

How Many Cooks Spoil the Soup?*

Othon Michail^{1,2} and Paul G. Spirakis^{1,2}

¹ Department of Computer Science, University of Liverpool, Liverpool, UK

² Computer Technology Institute and Press “Diophantus” (CTI), Patras, Greece
Email: Othon.Michail@liverpool.ac.uk, P.Spirakis@liverpool.ac.uk

Abstract. In this work, we study the following basic question: “*How much parallelism does a distributed task permit?*” Our definition of *parallelism* (or *symmetry*) here is not in terms of speed, but in terms of identical *roles* that processes have at the same time in the execution. We initiate this study in population protocols, a very simple model that not only allows for a straightforward definition of what a role is, but also encloses the challenge of isolating the properties that are due to the protocol from those that are due to the *adversary scheduler*, who controls the interactions between the processes. We (i) give a *partial characterization* of the set of predicates on input assignments that can be *stably computed with maximum symmetry*, i.e., $\Theta(N_{min})$, where N_{min} is the minimum multiplicity of a state in the initial configuration, and (ii) we turn our attention to the remaining predicates and prove a *strong impossibility result* for the *parity* predicate: the inherent symmetry of any protocol that stably computes it is *upper bounded by a constant that depends on the size of the protocol*.

1 Introduction

George Washington said “*My observation on every employment in life is, that, wherever and whenever one person is found adequate to the discharge of a duty by close application thereto, it is worse executed by two persons, and scarcely done at all if three or more are employed therein*”. The goal of the present paper is to investigate whether the analogue of this observation in simple distributed systems is true. In particular, we ask whether a task that can be solved when a single process has a crucial duty is still solvable when that (and any other) duty is assigned to more than one process. Moreover, we are interested in quantifying the degree of *parallelism* (also called *symmetry* in this paper) that a task is susceptible of.

Leader election is a task of outstanding importance for distributed algorithms. One of the oldest [Ang80] and probably still one of the most commonly used approaches [Lyn96, AW04, AAD⁺06, KLO10] for solving a distributed task in a given setting, is to execute a distributed algorithm that manages to elect

* Supported in part by the School of EEE/CS of the University of Liverpool, NeST initiative, and the EU IP FET-Proactive project MULTIPLEX under contract no 317532. The full version can be found at: <https://arxiv.org/abs/1604.07187>

a unique leader (or *coordinator*) in that setting and then compose this (either sequentially or in parallel) with a second algorithm that can solve the task by assuming the existence of a unique leader. Actually, it is quite typical, that the tasks of electing a leader and successfully setting up the composition enclose the difficulty of solving many other higher-level tasks in the given setting.

Due to its usefulness in solving other distributed tasks, the leader election problem has been extensively studied, in a great variety of distributed settings [Lyn96, AW04, FSW14, AG15]. Still, there is an important point that is much less understood, concerning whether an election step is *necessary* for a given task and *to what extent* it can be avoided. Even if a task T can be solved in a given setting by first passing through a configuration with a unique leader, it is still valuable to know whether there is a correct algorithm for T that avoids this. In particular, such an algorithm succeeds without the need to ever have less than k processes in a given “role”, and we are also interested in how large k can be without sacrificing solvability.

Depending on the application, there are several ways of defining what the “role” of a process at a given time in the execution is. In the typical approach of electing a unique leader, a process has the leader role if a *leader* variable in its local memory is set to true and it does not have it otherwise. In other cases, the role of a process could be defined as its complete local history. In such cases, we would consider that two processes have the same role after t steps iff both have the same local history after each one of them has completed t local steps. It could also be defined in terms of the external interface of a process, for example, by the messages that the process transmits, or it could even correspond to the branch of the program that the process executes. In this paper, as we shall see, we will define the role of a process at a given time in the execution, as the entire content of its local memory. So, in this paper, two processes u and v will be regarded to have the same role at a given time t iff, at that time, the local state of u is equal to the local state of v .

Understanding the parallelism that a distributed task allows, is of fundamental importance for the following reasons. First of all, usually, the more parallelism a task allows, the more efficiently it can be solved. Moreover, the less symmetry a solution for a given problem has to achieve in order to succeed, the more vulnerable it is to faults. For an extreme example, if a distributed algorithm elects in every execution a unique leader in order to solve a problem, then a single crash failure (of the leader) can be fatal.

1.1 Our Approach

We have chosen to initiate the study of the above problem in a very minimal distributed setting, namely in Population Protocols of Angluin *et al.* [AAD⁺06] (see Section 1.2 for more details and references). One reason that makes population protocols convenient for the problem under consideration, is that the role of a process at a given step in the execution can be defined in a straightforward way as the state of the process at the beginning of that step. So, for example, if we are interested in an execution of a protocol that stabilizes to the correct

answer without ever electing a unique leader, what we actually require is an execution that, up to stability, never goes through a configuration in which a state q is the state of a single node, which implies that, in every configuration of the execution, every state q is either absent or the state of at least two nodes. Then, it is straightforward to generalize this to any symmetry requirement k , by requiring that, in every configuration, every state q is either absent or the state of at least k nodes.

What is not straightforward in this model (and in any model with adversarially determined events), is how to isolate the symmetry that is *only* due to the protocol. For if we require the above condition on executions to be satisfied *for every* execution of a protocol, then most protocols will fail trivially, because of the power of the adversary scheduler. In particular, there is almost always a way for the scheduler to force the protocol to break symmetry maximally, for example, to make it reach a configuration in which some state is the state of a single node, even when the protocol does not have an *inherent* mechanism of electing a unique state. Moreover, though for computability questions it is sufficient to assume that the scheduler selects in every step a single pair of nodes to interact with each other, this type of a scheduler is problematic for estimating the symmetry of protocols. The reason is that even fundamentally parallel operations, necessarily pass through a highly-symmetry-breaking step. For example, consider the rule $(a, a) \rightarrow (b, b)$ and assume that an even number of nodes are initially in state a . The goal is here for the protocol to convert all a s to b s. If the scheduler could pick a perfect matching between the a s, then in one step all a s would be converted to b s, and additionally the protocol would never pass through a configuration in which a state is the state of fewer than n nodes. Now, observe that the sequential scheduler can only pick a single pair of nodes in each step, so in the very first step it yields a configuration in which state b is the state of only 2 nodes. Of course, there are turnarounds to this, for example by taking into account only equal-interaction configurations, consisting of the states of the processes after all processes have participated in an equal number of interactions, still we shall follow an alternative approach that simplifies the arguments and the analysis.

In particular, we will consider schedulers that can be maximally parallel. Such a scheduler, selects in every step a matching (of any possible size) of the complete interaction graph, so, in one extreme, it is still allowed to select only one interaction but, in the other extreme, it may also select a perfect matching in a single step. Observe that this scheduler is different both from the sequential scheduler traditionally used in the area of population protocols and from the fully parallel scheduler which assumes that $\Theta(n)$ interactions occur in parallel in every step. Actually, several recent papers assume a fully parallel scheduler implicitly, by defining the model in terms of the sequential scheduler and then performing their analysis in terms of parallel time, defined as the sequential time divided by n .

Finally, in order to isolate the *inherent* symmetry, i.e., the symmetry that is only due to the protocol, we shall focus on those schedules ³ that achieve as high symmetry as possible for the given protocol. Such schedules may look into the protocol and exploit its structure so that the chosen interactions maximize parallelism. It is crucial to notice that this restriction does by no means affect correctness. Our protocols are still, as usual, required to stabilize to the correct answer in *any* fair execution (and, actually, in this paper against a more generic scheduler than the one traditionally assumed). The above restriction is only a convention for estimating the *inherent* symmetry of a protocol designed to operate in an adversarial setting. On the other hand, one does not expect this *measure of inherent symmetry* to be achieved by the majority of executions. If, instead, one is interested in some *measure of the observed symmetry*, then it would make more sense to study an *expected observed symmetry* under some probabilistic assumption for the scheduler. We leave this as an interesting direction for future research (see Section 5 for more details on this).

For a given initial configuration, we shall estimate the symmetry breaking performed by the protocol not in any possible execution but an execution in which the scheduler tries to maximize the symmetry. In particular, we shall define the symmetry of a protocol on a given initial configuration c_0 as the maximum symmetry achieved over all possible executions on c_0 . So, in order to lower bound by k the symmetry of a protocol on a given c_0 , it will be sufficient to present a schedule in which the protocol stabilizes without ever “electing” fewer than k nodes. On the other hand, to establish an upper bound of h on symmetry, we will have to show that *in every* schedule (on the given c_0) the protocol “elects” at most h nodes. Then we may define the symmetry of the protocol on a set of initial configurations as the minimum of its symmetries over those initial configurations. The symmetry of a protocol (as a whole) shall be defined as a function of some parameter of the initial configuration and is deferred to Section 2.

Observation 1 *The above definition leads to very strong impossibility results, as these upper bounds are also upper bounds on the observed symmetry. In particular, if we establish that the symmetry of a protocol \mathcal{A} is at most h then, it is clear that under any scheduler the symmetry of \mathcal{A} is at most h .*

Section 2 brings together all definitions and basic facts that are used throughout the paper. In Section 3, we give a set of positive results. The main result here is a partial characterization, showing that a wide subclass of semilinear predicates is computed with symmetry $\Theta(N_{min})$, which is asymptotically optimal. Then, in Section 4, we study some basic predicates that seem to require much symmetry breaking. In particular, we study the *majority* and the *parity* predicates. For majority we establish a constant symmetry, while for parity we prove a strong impossibility result, stating that the symmetry of any protocol that stably computes it, is upper bounded by an integer depending only on the size of the protocol (i.e., a constant, compared to the size of the system). The latter

³ By “schedule” we mean an “execution” throughout.

implies that there exist predicates which can *only* be computed by protocols that perform some sort of leader-election (not necessarily a unique leader but at most a constant number of nodes in a distinguished leader role). In Section 5, we give further research directions that are opened by our work. All omitted details and proofs can be found in the full version.

1.2 Further Related Work

In contrast to static systems with unique identifiers (IDs) and dynamic systems, the role of symmetry in *static anonymous systems* has been deeply investigated [Ang80, YK96, Kra97, FMS98]. *Similarity* as a way to compare and contrast different models of concurrent programming has been defined and studied in [JS85]. One (restricted) type of symmetry that has been recently studied in systems with IDs is the existence of *homonyms*, i.e., processes that are initially assigned the same ID [DGFG⁺11]. Moreover, there are several standard models of distributed computing that do not suffer from a necessity to break symmetry globally (e.g., to elect a leader) like Shared Memory with Atomic Snapshots [AAD⁺93, AW04], Quorums [Ske82, MRWW01], and the LOCAL model [Pel00, Suo13].

Population Protocols were originally motivated by highly dynamic networks of simple sensor nodes that cannot control their mobility. The first papers focused on the computational capabilities of the model which have now been almost completely characterized. In particular, if the interaction network is complete (as is also the case in the present paper), i.e., one in which every pair of processes may interact, then the computational power of the model is equal to the class of the *semilinear predicates* (and the same holds for several variations) [AAER07]. Interestingly, the generic protocol of [AAD⁺06] that computes all semilinear predicates, elects a unique leader in every execution and the same is true for the construction in [CDS14]. Moreover, according to [AG15], all known generic constructions of semilinear predicates “fundamentally rely on the election of a single initial *leader* node, which coordinates phases of computation”. Semilinearity of population protocols persists up to $o(\log \log n)$ local space but not more than this [CMN⁺11]. If additionally the connections between processes can hold a state from a finite domain, then the computational power dramatically increases to the commutative subclass of **NSPACE**(n^2) [MCS11a]. The formal equivalence of population protocols to *chemical reaction networks* (CRNs), which model chemistry in a *well-mixed solution*, has been recently demonstrated [Dot14]. Moreover, the recently proposed *Network Constructors* extension of population protocols [MS16] is capable of constructing arbitrarily complex stable networks. Czyzowicz *et al.* [CGK⁺15] have recently studied the relation of population protocols to antagonism of species, with dynamics modeled by discrete Lotka-Volterra equations. Finally, in [CCDS14], the authors highlighted the importance of executions that necessarily pass through a “bottleneck” transition (meaning a transition between two states that have only constant counts in the population, which requires $\Omega(n^2)$ expected number of steps to occur), by proving that protocols that avoid such transitions can only compute existence

predicates. To the best of our knowledge, our type of approach, of computing predicates stably without *ever* electing a unique leader, has not been followed before in this area (according to [AG15], “[DH15] proposes a leader-less framework for population computation”, but this should not be confused with what we do in this paper, as it only concerns the achievement of dropping the requirement for a *pre-elected* unique leader that was assumed in all previous results for that problem). For introductory texts to population protocols, the interested reader is encouraged to consult [AR09] and [MCS11b].

2 Preliminaries

A *population protocol* (PP) is a 6-tuple (X, Y, Q, I, O, δ) , where X , Y , and Q are all finite sets and X is the *input alphabet*, Y is the *output alphabet*, Q is the set of *states*, $I: X \rightarrow Q$ is the *input function*, $O: Q \rightarrow Y$ is the *output function*, and $\delta: Q \times Q \rightarrow Q \times Q$ is the *transition function*.

If $\delta(a, b) = (a', b')$, we call $(a, b) \rightarrow (a', b')$ a *transition*. A transition $(a, b) \rightarrow (a', b')$ is called *effective* if $x \neq x'$ for at least one $x \in \{a, b\}$ and *ineffective* otherwise. When we present the transition function of a protocol we only present the effective transitions. The system consists of a population V of n distributed *processes* (also called *nodes*). In the generic case, there is an underlying *interaction graph* $G = (V, E)$ specifying the permissible interactions between the nodes. Interactions in this model are always pairwise. In this work, G is a *complete directed interaction graph*.

Let Q be the set of states of a population protocol \mathcal{A} . A configuration c of \mathcal{A} on n nodes is an element of $\mathbb{N}_{\geq 0}^{|Q|}$, such that, for all $q \in Q$, $c[q]$ is equal to the number of nodes that are in state q in configuration c and it holds that $\sum_{q \in Q} c[q] = n$. For example, if $Q = \{q_0, q_1, q_2, q_3\}$ and $c = (7, 12, 52, 0)$, then, in c , 7 nodes of the $7 + 12 + 52 + 0 = 71$ in total, are in state q_0 , 12 nodes in state q_1 , and 52 nodes in state q_2 .

Execution of the protocol proceeds in discrete steps and it is determined by an *adversary scheduler* who is allowed to be *parallel*, meaning that, in every step, it may select one or more pairwise interactions (up to a maximum matching) to occur at the same time. This is an important difference from classical population protocols where the scheduler could only select a single interaction per step. More formally, in every step, a non-empty matching $(u_1, v_1), (u_2, v_2), \dots, (u_k, v_k)$ from E is selected by the scheduler and, for all $1 \leq i \leq k$, the nodes u_i, v_i interact with each other and update their states according to the transition function δ . A *fairness condition* is imposed on the adversary to ensure the protocol makes progress. An infinite execution is *fair* if for every pair of configurations c and c' such that $c \rightarrow c'$ (i.e., c can go in one step to c'), if c occurs infinitely often in the execution then so does c' .

In population protocols, we are typically interested in computing predicates on the inputs, e.g., $N_a \geq 5$, being true whenever there are at least 5 *as* in the

input.⁴ Moreover, computations are *stabilizing* and not terminating, meaning that it suffices for the nodes to eventually converge to the correct output. We say that a protocol *stably computes* a predicate if, on any population size, any input assignment, and any fair execution on these, all nodes eventually stabilize their outputs to the value of the predicate on that input assignment.

We define the *symmetry* $s(c)$ of a configuration c as the *minimum multiplicity of a state that is present in c* (unless otherwise stated, in what follows by “symmetry” we shall always mean “inherent symmetry”). That is, $s(c) = \min_{q \in Q : c[q] \geq 1} \{c[q]\}$. For example, if $c = (0, 4, 12, 0, 52)$ then $s(c) = 4$, if $c = (1, \dots)$ then $s(c) = 1$, which is the minimum possible value for symmetry, and if $c = (n, 0, 0, \dots, 0)$ then $s(c) = n$ which is the maximum possible value for symmetry. So, the range of the symmetry of a configuration is $\{1, 2, \dots, n\}$.

Let $\mathcal{C}_0(\mathcal{A})$ be the set of all *initial configurations* for a given protocol \mathcal{A} . Given an initial configuration $c_0 \in \mathcal{C}_0(\mathcal{A})$, denote by $\Gamma(c_0)$ the set of all fair executions of \mathcal{A} that begin from c_0 , each execution being truncated to its prefix *up to stability*.⁵

Given any initial configuration c_0 and any execution $\alpha \in \Gamma(c_0)$, define the *symmetry breaking of \mathcal{A} on α* as the difference between the symmetry of the initial configuration of α and the minimum symmetry of a configuration of α , that is, the *maximum drop in symmetry* during the execution. Formally, $b(\mathcal{A}, \alpha) = s(c_0) - \min_{c \in \alpha} \{s(c)\}$. Also define the *symmetry of \mathcal{A} on α* as $s(\mathcal{A}, \alpha) = \min_{c \in \alpha} \{s(c)\}$. Of course, it holds that $s(\mathcal{A}, \alpha) = s(c_0) - b(\mathcal{A}, \alpha)$. Moreover, observe that, for all $\alpha \in \Gamma(c_0)$, $0 \leq b(\mathcal{A}, \alpha) \leq s(c_0) - 1$ and $1 \leq s(\mathcal{A}, \alpha) \leq s(c_0)$. In several cases we shall denote $s(c_0)$ by N_{min} .

The *symmetry breaking of a protocol \mathcal{A} on an initial configuration c_0* can now be defined as $b(\mathcal{A}, c_0) = \min_{\alpha \in \Gamma(c_0)} \{b(\mathcal{A}, \alpha)\}$ and:

Definition 1. We define the symmetry of \mathcal{A} on c_0 as $s(\mathcal{A}, c_0) = \max_{\alpha \in \Gamma(c_0)} \{s(\mathcal{A}, \alpha)\}$.

Remark 1. To estimate the *inherent* symmetry with which a protocol computes a predicate on a c_0 , we execute the protocol against an *imaginary* scheduler who is a *symmetry maximizer*.

Now, given the set $\mathcal{C}(N_{min})$ of all initial configurations c_0 such that $s(c_0) = N_{min}$, we define the *symmetry breaking of a protocol \mathcal{A} on $\mathcal{C}(N_{min})$* as $b(\mathcal{A}, N_{min}) = \max_{c_0 \in \mathcal{C}(N_{min})} \{b(\mathcal{A}, c_0)\}$ and:

Definition 2. We define the symmetry of \mathcal{A} on $\mathcal{C}(N_{min})$ as $s(\mathcal{A}, N_{min}) = \min_{c_0 \in \mathcal{C}(N_{min})} \{s(\mathcal{A}, c_0)\}$.

⁴ We shall use throughout the paper N_i to denote the number of nodes with input/state i .

⁵ In this work, we only require protocols to preserve their symmetry *up to stability*. This means that a protocol is allowed to break symmetry arbitrarily after stability, e.g., even elect a unique leader, without having to pay for it. We leave as an interesting open problem the comparison of this convention to the apparently harder requirement of maintaining symmetry forever.

Observe again that $s(\mathcal{A}, N_{min}) = N_{min} - b(\mathcal{A}, N_{min})$ and that $0 \leq b(\mathcal{A}, N_{min}) \leq N_{min} - 1$ and $1 \leq s(\mathcal{A}, N_{min}) \leq N_{min}$.

This means that, in order to establish that a protocol \mathcal{A} is at least $g(N_{min})$ symmetric asymptotically (e.g., for $g(N_{min}) = \Theta(\log N_{min})$), we have to show that for every sufficiently large N_{min} , the symmetry breaking of \mathcal{A} on $\mathcal{C}(N_{min})$ is at most $N_{min} - g(N_{min})$, that is, to show that for all initial configurations $c_0 \in \mathcal{C}(N_{min})$ there exists an execution on c_0 that drops the initial symmetry by at most $N_{min} - g(N_{min})$, e.g., by at most $N_{min} - \log N_{min}$ for $g(N_{min}) = \log N_{min}$, or that does not break symmetry at all in case $g(N_{min}) = N_{min}$. On the other hand, to establish that the symmetry is at most $g(N_{min})$, e.g., at most 1 which is the minimum possible value, one has to show a symmetry breaking of at least $N_{min} - g(N_{min})$ on infinitely many N_{min} s.

3 Predicates of High Symmetry

In this section, we try to identify predicates that can be stably computed with much symmetry. We first give an indicative example, then we generalize to arrive at a partial characterization of the predicates that can be computed with maximum symmetry, and, finally, we highlight the role of output-stable states in symmetric computations.

3.1 An Example: Count-to- x

Protocol *Count-to- x* : $X = \{0, 1\}$, $Q = \{q_0, q_1, q_2, \dots, q_x\}$, $I(\sigma) = q_\sigma$, for all $\sigma \in X$, $O(q_x) = 1$ and $O(q) = 0$, for all $q \in Q \setminus \{q_x\}$, and $\delta: (q_i, q_j) \rightarrow (q_{i+j}, q_0)$, if $i + j < x$, $(q_i, q_j) \rightarrow (q_x, q_x)$, otherwise.

Proposition 1. *The symmetry of Protocol Count-to- x , for any $x = O(1)$, is at least $(2/3)\lfloor N_{min}/x \rfloor - (x-1)/3$, when $x \geq 2$, and N_{min} , when $x = 1$; i.e., it is $\Theta(N_{min})$ for any $x = O(1)$.*

Proof. The scheduler⁶ partitions the q_1 s, let them be $N_1(0)$ initially and denoted just N_1 in the sequel, into $\lfloor N_1/x \rfloor$ groups of x q_1 s each, possibly leaving an incomplete group of $r \leq x-1$ q_1 s residue. Then, in each complete group, it performs a sequential gathering of $x-3$ other q_1 s to one of the nodes, which will go through the states q_1, q_2, \dots, q_{x-1} . The same gathering is performed in parallel to all groups, so every state that exists in one group will also exist in every other group, thus, its cardinality never drops below $\lfloor N_1/x \rfloor$. In the end, at step t , there are many q_0 s, $N_{x-1}(t) = \lfloor N_1/x \rfloor$, and $N_1(t) = \lfloor N_1/x \rfloor + r$, where $0 \leq r \leq x-1$ is the residue of q_1 s. That is, in all configurations so far, the symmetry has not dropped below $\lfloor N_1/x \rfloor$.

Now, we cannot pick, as a symmetry maximizing choice of the scheduler, a perfect bipartite matching between the q_1 s and the q_{x-1} s converting them all

⁶ Always meaning the *imaginary symmetry-maximizing scheduler* when lower-bounding the symmetry.

to the alarm state q_x , because this could possibly leave the symmetry-breaking residue of q_1 s. What we can do instead, is to match in one step as many as we can so that, after the corresponding transitions, $N_x(t') \geq N_1(t')$ is satisfied. In particular, if we match y of the (q_1, q_{x-1}) pairs we will obtain $N_x(t') = 2y$, $N_{x-1}(t') = \lfloor N_1/x \rfloor - y$, and $N_1(t') = \lfloor N_1/x \rfloor - y + r$ and what we want is

$$2y \geq \lfloor N_1/x \rfloor - y + r \Rightarrow 3y \geq \lfloor N_1/x \rfloor + r \Rightarrow y \geq \frac{\lfloor N_1/x \rfloor + r}{3},$$

which means that if we match approximately $1/3$ of the (q_1, q_{x-1}) pairs then we will have as many q_x as we need in order to eliminate all q_1 s in one step and all remaining q_{x-1} s in another step.

The minimum symmetry in the whole course of this schedule is

$$\begin{aligned} N_{x-1}(t') &= \lfloor N_1/x \rfloor - y = \lfloor N_1/x \rfloor - \frac{\lfloor N_1/x \rfloor + r}{3} \\ &= \frac{2}{3} \lfloor N_1/x \rfloor - \frac{r}{3} \geq \frac{2}{3} \lfloor N_1/x \rfloor - \frac{x-1}{3}. \end{aligned}$$

So, we have shown that if there are no q_0 s in the initial configuration, then the symmetry breaking of the protocol on the schedule defined above is at most $N_{min} - ((2/3)\lfloor N_1/x \rfloor - (x-1)/3) = N_{min} - ((2/3)\lfloor N_{min}/x \rfloor - (x-1)/3)$. Next, we consider the case in which there are some q_0 s in the initial configuration. Observe that in this protocol the q_0 s can only increase, so their minimum cardinality is precisely their initial cardinality N_0 . Consequently, in case $N_0 \geq 1$ and $N_1 \geq 1$, and if $N_{min} = \min\{N_0, N_1\}$, the symmetry breaking of the schedule defined above is $N_{min} - \min\{N_0, N_{x-1}(t')\}$. If, for some initial configuration, $N_0 \geq N_{x-1}(t')$ then the symmetry breaking is $N_{min} - N_{x-1}(t') \leq N_{min} - ((2/3)\lfloor N_1/x \rfloor - (x-1)/3)$. This gives again $N_{min} - ((2/3)\lfloor N_{min}/x \rfloor - (x-1)/3)$, when $N_1 \leq N_0$, and less than $N_{min} - ((2/3)\lfloor N_{min}/x \rfloor - (x-1)/3)$, when $N_1 > N_0 = N_{min}$. If instead, $N_0 < N_{x-1}(t') < N_1$, then, in this case, the symmetry breaking is $N_{min} - \min\{N_0, N_{x-1}(t')\} = N_0 - N_0 = 0$. Finally, if $N_0 = n$, then the symmetry breaking is 0. We conclude that for every initial configuration, the symmetry breaking of the above schedule is at most $N_{min} - N_{x-1}(t') \leq N_{min} - ((2/3)\lfloor N_{min}/x \rfloor - (x-1)/3)$, for all $x \geq 2$, and 0, for $x = 1$. Therefore, the symmetry of the *Count-to- x* protocol is at least $(2/3)\lfloor N_{min}/x \rfloor + (x-1)/3 = \Theta(N_{min})$, for $x \geq 2$, and N_{min} , for $x = 1$. \square

3.2 A General Positive Result

Theorem 1. *Any predicate of the form $\sum_{i \in [k]} a_i N_i \geq c$, for integer constants $k \geq 1$, $a_i \geq 1$, and $c \geq 0$, can be computed with symmetry more than $\lfloor N_{min}/(c/\sum_{j \in L} a_j + 2) \rfloor - 2 = \Theta(N_{min})$.*

Proof. We begin by giving a parameterized protocol (Protocol 1) that stably computes any such predicate, and then we shall prove that the symmetry of this protocol is the desired one.

Protocol 1 *Positive-Linear-Combination*

$Q = \{q_0, q_1, q_2, \dots, q_c\}$
 $I(\sigma_i) = q_{a_i}$, for all $\sigma_i \in X$
 $O(q_c) = 1$ and $O(q) = 0$, for all $q \in Q \setminus \{q_c\}$
 δ :

$$\begin{aligned}
 (q_i, q_j) &\rightarrow (q_{i+j}, q_0), \text{ if } i + j < c \\
 &\rightarrow (q_c, q_c), \text{ otherwise}
 \end{aligned}$$

Take now any initial configuration C_0 on n nodes and let $L \subseteq [k]$ be the set of indices of the initial states that are present in C_0 . Let also q_{min} be the state with minimum cardinality, N_{min} , in C_0 . Construct $\lfloor N_{min}/x \rfloor$ groups, by adding to each group $x = \lceil c / \sum_{j \in L} a_j \rceil$ copies of each initial state. Observe that each group has total sum $\sum_{j \in L} a_j x = x \sum_{j \in L} a_j = \lceil c / \sum_{j \in L} a_j \rceil (\sum_{j \in L} a_j) \geq c$. Moreover, state q_{min} has a residue r_{min} of at most x and every other state q_i has a residue $r_i \geq r_{min}$. Finally, keep $y = \lceil (N_{min} + r_{min}) / (x + 1) \rceil - 1$ from those groups and drop the other $\lfloor N_{min}/x \rfloor - y$ groups making their nodes part of the residue, which results in new residue values $r'_j = x(\lfloor N_{min}/x \rfloor - y) + r_j$, for all $j \in L$. It is not hard to show that $y \leq r'_j$, for all $j \in L$.

We now present a schedule that achieves the desired symmetry. The schedule consists of two phases, the *gathering* phase and the *dissemination* phase. In the dissemination phase, the schedule picks a node of the same state from every group and starts aggregating to that node the sum of its group sequentially, performing the same in parallel in all groups. It does this until the alarm state q_c first appears. When this occurs, the dissemination phase begins. In the dissemination phase, the schedule picks one after the other all states that have not yet been converted to q_c . For each such state q_i , it picks a q_c which infects one after the other (sequentially) the q_i s, until $N_c(t) \geq N_i(t)$ is satisfied for the first time. Then, in a single step that matches each q_i to a q_c , it converts all remaining q_i s to q_c .

We now analyze the symmetry breaking of the protocol in this schedule. Clearly, the initial symmetry is N_{min} . As long as a state appears in the groups, its cardinality is at least y , because it must appear in each one of them. When a state q_i first becomes eliminated from the groups, its cardinality is equal to its residue r'_i . Thus, so far, the minimum cardinality of a state is

$$\min\{y, \min_{j \in L} r'_j\} = y = \left\lceil \frac{N_{min} + r_{min}}{x + 1} \right\rceil - 1 > \left\lfloor \frac{N_{min}}{c / \sum_{j \in L} a_j + 2} \right\rfloor - 2.$$

It follows that the maximum symmetry breaking so far is less than $N_{min} - \left\lfloor \frac{N_{min}}{c / \sum_{j \in L} a_j + 2} \right\rfloor + 2$.

Finally, we must also take into account the dissemination phase. In this phase, the q_c s are $2y$ initially and can only increase, by infecting other states, until they become n and the cardinalities of all other states decrease until they all become

0. Take any state $q_i \neq q_c$ with cardinality $N_i(t)$ when the dissemination phase begins. What the schedule does is to decrement $N_i(t)$, until $N_c(t') \geq N_i(t')$ is first satisfied, and then to eliminate all occurrences of q_i in one step. Due to the fact that N_i is decremented by one in each step resulting in a corresponding increase by one of N_c , when $N_c(t') \geq N_i(t')$ is first satisfied, it holds that $N_i(t') \geq N_c(t') - 1 \geq N_c(t) - 1 \geq 2y - 1 \geq y$ for all $y \geq 1$, which implies that the lower bound of y on the minimum cardinality, established for the gathering phase, is not violated during the dissemination phase.

We conclude that the symmetry of the protocol in the above schedule is more than $\lfloor N_{\min}/(c/\sum_{j \in L} a_j + 2) \rfloor - 2$. \square

3.3 Output-stable States

Informally, a state $q \in Q$ is called *output-stable* if its appearance in an execution guarantees that the output value $O(q)$ must be the output value of the execution. More formally, if q is output-stable and C is a configuration containing q , then the set of outputs of C' must contain $O(q)$, for all C' such that $C \rightsquigarrow C'$, where ' \rightsquigarrow ' means *reaches in one or more steps*. Moreover, if all executions under consideration stabilize to an agreement, meaning that eventually all nodes stabilize to the same output, then the above implies that if an execution ever reaches a configuration containing q then the output of that execution is necessarily $O(q)$.

A state q is called *reachable* if there is an initial configuration C_0 and an execution on C_0 that can produce q . We can also define reachability just in terms of the protocol, under the assumption that if $Q_0 \subseteq Q$ is the set of initial states, then any possible combination of cardinalities of states from Q_0 can be part of an initial configuration. A *production tree* for a state $q \in Q$, is a directed binary in-tree with its nodes labeled from Q such that its root has label q , if a is the label of an internal node (the root inclusive) and b, c are the labels of its children, then the protocol has a rule of the form $\{b, c\} \rightarrow \{a, \cdot\}$ (that is, a rule producing a by an interaction between a b and a c in any direction)⁷, and any leaf is labeled from Q_0 . Observe now that if a path from a leaf to the root repeats a state a , then we can always replace the subtree of the highest appearance of a by the subtree of the lowest appearance of a on the path and still have a production tree for q . This implies that if q has a production tree, then q also has a production tree of depth at most $|Q|$, that is, a production tree having at most $2^{|Q|-1}$ leaves, which is a constant number, when compared to the population size n , that only depends on the protocol. Now, we can call a state q *reachable* (by a protocol \mathcal{A}) if there is a production tree for it. These are summarized in the following proposition.

Proposition 2. *Let \mathcal{A} be a protocol, C_0 be any (sufficiently large) initial configuration of \mathcal{A} , and $q \in Q$ any state that is reachable from C_0 . Then there is an initial configuration C'_0 which is a sub-configuration of C_0 of size $n' \leq 2^{|Q|-1}$ such that q is reachable from C'_0 .*

⁷ Whenever we use an unordered pair in a rule, like $\{b, c\}$, we mean that the property under consideration concerns both (b, c) and (c, b) .

Proposition 2 is crucial for proving negative results, and will be invoked in Section 4.

Proposition 3. *Let p be a predicate. There is no protocol that stably computes p (all nodes eventually agreeing on the output in every fair execution), having both a reachable output-stable state with output 0 and a reachable output-stable state with output 1.*

An output-stable state q is called *disseminating* if $\{x, q\} \rightarrow (q, q)$, for all $x \in Q$.

Proposition 4. *Let \mathcal{A} be a protocol with at least one reachable output-stable state, that stably computes a predicate p and let $Q_s \subseteq Q$ be the set of reachable output-stable states of \mathcal{A} . Then there is a protocol \mathcal{A}' with a reachable disseminating state that stably computes p .*

Theorem 2. *Let \mathcal{A} be a protocol with a reachable disseminating state q and let C_0^d be the subset of its initial configurations that may produce q . Then the symmetry of \mathcal{A} on C_0^d is $\Theta(N_{\min})$.*

Theorem 2 emphasizes the fact that disseminating states can be exploited for maximum symmetry. We have omitted its proof, because it is similar to the proofs of Proposition 1 and Theorem 1. This lower bound on symmetry immediately applies to single-signed linear combinations (where passing a threshold can safely result in the appearance of a disseminating state, because there are no opposite-signed numbers to inverse the process), thus, it can be used as an alternative way of arriving at Theorem 1. On the other hand, the next proposition shows that this lower bound does not apply to linear combinations containing mixed signs, because protocols for them cannot have output-stable states.

Proposition 5. *Let p be a predicate of the form $\sum_{i \in [k]} a_i N_i \geq c$, for integer constants $k \geq 1$, a_i , and $c \geq 0$ such that at least two a_i s have opposite signs. Then there is no protocol, having a reachable output-stable state, that stably computes p .*

4 Harder Predicates

In this section, we study the symmetry of predicates that, in contrast to single-signed linear combinations, do not allow for output-stable states. In particular, we focus on linear combinations containing mixed signs, like the *majority* predicate, and also on modulo predicates like the *parity* predicate. Recall that these predicates are not captured by the lower bound on symmetry of Theorem 2.

4.1 Bounds for Mixed Coefficients

We begin with a proposition stating that the majority predicate (also can be generalized to any predicate with mixed signs) can be computed with symmetry that depends on the difference of the state-cardinalities in the initial configuration.

Proposition 6. *The majority predicate $N_a - N_b > 0$ can be computed with symmetry $\min\{N_{\min}, |N_a - N_b|\}$, where $N_{\min} = \min\{N_a, N_b\}$.*

Remark 2. A result similar to Proposition 6 can be proved for any predicate $\sum_{i \in [k]} a_i N_i - \sum_{j \in [h]} b_j N'_j > c$, for integer constants $k, h, a_i, b_j \geq 1$ and $c \geq 0$.

Still, as we prove in the following theorem, it is possible to do better in the worst case, and achieve any desired constant symmetry.

Theorem 3. *For every constant $k \geq 1$, the majority predicate $N_a - N_b > 0$ can be computed with symmetry k .*

4.2 Predicates that Cannot be Computed with High Symmetry

We now prove a strong impossibility result, establishing that there are predicates that cannot be stably computed with much symmetry. The result concerns the *parity* predicate, defined as $n \bmod 2 = 1$. In particular, all nodes obtain the same input, e.g., 1, and, thus, all begin from the same state, e.g., q_1 . So, in this case, $N_{\min} = n$ in every initial configuration, and we can here estimate symmetry as a function of n . The parity predicate is true iff the number of nodes is odd. So, whenever n is odd, we want all nodes to eventually stabilize their outputs to 1 and, whenever it is even, to 0. If symmetry is not a constraint, then there is a simple protocol that solves the problem [AAD⁺06]. Unfortunately, not only that particular strategy, but any possible strategy for the problem, cannot achieve symmetry more than a constant that depends on the size of the protocol, as we shall now prove.

Theorem 4. *Let \mathcal{A} be a protocol with set of states Q , that solves the parity predicate. Then the symmetry of \mathcal{A} is less than $2^{|Q|-1}$.*

Proof. For the sake of contradiction, assume \mathcal{A} solves parity with symmetry $f(n) \geq 2^{|Q|-1}$. Take any initial configuration C_n for any sufficiently large odd n (e.g., $n \geq f(n)$ or $n \geq |Q| \cdot f(n)$, or even larger if required by the protocol). By definition of symmetry, there is an execution α on C_n that reaches stability without ever dropping the minimum cardinality of an existing state below $f(n)$. Call C_{stable} the first output-stable configuration of α . As n is odd, C_{stable} must satisfy that all nodes are in states giving output 1 and that no execution on C_{stable} can produce a state with output 0. Moreover, due to the facts that \mathcal{A} has symmetry $f(n)$ and that α is an execution that achieves this symmetry, it must hold that every $q \in Q$ that appears in C_{stable} has multiplicity $C_{\text{stable}}[q] \geq f(n)$.

Consider now the initial configuration C_{2n} , i.e., the unique initial configuration on $2n$ nodes. Observe that now the number of nodes is even, thus, the parity predicate evaluates to false and any fair execution of \mathcal{A} must stabilize to output 0. Partition C_{2n} into two equal parts, each of size n . Observe that each of the two parts is equal to C_n . Consider now the following possible finite prefix β of a fair execution on C_{2n} . The scheduler simulates in each of the two parts the previous execution α up to the point that it reaches the configuration C_{stable} .

So, the prefix β takes C_{2n} to a configuration denoted by $2C_{stable}$ and consisting precisely of two copies of C_{stable} . Observe that $2C_{stable}$ and C_{stable} consist of the same states with the only difference being that their multiplicity in $2C_{stable}$ is twice their multiplicity in C_{stable} . A crucial difference between C_{stable} and $2C_{stable}$ is that the former is output-stable while the latter is not. In particular, any fair execution of \mathcal{A} on $2C_{stable}$ must produce a state q_0 with output 0. But, by Proposition 2, q_0 must also be reachable from a sub-configuration C_{small} of $2C_{stable}$ of size at most $2^{|Q|-1}$. So, there is an execution γ restricted on C_{small} that produces q_0 .

Observe now that C_{small} is also a sub-configuration of C_{stable} . The reason is that (i) every state in C_{small} is also a state that exists in $2C_{stable}$ and, thus, also a state that exists in C_{stable} and (ii) the multiplicity of every state in C_{small} is restricted by the size of C_{small} , which is at most $2^{|Q|-1}$, and every state in C_{stable} has multiplicity at least $f(n) \geq 2^{|Q|-1}$, that is, C_{stable} has sufficient capacity for every state in C_{small} . But this implies that if γ is executed on the sub-configuration of C_{stable} corresponding to C_{small} , then it must produce q_0 , which contradicts the fact that C_{stable} is output-stable with output 1. Therefore, we conclude that \mathcal{A} cannot have symmetry at least $f(n) \geq 2^{|Q|-1}$. \square

Remark 3. Theorem 4 constrains the symmetry of any correct protocol for parity to be upper bounded by a constant that depends on the size of the protocol. Still, it does not exclude the possibility that parity is solvable with symmetry k , for any constant $k \geq 1$. The reason is that, for any constant $k \geq 1$, there might be a protocol with $|Q| > k$ that solves parity and achieves symmetry k , because $k < 2^{|Q|-1}$, which is the upper bound on symmetry proved by the theorem. On the other hand, the $2^{|Q|-1}$ upper bound of Theorem 4 excludes any protocol that would solve parity with symmetry depending on N_{min} .

5 Further Research

In this work, we managed to obtain a first partial characterization of the predicates with symmetry $\Theta(N_{min})$ and to exhibit a predicate (parity) that resists any non-constant symmetry. The obvious next goal is to arrive at an exact characterization of the allowable symmetry of all semilinear predicates.

Some preliminary results of ours, indicate that constant symmetry for parity can be achieved if the initial configuration has a sufficient number of *auxiliary nodes* in a distinct state q_0 . It seems interesting to study how is symmetry affected by auxiliary nodes and whether they can be totally avoided.

Another very challenging direction for further research, concerns networked systems (either static or dynamic) in which the nodes have memory and possibly also unique IDs. Even though the IDs provide an *a priori* maximum symmetry breaking, still, solving a task and avoiding the process of “electing” one of the nodes may be highly non-trivial. But in this case, defining the role of a process as its complete local state is inadequate. There are other plausible ways of defining the role of a process, but which one is best-tailored for such systems is still unclear and needs further investigation.

Finally, recall that in this work we focused on the *inherent* symmetry of a protocol as opposed to its *observed* symmetry. One way to study the observed symmetry would be to consider *random parallel schedulers*, like the one that selects in every step a maximum matching uniformly at random from all such matchings. Then we may ask “*What is the average symmetry achieved by a protocol under such a scheduler?*”. In some preliminary experimental results of ours, the expected observed symmetry of the *Count-to-5* protocol (i) if counted until the alert state q_5 becomes an absolute majority in the population, seems to grow faster than \sqrt{n} and (ii) if counted up to stability, seems to grow as fast as $\log n$ (see the full paper for more details).

Acknowledgements. We would like to thank Dimitrios Amaxilatis for setting up and running experiments for the evaluation of the observed symmetry.

References

- [AAD⁺93] Y. Afek, H. Attiya, D. Dolev, E. Gafni, M. Merritt, and N. Shavit. Atomic snapshots of shared memory. *Journal of the ACM (JACM)*, 40[4]:873–890, 1993.
- [AAD⁺06] D. Angluin, J. Aspnes, Z. Diamadi, M. J. Fischer, and R. Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18[4]:235–253, March 2006.
- [AAER07] D. Angluin, J. Aspnes, D. Eisenstat, and E. Ruppert. The computational power of population protocols. *Distributed Computing*, 20[4]:279–304, November 2007.
- [AG15] D. Alistarh and R. 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, 2015.
- [Ang80] D. Angluin. Local and global properties in networks of processors. In *Proceedings of the 12th annual ACM symposium on Theory of computing (STOC)*, pages 82–93. ACM, 1980.
- [AR09] J. Aspnes and E. Ruppert. An introduction to population protocols. In B. Garbinato, H. Miranda, and L. Rodrigues, editors, *Middleware for Network Eccentric and Mobile Applications*, pages 97–120. 2009.
- [AW04] H. Attiya and J. Welch. *Distributed computing: fundamentals, simulations, and advanced topics*, volume 19. Wiley-Interscience, 2004.
- [CCDS14] H.-L. Chen, R. Cummings, D. Doty, and D. Soloveichik. Speed faults in computation by chemical reaction networks. In *Proceedings of the 28th International Symposium on Distributed Computing (DISC)*, volume 8784 of *Lecture Notes in Computer Science*, pages 16–30, 2014.
- [CDS14] H.-L. Chen, D. Doty, and D. Soloveichik. Deterministic function computation with chemical reaction networks. *Natural computing*, 13[4]:517–534, 2014.
- [CGK⁺15] J. Czyzowicz, L. Gasieniec, A. Kosowski, E. Kranakis, P. G. Spirakis, and P. Uznański. On convergence and threshold properties of discrete lotka-volterra population protocols. In *42nd International Colloquium on Automata, Languages and Programming (ICALP)*, pages 393–405. Springer, 2015.

- [CMN⁺11] I. Chatzigiannakis, O. Michail, S. Nikolaou, A. Pavlogiannis, and P. G. Spirakis. Passively mobile communicating machines that use restricted space. *Theoretical Computer Science*, 412[46]:6469–6483, October 2011.
- [DGFG⁺11] C. Delporte-Gallet, H. Fauconnier, R. Guerraoui, A.-M. Kermarrec, E. Ruppert, and H. Tran-The. Byzantine agreement with homonyms. In *Proceedings of the 30th annual ACM SIGACT-SIGOPS symposium on Principles of distributed computing (PODC)*, pages 21–30. ACM, 2011.
- [DH15] D. Doty and M. Hajiaghayi. Leaderless deterministic chemical reaction networks. *Natural Computing*, 14[2]:213–223, 2015.
- [Dot14] D. Doty. Timing in chemical reaction networks. In *Proceedings of the 25th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 772–784, 2014.
- [FMS98] P. Flocchini, B. Mans, and N. Santoro. Sense of direction: Definitions, properties, and classes. *Networks*, 32[3]:165–180, 1998.
- [FSW14] K.-T. Förster, J. Seidel, and R. Wattenhofer. Deterministic leader election in multi-hop beeping networks. In *Proceedings of the 28th International Symposium on Distributed Computing (DISC)*, volume 8784 of *Lecture Notes in Computer Science*, pages 212–226. Springer, 2014.
- [JS85] R. E. Johnson and F. B. Schneider. Symmetry and similarity in distributed systems. In *Proceedings of the 4th annual ACM symposium on Principles of distributed computing (PODC)*, pages 13–22. ACM, 1985.
- [KLO10] F. Kuhn, N. Lynch, and R. Oshman. Distributed computation in dynamic networks. In *Proceedings of the 42nd ACM symposium on Theory of computing (STOC)*, pages 513–522. ACM, 2010.
- [Kra97] E. Kranakis. Symmetry and computability in anonymous networks: a brief survey. In *Proceedings of the 3rd International Conference on Structural Information and Communication Complexity*, pages 1–16, 1997.
- [Lyn96] N. A. Lynch. *Distributed Algorithms*. Morgan Kaufmann; 1st ed., 1996.
- [MCS11a] O. Michail, I. Chatzigiannakis, and P. G. Spirakis. Mediated population protocols. *Theoretical Computer Science*, 412[22]:2434–2450, May 2011.
- [MCS11b] O. Michail, I. Chatzigiannakis, and P. G. Spirakis. *New Models for Population Protocols*. N. A. Lynch (Ed), Synthesis Lectures on Distributed Computing Theory. Morgan & Claypool, 2011.
- [MRWW01] D. Malkhi, M. K. Reiter, A. Wool, and R. N. Wright. Probabilistic quorum systems. *Information and Computation*, 170[2]:184–206, 2001.
- [MS16] O. Michail and P. G. Spirakis. Simple and efficient local codes for distributed stable network construction. *Distributed Computing*, 29[3]:207–237, 2016.
- [Pel00] D. Peleg. *Distributed computing: a locality-sensitive approach*. SIAM monographs on discrete mathematics and applications. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.
- [Ske82] D. Skeen. A quorum-based commit protocol. Technical report, Cornell University, 1982.
- [Suo13] J. Suomela. Survey of local algorithms. *ACM Computing Surveys (CSUR)*, 45[2]:24, 2013.
- [YK96] M. Yamashita and T. Kameda. Computing on anonymous networks. i. characterizing the solvable cases. *Parallel and Distributed Systems, IEEE Transactions on*, 7[1]:69–89, 1996.