Durham Research Online

Deposited in DRO:
23 August 2018

Version of attached file:
Published Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:

Further information on publisher’s website:
https://doi.org/10.1093/mnras/sty2032

Publisher’s copyright statement:
This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society © 2018 The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

• a full bibliographic reference is made to the original source
• a link is made to the metadata record in DRO
• the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.
Explaining changing-look AGN with state transition triggered by rapid mass accretion rate drop

Hirofumi Noda\textsuperscript{1,2,3} and Chris Done\textsuperscript{3}

\textsuperscript{1}Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, 6-3 Aramaki-aoba, Aoba-ku, Sendai, Miyagi 980-8578, Japan
\textsuperscript{2}Astronomical Institute, Tohoku University, 6-3 Aramaki-aoba, Aoba-ku, Sendai, Miyagi 980-8578, Japan
\textsuperscript{3}Centre for Extragalactic Astronomy, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

ABSTRACT

We model the broad-band (optical/UV and X-ray) continuum spectrum of the ‘changing-look’ active galactic nucleus (AGN) Mrk 1018, as it fades from Seyfert 1 to 1.9 in \(\sim 8\) years. The brightest spectrum, with Eddington ratio \(L/L_{\text{Edd}} \sim 0.08\) has a typical type 1 AGN continuum, with a strong ‘soft X-ray excess’ spanning between the UV and soft X-rays. The dimmest spectrum, at \(L/L_{\text{Edd}} \sim 0.006\), is very different in shape as well as luminosity, with the soft excess dropping by much more than the hard X-rays. The soft X-ray excess produces most of the ionizing photons, so its dramatic drop leads to the disappearance of the broad-line region, driving the ‘changing-look’ phenomena. This spectral hardening appears similar to the soft-to-hard state transition in black hole binaries at \(L/L_{\text{Edd}} \sim 0.02\), where the inner disc evaporates into an advection dominated accretion flow, while the overall drop in luminosity appears consistent with the hydrogen ionization disc instability. None the less, both processes happen much faster in Mrk 1018 than predicted by disc theory. We critically examine scaling from Galactic binary systems and show that a major difference is that radiation pressure should be much more important in AGNs, so that the sound speed is much faster than expected from the gas temperature. Including magnetic pressure to stabilize the disc shortens the time-scales even further. We suggest that all changing-look AGNs are similarly associated with the state transition at \(L/L_{\text{Edd}} \sim \) a few per cent.

Key words: galaxies: active – galaxies: individual: Mrk 1018 – galaxies: Seyfert – X-rays: galaxies.

1 INTRODUCTION

The optical variability of active galactic nuclei (AGNs) is puzzling. The viscous time-scale for changing the mass accretion rate on to a supermassive black hole (SMBH) is extremely long for a standard thin disc at radii corresponding to those which produce the optical emission, at least thousands of years for typical AGN parameters. Yet AGNs typically show stochastic optical variability of a factor of a few on time-scales of months–years (e.g. Kelly, Bechtold & Siemiginowska 2009; MacLeod et al. 2010). This flux change could instead be produced from reprocessing of the X-ray emission from a central compact source. An X-ray tail accompanies the disc emission in all AGNs (e.g. Elvis et al. 1994; Lusso & Risaliti 2016), requiring that some of the gravitational power is dissipated in optically thin material rather than in the optically thick, geometrically thin disc. These X-rays vary on rapid time-scales, indicating that they come from a compact region around the SMBH, and the fast variable X-ray heating can result in fast optical variability (Clavel et al. 1992). However, while reprocessing can provide an overall explanation of the fast time-scales, reverberation studies of the optical continuum imply size scales that are generally a factor of a few larger than the thin disc predictions (e.g. Shappee et al. 2014; Noda et al. 2016; Troyer et al. 2016), and detailed models that use the observed X-ray emission to predict the optical flux do not match well to the observed light curves (e.g. Gardner & Done 2017).

None the less, a factor of a few discrepancy in size scale and/or issues with the detailed shape of the light curve is clearly less problematic than the original factor of \(\gtrsim 10^4\) difference in timescale. Hence, some form of X-ray reprocessing is the favoured model for the stochastic, small scale (a few tens of per cent) optical variability (see e.g. Lawrence 2012). However, this is much more difficult for systems that show larger amplitude variability. Some of these can be due to extrinsic effects, such as obscuration events, where material from the clumpy molecular torus moves into or out of the line of sight (e.g. Elitzur 2012; Gardner & Done 2017), or microlensing (e.g. Lawrence et al. 2016), but clearly there is an intrinsic change in luminosity in systems where the emission lines...
also respond. In particular, a small subset of AGNs change from types 1 (showing both broad and narrow lines) to 1.9 (where the broad lines almost disappear) or vice versa, associated with a change by more than a factor of 2 in continuum flux (‘changing‐look’ AGN; see e.g. LaMassa et al. 2015 and references therein). Some of these also show more extreme variability associated with the ‘changing look’, where the continuum changes by more than a factor of \( \gtrsim 10 \) over a time-scale of years (e.g. Graham et al. 2017). These objects are rare, but are now being increasingly discovered in large surveys with repeat cadence (e.g. MacLeod et al. 2016).

We focus here on one particularly well‐studied case of extreme variability associated with a ‘changing look’. The bare (low intrinsic absorption) nucleus Mrk 1018 changed from types 1 to 1.9 around 1980 (Cohen et al. 1986) and from types 1.9 to 1 around 2010 (McElroy et al. 2016). This event is unlikely to be associated with uncovering/covering of the central AGN by a dusty cloud from the torus as the time-scale does not correspond to the expected orbital velocity out at torus size scales, nor is there any progressive change in reddening as the source brightens/dims. Instead, for an intrinsic change in UV continuum, the broad‐line region (BLR) clouds should respond to the changing photoionizing flux (Baldwin et al. 1995; Korista et al. 1997; Korista & Goad 2000), while the narrow‐line region stays constant due to the much longer light travel time‐scales.

The broad lines are most sensitive to the extreme ultraviolet (EUV) part of the spectral energy distribution (SED), and this ionizing flux is mostly unobservable due to Galactic absorption. None of these studies shows that the continuum emission turns down below the expected disc emission, as the spectrum approaches the Lyman limit at \( \sim 1000 \) Å, while it turns up above the extrapolated 2–10 keV X‐ray power‐law emission in the soft X‐ray region. These two spectral breaks point towards each other, and can be fitted by a Comptonization spectrum dominating the ionizing flux in the far‐ultraviolet (FUV) region. The parameters of this Comptonization region are remarkably similar across multiple AGNs, indicating that there is ubiquitous warm (\( kT_w \sim 0.2 \) keV), optically thick material (\( \tau \sim 10–20 \); e.g. Czerny et al. 2003; Gierliński & Done 2004; Mehdipour et al. 2011; Noda et al. 2011, 2013; Matt et al. 2014; Petrucci et al. 2018; Porquet et al. 2018) as well as an inner hard X‐ray corona that is hot (\( kT_h \sim 50–100 \) keV) and optically thin (\( \tau \sim 1 \)). Thus to track the BLR ionizing flux, we need broad‐band, simultaneous UV and X‐ray data sets that can determine the warm and hot Comptonization components as well as any standard disc emission.

The decline of Mrk 1018 from 2008 was fairly well monitored with both optical/UV and X‐ray telescopes on Swift and XMM–Newton (McElroy et al. 2016; Husemann et al. 2016). Hence, we can use these data to model the change in soft X‐ray excess and hard X‐ray emission during the large drop in continuum as the source changed from types 1 to 1.9 over 8 years.

### 2 OBSERVATIONS AND DATA REDUCTION

We analyse the archival data sets of Mrk 1018 derived by the EPIC‐PN and OM on board XMM–Newton, and by the XRT and UVOT on board Swift. We perform the following data reductions to the individual satellite data sets.

**XMM–Newton**: The source was observed on 2008 August 7 (obsid: 0554920301), hereafter 2008 August. We reprocess the PN and OM data sets using ephchain and omichain in SAS‐16.0.0, respectively. We exclude PN events when the 10–12 keV background count rate is more than 0.35 cnt s\(^{-1}\). The PN source spectrum is extracted from a 40‐arcsec‐radius circle region centred at the AGN, while the PN background spectrum is from a 1‐arcmin‐radius circle region off the AGN. The PN response matrix and auxiliary response file are prepared by the rmfgen and arfgen in SAS‐16.0.0, respectively. The U, UVW1, and UVW2 spectra are created via cm2pha using count rates obtained by omdetect in the sequence of omichain and the response files distributed by the XMM–Newton Science Operations Centre.

**Swift**: The source was observed on 2008 June 11 (obsid: 00035776001), 2013 March 1 and June 7–8 (obsid: 00049654001 and 00049654002), and 2016 February 11 and 16 (obsid: 00080898001 and 00080898002). To extract X‐ray spectral files including background files, response matrices, and ancillary response files, we use the automated pipeline of the UK Swift Science Data Center (Evans et al. 2007, 2009). In the spectral extractions, we request to co‐add the two data sets (00049654001, 00049654002) in 2013, and the two (00080898001, 00080898002) in 2016 to get better signal‐to‐noise ratio, and select events with grade 0–12. These procedures provide three Swift/XRT spectra, and hereafter we call them the X‐ray spectra in 2008 June, 2013 and 2016. We extract the V, B, U, UVW1, UVW2, and the UVW2 count rates from data sets with obsid 0035776001, 00049654001, and 00080898001 to make optical–UV spectra in 2008 June, 2013, and 2016, respectively. The commands of uvotimsum and uvot2pha in heasoft 6.20 are used to merge multiple extensions, and to convert count rates into XSPEC readable files. The source region is a 5‐arcsec‐radius circle centred at the AGN, while the background is taken from a 50‐arcsec‐radius circle off the AGN, which includes no other bright sources. We use the response files of the UVOT bands distributed by the UK Swift Science Data Center.

By the grrpha command in heasoft 6.20, each X‐ray spectrum is binned so that a bin includes at least 20 photons to appropriately employ \( \chi^2 \) statistics in spectral fits. After the binning, an energy range up to \( \sim 10 \) keV is noticed in the XMM–Newton/EPIC‐PN spectrum, while we ignore the Swift/XRT spectra in an energy range of \( \gtrsim 6 \) keV where the number of photons is less than 20. Unless otherwise stated, errors shown in tables and figures refer to 90 per cent and 1σ errors, respectively.

### 3 DATA ANALYSIS

We first examine how the broad‐band spectrum of Mrk 1018 varies over time. Fig. 1 shows the spectra from optical to X‐ray in 2008 June (red) and August (black), 2013 (green), and 2016 (blue). It is clear that there is considerable difference in the variability across the spectrum. The optical remains fairly constant, indicating a large contribution from the host galaxy, while the UV drops by over an order of magnitude. The X‐rays also dim, but by a smaller factor, and change shape from that showing a strong soft X‐ray excess below 1 keV in the brightest spectrum (2008 August: black) that disappears in the dimmer spectra.

#### 3.1 Outer disc, soft Comptonization, hard Comptonization model

To quantify the spectral shape variations in optical, UV, and X‐ray, we first perform the broad‐band AGN spectral fit with a standard SMBH accretion model containing the disc emission, the soft X‐ray excess, and the hard Comptonization continuum optxagnf (Done et al. 2012), modified by the Galactic photoabsorption and reddening effects modelled by pihaw and redden, respectively. This assumes a standard disc emissivity (Novikov–Thorne), but that this...
energy is only dissipated in a standard disc structure from the outer radius $R_{\text{out}}$ to a coronal radius $R_{\text{cor}}$ that is a free parameter. Within $R_{\text{cor}}$, the gravitational energy is split, with a fraction $f_{\text{es}}$ powering the coronal X-ray power-law emission, characterized by a photon index $\Gamma_{\text{cor}}$, and the remainder dissipated in a warm Comptonization component, characterized by electron temperature $kT_e$ and optical depth $\tau$. This model has enough flexibility to fit to AGN spectra, but with the physical constraint that all the emission is assumed to be powered by a constant mass accretion rate, parameterized as $L_{\text{Edd}}/e$ for a given SMBH mass $M_{\text{BH}}$ and spin $a_*$. This additional energy conservation assumption allows the model components to be constrained across the unobservable EUV region. The model then is normalized by the physical parameters of $M_{\text{BH}}$, $L_{\text{Edd}}$, $a_*$, and co-moving distance $D$. Mrk 1018 has single-epoch black hole mass of $\log(M_{\text{BH}}/M_\odot) = 7.4-7.9$ (McElroy et al. 2016). We use $\log(M_{\text{BH}}/M_\odot) = 7.84$ in all that follows Ezhiyode et al. (2017) and assume $a_*=0$ following previous SED fits to the 2008 data with optxagnf. We use the co-moving distance of $D \approx 176$ Mpc, corresponding to the source redshift of $z = 0.042$.

The X-ray data also show a clear Fe-Kα at $\sim 6.4$ keV in the XMM–Newton spectrum of 2008 August. The remaining Swift data sets have insufficient signal-to-noise ratio to constrain this feature, but comparison of Chandra observations from 2010 to 2016 (not included here as these have no simultaneous optical/UV data) shows that the line is narrow, neutral, and consistent with an origin in the torus (LaMassa, Yaqoob & Kilgard 2017). Hence, from physically motivated torus reflection models called MYTorus, we include a Compton-scattered component and emission lines (hereafter we call MYTorus$S$ and MYTorus$L$, respectively), calculated with a Comptonized thermal intrinsic continuum with the electron temperature of 100 keV. We fix the optical depth for the intrinsic continuum, inclination angle, and the redshift at 0.8, 0°, and 0.0424, respectively, while the column density of the torus and the normalization are tied between MYTorus$S$ and MYTorus$L$, and left free. In 2008 June, 2013, and 2016, the torus reflection cannot be well constrained due to the usable energy range (see Section 2), we set the normalizations in MYTorus$S$ and MYTorus$L$ to zero.

The optical data clearly also show a constant host galaxy component. This is a late type merger, categorized as a lenticular galaxy (the S0 type in the Hubble classification, e.g. Winter et al. 2010). Following Ezhiyode et al. (2017), we add a model of S0-type host galaxy emission template named hostpol (Polletta et al. 2007). We tie the normalization of hostpol across the four spectra. The total model fit to the four data sets simultaneously is then $\text{phabs*redden*(optxagnf + MYTorusS + MYTorusL + hostpol)}$. The column density $N_{\text{H}}$ of phabs is fixed at the Galactic value of $2.43 \times 10^{22}$ cm$^{-2}$, while $E(B-V)$ of redden is tied to $1.5 \times N_{\text{H}}/(10^{22}$ cm$^{-2}$) utilizing the Galactic gas-dust ratio. The Swift data especially cannot constrain all the parameters of the optxagnf model. We tie the coronal outer radius $R_{\text{cor}}$ between 2008 June and August, and between 2013 and 2016.

In addition, we fix $\Gamma_{\text{H}}$ in 2016 at 1.66 following the 2016 phenomenological fit result without the MYTorus models in LaMassa et al. (2017), considering that the normalizations of the MYTorus models are fixed at zero.

We first fit the four spectra by making the electron temperature $kT_e$ and optical depth $\tau$ of the soft Comptonization in optxagnf free and tied among all. In the fit, the systematic errors of 3 per cent are employed to account for the calibration uncertainties mainly in absolute flux scaling between the OM and UVOT (e.g. Lobban et al. 2016). The fit is not acceptable, with $\chi^2$/d.o.f.$=347.5/269$, mainly because of large negative residuals in the soft X-ray band in 2008 June, showing that the soft X-ray excess spectral shape is different between 2008 June and August. Allowing $kT_e$ and $\tau$ in 2008 June to differ from those in 2008 Aug gives an almost acceptable fit with $\chi^2$/d.o.f.$=326.5/267$. Fig. 1 (left) shows the data fit to this model, together with the residuals, while Fig. 1 (right) shows the unabsorbed model components. The best-fitting parameter values and 90 per cent confidence uncertainties are given in Table 1.

### 3.2 No outer disc model

Recently, the optical/UV emission from an AGN has been interpreted by only a soft X-ray excess component without an outer standard disc (e.g. Mehdipour et al. 2015; Petrucci et al. 2018). However, there is still implicitly an outer disc as the model requires that there is some source of seed photons for the warm Comptonization. Petrucci et al. (2018) assume that these seed photons have a disc blackbody distribution, and this is a key to their optical/UV slope being able to fit many of their objects. The maximum
seed photon temperature of the disc blackbody, $kT_{\text{bb,S}}$, is left as a free parameter. This contrasts with the optxagnf model where the original picture was that the seed photons came from the disc underlying the warm Comptonization region (see appendix A in Done et al. 2012), but since these would affect the spectrum only in the unobservable EUV regime, the code approximated these by blackbody seed photons at the temperature of the hottest part of the standard disc i.e. $T(R_{\text{ISCO}})$. Thus, we try the Petrucci et al. (2018) model of phabs*redden*(nthcomp_S + nthcomp_H + MYTorusS + MYTorusL + hostpol). The warm Comptonization is modelled by nthcomp_S, while nthcomp_H models the coronal power-law emission. The warm Comptonization is assumed to have the seed photon distribution, characterized by a disc blackbody with maximum temperature $kT_{\text{bb,S}}$.

Similarly to the optxagnf fits above, the data strongly require that the soft X-ray excess shape in 2008 June is different from 2008 August. Hence, if we tie the photon index of the warm Comptonization between the four spectra then the data strongly require that its electron temperature is different. Conversely, if we tie the electron temperature then we require that the photon index of the warm Comptonization changes. We choose then to tie the warm Comptonization spectral index $\Gamma_S$ between the four spectra, as Petrucci et al. (2018) give some physical reasons for fixing this from reprocessing of the warm Comptonization in a passive, optically thick disc beneath the corona. Because the warm Comptonization component is not significantly detected in the soft X-ray emission in 2013 and 2016, we tie the nthcomp_S coronal temperature $kT_{\text{e,S}}$ across all the 2008 June, 2013, and 2016 data sets. We allow the seed photon temperature to be free in all data sets. The coronal Comptonization seed photon temperature is set to that of the soft Comptonization in all data sets, but it is assumed to be blackbody rather than disc blackbody. The photon index $\Gamma_H$ is fixed at those obtained by the fit with optxagnf to compare the reproducibility of the optical–UV band between the two models, and we fix the nthcomp_H coronal temperature $kT_{\text{e,H}} = 100\kT$ as in optxagnf. Reflection, host galaxy, and absorption are included as before.

This somewhat different spectral decomposition gives a similarly good fit with $\chi^2$/d.o.f.=314.2/269 (see Table 1), showing that the optical can be modelled either by assuming an outer standard disc or by assuming that there is optically thick material that reprocesses the emission from a warm Comptonizing layer above it. The best-fitting model is shown against the data in Fig. 2 (left) and as unabsorbed model components in Fig. 2 (right).

### 3.3 Changing broad-band continuum spectral shape

These models are both acceptable fits to the data, and allow us to characterize how the accretion flow emission changes as a function of $L/L_{\text{edd}}$. This is a direct parameter of the optxagnf modelling, but is not explicitly included in the warm and hot Comptonization description. Instead, we integrate the model components to get the luminosity in the warm and hard X-ray Comptonizing regions. This allows us to calculate $L/L_{\text{edd}}$ using the same mass and distance as before.

Both models have hard and soft Comptonization components, so Fig. 3 shows the change of $L_H(L_S + L_H)$, which is equivalent to $F_H$, when we use optxagnf (black) and the two Comptonization...
components with no outer disc (grey). Both models show similar behaviour in that the warm and hot Comptonization components show comparable luminosity at \(L/L_{\text{Edd}} \sim 0.08\), but that the ratio of the hot component gets larger when \(L/L_{\text{Edd}}\) becomes smaller, and it dominates the spectrum at \(L/L_{\text{Edd}} \sim 0.006\). Note that the errors in the case of the two Comptonization components are relatively small mainly because the spectral shapes are fixed.

### 3.4 Summary of results

(i) The observed 2–10 keV flux drops by a factor of \(\sim 6–7\) over \(\sim 8\) years, while the UV luminosity drops by a factor of \(\sim 20\).

(ii) Broad-band spectral models show that the total luminosity drops by a factor of \(13–15\), from \(L/L_{\text{Edd}} \sim 0.08\) to \(\sim 0.006\).

(iii) The warm Comptonization region (soft X-ray excess) is very obvious when the source is brightest, where the broad-band SED looks like a typical Seyfert type 1.

(iv) The warm Comptonization region may disappear completely as the source dims if there is residual emission from an outer standard disc (optxagnf model). Alternatively, the faint optical/UV emission may be produced entirely by the warm Comptonization, and not require any outer standard disc, in which case the warm Comptonization drops by a factor of \(\sim 60\).

Previous models of these data assumed a simple disc and power-law model for the optical/UV and X-ray flux, respectively, and did not specifically fit to the soft X-ray excess (Husemann et al. 2016). This is important as Figs 1 and 2 show that almost all of the EUV photons, the ones which most contribute to the BLR ionization, are in the soft excess emission. Therefore, the changing-look phenomenon in Mrk 1018 from types 1.9 to 1 around 1980 (Cohen et al. 1986) and from types 1 to 1.9 around 2010 (McElroy et al. 2016) is clearly controlled by the increase and decrease of the warm Comptonization region, respectively. We will explicitly explore the effect of this changing spectral shape as well as luminosity on the predicted BLR emission from an locally optimally-emitting cloud (LOC) model in a following paper.

The soft X-ray excess probably forms the seed photons for the hot Comptonization component, so a decrease in the ratio between soft and hard Comptonization means a decrease in seed photons for Compton cooling of the hard X-ray emitting electrons relative to the power injected into heating these electrons (e.g. Beloborodov 1999). Hence, the large decrease of the warm Comptonization flux is likely to lead to spectral hardening of the hot Comptonization continuum, although significant photon index decrease of the hot Comptonization component is not supported by the data. This can be tested by longer X-ray observations especially of faint epochs in the future.

However, the remaining results are much more difficult to explain, as discussed below.

### 4 DISCUSSION

#### 4.1 Standard accretion disc time-scales

Fundamentally, we see a drop in total accretion luminosity by a factor of \(13–15\) over a time-scale of \(\sim 8\) years. This is a problem for any standard disc model. There is a hierarchy of time-scales for a stan-
dard thin disc, with the dynamical time-scale being fastest, then the thermal time-scale and then the viscous. We calculate these for Mrk 1018, assuming that the optical/UV (typical temperature $3 \times 10^4$ K) is emitted from a standard thin disc at $R \sim 100R_g$. This implies $H/R = c_s/v_\phi \sim 5 \times 10^{-4}$ (where the sound speed $c_s = \sqrt{kT/m}$ and Keplerian velocity $v_\phi = \sqrt{GM/R} = \sqrt{R_g/R_c}$). The optical emission regions in AGN discs are much thinner than the optical parts.

The accretion rate is large enough. This requires a change in the effective viscosity, often parametrized as a jump in $\alpha$ from neutral to ionized, perhaps as an outcome of the MRI either requiring free electrons (Menou et al. 1999) or driving turbulence (Coleman et al. 2016). This results in a heating front propagating inwards, travelling at a speed which is shorter than the viscous time-scale by a factor of $H/R$ (Hameury et al. 1998; Dubus, Hameury & Lasota 2001; Lasota 2001; Hameury, Viallet & Lasota 2009).

The mass accretion rate through the disc is now larger than the mass accretion rate from the companion star, so it cannot maintain this steady-state structure. When the outer disc temperature falls below $10^4$ K, the instability works in reverse, triggering a cooling wave through the disc that switches hydrogen back to neutral and propagates at a speed faster than the heating front by a factor of $\alpha_{\text{in}}/\alpha_{\text{out}}$.

In accreting white dwarf systems, the cooling wave is triggered quite quickly after the heating wave, shutting off the outburst after only a small amount of the total disc material is accreted (Osaki 1996). Instead, in BHBs and neutron stars, the copious X-ray emission from the inner disc that is associated with the much larger gravitational potential irradiates the outer disc material, keeping it ionized. The disc material accretes viscously (exponential decline) until the X-ray luminosity drops enough for the outer disc to dip below the H-ionization temperature. The cooling front starts, but its inward speed is set instead by the viscous shrinking of the irradiated zone, giving a linear decline (King & Ritter 1998).

Thus, the mass accretion rate on to the black hole can increase by many orders of magnitude on time-scales of days for the heating front propagation (Dubus et al. 2001), while the decline is rather slow as it is controlled by the viscous time of the outer disc. This mechanism is well tested in BHBs and neutron stars (Shahbaz et al. 1999; Coriat, Fender & Dubus 2012), and in white dwarfs (Osaki 1996).

4.3 State change in BHBs

The dramatic variability associated with the H-ionization DIM can also trigger a spectral transition, which most probably represents a switch in the nature of the accretion flow from a hot, optically thin, geometrically thick flow (low/hard, Comptonization dominated e.g. the advection dominated accretion flow: Narayan & Yi 1995) to a cool, optically thick, geometrically thin disc (high/soft, thermal dominated). This transition can occur over a wide range in luminosity on the fast rise to outburst, but this may be due to the flow being far from steady state (Smith, Heindl & Swank 2002; Gladstone, Done & Gierliński 2007). Instead, the reverse transition on the slow decline occurs at a rather stable luminosity around 0.02$L_{Edd}$ (Maccarone 2003, see e.g. the review by Done, Gierliński & Kubota 2007).

Cygnus X-1 shows this spectral transition decoupled from the DIM as its outer disc is stable. This shows that the hard state is on average a factor of $3-4$ lower in bolometric luminosity than the soft state (Zdziarski et al. 2002). This could be entirely produced by the state change alone, with a constant mass accretion rate as the hard state should be somewhat radiatively inefficient, and some of the accretion energy is required to power the potentially substantial kinetic luminosity of the compact jet seen in the radio emission (Zdziarski et al. 2002). However, some change in mass accretion rate is required to trigger the state change, so it seems more reasonable to assume that there are long time-scale fluctuations in mass accretion rate of perhaps a factor of $1.5-2$ so that the bolometric luminosity at a single transition only changes by a factor of $\sim 1.5-2$. Thus, a
state change can give a drop in luminosity by a factor of $\sim 1.5$–$3$ without requiring any change in mass accretion rate.

The soft-to-hard transition involves a quite dramatic decrease in surface density and concomitant increase in temperature (see e.g. the review by Yuan & Narayan 2014, fig. 3). Changing the surface density radially requires a viscous time-scale but here the disc evaporates so it can lose mass vertically as it heats up to the virial temperature. Shaviv & Wehres (1986) show that even a standard Shakura–Sunyaev disc will have a thermal instability in its outer layers, as by definition the photosphere is rather inefficient at cooling, so any heat dissipated here heats up this material to the virial temperature. The next layer of the disc is then the photosphere, so it too heats up, and this process can lead to the entire disc switching into the hot solution (see e.g. Das & Sharma 2013; Hogg & Reynolds 2017). However, while this does involve vertical rather than radial motion, the time-scales are still the radial viscous time-scale (Czerny 2006) as can be seen most clearly from the reverse process of condensation at the soft-to-hard transition. The hot flow collapses down into the midplane, but its surface density is so much smaller than that of a standard disc that it requires a viscous time-scale for radial accretion from the outer disc to build up the material back into a standard disc. Indeed, this is seen in BHBs, where the time-scale for the soft-to-hard state transition is about 0.5 days which is close to the thin disc viscous time-scale at a few hundreds $R_g$ which is the expected radius of the evaporated region (Mayer & Pringle 2007).

### 4.4 Differences between AGNs and BHBs

Fundamentally, the behaviour of Mrk 1018 looks like a large drop in luminosity, similar to that seen in the transient BHBs that are triggered by the DIM, together with a state change as the luminosity drops below a few percent of Eddington. However, as has often been noted, the time-scales predicted by the thin disc models or by scaling from Galactic binaries are far too long. Here, we critically assess these time-scales, pointing out some significant differences between AGNs and BHBs.

A key difference between AGNs and BHBs is that the disc in AGNs is self-gravitating at around a few hundreds $R_g$ (e.g. Laor & Netzer 1989). This outer disc radius is much smaller than the disc in BHBs, which is truncated by tidal effects from the companion star at typical radii of $R \sim 10^3 R_g$. The combined effect of mass and radius predicts all time-scale scales as $M_{\text{out}}^{3/2}$ where $R_{\text{out}} = R_{\text{out}} R_g$.

Another key difference, connected to the smaller $R_{\text{out}}$ in AGNs, is in the importance of irradiation. This is effective only at large radii (Cunningham 1975; Shidatsu, Done & Ueda 2016) so becomes mostly unimportant in AGN discs (Gardner & Done 2017). Thus the DIM in AGNs should be more comparable to the DIM in white dwarf systems, where irradiation is not important due to the much weaker white dwarf gravitational potential (Hameury et al. 2009). Thus the rise and decay times in AGNs should be set by the heating/cooling front propagation speeds, not by the viscous decay time-scale.

We scale the DIM time-scale from the observed dwarf novae $\sim 1$ day time-scales of the heating/cooling front for a $1 M_\odot$ white dwarf with $R_{\text{out}} \sim 10^3$ up to an AGN with mass $7 \times 10^7 M_\odot$ and $R_{\text{out}} \sim 10^5$. This predicts a time-scale of $\sim 6$ years for the front propagation in AGNs, similar to that observed.

This however, assumed that $H/R$ is similar in the AGN disc as in CVs, whereas a standard thin disc has $H/R = c_s/R_{\text{in}} \sim (kT_{\text{gas}}/m_e c^2)^{1/2}/(1/r_{\text{opt}})^{1/2}$ which is much smaller in AGNs than in binaries (see Section 4.2). The front propagation time-scale depends on $R/H$ i.e. is a factor of 30 longer than a simple scaling from CVs, lengthening the predicted time-scale to $\sim 1000$ years. This is comparable to the time-scales seen in the AGN DIM simulations of Hameury et al. (2009).

There is one more break in the scaling between AGN and binaries, which is the importance of radiation pressure. The lower densities typical of AGN discs mean that gas pressure is lower at the same temperature, so radiation pressure is more likely to dominate. Thus, the sound speed is set by the radiation pressure rather than the gas pressure, and is faster than in gas pressure dominated discs. For Mrk 1018 at its peak luminosity, radiation pressure is a factor of $\sim 75$ larger than the gas pressure in the optical emission region (e.g. Laor & Netzer 1989). Radiation pressure was not included in the AGN DIM simulations of Hameury et al. (2009), as radiation-pressure-dominated discs are unstable (Shakura & Sunyaev 1973). This radiation pressure thermal-viscous instability is seen in all simulations where heating scales with total (gas plus radiation) pressure (e.g. Szuszkiewicz & Miller 2001), resulting in characteristic limit cycles between a slim, super-Eddington inner disc, and a much dimmer standard disc at all luminosities $L > 0.05L_{\text{Edd}}$.

We note that current MRI simulations do not show magnetic pressure dominating as they show the instability (Jiang, Stone & Davis 2013).

The large drop in luminosity could then be the DIM cooling front, propagating through a radiation and/or magnetic-pressure-dominated AGN disc. As the luminosity drops below a few percent of Eddington, it triggers the state transition as seen in BHBs. This occurs at more or less constant luminosity, but still requires a viscous time-scale to change the surface density by evaporation/condensation. Scaling the BHBs soft-to-hard transition time-scale of $\sim 0.5$ days by the mass difference alone (as both only affect the inner $100R_g$ gives $\sim 10^4$ years, but this does not take into account the difference in $H/R$ between the two systems, making the predicted time-scale even longer. Again, this can be solved if the disc is dominated by radiation and/or magnetic pressure instead of gas pressure. Once the system has made the transition to a hot accretion flow it should vary on the viscous time-scale of the hot flow, which is of order the thermal time-scale as $H/R$ is of order unity.

### 4.5 Observational predictions

As we suggest in Section 4.4, Mrk 1018 likely experiences both drop in mass accretion rate due to the cooling front propagation and spectral state transition by disc evaporation. Considering these processes individually, extremely variable AGNs are divided into three groups: (1) sources showing a state change due to disc evaporation/condensation associated with a factor of $1.5$–$3$ decrease/increase in luminosity, (2) sources with large mass accre-
tion rate change due to the thermal front propagation due to the H-ionization instability, and (3) sources showing both.

We suggest that many of changing-look AGNs are included in the group (1) or (3), and their variations commonly cross the state transition boundary of $L/L_\text{Edd}$ $\sim$ a few per cent. Sources with the variability amplitude of a factor of a few are probably categorized into group (1), while those with amplitude of a factor of more than 10 should be into group (3). Interestingly, many of the known changing-look AGNs, e.g. NGC 2617 (Shappee et al. 2014), Mrk 590 (Denney et al. 2014), and SDSS J0519+0033 (LaMassa et al. 2015) actually show $L/L_\text{Edd}$ $\sim$ a few per cent. A similar drop in the soft X-ray excess is seen by Rivers et al. (2012) in Mrk 590 between 2004 and 2011, which we suggest is due to a state change associated with the changing-look phenomena, as in Mrk 1018. Recently, several tens changing-look AGNs have been newly reported (MacLeod et al. 2016; Yang et al. 2018), but their Eddington ratios have not been reported yet. We thus predict that these new changing-look AGNs have $L/L_\text{Edd}$ $\sim$ a few per cent as well. Vice versa, we also predict that AGNs that exhibit drastic $L/L_\text{Edd}$ variation crossing a few per cent, e.g. NGC 3227 (Noda et al. 2014) and NGC 3516 (Noda et al. 2013) will experience changing-look phenomena simultaneously to drastic soft excess variation at $L/L_\text{Edd}$ $\sim$ a few per cent. We need further continuous optical spectroscopic monitoring spontaneously to soft X-ray monitoring on these sources for unconfirmed changing-look phenomena.

As MacLeod et al. (2010) reported, quasar variability cannot be explained by stochastic optical fluctuations driven by accretion rate, and there is additional variability depending on black hole mass and/or luminosity. If the thermal front propagation occurs as group (2), variability amplitude is determined by swept disc area that varies with black hole mass, i.e. luminosity with fixed accretion rate. Thus, the quasar variability may include that by the front propagation through a thin disc (e.g. Hameury et al. 2009), and some quasars are perhaps included into group (2) although they do not show the changing-look features. This can be tested by investigating detailed radial profile of disc temperature in quasars, which might include discontinuous drop if a thermal front is present. If their $L/L_\text{Edd}$ could cross a few per cent by the front propagation like Mrk 1018, they would show state transition, and enter group (3) being a changing-look quasar as reported by LaMassa et al. (2015).

5 CONCLUSIONS

In this paper, we show for the first time how the changing-look phenomenon is linked to a large change in the warm Comptonization component that forms the soft X-ray excess. The bolometric luminosity drops from $\sim 0.08L_\text{Edd}$ to $\sim 0.006L_\text{Edd}$, but the hard X-rays drop by only a factor of $\sim 7$ while the warm Comptonization decreases by at least a factor of $\sim 60$ and may disappear completely. This marked hardening of the spectrum looks very like the transition seen in stellar mass BHBS around $0.02L_\text{Edd}$, which is generally interpreted as the inner regions of an optically thick accretion disc evaporating into an hot accretion flow. We suggest that all the true changing-look AGNs make this spectral transition to/from an EUV bright accretion disc to a hot inner flow, and that the change in BLR properties is a mostly a consequence of the changing shape of the ionizing spectrum.

While the state transition describes the changing shape of the spectrum, the time-scale for disc evaporation/condensation should be the viscous time-scale of the disc. This predicts far too long a time-scale compared to that observed. There is a similar problem in explaining the background decline in flux that triggers the transition.

In Mrk 1018, the optical flux drops by more than a factor of 10, requiring a true change in mass accretion rate over a time-scale of $\sim 10$ years, orders of magnitude shorter than the viscous time-scale. We discuss all the factors that change between AGN and binary discs so as to properly scale between them, as sizes and scale heights should be different. However, one key difference that has not been considered so far is that AGN discs are dominated by radiation pressure even in standard disc models. Thus, the sound speed, and hence the viscous speed, is much faster for a given gas temperature. Including magnetic pressure to stabilize the disc again increases the sound speed, making it feasible that there can be correspondence between the observations and theory. We suggest the hydrogen ionization disc instability, where the heating/cooling front travels on a speed that is faster than the viscous time-scale by a factor of $H/R$, as the underlying mechanism for the flux drop. We speculate that this is the origin of the most extreme AGN variability. Proper modelling of this is required in order to predict the behaviour in more detail.

ACKNOWLEDGEMENTS

We thank the referee for his/her careful read of this manuscript and valuable comments. HN is supported by Program for Establishing a Consortium for the Development of Human Resources in Science and Technology, Japan Science and Technology Agency (JST). CD acknowledges the Science and Technology Facilities Council (STFC) through grant ST/P000541/1 for support. This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester.

REFERENCES

Done C., Gierlinski M., Kabota A., 2007, A&AR, 15, 1