Pulsar Glitch Detection with a Hidden Markov Model

A. Melatos1,2, L. M. Dunn1,2, S. Suvorova1,2,3, W. Moran3, and R. J. Evans2,3

1 School of Physics, University of Melbourne, Parkville, VIC 3010, Australia; amelatos@unimelb.edu.au
2 Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), Parkville, VIC 3010, Australia
3 Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia

Received 2019 December 18; revised 2020 May 5; accepted 2020 May 6; published 2020 June 16

Abstract

Pulsar timing experiments typically generate a phase-connected timing solution from a sequence of times of arrival (TOAs) by absolute pulse numbering, i.e., by fitting an integer number of pulses between TOAs in order to minimize the residuals with respect to a parameterized phase model. In this observing mode, rotational glitches are discovered, when the residuals of the no-glitch phase model diverge after some epoch, and glitch parameters are refined by Bayesian follow-up. Here, we present an alternative, complementary approach which tracks the pulse frequency \( f \) and its time derivative \( \dot{f} \) with a hidden Markov model (HMM), whose dynamics include stochastic spin wandering (timing noise) and impulsive jumps in \( f \) and \( \dot{f} \) (glitches). The HMM tracks spin wandering explicitly, as a specific realization of a discrete-time Markov chain. It discovers glitches by comparing the Bayes factor for glitch and no-glitch models. It ingests standard TOAs for convenience, and being fully automated, allows performance bounds to be calculated quickly via Monte Carlo simulations. Practical, user-oriented plots of the false-alarm probability and detection threshold (e.g., minimum resolvable glitch size) versus observational scheduling parameters (e.g., TOA uncertainty, mean delay between TOAs) and glitch parameters (e.g., transient and permanent jump sizes, exponential recovery timescale) are presented. The HMM is also applied to \( \sim \) yr of real data bracketing the 2016 December 12 glitch in PSR J0835–4510 as a proof of principle. It detects the known glitch and confirms that no other glitch exists in the same data with size \( \lesssim 10^{-7} f \).

Unified Astronomy Thesaurus concepts: Pulsars (1306); Neutron stars (1108); Stellar rotation (1629); Gravitational waves (678)

1. Introduction

The exceptional rotational stability of pulsars allows a terrestrial observer armed with an accurate clock to construct a phase-connected timing solution by absolute pulse numbering, even when the observations are irregularly spaced over decades and separated by many pulse periods (Lyne & Graham-Smith 2012). Traditionally, the timing solution is constructed in three stages. (i) The pulse train is folded by cross-correlating against a template profile in the frequency domain to generate a sequence of times of arrival (TOAs; Taylor 1992). (ii) A phase model is stipulated, which includes frame-of-reference terms (e.g., solar system barycenter), the pulsar’s intrinsic spin evolution (e.g., spin frequency and its time derivatives, packaged as the coefficients of a Taylor series), astrometric terms (e.g., sky position and proper motion), dispersion in the interstellar plasma, Keplerian orbital elements (if the pulsar is in a binary), and post-Keplerian corrections (Edwards et al. 2006). (iii) The parameters of the phase model are inferred by fitting the TOAs using a weighted least-squares algorithm, such that the residuals are white and minimized, if the model is perfect. The approach (i)–(iii) has proven highly successful. It forms the backbone of observational studies in pulsar astronomy on a wide range of topics, including tests of general relativity (Taylor 1992; Stairs 2003), magnetospheric electrodynamics and coherent emission (Michel 1991; Melrose 2017), interstellar scintillation (Rickett 1990), population synthesis and binary evolution (Faucher-Giguère & Kaspi 2006), and the search for nanoHz gravitational waves (Lentati et al. 2015; Shannon et al. 2015; Arzoumanian et al. 2016; Hobbs & Dai 2017).

One intriguing phenomenon revealed by phase-connected timing is rotational glitches: impulsive, erratically occurring, spin-up events which interrupt the secular, electromagnetic spin down of a rotation-powered pulsar. Traditionally, a glitch is discovered when the residuals with respect to a glitchless phase model diverge after a certain epoch, e.g., a jump in spin frequency causes a linear phase ramp. Once the glitch is discovered, two separate, glitchless phase models are fitted to the TOAs before and after the relevant epoch. The differences between the models define the parameters of the glitch, e.g., the jump in spin frequency and its derivatives (Lyne et al. 2000; Espinoza et al. 2011). It is hard to do this uniquely, because the phase evolution includes stochastic spin wandering, known as timing noise (Cordes & Downs 1985), which is often covariant with glitch-related features like post-glitch recoveries (Lyne et al. 1996). Moreover, the gaps between observations can be long and irregular, leading to degeneracies. Work has been undertaken recently to address these issues by applying Bayesian model selection to glitch detection, e.g., using software like TEMPO2 (Lentati et al. 2014; Shannon et al. 2016; Yu & Liu 2017; Lower et al. 2018, 2019). Bayesian methods are promising but relatively expensive; they have not been applied to most pulsars to date.

The physical mechanism that triggers glitch activity remains a mystery; see Haskell & Melatos (2015) for a recent review. Broadly speaking, however, it is thought to involve the sudden relaxation of spin-down-driven elastic stress and differential rotation by local, stick–slip processes such as starquakes (Middleditch et al. 2006; Chugunov & Horowitz 2010) and superfluid vortex avalanches (Warszawski & Melatos 2011). In this picture, glitches and their recoveries probe the material properties of bulk matter at nuclear densities, e.g., the shear modulus and superfluid energy gap, under physical conditions that cannot be replicated on Earth (Yakovlev et al. 1999;
Lattimer & Prakash 2007; van Eysden & Melatos 2010; Watts et al. 2015). In particular, the statistics of glitch sizes and waiting times carry important information (Melatos et al. 2008, 2018; Ashton et al. 2017; Fulgenzi et al. 2017; Carlin & Melatos 2019; Fuentes et al. 2019). Expanding the glitch database is essential for achieving a better understanding of the nuclear physics involved.

In this paper, we develop a fast approach to glitch detection and estimation, which complements the existing approach and contributes new insights into performance bounds and spin wandering, as explained in detail in Section 2. Standard TOAs, generated by cross-correlating the pulse train against a template profile, are still the starting point. Existing software like TEMPO2 and PSRCHIVE can be used unaltered. The TOAs are analyzed with a hidden Markov model (HMM), which tracks the underlying evolution of the pulsar’s rotation, including the secular and stochastic components associated with electromagnetic spin down and spin wandering, respectively. Glitches are detected by Bayesian model selection, by comparing the evidence for HMMs with and without glitches (as in TEMPO-EST). The paper is structured as follows. In Section 2, we motivate the algorithm by explaining clearly how it fits with existing approaches and what open issues it addresses. In Section 3, we define the logical components of the HMM-based phase tracker and map each component to its corresponding measurement or model variable in a pulsar timing experiment. In Section 4, we present and justify an algorithm for converting the HMM output into Bayesian evidence in order to select rigorously between phase models with and without glitches. The performance of the glitch-finding algorithm is then tested. Synthetic data are generated according to the procedure discussed in Section 5. Performance metrics such as receiver operating characteristic (ROC) curves are evaluated systematically as functions of the astrophysical and measurement noises, secular spin-down parameters, and glitch parameters in Section 6. The figures in Section 6 are designed to be practical. Together, they can be used to plan a glitch discovery campaign as a function of experimental variables such as TOA uncertainties and the desired measurement resolution, e.g., of glitch sizes and recovery timescales.

The paper is framed as a method paper. Most of the tests are done on synthetic data deliberately, to study the behavior of the algorithm under controlled conditions. The next step is to apply the HMM to real data, a larger project that is underway. A quick foretaste of what is possible is presented in Section 7 using public data from PSR J0835−4510 as a worked example. Theoretical aspects of the algorithm are explored further by Suvorova et al. (2018) in a general signal processing context.

2. Motivation

Before introducing the HMM in Section 3, we explain first what issues in glitch detection the new algorithm seeks to address, under what circumstances it proves useful (and when it does not), and how it complements traditional glitch detection methods. Existing approaches enjoy a long record of success, and it is important to articulate what specific contributions the new algorithm makes. The main contributions are (i) a fast

2 Electronic access to up-to-date glitch catalogs is available at the following locations on the World Wide Web: http://www.jb.man.ac.uk/pulsar/glitches/gTable.html (Jodrell Bank Centre for Astrophysics) and http://www.atnf.csiro.au/people/pulsar/psrcat/glitchTbl.html (Australia Telescope National Facility).

Some interesting questions remain unanswered about the performance bounds of traditional glitch searches based on software packages such as TEMPO2 (Hobbs et al. 2006; Edwards et al. 2006), PSRCHIVE (van Straten et al. 2012), TEMPO2EST (Lentati et al. 2014), and their relatives. Given the spin wandering amplitude and TOA measurement uncertainty in a particular pulsar, as well as a glitch size detection threshold, what is the false-alarm probability, when a traditional glitch search is performed? Are all cataloged glitches real (see footnote 4), or are some of the smaller events actually spin wandering (Jones 1990; D’Alessandro et al. 1995; Janssen & Stappers 2006; Yu & Liu 2017)? What is the smallest event that a traditional glitch search can detect, as a function of the false-alarm and false-dismissal probabilities? How does the detection limit vary between objects with different spin wandering amplitudes? Some work has been done to develop quantitative answers to these questions. Janssen & Stappers (2006) conducted Monte Carlo simulations to estimate the minimum glitch size resolvable in PSR J1740−3015; see also Watts et al. (2015) with reference to the Square Kilometer Array. Shannon et al. (2016) and Lower et al. (2018) reanalyzed TOAs from PSR J0835−4510 and PSR J1709−4429, respectively, within a Bayesian framework to look for false alarms and false dismissals, and a similar, multiobject project is underway using data collected by the Molonglo Observatory Synthesis Telescope (Jankowski et al. 2019; Lower et al. 2019). Yu & Liu (2017) have performed the largest study of glitch detection probabilities so far, again within a Bayesian framework, involving 165 pulsars timed by the Parkes Observatory between 1990 and 2011 (Yu et al. 2013; Yu & Liu 2017). The latter authors argued persuasively that the study should be extended to more pulsars. However, the task is not easy. One rigorous approach in signal processing is to construct an ROC curve for the search algorithm in question by plotting the detection probability against the false-alarm probability. This entails many Monte Carlo simulations, which are prohibitive to analyze, when traditional algorithms still rely on human supervision (e.g., by-eye inspection of post-fit residuals) even when aided by software like TEMPO2EST. Crowdsourcing offers one possible solution, perhaps by leveraging the infrastructure of the PULSE@Parkes project (Hobbs et al. 2009), but it brings its own logistical challenges. Consequently, few if any ROC curves have been published for traditional glitch-finding schemes.

How does the new algorithm relate to traditional methods of glitch detection? The HMM formulation shares some common features with recent work developing a new, Bayesian, pulsar timing infrastructure based on pulse domain analysis and/or model selection (Lentati et al. 2014, 2015, 2017b, 2018; Lentati & Shannon 2015; Lentati et al. 2017a; Ashton et al. 2019). The main similarity is that a glitch is discovered when the Bayes factor comparing glitch and no-glitch phase models surmounts a user-selected threshold, as with TEMPO2EST. However there are differences. (i) The HMM plugs into the traditional infrastructure for generating TOAs. It does not operate in the pulse domain, in order to maximize the use of existing software. It can be extended to the pulse domain in the future, if there is enough demand. (ii) The HMM does not treat spin

5 M. E. Lower (2020, private communication).

6 This is also true for many TEMPO2EST analyses to date.
wandering as “noise”; it tracks it explicitly. In other words, it evaluates the likelihood of the specific spin wandering pattern observed (i.e., a specific realization of a discrete-time Markov chain, in the language of stochastic processes), whereas TEMPO1EST and related algorithms analyze the ensemble statistics of the spin wandering (e.g., the timing noise power spectral density; Coles et al. 2011). (iii) The HMM is fast. It requires \( \sim 10^{12} \) floating point operations \((\sim 0.1 \text{ central processing unit (CPU) hours})\) per target per year of observations, starting from an approximate, glitchless timing solution generated by traditional methods.

We emphasize that the approach developed here does not supplant traditional timing methods nor the newer pulse domain approach. All three approaches complement each other and are more powerful when deployed in tandem. For example, when the goal is to measure a slow, secular phase evolution described faithfully by a Taylor expansion (e.g., in binary pulsar tests of general relativity), the HMM formulation is unnecessary, because there is no covariance between stochastic spin wandering and the secular dynamics (e.g., binary orbital decay). On the other hand, when spin wandering is covariant with other short-timescale phenomena like glitches and their recoveries, the HMM offers an alternative perspective on whether a glitch occurs by tracking the spin wandering directly within systematic performance bounds, while ingesting standard TOAs for the sake of convenience.

3. Phase Tracking

An HMM is a scheme for inferring the trajectory of a system through a sequence of unobservable (hidden) states by measuring observables related probabilistically to the hidden states. In the pulsar context, the observables are the TOAs, and the hidden state is the underlying rotational state of the pulsar (e.g., its spin frequency and instantaneous derivatives with respect to time), which cannot be measured uniquely from a single TOA or the interval between a TOA pair. In Section 3.1, we describe how to formulate the pulsar timing problem in terms of an HMM, which converts TOAs into a phase-connected timing solution. The state structure of the HMM is defined precisely in Section 3.2. We then relate the TOAs probabilistically to the pulsar’s rotational state in Section 3.3 and describe how the rotational state evolves stochastically under the action of electromagnetic spin down, timing noise, and glitches in Section 3.4. Resolution and gridding issues are discussed in Section 3.5. An efficient algorithm for solving the HMM numerically is set out in Appendix A. The presentation follows closely the formal derivation by Suvorova et al. (2018).

3.1. HMM Formulation

An HMM is a probabilistic finite-state automaton\(^7\) specified by a hidden-state variable \(q(t)\), which can take on \(N_q\) discrete values; an observation variable \(o(t)\), which is not necessarily discrete; and a sequence of times \(t_1 \leq \ldots \leq t_{N_f}\) when snapshots of the system are taken. In general, \(q(t_n)\) and \(o(t_n)\) are multidimensional vectors, and the times \(t_n\) are unequally spaced.

The probability for the system to jump from hidden state \(q_i\) at time \(t_n\) to hidden state \(q_j\) at time \(t_{n+1}\) is called the transition probability. It is given by

\[
A_{q_i q_j} = \Pr[q(t_{n+1}) = q_j | q(t_n) = q_i].
\]  

(1)

The probability of measuring the datum \(o(t_n)\) at time \(t_n\), if the system is in state \(q(t_n) = q_i\), is called the emission probability. It is given by

\[
L_{o(t_n) q_i} = \Pr[o(t_n) | q(t_n) = q_i].
\]

(2)

Writing \(Q_{1:N_f} = \{q(t_1), \ldots, q(t_{N_f})\}\) and \(O_{1:N_f} = \{o(t_1), \ldots, o(t_{N_f})\}\), we can express the total probability that the observed sequence \(O_{1:N_f}\) arises from the hidden sequence \(Q_{1:N_f}\) as

\[
\Pr(Q_{1:N_f} | O_{1:N_f}) = \Pi_{q(t_1)} L_{o(t_1) q(t_1)} \prod_{n=2}^{N_f} A_{q(t_n) q(t_{n-1})} L_{o(t_n) q(t_n)}.
\]

(3)

where

\[
\Pi_{q_i} = \Pr[q(t_i) = q_j]
\]

(4)

denotes the prior probability.

Three essential questions of practical value can be asked about an HMM of the above form (Rabiner 1989; Quinn & Hannan 2001). First, given the observed sequence \(O_{1:N_f}\) and a model \(M = \{A_{q_i q_j}, L_{o(t_n) q(t_n)}, \Pi_{q_i}\}\), what is \(\Pr(O_{1:N_f} | M)\), i.e., what is the Bayesian evidence for \(M\)? Knowing \(\Pr(O_{1:N_f} | M)\), one can select between different models. Second, given \(O_{1:N_f}\) and \(M\), what is the optimal hidden sequence \(Q_{1:N_f}\) which best explains the data according to some meaningful metric? Third, given \(O_{1:N_f}\), what model \(M\) maximizes \(\Pr(O_{1:N_f} | M)\)?

The first and second questions in the previous paragraph are fundamental to the glitch-finding problem studied in this paper. Efficient algorithms to solve them are presented in Appendix A and assembled into a systematic glitch-finding scheme in Section 4. There is no unique answer to the second question. One possible solution is \(Q_{1:N_f}^* = \text{arg max} \Pr(Q_{1:N_f} | O_{1:N_f}, M)\), which maximizes \(\Pr(Q_{1:N_f} | O_{1:N_f}, M)\) sequence-wise (Quinn & Hannan 2001). Another possible solution is \(\hat{q}(t_n) = \text{arg max} \Pr(q(t_n) | O_{1:N_f}, M)\) for \(1 \leq n \leq N_f\), which maximizes \(\Pr(q(t_n) | O_{1:N_f}, M)\) point-wise (Rabiner 1989). The third question, which corresponds here to learning a dynamical model of glitches statistically from the data, can be solved by iterative methods like the Baum–Welch algorithm (Rabiner 1989) but lies outside the scope of this work.

3.2. Summary of HMM Components

In the pulsar timing context, the components of the HMM are the following.

1. Hidden state. In this paper, we track the instantaneous frequency \(f(t)\) and its first time derivative \(f'(t)\). Future work can easily include higher-order derivatives, e.g., the secular component of the second derivative \(f''(t) = n f f' / f\) describing electromagnetic braking, where \(1 \leq n \leq 3\) is the electromagnetic braking index (Melatos 1997; Archibald et al. 2016). The stochastic component of \(f\), whose magnitude usually exceeds \(n f f' / f\) (Arzoumanian et al. 1994; Johnston & Galloway 1999), is absorbed in the wandering of \(f\). We also define (but do not track; see Sections 3.4 and 4) a Boolean variable, \(g(t)\), which equals unity if a glitch occurs at time \(t\) and zero otherwise.
In summary, therefore, the hidden state is \( q(t) = [f(t), \dot{f}(t), g(t)] \).

2. Observable. In this paper, the HMM time sequence \( \{t_1, ..., t_N\} \) is defined to map one to one onto the measured, unequally spaced TOAs, starting from the second TOA. The measurement variable at time \( t_n \) is defined to equal the displacement between consecutive TOAs, viz. \( o(t_n) = t_n - t_{n-1} \), where \( t_0 \) corresponds to the first TOA; henceforth we write \( x_n = t_n - t_{n-1} \) for brevity. Future refinements include augmenting \( o(t_n) \) with auxiliary information, e.g., tagging it with the pulse period measured locally at each TOA.

3. Emission probability. Given \( x_n \) and an associated measurement error, whose variance equals \( \sigma_{\text{TOA}}^2 \), there exists a limited but degenerate set of \( (f, \dot{f}) \) pairs, which produce an integer number of pulses in the interval \( x_n \). An explicit formula for the emission probability for arbitrary \( f \) and \( \dot{f} \) and Gaussian measurement errors is given in Section 3.3 in terms of the von Mises distribution. By way of illustration, in the artificial special case with \( \dot{f} = 0 \) and \( \sigma_{\text{TOA}} = 0 \), the emission probability is proportional to a sum of delta functions, \( \delta(f - 1/x_n) + \delta(f - 2/x_n) + \ldots \).

4. Transition probability. In this paper, we track the rotational phase on three timescales: (i) secular, electromagnetic braking on the longest timescale, \( f \gtrsim 10^3 \) yr, which greatly exceeds the total observation span, \( T_{\text{obs}} \lesssim 10^2 \) yr; (ii) spin wandering (timing noise) on an intermediate timescale, stretching from days to years (Cordes & Downs 1985; Price et al. 2012; Goncharov et al. 2019; Namkham et al. 2019; Parthasarathy et al. 2019; Lower et al. 2020); and (iii) glitches, i.e., unresolved jumps in \( f \) and \( \dot{f} \), whose rise times are much shorter than \( x_n \). The stochastic dynamics of \( q(t) = [f(t), \dot{f}(t), g(t)] \), which determine \( A_{\phi_{\text{gb}}} \), are modeled as biased Brownian motion with process variance per unit time \( \sigma^2 \) via a Langevin equation in Section 3.4 and Appendix B. Note that glitches are often followed by quasiexponential recoveries, which last days to years (van Eysden & Melatos 2010). The recoveries can be incorporated into the phase model in future work. Here we absorb them into the timing noise, which occurs on a similar timescale, and show a posteriori that this is an effective approach in practice, with the algorithm successfully detecting glitches in synthetic data containing recoveries (see Section 5).

5. Prior. A uniform prior is adopted on \( f \) and \( \dot{f} \) within a restricted domain, known as the domain of interest (DOI; see Section 3.5). Practically, the DOI for any pulsar is defined by traditional phase-connected timing methods, e.g., a standard TEMPO2 fit, as well as prior astrophysical knowledge, e.g., population-based constraints on glitch sizes (Melatos et al. 2008; Espinoza et al. 2011; Howitt et al. 2018). In general, \( \Pr(Q_{1:N_f}|O_{1:N_f}) \) is insensitive to the choice of a uniform prior, because \( \Phi_{\text{gb}} \) is just one factor out of \( 2N_f \gg 1 \) in what is usually a large product in Equation (3) (Suvorova et al. 2016, 2017; Abbott et al. 2017).

### 3.3. Emission Probability

Given a displacement \( x_n \), what can we say probabilistically about the rotational state of the pulsar at \( t_n \)? For \( \dot{f}(t_n) = 0 \), without measurement noise, we can infer the instantaneous frequency, \( f(t_n) \), to be an integer multiple of \( x_n^{-1} \). For \( \dot{f}(t_n) \neq 0 \), a particular combination of \( x_n, f(t_n), \) and \( \dot{f}(t_n) \) is inferred to be an integer. The combination is unique, as long as \( x_n \) is short enough (see below). When measurement noise is switched on, these statements continue to hold true, but the estimates are “fuzzy.” In the absence of measurement noise and discontinuous glitches, and with \( \dot{f} = 0 \) over a short-enough timescale, we can approximate the frequency evolution in the interval \( t_{n-1} \leq t \leq t_n \) as a backward Taylor series, \( f(t) = f(t_n) + (t - t_n)\dot{f}(t_n) \), and then integrate \( d\phi/dt = 2\pi f(t) \) to get the phase,

\[
\phi(t_n) = \phi(t_{n-1}) + 2\pi x_n f(t_n) - \pi x_n^2 \dot{f}(t_n). \tag{5}
\]

The minus sign in the last term arises because we use a backward difference scheme. Let \( N_{\phi} \) be the number of pulses between \( t_{n-1} \) and \( t_n \). By the definition of the TOAs, \( N_{\phi} \) is an integer, and we have \( N_{\phi} = \Phi(x_n) \) with \( \Phi(x_n) = x_n f(t_n) - x_n^2 \dot{f}(t_n)/2 \). This equation corresponds to a line in the \( f(t_n) - \phi(t_n) \) plane given \( x_n \) and \( N_{\phi} \).

If each TOA has a Gaussian measurement error with zero mean and variance \( \sigma_{\text{TOA}}^2 \), then \( x_n \) also has a Gaussian measurement error, denoted by \( w_n \), with twice the variance. We write the measurement equation as

\[
x_n = \Phi^{-1}(N_{\phi}) + w_n, \tag{6}
\]

where \( \Phi^{-1} \) is the inverse function of \( \Phi \), not its reciprocal. In practice, \( x_n \) is always short enough, i.e., \( x_n \ll 2\pi f(t_n)/|\dot{f}(t_n)| \), so that \( \Phi \) is uniquely invertible up to an integer multiple. The inversion is unique, even when the timing noise is strong \( |\dot{f}| \gg \sigma_{\text{TOA}}/f \), unlike higher-order Taylor expansions, where the inversion is multivalued (modulo the integer multiples) for \( x_n^2 \gtrsim 6f(t_n)/|\dot{f}(t_n)| \). In this paper, timing noise is tracked explicitly via the HMM transition probability, as described in Section 3.4.

The emission probability is proportional to the probability density function (PDF) of the observed variable \( x_n \). Suvorova et al. (2018) showed that the PDF of \( \Phi(x_n) \) is approximately a wrapped Gaussian, because the phase is \( 2\pi \) periodic; see Appendix A of the latter reference. Suvorova et al. (2018) also showed that the wrapped Gaussian can be approximated accurately by a von Mises distribution (Mardia & Jupp 2009), which is more convenient to evaluate numerically. Hence, one can write

\[
L_{x_n}(q(t_n)) = [2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos[2\pi \Phi(x_n)]\}, \tag{7}
\]

with

\[
\kappa = [2\sigma_{\text{TOA}}^2 f(t_n)^2]^{-1}. \tag{8}
\]

In Equation (7), \( I_0(\kappa) \) symbolizes a modified Bessel function of the first kind. It is approximated by \( I_0(\kappa) \approx (2\pi\kappa)^{-1/2} \exp(\kappa) \) in the regime \( \kappa \gg 1 \) to avoid underflow errors in the computation. Intuitively, \( \kappa^{-1/2} \) is the number of pulses squeezed into a time interval lasting as long as the uncertainty in \( x_n \). Note that the hidden state \( q(t_n) \) enters Equation (7) through \( \Phi(x_n) \), which depends on \( f(t_n) \) and \( \dot{f}(t_n) \). By contrast, \( N_{\phi} \) does not enter Equation (7) explicitly; the HMM does not count the number of pulses in the interval \( t_{n-1} \leq t \leq t_n \) explicitly, although this information can always be extracted post factum using Equation (6), once the HMM is solved to obtain \( Q_{1:N_f} \).
Figure 1 displays a sample of $L_{x=q(t)}$ contours in the $f(t_n)-\dot{f}(t_n)$ plane for two $x_n$ values. Each stripe corresponds to a peak of $L_{x=q(t)}$ along the line $x_n = \Phi(x_n) = (2\pi)^{-1}[x_n f(t_n) - x_n^2 \dot{f}(t_n)/2]$. Its slope, $2/x_n$, decreases as $x_n$ increases. Formally speaking, Equation (7) has an infinite number of equal-height peaks, each corresponding to an integer value of $N_w$. In practice, the number of peaks within the DOI (drawn arbitrarily here as the figure frame) is finite. Without extra information, e.g., a phase-connected solution constructed by traditional means, all the peaks are equally likely. As $x_n$ increases three-fold from the left panel to the right panel, two things happen: the minimum $f(t_n)$ (corresponding to $N_w = 1$) decreases, and the separation of the peaks decreases. On the one hand, therefore, there is greater ambiguity, because there are more peaks to interrogate in the DOI. On the other hand, the estimate of $q(t_n)$ is more accurate, once the HMM finds the optimal peak, because the peaks are narrower.\footnote{The peaks are narrower because they are more closely separated, not because their width decreases relative to their separation. The argument of the cosine in Equation (7) depends on $x_n$, but the factor $\kappa$ multiplying the cosine does not.} In Figure 1, the number of yellow stripes in the frame increases from 2 to 18, as $x_n$ increases from $1 \times 10^5$ s in the left panel to $3 \times 10^5$ s in the right panel. The FWHM per stripe projected on the $\dot{f}(t_n)$ axis decreases from $7.2 \times 10^{-11}$ Hz s$^{-1}$ in the left panel to $7.9 \times 10^{-12}$ Hz s$^{-1}$ in the right panel.

Equations (7) and (8) assume that the state space is continuous. In practice, the $f(t_n)-\dot{f}(t_n)$ plane is divided into a grid. A generalized version of Equation (8) that accounts for gridding and lets $\sigma_{\mathrm{TOA}}$ vary with $t_n$ is discussed in Section 3.5.

3.4. Transition Probability

The equation of motion obeyed by $q(t) = [f(t), \dot{f}(t), g(t)]$ in a real pulsar is unknown. Instead, we construct an idealized model for how $q(t)$ evolves during the HMM step $t_{n-1} \leq t \leq t_n$. Away from a glitch, we assume that the system obeys a continuous Wiener process described by the Langevin equation:

$$\frac{d^2 f}{dt^2} = \xi(t),$$

where $\xi(t)$ is a fluctuating torque derivative with white-noise statistics satisfying $\langle \xi(t) \rangle = 0$ and $\langle \xi(t) \xi(t') \rangle = \sigma^2 \delta(t-t')$, and $\sigma$ is a tunable parameter (units: Hz s$^{-3/2}$). At the instant when a glitch occurs, the continuous evolution is interrupted, and $f(t)$ and $\dot{f}(t)$ undergo impulsive permanent changes, $\Delta f_p$ and $\Delta \dot{f}_p$ respectively.

We emphasize that Equation (9) is designed mainly with the practical needs of the HMM in mind; it should not be viewed as a physical model of a pulsar. Nevertheless it does embody the three physical timescales discussed in point 4 in Section 3.3: long (electromagnetic braking), intermediate (timing noise), and short (glitches). Electromagnetic braking enters through the initial conditions; the secular spin-down torque sets $\dot{f}(t_{n-1})$. We ignore $\langle \dot{f} \rangle = n f^2 / f$ in Equation (9) as discussed in Section 3.2. Timing noise enters through the right-hand side of Equation (9). Its amplitude is set by $\sigma$, which satisfies $\langle (\dot{f}(t_n) - \dot{f}(t_{n-1}))^2 \rangle = \sigma^2 x_n$ as for any Wiener process. Glitch-driven jumps in the frequency and frequency derivative.
Glitches can occur anywhere within a TOA gap, because a discrete-time HMM only registers state changes at \( t_1, \ldots, t_N \) by definition. The TOA gap defines the uncertainty on the estimated glitch epoch, once the HMM detects a glitch. Note that \( \xi(t) \) corresponds to a fluctuating torque derivative, whereas theories of timing noise often invoke a fluctuating torque (Cordes 1980; Cordes & Downs 1985; Melatos & Peralta 2010; Melatos & Link 2014) as well as frequency and phase fluctuations (Cordes 1980). The distinction is unimportant in many HMM tracking problems. The degree to which it matters when doing model selection, as in this paper, is tested empirically in Section 6.

The forward Fokker–Planck equation corresponding to Equation (9) can be solved to find the PDF of \( q(t_n) \) given \( q(t_{n-1}) \) (Gardiner 1994), as required by Equation (3). The result, derived in Appendix B, is

\[
A_{q(t_n)q(t_{n-1})} = (2\pi)^{-1/2} \left| \Sigma \right|^{-1/2} N_G^{-1} \exp \left\{ -[q(t_n) - \mu]^T \Sigma^{-1} [q(t_n) - \mu] \right\},
\]

where the superscript \( T \) denotes the matrix transpose, the secular evolution is described by the mean vector \( \mu = (\mu_f, \mu_g) \), with

\[
\mu_f = f(t_{n-1}) + x_n \hat{f}(t_{n-1}) + g(t_{n-1}) (\Delta f_p + x_n \hat{\Delta f}_p),
\]

and

\[
\mu_g = \hat{f}(t_{n-1}) + g(t_{n-1}) \Delta g_p,
\]

the dispersion is described by the covariance matrix,

\[
\Sigma = \sigma^2 \begin{pmatrix}
\frac{x_n^3}{3} & \frac{x_n^2}{2} \\
\frac{x_n^2}{2} & x_n
\end{pmatrix},
\]

and \( \Sigma^{-1} \) is the matrix inverse of \( \Sigma \). In Equation (10), \( G = G[g(t_{n-1})] \) denotes the set of jump pairs \( (\Delta f_p, \Delta g_p) \) searched by the HMM at step \( t_{n-1} \), and \( N_G \) is the cardinality of \( G \). If a glitch does not occur, we have \( g(t_{n-1}) = 0, \Delta f_p = 0, \Delta g_p = 0, \) and \( N_G = 1 \). If a glitch does occur, \( \Delta f_p \) and \( \Delta g_p \) are constrained to lie within the DOI defined by astrophysical priors, e.g., historical glitch observations, and \( N_G \) is determined by the grid resolution within the DOI (see Section 3.5).

Figure 2 displays a sample of \( A_{q(t_n)q(t_{n-1})} \) contours in the \( f(t_n) - \hat{f}(t_n) \) plane for a representative choice of \( q(t_{n-1}) \) (centered on the red dot in each panel) and two values each of \( x_n \) and \( \sigma \). Every cross-centered ellipse in the figure corresponds to one term in Equation (10), i.e., one choice of \( \Delta f_p \) and \( \Delta g_p \) in \( G \) and the DOI. The principal axes of the ellipses are determined by \( x_n \) through the covariance matrix \( \Sigma \), as can be seen by comparing the left \( (x_n = 1 \times 10^5 \text{ s}) \) and right top \( (x_n = 2 \times 10^5 \text{ s}) \) panels. The ellipse marked with a red diamond corresponds to no glitch, i.e., \( g(t_{n-1}) = 0 \); the other ellipses have \( g(t_{n-1}) = 1 \). The DOI is drawn artificially small, so that the reader can see its boundaries within the figure while still making out the ellipses individually; in practice, one would typically expect \( \gtrsim 10^5 \) ellipses within the frame of Figure 2. For \( \sigma \) relatively low, as in the left panel, the ellipses are narrow and nearly disjoint. For \( \sigma \) relatively high, as in the right bottom panel, the ellipses broaden and overlap; \( \sigma \) increases five-fold in passing from the left to the right bottom panel. In the high-\( \sigma \) regime, \( A_{q(t_n)q(t_{n-1})} \) can be approximated as uniform across the DOI for \( g(t_{n-1}) = 1 \), with all glitch terms contributing equally to the sum in Equation (10), while the \( g(t_{n-1}) = 0 \) term stands apart.

The Boolean component \( g(t) \) of \( q(t) \) does not appear explicitly in Equation (9); in other words, we do not track it. This is partly because its true evolution is unknown. Are glitches a Poisson process, for instance, and what is the rate? Empirically, some pulsars show Poisson-like glitch activity, but others do not (Melatos et al. 2008; Espinoza et al. 2011; Howitt et al. 2018; Fuentes et al. 2019). Moreover, it is inefficient computationally to track \( g(t) \). Historical glitch data imply \( g(t) = 0 \) most of the time, with one glitch being discovered among every \( \gtrsim 10^5 \) TOAs, at least for glitches of a size that traditional timing methods can resolve (Janssen & Stappers 2006). In this paper, therefore, we do not assume anything about the distribution of glitch waiting times. Instead, we incorporate \( g(t) \) into the Bayesian model selection procedure described in Section 4. We run the HMM for a glitchless model with \( g(t_n) = 0 \) and a single-glitch model with \( g(t_n) = \delta_{nk} \) for fixed \( k (1 \leq n \leq N_T) \) and compute the odds ratio to test for the existence of a glitch at \( t_k \). We then repeat the exercise \( N_T \) times for \( 1 \leq k \leq N_T \). A full discussion of the procedure is given in Section 4 and Appendix A.

Quasixponential post-glitch recoveries do not feature in the hidden-state evolution described by Equations (9)–(13), even though they are observed in reality. In this paper, we include post-glitch recoveries in the synthetic data generated according to Section 5 and show that the HMM performs well at finding synthetic glitches with recoveries, even though the recoveries are not built into \( A_{q(t)} \). Of course, \( A_{q(t)} \) can be extended to include recoveries, at the expense of introducing at least two extra parameters into the HMM model, viz. the recovery fraction and recovery timescale, and weakening the Markov approximation. Empirically, some pulsars show Poisson-like glitch activity, but others do not (Melatos et al. 2008; Espinoza et al. 2011; Howitt et al. 2018; Fuentes et al. 2019). Moreover, it is inefficient computationally to track \( g(t) \). Historical glitch data imply \( g(t) = 0 \) most of the time, with one glitch being discovered among every \( \gtrsim 10^5 \) TOAs, at least for glitches of a size that traditional timing methods can resolve (Janssen & Stappers 2006). In this paper, therefore, we do not assume anything about the distribution of glitch waiting times. Instead, we incorporate \( g(t) \) into the Bayesian model selection procedure described in Section 4. We run the HMM for a glitchless model with \( g(t_n) = 0 \) and a single-glitch model with \( g(t_n) = \delta_{nk} \) for fixed \( k (1 \leq n \leq N_T) \) and compute the odds ratio to test for the existence of a glitch at \( t_k \). We then repeat the exercise \( N_T \) times for \( 1 \leq k \leq N_T \). A full discussion of the procedure is given in Section 4 and Appendix A.

3.5. Grid Resolution and DOI

The DOI relevant to a particular pulsar is the region in the \( f - \hat{f} \) plane that contains all possible hidden-state sequences \( Q_1 : N_T \) consistent with the observed TOAs. The set of hidden states is constructed by dividing the DOI into a grid, whose spacing is chosen to resolve essential features like electromagnetic spin down, timing noise, and glitches (see

\[ A_{q_{n-1}q_n} = A_{q_0q_1} = 1/3 \text{ successfully tracks various complicated random walks in gravitational wave applications (Suvorova et al. 2016); see Bayley et al. (2019).} \]

\[ 10 \text{ It is possible that the glitches observed to date, with } \Delta f_p \gtrsim 10^{-10}\text{f}, \text{ represent the “tip of the iceberg,” and there exists a (e.g., power-law) population of microglitches below the resolution limit of current experiments (Melatos et al. 2008; Onuchukwu & Chukwade 2016). Indeed, it has been argued that microglitches collectively add up to produce timing noise (D’Alessandro et al. 1995). On the other hand, there is evidence that the lower cutoff of the glitch size PDF is resolved observationally in PSR J0534+2200 (Espinoza et al. 2014). The existence of microglitches remains an open question at the time of writing.} \]
Section 3.2. In Appendix C, we offer one practical recipe for gridding the DOI. It is not unique; the reader is encouraged to modify it, as the experiment demands. We distinguish carefully between the DOI and the set \( G \) in Equation (10). The DOI encompasses the trajectories of the HMM, starting from the subset of the \( f - \dot{f} \) plane covered by the uniform prior (see Section 3.2). It is chosen at the outset so that it does not exclude any admissible HMM trajectory consistent with the observed TOAs and the phase model in Section 3.4. In contrast, \( G \) defines the set of glitch-related jumps in \( q(t) \) consistent with Section 3.4 and the requirement that \( q(t) \) stays within the DOI at all times. It is updated at each \( t_n \) and depends on \( q(t_{n-1}) \) via Equation (10). Appendix C describes how discretization affects \( G \) and modifies the Equations (7) and (8) for the emission probability.

4. Glitch Detection by Bayesian Model Selection

Once the phase tracker in Section 3 is implemented, the task of discovering a glitch reduces to comparing, given the data, the probability of a phase model with one or more glitches against the probability of a glitchless phase model. From a Bayesian perspective, the comparison reduces to calculating the evidence ratio (or marginal likelihood ratio) of the competing models. In Section 4.1 and Appendix A, we describe how to calculate the evidence ratio using the HMM forward algorithm. In Section 4.2, we generalize the evidence ratio calculation to multiple glitches. In Section 4.3 and Appendix A, we describe how to infer the optimal ephemeris using the HMM forward–backward algorithm, once the preferred model (the one with the highest evidence ratio) is identified. The preferred model may or may not contain a glitch. In Section 4.4, we present a preliminary survey of the computational cost. Finally, for the sake of completeness, we outline briefly in Appendix D a related approach to discovering glitches, known as a jump Markov model, and explain why it is not used here.

The HMM does not prefer a particular physical model of glitches. Any physical mechanism that conforms to the idealized transition probability, Equations (10)–(13), falls within the ambit of the HMM. Equations (10)–(13) take a generic form and are motivated observationally, so they automatically embrace many microphysical mechanisms, which have been developed to explain observed glitch activity, including superfluid vortex avalanches (Anderson & Itoh 1975; Warszawski & Melatos 2011), starquakes (Middleditch et al. 2006; Chugunov & Horowitz 2010), and hydrodynamic instabilities (Glampedakis & Andersson 2009); see Haskell & Melatos (2015) for a recent review. Equations (10)–(13) also embrace many microphysics-agnostic metamodels, which have been developed to make falsifiable predictions about long-term glitch statistics (Fulgenzi et al. 2017; Melatos et al. 2018; Carlin & Melatos 2019). In this sense, the HMM is robust toward physical mechanisms in the literature; it accommodates all the main classes. By the same token, it cannot discriminate between the classes; that is not its function—it is a glitch detector, not a sieve for physical mechanisms. In what follows,
the term “model” refers to a sequence \( g(t_0), \ldots, g(t_{N_g-1}) \) admissible by the Markov process, Equations (10)–(13), not a codification of a physical mechanism. The sequence preferred by the data is the one with the highest evidence ratio, as noted above and in Section 4.1. The reader is encouraged to experiment with alternatives to Equations (10)–(13) and explore their effect on glitch detection.

4.1. Model Evidence

Let \( M_0 \) denote the model where no glitch occurs in the interval \( t_0 \leq t \leq t_{N_g} \), i.e., we have \( g(t_{n-1}) = 0 \) for all \( 1 \leq n \leq N_g \). Let \( M_1(k) \) denote the model where one glitch occurs in the interval \( t_{k-1} \leq t \leq t_k \), i.e., we have \( g(t_{n-1}) = \delta_{nk} \) for all \( 1 \leq n \leq N_g \) \( (\delta_{nk} \text{ is the Kronecker delta symbol}) \). Let \( M_2(k, l) \) denote the model where one glitch occurs in the interval \( t_{k-1} \leq t \leq t_k \) and another glitch occurs in the nonoverlapping interval \( t_{l-1} \leq t \leq t_l \) i.e., we have \( g(t_{n-1}) = \delta_{nk} + \delta_{nl} \) with \( k \neq l \). In the tests in this paper, we consider a maximum of one glitch in the interval \( t_0 \leq t \leq t_{N_g} \), except in the worked example involving PSR J0835–4510 in Section 7, where we briefly consider a maximum of two glitches. In practice, when analyzing real data, one can generalize the model family to an arbitrary number of glitches using a greedy hierarchical algorithm (Suvorova et al., 2018), discussed in Section 4.2. Alternatively, one can subdivide the data into multiple segments, each of which is likely to contain one glitch at most, based on the history or the outcome of a preliminary TEMPO2 fit. The exact subdivision is left to the analyst’s discretion; e.g., for PSR J0534+2200 and PSR J0537–6910, one might choose segments of \( \approx 1 \) yr and \( \approx 0.3 \) yr, respectively.

We can compare the relative plausibility of two models by calculating their evidence ratio or Bayes factor. The evidence for a model \( M \) is defined as the probability \( \Pr(O_{1:N_g} | M) \) of measuring the data \( O_{1:N_g} \) given \( M \).\(^{11}\) In the HMM context, \( \Pr(O_{1:N_g} | M) \) equals the probability of measuring \( O_{1:N_g} \) given a hidden-state sequence \( Q_{1:N_g} \), multiplied by the probability of \( Q_{1:N_g} \), marginalized over all admissible sequences:

\[
\Pr(O_{1:N_g} | M) = \sum_{Q_{1:N_g}} \Pr(O_{1:N_g}|Q_{1:N_g}, M) \Pr(Q_{1:N_g}, M). \tag{14}
\]

There exist \( N_{Q}^{N_g} \) possible sequences \( Q_{1:N_g} \) in general, but they all pass through the same set of \( N_Q \) states at each HMM step. Therefore, the sum in Equation (14) can be computed efficiently from partial sums accounting for the \( N_Q \) possible transitions at each step. Appendix A explains how to do this using the HMM forward algorithm (Rabiner 1989), which calculates \( \Pr(q(t_i) = q_j, O_{1:n+1}|M) \) from \( \Pr(q(t_i) = q_j, O_{1:n}|M) \) by induction for \( 1 \leq i, j \leq N_Q \), with

\[
\Pr(q(t_{n+1}) = q_j, O_{1:n+1}|M) = L_{o(t_{n+1})} \sum_{q_j} A_{q_j q_j} \Pr(q(t_{n}) = q_j, O_{1:n}|M). \tag{15}
\]

The pseudocode for the HMM forward algorithm is presented in Appendix A.\(^{12}\)

If the Bayes factor exceeds a threshold, the model in the numerator is preferred. There is no unique way to set the threshold. On the popular Jeffreys scale (Jeffreys 1998), a Bayes factor above 10 counts as “strong” evidence, and a Bayes factor between \( 10^{1/2} \) and 10 counts as “substantial.” In this paper, we arbitrarily regard a glitch as having occurred in the interval \( t_{k-1} \leq t \leq t_k \) if we obtain \( \Pr(O_{1:N_g}|M_k) > 10^{1/2} \Pr(O_{1:N_g}|M_0) \). Tests with arbitrarily higher thresholds (up to 10\(^3 \)), which counts as “decisive” on the Jeffreys scale) yield qualitatively similar results.

4.2. Multiple Glitches

To search for multiple glitches in data which are not subdivided, Suvorova et al. (2018) proposed a greedy hierarchical algorithm, which works as follows. For \( m = 1, 2, \ldots \) in increasing order, construct the sequence of Bayes factors

\[
K_m(k) = \frac{\Pr(O_{1:N_g}|M_m(k_1, \ldots, k_{m-1}, k))}{\Pr(O_{1:N_g}|M_{m-1}(k_1, \ldots, k_{m-1}))}, \tag{16}
\]

where \( k_m^* \) indexes the TOA corresponding to the \( m \)’th detected glitch, and the no-glitch model \( M_0 \) has no arguments. In other words, \( K_m(k) \) evaluates, as a function of \( k \), the evidence for a model with \( m \) glitches at \( \{k_1^*, \ldots, k_m^*, k\} \) compared to the evidence for a model with \( m – 1 \) glitches at \( \{k_1^*, \ldots, k_{m-1}^*\} \). Starting from \( m = 1 \), if \( K_m(k) \) exceeds the user-selected threshold for some \( k \) (e.g., \( K_m(k) > 10^{1/2} \) for some \( k \)), we set \( k_m^* = \arg \max_k K_m(k) \) and increment \( m \). The iteration halts when we obtain \( K_m(k) < 10^{1/2} \) for all \( k \).

4.3. Optimal Ephemeris

Once the preferred model is identified out of the set \( \{M_0, M_1(k), M_2(k, l), \ldots\} \), the next step is to compute the ephemeris which fits the data best, given the preferred model. There is no unique definition of “best,” as discussed in Section 3.1 and Appendix A. In this paper, we stipulate that the optimal ephemeris is the one constructed from the most-probable state at each HMM step, given by

\[
\hat{q}(t_n) = \arg \max_{q(t_n)} \sum_{\tilde{q}(t_{n+1})} \prod_{\tilde{t} = t_{n+1}}^{t_n} L_{o(t_{\tilde{t}})\hat{q}(t_{\tilde{t}})} \times \prod_{\tilde{t} = t_{n+1}}^{t_{n+1}} A_{\hat{q}(t_{\tilde{t}})\hat{q}(t_{\tilde{t}})} L_{o(t_{\tilde{t}})\hat{q}(t_{\tilde{t}})} \times \prod_{\tilde{t} = t_{n+1}}^{t_{n+1}} A_{\hat{q}(t_{\tilde{t}})\hat{q}(t_{\tilde{t}})} L_{o(t_{\tilde{t}})\hat{q}(t_{\tilde{t}})}. \tag{17}
\]

Equation (17) takes sequences of the form \( \{Q_{1:n-1}, q(t_n) = q_i, Q_{n+1:N_g}\} \), calculates their probabilities according to Equation (3) for \( q \) fixed, sums the probabilities over \( Q_{1:n-1} \) and \( Q_{n+1:N_g} \) then maximizes over \( 1 \leq i \leq N_Q \). It can be evaluated efficiently by the HMM forward–backward algorithm, whose pseudocode is presented in Appendix A. The approach maximizes the number of most-probable states in the

\(^{11}\) The definition of the evidence depends on the form of Bayes’s theorem under consideration. If we consider \( \Pr(Q_{1:n} | O_{1:n}, M) = \Pr(O_{1:n}|Q_{1:n}, M) \Pr(Q_{1:n}|M) \Pr(O_{1:n}|M) \) for fixed \( M \), then \( \Pr(O_{1:n}|Q_{1:n}, M) \) is the likelihood and \( \Pr(O_{1:n}|M) \) in Equation (14) is the evidence, as in this paper. If we consider \( \Pr(M | O_{1:n}) = \Pr(O_{1:n}|M) \Pr(M) / \Pr(O_{1:n}) \) after marginalizing over \( Q_{1:n} \), then \( \Pr(O_{1:n}|M) \) is the likelihood, and \( \Pr(O_{1:n}) = \sum_M \Pr(O_{1:n}|M) \Pr(M) \) is the evidence.

\(^{12}\) To increase accuracy and avoid arithmetic underflow when computing products with many factors, such as Equation (3), we take advantage of the log-sum-exp approximation (Calafate & el Ghaoui 2014).
ephemeris. It also generates the PDF of \( q(t_n) \) automatically as a by-product, allowing one to examine the states in the neighborhood of the peak, to see how much \( \hat{q}(t_n) \) stands out. The results can be checked for broad consistency against \( Q_{t_n} \) (see Section 3.1). The subtle difference between \( Q_{t_n}^* \) and \( \{\hat{q}(t_1), \ldots, \hat{q}(t_N)\} \), along with the Viterbi algorithm which computes the former sequence efficiently, is described in Appendix A.

4.4. Computational Cost

From a practical standpoint, the computational cost of the glitch detector depends on what astrophysical experiment is being attempted. For example, a search for three glitches in a stretch of data with the greedy hierarchical algorithm in Section 4.2 involves passing the data through the HMM 3\( N_F \) times. \( N_F \) times to calculate \( K_1(k) \) for model \( M_1(k) \) and \( 1 \leq k \leq N_F \). \( N_F \) times to calculate \( K_2(k) \) for model \( M_2(k^*, k) \) and \( 1 \leq k \leq N_F \), and \( N_F \) times to calculate \( K_3(k) \) for model \( M_3(k^*, k^*, k) \) and \( 1 \leq k \leq N_F \). In order to embrace a variety of experiments, we present below a rough cost estimate for the key computational step that is common to all of them: a single pass of the HMM forward algorithm through the full data to calculate one Bayes factor, e.g., \( K_1(k) \). The HMM backward algorithm, which calculates the associated optimal ephemeris, costs roughly the same.

The cost of the HMM forward algorithm is of order \( N_G^2 N_F \), as described in Appendix A. Importantly, it does not depend on the data or the model parameters, with one exception which we discuss below. The HMM addresses each of the \( N_G^2 N_F \) links in the HMM trellis once without discretion and without reference to any tolerances; iterative convergence does not play a role.\(^{13}\) Preliminary benchmarking tests, characteristic of the computations in Section 6 and done on a consumer-grade, quad-core Intel CPU with 2.7 GHz clock speed, indicate that the run time for one pass of the HMM forward algorithm scales approximately as

\[
T_{CPU} = 9 \left( \frac{N_f}{10^3} \right)^2 \left( \frac{N_f}{10} \right)^2 \left( \frac{N_f}{10^2} \right)^2 \text{s},
\]

where \( N_f \) and \( N_f \) are the number of \( f \) and \( \dot{f} \) bins in the DOI, respectively (see Appendix C). Hence, from Equation (18), a typical experiment searching for a single glitch among \( N_F \) TOAs takes \( N_F T_{CPU} \approx 9 \times 10^2 (N_f/10)^3 (N_f/10) (N_f/10^2)^2 \text{s} \), independent of \( \sigma \) and \( x_n \).

In the transition probability \( A_{f_0}^{f_1} (t_{n-1}, \ldots, t_n) \) in Equation (10), each Gaussian term in the sum over \( G \) extends formally across the whole DOI. To accelerate the computation, we truncate \( A_{f_0}^{f_1} (t_{n-1}, \ldots, t_n) \) at three standard deviations along the \( f \) axis. If the truncated \( A_{f_0}^{f_1} (t_{n-1}, \ldots, t_n) \) spans multiple frequency bins, the computational cost scales according to Equation (18). If the truncated \( A_{f_0}^{f_1} (t_{n-1}, \ldots, t_n) \) fits wholly within one frequency bin, the scaling with \( N_f \) is linear instead, and one finds \( T_{CPU} \approx 0.8 (N_f/10)^3 (N_f/10) (N_f/10^2)^2 \text{s} \). The latter scaling prevails over Equation (18), when \( \approx 3\sigma (x_n)^3/2 \) drops below the frequency bin width. The latter dependence on \( \sigma \) and \( x_n \) is the exception foreshadowed in the previous paragraph. It stems from an implementation trick and is not fundamental to the HMM forward algorithm.

\(^{13}\) In other algorithms like Markov Chain Monte Carlo samplers, convergence is an issue, and the run time depends on the shape of the posterior distribution, the form of the proposal function, and the tolerance.

Further study of the computational cost is postponed to future work as it raises the role of graphics processing units (GPUs), a topic outside the scope of this paper. GPUs have proved effective in accelerating HMM-based searches for continuous gravitational wave signals with the Laser Interferometer Gravitational Wave Observatory (Abbott et al. 2019; Dunn et al. 2020). Acceleration by a factor of \( \sim 40 \) is achieved in the latter references.

5. Synthetic Data

We now quantify the performance of the HMM systematically through a suite of Monte Carlo tests based on synthetic data. Many valid recipes exist to generate the synthetic data; the physical origin and hence the statistics of the fluctuating torque are unknown from first principles and cannot be inferred uniquely from pulsar timing noise studies (Cordes & Downs 1985; Hobbs et al. 2004). In this paper, we take an empirical approach and generate data consistent with glitch templates derived from traditional pulsar timing studies (McCulloch et al. 1987; Wong et al. 2001), without seeking to relate the output to an underlying physical model. The TOAs are sampled according to a Poisson observing process for simplicity, as described in Appendix E, to ensure that they do not coincide artificially with a glitch, but any reasonable sampling algorithm (e.g., uniform spacing) does just as well. When analyzing real data, the TOAs are referred first to the solar system barycenter using standard methods (Taylor 1992). We do not consider the orbital motion of binary pulsars in this paper.

Let \( T > 0 \) be an epoch when a glitch occurs. Consider a time interval containing \( t = T \), which is short enough that spin wandering and the secular component of \( \dot{f} \) can be ignored temporarily, i.e., \( \dot{f} = 0 \). Traditional pulsar timing studies based on empirical fits to the data propose that the system evolves according to (McCulloch et al. 1987; Wong et al. 2001)

\[
f(t) = f(0) + \dot{f}(0)t + \Delta f_i + \Delta f_i(t - T) + \Delta f_i \exp[-(t - T)/\tau] H(t - T),
\]

where \( H(...) \) symbolizes the Heaviside step function, and \( \Delta f_i \) and \( \tau \) are the amplitude and \( e^{-1} \) recovery timescale, respectively, of the transient component of the frequency jump following the glitch. Now suppose that the time interval is long enough, that spin wandering cannot be ignored. Then, Equation (19) still describes the deterministic evolution before and after the glitch (ignoring \( \dot{f} \)); see Section 3.1) but with a random walk added. The random walk can be generated in many valid ways. In Appendix E, we present and justify a systematic recipe, which involves solving a system of two stochastic differential equations, one of which (see Equation (E1)) takes the form

\[
\frac{df}{dt} = \text{deterministic terms} + \zeta(t),
\]

with

\[
\langle \zeta(t)\zeta(t') \rangle = \sigma_{TN}^2 \delta(t - t').
\]

In Equations (20) and (21), the deterministic terms model secular spin down and glitch-related jumps and recoveries, \( \zeta(t) \) is a zero-mean, white-noise torque, \( \delta(...) \) is the Dirac delta function, and \( \sigma_{TN} \) is the timing noise amplitude (units:
Glitch detection by model selection involves asking if the Bayes factor relating two models exceeds a threshold. The threshold determines the false-alarm probability, \( P_{\text{fa}} \). Given \( P_{\text{fa}} \), the detection probability, \( P_{\text{d}} \), can be expressed as a function of the signal parameters, e.g., glitch size \( \Delta f_p \). One can set the threshold by fiat, as in Section 4.1, and infer \( P_{\text{fa}} \) or vice versa.

In this section, we evaluate the HMM’s performance by constructing ROC curves (\( P_{\text{d}} \) versus \( P_{\text{fa}} \), with signal parameters fixed) and detection probability curves (\( P_{\text{d}} \) versus one or more signal parameters, with \( P_{\text{fa}} \) fixed) for a range of representative values of the intrinsic astrophysical and measurement noises in the system (Section 6.1), secular spin-down parameters, e.g., \( \dot{f}_L \) and \( \ddot{f}_L \) (Section 6.2), and glitch parameters, e.g., size and recovery timescale (Section 6.3). A short, preliminary analysis of the impact on performance of the observational scheduling strategy, e.g., mean inter-TOA interval, is presented in Appendix G. Optimizing the observational schedule is a subtle exercise, which we will take up more fully in future work.

6. ROC Curves

The detectability of a glitch is connected to its size relative to the noise, which comes in two flavors. A glitch may be drowned out by TOA measurement errors; if \( \kappa \) is too small, the peaks in \( L_{\mu_{\ell}}(\kappa) \) in Equation (7) blur together. A glitch may also be obscured by astrophysical timing noise, if \( \Delta f_p \) is relatively small, \( \sigma_{\text{TN}} \) is relatively large, and there are long delays between TOAs. Conversely, a random walk with large \( \sigma_{\text{TN}} \) may masquerade as a step \( \Delta f_p \neq 0 \) during a subset of TOAs, triggering a false alarm.

Figure 3 illustrates how the task of detection is affected by \( \sigma_{\text{TOA}} \). Measurement errors enter through \( L_{\mu_{\ell}}(\kappa) \), which depends on \( \sigma_{\text{TOA}} \) through \( \kappa \) as defined by Equations (7) and (8) or Equation (C3) after gridding. The top panel in Figure 3 displays ROC curves for three values of \( \kappa \) ranging from \( \kappa^* \) to \( 10\kappa^* \), where \( \kappa^* \) is the fiducial value calculated according to the recipe in Section 3.3 and Appendix C. The results are encouraging. For \( \kappa = \kappa^* \), we obtain \( P_{\text{d}} \geq 0.8 \) for \( P_{\text{fa}} \geq 10^{-2} \) and \( P_{\text{d}} \geq 0.9 \) for \( P_{\text{fa}} \geq 10^{-1} \). The HMM’s performance varies mildly with \( \kappa \), e.g., \( P_{\text{d}} \) drops by \( \leq 0.15 \) across the ROC curve for \( \kappa = 10\kappa^* \); it does not depend sensitively on how one estimates \( \kappa^* \) from \( \sigma_{\text{TOA}} \). The bottom panel summarizes the behavior in a practical fashion by graphing \( P_{\text{d}} \) and \( P_{\text{fa}} \) for parameters matching the penultimate column in Table 1, with \( \kappa = \kappa^* \) updated according to Section 3.3 and Appendix C. The Bayes factor threshold is kept at \( 10^{1/2} \) and maintains \( P_{\text{fa}} \approx 10^{-2} \) across the plotted \( \sigma_{\text{TOA}} \) range; false alarms are not sensitive to \( \sigma_{\text{TOA}} \) when \( \kappa \) is updated. The detection probability drops off as \( \sigma_{\text{TOA}} \) increases, with \( P_{\text{d}} \leq 0.9 \) for \( \sigma_{\text{TOA}} \gtrsim 3 \times 10^{-5} \text{s} \). Roughly speaking, one requires \( \sigma_{\text{TOA}} \propto \Delta f_p \) to maintain a desired \( P_{\text{d}} \) value.

The HMM also contends with astrophysical timing noise. An important practical issue is how to select the HMM parameter \( \sigma \) for a particular astrophysical target. Again, there is no unique prescription, and the final conclusions concerning glitch detection are conditional on the choice made.\(^\text{14}\) A useful rule of thumb is to match the rms phase residual \( \langle b \phi(t_n) \rangle^2 \rangle^{1/2} \) accumulated by the random walk in the HMM, derived by integrating Equation (9), with the phase residual accumulated by the timing noise in the pulsar, derived by integrating Equations (20) and (21). The latter quantities are of order \( \sigma \langle x_n \rangle^{5/2} \) and \( \sigma_{\text{TN}} \langle x_n \rangle^{3/2} \), respectively, when integrated over the mean TOA gap, \( \langle x_n \rangle \), which implies \( \sigma \approx \langle x_n \rangle^{-1} \sigma_{\text{TN}} = \sigma^* \). In

\(^{14}\) The same applies to traditional timing methods or pulse domain analysis, where the conclusions concerning glitch detection are conditional on the phase model, e.g., Taylor expansion.
practice, \( \langle x_n \rangle \) is dominated by the TOA intervals between rather than within observation sessions.

Figure 4 illustrates how the HMM’s performance varies, as \( \sigma \) moves away from \( \sigma^* \). The top panel displays five ROC curves for \( 10^{-1} \leq \sigma/\sigma^* \leq 10^{1} \). For \( \sigma = \sigma^* \), we obtain \( P_f \geq 0.87 \) for \( P_d \geq 10^{-2} \) and \( P_d \geq 0.95 \) for \( P_{sa} \geq 10^{-1} \). The results do not change much near the optimum, with \( P_d \) changing by \( \lesssim 0.1 \) over the range \( 10^{-1} \leq \sigma/\sigma^* \leq 10^{1} \) for \( P_{sa} \geq 10^{-2} \). The bottom panel in Figure 4 summarizes the results in practical, observation-ready terms. Adjusting \( \sigma = \sigma^* \) as a function of \( \sigma_{TN} \) and \( \langle x_n \rangle \) as in the previous paragraph, we find that the detection probability stays roughly constant, as \( \sigma_{TN} \) increases, with \( P_d \gtrsim 0.9 \) (and \( P_{ia} \approx 10^{-2} \)) for \( \sigma_{TN} \leq 10^{-12} \). The blue curve; see Section 3.4, respectively, when integrated over the mean TOA gap, where.

Figure 4. HMM glitch detector performance as a function of HMM noise parameter \( \sigma \) and timing noise amplitude \( \sigma_{TN} \). (Top panel) ROC curve \( (P_d \) vs. \( P_f ) \) for five values of \( \sigma/\sigma^* \) in the range \( 10^{-1} \leq \sigma/\sigma^* \leq 10^{1} \), with \( \sigma^* \) set according to the recipe in Section 6.1 as a function of \( \sigma_{TN} \) and \( \langle x_n \rangle \). (Bottom panel) Detection probability \( P_d \) (blue curve) and false-alarm probability \( P_{fa} \) (red curve) vs. \( \sigma_{TN} \) (units: Hz s^{-1/2}), with the Bayes factor threshold held at 10^{-12}. The curves are restricted to \( \sigma_{TN} \leq 3 \times 10^{-12} \) Hz s^{-1/2} by the phase wandering mismatch described in Section 6.1. Parameters: as in Figure 3, except with \( 10^{-15} \leq \sigma_{TN}/(1 \text{ Hz s}^{-1/2}) \leq 10^{-11} \). Realizations: \( 1.5 \times 10^3 \) per ROC curve.

\( \eta \) is the grid spacing in \( \hat{f} \), implying \( \sigma \geq \sigma_{min} = \langle x_n \rangle^{-1/2} \eta \) and hence \( \sigma^* = \max(\sigma_{min}, \langle x_n \rangle^{-1} \sigma_{TN}) \). Additionally, the HMM is sensitive to the mismatch in phase wandering between the data (e.g., white noise in \( \hat{f} \); see Section 5) and the transition probabilities (white noise in \( \hat{f} \); see Section 3.4). The mean-square phase residuals arising from the two processes are given by \( \sigma_{TN}^2 \eta^2 \) and \( \sigma^2 \eta^2 \), respectively, when integrated over a specific TOA gap \( x_n \). Substituting the rule of thumb \( \sigma \approx \langle x_n \rangle^{-1} \sigma_{TN} \), we calculate the frequency mismatch to be \( \approx (\langle x_n \rangle^{-1} x_n - 1) \langle x_n \rangle^{1/2} \sigma_{TN} \), which exceeds the frequency bin size \( \eta_f \) for certain combinations of \( x_n \), \( \langle x_n \rangle \), and \( \sigma_{TN} \). For the parameters in the penultimate column of Table 1, with \( \eta_f = 6 \times 10^{-10} \) Hz, the effect becomes significant for \( \sigma_{TN} \approx 3 \times 10^{-14} \) Hz s^{-1/2}, which corresponds to the cutoff in the curves of the bottom panel of Figure 4.

### Table 1
Parameters Used to Generate Synthetic Data to Test the HMM, Classified as Astrophysical and Measurement Noise, Scheduling of Observations, Secular Spin Down, and Glitch Parameters

| Quantity | Symbol | Units | Min | Typical | Max |
|----------|--------|-------|-----|---------|-----|
| Noise    |        |       |     |         |     |
| Timing noise amplitude | \( \sigma_{TN} \) | Hz s^{-1/2} | \( 10^{-15} \) | \( 10^{-13} \) | \( 10^{-11} \) |
| TOA measurement uncertainty | \( \sigma_{TOA} \) | s | \( 10^{-6} \) | \( 10^{-5} \) | \( 10^{-3} \) |
| Scheduling |        |       |     |         |     |
| Mean waiting time | \( \langle x_n \rangle \) | days | \( 10^{-2} \) | 13 | 116 |
| Number of sessions | ... | ... | 5 | 51 | 10^2 |
| Secular spin down |        |       |     |         |     |
| Frequency | \( f_{ls} \) | Hz | \( 10^0 \) | \( 5.435 \times 10^2 \) | \( 10^3 \) |
| Frequency derivative | \( \dot{f}_{ls} \) | Hz s^{-1} | \( 10^{-15} \) | \( 10^{-15} \) | \( 10^{-11} \) |
| Glitch |        |       |     |         |     |
| Permanent frequency jump | \( \Delta f_p \) | Hz | \( 10^{-10} \) | \( 10^{-8} \) | \( 10^{-7} \) |
| Transient frequency jump | \( \Delta f_t \) | Hz | 0 | 0 | \( 10^{-8} \) |
| Recovery timescale | \( \tau \) | s | \( 10^3 \) | \( 10^{10} \) | \( 10^{10} \) |
| Permanent frequency derivative jump | \( \Delta f_p \) | Hz s^{-1} | \( 10^{-12} \) | \( 10^{-15} \) | \( 10^{-15} \) |

6.2. Secular Spin Down

Glitch detection is fundamentally an exercise in tracking fluctuations around the secular spin-down trend and
distinguishing statistically between a continuous random walk (timing noise) and discontinuous jumps (glitches). One therefore expects the HMM’s performance to be approximately independent of the secular trend itself, i.e., \( \tilde{f}_{LS} \) and \( \dot{f}_{LS} \), as long as \( \kappa \propto (\sigma_{TOA} \tilde{f}_{LS})^{-2} \) is held fixed,\(^\text{15}\) while the spin-down parameters vary. Figure 5 confirms that \( P_d \) stays approximately constant for \( 1 \leq \tilde{f}_{LS} / (1 \text{ Hz}) \leq 20 \) and drops away for \( \tilde{f}_{LS} \geq 20 \text{ Hz} \) for the parameters in Table 1, because \( \kappa \) decreases with \( \tilde{f}_{LS} \), when \( \sigma_{TOA} \) is held fixed. The rollover shifts right, as \( \sigma_{TOA} \) decreases, and depends on \( x_n, \eta_p, \eta_f \); there is nothing unique about \( \tilde{f}_{LS} \geq 20 \text{ Hz} \). Figure 5 also confirms that \( P_d \) > 0.9 stays approximately constant across the plotted range \( 10^{-15} \leq \tilde{f}_{LS} / (1 \text{ Hz s}^{-1}) \leq 10^{-11} \), with \( P_d \approx 10^{-2} \). Trials indicate that, for certain parameter combinations, \( \sigma_{TOA} \) is effectively underestimated when interpreted according to Equation (8), leading to high \( K_f(\bar{k}) \) values and false alarms. As a precaution, we correct this behavior by taking \( \sigma_{TOA} \) to be \( \approx 5 \times \) the fiducial TEMPO2 value. The correction factor is set empirically; it cannot be predicted analytically at present. It is conservative, as it reduces \( P_d \) marginally (by \( \approx 10\% \)) while nullifying the spike in \( P_{fa} \).

6.3. Glitch Parameters

The size of the smallest glitch detectable by the HMM is governed chiefly by the user-selected probabilities \( P_{fa} \) and \( P_d \). In general, the permanent jump \( \Delta f_p \) is partially covariant with other glitch parameters (e.g., \( \Delta f^', \Delta f^1 \), and \( \tau \)) as well as the nonglitch parameters discussed in Sections 6.1–6.2. However, we find that \( \Delta f_p \) affects \( P_d \) more strongly than the other parameters. Figure 6 illustrates this behavior. The top-left panel shows that \( P_d \) rises steeply to \( P_d \geq 0.9 \) for \( \Delta f_p \geq 8 \times 10^{-8} \text{ Hz} \) and the parameters in the penultimate column of Table 1. The top-right panel shows that \( P_d \) is roughly independent of \( \Delta f_p \) in the range \( -10^{-12} \leq \Delta f_p / (1 \text{ Hz s}^{-1}) \leq 10^{-15} \), where \( \Delta f_p \) values of both signs are tested.

The bottom panels in Figure 6 illustrate how the HMM’s performance depends on the form and duration of the glitch recovery. In the bottom-left panel, we observe that \( P_d \) rises to \( P_d \geq 0.5 \) for \( \tau \geq 2 \times 10^6 \text{ s} \), i.e., a glitch with a slower recovery is easier to detect. The plotted example involves a substantial transient component \( \Delta f_1 = \Delta f_p \), which explains why \( P_d \) depends on \( \tau \). The phase deviation produced by \( \Delta f_1 \) relative to the glitchless model builds up during the recovery and asymptotes to a constant value \( \approx \tau \Delta f_1 \), unlike the permanent component, whose phase deviation grows indefinitely as \( \approx \tau \Delta f_p \). The bottom-right panel graphs \( P_d \) as a function of the transient fraction, \( \Delta f^_1 / (\Delta f^_1 + \Delta f_p) \), holding \( \Delta f^_1 + \Delta f_p \) and \( \tau \) fixed. We find \( P_d \leq 0.5 \) for \( \Delta f^_1 \geq 0.6(\Delta f^_1 + \Delta f_p) \). That is, when the permanent fraction drops below some value, which depends on \( \tau \), the glitch ceases to be detectable, if the transient component cannot be detected in its own right, i.e., if \( \tau \Delta f_1 \) is too low. Conversely, if \( \Delta f_p \) is large enough, the phase deviation crosses the detection threshold eventually, irrespective of \( \Delta f_1 \) and \( \tau \).

7. Representative Worked Example: PSR J0835−4510

The tests in this method paper are restricted deliberately to synthetic data, in order to quantify the performance of the HMM under controlled conditions. We look forward to applying the HMM to real, astrophysical data in the near future. As a foretaste, we analyze a publicly available subset of 490 TOAs from the regularly timed object PSR J0835−4510 from MJD 57427 to MJD 57810 (Sarkissian et al. 2017a, 2017b). The data are preprocessed to cull the closest spaced TOAs (with \( x_n \leq 8.9 \times 10^7 \text{ s} \) for definiteness), as these TOA clusters exhibit excess white noise in TEMPO2. Results are presented below for the preprocessed data, comprising \( N_T = 212 \text{ TOAs} \), after checking for consistency against the 490 original TOAs.

\(^\text{15}\) The rough proportionality \( \kappa \propto (\sigma_{TOA} \tilde{f}_{LS})^{-2} \) governs how accurately \( N_c \) can be inferred through Equations (7) and (8) before the modifications introduced by gridding (see Appendix C).
7.1. 2016 December 12 Glitch

The results of applying the HMM to PSR J0835–4510 are presented in Figure 7. The first row displays the phase residuals arising from traditional TEMPO2 fits to the data. In the left panel, where the ephemeris does not incorporate a glitch, the phase wraps violently beyond the glitch epoch. In the right panel, where the ephemeris does incorporate a glitch, the phase wraps more slowly, because we do not correct for the quasiexponential post-glitch recovery in this panel. (The correction is performed by Sarkissian et al. 2017a, 2017b.) The second row of the figure displays the logarithm of the Bayes factor, \( K_t(k) = \text{Pr}(O_t|N_t|M_t(k))/\text{Pr}(O_t|N_t|M_0) \), for \( 1 \leq k \leq N_T \) and \( \sigma = 5 \times 10^{-16} \text{ Hz s}^{-3/2} \). The value of \( \sigma \) is estimated from the TEMPO2 residuals and the gridding bound \( \sigma \geq 1.3 \times 10^{-16} \text{ Hz s}^{-3/2} \) (dominated by \( H \) in Equation (C3)) plus a conservative safety factor. The one-glitch model \( M_t(173) \) is preferred strongly over \( M_0 \) and all \( M_t(k) \) with \( k = 173 \) \( (k = 383 \text{ before preprocessing}) \). The HMM glitch epoch, \( T = 57734.54 \text{ MJD} \), approaches that obtained by traditional methods, which yield \( T = 57734.4855(4) \text{ MJD} \) (Palfreyman 2016; Sarkissian et al. 2017b; Ashton et al. 2019). The maximum Bayes factor is huge, with \( \text{ln} K_t(173) = 1.1 \times 10^3 \), a testament to the discriminating power of the HMM. The third row displays \( \hat{f} (t) \) versus \( t \), inferred using the HMM forward-backward algorithm, for \( M_0 \) (left panel) and \( M_t(173) \) (right panel). The frequency step in the right panel is clearly visible. The fourth row displays the associated phase residuals, which wrap violently for \( M_0 \) at \( t > T \) while remaining roughly constant for \( M_t(173) \). The results confirm that \( M_t(173) \) offers a good description of the 2016 December 12 event despite modeling it as a step for simplicity, without a post-glitch recovery.

In order to check the robustness of the conclusion, that \( M_t(173) \) is the preferred model, we subdivide the data set into halves and quarters and plot \( K_t(k) \) versus \( k \) in Figure 8. The results for each subdivision are color-coded according to the legend. No matter how the data are subdivided, the conclusion is the same: \( M_t(173) \) is strongly preferred over \( M_0 \) and \( M_t(k) \) with \( k = 173 \) in the data segments that include \( t_{173} \), and \( M_t(173) \) is not rivaled by a better alternative in the data segments that do not include \( t_{173} \).

How does the recovered ephemeris compare with the traditional timing solution, now that the glitch is detected? Figure 9 presents the evolution of the posterior PDF \( \gamma_t(t_n) \) before and after the glitch. The first and second rows display contours of \( \gamma_t(t_n) \) marginalized over \( \hat{f} \) and \( f \), respectively, graphed as functions of \( t_n \), together with the point-wise optimal sequences \( \hat{f} (t_n) \) and \( f(t_n) \), respectively. Both marginalized posteriors are strongly and singly peaked around the optimal sequences. The jump in \( f(t_n) \) is visible in the top row. The third and fourth rows display orthogonal cross sections taken through the posterior PDF just before \( (t_{172}; \text{ third row}) \) and after \( (t_{174}; \text{ fourth row}) \) the glitch. In the left column, where \( \gamma_{t}(t_n) \) is marginalized over \( \hat{f} \), there is an upward shift in frequency, with \( \Delta f_0 = \hat{f} (t_{174}) - \hat{f} (t_{172}) = 1.596 \times 10^{-5} \text{ Hz} \). The cross sections are narrow, spanning \( \leq 4 \text{ bins} \) before and after the glitch. In the right column, where \( \gamma_t(t_n) \) is marginalized over \( f \), there is a downward shift in the frequency derivative, with \( \Delta f_0 = \hat{f} (t_{174}) - \hat{f} (t_{172}) = -4.4 \times 10^{-13} \text{ Hz s}^{-1} \). The shift is significant in the sense that it exceeds the dispersion, which actually decreases during the event (FWHM \( \leq 8 \text{ bins} \) before; see \( \leq 2 \text{ bins} \) after). The inferred jumps agree with traditional pulsar timing methods, which give \( \Delta f_0 = 1.6044 (2) \times 10^{-5} \text{ Hz} \) and \( \Delta f_0 = -1.21 (3) \times 10^{-13} \text{ Hz s}^{-1} \) (Palfreyman 2016; Sarkissian et al. 2017b), after allowing for the fact that the HMM transition probabilities do not include the post-glitch relaxation with \( \tau = 0.96 (17) \text{ days} \). (Including the relaxation is straightforward but lies outside the scope of this paper.) All in all, the optimal sequence stands out clearly above its nearest competitors.
7.2. Additional Glitches

A systematic search for multiple glitches lies outside the scope of this paper. Nonetheless, again as a foretaste of what is feasible, we search for a second glitch in PSR J0835−4510 from MJD 57427 to MJD 57810 by applying the greedy hierarchical algorithm introduced in Section 4.2 (Suvorova et al. 2018). The analysis is presented in Appendix H. We conclude that no statistically significant second event exists, in accordance with previous analyses (Sarkissian et al. 2017a, 2017b).

We look forward to applying the HMM to more real data sets. In particular, a fuller search for glitches in PSR J0835−4510 over several decades of continuous monitoring using the greedy hierarchical algorithm introduced in Section 4.2 will be undertaken in future work; the relevant data are not currently at our disposal. If additional events are found, they can be cross-checked in many ways. One can apply the HMM to data taken with a different telescope, e.g., the higher cadence Mount Pleasant Radio Observatory for PSR J0835−4510 (Palfreyman et al. 2018), just as when checking the output of traditional timing methods. The Mount Pleasant data were analyzed recently by Bayesian methods to study the pulse-to-pulse dynamics of the 2016 December 12 glitch (Ashton et al. 2019). One can also test how the Bayes factor changes as one tunes HMM parameters like $\kappa$, $\sigma$, and $P_{\text{gl}}$; see Section 6.1 for details. It is faster to do such tests systematically with the HMM than with traditional timing methods.

8. Conclusion

In this method paper, a new, systematic scheme is presented for detecting pulsar glitches given a sequence of standard TOAs. The scheme is structured around an HMM, which tracks the evolution of the pulse frequency and its first time derivative on long (electromagnetic spin down), intermediate (timing noise), and short (glitches) timescales. The emission probability of the HMM obeys a von Mises distribution. The transition probability obeys a Gaussian distribution derived from the Fokker–Planck equation for an unbiased Wiener process. The HMM forward algorithm is used to compute and compare the Bayesian evidence for models with and without glitches.

Figure 7. HMM analysis of 212 TOAs measured for PSR J0835−4510 from MJD 57427 to MJD 57810 (Sarkissian et al. 2017a, 2017b). (First row) Phase residuals $\phi(t_n)$ vs. TOA $t_n$ computed with TEMPO2 for no-glitch (left panel) and one-glitch (right panel) models, with $T = 57734.54$ MJD, $\Delta f_p = 1.596 \times 10^{-5}$ Hz, and $\Delta f_p = -4.4 \times 10^{-11}$ Hz s$^{-1}$ in the one-glitch model. (Second row) Logarithm of the Bayes factor, $K$, vs. the TOA index, $k$, computed with the HMM using $\sigma = 5 \times 10^{-14}$ Hz s$^{-1/2}$ and the DOI $f_{\text{LS}} = 11.1868550196$ Hz, $f_{\text{LS}} = 1.55886 \times 10^{-11}$ Hz, $-5.5 \times 10^{-4} \leq (f - f_{\text{LS}})/(1$ Hz s$^{-1}) \leq 1 \times 10^{-5}$, $-2 \times 10^{-12} \leq (f - f_{\text{LS}})/(1$ Hz s$^{-1}) \leq 2 \times 10^{-12}$, and $\eta = 5.606 \times 10^{-7}$ Hz (10 bins). (Third row) Recovered frequency $f(t_n)$ vs. TOA $t_n$ for the no-glitch (left panel) and one-glitch (right panel) HMM models $M_0$ and $M_1$. The vertical dashed line marks the glitch. (Fourth row) Unsummed per-gap phase residuals $\phi(t_n)$ vs. TOA $t_n$ for the HMM forward–backward sequences in the third row.
Once the preferred model is selected, the HMM forward–backward algorithm is used to compute the associated, point-wise optimal ephemeris, composed of the most-probable hidden state $q(t_n)$ at each $t_n$ given all the observations $O_{1:N_t}$. The algorithm and testing procedure are documented in Appendices A–H for the sake of reproducibility.

Monte Carlo simulations demonstrate that the HMM detects glitches accurately in synthetic data for a range of realistic intrinsic and measurement noises ($\sigma_{TN}, \sigma_{TOA}$; see Section 6.1), secular spin-down parameters ($f_{LS}, f_{dLS}$; see Section 6.2), glitch parameters ($\Delta f_p, \Delta f_p^d, \Delta \dot{f}, \tau$; see Section 6.3), and observational schedules ($\langle x_n \rangle, N_T$; see Appendix G). The performance of the HMM, in particular the trade-off between $P_a$ and $P_d$, is quantified systematically in terms of ROC curves constructed as functions of the above parameters. Success is achieved, even though (i) the HMM approximates glitches as instantaneous steps in $f$ and $\dot{f}$ without any post-glitch recovery, and (ii) the HMM models timing noise as white noise in the torque derivative (and hence red noise in the filtered torque), an approximation which applies to some but not all pulsars (Cordes 1980; Cordes & Downs 1985) and is violated deliberately when generating the synthetic data in this paper in order to challenge the robustness of the HMM. Several trends of practical utility are identified. (i) The HMM performs stably, neither overestimating nor underestimating the number of glitches, for $\sigma \approx \langle x_n \rangle^{-1} \sigma_{TN}$, with $\sigma_{TN}$ computed from the TEMPO2 phase residuals, as described in Section 6.1. (ii) In order to detect a glitch of size $\Delta f_p$, it is recommended to schedule observations with $\langle x_n \rangle^{1/2} \lesssim \sigma_{TOA}^2 \Delta f_p$, independent of the number of TOAs per continuous observing session. Roughly equal spacing is preferable, as false alarms occur more commonly adjacent to longer TOA gaps. (iii) Performance is essentially unaffected by $f_{dLS}$ and depends roughly on the product $\sigma_{TOA} f_{LS}$. (iv) The size of the smallest detectable glitch is governed mainly by $\Delta f_p$ and depends weakly on $\tau$, when the phase deviation produced by $\Delta f_p$ exceeds that produced by the transient ($\approx \tau \Delta f_p$). (v) Recipes for setting the DOI and grid resolution are set out in Appendix C.

The performance tests in this paper are restricted deliberately to synthetic data in order to establish performance bounds systematically under controlled conditions. Nevertheless, as a foretaste of what can be achieved with astronomical data, we also apply the HMM to 490 publicly available TOAs from PSR J0835–4510, covering the interval from MJD 57427 to MJD 57810 (Sarkissian et al. 2017a, 2017b). We confirm the existence of the large glitch on 2016 December 12, with log Bayes factor $\approx 1.1 \times 10^3$, and rule out with high statistical confidence the existence of a second glitch during the same interval. The inferred ephemeris, including $\Delta f_p^d$ and $\Delta \dot{f}$, agrees with that yielded by traditional timing methods, after allowing for the fact that the introductory HMM in this paper does not include post-glitch recoveries. We look forward to applying the HMM to other pulsars, both to detect glitches and to improve the sensitivity of nanohertz gravitational wave searches with pulsar timing arrays (Lentati et al. 2015; Shannon et al. 2015; Arzoumanian et al. 2016; Hobbs & Dai 2017).

In closing, we reaffirm that the HMM scheme developed in this paper complements—but does not replace—traditional glitch-finding approaches based on least-squares fitting of a Taylor-expanded phase model plus glitch template. Indeed, the HMM ingests standard TOAs and leverages the outputs of existing software (e.g., $f_{LS}, \dot{f}_{LS}$, and phase residuals $\delta \phi(t_n)$ from TEMPO2) to demarcate its state space (DOI). It complements existing Bayesian approaches, e.g., TEMPONEST (Lentati et al. 2014; Shannon et al. 2016; Lower et al. 2018), by tracking the observed spin wandering explicitly, as a specific realization of a discrete-time Markov Chain, instead of estimating its ensemble statistics (e.g., power spectral density). Every approach has advantages and disadvantages. The HMM is unsupervised, so its performance bounds (e.g., $P_a, P_d$) can be computed efficiently. It is fast, requiring $\sim 10^{12}$ floating point

Figure 8. Logarithm of the Bayes factor, $K_i(k)$, vs. the TOA index, $k$, for PSR J0835–4510 from MJD 57427 to MJD 57810, computed with the same HMM parameters as in Figure 7, but with the data segmented into halves (red and yellow curves) and quarters (purple, green, light blue, and brown curves). The blue curve, incorporating all the data, is copied from Figure 7 for comparison.
operations (∼0.1 CPU hours) per pulsar per year of observations. It discriminates accurately between spin wandering and glitches by tracking both phenomena explicitly with a Markov Chain. On the other hand, when spin wandering and glitches are negligible, the HMM is superfluous. Pulse domain methods ultimately promise the best sensitivity but they expend a lot of computational effort correcting for random pulse-to-pulse profile variations and do not ingest standard TOAs. They may be strongest when combined with an HMM similar to the one described here. If the problem allows, it is wise to apply several methods simultaneously. There is no purely objective answer to the question of whether or not a data set contains a glitch. The question is fundamentally statistical and can only be answered in the context of a user-selected false-alarm probability. The results in this paper show concretely and systematically how to define, compute, and set $P_{fa}$ for the HMM.

The authors thank Stefan Osłowski and Marcus Lower for pointing out important references and for providing access to data from the Molonglo Synthesis Radio Telescope for experimentation while developing the HMM algorithm. The PSR J0835−4510 data analyzed in Section 7 are described by Sarkissian et al. (2017a, 2017b). This research was supported by the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), grant No. CE170100004.
Appendix A
Solving the HMM

Let \( M = \{A_{\theta,q}, L_{\alpha(t_{n})\theta}, \Pi_{\theta}\} \) be an HMM with transition probability \( A_{\theta,q} \), emission probability \( L_{\alpha(t_{n})\theta} \), and prior probability \( \Pi_{\theta} \) defined according to Equations (1), (2), and (4) respectively. Let \( Q_{m,n} = \{q(t_{m}), \ldots, q(t_{n})\} \) and \( O_{m,n} = \{a(t_{m}), \ldots, a(t_{n})\} \) denote arbitrary, partial sequences of hidden and observed states, respectively, with \( 1 \leq m \leq n \leq N_{T} \). In this appendix, we present efficient numerical algorithms, which exploit recursion to solve the two fundamental HMM problems below.

1. What is the Bayesian evidence \( \Pr(O_{1:N_{T}}|M) \) for the model \( M \), given the full observed sequence, \( O_{1:N_{T}} \)? This question reduces to calculating
   \[
   \Pr(O_{1:N_{T}}|M) = \sum_{Q_{1:N_{T}}} \Pr(O_{1:N_{T}}|Q_{1:N_{T}}, M) \Pr(Q_{1:N_{T}}, M) \tag{A1}
   \]
   \[
   = \sum_{Q_{1:N_{T}}} \Pi_{\theta} \prod_{t=1}^{N_{T}} L_{\alpha(t_{t})\theta} q(t_{t}) \prod_{n=2}^{N_{T}} A_{\theta(t_{n-1})\theta} q(t_{n}) \tag{A2}
   \]

2. What is the optimal hidden sequence given \( M \) and \( O_{1:N_{T}} \)? This question reduces to calculating
   \[
   \hat{q}(t_{n}) = \arg \max_{q(t_{n})} \sum_{Q_{1:n-1}} \Pi_{\theta} \prod_{t=1}^{N_{T}} L_{\alpha(t_{t})\theta} q(t_{t}) \prod_{n=2}^{N_{T}} A_{\theta(t_{n-1})\theta} q(t_{n}) \tag{A3}
   \]
   \[\times \prod_{m=2}^{N_{T}} A_{\theta(t_{m-1})\theta} q(t_{m}) \prod_{m=n+1}^{N_{T}} A_{\theta(t_{m-1})\theta} q(t_{m}) \]
   for \( 2 \leq n \leq N_{T} \) and a uniform prior, if one wishes to maximize \( \Pr[q(t_{n})|O_{1:N_{T}}, M] \) point-wise, or
   \[
   Q_{1:N_{T}}^{*} = \arg \max_{Q_{1:N_{T}}} \sum_{Q_{1:n-1}} \Pi_{\theta} \prod_{t=1}^{N_{T}} L_{\alpha(t_{t})\theta} q(t_{t}) \prod_{n=2}^{N_{T}} A_{\theta(t_{n-1})\theta} q(t_{n}) \tag{A4}
   \]
   if one wishes to maximize \( \Pr(Q_{1:N_{T}}|O_{1:N_{T}}) \) sequence-wise. The difference between options (A3) and (A4) is explained below.

The above problems are essential building blocks of the glitch-finding algorithm in Section 4. A third fundamental problem—given \( O_{1:N_{T}} \), what model \( M \) maximizes the Bayesian evidence \( \Pr(O_{1:N_{T}}|M) \)?—amounts to learning the optimal model (here, the glitch dynamics) from the data. It is of great interest but lies outside the scope of this paper. The reader is referred to the excellent tutorial by Rabiner (1989) for a fuller treatment of the fundamental principles of HMMs.

A.1. Forward Algorithm

It may seem that evaluating the sum (A2) involves \( \sim N_{T} N_{Q}^{N_{T}} \) floating point operations, because each term is a product of \( 2N_{T} \) factors, and there are \( N_{Q}^{N_{T}} \) possible hidden sequences. Fortunately, recursive filtering offers a more efficient approach.

Consider the forward variable
   \[
   \alpha_{\theta}(t_{n}) = \Pr[q(t_{n}) = q_{i}, O_{1,n}|M], \tag{A5}
   \]
i.e., \( \alpha_{\theta}(t_{n}) \) equals the probability that one observes the partial data \( O_{1:n} \) during the interval \( t_{1} \leq t \leq t_{n} \), and the system occupies the state \( q_{i} \) at time \( t = t_{n} \). Notice that at \( t = t_{n} \), every one of the \( N_{Q} \) hidden states is reached from the same \( N_{Q} \) hidden states at \( t = t_{n-1} \). Hence, one can calculate (A2) by addressing every link in the trellis in Figure A1 once, instead of backtracking over every link multiple times while tracing all \( N_{Q}^{N_{T}} \) hidden sequences separately. The following algorithm achieves this economy by storing the partial results at each forward step through the trellis (Rabiner 1989; Quinn & Hannan 2001).

1. Initialization. For \( 1 \leq i \leq N_{Q} \), set
   \[
   \alpha_{\theta}(t_{1}) = \Pi_{\theta} L_{\alpha(t_{1})\theta}, \tag{A6}
   \]

Figure A1. Schematic of the HMM trellis. (Left panel) Subset of the links in the trellis. Every circle denotes a hidden state \( (q_{1}, q_{2}, q_{3}) \) at some time step (time increases to the right). The top rectangle contains the data \((X, Y)\). Every unbroken arrow corresponds to a nonzero transition probability, e.g., \( A_{12} \) is the probability of transitioning from \( q_{1} \) to \( q_{2} \). Every broken arrow corresponds to a nonzero emission probability, e.g., \( L_{2Y} \) is the probability of observing the data \( Y \) at the third time step while occupying hidden state \( q_{2} \). (Right panel) Sample of the links that go into evaluating the induction step for \( q_{2} \) (circles shaded blue) for the forward variable (red arrows; Equation (A7)) and the backward variable (blue arrows; Equation (A11)).
2. **Induction.** For $1 \leq n \leq N_F - 1$ and $1 \leq i \leq N_Q$, compute the forward variable by summing over its values at the previous HMM step:

$$\alpha_q(t_{n+1}) = L_{o(t_{n+1})} \sum_{j=1}^{N_Q} A_{q_jq} \alpha_j(t_n). \quad (A7)$$

3. **Termination.** The Bayesian evidence is the sum of the forward variable over the final states, viz.

$$\Pr(O_{1:N_F}|M) = \sum_{i=1}^{N_Q} \alpha_q(t_{N_F}). \quad (A8)$$

The trellis contains $N_Q^2$ links per HMM transition, and there are $N_F - 1$ transitions, so the computation involves $\sim N_F N_Q^2$ floating point operations in total, a large saving.

### A.2. Forward–Backward Algorithm

The maximization step in Equation (A3) can be executed with the help of recursive smoothing, without comparing the $N_Q$ possible hidden sequences. This is achieved by introducing a backward variable, analogous to the forward variable above, and then maximizing the product of the forward and backward variables.

Consider the backward variable

$$\beta_q(t_n) = \Pr[O_{n+1:N_F}|q(t_n) = q_i, M], \quad (A9)$$

i.e., $\beta_q(t_n)$ equals the probability that one observes the partial data $O_{n+1:N_F}$ during the interval $t_{n+1} \leq t \leq t_F$, conditional on the system occupying the state $q_i$ at time $t = t_n$. Every one of the hidden states at $t = t_n$ connects to the same set of hidden states at $t = t_{n+1}$, so one can express $\beta_q(t_n)$ inductively in terms of $\beta_q(t_{n+1}), \ldots, \beta_q(t_{N_F})$ by summing over the $N_Q$ possible transitions from $q(t_n)$ to $q(t_{n+1}), \ldots, q(t_{N_F})$ (Rabiner 1989).

1. **Initialization.** For $1 \leq i \leq N_Q$, set

$$\beta_q(t_{N_F}) = 1. \quad (A10)$$

2. **Induction.** For $1 \leq n \leq N_F - 1$ and $1 \leq i \leq N_Q$, compute the backward variable by summing over its values at the succeeding HMM step, starting from $n = N_F - 1$ and stepping back to $n = 1$:

$$\beta_q(t_n) = \sum_{j=1}^{N_Q} A_{q_jq} L_{o(t_{n+1})q} \beta_q(t_{n+1}). \quad (A11)$$

The backward algorithm (A9)–(A11) entails $N_F N_Q^2$ floating point operations like the forward algorithm.

We now ask what hidden state is most likely to be occupied at $t = t_n$, given the entire observed sequence $O_{1:N_F}$ and the model $M$. Define

$$\gamma_q(t_n) = \Pr[q(t_n) = q_i|O_{1:N_F}, M] \quad (A12)$$

$$= \left[ \sum_{j=1}^{N_Q} \alpha_j(t_n) \beta_j(t_n) \right]^{-1} \alpha_q(t_n) \beta_q(t_n), \quad (A13)$$

where Equation (A13) follows from Equation (A12), because the forward variable accounts for the hidden and observed sequences $Q_{1:n}$ and $O_{1:n}$ terminating at $q(t_n) = q_i$, and the backward variable accounts for the hidden and observed sequences $Q_{n:N_F}$ and $O_{n:N_F}$ originating at $q(t_n) = q_i$. Equation (A13) implies that the most likely state at each HMM step is given by

$$\hat{q}(t_n) = \arg \max_{1 \leq i \leq N_Q} \gamma_q(t_n) \quad (A14)$$

for $1 \leq n \leq N_F$. The denominator of Equation (A13) is a normalization factor which, for $t = N_F$, reduces to $\Pr(O_{1:N_F}|M)$ in Equation (A2) via Equations (A8) and (A10). It can be ignored when maximizing over $1 \leq i \leq N_Q$. Equations (A5)–(A14) together constitute the HMM forward–backward algorithm. The algorithm entails $N_F N_Q^2$ floating point operations, dominated by Equations (A11)–(A13); the final step, Equation (A14), reduces to $\sim N_F N_Q$ operations with binary maximization.

The above solution of the HMM optimization problem is not unique. It does maximize the number of the most-probable hidden states. On the other hand, there is no guarantee that the sequence generated thus is admissible, i.e., consistent with the transition probabilities. For example, if we have $A_{q_1q_i} = 0$ for some $q_j$ and $q_i$, it may not be possible to connect the sequence $\{\hat{q}(t_1), \ldots, \hat{q}(t_{N_F})\}$ generated by Equation (A14). In this sense, $\{\hat{q}(t_1), \ldots, \hat{q}(t_{N_F})\}$ differs from $Q_{1:N_F}$ in Equation (A4). The latter quantity is admissible by construction and maximizes the probability of the whole sequence rather than individual states along the sequence. In general, both approaches (and indeed others not discussed here) are valid. In this paper, we focus on $\{\hat{q}(t_1), \ldots, \hat{q}(t_{N_F})\}$ for three reasons. First, we wish to maximize the number of the most-probable hidden states when generating an ephemeris. Second, we find by trial and error that inadmissibility arises rarely in the glitch-finding application. Third, we wish to know the shape of the joint PDF of $f(t_n)$ and $\hat{t}_n$ at each $t_n$, in order to check how far the optimal sequence stands above its nearest competitors. This is done easily by plotting $\gamma_q(t_n)$ versus $q_i$, whereas $Q_{1:N_F}$ gives the best sequence only. Traditional, frequentist pulsar timing methods involve a mixture of point-wise and sequence-wise optimization by minimizing the squares of the point-by-point phase residuals summed over the entire sequence.

### A.3. Viterbi Algorithm

For the sake of completeness, we outline an algorithm for calculating $Q_{1:N_F}^*$ in Equation (A4). Known as the Viterbi algorithm and based on dynamic programming methods, it exploits the property that any subsequence of the optimal sequence is itself optimal in order to prune the trellis of admissible sequences (Rabiner 1989; Quinn & Hannan 2001). The pseudocode below matches closely the notation adopted by Suvorova et al. (2016); see Bayley et al. (2019).

Consider the variable

$$\delta_q(t_n) = \max_{Q_{1:n-1}} \Pr[q(t_n) = q_i, Q_{1:n-1}|O_{1:x}, M], \quad (A15)$$

$$= L_{o(t_n)q_i} \max_{q_j} A_{q_jq_i} \delta_q(t_{n-1}), \quad (A16)$$

which corresponds to the maximum probability that the HMM terminates in the hidden state $q_i$ at $t = t_n$ given the partial observation sequence $O_{1:x}$. Let $\delta_q(t_n)$ denote the hidden state at $t = t_{n-1}$ from which $q_i$ is reached at $t = t_n$, along the
sequence that maximizes \( \Pr[q(t_n) = q_n, Q_{1:n-1}|O_{1:n}, M] \) in Equation (A15), viz.

\[
\psi_q(t_n) = \arg\max_{q_n} A_{q_nq} \delta_q(t_n-1), \tag{A17}
\]

The Viterbi algorithm evaluates \( \delta_q(t_n) \) and \( \psi_q(t_n) \) for all of the \( N_T N_q \) nodes in the trellis in Figure A1 and then backtracks to reconstruct \( Q^*_{1:N_T} \). It resembles the forward algorithm, with the sum in Equation (A7) replaced by the maximization steps in Equations (A16) and (A17).

1. Initialization. For \( 1 \leq i \leq N_q \), set

\[
\delta_q(t_i) = \Pi_q L_{o(t_i)q}, \tag{A18}
\]

Note that \( \psi_q(t_i) \) is not initialized as it is never needed.

2. Forward recursion. For \( 2 \leq n \leq N_T \) and \( 1 \leq i \leq N_q \), implement the induction step, Equation (A16), via

\[
\delta_q(t_n) = L_{o(t_n)q} \max_{1 \leq j \leq N_q} A_{q_jq} \delta_q(t_{n-1}) \tag{A19}
\]

and

\[
\psi_q(t_n) = \arg\max_{1 \leq j \leq N_q} A_{q_jq} \delta_q(t_{n-1}). \tag{A20}
\]

3. Termination. Identify the state \( q^*(t_N) \), where the optimal sequence ends.

\[
\Pr(Q^*_{1:N_T}|O_{1:N_T}, M) = \max_{1 \leq j \leq N_q} \delta_q(t_N) \tag{A21}
\]

and

\[
q^*(t_N) = \arg\max_{1 \leq j \leq N_q} \delta_q(t_N). \tag{A22}
\]

4. Backward recursion. Backtrack through the trellis in Figure A1 to reconstruct the optimal sequence, guided by \( \psi_q(t_n) \). For \( 1 \leq n \leq N_T - 1 \), compute

\[
q^*(t_n) = \psi_q(t_{n+1}) \psi_q(t_{n+2}), \tag{A23}
\]

starting from \( n = N_T - 1 \) and stepping back to \( n = 1 \).

The algorithm involves \( N_T N_q \log N_q \) floating point operations with binary maximization (Quinn & Hannan 2001).

**Appendix B**

**Hidden-state Evolution via a Langevin Equation**

In this appendix, we derive the transition probabilities in Section 3.4 self-consistently by solving a stochastic differential equation for \( q(t) = [f(t), \dot{f}(t)] \) in the interstep interval \( t_{n-1} \leq t \leq t_n \). Although the system is measured at discrete instants \( t_n \), its state evolves stochastically between the TOAs due to timing noise (Cordes 1980). For now, we ignore the secular component of the torque derivative, \( \dot{f} \), a good approximation provided that \( \chi^2_{c} \leq 6\sigma(t_n)/|\dot{f}(t_n)| \) is satisfied; see Section 3.1. This leaves an approximately constant secular torque, which enters as an initial condition on \( \dot{f}(t_{n-1}) \), and a fluctuating torque derivative \( \xi(t) \), which drives the Langevin equation,

\[
\frac{df}{dt} = \xi(t), \tag{B1}
\]

with white-noise statistics,

\[
\langle \xi(t) \rangle = 0 \tag{B2}
\]

and

\[
\langle \xi(t)\xi(t') \rangle = \sigma^2 \delta(t - t'). \tag{B3}
\]

Angular brackets denote an ensemble average over noise realizations. Unlike \( \sigma^2_{\text{TOA}} \), the variance \( \sigma^2 \) is not a measurement uncertainty. It is a mean-square measure of the amplitude of the process noise driven by the fluctuating torque derivative, which may arise physically from starquakes and superfluid vortex avalanches, for example (Chugunov & Horowitz 2010; Warszawski & Melatos 2011; Haskell & Melatos 2015). Its units are Hz\(^2\) s\(^{-3}\); see \( \sigma^2_{\text{TOA}} \), which has units of s\(^2\). Likewise, \( \sigma^2 \) is not the same as the \( \sigma^2_{\text{TN}} \) in the synthetic data in Section 5, because \( \sigma^2_{\text{TN}} \) equals the variance in the autocorrelation function of the torque, not the torque derivative. White-noise fluctuations in the torque derivative, as in Equation (B1), are not necessarily physical; they are an artificial device to keep \( A_{q_0q} \) finite. By contrast, a fluctuating torque gives \( \langle \dot{f}(t)\dot{f}(t') \rangle \propto \delta(t - t') \), which diverges in the limit \( t \to t' \). The tests in Section 6 and Appendix F confirm that Equation (B1) works well empirically when tracking synthetic data generated by a fluctuating torque. This reflects a well-known property of HMMs, that they are insensitive to the exact form of \( A_{q_0q} \) as long as the dynamics during the interval \( t_{n-1} \leq t \leq t_n \) are captured broadly, e.g., \( A_{q_0q} = A_{q_0q} = A_{q_0q} = 1/3 \) often serves as an adequate model for more complicated Brownian motion (Quinn & Hannan 2001; Suvorova et al. 2016, 2017).

The PDF \( p[f(t_n), \dot{f}(t_n)|f(t_{n-1}), \dot{f}(t_{n-1})] \) at \( t = t_n \) describing the ensemble of Langevin trajectories starting from the state \( [f(t_{n-1}), \dot{f}(t_{n-1})] \) at \( t = t_{n-1} \) satisfies the Fokker–Planck equation (Gardiner 1994)

\[
\frac{\partial p}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial t^2}. \tag{B4}
\]

The coefficients in Equation (B4) are constant, so the solution is a Gaussian. It is defined by the first two moments, which can be calculated directly from the Langevin trajectory,

\[
f(t) = f(t_{n-1}) + (t - t_{n-1})\dot{f}(t_{n-1}) + g(t_{n-1}) \times [\Delta f_p(t_{n-1}) + (t - t_{n-1})\Delta f_p(t_{n-1})]
+ \int_{t_{n-1}}^{t} dt' \int_{t_{n-1}}^{t'} dt'' \xi(t'') . \tag{B5}
\]

For any stationary process, we have

\[
\int_{0}^{t} dt' \int_{0}^{t'} dt'' \langle \xi(t'')\xi(t'') \rangle = \sigma^2 \min(t, t'). \tag{B6}
\]

Combining Equations (B5) and (B6), we find that the first moments evolve according to

\[
\langle f(t) \rangle = f(t_{n-1}) + (t - t_{n-1})\dot{f}(t_{n-1}) + g(t_{n-1})[\Delta f_p(t_{n-1}) + (t - t_{n-1})\Delta f_p(t_{n-1})] . \tag{B7}
\]
\[ \langle \dot{f}(t) \rangle = \dot{f}(t_{n-1}) + g(t_{n-1}) \Delta \dot{f}_p(t_{n-1}) , \quad \text{(B8)} \]

and the second moments evolve according to
\[ \text{cov}[f(t), f(t)] = \frac{1}{3} \sigma^2(t - t_{n-1})^3, \quad \text{(B9)} \]
\[ \text{cov}[f(t), \dot{f}(t)] = \frac{1}{2} \sigma^2(t - t_{n-1})^2, \quad \text{(B10)} \]
\[ \text{cov}[\dot{f}(t), \dot{f}(t)] = \sigma^2(t - t_{n-1}), \quad \text{(B11)} \]

where \( \text{cov}(a, b) = \langle (a - \langle a \rangle)(b - \langle b \rangle) \rangle \) denotes the central covariance of \( a \) and \( b \). Equations (B7)–(B11) together define \( A_{fi(t)} \) through Equations (10)–(13).

We emphasize that the Wiener process, Equations (B1)–(B3), may not be realistic physically for every pulsar. Empirically speaking, pulsar timing noise does not display significant memory in the torque derivative over typical TOA gaps of days to weeks (Price et al. 2012), so the white noise in Equations (B2) and (B3) represents a fair approximation.\(^{16}\) However, for longer TOA gaps, redder timing noise, or long-lasting post-glitch recoveries (see Sections 3.4 and 7), Equations (B1)–(B3) need to be generalized.\(^{17}\) In this paper, we test the robustness of Equations (B1)–(B3) in two ways. First, we deliberately generate synthetic data with a different noise model, which is white in the torque instead of the torque derivative (see Equations (E1)–(E3)), yet still the HMM performs well in the tests in Section 6. Second, in Section 7, the assumption given by Equations (B1)–(B3) does not harm the HMM’s ability to locate accurately the 2016 December 12 glitch in PSR J0835–4510 and exclude the existence of a second glitch in its vicinity, in accordance with traditional analyses. This is comforting, because in PSR J0835–4510, the timing noise is relatively red, and the post-glitch recoveries are notoriously lengthy (Lyne et al. 1996). Generalizing the calculations in this appendix to redden Equations (B1)–(B3) is an interesting avenue for future work, once the HMM is validated against more real pulsars. Ultimately, glitch detection is an exercise undertaken conditionally with respect to a phase model; there is no model-independent answer to the question of whether or not a stretch of data contains a glitch. This is equally true of traditional methods, whether the model is simple (e.g., step changes in an otherwise smooth Taylor expansion) or complex (e.g., phase residuals with a power-law power spectral density (Shannon et al. 2016)).

### Appendix C

#### Defining the Grid and DOI

The DOI encompasses the point \( (f_{LS}, \dot{f}_{LS}) \) corresponding to the optimal (least squares) constant-coefficient phase model fitted with \( f = f_{LS} \), \( \dot{f} = \dot{f}_{LS}, \) and \( \ddot{f} = 0 \). The fit can be generated from the TOAs by running TEMPO2, for example. At any instant, the true, unknown \( f(t) \) and \( \dot{f}(t) \) deviate slightly from \( f_{LS} \) and \( \dot{f}_{LS}, \) respectively. One way to estimate the deviations is to attribute the phase residual \( \delta \phi(t_n) \) measured at each HMM step \( t_n \) \((1 \leq n \leq N_T)\) to a pure frequency fluctuation, \( \varepsilon_f(t_n) = \delta \phi(t_n)/\kappa_n \) (with \( \dot{f} = f_{LS} \)), or a pure frequency derivative fluctuation, \( \varepsilon_f(t_n) = 2 \delta \phi(t_n)/\kappa_n^2 \) (with \( f = f_{LS} \)). We then define the DOI to be the rectangular domain
\[
\min_{1 \leq n \leq N_T} \varepsilon_f(t_n) \leq S^{-1}(f - f_{LS}) \leq \max_{1 \leq n \leq N_T} \varepsilon_f(t_n), \quad \text{(C1)}
\]
\[
\min_{1 \leq n \leq N_T} \varepsilon_f(t_n) \leq S^{-1}(\dot{f} - \dot{f}_{LS}) \leq \max_{1 \leq n \leq N_T} \varepsilon_f(t_n), \quad \text{(C2)}
\]

where \( S \geq 1 \) is a dimensionless safety factor chosen by the user. Equations (C1) and (C2) are conservative, because in reality the fluctuations develop over multiple HMM steps (thereby reducing \( \varepsilon_f(t_n) \) and \( \varepsilon_f(t_n) \)) and occur in tandem \( \varepsilon_f(t_n) = 0 \) and \( \varepsilon_f(t_n) = 0 \), simultaneously).

The continuous physical variables \( f(t) \) and \( \dot{f}(t) \) are discretized for numerical purposes. Formally, the grid resolution is governed by the curvature of the likelihood function at its peak through the Cramér–Rao lower bound or related quantities like the parameter space metric in gravitational wave applications (Leaci & Prix 2015; Wette 2016). The peaks of \( \Lambda_{\nu(t_n)}, \nu = 7 \), sharpen, as \( \kappa \gg 1 \) increases. However, it is unclear how to apply such approaches to the problem at hand, because the distribution of the number of pulses between consecutive TOAs is unknown (Suvorova et al. 2018). It can be estimated, say as a Poisson or quasiperiodic process (Melatos et al. 2008; Fulgenzi et al. 2017; Howitt et al. 2018), with the relevant timescales determined iteratively (if the data are analyzed for the first time) or copied from the literature (if glitches have already been detected; Carlin et al. 2019). Alternatively, one can approximate the likelihood function (for the purpose of grid design only) assuming a constant \( \dot{f} \) and \( \sigma_{TOA} = 0 \), as discussed thoroughly by Suvorova et al. (2018). In this paper, for simplicity, we set the grid spacing to be the minimum \( \Delta f_p \) and \( \Delta f_{\ddot{f}} \) that we wish to resolve, limited only by computational cost. Other options that may deliver computational savings, such as logarithmic gridding, will be explored in future work.

The set \( G \) in Equation (10) is constructed as follows. For the frequency component, we allow all jumps with \( \Delta f_p > 0 \), such that \( f(t_{n-1}) + \Delta f_p \) is a valid state and lies in the DOI. For the frequency derivative component, we allow all jumps of either sign, such that \( f(t_{n-1}) + \Delta f_{\ddot{f}} \) is a valid state and lies in the DOI. Note that \( \Delta f_{\ddot{f}} > 0 \) does not imply \( f(t_n) > f(t_{n-1}) \) necessarily, because the spin down between TOAs may compensate for the glitch.

Gridding modifies the emission probability given by Equations (7) and (8), as noted in Section 3.3, by changing the effective value of \( \kappa \). This occurs because discretization introduces a state and hence a phase uncertainty proportional to the grid spacing, which adds in quadrature to the phase uncertainty arising from the intrinsic measurement uncertainty. Let \( \sigma_{TOA,n} \) be the measurement uncertainty in \( \nu \), and let \( \eta_p \) and \( \eta_{\ddot{f}} \) be the grid spacings in the frequency and frequency derivative variables, respectively. Then \( \kappa_n \), which equals the inverse square of the phase uncertainty accumulated over the interval \( x_n \), as in Equation (8), depends on \( t_n \) and takes the generalized form
\[
\kappa_n = \{(\sigma_{TOA,n})^2 + \sigma_{TOA,n}^2 f(t_n)^2 + x_n^2 \eta_p^2 + x_n^4 \eta_{\ddot{f}}^2/4\}^{-1}. \quad \text{(C3)}
\]
Equation (C3) reduces to Equation (8) for \( \sigma_{\text{TOA},n-1} = \sigma_{\text{TOA}} = \sigma_{\text{TOA}} \) and \( \eta_p = 0 = \eta_f \). It preserves the Markovian nature of the HMM, because \( L_{\text{HMM}}(t) \) depends only on the state and data at \( t_n \), now expanded to embrace \( \sigma_{\text{TOA},n}, \eta_p, \) and \( \eta_f \).

**Appendix D**

**Jump Markov Model**

Instead of relying on Bayesian model selection to detect glitches, as in Section 4, one can use the HMM to track the hidden Boolean variable \( g(t) \) introduced in Section 3.1. Glitches are sparse, so it is needlessly costly to sample all \( 2^N \) possible sequences \( \{g(t_0), ..., g(t_N-1)\} \). An approximation, known as a jump Markov model, involves replacing \( g(t_{n-1}) \) in Equations (11) and (12) by the hyperparameter \( g = \langle g(t) \rangle \), i.e., the time-averaged glitch probability per TOA. (A more sophisticated version assumes something about glitch statistics, e.g., a Poisson process, and relates \( g(t_{n-1}) \) to \( x_{n-1} \). Typically, we have \( g \ll 1 \). In the jump Markov model, \( A_{g|g} \) contains a simultaneous mixture of glitch and no-glitch evolution through Equations (11) and (12). Once the HMM generates an optimal sequence, it is important to check a posteriori the model evidence; in effect, \( g = \langle g(t) \rangle \) is a uniform prior on every \( g(t_n) \), which must be updated to estimate the posterior of \( g(t_n) \), once the HMM finishes its work.

Suvorova et al. (2018) investigated glitch finding with a jump Markov model and concluded that it does not work as well as the procedure described in Section 4 for small glitches. In short, the method finds too many false, small glitches, which makes sense; glitches are sparse, so we expect \( g(t_n) = 0 \) for most \( 1 \leq n \leq N_T \). The reader is referred to the detailed study by Suvorova et al. (2018) for more information.

**Appendix E**

**Generating Synthetic Data**

An infinite family of Langevin equations can generate solutions of the form (19) with a random walk added. In this paper, we solve

\[
\frac{df}{dt} = \frac{f_e - f}{2\tau} + f(0) + \left( \Delta f_p + \frac{\Delta f_p}{\tau} \right) H(t - T) \\
+ (\Delta f_e + \Delta f_s) \delta(t - T) + \zeta(t),
\]

\( \zeta(t) \) is a zero-mean, white-noise torque satisfying

\[
\langle \zeta(t) \zeta(t') \rangle = \sigma_{\text{TN}}^2 \delta(t - t'),
\]

\( \sigma_{\text{TN}} \) is the timing noise amplitude (units: Hz s\(^{-1/2}\)), \( \delta(\cdot) \) is the Dirac delta function, and \( f_e \) is an auxiliary variable, whose physical interpretation is irrelevant here (see below).  

Equations (E1) and (E2) are solved subject to the initial conditions \( f(0) = f(0) \) and \( f_e(0) = f(0) + f(0)\tau \).

---

**References**

Figure E1 displays a sample of the synthetic data generated by the above procedure. Overall, it comprises 250 TOAs sampled according to a Poisson process, whose waiting times \( \Delta t \) are distributed according to the probability density function \( p(\Delta t) = \lambda_{\text{TN}} \exp(-\lambda_{\text{TN}} \Delta t) \), with \( \lambda_{\text{TN}} = 0.864 \text{ day}^{-1} \). By assigning the TOAs randomly, we ensure that they do not coincide with the glitch in general. The top-left panel displays the time series \( f(t) \) sampled with high temporal resolution over a subinterval lasting \( 10^5 \text{s} \). The frequency fluctuations generated by the torque noise process in Equations (E1)–(E3) are clearly visible. Their rms amplitude, which reaches one part in \( \sim 10^3 \), is consistent with \( \sigma_{\text{TN}} = 5 \times 10^{-13} \text{Hz s}^{-1/2} \) over an interval of \( 10^5 \text{s} \). The top-right panel displays the phase evolution (including wrapping) within a short window lasting \( 1 \text{s} \) and indicates the TOAs of individual pulses within the window. The bottom-left panel shows the TOA residuals produced by a TEMPO2 fit to the whole data set, \( 0 \leq t/(1 \text{ day}) \leq 286 \), including a relatively large glitch with \( \Delta f_{\text{g}} = 5 \times 10^{-8} \text{Hz} \) at \( t = 144.67 \text{ days} \). The secular spin-down parameters \( f_{\text{LS}} \) and \( f_{\text{LS}} \) inferred from the fit agree well with the injected values of \( f(0) = f(0) \). The TOA residuals \( 10^{-5} \text{s} \) are consistent with \( \sigma_{\text{TN}} \) over 286 days and are consistent visually with the red phase noise produced by filtered white torque noise in line with Equations (E1)–(E3), except for a modest spike around the glitch epoch, because the glitch fit is imperfect. The bottom-right panel shows the autocorrelation function of the phase residuals in the bottom-left panel. The half-power point occurs at a lag of \( \approx 7 \text{ days} \). The oscillations at lags \( \leq 50 \text{ days} \) are characteristic of red phase noise, as observed in many pulsars (Price et al. 2012).

Equations (E1) and (E2) can be extended in several ways. As the system is linear and obeys the principle of superposition, it is easy to add more exponential recoveries with amplitudes \( \Delta f_e \) and timescales \( \tau_n (k > 1) \) by lifting the order of the system of differential equations and adding forcing terms proportional to \( H(t - T) \) and \( \delta(t - T) \) as in Equations (E1) and (E2). One can also include a secular second derivative of the form \( \langle \dot{f} \rangle \propto f^n \).

The reader may notice a similarity between Equations (E1) and (E2) and the two-component model of a neutron star interior (Baym et al. 1969), where \( f \) and \( f_e \) correspond to the spin frequencies of the rigid crust and neutron condensate, respectively. The analogy may prove useful in future work when physically interpreting the results of HMM-based glitch searches, but it is not pertinent to this paper. Here we merely exploit the mathematical correspondence, which renders Equations (E1) and (E2) with \( \zeta(t) = 0 \) equivalent to Equation (19), in order to generate synthetic data. We emphasize that the dynamical model for generating synthetic data differs deliberately from the dynamical model governing the hidden-state evolution in the HMM. For example, white-noise fluctuations enter through the torque in Equation (E1) (i.e., \( \xi(t) \)) and through the torque derivative in Equation (9) (i.e., \( \xi(t) \)). This reflects the situation in practice astrophysically, where the dynamical model for timing noise is unknown. It also confirms the robustness of the HMM in light of the encouraging results in Section 6. Likewise, the permanent jumps \( \Delta f_e \) and \( \Delta f_p \) in the HMM transition probability, Equations (10)–(12), do not match exactly the eponymous variables in the synthetic data generation model, Equations (E1) and (E2), although they are related.
Appendix F

Representative Worked Example: Synthetic Data

In this section, we illustrate how to apply the HMM in Section 3 and the model selection procedure in Section 4 to a sample of synthetic data generated according to the recipe in Section 5 and the parameters in the penultimate (typical) column in Table 1. The results are plotted in Figures F1 and F2. The worked example breaks out the steps in the analysis and introduces several useful diagnostics. It is a training run for the systematic performance tests in Section 6.

The top two panels of Figure F1 present the raw data before any analysis with the HMM. The top-left panel graphs the phase residuals $\delta \phi(t_n)$ as a function of time after subtracting a no-glitch spin-down model with $f_{LS} = 5.435 \text{ Hz} = 1.48 \times 10^{-8} \text{ Hz}$ and $f_{LS} = -1.41 \times 10^{-15} \text{ Hz s}^{-1}$, inferred by fitting the full data set with TEMPO2. The residuals diverge for $t > T = 253,371$ days quadratically (and the phase wraps at $n \approx 550$ days), as expected for a glitch with $\Delta f_b > 0$ and $\Delta t_p > 0$. Phase residuals are also plotted in the top-right panel after subtracting a no-glitch model with $f_{LS} = 5.435 \text{ Hz} = -1.2 \times 10^{-10} \text{ Hz}$, $f_{LS} = -9.164 \times 10^{-16} \text{ Hz s}^{-1}$, $T = 254.37$ days, $\Delta f_p = 8.535 \times 10^{-9} \text{ Hz}$, and $\Delta t_p = 1.09 \times 10^{-15} \text{ Hz s}^{-1}$, again fitted with TEMPO2. The residuals in the top-right panel do not diverge and have rms amplitude $\sim 10^{-3}$ rad, consistent with $\sigma_{TN} = 1 \times 10^{-12} \text{ Hz s}^{-1/2}$ integrated over $\sim 10^2$ days ($\sigma_{TN}$ dominates $\sigma_{TOA}$ in this example). The fitted parameters are close to the injected parameters quoted in the penultimate column in Table 1. Returning to the no-glitch fit, we convert $\delta \phi(t_n)$ into the DOI and grid spacing specified in the figure caption following the recipe in Appendix A.

Model selection is now performed. The second row of Figure F1 displays the Bayes factor, $K_1(k) = \text{Pr}[O_1; \Omega_1|M_1(k)]/\text{Pr}(O_1; \Omega_1|M_0)$, as a function of the TOA index, $k$. The Bayes factor peaks at $\ln K_1(k_1^* = 20) \sim 10^3$, well above the threshold $K_1(k) > 10^{1/2}$. In other words, a model featuring a glitch near the injection location ($k = 19$) is preferred categorically over the no-glitch model. The $\ln K_1(k)$ plateau near $k = k_1^*$ is typical of the HMM output for relatively large glitches, but the peak stands clearly above neighboring points, before the logarithm is taken. The single-TOA mismatch between the injected and recovered epochs is also typical. The HMM cannot say anything about the phase evolution between TOAs, so a single-TOA mismatch is always possible, even when a glitch is injected exactly at a TOA. Two-glitch models $M_2(k, l)$ are not considered here, as only one glitch is injected.

Finally, an ephemeris is constructed for the preferred model. The bottom four panels display $f(t_n)$ (third row) and $\delta \phi(t_n)$ (fourth row) as functions of $t_n$ for the point-wise optimal hidden sequence $q(t_n)$ found by the HMM forward–backward algorithm (blue curve) in Appendix A. For comparison, the sequence-wise...
optimal sequence $Q_{1:N_1}^*$ found by the Viterbi algorithm is also graphed (without residuals) as a red curve in the third row. Both HMM sequences lie close to each other and to the true, injected sequence, plotted as a dashed-dotted curve. The forward–backward sequence $\tilde{q}(t_n)$ yields an rms error of $\approx 1 \times 10^{-3}$ rad for the one-glitch model $M_1(k_1^*)$ (right column), which corresponds to $\approx 0.1 \max_n |\tilde{\delta}q(t_n)|$ for the no-glitch model $M_0$ (left column). The $M_0$ residuals are highest at $t \approx T$, as expected.

The posterior PDF of the hidden-state likelihood in the neighborhood of the optimal sequence is a useful diagnostic. It indicates how far the optimal sequence stands out above its nearest competitors. It also provides a way of estimating point-by-point confidence intervals for the optimal ephemeris in practical astrophysical applications, where the underlying, true ephemeris is unknown. The top two panels in Figure F2 display heat map contours of the posterior PDF computed by the forward–backward algorithm, $\gamma_q(t_n)$ in Equation (A13), marginalized over $\tilde{f}$ (first row) and $f$ (second row). The point-wise (forward–backward; blue curve) and sequence-wise (Viterbi; red curve) optimal-state sequences run through the middle of the high-probability (yellow) regions. The bottom four panels display cross sections of $\gamma_q(t_n)$ immediately before (third row) and after (fourth row) the recovered glitch. The optimal state stands out clearly and is localized precisely. The FWHM of the PDF marginalized over $\tilde{f}$ satisfies $\approx 1.4 \times 10^{-10}$ Hz and $\approx 4.8 \times 10^{-10}$ Hz before and after the glitch, respectively, while the FWHM of the PDF marginalized over $f$ satisfies $\approx 3 \times 10^{-16}$ Hz s$^{-1}$ and $\approx 6 \times 10^{-16}$ Hz s$^{-1}$ before and after the glitch, respectively. We can compute the jumps in $\tilde{f}$ and $f$ during the glitch by comparing the peaks in the third and fourth rows. The displacements are clearly visible, once enough time elapses. We find $\tilde{f}(t_{12}) - \tilde{f}(t_0) = 8.5 \times 10^{-9}$ Hz and $\tilde{f}(t_{12}) - f(t_0) = 3 \times 10^{-16}$ Hz s$^{-1}$, see the injected values $\Delta f_p = 1 \times 10^{-8}$ Hz and $\Delta \tilde{f}_p = 1 \times 10^{-15}$ Hz s$^{-1}$.
Appendix G

Schedule of Observations: Impact on Performance

When optimizing an observational campaign aimed at detecting glitches, it is important to plan how the spacing of observation sessions and the number of TOAs affect $P_{fa}$ and $P_d$. A typical observation session may last a few minutes, with ~104 pulses averaged to produce each TOA. Sessions are often separated by days to weeks, although of course there are exceptions; for instance, PSR J0835−4510 is monitored continuously for extended intervals (Palfreyman et al. 2016, 2018).

Figure G1 partially quantifies the above considerations. The left panel graphs $P_{fa}$ and $P_d$ as functions of the total number of observation sessions, after adjusting the Bayes factor threshold to achieve $P_d \geq 0.9$ for intervals between $\sim10^4$ and $\sim10^6$ s, which are readily achievable with dedicated or multibeam telescopes. For shorter intervals, the phase error due to $\sigma_{TOA}$, which is independent of $\langle x_n \rangle$, impairs the HMM’s performance. For longer intervals, the phase error due to binning (see Appendix C) dominates, because it scales $\propto \langle x_n \rangle$. The results in Figure G1 are generated for one TOA per observation session.

Tests show that $P_{fa}$ is roughly constant given between one and five TOAs per session for the parameters in Figure G1. A thorough study of multiple TOAs per session, including the related and important matter of pulse jitter (Helfand et al. 1975), is postponed to future work.

Figure G1 quantifies the above considerations. The left panel graphs $P_{fa}$ and $P_d$ as functions of the total number of observation sessions, after adjusting the Bayes factor threshold to achieve $P_d = 10^{-2}$ on average across the plotted range. A detection is highly probable in most realistic scenarios; we obtain $P_d \geq 0.9$ for $\geq 35$ sessions. The right panel graphs $P_{fa}$ and $P_d$ as functions of the mean interval between sessions averaged over the entire observation ($\sim1$ yr), after adjusting the Bayes factor threshold as in the left panel. We obtain $P_d \geq 0.9$ for intervals between $\sim10^4$ and $\sim10^6$ s.

We formulate a useful rule of thumb to predict how one should space observations to resolve glitches of a certain size. During a gap of duration $x_n$, phase deviations $\Delta f_n x_n$ and $\Delta \dot{f}_n x_n^2 / 2$ develop for frequency and frequency derivative jumps, respectively. Writing their ratio as $\approx (\Delta f / f_{LS})(\Delta \dot{f} / \dot{f}_{LS})^{-1}(x_n f_{LS} / \dot{f}_{LS})^{-1}$, we
see that the two contributions are comparable typically, e.g., for \( \Delta f / f_{LS} \approx 10^{-7} \), \( \Delta f / f_{LS} \approx 10^{-2} \), and \( x_n f / f_{LS} \approx 10^{-5} \). When the glitch-related phase deviations exceed those produced by TOA measurement errors (\( \sigma_{TOA, f} \)) and astrophysical timing noise (\( \sigma_{TN, x_n} \)), the glitch is discerned above the noise. This occurs for \( \Delta f_p \geq \sigma_{TOA, f} (x_n)^{-1} \) and \( \Delta f_p \geq \sigma_{TN, x_n}^{1/2} \) for \( \Delta f_p = 0 \), or \( \Delta f_p \geq 2\sigma_{TOA, f} (x_n)^{-2} \) and \( \Delta f_p \geq 2\sigma_{TN, x_n}^{1/2} \) for \( \Delta f_p = 0 \). Both special cases agree with the general expression for \( \Delta f_p = 0 \) and \( \Delta f_p = 0 \) presented by Suvorova et al. (2018).

Another way to gauge the impact of the observational schedule on the HMM is to note that, when false alarms occur, they correlate with large TOA gaps. Figure G2 displays, in orange, a histogram of \( x_n \) values adjacent to false alarms, along with a blue histogram of all the simulated \( x_n \) values, whether or not they are adjacent to a false alarm. The simulations are in the regime, where \( f \) can be neglected (see Section 3.2). As
expected, there is a clear trend: false alarms occur more frequently near larger gaps, because the number of peaks in $L_{x_n}(t_o)$ within the DOI increases with $x_n$, even though the peaks sharpen (see Section 3.3). The correlation strengthens, as $\kappa$ and hence the number of false alarms increase. The opposite trend applies to false dismissals: the HMM is more prone to reject a true glitch, when $x_n$ is relatively short, because the phase deviation across the gap is relatively small.

Appendix H

Is There a Second Glitch in PSR J0835$-$4510 between MJD 57427 and MJD 57810?

In this appendix, we apply the greedy hierarchical algorithm introduced in Section 4.2 (Suvorova et al. 2018) to test for the existence of a second glitch in PSR J0835$-$4510 in the interval from MJD 57427 to MJD 57810. Specifically, we assume that the glitch found at MJD 57734.54 is real and construct the model $M_2(173, k)$, which features a glitch at TOA interval $x_{173}$ and a second glitch at $x_k$.

Figure H1 presents the analysis of $M_2(173, k)$. The top panel displays the Bayes factor, $K_2(k) = \Pr(O_{1:2:2} | M_2(173, k)) / \Pr(O_{1:2:2} | M_1(173))$, as a function of $k$. A peak is observed at $k = 174$, with $\ln K_2(174) \approx 8$. Formally, this counts as a detection by the criterion in Section 4.1. (The detection threshold in Section 4.1 is roughly consistent with $P_{\text{fa}} \approx 1 \times 10^{-2}$ and $P_{\text{fa}} \approx 0.9$ throughout this paper.) However, it occurs at the TOA immediately following the first glitch and is likely to be associated with it, because the introductory version of the HMM in this paper treats a glitch as an instantaneous step with no quasiexponential recovery, whereas in reality a recovery with $\tau = 0.96(17)$ days is measured independently in the 2016 December 12 event (Palfreyman 2016; Sarkissian et al. 2017b). The second-highest peak in the top panel of Figure H1, which occurs at $k = 123$ with $\ln K_2(123) \approx 1$, does not count as a detection by the criterion in Section 4.1.

The bottom-left and -right panels in Figure H1 display the point-wise optimal-state sequences $\hat{f}(t_n)$ and $\hat{f}(t_n)$, respectively, calculated by the forward–backward algorithm. The plots zoom into the neighborhood of the second “glitch” at $t_{174}$ (dashed vertical line) and compare the one-glitch model $M_1(173)$ (blue curve) with the two-glitch model $M_2(173,174)$ (red curve). Both models handle the complicated, composite dynamics of the spin up and quasiexponential recovery with equal dexterity but in slightly different ways, e.g., $\hat{f}$ increases for $t > T$ in $M_2(173,174)$, decreases in $M_1(173)$, and asymptotes to its long-term, post-glitch value over $\sim 5$ days in both cases.

We note in closing that the three-glitch Bayes factor $K_3(k)$ does not exceed $10^{1/2}$ for any $k$, i.e., there is no evidence in the above data for a third glitch.

Figure H1. Search for a second glitch in PSR J0835$-$4510 between MJD 57427 and MJD 57810. (Top row) Bayes factor $K_2(k)$ vs. the TOA index $k$. (Bottom row) Point-wise optimal-state sequence $\hat{f}(t_n)$ (left panel) and $\hat{f}(t_n)$ (right panel) vs. $t_n$ for the one-glitch model $M_1(173)$ (blue curves) and two-glitch model $M_2(173,174)$ (red curves). Parameters: see Figure 7.
