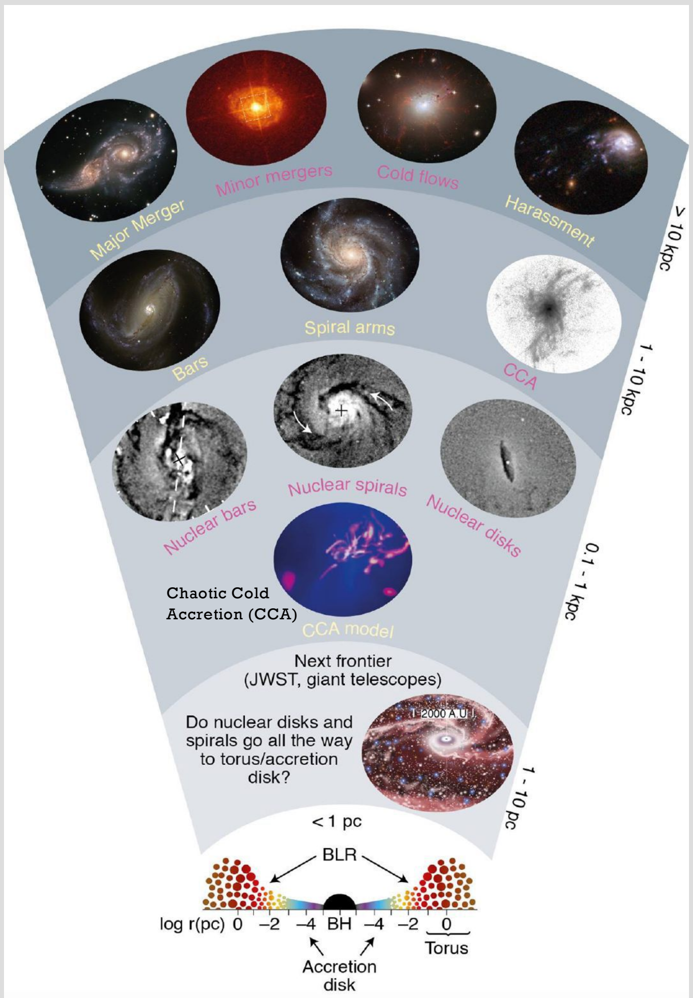
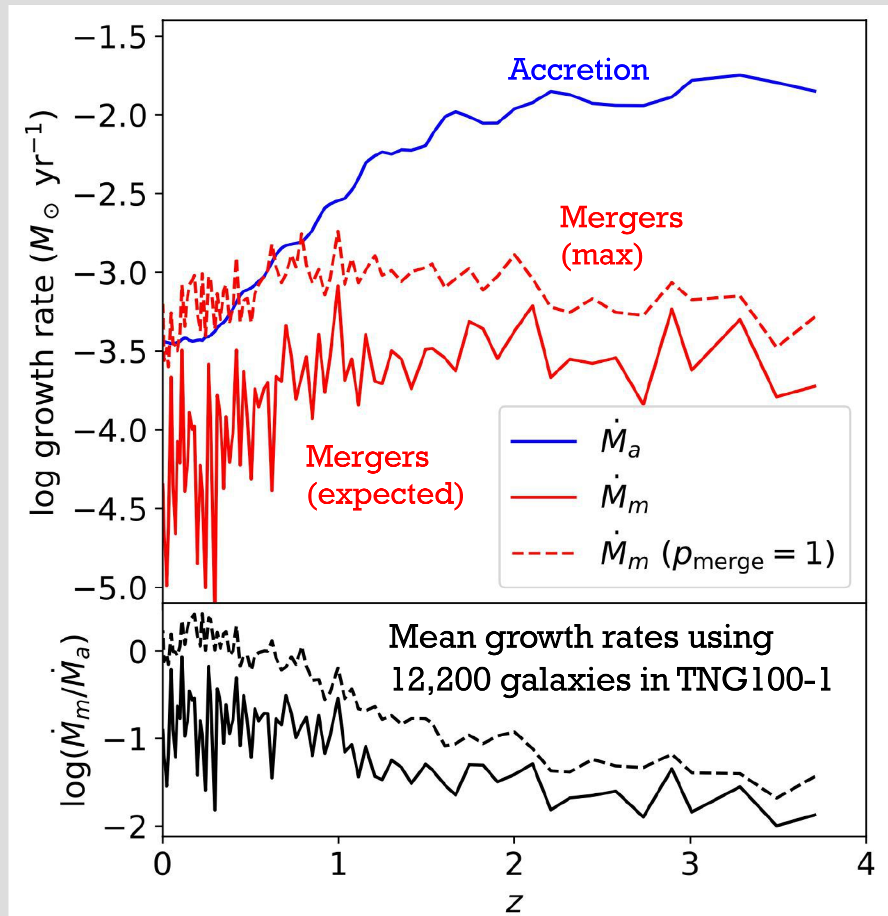
UV Insights into Supermassive Black Hole Growth

W.N. Brandt (Penn State)

Storchi-Bergmann & Schnorr-Müller (2019)





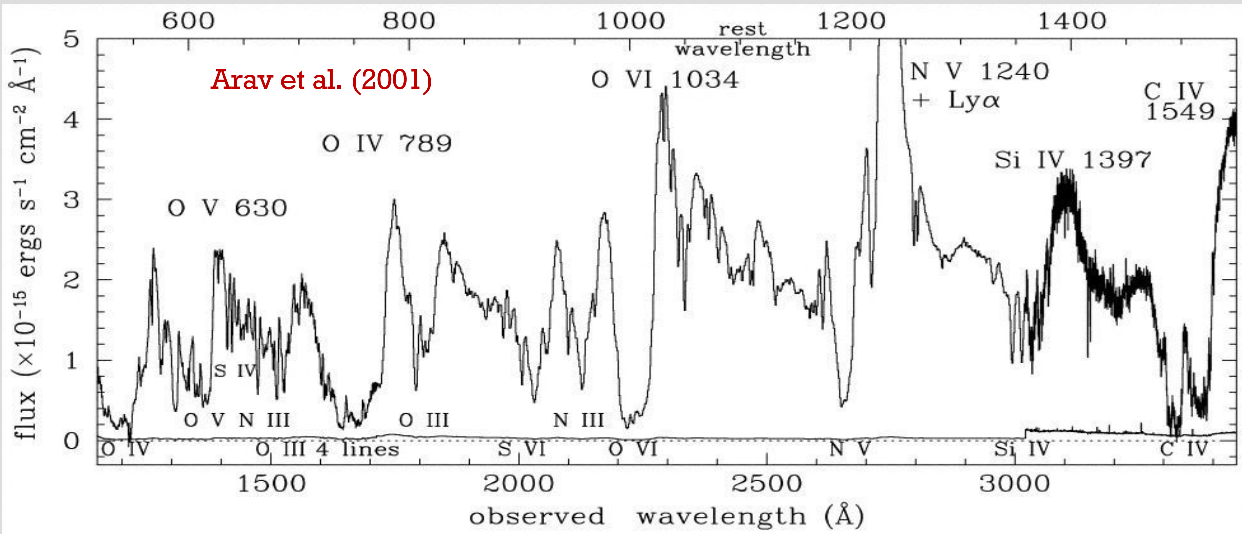
Zou et al. (2024)

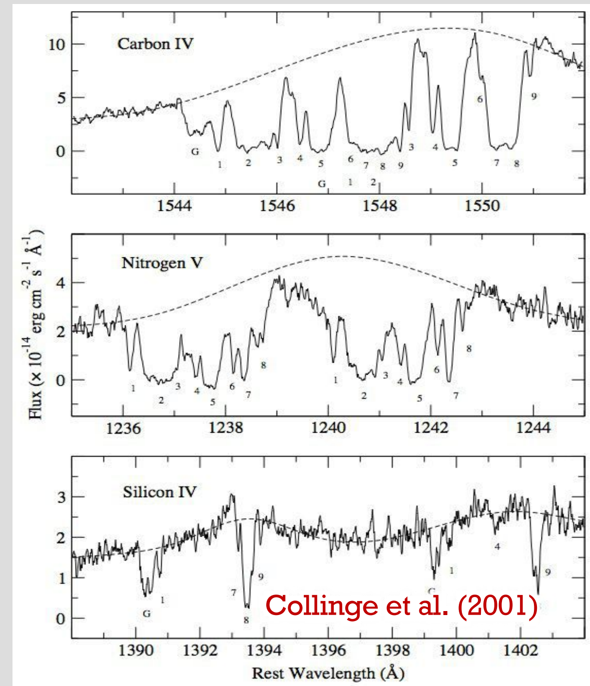
Given the dominant role of accretion of galactic gas, I will primarily focus on this topic.

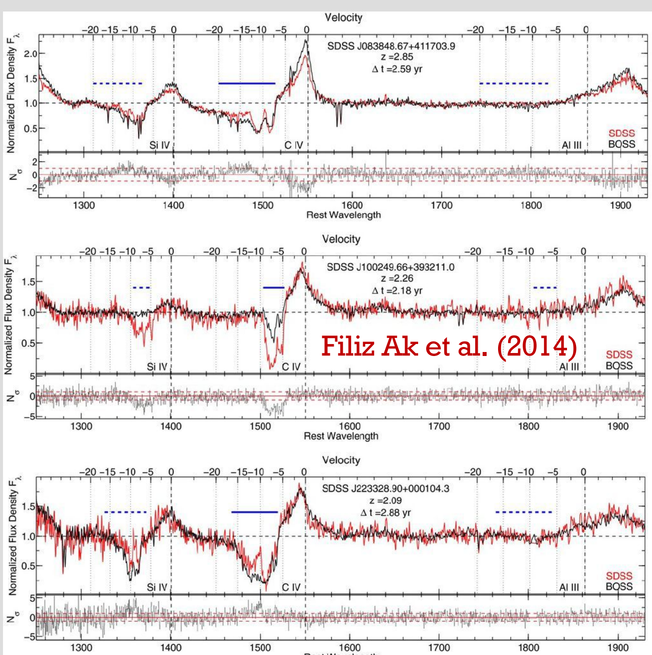
This process primarily produces broad-band UV radiation, so UV observations are highly informative about SMBH growth – though puzzling.

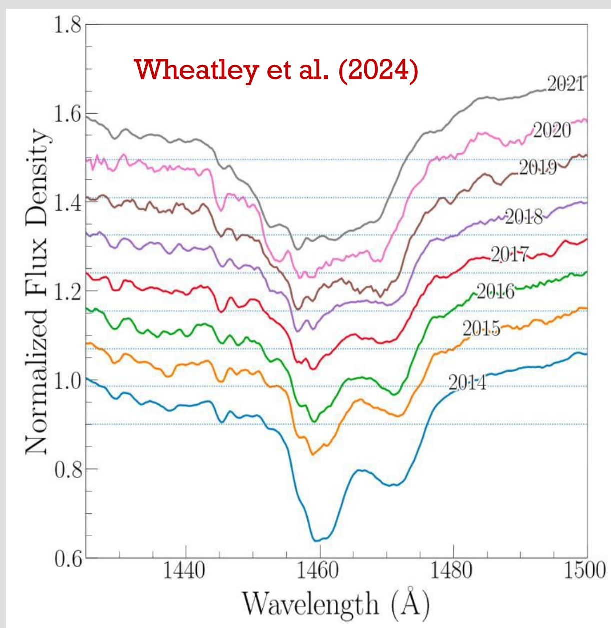
I'll cover insights from both *static* and *time-domain* UV observations.

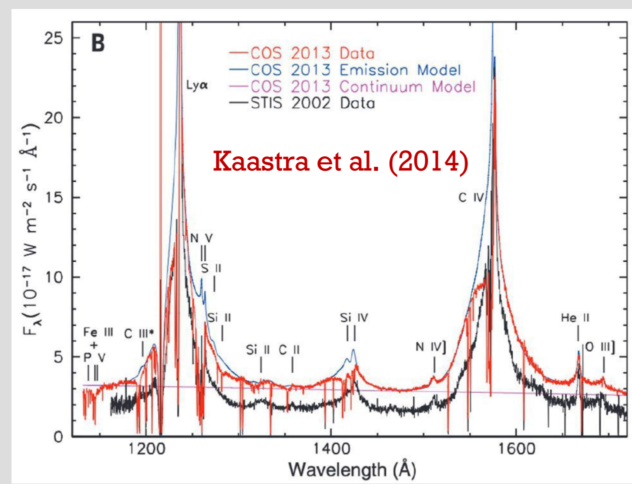
SMBHs Are Sloppy Eaters











Static UV Constraints
on
SMBH Growth

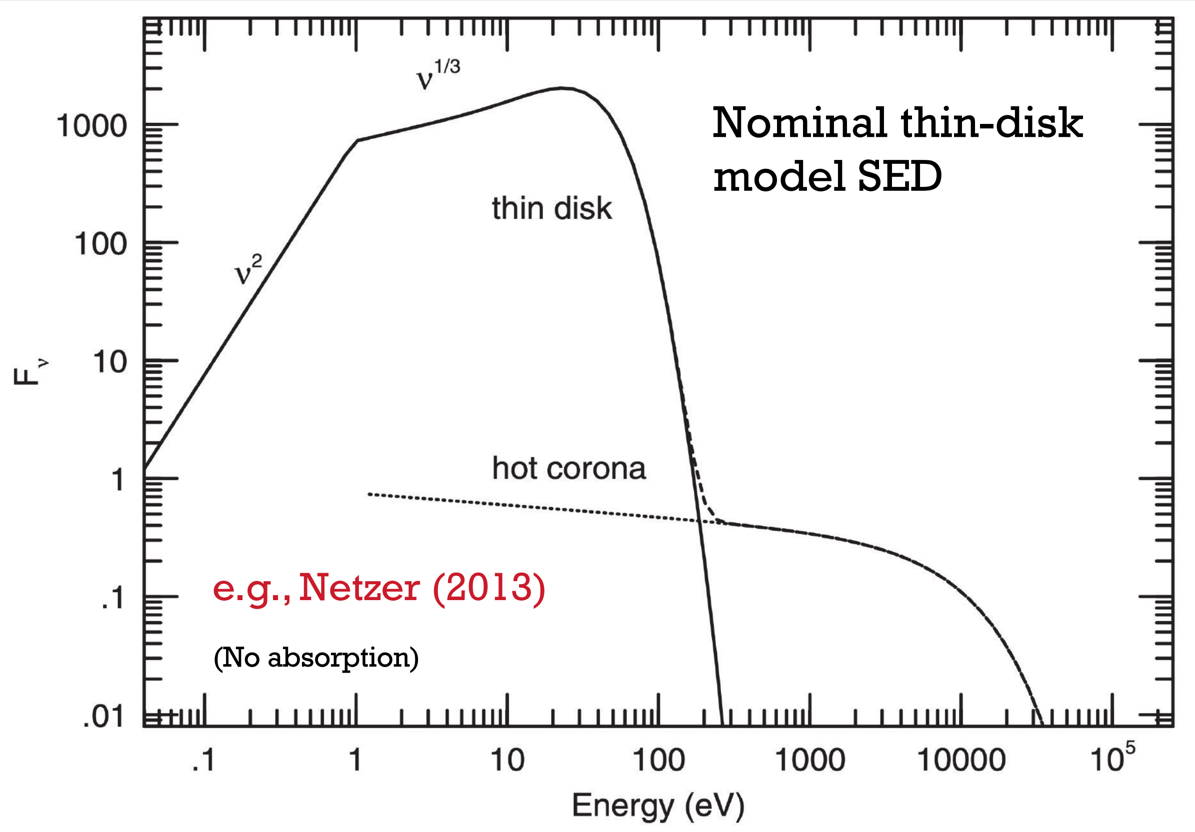
Typical AGN Spectral Energy Distribution

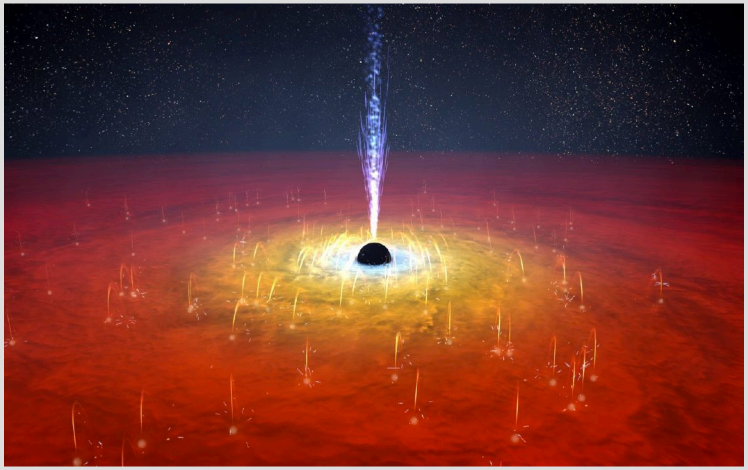
Standard thin accretion-disk model for a SMBH *roughly* predicts UV/optical continuum.

Hot corona added to explain the X-ray emission via Compton up-scattering.

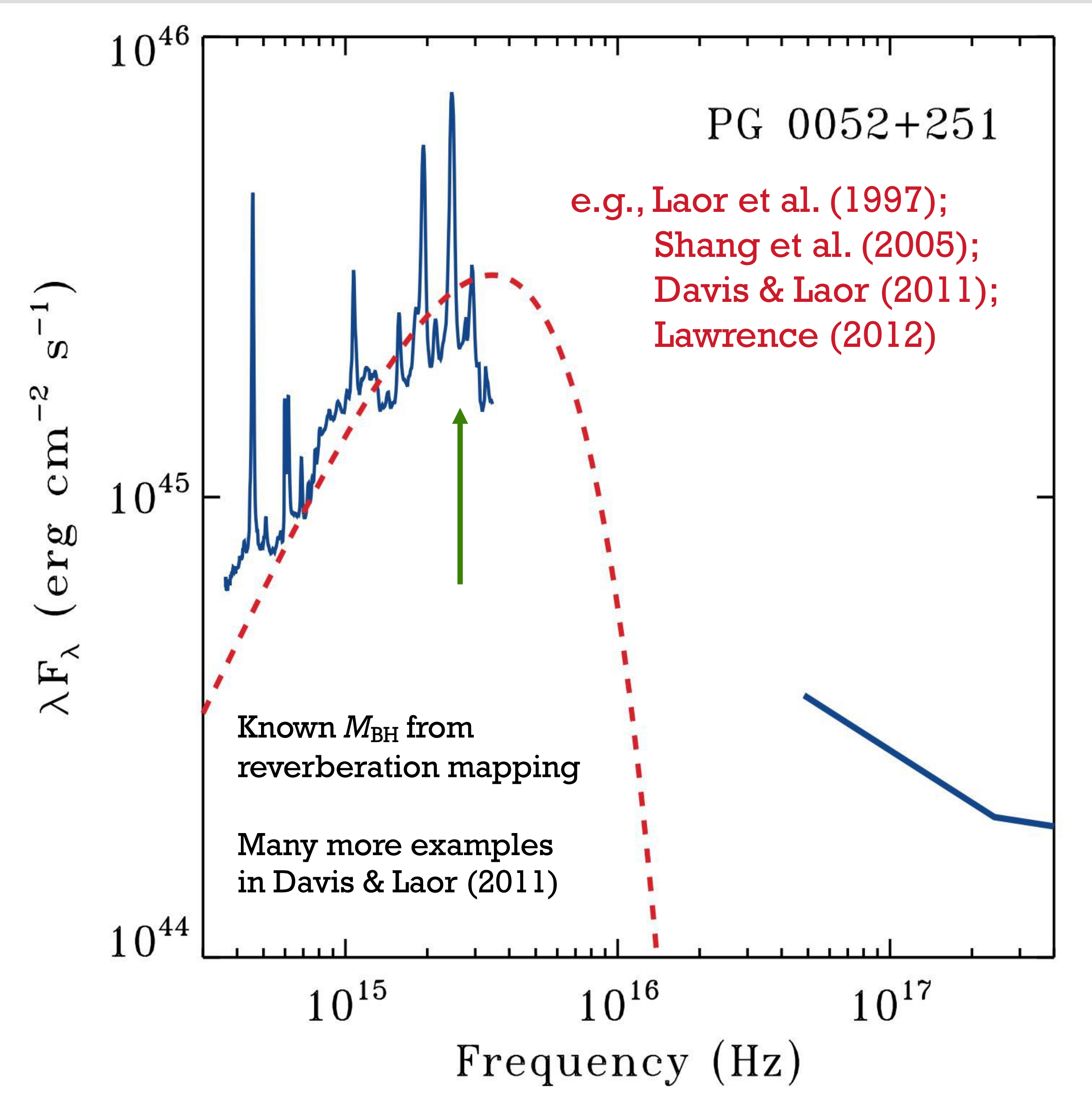
But challenges (and disputes!) emerge when detailed UV observational constraints are considered – much further work needed with hopefully big discoveries waiting.

"Epic magical thinking" – Senior AGN Researcher, 2023, arXiv:2308.04621





UV Temperature Problem



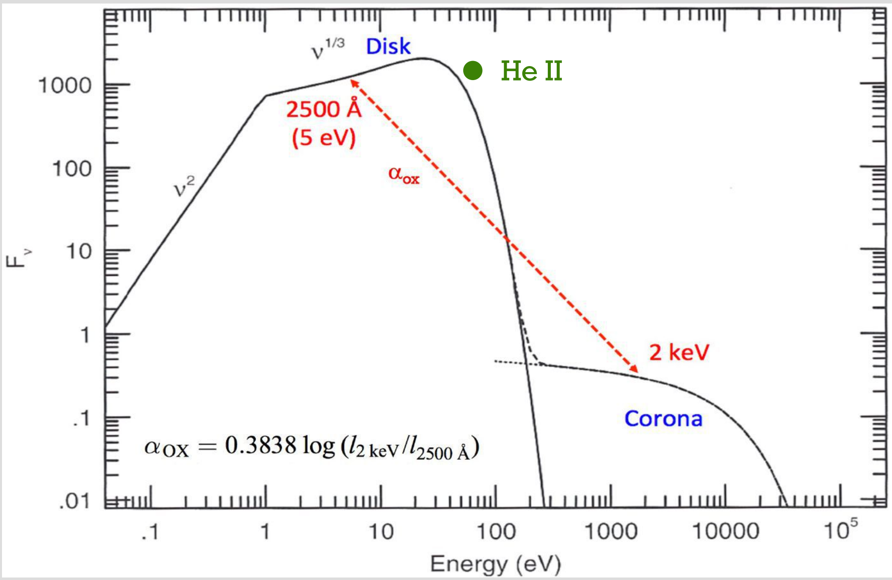
Model spectra generally exceed data above ~ 1000 Å.

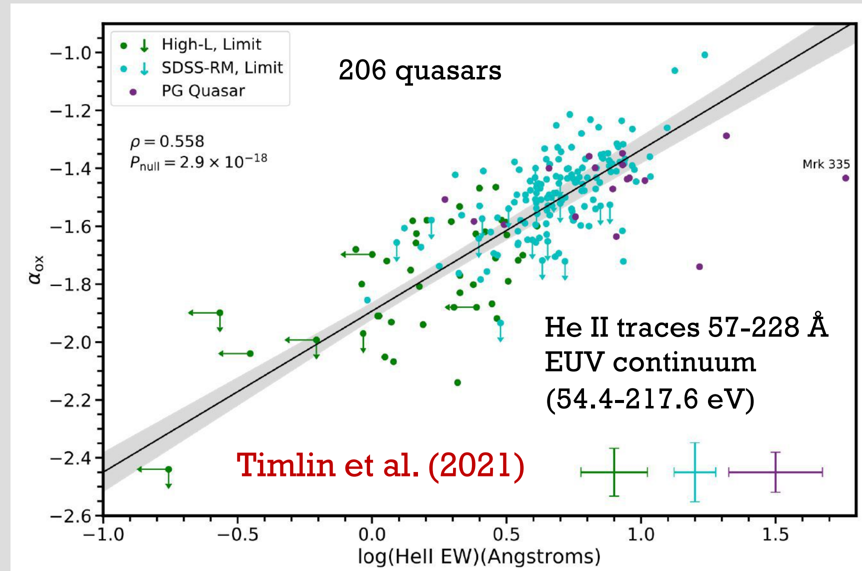
Observed T_{max} ~ 50,000 K, while accretion disks can have T_{max} up to ~ 250,000 K.

Cause of T_{max} discrepancy unclear:

- Outflows truncating disk?
- Advection?
- Reprocessing close to inner disk?
- Dust reddening?

UV-EUV-X-ray Correlations

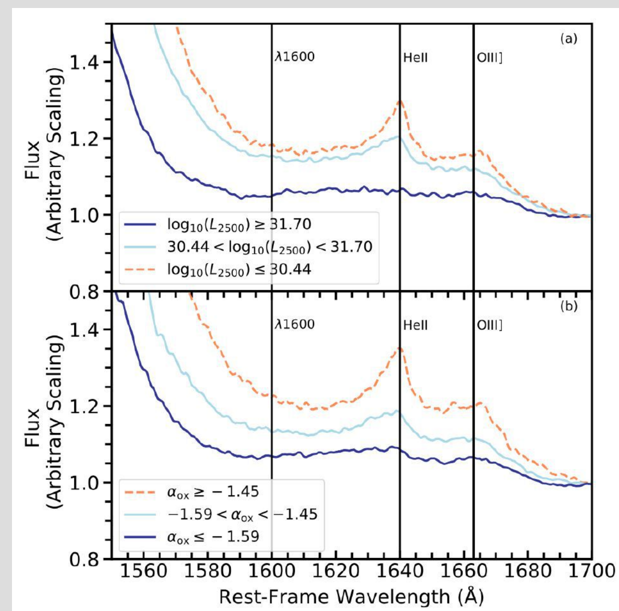




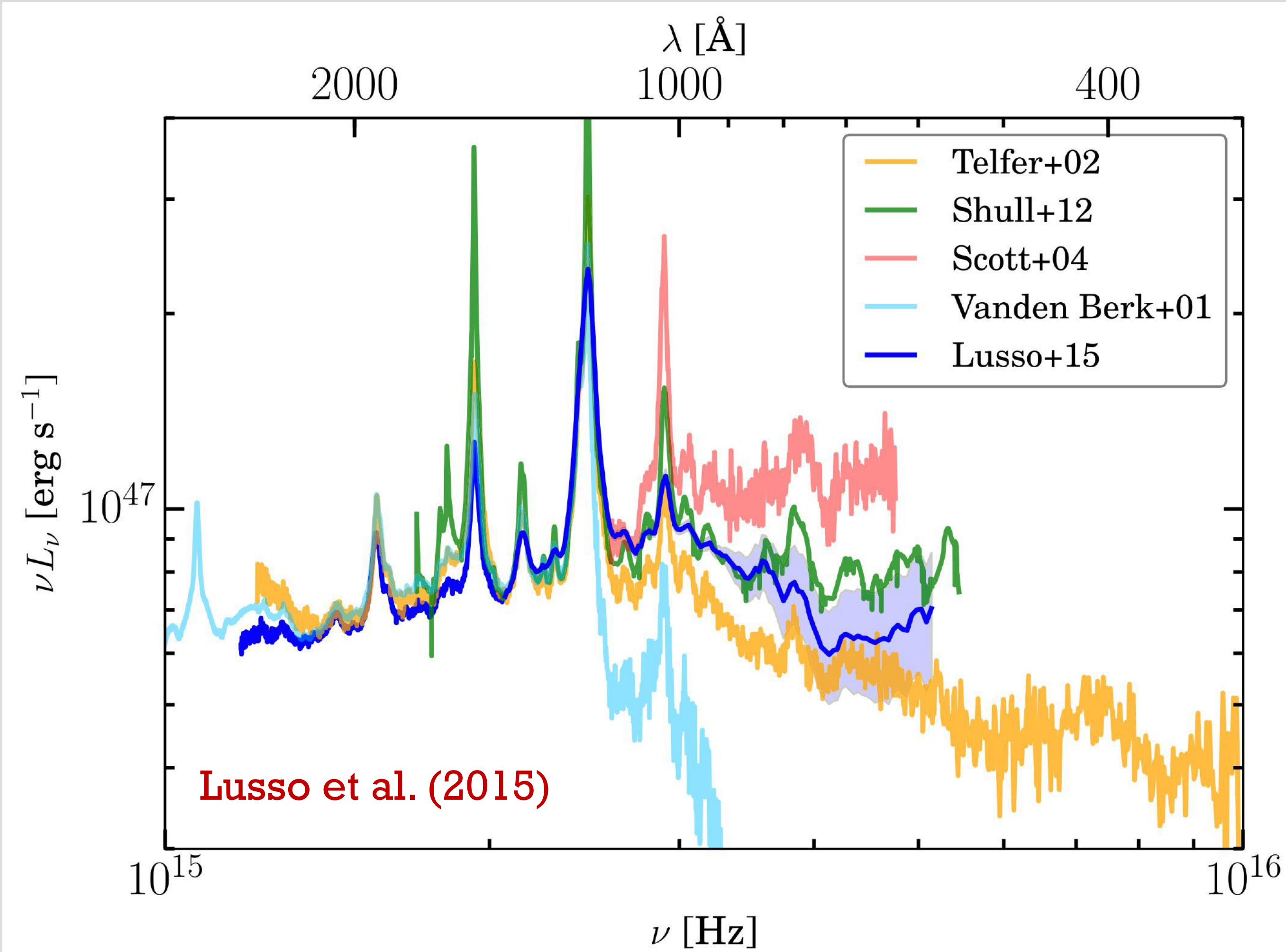
Observed correlations indicate the UV, EUV, and X-ray emissions are closely tied, and that the overall UV-EUV-X-ray spectral shape is largely set by luminosity.

Strong ties of three regions - UV disk, the EUV source (EUV warm corona?), and the X-ray hot corona.

Physical cause of these remarkable ties and luminosity dependence remains unclear.

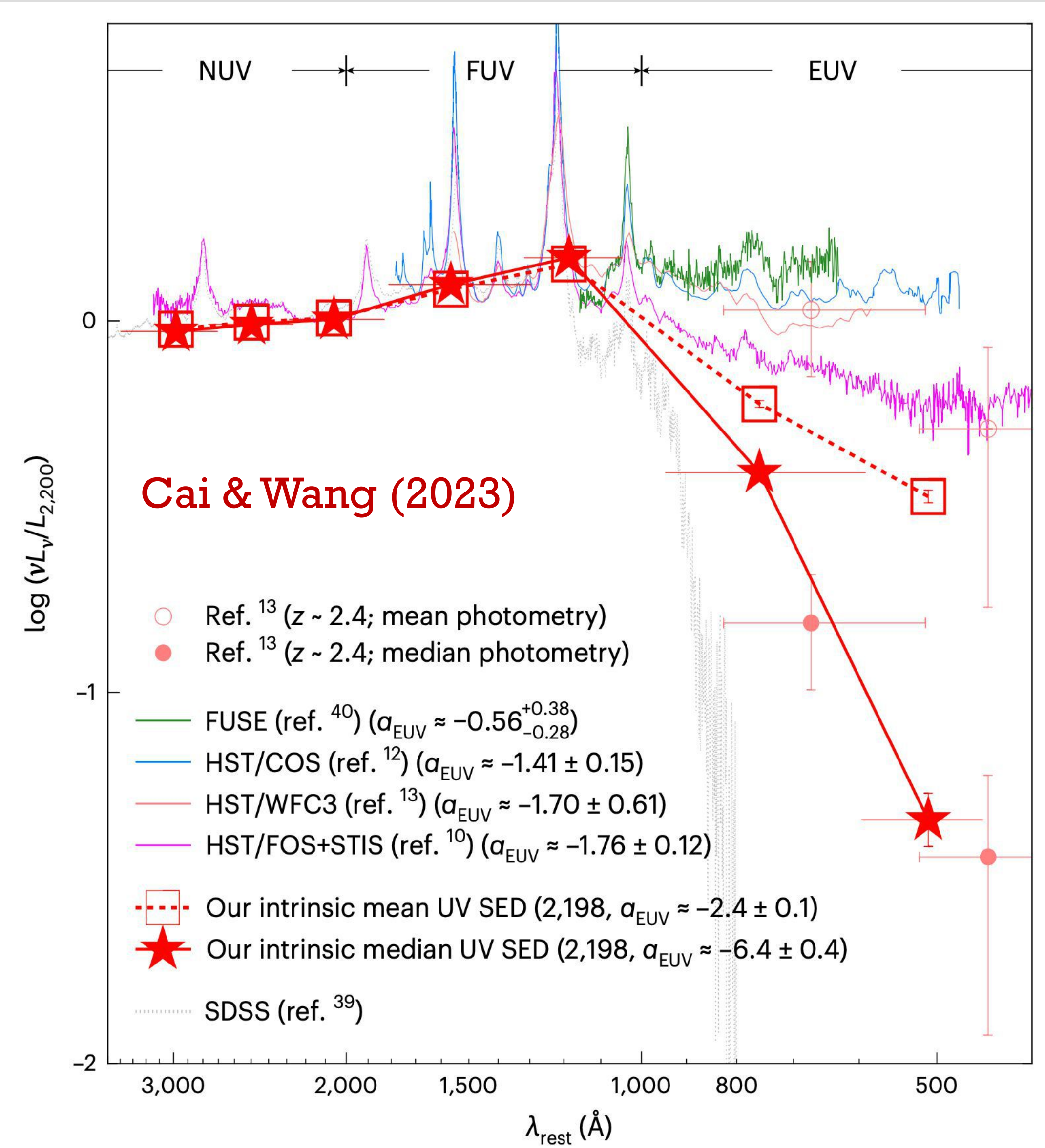


Quasar UV Composites



Significant differences in FUV-EUV composite spectra due to differing IGM absorption corrections, sample-selection effects, and perhaps luminosity effects.

Luminosity-Independent Average SED?



Attempt to correct for claimed EUV detection incompleteness.

Argue there is a luminosity-independent average SED down to ~ 500 Å – that is red in the EUV.

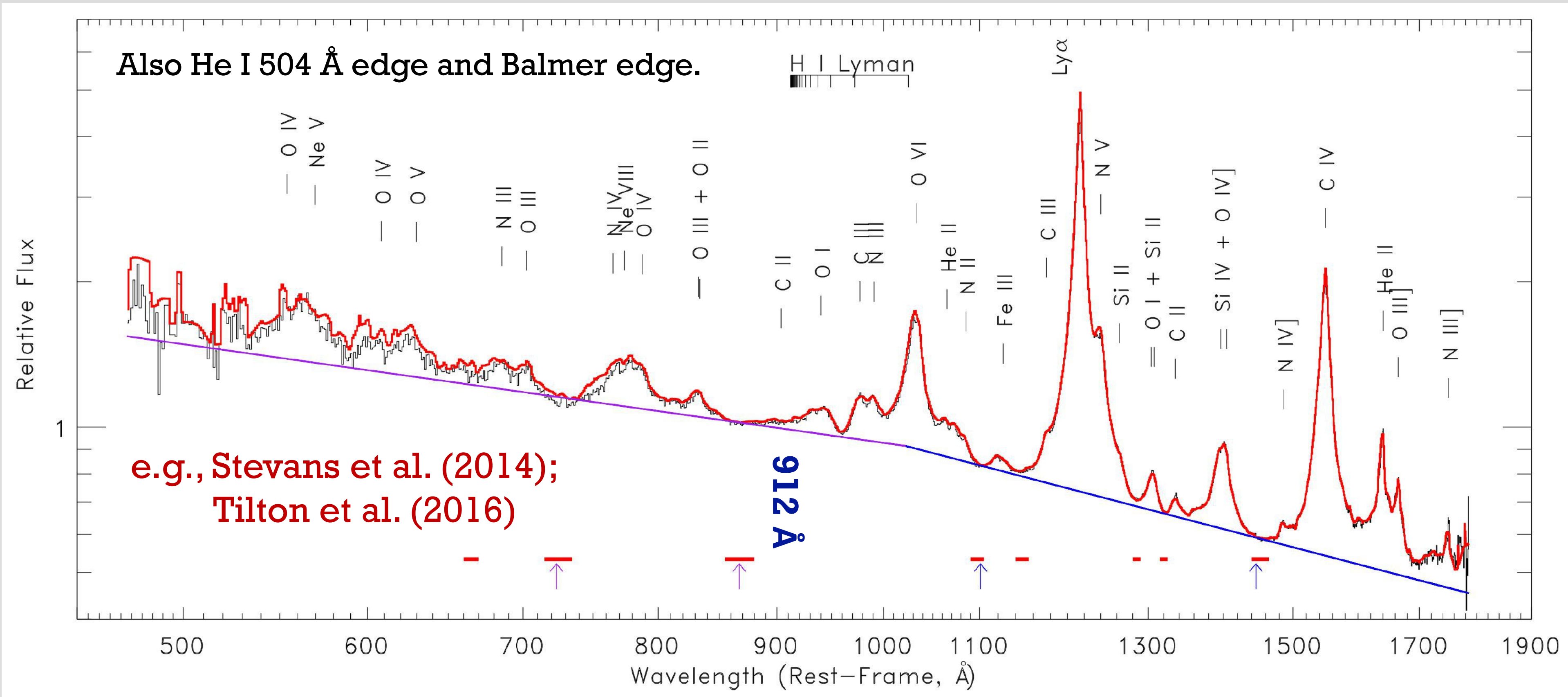
Data lie well below standard thin-disk predictions.

Red shape in the EUV might be explained by a disk "truncated" by strong mass outflow in a wind (e.g., Laor & Davis 2014).

Generally Don't See Lyman or Other Edges

Lyman and other edges (in absorption or emission) tend to be prominent in disk spectral models.

Their absence is not necessarily a problem but is an important constraint upon disk atmospheric structure.



UV Variability Constraints
on
SMBH Growth

Reverberation Measured Sizes

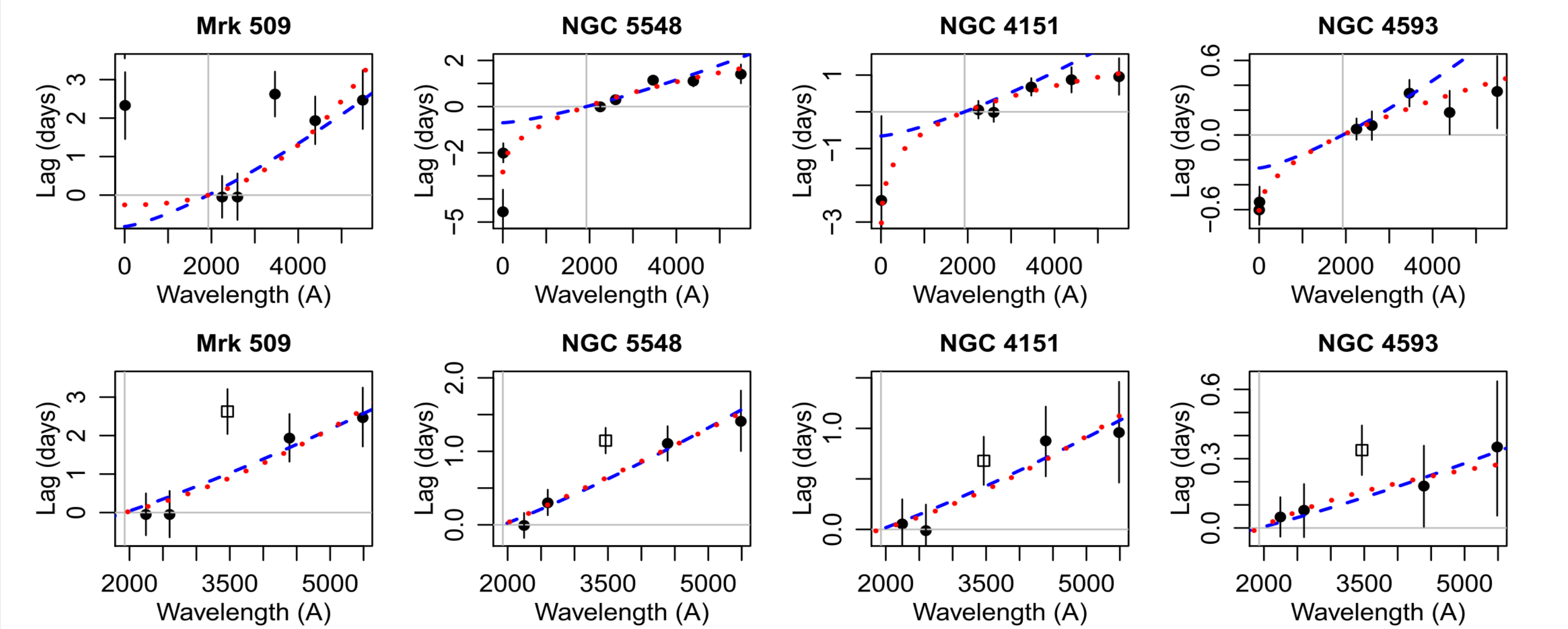


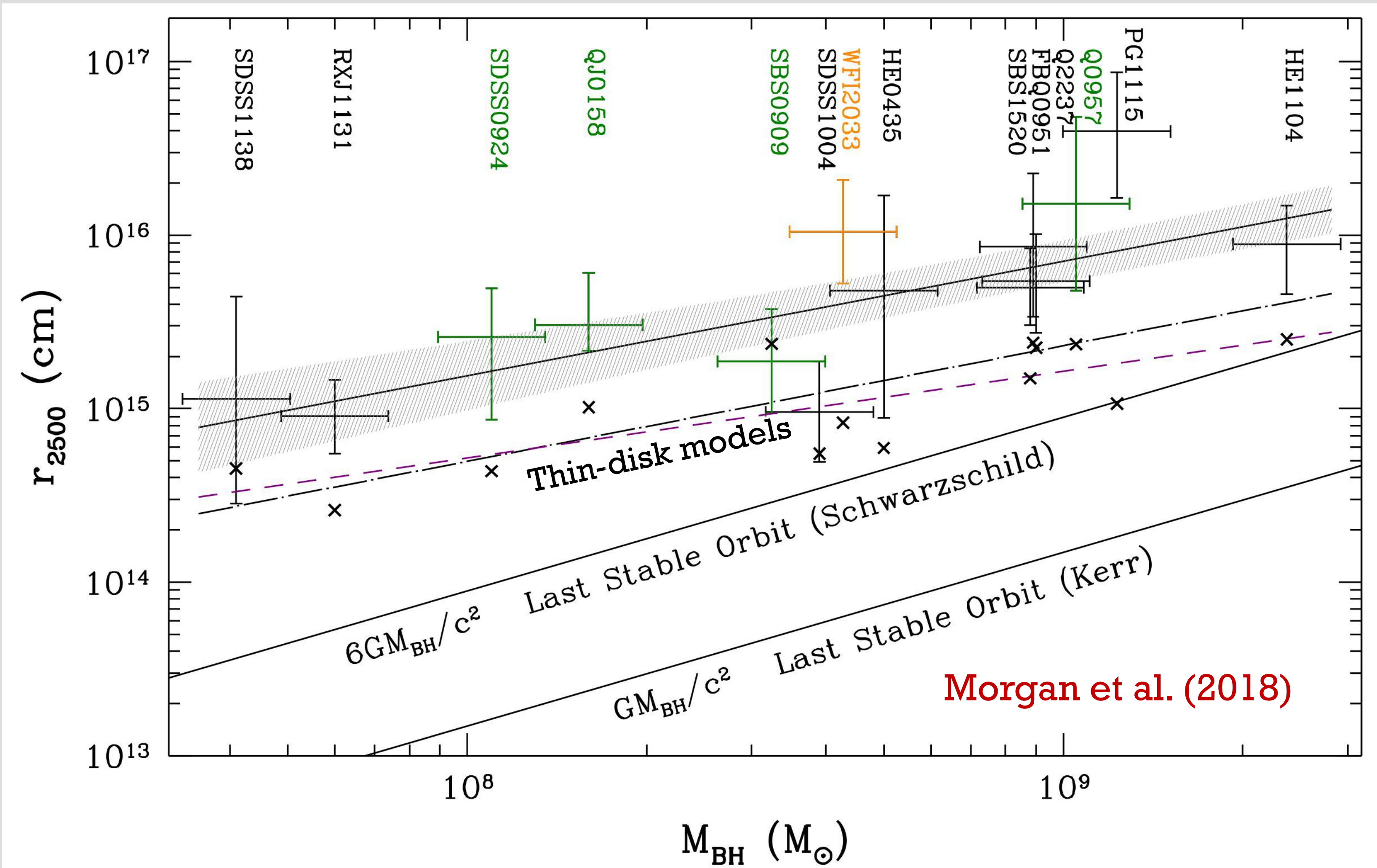
Figure 5. Plots of measured median ICCF centroid lag (τ) as a function of central wavelength (λ) for all bands in Figure 3. All lags are measured relative to the W2 band, so that autocorrelation point is not shown. The red dotted lines show the fit to the function $\tau = \tau_0[(\lambda/\lambda_0)^\alpha - 1]$, where τ_0 is the normalization, α is the power-law index, and λ_0 is the reference band wavelength, 1928 Å for the W2 band. The gray lines cross at $\lambda = \lambda_0$ because the ACF lag τ is identically zero. The blue dashed lines show the same fit but with fixed index $\alpha = 4/3$. The four top panels show the fits for the full data. In general, these functions yield poor fits due to a mismatch in the X-rays, an excess in the *U* band in all objects, and disagreements in *B* and *V* in two of the objects. Thus, the bottom four panels show fits restricted just to the UVOT data, excluding the *U* band. (The *U* band lags are shown as empty boxes because they do not participate in the fit.) These bottom panels show that once the X-rays and *U* band data are excluded, the fits are improved, with acceptable χ^2. We note that in one source (NGC 4593) all remaining points are within $\sim 1.2\sigma$ of zero lag and in another (NGC 4151) that only two of the remaining points (*B* and *V*) are significantly above zero.

Edelson et al. (2019)

The wavelength dependence of UV/optical interband lags is consistent with expectations for an accretion disk, though the disk size initially appears 2-3 times larger than expected.

Discrepancy mitigated if disk-ionization and relativity effects considered (Kammoun et al.)

Microlensing Measured Sizes



Quasar “Viscosity Crisis”

comment

Quasar viscosity crisis

Recent observations of extreme variability in active galactic nuclei have pushed standard viscous accretion disk models over the edge. I suggest either that some kind of non-local physics dominates accretion disks, or that the optical output we see comes entirely from reprocessing a central source.

Andy Lawrence

It is widely believed that active galactic nuclei (AGN), including the most luminous examples, quasars, are powered by accretion disks surrounding supermassive black holes. We have understood the general principles of accretion disks since the 1970s[1–3]. The disk rotates differentially, so neighbouring rings slip past each other. Some viscous process causes a drag between the rings, thereby transferring angular momentum outwards and producing local heating. If that local heating is also radiated thermally on the spot, this process determines the radial temperature profile ($T \propto R^{-3/4}$). A further simplifying assumption — that viscosity is proportional to the speed of sound — allows a complete solution of the disk structure. A well-known problem is that standard molecular viscosity, whereby particles from the fast lane slip into the slow lane and vice versa, is far too weak to explain the observed luminosities. From the 1970s onwards it was widely assumed that some of kind of turbulence and/or magnetic stresses would produce a viscosity-like effect. This idea was put on a sound footing in 1991, with the development of the theory of magneto-rotational instability (MRI)[4].

Accretion disk models nicely explain the luminosity and compactness of AGN, as well as the observed peak of the spectral energy distribution in the ultraviolet (UV) regime. Getting the details right has always been difficult[5], but these problems may be explained by effects that modify the spectral energy distribution, such as the presence of a Comptonizing atmosphere, or a system of clouds surrounding the disk[5,6]. However, by far the worst problem is variability. AGN vary significantly on timescales of weeks to months, whereas disks with the right degree of viscosity to explain the luminosity should take thousands of years to change their optical emission. Furthermore, variations at different wavelengths, from the optical through to the UV, vary simultaneously and have aligned peaks[7] (Fig. 1a), whereas in an accretion disk, different wavelengths come from different radii, which means changes should propagate through the disk.

This situation was rescued in the 1990s with the idea of X-ray reprocessing[8], whereby the central X-ray source, which can vary much more quickly than the part of the disk generating the optical light, shines on the disk and heats it. At any radius, heating has two causes: viscous heating, which changes only slowly; and X-ray heating, which can change quickly. Noticeably, although the shortest UV wavelengths might change by (say) a factor of two peak-to-trough, the redder optical wavelengths change only by a few per cent. There have been many papers arguing about whether or not X-ray reprocessing works in detail. The strongest argument in favour is the observation of delays between the variations at different wavelengths — on a timescale of hours to days, which is consistent with the travel-time delays of light[9,10].

However, the variability problem is now reaching a new crisis, thanks to the observation of extreme variability in some objects — factors of several over a decade or so, including, crucially, at optical wavelengths, not just in the extreme UV or X-ray regimes. Large changes have been known in a handful of nearby low-luminosity AGN for many years, but data comparison between the Sloan Digital Sky Survey and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) has revealed a large number of such objects[11] (Fig. 1b), including many at high luminosity. These objects have generally been referred to as ‘changing-look quasars’. The broad emission lines that normally accompany type I (that is, quasar-like) AGN seem to come and go along with the optical continuum; when the continuum and broad lines plummet, what is left behind is the narrow emission lines that dominate type II AGN. The varying broad emission lines tell us that the far-UV, as well as the optical emission must be changing dramatically.

Because these large changes occur in optical emission — not only in X-ray or far-UV emission — it seems difficult to avoid the conclusion that the outer region of the disk itself is undergoing a gross physical change on a timescale inconsistent with viscous heating. Furthermore, recent work, such as that comparing the Dark Energy Survey and the Sloan Digital Sky Survey, seems to suggest that extreme variability is not that unusual — possibly 30–50% of quasars sometimes vary by a large amount[12]. Studies of the variability structure function also suggest that the degree of optical variability for a typical quasar climbs inexorably at longer timescales[13]. Although some AGN have larger typical variability than others on any given timescale, it seems likely that all AGN vary dramatically if you wait long enough.

One might wonder whether some kind of variable obscuration, such as passing clouds in the clumpy torus, can explain the variability. However, studies of large changes usually conclude that this idea doesn’t fit the observations, because the timescales, the (lack of) colour changes and the relative line and continuum changes look wrong (Fig. 1b). It seems we really must confront the fact that accretion disk models are failing. Of course, good theorists have long-known that standard viscous accretion disk theory is just too simple, but it remains the observers’ paradigm. When interpreting data, researchers routinely assume that the standard theory is correct, and write optimistically of ‘accretion disk instabilities’ to explain outbursts. The problem is that the existence of common large-amplitude variability suggests that disks are in a state of permanent exception; it is not reasonable to describe them with standard viscous accretion disks at all. As Pringle said in 1981[14], ‘instability’ really means ‘inconsistency’.

We cannot solve this problem by simply cranking up the viscosity parameter. The rate of torque is closely related to the viscous scale length and therefore to the disk height, so the disk approximation breaks down completely. What can be done?

Non-local processes

Perhaps we must abandon the hope that the transfer of angular momentum, the

QUASARS IN CRISIS
THE ROYAL OBSERVATORY, EDINBURGH
AUGUST 6-9, 2019

correspondence

Old news on quasar viscosity

To the Editor — Much of the active galactic nuclei and quasar community has been fixated on a particular model for the energetically dominant ‘Big Blue Bump’ component of the spectral energy distribution for the past 40 years[1,2], despite the fact that the model is qualitatively incorrect. It’s a ‘quasi-static’ model, meaning that flow of matter through the disk is steady on human timescales, except for the very lowest luminosity cases, with gas elements migrating through a geometrically thin accretion disk towards the black hole; the rate is constant over periods that are long compared with any other timescale in the problem, or a human lifetime. But empirically, variations in flux were known to occur on timescales of weeks to months, nearly in phase throughout the optical and ultraviolet regions, so cognitive dissonance was a part of the theory from the get-go.

In a recent Comment titled ‘Quasar viscosity crisis’, Andy Lawrence[3] points out that the variability properties of quasars rule out the model. Enough was known about quasar variability to preclude application of the model at the time it was proposed, and the failings of the model were decisively documented and explained by Alloin et al.[4]. The new variability data alluded to in Lawrence’s article are immaterial; the arguments made in many papers over the decades, starting with Alloin et al., were robust, with orders of magnitude to spare. Hence, although Lawrence was correct[3], he wasn’t reporting any news. It would have been news in the early 1980s.

The situation is a little bit worse, actually, in that Lawrence[3] emphasizes that the new reports he highlights tracked the optical-band emission specifically, which is expected to vary even more slowly than the ultraviolet. But we showed[5] even before the work of Alloin et al.[4] that optical emission from the nucleus of the Seyfert NGC 4151 had varied and was down to an undetectable level at some epochs.

There is no disagreement on the science, but the record needs to be corrected. In an essay[6] written for the 50th anniversary of the discovery of quasars, I pointed out other very fundamental falsifications of the quasi-static disk model from the literature, some of which date back 30 years. In our field, theories are often falsified before publication; observers present their data in the context of debunked theories. And, of course, every generation makes the same discoveries over and over again. These arguments include the lack of the expected relationships of spectral energy distributions with mass and luminosity[7,8], both at single epochs and in difference-spectra (high state minus low state, which is crucial). And there’s the wee fact that gravitational microlensing mandates surface brightnesses (and hence thermodynamic emissivities) an order of magnitude below the theoretically expected value. Almost no one in the theory community tries to match that, and very few cite the sad fact, an exception being a toy model[9].

The culture is the same in most of the X-ray community. There, a seeming breakthrough was announced[10]: a four-day exposure taken by the Advanced Satellite for Cosmology and Astrophysics seemed to show an asymmetric horned profile — known as the signature of the Kerr metric — for the iron K-alpha fluorescence line. This paper was immediately followed by a closer look at the same data[11], broken into segments. It turns out that the magic profile never appears at a single epoch, but only in the four-day sum, so that if the observers had been given two days or eight days of observing time, the feature would not have appeared. This second paper presents the totally unexpected variability properties and then shows that with just a few more epicycles, everything works out fine.

The accepted geometry for an accretion disk is a thin, completely passive reprocessing disk on which stand upper and lower lampposts emitting X-ray continuum. Modellers routinely use thin disks for objects accreting at orders of magnitude above the Eddington limit[12]; they invoke fantastic and sometimes internally inconsistent iron abundances[12]. They ignore pesky things like the emission from the far side of the putative disks, even when inner gaps are employed and gravitational focusing causes the far-side emission to dominate the spectra[13].

Spin measurements completely rely on the assumption that the disks go from opaque to transparent instantaneously at the innermost stable circular orbit[13]. The lamps ride up and down as needed to scramble expected but unseen reverberation signals (how well this actually works is discussed in ref. [14] among others), they can move in radius and in azimuth as needed, and you don’t even have to put in the general relativity effects which Einstein would have thought belonged there[13]. Sometimes the lamps can even hover above the disk if a part of the profile lasts longer than a dynamical time (so, to be precise, it’s a ‘disk drive’ model). Critiques of reverberation and scattering models get short shrift[15] (indeed ref. [15] refutes the ‘clinching’ article[16] and accompanying celebratory commentary[17] on the unambiguous estimation of the spin of the black hole residing at the centre of NGC 1365).

While the cycle of news — as the name implies — is circular, our community cannot afford (in the literal sense of real-world cost) to forget (or even worse, disregard) important results that were already established decades ago. The maturity of our tools (both theoretical and observational) creates the scientific impetus for our community to go beyond simplifying assumptions about the accretion disks and tackle the problem head-on. ❐

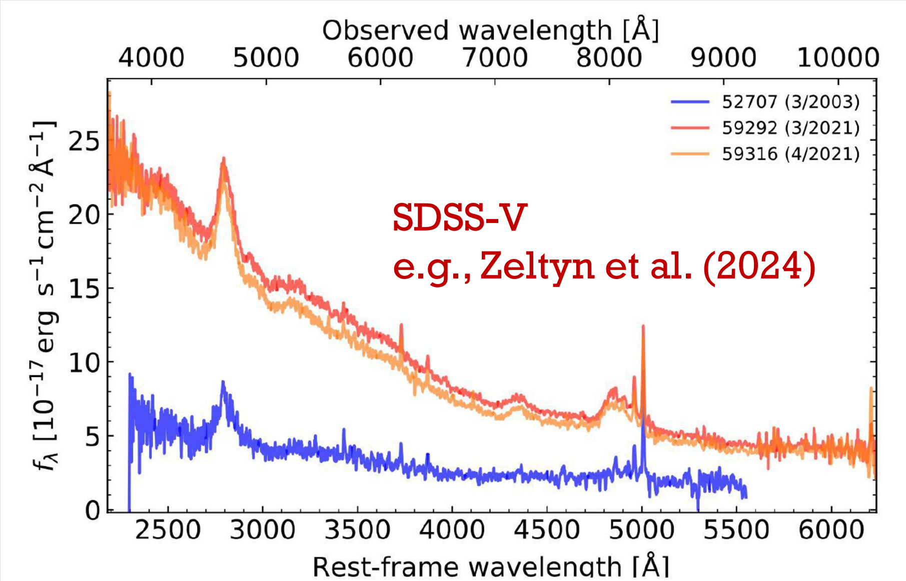
Robert Antonucci
Department of Physics, University of California, Santa Barbara, CA, USA.
e-mail: antonucci@physics.ucsb.edu

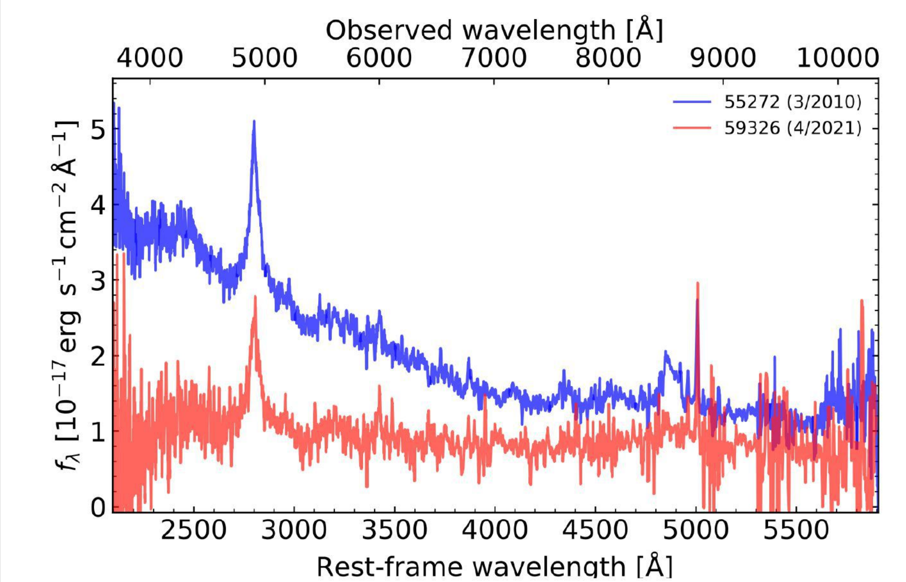
Published online: 3 July 2018
https://doi.org/10.1038/s41550-018-0521-1

References
1. Shields, G. A. *Nature* **273**, 519–520 (1978).
2. Malkan, M. A. *Astrophys. J.* **268**, 582–590 (1983).
3. Lawrence, A. *Nat. Astron.* **2**, 102–103 (2018).
4. Alloin, D., Pelat, D., Phillips, M. & Whittle, M. *Astrophys. J.* **288**, 205–220 (1985).
5. Antonucci, R. R. J. & Cohen, R. *Astrophys. J.* **271**, 564–574 (1983).
6. Antonucci, R. *Nature* **495**, 165–167 (2013).
7. Blaes, O. *The Central Engine of Active Galactic Nuclei* (eds. Ho, L. C. & Wang, J.-M.) 75–84 (Conference Series no. **373**, Astronomical Society of the Pacific, 2007).
8. Bonning, E. W., Shields, G. A., Stevens, A. C. & Salviander, S. *Astrophys. J.* **770**, 30 (2013).
9. Dexter, J. & Agol, E. *Astrophys. J. Lett.* **727**, L24 (2011).
10. Tanaka, Y. et al. *Nature* **375**, 659–661 (1995).
11. Iwasawa, K. et al. *Mon. Not. R. Astron. Soc.* **282**, 1038–1048 (1996).
12. Done, C. & Jin, C. *Mon. Not. R. Astron. Soc.* **460**, 1716–1724 (2016).
13. Niedźwiecki, A. & Zdziarski, A. A. *Mon. Not. R. Astron. Soc.* **477**, 4269–4273 (2018).
14. Mizumoto, M. et al. *Publ. Astron. Soc. Jpn* https://doi.org/10.1093/pasj/psy032 (2018).
15. Miller, L. & Turner, T. J. *Astrophys. J. Lett.* **773**, L5 (2013).
16. Risaliti, G. et al. *Nature* **494**, 449–451 (2013).
17. Reynolds, C. S. *Nature* **494**, 432–433 (2013).

“Crisis” is probably overblown – but something important is going on.

"Changing-Look" AGNs





CL AGNs show large UV/optical continuum and line variations over months-to-years.

Many appear to reflect intrinsic variations.

Observed timescales are much shorter than the nominal viscous/inflow timescale. Thermal instabilities acting on shorter timescales?

As expected, the "standard" viscous accretion disk model is too simple.

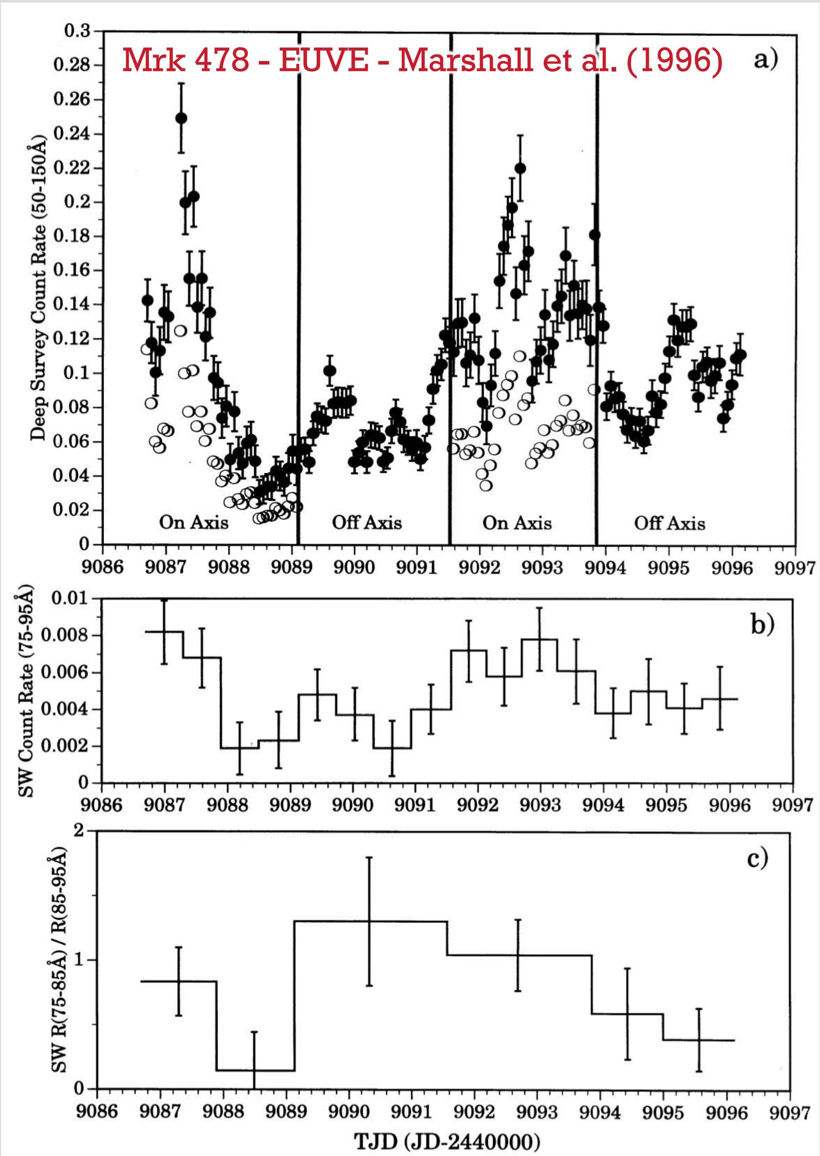
But what exactly is the replacement?

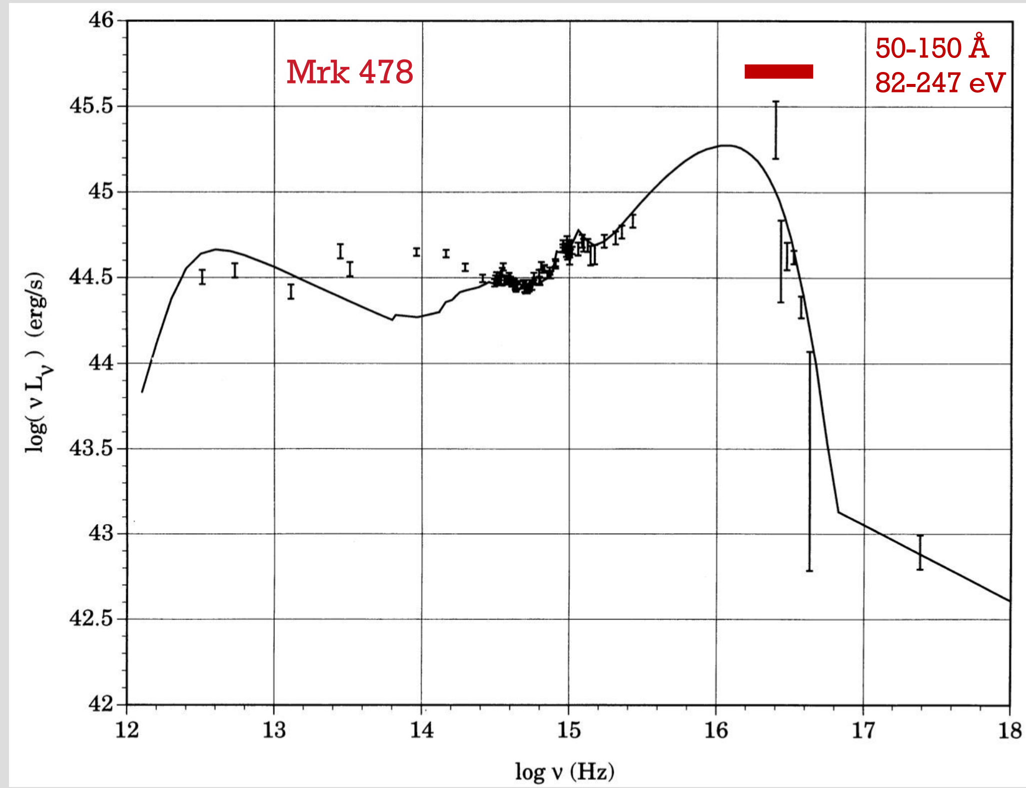
$$t_{\rm dyn} = \frac{1}{\Omega} = 1\ {\rm day}\,\frac{M}{10^8 M_\odot}\left(\frac{r}{30 r_g}\right)^{3/2}$$

$$t_{\rm th} = \frac{1}{\alpha\Omega} = 9\ {\rm day}\,\frac{M}{10^8 M_\odot}\,\frac{0.1}{\alpha}\left(\frac{r}{30 r_g}\right)^{3/2}$$

$$t_{\rm in} = \frac{1}{\alpha\Omega}\,\frac{r^2}{h^2} = 260\ {\rm yr}\,\frac{M}{10^8 M_\odot}\,\frac{0.1}{\alpha}\left(\frac{h}{0.01 r}\right)^2\left(\frac{r}{30 r_g}\right)^{3/2}$$

Strong, Rapid EUV Variability



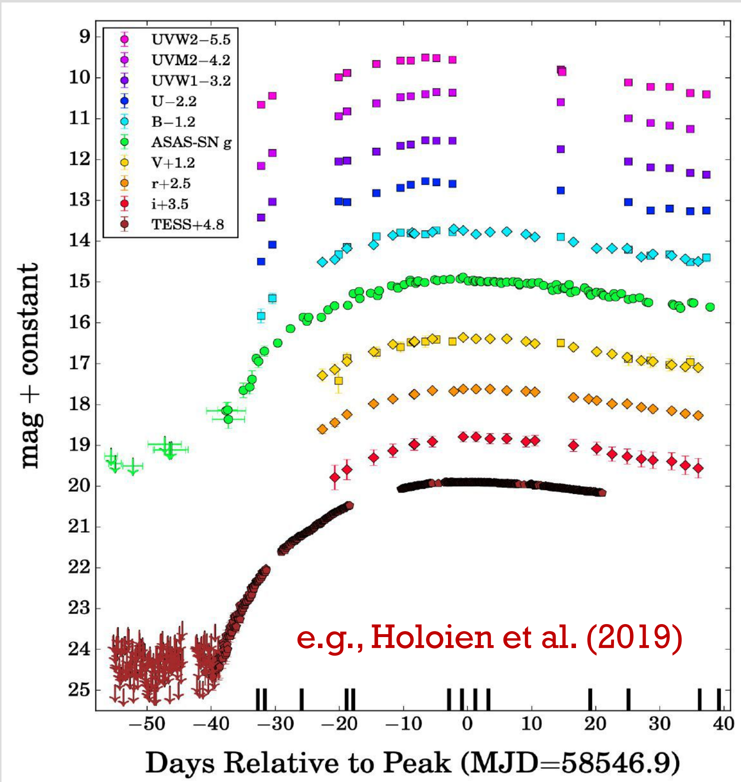


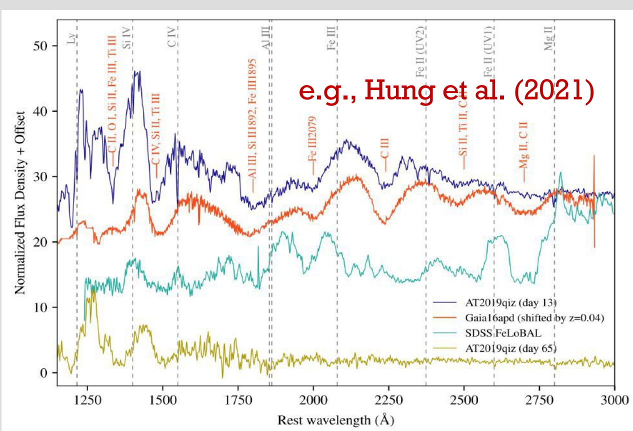
Day-timescale EUV variability is highly constraining for emission models – not a disk.

Suggests much of the EUV emission is sometimes from a "warm corona" that mildly Compton-upscatters UV/optical photons - and rapidly varies.

Exceptional Variability Phenomena

Tidal Disruption Events





TDEs provide critical insights into SMBH accretion physics and demographics at low $M_{\rm BH}$.

They likely only make a secondary contribution to cosmic SMBH growth overall.

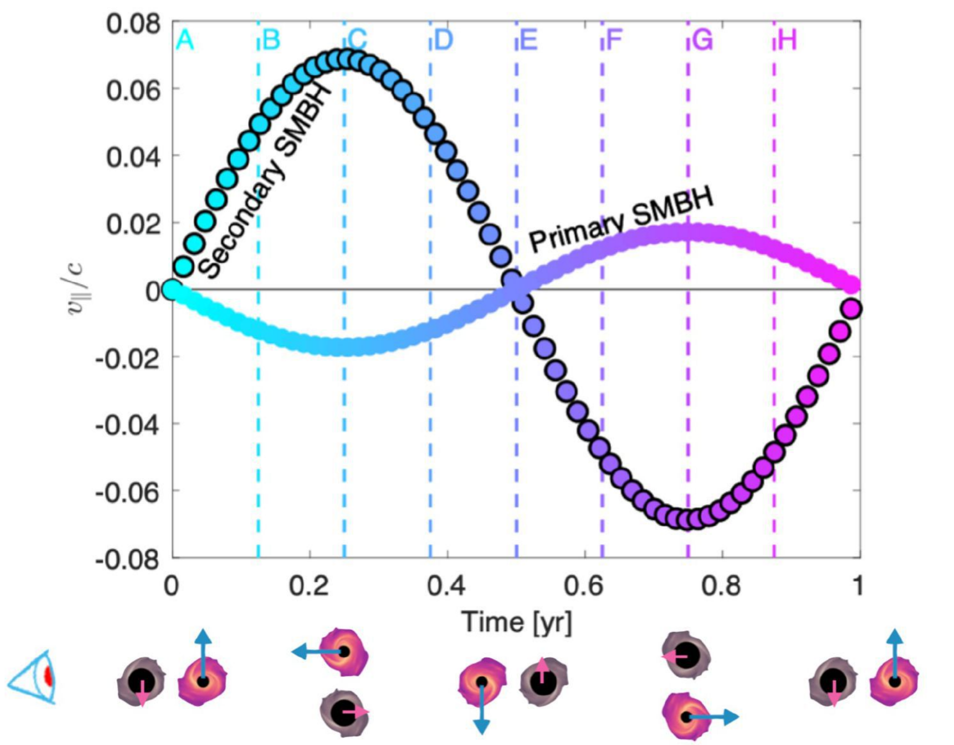
TDE emission generally peaks in the UV to X-ray, so TDEs are usually UV bright.

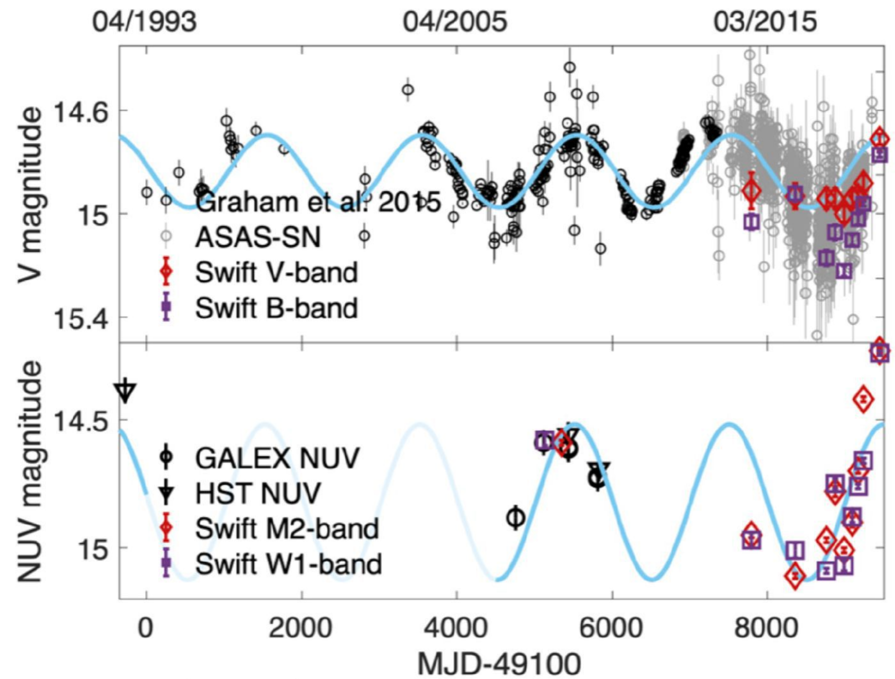
Rapid growth of TDE samples with, e.g., LSST and UVEX will help to understand their wide diversity:

- Early-time UV photometry
- UV tails of thermal emission from X-ray TDEs
- Apparent missing energy of optical TDEs
- UV spectroscopy of TDE outflows
- TDE unified model?

See Kulkarni et al. (2021) for further discussion.

SMBHB – Strong UV Doppler Boost





$$\frac{A_{\nu_1}}{A_{\nu_2}} = \frac{(3-\alpha_{\nu_1})}{(3-\alpha_{\nu_2})}$$

D'Orazio & Charisi (2023)

Figure 6. *Left*: Line-of-sight velocity for the primary and secondary SMBH in a binary with total mass $M = 10^9 M_\odot$, mass ratio $q = 1/4$, period $P = 1\,\mathrm{yr}$ and inclination $I = 60^\circ$. The orientation of the binary with respect to the observer at various phases of the orbit is shown at the bottom of this figure. We have highlighted the mini-disk of the secondary SMBH, since it is typically more luminous and moves faster, and thus dominates the variability. Adapted from Charisi et al. (2022). *Right*: The optical (top) and UV (bottom) lightcurve of PG 1302-102. The blue lines show the Doppler model given the optical and UV spectral indices. From Xin et al. (2020b).

Acts in addition to any direct modulation of accretion rate.

Hopefully, large-scale analyses can help clarify the frequency of SMBHBs.

Assorted Extreme AGN Variability in UV

UV/optical rejuvenations

Large-amplitude microlensing

Extinction events

Blazar flares

Stellar endpoints in disk/torus

The End

Thanks to Ari Laor (Technion) for constructive feedback.