A Statistical Study of Sub-Hour Flares of the VHE Gamma-Ray Emission of Markarian 421 During a High Flux State in 2001

R. M. Wagner*†, T. Schweizer* and E. Lorenz*†

*Max-Planck-Institut für Physik, D-80805 Munich, Germany
†Excellence Cluster "Universe", Technische Universität München, D-85748 Garching, Germany
‡ETH Zurich, CH-8093 Zurich, Switzerland

***Abstract*. From January to April 2001 the active galactic nucleus Markarian 421 showed a persistent high-flux state and frequent flares with rapid rise and fall times in the very-high energy domain. A deep observation of 249 hours has been carried out by the HEGRA telescope CT1. We analyze the flare rise and fall times in a large 56-night data sample with rather large statistics. A detailed time-structure analysis reveals typical rise and fall doubling times of about 20-30 minutes. The shortest detected rise and fall doubling times are ≈ 10 minutes. The flare duration was found to have a bimodal distribution with peaks around 25 and 175 minutes.**

***Keywords*: gamma rays: observations, BL Lacertae objects: individual (Mkn 421)**

I. Introduction

The active galactic nucleus (AGN) Markarian (Mkn) 421 was the first extragalactic γ-ray source detected in the TeV energy range [1]. So far flux variations larger than one order of magnitude and occasional flux doubling times as short as 15 minutes [2], [3] have been observed. Mkn 421 has been detected and studied in all wavelengths of the electromagnetic spectrum from radio waves to very-high energy (VHE; defined as $E > 100$ GeV) γ-rays. Its wide-range spectral energy distribution (SED) shows a synchrotron peak at keV energies, whereas the origin of the high energy (GeV-TeV) bump is still debated. The SED is commonly interpreted as being due to beamed, non-thermal emission of synchrotron and inverse-Compton radiation from ultrarelativistic electrons, accelerated by shocks moving along the jets at relativistic bulk speed. It can be described by synchrotron-self-Compton (SSC) models, e.g., [5], [6], or models involving external photon fields [7], [8]. Hadronic models [9], [10], however, can also explain the observed features. Ways to distinguish between the different emission scenarios are to determine the position and the structure of the second peak in the SED, to study correlations between wavebands, or to investigate the time variability of the emission.

In this work we study the $E > 1$ TeV γ-ray time variability of Mkn 421 on sub-hour time scales. We make use of an extensive data set, which covers a six-month epoch from 2001 January to June. During each of the observing nights long, uninterrupted observations of up to ≈ 7 hours were conducted. These observations allow us to perform a detailed study of light curve variations, in particular also the search for flares.

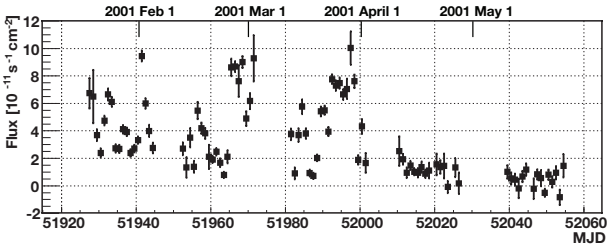

Fig. 1. The γ-ray flux of Mkn 421 above 1 TeV from 2001 January to 2001 June. The data have been binned on diurnal basis.

II. Observations and data analysis

The bright blazar Mkn 421 underwent a period of high VHE γ-ray emission from 2001 January to June, exhibiting numerous, large γ-ray flares (Fig. 1). These flares have been observed by various telescopes, e.g., [4], [11], [12], among them the $10\,\mathrm{m}^2$ stand-alone imaging air Cerenkov telescope CT1 [13] located on the Canary Island La Palma, Spain. With CT1, observations were conducted in 56 nights, amounting to 249 hours worth of data. Our work is based on the analysis in [14]. See [15] for a detailed description of the data analysis. An individual CT1 run lasts about 20 minutes. Typically, 8 to 60 γ-events were collected during such a run, which prevents binning the data in time much finer than the run duration, which is why we chose a bin width of 20 minutes for our study. While the HEGRA CT1 telescope had a trigger threshold of 700 GeV when observing at low zenith angles [16], a software threshold cut of 1 TeV was imposed in our analysis. This high energy threshold ensures, besides the high purity of the γ sample, the independence of the integral flux of the zenith angle. During the observation period in the first half of 2001 the weather conditions were exceptionally good. Non-physical reasons for rapid changes of the TeV flux can be excluded with high probability.

III. Statistical Flare Analysis

To identify nights with fast flux variations, a search procedure is applied: We fit a linear function to each light curve of at least three hours duration, assuming

a hypothesis of a constant or linearly increasing flux. Significant rapid flares are identified by a poor fit probability (derived from the χ^2/ndf value of the fit by $p_{\mathrm{line}} = 1 - P(\mathrm{ndf}, \chi^2)$, where $P(a,x) = \gamma(a,x)/\Gamma(a)$ is the regularized incomplete gamma function). We calculate the (negative) logarithm of the fit probability p_{line}. Nights with $-\log_{10}(p_{\mathrm{line}}) < N_{\mathrm{Cut}}$ are rejected from the flare analysis. In the following analysis, we use both $N_{\mathrm{Cut}} = 1.0$ (allowing for 10% false positives). In [15], we also employed $N_{\mathrm{Cut}} = 2.0$ (allowing for 1% false positives) to show that the final result is largely independent on the exact value of this cut.

A. The Flare Model

The limited statistics of the CT1 data makes pile-up of several small flares very probable; The sum of consecutive small flares is then perceived as a longer flare. The physical motivation for the pile-up assumption would be that many shock fronts from random turbulences are traversing the jet emitted by the AGN (see the discussion in Sect. IV). In leptonic acceleration models, each of the shock fronts would accelerate electrons and therefore emit high energy γ-rays produced by inverse Compton scattering. Also an inhomogeneous plasma blob hit by the shock fronts could lead to sub-flares. Our phenomenological flare model assumes exponential rise and fall times and allows to parameterize the time structure of a flare by these. The full width at half maximum (FWHM) flare duration can be derived from the model as well. The four free parameters of the model are: a flare amplitude a_0, a constant background F_{bg}, an exponential rise time τ_{rise}, and an exponential fall time τ_{fall}. The exponential rise time is expressed as doubling time in the limit of the fall time approaching infinity (likewise the fall time). The function used for fitting the nightly light curves thus is given by:

$$F(t) = F_{\mathrm{bg}} + a_0 \times \left(2^{t-t_0/\tau_{\mathrm{fall}}} + 2^{-(t-t_0)/\tau_{\mathrm{rise}}}\right)^{-1} \quad (1)$$

The background flux F_{bg}, on which the rapid flare is superimposed, may consist of other flares on longer timescales, particularly also of a pile-up of several flares. The exponential rise is motivated from the exponential increase of the energy of the underlying electron population that produces the photons in shock accelertion scenarios. The exponential falloff time is likewise motivated by an exponential electron cooling time. The electrons are cooled via synchrotron radiation and inverse Compton scattering. For simplicity we assume only one flare (on top of some background) per night. This assumption may not strictly hold true as discussed above, but within the temporal resolution of the CT1 telescope it should be adequate.

B. Quantitative Discrimination of False Positives

We need to make sure that the rise and fall times determined from fitting our flare model are statistically significant, as these nights have been selected among many nights with much slower rise and fall times. It has to be excluded that they arise just from large statistical fluctuations. In order to see how probable our results are, we analyzed a large sample of Monte Carlo-generated light curves and determined how often random fluctuations mimic rapid flares. To this end, 1,000 light curves for each of the nights under study have been simulated. For each time bin in the light curves we determined a flux value following a Gaussian distribution. The input parameters for the Gaussian distributions were fixed for each of the nights individually, using the nightly average fluxes as mean of the respective Gaussian distribution and the flux error as σ of the respective Gaussian distribution. For both the observed light curves and the simulated light curves we applied the same selection procedure: We performed a linear fit ($F = at + F_{\mathrm{bg}}$) to the light curve of each night.

As mentioned earlier, we calculate the (negative) logarithm of the fit probability and reject any light curves from further analysis with $-\log_{10}(p_{\mathrm{line}}) < N_{\mathrm{Cut}}$ with $N_{\mathrm{Cut}} = 1.0$. This is our selection criterion for nights with large rapid flares.

Fig. 2 shows the $-\log_{10}(p_{\mathrm{line}})$ distribution for the measured and the simulated constant-flux light curves. 37 out of 56 nights pass the "rapid-flare" requirement $N_{\mathrm{Cut}} = 1.0$ (28 nights survive $N_{\mathrm{Cut}} = 2.0$).

Our phenomenological flare model (Eq. 1) is applied to the surviving nights' light curves. From the model fit we determine the doubling (rise) time, the halving (fall) time and the FWHM (duration) of the flare. The FWHM is calculated numerically from the model function, and has been performed only for surviving nights' light curves that fulfill a number of requirements: for the FWHM analysis we require that the flare peak, determined from the fit, must lie within the observation window (30 and 17 nights remain, respectively, for $N_{\mathrm{Cut}} = 1.0$ and $N_{\mathrm{Cut}} = 2.0$) and we also exclude unreasonably short rise and fall times below the experimental resolution (smaller than 5 minutes; 29 and 28 nights remain, respectively, for the rise-time and fall-time analysis with $N_{\mathrm{Cut}} = 1.0$, and 16/17 with $N_{\mathrm{Cut}} = 2.0$).

The shape of chosen flare function describes the flares quite well, as the flare-function fits result typically in acceptable reduced χ^2 values and a corresponding $-\log_{10}(p_{\mathrm{flare}}) \lesssim 2$.

C. Results

Fig. 3 shows the rise and fall times as obtained from fits to all light curves passing the fast-flare selection criteria along with the MC expectation for the case of a Gaussian-distributed constant γ-flux passing the selection criteria of $-\log_{10}(p_{\mathrm{lin}}) > 1$. The doubling times and halving times range from about 10 minutes up to 140 minutes, with the rise times peaking at about $20 - 30$ minutes and the fall times peaking at about 10 minutes. Note that a fit for a single night does not have the significance to claim the above statement, while the combination of many nights together with a MC

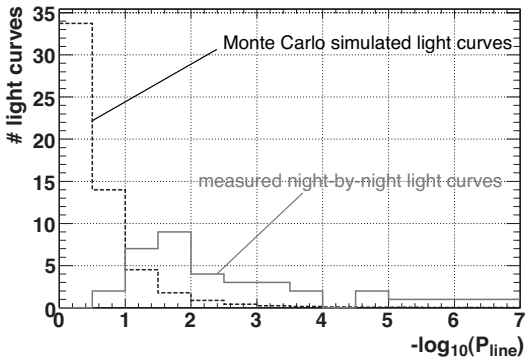

Fig. 2. The $-log_{10}(p_{lin})$ distribution of the linear fits to the sample of measured light curves in comparison to the linear fits to MC-simulated light curves with "constant" γ flux. The MC histogram has been normalized (divided by 1000) to the histogram of the measured points since for each real night 1000 light curves have been simulated. An excess at high χ^2/ndf is observed in comparison to the MC data. Light curves with $-\log_{10}(p_{\mathrm{line}}) > 1$ and $-\log_{10}(p_{\mathrm{line}}) > 2$ are assumed to originate from fast flares and are used for further analysis.

background study allows this conclusion. Our results show clear evidence for very small doubling (rise) time and halving (fall) times of $\lessapprox 10-15$ minutes. Ten nights out of the selected nights have flares with a rise time faster than 25 minutes, while 15 flares have faster fall times than 25 minutes. Nine of them are at the order of 10 minutes or faster. While we find that the shortest flare rise and fall times are equally small, overall there are more flares with a shorter fall time than rise time. The simulated (constant-flux) light curves exhibit identical rise-time and fall-time distributions, as expected.

This analysis is not sensitive to time scales below approximately 10 minutes, since the event statistics are not sufficient for a smaller bin width. Thus shorter doubling and halving times cannot be excluded. This analysis is sensitive to flare durations of typically below 140 minutes because of the maximum nightly observation time.

Fig. 3 also shows the corresponding FWHM duration distribution. Most measured flare durations differ significantly from the null-hypothesis expectation. They range from about 25 minutes up to 4 hours, with two pronounced peaks at $\approx$ 25 and 175 minutes FWHM. The time resolution due to event statistics implies a sensitivity threshold at $\approx$ 30 minutes FWHM. Our analysis therefore is able to detect rapid flares with durations from 30 minutes up to 250 minutes (maximum observation-night length). We do not find any correlation between the total observation time per night and the measured flare duration.

In conclusion we have demonstrated that the rise times and fall times in the selected nights are statistically significant (compared to the Monte-Carlo null hypothesis) and are not caused by background fluctuations. There is evidence for rise and fall times below 10 minutes.

IV. Discussion

While comprehensive studies of the flaring behavior of Mkn 421 exist for long-term X-ray observations, TeV flares during persistent high-flux period, like in the first half of 2001, have not yet been studied. Although single, individual TeV flares of about 20 minute time scales of Mkn 421 have been identified occasionally [2], [3], [12], our study reveals that such short-scale flares occurred rather persistently during the 2001 high-state of Mkn 421. We find flare durations from about 25 minutes up to 4 hours, while clear peaks in the flare duration distribution exist at ≈ 25 and 175 minutes FWHM. This bimodal distribution may point to a different nature of the flares with long (≈ 175 min) and short (≈ 25 min) duration. However, these timescales possibly constitute pile-ups of even faster flares, as the observations are limited by the statistics provided by the HEGRA CT1 detector.

It is difficult to explain these short-variability timescales. In order to obtain such a fast variability, causality requires the emission region to be sufficiently small,

$$R < c\Delta t D \tag{2}$$

where R is the size of the emission region, c the speed of light, Δt the variability time scale and D the Doppler beaming factor of the bulk jet motion,

$$D = \frac{1}{1+z}\frac{1}{\Gamma\left(1-\beta\cos(\theta)\right)}, \tag{3}$$

with the redshift of the source z, the Lorentz factor Γ, the jet viewing angle θ and $\beta = v/c$ the speed of the bulk jet motion. A reasonable assumption is $D = 10\ldots20$ [17], [18], [19]. For Mkn 421, the superluminal velocity of radio blobs was found not to exceed $2.5c$ [20]. This implies that the Doppler beaming factor cannot be much larger than about 20, conservatively assuming a very small observation angle to the emission direction of the jet. This estimate already takes into account that the jet has considerably cooled at the time when it becomes visible in the radio band. With these values we estimate the size of the emission region to be at most 5×10^{14} cm or ~ 30 AU. If the jet originates from accretion flows, as very often assumed, this would also be an upper limit on the inner boundary of the flows.

A possible explanation for short variability time scales is given by shock-in-jet-models [21], [22], in which shock fronts (rather small emission regions) are traversing the jet. Each of them accelerates electrons and thus may emit rapid γ-flares in a stochastic manner. Such shock fronts inside the jet occur naturally as a result of random turbulences when the jet expands into the interstellar medium. These blobs or shells inside the jet collide with each other at some distance from the central black hole (ranging from 10^{16} cm to 10^{18} cm or $10^3 - 10^5$ AU), which may create shock fronts.

The fact that fast flares challenge conventional leptonic acceleration models has also been found by [23]

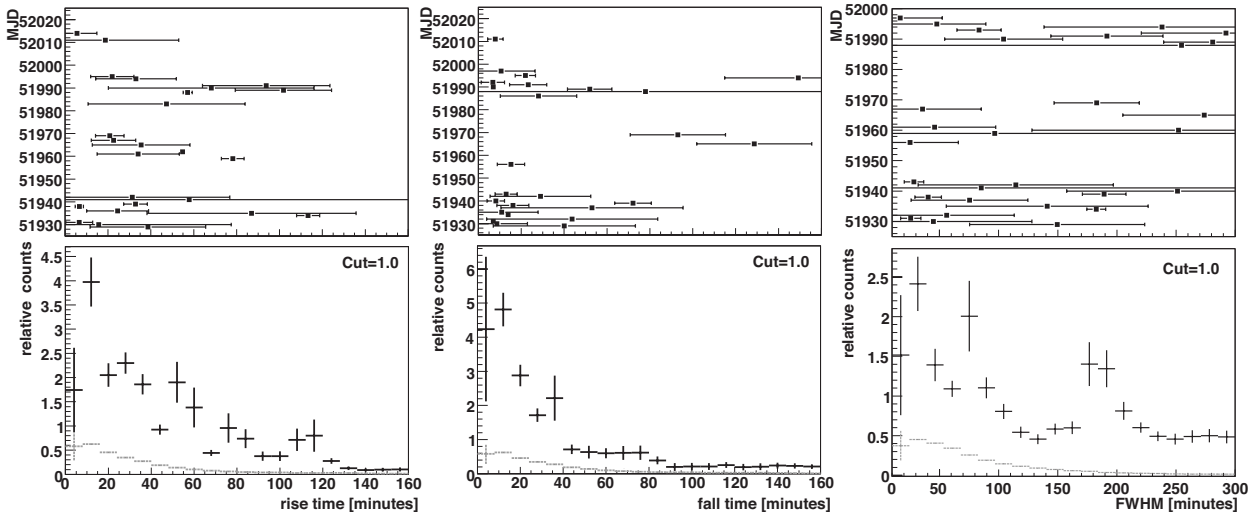

Fig. 3. Rise-time distribution (left), fall-time distribution (middle), and FWHM distribution (right) of the selected flares. The upper figure contains the doubling times for all individual nights passing the "rapid flare" criterion $N_{\rm Cut} > 1.0$. The lower panels compare their distribution (solid black line) with the MC constant-flux null hypothesis (dashed grey line). The MC counts are below the real data counts because of the background rejection cut.

and [24]. In simple SSC models the emission region is spherical and homogeneous. If a spherical emission region is assumed, the source easily becomes too compact, if not very high values for D are assumed [23], at odds with the measured jet speed in Mkn 421 (0.1 ± 0.02 c; [25]). Either the acceleration of the electrons producing the γ-rays happens rather close (some 100 Schwarzschild radii) to the central black hole [23], or they are accelerated somewhere in the AGN jet. [26] systematically study various particle acceleration and energy loss mechanisms in the jet and show that electron acceleration at shock fronts may happen quasi-instantaneously and thus may account for fast variability. Hadron acceleration, in contrast, may take several minutes and result in an evolving particle injection spectrum during the flare. "Instantaneous" second-order stochastic acceleration may also be possible in AGN jets.

Both the shortest rise and fall times in our analysis are $\approx 10-15$ minutes. Overall, we observe shorter flare fall times than rise times. This is in contrast to the long-term study of Mkn 421 in the X-ray band [27], where all resolvable flares showed a symmetric flare shape. Thus, for the TeV flares it seems as if the particle-acceleration time slightly dominates the process as compared to the IC cooling time. In addition to the flares with short fall-times, we found a wide shoulder between 40 and 90 minutes, again possibly pointing to two different classes of flares in the dataset.

V. Concluding Remarks

The present study has proven that persistent observations of bright TeV blazars in outburst are well suited for statistical analysis approaches. Particularly the existence of short-timescale flares has been proven on statistical basis, while no strong claims for periodicities in the X-ray and VHE γ-ray data under study here were possible.

Acknowledgements

We thank the HEGRA collaboration for the operation of the CT1 telescope on the Canary Island of La Palma. The financial support of the Max Planck Society is gratefully acknowledged. T.S.'s research was supported in part by the Humboldt University of Berlin.

References

[1] Punch, M., et al., 1992, Nature, 358, 477
[2] Gaidos, J. H., et al. 1996, Nature, 383, 31
[3] Aharonian, F., et al., 2002, Astron. Astrophys. 393, 89
[4] Aharonian, F., et al., 2003, Astron. Astrophys., 410, 813
[5] Marscher, A. P., & Gear, W. K., 1985, Astrophys. J., 298, 11
[6] Maraschi, L., Ghisellini, G., & Celotti, A., Astrophys. J., 397, L5
[7] Dar, A. & Laor, A., 1997, Astrophys. J., 478, L5
[8] Sikora, M., Begelman, M. C., & Rees, M. J., 1994, Astrophys. J., 421, 153
[9] Mücke, A. et al., 2003, Astropart. Phys, 18, 593
[10] Aharonian, F., 2000, New Astron., 5, 377
[11] Giebels, B., Dubus, G., & Khélifi, B., 2007, Astron. Astrophys., 462, 29
[12] Fossati, G., et al., 2008, Astrophys. J., 677, 906
[13] Mirzoyan, R., et al., 1994, Nucl. Instr. Phys. Res., 351, A513
[14] Schweizer, T., 2002, Ph.D. thesis, IFAE Barcelona
[15] Schweizer, T., Wagner, R. M., Lorenz, E., preprint (MPP-2008-80)
[16] Cortina, J., Barrio, J. A., & Rauterberg, G., 2000, AIP Conf. Ser., 515, 368
[17] Tavecchio, F., & Maraschi, L. 1998, Astron. Astrophys., 337, 25
[18] Bednarek, W., Protheroe, & R. J. 1999, MNRAS, 310, 577
[19] Albert, J., et al. 2007, Astrophys. J., 663, 125
[20] Marscher, A. P. 1999, Astropart. Phys., 11, 19
[21] Spada, M., Ghisellini, G., Lazzati, D., Celotti, A. 2001, MNRAS, 325, 1559
[22] Guetta, D., Ghisellini, G., Lazzati, D., & Celotti, A. 2004, Astrophys. J., 421, 877
[23] Begelman, M. C., Fabian, A. C., Rees, M. J. 2008, MNRAS, 384, 19
[24] Ghisellini, G., Tavecchio, F., Bodo, G., Celotti, A. 2009, MNRAS, 393, 16
[25] Piner, B. G., Edwards, P. G. 2005, Astrophys. J., 622, 168
[26] Tammi, J., Duffy, P. 2009, MNRAS, 393, 1063
[27] Cui, W. 2004, Astrophys. J., 605, 662