

Literature Summary

The Large Synoptic Survey Telescope (LSST) is a next generation telescope aimed at simultaneously probing many different regimes of astronomy and attempting to answer open questions from stellar science to cosmology. The main themes, as put forth by the LSST project are: probe the nature of Dark Matter (DM) and Dark Energy (DE), explore the phenomenology and new timescales of the variable and transient universe, study the Milky Way and local stellar populations, and create a catalog of as many solar system objects as possible (LSST Science Collaboration et al., 2017). In this short review, the focus will be on the general cadence structure of the LSST ten-year survey, its interplay with the science cases, specifically transient science, and addressing the challenge of making use of the LSST dataset for transient science through the lens of detection and classification. The bulk of the literature reviewed for this summary were the LSST Science Book vol.2 and the LSST Observing Strategy Whitepaper (LSST Science Collaboration et al., 2009, 2017).

Transient science with the LSST will be highly dependent on the ability to characterize the target populations of time-varying events. It will also be highly dependent on the cadence with which the survey area is sampled. The restrictions placed on these factors include our limited knowledge about the physics of these transient systems, the competition with other science cases for specific observing strategies and the physical limitations imposed by observing conditions and instrumentation. Before addressing the determination of the cadence from the perspective of transient science, it is useful to review what the baseline cadence is and how its features address the main science drivers behind the project.

1 General Cadence Considerations

The observational strategy, or cadence, that will be adopted for the LSST survey is still a work in progress. However, there are some main features which will be consistent throughout the optimization of cadence strategies. These features are primarily part of the Wide Fast Deep (WFD) survey strategy. These characteristics, which make the survey wide, fast, and deep, would seem to be something that any telescope would want the capability of having. In fact this is the first time a survey will certainly fulfill all three of these qualities, it signifies that we are now in an era of astronomy where all these goals can be achieved simultaneously and will become the norm of new astronomical surveys. Covering a large sky area at a pace that would allow for the detection of changes in the sky on short timescales out to a distance that is beyond our local galactic environment make this a WFD survey (LSST Science Collaboration et al., 2009).

One of the main features of LSST that allows this to be accomplished is the telescope's large etendue, which is a combined measure of an optical system's field of view (FOV) and light gathering power. LSST will have an etendue larger by order of magnitude than any previous survey. If etendue is a fiducial metric for science potential, as it is referred to in LSST Science Collaboration et al. (2009) and one measure of why the Sloan Digital Sky Survey was able to be so successful, then the LSST will be a prolific generator of science. For reference, the FOV of the LSST will be 9.6 deg^2 , or approximately 20 full moons. This large size was a requirement so that image quality would be atmospheric seeing limited, rather than optics limited. Additionally, the survey will operate in six filters, the u,g,r,i,z, and y filters similar to the Sloan Digital Sky Survey, with the addition of the sixth y filter (LSST Science Collaboration et al., 2009). A point that will be relevant if colors are needed is that the filter system for LSST can only accommodate five filters consecutively. Without expending

precious survey time on replacing one of the five filters for the sixth filter in a single night, the possible colors are limited to the filters chosen at the beginning of the night. Furthermore, switching between any of the active filters requires a two-minute pause in observing which decreases the efficiency of the survey (LSST Science Collaboration et al., 2017). If colors are determined to be necessary for a transient object, these factors will be important when considering identification and classification of transients.

The baseline cadence is not only the WFD survey, but it has been decided to set aside some of the available observing for mini-surveys such as the Deep Drilling Fields (DDF). While this is not the only mini survey, it is one with the highest priority of the additional surveys that will be completed. The DDF survey is also the only mini survey, out of those proposed in the current version of the Observing White Paper, which will explicitly help meet the science goals related to transient science. The DDF refer to a small selection of single telescope points spread throughout the sky that will be surveyed a much higher pace than the normal fields of the WFD survey. This is primarily to sample the area more frequently for early detection and better light curve sampling of transients, as will be discussed a bit more later on. Additionally, though not the primary intention, this enhanced sampling will provide a higher magnitude coadded five-sigma depth, probing the earlier universe (LSST Science Collaboration et al., 2017).

The current cadence, *minion_1016*, uses an approach that attempts to uniformly sample the entire visible sky area while accounting for many factors such as weather, the moon’s path, maintenance time, allocation for the mini-surveys, and other factors. This however is not the only strategy proposed for implementing the survey. Another strategy is what is known as a rolling cadence. This is a very interesting idea because it aims to cluster visits to a smaller subset of fields, rather than the fully viewable night sky, and rotate this clustered region around the full survey area over the ten-year period. This decrease in full-survey uniformity has the benefit of increased sampling frequency of a desired region. This has the affect of providing better characterized lightcurves for the transient events that are detected at the expense of detecting a higher number of events. This allows for better science to be done with the data that is collected rather than simply relying on large statistics. A lot is unknown how this will affect the other science cases, apart from degrading proper motion measurements, and much work is being done to explore the rolling cadence, but for transient science this would be a very suitable cadence strategy (LSST Science Collaboration et al., 2017).

One of the difficulties in evaluating a cadence is creating a consistent way to compare its impact on the diverse science cases that are being investigated with the LSST. This is still a work in progress, but much thought has been put into constructing a set of metrics for each case which use variables intrinsic to the survey, rather than the specific science case, for easy comparison when cadence features change. The idea is to compare a science case’s metrics, evaluated for the baseline cadence against other proposed cadences, and then summarize these metrics with what is being called a Figure of Merit (FOM). The suggested FOM will be percentage uncertainty in some measurement as a variation with respect to the number of objects for this science the cadence was able to observe. The computational requirements for running the Operations Simulator (OPSIM) are high enough that an incremental exploration of the parameter space would be very difficult. The comparison between cadence simulations explicitly, rather than the explicit variation with a specific cadence-feature, is more practical than attempting to explore the parameter space of each type of variation. Additionally, using a single FOM instead of using a variety of metrics is more productive for the comparison of drastically different science cases. While not the focus of discussion here, for the cases consid-

ered relevant metrics, or alternatively FOM, may be mentioned (LSST Science Collaboration et al., 2017).

Lastly, when concerning the current cadence, there are some issues that are being addressed and will hopefully have a positive impact on all science cases concerned. These issues include a bias of observations toward the western night sky, rather than being centered on the meridian in the ideal scenario. This degrades visit depth due to poorer seeing at higher airmass. Two other characteristics of the OPSIM which affect the predictions are the current input models for sky brightness and limits on moon avoidance. The last major issue is the tendency to bias a large number of visits, five or more, to single field in a given night, thus increasing the inter-night gap between field visits and reducing the number of fields visited in a night (LSST Science Collaboration et al., 2017). The first and last issues mentioned here highlight problems with the calculations that actually determine the observing strategy and the other problems are instead issues with inputs that influence these calculations.

2 Transient Science Cadence Demands

The discussion above about the baseline cadence and proposed rolling cadence address some of the issues that are relevant for transient science, specifically high sampling frequency, but now I aim to discuss more explicitly these requirements. Transient science is inherently dealing with limited timescales, and so increased sampling is a fairly straightforward requirement. Exploring the physics, or more appropriately the physics indirectly through lightcurve behavior, of the two transient scenarios of interest, Supernovae (SNe) and Kilonovae (KNe), sometimes referred to in the literature as Macronovae (MNe), will determine requirements on the implemented observing strategy to obtain data suitable for science. This is seen through intrinsic lightcurve characteristics and variation timescales that constrain detection and classification. Particularly of interest is early-time observations and identification which place quite stringent requirements on an executed observing proposal.

Furthermore, other transient science cases may place even more stringent constraints on an executed cadence. These are namely Gamma Ray Burst (GRB) afterglows, though these are quite likely the same KNe events as evidenced by the recent BNS merger discovery, stellar flares, and Cataclysmic Variables (CV) (Villar et al., 2017). These all have relevant timescales less than a day which make capturing their variability a very difficult task for the WFD survey. These would potentially benefit greatly from the rolling cadence proposal or a dedicated Target of Opportunity program with a DDF-like cadence. If we take the BNS merger discovery as typical of KNe events and the relevant timescales of SNe, then our cases of interest are on variational timescales from a couple hours to a few days (LSST Science Collaboration et al., 2017). A very general idea of the physics behind these timescales and the observations relevant for each case will be discussed next.

3 Supernovae

Supernovae make up one of the most important probes of cosmology. Particularly, through the ability to "standardize" type Ia SNe, quite precise measurements can be made for the distances to these objects and thus a local universe measurement of H_0 can be made. However, the precision of these distance measurements relies heavily on the standardization process which, in turn, relies on the characterization of the lightcurve, except in the case of spectroscopically sampled SNe. The LSST will not have a spectroscopic instrument but it will be

able to trigger other observatories for follow-up observations based on early detection and broad classification, i.e. SN as opposed to a local event like an asteroid or stellar flare. However, spectroscopic follow-up will be very limited due to the sheer amount of objects that the LSST will discover. For this reason, getting the photometric detection and classification correct is incredibly important for conducting science with the data that will be obtained.

Detection will be ultimately limited by the single-visit magnitude depth. More optimistically the coadded single night depth will be the limiting factor if we accept that the variation of a SN over one night is not too significant to be stacked with the other images. Estimates with the current baseline observing strategy place the number of SNe discovered on the order of hundreds of thousands to several million covering a redshift range out to approximately $z = 1$ (LSST Science Collaboration et al., 2017). This clearly provides a vast amount of opportunity for science if the SNe can be properly classified. Classification in this sense is actually a two step process for SNe. I am referring to that first an object needs to be differentiated from other transients as a SN and further the type of SN must be resolved.

First, the issue of correct identification of a SN must be addressed. Without having a full lightcurve this can be done in a few ways. SNe are simply the brightest objects that we observe in the sky relative to their local environments, reaching absolute magnitudes greater than -16 for most of their life time, and peak magnitudes in the range of -18 to -21 (LSST Science Collaboration et al., 2009). Secondly, relatively few observations should be able to distinguish SNe from other transients. Their large distances essentially makes them spatially static sources. This will eliminate any solar system bodies. More distinctly, the timescale of their brightness variation is a characteristic shared with only massive star eruptions, like Luminous Blue Variables, which are much more rare and have lower peak luminosity (LSST Science Collaboration et al., 2017). From a purely photometric point of view, then only a couple observations spread over a few days should be able to correctly identify any transient as a SN from sampling of a single lightcurve.

Identification of the type of SNe on the other hand requires more sampling of the object's lightcurve over a larger part of the lifetime of the transient. In all, the lightcurves for different types of SNe do not vary too greatly and thus to describe enough features of an object's lightcurve for reliable classification a well-sampled lightcurve is necessary. It should be immediately obvious that a rolling cadence would significantly help improve the size of such a sample of frequently sampled lightcurves (LSST Science Collaboration et al., 2017). How these lightcurves will be used to perform the SN type identification will be discussed more later in the section regarding photometric classification.

With the observations and stresses on the cadence reviewed broadly, it is beneficial to briefly state a bit about the underlying physics of SNe. Generally, SNe can be characterized into two scenarios, thermonuclear fusion and core-collapse. These refer roughly to Type Ia SNe and everything else, respectively. The detailed physics is incredibly complex but to get an understanding it can be said the thermonuclear fusion scenario is brought about by ignition of runaway carbon fusion in a white dwarf star, and the core-collapse scenario is the result of loss of radiation pressure supporting the core over very massive stars which causes a self-gravitating collapse and subsequent explosion (Heger et al., 2002; Mazzali et al., 2007). The actual mechanisms that drive these events are not very well known. SNe Ia generally show less variability between events which is why they have been so useful as standardizable objects for cosmology. There is hope that with a more accurate description of the physics, particularly for a subset of core-collapse known as plateau SNe, that they may also be used for cosmology.

4 Kilonovae

Kilonovae are not very well understood from the perspective of astronomical observations. There are a lot of predictions for their expected lightcurve behavior, through exploration of what at the time was only theorized to be their progenitor scenarios. It is worth noting that the predictions show very smooth, relatively featureless, lightcurves (Rosswog et al., 2016). However, only one such object has been observed that has been confirmed, via multi-messenger astronomy, to be a KN (Kasliwal et al., 2017). While one event is not enough to characterize a population, this event can be used to set limits on relevant time scales for better determination of lightcurves for futures observations and as a baseline for predictions about future KN events. Something to point out about this event is that the observed lightcurves seem to exhibit more features than those generated from numerical simulations. This will be important when extracting features for use with machine learning algorithms for identification and classification if multiple types of KNe are distinguished. Similar to SNe, they require high frequency sampling. In fact the timescales relevant for early lightcurve evolution are on the order of hours, a much more rapid development than a typical SNe Ia (Villar et al., 2017).

As seen in the light curves produced for the recent Binary Neutron Star (BNS) merger, at optical detection on the order of ten hours past the gravitational wave triggering event was not able to capture the lightcurve peak (Villar et al., 2017). This places a very difficult requirement on any potential observing strategy if a significant number of events with well-sampled lightcurves are desired for science. Additionally, these events are much more rare than SNe, their rates are thought to be low-enough that estimates usually use a rate of only 300 per year per gigaparsec cubed volume (Rosswog et al., 2016). However, one recent paper using a higher estimate of 1000 per year per gigaparsec cubed places an estimate on the total number that will be detected with the LSST to be approximately 75 (Scolnic et al., 2017).

In this scenario it becomes increasingly clear that a rolling cadence would be immensely improve detection efficiency with its increased frequency of sampling from night to night. The sampling frequency of the WFD survey in the baseline cadence, field re-visit every three nights, would miss the significant features due to variations in the immediately post-explosion lightcurve. Using the recent BNS merger discovery as the benchmark for lightcurve behavior of future KN events, which is an assumption worth revisiting later, within the first ten days there is a decline in brightness within the range of 3 magnitude in y band to 8 magnitude in the u band. Ten days is significant here because in the observed lightcurves from this recent event the lightcurves in all bands exhibit much less variation after this date post-trigger (Villar et al., 2017).

As expected, with only one observed event, there is a lot unknown about the physics of KNe that needs to be worked out. However, the basic physics have been quite well studied, beginning when stellar evolution indicated that the late stages of many stars is a neutron star and observations indicated that many stars evolve in binary systems. It is important to note that the physical scenario at play combines a lot of different terminology. The progenitors are BNS and potentially, though unobserved as of yet, neutron star blackhole binaries. Their cataclysmic merger immediately results in a gravitational wave signal and a gamma ray burst. The observed KN is then what is often referred to as the gamma ray burst afterglow. Modeling this merger results in a scenario comprised of jet and disc components around the merger remnant (Rosswog et al., 2016). Though this work was done with numerical simulations prior to the BNS merger discovery, the predictions from this physical picture agree incredibly well with the observations (Kasliwal et al., 2017).

5 Classification of Transients with LSST

It is easy to forget the distinction between detection of a transient and the classification of the event. Detection of an event is straightforward, using a standard template image of region, perform image differencing with the template and a new image and look at the remaining signal. If a signal exists beyond some threshold against the background noise this can confidently be labeled a detection (LSST Science Collaboration et al., 2009). However, upon seeing a new event in the sky we cannot immediately say that it is a supernova or any other transient. This is because all photons are the same regardless of the source, in the sense that a photon isn't distinctly labeled as a SN photon etc. However, they do vary in spectral energy distribution and flux over time. These two facts alone allow us to distinguish one type of an event from another.

The challenge is, when knowledge of the intrinsic variations between sources is limited, or the data describing the event is limited, how are different events able to be distinguished from one another? This is where machine learning becomes incredibly useful. As discussed in Lochner et al. (2016), there are many approaches to using machine learning for classification of transients, specifically SNe in their case, from the dataset that the LSST will generate. The process for identification and classification of KNe will need to be considered though it is easily imaginable that it will be nearly identical to that of SNe. The process to perform a classification is actually two-fold. First, given the information about the lightcurve, features must be extracted. Once a feature set is defined, application of a machine learning algorithm determines the classification object from the power in the different features which define the object's lightcurve (Lochner et al., 2016).

5.1 Photometric Classification with Machine Learning

For this discussion it is assumed that an adequately sampled light curve is available for each of the events being classified. What this actually means in terms of sampled data points across a lightcurve is yet to be determined for KNe, but for SNe there is some consensus on what observations are necessary for a properly sampled lightcurve. As discussed in the literature these criteria are, one observation earlier than five days before peak, one observation greater than thirty days post-peak, at least seven nights of observations from the time twenty days prior to peak brightness until sixty days post peak brightness, and the largest gap between observations not greater than fifteen days (LSST Science Collaboration et al., 2017). Further selection cuts can be placed on the required observations for an adequate lightcurve, for example a signal-to-noise requirement for a given number of observations could be required. However the above criteria were those that are directly affected by the sampling of the lightcurve.

An important point to appreciate is that the classification process is not definite. Through any method, you are not going to get a definite result to which class the object belongs. What will be returned is a probability that the considered object belongs to the classes that its features are compared against. Judging the effectiveness of a single classification is not appropriate because it does not capture the variability inherent in the sampling of multiple events. For this reason classifying systems are given a training dataset where the classes of the objects are already known. This allows a classifying system to generate a Receiver Operating Characteristic (ROC) curve which demonstrates the effectiveness of the approach by plotting False Positive Rate (FPR) versus True Positive Rate (TPR). Thus maximizing the Area Under Curve (AUC), where an AUC of one demonstrates perfect classification with

no false positives, is a very appropriate metric for comparing classification schemes as was specified in the work reviewed here (Lochner et al., 2016).

As mentioned, the general classification approach considered here is comprised of two sections: feature extraction and machine learning. Machine learning here takes the place of any mechanism that judges the relative importance of the extracted features towards placement of an object into a specific class. Feature extraction takes a data set and reduces its description from a full dataserie to a set of features which is inherently smaller in size. This is effectively a dimensionality reduction of the problem. This can be done in two ways, a model dependent approach or an approach agnostic of the particular problem at hand (Lochner et al., 2016). The model independent approach is particularly attractive for the case of KNe because there is no sample to create a similar template matching method with which has been developed for SNe. It could be argued that a template method could be constructed from simulated lightcurves, but then this would potentially induce a bias as this would be the same, likely non-representative set, that the machine learning algorithm would be trained on.

In the analysis of Lochner et al. (2016) a comparison of feature extraction methods, both model dependent and model independent, is done. Three model dependent approaches were considered, one that was a template fitting method developed from extensive datasets of existing SNe lightcurves, SALT2, and two parametric models based on knowledge of general features of SN lightcurves. The model independent approach uses wavelet decomposition via a type discrete wavelet transform known as á trous wavelet transform. Then to reduce the dimensionality further principal component analysis, which is a generalized eigendecomposition of the wavelet coefficients into uncorrelated features for classification, is used to create a finite feature set (Lochner et al., 2016).

Upon extraction of the features then judgement of those features against a baseline feature set for potential classes is performed. Here this is done via application of machine learning algorithms, of which five are compared in the work reviewed. These were a naive bayes method, boosted decision trees, support vector machines, k-nearest neighbors, and an artificial neural network approach to classification. These were then tested with the combinations of feature extraction methods to judge which performed the best using the AUC metric as the determining quantity. Boosted decision trees when combined with either template fitting method or wavelet decomposition methods of feature extraction performed the best, achieving AUC scores of within a few thousandths around 0.98. The use of support vector machines for the machine learning classifier also performed quite well and I believe would be an interesting choice to consider further due to their versatility. Although, they are computationally intensive and practically a boosted decision tree can be just as versatile for different classifying problems if enough decision trees are combined (Lochner et al., 2016). If we take this approach to classifying SNe we should be able apply it to identification, and potentially real-time identification of KNe. Additionally this could be used to explore the classification of KNe with the LSST, if classes exist from differing progenitor scenarios.

6 Potential Areas for Future Work

The preceding discussion only barely scratches the surface on many of the issues that underly the science being undertaken with transient science with the LSST, but it does give a starting point for beginning to investigate deeper into different issues affecting the potential for science. Specifically, through reading the broad material that I have tried to summarize here, without simply restating a series of facts, has led me to some related questions that I

believe would be very interesting to investigate.

If I roughly address my questions in the order of the material that I presented, the first question regards the general cadence structure. What would be the affect of implementing a Target of Opportunity (ToO) program as a mini-survey in the general cadence structure. Specifically the program would be in response to a gravitational wave or GRB triggering event. In effect, what I envision could be implemented is an observing strategy similar to the DDF, with multiple visits to a field on a given night, which covers the localization area as communicated by the triggering event detections. With a FOV of nearly ten square degrees the LSST offers a very unique opportunity to quickly find the source of such an event. If the sky-localization is on the order of the most recent BNS discovery, which was approximately thirty square degrees, then only a few pointings of LSST taking only a couple minutes will cover the entire region (Kasliwal et al., 2017). Even if this isn't representative of the localization we can expect from future events, an analysis of localizations from only gravitational wave detectors, including the KAGRA detector which should come online in 2018, shows that this network will be able to localize nearly seventy-five percent of all detected sources to thirty degrees or better (Fairhurst, 2014). This offers LSST a great opportunity impact the science of electromagnetic counterparts to gravitational wave sources.

Secondly, if we use machine learning to identify and classify SNe and KNe, what effect does the sampling quality of the objects lightcurve on the ability to classify the object? From reading Lochner et al. (2016) it is not clear to me that this is addressed. How far can we degrade the sampling of the lightcurve to still reliably classify an object? Which regions of the lightcurve are the most crucial for classification? These questions can help create a definitive set of selection criteria for an adequately sampled lightcurve. Maybe this doesn't change the criteria currently determined for SNe but it could potentially set the criteria for KNe.

Training a machine learning algorithm to correctly identify the class of an object is dependent on the dataset used for training being representative of the whole population of that class. How well can we define a representative set for KNe? As I mentioned many times, there has only been one confirmed observation a KNe, one data point can obviously not be an accurate representation. What can we learn from this observation about the whole population and apply this developing a training set for machine learning? A separate but similar question relating to classification of KNe, are there actually different types that we can consider separate classes? Do the progenitor scenarios cause distinguishable features in the lightcurve?

As part of processing potential KNe through the feature extraction and machine learning algorithms features need to be determined. With most predicted KNe lightcurves being relatively featureless, though as noted the observed lightcurves were more feature laden, could additional filters or color information be used in tandem with a single lightcurve to provide information relevant for classification? It's true, my definition of feature may not be appropriate, in the sense I am intending I am saying a feature is an exhibition of variational behavior.

Lastly, and of most interest to me, with a defined sample of KNe, what cosmology can be done with KNe? What if we consider a sample without gravitational wave information, will KNe be usable like type Ia SNe for cosmological measurements? How well can we measure distance/redshift information from only an optical KN event? It would seem that if we consider KNe resulting from BNS mergers, they would have a similarly stringent mass range and thus variability in the population that makes type Ia SNe standardizable for cosmology. What other scenarios for cosmology could KNe provide that differ or offer a complementary

measurement of the parameters as determined by SNe observations?

References

- Fairhurst, S. 2014, *Journal of Physics: Conference Series*, 484, 012007. <http://stacks.iop.org/1742-6596/484/i=1/a=012007?key=crossref.d2fa0a752cbd3c0273bff0334adee375>
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2002, arXiv:0212469. <http://arxiv.org/abs/astro-ph/0212469><http://dx.doi.org/10.1086/375341>
- Kasliwal, M. M., Nakar, E., Singer, L. P., et al. 2017, *Science*, eaap9455. <http://www.sciencemag.org/lookup/doi/10.1126/science.aap9455>
- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, arXiv:1603.00882. <http://arxiv.org/abs/1603.00882><http://dx.doi.org/10.3847/0067-0049/225/2/31>
- LSST Science Collaboration, L. S., Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201. <http://arxiv.org/abs/0912.0201>
- LSST Science Collaboration, L. S., Marshall, P., Anguita, T., et al. 2017, arXiv:1708.04058. <http://arxiv.org/abs/1708.04058><http://dx.doi.org/10.5281/zenodo.842713>
- Mazzali, P. A., Roepke, F. K., Benetti, S., & Hillebrandt, W. 2007, arXiv:0702351. <http://arxiv.org/abs/astro-ph/0702351><http://dx.doi.org/10.1126/SCIENCE.1136259>
- Rosswog, S., Feindt, U., Korobkin, O., et al. 2016, arXiv:1611.09822. <http://arxiv.org/abs/1611.09822><http://dx.doi.org/10.1088/1361-6382/aa68a9>
- Scolnic, D., Kessler, R., Brout, D., et al. 2017, arXiv:1710.05845. <http://arxiv.org/abs/1710.05845>
- Villar, V. A., Guillochon, J., Berger, E., et al. 2017, arXiv:1710.11576. <http://arxiv.org/abs/1710.11576>