

## CRITERIA FOR SAMPLE SELECTION TO MAXIMIZE PLANET SENSITIVITY AND YIELD FROM SPACE-BASED MICROLENS PARALLAX SURVEYS

JENNIFER C. YEE<sup>1,9</sup>, ANDREW GOULD<sup>2</sup>, CHARLES BEICHMAN<sup>3</sup>, SEBASTIANO CALCHI NOVATI<sup>3,4,5,10</sup>, SEAN CAREY<sup>6</sup>,  
B. SCOTT GAUDI<sup>2</sup>, CALEN B. HENDERSON<sup>2</sup>, DAVID NATAF<sup>7</sup>, MATTHEW PENNY<sup>2,9</sup>, YOSSI SHVARTZVALD<sup>8</sup>, AND WEI ZHU<sup>2</sup>

<sup>1</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>2</sup>Department of Astronomy, Ohio State University, 140 W. 18th Avenue, Columbus, OH 43210, USA

<sup>3</sup>NASA Exoplanet Science Institute, MS 100-22, California Institute of Technology, Pasadena, CA 91125, USA

<sup>4</sup>Dipartimento di Fisica “E. R. Caianiello,” Università di Salerno, Via Giovanni Paolo II, I-84084 Fisciano (SA), Italy

<sup>5</sup>Istituto Internazionale per gli Alti Studi Scientifici (IIASS), Via G. Pellegrino 19, I-84019 Vietri Sul Mare (SA), Italy

<sup>6</sup>*Spitzer*, Science Center, MS 220-6, California Institute of Technology, Pasadena, CA, USA

<sup>7</sup>Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia

<sup>8</sup>School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel

*Received 2015 April 29; accepted 2015 August 9; published 2015 September 9*

### ABSTRACT

Space-based microlens parallax measurements are a powerful tool for understanding planet populations, especially their distribution throughout the Galaxy. However, if space-based observations of the microlensing events must be specifically targeted, it is crucial that microlensing events enter the parallax sample without reference to the known presence or absence of planets. Hence, it is vital to define objective criteria for selecting events where possible and to carefully consider and minimize the selection biases where not possible so that the final sample represents a controlled experiment. We present objective criteria for initiating observations and determining their cadence for a subset of events, and we define procedures for isolating subjective decision making from information about detected planets for the remainder of events. We also define procedures to resolve conflicts between subjective and objective selections. These procedures maximize the planet sensitivity of the sample as a whole by allowing for planet detections even if they occur before satellite observations for objectively selected events and by helping to trigger fruitful follow-up observations for subjectively chosen events. This paper represents our public commitment to these procedures, which is a necessary component of enforcing objectivity on the experimental protocol. They will be implemented for the 2015 *Spitzer* microlensing campaign.

*Key words:* gravitational lensing; micro

## 1. INTRODUCTION

### 1.1. Measuring the Distances to Microlensing Planets

While more than 6000 planets (and strong planetary candidates) have been found within about 1 kpc of the Sun (the great majority discovered via the transit and radial velocity techniques), there are only a handful of confirmed planets with known distances that are greater than 4 kpc and only one confirmed planet in the Galactic bulge (Batista et al. 2014). All of these distant planets were found using gravitational microlensing, and in most cases the distances were determined using the “microlens parallax” technique (Gould 1992). Microlensing would therefore appear to be the most natural method to measure the Galactic distribution of planets, i.e., to determine planet frequency as a function of Galactic environment. Such a measurement would provide important constraints on planet formation theories. For example, Thompson (2013) has suggested that gas-giant formation may have been inhibited in the Galactic bulge due to the high intensity of ambient radiation during the main epoch of star formation.

However, while roughly half of the  $\sim 30$  published microlensing planets have measured distances, this sample is heavily biased toward nearby systems. The reasons for this are well understood and are closely related to the general biases in astronomy toward nearby objects. First, nearby lenses have larger lens-source trigonometric parallaxes,

$\pi_{\text{rel}} = \text{AU}(D_L^{-1} - D_S^{-1})$ , which gives rise to larger microlens parallaxes

$$\begin{aligned} \pi_E &\equiv \frac{\pi_{\text{rel}}}{\theta_E} \frac{\mu}{\mu}; & \theta_E^2 &\equiv \kappa M \pi_{\text{rel}}, \\ \kappa &\equiv \frac{4GM}{c^2 \text{AU}} \simeq 8.1 \frac{\text{mas}}{M_\odot}, \end{aligned} \quad (1)$$

where  $\mu$  is the lens-source relative proper motion (in either the heliocentric or geocentric frame),  $\theta_E$  is the angular Einstein radius, and  $M$  is the lens mass. As explained in some detail by Gould & Horne (2013), the magnitude of  $\pi_E$  quantifies the amplitude of the parallax distortion on the microlens light curve, so that all other things being equal, larger  $\pi_E$  implies easier detection. The most common method for measuring microlens parallax has been to observe the effect of Earth’s acceleration on the light curve (so-called orbital parallax). However, for typical Einstein timescales  $t_E \sim 20$  day, this effect is quite modest. This means that in addition to nearby lenses and low mass lenses, one is biased toward abnormally long duration events. It is difficult (though probably not impossible) to quantify these biases, but the main problem is that due to these biases, there are simply no microlens planets in the Galactic bulge with measured microlens parallaxes. Indeed, the one confirmed Bulge planet had its distance measured by other means.

This brings us to the other method of measuring lens distances: direct detection of the lens. The main difficulty is

<sup>9</sup> Sagan Fellow.

<sup>10</sup> Sagan Visiting Fellow.

that the lens is superposed on a (usually) substantially brighter source, and remains so for typically a decade or more after the event. If the lens is sufficiently bright, then it is possible to directly detect it by measuring the combined source and lens light using high-resolution imaging (adaptive optics or *Hubble Space Telescope*) and subtracting out the source contribution, which is known from the light curve model. This, in fact, is how the distance to the only planet known to be in the Galactic bulge was measured (MOA-2011-BLG-293Lb; Batista et al. 2014). At the present time, this method is primarily limited to lenses that are at least 15% as bright as the source: otherwise the excess light due to the lens cannot be reliably detected. Hence, it is biased toward luminous (i.e., massive) and nearby lenses.

The alternative is to wait until the source and lens separate due to their relative proper motions (typically a few  $\text{mas yr}^{-1}$ ) and can be individually resolved. Again, this method is more easily applied to brighter lenses and with current facilities one must wait  $\sim 10$  year for the source and lens to separate sufficiently. When the next generation of 30 m telescopes are available, it will be applicable to much fainter lenses because these will separate sufficiently from the sources to be resolved within a few years due to their relative proper motions (Alcock et al. 2001; Gould 2014; Batista et al. 2015; Bennett et al. 2015; Henderson 2015).

Therefore, the only path at present to routinely measure the distances to lenses (especially faint lenses), and hence to measure the Galactic distribution of planets, is via space-based microlens parallaxes. In this approach, one observes a microlensing event simultaneously from Earth and from a satellite in solar orbit, and derives  $\pi_E$  from the difference in the two light curves (Refsdal 1966). There are some challenges to this method (over and above the problem of gaining routine access to such a satellite). First, the results are subject to a four-fold degeneracy in  $\pi_E$ , including a two-fold degeneracy in  $\pi_E$ . However, Calchi Novati et al. (2015) showed that it is possible in practice to break this degeneracy in the great majority of cases. Second,  $\pi_E$  does not by itself yield distances and masses. Rather this requires knowledge of  $\theta_E$ ,

$$\pi_{\text{rel}} = \theta_E \pi_E, \quad M = \frac{\theta_E}{\kappa \pi_E}, \quad (2)$$

and of the source parallax  $\pi_S$  ( $\pi_L = \pi_{\text{rel}} + \pi_S$ ), although the latter is usually known quite adequately. Fortunately,  $\theta_E$  is usually measured for planetary events because the normalized source size  $\rho \equiv \theta_*/\theta_E$  can usually be measured from the source crossing of the planetary caustic, while the angular source size  $\theta_*$  is almost always known from its color and magnitude. Moreover, even for non-planetary (and non-binary) events, which generally lack such crossings, the lens distance (and so mass) can usually be estimated quite well from the measured  $\pi_E$  and kinematic arguments (Calchi Novati et al. 2015). Finally, for the case that the source proper motion can be measured, this estimate becomes even more accurate (Yee et al. 2015).

Hence, as shown by Calchi Novati et al. (2015), one can obtain an accurate estimate of the cumulative distribution of lens distances from a given sample, and can in principle compare this to the cumulative distance distribution of detected planets.

## 1.2. *Spitzer* and the Galactic Distribution of Planets

To determine the Galactic distribution of planets, however, the detected planets must be compared to the underlying distribution of planet sensitivities, not simply of events. Calchi Novati et al. (2015) did not attempt to do this because there was only one planet in their sample (Udalski et al. 2015b), making a meaningful comparison impossible. The small number of planet detections was rooted in the nature of the observing campaign, which was a 100-hr “pilot project” to determine the feasibility of making such microlens parallax measurements using *Spitzer*. Thus, the *Spitzer* observations were limited to the subset of events judged most likely to yield  $\pi_E$ , and no special effort was made to find planets within these events via, for example, intensive follow-up observations.

Calchi Novati et al. (2015) argued, nevertheless, that it would be possible to estimate the cumulative distribution of sensitivities, simply by measuring the sensitivity of each event in the standard fashion (Rhie et al. 2000; Gaudi & Sackett 2000; Gaudi et al. 2002) and multiplying these sensitivities by the distance distributions in their Figure 3, even though the selection function of the events was unknown (and probably unknowable). This argument rested critically on the fact that the events were monitored from the ground and chosen for *Spitzer* observations without regard to the presence or absence of planets. This is a very similar argument to the one made by Gould et al. (2010) in the first study to derive planet frequencies from microlensing planet detections. Calchi Novati et al. (2015) further argued that their sample could be concatenated with future space-based samples, regardless of whether these were carried out using *Spitzer* or other satellites such as *Kepler*, and regardless of whether the selection function was the same or different, known or unknown. The only proviso was that, as with the Gould et al. (2010) and Calchi Novati et al. (2015) samples, the events were monitored without regard to the presence or absence of planets.

## 1.3. This Paper

The goal of the *Spitzer* microlensing parallax program (and indeed any space-based microlens parallax program) is to create a sample of events with well-measured parallaxes. These events are selected for *Spitzer* observations from the sample of microlensing events detected by ground-based surveys. If these events are observed by the satellite without regard to whether or not they have planets, the final sample can be used to determine the Galactic distribution of planets, e.g., by comparing the frequency of planets in the Galactic bulge with the frequency of planets in the disk. Hence, achieving this scientific goal has three primary considerations. First and foremost, the decision to select an event for *Spitzer* observations must be independent of any knowledge of the presence or absence of a planet.<sup>11</sup> Second, these observations must lead to a measurable parallax. Finally, maximizing the constraints on the Galactic distribution of planets requires maximizing not only the number of planets detected but also the range of planets that *could* be detected (i.e., the planet sensitivity), since the

<sup>11</sup> While it is more natural to think about the presence of planets creating bias in the sample, the absence of planets is also important. For example, if an event does not show evidence of a planet, it could be excluded from selection for parallax measurements because it is “uninteresting.” Hence, the absence of planets has similar potential to create bias in the final sample but in the opposite direction as the presence of planets.

**Table 1**  
Glossary of Terms

Term	Definition
General:	
Planet sensitivity	The range of planets (as a function of $s$ and $q$ ) that could hypothetically lead to measureable signals in the light curve. Also called “detection efficiency.”
The “sample”	The sample of microlensing events used to measure the Galactic distribution of planets. Implicitly requires that a parallax be measured.
Characterization of an event:	
“Good” event	An event with significant sensitivity to planets and a high likelihood of yielding a parallax measurement.
“Bad” event	An event with poor sensitivity to planets or a low likelihood of yielding a parallax measurement.
General light curve terminology:	
The “peak” of an event	The part of the light curve encompassing $t_0$
The “wings” of an event	The part of the light curve with $ t - t_0  \gtrsim t_E$
“Rising”	An event is rising if $t < t_0$ . Likewise, the “rising” part of the light curve refers to the part of the light curve with $t < t_0$ .
“Falling”	An event is falling if $t > t_0$ . Likewise, the “falling” part of the light curve refers to the part of the light curve with $t > t_0$ .
“Over”	An event is over if it has returned to its baseline (unmagnified) magnitude.

**Table 2**  
Definitions of Relevant Times

Variable	Quantitative (approx.)	Qualitative
$j$	1–7	Week of observations (beginning Thursday and ending Wednesday)
$t_{\text{first}}$	HJD'7182.5 = 2015 June 9	First date of <i>Spitzer</i> observations (set by the allotted observing window or Sun-angle constraints)
$t_{\text{fin}}$	HJD' 7222.5 = 2015 July 19	Last date of <i>Spitzer</i> observations (set by the allotted observing window or Sun-angle constraints)
$t_{j,\text{dec}}$	Monday UT 9:00	Time when <i>Spitzer</i> observations are finalized for Week $j$
$t_{j,\text{next}}$	$t_{j,\text{dec}} + 3 \text{ days} = \text{Thursday}$	Time of first <i>Spitzer</i> observations for Week $j$
$t_{\text{sel}}$	$t_{j-1,\text{dec}} < t_{\text{sel}} \leq t_{j,\text{dec}}$	Time when event is selected for <i>Spitzer</i> observations
$t_{\text{com}}$	$\geq t_{\text{sel}}$	Time of public commitment to an event
$t_{\text{stop}}$	$\leq t_{\text{fin}}$	Time when <i>Spitzer</i> observations cease

**Note.** Although these terms are defined relative to the *Spitzer* campaign that is described in this paper, they are easily generalized for any targeted microlens parallax program.

detection efficiency is a crucial component to any measurement of the planet occurrence rate.

The primary goal of the present paper is to determine a strategy to monitor events with *Spitzer* to ensure the final sample of events with parallaxes is monitored without reference to the presence or absence of planets, because this is the property of the sample that is most difficult to control. At the same time, this strategy is driven by the additional goals of maximizing both the planet sensitivity of the monitored events and the likelihood of measuring parallaxes. By defining this strategy in advance of the observations, we can create a sample of events with measured parallaxes with maximal leverage for measuring the occurrence rate of planets as a function of Galactic distance.

We begin, in Section 2, with a general discussion of how events may be selected, either objectively or subjectively, and how that selection affects the resulting planet sensitivities of those events. Since much of that discussion is guided by planet sensitivity and the practical considerations of the *Spitzer* campaign, the reader may also wish to refer to Sections 4 and 5.2. Sections 3–5 cover the various ingredients necessary to define criteria for selecting events, namely the planet sensitivity

and the probability of measuring parallax. Then in Section 6, we formally define objective criteria for the *Spitzer* campaign to select events and also determine their observing cadences. Section 7 then discusses specific guidelines for subjectively choosing events for this campaign, and Section 8 specifies how the available observations will be distributed among the targets. Finally, we give a brief summary in Section 9. We have provided a glossary of terms in Table 1 to clarify some qualitative statements we may use and also colloquialisms that have arisen in microlensing.

## 2. OBJECTIVE VERSUS SUBJECTIVE SELECTION CRITERIA

There are many choices that must be made with respect to *Spitzer* observations of any individual event. One must decide when to begin making such observations, when to commit to the target,<sup>12</sup> with what initial cadence, whether and when to change this cadence, and whether and when to halt the observations. This entire chain must be carefully established in

<sup>12</sup> The distinction between deciding to observe a target and committing to a target is discussed in Section 2.2.

**Table 3**  
Ways to Select an Event for Satellite Parallax Observations

Type	This Event	Planets and Sensitivity from
Objective	Meets pre-defined criteria.	The entire event may be counted.
Subjective, Immediate	Does not meet pre-defined criteria, but is selected anyway.	The selection date ( $t_{sel} = t_{com}$ ) onwards may be counted.
Subjective, Conditional	Has unconstrained fits and may or may not be “good.” It is selected anyway but with specific conditions (defining event failure) for halting observations.	The selection date ( $t_{sel} = t_{com}$ ) onwards may be counted
Subjective, Secret	Has unconstrained fits and may or may not be “good.” It is selected anyway but may be dropped at any time until the event is publicly committed to.	The commitment date ( $t_{com} > t_{sel}$ ) onwards may be counted.

**Note.** There are four ways an event can be selected for satellite parallax observations, with three distinct modes of subjective selection. These simply reflect ways in which events may enter the sample. Any events that ultimately do not yield a parallax measurement will be removed from the sample, regardless of how they are selected.

order to ensure the fundamental requirement that the observational sequence be indifferent to the presence or absence of planets. Table 2 gives a brief overview of the relevant decision points, and Section 5.2 discusses how the specifics of *Spitzer* operations set the quantitative definitions.

The starting point is the choice to begin monitoring an individual event, i.e., “triggering” observations. This choice can be made either because the event meets some objective criteria (in which case the “choice” is automatic) or according to some subjective criteria of the team organizing the observations. Table 3 summarizes the various channels through which observations may be triggered. However, all other decisions about the monitoring are heavily influenced by the first dichotomy (objective versus subjective), so we divide the discussion according to it. As we will describe, optimal event selection requires a combination of objective and subjective selection. Because this is so, one must also decide what to do in advance if the objective and subjective selection procedures collide. That is, what should be done if an event is selected subjectively, but later meets the objective criteria for selection.

We discuss the architecture of the selection procedure before discussing the criteria themselves because the architecture is both non-trivial and logically independent of the criteria. Within the framework of this discussion, one must keep in mind that the overall goal is to maximize the sensitivity of the experiment to planets and that planet sensitivity rests primarily on ground-based observations (see Section 4). At the same time, after an event is selected, its entry into the final sample to measure the Galactic distribution of planets requires that its parallax is measured, which depends primarily on *Spitzer* observations.

In the following sections, we occasionally give examples to illustrate the points under discussion. For these examples, it may be helpful to keep in mind that some of the observables that affect planet sensitivity and parallax measurements include the time of the event peak  $t_0$ , the magnification of the event (larger is better), and the magnitude of the event as seen from the ground or from *Spitzer* (brighter is better). These observables and their relationship to planet sensitivity and parallax are discussed in detail in Sections 4 and 5. The final criteria are given in Section 6.

## 2.1. Objectively Chosen Events

### 2.1.1. Objective Selection

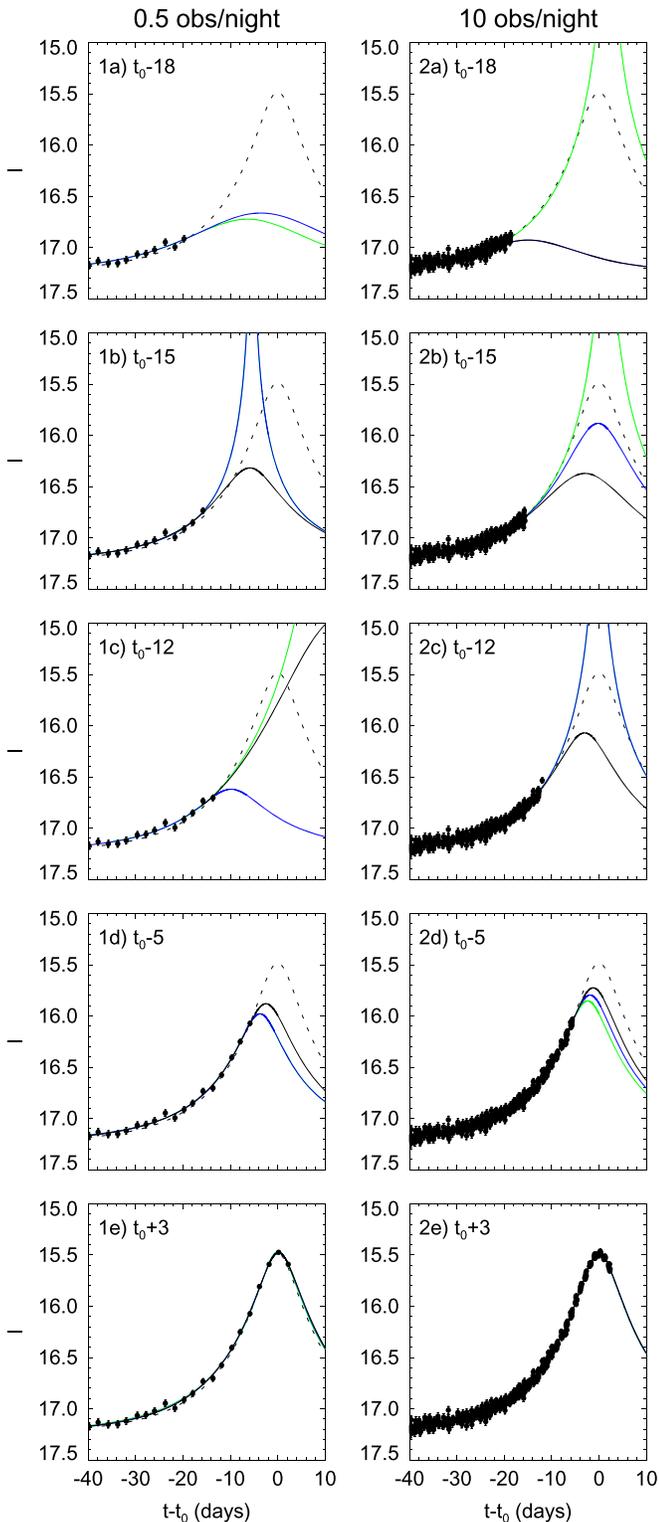
The great advantage of choosing events by objective criteria is that any planet that is discovered during an event that is so chosen can be included in the sample, and similarly, the planet sensitivity of the event over its entire duration can be included in the analysis as well. For example, suppose an event is announced by a survey group on May 1 but *Spitzer* observations cannot begin until June 8. The event undergoes a planetary deviation on May 15, peaks on May 28, and on June 3 is scheduled for *Spitzer* observations beginning June 8 because it is found to meet previously chosen objective criteria. Then the planet can be included in the sample, even though it was discovered before the *Spitzer* observations began, and even before it was known that it would eventually satisfy the objective conditions that triggered observations.

By contrast, in the absence of such criteria, the event could have been selected for *Spitzer* observations subjectively. In that case, neither the previously discovered planet, nor the planet sensitivity from the the entire pre-decision period, could be included in the analysis. Otherwise, the presence or absence of the planet could influence the event’s “selection” (i.e., inclusion in the final sample with measured parallaxes; Section 5.1 discusses what is meant by a “measured parallax”).

### 2.1.2. Objective Cadences

A large fraction of objectively chosen events will be similar to the hypothetical one described above in that most of their planet sensitivity will be in the data obtained before the time that the *Spitzer* observations begin. Therefore, it is absolutely essential that the cadence be chosen objectively as well. In order to enter the sample, the event must have a measured parallax. If the cadence is not chosen objectively, events with planets could receive extra observations to help ensure they have measured parallaxes. We will discuss specific algorithms to make this choice in Section 6.

Finally, we will just mention that to avoid wasting observations, there must also be a mechanism for halting this objectively determined *Spitzer* observation schedule when



**Figure 1.** Example fits to simulated event data. The dashed line shows the underlying point lens ( $u_0 = 0.2$ ,  $t_E = 20$  days) from which the data were generated with Gaussian errors. The right-hand panels show data sampled 10 times/night, every night, and the left-hand panels show the same data, but thinned to 1 observation every two nights. Starting from the top, the rows show the approximate  $1\text{-}\sigma$  range of fits (solid lines) to the data at  $t = 18, 15, 12, 5$  days before the peak and at  $t = 3$  days after the peak. These figures show that the fits to the high-cadence data (10 obs/night) are much better at capturing the true, underlying behavior of the events, whereas fits to the sparse data (0.5 obs/night) often give misleading fits that do not encompass the true event.

these observations are no longer useful. However, the only permitted reasons for doing so are that the microlens parallax has been measured or that the event (as seen from *Spitzer*) has already returned to baseline (i.e., is now essentially unmagnified), so that no further improvement is possible.

## 2.2. Subjectively Chosen Events

### 2.2.1. The Need for Subjective Selection

At first glance, the advantage conferred by the objective approach appears to be so great that one might wonder why one would consider the subjective approach at all. The main problem is that it is impossible to define objective criteria loosely enough to capture all events of interest without at the same time introducing a large number of events either with poor planet sensitivity or a low probability of yielding a parallax measurement (“bad” events). Hence, the objective criteria must be strictly defined so that all of the selected events are both highly sensitive to planets and have a high likelihood of yielding a parallax (“good” events). Otherwise, a large amount of observing time will be wasted on events of little value.

A second issue is whether or not an event chosen objectively will yield a parallax measurement. The central difficulty is that the event’s objectively chosen *Spitzer* observational sequence must yield a parallax measurement. It can be quite difficult to choose such observational sequences based solely on objective criteria, or to determine which events might be worth the observational effort to obtain parallaxes, or even to determine which might yield parallax measurements with any sort of effort. These difficulties can all be more effectively addressed by subjectively choosing the event, in which case one can also choose a cadence (or cadence algorithm) that is individually tailored to that event. For example, suppose that it is known that an event will meet objective criteria in 2 weeks, but this will allow for only 1 week of *Spitzer* observations. If we wait to start observations until this date, we risk the possibility that 1 week of data will be insufficient to measure a parallax, in which case the entire event and its sensitivity is lost. In contrast, we could select the event subjectively now to get 3 weeks of *Spitzer* observations and vastly improve the probability that those observations will yield a parallax.

Hence, there are two reasons that events might be chosen subjectively. First, because the objective criteria cannot capture all events of interest. Second, because an earlier subjective trigger may make the difference between measuring a parallax or not. As we discuss in the next section, not much planet sensitivity is lost by subjectively choosing events. We will also discuss in Sections 2.3 and 2.4 the resolution of conflicts for events that may be selected both subjectively and objectively.

### 2.2.2. Subjective Selection

If, for whatever reason, an event fails to meet the objective selection criteria, the team may decide to observe it anyway. The reasons one might want to do this are discussed in some detail below, but first we focus on the consequences of this decision, which leads to three types of subjective selection as defined in Table 3.

Actually, the key decision is not whether to begin *Spitzer* observations but when to (publicly) commit to a schedule of

*Spitzer* observations. Once such a decision is made, it must be accompanied by a public commitment of an observational sequence (or to an objective algorithm for determining that sequence). Otherwise, one could choose events without any knowledge of whether they would later show planets, discover that they do indeed host planets, and then be biased to observe them more frequently in order to preferentially increase the probability that their parallax will be measured, thus placing them in the sample for measuring the Galactic distribution of planets.

In the case of subjective decisions, all planets that show up after the public commitment date ( $t_{\text{com}}$ ), as well as all sensitivity to planets after this date, would be kept in the analysis, while all planets and planet sensitivity from before this date must be excluded.

We illustrate the need to separate the *decision* to observe from the *commitment* to observe with two examples.

First, suppose that at a given *Spitzer* decision time (typically  $t_{\text{sel}} = \text{Monday}$ , see Section 5.2 for details of the logistical constraints on *Spitzer* observations) an event that has not yet peaked has an ambiguous future, with some chance that it will rise in magnification sufficiently to have good sensitivity to planets and to enable a viable parallax measurement (e.g., panel (2a) of Figure 1). However, as of Monday, this cannot be established with any confidence, although ground-based data are likely to resolve this ambiguity within a few days. Based on this assessment, the team decides to observe this event once per day during the week beginning at the next upload three days hence, i.e.,  $t_{j,\text{next}} = \text{Thursday}$ . Even though the future of the event is uncertain, preemptive observations (rather than waiting for the next upload cycle, one week hence) could make the difference between a good parallax measurement and a meaningless upper limit. This is especially true since the *Spitzer* observations can only be updated once per week. Then on Sunday, ground-based data show that this event has risen sufficiently that its future behavior can be predicted well enough to determine that it is an interesting target (e.g., panel (2c) of Figure 1 as compared to panel (2a)). The team then announces that it is committed to monitoring this event and also announces its chosen cadence (or objective procedure for determining such cadence). All planets and planet sensitivity from  $t_{\text{com}} = \text{Sunday}$  forward can then be included in the analysis. Similarly, only *Spitzer* data beginning on Sunday can be used in the initial parallax measurement that determines whether the event enters the sample. Even though the Thursday–Saturday observations cannot be used, this preemptive decision to observe has resulted in extra observations (Sunday–Wednesday) that can be used to improve the parallax measurement as compared to waiting until the following Thursday to begin *Spitzer* observations. With regard to the specific role of *Spitzer* data in planet sensitivity and discovery: if the team has seen the first few days of *Spitzer* data prior to the Sunday commitment, then these cannot be included, but if they have not yet seen them, these can be (since in this case they would not have influenced the decision).

For a second example, consider the same case as above but with the event peaking at low magnification, and hence having both low sensitivity to planets and low likelihood of a measurable parallax (e.g., the data instead follow the solid black line in panel (2a) of Figure 1). The team then decides not to continue monitoring this event. Because the team never committed to the event, they have no obligation to continue

monitoring it, and so it is entirely dropped from the sample. By the same token, any planets discovered from this event cannot be included in the analysis. In contrast, if the decision and commitment were the same process, in order to avoid bias, the team would be required to continue monitoring an event that it recognized as worthless.

In the first example given above, the public commitment to observe the event was made after the *Spitzer* observation sequence was uploaded to the spacecraft and indeed after those observations began ( $t_{\text{com}} > t_{j,\text{next}} > t_{\text{sel}}$ ). This is the situation described as “Subjective, secret” in Table 3. However, the same principles apply to, for example, an event that is newly recognized a few days before the upload decision and is already recognized to be promising (e.g., panels (2c) or (2d) of Figure 1). The team could publicly commit to observing this event immediately with specified cadence (i.e., “Subjective, immediate” in Table 3), and then all planets and planet sensitivity from that date forward would be included (provided the parallax was measured well enough to put the event in the sample).

Finally, in the first example, the team might publicly commit immediately to observe the ambiguous event, but with explicitly stated criteria for halting such observations (Table 3, “Subjective, conditional”). For example, it might specify that observations would be discontinued if the event failed to reach some specified magnification before the time of the next upload. In this case, whether or not the observations were continued, the event might have sufficient *Spitzer* data to measure the parallax and enter the sample. If so, all planets and planet sensitivity from after  $t_{\text{sel}} = t_{\text{com}}$  could enter the sample.

At first sight, “Subjective, conditional” seems clearly superior to “Subjective, secret” because it enables inclusion of more planets and more planet sensitivity. However, this is not always the case. In fact, it often happens that the uncertainty in predicting rising events includes not only to their time of peak and peak magnification, but extends even to when this knowledge will be reliably available. For example, consider the evolution of the fits to the data in column 1 of Figure 1, which illustrates how the fits may change dramatically due to small fluctuations in the data and may not capture the true, underlying behavior of the event. If the team cannot reliably predict the future course of the event, it may not be possible to correctly pre-define criteria for halting observations. This creates the risk of being committed to observing many bad events or being forced to halt observations of an event that turns out to be good, but with different parameters than initially supposed.

Thus, as summarized in Table 3, subjective decisions can take a considerable variety of forms. The only constraint is that they must be constructed to avoid the possibility that the presence or absence of planets detected after  $t_{\text{com}}$  will influence the cadence of observations.

### 2.2.3. Subjective Cadences

In contrast to objectively chosen events, the cadences for subjectively chosen events can be chosen by the team. However, they must be fully specified at the time of the commitment to observations ( $t_{\text{com}}$ ). As with objectively chosen events, after  $t_{\text{com}}$  the only permitted reasons for halting the scheduled *Spitzer* observations are that the parallax has already been measured or the event has returned to baseline (as seen from *Spitzer*).

### 2.3. Collisions: Subjectively Chosen Events that Meet Objective Criteria

An event that has been subjectively chosen may, at a later date, meet the objective criteria. In this case, the objective selection of the event *must* take precedence. Otherwise, there is no point in having objective criteria since they could always be subjectively overridden. The objective criteria will specify an objectively determined cadence of observations. The more frequent of the subjective or objective cadences will take precedence for future observations. However, only the observations taken after the objective selection and at the objective cadence can be used to determine whether the parallax measurement is adequate to enter the sample. If it does enter, then all planets that are detected and all planet sensitivity (including from before the commitment to the event,  $t_{\text{com}}$ ) enter the analysis. The next section will clarify the reason for maintaining a higher, but subjective, cadence even if only a fraction of the data can be used for the objectively measured parallax.

Note that for these cases all planets discovered in *Spitzer* data (as well as all planet sensitivity from these data) should be included in the analysis, regardless of whether these data were taken before or after or in response to the commitment to the subjective selection of the event. The only exception would be if the planet were detected purely in the *Spitzer* data and subsequent *Spitzer* observations were increased because of it (see Section 2.5).

### 2.4. Reversions from Objective to Subjective

If an event has (1) been chosen subjectively, (2) subsequently satisfied the objective criteria and thus triggered a conversion to being objectively chosen, (3) fails to yield a parallax based on the objective portion of the *Spitzer* light curve, and (4) does yield a parallax based on all, post-commitment *Spitzer* data, then it automatically reverts to subjective status. In this case, only planets and planet sensitivity from the post-commitment part of the light curve can be included in the analysis.

The basic reason is that none of the decisions made after this commitment depended in any way on the presence or absence of a planet (other than possible planets before this date, which must be excluded from the calculation of planet occurrence).

### 2.5. Planet-characterization Observations

As discussed above, whether the decision to observe an event with *Spitzer* is objective or subjective, the cadence must be chosen without reference to the presence or absence of a planet. For the objectively chosen events, this cadence must be determined objectively by pre-determined criteria. For subjectively chosen events, they must be announced at the times of the decision, i.e., prior to the discovery of any planets.

However, once a planet is suspected, it can be important to increase the pace of observations from the ground and/or *Spitzer* in order to improve the *characterization* of the planet. In this case, one can increase the observational cadence, but only the observational data that would have been taken under the pre-determined schedule can be used to assess the *detectability* of planets and the measurability of the parallax. See, for example, Yee et al. (2012). Application of this rule is straightforward for *Spitzer* observations, which is the main focus in this section, because, as we have specified, the

observational cadence is in fact pre-determined. The situation is more complicated for ground-based observations of the same event, to which we now turn.

### 2.6. Ground-based Follow-up Observations

The majority of microlensing planets published to date were discovered by a combination of microlensing surveys that find large numbers of events, mostly with a low-to-moderate cadence of observations, and follow-up surveys that target individual events for more intensive monitoring in order to enhance the discovery and characterization of planets. In fact, survey groups sometimes go into “follow-up mode” by increasing the cadence in the survey field that contains a particularly interesting event (Yee et al. 2012) or even re-centering an existing field to incorporate a particular event (Yee et al. 2014). For present purposes, observations by specialized follow-up groups and survey teams in “follow-up mode” are equally considered as “follow-up.” The only exception would be follow-up observations that are determined by purely objective criteria.

Follow-up observations must be evaluated with respect to two questions: first, how do they impact planet sensitivity and planet detection, and second how do they impact the measurement of microlens parallax (and so entry into the sample).

#### 2.6.1. Follow-up and Planet Sensitivity/Detection

As with survey observations, follow-up observations contribute to planet sensitivity and detection for the entire duration of objectively chosen events and for all post-commitment observations of subjectively chosen events (which may be a time-span much longer than the window for *Spitzer* observations). Indeed, while the fundamental point of announcing subjective choices of events for *Spitzer* observations (i.e., the commitment to observe an event) is to establish a record of what planet sensitivity can be included, a major secondary goal is to encourage observations of these events, particularly those that are not well covered by the surveys, (see Section 4.1 for a general outline of standard survey strategies). Such subjective announcements automatically have the effect of encouraging follow-up observations because prior to such announcement, the planet detections can enter only if the event has already satisfied, or ultimately proves to satisfy, objective criteria.

The only question is whether changes in the adopted cadence of follow-up observations due to the perceived presence of planets influences the detectability of the planets. This can happen in principle if the planet generates an observed perturbation (in either survey or follow-up data) that is strong enough to trigger interest but not, by itself, sufficient to confirm the presence of the planet. In this case, follow-up observations aimed at characterizing the planet can make the difference between it being undetectable and being detectable.

This issue is not particular to parallax experiments: it pertains to any experiment that aims to make a statistical statement about planets using microlensing follow-up data. For example, Gould et al. (2010) noted that two of their six detected planets occurred in an event that showed an early (and in itself, not comprehensible) perturbation that ultimately proved to be due to a planet. This early perturbation did trigger additional observations, but these subsided over the next few days. Observations only intensified again when the event

approached high magnification, which was the standard trigger for high-cadence follow-up observations.

In brief, these issues arise in a minority of planetary events, and usually can be resolved based on records of the decision making. While it is possible that there may be unresolved cases in the future, the importance of characterizing planets is too great to allow this possibility to interfere with aggressive follow-up response to the tentative detection of planets.

### 2.6.2. Follow-up and Parallax

The other aspect that must be considered is the role of follow-up observations in measuring microlens parallaxes and so in putting individual events into the final sample used to measure the Galactic distribution of planets. There are two relatively distinct ways that this can happen. First, the planetary-induced features in the light curve may substantially increase the precision of the parallax measurement from *ground-based observations alone* and so make the difference between whether it is included in or excluded from the sample. Second, the follow-up observations may improve the precision of the non-planetary light-curve parameters ( $t_0, u_0, t_E$ )<sub>⊕</sub> (i.e., the time of maximum, the impact parameter, and the Einstein timescale) and thus improve the precision of the parallax measurement that comes from the comparison with *Spitzer* data. We discuss these in turn.

An & Gould (2001) argued that events with three peaks (features due to caustics induced by a companion) would gain significantly improved parallax measurements relative to otherwise similar point-lens events. However, in the intervening years, almost nothing has been done to investigate the role of the perturbations in the parallax measurements for detected planets. For example, early modeling showed that the immediate post-peak light curve of MOA-2009-BLG-266 yielded a surprisingly good parallax measurement, despite the fact that it is extremely rare for orbital parallax to be measurable before an event substantially returns to baseline. This was attributed to the sharp deviations in the light curve caused by a Neptune mass-ratio planet. Yet Muraki et al. (2011) say that the dominant source of the  $\Delta\chi^2 = 2789$  parallax signal derives from the MOA data in the wings and that very little parallax signal comes from the perturbed region. However, we find from fitting the MOA data alone (and excluding the perturbed region) that the parallax signal from these data is only  $\Delta\chi^2 = 205$ , implying that the perturbation could in fact play a major role in the strength of the parallax signal, as An & Gould (2001) anticipated.

It is very likely that planetary perturbations play a significant role in the strength of the parallax signal in many other events as well. This is likely to be a partial explanation for the fact that roughly half of all microlens planets published to date have measured parallaxes. Although historically, there has been a lack of interest in where the parallax signal comes from, with respect to *Spitzer* observations this question is of cardinal importance. If planetary events have more easily measured parallaxes than non-planetary events, then the sample of objects with measured parallaxes is biased.

Hence, in determining whether the event enters the sample, it is essential to ask whether it would have a sufficiently well-measured parallax even if there had been no planet. This means both eliminating follow-up observations that were triggered by the presence of a planet and also (for this purpose) replacing the actual light curve with a fake, point lens light curve based on

the event’s parameters ( $t_0, u_0, t_E$ ). This fake light curve could then be fit to determine the strength of the parallax signal from the point lens event. This procedure must also be applied in cases in which the planetary perturbation is seen from *Spitzer*.

A closely related issue is that an incipient planetary anomaly might be misinterpreted as evidence for an approach to high magnification, and hence trigger an “honest” (i.e., seemingly non-planet-related) decision to observe the event, either ground-based follow-up observations or observations by *Spitzer*. This occurred, for example, for ground-based observations of the second microlensing planet, OGLE-2005-BLG-071 (Udalski et al. 2005). Such observations by *Spitzer* might enable a parallax measurement that would not have been possible for a point-lens event. This must be checked in all cases, which again can be done through the use of fake light curves.

By contrast, it will be relatively rare for follow-up data to play a major role in the determination of the event’s point-lens parameters simply because the main features of the event that enter the parallax measurement usually derive from long-term observations and so are well measured from survey data. However, high-magnification events can be an exception, primarily because dense coverage of the peak is often required to determine the ground-based impact parameter, particularly if the surveys cover the field at low cadence. This was exactly the case for OGLE-2014-BLG-1049, one of the 21 events analyzed by Calchi Novati et al. (2015). However, in the great majority of cases (including this one), an improved parallax measurement is simply one of the benefits of conducting follow-up observations of parallax candidates. The primary motivation is generally the increased probability of detecting planets (Griest & Safizadeh 1998). The only exception would be if these critical near-peak observations were triggered by the known presence of a planet (rather than the hope of finding one). We expect such planet-triggered parallax-assisting measurements will be extremely rare and mention them primarily for completeness.

## 3. PRINCIPAL INGREDIENTS FOR DEVELOPING SELECTION CRITERIA

At a fundamental level, there are only two considerations in deciding whether to choose one event over another for *Spitzer* observations. First, how sensitive to planets is that event? Second, how likely is it that if *Spitzer* observations are undertaken, a microlens parallax will actually be measured? Thus we may schematically define a “quality factor”  $Q$  for the experiment

$$Q = \sum_i S_i P_i, \quad (3)$$

where  $S_i$  is the planet sensitivity of the  $i$ th event chosen, and  $P_i$  is the probability that it will yield a parallax measurement. Then the goal in developing selection criteria should be to maximize  $Q$ . One issue posed by Equation (3) is that planet sensitivity is actually a function of planet properties. However, this is easily resolved by adding two additional indices to  $S_i$  to specify these properties. We return to this complication in Section 4.3. A more fundamental challenge is that both  $S_i$  and  $P_i$  may be poorly known at the time that the decision must be made to initiate observations (c.f. Figure 1). The first step toward figuring out how to proceed in the face of these

uncertainties is to review what makes an event sensitive to planets and how much about this can be known at any given stage in its evolution, and to then address the corresponding questions about the measurability of its parallax.

#### 4. PLANET SENSITIVITIES

After an event is over, its “sensitivity” to planets<sup>13</sup> can be rigorously defined as a function of two variables, the planet–star mass ratio  $q$  and the planet–star separation  $s$  (in units of  $\theta_E$ ). See, for example, Gaudi & Sackett (2000), Gaudi et al. (2002), Gould et al. (2010), and Cassan et al. (2012). However, when choosing an event for additional observations while it is evolving, one must be guided by a more qualitative understanding of what properties make the event sensitive to planets and judge how likely it is that these will appear.

##### 4.1. Ground-based Microlensing Observations

Because microlensing events depend on the chance alignment of two stars at radically different distances, for the most part, these events cannot be predicted in advance. Hence, microlensing surveys monitor millions of stars toward the Galactic bulge, where the stellar density, and therefore the microlensing event rate, is highest. If the survey team sees a star brightening in a manner consistent with microlensing, they issue an alert announcing a new microlensing event.

Microlensing surveys monitor these fields at a variety of cadences (e.g., <http://ogle.astrouw.edu.pl/sky/ogle4-BLG/>). “High-cadence” fields are monitored one to several times per hour (e.g., column 2 of Figure 1), which is sufficient to characterize small planetary signals from terrestrial-mass planets. “Moderate-cadence” fields are observed several times per night, which can capture planetary signals from ice giants. “Low-cadence” fields are monitored once per night or less than once per night (e.g., column 1 of Figure 1) and are generally focused on producing alerts of ongoing microlensing events that can then be monitored more intensively by follow-up groups, although these survey observations themselves are occasionally sufficient to characterize large (gas-giant) planets.

In the context of this paper, we will focus on survey data from two sources. The primary data will come from the OGLE-IV survey (Udalski et al. 2015a), whose sky coverage is given in the above URL. In addition, we will consider data from the new Korea Microlensing Telescope Network (KMTNet) fields, which in 2015 will be conducting its first year of routine microlensing survey observations. Under the assumption that these observations are carried out, we will include in our evaluations data from the four core KMTNet fields, which will be observed many times per hour.

##### 4.2. Planet Sensitivity: Qualitative Features

Microlensing detects planets around the lens stars as perturbations to a standard (point-lens) microlensing event. That is, the microlensing event is overall dominated by the gravitational potential of the lens star (host), which splits the light into two images whose position and size change as a function of time. If a planet interacts with one of those images, it creates a perturbation that distorts the shape and total

magnification of that image, which can lead to a detectable signal. See Gaudi (2012), in particular Figures 4 and 5.

The total “planet sensitivity” of a given event depends on two factors. The first is the intrinsic sensitivity of the event to planets. Larger planets, and planets that are closer to the “Einstein ring” (circle with angular radius  $\theta_E$ ) are easiest to detect. The larger the images are, the more sensitive they are to planetary perturbations. Hence, more highly magnified events are more sensitive to planets, i.e., to smaller planets and to planets that are farther inside or outside the Einstein ring (Gould & Loeb 1992). The most extreme example would be the high magnification events (peak magnification of the underlying point lens event  $A(t_0) = A_{\max} \gtrsim 100$ ) in which the images form an almost perfect ring that probes a wide range of separations and is easily perturbed (Griest & Safizadeh 1998).

The second factor affecting planet sensitivity is how well the data cover potential perturbations. The quality of this coverage is defined by two principal characteristics: cadence and photometric precision. Regarding the first, planetary perturbations typically last between a few hours and a few days, so a cadence that is a factor of  $\sim 10$  more frequent than that is necessary to characterize those perturbations. Since such perturbations may occur at any time during the event, the sensitivity will be greatest if the data are continuous during this period. However, if observing resources are limited, then restricting continuous observations to the most highly magnified (hence most sensitive) parts of the light curve may be the most productive approach. Nevertheless, planets can appear in all parts of the light curve, even after the main event is over or before it began (e.g., Poleski et al. 2014a), so that observations are never “wasted.”

Photometric precision is mainly governed by source brightness. This factor therefore favors intrinsically bright sources, but highly magnified sources can also be bright (just at the moment that they are most sensitive to planets).

Even though we have not yet examined the other key determinant of  $Q$  (i.e., the probability of measuring  $\pi_E$ , Section 5), we can already draw a few general lessons from the above analysis of planet sensitivity. First, higher peak magnification is the best single indicator for choosing events (provided that the peak region can be intensively monitored from the ground). Second, for objectively chosen events, those that are in high-cadence survey fields are substantially more valuable than those that are not. This is because high-cadence events can yield planets long before the onset of *Spitzer* observations, or even before the event was recognized as microlensing, whereas low-cadence events generally cannot. Third, for events in low-cadence fields that achieve relatively high-magnification, it is important to mobilize follow-up observations prior to peak. If these are chosen subjectively, then the desirability of choosing this target will depend critically on the expectation that such follow-up observations will be carried out.

##### 4.3. Planet Sensitivity: Quantitative Determination

The question of quantifying the planet sensitivity  $S_i$  of each event is mainly outside the scope of the present paper because this can only be done after the event is largely over and hence after all of the observing decisions that are the subject of this paper have already been made. That is, these decisions must be made on the basis of the qualitative indicators discussed above.

<sup>13</sup> In other planet-finding contexts, e.g., radial velocities, “planet sensitivity” is often referred to as “detection efficiency.”

Nevertheless, for completeness we present here a broad overview of the relevant issues.

Planet sensitivity is measured as a function of two parameters ( $s$ ,  $q$ ) and so can be formally written  $S_i^{s,q}$ . These are two of the seven basic parameters that are the minimum needed to describe a planetary event. Three of the others are  $(t_0, u_0, t_E)$ , i.e., the parameters of the underlying point lens event. The remaining two are the angle between the source-lens trajectory and planet–star axis,  $\alpha$ , and the ratio of the source radius to the Einstein radius,  $\rho = \theta_*/\theta_E$ .

Historically there have been two approaches to determining planet sensitivity. In the first approach (Rhie et al. 2000), one constructs an ensemble of planetary light curves that vary in  $\alpha$  but are fixed in the remaining six parameters. The values of  $(t_0, u_0, t_E)$  are adopted from the best fit of the single-lens event. We address the choice of  $\rho$  further below. The remaining two parameters are just those being tested ( $s$ ,  $q$ ). For each light curve, one creates fake data points at the times of each of the real measurements, with values equal to those predicted by the model and error bars equal to the those of the real data points. With the adopted parameters, the fit to a fake curve without a planet would be “perfect,” i.e.,  $\chi^2 = 0$ , so any  $\chi^2$  in excess of this value must be due to the planet. One then fits a point-lens model to these fake data. Therefore, if the  $\chi^2$  is above some threshold (perhaps  $\Delta\chi^2 > 200$  for events with moderate magnification,  $10 < A_{\max} < 100$ , Yee et al. 2012) then the planet is said to be detectable. The fraction of all the  $\alpha$  at fixed  $(s, q)$  for which the planet is detectable is then said to be the sensitivity  $S_i^{s,q}$ .

In the other approach (Gaudi & Sackett 2000), one fits the actual data with planetary models with the same sampling of parameters, and measures the  $\Delta\chi^2$  improvement between the planetary model and the best-fit point lens model. This method is more time-consuming but has the advantage of simultaneously searching for all planets that may be lurking in the data down to the adopted threshold (Gaudi et al. 2002).

The choice of  $\rho$  is a subtle one. For planetary events, one often measures  $\rho$  from the smearing out of the light curve as the source passes the sharp edge of a caustic. Hence, when constructing fake planetary light curves, one must insert some value of  $\rho$  even though this quantity is very rarely measured in point-lens events. The problem is that while  $\theta_*$  is usually well-determined from the color and magnitude of the source (e.g., Yoo et al. 2004),  $\theta_E$  is not known. We do not review the various methods used to estimate  $\rho$  in the past but simply note that for *Spitzer* events,  $\theta_E = \pi_{\text{rel}}/\pi_E$  is usually known quite well because  $\pi_E$  is measured and  $\pi_{\text{rel}}$  is well constrained (Calchi Novati et al. 2015), allowing a well-constrained estimate of  $\rho = \theta_*/\theta_E$ .

Current microlensing experiments have far too few detections to constrain the full two-dimensional distribution of planets as a function of  $(s, q)$ . However, for comparing to data, one can marginalize over one or both indices, e.g.,

$$\begin{aligned} S_i^q &= \int_{-\infty}^{\infty} d(\log s) S_i^{s,q}; \\ S_i^{<\text{jup}} &= \int_{-\infty}^{-3} d(\log q) \xi(q) S_i^q, \end{aligned} \quad (4)$$

where  $\xi(q)$  is a prior on the mass-ratio function of planets. While this prior could in principle be any function, for example, a specific author’s best current estimate this function, we strongly suggest that a single, simple (if only crudely

correct), function be uniformly adopted until such time as the true function is much better measured than at present. Specifically, we suggest

$$\xi(q) = \frac{q}{10^{-5}}. \quad (5)$$

In this way, sensitivity estimates between different events and between different authors can be directly compared, without having to “back out” information before the prior was applied (which is usually difficult or impossible). This is similar to the adoption by the extragalactic distance-scale community of 50 kpc for the distance to the Large Magellanic Cloud. Every worker in the field had his or her own favorite number, but all agreed to use this one as a benchmark of comparison. Note that we have not specified such a prior for the first formula in Equation (4), i.e., the integral over  $\log s$ . However, this is equivalent to adopting a “prior” of unity, i.e., Öpik’s Law, which is in fact a similarly useful benchmark prior.

There is one further issue that has not been previously considered in the literature but is quite relevant here. In the above-described procedures, it is implicitly assumed that the observations were carried out without reference to the presence or absence of a planet. This has usually been the case, and in the one notable example that it was not, the authors took the trouble to remove the extra observations that were triggered by the presence of the planet (Yee et al. 2012). See also Poleski et al. (2014b). However, for *Spitzer* events that are chosen subjectively, a large fraction of the nominal planet sensitivity may be due to observations before the public announcement.

However, as discussed in Section 2.6.1, there is the possibility that a planetary perturbation in the early stages could trigger additional observations or selection before the perturbation is well-understood or even recognized, e.g., if the perturbation was mistaken for a rise toward high-magnification. To determine which hypothetical planets should be excluded from the planet sensitivity, we suggest that the following additional test be conducted for each hypothetical planet that is regarded as “detected” based on the full light curve: truncate the fake-data light curve at the decision date and fit only the points known as of this date<sup>14</sup> with a point-lens model. We suggest that if  $\Delta\chi^2 < 10$ , then the signal from the hypothetical planet would be too small to trigger either follow-up or *Spitzer* observations. Hence, the hypothetical-planet detection should be accepted, but otherwise it should be rejected. This limit is chosen because in our experience such  $\Delta\chi^2 < 10$  deviations are extremely common and so cannot possibly trigger resource-expanding actions. However, substantially higher  $\Delta\chi^2$  might well trigger an unconscious *Spitzer* decision announcement.

## 5. PROBABILITY OF MEASURING PARALLAX

While parallax measurements derive from a combination of ground-based and space-based data, the limiting factor will be *Spitzer* data in almost all cases. The main reason is that the *Spitzer* observations of any given event are restricted to 38 days by Sun-angle constraints,<sup>15</sup> and these 38 days fall in an arbitrary part of the light curve. Second, both the cadence and

<sup>14</sup> A subtle but important point: only data released as of this date should be considered. Data that were taken but not released, do not need to enter into this calculation since they cannot have influenced the outcome.

<sup>15</sup> In practice, the spread in targets over a few degrees in the Bulge allows us to stretch the time-frame of the campaign to 40 days.

quality of the data are very likely to be lower than the ground-based data. Finally, if the parallax is large enough, the event as seen from *Spitzer* may pass entirely outside the Einstein radius during the *Spitzer* observations and so be effectively unmagnified. Hence, the probability that  $\pi_E$  will be measurable reduces in essence to probability that adequate *Spitzer* observations can be obtained. Hence, most of the discussion about “measuring parallax” is rooted in the specific nature and procedures for *Spitzer* observations. However, we begin by briefly discussing what it means to “measure parallax.”

### 5.1. Meaning of “Parallax Measurement”

Quantities are usually said to be “measured” if a numerical value can be assigned to them with some error bar and if this value is determined to be inconsistent with zero with some specified level of confidence, e.g.,  $3\sigma$ . For space-based microlensing parallaxes, there are circumstances in which upper and lower limits are sufficiently constraining, and therefore the definition of a “measurement” requires explicit discussion.

For example, suppose that an event is observed from the ground with  $(u_0, t_E)_{\oplus} = (0.2, 10 \text{ day})$  and with  $t_{0,\oplus}$  within the window of *Spitzer* observations, but the *Spitzer* light curve is completely flat. Also suppose that the source is bright enough that 10% variations in its flux would have been detected. From a measured lack of variation, together with the fact that *Spitzer* was  $\sim 1$  AU from Earth at the time of observations, one could conclude that  $\pi_E > 1.5$ , but no specific value of  $\pi_E$  could be assigned. Although not a “measurement” by traditional standards, this lower limit would be highly constraining. That is, it would imply that the velocity projected on the observer plane would be  $\tilde{v} \equiv \text{AU}/(\pi_E t_E) > 115 \text{ km s}^{-1}$ , implying that the lens was very likely in the near disk,  $\pi_{\text{rel}} > 0.1 \text{ mas}$ . Hence, if there were no planet discovered, the distance would be statistically well enough constrained to enter the cumulative distance distribution function. If the event proved to have a planet, then it is likely that  $\rho$  (and so  $\theta_E$  would be measured), which would permit a strict lower limit on  $\pi_{\text{rel}} = \theta_E \pi_E$  and a strict upper limit on  $M = \theta_E / \kappa \pi_E$ .

At the opposite extreme, if for the same  $(t_0, u_0, t_E)_{\oplus}$ , the *Spitzer* and Earth-based lightcurves appeared identical, this would be consistent with  $\pi_E = 0$ , which in traditional terms might be considered as “no measurement.” However, if this consistency were quite tight, say  $\pi_E < 0.01$ , then the projected velocity would be constrained  $\tilde{v} > 17,300 \text{ km s}^{-1}$ , implying that the lens certainly lies in the bulge.

Hence, the final sample must be defined as events that yield true measurements or *either* upper or lower limits on  $\pi_E$  (or both). The exact limits cannot be defined in advance because there is not yet enough experience with *Spitzer* parallax measurements to determine what are reasonable limits.

### 5.2. Spitzer Procedure

*Spitzer* observation sequences can only be uploaded to the spacecraft once per week, and hence targets can only be changed on this timescale, and furthermore, the entire week of observations must be planned in advance. In addition, it takes several days to prepare the observation sequence for upload to the spacecraft even after the targets and observation sequence have been set. The net result is that the targets and sequence are

set 3–10 days before the observations are actually carried out. See Figure 1 of Udalski et al. (2015b).

In light of these considerations, and to facilitate the discussion, we define several variables summarized in Table 2. First, we define  $t_{j,\text{dec}}$  to be 6 hr prior to the time that observing choices must be forwarded to *Spitzer* operations for a given observing “week”  $j$ , i.e.,  $t_{j,\text{dec}} = \text{Monday UT } 15 - 6 \text{ hr}$ . Experience shows that this is the latest time that new information can be reliably incorporated into the observing request without risking the introduction of serious errors. Given the day-of-the-week constraints, the  $\sim 40$  day campaign, and a start date of June 3rd,  $j$  takes on values from 1 to 7.

The time of the first possible *Spitzer* observations of a given event is defined as  $t_{\text{first}}$ . For simplicity, we will let  $t_{\text{first}}$  be when those coordinates could first be observed by *Spitzer*, even if the event is not discovered until afterward this date. Finally, we define  $t_{j,\text{next}}$  as the time of the first possible observation that can be requested at  $t_{j,\text{dec}}$ , and we define  $t_{\text{fin}}$  as the final possible observation of an event before the *Spitzer* observing season ends (because of Sun-angle restriction and/or the end of the allocated observations). Note that  $t_{j,\text{dec}}$  and  $t_{j,\text{next}}$  change each week, while  $t_{\text{first}}$  and  $t_{\text{fin}}$  are defined for each particular event (set by the 38-day Sun-angle constraint). In general, for events selected for the first week of observations,  $t_{1,\text{next}} = t_{\text{first}}$ .

### 5.3. Spitzer Observation Cycles

Generally speaking, we expect most or all of the available *Spitzer* time to be devoted to this program during the  $\sim 40$  day observing window. Because there are almost no optical microlensing events with  $|b| < 1^\circ$ , the Bulge can be divided into north and south components. Hence, given continuous observing time, *Spitzer* observations in a given week can be carried out most efficiently if the targets are organized in concatenated “cycles” moving west-to-east through the southern bulge and then east-to-west through the northern bulge. Each event can then be given a priority  $n$ , which designates that it will be observed each  $1/n$  cycles. That is, if  $n = 1$  it will be observed every cycle through the Bulge, and if  $n = 8$  it will be observed every eighth cycle. We expect approximately eight cycles per day, each lasting  $\sim 2.4$  hr, with the exact number determined by the total observation time allotted and the total number of targets per cycle. This is discussed in more practical detail in Section 8.

### 5.4. Spitzer’s Role in Parallax Measurements

As originally conceived, the standard way to measure satellite parallax was to observe the peak of the light curve from the satellite and the full light curve from the ground. This requires only partial, but very specific, light curve coverage from space. Therefore, one of the goals of *Spitzer* observations is to try to capture this peak in as many cases as possible. Although this measurement is nominally still subject to the four-fold degeneracy in  $\pi_E$  (Refsdal 1966; Gould 1994), Gould (1995) showed that these degeneracies could be partially or fully broken by measuring the very small difference in  $t_E$  as seen from the two vantage points, and this idea was then investigated in extensive simulations (Boutreux & Gould 1996; Gaudi & Gould 1997). Hence, this goal of observing the peak of the event guided the 2014 *Spitzer* campaign.

For many years it was believed that because of the four-fold degeneracy, parallax measurements would not be possible if

the satellite observed only the rising or falling side of the event, but did not capture the peak. Coverage of the peak would in fact be required if one needed to derive independent point-lens parameters ( $t_0$ ,  $u_0$ ,  $t_E$ ) from ground-based and space-based observations (as is necessary in Gould 1995 to break the degeneracy). However, Calchi Novati et al. (2015) showed that the four-fold degeneracy can usually be broken by a combination of the so-called ‘‘Rich argument’’ and kinematic priors derived from a Galactic model. Once the problem of the four-fold degeneracy is removed, the requirements on the satellite data are drastically reduced.

First,  $t_{E,\text{sat}}$  can be regarded as ‘‘essentially known,’’ so that it is only necessary to determine two satellite parameters ( $t_{0,\text{sat}}$ ,  $u_{0,\text{sat}}$ ) to measure the parallax. Of course,  $t_E$  is actually slightly different as seen from Earth and the satellite because they have a relative motion of  $\sim 30 \text{ km s}^{-1}$  in the East direction. However, the resulting difference in  $t_E$  is directly determined by  $\pi_E$ , so while it is not strictly the case that  $t_E$  is irrelevant, it remains true that only two independent light curve parameters must be derived from the satellite light curve.

Second, the source flux parameter  $f_{s,\text{sat}}$  for the satellite can be determined independently of the satellite light curve using a color–color relation derived from field stars combined with the measured color ( $I - H$  or  $V - I$ ) derived from the ground-based light curve. Calchi Novati et al. (2015) obtained typical precisions for  $f_{s,\text{sat}}$  of 5% in the cases for which they had good  $H$  or  $V$  data. However, it remains necessary to determine  $f_{b,\text{sat}}$  from the light curve, which constitutes a third parameter that must be derived from the *Spitzer* data.

Then, from simple parameter counting, it is in principle enough to measure three non-colinear points on the light curve to measure the parallax (Dong et al. 2007), e.g., one point that is ‘‘known’’ to be at baseline and two others at different magnifications. In practice, more points are usually needed to have confidence in the measurement and to have checks against discrete degeneracies. However, it would be enough, for example, to track the falling part of the light curve from the time that the source exited the Einstein ring until it had dropped by 30% in magnification, i.e., approached baseline.

For events that are well before peak as the *Spitzer* window ends, the situation is less straightforward because there would probably not be any baseline and the short duration of the observations might not yield any measurable change of slope. However, such events could be recovered by post-event baseline observations, either six months later (when the Bulge is not visible from Earth, so ordinary satellite parallax observations are not feasible) or the following year. Hence,  $t_{\text{fin}}$  could be considered as a date in the distant future rather than the end of the current 38-day observing window.

Therefore, there are two different channels through which parallaxes can be measured with *Spitzer* for point lens events. First, *Spitzer* can observe just the peak of the light curve. Second, *Spitzer* can observe either the rising or the falling side of the event plus some measurement of the baseline. Finally, if the event has a binary lens, features from the binary may be used to measure the parallax. However, this situation is more complex since binary perturbations last long enough that they may not be fully captured by the *Spitzer* data and so the four-fold degeneracy may persist (Zhu et al. 2015).

### 5.5. *Spitzer* Photometric Pipeline Issues

The feasibility of measuring  $\pi_E$  from a given set of *Spitzer* observations obviously depends on the quality of the photometry that can be extracted from these observations. Remarkably, none of the wide range of publicly available *Spitzer*-specific photometry packages is well matched to the problem of time series of variable stars in crowded fields. As a result, the limits of what can be achieved from such photometry are not well understood.

All the main elements required to solve this problem are at hand, but they have not so far been combined. First, the *Spitzer* pixel response function (PRF) is extremely well understood. That is, if a point source has a known flux and known position relative to the optical axis, then the response of all pixels can be predicted to much higher precision than is relevant for the relatively faint sources that are studied in microlensing experiments. The positions and fluxes of the great majority of sources in the microlensing fields are known to be constant on the timescales of the 38-day *Spitzer* observing window. Moreover, the locations of all field sources that are bright enough to be relevant are known from ground-based optical astrometry at much higher precision than is needed for *Spitzer* photometry, while the location of the microlensed source is typically known with even higher precision in this optically based frame. Even the approximate  $3.6 \mu\text{m}$  fluxes (other than the microlensed source) are known from optical  $V/I$  photometry and fairly robust local-field  $V/I/3.6 \mu\text{m}$  color–color diagrams. Hence, a conceptually straightforward procedure would be to forward model the ensemble of  $n$  images with one flux parameter for each non-microlensed source and  $n$  flux parameters for the microlensed source. Intrinsically variable stars could be recognized as poor fits in this process and either ignored (if they were sufficiently far from the lensed source) or modeled with  $n$  parameters instead of just one.

We are working on such a pipeline, but since criteria for 2015 observations (beginning in June) are required several months in advance, we must assess likely *Spitzer* performance based on applying existing pipelines to 2014 *Spitzer* microlens data. These each contain some (but not all) of the advantages of the ideal pipeline outlined above. For example, the MOPEX pipeline fully incorporates the PRF but does not hold stellar positions constant, nor does it hold the flux of field stars constant. The well-known DoPhot (Schechter et al. 1993) pipeline can be applied to images formed by combining the six 30 s dithered images at each epoch. It can hold stellar positions constant but does not incorporate any information about the PRF. We also applied a variant of the ISIS pipeline, which uses image subtraction to the same combined images. Although this pipeline normally outperforms DoPhot for ground-based microlensing data (with some exceptions), we find that the lack of PRF information generally affects ISIS more adversely than DoPhot.

We conduct a purely empirical investigation, using 47 events from the 2014 *Spitzer* microlensing ‘‘pilot program’’ that have enough points to potentially construct a coherent light curve. We consider the photometry from MOPEX, DoPhot, and a preliminary version of our own pipeline. We create an *optically based* effective *Spitzer* (‘‘L-band’’) magnitude (since prior to obtaining *Spitzer* data we have no independent knowledge of the true *Spitzer* flux). This is defined

$$L_{\text{eff}} \equiv I - 0.93A_I - 1.3 + 0.5\Theta(I - A_I - 17.2), \quad (6)$$

where  $A_I$  is the extinction in  $I$ -band (Nataf et al. 2013) and  $\Theta$  is the Heaviside step function. We stress that no precise physical meaning should be attached to  $L_{\text{eff}}$ . It is simply an approximate predictor of the *Spitzer* flux based on optical data. The  $\Theta$  function divides all stars into two types: turnoff stars ( $\Theta = 1$ ) and low-luminosity giants ( $\Theta = 0$ ). The justification for this approximation is that significantly fainter (and redder) dwarfs generally will not enter our sample and significantly brighter (and redder) giants are very rare. Of course, by limiting ourselves to two classes of stars we are still ignoring evolution over the sub-giant branch. However, in the general case, it is not possible to make a finer distinction, particularly before a detailed investigation of an individual event has been made, as is almost always the case when one must make the decision about whether to monitor a particular event.

Using this proxy, we find that it is usually not possible to obtain good photometry with existing software unless there is at least one point with  $L_{\text{eff}} < 15.5$ . We therefore use this criterion as our principal guideline for deciding whether parallaxes can be measured for particular events. This may appear too conservative in that there will almost certainly be photometry improvements by the time that the data are analyzed. On the other hand, when making decisions about *Spitzer* observations, one must use the simplified “assumption” that *Spitzer* will see the same brightness source star as it would if it were observing from Earth because the true magnification as seen from *Spitzer* is unknown. That is, the whole point of the experiment is that the *Spitzer* and Earth-based light curves will differ by an intrinsically unpredictable amount. In particular, the source could be less magnified as seen from *Spitzer* than from the ground. Thus, we adopt  $L_{\text{eff}} < 15.5$  as a good balance between these two considerations.

An additional consideration, given that the IRAC pixels are  $1''.2$ , is that the target may be blended with other stars in the crowded field, which can affect the quality of the photometry. The severity of this blending depends both on the separation of the blend from the target and on their relative fluxes at  $3.6 \mu\text{m}$ . While the separation of potential blends can generally be determined from existing, higher-resolution, ground-based data, the relative fluxes cannot. Furthermore, in part because of the problems with the photometric pipelines, we have not been able to determine exactly what criteria can be used to assess whether or not a given blend star will cause a problem for the photometry. Hence, while we are aware of this issue, it is not possible to account for it at the present time.

## 6. OBJECTIVE CRITERIA

As discussed in Section 2.1, there is an extremely strong reason for choosing as many events as possible based on purely objective criteria: all planets (and planet sensitivity) from the entire event can be included in the sample. However, there is also a huge potential for wasted *Spitzer* observations if these criteria are not sufficiently restrictive. Hence, we have opted for a conservative approach.

An important point to keep in mind is that for events in low-cadence ground-based survey fields, there is no major advantage to selecting the event objectively because such events have very little sensitivity to planets in the absence of follow-up data (Section 4). Their sensitivity will only be substantial if higher-cadence (usually ground-based) observations are triggered. If this recognition also triggers *Spitzer*

observations (or rather, commitment to such observations) at the same time, then essentially no planet sensitivity is lost. Column 1 of Figure 1 demonstrates that these low-cadence events are also extremely hard to predict.

Another point to keep in mind is that it is substantially easier to predict the future course of events that have already peaked than those that are still rising (compare panels (1e) and (2e) to earlier panels in Figure 1), and hence to estimate accurately whether a successful parallax measurement can be made. This fact is especially important for events that have peaked before the *Spitzer* campaign has begun. For events that peak during the campaign, the probability of measuring  $\pi_E$  can be substantially enhanced if *Spitzer* observations are made over peak, i.e., before such secure information about the event parameters is available.

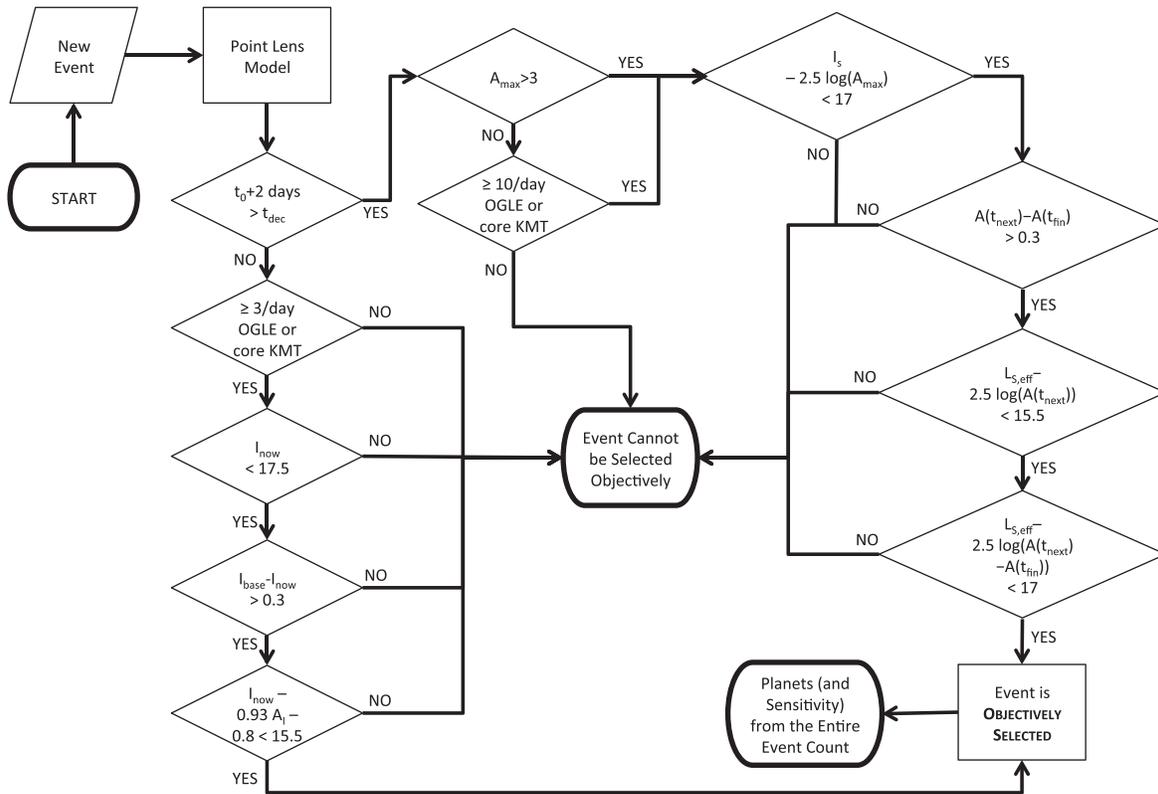
Guided by these considerations (and others related to subjective selection that are discussed below), J.C.Y. and A.G. developed some preliminary objective criteria, and then (independent of these criteria) each individually analyzed 242 events based on OGLE and MOA data obtained up through 2013 June 3, which is the analogous time to the first decision time in 2015. These events had been pre-selected based on very loose criteria from about 1000 events that had been found by these collaborations by this date. For each event, J.C.Y. and A.G. decided whether it should be chosen for hypothetical *Spitzer* observations to begin three days later, and if so at what cadence. All disagreements were discussed and the final joint decisions were subsequently evaluated based on comparison to the full 2013 light curves. The agreed-upon sample contained all nine events that were selected by the preliminary objective criteria, and also many that were not for a total of 44 events. Based on this detailed analysis J.C.Y. and A.G. refined the objective criteria for selection and the objective cadence choice, both of which are listed immediately below, and also developed general guidelines to subjectively choose events, which are discussed in Section 7.

Figure 2 summarizes the process for objectively selecting events. In this scheme, events may be divided into two categories: events that have already peaked ( $t = t_0 + 2$  days) and events “before” the peak ( $t < t_0 + 2$  days). In order to be objectively selected, the event must meet all criteria for the appropriate category. We begin by discussing events that have peaked, because they are generally better understood (i.e., the model fits have converged and their future behavior is well-constrained).

### 6.1. Events that have Already Peaked

- (A1)  $t_0 + 2 \text{ days} < t_{j,\text{dec}}$ .
- (A2)  $A_{\text{max}} > 3$  or in OGLE field w/ cadence  $\geq 10 \text{ day}^{-1}$  or in a “core” KMTNet field (approximately  $13 \text{ deg}^2$  and  $16 \text{ deg}^2$ , respectively, with substantial overlap).
- (A3)  $I_S - 2.5 \log(A_{\text{max}}) < 17$ .
- (A4)  $A(t_{j,\text{next}}) - A(t_{\text{fin}}) > 0.3$ .
- (A5)  $L_{S,\text{eff}} - 2.5 \log[A(t_{j,\text{next}})] < 15.5$ .
- (A6)  $L_{S,\text{eff}} - 2.5 \log[A(t_{j,\text{next}}) - A(t_{\text{fin}})] < 17$ .

Criterion (A1) is simply a practical definition of “post-peak.” Criterion (A2) selects for events that have significant planet sensitivity. For events in high-cadence fields this is essentially any event (provided it meets the other criteria) because planets can be discovered in these events far out into the wings and even at baseline. For other, low-cadence events, which



**Figure 2.** Flowchart illustrating the process for objectively selected events. An event may be objectively selected either before or after the peak, but must meet all of the selection criteria for that category.

generally need follow-up observations to gain substantial planet sensitivity, the criterion demands  $A_{\max} > 3$  as a minimum indicator that such follow-up observations will be regarded as worthwhile. However, triggering such follow-up observations will not be automatic. Hence, it will be important either to recognize in advance that the event will be objectively selected, or to subjectively select it prior to peak, with the understanding that its objective selection may subsequently take precedence (Section 2.3). Otherwise, since this mode of objective selection is triggered when the event is past peak, much of the planet sensitivity will be lost.

Criterion (A3) demands that the peak flux from the magnified source (not including any blended light) be  $I < 17$ . It increases the likelihood that the ground-based photometric precision will enable good planet sensitivity. To date, the overwhelming majority of planetary microlensing events have peaked  $I < 17$ . In addition, this criterion is important not only to secure an accurate fit to the light curve but to permit application of color-color relations to determine the *Spitzer* source flux, both of which affect the final parallax measurement.

Criterion (A4) is driven by the fact that parallax measurements require a well-measured change in magnification as seen from *Spitzer*. In practice, this means both measuring a flux change and independently determining the *Spitzer* source flux. For observations that begin past peak, it is impossible to reliably fit the *Spitzer* light curve for the source flux, so it must be determined from color-color relations, which can be done reliably to about 5%. Hence, we require at least a 0.3 change in magnification (i.e., 6-fold larger than 5%) based on our estimate of what will be required for reliable parallax measurements (Section 5.4). This criterion is equivalent to

demanding that the source is still in the Einstein ring (at the time of the next possible observation) for the case that the final possible observation is well outside the Einstein ring.

Criterion (A5) derives directly from the difficulty of extracting *Spitzer* photometry unless at least one point is brighter than  $L_{\text{eff}} < 15.5$  (Section 5.5). Criterion (A6) demands a minimal flux change in *Spitzer* flux. It will be automatically satisfied for the great majority of stars that satisfy criterion (A5) and is included to guard against including (at least automatically) events that are not predicted to change much over the remainder of the observations. As noted in the justification for criterion (A3), measurable flux changes are crucial for parallax measurement. Criterion (A6) ensures that not only the magnification changes, but the flux itself changes by a significant amount.

## 6.2. Rising Events

- (B1)  $t_0 > t_{\text{dec}} - 2$  days.
- (B2) in OGLE field w/ cadence  $\geq 3 \text{ day}^{-1}$  or in a “core” KMTNet field.
- (B3)  $I_{\text{now}} < 17.5$ .
- (B4)  $I_{\text{base}} - I_{\text{now}} > 0.3$ .
- (B5)  $L_{\text{eff,dwarf,now}} = I_{\text{now}} - 0.93A_I - 0.8 < 15.5$ .

Criterion (B1) is just the practical definition of a rising event<sup>16</sup> (i.e., the complement of criterion (A1)).

Criterion (B2) restricts this entire class of objective selection to fields with moderate-to-high cadence,  $\geq 3 \text{ day}^{-1}$ , to ensure

<sup>16</sup> Technically, this also includes events  $< 2$  days past peak. The reason for this choice is that it can be difficult to be confident that an event has indeed peaked unless there are data after the peak to demonstrate this explicitly.

that the selected events have good sensitivity to planets over the entire light curve. The primary concern is that once an event has met the full set of rather restrictive criteria, it will be close to the peak and a large fraction of the event will be over. As discussed in Section 4, without substantial follow-up data, low-cadence events have very little practical planet sensitivity. Hence, it is not worthwhile to observe them with *Spitzer* to try to measure a parallax because they will add very little to the final analysis of the planet occurrence rate as a function of Galactic distance. We note that most high-quality events that fail this criterion can be selected subjectively (Section 7) well before peak with the specific goal of triggering additional follow-up observations to improve the planet sensitivity.

The remaining three criteria make no reference to a light curve model and instead rely on purely empirical observables. Again, this is because experience shows that such models are not reliable for pre-peak light curves. All three make reference to “ $I_{\text{now}}$ ” which is the last measured OGLE point. In practice this means that rising events must be identified by OGLE although (B2) allows for those events to be in low-cadence fields if they are observed by KMTNet.

Criterion (B3) assures that the event will be bright enough for accurate measurements (necessary for both planet sensitivity and parallax). Criterion (B4) assures that even if the source is not blended, the event has risen at least 32% above baseline, i.e., the source must be (nearly) inside the Einstein ring. The combination of (B3) and (B4) ensure that even if the source turns out to be heavily blended, then at least the magnified flux will change significantly compared to its present value relative to a future baseline measurement, which will enable a measurement of the parallax. Finally, criterion (B5) attempts to assure that there will be at least one point above the photometric threshold for measuring a *Spitzer* light curve (Section 5.5). This is only “attempted” (rather than guaranteed) since the event may not be as magnified from *Spitzer*’s vantage as from Earth. Because the source magnitude is most likely not known at the time of this algorithmic selection, we conservatively assume it is a dwarf (i.e., relatively blue and so fainter as seen by *Spitzer* for fixed *I*-band brightness).

### 6.3. Objectively Determined Cadences

As we have discussed, events that are selected objectively must have objectively determined cadences. In practice, cadences are actually defined by “priorities,” where priority  $n$  means that the event is observed during  $1/n$  of the cycles through the microlensing fields (Section 5.3). However, we state these here in terms of cadences, since there is a clear-cut conversion from one to the other once the target sample is selected. We designate the following algorithm for setting the observation cadence for Week  $j$ .

- (C1) Default cadence:  $1 \text{ day}^{-1}$ .
- (C2)  $2 \text{ day}^{-1}$  provided that all of the following are true:
  - (a) this is first *Spitzer* observation period of the event, and
  - (b)  $t_{j,\text{next}} > t_0$ , and
  - (c)  $A(t_{j,\text{next}}) < 1.35$ .
- (C3)  $2 \text{ day}^{-1}$  for observations beginning when less than two full weeks remain.
- (C4) Stop observing the event, provided that all of the following are true:

- (a) at least 2 weeks of objectively determined observations are complete, and
- (b)  $t_{j,\text{next}} > t_0 + t_E$
- (c)  $t_{j+1,\text{next}} > t_0 + 2t_E$ .
- (C5) Stop observing the event, if either
  - (a) the parallax of an event has been measured from *Spitzer* data already collected, or
  - (b) the *Spitzer* light curve has already reached baseline (so no more parallax information could be extracted from additional observations).

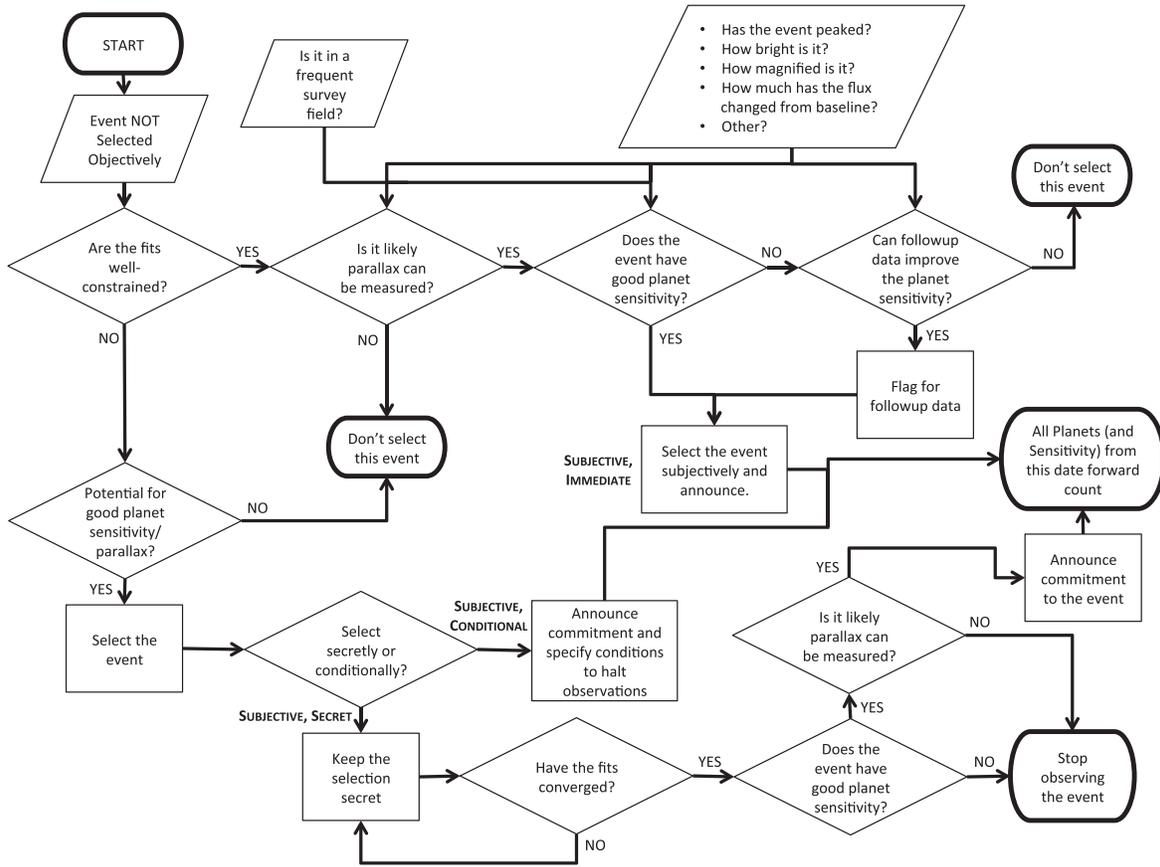
Criterion (C1) has been shown by Calchi Novati et al. (2015) to be generally adequate to make parallax measurements. However, for events that are leaving the Einstein ring as the *Spitzer* observations begin (C2) or for which there is only a short rising observational sequence at the end of the *Spitzer* window (C3), the cadence is doubled. These events have significantly more restricted light-curve coverage than the typical events analyzed by Calchi Novati et al. (2015) and therefore require higher cadence to obtain more points (so higher signal-to-noise ratio) while the event is still significantly magnified. (C4) imposes a reasonably conservative criterion for halting observations. In principle, this may cause *Spitzer* to miss a key portion of the light curve because it can in principle peak either earlier or later as seen from *Spitzer* than the Earth-based light curve would predict. However, because bulge lenses have small microlens parallaxes, the light curve peaks as seen from Earth and *Spitzer* are very close in time. On the other hand, for disk lenses, the *Spitzer* peak is usually earlier (or not much later) than from the ground because these disk lenses tend to move in the direction of Galactic rotation, i.e., about  $30^\circ$  east of north, whereas *Spitzer* is roughly due West of Earth. See Figure 2 of Calchi Novati et al. (2015). Finally, (C5) provides more specific conditions for halting observations if the *Spitzer* data can be reduced and analyzed in real-time.

### 6.4. Binary Star Events

For completeness, we also specify the objective selection criteria for binary events. Unlike planetary events, binary events show prominent anomalies that modify the single-lens light curve significantly. Therefore, most binary events can be recognized in advance, and the inability to model them with single-lens light curves makes all selection criteria based on single-lens modeling (i.e., Sections 6.1 and 6.2) fail in most cases. As a consequence, in the 2014 pilot program we subjectively selected binary events, such as OGLE-2014-BLG-1050 (Zhu et al. 2015), for observations because their nature as binary events had been confirmed.

Therefore, in order to enable statistical studies of stellar binaries one has to have objective selection criteria. After reviewing those binary events from the 2014 season, we decide to use the following criteria and cadence.

1. Begin *Spitzer* observations if
  - (a) the ground-based light curve is in a U-shaped trough, and
  - (b)  $L_{\text{eff, trough}} < 16$ .
2. End *Spitzer* observations either:
  - (a) one full week after the *Spitzer* light curve exits the caustic, or
  - (b) both:



**Figure 3.** Flowchart illustrating the process for subjectively selecting an event. By definition, this is less quantitative than the process for objectively selecting an event (see Figure 2), but the underlying considerations are the same in both cases, namely: “Does the event have good planet sensitivity?” and “Is this event likely to yield a *Spitzer* parallax?.”

- (i) one full week has passed since the ground-based light curve exits the caustic, and
- (ii) the *Spitzer* light curve is shown never to have entered the caustic.

3. Default cadence of 1 day<sup>-1</sup>.

### 7. GUIDELINES FOR SUBJECTIVELY CHOSEN EVENTS

We have already discussed in Section 2.2 the various reasons one might want to subjectively select an event and how that might affect the type of subjective selection (Table 3). Figure 3 summarizes the decision process that can lead to subjective selection of an event.

The most straightforward type of subjective selection is “Subjective, immediate” in which an event is immediately selected for observations, committed to, and has its cadence specified. Such a decision may be made at any time, including before the start of *Spitzer* observations. There are two primary types of events that might be selected this way. First, events that are discovered and well understood before  $t_{j,dec}$  but might not meet the objective criteria. Committing to observations immediately allows more of the planet sensitivity of that event to be captured. The second type of event is a well-understood, low-cadence event that requires follow-up observations in order to capture significant planet sensitivity.

However, we expect that the great majority of events will be selected under the categories “Subjective, conditional” or “Subjective, secret” because at time  $t_{j,dec}$  there is no way to

distinguish between events that will reach moderate or higher magnification ( $A_{max} > 3$ ) sometime before the start of the next cycle of *Spitzer* observations,  $t_{j+1,next}$  (i.e.,  $t_{j,dec} + 10$  day), and events that will turn over at low magnification and low flux levels (e.g., compare the two fits in panel (1b) of Figure 1). Moreover, even among those that are likely to achieve satisfactory magnification, it cannot be decided automatically whether suitable follow-up resources can be allocated to a specific event, given a wide variety of operational constraints. And finally, it may be impossible to determine which of these events will get to high or very high magnification based on routine survey data. However, additional investigation, including additional follow-up data and/or color information from survey and/or follow-up observations may resolve this question.

As we have already discussed above, subjective decisions may be made before or after  $t_{first}$ , but they must specify the cadence (or cadence algorithm) at the time of commitment. To be robust, any algorithm must be based on readily available data, such as the *I*-band light curve and the *I*-band field extinction.

Given the above factors, together with the fact that planet sensitivity is heavily skewed toward higher-magnification events, it is inevitable that *Spitzer* observations will be triggered for a large number of events of uncertain prospects, and therefore that the majority of these must be terminated promptly after the event fails to rise to the level that permits significant planet sensitivity and also permits its parallax to be

measured. Otherwise, a substantial fraction of observing time will be wasted on useless events.

Hence, the principal issue will be deciding how to frame this failure in either of the two cases (“Subjective, conditional” or “Subjective, secret”). This will have to be done on a case-by-case basis at the time the events are subjectively chosen, since the uncertain nature of such events makes it impossible to develop strict guidelines in advance. In the first case, this framework must provide a guide to terminating observations (e.g., “stop observations next week if event does not reach  $I < 16.2$  by the next decision point,  $t_{j+1,dec}$ ”). In the second, it must provide a guide as to whether or not the team should commit to observations of the event.

To the extent possible, the criteria governing such terminations (or continued observations) should be framed in terms of post-peak observables because these are more robust. They should then be formulated as proxies for the criteria (A1)–(A6), but with  $t_{j,next} \rightarrow t_0$ . However, this implicitly assumes that  $t_{j,dec} < t_0 < t_{j+1,dec}$  (essentially criterion (A1)). It will in general be necessary to specify what should be done if the event does not ultimately satisfy that criterion. For example, it could be stated that in this case, the observations continue at the same cadence and the decision is made during the next week.

## 8. OBJECTIVE ALLOCATION OF REMAINING OBSERVING TIME

### 8.1. General Considerations

An important aim of the *Spitzer* observations, as outlined in the proposal (Gould et al. 2014) is to detect and characterize planets from *Spitzer* observations themselves. Because planet sensitivity scales with magnification, this can best be done by monitoring higher-magnification events more intensively from *Spitzer*, in particular events that are at higher magnification as seen by *Spitzer*. Hence, after the fundamental goal is met by allocating enough observation time to reliably measure parallaxes (as outlined in Sections 6 and 7) and also to obtaining parallaxes of microlens binaries (which was also part of the proposal but is not the subject of the present study), the remaining time should be allocated to this purpose.

Because these additional planet-finding observations involve allocation of additional time to the *same* set of events that are the object of microlens parallax measurements, it is important to isolate the decisions about this allocation from the prospect of improving parallax measurements. Otherwise, events with known planets could receive additional measurements aimed at measuring parallax, making them more likely to have good parallax measurements relative to those that failed to show any planets. And since only events with good parallax measurements can be included in the sample to measure the Galactic distribution of planets, we must eliminate the potential for bias.

At the same time, it is difficult to develop completely objective criteria for these allocations because of the wide range of the possible quantities of available time and wide range of the possible quality of events to which they might be applied. In addition, the amount of additional time will depend on the precise number of events being monitored. Finally, the targets for a given week will have a wide range of magnifications, whose distribution cannot be predicted in advance.

### 8.2. Practical Execution

We propose the following algorithm to effectively separate these decisions while leaving adequate freedom to respond to potential planet sensitivity by obtaining additional *Spitzer* observations.

Recall from Section 5.3 that the most efficient way to observe these events is to cycle through the Bulge west-to-east and east-to-west. The priority of an event sets how often an event will be observed. Specifically, priority  $n = 1$  is observed every cycle whereas priority  $n = 8$  is observed every eighth cycle. The exact number of cycles depends on the the number of events, their priorities, and the amount of time available.

All events that have been chosen for *Spitzer* observations, *except newly selected subjective events* are rank ordered according to the  $2\sigma$  lower limit of their highest magnification ( $A_{j,max}$ ) in the observing window  $[t_{j,next}, t_{j+1,next}]$ . These are then assigned priorities that map monotonically to this ranking. The break points in this mapping are decided manually. For example,  $A_{j,max} > 20 \Rightarrow n = 1$ ,  $10 < A_{j,max} \leq 20 \Rightarrow n = 2$ ,  $5 < A_{j,max} \leq 10 \Rightarrow n = 3$ ,  $3 < A_{j,max} \leq 5 \Rightarrow n = 4$ . The only other rule is that the priorities and the number of events in each category result in a total number of observations that equals the number of observations available. Hence, there are likely many choices of break points and priorities that fulfill these two criteria (monotonic mapping and total observations) for any given rank-ordered set of events and amount of observing time. The final choice, which will set the number of cycles and their duration, is at the discretion of the *Spitzer* team (i.e., the authors of this paper).

While the manual decision on break points might seem to allow skewing (conscious or unconscious) of the strength of different parallax measurements, in fact this is virtually impossible. The events whose parallax is poorly measured with existing data, and therefore might require “saving” with additional observations, will be at low magnification and hence cannot be helped by any manual decision that is constrained by the monotonic mapping described above.

Here, the default meaning of “highest magnification” is highest as predicted for observations from the ground. However, it may be possible to download and process *Spitzer* data sufficiently quickly to make predictions about the course of the event as seen from *Spitzer*. In this case, the *Spitzer* magnifications should take precedence.

Why do we exclude the subjectively chosen events whose observations are just starting? There are two reasons. First, the future course of these events is often very poorly understood, so the  $2\sigma$  lower limit on the magnification is likely to be very low and hence unlikely to trigger the additional observations being considered here. More fundamentally, the cadence of these first-week observations is subjectively decided, so no objective procedure is required to allocate additional observations to planet hunting in these events if the team decides that is necessary.

The cadence (or algorithm for determining the cadence) of *Spitzer* observations must be specified at the time of the commitment to observe a subjectively selected event. However, in practice these cadences will almost always be set at  $t_{sel}$ , since even for “subjective, secret” events  $t_{com}$  will generally happen before the next decision point when the cadence could be changed ( $t_{j+1,dec}$ ). In subsequent weeks, any change in the cadence of these events must be through the process described above.

## 9. CONCLUSIONS

We have outlined an approach for maximizing the planet sensitivity of space-based microlens parallax surveys, for the particular case that the satellite targets are chosen based on ground-based identification of events. This applies to current observations by *Spitzer* and any narrow field of view, targeted observations from space.<sup>17</sup>

The basic principles are as follow.

First, objective criteria are quite easy to establish for events that have already peaked because their fits are well constrained. Because the criteria are objective, the entire time span of those events, with respect to both planets and planet sensitivity, can be included in the analysis. This includes, in particular, all of the time before peak, which lies before the onset of *Spitzer* observations and also before the event was even recognized as microlensing.

Second, it is also possible to establish objective criteria for a subset of pre-peak events. However, because these events are pre-peak, their model fits may not be reliable, so it is necessary to define these criteria in terms of observables.

Third, for objectively chosen events, cadences must also be determined by objective criteria.

Fourth, for those events in low-cadence survey fields, it is less important to define objective criteria, because the events have low sensitivity to planets unless additional follow-up observations are obtained.

Fifth, the remaining events, including those in low-cadence fields, can be chosen subjectively, but the full cadence (or prescription for determining the cadence) must be specified at the time that they are selected. Subjective selection can take several forms, but the most important aspect is when a commitment is made (and announced) to observe the event. It is this date that determines what planet sensitivity and planet detections are included in the analysis.

Sixth, in the case of events that were previously chosen on subjective grounds and that subsequently meet the objective criteria, their objective status must take precedence in evaluating the event as part of the sample. This assumes that a parallax measurement proved possible based only on the (more restricted subset of) objectively required *Spitzer* observations. If not, and if the full set of *Spitzer* observations yields a parallax measurement, then they revert to subjective status.

This paper constitutes a public announcement of our objective criteria and procedures. If there are any updates to

these, they will be posted on arXiv as a revision to or update of this paper.

Work by J.C.Y., A.G., and S.C. was supported by JPL grant 1500811. Work by J.C.Y. was performed under contract with the California Institute of Technology (Caltech)/Jet Propulsion Laboratory (JPL) funded by NASA through the Sagan Fellowship Program executed by the NASA Exoplanet Science Institute. D.M.N. was supported by the Australian Research Council grant FL110100012.

## REFERENCES

- Alcock, C., Allsman, R. A., Alves, D. R., et al. 2001, *Natur*, 414, 617  
 An, J. H., & Gould, A. 2001, *ApJL*, 563, L111  
 An, J. H., & Gould, A. 2014, *ApJ*, 780, 54  
 Batista, V., Beaulieu, J.-P., Bennett, D. P., et al. 2015, *ApJ*, 808, 170  
 Batista, V., Beaulieu, J.-P., Gould, A., et al. 2014, *ApJ*, 780, 54  
 Bennett, D. P., Bhattacharya, A., Anderson, J., et al. 2015, *ApJ*, 808, 169  
 Boutuereux, T., & Gould, A. 1996, *ApJ*, 462, 705  
 Calchi Novati, S., Gould, A., Udlaski, A., et al. 2015, *ApJ*, 804, 20  
 Cassan, A., Kubas, D., Beaulieu, J.-P., et al. 2012, *Natur*, 481, 167  
 Dong, S., Udalski, A., Gould, A., et al. 2007, *ApJ*, 664, 862  
 Gaudi, B. S. 2012, *ARA&A*, 50, 411  
 Gaudi, B. S., Albrow, M. D., & An, J. 2002, *ApJ*, 566, 463  
 Gaudi, B. S., & Gould, A. 1997, *ApJ*, 477, 152  
 Gaudi, B. S., & Sackett, P. 2000, *ApJ*, 528, 56  
 Gould, A. 1992, *ApJ*, 392, 442  
 Gould, A. 1994, *ApJL*, 421, L71  
 Gould, A. 1995, *ApJL*, 441, L21  
 Gould, A. 2014, *JKAS*, 47, 215  
 Gould, A., Carey, S., & Yee, J. C. 2014, *Spitzer Proposal ID 11006*  
 Gould, A., Dong, S., Gaudi, B. S., et al. 2010, *ApJ*, 720, 1073  
 Gould, A., & Horne, K. 2013, *ApJ*, 779, 28  
 Gould, A., & Loeb, A. 1992, *ApJ*, 396, 104  
 Griest, K., & Safizadeh, N. 1998, *ApJ*, 500, 37  
 Henderson, C. 2015, *ApJ*, 800, 58  
 Muraki, Y., Han, C., Bennett, D. P., et al. 2011, *ApJ*, 741, 22  
 Nataf, D. M., Gould, A., Fouqué, P., et al. 2013, *ApJ*, 769, 88  
 Poleski, R., Skowron, J., Udalski, A., et al. 2014a, *ApJ*, 795, 42  
 Poleski, R., Udalski, A., Dong, S., et al. 2014b, *ApJ*, 782, 47  
 Refsdal, S. 1966, *MNRAS*, 134, 315  
 Rhie, S. H., Bennett, D. P., Becker, A. C., et al. 2000, *ApJ*, 533, 378  
 Schechter, P. L., Mateo, M., & Saha, A. 1993, *PASP*, 105, 1342  
 Thompson, T. A. 2013, *MNRAS*, 431, 63  
 Udalski, A., Jaroszyński, M., Paczyński, B., et al. 2005, *ApJL*, 628, L109  
 Udalski, A., Szymański, M. K., & Szymański, G. 2015a, *AcA*, 65, 1  
 Udalski, A., Yee, J. C., Gould, A., et al. 2015b, *ApJ*, 799, 237  
 Yee, J. C., Han, C., Gould, A., et al. 2014, *ApJ*, 790, 14  
 Yee, J. C., Svartzvald, Y., Gal-Yam, A., et al. 2012, *ApJ*, 755, 102  
 Yee, J. C., Udalski, A., Calchi Novati, S., et al. 2015, *ApJ*, 802, 76  
 Yoo, J., DePoy, D. L., Gal-Yam, A., et al. 2004, *ApJ*, 603, 139  
 Zhu, W., Udalski, A., Gould, A., et al. 2015, *ApJ*, 805, 8

<sup>17</sup> It does not apply to a microlens survey by *Kepler* (K2) because the field will be a large, pre-selected region rather than having the observations targeted at specific, known microlensing events.