

**I. Ash Ejection and Exposure during Radius  
Expansion Type I X-ray Bursts,  
II. Stellar Dynamics at the Galactic Center,  
III. Weak Gravitational Lensing by  
Dark Matter Concentrations**

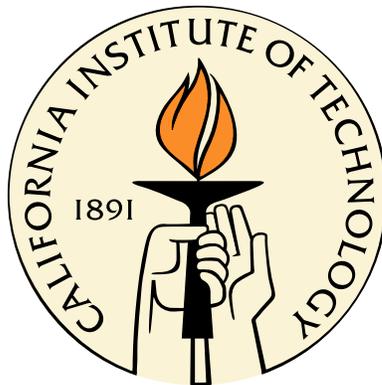
Thesis by

Nevin Nachum Weinberg

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2005

(Defended May 18, 2005)

© 2005

Nevin Nachum Weinberg

All Rights Reserved

# Acknowledgements

I would first like to thank my advisor, Marc Kamionkowski, for his gentle guidance and for encouraging me to explore different topics in astrophysics. I greatly appreciate his continual concern for my well-being and his admirable devotion to teaching. I am also very grateful for my friend, collaborator, and officemate, Miloš Milosavljević, who has taught me many of the most important skills I leave here with. His high standards and attention to detail significantly improved much of the work presented herein. I would also like to thank Lars Bildsten, who, in this last year, has selflessly taught me almost everything I now know about white dwarfs and neutron stars. I appreciate the patience and kindness he has shown me and my family.

These last several years would have been very difficult without the support of my parents, sisters, brother, and grandmother. My parents encouragement and interest in all my pursuits are invaluable. Lastly, and most of all, I would like to thank my incredible wife, Olga, who, after listening to my many complaints and worries over the years, still manages to treat each one as if it was of great importance. Between studying for medical school, searching for the cure for lung cancer, and taking Miko and me out for walks, she somehow remembers to put my wallet in my bag and chips in my lunch.

# Abstract

The studies presented herein are on three distinct topics in astrophysics. I first present an investigation of the possibility of ash ejection during X-ray bursts, then a study of stellar dynamics at the Galactic center, and finally I examine weak gravitational lensing by dark matter concentrations.

I. We solve the time-dependent entropy equation that describes the evolution of the vertical extent of the convective region of a neutron star atmosphere during a Type I X-ray burst. The convective region is well-mixed with ashes of nuclear burning due to the short turbulent mixing timescale, and its extent determines the rise time of the burst light curve. We show that the maximum vertical extent of the convective region during photospheric radius expansion (PRE) bursts can be sufficiently great that: (1) some ashes of burning are ejected by the radiation-driven wind during the PRE phase and, (2) some ashes of burning are exposed at the neutron star surface following the PRE phase. We calculate the expected column density of ashes in hydrogen-like states and determine the equivalent widths of the resulting photoionization edges. We find that the edges should be detectable with current high spectral resolution X-ray telescopes. A detection would probe the nuclear burning processes of a burst and might enable a measurement of the gravitational redshift of the neutron star.

II. We discuss physical experiments achievable via the monitoring of stellar dynamics near the massive black hole at the Galactic center with a diffraction-limited, next-generation, extremely large telescope (ELT). We use the Markov Chain Monte Carlo method to evaluate the constraints that the monitoring of these orbits will place on the matter content within the dynamical sphere of influence of the black hole. We compare these future constraints with the constraints obtained with the current data.

We also describe how the monitoring of stellar proper motions can be used to probe directly the masses of isolated stellar remnants near the MBH. We derive expressions for the rate at which perturbations from remnants deflect stellar orbits, and describe how the remnant masses can be extracted from the monitoring data.

III. We calculate the abundance of dark-matter concentrations that are sufficiently overdense to produce a detectable weak-gravitational-lensing signal. Most of these overdensities are virialized halos containing identifiable X-ray and/or optical clusters. However, a significant fraction are nonvirialized, cluster-mass overdensities still in the process of gravitational collapse—these should produce significantly weaker or no X-ray emission. Our predicted abundance of such dark clusters is consistent with the abundance implied by the detection of apparent dark lenses. Such weak lenses should help shed light on the process of cluster formation. We also examine the prospect of using the observed abundance of weak gravitational lenses to constrain the equation-of-state parameter  $w = p/\rho$  of the dark energy.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>I Ash Ejection and Exposure during Radius Expansion Type I X-ray Bursts</b>	<b>1</b>
<b>1 Exposing the Nuclear Burning Ashes of Radius Expansion X-ray Bursts</b>	<b>2</b>
1.1 Introduction . . . . .	3
1.2 Evolution of the atmosphere during a burst . . . . .	7
1.2.1 Thermal structure of the atmosphere . . . . .	7
1.2.2 Temporal evolution of the convective extent . . . . .	10
1.2.3 Light curve during burst rise and radiative winds . . . . .	16
1.2.4 Dependence of the convective extent on burst parameters . . . . .	18
1.3 Detecting the nuclear burning ashes . . . . .	20
1.3.1 Composition of ejected and exposed ashes . . . . .	21
1.3.2 Ejection of p-nuclei . . . . .	23
1.3.3 Spectral edges in wind outflow . . . . .	24
1.3.4 Spectral edges from NS surface . . . . .	27
1.4 Summary and conclusions . . . . .	28

<b>II</b>	<b>Stellar Dynamics at the Galactic Center</b>	<b>35</b>
<b>2</b>	<b>Stellar Dynamics at the Galactic Center with an Extremely Large Telescope</b>	<b>36</b>
2.1	Introduction . . . . .	37
2.2	Observing stars in the central arcsecond with an ELT . . . . .	42
2.2.1	Astrometric and spectroscopic limit . . . . .	43
2.2.2	Confusion limit . . . . .	44
2.2.3	Central point source — Sgr A* . . . . .	48
2.3	Orbital dynamics . . . . .	49
2.3.1	Equations of motion . . . . .	50
2.3.2	Newtonian retrograde precession . . . . .	51
2.3.3	Relativistic prograde precession . . . . .	53
2.3.4	Frame dragging . . . . .	53
2.3.5	The Roemer time delay . . . . .	55
2.3.6	Interstellar interactions . . . . .	57
2.4	Method . . . . .	61
2.4.1	Parameter estimation . . . . .	61
2.4.2	Mock data . . . . .	64
2.5	Results . . . . .	68
2.5.1	Measuring $M_{\text{bh}}$ and $R_0$ . . . . .	69
2.5.2	Measuring the extended matter distribution . . . . .	70
2.5.3	Measuring relativistic effects . . . . .	71
2.6	Constraints on Galactic structure from measurements of $R_0$ . . . . .	73
2.7	Conclusions . . . . .	74
2.A	Current constraints on $R_0$ and the extended matter distribution . . . . .	76
<b>3</b>	<b>Finding Isolated Stellar Remnants at the Galactic Center</b>	<b>84</b>
3.1	Introduction . . . . .	85
3.2	Gravitational interactions . . . . .	86
3.3	Parameter estimation . . . . .	92

3.4	Discussion . . . . .	94
<b>III</b>	<b>Weak Gravitational Lensing by Dark Matter Concentrations</b>	<b>100</b>
<b>4</b>	<b>Weak Gravitational Lensing by Dark Clusters</b>	<b>101</b>
4.1	Introduction . . . . .	102
4.2	Minimum overdensity required to produce weak-lensing signal . . . . .	104
4.3	Dynamics . . . . .	108
4.4	Abundances . . . . .	113
4.5	Results . . . . .	115
4.5.1	Minimum overdensity as a function of redshift . . . . .	115
4.5.2	The abundance of dark and virialized lenses . . . . .	118
4.5.3	The effect of increasing the image size on the lensing signal . . . . .	122
4.5.4	Estimating $\sigma_8$ from the abundance of weak lenses . . . . .	125
4.6	Discussion and conclusions . . . . .	126
4.A	Derivation of the signal-to-noise relation for various density profiles . . . . .	129
<b>5</b>	<b>Constraining Dark Energy from the Weak Gravitational Lens Abundance</b>	<b>137</b>
5.1	Introduction . . . . .	138
5.2	Minimum overdensity needed to produce detectable lensing signal . . . . .	141
5.3	Spherical collapse in dark energy cosmologies . . . . .	145
5.3.1	Dynamics . . . . .	146
5.3.2	Abundances . . . . .	151
5.3.3	Normalizing the power spectrum . . . . .	155
5.4	Results . . . . .	157
5.4.1	Preliminaries . . . . .	158
5.4.2	Weak lens abundances . . . . .	160
5.4.3	Fraction of lenses that are dark . . . . .	163

5.5	Discussion and conclusions . . . . .	167
5.A	Phantom energy: dark energy with $w < -1$ causes a cosmic doomsday	170

# List of Figures

1.1	Thermal timescale and growth timescale of the convective region . . .	14
1.2	Evolution of the temperature profile . . . . .	15
1.3	Rising portion of the burst light curve . . . . .	17
1.4	Evolution of the top of the convective zone . . . . .	19
1.5	Maximum extent of the convective region . . . . .	20
1.6	Rise time of the burst light curve . . . . .	21
1.7	Ash composition as a function of column depth . . . . .	22
1.8	Ash composition for three different models . . . . .	23
1.9	Equivalent width of the photoionization edge due to ashes in wind . . .	25
1.10	Equivalent width of the photoionization edge due to ashes at surface .	28
2.1	$K$ -band magnitude limit and number of detectable stars . . . . .	47
2.2	Astrometric positions of the 20 synthesized orbits . . . . .	66
2.3	Constraint on $M_{\text{bh}}$ and $R_0$ obtainable with an ELT . . . . .	69
2.4	Constraint on the extended matter distribution obtainable with an ELT	70
2.5	An ELT's sensitivity to post-Newtonian effects . . . . .	72
2.6	Constraint on $M_{\text{bh}}$ and $R_0$ from current Keck data . . . . .	77
2.7	Constraint on the extended matter distribution from current Keck data	78
3.1	Probability that a star experiences detectable gravitational deflection .	91
3.2	Constraint on stellar remnant masses from detected deflections . . . .	93
4.1	Radial evolution of density perturbation in STHC model . . . . .	111
4.2	Nonlinear overdensity as a function of linear-theory overdensity . . . .	113
4.3	Mean tangential shear as a function of distance from lens center . . . .	116

4.4	Minimum nonlinear overdensity needed to detect a weak lens . . . . .	117
4.5	Redshift distribution of dark lenses and virialized lenses . . . . .	118
4.6	Number counts of dark and virialized lenses . . . . .	119
4.7	Predicted weak-lensing mass distribution . . . . .	120
4.8	Fraction of weak lenses that are dark lenses as a function of redshift . .	121
4.9	Signal-to-noise ratio within angular radius $\theta$ from lens center . . . . .	123
4.10	Effect of increasing image size on lens abundance estimates . . . . .	124
4.11	Predicted sky density of weak lenses . . . . .	125
5.1	Nonlinear overdensity versus linear overdensity for three $w$ models . . .	148
5.2	Radial evolution of density perturbation in quintessence STHC model .	149
5.3	Nonlinear overdensity at virialization for three $w$ models . . . . .	150
5.4	Linear-theory threshold for collapse in quintessence STHC model . . .	151
5.5	Dependence of $\sigma_8$ on $w$ . . . . .	156
5.6	Constraint on $\sigma_8$ from X-ray cluster abundance and COBE . . . . .	157
5.7	Comoving number density of virialized objects for three $w$ models . . .	159
5.8	Fraction of non-virialized overdensities to virialized overdensities . . . .	160
5.9	Redshift distribution of virialized lenses and non-virialized lenses . . .	161
5.10	Number of lenses per square degree as a function of $w$ . . . . .	163
5.11	Fraction of lenses that are dark as a function of $w$ . . . . .	164
5.12	Number of lenses per square degree as a function of image aperture size	165
5.13	Observed angular size of $\theta_{\text{vir}}$ and $\theta_{\text{max}}$ as a function of redshift . . . . .	166
5.14	Current constraints to the $w$ - $\Omega_m$ parameter space assuming $w > -1$ . .	171
5.15	Current constraints to the $w$ - $\Omega_m$ parameter space extended to $w < -1$	173

# List of Tables

1.1	Burst ignition models . . . . .	13
2.1	Stellar orbital parameters . . . . .	67
5.1	History and future of the Universe with $w = -3/2$ phantom energy . .	175

## Part I

# Ash Ejection and Exposure during Radius Expansion Type I X-ray Bursts

# Chapter 1

## Exposing the Nuclear Burning Ashes of Radius Expansion X-ray Bursts

NEVIN N. WEINBERG

Theoretical Astrophysics, California Institute of Technology, Pasadena, CA 91125

LARS BILDSTEN

Kavli Institute for Theoretical Physics, Kohn Hall, University of California, Santa Barbara, CA  
93106

HENDRIK SCHATZ

Department of Physics and Astronomy, National Superconducting Cyclotron Laboratory & Joint  
Institute of Nuclear Astrophysics, Michigan State University, East Lansing, MI 48824

## Abstract

We solve the time dependent entropy equation that describes the evolution of the vertical extent of the convective region of a neutron star atmosphere during a Type I X-ray burst. The convective region is well-mixed with ashes of nuclear burning due to the short turbulent mixing time scale and its extent determines the rise time of the burst light curve. We show that the maximum vertical extent of the convective region during photospheric radius expansion (PRE) bursts can be sufficiently great that: (1) some ashes of burning are ejected by the radiation driven wind during the PRE phase and, (2) some ashes of burning are exposed at the neutron star surface following the PRE phase. Depending on the ignition conditions, ashes with mass number in the range  $A \sim 30 - 60$  are mixed in with the ejected material. If bursts ignite in the ashes of previous bursts, as recent numerical simulations suggest, even heavier elements are ejected, possibly including some light p-nuclei in the  $A = 80 - 100$  region whose origin in the solar system is not understood. As the ejected material cools during the PRE phase some of the ejected heavy-element ashes will cease to be fully ionized. In addition, those ashes that remain bound to the neutron star will temporarily reside in the photosphere even after it has settled back down to the neutron star surface. If sufficiently heavy, some of these photospheric ashes will also cease to be fully ionized. We calculate the expected column density of ashes in hydrogen-like states and determine the equivalent widths of the resulting photoionization edges. We find that with current high spectral resolution X-ray telescopes, which have not yet examined PRE bursts, one should be able to detect the line edges in both the wind and from the neutron star surface. A detection would probe the nuclear burning processes of a burst and might enable a measurement of the gravitational redshift of the neutron star.

---

## 1.1 Introduction

Type I X-ray bursts are produced by the unstable nuclear burning of freshly accreted hydrogen- and/or helium-rich material on the surface of a neutron star (NS) in a low-mass X-ray binary (for reviews, see Lewin, van Paradijs, & Taam 1995; Bildsten 1998; Strohmayer & Bildsten 2004). The burst energies ( $10^{39} - 10^{40}$  erg), durations ( $\sim 10 - 100$  s), and recurrence times (hours to days), depend strongly on the composition of the accreted matter and on the accretion rate,  $\dot{M}$ , which can range from  $10^{-11}$  to  $10^{-8} M_{\odot} \text{ yr}^{-1}$ . The burst properties are also sensitive to the composition of the ashes

of burning from preceding bursts as underscored by recent burst simulations that implement large nuclear reaction networks for the energy generation (Schatz et al., 2001; Brown et al., 2002; Woosley et al., 2004).

The sensitivity of the nuclear energy generation rate to temperature and density concentrates the burning during a burst to a thin layer at the base of the accreted atmosphere (e.g., Fujimoto, Hanawa, & Miyaji 1981; Fushiki & Lamb 1987). Since the thermal timescale there is  $\sim 1 - 10$  s while the dynamical timescale is  $\sim 10^{-6}$  s, the temperature gradient near the burning layer is nearly adiabatic resulting in a region of highly efficient convection. The short turbulent mixing timescale of the convection ensures that the ashes of burning are well-mixed throughout the convective region.

The vertical extent of the convective region evolves during the burst, as demonstrated in time-dependent numerical simulations (e.g., Joss 1978; Taam 1980; Ayasli & Joss 1982; Woosley & Weaver 1984; Woosley et al. 2004). In this paper, we carry out a thorough survey of the dependence of the convection region's extent on  $\dot{M}$ , the composition of the accreted material, and the pre-burst thermal state of the atmosphere. We also show how the evolution of the convective extent influences the observed burst rise times.

We demonstrate that for photospheric radius expansion (PRE) bursts, in which the super-Eddington luminosity drives a radiation-driven wind, the convective region extends out to sufficiently low pressures that ashes can be ejected by the wind. Depending on the burst parameters, as much as  $\sim 1\%$  of the accreted mass can be ejected by the radiative wind (Paczynski & Proszynski, 1986; Joss & Melia, 1987; Nobili, Turolla, & Lapidus, 1994), corresponding roughly to the ratio of nuclear energy release to gravitational binding energy  $\sim E_{\text{nuc}}/E_{\text{grav}} \simeq 2 \text{ MeV nucleon}^{-1}/200 \text{ MeV nucleon}^{-1}$ . The ejected ashes may be detectable with spectroscopy during the PRE phase and afterwards, when the photosphere, laced with heavy element ashes, settles back down to the NS surface.

The sequence of events describing how the convective region evolves during a burst is as follows. As the base temperature rises and the energy flux increases during the early stages of a burst, the entropy in the convective region increases.

Initially, only a negligible amount of thermal energy is lost to radiation diffusing into the overlying radiative region and underlying crust. Since the timescale for radiative diffusion across the convective-radiative interface is longer than the burning timescale during these early stages, the convective region extends vertically outward to lower pressures (Fujimoto, Hanawa, & Miyaji, 1981). Eventually, the burning rate becomes sufficiently slow that the energy flux from the base is carried most efficiently by radiation rather than by convection. At that point the convective region recedes back to higher pressures.

Regardless of whether the convective region is expanding outwards or receding downwards, its extent at a given time is effectively set by the radial location in the atmosphere where the constant entropy of the convective region equals the radially increasing entropy of the overlying radiative region. Based on this argument, Joss (1977) showed that the convective region never acquires a high enough entropy to reach the photosphere located at column depth  $y \sim 1 \text{ g cm}^{-2}$ . Nonetheless, the convective region can reach pressures  $\lesssim 1\%$  of that at the base of the accreted layer (Taam, 1981; Ayasli & Joss, 1982; Hanawa & Sugimoto, 1982; Hanawa & Fujimoto, 1984). Woosley et al. (2004) showed that in a pure He burst, the peak flux exceeded the Eddington limit and the convective region extended beyond their numerical surface (i.e., the resolution limit of their grid located at a pressure  $\sim 0.3\%$  that at the base).

In this paper, we solve the time-dependent entropy equation that describes the evolution of the thermal structure of the atmosphere and the growth of the convective region. We calculate the minimum pressure reached by the convective region  $y_{c,\min}$  for a range of burst parameters. We show that the maximum extent of the convective region is sensitive to the burst ignition conditions and that, in general, the larger the burst peak flux and the smaller the entropy of the pre-burst atmosphere, the greater the convective region's maximum extent. The maximum convective extent also increases with decreasing  $\dot{M}$ . We find that in systems where the accreted material is helium-rich, such as 4U 1820–30 (see Cumming 2003 and references therein), or in systems accreting solar abundances at  $\dot{M} \sim 10^{-10} M_{\odot} \text{ yr}^{-1}$ ,  $y_{c,\min} \ll y_{\text{wind}}$  during

PRE bursts, where  $y_{\text{wind}}$  is the column depth above which mass is ejected by the radiation-driven wind. As a result, some of the nuclear burning ashes are ejected by the wind while some remain bound to the NS and are exposed at the photosphere after the wind turns off.

In § 1.2 we describe our analytic prescription for calculating the evolution of the thermal structure of the radiative and convective regions during fuel accumulation and the X-ray burst. We examine how the maximum extent of the convective region depends on the burst ignition conditions and explore those conditions most conducive to ash ejection and exposure. Since the burst rise time is effectively set by the thermal diffusion time at the maximum convective extent, we also evaluate the dependence of the burst rise time on burst parameters such as accretion rate and accreting composition. In § 1.3 we describe the composition of the ashes and the observational consequences of heavy-element ash ejection and surface exposure. Depending on the ignition conditions, nuclei as heavy as  $A \sim 60$  are ejected by the wind. If bursts ignite in the ashes of previous bursts, as Woosley et al. (2004) suggest, even heavier elements are ejected. These may include some light p-nuclei whose origins are not understood, as they are systematically underproduced in all standard p-process scenarios. We show that a sufficient amount of these nuclei can be ejected during PRE bursts to account for at least 1 – 10% of the observed solar system abundances. We also calculate the column density of ashes in hydrogen-like states in both the wind and the surface photosphere and discuss the prospects for detecting absorption line features of the ashes from high spectral resolution observations of PRE bursts. Such line features offer a direct probe of the nuclear burning and may help constrain the NS equation of state. In § 1.4 we summarize our results, argue that ash ejection may be responsible for the unusual Ne/O ratios observed in four candidate ultracompact binaries, and mention the possibility of ash ejection during superbursts.

## 1.2 Evolution of the atmosphere during a burst

In this section we consider the evolution of the atmosphere's thermal structure during a burst. We start in § 1.2.1 with a general description of the atmosphere's structure and explain the boundary conditions at the convective-radiative interface. The evolution of the convective extent is described in § 1.2.2 and in § 1.2.3 we illustrate how our prescription for determining this evolution yields an estimate of the light curve during burst rise. In § 1.2.4 we explore the dependence of the convective extent and the rise times on burst parameters such as accretion rate, composition of accreted material, and column depth of ignition.

### 1.2.1 Thermal structure of the atmosphere

The NS atmosphere maintains hydrostatic equilibrium throughout the burst so that the pressure varies with height as  $dP/dr = -\rho g$  where  $\rho$  is the density and  $g$  is the surface gravity. We assume an  $M = 1.4M_{\odot}$  NS with radius  $R = 10$  km giving  $g = (1 + z)GM/R^2 = 2.43 \times 10^{14}$  cm s<sup>-2</sup> where the gravitational redshift  $z = (1 - 2GM/Rc^2)^{-1/2} - 1 = 0.31$ . Since the atmosphere is thin compared with the NS radius,  $g$  is effectively constant throughout the accreted layer. Hydrostatic balance therefore yields  $P = gy$ , where the column depth  $y$  is defined by  $dy = -\rho dr$ . Parameterizing the spatial coordinate by  $y$ , we now calculate the evolution of the thermal structure of the atmosphere before and during a burst. We are interested in determining the extent of the convective region over a broad range of burst parameters and thus consider only one-dimensional models in our calculations. We do not account for the affect of a spreading burning front during burst rise nor the influence of rotation on the convective structure, though these effects may be important (see Spitkovsky et al. 2002; Zingale et al. 2003).

Just before the thermally unstable helium ignition the accumulating atmosphere is entirely radiative and its thermal profile is described by the heat equation

$$\frac{dT}{dy} = \frac{3\kappa F}{4acT^3}, \quad (1.1)$$

where  $F$  is the outward heat flux and the opacity  $\kappa$  has contributions from electron scattering and free-free absorption and is calculated using the approximation given by Schatz et al. (1999). The pre-burst flux  $F = F_H + F_{\text{crust}}$ , where  $F_H$  is the flux from stable hydrogen burning via the hot CNO cycle and  $F_{\text{crust}}$  is the flux from heat released by electron captures and pycnonuclear reactions deep in the crust (Brown & Bildsten, 1998; Brown, 2000, 2004). Following the burst ignition calculations of Cumming (2003, hereafter C03),  $F_H = \epsilon_H \min(y_H, y_b)$  where  $y_H$  is the column depth of the layer that is burning hydrogen,  $y_b \simeq 3 \times 10^8 \text{ g cm}^{-2}$  is the column depth at the base, and  $\epsilon_H = 5.8 \times 10^{13} \text{ ergs g}^{-1} \text{ s}^{-1} (Z/0.01)$  is the hot CNO energy production rate for a CNO mass fraction  $Z$ . For a given local accretion rate  $\dot{m}$  (in units of  $\text{g cm}^{-2} \text{ s}^{-1}$ ) and accreted hydrogen mass fraction  $X_0$ , the hydrogen burning depth is  $y_H = 6.8 \times 10^8 \text{ g cm}^{-2} (\dot{m}/0.1\dot{m}_{\text{Edd}})(0.01/Z)(X_0/0.71)$ . Here  $\dot{m}_{\text{Edd}} = 2m_p c / (1 + X_H) R \sigma_{\text{Th}}$  is the local Eddington accretion rate where  $m_p$  is the proton mass,  $c$  the speed of light, and  $\sigma_{\text{Th}}$  the Thomson scattering cross section. For  $\dot{m} < \dot{m}_{\text{crit}} \simeq 0.04\dot{m}_{\text{Edd}}$  there is enough time to burn all the hydrogen before the helium burning becomes unstable, and the burst ignites in a pure helium layer. As in C03, we assume  $F_{\text{crust}} = \dot{m} Q_{\text{crust}} = 10^{21} \text{ ergs cm}^{-2} \text{ s}^{-1} \dot{m}_4 Q_{17}$  where  $\dot{m}_4 = 10^4 \text{ g cm}^{-2} \text{ s}^{-1}$  and  $Q_{\text{crust}} = Q_{17} \times 10^{17} \text{ ergs g}^{-1}$  is the radiative energy emerging from the crust per unit mass.

The thermal evolution of the NS atmosphere during a burst is described by the entropy equation

$$T \frac{ds}{dt} = \frac{dF}{dy} + \epsilon, \quad (1.2)$$

where  $\epsilon$  is the energy release rate from nuclear burning. During the burst the entropy grows with time due to nuclear burning and we neglect the advective accretion flow. Therefore,  $T ds = C_p dT$  where  $C_p$  is the specific heat at constant pressure. Integrating equation (1.2) over column depth then gives

$$\int_{y_1}^{y_2} C_p \frac{dT}{dt} dy = F(y_2) - F(y_1) + \int_{y_1}^{y_2} \epsilon dy. \quad (1.3)$$

We assume the atmosphere is composed of two regions: a completely convective region between  $y_c < y < y_b$ , and a completely radiative region for  $y < y_c$ . During a burst  $y_b$

is constant while  $y_c$  evolves from an initial value  $y_c = y_b$  to a minimal value  $y_c = y_{c,\min}$  and finally back to  $y_c = y_b$ . Demarcating the atmosphere in this way is reasonable given that the convective eddies are highly subsonic over most of the convective zone, i.e., near the base  $v_{\text{conv}} \simeq (F/\rho)^{1/3} \sim 10^7 \text{ cm s}^{-1} \ll c_s \simeq (gy_b/\rho)^{1/2} \sim 2 \times 10^8 \text{ cm s}^{-1}$ .

Although the entropy in the convective region grows with time, at a given instant it is nearly spatially constant. This, in addition to the subsonic motion of the convective eddies, suggests that the temperature in the convective region very nearly follows an adiabat so that  $(d \ln T / d \ln y)_{\text{conv}} = (d \ln T / d \ln y)_{\text{ad}} \equiv n(y)$ , i.e.,  $T(y_c < y < y_b) = T_b (y/y_b)^{n(y)}$ , where the adiabatic index  $n(y)$  varies with column depth. For the equation of state we use the interpolation formulae of Paczynski (1983) to account for the partially degenerate electrons. Using his notation, the specific heat and adiabatic index are given by

$$C_p = \frac{1}{\rho T} \left[ \frac{3}{2} P_i + 12 P_r + \frac{P_{\text{end}}^2}{(f-1)P_e} + \frac{P \chi_T^2}{\chi_\rho} \right], \quad (1.4)$$

$$n = \frac{P}{C_p \rho T} \frac{\chi_T}{\chi_\rho}, \quad (1.5)$$

where  $P_i$ ,  $P_r$ , and  $P_e = (P_{\text{end}}^2 + P_{\text{ed}}^2)^{1/2}$  are the pressure due to ions, radiation, and electrons, respectively,  $P_{\text{end}}$  and  $P_{\text{ed}}$  are an approximation to the degenerate and non-degenerate components of the electron pressure,  $P = P_i + P_e + P_r$ ,  $f = d \ln P_{\text{ed}} / d \ln \rho$ , and

$$\chi_T \equiv \left( \frac{\partial \ln P}{\partial \ln T} \right)_\rho = \frac{1}{P} \left[ P_i + 4 P_r + \frac{P_{\text{end}}^2}{P_e} \right], \quad (1.6)$$

$$\chi_\rho \equiv \left( \frac{\partial \ln P}{\partial \ln \rho} \right)_T = \frac{1}{P} \left[ P_i + \frac{P_{\text{end}}^2 + f P_{\text{ed}}^2}{P_e} \right]. \quad (1.7)$$

At burst onset the pressure is nearly that of an ideal gas and  $n \simeq 2/5$  while at late times radiation pressure contributes significantly, which in the limit  $P = P_r$  gives  $n = 1/4$ .

The thermal profile in the radiative region satisfies equation (1.1). During the early stages of a burst the thermal time at the base of the radiative region is much

longer than the growth time of the convective region. The flux in the radiative region during this time is therefore the pre-burst flux (i.e.,  $F_H + F_{\text{crust}}$ ). Eventually, the convective region reaches sufficiently low pressures that the thermal time becomes comparable to the growth time and the flux through the radiative region begins to rise. Since the radiative region is composed primarily of freshly accreted hydrogen and/or helium, the main opacity is Thomson scattering  $\kappa \simeq \kappa_{\text{es}} = \sigma_{\text{Th}}(1 + X)/2m_p$ . The opacity varies only slightly with column depth so that over much of the radiative region  $d \ln T / d \ln y \simeq 1/4$ . For mixed hydrogen/helium accretion at  $\dot{m} < \dot{m}_{\text{crit}}$ , a pure helium layer develops over the region  $y_H < y < y_b$ . In this region  $F_H = 0$  and since  $F_{\text{crust}}$  is small at low  $\dot{m}$ , the pre-burst profile there is nearly isothermal.

We define the column depth of the convective-radiative interface  $y_c$  as the location where the density of the radiative solution just exceeds that of the convective solution. For models accreting hydrogen, there is a large compositional contrast between the helium-rich matter that is burning and the outer hydrogen-rich material. This inhibits the outward progress of the convective zone as the burning material must get even hotter to become buoyant in the overlying hydrogen-rich envelope. Since the particle diffusion timescale is much longer than the growth time of the convective zone  $t_{\text{gr}} \equiv dt/d \ln T_b$ , we approximate the composition gradient as a discontinuous step function. For this reason, we also assume there is not adequate time for semi-convection to develop. As we will show, the steeper composition gradient in systems accreting hydrogen-rich material results in a significant suppression of the convective extent as compared with those systems accreting pure helium. This is true even when the former accretes at  $\dot{m} < \dot{m}_{\text{crit}}$  and thus also burns in a pure helium environment (see also Cumming & Bildsten 2000).

### 1.2.2 Temporal evolution of the convective extent

The evolution of the convective extent  $y_c$  depends on the rate at which the base temperature rises  $dT_b/dt$  and the rate at which the thermal energy of the overlying radiative region increases. The rate of temperature change in the convective region is

$dT/dt = (y/y_b)^n [dT_b/dt + T_b \ln(y/y_b)dn/dt]$ . The second term is negligible compared with the first term so that by equation (1.3),

$$\frac{dT_b}{dt} = \frac{\int_{y_c}^{y_b} \epsilon dy + F_H + F_{\text{crust}} - F_{\text{loss}}(y_c)}{\int_{y_c}^{y_b} C_p(y/y_b)^{n(y)} dy}, \quad (1.8)$$

where  $F_{\text{loss}}(y_c)$  is the radiative flux escaping from the convective region into the overlying radiative region. Physically, the rate at which  $T_b$  changes is determined by the competition between the net energy input into the convective region (i.e., the energy generated by nuclear burning and crustal heating minus the energy lost to radiation) and the energy expended in heating up the growing convective region.

We determine  $dX_i/dt = R_i(\rho, T_b, X_1, \dots, X_N)$ , the rate at which the mass fraction  $X_i$  of species  $i$  changes, assuming  $X_H = X_0(1 - y/y_H)$  for  $y < y_H$  and  $X_H = 0$  for  $y > y_H$ . We average over the convective region according to  $\langle X_i \rangle = \int dy X_i(y)/(y_b - y_c)$  for each species  $i$ . For most models, the nuclear energy generation rate  $\epsilon$  is determined using an  $\alpha$ -capture reaction network with rates  $R_i$  given by the NACRE group. For models HeA1full and HHeB1full (see Table 1.1) we used an updated reaction network that includes all proton-rich nuclei from hydrogen to xenon. The  $\alpha$ -capture reaction network includes a multiplicative factor of 1.9 in the triple alpha energy production rate to account for two proton captures on  $^{12}\text{C}$  when hydrogen is present (CB00).

When we include the full network and not just the  $\alpha$ -capture reactions we find that protons are produced via  $(\alpha, p)$  branches off the  $\alpha$ -chain, so that proton captures occur even when burning initiates in a pure helium layer (i.e., those systems accreting mixed hydrogen and helium at  $\dot{m} \lesssim 0.04\dot{m}_{\text{Edd}}$  or systems accreting pure helium). The dominant proton source is the reaction  $^{24}\text{Mg}(\alpha, p)^{27}\text{Al}(\alpha, p)$  though others contribute somewhat as well. The proton abundances achieved by these reactions are high enough to allow the reaction  $^{12}\text{C}(p, \gamma)^{13}\text{N}(\alpha, p)^{16}\text{O}$  to dominate over the  $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$  reaction. The latter reaction acts as a bottleneck in the  $\alpha$ -chain, so in bypassing it the energy generation rate increases and the convective zone reaches even lower pressures. Furthermore, some protons are captured by heavier nuclei, which together

with the  $(\alpha, p)$  branches leads to a broadening of the abundance pattern.

We use mixing length theory to estimate  $F_{\text{loss}}(y_c)$ . According to mixing length theory the fraction of the total flux transported by convection at column depth  $y$  is (Hansen & Kawaler, 1994),

$$F_{\text{conv}}/F = \frac{\nabla_{\text{rad}} - \nabla_{\text{ad}}}{\nabla_{\text{rad}}} \left( 1 - \frac{1}{Nu} \right), \quad (1.9)$$

where  $\nabla_{\text{ad}} \equiv n(y)$ ,  $\nabla_{\text{rad}} = 3\kappa F y / 4acT^4$ , and  $Nu$  is the Nusselt number describing the efficiency of convection. During the burst, convection is very efficient over most of the convective region. Only when  $y_c \simeq y_{c,\text{min}}$  does convection become inefficient near the top of the convective region, though it always remains subsonic ( $v_{\text{conv}} \lesssim c_s$ ). Because mixing length theory only provides an order of magnitude estimate of  $F_{\text{loss}}(y_c)$ , we introduce a scaling prefactor  $\lambda$  to parameterize the uncertainty in its exact value. We thus have  $F_{\text{loss}}(y_c) = F - F_{\text{conv}}(y_c) \approx \lambda F \nabla_{\text{ad}} / \nabla_{\text{rad}} = \lambda 4acT_c^4 \nabla_{\text{ad}} / 3\kappa y_c$ . Unless otherwise stated, we assume  $\lambda = 1$ . As we show below, the maximum extent of the convective region is sensitive to the value of  $\lambda$ .

The rate of temperature change at the convective-radiative interface  $dT_c/dt$  depends on the growth time of the convective region  $t_{\text{gr}} = dt/d\ln T_b$  and the thermal timescale at the interface  $t_{\text{th}} = C_p T_c y_c / F_{\text{loss}}(y_c)$  (see Hanawa & Sugimoto 1982). At early times, following the onset of unstable helium burning, the interface is located at large column depths and  $t_{\text{th}} \gg t_{\text{gr}}$ . Thus, the radiative region cannot thermally adjust to the growing convective region and the thermal profile in the radiative region is unchanged from the pre-ignition profile. The initial entropy of the atmosphere, which is set by  $F_{\text{initial}} = F_H + F_{\text{crust}}$ , is therefore important in determining the evolution of the convective region.<sup>1</sup>

Eventually the convective region reaches a low enough column depth that  $t_{\text{th}} = t_{\text{gr}}$ .

---

<sup>1</sup>When the density and temperature first get high enough for helium to ignite the burning timescale may be longer than or comparable to the thermal timescale. The flux from this early burning may change the pre-burst profile of the radiative region slightly, though eventually the burning becomes non-linear and the burning time becomes much shorter than the thermal time. To examine how this might affect the growth of the convective region we computed  $y_c(t)$  assuming an artificially high pre-burst flux (e.g.,  $5 \times F_{\text{initial}}$ ). We find that this effect may decrease the maximum extent of the convective region by as much as a factor of 3.

Table 1.1. Burst ignition models

Model	$\dot{m}$	$y_{b,8}$	$X_0$	$Z_{\text{CNO}}$	$Q_{\text{crust}}$	network
HeA1full	0.1	3.0	0.0	0.01	0.1	full
HeA1	0.1	3.0	0.0	0.01	0.1	$\alpha$
HeA2	0.1	5.0	0.0	0.01	0.1	$\alpha$
HeA3	0.1	3.0	0.0	0.01	0.2	$\alpha$
HeC1	0.01	3.0	0.0	0.01	0.1	$\alpha$
HHeA1	0.1	3.0	0.71	0.01	0.1	$\alpha$
HHeB1full	0.05	3.0	0.71	0.01	0.1	full
HHeC1	0.01	3.0	0.71	0.01	0.1	$\alpha$
HHeC2	0.01	3.0	0.1	0.01	0.1	$\alpha$
HHeC3	0.01	3.0	0.1	0.0001	0.1	$\alpha$

Note.—Col. (1): Model. Col. (2): Local accretion rate in units of  $\dot{m}_{\text{Edd}}$ . Col. (3): Ignition column depth  $y_{b,8} = y_b/10^8 \text{ g cm}^{-2}$ . Col. (4): Accreted hydrogen fraction  $X_0$ . Col. (5): CNO mass fraction  $Z_{\text{CNO}}$ . Col. (6):  $Q_{\text{crust}}$  in units of  $\text{MeV nucleon}^{-1}$ . Col. (7): Reaction network used.

The radiative flux escaping the convective region  $F_{\text{loss}}(y_c)$  can then finally diffuse through the entire radiative region without being overtaken by the growing convective region. Some of this flux will heat up the radiative region, while the remainder escapes out through the photosphere. The subsequent evolution of the convective-radiative interface is determined by the column depth at which  $t_{\text{th}} = t_{\text{gr}}$ , i.e., the radiative region continuously adjusts to  $F_{\text{loss}}(y_c)$ , which varies due to changes in the burning rate.

The evolution of  $t_{\text{th}}$  and  $t_{\text{gr}}$  during a burst is illustrated in Figure 1.1 for accretion of pure helium at a rate  $\dot{m}/\dot{m}_{\text{Edd}} = 0.1$  and solar abundance accretion at a rate  $\dot{m}/\dot{m}_{\text{Edd}} = 0.01$  and  $0.1$  ( $T_b$  is used to mark the passage of time; see Table 1.1 for a summary of the model parameters). The pure helium model is similar to the ignition models found by C03 in fits to the burst properties of 4U 1820-30. For hydrogen-rich accreted material, the lower  $\dot{m}$  the shorter the growth time at a given  $T_b$ . This is because the lower  $\dot{m}$  the larger the mass fraction of helium at the burning layer and

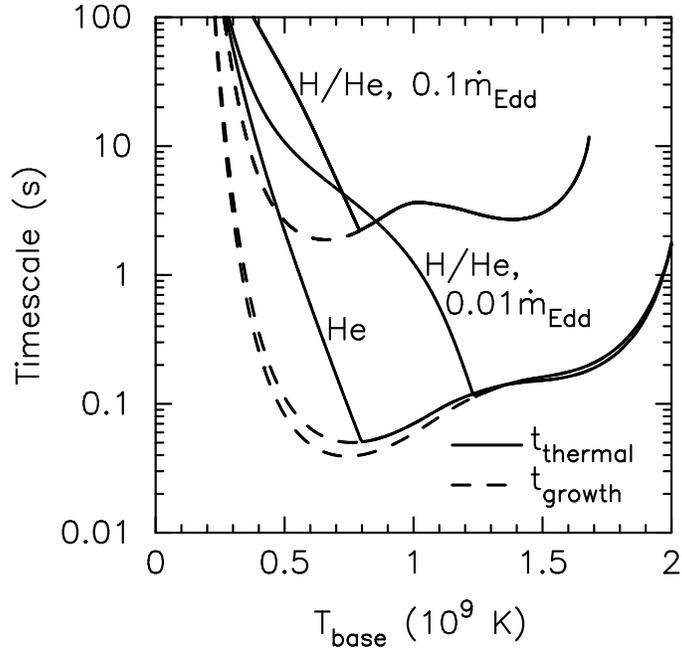


Figure 1.1 Thermal timescale at the convective-radiative interface  $t_{\text{th}}(y_c)$  (*solid line*) and the growth timescale of the convective region  $t_{\text{gr}} \equiv dt/d \ln T_b$  (*dashed line*) as a function of the base temperature  $T_b$  for nearly pure helium accretion with  $\dot{m}/\dot{m}_{\text{Edd}} = 0.1$  (model HeA1; see Table 1.1) and solar abundance accretion ( $X_0 = 0.71$ ) with  $\dot{m}/\dot{m}_{\text{Edd}} = 0.1$  and  $0.01$  (models HHeA1 and HHeC1). All three models assume  $y_b = 3 \times 10^8 \text{ g cm}^{-2}$ ,  $Z_{\text{CNO}} = 0.01$ , and  $Q_{\text{crust}} = 0.1 \text{ MeV nucleon}^{-1}$ . The maximum extent of the convective region occurs approximately when the equality  $t_{\text{th}} = t_{\text{gr}}$  is first satisfied.

therefore the higher the triple alpha energy generation rate  $\epsilon_{3\alpha} \propto X_{\text{He}}^3$ .

In Figure 1.2 we show the temperature profile at different stages of a burst for the same ignition models. At early times  $t_{\text{th}} > t_{\text{gr}}$  and the convective region moves outward while the radiative region remains unchanged from its initial profile. When the top of the convective region reaches a low enough pressure that  $t_{\text{th}} = t_{\text{gr}}$ , the radiative region heats up. Eventually the entropy in the radiative region gets sufficiently high and the burning rate sufficiently slow that the convective region recedes. The slope of the convective adiabat changes during the burst, with radiation pressure becoming increasingly important ( $n \rightarrow 0.25$ ) as  $T_b$  rises. Ultimately, the entire atmosphere becomes radiative.

The temperature jump at the convective-radiative interface is less drastic for the

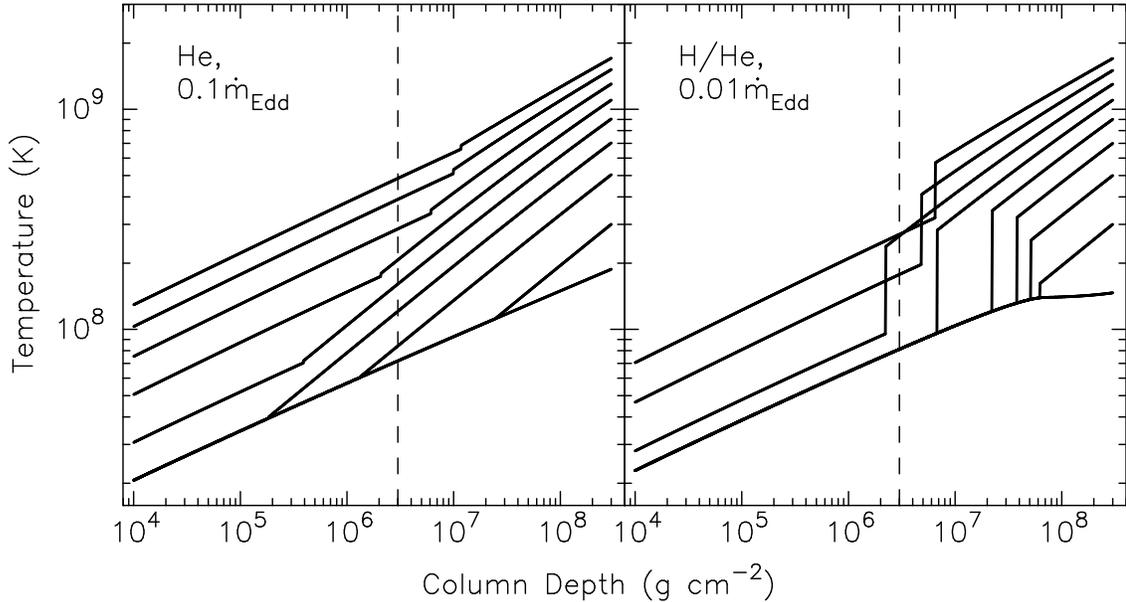


Figure 1.2 Evolution of the temperature profile for the pure helium accretion model HeA1 (*left panel*) and the mixed hydrogen/helium accretion model HHeC1 (*right panel*) at nine different times corresponding to  $T_b/10^9 \text{ K} = T_0, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5,$  and  $1.7$ , where  $T_0$  is the base temperature set by the radiative solution with  $F = F_H + F_{\text{crust}}$ . For both models we assume  $Z_{\text{CNO}} = 0.01$ ,  $y_b = 3 \times 10^8 \text{ g cm}^{-2}$ , and  $Q_{\text{crust}} = 0.1 \text{ MeV nucleon}^{-1}$ . The convective radiative interface is located at the temperature (i.e., composition) discontinuity. The vertical *dashed line* denotes the column depth where  $y = 0.01y_b$ , corresponding approximately to the column depth  $y_{\text{wind}}$  below which material is ejected by the radiative wind of a PRE burst. The slope of the convective region becomes more radiative-like ( $n \rightarrow 0.25$ ) as the temperature rises and the ratio of gas pressure to radiation pressure decreases.

pure helium model compared to the mixed hydrogen/helium model because the compositional contrast between the ashes and unburnt material is smaller in the former model. Thus, although both models have a pure helium burning layer, the ashes of the hydrogen/helium model must get hotter to become buoyant in the overlying hydrogen-rich matter and therefore they do not reach pressures as low as those of the pure helium model. As we show below, both models achieve a super-Eddington luminosity that drives a radiative wind capable of ejecting material located at column depths  $y < y_{\text{wind}} \simeq 0.01y_b$ . For the pure helium model the convective zone reaches pressures  $y_{c,\text{min}} \ll y_{\text{wind}}$  and one expects ashes of burning to be amongst the wind ejecta. This condition is only marginally satisfied for the mixed hydrogen/helium

model, though as we show in § 1.2.4 these results are sensitive to the value of  $\lambda$ .

### 1.2.3 Light curve during burst rise and radiative winds

We determine the rising portion of the burst light curve by calculating the radiative flux loss at the photosphere  $F_{\text{ph}}$  as  $T_b$  increases. To obtain  $F_{\text{ph}}$  we integrate equation (1.1) inwards (assuming the radiative zero solution at the outer boundary), varying the flux at the top until the radiative solution intersects the convective solution at the column depth  $y_c$ . Initially,  $t_{\text{th}} > t_{\text{gr}}$  and  $F_{\text{ph}} = 0$ . When  $t_{\text{th}}$  first equals  $t_{\text{gr}}$  the flux  $F_{\text{ph}}$  begins to rise, and increases thereafter as the radiative region heats up due to the rise in  $F_{\text{loss}}(y_c)$ . In Figure 1.3 we show the rising portion of the light curve for the same models as Figure 1.1 and also for the pure helium model using the full reaction network. We plot the luminosity  $L = 4\pi R^2 F_{\text{ph}}$  as a function of  $t - t_0$ , where  $t_0$  corresponds to the time when the equality  $t_{\text{th}} = t_{\text{gr}}$  is first satisfied. The rise time is effectively set by  $t_{\text{th}}(y_c = y_{c,\text{min}})$ . Thus, low  $\dot{m}$  bursts have shorter rise times because the convective region extends to lower pressures where the thermal timescale is shorter. As the differences between the models HeA1 and HeA1full illustrate, the shape and rise time of the light curves are sensitive to the energy generation rate. Depending on the ignition conditions, the base temperature can get sufficiently high that the flux becomes super-Eddington. We find that a peak surface luminosity  $L > L_{\text{Edd}}$  where  $L_{\text{Edd}} = 4\pi GMc(1+z)/\kappa$  is attained when  $\dot{m} \lesssim 0.1\dot{m}_{\text{Edd}}$  for accreted material with solar abundance ( $\dot{m}$  can be higher if the material is helium-rich). Such systems are likely to develop a radiative wind. Eventually, the flux loss exceeds the flux from nuclear burning and the atmosphere begins to cool, though we do not calculate this portion of the light curve.

The radiative wind generated by the super-Eddington luminosity of a PRE burst will eject some ashes of burning if: (1) the fraction of the accreted mass ejected during the wind satisfies  $\Delta M_w/M_{\text{acc}} = y_{\text{wind}}/y_b > y_{c,\text{min}}/y_b$ , and (2) the wind is generated subsequent to the convective region reaching  $y_c < y_{\text{wind}}$ . As Figure 1.3 illustrates,  $y_c = y_{c,\text{min}}$  well before the burst becomes super-Eddington and therefore condition

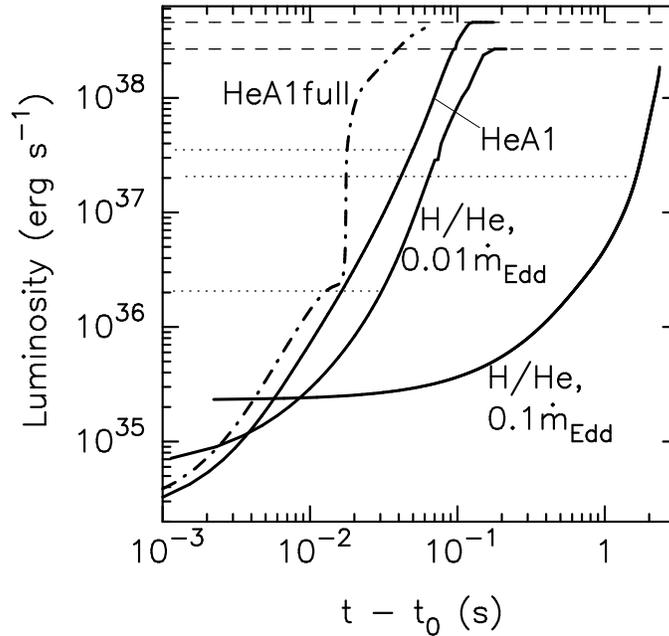


Figure 1.3 Rising portion of the burst light curve for the same three models shown in Figure 1.1 (*solid lines*). The *dash-dot line* is the light curve for the HeA1full model calculated using the full reaction network. The luminosity is plotted as a function of the offset time  $t - t_0$  where  $t_0$  corresponds to the time when the radiative region first begins to evolve away from its pre-burst state (i.e., when the equality  $t_{\text{th}} = t_{\text{gr}}$  is first satisfied). The *dotted lines* denote the constant accretion luminosity for each model and the *dashed lines* denote the Eddington luminosity at the photosphere.

(2) is satisfied if condition (1) holds. Since the gravitational binding energy at the surface is  $\sim 100$  times greater than the helium burning energy release per unit mass, at most  $\sim 1\%$  of the atmosphere is ejected by the wind. To obtain an estimate of  $\Delta M_w / M_{\text{acc}}$ , suppose the surface luminosity  $L$  is super-Eddington so that matter is lifted off the neutron star surface with a force  $f = (L - L_{\text{Edd}})/c$ . Since  $f \approx v_{\text{esc}} \dot{M}_w$ , where  $v_{\text{esc}} = (2GM/R)^{1/2}$  is the escape velocity at the surface and  $\dot{M}_w$  is the mass-loss rate due to the wind,

$$\begin{aligned} \dot{M}_w \simeq \frac{L - L_{\text{Edd}}}{v_{\text{esc}} c} &= 8.6 \times 10^{17} \text{ g s}^{-1} \left( \frac{0.2 \text{ cm}^2 \text{ g}^{-1}}{\kappa} \right) \\ &\times \left( \frac{L}{L_{\text{Edd}}} - 1 \right) \end{aligned} \quad (1.10)$$

(Wallace, Woosley, & Weaver 1982; Paczynski & Proszynski 1986, see also Yahel,

Brinkmann, & Braun 1984; Joss & Melia 1987; Nobili, Turolla, & Lapidus 1994). The burst duration  $\Delta t \sim M_{\text{acc}} Q_{\text{nuc}}/L$ , where  $Q_{\text{nuc}} = 1.6 + 4.0\langle X_H \rangle$  MeV nucleon<sup>-1</sup> is the nuclear energy release and  $\langle X_H \rangle$  is the mass-weighted mean  $X_H$  in the burning layer (C03), so that for a pure helium burst  $\Delta M_w/M_{\text{acc}} \simeq \dot{M}_w \Delta t/M_{\text{acc}} \sim (1 - L_{\text{Edd}}/L) Q_{\text{nuc}}/v_{\text{esc}} c = 0.003(1 - L_{\text{Edd}}/L)$ .

## 1.2.4 Dependence of the convective extent on burst parameters

In Figure 1.4 we show the time variation of the convective extent  $y_c$  during a burst, with the progression marked by the temperature at the base of the burning layer  $T_b$  rather than by time. We consider burst ignition models assuming both mixed hydrogen/helium accretion and nearly pure helium accretion (see Table 1.1); the latter models are similar to those obtained by CO3 in fits to the burst properties of 4U 1820-30. The burning layer is pure helium in all models except in HHeA1, HHeB1full, and HHeC3, which have  $X_H \simeq 0.4, 0.1,$  and  $0.1,$  respectively. We show results for models with  $\alpha$ -only and full reaction networks and consider a range of accretion rate, ignition column depth, metallicity, and crustal heat flux.

The evolution of  $y_c$  is sensitive to the assumed ignition conditions and to the energy production due to non- $\alpha$ -capture reactions (see § 1.2.2). For the models considered, the fraction of accreted mass lying above the convective region at its maximum extent is in the range  $10^{-4} \lesssim \Delta M/M_{\text{acc}} = y_{c,\text{min}}/y_b \lesssim 10^{-2}$ , where  $M_{\text{acc}} \simeq 4\pi R^2 y_b \approx 4 \times 10^{21}$  g ( $y_b/3 \times 10^8$  g cm<sup>-2</sup>). Thus, the possibility exists that  $y_{c,\text{min}}/y_b < \Delta M_w/M_{\text{acc}}$ . Ignition conditions that maximize the peak flux of a burst and minimize the entropy of the atmosphere during the fuel accumulation stage yield bursts with the greatest convective extents. The larger the peak flux, the faster  $T_b$  rises, and hence the shorter the growth timescale  $t_{\text{gr}}$  at a given  $T_b$ . The convective region must therefore reach out to lower column depths before  $t_{\text{th}}(y_c) = t_{\text{gr}}$ . As illustrated in Figure 1.5, such ignition conditions are best satisfied at low  $\dot{m}$ . The reason is twofold. First, for  $\dot{m} < \dot{m}_{\text{crit}}$ , the pure helium layer that develops at the base of the atmosphere yields a peak luminosity

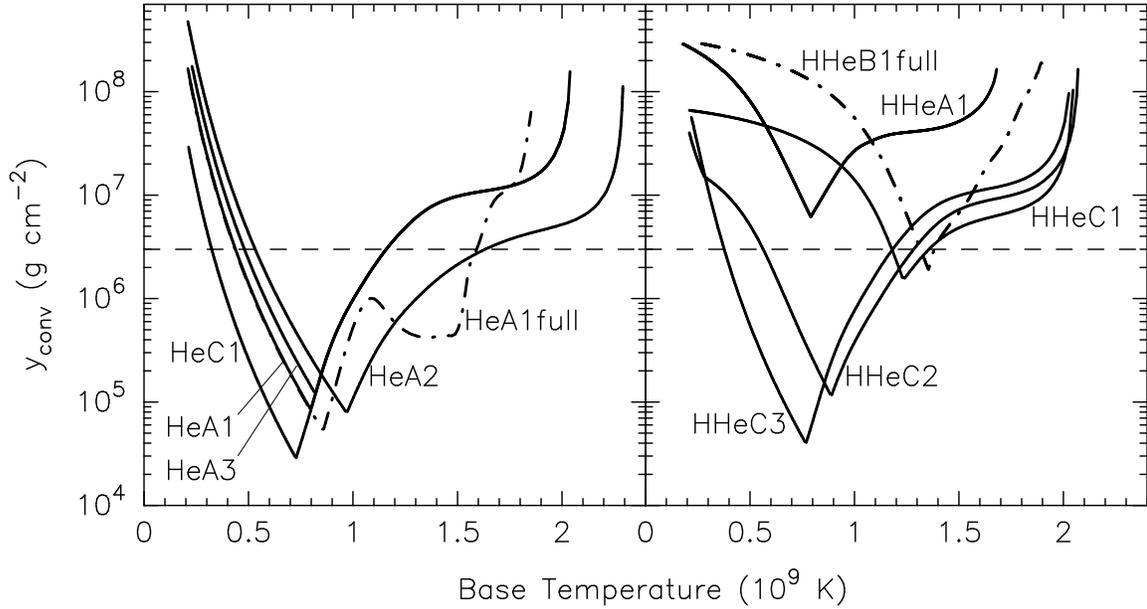


Figure 1.4 Evolution of the top of the convective zone  $y_c$  as a function of base temperature  $T_b$  for the five burst models with pure helium accreted material (*left panel*) and the five models with mixed hydrogen-helium accreted material (*right panel*) as described in Table 1.1. The full reaction network models are shown as *dash-dot lines*. The horizontal *dashed line* denotes the column depth where  $y = 0.01y_b$ , corresponding approximately to the column depth  $y_{\text{wind}}$ .

considerably greater than that of bursts with mixed hydrogen/helium burning layers. Second, because  $F_H$  and  $F_{\text{crust}}$  are both  $\propto \dot{m}$ , the lower  $\dot{m}$  the smaller the pre-burst flux and hence the lower the initial atmospheric entropy.

Since the rise time is determined by the thermal time at the convective-radiative interface, bursts with a greater convective extent and thus lower  $\dot{m}$  have shorter rise times, as shown in Figure 1.6. For accretion of helium-rich material, the rise times are nearly independent of accretion rate (the weak sensitivity is due to the dependence of  $F_{\text{crust}}$  on  $\dot{m}$ ). By contrast, for accretion of hydrogen-rich material the rise times increase by a factor of  $\sim 100$  between  $\dot{m}/\dot{m}_{\text{Edd}} = 0.02$  and  $0.2$ . In their multizone numerical simulations of X-ray bursts, Woosley et al. (2004) consider models that accrete mixed hydrogen/helium with solar abundance at  $\dot{m}/\dot{m}_{\text{Edd}} = 0.02$  and  $0.1$ . They follow several sequences of bursts for each model and obtain rise times in the range  $0.51 \times 10^{-3} - 32.1 \times 10^{-3}$  s for  $\dot{m}/\dot{m}_{\text{Edd}} = 0.02$  and  $0.51 - 0.66$  s for  $\dot{m}/\dot{m}_{\text{Edd}} = 0.1$ ,

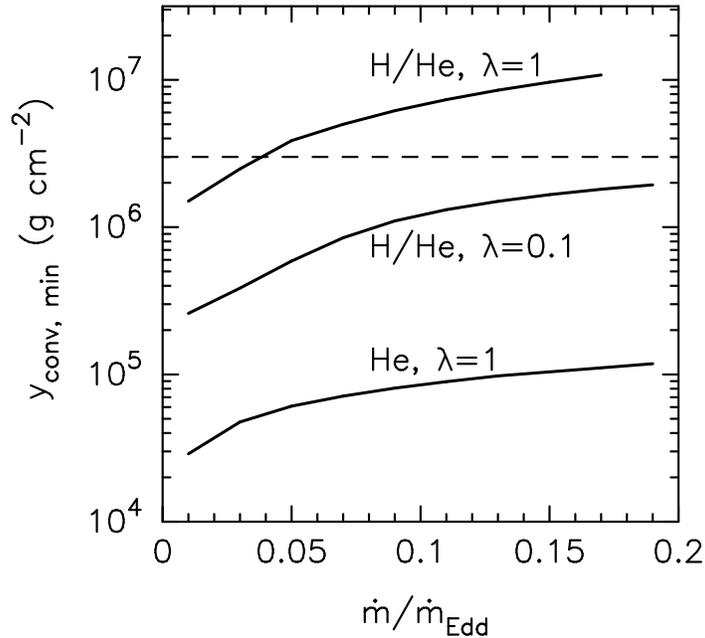


Figure 1.5 Maximum extent of the convective region  $y_{c,\min}$  as a function of  $\dot{m}$  in units of the Eddington rate. Results are shown for mixed hydrogen/helium accretion with  $F_{\text{loss}}$  pre-factor  $\lambda = 0.1$  and  $1$  and pure helium accretion with  $\lambda = 1$  assuming  $y_b = 3 \times 10^8 \text{ g cm}^{-2}$ ,  $X_0 = 0.71$ ,  $Z_{\text{CNO}} = 0.01$ ,  $Q_{\text{crust}} = 0.1 \text{ MeV nucleon}^{-1}$ . The horizontal *dashed line* denotes the column depth where  $y = 0.01y_b$ , corresponding approximately to the column depth  $y_{\text{wind}}$ .

both in very good agreement with our estimates. Our rise times are also in broad agreement with the relevant models of earlier numerical simulations (e.g., Taam 1981; Ayasli & Joss 1982; Wallace, Woosley, & Weaver 1982).

### 1.3 Detecting the nuclear burning ashes

In § 1.2 we showed that the convective zone reaches sufficiently low pressures during PRE bursts that  $y_{c,\min} \ll y_{\text{wind}}$  over a broad range of burst parameters. Thus, ashes of nuclear burning can be ejected by the radiative wind of PRE bursts. When the photosphere settles back down to the NS surface following the PRE phase, it too will be laced with heavy-element ashes. In this section we describe the composition of the ejected and exposed ashes (§ 1.3.1) and address whether these ashes can be detected. In § 1.3.2 we discuss the possibility that some of the p-nuclei observed in the solar

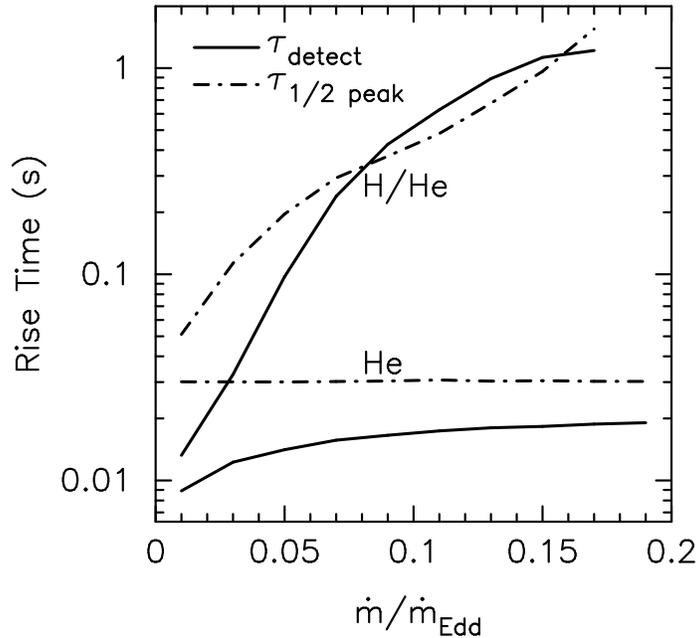


Figure 1.6 Rise time  $dt/d \ln L$  of the burst light curve as a function of  $\dot{m}$  in units of the Eddington rate. Two sets of rise times are computed for the mixed hydrogen/helium and pure helium models of Figure 1.5 with  $\lambda = 1$ :  $\tau_{\text{detect}}$  (*solid lines*) is the rise time when the luminosity first exceeds the accretion luminosity, corresponding to when an observer can first detect the burst;  $\tau_{1/2\text{peak}}$  (*dash-dot lines*) is the rise time when the luminosity is half its peak value.

system owe their origin to ash ejection during PRE bursts. In §§ 1.3.3 and 1.3.4 we determine the expected strength of spectral line features from ashes ejected in the wind and those exposed at the NS surface.

### 1.3.1 Composition of ejected and exposed ashes

Just prior to the onset of the wind, the convective zone has receded to the base and the atmosphere has a stratified compositional structure. Throughout the region  $y < y_{c,\text{min}}$ , the composition is that of the unprocessed accreted material. As we show in Figure 1.7, for  $y > y_{c,\text{min}}$ , the composition is determined by the burning stage at the moment  $y_c(t) = y$  during the convective zone’s retreat to the base. Results are shown for model HeA1, which accounts for only  $\alpha$ -capture reactions in the nuclear network. The composition at the base of the wind is approximately that

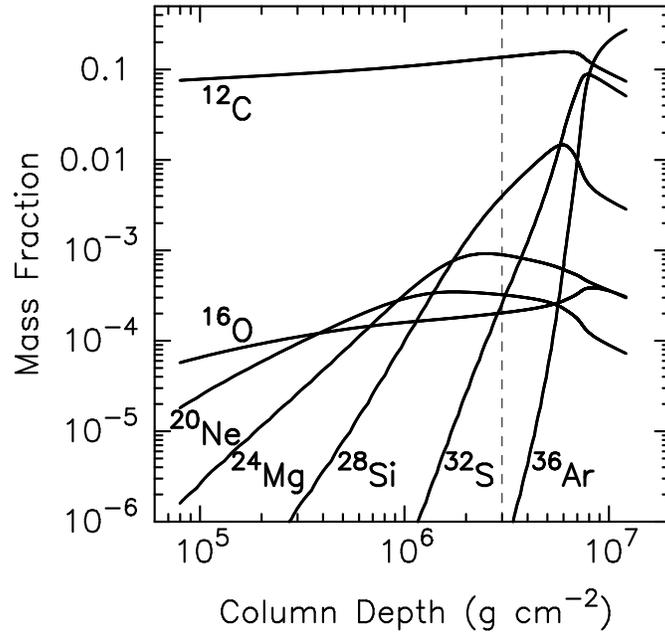


Figure 1.7 Ash composition as a function of column depth after the convective zone has completely receded (i.e.,  $y_c = y_b$ ). Results are shown for the  $\alpha$ -only reaction network model HeA1. The helium mass fraction lies above the displayed region. For  $y < y_{c,\min} \simeq 8 \times 10^4 \text{ g cm}^{-2}$ , the composition is that of the unprocessed accreted material. The vertical *dashed line* denotes  $y = 0.01y_b \simeq y_{\text{wind}}$ .

at  $y = 0.01y_b = 3 \times 10^6 \text{ g cm}^{-2}$ .

In Figure 1.8 we show the ash composition at  $y/y_b = 0.01 \simeq \Delta M_w/M_{\text{acc}}$  for models HeA1, HeA1full, and HHeB1full. While helium comprises the largest fraction of the mass in all three models, the models' overall abundance distributions differ significantly from one another. Compared to the HeA1 model, the HeA1full model has a much higher mass fraction of nuclei with  $A \sim 30$ , with significant peaks at  $^{28}\text{Si}$  and  $^{32}\text{S}$ . The difference is a result of the accelerated nuclear processing due to the bypass of the  $^{12}\text{C}(\alpha,\gamma)^{16}\text{O}$  reaction via the  $^{12}\text{C}(\text{p},\gamma)^{13}\text{N}(\alpha,\text{p})^{16}\text{O}$  reaction (see § 1.2.2). In model HHeB1full solar abundance material is accreted at a sufficiently high rate ( $\dot{m} = 0.05\dot{m}_{\text{Edd}} > \dot{m}_{\text{crit}}$ ) that when helium ignites  $X_H \simeq 0.1$  at the base. The resulting proton captures yield a substantial amount of nuclei with  $A \sim 60$  by the time  $y_c = 0.01y_b$ .

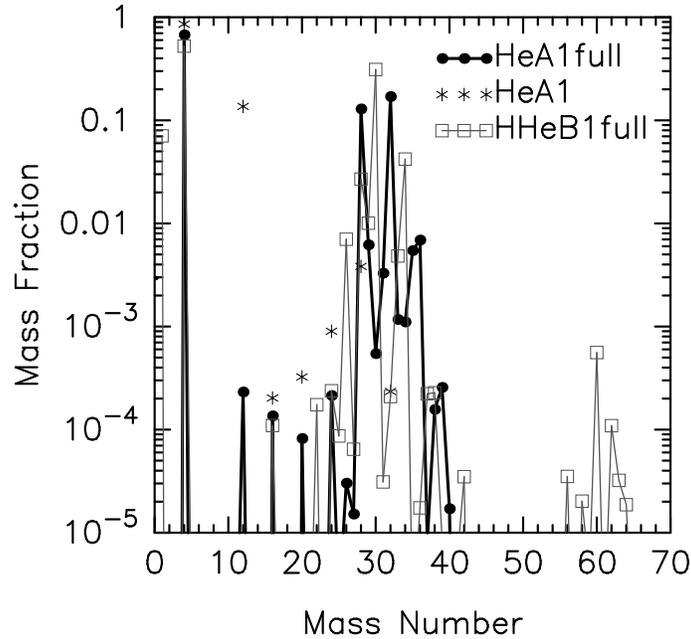


Figure 1.8 Composition of material processed during a burst at the time the convective zone is located at  $y_c = 0.01y_b \simeq y_{\text{wind}}$  and is receding to higher pressures. Results are shown for the pure helium accretion model HeA1, which accounts for only the  $\alpha$ -capture processes, the pure helium accretion model HeA1full and the mixed hydrogen/helium accretion model HHeB1full, which account for both the  $\alpha$ -capture processes and proton-capture processes (see § 1.2.2).

### 1.3.2 Ejection of p-nuclei

In their numerical simulations of X-ray bursts, Woosley et al. (2004) find that bursts ignite in the ashes of previous ones. These endpoint ashes are heavier than those processed during the ongoing burst at the time  $y_c/y_b = 0.01$  and are also likely mixed throughout the convective region (though we do not show these ashes in Figures 1.7 and 1.8). Using a reaction network that extends up to Xe, Schatz et al. (2001) find that for  $\dot{m} = 0.3\dot{m}_{\text{Edd}}$  and accreted material with solar abundance (i.e., burning in a hydrogen-rich environment), the endpoint of the rp process burning is a closed SnSbTe cycle that naturally limits rp process nucleosynthesis to light p nuclei. They find overproduction factors (relative to solar abundance) of  $\sim 10^8$  for the p nuclei  $^{92}\text{Mo}$  and  $^{96}\text{Ru}$  and  $\sim 10^9$  for the p nucleus  $^{98}\text{Ru}$  (see also Schatz et al. 1998). Standard p-process scenarios are unable to adequately explain the observed solar system abundances of

these p-nuclei (for a review, see Wallerstein et al. 1997).

Whether ash ejection during PRE bursts can account for the observed solar system abundances depends on the amount of p-process material ejected into the interstellar medium by a PRE burst and the event rate of such bursts over the Galaxy lifetime. In order to produce large amounts of p-nuclei, the burning layer must be hydrogen-rich ( $\dot{m} \sim 0.3\dot{m}_{\text{Edd}}$ ) while PRE bursts require helium-rich burning layers and thus  $\dot{m} \lesssim 0.05\dot{m}_{\text{Edd}}$ . However, this does not preclude PRE bursts from ejecting the nuclei, as accretion rates in bursting low-mass X-ray binaries are observed to vary by factors of a few, with individual systems undergoing transitions from hydrogen-rich to helium-rich burning over year timescales (as evidenced by variations in burst duration and peak fluxes; see e.g., Cornelisse et al. 2003). Thus, a system undergoing a PRE burst may ignite in ashes rich with p-nuclei. To determine the fractional amount  $\eta$  of ashes that must be ejected in order to account for the observed solar system abundance of p-nuclei, assume a p-nuclei overproduction factor  $\xi = 10^9$  and a galaxy disk mass  $M_{\text{disk}} = 4 \times 10^{10}$  (Klypin et al., 2002). Currently, there are  $\sim 10$  active X-ray burst systems at  $\dot{M} \sim 10^{-9}M_{\odot} \text{ yr}^{-1}$  and  $\sim 100$  at  $\dot{M} \sim 10^{-10}M_{\odot} \text{ yr}^{-1}$  (Lewin, van Paradijs, & Taam, 1995). If we assume this is representative of the population count over the galaxy lifetime, then the total amount of mass accreted by all PRE burst systems over  $10^{10}$  yr is  $M_{\text{acc,tot}} \sim 100M_{\odot}$ . Then  $\eta = 0.4(M_{\text{disk}}/4 \times 10^{10}M_{\odot})(\xi/10^9)^{-1}(M_{\text{acc,tot}}/100M_{\odot})^{-1}$ . Thus,  $\eta$  is a factor of  $\sim 10 - 100$  too high given that only  $\sim 1\%$  of all accreted matter is ejected. The discrepancy can be overcome if, for example, the Galactic distribution of p-nuclei is inhomogeneous (i.e., the solar abundance is higher than the Galactic mean by a factor of  $\sim 10 - 100$ ) or the population of bursting systems was larger in the past. Assessing the plausibility of either of these scenarios is beyond the scope of this paper.

### 1.3.3 Spectral edges in wind outflow

During the PRE phase, an optically thick, transonic, radiation-driven wind forms. The sonic point of the wind lies  $10 - 100$  km above the NS surface and the photo-

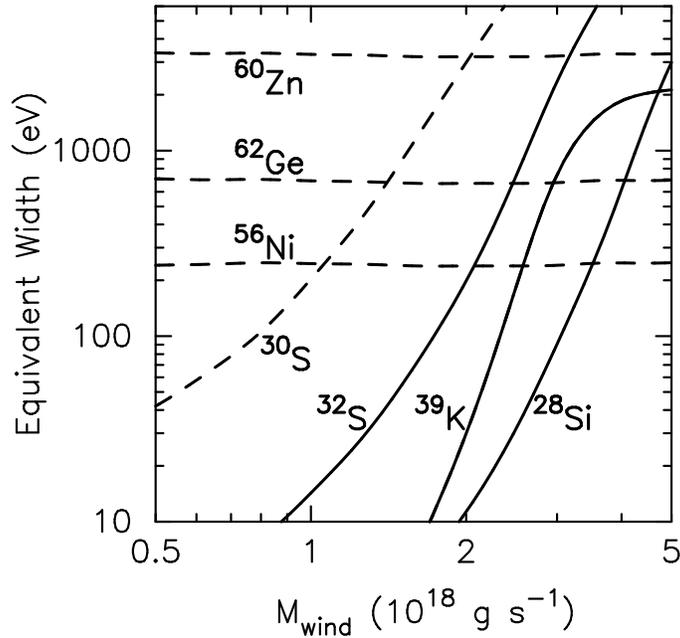


Figure 1.9 Equivalent width of the photoionization edge due to ashes ejected in a wind as a function of the mass outflow rate for the PRE burst models HeA1full (*solid lines*) and HHeB1full (*dashed lines*).

sphere, defined as the location where the effective optical depth  $\tau_* \equiv \kappa \rho r$  is near unity, is a factor of  $\sim 10$  farther out (Paczynski & Proszynski, 1986; Joss & Melia, 1987; Nobili, Turolla, & Lapidus, 1994). Once matter reaches the sonic point it is essentially unbound from the NS, and is ejected to infinity. The ejected matter cools during the PRE phase allowing some heavy element ashes to bind with one or more electrons. Given that the luminosity always remains near Eddington at the photosphere and  $r_{\text{ph}} \sim 100$  km, the effective temperature at the photosphere is  $T_{\text{eff}} \sim 10^6 - 10^7$  K. The recombinations are fast compared to the flow timescale (see below), and the fraction of heavy ions in a hydrogen-like state is of order unity by Saha equilibrium. The total column density of hydrogen-like ions above the photosphere is thus  $N_{\text{wind}} \sim f(A, Z)N_e \simeq 10^{20} \text{ cm}^{-2}(f/10^{-4})$  where  $N_e \approx \sigma_{\text{Th}}^{-1}$  is the electron column density and  $f(A, Z)$  is the abundance by number of element  $Z$  with mass number  $A$ . This corresponds to an optical depth to a photoionization edge due to hydrogen-like ions of  $\tau \approx N_{\text{wind}}\sigma_{\text{bf}} \sim 1$ , where the bound-free cross section

$\sigma_{\text{bf}}(E) = 6.32 \times 10^{-18} \text{ cm}^2 (E_e/E)^3 Z^{-2}$  and  $E_e \simeq 13.6Z^2 \text{ eV}$  is the edge energy (Rybicki & Lightman, 1979). It is worth noting that the effective temperature is low enough at the photosphere that even at solar abundances  $\tau \simeq 1$ . That the ejected ashes have abundances much larger than solar only improves the likelihood of detection.

We now calculate  $N_{\text{wind}}$  more exactly and determine the resulting equivalent width (EW) of the photoionization edge. We leave the calculation of the H $\alpha$  EW to future work. The column density of hydrogen-like ions is given by

$$N_{\text{wind}} = \frac{f(A, Z)}{m_p} \int_{r_{\text{ph}}}^{\infty} \zeta [T(r), \rho(r)] \rho(r) dr \quad (1.11)$$

where  $\zeta$  is the fraction of that species in the hydrogen-like state at a given temperature and density from Saha equilibrium. We use the results of Paczynski & Proszynski (1986), who calculate general relativistic models of radiation-driven winds from NSs, to obtain values for  $r_{\text{ph}}$ ,  $T(r_{\text{ph}})$ , and  $\rho(r_{\text{ph}})$  as a function of the mass outflow rate  $\dot{M}_w$ . For example, at  $\dot{M}_w = 10^{18} \text{ g s}^{-1}$ ,  $r_{\text{ph}} = 3 \times 10^7 \text{ cm}$ ,  $T_{\text{ph}} = 5 \times 10^6 \text{ K}$ , and  $\rho_{\text{ph}} = 5 \times 10^{-7} \text{ g cm}^{-3}$ . Over the relevant range of  $\dot{M}_w$ , the wind is quasi-static in that the flow timescale  $t_{\text{flow}} \sim r/v \sim 0.1 \text{ s}$  is shorter than the characteristic timescale of the wind  $t_w = \Delta M_w / \dot{M}_w \sim 1 \text{ s}$ , where  $v(r)$  is the fluid velocity at  $r$  (see e.g., Joss & Melia 1987). Thus, the equation of continuity gives  $\dot{M}_w \simeq 4\pi r^2 \rho v$ . Paczynski & Proszynski (1986) show that the velocity is nearly constant beyond the photosphere, i.e.,  $v(r > r_{\text{ph}}) \simeq v_{\text{ph}} \sim 10^8 \text{ cm s}^{-1}$ , so that  $\rho(r > r_{\text{ph}}) \approx \dot{M}_w / 4\pi r^2 v_{\text{ph}}$ . The gas above the photosphere is Compton-heated by the hot photons originating in the photosphere. Since the Compton heating timescale  $t_c \sim kT m_e c^2 / E_\gamma F \sigma_{\text{Th}} \sim 10^{-7} \text{ s} \ll t_{\text{flow}}$ , where  $E_\gamma$  is the photon energy and  $F$  the flux, the gas temperature is nearly constant out to radii well above the photosphere (Joss & Melia, 1987; Nobili, Turolla, & Lapidus, 1994). For  $r \gg r_{\text{ph}}$ ,  $t_c \lesssim t_{\text{flow}}$  and the gas cools adiabatically, though the density in this region is so low that in calculating  $N_{\text{wind}}$  we assume  $T(r > r_{\text{ph}}) = T_{\text{ph}}$ . For the more abundant species shown in Figure 1.8, we obtain column densities in the range  $N_{\text{wind}} \sim 10^{16} - 10^{21} \text{ cm}^{-2}$ .

To determine the EW of the photoionization edge we integrate over the optical depth above the edge,  $\tau_e = N_{\text{wind}}\sigma_{\text{bf}}(E)$ . Since the recombination timescale  $t_{\text{rec}} \equiv 1/n_e\langle\sigma_{\text{fb}}v\rangle \sim 0.01Z^{-4}\rho_{-6}^{-1}T_7^{-3/2}B(Z, T)$  s  $\ll t_{\text{flow}}$  where  $\rho_{-6} = \rho/10^{-6}$  g cm $^{-3}$ ,  $T_7 = T/10^7$  K,  $B(Z, T) = \exp(x) \int_x^\infty \exp(-t) d \ln t$ ,  $x \equiv E_e/kT$ , and  $\sigma_{\text{fb}}$  is related to  $\sigma_{\text{bf}}$  through the Milne relation (Rybicki & Lightman, 1979), the recombinations are nearly instantaneous on the timescale of the wind flow. Thus, assuming an effectively cold atmosphere,  $\text{EW}_e = \int \{1 - \exp[-N_{\text{wind}}\sigma_{\text{bf}}(E)]\} dE$  (Bildsten et al., 2003). In Figure 1.9 we show  $\text{EW}_e$  as a function of  $\dot{M}_w$  for some of the ejected ashes of models HeA1full and HHeB1full. As  $\dot{M}_w$  increases,  $T_{\text{ph}}$  decreases and the hydrogen-like fraction  $\zeta$  increases. For the heaviest ashes (e.g.,  $^{56}\text{Ni}$ ,  $^{60}\text{Zn}$ , and  $^{62}\text{Ge}$ )  $\zeta \sim 1$  even at low  $\dot{M}_w$  so that  $N_{\text{wind}}(Z \sim 30)$  and  $\text{EW}_e(Z \sim 30)$  are essentially set by  $f(A, Z)$ , as noted above. The computed values of  $\text{EW}_e$  are within the range accessible by current high resolution X-ray telescopes.

### 1.3.4 Spectral edges from NS surface

Following the burst peak, the atmosphere begins to cool, the flux becomes sub-Eddington, and the wind turns off. The photosphere, still laced with heavy-element ashes, settles back down to the NS surface. As in the case of line formation in the wind outflow, whether these surface ashes can be detected depends on the column density  $N_{\text{surf}}$  of ashes that are not fully ionized. Here  $N_{\text{surf}} \simeq f(A, Z)\zeta(T, \rho)y_{\text{ph}}/m_p$ , where  $y_{\text{ph}} \approx 1$  g cm $^{-2}$ . As the surface temperature  $T$  decreases,  $N_{\text{surf}}$  increases, although lines cannot be detected once  $L = 4\pi R^2\sigma T^4 \lesssim L_{\text{acc}}$ . The decay timescale of PRE bursts (i.e., the time during which  $L_{\text{acc}} < L < L_{\text{Edd}}$ ) is typically around  $\sim 10$  s. The downward drift speed of a nucleus with  $Z \sim 30$  in a pure H atmosphere is  $v \approx 1$  cm s $^{-1}T_7^{3/2}/\rho$  so that it takes  $t_s \approx 1$  s  $(y/1 \text{ g cm}^{-2})/T_7^{3/2}$  for such a nucleus to fall a column depth  $y$  (Bildsten et al., 1992, 2003). Thus, the residence time of ashes in the photosphere also limits the detectability of lines as will high NS rotation rates (Özel & Psaltis, 2003; Bhattacharyya et al., 2004; Chang et al., 2005).

In Figure 1.10 we plot  $\text{EW}_e = \int \{1 - \exp[-N_{\text{surf}}\sigma_{\text{bf}}(E)]\} dE$  as a function of  $L/L_{\text{acc}}$

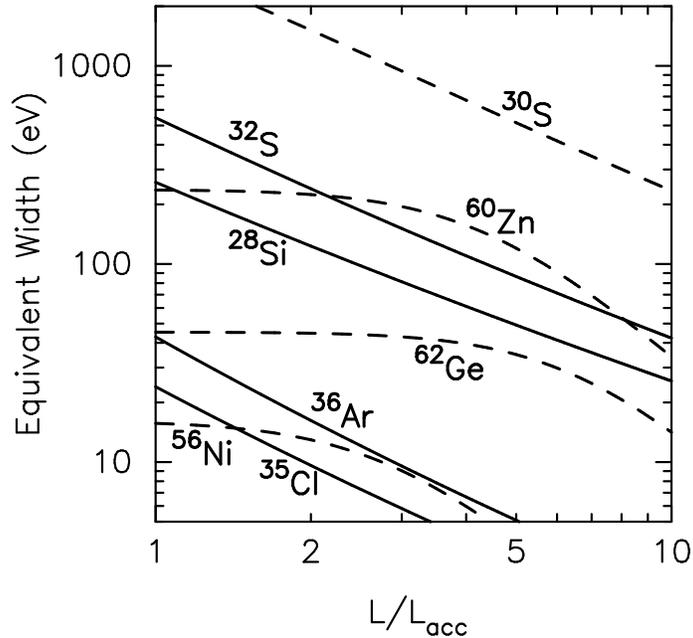


Figure 1.10 Equivalent width of the photoionization edge due to ashes residing in the photosphere at the NS surface for models HeA1full (*solid lines*) and HHeB1full (*dashed lines*). The EW is plotted as a function of the ratio of surface luminosity to accretion luminosity for a cooling NS atmosphere following the PRE phase. Detecting such lines might allow for a measurement of the NS gravitational redshift.

for the same models as Figure 1.9. For reference,  $L_{\text{Edd}}/L_{\text{acc}} \simeq 10$  and 30 for models HeA1full and HHeB1full, respectively. The values of  $\text{EW}_e$  are again within the range accessible by current X-ray telescopes and a measurement of the NS gravitational redshift may be possible.

## 1.4 Summary and conclusions

We have shown that during a radius expansion Type I X-ray burst the convective region of the atmosphere reaches sufficiently low pressures that ashes of nuclear burning, mixed throughout the convective zone, can be ejected by the burst's radiative wind. Specifically, we used an analytic prescription to solve for the evolution of the atmosphere's thermal structure and found that systems that accrete pure helium, such as 4U 1820-30, and those that accrete mixed hydrogen and helium at  $\dot{m} \lesssim 0.05\dot{m}_{\text{Edd}}$ ,

have convective zones that reach minimum column depths  $y_{c,\min} \ll y_{\text{wind}} \simeq 0.01y_b$ . Previous studies have also found that the convective zone can reach pressures  $\lesssim 1\%$  that at the base for low  $\dot{m}$  bursts. However, these studies focused on numerical simulations of bursts that were unable to resolve the low-pressure regions reached by the convective zone. Furthermore, such studies only explored a limited range of burst parameter space while our analytic treatment enabled us to carry out a survey of the dependence of  $y_{c,\min}$  on a variety of burst parameters such as  $\dot{m}$ , the composition of accreted matter, and the pre-burst thermal state of the atmosphere. We have compared some of the results of our analysis, such as the burst rise times, to those of numerical simulations and found excellent agreement.

We did not account for the affect of a laterally spreading burning front during burst rise nor the influence of rotation, though these may significantly alter the convective structure. We also did not account for the suppression of convective motions by a magnetic field even though the magnetic energy density  $B^2/8\pi$  can be comparable to the convective energy density  $\rho v_{\text{conv}}^2$  at low pressures. Such effects might alter the conditions under which ashes are ejected.

For specific burst models we determined the composition of ejected ashes and calculated the expected column density of hydrogen-like nuclei using models of relativistic radiation-driven winds. We then computed the EW of the photoionization edge for the most abundant hydrogen-like nuclei in the wind. We carried out a similar procedure to determine the EW for those ashes that remain bound to the NS and thus reside in the photosphere after it settles back down to the NS surface. We found EWs in the range 10 – 1000 eV suggesting that the line features can be detected with current high resolution X-ray telescopes. Detecting the lines would directly probe the nuclear burning. Those lines formed at the surface may allow for a measurement of the gravitational redshift and thus help constrain the NS equation of state.

If bursts ignite in the ashes of previous bursts then some of these old ashes, which are thought to contain large overabundances of p-nuclei relative to solar, are also ejected in the wind. We showed that at least  $\sim 1 - 10\%$  of the p-nuclei observed in the solar system may originate in X-ray bursts.

Although low  $\dot{m}$  systems yield highly extended convective regions, their burst recurrence times are long and often irregular making them difficult targets to monitor given a narrow window of observing time. More promising are systems in which the neutron star accretes helium-rich material ( $X_{\text{He}} \sim 0.9 - 1$ ) from an evolved companion, such as a cold helium white dwarf. The ultracompact binary 4U 1820-30 ( $P_{\text{orb}} = 11.4$  min; Stella, Priedhorsky, & White 1987) is thought to reside in such a system (Cumming 2003, and references therein). In particular, the system exhibits radius expansion bursts with fairly regular burst recurrence times of only a few hours despite accreting at  $\sim 0.1\dot{m}_{\text{Edd}}$  (see e.g., Cornelisse et al. 2003, Kuulkers et al. 2003). The interpretation of this system as a helium accretor is supported by its ultracompact nature and the bursts' fast rise and decay times and  $\alpha$ -values.

Observations of four candidate ultracompact binaries have shown an unusual Ne/O ratio in the absorption along the line of sight, with number ratios several times the interstellar medium (ISM) value (Juett et al., 2001; Juett & Chakrabarty, 2003). Two of the systems, 4U 1543-624 and 4U 1850-087, have shown variations in the Ne/O ratio in follow-up observations (Juett & Chakrabarty, 2003, 2005), suggesting either variations in the ionization state of the Ne and O or variations in the intrinsic abundances. Both possibilities imply the absorption is from material local to the binaries. One explanation for the unusual ratios is that the intrinsic abundance of Ne and O is the same as the ISM but the O is in a higher ionization state than Ne, leading to an apparent enhancement of the Ne/O ratio. Another possibility is that the degenerate donors in these ultracompact binaries are neon-rich, perhaps the chemically fractionated cores of C-O-Ne or O-Ne-Mg white dwarfs which have previously crystallized (Schulz et al., 2001; Juett & Chakrabarty, 2003). A third possibility we now propose is that the donors in these ultracompact systems are He white dwarfs, as in 4U 1820-30, and that the accretion of helium rich matter results in PRE bursts and ejection of ashes with highly non-solar abundances of Ne and O. A radius expansion burst was indeed seen in one of the systems with an unusual Ne/O ratio (2S 0918-549; Cornelisse et al. 2002). Though only a single realization, the Ne/O ratio by number at  $y = 0.01y_b$  for model HeA1 shown in Figure 1.7 is  $\sim 1.2$ ,

comparable to that seen in the four systems. In order for this explanation to work, the ejected ashes must either somehow remain in the environment of the binary or be continuously replenished by periodic PRE bursts. The variations in the observed Ne/O ratio in observations of 4U 1543-624 and 4U 1850-087 could just reflect the time since the last PRE burst.

Finally, we note that ash ejection may also occur during superbursts that undergo photospheric radius expansion. The amount of mass ejected by superbursts may be much larger than ordinary PRE bursts, suggesting even higher column densities of ejected ashes.

# Bibliography

Ayasli, S. & Joss, P. C. 1982, *ApJ*, 256, 637

Bhattacharyya, S., Miller, M. C., & Lamb, F. K. 2004 (astro-ph/0412107)

Bildsten, L., Salpeter, E. E., & Wasserman, I. 1992, *ApJ*, 384, 143

Bildsten, L. 1998, in *The Many Faces of Neutron Stars*, ed. R. Buccheri et al. (Dordrecht: Kluwer), 419

Bildsten, L., Chang, P., & Paerels, F. 2003, *ApJL*, 591, L29

Brown, B. A., Clement, R. R. C., Schatz, H., & Volya, A. 2002, *Phys. Rev. C*, 65, 045802

Brown, E. F. & Bildsten, L. 1998, *ApJ*, 496, 915

Brown, E. F. 2000, *ApJ*, 531, 988

Brown, E. F. 2004, *ApJL*, 614, L57

Chang, P., Bildsten, L., & Wasserman, I. 2005 (astro-ph/0505062)

Cornelisse, R., et al. 2002, *A&A*, 392, 885

Cornelisse, R., et al. 2003, *A&A*, 405, 1033

Cumming, A. & Bildsten, L. 2000, *ApJ*, 544, 453 (CB00)

Cumming, A. 2003, *ApJ*, 595, 1077 (C03)

Fujimoto, M. Y., Hanawa, T., & Miyaji, S. 1981, *ApJ*, 247, 267

- Fushiki, I. & Lamb, D. Q. 1987, *ApJL*, 323, L55
- Galloway, D. K., Psaltis, D., Chakrabarty, D., & Munro, M. P. 2003, *ApJ*, 590, 999
- Hanawa, T. & Sugimoto, D. 1982, *Publ. Astron. Soc. Japan*, 34, 1
- Hanawa, T. & Fujimoto, M. Y. 1984, *Publ. Astron. Soc. Japan*, 36, 199
- Hansen, C. J. & Kawaler, S. D. 1994, *Stellar Interiors: Physical Principles, Structure, and Evolution*, (Springer-Verlag)
- Joss, P. C. 1977, *Nature*, 270, 310
- Joss, P. C. 1978, *ApJL*, 225, L123
- Joss, P. C. & Melia, F. 1987, *ApJ*, 312, 700
- Juett, A. M., Psaltis, D., & Chakrabarty, D. 2001, *ApJL*, 560, L59
- Juett, A. M., & Chakrabarty, D. 2003, *ApJ*, 599, 498
- Juett, A. M., & Chakrabarty, D. 2005, (astro-ph/0501472)
- Klypin, A., Zhao, H., & Somerville, R. S. 2002, *ApJ*, 573, 597
- Kuulkers, E., den Hartog, P. R., in't Zand, J. J. M., Verbunt, F. W. M., Harris, W. E., & Cocchi, M. 2003, *A&A*, 399, 663
- Lewin, W. H. G., van Paradijs, J., & Taam, R. E. 1995, in *X-ray Binaries*, ed. W. H. G. Lewin et al. (Cambridge: Cambridge Univ. Press), 175
- Nobili, L., Turolla, R., & Lapidus, I. 1994, *ApJ*, 433, 276
- Özel, F., & Psaltis, D. 2003, *ApJL*, 582, L31
- Paczynski, B. 1983, *ApJ*, 267, 315
- Paczynski, B. & Proszynski, M. 1986, *ApJ*, 302, 519

- Rybicki, G. B., & Lightman, A. P. 1979, *Radiative Processes in Astrophysics* (New York:Wiley)
- Schatz, H., et al. 1998, *Phys. Rep.*, 294, 167
- Schatz, H., Bildsten, L., Cumming, A., & Wiescher, M. 1999, *ApJ*, 524, 1014
- Schatz, H., et al. 2001, *Physical Review Letters*, 86, 3471
- Schulz, N. S., Chakrabarty, D., Marshall, H. L., Canizares, C. R., Lee, J. C., & Houck, J. 2001, *ApJ*, 563, 941
- Spitkovsky, A., Levin, Y., & Ushomirsky, G. 2002, *ApJ*, 566, 1018
- Stella, L., Priedhorsky, W., & White, N. E. 1987, *ApJL*, 312, L17
- Strohmayer, T., & Bildsten, L. 2004, in *Compact Stellar X-Ray Sources*, ed. W. H. G. Lewin & M. van der Klis (Cambridge: Cambridge Univ. Press), in press (astro-ph/0301544)
- Taam, R. E. 1980, *ApJ*, 241, 358
- Taam, R. E. 1981, *ApJ*, 247, 257
- Wallace, R. K., Woosley, S. E., & Weaver, T. A. 1982, *ApJ*, 258, 696
- Wallerstein, G., et al. 1997, *Reviews of Modern Physics*, 69, 995
- Woosley, S. E. & Weaver, T. A. 1984, *High Energy Transients in Astrophysics*, 273
- Woosley, S. E., et al. 2004, *ApJS*, 151, 75
- Yahel, R. Z., Brinkmann, W., & Braun, A. 1984, *A&A*, 139, 359
- Zingale, M., et al. 2003, *Astronomical Society of the Pacific Conference Series*, 293, 329

## Part II

# Stellar Dynamics at the Galactic Center

## Chapter 2

# Stellar Dynamics at the Galactic Center with an Extremely Large Telescope<sup>†</sup>

NEVIN N. WEINBERG, MILOŠ MILOSAVLJEVIĆ

Theoretical Astrophysics, California Institute of Technology, Pasadena, CA 91125

ANDREA M. GHEZ

Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1562

---

<sup>†</sup> A version of this chapter was first published in *The Astrophysical Journal*, 622, p. 878–891 (2005).

# Abstract

We discuss physical experiments achievable via the monitoring of stellar dynamics near the massive black hole at the Galactic center with a diffraction-limited, next-generation, extremely large telescope (ELT). Given the likely observational capabilities of an ELT and what is currently known about the stellar environment at the Galactic center, we synthesize plausible samples of stellar orbits around the black hole. We use the Markov Chain Monte Carlo method to evaluate the constraints that the monitoring of these orbits will place on the matter content within the dynamical sphere of influence of the black hole. We compare these future constraints with those obtained with the current data. We express our results as functions of the number  $N$  of stars with detectable orbital motions and the astrometric precision  $\delta\theta$  and spectroscopic precision  $\delta v$  at which the stellar proper motions and radial velocities are monitored. Our results are easily scaled to different telescope sizes and precisions. For  $N = 100$ ,  $\delta\theta = 0.5$  mas, and  $\delta v = 10$  km s<sup>-1</sup>—a conservative estimate of the capabilities of a 30 meter telescope—we find that if the extended matter distribution enclosed by the orbits at 0.01 pc has a mass greater than  $\sim 10^3 M_\odot$ , it will produce measurable deviations from Keplerian motion. Thus, if the concentration of dark matter at the Galactic center matches theoretical predictions, its influence on the orbits will be detectable. We also estimate the constraints that will be placed on the mass of the black hole and on the distance to the Galactic center, and find that both will be measured to better than  $\sim 0.1\%$ . We discuss the significance of knowing the distance to within a few parsecs and the importance of this parameter for understanding the structure of the Galaxy. We demonstrate that the lowest-order relativistic effects, such as the prograde precession, will be detectable if  $\delta\theta \lesssim 0.5$  mas. Barring the favorable discovery of a star on a highly compact, eccentric orbit, the higher-order effects, including the frame dragging due to the spin of the black hole, will require  $\delta\theta \lesssim 0.05$  mas. Finally, we calculate the rate at which monitored stars experience detectable nearby encounters with background stars. The encounters probe the mass function of stellar remnants that accumulate near the black hole. We find that  $\sim 30$  such encounters will be detected over a ten year baseline for  $\delta\theta = 0.5$  mas.

---

## 2.1 Introduction

Observational programs with ten meter class telescopes, including the W. M. Keck Observatory and the Very Large Telescope (VLT), have yielded a wealth of information on the stellar content inside the sphere of influence of the massive black hole at

the Galactic center (GC; Ghez et al. 1998; Gezari et al. 2002; Hornstein et al. 2002; Figer et al. 2003; Genzel et al. 2003; Ghez et al. 2005; Schödel et al. 2003). The black hole is located at the center of a compact stellar cluster that has been the target of observational surveys for a decade (e.g., Krabbe et al. 1995; Figer et al. 2000; Gezari et al. 2002). Near-infrared monitoring with speckle and adaptive optics techniques has recently enabled complete orbital reconstruction of several stellar sources orbiting the black hole (Eckart et al., 2002; Schödel et al., 2002, 2003; Ghez et al., 2005). Sources have been monitored with astrometric errors of a few milli-arcseconds (Ghez et al., 2003; Schödel et al., 2003), and radial velocity errors  $< 50 \text{ km s}^{-1}$  (Eisenhauer et al., 2003; Ghez et al., 2003), allowing the detection of the accelerated proper motions of  $\sim 10$  stars. One of these stars has an orbital period of only  $\sim 15 \text{ yr}$  (Ghez et al., 2005; Schödel et al., 2003).

The presence of a dark mass at the center of the Galaxy could in principle be inferred from the static nature of the radio source Sgr A\* located at the center of the stellar cluster (Backer & Sramek, 1999; Reid et al., 1999). Nevertheless, it is the stars with the shortest orbital periods that have provided unequivocal proof of the existence of a massive black hole and a measurement of its mass of  $\sim 4 \times 10^6 M_\odot$  (Ghez et al., 2005; Schödel et al., 2003). Since, for a fixed angular scale, the orbital periods are proportional to  $R_0^{3/2} M_{\text{bh}}^{-1/2}$  and the radial velocities are proportional to  $R_0^{-1/2} M_{\text{bh}}^{1/2}$  where  $R_0$  is the heliocentric distance to the black hole and  $M_{\text{bh}}$  is its mass, the two parameters are not degenerate and can be determined independently (see Eisenhauer et al. 2003 and appendix to this chapter).

In spite of the quality of elementary data available about the black hole and the bright stellar sources, the matter content in the vicinity of the black hole remains unknown. The observed stellar sources may represent only a fraction of the total matter content. Since the radial diffusion time  $\sim 10^{8-9} \text{ yr}$  is shorter than the age of the bulge, a large number of massive compact remnants ( $5 - 10 M_\odot$  black holes) could have segregated into, and may dominate the matter density inside, the dynamical sphere of influence of the black hole (Morris, 1993; Miralda-Escudé & Gould, 2000). Furthermore, adiabatic growth of the massive black hole could have compressed a

pre-existing distribution of cold dark matter (CDM) (Ipser & Sikivie, 1987; Quinlan, Hernquist, & Sigurdsson, 1995; Gondolo & Silk, 1999) and stars (Peebles, 1972; Young, 1980) into a dense “spike.” A variety of dynamical processes, however, are capable of destroying such a spike (Ullio, Zhao, & Kamionkowski, 2001; Merritt et al., 2002; Gnedin & Primack, 2003; Merritt, 2003). A sustained CDM spike would have implications for the detection of annihilation radiation for the CDM models in which the CDM consists of weakly interacting massive particles (WIMPs).

The most complete catalogue of stars in the central parsecs was compiled by Genzel et al. (2000) and Schödel et al. (2003). In a survey of the stellar sources, Genzel et al. (2003) infer a spatial number density of  $n(r) \propto r^{-1.4}$  over the radial range  $0.004 < r < 0.4$  pc. Their sample was 50% complete for stars brighter than  $K \sim 18$ , where completeness is defined as the percentage of stars in the field of view that are detectable and thus included in the sample. Expressed in terms of stellar mass, the sample is 50% complete for masses  $m \gtrsim 3M_{\odot}$ , assuming stars on the main sequence, a distance to the GC of 8.0 kpc (Reid, 1993) and  $K$ -band extinction of 3.3 mag (Rieke, Rieke, & Paul, 1989). A picture is emerging in which the brightest stars in the central cluster ( $< 0.03$  pc) are young, main-sequence stars with apparent magnitudes  $K > 13$  and masses  $10 - 15M_{\odot}$ . The stars outside 0.03 pc appear to be spectroscopically and kinematically distinct. They span a larger range of magnitudes  $K \gtrsim 10$  and contain  $\sim 40$  mass-losing Wolf-Rayet stars (e.g., Genzel et al. 2003 and R. Genzel, private communication). Unlike the central cluster, these stars appear to belong to twin, misaligned stellar disks (Levin & Beloborodov, 2003; Genzel et al., 2003).

The formation of the observed young stars with  $\times 100$  larger specific binding energies relative to the black hole than that of the nearest observed accumulation of molecular gas (e.g., Jackson et al. 1993) presents a challenge to star formation theories and is a persistent puzzle (e.g., Morris 1993; Ghez et al. 2003; Genzel et al. 2003). A number of mechanisms for the formation and migration of stars in the tidal field of the massive black hole have been proposed (Gerhard, 2001; Gould & Quillen, 2003; Hansen & Milosavljević, 2003; Levin & Beloborodov, 2003; Kim &

Morris, 2003; Milosavljević & Loeb, 2004). While the mechanisms have important implications, they are also each deficient in at least one way.

There is a dearth of giants in the GC region (Eckart et al., 1995). Recently, Figer et al. (2003) measured the radial velocities of 85 cool, normal giant stars with projected distances from the central region between 0.1 – 1 pc. They find nearly complete deficiency of giants with large radial velocities ( $V_{\text{rad}} > 200 \text{ km s}^{-1}$ ). Since a star in a circular orbit at a distance of 0.1 pc from the black hole has velocity  $\sim 400 \text{ km s}^{-1}$ , the absence of any such stars with comparable radial velocities indicates that the observed giants are indeed limited to the region outside the central  $\sim 0.5 \text{ pc}$ .

While the measured stellar density profile of the Galactic bulge is consistent with that of a singular isothermal sphere (Becklin, 1968), the profile in the central parsec is not well-known, especially for the lower-mass stellar populations. Assuming relaxation that is driven by two-body processes, Bahcall & Wolf (1976) showed that the equilibrium phase space distribution for a population of equal mass stars is a power law in density  $\rho \propto r^{-7/4}$ . For a multimass distribution the lighter stars are less centrally concentrated, resulting in a power-law profile that ranges from  $r^{-3/2}$  for the least massive species to  $r^{-7/4}$  for the most (Bahcall & Wolf, 1977; Murphy, Cohn, & Durisen, 1991). A coeval family of stars in the central region has reached equilibrium only if it is older than the relaxation time

$$\begin{aligned}
 t_E &\sim \frac{\sigma^3}{G^2 m_\star \rho \ln \Lambda} \\
 &\approx 2 \times 10^8 \text{ yr} \left( \frac{r}{1 \text{ pc}} \right)^{1/4} \left( \frac{M_{\text{bh}}}{4 \times 10^6 M_\odot} \right)^{3/2} \\
 &\quad \times \left( \frac{m_\star}{10 M_\odot} \right)^{-1} \left( \frac{\rho_{1\text{pc}}}{2 \times 10^5 M_\odot \text{ pc}^{-3}} \right)^{-1} \left( \frac{\ln \Lambda}{10} \right)^{-1}
 \end{aligned} \tag{2.1}$$

where  $\sigma$  is the local linear stellar velocity dispersion,  $m_\star$  is the mass of a typical field star,  $\rho$  is the local stellar density, and  $\ln \Lambda$  is the Coulomb logarithm.

Since the main sequence lifetime of stars more massive than  $\sim 2M_\odot$  is shorter than  $t_E$ , young massive stars in the GC are not relaxed; their distribution is primarily a reflection of their formative conditions. While lower-mass dwarf stars are sufficiently

old to be relaxed in the central potential, their distribution in the innermost region could be affected by an abundance of stellar-mass black holes ( $5-10M_{\odot}$ ). As products of normal stellar evolution, stellar mass black holes sink in the potential of the massive black hole (Morris, 1993; Miralda-Escudé & Gould, 2000) and displace the less massive stars and remnants.

Speckle imaging and more recently adaptive optics with the Keck and VLT have provided several milliarcsecond astrometry, enabling the detection of proper motions within the inner 0.5 pc and accelerated proper motions of  $\sim 10$  stars within the inner 0.05 pc. Radial velocities with spectroscopic precisions of  $\delta v \sim 30 \text{ km s}^{-1}$  have also been obtained for the star S0-2, which has been monitored for over 70% of its orbit including pericenter passage at  $\sim 130 \text{ AU}$  from the black hole. These observations have enabled the black hole mass and GC distance to be measured to within  $\sim 10\%$  (Ghez et al., 2005; Schödel et al., 2003).

Here we examine the extent to which one can probe the GC potential by monitoring stars with a diffraction-limited, next-generation, extremely large telescope (ELT). As compared with current 10 m class telescopes, the finer angular resolution of an ELT enables the orbital motions of many more stars to be detected, each at greater astrometric precision,  $\delta\theta$ , and spectroscopic precision,  $\delta v$ . Given the range of possible sizes of future telescopes and given the uncertainties in the ultimate capabilities of a specific telescope class (e.g., 30 meter telescopes) we choose to express our results not as functions of the ELT aperture but rather as functions of  $\delta\theta$ ,  $\delta v$ , and the number  $N$  of stars with detectable orbital motions. We take  $\delta\theta = 0.5 \text{ mas}$  and  $\delta v = 10 \text{ km s}^{-1}$  as our fiducial model, corresponding to a conservative estimate of the capabilities of a telescope with a  $D = 30 \text{ m}$  aperture. We show that  $N$  scales with telescope aperture as  $N \simeq 100(D/30 \text{ m})^2$ . We demonstrate that with an ELT one can measure the density profile of a dark matter spike and those general relativistic effects that scale as  $(v/c)^2$ , where  $v$  is the speed of a star and  $c$  is the speed of light. Furthermore, we show that the distance to the GC will be measured to remarkable precision. This will help place tight constraints on models of the overall Galactic structure. We also show that with an ELT one can detect the gravitational interactions between monitored

stars and the background massive stellar remnants that accumulate near the central black hole. Such interactions may probe the mass function of the stellar-mass black holes thought to dominate the matter density in the region.

The paper is organized as follows. In § 2.2 we calculate the number of stars with accelerated proper motions that can be monitored with a given ELT based on its astrometric, spectroscopic, and confusion limits. We also describe a realistic monitoring program and demonstrate that confusion with the infrared emission from Sgr A\* is unlikely to affect an ELT’s ability to measure stellar motions. In § 2.3 we model the orbital data and estimate the magnitude of various non-Keplerian effects including Newtonian retrograde precession due to extended matter, relativistic prograde precession, precession induced by the coupling of orbits to the spin of the black hole, and the Roemer time delay. In § 2.3.6 we consider the effect of stellar interactions on the motion of the monitored stars. Specifically, we estimate the rate at which discrete stellar encounters result in detectable changes of orbital motions. In § 2.4 we discuss a method for generating mock ELT orbital data and describe a computational technique for estimating uncertainties in the orbital parameters. The results of our calculations are given in § 2.5. In § 3.4 we discuss astrophysical applications of the proposed observations. Finally, in an appendix, we show the current constraints on  $M_{\text{bh}}$ ,  $R_0$ , and the extended matter distribution, from fits to the Keck orbital data.

## 2.2 Observing stars in the central arcsecond with an ELT

The purpose of this section is to estimate the number and distribution of stars whose orbital motions can be detected with an ELT and to determine the astrometric and spectroscopic precision to which their motions can be measured (§ 2.2.1). The latter are determined by the specifications of the telescope and the properties of the stellar population at the GC. Several factors complicate the monitoring of orbits within the central arcsecond. The greatest obstacle to detecting and following hitherto unseen

stars is stellar crowding. Light contamination from nearby bright stars as well as the light from underlying faint stars flood the pixel elements and impose a limit to the faintest detectable star. In § 2.2.2 we estimate the minimum luminosity permitted by the crowding and thereby obtain an estimate of the number count of stars with observable orbital motions.

### 2.2.1 Astrometric and spectroscopic limit

With adaptive optics, an ELT will operate near its diffraction limit in the  $K$ -band. By determining the centroid of images, the measured relative positions of stars are a factor of  $\sim 20 - 40$  more definite than the images' diffraction limit. For instance, the diffraction limit of Keck is  $\sim 50$  mas while the astrometric error of a bright star near the GC as seen by Keck is  $\sim 1 - 2$  mas. Naively, the expected astrometric limit of an ELT with  $D = 30$  m is therefore  $\delta\theta_{30} \sim 0.5$  mas.

In practice, the astrometric limit achievable with adaptive optics depends on whether atmospheric fluctuations or centroid measurement errors dominate the signal. At the GC the separation between the guide star needed for the adaptive optics infrared wavefront sensor and the star under study is typically  $\sim 5''$ , corresponding to a separation of 0.25 m at the top ( $\sim 10$  km) of the atmosphere. As long as the telescope aperture is larger than this separation, as is the case for Keck and an ELT, the atmosphere dominates and the astrometric precision scales with the telescope diameter as  $D^{2/3}$  (Shao & Colavita, 1992). A 30 meter ELT is expected to have an astrometric limit that is  $3^{2/3} \approx 2$  times smaller than Keck's for  $K \lesssim 24$ . We therefore adopt  $\delta\theta_{30} \sim 0.5$  mas in our calculations, though we consider this a conservative estimate; a 30 meter ELT may attain an astrometric limit as small as 0.1 mas.<sup>1</sup>

With an adaptive-optics-fed spectrometer on Keck, Ghez et al. (2003) detected spectral absorption lines in the star S0-2 at a spectral resolution of  $R = \lambda/\Delta\lambda \sim 4000$ , yielding a radial velocity measurement with an error of  $40 \text{ km s}^{-1}$  (see also Eisenhauer et al. 2003). Integral field spectroscopy in the near-IR with a 30 meter ELT is expected

---

<sup>1</sup>[www.astro.caltech.edu/mirror/celt/participants/AOWG/SRD\\_v9.pdf](http://www.astro.caltech.edu/mirror/celt/participants/AOWG/SRD_v9.pdf)

to enable measurements with  $R \sim 1 - 2 \times 10^4$ , suggesting that velocity errors of  $\delta v_{30} \sim 10 \text{ km s}^{-1}$  are attainable. This too is a conservative estimate as a 30 meter ELT may achieve velocity errors more than an order of magnitude smaller (D. Figer, private communication). As we discuss below, an ELT will be able to detect stars that are fainter and hence cooler than those currently detectable. Cool stars exhibit rich spectral features including possible molecular lines, enabling high spectral resolution studies. For example, Figer et al. (2003) obtained radial velocities for 85 cool stars in the central parsec of the Galaxy with velocity errors of  $\sim 1 \text{ km s}^{-1}$ .

Although an ELT's astrometric and spectroscopic limits may differ from the above estimates, we show in § 2.4.2 that the uncertainties in the model parameters extracted from the monitoring data, such as the distance to the GC and the extended matter profile, scale almost linearly with the measurement errors. The constraints on the parameters for different values of  $\delta\theta$  and  $\delta v$  can therefore be readily inferred from our results.

### 2.2.2 Confusion limit

The brighter stars wash out the signal of fainter stars, thereby limiting the luminosity of the faintest observable star. This limit depends on the telescope optics (e.g., angular resolution) and on the stellar luminosity function (LF). Using measurements of stellar photometry near the GC, we now estimate the minimum luminosity that a star at the GC can have and still be identified and monitored with an ELT. For a given star of luminosity  $l$  and for a given  $K$ -band stellar LF, we determine the integrated flux from all nearby background stars with luminosity  $< l$ . At some minimum luminosity, the emission from a single star is comparable to the background emission; this luminosity sets the confusion limit.

Following Takeuchi et al. (2001) and references therein, let  $x_S = S h(\theta, \phi)$  be the response of the telescope to a source of flux density  $S$  at an angular position  $(\theta, \phi)$  from the line-of-sight axis to the center of the source.  $h(\theta, \phi)$  is the point-spread function (PSF) of the telescope, normalized to unity at the center. Since all sources at the

GC are essentially at the same distance, we can instead express the response in terms of stellar luminosity  $l$ , i.e., let  $x = lh(\theta, \phi)$ . The variance in the telescope response due to crowding is the confusion noise  $\sigma$ . To detect a source with high statistical significance, its luminosity must be greater than some cutoff  $l_c$ , or equivalently,  $x$  must be greater than a response cutoff  $x_c$ . Defining  $q = x_c/\sigma$ , a source is above the confusion limit if its signal-to-noise  $S = x/\sigma > q$ , where we take  $q = 5$ .

If the number of stars per square arcsec with luminosity in the range  $(l, l + dl)$  is  $dN = \alpha\Phi(l)dl$ , where  $\alpha$  is the normalization of the LF and  $\Phi(l)$  is its shape, then the mean number of source responses of intensity  $x$  is

$$\begin{aligned} R(x)dx &= \int \frac{dN(l)}{dl} dl d\Omega \\ &= \int \alpha\Phi\left(\frac{x}{h(\theta, \phi)}\right) \frac{d\Omega}{h(\theta, \phi)} dx, \end{aligned} \quad (2.2)$$

where the integral is over the solid angle of the PSF. The confusion noise  $\sigma$  due to all sources fainter than  $x_c$  is then

$$\sigma^2 = \int_0^{x_c} x^2 R(x) dx. \quad (2.3)$$

Since we are interested in calculating the cutoff response of a given detector for a given LF, we need to solve for the confusion noise. Assuming a power law LF of the form  $dN/dl = \alpha\Phi(l) = \alpha l^{-\eta}$  we have

$$R(x) = \int \alpha \left[ \frac{x}{h(\theta, \phi)} \right]^{-\eta} \frac{d\Omega}{h(\theta, \phi)} = \alpha x^{-\eta} \Omega_{\text{eff}}, \quad (2.4)$$

where

$$\Omega_{\text{eff}} = \int h(\theta, \phi)^{\eta-1} d\Omega. \quad (2.5)$$

Therefore

$$\sigma^2 = \alpha \Omega_{\text{eff}} \int_0^{x_c} x^{2-\eta} dx = \left( \frac{q^{3-\eta} \alpha \Omega_{\text{eff}}}{3-\eta} \right)^{2/(\eta-1)}. \quad (2.6)$$

In this paper we only consider power-law LFs, though one can obtain an expression for  $\sigma$  for general forms of the LF (Franceschini et al., 1989).

For a Gaussian PSF  $h(\theta, \phi) = h(\theta) = \exp[-(4 \ln 2)(\theta/\theta_0)^2]$ , where  $\theta_0$  is the PSF's full-width at half-maximum. This gives  $\Omega_{\text{eff}} = \pi\theta_0^2 / [(4 \ln 2)(\eta - 1)]$  and thus

$$\sigma = \left( \frac{q^{3-\eta}}{3-\eta} \right)^{1/(\eta-1)} \left[ \frac{\pi\theta_0^2\alpha}{(4 \ln 2)(\eta-1)} \right]^{1/(\eta-1)}. \quad (2.7)$$

We now estimate the value of  $\sigma$  for the Keck and an ELT. In the  $K$ -band,  $\theta_0 \simeq 50$  mas for Keck and  $\theta_0 \simeq 15$  mas for a  $D = 30$  m ELT. We also need the  $K$ -band luminosity function (KLF) of stars at the GC. Genzel et al. (2003) find that the KLF within  $1.5''$  of the GC is well-described by a power-law with slope  $\beta = d \log N / dK = 0.21 \pm 0.03$  where  $K$  is the apparent magnitude in the  $K$ -band. We consider KLFs with slopes within the  $\sim 2\sigma$  range  $0.15 < \beta < 0.27$ , which in terms of  $\eta = 1 + \beta/0.4$  corresponds to the range  $1.38 < \eta < 1.68$ .

Schödel et al. (2003) measured the photometry of more than 40 stars in the central arcsec, 29 of which reside within  $0.8''$  ( $\sim 6000$  AU). We normalize the KLF to these 29 stars. We limit our analysis to these innermost stars since the KLF inside  $0.8''$  appears to differ from that outside this region (see § 3.1). We do not attempt to account for a possible radial dependence but instead assume the KLF is constant.

Of the 29 stars, the brightest has apparent magnitude  $K = 13.4$  and the dimmest  $K = 17.3$ . Assuming a  $K$ -band extinction of 3.3 mag (Rieke, Rieke, & Paul, 1989) and a distance to the GC  $R_0 = 8$  kpc, these apparent magnitudes correspond to  $K$ -band luminosities of  $l_{\text{min}} = 0.8L_\odot$  and  $l_{\text{max}} = 28L_\odot$ . For a given  $\eta$  we calculate  $\alpha = N_{\text{obs}}(1 - \eta) / (l_{\text{max}}^{1-\eta} - l_{\text{min}}^{1-\eta})$  where  $N_{\text{obs}} = 29/\pi(0.8'')^2$  and by equation (2.7) solve for  $\sigma$ . Integrating the luminosity function over stars brighter than  $x_c = q\sigma$  yields the number count of detectable stars  $N(l > l_c) \sim \alpha x_c^{1-\eta} / (\eta - 1)$ .

In Figure 2.1 we show how the  $K$ -band magnitude limit and number  $N$  of stars with detectable orbital motions (those within 3000 AU of the GC; see below) scale with the aperture of a diffraction limited ELT assuming  $\beta = 0.21 \pm 0.06$  ( $\eta = 1.53 \pm 0.15$ ). Since by equation (2.7)  $N \propto x_c^{1-\eta} \propto \theta_0^{-2}$ , we find that  $N \simeq 100(D/30 \text{ m})^2$ .

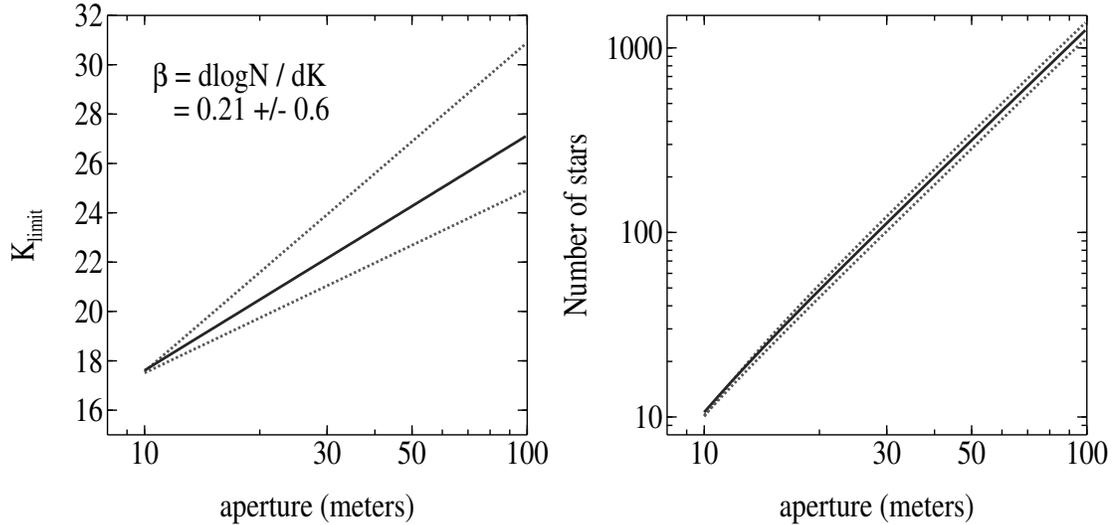


Figure 2.1  $K$ -band magnitude limit and number of stars with detectable orbital motions as a function of the aperture of a diffraction-limited ELT. Results are shown for power-law  $K$ -band luminosity functions normalized to observations by Schödel et al. (2003) with slopes matching the  $\sim 2\sigma$  range found by Genzel et al. (2003). The uncertainty in the number counts is considerably smaller than the uncertainty in the magnitude limits.

Furthermore, because  $x_c \propto \alpha^{1/(\eta-1)}$ ,  $N(l > l_c)$  is not very sensitive to the value of  $x_c$  for a fixed  $\eta$ . Therefore, when the above analysis is performed on a subset of the 29 stars within  $0.8''$  (e.g., stars within  $0.4''$  or alternatively stars brighter than  $K = 16$ ), the derived number counts, unlike the magnitude limits, do not change significantly. The number counts we derive for an ELT are therefore robust even though the magnitude limits are subject to some uncertainty.

To extract orbital parameters the acceleration of a star in the plane of the sky must be detected, i.e., it must be greater than the threshold acceleration  $\xi_t$ . For Keck  $\xi_t \sim 1 - 2 \text{ mas yr}^{-2}$  while for a 30 meter ELT  $\xi_t \sim 0.5 \text{ mas yr}^{-2}$ . The accelerated proper motion is detectable over the entire orbit if the acceleration at apocenter exceeds the threshold. For a face on orbit this requires  $a(1 + e) < (GM/\xi_t R_0)^{1/2}$ . Thus the acceleration will be detectable with a 30 meter ELT over the entire orbit if  $a \lesssim 3000 \text{ AU}$  (period  $\lesssim 80 \text{ yr}$ ). To construct our mock stellar orbits to simulate

observations that can be made with an ELT, we only consider orbits satisfying this constraint. As Figure 2.1 shows, within 3000 AU, approximately 100 stars are brighter than a 30 meter ELT’s confusion limit. Furthermore, since the surface density of stars is  $\sim 200 \text{ arcsec}^{-2}$ , one does not expect to observe stars with apocenters smaller than  $\sim 300 \text{ AU}$  ( $\sim 0.04''$ ) with such an ELT. We therefore conclude that a 30 meter ELT will detect the accelerated motion of  $\sim 100$  stars with semi-major axes  $\lesssim 3000 \text{ AU}$  and apocenter distances  $\gtrsim 300 \text{ AU}$ .

Another related issue is the frequency with which an ELT will measure positions and radial velocities for the  $N$  monitored stars, given a reasonable commissioning of  $\sim 10$  GC exposures per year. As it will be equipped with an integral field unit spectrometer an ELT can obtain simultaneous spectral and spatial data over a relatively large region of sky. It is possible for it to measure positions and velocities for all  $N$  stars in a single image. This suggests that a dedicated observing program can reasonably obtain ten measurements per star per year.

At the high levels of precision obtainable with an ELT, orbital parameter constraints should scale with the measurement errors  $\sigma$  (i.e.,  $\delta\theta$  and  $\delta v$ ) and the number of stars  $N$  as  $\sigma/N^{1/2}$ . Based on the above discussion we therefore expect the parameter constraints to scale with telescope aperture as  $\sim D^{-2/3}/D \sim D^{-5/3}$ . We verify this relation in our numerical simulation results described in § 2.5.

### 2.2.3 Central point source — Sgr A\*

At radio wavelengths Sgr A\* is detected as a nonthermal (Beckert et al., 1996; Serabyn et al., 1997), compact (Rogers et al., 1994), static (Backer & Sramek, 1999; Reid et al., 1999), variable (Zhao, Bower, & Goss, 2001) source. An X-ray source coincident with Sgr A\* has also been detected (Baganoff et al., 2003) and consists of a resolved, steady-state component with size  $\sim 1''$  and an unresolved flaring component that increases in flux density by an order of magnitude over the course of a few hours roughly once per day (Baganoff et al., 2001; Goldwurm et al., 2003; Porquet et al., 2003).

Recently, a near-infrared counterpart to Sgr A\* , located within a few mas of the dynamically determined black hole position, has been observed in the  $H$  band ( $1.7\mu\text{m}$ ; Genzel et al. 2003b) and the  $L'$  band ( $3.8\mu\text{m}$ ; Ghez et al. 2004). Like the X-ray emission, the infrared emission consists of a quiescent component and a variable component. The latter exhibits flux densities that increase by a factor of a few over the course of tens of minutes to one week, with possible signs of periodicity (Genzel et al., 2003b). The observed  $L'$ -magnitudes are in the range  $12.2 - 13.8$ , corresponding to  $L'$ -luminosities  $\sim 10 - 100L_{\odot}$ .

Although the current sample of stars with detected accelerated motion are brighter than the Sgr A\* infrared emission, the stars detectable with an ELT will have comparable luminosities. A star that passes near the black hole can therefore be confused with the emission of Sgr A\* . Conservatively, such confusion limits monitoring when the projected separation between a star and Sgr A\* is smaller than the resolution of the detector. For a 30 meter ELT operating at the diffraction limit in the  $K$  band, this corresponds to  $\sim 15$  mas (120 AU). However, as we found in § 2.2.2, confusion with nearby stars precludes such a telescope from detecting orbits with apocenters smaller than 300 AU ( $\sim 40$  mas) and most of the monitored stars do not therefore pass within 15 mas of Sgr A\* . Of those that do, most spend only a small fraction of their total orbital period that close to the black hole; e.g., a star with a semi-major axis of 200 AU and eccentricity 0.9 is within 15 mas of the black hole for only 10% of its orbital period. The same arguments hold for other ELT apertures. Therefore, the infrared emission from the black hole will not significantly impair orbital monitoring with an ELT.

## 2.3 Orbital dynamics

While current observations of stellar proper motions near the black hole at the GC are consistent with motion around a Newtonian point mass, we show that with an ELT non-Keplerian motions are going to be detectable. There are various effects that cause deviations from Keplerian motion, including the Newtonian retrograde

precession (NRP) of an orbit due to the presence of an extended matter distribution (§ 2.3.2), the relativistic prograde precession (RPP; § 2.3.3), and the frame dragging effects related to the black hole spin (§ 2.3.4). In addition, we account for an apparent deformation of the observed proper motion (“Roemer effect”; § 2.3.5) due to the differences in light travel times at different locations along the orbit. A discussion of the effects of encounters between monitored stars and background stars is given in § 2.3.6. We now describe the orbital equations of motion and estimate the magnitude of the various non-Keplerian effects. A number of relativistic effects, including those we consider below, are discussed in Pfahl & Loeb (2003) in connection with long-term timing observations of a radio pulsar that might be detected in a  $\lesssim 100$  year orbit about the GC.

### 2.3.1 Equations of motion

We found in § 2.2 that we do not expect a 30 meter ELT to detect orbits with apocenter smaller than  $\sim 300$  AU, due to confusion noise. Assuming orbits uniformly distributed in  $e^2$ , the probability that a given star has  $e > 0.99$  is 2%. Since most of the  $\sim 100$  stars such an ELT monitors will have semi-major axes  $> 1000$  AU, it is unlikely that any will have pericenter distance smaller than a few AU. As a result, the ratio of the Schwarzschild radius to the pericenter distance of the stars will satisfy  $R_s/r_p \lesssim 0.05$ , or expressed in terms of the stellar velocity at pericenter,  $v_p/c \lesssim 0.2$ . The post-Newtonian approximation to the geodesic equations that is accurate to order  $(v/c)^2$  provides an adequate description of the stellar orbits given the observational precision expected with an ELT.

The geodesic equation for test particles orbiting a spherically symmetric mass is, in the post-Newtonian approximation (Weinberg, 1972; Rubilar & Eckart, 2001),

$$\frac{d\mathbf{v}}{dt} = -\nabla\Phi - [2\nabla\Phi^2 + v^2\nabla\Phi - 4\mathbf{v}(\mathbf{v} \cdot \nabla)\Phi - \mathbf{v} \times (\nabla \times \zeta)]/c^2, \quad (2.8)$$

where  $\mathbf{v} = d\mathbf{x}/dt$  is the velocity vector,  $\Phi$  is a time-independent gravitational potential, and  $\zeta = 2G(\mathbf{x} \times \mathbf{J})/r^3$  is a vector potential associated with the spin  $\mathbf{J}$  of the

gravitating mass, which we assume is constant with time. We assume the density distribution of the extended matter at radii traversed by the stars and smaller is a power-law profile  $\rho(r) = \rho_0(r/r_0)^{-\gamma}$ . Input model parameters are described in § 2.4.2.

The relativistic effects include corrections to the orbital dynamics and to the observed motion due to propagation effects. The former class includes the RPP and frame dragging while the latter class includes the lowest order  $(v/c)$  Roemer time delay and such  $(v/c)^2$  effects as time dilation, gravitational redshift, and the Shapiro time delay. Since the  $(v/c)^2$  propagation effects each have different functional dependences on the orbital parameters (see e.g., Pfahl & Loeb 2003), including them may break degeneracies, though they may also weaken the sensitivity to some parameters. However, our interest in the  $(v/c)^2$  relativistic effects is primarily connected with the ability of an ELT to probe general relativity on the scale of a massive black hole rather than with the effects' potential utility for parameter estimation. Since  $v_p/c < 0.2$ , relativistic effects help to constrain the orbital parameters by at most a few percent. We therefore chose not to include the  $(v/c)^2$  propagation effects in our analyses. The analysis of actual data obtained from an ELT must, however, account for all the relativistic effects.

### 2.3.2 Newtonian retrograde precession

The NRP was discussed in the context of the GC by Rubilar & Eckart (2001). An extended matter distribution causes stellar orbits to precess due to differences in the amount of mass that is contained between the apocenter and the pericenter radii. In the numerical calculations that follow, we determine how much an orbit precesses due to the extended matter by solving equation (2.8). Here, however, we obtain an estimate of the magnitude of the precession for stars at the GC by considering the potential of the extended matter to be a small correction  $\delta\Phi$  to the potential of the black hole. Expanding the total potential to linear order in  $\delta\Phi$ , the angular shift per

period is (Landau & Lifshitz, 1960)

$$\Delta\phi_{\text{Newt}} = \frac{\partial}{\partial L} \left( \frac{2}{L} \int_0^\pi r^2 \delta\Phi d\varphi \right), \quad (2.9)$$

where  $L$  is the orbital angular momentum per unit mass,  $r = a(1 - e^2)/(1 + e \cos \varphi)$ , and  $\varphi$  is the phase of the orbit. If  $\gamma < 2$  we have  $\delta\Phi = \beta r^{2-\gamma}$ , where  $\beta = GM_{\text{ext}}/(2 - \gamma)r_0^{3-\gamma}$  is a constant and  $M_{\text{ext}} \equiv M_{\text{ext}}(r < r_0)$  is the extended matter mass within  $r_0$ . The orbital precession is then given by (Munyanza & Viollier, 1999)

$$\Delta\phi_{\text{Newt}} = \frac{2\beta}{GM_{\text{bh}}} [a(1 - e^2)]^{3-\gamma} g(\gamma, e), \quad (2.10)$$

where

$$g(\gamma, e) = \frac{1 - e^2}{e^2} (4 - \gamma) [I_{4-\gamma}(e) - I_{5-\gamma}(e)] + (7 - 2\gamma)I_{4-\gamma}(e) \quad (2.11)$$

and

$$I_n(e) \equiv \int_0^\pi \frac{d\varphi}{(1 + e \cos \varphi)^n}. \quad (2.12)$$

Assume that the extended matter consists of stars with  $\gamma = 7/4$  and  $\rho_{1\text{pc}} = 2 \times 10^5 M_\odot \text{pc}^{-3}$  (see § 3.1). Consider an S0-2-like orbit with a semi-major axis of 0.005 pc and eccentricity  $e = 0.9$ . The enclosed stellar mass at apocenter and pericenter are  $6000M_\odot$  and  $150M_\odot$ . Solving equation (2.10) yields a precession per revolution of  $\Delta\phi_{\text{Newt}} \approx 0.08^\circ$ , corresponding to an apparent angular apocenter shift of roughly  $\Delta\phi_{\text{Newt}} a(1 + e)/R_0 \approx 0.3 \text{ mas}$ . Thus, a few S0-2-like orbits with astrometric errors of 0.5 mas provide a meaningful constraint on the stellar distribution within the inner few milliparsecs. If the density of the dark matter cusp at the stellar positions exceeds  $\sim 10^8 M_\odot \text{pc}^{-3}$ , then it too will produce a detectable precession; it will not be easily distinguished from the stellar contribution (however see § 2.3.6).

### 2.3.3 Relativistic prograde precession

The RPP causes a pericenter advance per revolution of (see Weinberg 1972)  $\Delta\phi_{\text{pro}} = 3\pi R_s/a(1 - e^2)$ , where  $R_s = 2GM_{\text{bh}}/c^2$  is the Schwarzschild radius of the black hole. The magnitude of the effect is  $\propto (v/c)^2$ . The apparent apocenter shift per revolution caused by the RPP is  $\Delta s \approx \Delta\phi_{\text{pro}}a(1 + e)/R_0 = 3\pi R_s/R_0(1 - e)$ , which corresponds to an apparent shift of  $\sim 1$  mas for the star S0-2. Although the RPP has an additional factor of  $(v/c)$  relative to the Roemer effect (§ 2.3.5), this attenuation can be compensated for by having a few high eccentricity stars in the sample. Furthermore, unlike the Roemer effect, the RPP shift is to first order independent of the semi-major axis and is therefore equally sensitive to stars at all radii (although stars at large radii also have long periods). Consider an orbit seen face-on and observed for  $N_{\text{orb}}$  complete periods. Since the precession angles per revolution add linearly, the signal-to-noise from the RPP is  $S_{\text{pro}} \sim \Delta s N_{\text{orb}}/\delta\theta$ , or

$$S_{\text{pro}} \sim 0.1 \frac{N_{\text{orb}}}{1 - e} \left( \frac{M_{\text{bh}}}{4 \times 10^6 M_{\odot}} \right) \left( \frac{R_0}{8 \text{ kpc}} \right)^{-1} \left( \frac{\delta\theta}{1 \text{ mas}} \right)^{-1}. \quad (2.13)$$

In a sample of 100 stars observed with astrometric errors of 0.5 mas and having an eccentricity distribution uniform over  $e^2$ , we expect on average eight stars with  $e > 0.96$ . If only one such star is followed over just a single period, the RPP shift will be measured to  $5\text{-}\sigma$  accuracy.

### 2.3.4 Frame dragging

For a spinning black hole, frame dragging effects also cause a precession of the pericenter. The spin precession per revolution for a star orbiting a black hole with spin angular momentum  $J$  is given approximately by (see Weinberg 1972, equation (9.5.22);

note different notation)

$$\begin{aligned}\Delta\phi_{\text{spin}} &\approx -8\pi j \left(\frac{GM_{\text{bh}}}{cL}\right)^3 \cos\psi \\ &= -\frac{2\sqrt{2}j\Delta\phi_{\text{pro}}}{3} \sqrt{\frac{R_s}{a(1-e^2)}} \cos\psi,\end{aligned}\tag{2.14}$$

where  $\psi$  is the angle between the orbital angular momentum vector and the black hole spin axis and  $0 \leq j \equiv cJ/GM_{\text{bh}}^2 \leq 1$  is the black hole spin parameter.

The black hole spin induces an apocenter shift that is smaller than the RPP shift by a factor of  $\sim v/c$ . Even if the black hole is maximally spinning ( $j = 1$ ), the shift represents only a 5% contribution on top of the RPP for a star with  $a = 200$  AU and eccentricity  $e = 0.92$ . For an orbit observed face-on the signal-to-noise from a spin-induced apocenter shift is

$$\begin{aligned}S_{\text{spin}} &\approx \frac{2\pi\sqrt{2}j}{R_0\sqrt{a(1+e)}} \left(\frac{R_s}{1-e}\right)^{3/2} \frac{N_{\text{orb}}}{\delta\theta} \cos\psi \\ &\approx 0.001 \frac{jN_{\text{orb}} \cos\psi}{\sqrt{(1+e)(1-e)^3}} \left(\frac{M_{\text{bh}}}{4 \times 10^6 M_{\odot}}\right)^{3/2} \\ &\quad \times \left(\frac{a}{1000 \text{ AU}}\right)^{-1/2} \left(\frac{R_0}{8 \text{ kpc}}\right)^{-1} \left(\frac{\delta\theta}{1 \text{ mas}}\right)^{-1}.\end{aligned}\tag{2.15}$$

For example, a  $5\text{-}\sigma$  detection is achieved with an ELT with  $\delta\theta = 0.5$  mas if a star with  $a = 300$  AU and  $e = 0.99$  is monitored for three complete orbits. We expect a 30 meter ELT to detect one star with a semi-major axis that small (§ 2.2). Assuming eccentricities uniformly distributed in  $e^2$  the probability that star has  $e > 0.99$  is only  $\sim 2\%$ . If  $\delta\theta = 0.05$  mas, a star with  $a = 300$  AU and  $e = 0.95$  will yield a  $5\text{-}\sigma$  detection after being monitored for three complete orbits. Since so high a resolution requires an ELT with aperture  $D \sim 100$  m, there will be several stars with  $a \lesssim 300$  AU and of these  $\sim 10\%$  will have  $e > 0.95$ . Detecting such a stellar orbit is therefore not unlikely. The spin-induced orbital precession thus requires an astrometric precession of  $\delta\theta \lesssim 0.05$  mas (see also Jaroszynski 1998; Fragile & Mathews 2000).

### 2.3.5 The Roemer time delay

For orbits with non-zero inclination, the distance between the earth and star, and hence the difference in time between stellar emission and observation, varies with orbital phase. This time delay, given by  $\Delta t = t_{\text{obs}} - t_{\text{em}} = z(t_{\text{em}})/c$ , where  $z(t)$  is the relative distance between the star and the massive black hole, was first recognized by Roemer in 1676 in application to the phases of Jupiter's moons. Unlike the relativistic Doppler effect, which includes corrections of order  $(v/c)^2$  and higher, the Roemer delay is the classical Doppler effect, which only includes terms up to order  $v/c$  (see, e.g., Loeb 2003). The delay has a magnitude corresponding to a few percent of a year for an S0-2-like orbit, and is observed as an additional shift in the apparent stellar position with time,  $\Delta s(t)$ . For a circular orbit seen edge-on the stellar positions  $z(t)$  and  $s(t)$  are sinusoidal so that

$$\begin{aligned} \Delta s(t)/a &= \cos(\omega t_{\text{obs}}) - \cos(\omega t_{\text{em}}) \\ &= \cos[\omega(t_{\text{em}} + a/c \sin(\omega t_{\text{em}}))] - \cos(\omega t_{\text{em}}) \\ &\simeq -\frac{\omega a}{c} \sin^2(\omega t_{\text{em}}), \end{aligned} \quad (2.16)$$

where  $\omega = 2\pi/P$  and we used the fact that for orbits at the GC  $v \ll c$ . The maximum shift, in units of the semi-major axis, is therefore  $v/c$ . For non-zero eccentricity and arbitrary inclination the star's projected position and distance as a function of time are (see e.g., Murray & Dermott 1999)

$$\begin{Bmatrix} x(t) \\ y(t) \\ z(t) \end{Bmatrix} = r(t) \begin{Bmatrix} \cos[\varphi(t) + \alpha] \\ \sin[\varphi(t) + \alpha] \cos i \\ \sin[\varphi(t) + \alpha] \sin i \end{Bmatrix}, \quad (2.17)$$

where  $i$  is the inclination,  $\varphi(t)$  is the orbital-phase (i.e., the true anomaly),  $\alpha$  is the argument of pericenter, and we chose the reference direction so that the  $x$ -axis coincides with the longitude of ascending node. The Roemer shift is then  $|\Delta s| = (\Delta x^2 + \Delta y^2)^{1/2}$  where  $\Delta x = x(t_{\text{obs}}) - x(t_{\text{em}})$  and similarly for  $\Delta y$ . To linear order in

$v/c$  the orbit-averaged Roemer shift can be written as

$$\langle |\Delta s(t)|/a \rangle \simeq \frac{\omega a}{c} \sin i \sqrt{f(e, i, \alpha)} \quad (2.18)$$

where  $f(e, i, \alpha)$  is a factor of order unity. For the two extreme cases  $\alpha = 0$  and  $\alpha = \pi/2$  (corresponding to the line-of-node along the major-axis and minor-axis, respectively),  $f(e, i, \alpha)$  is given by

$$\begin{aligned} f(\alpha = 0) = & \frac{\sqrt{1-e^2}}{2e^4} \left\{ 2e^4 \cos^2 i \left( 3 - \sqrt{1-e^2} \right) \right. \\ & + 3e^2 \left[ 2 - \sqrt{1-e^2} - \cos^2 i \left( 4 - 3\sqrt{1-e^2} \right) \right] \\ & \left. - 6 \sin^2 i \left( 1 - \sqrt{1-e^2} \right) \right\}, \end{aligned} \quad (2.19)$$

$$\begin{aligned} f(\alpha = \pi/2) = & \frac{\sqrt{1-e^2}}{2e^4} \left\{ -2e^4 \left( 2 - \sqrt{1-e^2} \right) \right. \\ & + e^2 \left[ 10 - 7\sqrt{1-e^2} - \cos^2 i \left( 4 - \sqrt{1-e^2} \right) \right] \\ & \left. - 6 \sin^2 i \left( 1 - \sqrt{1-e^2} \right) \right\}. \end{aligned} \quad (2.20)$$

An ELT will be able to detect the effect of the Roemer delay in orbits at the GC. The signal-to-noise from  $N_{\text{obs}}$  observations of an orbit measured with astrometric errors  $\delta\theta$  is approximately  $S_{\text{delay}} \sim \langle \Delta s \rangle N_{\text{obs}}^{1/2} / R_0 \delta\theta$ , or

$$\begin{aligned} S_{\text{delay}} \approx & 0.8 \sqrt{N_{\text{obs}}} \sin i \sqrt{f(e, i, \alpha)} \left( \frac{a}{1000 \text{ AU}} \right)^{1/2} \\ & \times \left( \frac{M_{\text{bh}}}{4 \times 10^6 M_{\odot}} \right)^{1/2} \left( \frac{R_0}{8 \text{ kpc}} \right)^{-1} \left( \frac{\delta\theta}{1 \text{ mas}} \right)^{-1}. \end{aligned} \quad (2.21)$$

If, e.g., we pick  $i \sim \pi/3$ ,  $\alpha = 0$ , and  $e \sim 1/\sqrt{2}$ , an astrometric error of 0.5 mas, a mean semi-major axis of 1000 AU, and 10 observations per star, then we can detect the delay to  $S_{\text{delay}} \sim 5$  with roughly ten stars. We therefore expect the Roemer delay to be detectable in such an ELT's sample of  $\sim 100$  stars and the effect must be taken into account during parameter estimation.

### 2.3.6 Interstellar interactions

In the previous sections we described the motion of a star in the potential of a black hole and a smooth distribution of extended matter, including stars, remnants, and dark matter. This approximation ignores the fact that the potential due to stars and remnants is the sum of discrete point-mass potentials and is therefore not perfectly smooth. The stars experience perturbations due to nearby encounters with individual stars and due to fluctuations in the potential arising from all stars. These perturbations cause a star's orbital parameters to change with time. The magnitude and the rate of these changes depend on the stellar mass function, since the perturbations are sensitive to the characteristic mass of the field stars. Thus, measuring the effects of stellar encounters is a probe of the mass function in the central parsec. It also breaks the degeneracy between the contributions of stellar matter and dark matter to the Newtonian orbital precession. Encounters may also be a source of noise in measurements of orbital parameters such as the black hole mass and distance to the GC. While we do not include the effects of encounters in our numerical calculations presented in § 2.5, we now estimate their magnitude and demonstrate that the encounters might be detectable with an ELT and present a powerful probe of the mass function of stellar remnants at the GC.

An encounter between a test star of mass  $m_j$  and a field star of mass  $m_i$  with impact parameter  $b$  induces a change in the test star velocity given by (see, e.g., Spitzer 1987)

$$\delta v = \frac{2m_i v_{\text{rel}}}{m_i + m_j} \left[ 1 + \left( \frac{b}{b_0} \right)^2 \right]^{-1/2}, \quad (2.22)$$

where  $b_0 = G(m_i + m_j)/v_{\text{rel}}^2$  and  $v_{\text{rel}}$  is the initial relative velocity of the stars. The encounter induces a change in the test star's velocity distinct from that due to orbital motion around the black hole. We solve for the maximum impact parameter  $b_{\text{max}}$  such that an encounter induces a change in velocity of the test star larger than the minimum detectable change  $\delta v_{\text{min}}$ . For uncorrelated position measurements the

minimum detectable change in velocity is  $\delta v_{\min} \sim \delta\theta R_0/\sqrt{N_{\text{obs}}}T$ , where  $T$  is the time baseline over which the orbit is monitored, and  $N_{\text{obs}}$  is the number of position measurements taken in time  $T$ . Assuming  $\delta\theta = 0.5$  mas (§ 2.2.1),  $T = 10$  yr, and  $N_{\text{obs}} = 100$  yields  $\delta v_{\min} \sim 0.2$  km s<sup>-1</sup>. By equation (2.22) we have

$$b_{\max} = b_0 \sqrt{\left(\frac{2m_i}{m_i + m_j} \frac{v_{\text{rel}}}{\delta v_{\min}}\right)^2 - 1} \approx \frac{2Gm_i}{v_{\text{rel}}\delta v_{\min}}, \quad (2.23)$$

where the approximation assumes  $v_{\text{rel}} \gg \delta v_{\min}$  and  $m_j \lesssim m_i$ . Assume  $v_{\text{rel}} \sim v_p = (GM(1+e)/(1-e)a)^{1/2}$  for an encounter near pericenter and  $v_{\text{rel}} \sim (GM/a)^{1/2}$  for an encounter near apocenter. If we take  $\delta v_{\min} \sim 0.2$  km s<sup>-1</sup> and  $m_i = 10M_{\odot}$  then for an S0-2-like orbit,  $b_{\max} \sim 10$  AU at pericenter and  $b_{\max} \sim 50$  AU at apocenter.

We ignore the effect of the black hole on the encounter and treat the interaction between the stars as a two-body problem. This is a fair approximation as long as the duration of the encounter is much shorter than the time scale over which the orbital velocity changes significantly due to the influence of the black hole. At pericenter passage, where the orbital acceleration is greatest, the orbital time scale is  $t_p \sim (1-e)^{3/2}P$ , where  $P$  is the orbital period. The two-body approximation is valid as long as the duration of the encounter satisfies  $t_{\text{enc}} \sim b_{\max}/v_{\text{rel}} \ll t_p$ . For an S0-2-like orbit  $t_p \sim 0.5$  yr while by equation (2.23)  $t_{\text{enc}} \lesssim 0.01$  yr even for  $m_i = 20M_{\odot}$ .

Next, we estimate the rate at which encounters  $b < b_{\max}$  occur for a star on a given orbit (see, e.g., Yu 2003). Let  $\Gamma_{ij}(\mathbf{r}, \mathbf{v}_j, t) dm_i$  be the rate at which a star with mass  $m_j$  at position  $\mathbf{r}$  with velocity  $\mathbf{v}_j$  at time  $t$  encounters stars with masses in the range  $m_i \rightarrow m_i + dm_i$ . Assume the number density of stars is spherically symmetric and follows a power law  $\nu(r) = \nu_0(r/r_h)^{-\alpha}$ . The phase-space distribution function of the stars is given by (Magorrian & Tremaine, 1999)  $f(\mathcal{E}) = h(\alpha)\mathcal{E}^{\alpha-3/2}$ , where  $\mathcal{E} = \Psi(r) - v^2/2$  and  $\Psi(r)$  is the relative gravitational potential at  $r$ , while

$$h(\alpha) = (2\pi\sigma_h^2)^{-3/2}\nu_0 \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-1/2)}\sigma_h^{-2\alpha+3}, \quad (2.24)$$

with  $\sigma_h$  the linear stellar velocity dispersion outside the sphere of influence of the BH

$r_h \simeq 1$  pc. The rate of detectable encounters in the mass bin is then

$$\begin{aligned}\Gamma_{ij} &= \int_0^\infty dv_i \frac{2\pi v_i}{v_j} f(\mathbf{r}, v_i) \int_{|v_i - v_j|}^{v_i + v_j} dv_{\text{rel}} v_{\text{rel}}^2 \Sigma(v_{\text{rel}}) \\ &= 2\pi K h(\alpha) \int_0^{\sqrt{2\Psi}} \left( \Psi - \frac{v_i^2}{2} \right)^{\alpha - 3/2} \\ &\quad \times \frac{v_i}{v_j} (v_i + v_j - |v_i - v_j|) dv_i,\end{aligned}\tag{2.25}$$

where the cross section for detectable encounters  $\Sigma = \pi b_{\text{max}}^2$ , and  $K = 4\pi G^2 m_i^2 / \delta v_{\text{min}}^2$ .

We now determine the rate at which stars that will be monitored with an ELT undergo detectable encounters. The integral in equation (2.25) is most easily evaluated in the special case  $\alpha = 3/2$ , which is compatible with current observational constraints (Genzel et al., 2003). To obtain a rough estimate of the rates, consider the case  $\alpha = 3/2$  and assume the background stars all have identical mass not smaller than that of the test star (e.g., they are a population of stellar mass black holes). By equation (2.25)

$$\Gamma_j(\mathbf{r}, \mathbf{v}_j, t) = \frac{3\sqrt{2} K \nu_0}{8 \sigma_h^3} \left( 2\Psi - \frac{1}{3} v_j^2 \right),\tag{2.26}$$

and upon averaging over the orbital phase

$$\begin{aligned}\Gamma_j(e, a) &= \frac{1}{P} \int_0^P \Gamma_j(\mathbf{r}, \mathbf{v}_j, t) dt \\ &= \frac{5\pi\sqrt{2}\nu_0}{2\sigma_h^3} \left( \frac{Gm_i}{\delta v_{\text{min}}} \right)^2 \frac{GM_{\text{bh}}}{a}.\end{aligned}\tag{2.27}$$

Assume the  $N$  stars monitored with an ELT have an eccentricity distribution uniform in  $e^2$  (isotropic velocity ellipsoid) so that  $dN/de da \propto ea^{2-\alpha} = e\sqrt{a}$ . Integrating over these distributions and normalizing to  $N = N(< a_2) \propto a_2^{3/2}$  yields the total rate at which encounters are detected with an ELT:

$$\begin{aligned}\Gamma(\alpha = 3/2) &= \int_0^1 \int_{a_1}^{a_2} \frac{dN(e, a)}{de da} \Gamma_j(e, a) de da \\ &= \frac{15\pi\sqrt{2}\nu_0}{2\sigma_h^3} \left( \frac{Gm_i}{\delta v_{\text{min}}} \right)^2 \frac{GM_{\text{bh}} N}{a_1 + a_2 + \sqrt{a_1 a_2}},\end{aligned}\tag{2.28}$$

where  $a_1$  and  $a_2$  define the range in semi-major axis that is accessible to observations.

Given the above expression for the encounter rate for  $\alpha = 3/2$ , we rely on scaling relations to estimate the rate for different  $\alpha$ . Since the encounter rate is proportional to the stellar density,  $\Gamma(\alpha) \simeq \Gamma(3/2)(r_h/r)^{\alpha-3/2}$ . Thus, if  $\alpha = 7/4$ , the rate of encounters is  $\sim 3$  times larger than for  $\alpha = 3/2$ . The time scale for detectable encounters is therefore

$$\begin{aligned} \Gamma^{-1} \sim & 0.3 \text{ yr} \left( \frac{m_i \nu_0}{2 \times 10^5 M_\odot \text{ pc}^{-3}} \right)^{-1} \left( \frac{m_i}{10 M_\odot} \right)^{-1} \left( \frac{N}{100} \right)^{-1} \\ & \times \left( \frac{a_2}{3000 \text{ AU}} \right) \left( \frac{\sigma_h}{100 \text{ km s}^{-1}} \right)^3 \left( \frac{\delta\theta}{0.5 \text{ mas}} \right)^2 \\ & \times \left( \frac{T}{10 \text{ yr}} \right)^{-2} \left( \frac{N_{\text{obs}}}{100} \right) \left( \frac{a_2}{r_h} \right)^{\alpha-3/2}, \end{aligned} \quad (2.29)$$

where we use the results of § 2.2 that  $N \approx 100$ ,  $a_1 \simeq 200 \text{ AU}$ , and  $a_2 \simeq 3000 \text{ AU}$ , and have also assumed that the mass density of background particles  $m_i \nu_0$  is constant and independent of  $m_i$ .

Therefore, assuming a density cusp dominated by  $\sim 10 M_\odot$  black holes,  $\sim 30$  nearby stellar encounters will be detectable during ten years of monitoring with an ELT with  $\delta\theta = 0.5 \text{ mas}$ . Measurement of the frequency of detectable orbital deflections  $\Gamma \propto m_i$  is a direct test of the average mass of the dark remnants that probably dominate the mass density near the black hole (Morris, 1993; Miralda-Escudé & Gould, 2000) but are otherwise not directly detectable. Since  $N(< a) \propto a^{3/2}$ , then by equation (2.28),  $\Gamma \propto a^{1/2}$ , i.e., the encounter rate increases with distance from the massive black hole. The stars at  $a > 3000 \text{ AU}$  with detectable linear proper motion may therefore yield the strongest constraint on the mass function of stellar remnants, despite being below the threshold for detecting accelerated motion due to the massive black hole.

## 2.4 Method

In this section we describe how we generate mock ELT orbital data. We also describe our implementation of the Markov Chain Monte Carlo (MCMC) method, which we use to estimate the uncertainties in the orbital parameters and map the shape of the likelihood surface.

### 2.4.1 Parameter estimation

We are interested in estimating the uncertainties in the parameters given proper motion and radial velocity information for a sample of  $N$  stars orbiting the massive black hole at the GC. Each star’s projected orbit is described by six phase-space parameters. The black hole mass, its three-dimensional position, and the normalization and slope of the extended matter distribution, contribute an additional six parameters. The entire parameter space of our model therefore has dimension  $J = 6N + 6$ .

Parameter estimation on a  $J$ -dimensional grid is not practical. Since the computational cost of the grid-based approach increases exponentially with  $J$ , the parameter space becomes prohibitively large for even just two or three stars. By contrast, the cost of the MCMC method scales almost linearly with  $J$ .

We now briefly describe the basic ideas of the MCMC method and our choice of implementation. A general discussion of the theory and application of the MCMC approach is given in Gilks, Richardson, & Spiegelhalter (1996). Readers not interested in the details of our parameter estimation scheme can skip ahead to § 2.4.2.

Let  $D$  denote the observed data,  $\theta$  the model parameters,  $P(\theta)$  the prior distribution (which is uniform here), and  $L(D|\theta)$  the likelihood of detecting the data for a given set of parameter values. By Bayes’s theorem the distribution of  $\theta$  conditioned on  $D$  is given by

$$\pi(\theta|D) = \frac{P(\theta)L(D|\theta)}{\int P(\theta)L(D|\theta)d\theta}, \quad (2.30)$$

and is called the posterior distribution of  $\theta$ . The statistical properties of the parameters such as means, moments, and confidence contour levels, are entirely specified by

$\pi(\theta|D)$ .

Explicit evaluation of the integral in the denominator of equation (2.30) is not practical in large dimensional models. The MCMC method avoids evaluating the integral by instead generating a Markov chain of parameter points the distribution of which converges to the posterior distribution  $\pi(\theta|D)$ . The Markov aspect refers to the property that the probability distribution of the  $n$ th state (i.e., point) in the chain  $\theta_n$  depends only on the previous state  $\theta_{n-1}$ . It can be shown (e.g., Gilks, Richardson, & Spiegelhalter 1996) that the density of points in a Markov chain converges to  $\pi(\theta|D)$  if the following criteria are satisfied: (1) the chain is irreducible, namely from any starting state  $\theta_0$  the chain can reach any non-empty set with positive probability in some finite number of iterations; (2) the chain is aperiodic in that it does not oscillate between different sets of states in a regular periodic fashion; (3) the chain is positive recurrent, meaning that if the initial value  $\theta_0$  is sampled from the posterior then the expected time (i.e., number of iterations) to return arbitrarily close to state  $\theta_0$  is finite. There are several algorithms for generating Markov chains that satisfy the above properties. We use the Metropolis algorithm (Metropolis et al., 1953) in our numerical calculations.

Our implementation of the Metropolis algorithm is as follows:

1. Start a chain at  $t = 0$  with some initial state  $\theta_0$ .
2. Generate a trial state  $\theta'$  according to the jump proposal distribution  $q(\theta'|\theta_t)$  (see below). Compute

$$\alpha(\theta_t, \theta') = \min \left[ 1, \frac{L(D|\theta')}{L(D|\theta_t)} \right]. \quad (2.31)$$

3. Sample a uniform random variable  $U$  that lies between  $(0, 1)$ .
4. If  $U \leq \alpha(\theta_t, \theta')$  then set  $\theta_{t+1} = \theta'$  (i.e., accept the jump). If  $U > \alpha(\theta_t, \theta')$  then set  $\theta_{t+1} = \theta_t$ .
5. Increment  $t$ .

6. Go to step #2.

If the observational errors follow a normal distribution,  $L(D|\theta) \simeq \exp[-\chi^2(\theta, D)/2]$ . The  $\chi^2(\theta, D)$  statistic for a single star is given by

$$\begin{aligned} \chi^2(\theta, D) = & \sum_{i=1}^M \left\{ \frac{[x_i(\theta) - x_i(D)]^2}{\sigma_{x,i}^2} + \frac{[y_i(\theta) - y_i(D)]^2}{\sigma_{y,i}^2} \right\} \\ & + \sum_{j=1}^K \frac{[v_j(\theta) - v_j(D)]^2}{\sigma_{v,j}^2}, \end{aligned} \quad (2.32)$$

where  $(x, y)$  is the astrometric position of the star,  $v$  its radial velocity, and  $\sigma$  the corresponding measurement errors (i.e.,  $\sigma_{x,y} = \delta\theta$  and  $\sigma_v = \delta v$ ). We simultaneously fit to multiple stars by summing each star's  $\chi^2$  to form a cumulative  $\chi^2$  for the model.

The jump proposal distribution  $q(\theta'|\theta_t)$  is the probability of selecting a trial state  $\theta'$  given the current state  $\theta_t$ . For the Metropolis algorithm one considers only symmetric proposals of the form  $q(\theta'|\theta_t) = q(\theta_t|\theta')$ . We choose to model the jump distribution as a multivariate normal distribution with mean  $\theta_t$  and constant covariance matrix  $C$ .

Although the distribution of points in a chain is independent of the form of the jump distribution once the Markov chain has converged, the time it takes a chain to converge is sensitive to the jump distribution. To ensure an efficient run one must carefully choose the shape and step size of the jump distribution. An ideal jump distribution has a shape and step size that not only minimizes the convergence time but also samples the entire posterior distribution efficiently. In our implementation the shape of the jump distribution is determined by  $C$  and the step size is determined by a constant scale factor multiplying  $C$ .

$C$  is chosen such that the shape of the jump distribution is similar to that of the posterior distribution, although we again emphasize that the shape is only important for the efficiency of convergence. This ensures that the chain mixes well even in regions of degeneracy. To this aim, we compute the covariance matrix that describes the shape of the  $\chi^2$  surface in the neighborhood of its minimum. We first compute the approximate best-fit parameter state  $\theta_{\text{bf}}$  by minimizing  $\chi^2$ . We then specify a pilot

covariance matrix  $C_p$  that is purely diagonal with variances given by a reasonable guess of the  $1\sigma$  uncertainties for individual parameters. We draw a number ( $\sim 1000$ ) of pilot points from a multivariate normal distribution with mean  $\theta_{\text{bf}}$  and covariance  $C_p$ . Since the pilot points are within  $\sim 1\sigma$  of the  $\chi^2$  minimum, the shape of the  $\chi^2$  surface in the region of the points is approximately quadratic. We solve the linear least-squares problem by fitting a quadratic  $\chi^2$  model to the points and obtain the approximate Fisher matrix that describes the curvature of the  $\chi^2$  surface. We then determine the eigenvalues and eigenvectors of the Fisher matrix. If any of the eigenvalues are negative, indicating that the shape of the surface is unconstrained in some direction, we generate more pilot points and redo the linear least-squares fit. Finally, we invert the resulting Fisher matrix to obtain the covariance matrix  $C$ .

The constant scale factor that determines the step sizes must also be carefully chosen to ensure that the chain effectively explores the parameter space. If the steps are too small the chain does not mix well, as it stays in one region of the parameter space for long periods of time. If the steps are too large, the trial states are rejected frequently. For a multivariate normal jump distribution the most efficient step sizes are those for which  $\sim 25\%$  of the jump proposals are accepted (see Gelman 1995). We chose the (constant) jump scale factor to optimize the acceptance rate.

The first steps in a chain may be sensitive to the starting state  $\theta_0$  and are therefore not sampled from the posterior distribution. We discard these initial “burn in” points. To ensure that a chain has converged and is sampling the full posterior distribution we run multiple chains each starting at widely dispersed states. We tested for convergence with the Gelman-Rubin test statistic (Gilks, Richardson, & Spiegelhalter, 1996).

### 2.4.2 Mock data

To generate a realistic set of orbital data we must determine: (i) the number of stars  $N$  we can detect and monitor with an ELT, (ii) the spatial distribution of these stars, (iii) the number of observations per year per star and, (iv) the observational errors in the stellar positions and velocities. In § 2.2 we showed that with a 30 meter

ELT the position of the stars can be centroided to an astrometric precision  $\delta\theta_{30}$  between  $0.1 - 0.5$  mas and the radial velocities measured to accuracies  $\delta v_{30}$  between  $1 - 10$  km s<sup>-1</sup>. We found that with such an ELT we can detect the accelerated proper motion of approximately 100 stars. We estimate that an integral-field spectrograph on an ELT enables a dedicated GC observing program to obtain the positions and velocities of each of the 100 stars roughly ten times per year.

A realistic mock data set might therefore consist of  $N = 100$  stars, observed over a ten year baseline with ten observations per year per star, with position and velocity measurements for each star accurate to 0.5 mas and 10 km s<sup>-1</sup>. Unfortunately, running our MCMC simulation on such a large data set was not feasible due to limits in computational speed. A typical run requires  $\sim 10^7$  iterations (i.e., jumps) in order to fully sample the posterior distribution. This corresponds to a minimum of  $\sim 3$  days on a desktop machine for just 20 stars ( $J \simeq 126$ ) with 100 points per star; a simulation with 100 stars takes approximately five times longer. However, one can obtain realistic results from a reduced sample size by properly scaling the  $\chi^2$  values (see equation 2.32) to emulate the full sample size. In particular, we construct a mock data set with  $N = 20$  and multiply the  $\chi^2$  of each star by a factor of five.

This approach yields realistic estimates of parameter uncertainties as long as the mock data set with  $N = 20$  stars fairly represents the full data set with 100 stars. We minimize the effects of sample variance as follows. We first generate data for 1000 synthetic orbits. These orbits are drawn from the distribution function of the power-law density profile assuming randomly oriented orbits and considering only those orbits with semi-major axes in the range detectable with an ELT (see § 2.2). We generate mock data for these orbits assuming Gaussian position and velocity errors with dispersions  $\delta\theta$  and  $\delta v$  and a specific input model for the potential (e.g., black hole plus extended matter). For each individual star, we compute the difference in  $\chi^2$  between the best-fit model (essentially the model used to generate the data) and the null hypothesis model (e.g., no extended matter). We then rank the stars by the size of this  $\chi^2$  difference. We bin the 1000 stars into  $N$  bins according to their rank and randomly select one star from each bin. The resulting  $N$  stars form the set of

orbits to which we fit.

Table 2.1 lists the orbital parameters for one realization of a sample of 20 stars to which we fit. Given the orbital parameters, we generate mock data by solving the equation of motion for each star (see equation 2.8). In Figure 2.2 we show the astrometric positions of the 20 stars over the ten year observational baseline with ten epochs per year. The values of the input model parameters describing the potential are:  $M_{\text{bh}} = 4 \times 10^6 M_{\odot}$ ,  $R_0 = 8$  kpc,  $(x_{\text{bh}}, y_{\text{bh}}) = (0, 0)$ ,  $M_{\text{ext}}(r < 0.01 \text{ pc}) = 6000 M_{\odot}$  and either  $\gamma = 1.5$  or 2.

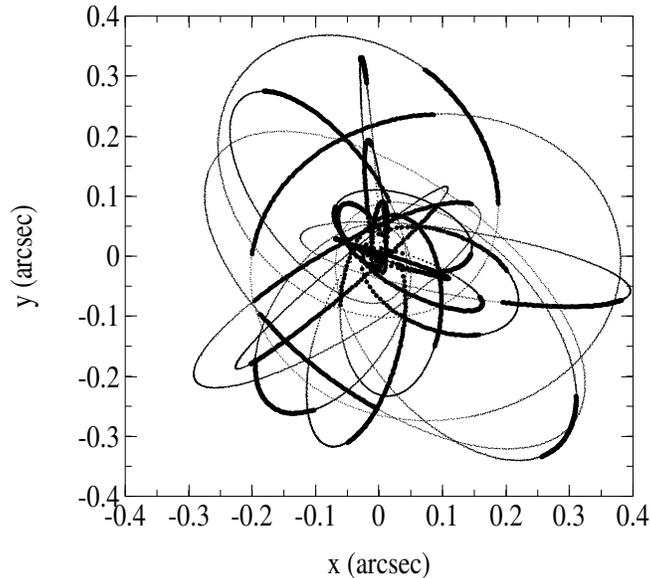


Figure 2.2 Astrometric positions of the 20 synthesized orbits to which we fit the model. The thick-lined portion of each orbit is the proper motion over the fitted 10 year baseline assuming 10 observations per year.

To test that the parameter uncertainty estimates are not affected by sample variance we ran simulations on several different draws of 20 stars. As we show in § 2.5, the parameter uncertainties obtained are similar amongst the different data sets, suggesting that sample variance does not affect the results. Thus, given the current uncertainties in an ELT’s ultimate capabilities, as well as the uncertainty in the exact nature of the stellar distribution at the GC, we conclude that to a reasonable

Table 2.1. Stellar orbital parameters

Star	$P$ (yr)	$a$ (AU)	$e$	$r_{\min}$ (AU)	$i$ (deg)
1	2.9	325	0.852	48	89
2	5.2	476	0.587	197	104
3	5.4	488	0.783	106	46
4	9.5	712	0.560	313	86
5	13.4	895	0.800	179	114
6	14.2	931	0.772	212	81
7	17.7	1078	0.800	216	15
8	23.2	1289	0.605	510	126
9	25.2	1363	0.903	132	25
10	26.6	1413	0.730	382	142
11	36.8	1755	0.705	517	56
12	40.4	1869	0.890	206	93
13	42.3	1928	0.885	223	105
14	46.6	2056	0.703	611	42
15	50.6	2172	0.552	973	104
16	51.5	2197	0.567	951	145
17	52.2	2218	0.439	1245	106
18	52.5	2225	0.636	810	95
19	57.8	2372	0.319	1617	26
20	78.1	2901	0.220	2264	60

Note. — The listed parameters are orbital period ( $P$ ), semi-major axis ( $a$ ), eccentricity ( $e$ ), pericenter distance ( $r_{\min}$ ), and inclination ( $i$ ). They are not all independent variables.

approximation a mock data set comprising  $N = 20$  stars with  $\chi^2$  values increased fivefold yields parameter uncertainties similar to that expected with observations by a 30 meter ELT.

We also show in the next section that the orbital parameter constraints scale with the measurement errors  $\sigma$  and number of stars  $N$  as  $\sigma/N^{1/2}$ . Results for a wide range of assumed ELT capabilities (i.e., different  $\delta\theta$  and  $\delta v$ , different aperture, etc.) can therefore be computed by scaling the results of our fiducial 30 meter ELT model, using the relations between  $N$ , aperture  $D$ ,  $\delta\theta$ , and  $\delta v$ , given in § 2.2.

## 2.5 Results

In this section we investigate how well observations with an ELT constrain the structure of the GC. Our model of the GC and the orbits was described in § 2.3. We draw stellar orbital parameters from a phase-space distribution determined by the model and use these orbits to synthesize mock ELT data (see § 2.4). We then fit a model to the mock data and calculate the uncertainties in the parameters using the MCMC technique discussed in § 2.4.1. We show results for a 30 meter ELT with  $(\delta\theta, \delta v) = (0.5 \text{ mas}, 10 \text{ km s}^{-1})$  and  $(\delta\theta, \delta v) = (0.1 \text{ mas}, 2 \text{ km s}^{-1})$ . However, since the parameter uncertainties scale with measurement error  $\sigma$  and number of monitored stars  $N$  as  $\sigma/N^{1/2}$ , the results can be used to describe the capabilities of an ELT with different specifications. For example, a 100 meter ELT will detect  $\sim 10\times$  as many stars (§ 2.2.2); if the telescope has astrometric and spectroscopic errors that are smaller than those of a 30 meter telescope by a factor of 5 the parameter uncertainties will be  $\sim 10\times$  smaller. In this section, we estimate the limits that can be placed on the parameters associated with the black hole including  $M_{\text{bh}}$  and  $R_0$  (§ 2.5.1), as well as on the extended distribution of (dark) matter near the black hole (§ 2.5.2). We discuss the dependence of the limits on the astrometric and spectroscopic precision of the observations. We also investigate whether relativistic corrections to the Keplerian motion can be detected at the GC (§ 2.5.3).

### 2.5.1 Measuring $M_{\text{bh}}$ and $R_0$

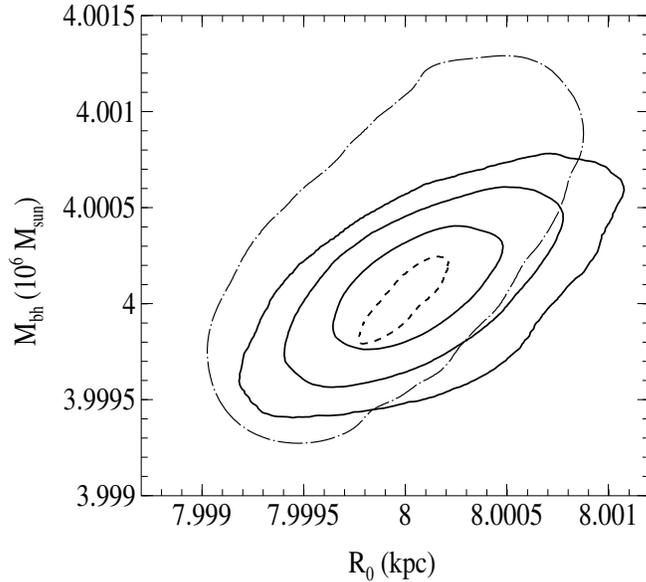


Figure 2.3 Constraint on  $M_{\text{bh}}$  and  $R_0$  obtainable with an ELT assuming an extended matter distribution with  $\gamma = 1.5$  (results are similar for  $\gamma = 2$ ). The solid contours show the 68%, 95%, and 99.7% confidence levels assuming an astrometric limit of  $\delta\theta = 0.5$  mas and a spectroscopic limit of  $\delta v = 10 \text{ km s}^{-1}$  for the draw of 20 stars shown in Table 2.1. The line-dot contour shows the 99.7% confidence level for a different draw of 20 stars. The dashed contour shows the 99.7% confidence level for smaller astrometric and spectroscopic limits of  $\delta\theta = 0.1$  mas and  $\delta v = 2 \text{ km s}^{-1}$ . An ELT will constrain both  $M_{\text{bh}}$  and  $R_0$  to better than 0.1%.

In Figure 2.3 we show the constraints an ELT will place on  $M_{\text{bh}}$  and  $R_0$ . For an astrometric limit of  $\delta\theta = 0.5$  mas and a spectroscopic limit of  $\delta v = 10 \text{ km s}^{-1}$  (see § 2.2.1) the fractional uncertainties in  $M_{\text{bh}}$  and  $R_0$  are less than 0.1% at the 99.7% level. This is a factor of  $\sim 100$  better than present uncertainties. The result is robust in that simulations with distinct mock data sets of 20 stars, drawn in the fashion described in § 2.4.2, produce similar uncertainties in the model parameters.

For astrometric and spectroscopic limits that are a factor of 5 smaller the fractional uncertainties in  $M_{\text{bh}}$  and  $R_0$  are smaller by almost a factor of 5. The uncertainties in  $M_{\text{bh}}$  and  $R_0$  scale almost linearly with the measurement errors for observations at

this precision. We also verified that the uncertainties scale with  $N$  as roughly  $N^{-1/2}$ .

Observations with a 30 meter ELT will therefore constrain the distance to the GC to within a few parsecs and the mass of the black hole to within a few thousand solar masses. We discuss the implications of measuring  $R_0$  to such high accuracy in § 3.4.

## 2.5.2 Measuring the extended matter distribution

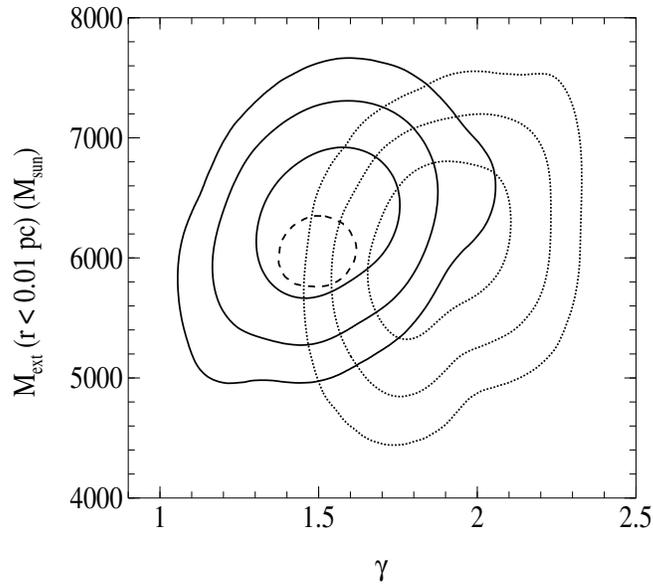


Figure 2.4 Constraint on the extended matter distribution obtainable with an ELT. Shown are the 68%, 95%, and 99.7% confidence levels on the enclosed mass and slope of an extended matter distribution assuming an astrometric limit of  $\delta\theta = 0.5$  mas and a spectroscopic limit of  $\delta v = 10 \text{ km s}^{-1}$ . The input models have power-law slope of  $\gamma = 1.5$  and  $\gamma = 2$  and an input enclosed mass of  $6000M_{\odot}$  within 0.01 pc. The dashed contour is the constraint at the 99.7% level for measurement errors that are a factor of 5 smaller.

In Figure 2.4 we show the constraints an ELT will place on the extended matter distribution for input power-law models with  $M_{\text{ext}}(r < 0.01 \text{ pc}) = 6000M_{\odot}$  and  $\gamma = 1.5$  or  $\gamma = 2$ . We chose these distributions in order to conform to the extrapolation of the observed stellar density distribution and to theoretical estimates of dark matter clustering (see § 3.1). We find that for an ELT with  $\delta\theta = 0.5$  mas and  $\delta v = 10 \text{ km s}^{-1}$

one can detect such extended matter distributions, yielding measurements of  $M_{\text{ext}}$  and  $\gamma$  that are accurate to 20 – 30% (i.e.,  $\delta M_{\text{ext}} \sim 1500M_{\odot}$  and  $\delta\gamma \sim 0.5$ ). Since the amplitude of the Newtonian retrograde precession varies linearly with  $M_{\text{ext}}$  (§ 2.3.2), the fractional uncertainty is  $\delta M_{\text{ext}}/M_{\text{ext}} \propto \delta\Delta\phi_{\text{Newt}}/\Delta\phi_{\text{Newt}} \propto M_{\text{ext}}^{-1}$ , where  $\delta\Delta\phi_{\text{Newt}}$  is set by the astrometric precision. Thus  $\delta M_{\text{ext}}$  is independent of  $M_{\text{ext}}$  so that an extended matter distribution is detectable (i.e., observations yield a lower bound) for  $\delta\theta = 0.5$  mas and  $\delta v = 10 \text{ km s}^{-1}$  as long as  $M_{\text{ext}}(r < 0.01 \text{ pc}) \gtrsim \delta M_{\text{ext}} \sim 1500M_{\odot}$ . Such an ELT will therefore place interesting constraints on the extended matter at the GC.

### 2.5.3 Measuring relativistic effects

As discussed in § 2.3.3, order of magnitude estimates suggest that post-Newtonian corrections to the equations of motion, involving terms of order  $(v/c)^2$ , are measurable with an ELT with astrometric resolution of  $\delta\theta \lesssim 0.5$  mas. In an effort to demonstrate this more quantitatively, we allow the speed of light to be a parameter in our model and examine how well we recover its value. We purposely do not include relativistic corrections to the observed motion associated with propagation effects (e.g., the Roemer time delay and other higher-order corrections) so that we can examine the detectability of  $(v/c)^2$  general relativistic corrections to the orbital dynamics such as the prograde precession of the major axis position. In Figure 2.5 we show the constraint on  $c$  as a function of  $M_{\text{enc}}$ . Post-Newtonian effects are observable, as  $c$  is measured to  $\sim 5\%$  accuracy. Since  $v/c \lesssim 0.2$  for all stars in the sample (§ 2.3.1) the few percent constraint on  $c$  suggests that while the  $(v/c)^2$  effects are measurable, the  $(v/c)^3$  effects are not. The orbital precession due to black hole spin is of order  $(v/c)^3$  (§ 2.3.4) and detecting it with an ELT with  $\delta\theta = 0.5$  mas requires the favorable discovery of a star on a compact and highly eccentric orbit. Based on estimates of the signal-to-noise from a spin-induced apocenter shift (equation [2.15]), an astrometric precision of  $\sim 0.05$  mas is needed to reliably detect the black hole spin.

The degeneracy between  $c$  and  $M_{\text{enc}}$  is a consequence of the degeneracy between

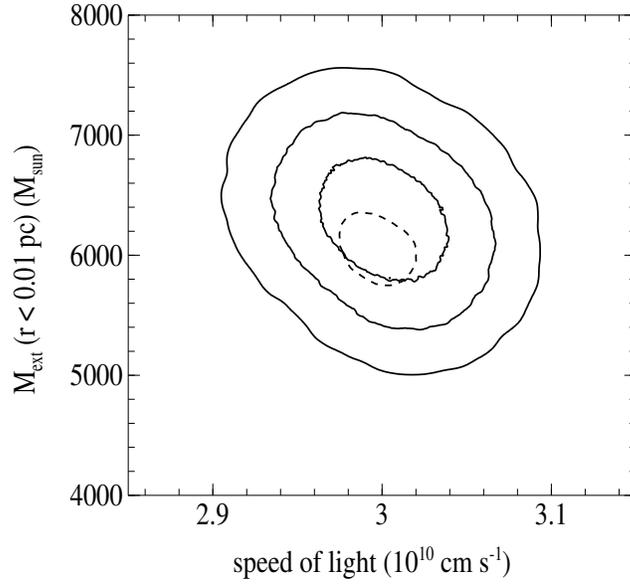


Figure 2.5 An ELT’s sensitivity to post-Newtonian effects assuming an astrometric limit of  $\delta\theta = 0.5$  mas and a spectroscopic limit of  $\delta v = 10$  km s<sup>-1</sup>. Shown is the uncertainty in the speed of light and the extended matter mass as obtained by including post-Newtonian corrections to the equations of motion. The Roemer delay and special relativistic effects are not included in the model in order to demonstrate that general relativistic effects of order  $(v/c)^2$ , including the prograde precession, are detectable with an ELT. The line styles are the same as in Figure 2.3.

the prograde relativistic precession and the retrograde Newtonian precession. Decreasing  $c$  increases the amount of prograde motion  $\Delta\phi_{\text{pro}}$  while increasing  $M_{\text{enc}}$  increases the amount of retrograde motion  $\Delta\phi_{\text{Newt}}$ . The two effects compensate for one another over a range of  $c$  and  $M_{\text{enc}}$ . The degeneracy is broken at sufficiently extreme values of  $M_{\text{enc}}$  because the relativistic and Newtonian effects each induce a distinct precessional shape.

## 2.6 Constraints on Galactic structure from measurements of $R_0$

The distance to the GC,  $R_0$ , is a fundamental parameter in models of the Milky Way structure. As Olling & Merrifield (2001) note, models of the Milky Way exhibit strong interrelations between the Galactic constants ( $R_0$  and the local Galactic rotation speed  $\Theta_0$ ), the shortest-to-longest axis ratio,  $q = c/a$ , of the dark matter halo, and the local stellar column density  $\Sigma_*$ . The determination of  $q$  is of particular interest since different models of dark-matter and structure formation scenarios predict different values for  $q$ . Cold dark-matter simulations typically produce galactic halos that are triaxial (Warren et al., 1992; Jing & Suto, 2002) although these tend to become oblate under the influence of the dissipative infall of gas resulting in halos with  $q \simeq 0.5$  (Dubinski, 1994). Alternatively, hot dark-matter models predict round halos with  $q \sim 0.8$  (Peebles, 1993) while some baryonic dark-matter models imply  $q \sim 0.2$  (Pfenniger, Combes, & Martinet, 1994). As we now discuss, determining  $R_0$  to 0.1% via monitoring of stellar orbits at the GC with an ELT enables an extremely precise measurement of  $q$  in the Milky Way.

Olling & Merrifield (2000) demonstrate that there is significant uncertainty in existing estimates of  $q$  in galaxies due to both the limited amount of data available for measuring  $q$  and the fact that different measurement techniques have yielded systematically different values. Presently, the situation is not any better for our own Galaxy, with plausible values lying in the range  $0.3 \lesssim q \lesssim 1$ .

The measurement of  $q$  in the Milky Way entails measuring the Galaxy's radial mass distribution and the degree to which this mass distribution is flattened. Olling & Merrifield (2000) show that the uncertainty in  $q$  in the Milky Way is almost entirely due to the large errors in the Galactic constants  $\Theta_0$  and  $R_0$ . Indeed, Olling & Merrifield (2001) show that the fractional uncertainty in  $q$  is nearly twice the fractional uncertainty in  $\Theta_0$ . Therefore, a precision measurement of the Sun's proper motion with respect to the GC in combination with a precision measurement of  $R_0$  tightly constrains  $\Theta_0$  and hence  $q$ . According to Salim, Gould, & Olling (2002) future astrometric

surveys will be able to measure the Sun’s proper motion  $\mu = V/R_0$  to within several microarcseconds, corresponding to 0.1% accuracy. Here  $V = \Theta_0 + V_\odot \simeq 220 \text{ km s}^{-1}$  is the sum of the rotation speed of the local standard of rest and the Sun’s motion relative to it. The uncertainty in  $\Theta_0$  will be the dominant error in  $V$ ;  $V_\odot$  is already known to an accuracy of  $0.6 \text{ km s}^{-1}$  from the *Hipparcos* catalogue (Dehnen & Binney, 1998). Thus, the monitoring of stellar orbits at the GC with an ELT in conjunction with future astrometric survey missions will constrain the Milky Way’s dark-matter halo shape parameter  $q$  to a few tenths of a percent.

## 2.7 Conclusions

We have examined a variety of experiments that can be achieved through the infrared monitoring with an ELT of stars within a few thousand AU of the GC. The astrometric limit of a 30 meter ELT is conservatively 0.5 mas and possibly as high as 0.1 mas. By comparison, the astrometric limit of current observations is 1 – 2 mas.

The greater point-source sensitivity and spectral resolution of an ELT enables the measurement of radial velocities with errors  $\lesssim 10 \text{ km s}^{-1}$ . At present, of the  $\sim 10$  stars with measured accelerated proper motions, spectral lines have been detected only in S0-2, with radial velocity uncertainties of  $\sim 30 \text{ km s}^{-1}$ . Measuring the radial velocities of stars breaks the degeneracy between mass and distance and thus yields a direct measurement of the distance to the GC. If the spectra of fainter stars can be obtained, the detection of deep molecular lines will improve upon the velocity estimates by an additional factor  $\times 10$ . The solar type stars that will be detectable with an ELT may therefore yield radial velocity uncertainties considerably smaller than  $10 \text{ km s}^{-1}$ .

A 30 meter ELT will be able to detect stars down to a  $K$ -band magnitude of  $K \sim 22$ , approximately four magnitudes fainter than currently possible. Due to confusion, it will be difficult to detect still fainter stars. Using measurements of the  $K$ -band luminosity function within the inner  $1''$  of the GC, we estimate that such an ELT will detect the accelerated motion of  $\sim 100$  stars with semi-major axes in

the range  $200 \lesssim a \lesssim 3000$  AU. Current observations are limited to the detection of  $\sim 10$  stars, all with  $a \gtrsim 1000$  AU. We find that the number of stars with detectable accelerated motion scales with the aperture of an ELT as  $N \simeq 100(D/30 \text{ m})^2$ .

Given the observational capabilities of an ELT and the likely, albeit at low masses largely uncertain, stellar environment at the GC, we constructed a plausible sample of stellar orbits. The model includes the dynamical contribution of an extended distribution of dark matter around the black hole that is composed of stellar remnants and CDM. We find that for measurements at the precision obtainable with an ELT the uncertainty in the model parameters scale with the measurement errors  $\sigma$  (i.e.,  $\delta\theta$ ,  $\delta v$ ) and the number of monitored stars  $N$  as roughly  $\sigma/N^{1/2}$ . Thus, while we focus on the capabilities of a diffraction limited 30 meter ELT with  $\delta\theta = 0.5$  mas and  $\delta v = 10 \text{ km s}^{-1}$ , our results can be used to determine the capabilities of an ELT with different specifications. For example, a 100 meter ELT will detect  $\sim 10\times$  as many stars so that if it has astrometric and spectroscopic errors that are smaller by a factor of 5, the measurement accuracy in the parameters will improve by a factor of approximately ten.

We find that with a 30 meter ELT the parameters  $M_{\text{bh}}$  and  $R_0$  will be measured to an accuracy better than 0.1%. Determining  $R_0$  to within a few parsecs will significantly constrain models of the Galactic structure as it aids the precise measurement of the dark matter halo shape.

While current observations of stellar proper motions are compatible with Keplerian motion, a number of dynamical effects produce significant deviations, including the Newtonian retrograde precession, the relativistic prograde precession, frame dragging due to the black hole spin, and interstellar interactions involving nearby encounters. All but the frame dragging effect produce non-Keplerian motions that are detectable with a 30 meter ELT. Unfortunately, the spin of the massive black hole at the GC will probably be out of reach to kinematic studies unless an astrometric precision of  $\sim 0.05$  mas is achieved.

The presence of an extended distribution of matter results in a Newtonian retrograde precession due to differences in the amount of mass enclosed within an orbit's

pericenter and apocenter. We considered extended matter density profiles consistent with current observations of the stellar distribution at the GC. We modeled the distribution as a power-law profile normalized such that  $M_{\text{ext}}(r < 0.01 \text{ pc}) = 6000M_{\odot}$  and with slope  $\gamma = 1.5$  or 2. Standard models of dark matter clustering about a massive black hole predict similar profiles. An orbit monitoring program with a 30 meter ELT will constrain the mass and slope of such profiles to  $\sim 30\%$  accuracy. Thus, monitoring orbits with an ELT provides a probe of the extended matter distribution within  $\sim 10^4$  Schwarzschild radii of the massive black hole at the GC.

We also calculated the rate at which the monitored stars experience detectable deflections due to stellar gravitational scattering encounters with background compact remnants. We considered a detection threshold set by the minimum detectable change in the velocity of a monitored star. For a density cusp dominated by  $\sim 10M_{\odot}$  black holes,  $\sim 30$  nearby stellar encounters will be detected by a 30 meter ELT over a ten year observing baseline. This will confirm the presence of a cusp of compact remnants at the GC and enable the measurement of the remnants' masses.

The authors would like to thank M. Kamionkowski, D. Figer, K. Matthews, and E. Pfahl for comments, and E. Agol, S. Phinney, and J. Graham for helpful discussions. We also thank the referees for their helpful comments. NNW acknowledges the support of an NSF graduate fellowship and DoE DE-FG03-92ER40701. MM was supported at Caltech by a postdoctoral fellowship from the Sherman Fairchild Foundation.

## 2.A Current constraints on $R_0$ and the extended matter distribution

In collaboration with Andrea Ghez's GC group at UCLA, we present here estimates of  $R_0$  and limits on the extended matter distribution based on an analysis of the orbital

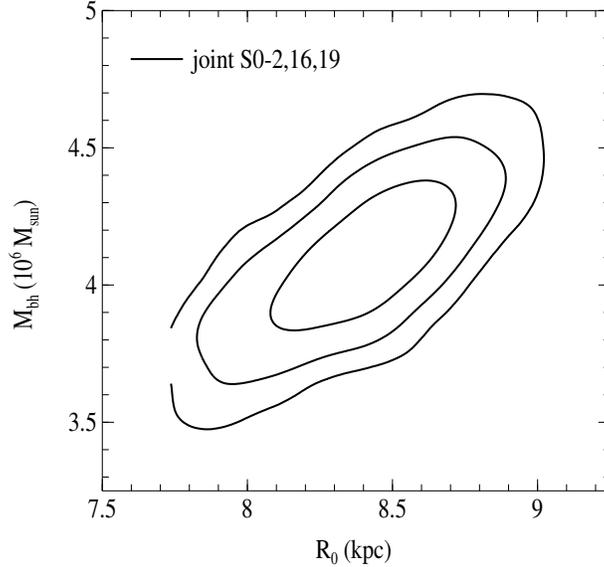


Figure 2.6 Constraint on  $M_{\text{bh}}$  and  $R_0$  from joint fits to the orbital motions of the stars S0-2, S0-16, and S0-19.

motion measurements obtained with Keck over the last decade. The astrometric and radial velocity measurements are described in part in Ghez et al. (2005) and Ghez et al. (2003), respectively; the most recent data along with the constraints presented here are described in an upcoming paper. We perform a joint fit to the orbital motions of the stars S0-2, S0-16, and S0-19 using the methods described in § 2.4.

In Figure 2.6 we show the current constraint on  $M_{\text{bh}}$  and  $R_0$  as obtained with the joint orbital fits. The measured value for  $R_0$  is in agreement with the values obtained using other, less direct, techniques and the uncertainty is comparable in magnitude (cf., Reid 1993). Another group using VLT observations of orbital motions at the GC have obtained a similar measurement of  $R_0$  (Eisenhauer et al., 2005).

In Figure 2.7 we show the limits on  $M_{\text{ext}}(r < 0.01 \text{ pc})$  and  $\gamma$  assuming the power-law model described in § 2.3.2. Currently, one can only place an upper-bound on the extended mass corresponding to  $\sim 10\%$  of the black hole mass. Theoretical estimates predict an extended mass of  $\sim 10^3 M_{\odot}$  within 0.01 pc, well below the current upper-bound.

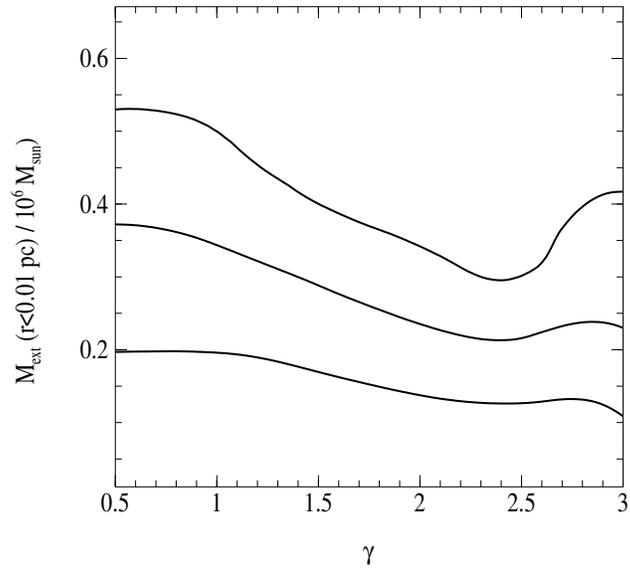


Figure 2.7 Constraint on the extended matter distribution from joint fits to the orbital motions of the stars S0-2, S0-16, and S0-19.

As one expects given the current data, the orbital motions are fit equally well by relativistic and non-relativistic models.

# Bibliography

- Alexander, T. & Sternberg, A. 1999, *ApJ*, 520, 137
- Backer, D. C. & Sramek, R. A. 1999, *ApJ*, 524, 805
- Baganoff, F. K. et al. 2001, *Nature*, 413, 45
- Baganoff, F. K. et al. 2003, *ApJ*, 591, 891
- Bahcall, J. N. & Wolf, R. A. 1976, *ApJ*, 209, 214
- Bahcall, J. N. & Wolf, R. A. 1977, *ApJ*, 216, 883
- Beckert, T., Duschl, W. J., Mezger, P. G., & Zylka, R. 1996, *A&A*, 307, 450
- Becklin, E. E. 1968, Ph.D. Thesis,
- Dehnen, W. & Binney, J. J. 1998, *MNRAS*, 298, 387
- Dubinski, J. 1994, *ApJ*, 431, 617
- Eckart, A., Genzel, R., Hofmann, R., Sams, B. J., & Tacconi-Garman, L. E. 1995, *ApJL*, 445, L23
- Eckart, A., Genzel, R., Ott, T., & Schödel, R. 2002, *MNRAS*, 331, 917
- Eisenhauer, F., Schödel, R., Genzel, R., Ott, T., Tecza, M., Abuter, R., Eckart, A., & Alexander, T. 2003, *ApJL*, 597, L121
- Eisenhauer, F., et al. 2005 (astro-ph/0502129)
- Figer, D. F. et al. 2000, *ApJL*, 533, L49

- Figer, D. F. et al. 2003, *ApJ*, 599, 1139
- Fragile, P. C. & Mathews, G. J. 2000, *ApJ*, 542, 328
- Franceschini, A., Toffolatti, L., Danese, L., & de Zotti, G. 1989, *ApJ*, 344, 35
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 1995, *Bayesian Data Analysis* (New York: Chapman & Hall/CRC)
- Genzel, R., Pichon, C., Eckart, A., Gerhard, O. E., & Ott, T. 2000, *MNRAS*, 317, 348
- Genzel, R. et al. 2003a, *ApJ*, 594, 812
- Genzel, R., Schödel, R., Ott, T., Eckart, A., Alexander, T., Lacombe, F., Rouan, D., & Aschenbach, B. 2003b, *Nature*, 425, 934
- Gerhard, O. 2001, *ApJL*, 546, L39
- Gezari, S., Ghez, A. M., Becklin, E. E., Larkin, J., McLean, I. S., & Morris, M. 2002, *ApJ*, 576, 790
- Ghez, A. M., Klein, B. L., Morris, M., & Becklin, E. E. 1998, *ApJ*, 509, 678
- Ghez, A. M. et al. 2003a, *ApJL*, 586, L127
- Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M., Becklin, E. E., & Duchêne, G. 2005, *ApJ*, 620, 744
- Ghez, A. M., Wright, S. A., Matthews, K., Thompson, D., Le Mignant, D., Tanner, A., Hornstein, S. D., Morris, M., Becklin, E. E., Soifer, B. T. 2004, *ApJ*, 601, L159
- Gilks, W.R., Richardson, S., & Spiegelhalter, D.J. 1996, *Markov Chain Monte Carlo in Practice* (New York: Chapman & Hall/CRC)
- Gnedin, O. Y. & Primack, J. R. 2003, preprint (astro-ph/0308385)

- Goldwurm, A., Brion, E., Goldoni, P., Ferrando, P., Daigne, F., Decourchelle, A., Warwick, R. S., & Predehl, P. 2003, *ApJ*, 584, 751
- Gondolo, P. & Silk, J. 1999, *Physical Review Letters*, 83, 1719
- Gould, A. & Quillen, A. C. 2003, *ApJ*, 592, 935
- Hansen, B. M. S. & Milosavljević, M. 2003, *ApJL*, 593, L77
- Hornstein, S. D., Ghez, A. M., Tanner, A., Morris, M., Becklin, E. E., & Wizinowich, P. 2002, *ApJL*, 577, L9
- Ipsier, J. R. & Sikivie, P. 1987, *Phys. Rev. D*, 35, 3695
- Jackson, J. M., Geis, N., Genzel, R., Harris, A. I., Madden, S., Poglitsch, A., Stacey, G. J., & Townes, C. H. 1993, *ApJ*, 402, 173
- Jaroszynski, M. 1998, *Acta Astronomica*, 48, 653
- Jing, Y. P. & Suto, Y. 2002, *ApJ*, 574, 538
- Kim, S. S. & Morris, M. 2003, *ApJ*, 597, 312
- Krabbe, A. et al. 1995, *ApJL*, 447, L95
- Landau, L. & Lifshitz, D. 1960, *Mechanics* (Oxford: Pergamon Press)
- Levin, Y. & Beloborodov, A. M. 2003, *ApJL*, 590, L33
- Loeb, A. 2003, preprint (astro-ph/0309716)
- Magorrian, J. & Tremaine, S. 1999, *MNRAS*, 309, 447
- Merritt, D., Milosavljević, M., Verde, L., & Jimenez, R. 2002, *Physical Review Letters*, 88, 191301
- Merritt, D. 2003, preprint (astro-ph/0311594)
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. 1953, *J. Chem. Phys.*, 21, 1087

- Milosavljević, M. & Loeb, A. 2004, 604, L45
- Miralda-Escudé, J. & Gould, A. 2000, ApJ, 545, 847
- Morris, M. 1993, ApJ, 408, 496
- Munyanenza, F., Viollier, R. D. 1999, preprint (astro-ph/9910566)
- Murphy, B. W., Cohn, H. N., & Durisen, R. H. 1991, ApJ, 370, 60
- Murray, C. D., Dermott, S. F. 1999, Solar System Dynamics (Cambridge: Cambridge University Press)
- Olling, R. P. & Merrifield, M. R. 2000, MNRAS, 311, 361 (OM00)
- Olling, R. P. & Merrifield, M. R. 2001, MNRAS, 326, 164
- Peebles, P. J. E. 1972, ApJ, 178, 371
- Peebles, P. J. E. 1993, Principles of Physical Cosmology (Princeton, NJ: Princeton University Press)
- Pfahl, E. & Loeb, A. 2003, preprint (astro-ph/0309744)
- Pfenniger, D., Combes, F., & Martinet, L. 1994, A&A, 285, 79
- Porquet, D., Predehl, P., Aschenbach, B., Grosso, N., Goldwurm, A., Goldoni, P., Warwick, R. S., & Decourchelle, A. 2003, A&A, 407, L17
- Quinlan, G. D., Hernquist, L., & Sigurdsson, S. 1995, ApJ, 440, 554
- Reid, M. J. 1993, ARAA, 31, 345
- Reid, M. J., Readhead, A. C. S., Vermeulen, R. C., & Treuhaft, R. N. 1999, ApJ, 524, 816
- Rogers, A. E. E. et al. 1994, ApJL, 434, L59
- Salim, S., Gould, A., & Olling, R. P. 2002, ApJ, 573, 631

- Rieke, G. H., Rieke, M. J., & Paul, A. E. 1989, *ApJ*, 336, 752
- Rubilar, G. F. & Eckart, A. 2001, *A&A*, 374, 95
- Sanders, R. H. & Lowinger, T. 1972, *AJ*, 77, 292
- Schödel, R. et al. 2002, *Nature*, 419, 694
- Schödel, R., Ott, T., Genzel, R., Eckart, A., Mouawad, N., & Alexander, T. 2003, *ApJ*, 596, 1015
- Serabyn, E., Carlstrom, J., Lay, O., Lis, D. C., Hunter, T. R., & Lacy, J. H. 1997, *ApJL*, 490, L77
- Shao, M. & Colavita, M. M. 1992, *A&A*, 262, 353
- Spitzer, L. 1987, *Dynamical evolution of globular clusters* (Princeton, NJ: Princeton University Press)
- Takeuchi, T. T., Kawabe, R., Kohno, K., Nakanishi, K., Ishii, T. T., Hirashita, H., & Yoshikawa, K. 2001, *Publ. Astron. Soc. Pacific*, 113, 586
- Warren, M. S., Quinn, P. J., Salmon, J. K., & Zurek, W. H. 1992, *ApJ*, 399, 405
- Weinberg, S. 1972, *Gravitation and cosmology: Principles and applications of the general theory of relativity* (New York: Wiley)
- Ullio, P., Zhao, H., & Kamionkowski, M. 2001, *Phys. Rev. D*, 64, 43504
- Young, P. 1980, *ApJ*, 242, 1232
- Yu, Q. 2003, *MNRAS*, 339, 189
- Zhao, J., Bower, G. C., & Goss, W. M. 2001, *ApJL*, 547, L29

## Chapter 3

# Finding Isolated Stellar Remnants at the Galactic Center

MILOŠ MILOSAVLJEVIĆ AND NEVIN N. WEINBERG

Theoretical Astrophysics, California Institute of Technology, Pasadena, CA 91125

### Abstract

Although the central parsec of the Galaxy was host to intense star formation activity in the recent  $(1 - 3) \times 10^7$  yr, the stellar mass function inside the sphere of dynamical influence of the massive black hole (MBH) Sagittarius A\* at the Galactic center is completely unknown. A measurement of the mass function is particularly desirable as the mass density is likely dominated by massive remnants, including stellar-mass black holes, that collect near the MBH because of dynamical segregation. The cluster of massive remnants is expected theoretically but has eluded observations. Its existence is of critical importance for the prospects for detecting gravitational radiation from the capture of compact objects by central MBHs in other galaxies. Here we describe how the monitoring of stellar proper motions can be used to directly probe the masses of isolated stellar remnants near the MBH. We derive expressions for the rate at which perturbations from remnants deflect stellar orbits and describe how the remnant masses can be extracted from the monitoring data. We discuss alternative astronomical signatures of the concentration of remnants near Sgr A\* and suggest that the X-ray transients discovered by Muno et al. (2005) are isolated stellar-mass black holes accreting from dense molecular clouds.

---

### 3.1 Introduction

The monitoring of bright stars near the massive black hole (MBH) at the Galactic center with speckle imaging and adaptive optics has yielded stellar proper motions and revealed orbital solutions (e.g., Schödel et al. 2002; Ghez et al. 2003). A combination of proper motions and radial velocity data were used to independently constrain the mass of, and distance to, the MBH (Eisenhauer et al., 2003). The central parsec of the Galaxy was recently host to intense star formation activity. A number of B stars (the “Galactic center cluster”), Wolf-Rayet stars, and luminous supergiants orbit within the dynamical sphere of influence of the MBH (e.g., Krabbe et al. 1995; Genzel et al. 2000; Gezari et al. 2002; Ghez et al. 2003; Genzel et al. 2003). These massive stars might be a frosting on an invisible cake: The mass density should be dominated by massive stellar remnants that have eluded detection.

The most massive stars and remnants sink toward the MBH via dynamical friction, displacing the less massive ones to larger radii. Although the OB-type stars may be the most massive objects while on the main sequence, the stellar-mass black holes are the most massive species on average. The black holes form a cusp in which the density is a power law in radius  $\rho \propto r^{-7/4}$  (Bahcall & Wolf, 1976, 1977; Murphy, Cohn, & Durisen, 1991; Preto, Merritt, & Spurzem, 2004). Miralda-Escudé & Gould (2000) estimate that  $\sim 20,000$  stellar-mass black holes could have collected within the central parsec (also see early analysis by Phinney 1989 and Morris 1993). The rate at which the stellar-mass black holes are captured by the MBH is  $\sim 10^{-6} - 10^{-7} \text{ yr}^{-1}$  (Sigurdsson & Rees, 1997; Freitag, 2003; Gair et al., 2004), much too low to deplete them.

A direct measurement of the stellar black hole masses would be of immense benefit given current uncertainties in the black hole mass function. The mass of the black hole that is produced in the core collapse of a very massive star with solar or super-solar metallicity is sensitive to mass loss after the star becomes a Wolf-Rayet star (Woosley, Heger, & Weaver, 2002). The mass loss is not severe for the first generation of metal-poor, “Population III” stars, and these could have left behind a population

of  $\sim 100M_{\odot}$  black holes (e.g., Madau & Rees 2001). Only 19 X-ray binaries have been dynamically confirmed to contain compact objects with masses above  $3M_{\odot}$ ; some of these have masses uncertain by a factor of 2 (Orosz, 2003).

The Galactic center is an ideal place to look for exceptionally massive remnants. The region is host to star-formation as evidenced by the young stars in the Galactic center cluster and in the nearby Arches and Quintuplet clusters. Quintuplet contains one of the most luminous stars known, the ‘‘Pistol star’’ (Figer et al., 1998), which is a candidate progenitor of a black hole more massive than found in X-ray binaries. Compact star clusters may be sites of runaway stellar coalescence leading to the formation of supermassive stars that collapse into intermediate-mass black holes (IMBHs; Portegies Zwart & McMillan 2002; Gürkan, Freitag, & Rasio 2004). The clusters sink in the gravitational potential of the Galactic bulge via dynamical friction and are disrupted by the MBH, thereby delivering the IMBHs into the MBH’s sphere of dynamical influence (Hansen & Milosavljević, 2003; Yu & Tremaine, 2003; Kim, Figer, & Morris, 2004; Levin et al., 2005).

We here describe how the monitoring of stellar proper motions can be used to directly probe the mass function of stellar remnants near the MBH. A similar technique was discussed by Ashurov (2004) in the context of globular clusters. In § 3.2, we derive expressions for the rate at which perturbations from remnants deflect stellar orbits. In § 3.3, we describe the procedure by which the remnant masses and densities can be extracted from the monitoring data. In § 3.4, we discuss some other observable signatures of the remnants.

## 3.2 Gravitational interactions

Gravitational interactions deflect neighboring bodies. We work under the approximation that the interaction timescale is much shorter than the orbital period and the period over which the motion of a star is being monitored. Both conditions are satisfied for stars that traverse a large fraction of an orbit over a decade. An exception are the stars in a secular resonant relation with a perturber; these resonances, however,

do not affect the semimajor axis (Rauch & Tremaine, 1996).

Discreteness of the stellar gravitational potential, which is a sum of a large number of point-mass potentials, induces diffusion of the orbital elements of a test star. Occasional nearby encounters (“large-angle scattering”) induce instantaneous, large deflections that are not described by diffusion. We here calculate the evolution of the semimajor axis probability distribution due to orbital diffusion and nearby encounters.

The focus on the semimajor axis is arbitrary. Other orbital elements are also perturbed; Similar expressions to those for the semimajor axis can be derived for the eccentricity, the inclination, etc.

Near the black hole, the semimajor axis is related to the orbital energy via  $a = GM_{\text{bh}}/2E$ , where  $M_{\text{bh}} \approx 4 \times 10^6 M_{\odot}$  is the mass of the MBH; we use the semimajor axis and the energy interchangeably in what follows. We first calculate the orbital diffusion due to multiple, small-angle deflections, induced by the discreteness of the stellar gravitational potential. The diffusion of the probability  $N(E, t)dE$  that a star subject to such encounters has energy between  $E$  and  $E + dE$  at time  $t$  is described by the orbit-averaged Fokker-Planck equation. Since we are interested in the evolution of the semimajor axis, we ignore the angular momentum terms in the Fokker-Planck equation. This simplifies the calculation and does not significantly affect our estimates. It also simplifies the forthcoming analysis to assume that the mass of the test star equals the mass  $m$  of the perturbers. Though artificial, this assumption introduces small inaccuracies in our estimates; generalization to a spectrum of stellar and perturber masses is straightforward but tedious.

The Fokker-Planck equation reads

$$\frac{\partial N}{\partial t} = -\frac{\partial}{\partial E} (\langle \Delta E \rangle_t N) + \frac{1}{2} \frac{\partial^2}{\partial E^2} (\langle (\Delta E)^2 \rangle_t N), \quad (3.1)$$

where  $\langle \Delta E \rangle_t$  and  $\langle (\Delta E)^2 \rangle_t$  are the orbit-averaged diffusion coefficients; the averaging is carried out as in equation (3.7).

Expressions for the diffusion coefficients in the case of an isotropic background of

perturbers are given in equations (A20) and (A21) of Cohn (1979), which are written in terms of the phase space distribution of the perturbers  $f(E_p)$ . For a general power-law profile, the isotropic distribution function is given by (Magorrian & Tremaine, 1999)

$$f_\gamma(E_p) = \frac{\nu_0}{(2\pi\sigma_0^2)^{3/2}} \frac{\Gamma(\gamma+1)}{\Gamma(\gamma-1/2)} \left(\frac{E_p}{\sigma_0^2}\right)^{\gamma-3/2}, \quad (3.2)$$

where  $\nu_0$  is the number density of the perturbers at the radius  $r_0 \sim GM_{\text{bh}}/\sigma_0^2 \sim 1$  pc and  $\sigma_0 \sim 100$  km s<sup>-1</sup> is the one-dimensional stellar velocity dispersion just outside of  $r_0$ . A realistic choice for the number density is  $\nu_0 \sim 2 \times 10^5 (m/M_\odot)^{-1}$  pc<sup>-3</sup>.

The distribution function attains the simplest form, that of a constant  $f_{3/2} = 3\sqrt{2}\nu_0/16\pi\sigma_0^3$ , for a spatial density profile that is a power law  $\rho \propto r^{-\gamma}$  with  $\gamma = 3/2$ . This profile is compatible with the observed stellar number density distribution (Genzel et al., 2003), and is slightly shallower than the density profile  $\gamma = 7/4$  that the most massive bodies would settle into as a result of dynamical friction and diffusion (see § 3.1). The constant is a conservative choice for the distribution function that further simplifies the calculation.

Substituting  $f_{3/2}$  into the expressions for the diffusion coefficients we obtain  $\langle \Delta E \rangle = (4/3)\pi\Gamma f_{3/2}(2\varphi - 5E)$  and  $\langle (\Delta E)^2 \rangle = (8/15)\pi\Gamma f_{3/2}v^2(2\varphi + 3E)$ , where  $\Gamma = 4\pi G^2 m^2 \ln \Lambda$ ,  $\ln \Lambda$  is the Coulomb logarithm, and  $\varphi(r) = GM_{\text{bh}}/r$  is the gravitational potential of the black hole. Upon orbit averaging,  $\langle \Delta E \rangle_t = -4\pi\Gamma E f_{3/2}/3$  and  $\langle (\Delta E)^2 \rangle_t = -16\pi\Gamma E^2 f_{3/2}(1 - 8/\sqrt{1 - e^2})/15$ .

The solution to equation (3.1), written as a probability  $N(a_0, a, t)$  that a star with initial semimajor axis  $a_0 = GM_{\text{bh}}/2E_0$  ends up on an orbit with semimajor axis  $a = GM_{\text{bh}}/2E$  after time  $t$  is given by

$$N(a_0, a, t) = \frac{\exp\{-[\ln(a_0/a) + b_2\lambda t]^2/4b_1\lambda t\}}{a\sqrt{4\pi b_1\lambda t}}, \quad (3.3)$$

where  $\lambda = 4\pi\Gamma f_{3/2}/15$ ,  $b_1 = 16(1 - e^2)^{-1/2} - 2$ , and  $b_2 = 16(1 - e^2)^{-1/2} + 3$ .

Next, we turn to an alternative statistical description of stellar encounters, which explicitly accounts for single scattering events that result in large deflections. We

calculate the orbit-averaged rate  $K(E, E')dE'$  at which a star with specific binding energy  $E$  experiences an encounter in which its energy changes to a value between  $E'$  and  $E' + dE'$ . The rate is evaluated by time-averaging the local rate  $K(E, E', r)$  along the initial orbit. The local rate was derived by Goodman (1983) for globular clusters but is equally applicable to objects orbiting a massive black hole. In the notation of Goodman,  $\Delta E = E' - E$  is the star's energy change,  $v = \sqrt{2(\varphi - E)}$  and  $v' = \sqrt{2(\varphi - E')}$  are the velocities of the star before and after the encounter,  $E_p$  is the energy of the perturber, while  $v_p = \sqrt{2(\varphi - E_p)}$  and  $v'_p = \sqrt{2(\varphi - E_p + \Delta E)}$  are the velocities of the perturber before and after the encounter. The rate equals

$$K(E, E', r) = \frac{2\pi^2 G^2 m^2}{v|\Delta E|^3} \begin{cases} A(E, E', r) + B(E, E', r) \\ C(E, E', r) + D(E, E', r) \end{cases}, \quad (3.4)$$

where the upper term should be used for  $\Delta E < 0$ , the lower for  $\Delta E > 0$ , and

$$\begin{aligned} A(E, E', r) &= \int_0^{E'} f(E_p) \left( \frac{8}{3}v^2 - 4\Delta E \right) v dE_p, \\ B(E, E', r) &= \int_{E'}^{\varphi(r)+\Delta E} f(E_p) \left( \frac{8}{3}v_p'^2 - 4\Delta E \right) v'_p dE_p, \\ C(E, E', r) &= \int_0^{E'} f(E_p) \left( \frac{8}{3}v'^2 + 4\Delta E \right) v' dE_p, \\ D(E, E', r) &= \int_{E'}^{\varphi(r)} f(E_p) \left( \frac{8}{3}v_p^2 + 4\Delta E \right) v_p dE_p. \end{aligned} \quad (3.5)$$

Substituting  $f_{3/2}$  into the integrals yields the transition rate

$$\begin{aligned} K(E, E', r) &= \frac{\sqrt{2}\pi}{10} \frac{G^2 m^2 \nu_0}{\sigma_0^3 |E' - E|^3} \\ &\times \begin{cases} 10\varphi E' + 5E'E - 15E'^2 + 8\varphi^2 - 6\varphi E - 2E^2 & (\text{if } E' < E), \\ \sqrt{\frac{\varphi - E'}{\varphi - E}} (14\varphi E' - 5E'E - 7E'^2 + 8\varphi^2 - 10\varphi E) & (\text{if } E' > E). \end{cases} \end{aligned} \quad (3.6)$$

For an eccentric orbit with eccentricity  $e$ , the radial dependence is removed by taking

the orbital average,

$$\begin{aligned} K(E, E') &= P(E)^{-1} \oint K(E, E', r(t)) dt \\ &= 2P(E)^{-1} \int_{r_-}^{r_+} \frac{K(E, E', r)}{\sqrt{2(\varphi - E) - (1 - e^2)\varphi^2/2E}} dr, \end{aligned} \quad (3.7)$$

where  $r_-$  and  $r_+$  are the turning points at which the argument of the square root in equation (3.7) vanishes and  $P(E) = 2^{-1/2}\pi GM_{\text{bh}}E^{-3/2}$  is the orbital period of the star.

The probability  $N(E, t)dE$  that a star has energy between  $E$  and  $E + dE$  at time  $t$  evolves according to

$$\frac{\partial N(E, t)}{\partial t} = \int_0^{2E} [N(E', t)K(E', E) - N(E, t)K(E, E')] dE'. \quad (3.8)$$

The first term on the right hand side of equation (3.8) represents the increase due to scattering from another energy into  $E$ , and the second represents the decrease due to scattering from  $E$  into another energy. The integral in equation (3.8) is logarithmically divergent which is a disguised form of the Coulomb logarithm (Goodman, 1983).

Assuming the initial energy of the star  $E_0$  is precisely known, for times short compared to the relaxation time  $\sim E/\langle(\Delta E)^2\rangle^{1/2}$  we expand  $N(E, t)$  in order of  $t$  and keep only the linear order in equation (3.8) to obtain

$$N(E, t) \approx tK(E_0, E) \quad (3.9)$$

The expansion in equation (3.9) is valid only where  $|E - E_0|$  is larger than the diffusion width  $\sim t\langle(\Delta E)^2\rangle^{1/2}$ .

For the special case of a circular orbit of energy  $E$  scattered onto an eccentric orbit with energy  $E'$ ,  $\varphi(r) = 2E$ . Changing to semimajor axes in the place of energies we obtain

$$N(a_0, a, t) = \frac{\sqrt{2}\pi}{10} \frac{a_0 t}{a^2} \frac{G^2 m^2 \nu_0}{\sigma_0^3} \tilde{K}_{\text{circ}}(a_0/a) \quad (\text{circular}), \quad (3.10)$$

where

$$\tilde{K}_{\text{circ}}(x) = |1 - x|^{-3} \times \begin{cases} 18 + 25x - 15x^2 & (\text{if } x < 1), \\ \sqrt{2 - x}(12 + 23x - 7x^2) & (\text{if } x > 1). \end{cases} \quad (3.11)$$

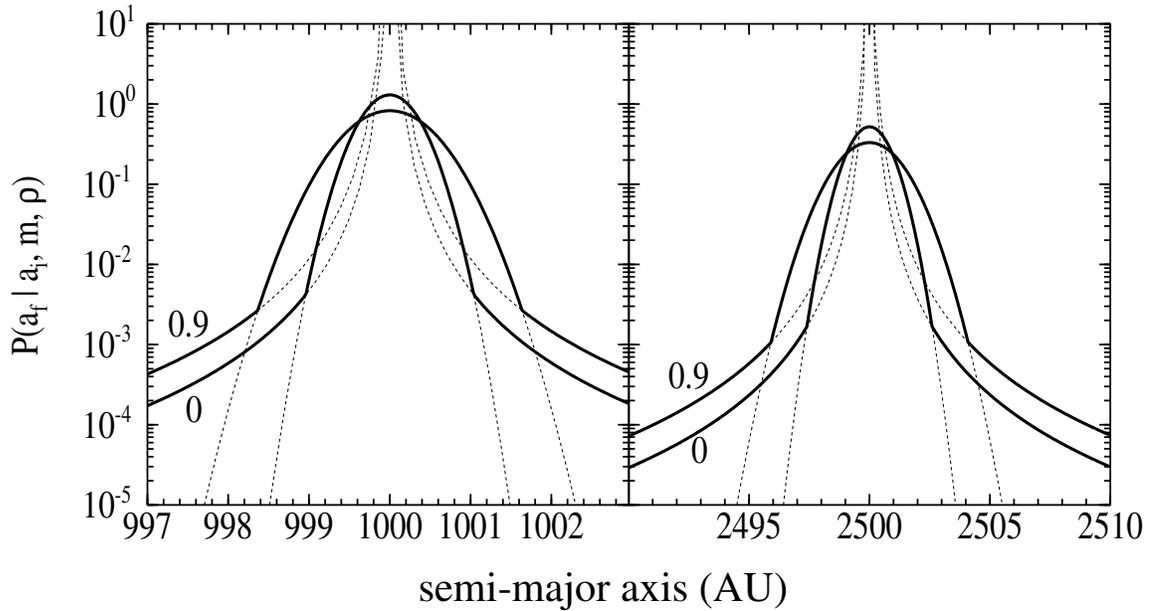


Figure 3.1 Probability that after  $T = 10$  yr the semi-major axis of a star is changed from an initial value  $a_i$  to a final value  $a_f$  due gravitational encounters encounters for  $a_i = 1000$  AU (left panel) and  $a_i = 2500$  AU (right panel) with  $e = 0$  and  $e = 0.9$  assuming  $m = 10M_{\odot}$ .

In Figure 3.1 we show the probability per unit semimajor axis that a star with initial semi-major axis  $a_0$  ends up on an orbit with semi-major axis  $a$  in a time  $T = 10$  yr. For small  $|\Delta a| = |a - a_0|$ , where the linear expansion in equation (3.9) is invalid, diffusion due to multiple, small-angle deflections is the correct description and the probability is given by equation (3.3). For large  $|\Delta a|$ , large-angle deflections dominate over diffusion and the solution is given by equation (3.9), or by equation (3.10) for the special case of a circular orbit. In both regimes the dependence on the density of scatterers, their mass, and the time, occurs strictly through the product

*mpt.*

### 3.3 Parameter estimation

We here derive the posterior likelihood on the mass and the density of perturbers  $P(m, \rho | \{\tilde{a}_{i,k}, \tilde{a}_{f,k}\})$  given a set of initial and final measured semimajor axes  $\tilde{a}_{i,k}$  and  $\tilde{a}_{f,k}$ . We use tilde to denote measured quantities and the index  $k$  ranges from 1 to  $N_{\text{star}}$ , where  $N_{\text{star}}$  is the number of monitored stars. Lacking prior knowledge of  $m$  and  $\rho$ , Bayes's theorem tells us that the posterior likelihood equals

$$P(m, \rho | \{\tilde{a}_{i,k}, \tilde{a}_{f,k}\}) = \prod_{k=1}^{N_{\text{star}}} \frac{P(\tilde{a}_{i,k}, \tilde{a}_{f,k} | m, \rho)}{\int \int P(\tilde{a}_{i,k}, \tilde{a}_{f,k} | m, \rho) dm d\rho}. \quad (3.12)$$

Let  $P(\tilde{a}|a)$  be the probability that a measurement of the semimajor axis yields the value  $\tilde{a}$  given that the true semimajor axis equals  $a$ . Then the probability that the measured semimajor axis changes from  $\tilde{a}_i$  to  $\tilde{a}_f$  is given by (for clarity we omit index  $k$ )

$$P(\tilde{a}_i, \tilde{a}_f | m, \rho) = \int \int P(\tilde{a}_i | a_i) P(\tilde{a}_f | a_f) P(a_i, a_f | m, \rho) da_i da_f. \quad (3.13)$$

The probability that a star with true semimajor axis  $a_i$  is scattered onto  $a_f$  is the product of the star having semimajor axis  $a_i$  with the transition probability calculated in § 3.2,  $P(a_i, a_f | m, \rho) = P(a_i)N(a_i, a)$ , where  $P(a)$  is the distribution of test star semimajor axes, and  $N(a_i, a)$  was defined in § 3.2 and depends on  $m$  and  $\rho$ .

Finally, the uncertainty in the measurement of the semimajor axis must be quantified. It depends on the eccentricity and the inclination of the orbit. We here ignore these details and model the uncertainty via

$$P(\tilde{a}|a) = \delta^{-1} \sqrt{N_{\text{obs}}/2\pi} e^{-N_{\text{obs}}(a-\tilde{a})^2/2\delta^2}, \quad (3.14)$$

where  $N_{\text{obs}}$  is the number of observations of a star, and  $\delta$  is the characteristic astrometric uncertainty in determining the stellar position in three dimensions. This picture applies only to short-period stars with orbits that can be monitored over at

least a large fraction of the orbit.

An estimate of  $\delta$  for a 30 meter class telescope is 0.5 mas (Weinberg, Milosavljević, & Ghez, 2004), which at the Galactic center translates to  $\delta \sim 4$  AU. While keeping in mind that the method is sensitive only to the product  $m\rho$ , we set the density of the perturbers to  $\rho = 2 \times 10^5 (r/1 \text{ pc})^{-3/2} M_\odot \text{ pc}^{-3}$ . We calculate  $P(m|\rho, \{\tilde{a}_{i,k}, \tilde{a}_{f,k}\})$  with  $t = 10$  yr,  $N_{\text{obs}} = 100$ ,  $a_0 = 3000$  AU, and a prior  $m > 0.5M_\odot$ . We generate mock data sets  $\{\tilde{a}_{i,k}, \tilde{a}_{f,k}\}$  using  $m = (5, 10)M_\odot$ . For given  $m$ , data sets drawn from the same distribution yield slightly different posterior likelihoods. To obtain unique likelihoods, we carried out a geometric averaging over a large number of realizations. The resulting posterior likelihoods are shown in Figure 3.2. If only 10 stars are available at  $a_0 = 1000$  AU, no constraints can be obtained on the mass.

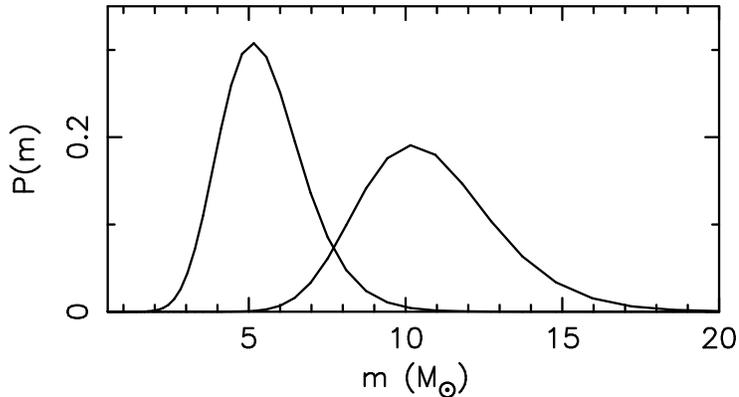


Figure 3.2 Posterior likelihood distributions  $P(m|\rho, \{\tilde{a}_{i,k}, \tilde{a}_{f,k}\})$  for a sample of 100 stars on circular orbits at  $a_0 = 3000$  AU generated for  $m = (5, 10)M_\odot$ . For details see § 3.3.

To measure the mass independently one must either assume prior knowledge of the density, or measure the density via another method. The total mass density can be determined via accurate astrometry of stars orbiting the MBH, since the extended mass distribution in remnants induces Newtonian retrograde precession in the stellar orbits. Its amplitude is proportional to the density (Rubilar & Eckart, 2001), and will be measurable with a 30 meter class telescope (Weinberg, Milosavljević, & Ghez, 2004).

We now review the limitations and the caveats of our simplified calculation. The

choice to restrict to the evolution of the semimajor axis is conservative. Other stellar orbital elements also incur changes and this enhances the power to constrain the perturber mass. In § 3.2, the choice to equate the perturber mass to that of the test star is valid for  $5 - 10M_{\odot}$  black hole perturbers and B-type stars, which lie in the same mass range. In § 3.3, the choice of Gaussian measurement uncertainty on the semimajor axis (equation 3.14) is ad hoc. The correct uncertainty would differ between stars and would be a function of the inclination of the orbit and of the fraction of the orbit covered by the monitoring. The choice to restrict to circular orbits while evaluating the posterior likelihood is conservative, since eccentric orbits diffuse a little faster, as evident in Figure 3.1. Finally, the choice of density profile with logarithmic slope  $\gamma = 3/2$  is conservative given that  $\gamma = 7/4$  is expected; the latter slope would imply  $\sim 3$  times larger rate of encounters, although the overall normalization of the density is sufficiently uncertain to preclude a definite prediction.

### 3.4 Discussion

Some other methods might also help detect isolated stellar remnants at the Galactic center. Chanamé, Gould, & Miralda-Escudé (2001) suggested that the cluster of black holes could be detected by monitoring pairs of images of background bulge stars that are lensed by the MBH and looking for microlensing events induced by the black holes. Required for this is photometric completeness at  $K = 23$  mag, which might be within reach of a 30 meter class telescope. Chanamé & Gould (2002) also suggest that one could identify a population of millisecond pulsars and look for a signature of their past dynamical ejection by stellar-mass black holes in the cluster. Unfortunately, pulse broadening makes the detection of any pulsars in the central parsec difficult.

The remnants accrete from the hot plasma permeating the Galactic center region; some of the remnants may also accrete from molecular clouds. The X-ray luminosity of a remnant accreting from the ISM is given by  $L = \epsilon \dot{m} c^2$  where  $\epsilon$  is the X-ray efficiency and  $\dot{m}$  the accretion rate. The plasma temperature at  $1''$  from Sgr A\* is

$\sim 1$  keV (Baganoff et al., 2003) and the relative velocity of the remnants through the hot gas is  $\sim 700$  km s $^{-1}$ . The accretion rate is  $\dot{m} = 4\pi\lambda G^2 m^2 \rho / (v^2 + c_s^2)^{3/2}$  with  $\lambda \lesssim 1$ , where  $c_s \sim 300$  km s $^{-1}$  is the sound speed. The size of the accretion disk that forms around the remnant depends on the pressure gradients across the accretion radius  $r_a = Gm / (v^2 + c_s^2) \sim 2 \times 10^{11} m_{10} / v_7^2$  cm, where  $m = 10 m_{10} M_\odot$  and  $v = 700 v_7$  km s $^{-1}$ . In the Galactic disk, electron density gradients on such scales are  $\ll 1$  (Armstrong et al. 1995; see also Agol & Kamionkowski 2002). However at the Galactic center interstellar turbulence could be much stronger than in the Galactic disk and the gradients could be significant. The luminosity associated with accretion is  $L \sim 10^{30} \epsilon_{-1} \lambda m_{10}^2 / v_7^3$  ergs s $^{-1}$ , where  $\epsilon = 0.1 \epsilon_{-1}$ . The radiative efficiency of Sgr A\* in the X-rays, relative to the estimated mass flow across the Bondi radius at 1'', is  $\epsilon \sim 10^{-10}$ ; such efficiency would render direct detection of remnants in the X-rays hopeless.

A remnant traversing a static, dense molecular cloud with  $n = 10^6$  cm $^{-3}$  will emit at  $\sim 10^{33} \epsilon_{-1} \lambda m_{10}^2 / v_7^3$  ergs s $^{-1}$ . This is consistent with the estimated time-averaged luminosities of the X-ray transients identified with *Chandra* by Munro et al. (2005). These authors suggested that the transients are black hole binaries formed by three-body exchange between a stellar binary and a stellar-mass black hole in the central parsec. If one assumes a high accretion efficiency then the X-ray transients could instead be isolated stellar-mass black holes traversing dense molecular clouds. The emission will be variable on a time scale equal to the orbital time at the accretion radius,  $t_{\text{var}} \sim r_a^{3/2} / \sqrt{Gm} \sim m_{10} / v_7^3$  h, which is compatible with the apparent variability in the sources of Munro et al. Using opacities in Morrison & McCammon (1983), the clouds of size 0.1 pc are marginally optically thin to 1 keV photons.

In §3.2, we limited our analysis to encounters between stellar remnants and stars with measurable orbital motion about the MBH. However, encounters also produce anomalous motion in stars at larger radii ( $r \sim 0.05$  pc), for which accelerated motion due to the MBH is not detectable. The rate of detectable encounters among all stars with measurable linear motion is  $K \sim 3(N_{\text{stars}}/100)(r/1 \text{ pc})^{1/2-\gamma} \text{ yr}^{-1}$  (Weinberg, Milosavljević, & Ghez, 2004). Since  $N_{\text{stars}} \propto r^{3-\gamma}$ , if  $\gamma < 7/4$  the rate increases with

radius. Thus, stars at large radii may also be sensitive probes of the stellar remnant mass function. However, given the probable inhomogeneity of the mass distribution at large radii, it may be difficult to disentangle encounters with remnants from other gravitational interactions experienced by a monitored star.

M. M. was supported at Caltech by a postdoctoral fellowship from the Sherman Fairchild Foundation. N. N. W. acknowledges the support of an NSF graduate fellowship and DoE DE-FG03-92ER40701.

# Bibliography

Agol, E., & Kamionkowski, M. 2002, MNRAS, 334, 553

Armstrong, J. W., Rickett, B. J., & Spangler, S. R. 1995, ApJ, 443, 209

Ashurov, A. E. 2004, AJ, 127, 2154

Baganoff, F. K., et al. 2003, ApJ, 591, 891

Bahcall, J. N. & Wolf, R. A. 1976, ApJ, 209, 214

Bahcall, J. N. & Wolf, R. A. 1977, ApJ, 216, 883

Chanamé, J. & Gould, A. 2002, ApJ, 571, 320

Chanamé, J., Gould, A., & Miralda-Escudé, J. 2001, ApJ, 563, 793

Cohn, H. 1979, ApJ, 234, 1036

Eisenhauer, F., Schödel, R., Genzel, R., Ott, T., Tecza, M., Abuter, R., Eckart, A.,  
& Alexander, T. 2003, ApJL, 597, L121

Figer, D. F., Najarro, F., Morris, M., McLean, I. S., Geballe, T. R., Ghez, A. M., &  
Langer, N. 1998, ApJ, 506, 384

Freitag, M. 2003, ApJL, 583, L21

Gair, J. R., Barack, L., Creighton, T., Cutler, C., Larson, S. L., Phinney, E., &  
Vallisneri, M. 2004, preprint (gr-qc/0405137)

Genzel, R., Pichon, C., Eckart, A., Gerhard, O. E., & Ott, T. 2000, MNRAS, 317,  
348

- Genzel, R. et al. 2003a, ApJ, 594, 812
- Aschenbach, B. 2003b, Nature, 425, 934
- Gezari, S., Ghez, A. M., Becklin, E. E., Larkin, J., McLean, I. S., & Morris, M. 2002, ApJ, 576, 790
- Ghez, A. M. et al. 2003, ApJL, 586, L127
- Goodman, J. 1983, ApJ, 270, 700
- Gürkan, M. A., Freitag, M., & Rasio, F. A. 2004, ApJ, 604, 632
- Hansen, B. M. S. & Milosavljević, M. 2003, ApJL, 593, L77
- Kim, S. S., Figer, D. F., & Morris, M. 2004, ApJL, 607, L123
- Krabbe, A. et al. 1995, ApJL, 447, L95
- Levin, Y., Wu, A. S. P., & Thommes, E. W. 2005, preprint (astro-ph/0502143)
- Maccarone, T. J. 2004, MNRAS, 351, 1049
- Madau, P. & Rees, M. J. 2001, ApJL, 551, L27
- Magorrian, J. & Tremaine, S. 1999, MNRAS, 309, 447
- Miralda-Escudé, J. & Gould, A. 2000, ApJ, 545, 847
- Morris, M. 1993, ApJ, 408, 496
- Morrison, R., & McCammon, D. 1983, ApJ, 270, 119
- Muno, M. P., Pfahl, E., Baganoff, F. K., Brandt, W. N., Ghez, A., Lu, J., & Morris, M. R. 2005, preprint (astro-ph/0412492)
- Murphy, B. W., Cohn, H. N., & Durisen, R. H. 1991, ApJ, 370, 60
- Orosz, J. A. 2003, IAU Symposium, 212, 365
- Phinney, E. S. 1989, IAU Symp. 136: The Center of the Galaxy, 136, 543

- Portegies Zwart, S. F., & McMillan, S. L. W. 2002, *ApJ*, 576, 899
- Preto, M., Merritt, D., & Spurzem, R. 2004, *ApJL*, 613, L109
- Rauch, K. P., & Tremaine, S. 1996, *New Astronomy*, 1, 149
- Rubilar, G. F. & Eckart, A. 2001, *A&A*, 374, 95
- Schödel, R. et al. 2002, *Nature*, 419, 694
- Sigurdsson, S. & Rees, M. J. 1997, *MNRAS*, 284, 318
- Weinberg, N. N., Milosavljević, M., & Ghez, A. M. 2004, preprint (astro-ph/0404407)
- Woosley, S. E., Heger, A., & Weaver, T. A. 2002, *Rev. Mod. Phys.*, 74, 1015
- Yu, Q. 2003, *MNRAS*, 339, 189
- Yu, Q., & Tremaine, S. 2003, *ApJ*, 599, 1129

## Part III

# Weak Gravitational Lensing by Dark Matter Concentrations

## Chapter 4

# Weak Gravitational Lensing by Dark Clusters<sup>†</sup>

NEVIN N. WEINBERG AND MARC KAMIONKOWSKI

California Institute of Technology, Mail Code 130-33, Pasadena, CA 91125 USA

---

<sup>†</sup> A version of this chapter was first published in *MNRAS*, 337, p. 1269-1281 (2002).

### Abstract

We calculate the abundance of dark-matter concentrations that are sufficiently overdense to produce a detectable weak-gravitational-lensing signal. Most of these overdensities are virialized halos containing identifiable X-ray and/or optical clusters. However, a significant fraction are nonvirialized, cluster-mass, overdensities still in the process of gravitational collapse—these should produce significantly weaker or no X-ray emission. Our predicted abundance of such dark clusters is consistent with the abundance implied by the detection of apparent dark lenses. Weak lensing by these nonvirialized objects will need to be considered when determining cosmological parameters with the lens abundance in future weak-lensing surveys. Such weak lenses should also help shed light on the process of cluster formation.

---

## 4.1 Introduction

Weak gravitational lensing due to the deep gravitational potential of a galaxy cluster gives rise to a detectable weak distortion of the images of background galaxies. This weak shear has now been detected around roughly 30 clusters and been used to map the total dark-matter mass in the clusters as well as the dark-matter distributions within the clusters (see Bartelmann & Schneider 2001, Mellier 1999). Weak lensing also has the potential to map the mass distribution on even larger scales (Miralda-Escudé 1991; Blandford et al. 1991; Kaiser 1992; Bartelmann & Schneider 1992; Stebbins 1997; Kamionkowski et al. 1998). Just last year, four groups independently reported detection of cosmic shear, distortions to background galaxies induced by weak gravitational lensing by mass inhomogeneities on *few*-Mpc scales along the line of sight (Bacon et al. 2000; Kaiser et al. 2000; Wittman et al. 2000; Van Waerbeke et al. 2000). It is apparent that in the future, such cosmic-shear surveys will have the sensitivity to identify galaxy clusters in the field. Since such surveys will probe the *total* mass directly, it could provide a powerful new technique for determining the cluster-halo abundance and thus the power-spectrum amplitude  $\sigma_8$  and matter density  $\Omega_m$  (e.g., Kruse & Schneider 1999; Reblinsky et al. 1999).

In fact, one spectroscopically confirmed cluster has already been detected via its gravitational-lensing effect on background galaxies (Wittman et al. 2001). More intriguing is the apparent dark lens discovered by Erben et al. (2000). This lensing signal corresponds to a  $\sim 10^{14} M_\odot$  mass concentration, but there is no obvious corresponding galaxy overdensity (Gray et al. 2001) and only faint (if any) X-ray emission. Evidence for other apparent dark lenses has been reported by Miralles et al. (2002) and Koopmans et al. (2000), the latter involving a detection through strong, rather than weak, lensing.

In retrospect, the existence of such dark concentrations should not come as too much of a surprise. Galaxy clusters form at rare (e.g.,  $> 3\sigma$ ) high-density peaks of a Gaussian primordial distribution. Thus, for every virialized cluster, there should be a significant number of proto-clusters (e.g.,  $2\sigma - 3\sigma$  peaks), mass overdensities that

have not yet undergone gravitational collapse and virialized, but which have begun to break away from the cosmological expansion. The timescale for collapse of cluster-mass objects is large, and the overdensities can be very large even before they have virialized. It should thus not be too surprising if such objects produce a weak-lensing signal that resembles that from virialized clusters.

These proto-clusters should contain galaxies and maybe a few groups that later merge to form the cluster. Since the X-ray luminosity is a very rapidly varying function of the virialized mass, the summed X-ray emission from these objects should be much smaller than that from a fully virialized cluster of the same mass. When we refer to these proto-clusters as “dark,” we thus mean that they should be X-ray underluminous. Strictly speaking, the mass-to-light ratios of these clusters should be comparable to those for ordinary clusters. However, high-redshift clusters may be difficult to pick out in galaxy surveys, and these proto-clusters should have a sky density a few times smaller. Thus it would not be surprising if these dark lenses had no readily apparent corresponding galaxy overdensity.

In this paper we calculate the abundance of dark and virialized lenses. To do so, we first determine the overdensity required to produce a detectable weak-lensing signal as a function of redshift. We consider several different density profiles including a homogeneous sphere, an isothermal sphere, a Navarro, Frenk, & White (Navarro et al. 1997, 1996, 1995) profile and a Hernquist (Hernquist 1990) profile. We then use the spherical-top-hat-collapse (STHC) model to determine the differential abundance of overdensities as a function of position along their evolutionary cycle. Using the aperture mass technique (Schneider 1996) we can then determine the sky density and redshift distribution of halos that are sufficiently overdense to produce a detectable weak-lensing signal.

As our results below will show, there should be roughly one dark lens for every 5–10 virialized lenses discovered by weak lensing. It is worthwhile to point out that this result is robust in that the ratio of dark to virialized lenses is not expected to be very sensitive to the amount of *observational* noise in the lensing map, i.e., observational noise will equally affect the detectability of both types of lenses. Therefore, although

our results are obtained by assuming the only source of noise is the intrinsic ellipticity distribution of the source galaxies—in accordance with other such theoretical weak-lensing studies found in the literature—the predicted relative abundance of dark and virialized lenses will not change very much if we made a more exact estimate of the total noise in weak-lensing maps. It is also encouraging to note that given the sky coverage and average image size of weak-lensing maps to date, the number of dark lenses we would expect to have seen is of order unity and therefore consistent with the detection (Erben et al. 2000, Miralles et al. 2002) of one or two dark lenses.

## 4.2 Minimum overdensity required to produce weak-lensing signal

In this section we provide the conditions for an overdensity of mass  $M$  and radius  $R$  at redshift  $z$  to produce a detectable weak-lensing signal. Following the procedure of Bartelmann & Schneider (2001) (see also Schneider 1996, Seitz & Schneider 1997, Kruse & Schneider 1999) we determine the dependence of a lensing system’s signal-to-noise ratio on that system’s overdensity and redshift.

In a weak-lensing map, a mass overdensity causes the image of the background source galaxies to be tangentially sheared. Noise is introduced by both the intrinsic ellipticity of these background galaxies as well as by the presence of foreground galaxies in the image. To arrive at a signal-to-noise relation for a weak-lensing system, consider  $N$  galaxy images each at angular position  $\boldsymbol{\theta}_i = (\theta_i \cos \phi_i, \theta_i \sin \phi_i)$  with tangential ellipticity  $\epsilon_t(\boldsymbol{\theta}_i)$  and within a lens-centered annulus that is bounded by angular radii  $\theta_{\text{in}} \leq \theta_i \leq \theta_{\text{out}}$ . The shear  $\gamma$  is related linearly to the dimensionless surface mass density of the lens, which is the physical surface mass density  $\Sigma(\boldsymbol{\theta})$  divided by the critical surface mass density  $\Sigma_{\text{crit}}$ . For a lens at redshift  $z_d$  and a source at redshift  $z_s$ ,

$$\Sigma_{\text{crit}}(z_d; z_s) = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}}, \quad (4.1)$$

where  $D_d$ ,  $D_s$  and  $D_{ds}$  are the angular-diameter distances between the lens and the

observer, the source galaxy and the observer, and the lens and the source, respectively. To account for the redshift distribution of the source galaxies, define (Seitz & Schneider 1997)

$$Z(z_s; z_d) \equiv \frac{\lim_{z_s \rightarrow \infty} \Sigma_{\text{crit}}(z_d; z_s)}{\Sigma_{\text{crit}}(z_d; z_s)} = \frac{\Sigma_{\text{crit}\infty}(z_d)}{\Sigma_{\text{crit}}(z_d; z_s)}. \quad (4.2)$$

Then the dimensionless surface mass density is given by

$$\kappa(\boldsymbol{\theta}, z_s) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{\text{crit}}} = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{\text{crit}\infty}} \frac{\Sigma_{\text{crit}\infty}}{\Sigma_{\text{crit}}} \equiv \kappa(\boldsymbol{\theta}) Z(z_s; z_d). \quad (4.3)$$

Furthermore, the linear relation between the shear and surface mass density implies that they have the same dependence on source redshift so that  $\gamma(\boldsymbol{\theta}, z_s) \equiv Z(z_s; z_d)\gamma(\boldsymbol{\theta})$ . For the rest of this paper any reference to  $\kappa$  or  $\gamma$  refers to  $\kappa(\boldsymbol{\theta})$  and  $\gamma(\boldsymbol{\theta})$ , respectively. Assuming the intrinsic orientation of galaxy sources is random, the expectation value of the image ellipticity is (Seitz & Schneider 1997; Bartelmann & Schneider 2001)

$$E(\epsilon) \approx \langle Z \rangle \gamma(\boldsymbol{\theta}), \quad (4.4)$$

where

$$\langle Z \rangle = \int dz_s p_z(z_s) Z(z_s; z_d), \quad (4.5)$$

and  $p_z(z_s)$  is the redshift distribution of source galaxies. The function  $\langle Z \rangle = \langle Z \rangle(z_d)$ —of order unity for the redshifts considered—allows a source redshift distribution to be collapsed onto a single redshift  $z_s$  satisfying  $Z(z_s) = \langle Z \rangle$  (see Bartelmann & Schneider 2001).

Using the  $M_{\text{ap}}$ -statistics introduced by Schneider (1996), define a discretized estimator for the spatially filtered mass inside a circular aperture of angular radius  $\theta$ ,

$$M_{\text{ap}} \equiv \frac{1}{n} \sum_{i=1}^N \epsilon_t(\boldsymbol{\theta}_i) Q(|\boldsymbol{\theta}_i|), \quad (4.6)$$

where  $n$  is the number density of galaxy images and  $Q$  is a weight function that will be chosen later so as to maximize the signal-to-noise ratio of the estimator. Assuming

the ellipticities of different images are uncorrelated, the dispersion of  $M_{\text{ap}}$  can be obtained by squaring (6) and taking the expectation value, yielding

$$\sigma^2 = \frac{\sigma_\epsilon^2}{2n^2} \sum_{i=1}^N Q^2(|\boldsymbol{\theta}_i|), \quad (4.7)$$

where  $\sigma_\epsilon$  is the dispersion of the two component ellipticity. By (4) the expectation value of  $M_{\text{ap}}$  is,

$$\langle M_{\text{ap}} \rangle = \frac{\langle Z \rangle}{n} \sum_{i=1}^N \gamma_t(\boldsymbol{\theta}_i) Q(|\boldsymbol{\theta}_i|), \quad (4.8)$$

where  $\gamma_t$  is the tangential shear. Taking the ensemble average of (8) over the probability distribution for the galaxy positions gives,

$$\langle M_{\text{ap}} \rangle_c = 2\pi \langle Z \rangle \int_{\theta_{\text{in}}}^{\theta_{\text{out}}} d\theta \theta \langle \gamma_t \rangle(\theta) Q(\theta), \quad (4.9)$$

where  $\langle \gamma_t \rangle(\theta)$  is the mean tangential shear on a circle of angular radius  $\theta$  and the subscript ‘‘c’’ stands for continuous. Similarly, we can take the ensemble average of the dispersion (7), to obtain

$$\sigma_c^2 = \frac{\pi \sigma_\epsilon^2}{n} \int_{\theta_{\text{in}}}^{\theta_{\text{out}}} d\theta \theta Q^2(\theta). \quad (4.10)$$

The ensemble-averaged signal-to-noise ratio is then,

$$\frac{S}{N} = \frac{\langle M_{\text{ap}} \rangle_c}{\sigma_c} = \frac{2\langle Z \rangle \sqrt{\pi n}}{\sigma_\epsilon} \frac{\int_{\theta_{\text{in}}}^{\theta_{\text{out}}} d\theta \theta \langle \gamma_t \rangle(\theta) Q(\theta)}{\sqrt{\int_{\theta_{\text{in}}}^{\theta_{\text{out}}} d\theta \theta Q^2(\theta)}}. \quad (4.11)$$

By the Cauchy-Schwarz inequality the signal-to-noise ratio of the estimator is maximized if

$$Q(\theta) \propto \langle \gamma_t \rangle(\theta). \quad (4.12)$$

Since

$$\langle \gamma_t \rangle(\theta) = \bar{\kappa}(\theta) - \langle \kappa \rangle(\theta), \quad (4.13)$$

(Bartelmann 1995) where  $\langle\kappa\rangle(\theta)$  is the dimensionless mean surface mass density on a circle of radius  $\theta$  and  $\bar{\kappa}(\theta)$  is the dimensionless mean surface mass density within a circle of radius  $\theta$ , the maximized signal-to-noise ratio becomes

$$\frac{S}{N} = \frac{2\langle Z \rangle \sqrt{\pi n}}{\sigma_\epsilon} \sqrt{\int_{\theta_{\text{in}}}^{\theta_{\text{out}}} d\theta \theta [\bar{\kappa}(\theta) - \langle\kappa\rangle(\theta)]^2}. \quad (4.14)$$

If instead of using a maximized weight function  $Q$  we chose one of the often used generic weight functions given in Schneider et al. (1998), our estimate of the signal-to-noise ratio for a given lens would be slightly smaller. In an upcoming paper (Weinberg & Kamionkowski 2002) we show that although using such a weight function reduces the predicted abundance of dark and virialized lenses somewhat, the principle result of this paper, namely that the relative abundance of dark to virialized lenses is 10–20%, is virtually unchanged so long as  $\theta_{\text{out}} \gtrsim 3$  arcmin.

To compute the signal-to-noise ratio for a lens with a given density profile we need to determine the mean tangential shear of the source galaxies. Different density profiles will in general produce different shear patterns. In particular, the more cuspy a profile, the stronger its lensing signal. Of course this becomes more complicated when considering profiles with power-law breaks. For instance, although the NFW profile goes as  $r^{-1}$  at small radii while the isothermal sphere goes as  $r^{-2}$ , at larger radii the former varies as  $r^{-3}$  while the latter remains at  $r^{-2}$ . The net effect, as we will show, is that the NFW profile yields a stronger signal compared to the isothermal sphere for lenses at reasonable redshifts. That said, we consider a variety of profiles to account for the full range of possibilities and to study the dependence of our results on these profiles. Specifically, we compare the calculated abundances assuming the overdensity is a point mass, a uniform-density sphere, an isothermal sphere, an NFW profile, and a Hernquist profile. For an object of a given mass, mean overdensity, and density profile, we can solve for the parameters of the given profile (e.g., the radius, the velocity dispersion, the scale radius, the scale density, etc.) and determine, using equation (14), whether such an object produces a sufficiently large weak-lensing signal-to-noise ratio so as to be detectable. Note that for an overdensity with angular

radius smaller than the size of the lensing image, the shear pattern beyond the radius of the overdensity will be that of a point mass. Furthermore, if the angular radius is larger than the image size then the lensing signal is determined by just the mass  $M_P$  within the projected image radius  $P = \theta_{\text{out}} D_d$  and not the mass outside this radius. The derivation of the signal-to-noise relation for each of these profiles is given in the Appendix.

To produce a detectable signal, an overdensity must be large enough to yield a signal-to-noise ratio greater than some minimum value. For the calculations done in this paper we adopt the following fiducial values, unless stated otherwise:  $(S/N)_{\text{min}} = 5$ ,  $\theta_{\text{out}} = 5$  arcmin, the number density of galaxy images is  $n = 30$  arcmin $^{-2}$ , and  $\sigma_\epsilon = 0.2$ . The minimum nonlinear overdensity corresponds to a particular position along the linear-theory evolutionary cycle. In the next section we discuss how we relate the minimum *nonlinear* overdensity to a corresponding minimum *linear*-theory overdensity. This will enable us to apply the Press-Schechter formalism to obtain an estimate of the abundance of overdensities that produce a weak-lensing signal as a function of redshift.

### 4.3 Dynamics

We use the STHC model to relate the minimum nonlinear overdensity needed for a detectable weak-lensing signal at a given redshift to a minimum linear-theory overdensity. According to STHC the nonlinear evolution of cosmic density fluctuations is approximated by a dynamical model in which the initial linear perturbation is an isolated, uniform sphere surrounded by unperturbed matter. Gravitational instability causes the initially small linear perturbation to grow and enter the nonlinear regime, ultimately forming a virialized object that is decoupled from the cosmological background. In order to avoid the collapse to infinite density predicted by the solution of STHC, we invoke a simple smoothing scheme that allows us to map a linear overdensity greater than the critical linear density contrast,  $\delta_c \sim 1.69$ , to a finite nonlinear overdensity. In what follows we shall consider the STHC model in a  $\Lambda$ CDM universe.

Following the derivation of the relevant STHC formula, we present our smoothing scheme. Finally, we discuss how we distinguish “virialized” clusters from those that have not yet collapsed.

For a flat cosmology with a cosmological constant, the change in the proper radius,  $r$ , with scale factor  $a$  for a uniform spherical overdensity of fixed mass  $M$  is given by (see Peebles 1984, Eke et al. 1996)

$$\left(\frac{dr}{da}\right)^2 = \frac{r^{-1} + \omega r^2 - \beta}{a^{-1} + \omega a^2}, \quad (4.15)$$

where  $a = (1 + z)^{-1}$ ,  $\beta$  is a constant that is positive for overdensities and

$$\omega = (\Omega_0^{-1} - 1), \quad (4.16)$$

where  $\Omega_0$  is the cosmological density parameter. Note that the units of  $r$  are such that  $(3M/4\pi\rho_0)^{1/3} \equiv 1$  where  $\rho_0$  is the cosmological background density at  $z = 0$ . Separating the variables in equation (15) and integrating gives

$$\int_0^r \frac{r'^{1/2}}{(\omega r'^3 - \beta r' + 1)^{1/2}} dr' = \int_0^a \frac{a'^{1/2}}{(\omega a'^3 + 1)^{1/2}} da'. \quad (4.17)$$

Solving for the root of the numerator in equation (15) gives the turnaround radius (i.e., radius at maximum expansion),  $r_{ta}$ , as a function of the density parameter  $\omega$  and perturbation amplitude  $\beta$ . An exact solution for  $r_{ta}$  is given in Appendix A of Eke et al. (1996). For overdensities that are past turnaround the left-hand side of equation (17) is integrated from zero to  $r_{ta}$  and added to the integral from  $r$  to  $r_{ta}$ . The evolution of the radius of an overdensity as a function of time is illustrated in Figure 1. Note that the cosmological constant has the effect of slowing the collapse as compared to a CDM universe.

The nonlinear overdensity is given by

$$1 + \delta^{\text{NL}} = \frac{\rho_{\text{pert}}}{\rho_b}, \quad (4.18)$$

where  $\rho_{\text{pert}}$  is the mean density of the perturbed region and  $\rho_b$  is the background density at the given redshift. Since  $\rho_{\text{pert}} = \rho_0/r^3$  and  $\rho_b = \rho_0/a^3$ , the nonlinear overdensity becomes

$$1 + \delta^{\text{NL}} = \left(\frac{a}{r}\right)^3. \quad (4.19)$$

For a given nonlinear overdensity of mass  $M$  at redshift  $z$  we can find the radius of the perturbation  $r$  such that  $(S/N) \geq (S/N)_{\text{min}}$ . We can then solve equation (17) for  $\beta$ .

We now relate this same  $\beta$  to the linear-theory perturbation amplitude. Eke et al.(1996) showed that

$$\beta = \frac{a_0(2\omega)^{1/3}}{3A(a_0(2\omega)^{1/3})} \delta_0^{\text{lin}}, \quad (4.20)$$

where  $a_0$  is the scale factor today,  $\delta_0^{\text{lin}}$  is the linear-theory overdensity extrapolated to the present and

$$A(x) = \frac{(x^3 + 2)^{1/2}}{x^{3/2}} \int^x \left(\frac{u}{u^3 + 2}\right)^{3/2} du, \quad (4.21)$$

(Peebles 1980). The linear-theory overdensity at redshift  $z$  is given by

$$\delta^{\text{lin}}(z) = \delta_0^{\text{lin}} D(a), \quad (4.22)$$

where  $D(a)$ , the linear theory growth factor for a  $\Lambda$ CDM cosmology, is

$$D(a) = \frac{A(a(2\omega)^{1/3})}{A(a_0(2\omega)^{1/3})}. \quad (4.23)$$

Using equations (20), (22), and (23) we get the desired relation between the linear-theory overdensity and  $\beta$ :

$$\delta^{\text{lin}}(z) = 3\beta \frac{A(a(2\omega)^{1/3})}{a_0(2\omega)^{1/3}}. \quad (4.24)$$

Equations (17), (19) and (24) therefore provide a map between the nonlinear overdensity and the linear-theory overdensity at a given redshift.

It can be shown that  $r \rightarrow 0$  in the limit that  $\delta^{\text{lin}} \rightarrow \delta_c$ , corresponding to the well-known infinite density predicted by the solution of STHC. An actual overdensity will, of course, virialize before reaching the singular solution. To properly account for this we introduce the following smoothing scheme.

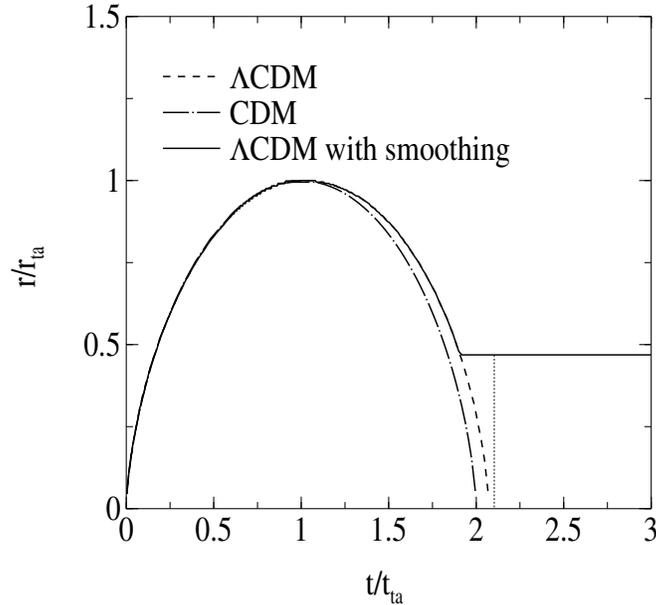


Figure 4.1 Radial evolution of a density perturbation according to the STHC model. At the turnaround time,  $t = t_{ta}$ , the perturbation reaches a maximum-expansion radius and begins to collapse. As expected, in a  $\Lambda$ CDM cosmology (*dashed* curve) the collapse takes somewhat longer than in a CDM cosmology (*line-dot* curve). The collapse to a singularity predicted by the solution of the STHC model is avoided by the smoothing scheme (*solid* curve), which yields a constant radius once the virialized overdensity is reached.

Rather than assuming that an overdensity satisfies equation (19) throughout its evolution, assume it satisfies it only until it reaches the virialized overdensity  $1 + \delta_{\text{vir}}^{\text{NL}}(z)$ . Once the perturbation reaches the virialized overdensity take its radius to be a constant with time so that the overdensity continues to grow only because the cosmological background density keeps decreasing. The nonlinear overdensity is

therefore given by

$$1 + \delta^{\text{NL}} = \begin{cases} \left(\frac{a(t)}{r}\right)^3, & \text{if } \left(\frac{a(t)}{r}\right)^3 \leq 1 + \delta_{\text{vir}}^{\text{NL}}(z); \\ (1 + \delta_{\text{vir}}^{\text{NL}}) \left(\frac{a(t)}{a_{\text{vir}}}\right)^3, & \text{otherwise,} \end{cases} \quad (4.25)$$

where  $a_{\text{vir}}$  is the scale factor at virialization. Since  $\delta^{\text{lin}}(t_2) = \delta^{\text{lin}}(t_1)D(a_2)/D(a_1)$ , the linear-theory overdensity then becomes

$$\delta^{\text{lin}} = \begin{cases} 3\beta \frac{A(a(2\omega)^{1/3})}{a_0(2\omega)^{1/3}}, & \text{if } 1 + \delta^{\text{NL}} \leq 1 + \delta_{\text{vir}}^{\text{NL}}(z); \\ \delta^{\text{lin}}(a_{\text{vir}}) \frac{D(a(t))}{D(a_{\text{vir}})}, & \text{otherwise.} \end{cases} \quad (4.26)$$

Therefore, if the minimum nonlinear overdensity needed to produce a detectable weak-lensing signal at redshift  $z$  is larger than the virialization overdensity, we evaluate  $a_{\text{vir}}$  using equation (25) and then compute the minimum linear-theory overdensity using the lower expression in equation (26). In Figure 1 we plot the radius of an overdensity as a function of time using this smoothing scheme. In Figure 2 we show the nonlinear overdensity as a function of the linear-theory overdensity. Note that the value of the overdensity at virialization can be obtained by assuming  $r = r_{\text{vir}}$ , the virialized radius, in equation (19), and using the expression from Lahav et al. (1991), which gives the ratio between the turnaround radius and the virialization radius. For convenience we use the Kitayama & Suto (1996) approximation to  $1 + \delta_{\text{vir}}^{\text{NL}}(z)$ , as well as their approximation to  $\delta_c(z)$ . We independently verified that both approximations matched the solution of the exact formalism described above.

In summary, given the minimum nonlinear overdensity needed to produce a detectable weak-lensing signal,  $\delta_{\text{min}}^{\text{NL}}$ , of an object of mass  $M$  at redshift  $z$ , we use equations (17), (25), and (26) to compute the corresponding minimum linear-theory overdensity,  $\delta_{\text{min}}^{\text{lin}}$ , needed to produce a detectable weak-lensing signal. If  $\delta_{\text{min}}^{\text{lin}} < \delta_c(z)$  then the object can produce a detectable weak gravitational lens, even though it is not yet virialized.

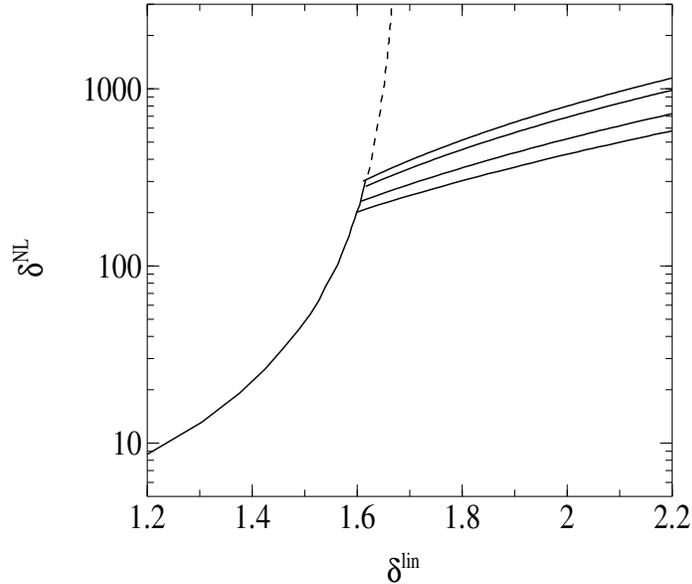


Figure 4.2 Nonlinear overdensity as a function of the linear-theory overdensity according to the STHC model. The full solution of the STHC model predicts collapse to an infinite overdensity as  $\delta_{\text{lin}} \rightarrow 1.69$  (*dashed* curve). According to the smoothing scheme, however, once a mass concentration reaches the virialization overdensity  $1 + \delta_{\text{vir}}^{\text{NL}}(z)$ , its radius remains constant so that the overdensity increases in proportion to the decrease in the background density. The *solid* curves show the smoothing scheme solution for mass concentrations that reach the virialization overdensity at  $z = 0.1, 0.2, 0.5$ , and  $1.0$ , from top to bottom. In an Einstein–de Sitter Universe, the virialization overdensity is independent of redshift and therefore all of the solid curves would be the same.

## 4.4 Abundances

To calculate the abundance of overdensities that produce a detectable weak-lensing signal as a function of redshift, we use Press-Schechter theory assuming Gaussian statistics for the initial linear-theory density field. The differential number count of lensing objects per steradian, per unit redshift interval is

$$\frac{dN(\delta_{\text{min}}^{\text{lin}})}{dzd\Omega} = \frac{dN(\delta_{\text{min}}^{\text{lin}})}{dV} \frac{dV}{dzd\Omega}, \quad (4.27)$$

where

$$\frac{dV}{dzd\Omega} = \frac{c}{H_0} \frac{(1+z)^2 D_A(z)^2}{\sqrt{\Omega_0(1+z)^3 + 1 - \Omega_0}}, \quad (4.28)$$

is the comoving-volume element,  $c$  is the speed of light,  $H_0$  is Hubble's constant, and  $D_A(z)$  is the angular-diameter distance at redshift  $z$ . The total number density of weak lenses is given by

$$\frac{dN(\delta_{\min}^{\text{lin}})}{dV} = \int_0^\infty f(M; \delta_{\min}^{\text{lin}}) \frac{dn}{dM}(M) dM, \quad (4.29)$$

where  $dn(M)/dM$ , the comoving number density of virialized objects of mass  $M$  in the interval  $dM$ , is (Press & Schechter 1974)

$$\frac{dn}{dM}(M) = \sqrt{\frac{2}{\pi}} \frac{\rho_0}{M^2} \frac{\delta_c(z)}{\sigma(M, z)} \left| \frac{d \ln \sigma}{d \ln M} \right| \exp \left[ -\frac{\delta_c(z)^2}{2\sigma^2} \right]. \quad (4.30)$$

In this paper we use the Viana & Liddle (1999) fits to the dispersion of the density field,  $\sigma(M, z)$ , obtained from the galaxy cluster X-ray temperature distribution function. The function  $f(M; \delta_{\min}^{\text{lin}})$  is the fraction of objects, either dark or virialized, that can lens ( $\delta > \delta_{\min}^{\text{lin}}$ ) relative to those that are virialized ( $\delta > \delta_c$ ). The probability that an object's linear overdensity is in the range  $\delta_1 < \delta < \delta_2$  is

$$P(\delta_1 < \delta < \delta_2) = \text{erf} \left( \frac{\delta_2}{\sqrt{2}\sigma(M, z)} \right) - \text{erf} \left( \frac{\delta_1}{\sqrt{2}\sigma(M, z)} \right), \quad (4.31)$$

where ‘‘erf’’ is the error function. Therefore, for dark lenses (i.e., those objects with  $\delta_{\min}^{\text{lin}} < \delta < \delta_c$ )

$$f_{\text{dark}}(M, z) = \begin{cases} \frac{P(\delta_{\min}^{\text{lin}} < \delta < \delta_c)}{P(\delta > \delta_c)}, & \delta_{\min}^{\text{lin}} < \delta_c; \\ 0, & \text{otherwise,} \end{cases} \quad (4.32)$$

while for virialized lenses ( $\delta > \delta_c$  and  $\delta > \delta_{\min}^{\text{lin}}$ ),

$$f_{\text{vir}}(M, z) = \begin{cases} \frac{P(\delta > \delta_{\min}^{\text{lin}})}{P(\delta > \delta_c)}, & \delta_{\min}^{\text{lin}} > \delta_c; \\ 1, & \text{otherwise.} \end{cases} \quad (4.33)$$

For low enough masses, the minimum overdensity needed to produce a detectable weak-lensing signal becomes so large that both  $f_{\text{dark}}$  and  $f_{\text{vir}}$  approach zero, thereby imposing an effective weak-lensing mass threshold (see Section 5.2). Integrating equation (27) over redshift assuming  $f = f_{\text{dark}}$  yields the number count of dark lenses per unit area on the sky and similarly for virialized lenses when  $f = f_{\text{vir}}$ .

## 4.5 Results

### 4.5.1 Minimum overdensity as a function of redshift

We can now compute the sky density of weak lenses. To gain physical insight into the results as well as illustrate the calculational procedure discussed above we first show the redshift dependence of the minimum nonlinear overdensity. As noted earlier, the result is sensitive to the lens density profile on account of the minimum overdensity's dependence on the shear. Since the shear is proportional to the surface mass density, the NFW and Hernquist profiles (for whom  $\rho \propto r^{-1}$  as  $r \rightarrow 0$ ) have a constant shear at small radii while the isothermal sphere profile ( $\rho \propto r^{-2}$ ) has a shear that goes as  $r^{-1}$  for all radii. This is shown in Figure 3, where we plot the radial dependence of the mean tangential shear for these different profiles.

In Figure 4, the minimum nonlinear overdensity as a function of redshift for a  $10^{14} M_{\odot}$  object is plotted for the various profiles. All the profiles show the same general trend: a minimum at  $z \sim 0.3$  and monotonic rises at lower and higher redshifts. This is a consequence of the source-galaxy redshift distribution, which we assume is given

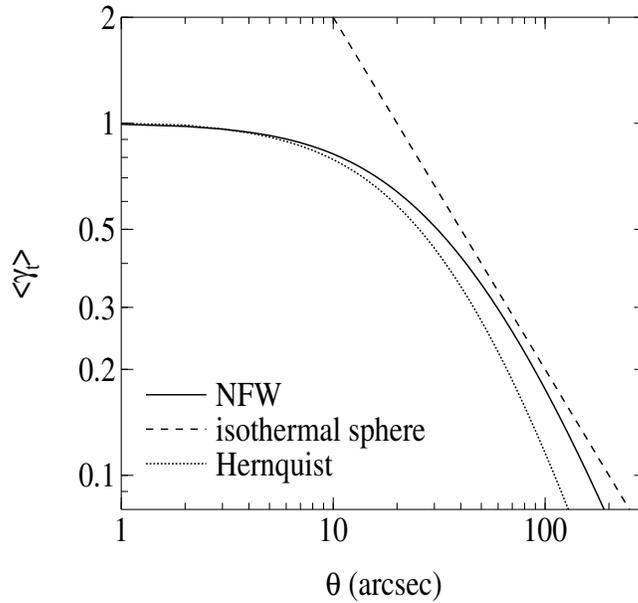


Figure 4.3 Mean tangential shear as a function of angular distance from the lens center for an NFW (*solid curve*), Hernquist (*dotted curve*) and isothermal sphere (*dashed curve*) density profile. The normalization is arbitrary.

by a function of the form

$$p_z(z_s) = \frac{\beta z_s^2}{\Gamma(3/\beta) z_0^3} \exp[-(z_s/z_0)^\beta], \quad (4.34)$$

with  $\beta = 1.5$  and mean redshift  $\langle z_s \rangle \approx 1.5z_0 = 1.2$  (cf. Smail et al. 1995; Brainerd et al. 1996; Cohen et al. 2000). Since lenses are most effective when they lie midway between the source and the observer (i.e., the factor  $D_d D_{ds}/D_s$  peaks when  $D_d \simeq D_{ds}$ ), an overdensity at  $z \sim 0.3$  is ideally positioned to lens source galaxies that are primarily located at  $z = \langle z_s \rangle \sim 1$ , thereby accounting for the minimum in the curves. Accordingly, overdensities located at lower and higher redshifts than  $z \sim 0.3$  are less effective at lensing so that a larger overdensity is needed to produce a detectable lens. In addition, for an overdensity with redshift approaching unity, there are fewer background galaxies to lens (less signal) as well as more foreground galaxies in the image (greater noise), further decreasing the observed lensing signal-to-noise

ratio.

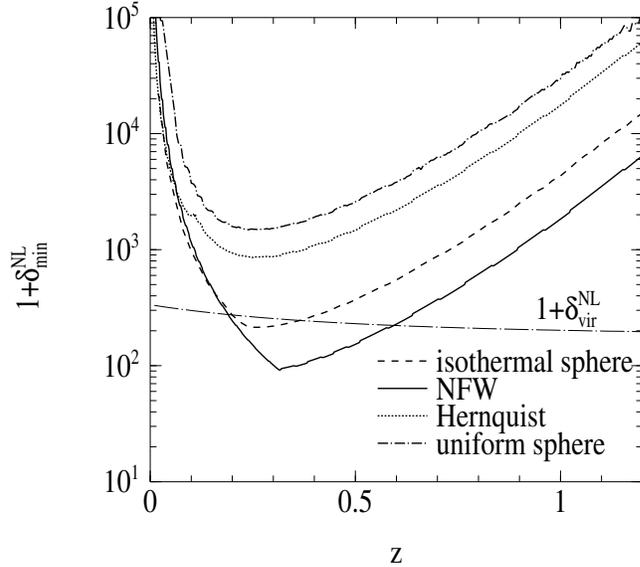


Figure 4.4 Minimum nonlinear overdensity needed to produce a detectable weak lensing signal as a function of redshift for a  $10^{14}M_{\odot}$  object with a density profile that is a uniform-density sphere (*line-dot* curve), a truncated isothermal sphere (*dashed* curve), a Hernquist profile (*dotted* curve), and an NFW profile (*solid* curve). An overdensity with a larger mass will displace these curves downwards. The thin, *long-dash-dot* curve is the overdensity at virialization in the STHC model.

Another feature to note in Figure 4 is the difference in amplitude of  $1 + \delta_{\min}^{\text{NL}}$  between the different profiles. Over most of the redshift range, the NFW profile requires the smallest overdensity in order to produce a detectable weak-lensing signal while the uniform-density sphere requires the largest. This is because the NFW profile has its mass much more centrally concentrated as compared to the uniform-density sphere. A source galaxy at some angular radius near the lens center will therefore be sheared more strongly by the former and hence produce a larger signal. A similar explanation accounts for the differences in amplitude of  $1 + \delta_{\min}^{\text{NL}}$  between the non-uniform profiles.

### 4.5.2 The abundance of dark and virialized lenses

In Figure 5 we show the redshift distribution (normalized to unity) of dark and virialized lenses for the NFW, Hernquist, and isothermal-sphere profiles. Because the

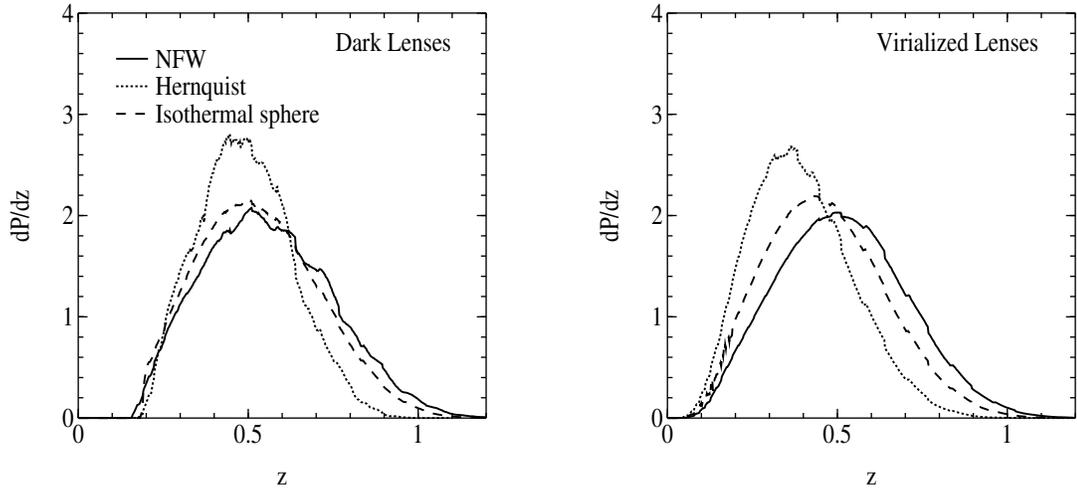


Figure 4.5 Redshift distribution of dark lenses (*left* panel) and virialized lenses (*right* panel) for a truncated isothermal sphere (*dashed* curve), a Hernquist profile (*dotted* curve), and an NFW profile (*solid* curve). The ordinate gives the normalized probability distribution per unit redshift interval.

minimum overdensity for the uniform-density sphere was so large, the probability of detecting a lens with such a profile is negligible and hence no longer considered. For all three profiles the distribution peaks at  $z \approx 0.5$  and has a full-width at half-maximum of  $\Delta z \approx 0.5$ . The distribution drops off at  $z \approx 1$  for two reasons: the minimum overdensity is becoming increasingly large since  $\langle z_s \rangle \simeq 1$ , and the STHC dynamics predicts fewer and fewer massive, large overdensities at these higher redshifts.

The sky density of dark lenses as a function of redshift for the same three profiles is shown in Figure 6. Depending on the density profile, we expect to find between 1–20 dark lenses per square degree out to  $z = 1$ , and virtually none at higher redshifts. The reason the Hernquist profile predicts a smaller dark-lens sky density compared with the isothermal sphere and NFW is that such a profile requires a larger overdensity to produce a detectable weak-lensing signal (see Figure 3). Finally, note that although

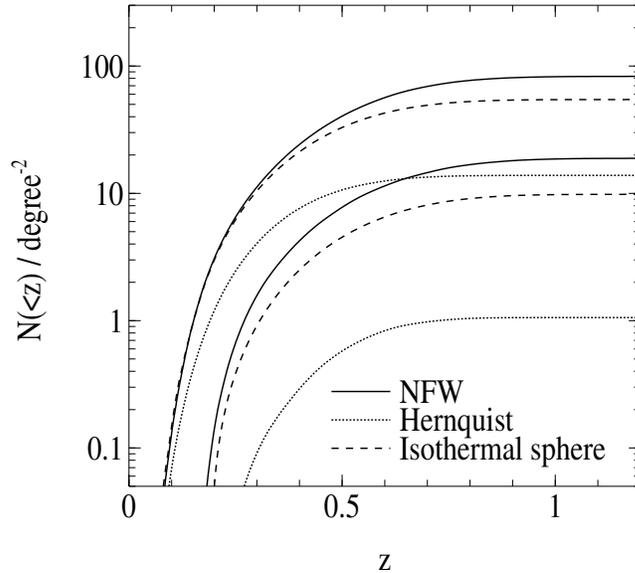


Figure 4.6 Number counts of dark and virialized lenses for a truncated isothermal sphere (*dashed* curve), an Hernquist profile (*dotted* curve), and an NFW profile (*solid* curve). The ordinate gives the sky density of lenses at redshifts less than  $z$ . The top curve for a given density profile corresponds to the sky density of virialized lenses and the bottom curve to the sky density of dark lenses.

this distribution is integrated over dark lenses of all masses, the minimum overdensity as a function of redshift becomes so large for  $M \lesssim 5 \times 10^{13} M_{\odot}$  that there are virtually no dark lenses with such small masses. This point is illustrated in Figure 7, where we plot the weak-lensing mass distribution (i.e., the integrand of equation (29) times the mass) for both dark and virialized lenses at  $z = 0.5$ . Furthermore, since the Press-Schechter mass function falls off steeply with mass, there will be very few dark lenses with  $M \gtrsim 10^{15} M_{\odot}$  despite the lower value of the minimum overdensity at these masses.

In Figures 5 and 6 we also show the redshift distribution and sky density of virialized lenses for the three different density profiles. Although the weak-lensing mass threshold is, as expected, somewhat smaller for virialized lenses than for dark lenses (see Figure 7) their normalized distributions are not very different. Nonetheless, the sky density of virialized lenses is  $10 - 80 \text{ degree}^{-2}$  and hence a factor of  $4 - 10$

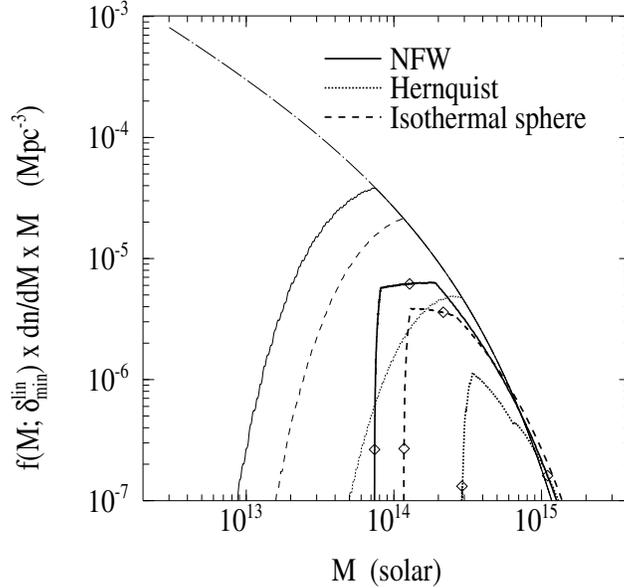


Figure 4.7 Predicted weak-lensing mass distribution at redshift  $z = 0.5$ . Shown is the comoving number density as a function of mass of dark lenses (*thick* lines) and virialized lenses (*thin* lines) of mass  $M$  in the interval  $d \ln M$ . Plotted for comparison is the virialized mass distribution, i.e.,  $\frac{dn}{dM}(M) \times M$  (*thin, dash-dot* line). Since the minimum mass overdensity needed to produce a detectable lens (see Figure 4) is lowest for the NFW profile (*solid* curves), such a profile predicts a smaller weak-lensing mass threshold as compared to the Hernquist profile (*dotted* curve) and the truncated isothermal sphere (*dashed* curve). The two diamonds on each dark lens mass distribution curve mark the mass at which the minimum overdensity needed to produce a detectable lens is 275 and 100. The sharp lower-mass cutoff in the dark lens mass distribution is a consequence of the heaviside step-function nature of  $f_{\text{dark}}$ .

larger than the sky density of dark lenses. This is because by redshifts of  $z \approx 0.5$  (where the distributions peak) a majority of objects in the mass range that can lens will have already virialized.

Having computed the redshift distribution and sky density of dark and virialized lenses we now determine their relative abundances. The fraction of weak lenses that are caused by dark, non-virialized objects as a function of redshift is shown in the left panel of Figure 8. Out to  $z \approx 1$  the fraction is nearly constant with about 20% of all weak lenses arising from dark objects. We again emphasize that the predicted

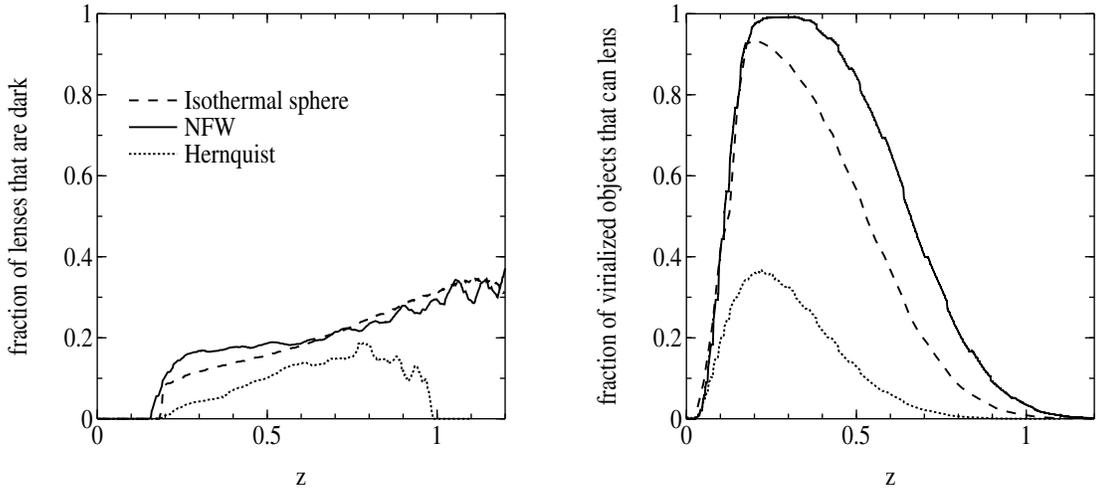


Figure 4.8 Left panel: The fraction of weak lenses that are dark lenses as a function of redshift for a truncated isothermal sphere (*dashed* curve), an Hernquist profile (*dotted* curve), and an NFW profile (*solid* curve). The fraction is relatively constant between redshifts  $z = 0.2$  and  $z = 1.0$ , beyond which the abundance of both dark and virialized lenses drops to zero. The coarseness of the curves for  $z > 1$  is an artifact of numerical noise that is a result of this dropoff in both abundances. Right panel: The fraction of virialized objects with  $M > 5 \times 10^{13} M_{\odot}$  that are able to weak lens as a function of redshift for the same density profiles as above.

abundance of dark lenses relative to virialized lenses is significant not because dark lenses comprise the lower-mass end of the mass function; on the contrary, virialized lenses have a lower mass threshold than dark lenses as shown in Figure 7. Rather, it is significant because according to the STHC model, a substantial fraction of cluster-mass objects are sufficiently overdense to produce a detectable lensing signal despite not having reached the virialization overdensity. For  $z > 1$  the abundance of weak lenses of all types (both virialized and dark) drops off significantly. This, again, is because  $\langle z_s \rangle \sim 1$  and because the evolution of overdensities has not yet had enough time to produce sufficiently large overdensities. This is also illustrated in the right panel of Figure 8, where we show the fraction of virialized objects that can lens as a function of redshift. For  $0.2 \lesssim z \lesssim 0.5$  a large fraction of virialized objects with  $M > 5 \times 10^{13} M_{\odot}$  can produce a detectable weak-lensing signal but, for the same reason as above, by  $z = 1$  this fraction is nearly zero.

Finally we would like to point out that given the above results for the weak-lensing mass distribution, it is not surprising that in their study of weak lensing by low-mass galaxy groups, Hoekstra et al. (2001) could only (just barely) detect a weak-lensing signal by stacking 50 such groups together. Namely, the groups in Hoekstra et al.'s sample, which were at a mean redshift of 0.3, had a mean overdensity of only  $\sim 75$  and a mean mass of just  $\sim 4 \times 10^{13} M_{\odot}$ , assuming an isothermal density profile and using their measured value of  $\sim 275 \text{ km s}^{-1}$  for the lensing-inferred velocity dispersion. Therefore, as Figure 7 suggests, individual groups from their sample were neither massive enough nor sufficiently overdense to produce a detectable weak-lensing signal.

### 4.5.3 The effect of increasing the image size on the lensing signal

In the above calculations we assume that the lensing images are 5 arcmin in radius, roughly the size of lensing maps to date. However, if a lens is relatively nearby or has a large radial extent it is possible that a large fraction of the total lensing signal is missed. This effect might be especially troublesome for the detection of dark lenses, given that they are not yet virialized and hence have larger radii. We now address this issue by determining the extent to which increasing the image size alters the predicted abundance of dark lenses.

Before moving on, however, we note that while we examined the predicted distribution and sky density of weak lenses for a variety of profiles, there is good reason to regard the NFW profile as the most plausible. For virialized lenses this is clearly the case as N-body simulations show that the halo density profiles are well-fit by the NFW form. Though it is difficult to be as certain in the case of dark, non-virialized, lenses (N-body simulation fits to profiles have so far only been for virialized systems), because most of the dark lenses are well past turnaround ( $1 + \delta^{\text{NL}} \gtrsim 50$ ) and because it is unlikely that the STHC model perfectly describes the evolution of overdensities all the way to virialization, assuming an NFW profile for dark lenses is a fair ap-

proximation. Furthermore, since virialization is expected to occur from inside-out, the centers of dark lenses, where most of the lensing signal is coming from (as we show quantitatively below), are likely near virialization and hence well described by the NFW profile. For these reasons (and also to avoid overly cluttered figures), the rest of the figures in this paper show results only for the NFW profile. To obtain approximate results for the other profiles, simply scale by the relative abundances shown in Figure 6.

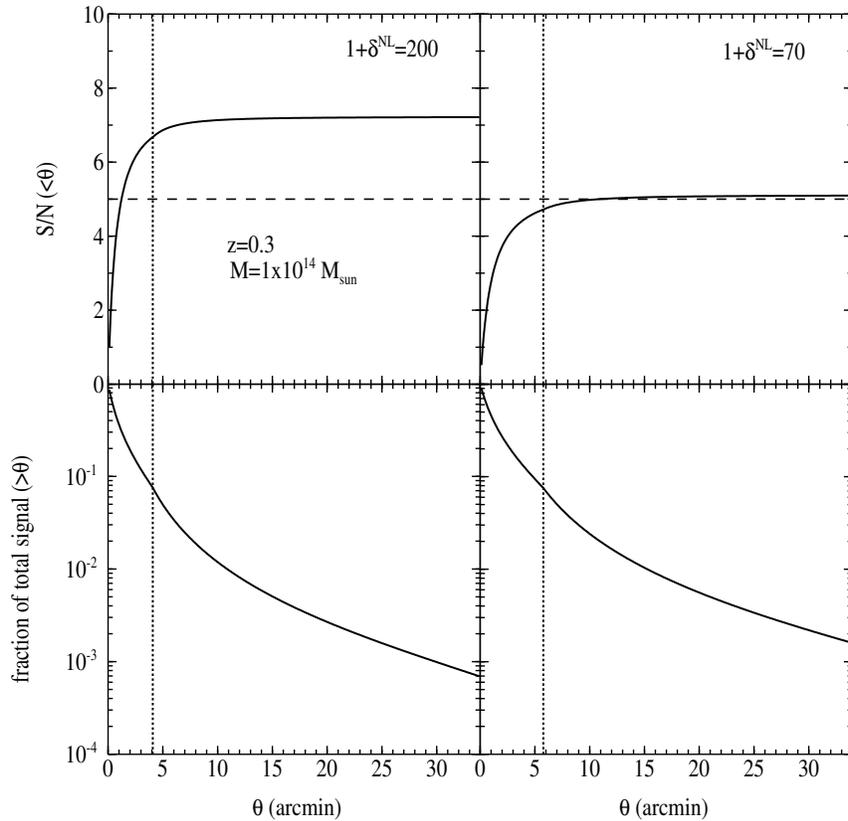


Figure 4.9 Upper panels: The signal-to-noise ratio within an angular radius  $\theta$  from the lens center as a function of  $\theta$  for an NFW profile. Lower panels: The fraction of the total lensing signal that comes from outside the angular radius  $\theta$ . Approximately 90% of the lensing signal comes from a region smaller than the lensing halo radius. All four panels correspond to an object at redshift  $z = 0.3$  and mass  $M = 10^{14} M_{\odot}$ . The left-hand-side plots are for a nonlinear overdensity of 200 and the right-hand-side plots for a nonlinear overdensity of 70.

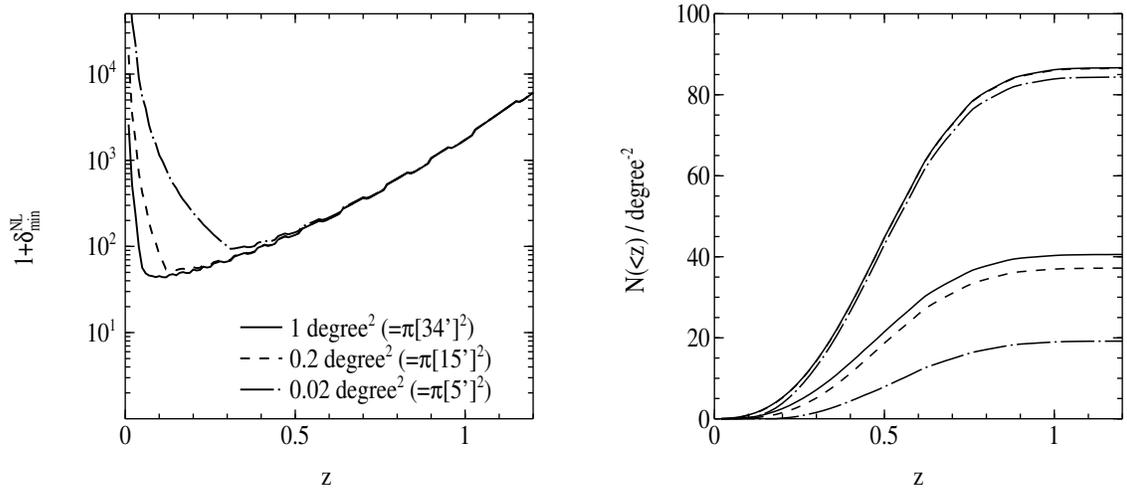


Figure 4.10 Effect of increasing the image size on the abundance estimates. Plotted are the minimum overdensity needed to produce a detectable lens (*left* panel) and the sky density of weak lenses (*right* panel) as functions of redshift for image sizes of  $0.02 \text{ degree}^2$  (*line-dot* curve),  $0.2 \text{ degree}^2$  (*dashed* curve), and  $1 \text{ degree}^2$  (*solid* curve). The top three curves in the right panel correspond to virialized lenses and the bottom three to dark lenses. An NFW profile is assumed. Note that increasing the image size beyond  $1 \text{ degree}^2$  barely increases the predicted sky density since the signal becomes increasingly small at larger angular distances from the lens center (see Figure 9).

In Figure 9 we plot the cumulative signal-to-noise ratio and the fraction of the total signal as a function of the angular distance from the lens center for a lens with an NFW profile at redshift  $z = 0.3$  with mass  $M = 10^{14} M_{\odot}$ . Although the fraction of the signal that comes from within 5 arcmin is  $\sim 90\%$  for lenses of overdensity  $1 + \delta_{\min}^{\text{NL}} = 200$  and  $1 + \delta_{\min}^{\text{NL}} = 70$ , the lens with overdensity 200 requires an image size of just  $\sim 2$  arcmin to be detectable ( $S/N = 5$ ) while the lens with overdensity 70 requires  $\sim 10$  arcmin to be detectable. In general we find that in order to detect nearly all dark lenses with  $S/N \geq 5$  in a given field the image area must be at least  $\sim \pi(15')^2 \approx 0.2 \text{ degree}^2$ , as shown in Figure 10. Larger image sizes will not significantly increase the number of dark lenses detected as very little signal comes from radii larger than 15 arcmin. Also note that although lensing geometry favors a lens midway between observer and source, this effect is somewhat countered by the fact that the closer a weak lens is to the observer, the closer the source galaxy images

pass to the highly overdense lens center (i.e., the solid angle subtended by the lens is larger). As a result, if the image size is large enough to enclose a large portion of the lens core, the lensing signal will be strongest when the lens-observer distance is slightly smaller than the lens-source distance. This accounts for the shift, shown in Figure 10, of the minimum of  $1 + \delta_{\min}^{\text{NL}}(z)$  toward smaller redshift as the image size is increased.

#### 4.5.4 Estimating $\sigma_8$ from the abundance of weak lenses

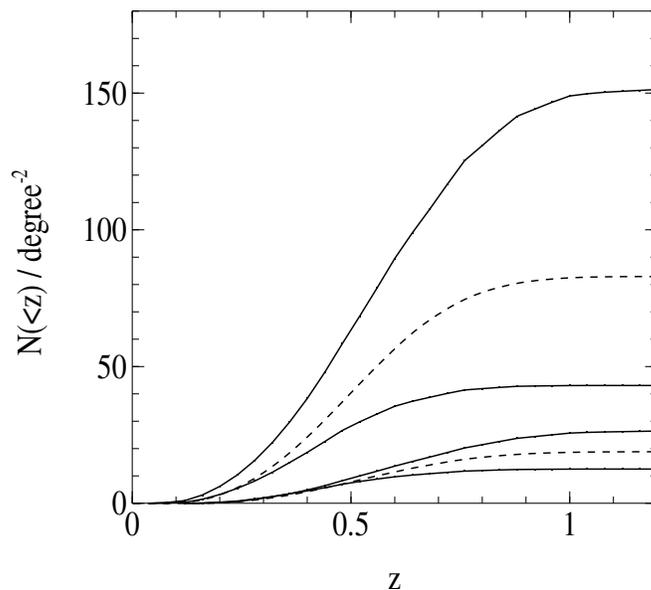


Figure 4.11 Predicted sky density of weak lenses at the 95% confidence limits of  $\sigma_8$  (*solid* curves) given by Viana & Liddle (1999). The *dashed* curves are the predicted sky densities for the mean value of  $\sigma_8$ . The top three curves correspond to virialized lenses, and the bottom three to dark lenses. An NFW profile is assumed.

The present-day abundance of rich, X-ray clusters has been used to constrain the value of  $\sigma_8$ , the amplitude of mass fluctuations in spheres of radius  $8h^{-1}$  Mpc (Evrard 1989; Henry & Arnaud 1991; White, Efstathiou & Frenk 1993; Viana & Liddle 1996; Eke, Cole & Frenk 1996; Kitayama & Suto 1997, Viana & Liddle 1999). In Figure 11 we show the extent to which the measured abundance of weak lenses (from future

cosmic-shear surveys, say) can further constrain  $\sigma_8$ . Here we have plotted the sky density of weak lenses (both virialized and dark) at the 95% confidence limits of  $\sigma_8$  given by Viana & Liddle (1999). Since weak lenses are produced by only relatively rare objects, their abundance is very sensitive to the value of  $\sigma_8$ , suggesting the usefulness of weak lenses in measuring the amplitude of mass fluctuations.

Another benefit of using weak lenses to measure  $\sigma_8$  is their broad redshift distribution. In particular, a systematic uncertainty in measuring  $\sigma_8$  by measuring rich cluster abundances is the degeneracy between  $\sigma_8$  and  $\Omega_m$  that arises from the limited range in redshift in which rich clusters are observed. To break this degeneracy, substantial effort is made to measure not only the present-day rich cluster abundance but also the rich cluster abundance at higher redshifts ( $z \approx 0.3$ ; e.g., Henry 1997). This, in turn, gives an estimate of the evolution of the cluster mass function and hence an estimate of  $\Omega_m$ . However, because they are faint, detecting high redshift ( $z > 0.3$ ) rich clusters is difficult. Weak lenses, on the other hand, do not suffer from this limitation and in fact are expected to have a broad redshift distribution and be most abundant at  $z \sim 0.5$  (see Figure 5). As a result, detecting weak lenses provides an excellent means of measuring the evolution of the mass function and hence measuring  $\Omega_m$ . By thereby breaking the degeneracy between  $\sigma_8$  and  $\Omega_m$ , weak-lensing surveys are also well-suited to constrain the power-spectrum amplitude,  $\sigma_8$ .

## 4.6 Discussion and conclusions

In this paper, we have calculated the abundance of dark and virialized lenses. This was accomplished by first expressing the lensing signal strength as a function of the dark-matter overdensity and redshift. Having determined the overdensity required to produce a detectable weak-lensing signal we used the STHC model to calculate the differential abundance of overdensities as a function of position along their evolutionary cycle. Overdensities whose lensing signal yielded  $S/N \geq 5$  were divided into two classes: those with  $1 + \delta^{\text{NL}} < 1 + \delta_{\text{vir}}^{\text{NL}}$  were dark lenses while those with  $1 + \delta^{\text{NL}} > 1 + \delta_{\text{vir}}^{\text{NL}}$  were virialized lenses.

The distinction between dark and virialized lenses was based on the former being at an unrelaxed, and hence earlier, stage in the overdensity evolutionary cycle. This distinction is not arbitrary but rather is expected to result in observational features that definitively separate the two classes of lenses. For instance, since dark lenses will typically have overdensities of  $1 + \delta^{\text{NL}} \sim 100$  while virialized lenses have  $1 + \delta^{\text{NL}} \sim 300$  (see Figure 4), the projected surface density of a dark lens is smaller than that of a virialized lens by a factor of  $3^{2/3} \sim 2$ . The sky density of galaxies in a dark lens will therefore be about two times smaller than in a virialized lens. As it is difficult to detect a significant galaxy overdensity for even a virialized, lensing, cluster at redshifts of  $z \sim 0.5$ , it will be all the more difficult to do so for a dark cluster. Another distinctive observational feature expected of dark lenses is a low X-ray luminosity as compared with virialized lenses, a consequence of the X-ray luminosity function's steep dependence on total virialized mass. This effect might also account for the low X-ray luminosities observed by Postman et al. (2001) in three high-redshift clusters; namely, these objects are in fact proto-clusters that have not yet completely virialized.

Although we considered a variety of density profiles in our calculations of the predicted distribution and sky density of weak lenses, as we noted in Section 5.3, there is good reason to regard the NFW profile as the most plausible form for both virialized and dark lenses. Nonetheless, while N-body simulations show that virialized systems are well-fit by the NFW form, testing whether non-virialized, cluster-mass, halos in N-body simulations are also well-described by the NFW profile is a worthwhile investigation that has not yet been performed. That said, we have shown that the redshift distribution of dark and virialized lenses for all the considered profiles is fairly broad with an average around  $z = 0.5$  and an FWHM of  $\Delta z \approx 0.5$ . The sky density of dark lenses for the NFW profile was calculated to be  $\sim 20 \text{ degree}^{-2}$  (and  $\sim 10 \text{ degree}^{-2}$  for an isothermal sphere profile) and should therefore be readily detectable by upcoming cosmic shear surveys. For virialized lenses, we found a sky density of  $\sim 80 \text{ degree}^{-2}$  assuming an NFW profile (and  $\sim 50 \text{ degree}^{-2}$  for an isothermal sphere profile), a factor of 4 to 5 larger than that of dark lenses. This difference is due to the fact that most of the weak lenses are at redshift  $z \approx 0.5$  and have masses of  $\sim 10^{14} M_{\odot}$ .

so that the majority are, according to the STHC model, virialized. It is important to note that while the aperture mass weight function used here was chosen to match a specific density profile, as we show in an upcoming paper (Weinberg & Kamionkowski 2002), a more general, non-optimal weight function, such as that given by Schneider et al. (1998), would lower the overall abundance of both types of lenses equally. The principle result of this paper, namely the expectation that  $\sim 10 - 20\%$  of weak-lenses are dark, would not change.

We find it encouraging that given the sky coverage of weak-lensing maps to date ( $\sim 1000 \text{ arcmin}^2$ ) and the average size of the individual lensing maps ( $\sim 30 \text{ arcmin}^2$ ), the number of dark lenses we would expect to have seen is of order unity and thus consistent with the detection (Erben et al. 2000, Miralles et al. 2002) of one or two dark lenses. Furthermore, in mock observations of numerical simulations, White, van Waerbeke, & Mackey (2001) showed that a weak-lensing search for clusters will likely suffer from serious line-of-sight projection effects due to the fact that clusters preferentially live in larger structures. These structures on larger scales, which perhaps correspond to  $2\sigma - 3\sigma$  peaks in the primordial distribution, may well be the type of systems that we find give rise to dark lenses.

Finally, we have also shown that measuring the abundance of weak lenses can substantially help to constrain  $\sigma_8$ , the rms mass fluctuation in spheres of radius  $8h^{-1}$  Mpc. This is a consequence of the broad redshift distribution of weak lenses and the fact that they correspond to high-density peaks in the Gaussian primordial distribution. Cosmic-shear surveys, with their ability to detect cluster mass weak lenses over large areas of sky, should therefore provide a powerful new technique for determining the power-spectrum amplitude.

We thank R. Ellis for useful comments, and an anonymous referee for very constructive suggestions and comments that have improved the presentation of this paper. NNW acknowledges the support of an NSF Graduate Fellowship. This work was supported in part by NSF AST-0096023, NASA NAG5-8506, and DoE DE-FG03-92-ER40701.

## 4.A Derivation of the signal-to-noise relation for various density profiles

Starting from equation (14), we derive the signal-to-noise relation for a point mass, a uniform-density sphere, a truncated isothermal sphere, an NFW profile and a Hernquist profile. Since these profiles are all axially symmetric,  $\langle \kappa \rangle(\theta) = \kappa(\theta)$ .

1. *Point Mass*: The dimensionless mean surface mass density within a circle of radius  $\theta$  for a deflecting lens of point mass  $M$  at angular diameter distance  $D_d$  is

$$\bar{\kappa} = \frac{1}{\Sigma_{\text{crit}\infty}} \frac{M}{\pi P^2}, \quad (4.35)$$

where  $P = \theta D_d$ . The quantity  $\kappa$ , the dimensionless mean surface mass density on a circle of radius  $\theta$ , is  $\propto \delta(\theta)$ , the Dirac delta function. Therefore, by equation (14), the signal-to-noise relation for a point mass is given by

$$\frac{S}{N} = \frac{\sqrt{2}\langle Z \rangle M}{\sigma_\epsilon \Sigma_{\text{crit}\infty} \pi D_d^2} \frac{\sqrt{\pi n}}{\theta_{\text{out}}} \sqrt{\left(\frac{\theta_{\text{out}}}{\theta_{\text{in}}}\right)^2 - 1}. \quad (4.36)$$

This can be expressed as a minimum mass needed to produce a detectable weak-lensing signal, which in useful units is

$$M_{\text{min}} = 3.7 \times 10^{13} \left(\frac{S/N}{5}\right) \left(\frac{D_d}{0.3 D_H}\right)^2 \left(\frac{\sigma_\epsilon}{0.2}\right) \left(\frac{\theta_{\text{out}}}{5 \text{ arcmin}}\right) \left(\frac{n}{30 \text{ arcmin}^{-2}}\right)^{-1/2} \\ \times \left(\frac{(\theta_{\text{out}}/\theta_{\text{in}})^2 - 1}{100}\right)^{-1/2} \Sigma_{\text{crit}\infty} \langle Z \rangle^{-1} M_\odot,$$

where  $D_H = c/H_0$  is the Hubble distance.

2. *Uniform Density Sphere*: Repeating the same procedure as above but for a sphere of uniform density  $\rho(r) = \rho_c$  and mass  $M$  yields the following for the surface mass density (where we use the Abel integral equation to relate volume

mass density to surface mass density);

$$\kappa = \frac{1}{\Sigma_{\text{crit}\infty}} \int_{-\infty}^{\infty} dz \rho(r) = \frac{2\rho_c}{\Sigma_{\text{crit}\infty}} \sqrt{R^2 - P^2}, \quad (4.37)$$

$$\begin{aligned} \bar{\kappa} &= \frac{1}{\pi P^2} \int_0^P \kappa(P') 2\pi P' dP' \\ &= \frac{4\rho_c}{3\Sigma_{\text{crit}\infty}} \left( \frac{R^3 - (R^2 - P^2)^{3/2}}{P^2} \right), \end{aligned} \quad (4.38)$$

where  $dz$  is along the line of sight and  $R = (3M/4\pi\rho_c)^{1/3}$  is the radius of the sphere. The signal-to-noise ratio is then computed by solving equation (14) with the above relations for  $\kappa$  and  $\bar{\kappa}$ .

3. *Truncated Isothermal Sphere*: The radial density profile of an isothermal sphere is

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2}, \quad (4.39)$$

where  $\sigma_v$  is the line-of-sight velocity dispersion of the particles (i.e., galaxies) in the system. The surface mass density is then given by

$$\kappa = \frac{1}{2}\bar{\kappa} = \frac{\theta_E}{2\theta}, \quad (4.40)$$

where  $\theta_E = \sigma_v^2/GD_d\Sigma_{\text{crit}\infty}$ . For a truncated isothermal sphere of mass  $M$  and radius  $R$ ,  $M = \int_0^R dr \rho(r) 4\pi r^2 = 2\sigma_v^2 R/G$ , so that

$$\theta_E = \frac{M}{2R} \frac{1}{D_d \Sigma_{\text{crit}\infty}}.$$

Equation (14) then gives

$$\frac{S}{N} = \frac{\langle Z \rangle M}{\sigma_\epsilon \Sigma_{\text{crit}\infty} D_d} \frac{\sqrt{\pi n}}{2R} \sqrt{\ln(\theta_{\text{out}}/\theta_{\text{in}})}. \quad (4.41)$$

4. *NFW Profile*: The NFW density profile is given by

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}, \quad (4.42)$$

where  $r_s$  and  $\rho_s$  are the scale radius and density, respectively. The mass within radius  $r$  is then

$$M(r) = 4\pi\rho_s r_s^3 \left( \ln(1 + r/r_s) - \frac{r/r_s}{1 + r/r_s} \right). \quad (4.43)$$

Bartelmann (1996) (see also Wright & Brainerd 2000) showed that the radial dependence of the tangential shear for an NFW profile is

$$\gamma_{\text{nfw}}(x) = \bar{\kappa}(x) - \kappa(x) = \frac{\rho_s r_s}{\Sigma_{\text{crit}\infty}} g(x), \quad (4.44)$$

where  $x = \theta D_d / r_s$  and

$$g(x) = \begin{cases} \frac{8 \operatorname{arctanh} \sqrt{\frac{1-x}{1+x}}}{x^2 \sqrt{1-x^2}} + \frac{4}{x^2} \ln\left(\frac{x}{2}\right) - \frac{2}{(x^2-1)} + \frac{4 \operatorname{arctanh} \sqrt{\frac{1-x}{1+x}}}{(x^2-1)(1-x^2)^{1/2}}, & (x < 1) \\ \frac{8 \arctan \sqrt{\frac{x-1}{1+x}}}{x^2 \sqrt{x^2-1}} + \frac{4}{x^2} \ln\left(\frac{x}{2}\right) - \frac{2}{(x^2-1)} + \frac{4 \arctan \sqrt{\frac{x-1}{1+x}}}{(x^2-1)^{3/2}} & (x > 1). \end{cases} \quad (4.45)$$

The signal-to-noise ratio is then

$$\frac{S}{N} = \frac{2\sqrt{\pi n} \langle Z \rangle}{\sigma_\epsilon \Sigma_{\text{crit}\infty} D_d} \rho_s r_s^2 \sqrt{\int_{x_{\text{in}}}^{x_{\text{out}}} dx x g(x)^2}. \quad (4.46)$$

There are thus three unknowns if given an overdensity of mass  $M$ :  $r_s$ ,  $\rho_s$ , and  $R$ . We therefore need a third relation in addition to equations (A10) and (A13) in order to break the degeneracy. It is obtained via the following conservation of energy argument, first put forth by Dalcanton et al. (1997) for the case of disk formation.

Assume the mass profile before collapse is a uniform sphere of radius  $R_i$  and assume that at this initial stage the system's energy is entirely gravitational ( $E = -3GM^2/5R_i$ ). As noted by Dalcanton et al. (1997), this assumption is well-motivated in the context of disk formation by the observed similarity between disk angular momentum distributions and the angular momentum dis-

tribution of a uniformly rotating sphere. It is natural to assume a similar initial condition occurs for systems at larger scales, i.e., cluster masses. As the overdensity collapses and approaches virialization, the mass distribution evolves into an NFW profile, as suggested by numerical simulations. At this stage the systems potential energy within a radius  $r$  is

$$\Phi(y) = -8\pi^2 G \rho_s^2 r_s^5 \left( 1 - \frac{2y \ln y + 1}{y^2} \right), \quad (4.47)$$

where  $y \equiv 1 + r/r_s$ . Assuming the energy of the overdensity within  $R_i$  is conserved during collapse and that the system is near virialization so that  $E \approx |\Phi|/2$  then gives  $R_i = 8.74r_s$ . Since the truncation radius is given by the radius that contains mass  $M$ , by conservation of mass  $R = R_i = 8.74r_s$  (i.e., though the mass is redistributed as the overdensity evolves the size of the sphere containing mass  $M$  is constant in time). When we include the effects of the cosmological constant in the conservation of energy argument there is little change in the result. The above relation between  $R$  and  $r_s$  thus provides the sought-after third equation needed to break the degeneracy between  $r_s$ ,  $\rho_s$ , and  $R$ . In an upcoming paper (Weinberg & Kamionkowski 2002) we show that the above approach yields concentration parameters that are slightly different from those obtained by N-body simulations (i.e., Bullock et al. 2001). Nonetheless, the concentration parameters obtained by the two approaches predict a similar abundance of virialized lenses. Note that since the N-body simulations fit the concentration parameters to virialized objects, the above analytic approach must be used in order to compute the abundances of dark lenses.

5. *Hernquist Profile*: The Hernquist profile is given by

$$\rho(r) = \frac{M_\infty}{2\pi} \frac{1}{(r/r_s)(r+r_s)^3}, \quad (4.48)$$

where  $r_s$  is the scale radius and  $M_\infty$  is the mass enclosed at infinity. The mass

within radius  $r$  is then

$$M(r) = M_\infty \left( \frac{r/r_s}{1 + r/r_s} \right)^2. \quad (4.49)$$

Using the Abel integral equation it can be shown that the dimensionless surface mass density for the Hernquist profile is

$$\kappa(x) = \frac{M_\infty}{\pi r_s^2 \Sigma_{\text{crit}\infty}} f(x), \quad (4.50)$$

where  $x = \theta D_d / r_s$  and

$$f(x) = \begin{cases} \frac{1}{(x^2-1)^2} \left[ \frac{(2+x^2) \operatorname{arctanh} \sqrt{\frac{1-x}{1+x}}}{\sqrt{1-x^2}} - \frac{3}{2} \right], & (x < 1) \\ \frac{1}{(x^2-1)^2} \left[ \frac{(2+x^2) \operatorname{arctan} \sqrt{\frac{x-1}{1+x}}}{\sqrt{x^2-1}} - \frac{3}{2} \right], & (x > 1) \end{cases} \quad (4.51)$$

The dimensionless surface mass density within  $x$  is then,

$$\bar{\kappa}(x) = \frac{2}{x^2} \frac{M_\infty}{\pi r_s^2 \Sigma_{\text{crit}\infty}} \int_0^x dx' x' f(x'). \quad (4.52)$$

The signal-to-noise ratio is then obtained by inserting the above relations into equation (14).

As in the case of the NFW profile, given an overdensity of mass  $M$ , there are three unknowns. We therefore apply the same energy conservation argument as above, assuming the overdensity is initially a uniform density sphere of radius  $R_i$  and upon collapse relaxes to a Hernquist profile. The potential energy upon collapse is

$$\Phi(y) = -\frac{GM_\infty^2}{6r_s} \left( 1 - \frac{6y^2 - 8y + 3}{y^4} \right), \quad (4.53)$$

where  $y = 1 + r/r_s$ . Assuming energy conservation and a nearly virialized overdensity yields  $R = R_i = 3.2r_s$ , allowing us to solve the signal-to-noise relation.

# Bibliography

Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625

Bartelmann M., 1996, A&A, 313, 697

Bartelmann M., 1995, A&A, 303, 643

Bartelmann M., Schneider P., 1992, A&A, 259, 413

Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291

Blandford R. D., Saust A. B., Brainerd T. G., Villumsen J. V., 1991, MNRAS, 251, 600

Brainerd T. G., Blandford R. D., Smail I., 1996, ApJ, 466, 623

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559

Cohen J. G., Hogg D. W., Blandford R. D. et al. 2000, ApJ, 538, 29

Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659

Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263

Erben T., van Waerbeke L., Mellier Y., Schneider P., Cuillandre J.-C., Castander F. J., Dantel-Fort M., 2000, A&A, 355, 23

Evrard A. E., 1989, ApJ, 341, L71

Gray M. E., Ellis R. S., Lewis J. R., McMahon R. G., Firth A. E., 2001, MNRAS, 325, 111

- Henry J. P., 1997, ApJ, 489, L1
- Henry J. P., Arnaud K. A., 1991, ApJ, 372, 410
- Hernquist L., 1990, ApJ, 356, 359
- Hoekstra H., Franx M., Kuijken K., Carlberg R. G., Yee H. K. C., Lin H., Morris S. L., Hall P. B., Patton D. R., Sawicki M., Wirth G. D., 2001, ApJ, 548, L5
- Kaiser N., 1992, ApJ, 388, 272
- Kaiser N., Wilson, G. Luppino G. A., 2000, astro-ph/0003338
- Kamionkowski M., Babul A., Cress C. M., Refregier A., 1998, MNRAS, 301, 1064
- Kitayama T., Suto Y., 1997, ApJ, 490, 557
- Kitayama T., Suto Y., 1996, ApJ, 469, 480
- Koopmans L. V. E. et al. (the CLASS collaboration), 2000, A&A, 361, 815
- Kruse G., Schneider P. 1999, MNRAS, 302, 821
- Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, MNRAS, 251, 128
- Mellier Y., 1999, ARAA, 37, 127
- Miralda-Escudé J., 1991, ApJ, 380, 1
- Miralles J. M., Erben T., Haemmerle H., Schneider P., Fosbury R. A. E., Freudling W., Pirzkal N., Jain B., White S. D. M., 2002, A&A, 388, 68
- Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
- Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 56
- Peebles P. J. E., 1984, ApJ, 284, 439

- Peebles P. J. E., 1980, *The Large Scale Structure of the Universe*, Princeton Univ. Press, Princeton, N.J.
- Postman M., Lubin L. M., Oke J. B., 2001, *AJ*, 122, 1125
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Reblinsky K., Kruse G., Jain B., Schneider P. 1999, *A&A*, 351, 815
- Schneider P., 1996, *MNRAS*, 283, 837
- Schneider P., Van Waerbeke L., Jain B., Kruse G., 1998, *MNRAS*, 296, 873
- Seitz C., Schneider P., 1997, *A&A*, 318, 687
- Smail I., Hogg D. W., Yan L., Cohen J. G., 1995, *ApJ*, 449, L105
- Stebbins A., 1997, astro-ph/9609149
- Van Waerbeke L., Mellier Y., Erben T., et al., 2000, *A&A*, 358, 30
- Viana P. T. P., Liddle A. R., 1999, *MNRAS*, 303, 535
- Viana P. T. P., Liddle A. R., 1996, *MNRAS*, 281, 323
- Weinberg N. N., Kamionkowski M., 2002, *MNRAS*, in preparation
- White M., Van Waerbeke L., Mackey J., 2002, *ApJ* in press, astro-ph/0111490
- White S. D. M., Efstathiou G., Frenk C. S., 1993, *MNRAS*, 262, 1023
- Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, *Nature*, 405, 143
- Wittman D., Tyson J. A., Margoniner V. E., Cohen J. G. Dell'Antonio I. P. 2001, *ApJL*, 557, L89
- Wright C. O., Brainerd T. G., 2000, *ApJ*, 534, 34

## Chapter 5

# Constraining Dark Energy from the Weak Gravitational Lens Abundance<sup>†</sup>

NEVIN N. WEINBERG AND MARC KAMIONKOWSKI

California Institute of Technology, Mail Code 130-33, Pasadena, CA 91125 USA

---

<sup>†</sup> A version of this chapter was first published in *MNRAS*, 341, p. 251–262 (2003).

## Abstract

We examine the prospect of using the observed abundance of weak gravitational lenses to constrain the equation-of-state parameter  $w = p/\rho$  of the dark energy. Dark energy modifies the distance-redshift relation, the amplitude of the matter power spectrum, and the rate of structure growth. As a result, it affects the efficiency with which dark-matter concentrations produce detectable weak-lensing signals. Here we solve the spherical-collapse model with dark energy, clarifying some ambiguities found in the literature. We also provide fitting formulas for the nonlinear overdensity at virialization and the linear-theory overdensity at collapse. We then compute the variation in the predicted weak-lens abundance with  $w$ . We find that the predicted redshift distribution and number count of weak lenses are highly degenerate in  $w$  and the present matter density  $\Omega_0$ . If we fix  $\Omega_0$  the number count of weak lenses for  $w = -2/3$  is a factor of  $\sim 2$  smaller than for the  $\Lambda$ CDM model  $w = -1$ . However, if we allow  $\Omega_0$  to vary with  $w$  such that the amplitude of the matter power spectrum as measured by the Cosmic Background Explorer (COBE) matches that obtained from the X-ray cluster abundance, the decrease in the predicted lens abundance is less than 25% for  $-1 \leq w < -0.4$ . We show that a more promising method for constraining the dark energy—one that is largely unaffected by the  $\Omega_0 - w$  degeneracy as well as uncertainties in observational noise—is to compare the relative abundance of virialized X-ray lensing clusters with the abundance of non-virialized, X-ray underluminous, lensing halos. For aperture sizes of  $\sim 15$  arcmin, the predicted ratio of the non-virialized to virialized lenses is greater than 40% and varies by  $\sim 20\%$  between  $w = -1$  and  $w = -0.6$ . Overall, we find that if all other weak lensing parameters are fixed, a survey must cover at least  $\sim 40$  square degrees in order for the weak lens number count to differentiate a  $\Lambda$ CDM cosmology from a dark-energy model with  $w = -0.9$  at the  $3\sigma$  level. If, on the other hand, we take into account uncertainties in the lensing parameters, then the non-virialized lens fraction provides the most robust constraint on  $w$ , requiring  $\sim 50$  square degrees of sky coverage in order to differentiate a  $\Lambda$ CDM model from a  $w = -0.6$  model to  $3\sigma$ .

---

## 5.1 Introduction

Observations of distant type Ia supernovae (SNIa) indicate that the universe is undergoing a phase of accelerated expansion (Perlmutter et al. 1999, Riess 1998). This, combined with the flat geometry favored by the cosmic microwave background (CMB) measurements (Miller et al. 1999, de Bernardis et al. 2002, Halverson et al. 2002,

Sievers et al. 2002, Lee et al. 2001) and the evidence for a low matter-density with  $\Omega_0 \sim 0.3$  (Peacock 2001, Percival et al. 2001), suggests that the bulk of the total energy density of the universe is in the form of some exotic dark energy with a negative equation of state. One of the primary objectives of cosmology today is to uncover the origin and nature of this dark energy.

A possible candidate for the dark energy is a cosmological constant  $\Lambda$ , with an equation of state  $w = p/\rho$  (where  $p$  is the pressure and  $\rho$  is the energy density of the dark energy) strictly equal to  $-1$ . Another possibility, and one that may find favor from a particle-physics point of view, is a dynamical scalar field, termed quintessence,  $Q$ . Unlike the cosmological constant, the  $Q$ -component is both time-dependent and spatially inhomogeneous with an equation of state  $w > -1$  that is likely to be redshift-dependent. Determining the value of  $w$  and how it changes with time are key to constraining the nature of the dark energy.

While the accelerating expansion implies only that  $w < -1/3$ , combinations of CMB data, SNIa data, and large-scale-structure data suggest that  $w$  is most likely in the range  $-1 \leq w < -0.6$  (Wang et al. 2000, Huterer & Turner 2001, Bean & Melchiorri 2002, Baccigalupi et al. 2002; also see appendix). Though combining these different data sets have provided some constraint on  $w$ , how  $w$  should vary with redshift is largely unknown. Particle physics offers several possible functional forms for the quintessence field's potential  $V(Q)$  and hence possible scenarios for the time history of  $w$ . Nonetheless, determining  $w$ 's redshift evolution observationally is likely to be very challenging (Barger & Marfatia 2001, Maor et al. 2001, Weller & Albrecht 2001).

Strengthening the measured constraint on  $w$  and perhaps excluding the cosmological constant as the source of the dark energy appear, however, to be attainable goals within the near future. Since the dark-energy dynamics influences both the evolution of the background cosmology and the growth of structure, it directly affects many observables. Its modification of the angular-diameter distance, the luminosity distance, and the amplitude of the matter power spectrum, are the primary sources of dark-energy constraint in measurements of CMB anisotropies, SNIa, and local cluster

abundances, respectively.

In this paper we consider another possible means of constraining  $w$ : measurement of weak gravitational-lens abundances. Weak-lensing—the weak distortion of background-galaxy images due to the deep gravitational potential of an intervening overdensity—provides a powerful technique for mapping the distribution of matter in the universe (see reviews by Bartelmann & Schneider 2001, Mellier 1999). Here we study the impact of the dark energy on the predicted redshift distribution and sky density of weak lenses. Dark energy affects the abundance of weak lenses by not only modifying the distance-redshift relation and the matter power spectrum but also by altering the rate of structure growth. In particular, the larger  $w$  is the faster and earlier objects collapse. An interesting consequence of this is that if we separate weak lenses into the two observational classes—those that have collapsed and reached virial equilibrium and are therefore X-ray luminous and those that are non-virialized and hence X-ray underluminous (Weinberg & Kamionkowski 2002; hereafter WK02)—the abundance of one class evolves slightly differently from the other. Therefore the relative fraction of these two types of lenses varies with  $w$ . This observable is especially promising as compared to measurements of absolute abundances because it is less sensitive to uncertainties in both the cosmological parameters and the noise in the lensing map.

This paper is organized as follows. In Section 2 we briefly summarize the weak-lensing signal-to-noise estimator and discuss how we determine the mass- and redshift-dependent minimum overdensity required to produce a detectable weak-lensing signal. Section 3 is devoted to the spherical-collapse model in quintessence cosmologies. We provide fitting formulas for the nonlinear overdensity at virialization and the linear-theory density at collapse, and describe our approach to normalizing the matter power spectrum. In Section 4 we show the resulting effect the dark energy has on the weak-lens abundances, and in Section 5 we present our conclusions. In an appendix, we discuss the current constraints on  $w$  without the restriction  $w > -1$  and consider the consequences that follow if the dark energy is phantom energy  $w < -1$ .

Finally, we note that a similar analysis has recently been performed by Bartel-

mann, Perrotta & Baccigalupi (2002), although not for the case of non-virialized lenses. Although we agree with their general conclusion that the weak lens abundance is a potentially sensitive probe of the dark energy, our results differ from their results in important details. We discuss these differences in Section 4.2.

## 5.2 Minimum overdensity needed to produce detectable lensing signal

In order to compute the abundance of weak gravitational lenses for dark-energy cosmologies we must first determine the necessary conditions for a halo of a given density profile and redshift to produce a detectable weak-lensing signal. Of course the more overdense a halo is relative to the background density the more it coherently distorts the nearby background galaxies and hence the stronger its lensing signal. The detectability of this signal is hampered, however, by noise in the weak-lensing map, primary of which is the intrinsic ellipticity distribution of the background galaxies. The goal is therefore to determine the minimum overdensity a halo must have such that it produces a sufficiently large signal relative to the noise so as to be detectable. A convenient method for computing this minimum overdensity is provided by Schneider's (1996) aperture-mass technique.

Consider a lens at redshift  $z_d$  of surface mass density  $\Sigma(\vartheta)$  within an angular radius  $\vartheta$ . For a source at redshift  $z_s$  the convergence  $\kappa$  is given by

$$\kappa(\vartheta) = \frac{\Sigma(\vartheta)}{\Sigma_{\text{crit}}}, \quad \Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}}, \quad (5.1)$$

where  $D_d$ ,  $D_s$ , and  $D_{ds}$  are the angular-diameter distances between the lens and the observer, the source galaxy and the observer, and the lens and the source, respectively. Following Schneider (1996), define a spatially filtered mass inside a circular aperture of angular radius  $\theta$ ,

$$M_{\text{ap}}(\theta) \equiv \int d^2\boldsymbol{\vartheta} \kappa(\boldsymbol{\vartheta}) U(|\boldsymbol{\vartheta}|), \quad (5.2)$$

where  $U(\vartheta)$  is a continuous weight function that vanishes for  $\vartheta > \theta$ . If  $U(\vartheta)$  is a compensated filter function,

$$\int_0^\theta d\vartheta \vartheta U(\vartheta) = 0, \quad (5.3)$$

then  $M_{\text{ap}}$  can be expressed in terms of the tangential component of the observable shear,  $\gamma_t$ ,

$$M_{\text{ap}}(\theta) = \int d^2\vartheta \gamma_t(\vartheta) Q(|\vartheta|), \quad (5.4)$$

where the function  $Q$  is related to  $U$  by

$$Q(\vartheta) = \frac{2}{\vartheta^2} \int_0^\vartheta d\vartheta' \vartheta' U(\vartheta') - U(\vartheta). \quad (5.5)$$

In this paper we use the  $l = 1$  radial filter function from the family given in Schneider et al. (1998):

$$U(\vartheta) = \frac{9}{\pi\theta^2}(1-x^2)\left(\frac{1}{3}-x^2\right), \quad Q(\vartheta) = \frac{6}{\pi\theta^2}x^2(1-x^2), \quad (5.6)$$

where  $x = \vartheta/\theta$ . Taking the expectation value over galaxy positions and taking into account the redshift distribution of source galaxies then gives

$$M_{\text{ap}}(\theta) = \langle Z \rangle \int d^2\vartheta \langle \gamma_t \rangle(\vartheta) Q(|\vartheta|), \quad (5.7)$$

where  $\langle \gamma_t \rangle(\vartheta)$  is the mean tangential shear on a circle of angular radius  $\vartheta$ . The function  $\langle Z \rangle$ , given by

$$\langle Z \rangle = \int dz_s p_z(z_s) Z(z_s; z_d), \quad (5.8)$$

where  $p_z(z_s)$  is the redshift distribution of source galaxies and (Seitz & Schneider 1997)

$$Z(z_s; z_d) \equiv \frac{\lim_{z_s \rightarrow \infty} \Sigma_{\text{crit}}(z_d; z_s)}{\Sigma_{\text{crit}}(z_d; z_s)} = \frac{\Sigma_{\text{crit}\infty}(z_d)}{\Sigma_{\text{crit}}(z_d; z_s)} \quad (5.9)$$

allows a source with a known redshift distribution to be collapsed onto a single redshift  $z_s$  satisfying  $Z(z_s) = \langle Z \rangle$  (Seitz & Schneider 1997; Bartelmann & Schneider 2001).

The source-redshift distribution is taken to be

$$p_z(z_s) = \frac{\beta z_s^2}{\Gamma(3/\beta)z_0^3} \exp[-(z_s/z_0)^\beta], \quad (5.10)$$

with  $\beta = 1.5$  and mean redshift  $\langle z_s \rangle \approx 1.5z_0 = 1.2$  (cf., Smail et al. 1995; Brainerd et al. 1996; Cohen et al. 2000). Finally, assuming the ellipticities of different images are uncorrelated it can be shown (cf., Kruse & Schneider 1999) that the dispersion  $\sigma_M(\theta)$  of  $M_{\text{ap}}$  is

$$\sigma_M^2(\theta) = \frac{\pi\sigma_\epsilon^2}{n} \int_0^\theta d\vartheta \vartheta Q^2(\vartheta), \quad (5.11)$$

where  $n$  is the number density of galaxy images and  $\sigma_\epsilon$  is the dispersion in the galaxies' intrinsic ellipticity. In this paper we assume  $n = 30 \text{ arcmin}^{-2}$  and  $\sigma_\epsilon = 0.2$ . The signal-to-noise ratio  $S$  within an aperture radius  $\theta$  is then given by

$$S = \frac{M_{\text{ap}}}{\sigma_M} = \frac{2\langle Z \rangle \sqrt{\pi n} \int_0^\theta d\vartheta \vartheta \langle \gamma_t \rangle(\vartheta) Q(\vartheta)}{\sigma_\epsilon \sqrt{\int_0^\theta d\vartheta \vartheta Q^2(\vartheta)}}. \quad (5.12)$$

The tangential shear at  $\vartheta$ ,  $\langle \gamma_t \rangle(\vartheta)$ , depends on the amplitude and shape of the lensing halo's density profile. Bartelmann (1995) showed that  $\langle \gamma_t \rangle(\vartheta) = \bar{\kappa}(\vartheta) - \langle \kappa \rangle(\vartheta)$ , where  $\langle \kappa \rangle(\vartheta)$  is the dimensionless mean surface mass density on a circle of radius  $\vartheta$  and  $\bar{\kappa}(\vartheta)$  is the dimensionless mean surface mass density within a circle of radius  $\vartheta$ . In this paper we describe the mass density of lensing halos with the universal density profile introduced by Navarro, Frenk & White (1996; 1997; hereafter NFW). Thus, for an NFW halo at a given redshift with a given mass and mean overdensity relative to the background ( $\Delta \equiv \langle \rho_{\text{pert}} \rangle / \rho_b$ ), we can solve for the parameters of the profile (i.e., the scale radius and the scale density) and obtain an estimate of  $\langle \gamma_t \rangle(\vartheta)$ .

Since the details of how we solve for the NFW-profile parameters are given in the Appendix of WK02 we do not repeat them here. Briefly describing the key points, we assume a collapse process analogous to that used by Dalcanton et al. (1997), in which the mass profile before collapse is a uniform sphere and that as the overdensity collapses and approaches virialization, the mass distribution evolves into an NFW

profile. Based on conservation of energy and mass we then obtain a halo concentration expressed in terms of the ratio between the turnaround radius and the scale radius. The benefit of such a procedure is that it enables us to solve for the NFW-profile parameters for both virialized and non-virialized systems, despite the fact that the concentration parameters quoted in the literature are for virialized systems only. We have compared the virialized halo concentration parameters inferred using this procedure with those obtained from numerical simulations and found that the two are roughly consistent with each other. For example, in our approach a  $10^{14}M_{\odot}$  object at  $z = 0$  has an NFW halo concentration  $c$  (defined as the ratio of the radius enclosing an overdensity of 200 to the scale radius) of 8.74 if  $\Delta = 200$  and 6.53 if  $\Delta = 100$ , while numerical simulations typically yield values of  $c \sim 7$  for virialized halos.

Note too that although N-body simulation fits to profiles have so far only been for virialized haloes, because most of the dark lenses are well past turnaround ( $\Delta \gtrsim 100$ ) and because the STHC model likely breaks down at some point before virialization, assuming an NFW profile for dark lenses is a fair approximation. Furthermore, since virialization occurs from inside-out, the central regions of a dark lens, where most of the weak-lensing signal comes from, are likely to be near virialization and thus well-described by the NFW form. Lastly, although we only consider the NFW profile in this paper, in WK02 we computed virialized and dark lens abundances assuming various types of other profiles including a uniform density sphere, the Hernquist (Hernquist 1990) profile, and the Isothermal Sphere profile. Although the total number count of weak lenses does change for these different profiles the normalized redshift distributions and the number count ratios of dark lenses to virialized lenses are largely unaffected. In this paper, we are chiefly concerned with the possibility of constraining the equation of state of the dark energy via these differential, rather than cumulative, abundances. Assuming the weak lenses have an NFW profile is therefore not crucial to the arguments or conclusions made herein.

With the density profile known we can determine, using equation (12), the expected value of  $S$ . The minimum mean overdensity,  $\Delta_{\min}$ , needed to produce a detectable lens is then given by that overdensity for which  $S > S_{\min}$ . In this paper

we assume  $S_{\min} = 5$  and  $\theta = 5'$ , unless stated otherwise.

### 5.3 Spherical collapse in dark energy cosmologies

According to the spherical model of gravitational collapse a density perturbation with a nonlinear overdensity  $\Delta$  corresponds to a particular position along the linear-theory evolutionary cycle. Thus the minimum nonlinear overdensity  $\Delta_{\min}$  described above corresponds to a minimum linear-theory overdensity  $\delta_{\min}$ ; if an object of mass  $M$  at redshift  $z$  has a linear-theory overdensity  $\delta > \delta_{\min} = \delta_{\min}(M, z)$ , then it is sufficiently overdense to produce a detectable weak-lensing signal. By determining  $\delta_{\min}$  from the computed  $\Delta_{\min}$  we can apply the Press-Schechter (1974) theory to calculate the number of halos per unit mass and redshift with  $\delta > \delta_{\min}$  and hence  $S > S_{\min}$ . We can then find the redshift distribution and sky density of weak lenses and how these observables vary with  $w$ . We will show that for a broad range of dark-energy cosmologies a substantial fraction of detectable weak gravitational lenses have  $\delta_{\min} < \delta_c \approx 1.69$ , where  $\delta_c$  is the critical density threshold for collapse. Those objects with  $\delta < \delta_c$  are commonly thought to be density perturbations that have not yet reached virialization and are therefore expected to have observational properties that are very different from typical virialized lensing clusters.

In this Section, we present the approach used to map the minimum nonlinear overdensity  $\Delta_{\min}$  to a minimum linear-theory overdensity  $\delta_{\min}$  for quintessence models (QCDM). We describe the dynamical equations of gravitational collapse in QCDM and give fitting formulas for the nonlinear overdensity at virialization,  $\Delta_{\text{vir}}(z)$ , and the critical density  $\delta_c$ . We then discuss how we calculate the abundances of weak gravitational lenses, both those with  $\delta < \delta_c$  and those with  $\delta > \delta_c$ . Below we assume a flat cosmology with a Hubble parameter  $h = 0.65$ , a spectral index  $n_s = 1$ , a baryon density  $\Omega_b h^2 = 0.02$ , and  $\Omega_0 = 0.3$ , unless stated otherwise.

### 5.3.1 Dynamics

In quintessence the dark energy is a dynamical, time-dependent component,  $Q$ , with an equation of state parametrized by  $w \equiv p_Q/\rho_Q$ , the pressure divided by the energy density. The evolution of the energy density with the cosmological scale factor goes as  $\rho_Q \propto a^{-3(1+w)}$ , so that for  $w = -1$  the standard cosmological-constant model,  $\Lambda$ CDM, is recovered. Current observational evidence cannot yet rule out a  $w$  in the range  $-1 \leq w \lesssim -0.5$ .

In order to relate a nonlinear overdensity to a linear-theory overdensity in QCDM we must first solve for the evolution of the overdensity's radius,  $R$ , with time. For a spherical overdensity patch with uniform matter density  $\rho_{\text{pert}} = 3M/4\pi R^3$  the evolution is described by the momentum component of the Einstein equations (Wang & Steinhardt 1998; hereafter WS98):

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3} [\rho_{\text{pert}} + (1 + 3w)\rho_Q]. \quad (5.13)$$

As WS98 pointed out, for  $w \neq -1$  the space curvature  $k_{\text{pert}}$  inside the overdensity patch is time-dependent. Physically, this is because the evolution of the energy density in the  $Q$  component is evolving independently of the change in radius of the overdensity patch. As a result, one cannot assume that within the collapsing overdensity the rate of change of the internal energy in the  $Q$ -component,  $u_Q$ , equals the rate of work done by the  $Q$ -component. That is, because  $d\rho_Q/dt$  is nonzero unless the  $Q$ -component is the cosmological constant,

$$\begin{aligned} \frac{du_Q}{dt} &= \frac{d}{dt}(\rho_Q V) \\ &\neq -p_Q \frac{dV}{dt}, \end{aligned} \quad (5.14)$$

where  $V \propto R^3$  is the volume of the overdensity patch. Therefore equation (13) cannot be cast in the form of a first-order differential equation as is often done when going from an acceleration equation to a Friedman-like energy equation. Assuming a constant  $k_{\text{pert}}$ , as was done in the version 1 preprint of Łokas & Hoffman (2001),

yields significantly different solutions for the evolution of the radius,  $R(t)$ , and hence for  $\Delta_{\text{vir}}(z)$  and  $\delta_c$ .

If we combine equation (13) with the Friedman equation for the background,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\rho_b + \rho_Q), \quad (5.15)$$

and impose the boundary conditions  $dR/da|_{a=a_{\text{ta}}} = 0$  and  $R|_{a=0} = 0$ , where  $a_{\text{ta}}$  is the scale factor at turn-around, then for a spherical density perturbation with a given  $\Delta$  and redshift  $z$ , the unique temporal evolution of the overdensity, from linearity to nonlinearity, can be solved (cf., Appendix A in WS98). We then have a one-to-one map from  $\Delta(z)$  to  $\delta(z)$ , as shown in Figure 1 for the cases  $w = -1, -2/3$  and  $-1/3$ . The map has a mild  $w$  dependence, with a given  $\delta$  corresponding to a slightly larger  $\Delta$  as  $w$  increases. This is a consequence of the earlier formation of structure in QCDM models relative to  $\Lambda$ CDM models; overdensities collapse faster and are therefore more concentrated for  $w > -1$ . This point is well illustrated in Figure 2, where we show the growth of a spherical perturbation for the same quintessence models. As expected, the larger  $w$  is, the earlier structures reach turnaround and collapse.

It can be shown that in the limit  $\delta \rightarrow \delta_c$  the spherical-collapse model predicts that the radius,  $R$ , of the overdensity goes to zero and hence  $\Delta \rightarrow \infty$ . Of course well before reaching the singular solution an actual overdensity will virialize, thereby halting its collapse. To account for this fact we invoke a simple smoothing scheme in which the radius of the matter perturbation is constant with time upon reaching the virialized overdensity (see Figure 2). We refer the reader to WK02 for details of the smoothing method.

As described in WS98, the value of  $\Delta_{\text{vir}}(z)$  for quintessence models, needed here in order to implement the smoothing scheme, can be obtained via the virial theorem, energy conservation, and solving equations (13) and (15) for the overdensity at turnaround. In Figure 3 we show the resulting numerical solution to  $\Delta_{\text{vir}}(z)$ . We find that an accurate fitting function to  $\Delta_{\text{vir}}(z)$  for  $-1 \leq w \leq -0.3$ , modeled after the

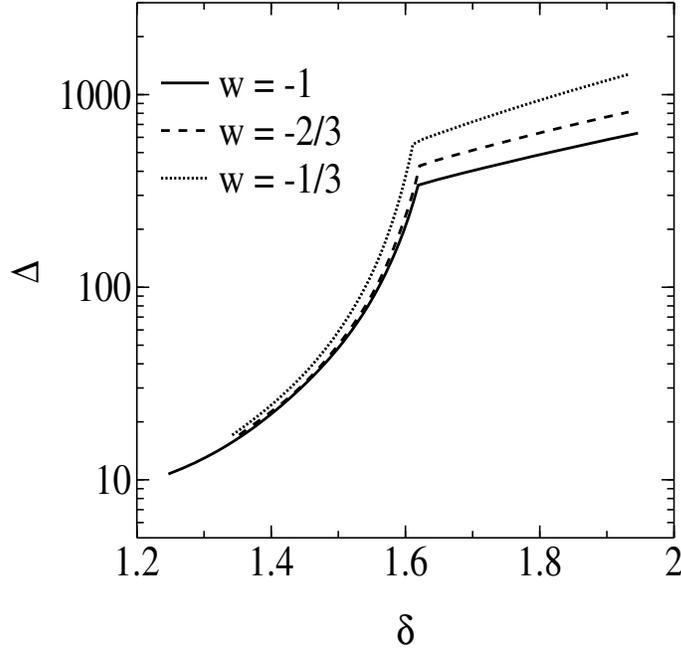


Figure 5.1 Nonlinear overdensity as a function of the linear-theory overdensity for three constant- $w$  models. The full solution of the spherical-collapse model predicts collapse to an infinite overdensity as  $\delta \rightarrow \delta_c \sim 1.69$ . According to the smoothing scheme, however, once a mass concentration reaches the virialization overdensity  $\Delta_{\text{vir}}(z)$ , its radius remains constant so that the overdensity increases in proportion to the decrease in the background density. Shown are the smoothing-scheme solutions for mass concentrations that reach the virialization overdensity at  $z = 0$ .

approximation given in Kitayama & Suto (1996) for a  $\Lambda$ CDM cosmology, is

$$\Delta_{\text{vir}}(z) = 18\pi^2 [1 + a\Theta^b(z)], \quad (5.16)$$

where

$$\begin{aligned} a &= 0.399 - 1.309(|w|^{0.426} - 1), \\ b &= 0.941 - 0.205(|w|^{0.938} - 1), \end{aligned} \quad (5.17)$$

and  $\Theta(z) = 1/\Omega_m(z) - 1 = (1/\Omega_0 - 1)(1+z)^{3w}$ . Since structures start to form earlier the larger  $w$  is, the mean gas temperature in collapsing objects is higher in larger- $w$  models. As a result, a greater overdensity is required in order for such objects to

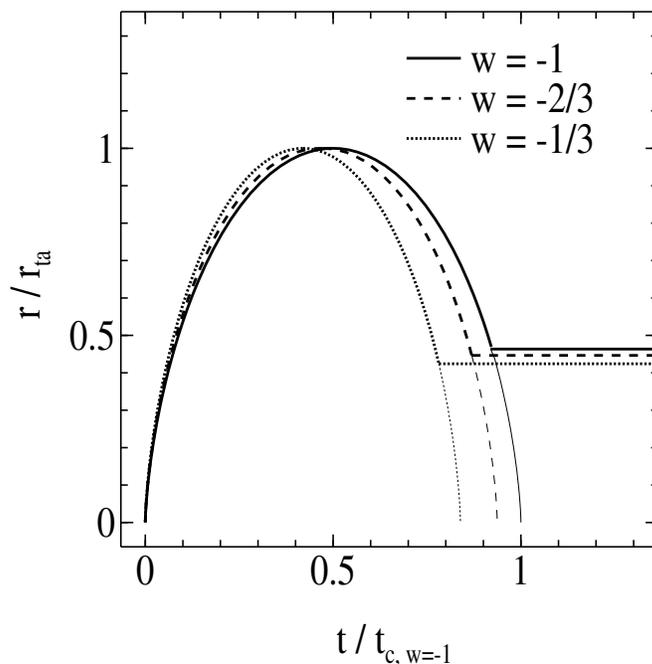


Figure 5.2 Radial evolution of a density perturbation that is collapsing today according to the spherical-collapse model. The ordinate gives the radius,  $r$ , in units of the turnaround radius,  $r_{\text{ta}}$ , and the abscissa gives the time,  $t$ , in units of the overdensity collapse-time for the  $\Lambda$ CDM model. As  $w$  increases perturbations reach turnaround and collapse earlier, although growth is suppressed earlier as well. The collapse to a singularity predicted by the solution of the spherical-collapse model is avoided by the smoothing scheme (*thick curves*), which yields a constant radius once the virialized overdensity is reached.

become bound and virialized, explaining why  $\Delta_{\text{vir}}$  rises with increasing  $w$ . Note, however, that for  $\Delta(z) < \Delta_{\text{vir}}(z)$  the map from nonlinear to linear overdensity has a weak dependence not only on  $w$  but on  $\Omega_0$  and redshift as well. The critical threshold for collapse today  $\delta_c(z=0) = \delta_c(z) D(0, \Omega_0, w)/D(z, \Omega_0, w)$ , where  $D(z, \Omega_0, w)$  is the linear growth factor (see WS98), also has a weak dependence on  $\Omega_0$  and  $w$ , as shown in Figure 4. For  $0.1 \leq \Omega_0 \leq 1$  and  $-1 \leq w \leq -0.3$ , we find that an accurate fitting function to  $\delta_c(z)$ , also modeled after the approximation given in Kitayama & Suto

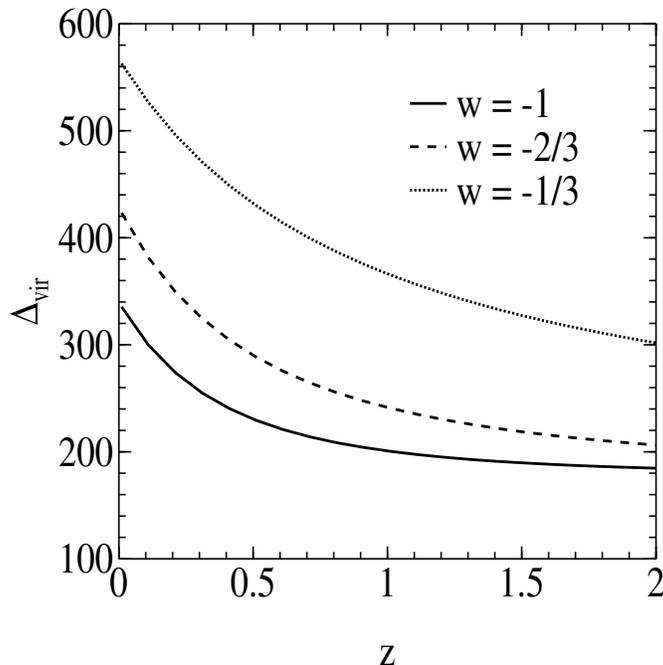


Figure 5.3 Nonlinear overdensity at virialization as a function of redshift for three constant- $w$  models. As  $w$  increases  $\Delta_{\text{vir}}$  increases because overdensities collapse earlier, when the mean gas temperature was higher. For all models  $\Delta_{\text{vir}}$  asymptotes to the Einstein–de Sitter value of 178 at high redshift.

(1996) for a  $\Lambda$ CDM cosmology, is

$$\begin{aligned} \delta_c(z) &= \frac{3(12\pi)^{2/3}}{20} [1 + \alpha \log_{10} \Omega_m(z)], \\ \alpha &= 0.353w^4 + 1.044w^3 + 1.128w^2 \\ &\quad + 0.555w + 0.131. \end{aligned} \tag{5.18}$$

Incorrectly assuming that  $k_{\text{pert}}$  is constant, however, yields a  $\delta_c(z=0)$  with a much stronger dependence on these parameters, with inferred values for  $\Omega_0 = 0.3$  of  $\delta_c(z=0) \sim 1.5$  and  $\sim 1.0$  for  $w = -2/3$  and  $w = -1/3$ , respectively (Lokas & Hoffman 2001).

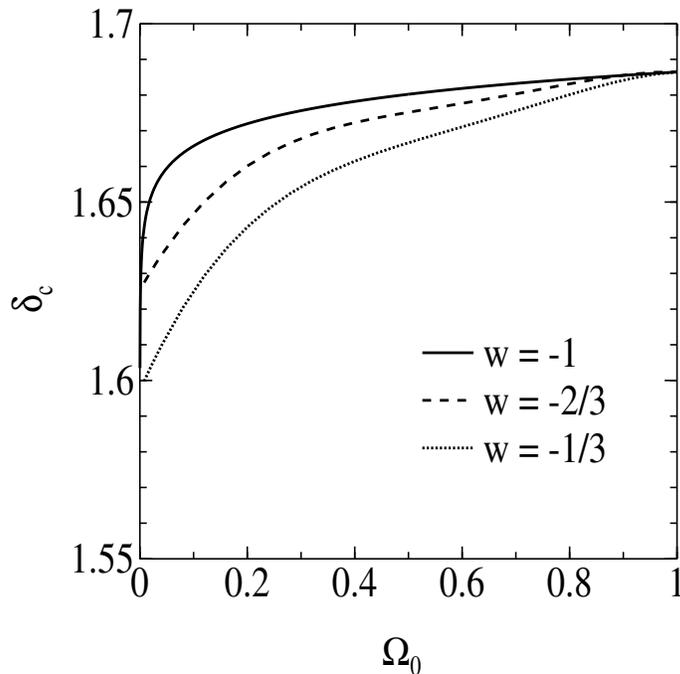


Figure 5.4 Linear-theory critical threshold for collapse,  $\delta_c$ , at  $z = 0$  as a function of  $\Omega_0$  for three constant- $w$  models.  $\delta_c$  does not vary significantly over a wide range in  $w$  or  $\Omega_0$ .

### 5.3.2 Abundances

Since we are interested in computing the abundances of both virialized weak lenses and non-virialized weak lenses we consider two ranges of overdensity in our lens-abundance calculations: (1)  $\delta_{\min} < \delta < \delta_c$ , the non-virialized lenses, and, (2)  $\delta > \delta_c \geq \delta_{\min}$ , the virialized lenses. As we showed in WK02, the mass distribution of both the dark and virialized lenses peaks near  $\sim 10^{14} M_\odot$ , though the virialized lenses' mass threshold of  $\sim 10^{13} M_\odot$  is several times *smaller* than that of the dark lenses (see WK02, Figure [7]). Although both types of lenses correspond to overdensities in a similar mass range, the virialized lenses are typically virialized clusters that form at rare (e.g.,  $> 3\sigma$ ) high-density peaks of a Gaussian primordial distribution, while the non-virialized lenses correspond to proto-clusters (e.g.,  $2\sigma - 3\sigma$  peaks)—mass overdensities that have not yet undergone gravitational collapse and virialized, but that have begun to break away from the cosmological expansion. These proto-clusters should contain galaxies and perhaps a few groups that later merge to form the cluster (cf., White, van Waerbeke,

& Mackey 2002). The timescale for collapse of cluster-mass objects is large, and the overdensities can be very large even before they have virialized. It should therefore not be too surprising that proto-clusters produce a weak-lensing signal that resembles that from virialized clusters.

Though the lensing signals may be similar, the two lens types are expected to have different observational features. In particular, since the X-ray luminosity is a very rapidly varying function of the virialized mass, the summed X-ray emission from a non-virialized lens should be much smaller than that from a fully virialized lensing cluster of the same mass. In referring to these proto-clusters as “dark,” we thus mean that they should be X-ray underluminous. Although the mass-to-light ratio of these clusters should be comparable to those for ordinary clusters, since (1) high-redshift clusters may be difficult to pick out in galaxy surveys and, (2) proto-clusters will typically have a sky density a few times smaller than ordinary clusters, it would also not be surprising if these dark lenses had no readily apparent corresponding galaxy overdensity. Observational evidence of such dark lenses has been reported in detections by Erben et al. (2000), Umetsu & Futamase (2000), Miralles et al. (2002), Dahle et al. (2002), and Koopmans et al. (2000), the latter involving a detection through strong, rather than weak, lensing. A more detailed discussion of the features that may distinguish dark and virialized weak lenses is given in WK02.

Since the weak-lensing signal reveals only the projected mass distribution, line-of-sight projection effects can lead to false halo detections. Using mock observations of numerical simulations White, van Waerbeke, & Mackey (2002) showed that the presence of large-scale structure results in projection effects that significantly limit the efficiency of a weak lensing survey. Similarly, Hoekstra (2002) showed that the combined effect of large-scale structure and the intrinsic ellipticities of background galaxies can lead to 20 – 40% errors in the determination of a lensing halo’s mass. Such effects will obviously hamper efforts to constrain cosmological parameters from cumulative number counts of weak lenses. However, it is important to note that because the mass distributions of both dark lenses and virialized lenses are similar, false halo detections are not expected to affect the dark lenses anymore than they do

the virialized lenses. In fact, because the dark lenses have a larger mass threshold, a weak-lensing survey might be more efficient at detecting dark lenses than virialized lenses. Complementary observations in the optical and X-ray (and perhaps strong lensing and the Sunyaev-Zel'dovich effect) will clearly help to minimize the number of false detections. Although virialized lenses might benefit more from such follow-up observations, dark lenses will likely benefit as well, as they are expected to have a slight overdensity of galaxies as compared to the field, and also perhaps emit a weak X-ray signal. Therefore, even though systematic uncertainties like false halo detections might render the absolute number count of dark or virialized lenses impractical as a means of constraining cosmological parameters, the relative number count of dark to virialized lenses remains a viable option since both lens types are, for the most part, equally affected by such systematics. Of course these conclusions are based on the simplifying assumptions inherent in the STHC model. It would therefore be very interesting to compare the analytic weak-lensing results obtained in this paper with numerical simulation predictions, in which non-virialized haloes are considered in addition to the oft-considered virialized haloes. This would provide an independent means of quantifying the completeness of future weak-lensing halo searches for both virialized and dark lenses. However, that comparison is beyond the scope of the present paper.

Turning now to the details of the calculational method, in order to compute the abundances of virialized and dark lenses we need to know the probability that an object of a given mass at a given redshift is in one of the above mentioned ranges in overdensity. If we assume Gaussian statistics for the initial linear-theory density field, then the probability that an object's overdensity is in the range  $\delta_1 < \delta < \delta_2$  is

$$P(\delta_1 < \delta < \delta_2) = \text{erf}\left(\frac{\nu_2}{\sqrt{2}}\right) - \text{erf}\left(\frac{\nu_1}{\sqrt{2}}\right), \quad (5.19)$$

where “erf” is the error function,  $\nu = \delta/\sigma$ , and  $\sigma = \sigma(M, z)$  is the rms density fluctuation of an object of mass  $M$  at redshift  $z$ . From Press-Schechter theory, we have that the comoving number density of virialized objects (those with  $\delta > \delta_c$ ) of

mass  $M = 4\pi R^3 \rho_0/3$  in the interval  $dM$  that are at redshift  $z$  in a universe with comoving background density  $\rho_0$  is,

$$\frac{dn}{dM}(M, z) = \sqrt{\frac{2}{\pi}} \frac{\rho_0}{M^2} \frac{\delta_c(z)}{\sigma(M, z)} \left| \frac{d \ln \sigma(M, z)}{d \ln M} \right| \times \exp \left[ -\frac{\delta_c(z)^2}{2\sigma^2(M, z)} \right]. \quad (5.20)$$

We can therefore compute the abundance of objects in the overdensity range  $\delta_1 < \delta < \delta_2$  by convolving the above mass function of virialized objects with  $P(\delta_1 < \delta < \delta_2)/P(\delta > \delta_c)$ . Specifically, the fraction of objects that can lens relative to those that are virialized is, for dark lenses,

$$f_{\text{dark}}(M, z) = \begin{cases} \frac{P(\delta_{\min} < \delta < \delta_c)}{P(\delta > \delta_c)}, & \delta_{\min} < \delta_c; \\ 0, & \text{otherwise,} \end{cases} \quad (5.21)$$

and for virialized lenses,

$$f_{\text{vir}}(M, z) = \begin{cases} \frac{P(\delta > \delta_{\min})}{P(\delta > \delta_c)}, & \delta_{\min} > \delta_c; \\ 1, & \text{otherwise.} \end{cases} \quad (5.22)$$

As noted in WK02, the lower the mass of the object the larger the minimum overdensity needed to produce a detectable weak-lensing signal. For low enough masses the minimum overdensity becomes so large that both  $f_{\text{dark}}$  and  $f_{\text{vir}}$  approach zero, thereby imposing an effective weak-lensing mass threshold. It is worthwhile to note that we have also considered the mass function suggested by Sheth & Tormen (1999), a variant of the Press-Schechter mass function that more accurately reproduces the mass functions found in numerical simulations. However, since the resulting lens abundances are essentially the same for both mass functions and because it is useful to compare our results with previous theoretical investigations of weak lens abundances, which often used the Press-Schechter mass function (e.g., Kruse & Schneider

1999), we only show results for the Press-Schechter mass function.

To summarize, given  $f$  and equation (20) we can compute the total comoving number density of weak lenses of a particular type. Multiplying by the comoving volume element  $dV_c/dz d\Omega(w)$  then gives the differential number count of lensing objects per steradian, per unit redshift interval:

$$\frac{dN}{dz d\Omega} = \frac{dV_c}{dz d\Omega} \int_0^\infty f(M) \frac{dn}{dM}(M) dM. \quad (5.23)$$

By integrating over redshift we can then compute the number of dark and virialized lenses we expect to see per unit area of sky for a given QCDM model.

### 5.3.3 Normalizing the power spectrum

In equation (23) the volume term and the two terms within the integrand are all functions of  $w$ . While the predicted abundance of weak lenses will therefore vary with  $w$ , the degree to which it will vary depends on the shape and normalization of the power spectrum of density fluctuations. In particular, to compute the abundance of weak-lenses we need to know  $\sigma(M, z)$ .

For the *shape* of the power spectrum we use the fitting formulas given in Ma et al. (1999) for QCDM models with the transfer function and shape parameter for  $\Lambda$ CDM models given by Bardeen et al. (1986) and Hu & Sugiyama (1996, eqs. [D-28] and [E-12]), respectively. Since the Q-component does not cluster on scales less than  $\sim 100$  Mpc (Caldwell, Dave & Steinhardt 1998), at the weak-lensing scales the shape of the spectrum does not differ significantly from the well studied  $\Lambda$ CDM shape.

The *normalization* of the power spectrum, often expressed in terms of  $\sigma_8$ , the rms fluctuation today at a scale of  $8 h^{-1}$  Mpc, is not as well-constrained as its shape and will in general be a function of  $w$ . There are two different methods commonly used to obtain the normalization: to fix it by the observed X-ray cluster abundance or to fix it by the CMB large-scale anisotropies observed by the COBE satellite. Both approaches have comparable uncertainties; the cluster abundance constraint on  $\sigma_8$  has a 20% uncertainty at the  $2\sigma$  level (WS98) while the COBE constraint has a 7%

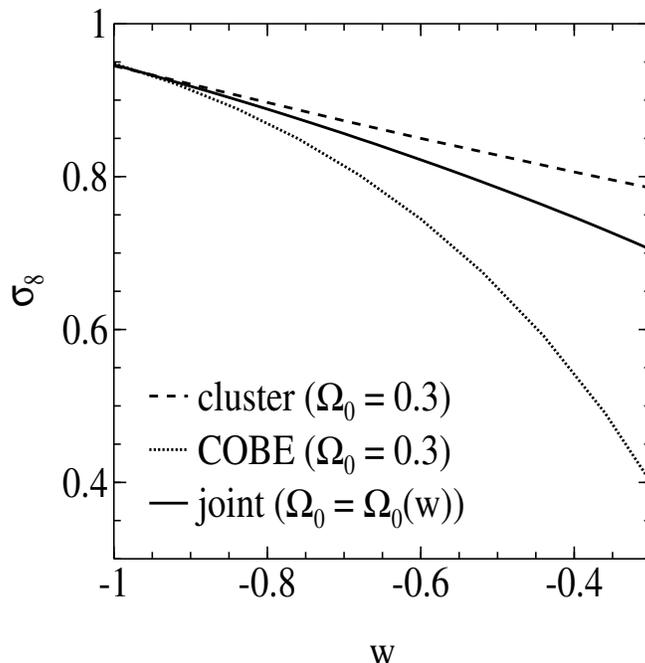


Figure 5.5 Dependence of  $\sigma_8$  on  $w$  as obtained using three different approaches: fixing  $\Omega_0 = 0.3$  and normalizing to the observed X-ray cluster abundance (*dashed* line), fixing  $\Omega_0 = 0.3$  and normalizing to COBE (*dotted* line), and allowing  $\Omega_0$  to vary with  $w$  such that the cluster abundance constraint matches the COBE constraint (*solid* line).

uncertainty at the  $1\sigma$  level (Bunn & White 1997). To obtain an estimate of how  $\sigma_8$  varies with  $w$  so that we may, in turn, determine how  $dN/dzd\Omega$  varies with  $w$  for dark and virialized lenses, we will consider three possible approaches. The first two involve fixing the cosmological parameters (e.g.,  $\Omega_0$ ,  $h$ ,  $\Omega_b$ ,  $n_s$ ) and using either the cluster-abundance constrained  $\sigma_8(w)$  or the COBE constrained  $\sigma_8(w)$ . For the former we will use the fit given in WS98, and for the latter the fit given by Ma et al. (1999); see Figure 5. The third approach is to allow the cosmological parameters to be free parameters and then jointly match the cluster-abundance constraint with the COBE constraint so that each gives the same  $\sigma_8(w)$ . Since measurements of  $\sigma_8$  are most degenerate with  $\Omega_0$ , we will let  $\Omega_0$  be the parameter that varies. In Figure 6 we show the region in the  $\Omega_0$ - $w$  plane where the X-ray cluster-abundance constraint, at the 95% confidence level, overlaps the COBE constraint. The solid curve shows where the central values match, with the resulting range in  $\Omega_0$  ( $0.3 \lesssim \Omega_0 \lesssim 0.4$

for  $-1 < w < -0.4$ ) within observational uncertainties (Wang et al. 2000). The corresponding  $\sigma_8(w)$  curve is shown in Figure 5. As we will show, the predicted weak-lens abundances and how they vary with  $w$  strongly depend on which  $\sigma_8(w)$  normalization approach is chosen.

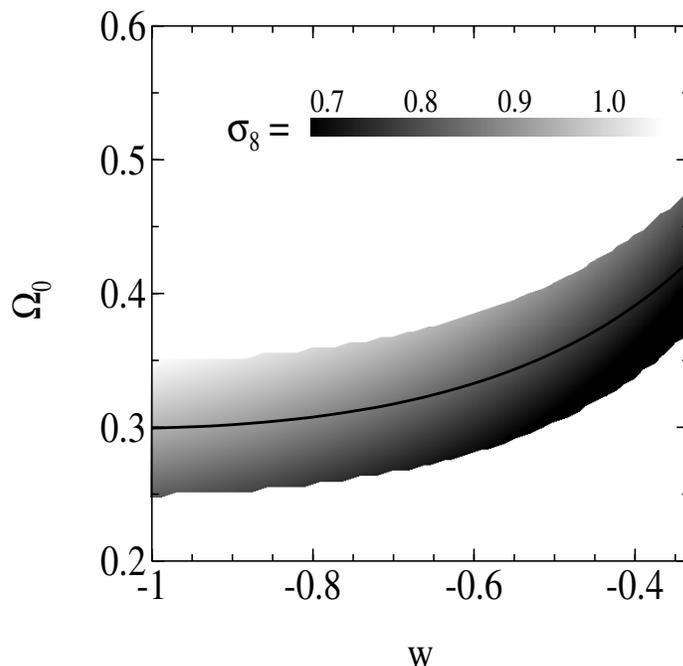


Figure 5.6 Region in the  $\Omega_0$ - $w$  plane where the X-ray cluster abundance constraint of  $\sigma_8$ , at 95% confidence, overlaps the COBE constraint of  $\sigma_8$ . The gray scale gives the corresponding  $\sigma_8$  values and the *solid* line shows wheres the central values match.

## 5.4 Results

We are interested in determining whether the number count and redshift distribution of both dark and virialized weak lenses have the potential to constrain  $w$ . Another possibly useful observable for this purpose is the number count of dark lenses relative to virialized lenses. Since dark lenses are at an earlier stage of their dynamical evolution as compared to virialized lenses, those cosmologies that favor a faster growth of structure (i.e., QCDM models with larger  $w$ ) will, for a given  $\sigma_8$ , have fewer dark lenses and more virialized lenses. The ratio of the two is therefore expected to vary

with  $w$ . A priori, this latter observable seems particularly promising. As discussed in WK02, the ratio of dark to virialized lenses is not very sensitive to observational noise in the weak-lensing maps since observational noise equally affects the detectability of both types of lenses. Contrastingly, uncertainties in observational noise will make it difficult to constrain  $w$  by simply comparing predicted weak-lens number counts with observed weak-lens number counts.

Before presenting how the above observables are modified by the dark energy we first discuss how each of the factors that determine the observed abundance is affected by changes in  $w$ . Doing so provides both physical insight into the results and illustrates the calculational procedure discussed in the previous sections.

### 5.4.1 Preliminaries

As noted above, the predicted abundance of weak lenses will vary with  $w$  on account of three factors: the comoving volume element, the Press-Schechter comoving number density of virialized objects, and the value of  $f_{\text{dark/vir}}$  [equations (21) and (22)]. The degree to which each varies depends on the chosen  $\sigma_8(w)$  normalization. As Figure 7 shows,  $dV_c/dz d\Omega$  decreases monotonically with increasing  $w$  for both fixed  $\Omega_0$  and  $\Omega_0 = \Omega_0(w)$  as given by jointly normalizing  $\sigma_8$  to COBE and the cluster abundance. However, because the joint normalization yields a larger  $\Omega_0$  with  $w$  and a less significant decline in  $\sigma_8$  for  $w > -1$  as compared to the COBE normalization with  $\Omega_0$  fixed, the former approach predicts a nearly constant virialized object number density with increasing  $w$  while the latter predicts a significant decrease in the number density.

A similar trend is seen in the functions  $f_{\text{vir}}$  and  $f_{\text{dark}}$ , as Figure 8 demonstrates. Here we plot the fraction of objects that have not yet reached turnaround ( $0 < \Delta < \Delta_{\text{ta}}$ ) and the fraction of objects that are between turnaround and virialization ( $\Delta_{\text{ta}} < \Delta < \Delta_{\text{vir}}$ ) relative to those objects that are virialized ( $\Delta > \Delta_{\text{vir}}$ ). The figure illustrates several key elements of structure formation according to the spherical-collapse model for dark-energy cosmologies. First, the fraction,  $\chi$ , of objects in both of these lower-overdensity ranges increases with mass in accordance with the hierar-

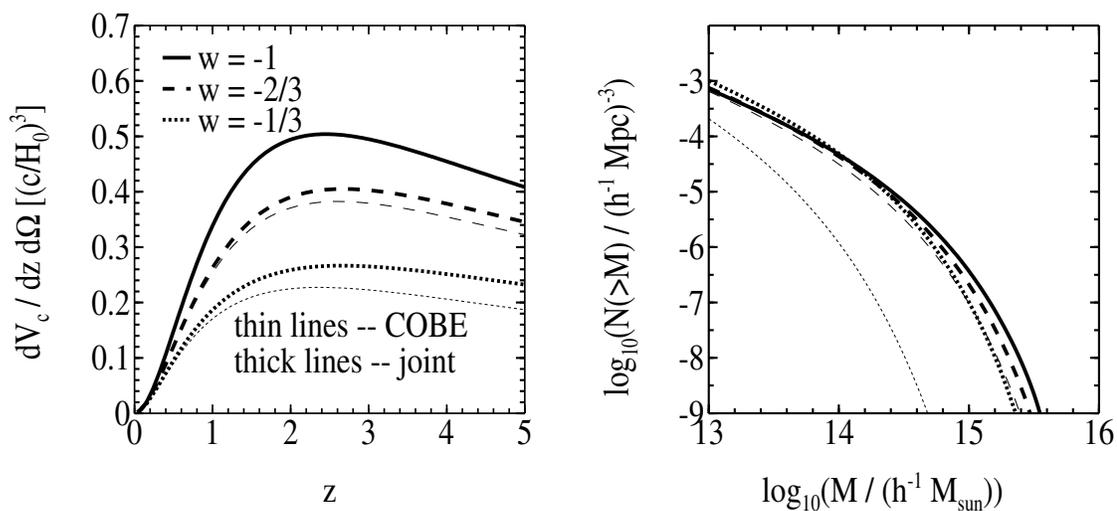


Figure 5.7 Comoving volume element as a function of redshift (*left* panel) and the comoving number density of virialized objects as a function of mass (*right* panel) for three constant- $w$  models. Results are shown for both the COBE normalization of  $\sigma_8$  with fixed  $\Omega_0 = 0.3$  (*thin* lines) and the joint cluster abundance–COBE normalization of  $\sigma_8$  with  $\Omega_0 = \Omega_0(w)$  (*thick* lines).

chical growth of structure. The fraction also increases with redshift since objects are collapsing and evolving toward virialization. It is also interesting to note that objects with  $\Delta_{\text{ta}} < \Delta < \Delta_{\text{vir}}$  evolve more rapidly as compared to objects with  $0 < \Delta < \Delta_{\text{ta}}$ . This is demonstrated by the fact that at  $z = 1$  the fraction of both types of objects is nearly the same though by  $z = 0$  there are more objects that have not reached turnaround. Furthermore, the larger  $w$  is, the greater the difference between the rates of evolution. These effects are a consequence of the suppression of structure growth in cosmologies with dark energy; namely, growth slows down earlier for larger  $w$  and those objects that are less overdense at a given redshift have greater difficulty overcoming the repulsive effects of the dark energy and collapsing. Finally, the plots show how strongly the fraction depends on the chosen  $\sigma_8$  normalization, with a significant variation with  $w$  for the COBE normalization and a fairly small variation for the cluster-abundance normalization. This, in turn, means that the degree to which the functions  $f_{\text{vir}}$  and  $f_{\text{dark}}$  vary with  $w$  is highly dependent on the assumed

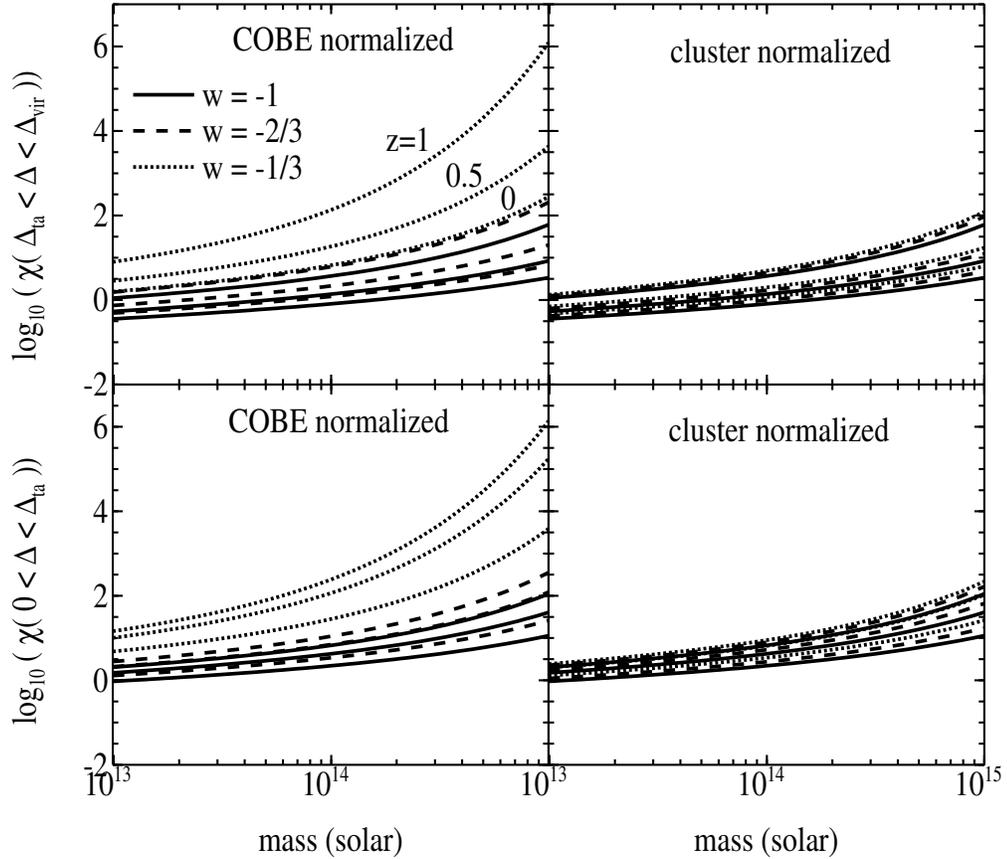


Figure 5.8 Fraction  $\chi$  of objects with overdensities in the range  $\Delta_{\text{ta}} < \Delta < \Delta_{\text{vir}}$  (*upper* panels) and  $0 < \Delta < \Delta_{\text{ta}}$  (*lower* panels) relative to those objects that are virialized,  $\Delta > \Delta_{\text{vir}}$ , as a function of mass. The *left* panels correspond to the COBE normalization of  $\sigma_8$  with  $\Omega_0 = 0.3$  and the *right* panels correspond to the X-ray cluster abundance normalization of  $\sigma_8$  with  $\Omega_0 = 0.3$ . For each constant- $w$  model we show the fraction  $\chi$  at  $z = 0$  (bottom curve),  $z = 1/2$  (middle curve), and  $z = 1$  (top curve).

normalization approach.

#### 5.4.2 Weak lens abundances

In Figure 9 we show the predicted redshift distribution of virialized lenses and dark lenses for three constant  $w$  models. For the COBE-normalized  $\sigma_8$  with fixed  $\Omega_0$  the distributions show a fairly strong sensitivity to  $w$ . As  $w$  increases from  $-1$  to  $-1/3$  the peak of the distributions shifts toward lower redshifts. Although one might expect

the trend to be in the opposite direction given that structures form faster for larger  $w$  models, the effect is counteracted by the decrease in  $\sigma_8$  with increasing  $w$ . That the decrease in  $\sigma_8$  so overwhelms any tendency for structure to form faster for  $w > -1$  is not surprising given the weak  $w$  dependence in the  $\Delta - \delta$  map (Figure 1) and in the function  $\delta_c(z)$  (Figure 4). Note, however, that the shift in the distributions with  $w$  becomes much less significant if a joint COBE-cluster abundance normalization is assumed. Finally, given that dark lenses are likely progenitors of virialized clusters, it is not surprising that both normalization approaches predict that the dark lenses have a larger mean redshift than the virialized lenses.

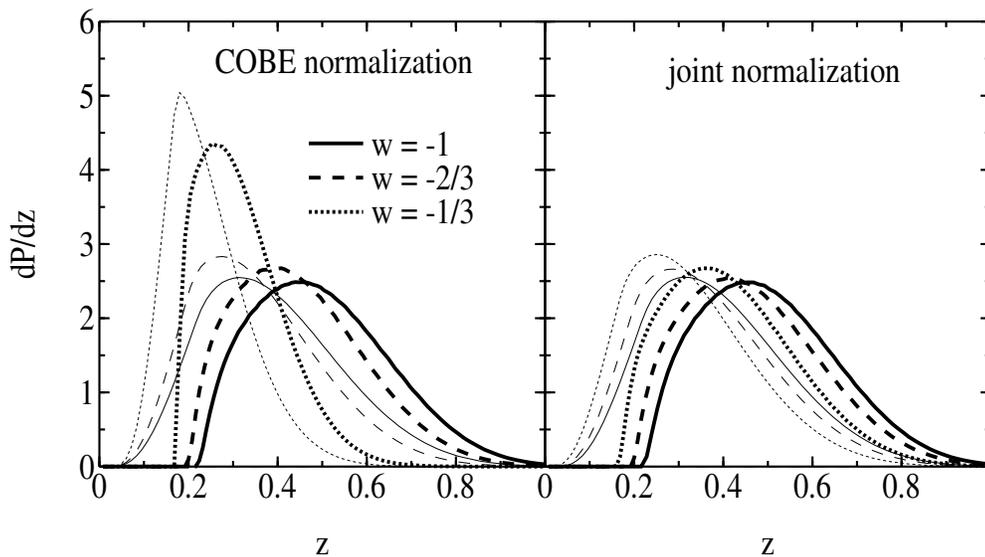


Figure 5.9 Normalized redshift distribution of virialized lenses (*thin* lines) and non-virialized lenses (*thick* lines) for three constant- $w$  models. The *left* panel shows results obtained when  $\sigma_8$  is normalized to COBE with  $\Omega_0 = 0.3$  and the *right* panel when  $\sigma_8$  is jointly normalized to COBE and the X-ray cluster abundance with  $\Omega_0 = \Omega_0(w)$ . The peak of the redshift distributions shifts toward lower redshifts as  $w$  increases because  $\sigma_8$  decreases with  $w$ . The shift in the peaks is less drastic, however, when the joint normalization is assumed.

To determine how well the weak lens redshift distributions can constrain  $w$  we generated mock redshift data and determined (using the Kolmogorov-Smirnov test) the probability of differentiating two different constant- $w$  models as a function of

the number of lenses detected. We found that to differentiate a  $\Lambda$ CDM model from both a  $w = -0.6$  model and a  $w = -0.9$  model at the  $3\sigma$  level required, on average, approximately 200 weak lenses and 2000 weak lenses, respectively. As we show below, this corresponds to a survey coverage of  $\sim 15$  and  $\sim 150$  square degrees. Note, however, that for sufficiently wide surveys systematic uncertainties such as mass-redshift selection effects and lens density profiles might dominate the errors.

By integrating over the redshift distribution we obtain the total number of virialized and dark lenses expected per square degree on the sky. As Figure 10 shows, the COBE normalization with  $\Omega_0 = 0.3$  shows a significant decline in the number count as  $w$  increases. By  $w = -2/3$  the number count of both virialized and dark lenses has dropped by a factor of two from the  $\Lambda$ CDM value. The joint normalization, in which we allow  $\Omega_0$  to vary with  $w$ , predicts a much more mild dependence on  $w$  with the number count dropping by only  $\sim 20\%$  from  $w = -1$  to  $w = -2/3$  for both lens types. Therefore, while the COBE-only normalization approach predicts that the sky coverage needed to distinguish the  $\Lambda$ CDM model from a  $w = -0.6$  model to  $3\sigma$  is only  $\sim 2$  degree<sup>2</sup>, the joint approach requires  $\sim 15$  degree<sup>2</sup>. Similarly, to distinguish the  $\Lambda$ CDM model from  $w = -0.9$  requires  $\sim 40$  degree<sup>2</sup> and  $\sim 100$  degree<sup>2</sup>, respectively. The systematic uncertainties affecting absolute sky density measurements, such as noise in the lensing maps and uncertainties in the lens density profiles, are expected to add further complications. This suggests that it will be very difficult to constrain  $w$  using just the number count of either virialized or dark lenses without, at the very least, a tighter constraint on  $\Omega_0$ .

We also note that our results do not agree with the results found by Bartelmann, Perrotta, & Baccigalupi (2002; hereafter BPB). They found that from  $w = -1$  to  $w \approx -0.6$ , the number of virialized weak lenses per square degree *increases* by nearly a factor of two. The increase is roughly linear up to the maximum after which the number count declines steeply. In obtaining these results, however, they use the formulas for  $\Delta_{\text{vir}}$  and  $\delta_c$  given in Lokas & Hoffman (2001), who assume that the space curvature within a collapsing overdensity patch is time-independent. As we showed in Section 3.1, this assumption is invalid for  $w \neq -1$  and leads to incorrect values for

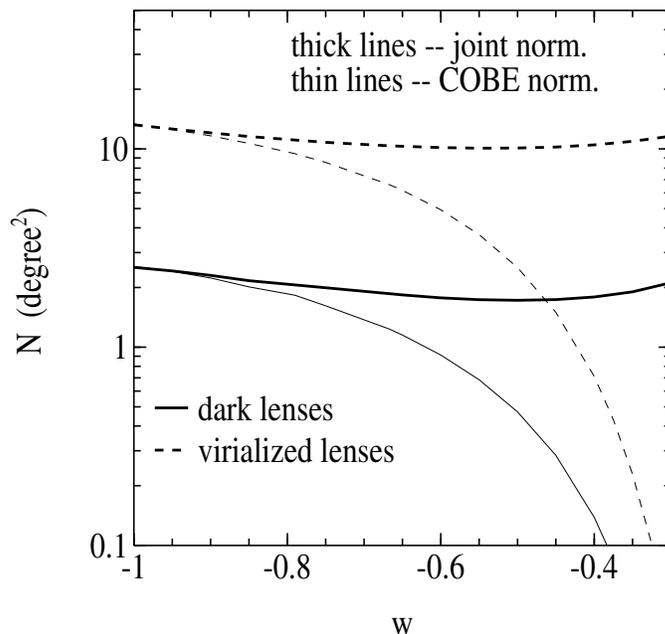


Figure 5.10 Total number of virialized lenses (*dashed* curves) and non-virialized lenses (*solid* curves) per square degree as a function of  $w$ . *Thin* lines correspond to the COBE normalized  $\sigma_8$  with  $\Omega_0 = 0.3$  and *thick* lines to the joint COBE–cluster abundance normalized  $\sigma_8$  with  $\Omega_0 = \Omega_0(w)$ . While the number count drops by a factor of two between  $w = -1$  and  $w = -2/3$  for the COBE-only normalization, the drop is much less significant for the joint normalization.

$\Delta_{\text{vir}}$  and  $\delta_c$ . To confirm that this is the source of our differences, we recomputed the number count of weak lenses as a function of  $w$  using the algorithm described in BPB (which differs somewhat from ours because we are interested in separating lenses into virialized and non-virialized types). When we assume the incorrect Lokas & Hoffman (2001) values for  $\Delta_{\text{vir}}$  and  $\delta_c$  we recover the results found by BPB; however, when we assume the values for  $\Delta_{\text{vir}}$  and  $\delta_c$  predicted by solving the spherical-collapse equations of Section 3.1, we obtain results very similar to those described in the preceding paragraphs.

### 5.4.3 Fraction of lenses that are dark

As mentioned above, the number-count ratio of dark to virialized lenses is an observable that is much less sensitive to observational noise than is the redshift distribution

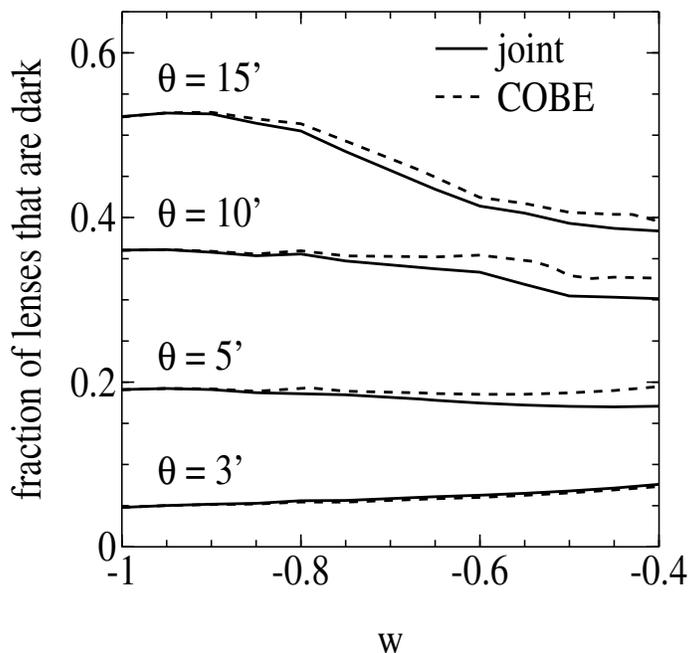


Figure 5.11 Fraction of lenses that are dark as a function of  $w$  for aperture sizes  $\theta = 3, 5, 10$  and  $15$  arcmin. Results are shown for both the COBE normalization with  $\Omega_0 = 0.3$  (*dashed* lines) and the joint COBE–cluster abundance normalization with  $\Omega_0 = \Omega_0(w)$  (*solid* lines). As  $\theta$  increases, the fraction of lenses that are dark rises significantly.

and number count of weak lenses. Unfortunately, for aperture sizes  $\theta$  (defined in Section 2) less than  $10'$  in radius the ratio is fairly constant over a broad range in  $w$ , as we show in Figure 11. The ratio varies more strongly if the aperture size is increased to  $15'$ . In particular, for  $\theta = 15'$  there is a  $\sim 20\%$  difference between the  $\Lambda$ CDM model and  $w = -0.6$ , so that differentiating the two models to a  $3\sigma$  significance requires the detection of  $\sim 600$  virialized lenses or equivalently a sky coverage of  $\sim 50$  degree<sup>2</sup>. Although using the non-virialized lens fraction requires large survey coverage for modest constraints on  $w$ , its principal advantage (in addition to being relatively insensitive to observational noise) is that it is not very sensitive to the chosen method of normalization; for any aperture size both the joint normalization and the COBE normalization with fixed  $\Omega_0$  yield similar dependences on  $w$ . Therefore, unlike the case for weak-lens sky-density or redshift distribution predictions, uncertainties in  $\sigma_8$  and  $\Omega_0$  do not strongly affect the predicted ratio of dark to virialized lenses. Inci-

dentially, although aperture sizes greater than  $\sim 15'$  yield ratios with even stronger  $w$  dependences, noise contributions from large-scale structure become significant at such large angular distances from the lens center (Hoekstra 2002). It is therefore not practical to make measurements at radii well beyond  $15'$ .

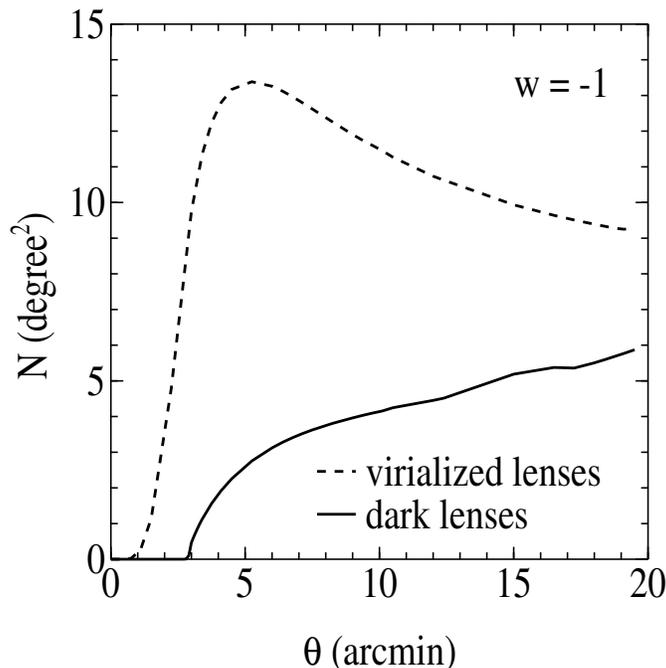


Figure 5.12 Total number of virialized lenses (*dashed* line) and non-virialized lenses (*solid* line) per square degree as a function of the aperture size  $\theta$  for a  $\Lambda$ CDM cosmology. While the number count of virialized lenses peaks at  $\theta = 5$  arcmin and declines thereafter, the number count of non-virialized lenses increases almost linearly for  $\theta > 5$  arcmin.

As an aside, while the ratio of dark to virialized lenses does not have a particularly strong  $w$  dependence, it does have a strong  $\theta$  dependence; only  $\sim 5\%$  of lenses are dark when  $\theta = 3'$  but  $\sim 50\%$  are when  $\theta = 15'$ . In Figure 12 we plot the number of virialized and dark lenses as a function of  $\theta$  for the  $\Lambda$ CDM model. As  $\theta$  increases from  $3'$  to  $15'$  the sky density of dark lenses increases from zero to five per square degree while the sky density of virialized lenses peaks at  $\theta = 5'$  and gradually declines for larger aperture sizes. Figure 13 explains this trend. For an overdensity of mass  $M = 5 \times 10^{14} M_{\odot}$  we plot, as a function of redshift,  $\theta_{\text{vir}}$ , the projected angular size of the virialization radius, and  $\theta_{\text{max}}$ , the projected angular size of the maximum

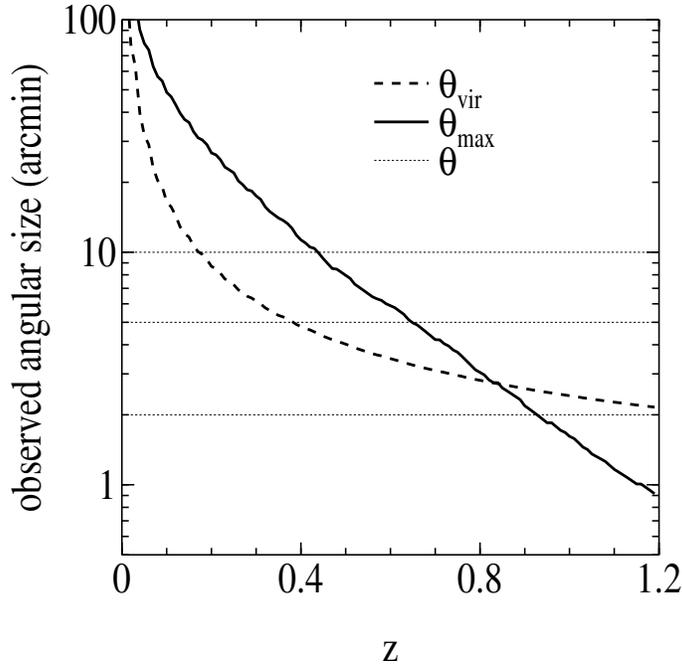


Figure 5.13 Observed angular size of  $\theta_{\text{vir}}$  (*dashed* line) and  $\theta_{\text{max}}$  (*solid* line) as a function of redshift for an overdensity of mass  $M = 5 \times 10^{14} M_{\odot}$  for a  $\Lambda$ CDM cosmology. If  $\theta_{\text{max}} > \theta_{\text{vir}}$  an overdensity need not be virialized to produce a detectable lensing signal. However, the range in redshift over which a non-virialized lens can be detected is limited by the aperture size  $\theta$  (e.g., *thin, dotted* lines), which defines the maximum observable angular scale. For  $\theta \lesssim 3$  arcmin virtually no dark lenses can be detected.

radius that produces a detectable lens (i.e.,  $\theta_{\text{max}} = R_{\text{max}}(z)/D_{\text{d}}(z)$  where  $R_{\text{max}}^3 = 3M/4\pi\Delta_{\text{min}}(z)$ ). For  $\theta_{\text{max}} > \theta_{\text{vir}}$  an overdensity can be non-virialized and still produce a detectable lensing signal (i.e., a dark lens). However, since  $\theta$  defines the maximum observable angular scale, for sufficiently small  $\theta$  there is no range in redshift such that  $\theta > \theta_{\text{max}} > \theta_{\text{vir}}$ , in which case non-virialized overdensities cannot produce a detectable lens. In general we find that the minimum aperture size needed to detect dark lenses is  $\sim 3'$ . For larger  $\theta$ , the area below  $\theta_{\text{max}}$  and above  $\theta_{\text{vir}}$  has a substantial relative increase while the area below  $\theta_{\text{vir}}$  has just a mild relative increase. After taking into account the fact that the aperture mass  $M_{\text{ap}}(\theta)$  decreases with increased  $\theta$ , this translates to an increase in the sky density of dark lenses and a decrease in the sky density of virialized lenses for  $\theta > 5'$ . The fraction of lenses that are dark therefore increases with aperture size.

## 5.5 Discussion and conclusions

We have examined the possibility of using the measured abundance of weak gravitational lenses to constrain a principal property of the dark energy, its equation-of-state parameter  $w$ . Since dark energy modifies both the background cosmology of the universe and the growth of structure, it will necessarily have an effect on the efficiency of weak-lensing. The goal of this paper was to determine the nature and strength of the effect.

The change in the background cosmology with  $w$  influences the predicted weak lens abundance in essentially three ways. First, the size of comoving volume elements shrink with increasing  $w$ . Second, the distance-redshift relation is modified, thereby shifting the location of the lensing-kernel maximum (i.e., where the combination of angular diameter distances  $D_{ds}D_d/D_s$  peaks). Third, since the evolution of the background matter density is modified by the dark energy, the density of a given halo relative to the background density changes with  $w$ . This, in turn, affects the strength of a halo's lensing signal; the larger the overdensity the stronger the signal. While the volume term is explicitly factored into the expression for the weak-lens sky density [equation (22)], the latter two effects are incorporated into the signal-to-noise estimator for which we use the aperture-mass technique introduced by Schneider (1996).

The change in the growth of structure with  $w$  is somewhat more subtle. The dark energy modifies both the rate of structure growth and the amplitude of the matter power spectrum. To determine the former we solved the spherical-collapse model with dark energy included. Though growth occurs more rapidly as  $w$  increases, the overall effect on the  $\Delta - \delta$  map, needed to relate the minimum overdensity required to produce a detectable lens,  $\Delta_{\min}$ , to a corresponding linear-theory overdensity  $\delta_{\min}$ , is fairly small. Similarly, the linear-theory overdensity at collapse  $\delta_c$  does not vary much with  $w$ . The effect on  $\Delta_{\text{vir}}$  is more significant, however. As  $w$  increases, structures require substantially greater overdensities in order to reach virial equilibrium because they collapse sooner, when the universe is younger and hotter.

To determine how the power-spectrum amplitude,  $\sigma_8$ , varies with  $w$  we considered

three possible approaches. One was to normalize to the X-ray cluster abundance as was done in WS98. Another was to normalize to the COBE measurements of CMB anisotropies on large angular scales. These two approaches predict similar values of  $\sigma_8$  for the  $\Lambda$ CDM model. However, if all cosmological parameters are held fixed as  $w$  varies, the values of  $\sigma_8$  are no longer in accordance. This is because the cluster abundance approach is accounting for the earlier forming, and hence hotter, galaxy clusters in models with  $w > -1$ . The COBE normalization, on the other hand, is accounting for the increase in the Integrated Sachs-Wolfe (ISW) effect as  $w$  increases (cf. BPB). Given these differing influences, the two approaches are not expected to yield the same  $\sigma_8$  when all the cosmological parameters are held fixed to those of the  $\Lambda$ CDM model while  $w$  is varied. This suggests a third approach to normalizing the power spectrum; namely, let the parameters vary with  $w$  such that the cluster abundance normalization matches the COBE normalization. In practice we accomplished this by letting just  $\Omega_0$  vary with  $w$ , as it is the parameter most degenerate with  $\sigma_8$ . The resulting range in  $\Omega_0$  for  $-1 < w \lesssim -0.4$  was found to be  $0.3 < \Omega_0 < 0.4$  and hence within observational uncertainties. Though all three normalization approaches predict that  $\sigma_8$  decreases with  $w$ , the difference in the magnitude of the decrease between the approaches is significant. As a result, each predicts substantially different variations in the weak-lens abundance with  $w$ .

Having determined all the dark energy effects, we computed the redshift distribution and sky density of weak lenses as a function of  $w$ . As in WK02, we distinguished between two classes of lenses, those that have collapsed and virialized and those that have not. This distinction is based on the expectation that the virialized lenses, being in a relaxed state, are X-ray and/or optically luminous. The non-virialized lenses, being at an earlier stage in the overdensity evolutionary cycle, are expected to be X-ray underluminous because the observed X-ray luminosity function has a steep dependence on the total virialized mass within a halo. Furthermore, though the typical mass of both lens types is  $\sim few \times 10^{14} M_\odot$ , the sky density of galaxies within the non-virialized lenses is expected to be smaller than in the virialized lenses because they have not yet collapsed and hence have larger radii (see WK02 for more details).

We found that the variation in the redshift distribution and the sky density of both lens types with  $w$  depends strongly on the power-spectrum-normalization approach. If  $\Omega_0$  is fixed and  $\sigma_8$  is normalized to the COBE measurements, there is a significant variation in the abundances with  $w$ . In particular, the sky density of both virialized lenses and non-virialized lenses drops by a factor of two from  $w = -1$  to  $w = -2/3$ . This decline, a result of the significant decrease in  $\sigma_8$  with  $w$ , occurs despite the faster formation of structure for  $w > -1$ . If, on the other hand,  $\Omega_0$  is allowed to vary with  $w$  such that the COBE normalization matches the cluster-abundance normalization, the redshift distributions and sky density change very little with  $w$ ; between  $w = -1$  and  $w = -2/3$  the sky density of both lens types varies by just  $\sim 20\%$ . This insubstantial variation is the result of an increase in  $\Omega_0$  with  $w$  and a less significant drop in  $\sigma_8$  with  $w$  as compared to the COBE normalization with  $\Omega_0$  fixed. Obtaining a strong constraint on  $w$  from the sky density or redshift distribution of weak lenses therefore appears to be contingent on improved measurements of  $\Omega_0$  from independent observations.

Perhaps more promising is the possibility of utilizing the observed ratio of dark lenses to virialized lenses. Unlike measurements of the absolute sky density of weak lenses, the ratio is not very sensitive to the amount of observational noise in the weak-lensing maps since the abundance of both dark lenses and virialized lenses are equally affected by noise. Similarly, the ratio does not vary significantly over a wide range of cosmological parameters so that uncertainties due to the  $\Omega_0 - w$  degeneracy are minimized. We found that for aperture sizes of  $\sim 15'$  the ratio varies by about 20%, dropping from 0.5 to 0.4, between the  $\Lambda$ CDM model and  $w = -0.6$ . We also showed that the ratio of dark to virialized lenses increases with aperture size, in effect because larger apertures enable the detection of the more extended radii of the non-virialized lenses.

Weak-lensing has already been shown to be a powerful probe of the matter distribution in the universe (see e.g., Bartelmann & Schneider 2001). It also has the potential to help constrain the amount and nature of the dark energy. Huterer (2002) showed that given reasonable prior information on other cosmological parameters, the

weak-lensing convergence power spectrum can impose constraints on the dark energy comparable to those of upcoming type Ia supernova and number-count surveys of galaxies and galaxy clusters. Constraining the dark energy from absolute measurements of weak-lens abundances will likely prove difficult, however. The variation in the weak-lens sky density with  $w$  is sufficiently small that modest uncertainties in  $\Omega_0$  (and observational noise) can mask the effect of the dark energy. More auspicious is the possibility of utilizing the relative abundance of dark lenses to virialized lenses to constrain  $w$ . Future weak-lensing projects such as the the VISTA survey, the SNAP mission, and LSST (see Tyson et al. 2002 for a discussion of its great promise as a probe of dark energy) are expected to provide the wide-field surveys needed for this technique to be viable.

We thank R. Caldwell for helpful suggestions, and an anonymous referee for useful comments that have improved the presentation of this paper. NNW acknowledges the support of an NSF Graduate Fellowship. This work was supported by NSF AST-0096023, NASA NAG5-9821, and DoE DE-FG03-92-ER40701.

## 5.A Phantom energy: dark energy with $w < -1$ causes a cosmic doomsday<sup>†</sup>

The simplest explanation for dark energy is a cosmological constant, for which  $w = -1$ . However, this cosmological constant is 120 orders of magnitude smaller than expected from quantum gravity. Thus, although we can add this term to Einstein's equation, it is really only a placeholder until a better understanding of this negative pressure arises. Another widely explored possibility is quintessence (Caldwell, Dave, & Steinhardt, 1998; Ratra & Peebles, 1988; Wetterich, 1995; Coble, Dodelson, & Frieman, 1997; Turner & White, 1997; Boyle, Caldwell, & Kamionkowski, 2002), a cosmic

---

<sup>†</sup>A version of this section was first published as R. Caldwell, M. Kamionkowski, and N. Weinberg *Physical Review Letters*, 91, 071301 (2003).

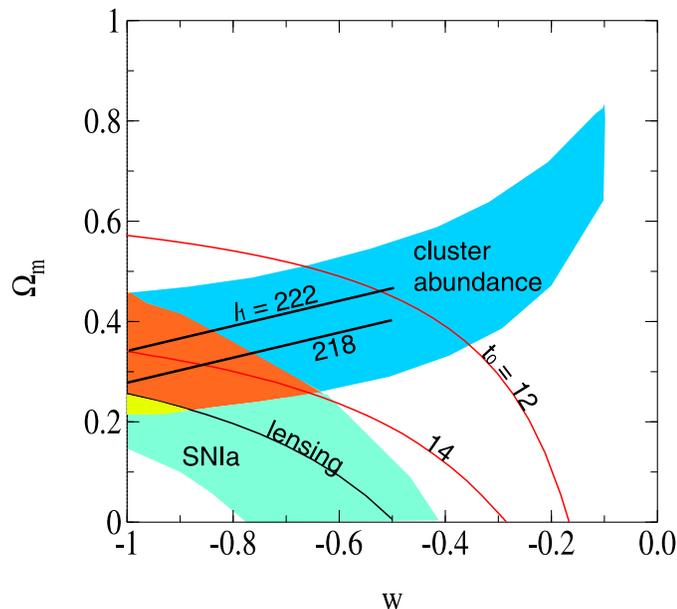


Figure 5.14 Current constraints to the  $w$ - $\Omega_m$  parameter space. The red solid curves show the age (in Gyr) of the Universe today (assuming a Hubble parameter  $H_0 = 70$  km sec $^{-1}$  Mpc $^{-1}$ ). The light shaded regions are those allowed (at  $2\sigma$ ) by the observed cluster abundance and by current supernova measurements of the expansion history. The dark orange shaded region shows the intersection of the cluster-abundance and supernova curves, additionally restricted (at  $2\sigma$ ) by the location of the first acoustic peak in the cosmic-microwave-background power spectrum and quasar-lensing statistics.

scalar field that is displaced from, but slowly rolling to, the minimum of its potential. In such models, the equation-of-state parameter is  $-1 < w < -1/3$ , and the dark-energy density decreases with scale factor  $a(t)$  as  $\rho_Q \propto a^{-3(1+w)}$ .

Fig. 5.14 shows constraints to the  $w$ - $\Omega_m$  parameter space (where  $\Omega_m$  is the pressureless matter density in units of the critical density) from the cluster abundance, supernovae, quasar-lensing statistics (see refs. Wang et al. 2000; Perlmutter, Turner, & White 1999 and references therein), and the first acoustic peak in the CMB power spectrum (values taken from Ref. Page et al. 2003). As the figure shows,  $w$  seems to be converging to  $w = -1$ .

But what about  $w < -1$ ? Might the convergence to  $w = -1$  actually be indicating that  $w < -1$ ? Why restrict our attention exclusively to  $w \geq -1$ ? Matter with  $w < -1$ , dubbed “phantom energy” (Caldwell, 2002), has received increased

attention among theorists recently. It certainly has some strange properties. For example, the energy density of phantom energy increases with time. It also violates the dominant-energy condition (Hawking & Ellis, 1973; Carroll, Hoffman, & Trodden, 2003a), a cherished notion that helps prohibit time machines and wormholes. However, it is hard to see how time machines and wormholes would arise with phantom energy. Although sound waves in quintessence travel at the speed of light, it does not automatically follow that disturbances in phantom energy must propagate faster than the speed of light; in fact, there are already several scalar-field models for phantom energy in which the sound speed is subluminal (Caldwell, 2002; Parker & Raval, 2001; Armendariz-Picon et al., 1999; Chiba, Okabe, & Yamaguchi, 2000; Faraoni, 2002; Carroll, Hoffman, & Trodden, 2003b). It is true that these models feature unusual kinetic terms in their Lagrangians, but such terms may arise in supergravity (Nilles, 1984) or higher-derivative-gravity theories (Pollock, 1988). Theorists have also discussed stringy phantom energy (Frampton, 2002) and brane-world phantom energy (Sahni & Shtanov, 2002). Connections with the dS/CFT correspondence have also been made (McInnes, 2002). To be sure, phantom energy is not something that any theorist would have expected; on the other hand, not too many more theorists anticipated a cosmological constant! Given the limitations of our theoretical understanding, it is certainly reasonable to ask what empirical results have to say.

In Fig. 5.15 we generalize the analysis of cosmological constraints to a parameter space that extends to  $w < -1$ . As indicated here, there is much acceptable parameter space in regions with  $w < -1$  (see also refs. Hannestad 2002; Schuecker et al. 2003). With certain prior assumptions, the best fit is actually at  $w < -1$ .

As we now show, if  $w < -1$  persists, then the fate of the Universe is quite fantastic and completely different than the possibilities previously discussed. To begin, let us review these other fates. In a flat or open Universe *without* dark energy, the expansion continues forever, and the horizon grows more rapidly than the scale factor; the Universe becomes colder and darker, but with time the comoving volume of the observable Universe evolves so that the number of visible galaxies grows. If the expansion is accelerating, as a consequence of dark energy with  $-1 \leq w < -1/3$ ,

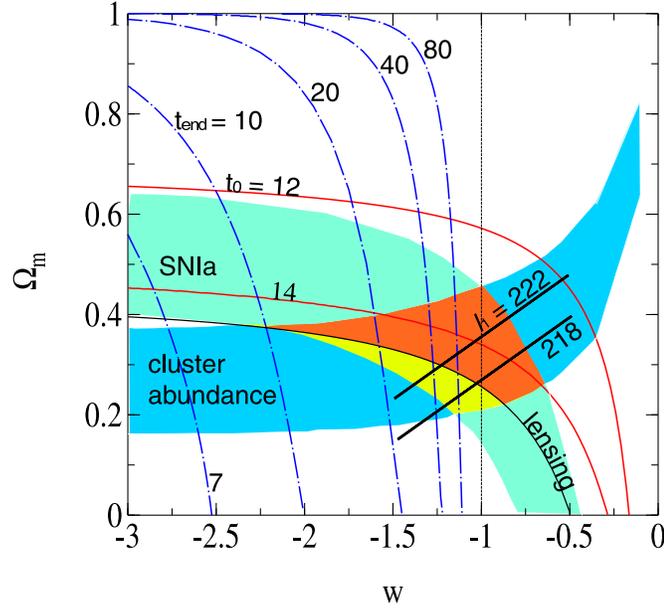


Figure 5.15 Same as in Fig. 5.14, except extended to  $w < -1$ . Here, the blue dot-dash curves show for phantom-energy ( $w < -1$ ) models the time (in Gyr) remaining in the Universe (assuming a Hubble parameter  $H_0 = 70 \text{ km sec}^{-1} \text{ Mpc}^{-1}$ ).

then the expansion again continues forever. However, in this case, the scale factor grows more rapidly than the horizon. As time progresses, galaxies disappear beyond the horizon, and the Universe becomes increasingly dark. Still, structures that are currently gravitationally bound, such as the Milky Way and perhaps the Local Group, remain unaffected. Thus, although extragalactic astronomy becomes less interesting, Galactic astronomy can continue to thrive.<sup>1</sup>

With phantom energy, the Friedmann equation governing the time  $t$  evolution of the scale factor  $a(t)$  becomes  $H^2 \equiv (\dot{a}/a)^2 = H_0^2[\Omega_m/a^3 + (1 - \Omega_m)a^{-3(1+w)}]$ , where  $H_0$  is the Hubble parameter, and the dot denotes a time derivative. If  $\Omega_m \simeq 0.3$ , then the Universe is already dark energy-dominated, and for  $w < -1$  it will become increasingly dark-energy-dominated in the future. We thus approximate the subsequent evolution of the scale factor by neglecting the first term on the right-hand side. Doing so, we find that the scale factor blows up in a time  $t_{rip} - t_0 \simeq$

<sup>1</sup>There is another possibility: if the quintessence potential at some point becomes negative, then the Universe can reach a point of maximum expansion and then re-collapse (Steinhardt & Turok, 2002; Kallosh, 2002).

$(2/3)|1+w|^{-1}H_0^{-1}(1-\Omega_m)^{-1/2}$  from the current time  $t_0$ . For example, for  $w = -3/2$  and  $H_0 = 70 \text{ km sec}^{-1} \text{ Mpc}^{-1}$ , the time remaining before the Universe ends in this “Big Rip” (McInnes, 2002) is 22 Gyr.

As in a cosmological-constant Universe, the scale factor grows more rapidly than the Hubble distance  $H^{-1}$  and galaxies will begin to disappear beyond the horizon. With phantom energy, the expansion rate  $H$  grows with time, the Hubble distance decreases, and so the disappearance of galaxies is accelerated as the horizon closes in on us. More intriguing is that the increase in the dark-energy density will ultimately begin to strip apart gravitationally bound objects. According to general relativity, the source for the gravitational potential is the volume integral of  $\rho + 3p$ . So, for example, a planet in an orbit of radius  $R$  around a star of mass  $M$  will become unbound roughly when  $-(4\pi/3)(\rho + 3p)R^3 \simeq M$ . With  $w \geq -1$ ,  $-(\rho + 3p)$  is decreasing with time so if  $-(4\pi/3)(\rho + 3p)R^3$  is smaller than  $M$  today, then it will remain so ever after. Thus, any system that is currently gravitationally bound (e.g., the solar system, the Milky Way, the Local Group, galaxy clusters) will hereafter remain so.

With phantom energy,  $-(\rho + 3p)$  increases, and so at some point in time every gravitationally bound system will be dissociated. With the time evolution of the scale factor and the scaling of the phantom-energy density with time, we find that a gravitationally bound system of mass  $M$  and radius  $R$  will be stripped at a time  $t \simeq P\sqrt{2|1+3w|}/[6\pi|1+w|]$ , where  $P$  is the period of a circular orbit around the system at radius  $R$ , before the Big Rip (see Table 5.1). Interestingly, this time is independent of  $H_0$  and  $\Omega_m$ .

Thus, for example, for  $w = -3/2$ , the interval is  $t \simeq 0.3P$  before the end of time. In this case, clusters will be stripped roughly a billion years before the end of time. In principle, if  $w$  were sufficiently negative, the Andromeda galaxy would be torn from the Local Group before it could fall into the Milky Way; however, given current upper limits to  $-w$ , this is unlikely. For  $w = -3/2$ , the Milky Way will get stripped roughly 60 million years before the Big Rip. Curiously, when this occurs the horizon will still be  $\sim 70 \text{ Mpc}$ , so there may still be other observable galaxies that we will also see stripped apart (although given the time delay from distant objects,

Table 5.1 The history and future of the Universe with  $w = -3/2$  phantom energy.

Time	Event
$\sim 10^{-43}$ s	Planck era
$\sim 10^{-36}$ s	Inflation
First Three Minutes	Light Elements Formed
$\sim 10^5$ yr	Atoms Formed
$\sim 1$ Gyr	First Galaxies Formed
$\sim 15$ Gyr	<i>Today</i>
$t_{rip} - 1$ Gyr	Erase Galaxy Clusters
$t_{rip} - 60$ Myr	Destroy Milky Way
$t_{rip} - 3$ months	Unbind Solar System
$t_{rip} - 30$ minutes	Earth Explodes
$t_{rip} - 10^{-19}$ s	Dissociate Atoms
$t_{rip} = 35$ Gyrs	Big Rip

we will see the Milky Way destroyed first). A few months before the end of time, the earth will be ripped from the Sun, and  $\sim 30$  minutes before the end the earth will fall apart. Similar arguments also apply to objects bound by electromagnetic or strong forces. Thus, molecules and then atoms will be torn apart roughly  $10^{-19}$  seconds before the end, and then nuclei and nucleons will get dissociated in the remaining interval. In all likelihood, some new physics (e.g., spontaneous particle production or extra-dimensional, string, and/or quantum-gravity effects) may kick in before the ultimate singularity, but probably after the sequence of events outlined above.

The end of structure, from cosmic, macroscopic scales down to the microscopic, leads us to remark that our present epoch is unique from the viewpoint that at no other time are non-linear structures possible. When the phantom energy becomes strong enough, gravitational instability no longer works and the Universe becomes homogeneous. Eventually, individual particles become isolated: points separated by a distance greater than  $3\delta t(1+w)/(1+3w)$  at a time  $t_{rip} - \delta t$  cannot communicate before the Big Rip. Therefore, the dominance of the phantom energy signals the end of our brief era of cosmic structure, which began when the non-relativistic matter emerged from the radiation. In such a Universe, certain cosmic questions have new significance. It is natural to find ourselves—or more generally, non-linear structure—

living close to the onset of acceleration if the structure is soon destroyed and the Universe does not survive much longer afterwards (McInnes, 2002). A Big Rip renders the “why now?”, or question of cosmic coincidence, irrelevant.

The current data indicate that our Universe is poised somewhere near the razor-thin separation between phantom energy, cosmological constant, and quintessence. Future work, and the longer observations by WMAP, will help to determine the nature of the dark energy. In the meantime we are intrigued to learn of this possible new cosmic fate that differs so remarkably from the re-collapse or endless cooling considered before. It will be necessary to modify the adopted slogan among cosmic futurologists—“*Some say the world will end in fire, Some say in ice*” (Frost, 1916)—for a new fate may await our world.

RRC thanks the UCSB KITP for hospitality. This work was supported at Caltech by NASA NAG5-9821 and DoE DE-FG03-92-ER40701, at the KITP by NSF PHY99-07949, and at Dartmouth by NSF grant PHY-0099543. NNW was supported by a NSF graduate fellowship.

# Bibliography

- Armendariz-Picon, C., Damour, T. & Mukhanov, V., *Phys. Lett. B* **458**, 209–218 (1999).
- Baccigalupi C., Balbi A., Matarrese S., Perrotta F., Vittorio N., 2002, *Phys. Rev. D*, 65, 063520
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
- Barger V., Marfatia D., 2001, *Phys. Lett. B*, 498, 67
- Bartelmann M., 1995, *A&A*, 303, 643
- Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
- Bartelmann M., Perrotta F., Baccigalupi C., 2002, submitted to *A&A*, astro-ph/0206507 (BPB)
- Bean R., Melchiorri A., 2002, *Phys. Rev. D*, 65, 041302
- Benoit, A. et al., astro-ph/0210306 (2002).
- Boyle, L., Caldwell, R. R. & Kamionkowski, M., *Phys. Lett. B* **545**, 17–22 (2002).
- Brainerd T. G., Blandford R. D., Smail I., 1996, *ApJ*, 466, 623
- Bunn E. F., White M., 1997, *ApJ*, 480, 6
- Caldwell R. R., Dave R., Steinhardt P. J., 1998, *Phys. Rev. Lett.*, 80, 1582
- Caldwell, R.R., *Phys. Lett. B* **545**, 23–29 (2002); astro-ph/9908168.

- Carroll, S. M., Hoffman, M. & Trodden, M., astro-ph/0302067 (2003a).
- Carroll, S. M., Hoffman, M. & Trodden, M., astro-ph/0301273 (2003b).
- Chiba, T., Okabe, T. & Yamaguchi, M., *Phys. Rev. D* **62**, 023511 (2000).
- Coble, K., Dodelson, S. & Frieman, J., *Phys. Rev. D* **55**, 1851–1859 (1997).
- Cohen J. G., Hogg D. W., Blandford R. D. et al. 2000, ApJ, 538, 29
- Dahle H., Pedersen K., Lilje P. B., Maddox S. J., Kaiser N., 2002, submitted to ApJ, astro-ph/0208050
- Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659
- de Bernardis, P. *et al.*, *Nature* **404**, 955–959 (2000).
- de Bernardis P. et al., 2002, ApJ, 564, 559
- Erben T., van Waerbeke L., Mellier Y., Schneider P., Cuillandre J.-C., Castander F. J. Dantel-Fort M.,(2000, A&A, 355, 23
- Faraoni, V., *Int. J. Mod. Phys. D* **11**, 471–481 (2002).
- Frampton, P., astro-ph/0209037 (2002).
- Frost, R. “Fire and Ice” (1916), from *Collected Poems of Robert Frost* (Holt & Co., 1930).
- Halverson N. W. et al., 2002, ApJ, 568, 38
- Hanany, S. *et al.*, *Astrophys. J. Lett.* **545**, L5–L9 (2000).
- Hannestad, S. & Mortsell, E., *Phys. Rev. D* **66**, 063508 (2002).
- Hawking, S. W. & Ellis, G. F. R. *The large scale structure of space-time* (Cambridge University Press, Cambridge, 1973).
- Hernquist L., 1990, ApJ, 356, 359

- Hoekstra H., 2002, submitted to MNRAS, astro-ph/0208351
- Hu W., Sugiyama N., 1996, ApJ, 471, 542
- Huterer D., Turner M. S., 2001 Phys. Rev. D, 64, 123527
- Huterer D., 2002, Phys. Rev. D., 65, 063001
- Kallosh, R. *et al.*, *Phys. Rev. D* **66**, 123503 (2002).
- Kitayama T., Suto Y., 1996, ApJ, 469, 480
- Koopmans L. V. E. et al. (the CLASS collaboration), 2000, A&A, 361, 815
- Kruse G., Schneider P. 1999, MNRAS, 302, 821
- Lee A. T. et al., 2001, ApJ, 561, L1
- Lokas E. L., Hoffman Y., 2001, astro-ph/0108283, v1
- Ma C.-P., Caldwell R. R., Bode P., Wang L., 1999, ApJ, 521, L1
- Maor I., Brustein R., Steinhardt P. J., 2001, Phys. Rev. Lett., 86, 6
- Mason, B. S. *et al.*, astro-ph/0205384 (2002).
- McInnes, B., *JHEP* **0208**, 029 (2002).
- Mellier Y., 1999, ARAA, 37, 127
- Miller A. D. et al. , 1999, ApJ, 524, L1
- Miralles J. M., Erben T., Haemmerle H., Schneider P., Fosbury R. A. E., Freudling W., Pirzkal N., Jain B., White S. D. M., 2002, astro-ph/0202122
- Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493 (NFW)
- Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563 (NFW)
- Nilles, H.-P., *Phys. Rep.* **110**, 1–162 (1984).

- Page, L. *et al.*, astro-ph/0302220 (2003).
- Parker, L. & Raval, A., *Phys. Rev. Lett.* **86**, 749 (2001).
- Peacock J. A., *Nature*, 2001, 410, 169
- Percival W. J. et al., 2001, *MNRAS*, 327, 1297
- Perlmutter S. et al., 1999 *ApJ*, 517, 565
- Perlmutter, S., Turner, M. S. & White, M., *Phys. Rev. Lett.* **83**, 670–673 (1999).
- Pollock, M. D., *Phys. Lett. B* **215**, 635–641 (1988).
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Ratra, B. & Peebles, P. J. E., *Phys. Rev. D* **37**, 3406–3427 (1988)
- Riess A. G. et al., 1998, *AJ*, 116, 1009
- Sahni, V. & Shtanov, Y., astro-ph/0202346 (2002).
- Schneider P., 1996, *MNRAS*, 283, 837
- Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, *MNRAS*, 296, 873
- Schuecker, P., Caldwell, R. R., Böhringer, H., Collins, C. A., Guzzo, L., & Weinberg, N. N. 2003, *A&A*, 402, 53
- Seitz C., Schneider P., 1997, *A&A*, 318, 687
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- Sievers J. L. et al., 2002, submitted to *ApJ*, astro-ph/0205387
- Smail I., Hogg D. W., Yan L., Cohen J. G., 1995, *ApJ*, 449, L105
- Spergel, D. N. *et al.*, astro-ph/0302209 (2003).
- Steinhardt, P. J. & Turok, N., *Science* **296**, 1436–1439 (2002).

- Turner, M. S. & White, M., *Phys. Rev. D* **56**, 4439–4443 (1997).
- Tyson J. A., Wittman D. M., Hennawi J. F., Spergel D. N., 2002, astro-ph/0209632
- Umetsu K., Futamase T., 2000, *ApJ*, 539, 5
- Wang L., Steinhardt P. J., 1998, *ApJ*, 508, 483 (WS98)
- Wang L., Caldwell R. R., Ostriker J. P., Steinhardt P. J., 2000, *ApJ*, 530, 17
- Wang, L. *et al.*, *Astrophys. J.* **530**, 17–35 (2000).
- Weinberg N. N., Kamionkowski M., 2002, *MNRAS*, in press, astro-ph/0203061 (WK02)
- Weller J., Albrecht A., 2001, *Phys. Rev. Lett.*, 82, 896
- Wetterich, C., *Astron. Astrophys.* **301**, 321 (1995).
- White M., van Waerbeke L., Mackey J., 2002, *ApJ*, 575, 640