Consensus-based Page's test in sensor networks

Paolo Braca\textsuperscript{a}, Stefano Marano\textsuperscript{a,}\textsuperscript{*}, Vincenzo Matta\textsuperscript{a}, Peter Willett\textsuperscript{b,1}

\textsuperscript{a} DIIIE, University of Salerno, via Ponte don Melillo, 84084 Fisciano (SA), Italy
\textsuperscript{b} ECE Department, U-157, University of Connecticut, Storrs, CT 06269, USA

\begin{abstract}
Page's test is a well-known statistical technique to approach quickest detection problems, namely the detection of an abrupt change in the statistical distribution of a certain monitored phenomenon. Running consensus is a recently proposed signal processing procedure aimed at reaching agreement among the nodes of a fully flat network, and its peculiar feature is the simultaneity of two stages: that of acquiring new measurements by the sensors, and that of data fusion involving inter-sensor communications.

In this paper we study a quickest detector based on the running consensus scheme, and compare it to a bank of independent Page's tests. Exploiting insights from previous studies, we propose closed-form analytical approximations of the performances of these detection schemes and address a comparison in terms of relative efficiencies. The approximated performance figures are then checked by simulation to validate the analysis and to investigate non-asymptotic scenarios.
\end{abstract}

\section{Introduction}

Quickest detection has important applications in all those areas where an abrupt change in the state of nature is to be reported as soon as possible, and Page's test is a well-known signal processing procedure for quickest change detection based upon the so-called CUSUM (cumulative sum) statistic, see, e.g., [1--3]. Page's test was first proposed in [1] in a centralized scenario where all the observations are available at a given site. However, much recent interest in the detection community has focused on decentralized cases where the state of the nature is monitored by a network made of a large number of small sensors deployed over a surveyed area, sensors that collect data to be delivered to a central unit, usually called a fusion center. In these inference systems, the problem of quickest detection is to be faced with the communication issue that rules the amount of information to be delivered to the fusion center.

The application domains of these systems include the fields of the environment monitoring, quality control, failure detection of mechanical/electronic devices, medical instrumentation, homeland security, etc. In all these applications the network is designed to detect as soon as possible a certain event, in order to take adequate actions in a very timely manner. For example, in a typical environment monitoring application, the network monitors a number of early indicators of an earthquake: as soon as an anomalous change of certain physical/chemical characteristics is detected, an alert message is produced and the network manager can thus activate the emergency protocols. More details and discussions can be found in [2].

Nowadays, the literature on decentralized change detection is abundant and well-assessed, and the original issue of detecting an abrupt change has been extended in several interesting directions: the scenario where the two statistical distributions of the data before and after the change are only partially known, the case where the
change is temporary and the length of the transient is of interest, the case of dependent data, the case of optimizing the system under different criteria (information theoretic, for instance) and different constraints (e.g., energy), scenarios with quantized data, the presence of different kinds of feedback channels (from the fusion center towards the sensors), and so forth. In particular, in [4] the authors compare a binary decentralized system with Page’s test, showing that the two systems exhibit equivalent performance. In [5] the authors first quantize the observations and then deliver the quantized information to the fusion center for the final decision. In a similar scenario, in [6] the utility of a feedback channel is investigated, while the authors of [7,8] study the asymptotic optimality of the decentralized system. In [9] the performance of the CUSUM test for a given quantization scheme is addressed.

In inference applications, very recently, there has also been a growing interest in fully decentralized (also referred to as fully flat) architectures. In these systems, different from the above described scenarios, there is no fusion center that collects data and its role is replaced by the sensors themselves, equipped with further functionalities such as a certain amount of signal processing capability and communication connections allowing the exchange of data between neighboring nodes. The typical mode of operation of a fully flat decentralized system assumes that the sensors, after having terminated the sensing stage in which data are collected, exchange local information that, suitably processed, leads to a final global agreement about a common value such as the arithmetic mean of all the measurements. When the sensor network is engaged in an inference task, this final value that all the nodes asymptotically (with time) share represents the inference. Gossip algorithms (see, e.g., [10]) provide a common paradigm to analyze and design how these fully decentralized schemes reach consensus (synonymous with agreement). Useful surveys on the consensus literature are [11–13]. In [14] the authors use a procedure, called parley due to its similarity to the process of negotiation by a team of human decision makers, to achieve the consensus on the final decision in a hypothesis testing problem. In [15] the convergence and the asymptotic agreements of the agents’ decisions is discussed. In [16] the communication architecture for reaching consensus is investigated, while in [17] the concept of “fast” consensus is introduced. In [18] a distributed Kalman filter based on a consensus procedure is proposed.

Running consensus is a variation of the classical gossip paradigm in which the sensing and the data-exchange stages are interleaved and simultaneous. Namely, in each time slot, new measurements are taken by the nodes, and a gossip stage among the nodes is also implemented. The new challenge, therefore, is to reach consensus even with the time-varying dataset that the ensemble of nodes collects [19–22]. This should be faced with the classical consensus procedure where the data are all collected before than the gossip stage begins, so that the classical consensus algorithm runs on a static dataset.

Page’s test is a well-known statistical procedure to detect an abrupt change in the statistical distribution of observed data. It provides the best performance, in terms of detection delay for a given false alarm rate, under standard conditions of independent and identically distributed data (before and after the change in distribution). While other test can be certainly conceived, there is no doubt that Page’s test is a natural candidate for detecting abrupt changes also in distributed environments, and the main idea behind this manuscript is simply to fuse the running consensus update rule with Page’s testing recursion.

As opposed to its counterpart equipped with a fusion center, quickest change detection by means of a fully decentralized architecture has not received much attention in the literature. In particular, no reference is available that reports on a marriage between Page’s approach and running consensus; therefore, we investigate this issue in the simplest setting, leaving the many generalizations of the change detection problem, as earlier quoted, for the future.

Another decentralized architecture we consider in this paper is a bank of independent Page’s processors, where each node of the system implements a CUSUM test, independent of all other sensors. In this case the only interaction among the sensors is a stopping message that the first node that detects a change broadcasts to all other sensors in the system: to declare the change and to “sleep” (shut down) the system.

Note that the two architectures have very different demands in terms of communication and processing tasks, and each may be appropriate in a different application. In this regard, one should observe that the bank strategy does not use the full information available in the network, but at the same time does not require inter-sensor communication. Conversely, the approach based on running consensus does use the complete network information and, accordingly, one may expect that it outperforms the bank of independent Page’s test in terms of detection quickness. This is in fact the case, as shown in the following sections. On the other hand, it should be clear at the outset that the bank should be preferred in sensor networks where the sensors are widely spaced and/or where the communication channel is very “non-ideal” (very noisy or affected by strong fading, etc.). Conversely, when the energy cost of a reliable communication is affordable, the running consensus scheme may be preferable.

The remainder of the paper is organized as follows. In Section 2 the problem is formalized, and classical results on Page’s test are summarized. Section 3 deals with the proposed decentralized counterparts thereof. In Section 4, the algorithms are tested on typical change detection problems, and the theoretical formulas are compared to the results of Monte Carlo simulations. Summary and final comments are provided in Section 5.

2. Problem formalization

We assume that a wireless sensor network is engaged in solving a change detection problem, which can be modeled as follows. Let \( n \geq 1 \) be a discrete time index, and let \( j \in \{1, 2, \ldots, M\} \) denote the sensor of the network,
with $M$ the number of sensors. The $n$th observation $x_{nj}$ collected by the $j$th remote node obeys a given probability density function (pdf) $f_0(x)$ until an unknown time $n_0$. From $n_0$ (included) on, the pdf of the data changes to $f_1(x)$ at all the sensors. The goal of the network is to detect such change as soon as possible, with a constraint on the average time between false alarms. Namely, for any given average time between false alarms, we want to minimize the detection delay, i.e., the time (number of samples, since our model is discrete time) between the change in distribution occurring at $n_0$ and the instant at which such change is actually declared. Formally, $\forall j$, we have

$$f_0(x) : x_{1,j}, x_{2,j}, \ldots, x_{n_0-1,j}$$

$$f_1(x) : x_{n_0,j}, x_{n_0+1,j}, \ldots$$

(1)

The random variables are assumed to be statistically independent. According to this model, at each time slot $n$, the network globally collects a vector of observations $x_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]^T$, where $T$ stands for transposition. Note that, since the network works without fusion center, there is no single place in the system where $x_n$ is fully available; rather, each sensor observes just one entry of the vector $x_n$.

Were this vector be actually available to an ideal centralized entity, this latter would operate according to the well-known Page’s test [1], which works as follows. The so-called CUSUM log-likelihood of the data is computed

$$S_n = \sum_{i=1}^{n} l(x_i) = \sum_{i=1}^{n} \sum_{j=1}^{M} \log \frac{f_1(x_{ij})}{f_0(x_{ij})},$$

(2)

where $l(x)$ denotes the log-likelihood pertaining to vector $x$. Then, the standard recursion for implementing Page’s test is

$$S_n = \max(0, S_{n-1} + l(x_n)),$$

(3)

and the associated (random) stopping time is

$$N = \arg \min_n \{S_n \geq \gamma\}.$$  

(4)

This defines a decision rule: a change of distribution is declared as soon as $S_n$ crosses the threshold $\gamma$ for the first time. Note also that (3) implies that the log-likelihood resets each time it falls below zero, which is thus the point from which Page’s test restarts.

The usual optimality criterion in change detection is to minimize the delay needed for detecting the change, while diluting as much as possible the false alarms. Accordingly, the main performance metric will be the average delay for detection, corresponding to a prescribed average false alarm rate. This latter is usually defined as the reciprocal of the average sample size under the null hypothesis, that is, $1/E_l[N]$, where $E_l[\cdot]$ denotes the statistical expectation computed under density $f_l(x), i=0, 1$, respectively [2].

As for the detection delay $E[N]$, a useful upper bound can be analytically obtained by assuming that the CUSUM is zero at time $n_0$ at which the state of the nature changes. In other words, since accounting for the exact value taken by the CUSUM statistic at $n_0$ is usually intractable, we set this unknown value to zero. The rationale is that under $f_0(x)$ the CUSUM remains close to zero, and the fact that this gives an upper bound on the detection delay follows from observing that the true path toward the threshold crossing can be only shorter than the path starting from zero [2].

Approximate formulas for the above two performance indices are available in the literature, and are obtained by neglecting the “excess over the threshold” of the test statistic at the stopping time, see, e.g., [1,2,23]. From these references, we have that the relationship between the detection threshold $\gamma$ and the false alarm rate, which we denote by $R_c$, can be written as

$$R_c(\gamma) \approx \frac{M\lambda_{01}}{e^\gamma - 1} \approx \frac{M\lambda_{01}e^{-\gamma}}{\gamma},$$

(5)

where $\lambda_{01}$ is the Kullback-Leibler (KL) divergence between $f_0(x)$ and $f_1(x)$, see [24]. Inverting the previous relationship defines the function $\gamma_c(R)$, which, for large thresholds, simplifies to

$$\gamma_c(R) \approx \log(M\lambda_{01}/R).$$

(6)

Let $D_c$ be the average number of samples from the change point from which Page’s test restarts. This defines a decision rule: a change of distribution is declared as soon as $S_n$ falls below zero, which is thus the point from which Page’s test restarts.

3. Proposed schemes for quickest detection

3.1. Page’s test with running consensus

Let us now switch to a genuinely decentralized scenario, and suppose, in particular, that a wireless sensor...
network engaged in change detection has a fully flat architecture without fusion center. A running consensus protocol is implemented to reach agreement about the ideal centralized detection statistic, as the number of consensus steps grows. In-depth description of the running consensus procedure can be found in [19–22] and will not be repeated here; in the following we only recall the basic elements needed to make this paper self-consistent.

With reference to the running consensus strategy, we assume that sensors are able simultaneously to acquire, to exchange and to process data. Specifically, during time slot $n$, observations are first collected by the sensors, and then shared according to a standard gossip algorithm [10]. The flexibility of the standard gossip algorithm allows us to model arbitrary network topologies, in the sense that the information sharing may involve neighboring sensors when the sensors’ transmission range is very limited, or arbitrarily selected nodes in the case of one-hop fully connected architectures, or whatever else. The exchanged data are not simply the measurements, but rather the current detection statistic available to any individual sensor. The local detection statistic that summarizes the state of knowledge of the sensor is also referred to as its state, and will be denoted by $S_n^j$.

What one would like, of course, is that $S_n^j$ for all $j$ would converge to the centralized detection statistic $S_n$ as the time index $n$ diverges. To aim this running consensus protocol prescribes to use the following update rule [19–22]:

$$
\begin{pmatrix}
S_{n,1} \\
S_{n,2} \\
\vdots \\
S_{n,M}
\end{pmatrix} = W_n \begin{pmatrix}
S_{n-1,1} \\
S_{n-1,2} \\
\vdots \\
S_{n-1,M}
\end{pmatrix} + MW_p\begin{pmatrix}
\ell(x_{n,1}) \\
\ell(x_{n,2}) \\
\vdots \\
\ell(x_{n,M})
\end{pmatrix}. 
$$

(12)

In the above $\ell(x) = \log f_i(x)/f_0(x)$ represents the log-likelihood pertaining to a single sample $x$, and the $M$ matrices $W_n$, $n=1,2,…$, are iid (independent identically distributed) and doubly stochastic, i.e., both the sum of the entries over the columns and the sum of the entries over the rows are equal to one.

Let $E[(S_n^j - S_n^j)^2]$ the average square error between the state $S_n^j$ of the $j$th sensor and that of the centralized system $S_n$. It can be shown that [20]\(^2\)

$$
E[(S_n^j - S_n^j)^2] \leq M^2\text{VAR}[(x)] \frac{\lambda_M}{1-\lambda_M},
$$

(13)

where $\text{VAR}[(x)]$ denotes the variance computed under density $f_i(x)$, $i=0,1$. Note that the bound is independent of the time instant $n$. In the above expression $\lambda_M \in (0,1)$ is the second largest eigenvalue of the matrix $E[W_n W_n^T]$ and takes account of the connection properties (i.e., the topology) of the network [20]. It should be emphasized how the connection properties of the network, that model the geographic distribution of the nodes, their transmission range, the channel characteristics and so forth, play their role only through $\lambda_M \in (0,1)$. Therefore, the analysis that follows is valid under general topologies (under the assumption of a connected network: any sensor can reach any other sensor by multiple hops).

To fix ideas, let us consider a simple example of matrices $W_n$ that correspond to the so-called pairwise averaging algorithm [10]. Suppose that, at time $n$, a pair $(h,k)$ of sensors is uniformly and randomly selected among the total number of pairs. The realization of $W_n$ is

$$
W_n = I - \frac{(u_k - u_h)(u_k - u_h)^T}{2},
$$

(14)

where $I$ is the identity matrix, and $u_k$ is a vector of all zeros, but for the $k$th entry which is unity. Using this matrix in (12) we see that the state of the sensors $h$ and $k$ is simply replaced by the corresponding arithmetic average.

Let us compactly denote by $U(t)$ the updating rule for the sensors’ state described by (12), namely let us rewrite (12) in the compact form $S_n^j = U(t(S_{n-1,1}^j | U^{n-1})).$ Combining that with the recursion (3), the overall recursion at the $j$th sensor can be written as follows:

$$
S_n^j = \max(0, U(t(S_{n-1,1}^j | U^{n-1}))).
$$

(15)

Note that, while the update rule $U$ is linear, the addition of Page’s reset rule has introduced a nonlinear effect not present in classical gossip algorithms.

A key observation should be made at this point. The nature of the running consensus scheme implies that all the nodes asymptotically share one and the same state (statistic). Thus, what one expects is that all the nodes report a detected change approximately at the same time instant. In fact, the running consensus introduces a strong correlation among nodes by continuously propagating information across the system, and this correlation is responsible for detecting the change at almost equal times. Therefore, provided that the system evolves for a sufficiently long time, the estimated time at which the distribution change took place can be recovered by any of the $M$ nodes of the network, and there is no need to implement some “broadcast halt” in the system: All the nodes will detect the change and their task of computing the decision statistic and of gossiping is automatically terminated at that time, without the need of a halting command to be sent to all the sensors. Moreover, the fact that the system eventually reaches consensus also implies that the performance of the network can be computed with reference to any sensor (the state of any sensor is asymptotically one and the same of all other sensors).

Since the running consensus protocol is designed to reach agreement about the ideal centralized statistic $S_n$, (see [20,22]), we expect that the statistic $S_n^j$ (any $j$) tends to track $S_n$. Accordingly, Fig. 1 shows the behavior of $S_n$ and of $S_n^j$ for three values of $j$, i.e., for three different nodes. We indeed see that they behave similarly. Specifically, initially all the statistics often reset to zero while after the change in distribution they tend to increase up to eventually declaring a change as the threshold is crossed. The times at which the threshold is crossed almost coincide for the three running consensus statistics in that, as seen in the zoomed plot, they always...
behavior quite similarly. We also see that the running consensus statistics closely track the statistic of the ideal system (continuous curve): their detection delays are close to that of the ideal system.

One can hope that the performance of the running consensus detector stays close to the theoretical limit represented by the centralized ideal system. To elaborate on the issue, it is convenient to write the detection statistic available at the $j$th sensor as

$$S_{nj} = S_n + e_{nj},$$

emphasizing the distance of the current state $S_{nj}$ from its asymptotic value $S_n$, measured by the error term $e_{nj}$. The corresponding stopping time at the $j$th sensor becomes

$$N_j = \arg \min_n (S_{nj} \geq \gamma),$$

which implicitly defines the decision rule.

Approximate performance evaluation of the decentralized Page’s test with running consensus is now in order, via the behavior of the error term $e_{nj}$. To this aim, we shall exploit results from our previous work on running consensus [19–22], and try to apply them to the nonlinear update (15). We would like to stress that the arguments below are only heuristic, and their validity is checked by computer simulations in a later section.

That said, let us first work under the assumption that the absolute error is strictly bounded, that is, $|e_{nj}| \leq \epsilon$, $\forall n$ and $\forall j$. The behavior of $\epsilon$ with the network topology and with $M$ is expected to be similar to that shown in (13). The system initially collects data distributed according to $f_0(x)$ and, until a threshold crossing occurs (because a real change or a false alarm has occurred), the $j$th sensor may have made a certain number of resets to zero, depending upon both $S_n$ and $e_{nj}$. It is reasonable to assume that, for large values of $\gamma$ (i.e., $\gamma \gg \epsilon$), the threshold crossing will be determined by the behavior of the centralized statistic $S_n$, rather than by the error term. Otherwise stated, when $S_n$ starts driving the random walk toward the upper threshold (either a real change or a false alarm), the error terms at the various sensors will become less and less influential compared to $S_n$ and the sensors will tend to agree about the need to declare a change.

These arguments can be formalized as follows. Since $S_n-\epsilon \leq S_n + e_{nj} \leq S_n + \epsilon$, we can write (assuming $\gamma > \epsilon$):

$$\arg \min_n (S_n + e \geq \gamma) \leq N_j \leq \arg \min \frac{S_n - e \geq \gamma}{a.s.},$$

where a.s. means almost surely. We can thus define

$$N = \arg \min_n (S_n \geq \gamma - \epsilon), \quad N = \arg \min_n (S_n \geq \gamma + \epsilon),$$

which are recognized as the stopping times associated with classical centralized Page’s tests having thresholds $\gamma - \epsilon$ and $\gamma + \epsilon$, respectively. From (18), we can thus write (where $i=0,1$, represents the pertinent statistical distribution)

$$E_i[N] \leq E_i[N] \leq E_i[N], \quad \forall j = 1, 2, \ldots, M.$$  

Recall that the adopted optimality criterion is to minimize the detection delay for a given false alarm rate. The performance of the running consensus test can be therefore approximated as follows. As to the false alarm rate $R_j$ at sensor $j$, we use the lower bound $E_{0} \{N\}$ that yields

$$R_j \leq R_{c}(\gamma - \epsilon),$$

with the function $R_{c}(\cdot)$ defined in (5). Now, let us fix $R$ and consider for the running consensus the threshold value $\gamma = \gamma_{c}(R) + \epsilon$, namely the threshold of the centralized system plus $\epsilon$. In this way we get $R_j \leq R_{c}(\gamma - \epsilon) = R$. 

---

**Fig. 1.** Behavior of the running consensus statistics $S_{nj}$, $j=1, 2, 3$, for quickest detection in a simple network made of three sensors (three dashed lines, almost superimposed upon each other). The continuous curve shows the statistic $S_n$ of an ideal centralized system.
implying that this threshold setting provides a conservative choice ensuring that the prescribed false alarm $R$ is not exceeded.

Switching to the analysis of the average delay at sensor $j$, say $D_j$, using (19), we have
$$D_j(\gamma - \varepsilon) \leq D_j \leq D_j(\gamma + \varepsilon),$$
and using again $\gamma = \gamma_j(R) + \varepsilon$, yields
$$D_j(\gamma_j(R)) \leq D_j \leq D_j(\gamma_j(R) + 2\varepsilon).$$

All these delays diverge when the threshold grows. However, since $D_j(\gamma)$ depends essentially linearly upon $\gamma$, we have that, in the regime of small $R$ (that is, large $\gamma_j(R)$), the ratio between $D_j$ and the average delay of a centralized test with false alarm rate $R$ approaches unity, for any $j$. Accordingly, the overall performance of the running consensus test is well-described by the same operating characteristic of the centralized system:
$$D_r(R) \approx D_c(R) \approx \frac{\log(MA_{01}/R)}{MA_{10}},$$
(20)
where the subscript “r” stands for “running”, versus “c” for “centralized”.

We should emphasize that the $\varepsilon$ appearing in (18) is unknown and, for this reason, the threshold setting $\gamma = \gamma_j(R) + \varepsilon$ is in principle precluded. However, although we cannot fix the threshold with high accuracy and confidence, any choice of the threshold leads to an operating point belonging to the optimal operating characteristic (20): The functional relationship (20) between the false alarm rate and the detection delay is accurate enough for practical purposes, as will be confirmed by simulation in Section 4.

In practice, one might simply neglect the $\varepsilon$ and set the threshold of the running consensus detector equal to that of the centralized scheme, provided by relationship (5), or by approximation (6); note that this does not ensure that the design is conservative. Thus, the system designer uses a desired false alarm rate to fix the threshold according to (5) and (6). Then the detector will exhibit a false alarm different from that assumed at the design stage but the pair of actual performance indices, false alarm rate and detection, corresponds with good accuracy to an optimal pair ruled by (20).

We have assumed so far that the error is bounded. This turns out to be reasonable in many practical applications (e.g., with quantized data giving rise to likelihood ratios with bounded support), but cannot be considered as a general rule. What if the bounded support condition is not met? Capitalizing on sophisticated mathematical tools, some rigorous asymptotic results have been obtained under the weaker assumption of boundedness on the average [25], a condition which seems suited to our problem, see Eq. (13). However, these latter studies refer to specific detection examples, and cannot be directly applied to our setup; but attempting to tailor the results in [25] to the problem we address here is well beyond the scope of this paper. From an engineering perspective the approximation of strictly bounded error is commonly used in sequential analysis to grasp the essential features of the detectors [23].

3.2. Bank of parallel Page’s detectors

In this section we study the case that each sensor runs a Page’s test, without exchanging data on-the-fly with other sensors. Such a bank of parallel Page’s processors evolves until the first of the $M$ sensors detects a change, and when this happens a broadcast message from the “firing” sensor is sent through the network in order to terminate the inference task. This is a form of cooperation that exploits the diversity among sensors’ states, and might provide performance improvements with respect to a single Page’s test.

The false alarm rate of the bank can be easily characterized. Let $L_j$ be the duration of the $j$th Page’s test, i.e., the time at which the $j$th sensor would detect a change. The overall stopping time of the bank is clearly given by
$$N^{(bank)} = \min_{j \in \{1, \ldots, M\}} L_j.$$  (21)
Assuming $f_0(\cdot)$, we argue as follows. For $\gamma$ large enough, we can suppose that the time interval between two successive false alarms is essentially ruled by the number of resets (a reset happens when the value of the CUSUM statistic is zero) times the length of a single path of the statistics evolving between two successive resets; the length of the path from the last reset to the threshold crossing may be neglected in this approximation (see also the similar arguments provided in [2]). Therefore, we compute the time interval between false alarms as the random number of resets multiplied by the average length of a single inter-reset path.

The number of resets in a single Page’s test can be shown to follow a geometric distribution [2]. Indeed after a reset the process probabilistically restarts as new, and the number of resets can be thought as the number of “failures” before observing a “success” in a sequence of iid Bernoulli trials, which leads exactly to the geometric law [26]. Therefore, let $(1 - p)p^k$ be the probability of occurrence of $k$ resets, $p$ being the characteristic parameter of the geometric.

In the case of a bank of $M$ Page’s tests, the false alarm rate is thus ruled by the minimum of $M$ iid geometric variables. Now, the minimum of $M$ iid geometric random variables of parameter $p$ is again a geometric random variable but with parameter $p^M$, as it can be straightforwardly verified [26]. At this point, we can directly exploit [2, Eq. (5.2.22)] by replacing in that equation $p$ with $p^M$, thus obtaining the relationship between the threshold $\gamma$ of the bank and the false alarm rate $R_b$. We have
$$R_b(\gamma) \approx R_c(\gamma) \approx \frac{MA_{01}}{e^{-\gamma} - 1} \approx \frac{MA_{01}e^{-\gamma}}{\gamma}.$$  (22)
Remarkably, the relationship is exactly the same as already obtained with the ideal centralized system and with the running consensus scheme.

Let us consider now the average delay $D_b = E_1[N^{(bank)}]$, whose approximate evaluation is now in order. Let $F_1(x)$ be the cumulative distribution function (CDF, hereafter) of $L_j$ when the observation model is $f_1(\cdot)$. Let us approximate this CDF as that of the stopping time of a random walk.
with positive drift and a single barrier $\gamma > 0$. This approximation involves the so-called Wald or inverse Gaussian distribution, defined as

$$F_W(x; z) = \left[ 1 - Q\left(\frac{x-1}{\sqrt{x}}\right) \right] + e^{2z}Q\left(\frac{x+1}{\sqrt{x}}\right),$$

where $Q(\cdot)$ is the standard Gaussian exceedance probability function. Indeed, we have [27]

$$F_1(\zeta; E_1[L_j]) \approx F_W(\zeta; \gamma \delta) \quad \text{with} \quad \delta = \frac{A_{10}}{\text{VAR}_1[\zeta(x)]},$$

where $\text{VAR}_1[\cdot]$ is the variance computed under $f_1(\cdot)$, and where for $E_1[L_j]$ we use (10).

Now, since $N^{\text{bank}}$ in (21) is the minimum of $M$ iid random variables $L_i$, we have that the complementary CDF of $N^{\text{bank}}$ is the complementary CDF of $L_i$ raised to the $M$th power [26]. Then, recalling that the expectation of a non-negative random variable can be computed as the integral of its complementary CDF, we can exploit the above approximation of $F_1(\cdot)$ to get

$$D_b(\gamma) \approx \gamma + e^{\gamma - 1 - \frac{1}{A_{10}}} \int_{0}^{\infty} (1 - F_W(\zeta; \gamma \delta))^M d\zeta$$

that can be easily used for numerical evaluation.

An explicit expression for the operating characteristic of the bank can be found by approximating $\gamma + e^{\gamma - 1} \approx \gamma$ in (23), and exploiting the last expression in (22), yielding

$$D_b(R) = \frac{\text{log}(M A_{10} / R)}{g(M/R) A_{10}},$$

where the function $g(M/R)$ is defined as

$$\int_{0}^{\infty} \left[ 1 - F_W\left(\zeta; \frac{A_{10}}{\text{VAR}_1[\zeta(x)]} \log \left(\frac{M A_{10}}{R}\right)\right) \right]^M d\zeta.$$

In this paper we mainly use the approximations of (23) and (24). Nonetheless, a brief digression on how a simpler formula can be also obtained for moderately large $M$, is now provided. Given $M$ iid random variables $L_i$, a simple application of the theory of the transformation of random variables reveals that

$$\min_{j=1,...,M} L_i = \frac{d}{1 - \min_{j=1,...,M} U_i},$$

where $U_i$ are iid $(0, 1)$-uniform variates, and $F_1^{-1}(\cdot)$ is the inverse function of the CDF of $L_i$. In the above $X \equiv Y$ means that the two random variables $X$ and $Y$ have the same distribution. From (21), this implies

$$E_1[N^{\text{bank}}(M)] = E \left[ F_1^{-1} \left( \frac{d}{1 - \min_{j=1,...,M} U_i} \right) \right].$$

For large $M$, a legitimate approximation consists in exchanging the expectation operator with the function $F_1^{-1}(\cdot)$ (see, e.g., Castillo [28, p. 76]), yielding

$$E_1[N^{\text{bank}}(M)] \approx F_1^{-1} \left( \frac{1}{M+1} \right),$$

where we used the fact that the expected value of the minimum of $M$ iid $(0, 1)$-uniform random variables is $1/(M+1)$. Via the approximation in terms of Wald’s distribution $F_1(x) \approx F_W(x/E_1[L_j]; \gamma \delta)$, the above yields $F_1^{-1}(\gamma) \approx E_1[L_j] F_W^{-1}(\gamma; \gamma \delta)$. By substituting in (25), the average delay can be finally expressed as

$$D_b(\gamma) \approx \frac{\gamma - \frac{1}{M+1}}{A_{10}} F_W^{-1}\left(\frac{1}{M+1}; \gamma \delta\right),$$

which, in addition to (23), is a further approximation of the delay that could be useful in the regime of large $M$.

### 3.3. Relative efficiencies

We use as a proxy to compare different detection schemes the so-called relative efficiency of strategy 1 with respect to strategy 2, for a given false alarm rate $R$. This is defined as

$$\eta_{1,2}(R) = \frac{D_2(R)}{D_1(R)},$$

$D_1(R)$ and $D_2(R)$ being the average delays of detectors 1 and 2, respectively, when their false alarm rate is fixed to the same value $R$. To avoid possible misunderstanding, we explicitly note that the term efficiency is not used in connection with energy considerations, but rather in connection with the detection performance. With this definition, $\eta_{1,2}(R) > 1$ implies that strategy 1 outperforms strategy 2, in the sense that it exhibits a smaller delay, for the same false alarm rate, and vice versa for $\eta_{1,2}(R) < 1$. As to the notation, recall that we use the subscript “$r$” to denote the decentralized Page’s test with running consensus, “$b$” refers to the bank of parallel Page’s processors, while “$s$” and “c” refer to the single Page’s test and to the ideal centralized entity, respectively. The approximations derived in the previous sections allow us to write the following relationships.

- The efficiency of the optimal centralized Page’s detector with respect to its running consensus counterpart, thanks to (20), is simply:

$$\eta_{1,2}(R) = \frac{D_2(R)}{D_1(R)} \approx 1,$$

meaning that the two strategies are essentially equivalent. This result merits emphasis: within the limits of the approximation, the decentralized scheme is as efficient as the ideal centralized detector, in terms of detection speed.

In our context of a flat architecture, the best option would be that each sensor had the same detection performance as the ideal centralized system. There is no doubt that, to achieve this goal, a huge amount of information must be shared and exchanged among the sensors. Therefore, the basic message of (27) is that the goal of mimicry of the ideal centralized system can be approximately achieved by the running consensus protocol based using gossip, which is quite parsimonious in terms of energy consumption for communications. Therefore, a considerable gain in terms of energy is paid in the coin of a very little, asymptotically negligible, performance degradation; some further comment on the cost in terms of energy will follow.

- Recalling that the operating characteristic of the running consensus coincides approximately with that of the centralized system, the efficiency of the single-sensor test with respect to the running consensus strategy can be computed by combining (5) with (7), and (9) with (10).
When \( \gamma \) is large enough, using the second approximation of (8) and (11), one gets

\[
\eta_{b_{s}}(R) = \frac{D_{s}(R)}{D_{b}(R)} \approx \frac{1}{M} \left( 1 + \frac{\log M}{\log(A_{01}/R)} \right).
\] (28)

For relatively small false alarm rate \( R \), we see that using \( M \) units essentially reduces the delay by a factor \( M \) with respect to the use of a single detector. We note explicitly that the gain in terms of detection performance is paid in the coin of the communication: the single sensor, by definition, requires no communication at all; the running consensus, whose energy parsimony with respect to the centralized system has been emphasized above, actually requires a certain amount of communication among the sensors. Computing the exact communication cost requires specifying, among other things, the exact gossip procedure (i.e., \( W_{n} \)) adopted.

- The comparison between the two decentralized architectures, i.e., the bank of Page’s tests and the running consensus, follows by combining (5) with (7), and (22) with (23). Assuming \( \gamma \) large, the relative efficiency can be derived from the second of (8) and (24) in the form:

\[
\eta_{b_{s}}(R) = \frac{D_{s}(R)}{D_{b}(R)} \approx \frac{g(M,R)}{M}. 
\] (29)

Note that the energy spent for communication by the bank is essentially negligible, amounting to a single broadcast at the end of the detection task. In this respect, it is almost equivalent to a single sensor. Conversely, running consensus does involve inter-sensor communication, and thus one may expect that the bank is outperformed by the running consensus in terms of speed.

- It is also interesting to compare the behavior of the bank against the use of a single Page’s test. From approximations (11) and (24) we have

\[
\eta_{b_{s}}(R) = \frac{D_{s}(R)}{D_{b}(R)} \approx \frac{g(M,R)}{1 + \log M / \log(A_{01}/R)}. 
\] (30)

while a more accurate expression would involve Eqs. (9), (10), (22), and (23). It turns out, as we will show in the next section, that a certain value of \( M \) exists at which \( \eta_{b_{s}}(R) \) is maximized; also, there is a certain value of \( M \) beyond which \( \eta_{b_{s}}(R) \) falls below unity, meaning that using a bank of Page’s filters is less effective than using a single unit.

### 3.4. Energy cost

A few comments on the cost in terms of energy that pertains to the different detectors are now in order. An exact evaluation is difficult, because the cost in terms of energy depends substantially upon the characteristic of the channels, the kind of modulations, the complexity of the radio devices, the availability of error control codes, and so on. At any rate, a rough computation of the energy cost, which is enough for comparing the different systems, amounts to count the number of communication sessions to be activated, i.e., we count the number of transmitted/received messages that each detector uses. It is convenient to consider the number of such messages per time sample, i.e., the number of connections per unit time.

Since the ideal centralized system requires that all the data collected by the sensors are made available to a common fusion center, this requires, for each time instant, a number of delivering equal to the number of sensors \( M \). On the other hand, the running consensus scheme requires a different number of messages for different connection matrices \( W_{n} \). For instance, with the so-called pairwise averaging matrix in (14), at each time instant only one pair of sensors exchange data: this can be counted as a single (bidirectional) connection and implies an energy cost per unit time equal to 1. In the examples that follows we also consider repeated pairwise averaging in which, during each time slot, the sensors have the chance of making \( v \) successive pairwise exchanges; then the energy cost is \( v \). Finally, the bank of Page’s filters implies no energy cost, since the communication burden is simply due to a single broadcast signal at the end of the task.

### 4. Examples and numerical experiments

#### 4.1. Gaussian detection

The previous analysis is now corroborated by simulation. We begin with a case study commonly used as benchmark in the context of model change detection. Consider hence the detection problem formalized in (1) and suppose that two zero-mean Gaussian pdfs, featuring different variances, are involved:

\[
f_{0}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^{2}/2}, \quad f_{1}(x) = \frac{1}{\sqrt{2\pi\sigma^{2}}} e^{-x^{2}/2\sigma^{2}}.
\]

As to the specific running consensus protocol employed in the decentralized Page’s test, we refer to a standard repeated pairwise averaging in which, during each time slot, the sensors have the chance of making \( v = 5 \) successive pairwise exchanges; details can be found in, e.g., [20]. We report the results from \( 10^{4} \) Monte Carlo simulations for \( M = 10 \) sensors, and \( \sigma = 1.032 \). This rather small value of \( \sigma \) implies that the two hypotheses are quite close, namely, that detecting the change in the statistical distribution is really a difficult task; we want to test our systems just in this challenging scenario.

Fig. 2 reports the comparison between the simulated false alarm rate and that obtained by exploiting the approximate analytical relationships between the threshold \( \gamma \) and the false alarm \( R \). The theoretical relationships for the ideal centralized, the running consensus and the bank coincide, as seen by (5) and (22), and by the discussion on the threshold setting for the running consensus detector, while the curve of the single sensor is scaled by a factor \( 1/M \). In Fig. 2 the lines in bold represent the more precise approximations contained in the quoted equations, while the simplifications for large \( \gamma \) are reported as thinner curves.

The simulation points of the systems studied (no simulation results are given for the case of a single Page’s test) show a satisfying accuracy, except for the case of the running consensus. Indeed, we have already pointed out that using the threshold setting as for the centralized scheme is
not very accurate, and at least a corrective term should be added. This could be certainly estimated by simulation, but a case-by-case analysis is required to address change detection problems with different distributions.

In absence of such a correction, the practical impact is that one is not able to set accurately the threshold, to guarantee a precise \( R \). On the other hand, we have already seen that, once a threshold has been selected, the corresponding pair of detection delay and false alarm rate stays quite close to the optimal curve. This is shown in Fig. 3, where we run the different tests over a range of detection thresholds and report the false alarm rate and of the detection delay.

The curves pertaining to the theoretical formulas in Fig. 3 have been drawn in the following way. The bold solid curve results by combining the first approximation of (5) with that in (7), while the thinner solid line is obtained when the corresponding expressions for large \( \gamma \) are used. As to the bank detector, the bold curve with diamond markers follows by combining (23) with the first (more accurate) approximation in (22); the rougher (24) is depicted as the thinner curve with diamond markers. Finally, as to the single sensor case, the bold dashed curve follows from the first expressions in (9) and in (10), while the thinner dashed curve depicts (11). From Fig. 3 we see that all the simulation points stay very close to those obtained by theory. Fig. 3 also shows that the small discrepancies between theory and simulation tend to vanish as \( R \) decreases, as the arguments exploited in deriving the approximate formulas anticipated.

It is important to note that the operating characteristic of the bank approaches that of a single sensor when the false alarm rate increases: with large \( R \), using \( M \) processors organized in a bank architecture provides negligible improvements with respect to the simple single test. On the other hand, for any value of the false alarm rate, we note that the running consensus provides substantial improvement with respect to the bank, in terms of detection performances. We have already pointed out that the situation is reversed in terms of communication burden: the bank might represent a viable choice whenever the usage of the running scheme is hampered by the energy spent for communication.

In Figs. 4 and 5, we display the theoretical relative efficiencies \( \eta_{s,r}(R) \), \( \eta_{b,r}(R) \) and \( \eta_{b,s}(R) \), as function of the number of sensors \( M \), for several values of the false alarm rate \( R \). We see that \( \eta_{s,r}(R) \) is always less than unity as we expected, and the curves approach that of \( 1/M \), also reported for comparison, for small false alarm rates. As a matter of fact, this behavior can be easily predicted by inspecting directly Eq. (28).

From \( \eta_{b,r}(R) \) we observe that the running scheme always outperforms the bank of filters, as the arguments below (29) suggested. It is also worth noting how \( \eta_{b,r}(R) \) approaches zero for any choice of \( R \), thus revealing that the fusion of the sensors’ data during the detection test significantly improves the performances.

The behavior of \( \eta_{b,s}(R) \) in Fig. 5 reveals that the bank of Page processors exhibits an optimum value of \( M \) at which the detection efficiency with respect to a single sensor is maximized. There also exists a maximum value of \( M \) beyond which \( \eta_{b,s}(R) \) falls below unity: interestingly, for moderately large \( M \) there is no benefit at all in using a bank of Page’s detectors with respect to a single unit, in terms of system performances. As an example, with a false alarm rate of \( 10^{-4} \), it turns out that a bank made of only five filters provides the best performance improvements with respect to a single filter. The relative efficiency decreases for \( M > 5 \) and at about \( M=220 \) falls below unity, meaning that a bank of more than 220 filters performs worse than a single detection unit. This limiting value of \( M \) seems to increase significantly with decreasing false alarm rate: at false alarm rates below \( 10^{-5} \) it is always better to use the bank, and the optimal \( M \) that maximizes the performances takes values in the hundreds.

The curves in Figs. 4 and 5 were obtained with the more accurate expressions derived in the previous
sections. Actually, using the simpler relationships (28) and (29) in Fig. 4, would change nothing since the curves would be practically indistinguishable from those given in the figure. As to \( Z_b \), in Figs. 5, Eq. (30) does give the precise behavior of the efficiency but, to address the quantitative analysis of the limiting values of \( M \), the more accurate formulas are certainly preferable.

4.2. Bernoulli observations

We now consider a completely different example with discrete observations. Let us assume that the data collected by the sensors are iid binary variables taking value in \( \{0,1\} \), drawn from a Bernoulli distribution. Initially, the probability of observing 1 is 0.5, i.e., the binary variables are equiprobable, whereas after the change in distribution the probability of 1 becomes 0.505. Therefore, we are faced with detecting as quickly as possible a slight imbalance in the occurrence of the symbols. As for the previous example, the two hypotheses are quite close and the detection task is challenging.

In this subsection we also consider a different network topology—actually we recall that the analysis developed in this paper is valid for arbitrary network topologies and connection matrices, provided that the system is connected in the sense that any sensor can reach any other sensor by multiple hops. Up to now we have considered the connection matrix in (14), which models a pairwise gossip where each pair of communicating sensors is selected at random among all the possible pairs. We now assume, instead, that sensors can communicate only with their direct neighbors: we refer to Fig. 6 where neighboring sensors are connected by straight lines. It is also assumed that at each time step \( v \) pairs of neighboring sensors are selected to average their state, and in the worked example we use \( v=1 \) and 5.

The results of computer simulations are shown in Fig. 7 and comments similar to those of the Gaussian example apply. The relationship between the threshold and the false alarm rate is compared to the analytical approximation of (5). Note that by increasing \( v \) the accuracy improves, since convergence toward the consensus speeds up [20]. The operational characteristic obtained by computer experiments is compared to the theoretical curve resulting from the first approximation of (5) with the first approximation in (7). The match is excellent.
5. Summary

In a fully decentralized scenario where a sensor network is not equipped with a central unit, and where the sensors are given some capability of communication, we consider the simplest version of quickest detection problems, with fully known distributions and permanent change. The main aim of the work is to explore the feasibility of a consensus-based decentralized quickest detector when the system architecture is fully flat (i.e., without fusion center). The bottom line of this analysis is that the running consensus detector is not only implementable, but its performance is very close to that of an ideal centralized entity.

We also investigate the behavior of a bank of detectors, each implementing a Page procedure, whose only interaction consists of halting the whole detection task whenever the first of the $M$ nodes declares a change. This can be viewed as a kind of sensor network subject to an OR stopping rule. For comparison purposes, we also report results of a single, well-studied, Page test. There are, therefore, four actors on the scene: a single test, the ideal centralized entity, the bank of filters, and the running consensus-based detector. The former two are used only as baselines for comparison, investigating the last was the original motivation of our work, and the bank also provides non-trivial insights.

Analytical approximations are derived for the detection delay and the false alarm rate of these systems, and their quantitative comparison is addressed in terms of relative efficiencies. In addition, we also briefly discuss the energy that each architecture requires to work, even though our analysis here is only qualitative. The effect of the network topology is discussed. A Gaussian change-in-variance example is studied in some detail and an example with discrete observations is also addressed by computer simulations that corroborate the theoretical analysis. Our main findings and insights are summarized as follows.

(i) The bank architecture consists of $M$ Page’s tests, and exploits diversity among them by considering only the quickest one. Conversely, the running consensus detector fuses not the decisions but the original observations. Such a pre-decision fusion rule outperforms the post-decision fusion of the bank, as quantitatively measured by the relative efficiency $R_{bc}$ which is always less than unity, and that rapidly approaches zero when the number $M$ of
elementary units grows. Indeed, the next comment applies.

(ii) The performance of the running consensus detector approaches that of the ideal centralized system: \( \eta_{cr} \approx 1 \). In particular, for low false alarm rates, the running consensus detector made of \( M \) units performs approximately \( M \) times better than a single unit: \( \eta_{cr} \to 1/M \), meaning that a change in the state of the nature is detected \( M \) times faster.

(iii) Energy considerations may suggest resorting to the more parsimonious bank, in place of the running consensus detector. In this case it is worth noting that the bank usually outperforms a single test, as we expect. However, what seems not obvious, there exists a maximum number of units beyond which the Page-bank system is outperformed by one any of its constituent units. There also exists an optimal system size that maximizes the performance of the bank: \( \eta_{bs}(M) \) admits a maximum at some \( M \).

Further investigations might include in the analysis the case of transient changes, partially unknown statistical distributions, scenarios with quantized data, the adoption of optimality criteria encompassing energy constraints, and so forth. As to the bank architecture, considering AND or \( k \)-out-of-\( M \) halting rule, in place of the OR here addressed, may be of some interest.

References