a\right] \\ =\ &\Pr[K(\log n+1/2+\gamma/\ln 2-\epsilon_2)-\mathbf{S}\geq aK] \\ =\ &\Pr[K(\log n+1/2+\gamma/\ln 2+\epsilon_1/\ln 2+\epsilon_2)-\mathbf{S}\geq aK+(\epsilon_1/\ln 2+2\epsilon_2)K] \\ <\ &\Pr[\mathrm{E}[\mathbf{S}]-\mathbf{S}\geq(a+\epsilon_1/\ln 2+2\epsilon_2)K] \quad \text{Since } E[\mathbf{S}]<K\left(\log n+1/2+\frac{\gamma+\epsilon_1}{\ln 2}+\epsilon_2\right) \\ <\ &\Pr[\mathrm{E}[\mathbf{S}]-\mathbf{S}\geq aK]. \end{aligned}$$

Since the events $\frac{\mathbf{S}}{K} - \log n - \delta_0 \geq a$ and $\log n + \delta_0 - \frac{\mathbf{S}}{K} \geq a$ are disjoint, and the events $\mathbf{S} - \mathrm{E}[\mathbf{S}] \geq aK$ and $\mathrm{E}[\mathbf{S}] - \mathbf{S} \geq aK$ are disjoint, the union bound holds with equality, so

$$\begin{aligned}
\Pr\left[\left|\frac{\mathbf{S}}{K} - \log n - \delta_0\right| \geq a\right] &= \Pr\left[\frac{\mathbf{S}}{K} - \log n - \delta_0 \geq a\right] + \Pr\left[\log n + \delta_0 - \frac{\mathbf{S}}{K} \geq a\right] \\
&< \Pr[\mathbf{S} - \mathrm{E}[\mathbf{S}] \geq aK] + \Pr[\mathrm{E}[\mathbf{S}] - \mathbf{S} \geq aK] \\
&= \Pr[|\mathbf{S} - \mathrm{E}[\mathbf{S}]| \geq aK].
\end{aligned}$$

Let $t = aK$. Applying Lemma C.8 with these values of K and t,

$$\begin{aligned}
\Pr\left[\left|\frac{\mathbf{S}}{K} - \log n - \delta_0\right| \geq a\right] &< \Pr[|\mathbf{S} - \mathrm{E}[\mathbf{S}]| \geq aK] \\
&= \Pr[|\mathbf{S} - \mathrm{E}[\mathbf{S}]| \geq t] \\
&\leq 2 \cdot e^{K - t/4} \\
&= 2 \cdot e^{K(1 - \frac{a}{4})} \\
&= 2 \cdot e^{-K(\frac{a}{4} - 1)} \\
&\leq 2 \cdot e^{-\frac{\ln n}{(\frac{a}{4} - 1)}(\frac{a}{4} - 1)} \\
&= 2 \cdot e^{-\ln n} \\
&= \frac{2}{n}. \quad \blacktriangleleft
\end{aligned}$$

For example, choosing $a = \ln 2 + 4 < 4.7$ means we can choose $K \geq \frac{\ln n}{\frac{a}{4} - 1} = \frac{\ln n}{(\ln(2) + 4)/4 - 1} = \frac{\ln n}{\ln(2)/4} = 4 \log_2 n$:

▶ **Corollary C.10.** *Let $n \in \mathbb{N}^+$, $n \geq 50$, $K \geq 4 \log n$. Let $\mathbf{M}_1, \ldots, \mathbf{M}_K$ be i.i.d. random variables, each of which is the maximum of n i.i.d. $\frac{1}{2}$-geometric random variables. Define $\mathbf{S} = \sum_{i=1}^{K} \mathbf{M}_i$. Then*

$$\Pr\left[\left|\frac{\mathbf{S}}{K} - \log n\right| \geq 4.7\right] \leq \frac{2}{n}.$$

D Timer lemma

▶ **Lemma D.1.** *Let $0 < \delta \leq \frac{1}{2}$. Let $0 < k \leq n$ and m be positive integers. Suppose we have n bins, of which k are initially empty, and we throw m additional balls randomly into the n bins. Then* $\Pr[\leq \delta k \text{ bins remain empty}] < (2\delta e^{m/n})^{\delta k}$.

Proof. When there are i bins empty, the probability that the next ball fills an empty bin is $\frac{i}{n}$. Thus, the number of balls needed until $\leq \delta k$ bins are empty is a sum $\mathbf{S} = \sum_{i=\delta k+1}^{k} \mathbf{G}_i$ of independent geometric random variables $\mathbf{G}_{\delta k+1}, \ldots, \mathbf{G}_k$, where $\mathbf{G}_i$ has $p_i = \Pr[\text{success}] = \frac{i}{n}$, "success" representing the event of throwing a ball into one of the k initially empty bins.

The moment-generating function of a geometric random variable $\mathbf{G}$ with $\Pr[\text{success}] = p$, defined whenever $\theta < -\ln(1-p)$ [29], is

$$\mathrm{E}\left[e^{\theta \mathbf{G}}\right] = \frac{pe^{\theta}}{1-(1-p)e^{\theta}} = \frac{p}{e^{-\theta}-1+p} \leq \frac{p}{p-\theta},$$

where the last inequality follows from $e^x - 1 \geq x$ for all $x \in \mathbb{R}$. Thus for each $i \in \{\delta k, \ldots, k\}$,

$$\mathrm{E}\left[e^{\theta \mathbf{G}_i}\right] \leq \frac{\frac{i}{n}}{\frac{i}{n}-\theta} = \frac{i}{i-\theta n}.$$

By independence of the $\mathbf{G}_i$'s, the moment-generating function of the sum $\mathbf{S}$ is

$$\mathrm{E}\left[e^{\theta \mathbf{S}}\right] = \mathrm{E}\left[e^{\theta \sum_{i=\delta k+1}^{k} \mathbf{G_i}}\right] = \prod_{i=\delta k+1}^{k} \mathrm{E}\left[e^{\theta \mathbf{G_i}}\right] \leq \prod_{i=\delta k+1}^{k} \frac{i}{i-\theta n}.$$

Setting $\theta = -\frac{\delta k}{n}$, and using the fact that $\delta \leq \frac{1}{2}$ to cancel terms, we have

$$\begin{aligned}
\mathrm{E}\left[e^{\theta \mathbf{S}}\right] &\leq \prod_{i=\delta k+1}^{k} \frac{i}{i+\delta k} = \left(\frac{\delta k+1}{\delta k+1+\delta k}\right)\left(\frac{\delta k+2}{\delta k+2+\delta k}\right)\cdots\left(\frac{k}{k+\delta k}\right) \\
&= \frac{(\delta k+1)\ldots(\delta k+\delta k)}{1} \cdot \frac{(\delta k+1+\delta k)\ldots(k)}{(\delta k+1+\delta k)\ldots(k)} \cdot \frac{1}{(k+1)\ldots(k+\delta k)} \\
&= \frac{(\delta k+1)\ldots(\delta k+\delta k)}{(k+1)\ldots(k+\delta k)} < \left(\frac{2\delta k}{k}\right)^{\delta k} = (2\delta)^{\delta k}.
\end{aligned}$$

The event that throwing m balls results in at most δk empty bins is equivalent to the event that $\mathbf{S} \leq m$. By Markov's inequality, since $\theta = -\frac{\delta k}{n} < 0$,

$$\Pr[\mathbf{S} \leq m] = \Pr\left[e^{\theta \mathbf{S}} \geq e^{\theta m}\right] \leq \frac{\mathrm{E}\left[e^{\theta S}\right]}{e^{\theta m}} < (2\delta)^{\delta k} e^{\frac{\delta k}{n} m} = (2\delta e^{m/n})^{\delta k}. \qquad \blacktriangleleft$$

We say a transition *consumes* a state s if executing the transition strictly reduces the count of s, and that the transition *produces* s if it strictly increases the count of s. The next lemma bounds the rate of consumption of s, showing that the count of s cannot decrease too quickly. It also makes the observation that, since we are reasoning about s assuming that it is only consumed, we can upper-bound the probability of the count of s dropping below δk at *any* time $t \in [0, T]$, not just at time $t = T$.

▶ **Lemma D.2.** *Let s be a state in a population protocol, let $0 < \delta \leq \frac{1}{2}$, and let k be the count of s at time 0. Let $\mathbf{C}_{t,s}$ denote the count of s at time t. Then for all $T > 0$,*

$$\Pr[(\exists t \in [0, T])\ \mathbf{C}_{t,s} \leq \delta k] \leq (2\delta e^{3T})^{\delta k}.$$

Proof. s may be produced and consumed. To establish that the count of s remains large for a constant amount of time, in the worst case we assume that s is only consumed. We also make the worst case assumption that each time an agent in state s is picked for a transition, it changes state and we consume that copy of s. We further make the worst-case assumption that if both agents are in state s, both change to a different state.

The following *almost* works: model each transition as throwing *two* balls into bins, where each agent is a bin, considered "empty" if it is in state s. Each time a transition picks an agent, this puts a ball into the bin that agent represents. Thus, the number of balls in a bin represents the total number of times that the agent interacts. However, these are not identically distributed processes, since a bin may be picked twice consecutively in the balls-and-bins distribution, whereas when agents are picked two at a time, the two agents are guaranteed to be unequal. Thus the actual distribution has slightly higher probability of fewer empty bins than the simplified "throw-two-balls-every-transition" approximation.

So instead, consider the distribution of empty bins given by throwing *three* balls for every transition. Suppose $p \in \{\delta k + 1, \ldots, k\}$ bins out of n are currently empty. After the next three balls, the number of empty bins $\mathbf{E}_3$ will be p, $p-1$, $p-2$, or $p-3$. We have that

$$\begin{aligned}
\Pr[\mathbf{E}_3 = p] &= \left(\frac{n-p}{n}\right)^3, \\
\Pr[\mathbf{E}_3 = p-1] &= \left(\frac{n-p}{n}\right)^2 \cdot \frac{p}{n} \\
&\quad + \frac{n-p}{n} \cdot \frac{p}{n} \cdot \frac{n-p-1}{n} \\
&\quad + \frac{p}{n} \cdot \left(\frac{n-p-1}{n}\right)^2, \\
\Pr[\mathbf{E}_3 = p-2] &= \frac{n-p}{n} \cdot \frac{p}{n} \cdot \frac{p-1}{n} \\
&\quad + \frac{p}{n} \cdot \frac{n-p-1}{n} \cdot \frac{p-1}{n} \\
&\quad + \frac{p}{n} \cdot \frac{p-1}{n} \cdot \frac{n-p-2}{n}, \\
\Pr[\mathbf{E}_3 = p-3] &= \frac{p}{n} \cdot \frac{p-1}{n} \cdot \frac{p-2}{n}.
\end{aligned}$$

Compare this to the true distribution $\mathbf{E}_2$ of the number of empty bins after one interaction, where two unequal bins are picked at random each to get a ball. Then

$$\begin{aligned}
\Pr[\mathbf{E}_2 = p] &= \frac{\binom{n-p}{2}}{\binom{n}{2}} = \frac{(n-p)(n-p-1)}{n(n-1)}, \\
\Pr[\mathbf{E}_2 = p-1] &= \frac{n-p}{n} \cdot \frac{p}{n}, \\
\Pr[\mathbf{E}_2 = p-2] &= \frac{\binom{p}{2}}{\binom{n}{2}} \cdot \frac{\binom{p-1}{2}}{\binom{n}{2}} = \frac{p(p-1)}{n(n-1)} \cdot \frac{(p-1)(p-2)}{n(n-1)} \\
\Pr[\mathbf{E}_2 = p-3] &= 0
\end{aligned}$$

It can be verified by inspection that for each $\ell \in \{p-3, p-2, p-1, p\}$,

$$\Pr[\mathbf{E}_2 \leq \ell] = \sum_{\ell' \in \{p-3,\ldots,\ell\}} \Pr[\mathbf{E}_2 = \ell'] < \sum_{\ell' \in \{p-3,\ldots,\ell\}} \Pr[\mathbf{E}_3 = \ell'] = \Pr[\mathbf{E}_3 \leq \ell].$$

Thus, the distribution of empty bins given by throwing three balls independently at random stochastically dominates the true distribution of empty bins after one interaction.

The number of interactions in time T is Tn. Using the stochastically dominating distribution above, we model this as throwing $m = 3Tn$ balls independently. By Lemma D.1,

$$\Pr\left[(\exists t \in [0,T])\ \mathbf{C}_{t,s} \leq \delta k\right] \leq (2\delta e^{m/n})^{\delta k} = (2\delta e^{3T})^{\delta k}. \quad \blacktriangleleft$$

The following corollary with $\delta = \frac{1}{81}$ and $T = 1$ states that within time 1, it is unlikely for the count of any state to decrease by more than factor 81 from k to $k/81$.

▶ **Corollary D.3.** *Let s be a state in a population protocol, and let k be the count of s at time 0. Let $\mathbf{C}_{t,s}$ denote the count of s at time t. Then*

$$\Pr\left[(\exists t \in [0,1])\ \mathbf{C}_{t,s} \leq k/81\right] \leq 2^{-k/81}.$$

Proof. Note that $2e^3 < 40.2$, so setting $\delta = \frac{1}{81}$ and $T = 1$ implies that $2\delta e^{3T} < \frac{1}{2}$. Applying Lemma D.2 with $\delta = \frac{1}{81}$ and $T = 1$, we have

$$\Pr\left[(\exists t \in [0,1])\ \mathbf{C}_{t,s} \leq k/81\right] < (2\delta e^{3T})^{\delta k} < 2^{-k/81}. \quad \blacktriangleleft$$

Recall the timer lemma used in Section 4. The proof follows the same structure as the main theorem of [19], but uses the discrete-time model of population protocols rather than the continuous-time model of chemical reaction networks. Additionally, care must be taken to show that although the number of states is infinite (so clearly only a finite number can appear in finite time), those states producible via a constant number of transitions, whose probabilities are bounded below by a positive constant, in sufficiently large dense configurations are all produced in large quantity in constant time.

Lemma 4.2. *Let $\alpha > 0$, $m \in \mathbb{N}^+$, $\rho \in (0,1]$, and P be a population protocol. Then there are constants $\epsilon, \delta, n_0 > 0$ such that, for all $n \geq n_0$, for all α-dense configurations $\vec{c}$ of P with $n = ||\vec{c}||$, the following holds. Let Λ_ρ^m be the set of states m-ρ-producible from $\vec{c}$. For $s \in \Lambda$ and $t > 0$, let $\mathbf{C}_{t,s}$ be the random variable denoting the count of s at time t, assuming at time 0 the configuration is $\vec{c}$. Then $\Pr\left[(\forall s \in \Lambda_\rho^m)\ \mathbf{C}_{1,s} \geq \delta n\right] \geq 1 - 2^{-\epsilon n}$.*

Proof. We need that $|\Lambda_\rho^m| < \infty$. To see this holds, assume otherwise. Note that if all pairs of states a, b have a finite number of transitions of the form $a, b \to \ldots$, then this implies by induction that each Λ_ρ^m is finite. So assume there are $a, b \in \Lambda$ and an infinite set of transitions $a, b \xrightarrow{\rho_i} \ldots$ for $i \in \mathbb{N}$. Because these are probabilities, $\sum_{i=0}^{\infty} \rho_i \leq 1$. Then $\lim_{i \to \infty} \rho_i = 0$, so for all but finitely many i, we have $\rho_i < \rho$. Transitions with $\rho_i < \rho$ cannot be used to produce states in Λ_ρ^m, as their rate constants smaller than the definition of Λ_ρ^m allows. This shows that $|\Lambda_\rho^m| < \infty$.

By hypothesis all $s \in \Lambda^0$ satisfy $\vec{i}(s) \geq \alpha n$. Fix a particular $s \in \Lambda^0$. Let $k = \alpha n$ in Corollary D.3; then

$$\Pr\left[(\exists t \in [0,1])\ \mathbf{C}_{t,s} < \frac{\alpha n}{81}\right] \leq 2^{-\alpha n/81}.$$

By the union bound,

$$\Pr\left[(\exists s \in \Lambda^0)(\exists t \in [0,1])\ \mathbf{C}_{t,s} < \frac{\alpha n}{81}\right] \leq |\Lambda^0| 2^{-\alpha n/81}. \tag{1}$$

That is, with high probability, all states in Λ^0 have "large" count (at least $\frac{\alpha n}{81}$) for the entire first unit of time. Call this event $H(\Lambda^0)$ (i.e., the complement of the event in (1)).

We complete the proof by a "probabilistic induction" on $i \in \{0, 1, \ldots, m\}$ as follows. We show a sequence $\delta_0 > \delta_1 > \ldots > \delta_m > 0$ such that the following holds. Inductively assume

that for all $s \in \Lambda_\rho^i$ and all $t \in \left[\frac{i}{m+1}, 1\right]$, $\mathbf{C}_{t,s} \geq \delta_i n$. Call this event $H(\Lambda_\rho^i)$. Then we will show that assuming $H(\Lambda_\rho^i)$ holds, with high probability $H(\Lambda_\rho^{i+1})$ holds, i.e., for all $s \in \Lambda_\rho^{i+1}$ and for all $t \in \left[\frac{i+1}{m+1}, 1\right]$, $\mathbf{C}_{t,s} \geq \delta_{i+1} n$. The base case is established by (1) for $\delta_0 = \frac{\alpha}{81}$.

We use Chernoff bounds for binomial random variables, which state that, for $1 \leq i \leq k$, if each $\mathbf{X}_i$ is an independent 0/1-random variable with $\Pr[\mathbf{X}_i = 1] = p$, defining $\mathbf{X} = \sum_{i=1}^k \mathbf{X}_i$ and $\mu = \mathrm{E}[\mathbf{X}] = kp$, then for $0 < \beta \leq 1$, $\Pr[\mathbf{X} \leq (1-\beta)\mu] \leq e^{-\mu\beta^2/2}$ and $\Pr[\mathbf{X} \geq (1+\beta)\mu] \leq e^{-\mu\beta^2/3}$.

To see that the inductive case holds, fix a particular state $s \in \Lambda_\rho^{i+1} \setminus \Lambda_\rho^i$; then $s \in \mathrm{PROD}(\Lambda_\rho^i)$. By the definition of $\mathrm{PROD}(\Lambda_\rho^i)$, s is produced by a transition of the form $x, y \xrightarrow{\rho'} s, s'$ where $x, y \in \Lambda_\rho^i$ and $\rho' \geq \rho$. At time t, the given transition has probability $\rho' \cdot \mathbf{C}_{t,x} \cdot \mathbf{C}_{t,y} / \binom{n}{2}$ (if $x \neq y$) or $\rho' \cdot (\mathbf{C}_{t,x} \cdot (\mathbf{C}_{t,x} - 1)/2) / \binom{n}{2}$ (if $x = y$) of occurring in the next interaction. We make the worst-case assumption that the probability is the latter probability, which is smaller when we also make the worst-case assumption $\mathbf{C}_{t,x} = \mathbf{C}_{t,y} = \delta_i n$, and substitute ρ for ρ' since $\rho' \geq \rho$.

By the induction hypothesis $H(\Lambda_i)$, for all $t \in \left[\frac{i}{m+1}, 1\right]$, $\mathbf{C}_{t,x} \geq \delta_i n$ and $\mathbf{C}_{t,y} \geq \delta_i n$. So for each interaction between time $\frac{i}{m+1}$ and $\frac{i+1}{m+1}$, the probability that it executes transition $x, y \xrightarrow{\rho'} s, s'$ is at least

$$\frac{\rho\delta_i n(\delta_i n - 1)/2}{\binom{n}{2}} = \frac{\rho\delta_i n(\delta_i n - 1)/2}{n(n-1)/2} = \frac{\rho\delta_i(\delta_i n - 1)}{n-1} > \frac{\rho\delta_i(\delta_i n)}{n} = \rho\delta_i^2.$$

There are $\frac{n}{m+1}$ interactions in that time interval. Thus the number of times $x, y \xrightarrow{\rho'} s, s'$ executes in that interval is stochastically dominated by a binomial random variable $\mathbf{X}^+$, with $k = \frac{n}{m+1}$ trials and probability of success $p = \rho\delta_i^2$, and $\mu = \mathrm{E}[\mathbf{X}^+] = kp = \frac{n\rho\delta_i^2}{m+1}$. By the Chernoff bound, setting $\beta = \frac{1}{2}$,

$$\Pr\left[\mathbf{X}^+ \leq \frac{n\rho\delta_i^2}{2(m+1)}\right] = \Pr\left[\mathbf{X}^+ \leq (1-\beta)\mu\right] \leq e^{-\mu\beta^2/2} = \exp\left(-\frac{n\rho\delta_i^2}{8(m+1)}\right).$$

The above analysis lower bounds how many times transition $x, y \xrightarrow{\rho'} s, s'$ executes, producing s each time. We now upper bound how many times s is consumed in this same interval. We are trying to show that s gets to a large count, so we make the worst-case assumption that its count starts at 0. Any any time, out of $\binom{n}{2}$ pairs of agents, at most $s(n-1)$ of those pairs have at least one agent in state s. So at time t, the probability that the next transition consumes s is at most $\frac{s(n-1)}{\binom{n}{2}} = 2\frac{\mathbf{C}_{t,s}}{n}$.

Prior to s reaching count $n\rho\delta_i^2/32$, we can make the worst case assumption that the probability of each transition consuming s is exactly $2\frac{n\rho\delta_i^2/32}{n} = \rho\delta_i^2/16$. In this worst case the number of transitions consuming s in the $k = n/m$ interactions in the time interval $\left[\frac{i}{m+1}, \frac{i+1}{m+1}\right]$ is stochastically dominated by $2\mathbf{X}^-$, where $\mathbf{X}^-$ is a binomial random variable with $k = n/(m+1)$ trials and probability of success $p = \rho\delta_i^2/16$. (We consider $2\mathbf{X}^-$ instead of $\mathbf{X}^-$ to account for the fact that each transition in the worst case could consume 2 copies of s.) Apply the Chernoff bound with $\mu = \mathrm{E}[\mathbf{X}^-] = kp = \frac{n\rho\delta_i^2}{16(m+1)}$ and $\beta = 1$ to give

$$\Pr\left[\mathbf{X}^- \geq \frac{n\rho\delta_i^2}{8(m+1)}\right] \quad = \quad \Pr\left[\mathbf{X}^- \geq (1+\beta)\mu\right] \leq e^{-\mu\beta^2/3} = \exp\left(-\frac{n\rho\delta_i^2}{48(m+1)}\right).$$

Thus

$$\Pr\left[2\mathbf{X}^- \geq \frac{n\rho\delta_i^2}{4(m+1)}\right] \leq \exp\left(-\frac{n\rho\delta_i^2}{48(m+1)}\right).$$

Note that this count threshold is half the count threshold we derived for the lower bound on the number of transitions $x, y \xrightarrow{\rho} s, s'$ producing s. Thus, applying the union bound to these two events to bound $\mathbf{X}^+ - 2\mathbf{X}^-$, the *net* production of s (number produced minus number consumed), we have that

$$\begin{aligned}
\Pr\left[\mathbf{X}^+ - 2\mathbf{X}^- \leq \frac{n\rho\delta_i^2}{4(m+1)}\right] &\leq \Pr\left[\mathbf{X}^+ \leq \frac{n\rho\delta_i^2}{2(m+1)} \text{ or } 2\mathbf{X}^- \geq \frac{n\rho\delta_i^2}{4(m+1)}\right] \\
&\leq \exp\left(-\frac{n\rho\delta_i^2}{8(m+1)}\right) + \exp\left(-\frac{n\rho\delta_i^2}{48(m+1)}\right) \\
&< 2 \cdot \exp\left(-\frac{n\rho\delta_i^2}{48(m+1)}\right).
\end{aligned}$$

So with probability at least $1 - 2 \cdot \exp\left(-\frac{n\rho\delta_i^2}{48(m+1)}\right)$, at least $n\rho\delta_i^2/(4(m+1))$ net copies of s are produced at some point in the time interval $\left[\frac{i}{m+1}, \frac{i+1}{m+1}\right]$.

Letting $k = n\rho\delta_i^2/(4(m+1))$ in Corollary D.3, with probability at least $1 - 2^{-k/81} = 1 - 2^{-n\rho\delta_i^2/(324(m+1))}$, we have that $\mathbf{C}_{t,s} \geq \delta_i n/81$ for all times $t \in \left[\frac{i+1}{m+1}, 1\right]$. Setting $\delta_{i+1} = (k/n)/81 = \rho\delta_i^2/(324(m+1))$ proves the inductive case with probability of failure at most

$$2 \cdot \exp\left(-\frac{n\rho\delta_i^2}{48(m+1)}\right) + 2^{-\delta_{i+1}n}$$

By the union bound over all $|\Lambda_\rho^m|$ states in all levels of induction, setting $\delta = \delta_m = \rho^m(\alpha/2)^{2^m}/(324(m+1))^m$, noting that $\delta \leq \delta_i$ for all $0 \leq i \leq m$, with probability most

$$|\Lambda_\rho^m| \left(2 \cdot \exp\left(-\frac{n\rho\delta^2}{48(m+1)}\right) + 2^{-\delta n}\right),$$

the count of all states in Λ_ρ^m fails to reach at least δn by time $t = 1$. By setting n_0 sufficiently large, the above probability is < 1 for all $n = n_0$ (and therefore for all greater n as well). By setting $\epsilon > 0$ sufficiently small, this probability is at most $2^{-\epsilon n}$. ◀