Math. Res. Lett. **13** (2006), *no.* 4, 557–570

A PROOF OF A THEOREM OF LUTTINGER AND SIMPSON ABOUT THE NUMBER OF VANISHING CIRCLES OF A NEAR-SYMPLECTIC FORM ON A 4-DIMENSIONAL MANIFOLD

Clifford Henry Taubes

Abstract. A proof is given of a theorem announced some years ago by Luttinger and Simpson to the effect that a compact 4-manifold that has a near-symplectic form in a given cohomology class admits one in the same class whose zero locus consists of any given, but strictly positive number of disjoint, embedded circles.

1. Introduction

A symplectic form on a smooth, oriented 4-manifold is a closed, 2-form whose square is nowhere vanishing and positive. A near symplectic form is a non-trivial, closed form whose square is non-negative and, on its zero locus, has rank 3 derivative. A number of years ago, Karl Luttinger and Carlos Simpson announced the following theorem:

Theorem 1. *Let n denote a positive integer. A smooth, oriented compact, connected 4-manifold that admits a near symplectic form in a given cohomology class admits one in the same class whose zero locus consists of the disjoint union of n embedded circles.*

No proof has been published. Having been asked at times about this theorem, and having quoted it on various occasions (see, e.g., [T1]), the author set about the task of providing a proof. This paper contains the author's proof of this theorem of Luttinger and Simpson. After writing this paper, the author learned that Tim Perutz [P] has recently proved the Luttinger-Simpson theorem along somewhat different lines.

To give some context to this theorem, note first that Hodge theory can be used to construct a non-trivial 2-form with non-negative square on any 4-manifold with postive self-dual, 2nd Betti number. Meanwhile, a folk theorem known to gauge theory afficianados from the work of Simon Donaldson in the 1980's (see [DK]) asserted that such a 2-form can be found with rank 3 derivative on its zero locus. Thus, its zero locus consists of some number of disjoint, embedded circles. A proof of this folk theorem was published by Honda [Ho].

By definition, a near symplectic form with no vanishing locus is symplectic. However, there are compact 4-manifolds with near symplectic forms but no symplectic ones [T2], [T3], [K]. Thus, any near symplectic form on such a manifold must have at least one component circle to its zero locus.

Received by the editors November 14, 2005.
Supported in part by the National Science Foundation.

Theorem 1 has an analog for non-compact, but asymptotically Euclidean 4-manifolds. In this regard, a manifold is said to be asymptotically Euclidean when the complement of a compact set is diffeomorphic to the complement in $\mathbb{R}^4$ of a ball. A closed 2-form on such a manifold is deemed asymptotically standard when such a diffeomorphism pulls it back as $dx^1 \wedge dx^2 + dx^3 \wedge dx^4$. Hodge theory with arguments like those used in [Ho] can be used to prove that every asymptotically Euclidean 4-manifold has an asymptotically standard, exact near-symplectic form.

Theorem 2. *Let n be a positive integer. Every smooth, oriented, asymptotically Euclidean 4-manifold has an asymptotically standard, exact, near symplectic form whose zero locus consists of the disjoint union of n embedded circles.*

A celebrated theorem of Gromov [G] asserts that $\mathbb{R}^4$ is the only asymptotically Euclidean 4-manifold with an asymptotically standard symplectic form.

By the way, recent works of Kirby and Gay [KG] for the compact manifold case and Scott [S] for the case of an asymptotically Euclidean manifold construct near symplectic forms whose zero loci are determined apriori from a Kirby calculus presentation of the manifold. Near symplectic forms are used in [ADK] to study the differential topology of 4-manifolds. Applications towards this same end are conjectured in [T1].

The remainder of this article contains the proofs of Theorems 1 and 2.

2. The birth of circles

The purpose of this section is to prove the following proposition:

Proposition 2.1. *Let n denote a non-negative integer and let X denote a smooth oriented 4-manifold with a near symplectic form whose zero locus consists of n disjoint, embedded circles. Then X has a cohomologically equivalent near symplectic form whose zero locus consists of $n+1$ disjoint, embedded circles.*

Proof of Proposition 2.1. A newborn circle is constructed here by changing the original 2-form on a compact set in a coordinate chart that is disjoint from all zeros of the original form. As a theorem of Moser [M] asserts that all symplectic forms are locally symplectomorphic, the construction of a new vanishing circle is needed only for the case when the original form is the standard symplectic form on a ball in $\mathbb{R}^4$. This standard form is denoted in what follows as ω; it is the form on $\mathbb{C}^2 = \mathbb{R}^4$ that is given with respect to complex coordinates (z, w) by

$$\omega = \frac{i}{2}(dz \wedge d\bar{z} + dw \wedge d\bar{w}). \tag{2.1}$$

The plan is to modify ω only near the origin so that the result, ω', is a near symplectic form that vanishes on a single circle. The description of this modification is given in five parts. The first part describes ω' in a small ball about the origin, and the subsequent parts describe ω' at successively larger distances from the origin through a distance beyond which $\omega' = \omega$

Part 1: Let ε denote a real number with absolute value less than 1 and suppose that $R \gg 1$. Now introduce the form $\mu = \mu(\varepsilon, R)$ given by

$$\mu = \frac{i}{2}(-\varepsilon+|z|^2-|w|^2)(dz\wedge d\bar{z}+dw\wedge d\bar{w})+\frac{i}{2}(zw-R\bar{w})d\bar{z}d\bar{w}-\frac{i}{2}(\bar{z}\bar{w}-Rw)dzdw. \tag{2.2}$$

This form is closed, and is such that $\mu \wedge \mu \geq 0$. When $\varepsilon > 0$, the zero locus of μ consists of two circles, these being

$$\begin{aligned} &\bullet \quad Z_\varepsilon = \{(z,w) : |z|^2 = \varepsilon \ \text{ and } \ w = 0\}. \\ &\bullet \quad Z_{\mathbb{R}} = \{(z,w) : |w|^2 = R^2 - \varepsilon \ \text{ and } \ z = R\bar{w}/w\}. \end{aligned} \tag{2.3}$$

In the case $\varepsilon < 0$, there is only one vanishing circle; this is the locus

$$\{(z,w) : |w|^2 = R^2 - \varepsilon \ \text{ and } \ z = R\bar{w}/w\}. \tag{2.4}$$

Thus, as long as $R^2 \gg |\varepsilon|$, the 1-parameter family $\varepsilon \to \mu(\varepsilon, R)$ sees the birth of a vanishing circle as ε crosses zero from negative values to positive values. Of course, reversing the motion of ε sees the death of a vanishing circle.

Now, fix $\varepsilon > 0$ but much less than 10^{-10}, and then fix R so that $R\varepsilon^{1/2} \gg 1$. Fix some large number, T. A lower bound for T is $10^{12}R\varepsilon^{1/2}$. The form ω' on the ball of radius $\delta = (2\varepsilon)^{1/2}$ about the origin is given by

$$\omega' = T^{-1}\mu. \tag{2.5}$$

Part 2: Let $r = (|z|^2 + |w|^2)^{1/2}$. This part describes ω' where $\delta \leq r \leq 4\delta$. For this purpose, fix a smooth, non-decreasing function on $[0, \infty)$ that has value 0 on $[0, 1]$ and value 1 on $[2, \infty)$. Let β denote the chosen function. When $\kappa > 0$ has been specified, then β_κ denotes the function on $\mathbb{R}^4$ that maps (z, w) to $\beta(r^2/\kappa^2)$. Thus, β_κ is zero where $r \leq \kappa$ and is equal to 1 where $r \geq 2^{1/2}\kappa$.

The form ω is exact; in particular, $\omega = d\alpha$ where

$$\alpha = \frac{i}{4}(zd\bar{z} - \bar{z}dz) + \frac{i}{4}(wd\bar{w} - \bar{w}dw). \tag{2.6}$$

With α understood, define ω' where $\delta \leq r \leq 4\delta$ to equal

$$\omega' = T^{-1}(\mu + d(\beta_\delta\alpha)). \tag{2.7}$$

The form depicted in (2.7) is closed. Moreover, its square can be written as the product of the Euclidean volume form and the function $T^{-2}\sigma_0$ where

$$\sigma_0 = 2\theta^2 + 2\delta^{-2}r^2\theta\beta'_\delta + 2|(\delta^{-2}\beta'_\delta + 1)zw - R\bar{w}|^2. \tag{2.8}$$

Here, $\theta = \beta_\delta + (-\varepsilon + |z|^2 - |w|^2)$ and β'_δ is shorthand for the function $(\frac{d}{dt}\beta)(r^2/\delta^2)$. As explained momentarily, (2.8) depicts a positive function where $\delta \leq r \leq 4\delta$.

To see that $\sigma_0 > 0$ where $\delta \leq r \leq 4\delta$, note first that σ_0 is positive if θ is positive, and θ is positive unless $|w|^2 + \varepsilon > |z|^2$. Since $\varepsilon = \frac{1}{2}\delta^2$ and $|w|^2 + |z|^2 \geq \delta^2$, this implies that θ is negative only where $|w|^2 = \frac{1}{4}\delta^2$. This understood, $\theta \geq -3|w|^2$ in any event. Now suppose that the right most term in (2.8) is greater than $\frac{1}{2}R^2|w|^2$. When such is the case, the previous lower bound for θ implies that the expression in (2.8) is no less than

$$2\theta^2 - 48|w|^2 + \frac{1}{2}R^2|w|^2. \tag{2.9}$$

This last expression is positive granted that $R^2 > 96$. As $R \geq 10$, the preceding inequality holds.

Meanwhile, the right most term in (2.8) is greater than $\frac{1}{2}R^2|w|^2$ unless

$$|(\delta^{-2}\beta'_\delta + 1)|z| - R| \leq \frac{1}{2}R. \tag{2.10}$$

This then requires that $|z| \geq \frac{1}{2}R\delta^2$. Under the circumstances, this last condition does not hold on the domain of interest because $\frac{1}{2}R\delta^2 = 2^{-1/2}(R\varepsilon^{1/2})\delta$. Indeed, this is much greater than 4δ given that $R\varepsilon^{1/2}$ is much greater than 1.

Part 3: This part describes the form ω' where $4\delta \leq r \leq 8\delta$. To this end, note that where $r \geq 2\delta$, the form that is depicted in (2.7) can be written as $T^{-1}\omega_1$ where

$$\omega_1 = (1+\varepsilon)\omega - \frac{i}{2}R(\bar{w}d\bar{z}d\bar{w} - wdzdw) + \left[(|z|^2 - |w|^2)\omega + \frac{i}{2}zwd\bar{z}d\bar{w} - \frac{i}{2}\bar{z}\bar{w}dzdw\right]. \tag{2.11}$$

Note in particular that the form that is depicted in the brackets on the far right in (2.11) has norm where $4\delta \leq r \leq 8\delta$ that is bounded by $10^3\delta^2$. Moreover, it can be written as $d\tau$ where $|\tau| = 10^5\delta^3$. As a consequence, the norm of $d[(1-\beta_{4\delta})\tau]$ is bounded by $10^6\delta^2$.

Granted these last points, define ω' where $4\delta \leq r \leq 8\delta$ to be

$$\omega' = (1+\varepsilon)\omega - \frac{i}{2}R(\bar{w}d\bar{z}d\bar{w} - wdzdw) + d[(1-\beta_{4\delta})\tau]. \tag{2.12}$$

Because the norm of the right most term in (2.12) is smaller than $10^6\delta^2$, the square of the form that is depicted in (2.12) is nowhere zero where $4\delta \leq r \leq 8\delta$ if $\delta < 10^{-4}$.

Part 4: This part describes ω' where $8\delta \leq r \leq 16\delta$. To this end, choose a nondecreasing function, χ, on $[0,\infty)$ with value $T^{-1}(1+\varepsilon)$ on $[0,1]$ and value 1 on $[2,\infty)$. Let $\chi_{8\delta}$ denote the function $\chi(r^2/(8\delta)^2)$ on $\mathbb{R}^4$. Let α denote the 1-form in (2.6). Noting that

$$\omega' = T^{-1}(1+\varepsilon)d\alpha - \frac{i}{2}T^{-1}R\bar{w}d\bar{z}d\bar{w} + \frac{i}{2}T^{-1}Rwdzdw \tag{2.13}$$

where $r \sim 8\delta$, the definition of ω' extends to where $8\delta \leq r \leq 16\delta$ as

$$\omega' = d(\chi_{8\delta}\alpha) - \frac{i}{2}T^{-1}R\bar{w}d\bar{z}d\bar{w} + \frac{i}{2}T^{-1}Rwdzdw \tag{2.14}$$

The square of the form that is depicted in (2.14) is a nowhere zero multiple of the Euclidean volume form. Indeed, the square is obtained by multiplying the volume form by the function

$$2\chi_{8\delta}^2 + 2(8\delta)^{-2}r^2\chi_{8\delta}\chi'_{8\delta} + 2|(8\delta)^{-2}zw\chi'_{8\delta} - T^{-1}R\bar{w}|^2; \tag{2.15}$$

and this function is positive because $\chi'_{8\delta} = (\frac{d}{dt}\chi)(r^2/(8\delta)^2)$ is non-negative.

Part 5: This last part describes ω' where $r \geq 16\delta$. To this end, note that the 2-form $-\frac{i}{2}\bar{w}d\bar{z}d\bar{w} + \frac{i}{2}wdzdw$ can be written as $d\nu$ with $\nu = -\frac{i}{2}(\bar{w}\bar{z}d\bar{w} - wzdw)$. This understood, extend the definition of ω' to where $r \geq 16\delta$ using

$$\omega' = \omega + T^{-1}Rd[(1-\beta_{16\delta})\nu]. \tag{2.16}$$

This form has everywhere positive square provided that $T^{-1}R\delta \ll 10^{-6}$. Thus, as long as $T \gg 10^6R\delta = 2^{1/2}10^6R\varepsilon^{1/2}$, the form depicted in (2.16) is symplectic where $r \geq 16\delta$. By design, it is equal to ω where $r \geq 32\delta$.

3. Melding component circles

Suppose that ω is a near symplectic form on a given 4-manifold whose zero locus is a smooth, embedded, union of circles. Let Z denote this zero locus. Suppose that $p \neq p'$ are points in Z. The purpose of the subsequent discussion is to describe a second near symplectic form that is cohomologous to ω and whose zero locus, Z', is a disjoint union of embedded circles with the following property: There exists a ball, B_0, that contains p and p' and is such that

- $Z \cap B_0$ is the disjoint union of two intervals, I and I', with $p \in I$ and $p' \in I'$.
- $Z \cap (M - B_0) = Z' \cap (M - B_0)$.
- $Z' \cap B_0$ is the disjoint union of two arcs that connect ∂I to $\partial I'$.

The new form is denoted in what follows by ω'.

Theorems 1 and 2 follow directly given Proposition 2.1 and the existence of the form ω' as just described.

There are six parts to the construction of B_0 and ω'.

Part 1: Here is the strategy: The ball B_0 contains a smaller, closed ball, B_1, that is chosen to have the following four properties: First, p and p' are in B_1 and $Z \cap B_1$ is a pair of disjoint arcs, one containing p and the other p'. Second, the inclusion $\iota : \partial B_1 \to X$ is transversal to Z and so ∂B_1 intersects Z at four points. Third, $\iota^*\omega$. can be written as $d\alpha$ and $\alpha \wedge d\alpha \geq 0$ with equality only at the four points in $Z \cap \partial B_1$. In particular, α is a contact form on the complement in ∂B_1 of $Z \cap \partial B_1$. Finally, the contact structure that α defines on the complement of $Z \cap \partial B_1$ is overtwisted.

With the preceding understood, a diffeomorphism is constructed from B_1 to itself with certain special properties. First, the diffeomorphism interchanges two of the four points that comprise $Z \cap \partial B_1$; it fixes one boundary of I and one of I' while interchanging the other boundary component of I with that of I'. Second, the diffeomorphism pulls ω back as itself in a neighborhood of four small radius balls that are centered on the four points of $Z \cap \partial B_1$. Third, the diffeomorphism pulls α back as itself near these same four point. Let ψ denote this diffeomorphism. With the preceding understood, the form $\psi^*\alpha$ agrees with α near $Z \cap \partial B_1$ and is a contact form on the complement in ∂B_1 of $Z \cap \partial B_1$. Note that $\psi^*\alpha$ is overtwisted since α is overtwisted. Finally, the 2-plane fields kernel(α) and kernel($\psi^*\alpha$) are homotopic as 2-plane fields with fix boundary values.

Theorems of Eliashberg and Gray are invoked next to find a diffeomorphism, $\phi : \partial B_1 \to \partial B_1$ that restricts as the identity on a neighbhood of $Z \cap \partial B_1$, and is such that $\phi^*\psi^*\alpha = g\alpha$ with $g > 0$ a function on ∂B_1 that equals 1 near $Z \cap \partial B_1$. With ϕ in hand, define a new manifold, X', by surgery on X:

$$X' = (X - B_1) \cup_\phi B_1. \tag{3.1}$$

Thus, B_1 is removed and then glued back using the diffeomorphism ϕ. A theorem of Hatcher [Ha] says that ϕ is homotopic to the identity map of ∂B_1 via a 1-parameter family of diffeomorphisms of B_1. As a result, X is diffeomorphic to X'. Here is another consequence: Use (3.1) to define a canonical embedding, $\upsilon : X - B_1 \to X'$. Let $B_0 \subset X$ denote a ball that contains B_1 in its interior. There exists a diffeomorphism $\lambda : X \to X'$ such that $\lambda = \upsilon$ on $X - B_0$.

The final step argues that the form ω on $X - B_1$ can be modified in a small neighborhood of B_1 so as to match up smoothly across ∂B_1 with a small, positive, constant multiple of the form $\phi^*\psi^*\omega$. The two thus define a smooth, near-symplectic form on X' whose zero locus is a disjoint union of embedded circles. Take ω' to be the pull back via the diffeomorphism λ of this near symplectic form on X'.

Part 2: This part of the story describes the ball B_1. This ball is obtained by smoothing a C^1 ball that is obtained as the union of five parts. Two of these parts consist of a pair that lie near p and two consist of an analogous pair near p'. The remaining part is a small radius, tubular neighborhood of a path between p and p'. The interior of this path should be disjoint from Z, and it is constrained near its endpoints. These five parts are described in turn.

To consider the parts near p, remark first that the form ω can be modified near its zero locus so that the modification has the same zero locus as the original, and such that any given point on the zero locus has a neighborhood with coordinates (t, x, y, z) for which the modified form appears as:

$$dt(xdx + ydy - 2zdz) + xdydz - ydxdz - 2zdxdy. \tag{3.2}$$

Thus, the zero locus is the t-axis in this coordinate chart. The Euclidean metric defines a metric for such a coordinate chart. Such modifications near the points p and p' are assumed implicitly in what follows; the modified form is denoted by ω as was the original.

Suppose now that (t, x, y, z) are coordinates as just described that are centered on the point p. Fix a small real number, $\varepsilon > 0$, and let C_- denote the half ball in the coordinate system given by the conditions

$$C_- = \{(t, x, y, z) : t^2 + x^2 + y^2 + z^2 = \varepsilon^2 \text{ and } z \leq 0\}. \tag{3.3}$$

This C_- is the first of the pair of components that defines the ball B_1.

The second component is denoted as C_+. The specification of C_+ requires the choice of a small, positive number, $\delta > 0$. I shall take $\delta \gg \varepsilon$, and this will require that ε be very small. Here is C_+:

$$C_+ = \{(t, x, y, z) : t^2 + x^2 + y^2 \leq \varepsilon^2 \text{ and } 0 \leq z \leq \delta\}. \tag{3.4}$$

Thus, $C_- \cup C_+$ is a half-ball that extends a distance δ along the positive z axis.

Analogous versions of C_- and C_+ are defined near p' using the p' version of the coordinates (t, x, y, z). These are denoted in what follows by C'_- and C'_+.

To define the final component of B_1, choose an embedded arc, $\gamma : [-1, 1] \to X$ such that $\gamma(-1) = p$ and $\gamma(1) = p'$. Require that γ have the following properties: First, its interior is disjoint from Z. Second γ coincides where its affine parameter is near -1 with a segment of the positive z axis. To be precise, assume that

$$\gamma(s) = (t = 0,\ x = 0,\ y = 0,\ z = s + 1) \tag{3.5}$$

for $-1 \leq s \leq -1 + \delta$. An analogous constraint as defined using the p' version of the coordinates (t, x, y, z) is required near $s = 1$. In this case, γ should coincide where its affine parameter is near 1 with a segment of the negative z axis; thus $\gamma(s) = (0, 0, 0, s - 1)$ when $1 - \delta \leq s \leq 1$.

Granted (3.5), the fifth part of B_1 consists of the radius ε tubular neighborhood of the portion of γ where $-1 + \delta \leq s \leq 1 - \delta$. Use C_0 to denote this portion of B_1.

Note that ∂B_1 is a C^1 submanifold, but not C^2. The failure of differentiablility occurs where C_+ joins to C_-, thus on the 2-sphere where $z = 0$ and $t^2 + x^2 + y^2 = \varepsilon^2$. One way to rectify this is to replace C_- as follows: Fix some $\varepsilon_1 > 0$ but very small so that $\varepsilon_1 \ll \varepsilon$. Let $f : [0, 1] \to [0, \infty)$ denote a smooth, non-decreasing function with $f(t) = t$ where $t \geq \varepsilon_1^2$ and $f(t) = 0$ where $t \leq \frac{1}{2}\varepsilon_1^2$. Now set

$$C'_- = \{(t, x, y, z) : t^2 + x^2 + y^2 + f(z^2) \leq \varepsilon^2 \text{ and } z \leq 0\}. \tag{3.6}$$

Part 3: This part describes the 1-form α. To this end, I first define α on the part of ∂B_1 in C_+. To do so, I write $t = r\cos(\lambda)$, $x = r\sin(\lambda)\cos(\varphi)$ and $y = r\sin(\lambda)\sin(\varphi)$; I then observe that ω on C_+ is given by

$$\begin{aligned}\omega = r^2 \sin(\lambda)drd\lambda - 2z\cos(\lambda)drdz + 2zr\sin(\lambda)d\lambda dz \\ + r^2\sin^2(\lambda)d\varphi dz - 2zr\sin^2(\lambda)drd\varphi - 2zr^2\sin(\lambda)\cos(\lambda)d\lambda d\varphi.\end{aligned} \tag{3.7}$$

This understood, the pull-back of ω to the part of ∂B_1 in C_+ is

$$\omega_+ = 2rz\sin(\lambda)d\lambda dz - r^2\sin^2(\lambda)dzd\varphi - 2r^2z\sin(\lambda)\cos(\lambda)d\lambda d\varphi. \tag{3.8}$$

Granted (3.8), note that $\omega = d\alpha_+$ with

$$\alpha_+ = -r^2z\sin^2(\lambda)d\varphi - 2r\cos(\lambda)zdz + \frac{1}{3}r^3\sin(\lambda)d\lambda. \tag{3.9}$$

Thus, upon restriction to ∂B_1, one has $\omega_+ = d\alpha_+$. In addition:

$$\alpha_+ \wedge d\alpha_+ = \left[\frac{1}{3}r^2\sin^2(\lambda) + 2z^2(1 + \cos^2(\lambda))\right] r^3\sin(\lambda)d\lambda d\varphi dz. \tag{3.10}$$

This 3-form vanishes only where $x = y = z = 0$.

Consider next the story for C'_-. On the portion of ∂B_1 in C_- near where $z > -\varepsilon$, the function r becomes a function of z via

$$r = (\varepsilon^2 - f(z^2))^{1/2}. \tag{3.11}$$

As a consequence, the pull-back of ω to the portion of $\partial B_1 \cap C'_-$ were $z > -\varepsilon$ is

$$\begin{aligned}\omega_- = (-r^2r_z + 2zr)\sin(\lambda)d\lambda dz - (r^2 + 2rzr_z)\sin^2(\lambda)dzd\varphi \\ - 2zr^2\sin(\lambda)\cos(\lambda)d\lambda d\varphi.\end{aligned} \tag{3.12}$$

Note that I can write ω_- as $d\alpha_-$ with

$$\alpha_- = -r^2z\sin^2(\lambda)d\varphi - 2r\cos(\lambda)zdz + \frac{1}{3}r^3\sin(\lambda)d\lambda. \tag{3.13}$$

A calculation finds that

$$\alpha_- \wedge d\alpha_- = \left[\frac{1}{3}((r^2 + z^2f')\sin^2(\lambda) + 2z^2(1 + \cos 2(\lambda))\right] r^3\sin(\lambda)d\lambda d\varphi dz. \tag{3.14}$$

A change of coordinates near where $z = -\varepsilon$ finds that α_- is smooth on this locus also. The 3-form that is depicted in (3.14) vanishes only on the locus where $z = x = y = 0$.

A comparison between (3.13) and (3.9) finds that α_- smoothly extends α_+ from the C_+ portion of ∂B_1 to the C_- portion.

Here is one last observation: The contact structure just described in *overtwisted.* Indeed, the circle where $z = 0$ and $\cos(\lambda) = 0$ is tangent to the contact plane field, but bounds a disk in the portion of ∂B_1 in $C_- \cup C_+$ that avoids the two points where the contact form vanishes and is transverse to the contact plane field along its boundary. This disk is obtained as a perturbation of the disk that is defined by the conditions $z = 0$ and $\frac{1}{2}\pi \le \lambda \le \pi$; the pertubation has z becoming slightly negative as λ approaches π. The existence of a disk with these properties characterizes an overtwisted contact structure.

An analogous contact form should be defined on the part of ∂B_1 in $C'_- \cup C'_+$. The C'_+ part of the latter is denoted below by α'_+.

The next task is to extend the contact structure just described to the portion of ∂B_1 that lies in C_0. To this end, observe that for $z \sim \delta$ on the part of ∂B_1 in C_+, the contact form α_+ can be written as

$$-\tau dk - zb^2 d\varphi + \frac{1}{3} r^3 \sin(\lambda) d\lambda \tag{3.15}$$

where $b = r\sin(\lambda)$, $\tau = r\cos(\lambda)$ and $k = z^2 - 1$. Meanwhile, for $z' \sim -\delta$ on the part of ∂B_1 in C'_+, the contact form α'_+ can be written as

$$-\tau' dk' - |z'|b'^2 d(-\varphi') + \frac{1}{3} r^3 \sin(\lambda') d\lambda' \tag{3.16}$$

where $b' = r\sin(\lambda')$, $\tau' = r\cos(\pi - \lambda')$, and $k' = 1 - z'^2$. Note in this regard that the change of coordinates from (z', λ', φ') to $(z', \pi - \lambda', -\varphi')$ defines an oriented map that extends over $C'_+ \cup C'_-$ as the map $(t', z', x', y') \to (-t', z', x', -y')$.

Here is a crucial point: The function $k = z^2 - 1$ restricts to the $-1 < s \le -1 + \delta$ part of γ as an increasing function of s. It is also the case that $k' = 1 - z'^2$ restricts to the part of γ where $1 - \delta \le s < 1$ as an increasing function of s. As a consequence, there is an oriented diffeomorphism, $s \to \sigma(s)$, that sends $(-1, 1)$ to $(-1, 1)$ and is such that $k = \sigma$ where $\sigma \sim -1 + \delta^2$, and such that $k' = \sigma$ where $\sigma \sim 1 - \delta^2$.

To proceed, it is worth considering what ω looks like on a neighborhood of an embedded path in X that avoids Z. I'll denote this path by γ. Choose a coordinate system (σ, v_1, v_2, v_3) on a tubular neighborhood of γ so that $\gamma = \{(-, 0, 0, 0)\}$. Then

$$\omega|_\gamma = d\sigma \wedge A_i(\sigma) dv_i + \varepsilon_{ijk} B_k(\sigma) dv_i \wedge dv_j, \tag{3.17}$$

where $B_i A_i > 0$ at all values of σ. I can change coordinates $w_1 = A_i(\sigma) v_i$ so that

$$\omega|_\gamma = d\sigma \wedge dw_1 + 2B_1 dw_2 \wedge dw_3 + (B_2 dw_3 - B_3 dw_2) \wedge dw_1 \tag{3.18}$$

after redefining the collection $\{B_i\}$. Note that $B_1 > 0$. Change coordinates again, this time to coordinates (σ, c_1, c_2, c_3) with (c_1, c_2, c_3) given by

$$w_1 = c_1,\ w_2 = -mB_1^{-1/2}(c_2 + (B_2/B_1)c_1),\ w_3 = mB_1^{-1/2}(c_3 + (B_3/B_3)c_1); \tag{3.19}$$

here $\sigma \to m(\sigma)$ is a favorite, strictly positive function on the domain of σ. With respect to these new coordinates,

$$\omega|_\gamma = d\sigma \wedge dc_1 - 2m(\sigma) dc_2 \wedge dc_3. \tag{3.20}$$

Now consider extending ω off of γ. To this end, fix in advance a smooth function $\sigma \to n(\sigma)$ and then use Moser's procedure [M] to find a tubular neighborhood of γ with coordinates (σ, c_1, c_2, c_3) where (3.20) can be extended as

$$\begin{aligned} \omega = {} & d\sigma \wedge dc_1 + m'(c_2 dc_3 - c_3 dc_2) \wedge d\sigma \\ & - 2m dc_2 \wedge dc_3 + \frac{1}{2} dc_1 \wedge d[n(\sigma)(c_2^2 + c_3^2)]. \end{aligned} \tag{3.21}$$

These coordinates are unique up to a Hamiltonian diffeomorphism that preserves γ.

To say more, write ω as in (3.21) on the tubular neighborhood and note that writing $c_1 = r\cos(\theta)$, $c_2 = r\sin(\theta)\sin(\phi)$ and $c_3 = r\sin(\theta)\cos(\phi)$, finds $\omega = da$ with

$$a = -r\cos(\theta)d\sigma - m(\sigma)r^2\sin^2(\theta)d\phi + \frac{1}{3}n(\sigma)r^3\sin(\theta)d\theta. \tag{3.22}$$

This looks very much like what is written in (3.15) with the following identifications: First, $k = \sigma$ while $m(\sigma) = z$ and $n(\sigma) = 1$ when $\sigma \sim -1 + \delta^2$. Second, $\theta = \lambda$ and $\phi = \varphi$ when $\sigma \sim -1 + \delta^2$. It also looks much like what appears in (3.16) with the identifications $k' = \sigma$, $m(\sigma) = |z'|$, $n(\sigma) = -1$, $\theta = -\lambda'$ and $\phi = -\varphi'$ when $\sigma \sim 1+\delta^2$.

These identifications can be made if $c_1 = t$ near p and $c_1 = -t'$ near p' and $(c_2, c_3) = (x, y)$ near p; meanwhile $c_1 = -t'$ and $(c_2, c_3) = (-x', y')$ near p'. This last set of identifications can be arranged with no difficulties when the tubular neighborhood has small radius (thus ε is very small).

Note finally that the restriction of ω as depicted in (3.21) to the locus where $r = (c_1^2 + c_2^2 + c_3^2)^{1/2} = \varepsilon$ is the form

$$\omega_0 = \left(1 - \frac{1}{3}n'r^2\right) r\sin(\theta)d\theta \wedge d\sigma - m'r^2\sin^2(\theta)d\sigma \wedge d\phi - 2mr^2\sin(\theta)\cos(\theta)d\theta \wedge d\phi. \tag{3.23}$$

In addition,

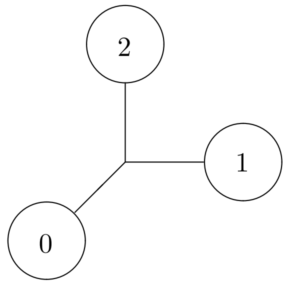
$$a \wedge \omega_0 = \left(2m\cos^2(\theta) + m\left(1 - \frac{1}{3}n'r^2\right)\sin^2(\theta) + \frac{1}{3}nm'r^2\sin^2(\theta)\right) r^3\sin(\theta)d\sigma d\theta d\phi, \tag{3.24}$$

which is positive provided that r is small.

Thus, taking $\alpha = a$ on the part of ∂B_1 in C_0 extends the definition of α to the whole of ∂B_1 as a contact form for the restriction of ω to B_1.

Part 4: This part of the discussion concerns the diffeomorphism ψ. As noted at the very outset, each point on the zero locus of ω has a neighborhood with coordinates in which ω appears as depicted in (3.2). In particular, this is the case for the four points where the zero locus intersects ∂B_1. This understood, there is no obstruction to demanding that ψ, as it permutes these points, maps balls about the points so as to identify the version of (t, x, y, z) in the domain ball with the corresponding version in the range ball. This will insure that $\psi^*\omega = \omega$ and $\psi^*\alpha = \alpha$ on these four balls.

It is proves convenient when discussing the 2-plane fields kernel(α) and kernel($\psi^*\alpha$) to make a further constraint on ψ beyond the requirement that it permute the four balls described above. To describe this additional constraint, agree to fix a small ball in ∂B_1 about each of the zeros of ω where the coordinates used in (3.2) are valid. Let $S \subset \partial B_1$ denote the complement of these four balls. This S can be viewed as the complement inside the unit ball of $\mathbb{R}^3$ of three smaller open balls, b_0, b_1 and b_2. To be explicit, take these smaller balls to have radius $\frac{1}{100}$, with the center of b_0 at $(-\frac{1}{25}, -\frac{1}{25}, 0)$, the center of b_1 at $(\frac{1}{25}, 0, 0)$ and the center of b_2 at $(0, \frac{1}{25}, 0)$. The map ψ will send b_1 to b_2 while fixing b_0 and the boundary of the 3-ball. Let L denote the 2-complex that is obtained from $\partial b_0 \cup \partial b_1 \cup \partial b_2$ by adjoining three arcs, these the respective shortest arcs between the origin and ∂b_0, ∂b_1 and ∂b_2. Let L denote this 2-complex. Here is a picture of L:



A sketch of L.

The map ψ can be constructed so as to map L to itself and so that $\psi^*(\alpha) = \alpha$ on a neighborhood of $b_0 \cup b_1 \cup b_2$ and on a neighborhood of the boundary of the unit ball in $\mathbb{R}^3$.

Keeping in mind that α and $\psi^*\alpha$ agree on the boundary of S, the final point here is made by

Lemma 3.1. *The kernel of $\psi^*\alpha$ can be homotoped as a 2-plane field to the kernel of α with no change along the homotopy on the boundary of S.*

Proof of Lemma 3.1. Frame T^*S using the standard coordinate framing from $\mathbb{R}^3$. As an oriented 2-plane field on S is the kernel of a nowhere zero 1-form on S, this fixed framing identifies any given 2-plane field with a map from S to S^2, and it identifies any given map from S to S^2 with a 2-plane field. This understood, the kernels of α and $\psi^*\alpha$ are homotopic as 2-plane fields (rel ∂S) if and only if the corresponding maps to S^2 are homotopic rel ∂S. To prove that such a homotopy exists, note that S can be written as $((S^2 \times [0,1]) \cup L)/\sim$, where the equivalence relation identifies $S^2 \times \{0\}$ with L via a surjective map $f : S^2 \to L$. This follows from the fact that L has a regular neighborhood whose boundary is a 2-sphere. This picture of S identifies the boundary of the unit ball with $S^2 \times \{1\} \subset S$.

Let $\sigma : S \to S^2$ denote a given map to S^2. Agree to identify the image S^2 with the unit sphere in the Lie algebra of $SU(2)$. The pull-back of σ from S to $S^2 \times [0,1]$ can be lifted as a map to $SU(2)$ in the following manner: The lift, $h : S^2 \times [0,1] \to SU(2)$ is such that $\sigma(z,t) = h(z,t)\sigma(z,0)h^{-1}(z,t)$ and such that $h(z,0) = \mathbb{I} \in SU(2)$. Note that h is not unique; it can be modified by $h \to h\exp(u\sigma)$ where $u : S^2 \times [0,1] \to \mathbb{R}/(2\pi\mathbb{Z})$.

If σ and σ' are two maps from S to S^2 that agree on $S^2 \times \{1\}$, then the corresponding lifts h and h' can be chosen so as to agree on $S^2 \times \{1\}$ as well. This being the case, then the map from $S^2 \times [0,1]$ to S^2 that sends (z,t) to $h'(z,2t)$ for $t \leq \frac{1}{2}$ and to $h(z,2-2t)$ for $t \geq \frac{1}{2}$ is a map from $S^2 \times [0,1]$ to $SU(2)$ that restricts to both $S^2 \times \{0\}$ and $S^2 \times \{1\}$ as the identity. Let μ denote a volume form on $SU(2)$ with volume 1. Then the integral over $S^2 \times [0,1]$ of the pull-back of this form by g is an integer, this denoted in what follows by $n_{h,h'}$. It is relevant only by virtue of the following fact: Given maps h and h' as just described are homotopic if and only if $n_{h,h'} = 0$. Note for reference below that this integer is zero in the case where h is defined by α and h' by $\psi^*\alpha$; this is because h' can be taken to equal ψ^*h. This last conclusion requires that ψ map L to itself.

The preceding observation does not imply that the respective maps from S to S^2 defined by a and $\psi^*\alpha$ are homotopic. One further item is needed. To explain, let $\sigma : S \to S^2$ denote the map that is defined by α and let σ' denote the one defined $\psi^*\alpha$. A homotopy from h to h' would give one from σ and σ' were $\sigma = \sigma'$ on L. However, such need not be the case. To address this concern, note first that σ and σ' agree on the boundary of the unit ball, and also on the parts of L that comprise boundaries of the balls b_0, b_1 and b_2. The maps σ and σ' can only differ on the three arcs that connect ∂b_0, ∂b_1 and ∂b_2 to the origin. Of course, σ and σ' agree on the endpoints of these arcs. In any event, the map σ' can be homotoped rel ∂S so that the result, σ'', differs from σ' only in a small tubular neighborhood of each arc from L and agrees with σ on L. This is because the image space S^2 is simply connected. This map σ'' produces a corresponding map, $h'' : S^2 \times [0, 1] \to S^2$. The issue here is whether h and h'' are homotopic.

To see that such is the case, σ'' will be constructed in steps, and each step will results in a corresponding map from $S^2 \times [0, 1]$ that is homotopic to h' rel $S^2 \times \{0, 1\}$. The first step homotopes σ' to a map, σ_1, that agrees with σ' on the complement of a very small radius ball about the origin and also at the origin; but differs in being constant on a small neighborhood of the origin. This can be done so that σ_1 is as close as desired to σ' in the C^0 topology. The result gives a map, h_1, that is C^0 close to h' and so is homotopic to h' rel ∂S.

The second step changes σ_1 to σ_2. The map σ_2 can be made as close as desired to σ_1 in the C^0 topology. In particular, it agrees with σ_1 on the complement of the union of a small radius tubular neighborhood of each arc in L and it agrees with σ_1 on each such arc. However, in a very small radius neighborhood of each such arc from L, the map σ_2 depends only on the affine coordinate along the central arc. The fact that σ_2 is C^0 close to σ_1 implies that the corresponding h_2 is homotopic to h_1.

The final step changes σ_2 to σ''. The map σ'' agrees with σ_2 on the complement of a very small tubular neighborhood of each arc in L. It is assumed in what follows that the radius of these neighborhoods is chosen so that in any such neighborhood, the map σ_2 depends only on the affine coordinate along the central arc. Inside such a tubular neighborhood, the map σ'' depends only on the affine coordinate along the central arc and on the radial coordinate on the transverse disks to the arc. Meanwhile, σ'' agrees with σ along each arc in L. Now, there is no reason for σ'' to be C^0 close to σ_2 since there is no apriori reason for α and $\psi^*\alpha$ to be close along these arcs. However, where σ'' differs from σ_2, both maps factor through a two dimensional space. Here is why: Where $\sigma_2 \neq \sigma''$, the map σ_2 depends only on the affine coordinate along each arc of L, and σ'' depends only on the latter coordinate and on the radial coordinate on the transverse disks. As a consequence, the corresponding maps h_2 and h'' can be taken so as to differ only where they both factor through a 2-dimensional space. This implies that the volume form on $SU(2)$ is pulled back as zero by both h_2 and h'' where these two maps differ. As a consequence, $n_{h_2,h''} = 0$ and so h_2 and h'' are homotopic as desired.

Part 5: This part concern the existence of the diffeomorphism ϕ. To begin, recall that Lemma 3.1 asserts that α and $\psi^*\alpha$ define 2-plane fields on S that are homotopic rel ∂S. As both define overtwisted contact plane fields, a theorem of Eliashberg (Theorem 3.1.1 in [E]) asserts that these contact 2-plane fields are homotopic as contact fields on S via an homotopy that restricts to the identity on ∂S. This understood, a theorem of Gray [Gr] asserts that there exists a diffeomorphism, this being ϕ, that restricts as the identity on a neighborhood of the 4 points in ∂B_1 where $\omega = 0$, and pulls back the kernel of $\psi^*\alpha$ as the kernel of α where these 1-forms are non-zero. The fact that $\phi^*\psi^*\text{kernel}\,(\alpha) = \text{kernel}\,(\alpha)$ and $\phi^*\psi^*\alpha = \alpha$ near the zeros of ω imply that

$$\phi^*\psi^*\alpha = g\alpha \tag{3.25}$$

where g is a strictly positive function on ∂B^1.

Part 6: Fix some very small but positive number ε_2 and use the exponential map from the metric to trivialize a tubular neighborhood of ∂B_1 as $(-\varepsilon_2, \varepsilon_2) \times \partial B_1$ so that ∂B_1 identified with $\{0\} \times \partial B_1$. This trivialization can and should be chosen with the following property: Near each zero of ω in ∂B_1, the fibers of the projection to ∂B_1 appear in the coordinates (r, λ, φ, z) that are used in (3.9) as the loci where (λ, φ, z) is constant. In what follows, s denotes the coordinate on $(-\varepsilon_2, \varepsilon_2)$. Near each zero of ω where the projection sends (r, λ, φ, z) to (λ, φ, z), the coordinate s is taken to be $r - \varepsilon$.

Extend ϕ to this tubular neighborhood as the identity on the $(-\varepsilon_2, \varepsilon_2)$ factor. Doing so produces two symplectic forms on the $(-\varepsilon_2, 0)$ portion of the tubular neighborhood; the first being ω, and the second $\phi^*\psi^*\omega$. Note that they agree on the portion of the tubular neighborhood that lies over any small radius ball in ∂B_1 about a zero of ω.

Let s now denote the coordinate on $(-\varepsilon_2, \varepsilon_2)$ and write

$$\omega = d(\alpha + sb) \tag{3.26}$$

where b is a smooth s-valued 1-form on ∂B_1. In this regard,

$$b|_{s=0} \wedge d\alpha \geq 0 \tag{3.27}$$

with equality only on the four zeros of ω in ∂B_1. Note that over a small radius ball in ∂B_1 about a zero of ω, one can assume without loss of generality that $\alpha + sb = \alpha_+$, this the form that is depicted in (3.9).

Meanwhile, $\phi^*\psi^*\omega$ can be written as

$$\phi^*\psi^*\omega = d(g \cdot \alpha + sb') \tag{3.28}$$

where b' is another smooth, s-valued 1-form on ∂B_1. In this case,

$$b'|_{s=0} \wedge (dg \wedge \alpha + g d\alpha) \geq 0 \tag{3.29}$$

with equality only on the zeros of ω in ∂B_1. As in the case with ω, there is no generality lost by taking $g \cdot \alpha + sb' = \alpha_+$ over a small radius ball about a zero of ω.

Now, let β denote a non-decreasing, smooth function on $[-\varepsilon_2, \varepsilon_2]$ that is 0 near $-\varepsilon_2$ and 1 on $[0, \varepsilon_2]$. In particular, given some positive $\varepsilon_3 \ll \varepsilon_2$, choose β so that $\beta' = 1/\varepsilon_2$ where $-\varepsilon_2 + \varepsilon_3 \leq s \leq -\varepsilon_3$ and $\beta' < 2/\varepsilon_2$ everywhere. Let κ denote a positive number that is less than 1 and consider

$$\begin{aligned} \mu &= d[\beta(\alpha + sb) + \kappa(1 - \beta)(g\alpha + sb')] \\ &= ds \wedge [\beta b + \kappa(1 - \beta)b' + \beta'((1 - \kappa g)\alpha + sb - \kappa sb')] \\ &\quad + [\beta + \kappa g(1 - \beta)]d\alpha + \kappa(1 - \beta)dg \wedge \alpha + \cdots, \end{aligned} \tag{3.30}$$

where the unwritten terms are $\mathcal{O}(\varepsilon_2)$ in size. The important point here is that the form depicted in (3.30) is symplectic except at the zeros of ω if $\kappa < \sup(g)$, ε_2 is very small, and then ε_3 very much smaller than ε_2. To see why this is, consider first the story near a zero of ω. Near such a point, $g = 1$, the coordinate $s = r - \varepsilon$, and

$$\mu = (\beta + \kappa(1-\beta))\omega + dr \wedge \beta'(1-\kappa)\alpha_+. \tag{3.31}$$

Given (3.7), this finds

$$\begin{aligned}\mu = {} & dr \wedge \left[(\beta + \kappa(1-\beta)) + \frac{1}{3} r\beta'(1-\kappa)\right] r^2 \sin(\lambda) d\lambda \\ & - dr \wedge [(\beta + \kappa(1-\beta)) + r\beta'(1-\kappa)] 2z\cos(\lambda) dz \\ & - dr \wedge [(\beta + \kappa(1-\beta)) + r\beta'(1-\kappa)] z \sin^2(\lambda) d\varphi \\ & + 2zr\sin(\lambda) d\lambda dz + r^2 \sin^2(\lambda) d\varphi dz - 2zr^2 \sin(\lambda)\cos(\lambda) d\lambda d\varphi.\end{aligned} \tag{3.32}$$

When $\kappa < 1$, the square of this form is positive except on the line segments where both z and $\sin(\lambda)$ are 0. This follows from the fact that $\beta' \geq 0$. In particular, (3.32) defines a near symplectic form when $\kappa < 1$.

The next point to make is that $\mu \wedge \mu$ is strictly positive away from the zero locus of ω. To see why, remark that if $\kappa < \sup(g)$ and ε_2 is very small, then $\mu \wedge \mu$ is dominated by the following part of μ

$$\beta(1-\kappa g) ds \wedge \alpha + [\beta + \kappa g(1-\beta)] d\alpha + \kappa(1-\beta) dg \wedge \alpha \tag{3.33}$$

except where β' is on the order of unity or smaller. If ε_3 is much smaller than ε_2, then the latter region has β very close to 1 and $(1-\beta) \sim \varepsilon_3/\varepsilon_2$ or $(1-\beta)$ very close to 1 and $\beta \sim \varepsilon_3/\varepsilon_2$. In the former case, terms with $(1-\beta)$ are negligible, and

$$\mu \sim ds \wedge [\beta b + \beta'((1-\kappa g)\alpha + sb - \kappa s b')] + \beta d\alpha. \tag{3.34}$$

This understood, the fact that μ is near symplectic here follows from (3.27). In the case where β is very small,

$$\mu \sim ds \wedge [\kappa(1-\beta)b' + \beta'((1-\kappa g)\alpha + sb - \kappa s b')] + \kappa(1-\beta)(g d\alpha + dg \wedge \alpha). \tag{3.35}$$

Here, the fact that μ is near symplectic follows from (3.29).

Note that $\mu = \omega$ where $s \geq 0$ and $\mu = \phi^* \psi^* \omega$ where s is very near $-\varepsilon_2$.

Acknowledgement. The author thanks Rosa Sena-Dias for discussions about the theorem of Luttinger and Simpson.

References

[ADK] D. Auroux, S.K. Donaldson and L. Katzarkov, *Singular Lefschetz pencils*, Geom. Topol. **9** (2005) 1043–1114.

[DK] S.K. Donaldson and P.B. Kronheimer, The geometry of 4-manifolds, Oxford University Press, 1990.

[E] Y. Eliashberg, *The classification of overtwisted contact structures on 3-manifolds*, Invent. Math. **98** (1989) 623–637.

[Gr] J. Gray, *Some global properties of contact structures*, Ann. Math. **69** (1959) 421–450.

[G] M. Gromov, *Pseudo-holomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985) 307–347.

[Ha] A. Hatcher, *A proof of a Smale conjecture,* Diff $(S^3) \simeq SO(4)$, Ann. of Math. **117** (1983) 553–607.

[Ho] K. Honda, *An openness theorem for harmonic 2-forms on 4-manifolds*, Illinois J. Math. **44** (2000) 479–495.

[K] P.B. Kronheimer, *Minimal genus in $S^1 \times M^3$*, Invent. Math. **135** (1999) 45–61.

[KG] D. Gay and R. Kirby, *Constructing symplectic forms on 4-manifolds which vanish on circles*, Geom. Topol. **8** (2004) 743–777.

[M] J.K. Moser, *On the volume of manifolds*, Transactions of the A.M.S. **120** (1965) 280–296.

[P] T. Perutz, *Zero sets of near symplectic forms*, math.SG/0601320.

[S] R.H. Scott, *Closed, self-dual two-forms on four-dimensional handlebodies*, Ph.D. Thesis, Harvard Univerity 2003.

[T1] C.H. Taubes, *The geometry of the Seiberg-Witten invariants*, in the Proceedings of the International Congress of Mathematicians, Berlin 1998, Vol II, Documenta Mathematica Extra Volume ICM 1998, 493–504.

[T2] ———, *The Seiberg-Witten invariants and symplectic forms*, Math. Res. Lett. **1** (1994) 809–822.

[T3] ———, *More constraints on symplectic forms from the Seiberg-Witten invariants*, Math. Res. Lett. **2** (1995) 9–13.

Department of Mathematics, Harvard University, Cambridge, MA 02138
E-mail address: chtaubes@math.harvard.edu