Physical cryptographic verification of nuclear warheads

Citation: Kemp, R. Scott et al. "Physical Cryptographic Verification of Nuclear Warheads." Proceedings of the National Academy of Sciences 113, 31 (July 2016): 8618–8623 © 2016 National Academy of Sciences

As Published: http://dx.doi.org/10.1073/PNAS.1603916113

Publisher: National Academy of Sciences (U.S.)

Persistent URL: http://hdl.handle.net/1721.1/114901

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Physical cryptographic verification of nuclear warheads

R. Scott Kemp[a,1,2], Areg Danagoulian[a,1], Ruaridh R. Macdonald[a,1], and Jayson R. Vavrek[a,1]

[a]Laboratory for Nuclear Security and Policy, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Richard L. Garwin, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, and approved June 3, 2016 (received for review March 8, 2016)

How does one prove a claim about a highly sensitive object such as a nuclear weapon without revealing information about the object? This paradox has challenged nuclear arms control for more than five decades. We present a mechanism in the form of an interactive proof system that can validate the structure and composition of an object, such as a nuclear warhead, to arbitrary precision without revealing either its structure or composition. We introduce a tomographic method that simultaneously resolves both the geometric and isotopic makeup of an object. We also introduce a method of protecting information using a provably secure cryptographic hash that does not rely on electronics or software. These techniques, when combined with a suitable protocol, constitute an interactive proof system that could reject hoax items and clear authentic warheads with excellent sensitivity in reasonably short measurement times.

isotopic tomography | nuclear weapons | disarmament | verification

The United States and Russia together retain more than 10,000 nuclear weapons, of which about 5,500 are currently designated to be removed from the stockpile (1). More designations are expected in the future: The 2010 New Strategic Arms Reduction Treaty requires both countries to reduce to a maximum of 1,550 deployed warheads each by February 2018, and the most recent US Nuclear Posture Review states, "The United States and Russia have deeply reduced their nuclear forces from Cold War levels, but both still retain many more nuclear weapons than needed" (2).

The treaties behind arms reductions typically require verification measures to assure compliance with obligations. However, despite nearly five decades of research by US and Russian laboratories, no method for verifying the dismantlement of nuclear warheads without unacceptable intrusiveness has been found. In the absence of a solution, past arms-control treaties have verified the dismantlement of delivery systems instead, an effective substitute provided the majority of nuclear forces are deployed in dedicated systems. Direct verification of nondeployed warheads will ultimately be needed to certify warhead destruction, which is critical to reducing their influence on the strategic calculus and the risk of theft.

Attempts to verify warhead dismantlement began with Project Cloud Gap in 1963 (3). This and subsequent approaches were built upon traditional assay methods, but it was soon discovered that even very invasive techniques could be readily defeated with a hoax, whereas the least invasive methods still revealed classified design information (4–6). As a result, proposals for verifying warheads grew progressively less intrusive. The epistemology of certainty came to be defined, not in mathematical terms, but by a view that human interactions during verification could create confidence. In an acknowledgment of the seemingly intractable confidence–secrecy tradeoff, a high-level scientific review panel stated that verification "will of necessity be less than perfect," and "must rely on difficult political/strategic judgments" (4). History has shown, however, that staking the verification of controversial arms-control agreements on trust has led to outsized fears of cheating that can facilitate a failure to ratify treaties in national legislatures (7, 8). A rigorous approach to warhead verification is therefore desirable if it can simultaneously satisfy secrecy requirements.

In an attempt to solve this dilemma, information barriers were introduced in the 1990s. These are software programs that analyze measured attributes of warheads while hiding data from human inspectors (5, 6, 9). In principle, the information barrier would allow for more intrusive measurements while still protecting secrets. It was later recognized that this approach merely shifted the challenge of protecting classified information into the electronic domain: Potentially undiscovered backdoors, encryption vulnerabilities, or hardware flaws could allow classified information to leak through the barrier. Integrity suffered as well: Trapdoor functionalities could be designed into the system to pass specially made hoax objects as if they were authentic warheads (5). The sensitivity of measurements also suffered because secrecy requirements forced measurement objectives to be so loosely defined that cheating became trivial. Whereas work on electronic information barriers continues, there is at present no accepted method for proving the absence of vulnerabilities in an electronic system, nor has a necessary and sufficient test of authenticity been defined (10).

Necessary and Sufficient Conditions for Warhead Verification

To prove a statement requires a system that is complete, meaning the system must be capable of proving that any true warhead is true; and the system must be sound, meaning the system must reject all hoax warheads as false (11). The politics of warhead verification imposes a third requirement, that there be zero disclosure of sensitive information, either measured directly or inferable by statistical reconstructions.

Because "warheadness" is not a physically measurable property, a special axiom must be constructed to bridge from the domain of physical measurements to statements about authenticity. Although such axioms are rarely made explicit, the strength of any verification system hangs on the veracity of this axiom. We adopt the following: If every manifold (a connected logical body) in a candidate warhead is identical (within allowed tolerances) in size, shape, spatial relation, density, and isotopic composition to a corresponding manifold in an authentic warhead, then the candidate warhead is authentic.

This is not a perfect bridge axiom because it is insensitive to microstructure, such as grain size in metals, surface polish, bonding interfaces, and chemical isomers with identical densities. Variations

Significance

We provide a method for potentially solving one of the central problems in arms control: proving the authenticity of a nuclear warhead without revealing information about its construction. The proposal is the first, to our knowledge, to be proof-theoretically sound (i.e., unspoofable) and information-theoretically secure. The method uses a tomography technique which can resolve the spatial distribution of mass with isotopic specificity. The soundness and security emerge from an integral-geometric transform that is able to uniquely map 3D geometries from zero-dimensional point measurements.

Author contributions: R.S.K., A.D., R.R.M., and J.R.V. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]R.S.K., A.D., R.R.M., and J.R.V. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: rsk@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603916113/-/DCSupplemental.

in these properties can affect warhead function, and in principle it would be desirable to validate all these properties during the verification process. However, we suggest that a hoax warhead that was manufactured to be identical in isotopic composition and geometry to a real warhead, but which deviated only by some microstructural property, would be so nearly difficult to manufacture as a real warhead that it would not be a valuable cheating strategy. As such, this axiom provides in our estimation a sufficient test of authenticity.

A physical test able to satisfy the axiom is nontrivial. It must be capable of mapping the spatial distribution of every isotope of interest in the warhead: in essence, an isotopic tomogram is required. This work introduces a method for making such a tomogram.

Concept and Protocol

We extend the previously proposed template approach to verification, whereby a candidate warhead is compared with a known-authentic warhead called a template. To verify that the candidate is essentially identical to the template, we propose to radiograph both objects at a set of identical but randomly chosen orientations. Unlike a traditional imaging radiograph, the process uses a single-pixel detector that cannot resolve spatial details but nonetheless ensures the entire 3D geometry matches with high confidence. Simultaneously, reconstruction of warhead composition and geometry is prevented by scattering "foils" that serve as physical analogs to encryption keys. The overall interaction would proceed as follows:

i) A template warhead is established by a joint exercise between the host (the country offering candidate warheads for verification) and the inspector.
ii) The host submits additional candidate warheads to be tested into joint custody.
iii) The host installs an encrypting foil into the measurement apparatus (Fig. 1). The foil acts as a physical secret key that protects classified information during a measurement.
iv) The inspector supervises tomographic measurements of the template and candidate according to randomized parameters of the inspector's choosing. Inspector's choice renders cheating impractical for the host.
v) Although the measured data are physically encrypted, if the candidate-warhead data match the template-warhead data, then the inspector has very high confidence the candidate warhead is a physical copy of the template warhead.

In our system, a template is an actual nuclear warhead. One template must be established for each warhead design being dismantled. The template must inherit the presumption of authenticity by a mechanism acceptable to the inspector. It has been previously suggested that the authenticity of a template might derive from its situational context (6). For example, a template warhead could be chosen at random from the fleet of intercontinental ballistic missiles. Because a country's nuclear deterrent hinges on the presence of real warheads on its missiles, the randomly selected warhead is very likely to be real. Other mechanisms, such as a credible record of custody, could be used to establish the authenticity of a nondeployed warhead template. Because the proposed technique hinges on the authenticity of the template, it should be recognized that any uncertainties associated with it are transferred to all verified items. The method by which one selects a template is therefore of crucial importance and deserves additional study.

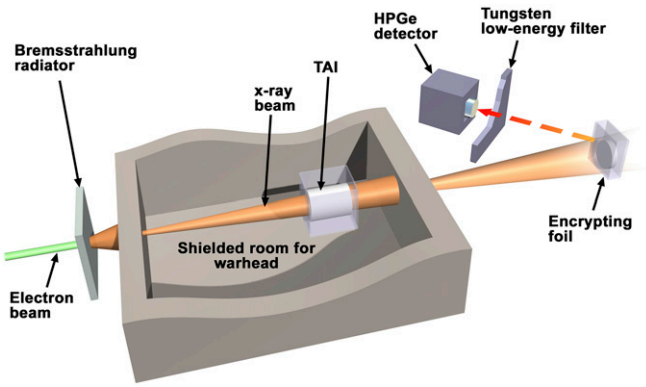

Fig. 1. Schematic of measurement apparatus. TAI is the treaty accountable item, either the template nuclear warhead or candidate warhead and its packaging.

Measurement

The measurement consists of a series of projections that can be used to create a tomogram of the warhead's geometry. Each projection is based on transmission nuclear-resonance fluorescence, which allows the system to resolve the geometric distribution of each isotope independently.

The measurement apparatus is illustrated in Fig. 1. A high-energy X-ray beam (2–9 MeV) passes through the object being tested. An ideal beam would be monoenergetic, such as produced by a laser Compton light source, as it greatly simplifies the proof of information security (12). In this paper, we simulate a bremsstrahlung X-ray source producing a wide continuum of photon energies, which is a more practical and affordable alternative, but one that requires the imposition of additional constraints to ensure information security. Bremsstrahlung will also give measured objects a significantly greater radiation dose, the implications of which would need to be evaluated by the host.

Nuclei in the test object will resonantly absorb photons from the interrogating beam if the photon energy matches an allowed transition of the nucleus. The widths of these absorption resonances are ∼1 eV FWHM and are described by a Doppler-broadened Breit–Wigner formulation (13). Most of the resonances are ∼10 keV or more apart, with the energies of each resonance being unique to each isotope. The sparseness of these resonances (visible in Fig. 2) prevents isotopic confusion, whereas the uniqueness of the resonant energy provides a one-to-one mapping between energy and the identity of the material.

As the beam traverses the test object, the resonant absorption of photons acts like a filter. The flux of photons at a nuclear-resonant energy is depleted in proportion to the amount of that isotope present, creating narrow absorption lines in the spectrum of the transmitted beam. The position and depth of these notches depends on the integrated density of each isotope along the path of the beam. By rotating the warhead and taking projections at different angles, the internal geometry of the warhead can be tomographically sampled.

The transmitted beam is not measured directly. Instead, the transmitted beam is incident upon a packet of thin foils, through which the large majority of the beam's X-rays will pass into a beam dump. However, a subset of photons that have passed through the warhead will still be at nuclei-resonant energies. If a resonant isotope is present in the foil packet, the on-resonance photons will now be absorbed by foil nuclei, exciting the nuclei into short-lived states. The excited nuclei of the foil rapidly decay (typically in < 1ps), emitting characteristic gammas in all directions, a process known as nuclear-resonance fluorescence (NRF). The rate at which these gammas are emitted depends on the fluence remaining in the beam leaving the warhead that is incident upon the foil, and on the areal thickness of the isotope in the foil (Eq. **1**). It is this fluorescent signature of the foil that is measured with detectors. Because there is no other process in materials that can mimic these characteristic gammas, we believe cheating by material substitution to be impossible. This might not be the case for other nuclear and atomic processes, such as neutron scattering, for which combinations of materials can be made to simulate the behavior of a specific isotope.

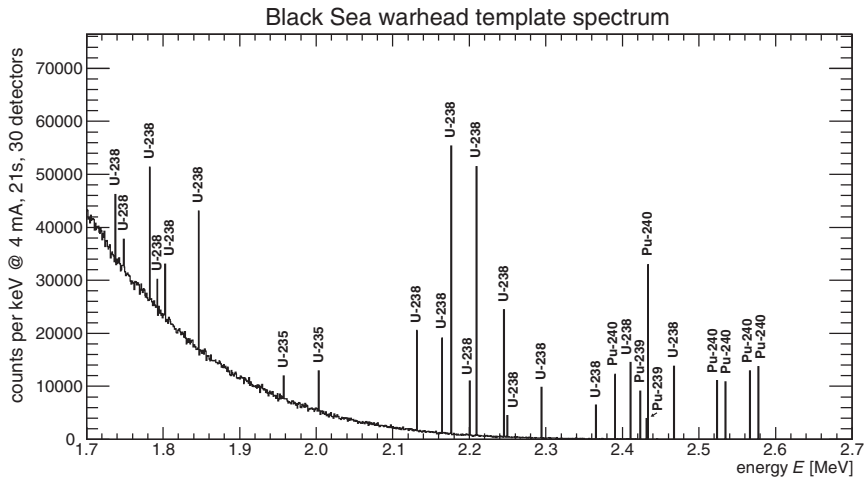

Fig. 2. Geant4 simulation of the spectrum generated by the Black Sea warhead and an encrypting foil, at 1 keV per bin. The spectrum represents the combined output of 30 idealized HPGe detectors placed at the distance of 1 m from the foil, using bremsstrahlung generated by a 4-mA 2.7-MeV electron beam and an exposure time of 21 s. The low-energy continuum is primarily the result of charged particles in the foil creating secondary bremsstrahlung.

Theory of a Single Projection. The spectral exitance (photons emitted per unit energy per unit surface area) $M(E,w)$ from some point w on the foil is closely described by Eq. **1**. The integral term represents the projection made when the beam flux passes through the object, which is described by a 3D density map $\rho_{\mathrm{obj},i}(r+s\theta)$ of the object for each isotope i. The integral runs through the object along a ray $r+\theta=\overline{rw}$ connecting the X-ray source at point r to a point w on the foil along some direction vector θ, the length of which is parameterized by s. The rate of photon attenuation depends on the NRF cross-section $\sigma_{\mathrm{NRF},i}$ for isotope i at energy E. The second bracketed term describes the response of the encrypting foil based on the foil's composition as described by the density $\rho_{\mathrm{foil},i}(w)$ of isotope i at point w, and its effective thickness z_i (*SI Appendix*, Eq. **S10c**).

$$M(E,w)\approx\phi_o A_{\mathrm{nr}}\exp\left[\sum_i\left(-\sigma_{\mathrm{NRF},i}\int_{\overline{rw}}\rho_{\mathrm{obj},i}(r+s\theta)\,ds\right)\right]\times\left[1-\exp\left(\sum_i-\sigma_{\mathrm{NRF},i}\,\rho_{\mathrm{foil},i}(w)z_i\right)\right]. \quad [1]$$

Note that the spectrum of photons exiting from the foil is a function of both $\rho_{\mathrm{obj},i}$ and $\rho_{\mathrm{foil},i}$. By changing the object being measured (ρ_{obj}) but holding the foil (ρ_{foil}) constant, the inspector can compare two measurements of M to determine whether the projection for a candidate warhead is identical to the same projection for a template warhead. Note that it is impossible to solve for the sensitive information in ρ_{obj} without knowledge of ρ_{foil}. Because the host can provide the foil without revealing its isotopic composition to the inspector, the vector of isotopes in the foil acts like a one-time-pad encryption key that prevents the inspector from learning the warhead's composition. Although the inspector does not know the amount of each isotope in the foil, each of which can vary by orders of magnitude, the inspector can still be certain that the foil is sensitive to each isotope of interest, for if the foil lacks a sufficient amount of an isotope, the corresponding fluorescence signal will not be adequately observed within the agreed measurement duration. In this respect, the foil is self-validating. The foil's gamma fluorescence can be measured using high-purity germanium (HPGe) detectors. Placing the detectors in shielded cavities that view the surface of the foil packet from angles that are large (typically >135°) relative to the beam direction will minimize the background flux from nonresonant processes in the foil, such as secondary bremsstrahlung, multiple Compton scatters, and X-ray fluorescence (XRF). The information security aspects of these non-NRF processes are discussed below.

Simulating a Single Projection. A measurement was simulated using the Geant4 Monte Carlo simulation toolkit with the G4NRF NRF-physics package (14, 15), which tracks each photon and its interaction daughters through the warhead, foil, and detector geometries and simulates all relevant photon interactions in these materials. The test objects were illuminated with a pencil beam of bremsstrahlung X-rays. A total of 8×10^{11} photons with a 2.7-MeV endpoint bremsstrahlung energy distribution was incident on each test object (*SI Appendix*, section 5.1). The template-warhead (Table 1) model is loosely based on approximate Soviet design information revealed during the Black Sea experiment, taking the median estimate for the thicknesses of weapons-grade plutonium, high-explosive, and highly enriched uranium; and applying the simplifying assumption that these materials are arranged as concentric spherical shells (16). The result is an object with the correct materials in approximately the correct quantities; it does not reflect the design of a real warhead.

The photons emitted from the encrypting foil at angles >135° from the forward-going beam are tallied. The resulting full-spectrum histogram as seen by the detector is shown in Fig. 2. For the simulation, the "foil" used was 2 cm thick and made from equal parts by mass ^{235}U, ^{238}U, ^{239}Pu, and ^{240}Pu, at a homogeneous overall density of 19 g/cm^3. This choice was made strictly to conserve computational time. In a practical environment, the foil would be thinner, varying from micrometers to centimeters, providing about 4 orders of magnitude in dynamic range while staying within reasonable measurement times (*SI Appendix*, section 5.4). This dynamic range is what allows the foil to perform a meaningful encrypting function.

Table 1. Assumed parameters for the Black Sea warhead template

Material name and isotopic composition	Density, g/cm^3	Inner radius, cm	Outer radius, cm
Weapon-grade plutonium (WGPu): 94% Pu-239, 6% Pu-240	19.84	6.27	6.7
HMX high explosive: 3% H, 16% C, 38% N, 43% O	1.89	6.7	13.2
Highly enriched uranium (HEU): 95% U-235, 5% U-238	18.7	13.2	13.45

Tomographic Uniqueness from Projections. The bridge axiom between observables and the conclusion of authenticity requires more than a statement about integrated densities along a single line of projection; it requires that all of the materials in the candidate object be identically distributed to those in the template object. The internal geometry must be somehow sampled to satisfy the bridge axiom, otherwise the system admits the possibility of a geometric hoax, an object that has all of the correct materials and is able to elicit the same NRF spectrum, but with a different geometric configuration. Such a hoax is illustrated in *SI Appendix*, Fig. S5.

The projection-slice theorem, which underpins tomographic imaging, holds that all 3D manifolds can be uniquely mapped from an infinite set of 2D projections, although a finite set of projections is usually adequate in practice (17). However, projection images of this type would reveal classified information about the warhead's structure, such as the diameter of the plutonium pit. To avoid this problem, we limit the system to a uniform foil that operates as a single-pixel camera with zero spatial-dimension sensitivity. However, this presents a challenge: the mathematical formalism for tomographic imaging from dimensionless single-pixel images has not been previously developed.

We devised an integral transform that is able to uniquely map a 3D geometry from a set of zero-dimensional projections of the type made by the single-pixel camera (*SI Appendix*, section 2). This transform is shown in Eq. **2**.

$$\begin{aligned}\mathcal{K}\rho_{\text{obj}}(r,\theta) &\equiv \int_{\mathbb{S}^2} \exp\left(-\int \rho_{\text{obj}}(r+s\theta)\, ds\right) d\theta \\ &\propto \int_{\text{foil}} M(E,w)\, dw,\end{aligned} \qquad [2]$$

where $\rho_{\text{obj}}(r+s\theta)$ is the density along the ray $r+\theta$ running through the object being mapped. The exterior integral runs over the entire 2D projection space $\mathbb{S}^2$, collapsing the 2D image into a scalar value—the same process as occurs in a single-pixel camera when the detector integrates photons from the entire foil. The inner integral over ds is the X-ray transform used in conventional tomography. The transform is unique and invertible for any object ρ_{obj}, which implies that any difference in the candidate object's structure can be resolved and no two structures will produce the same transform signal (see proof in *SI Appendix*, section 2.2). Although no inversion formula has been identified for the $\mathcal{K}$ transform, the inversion formula is not needed as the candidate and template geometries can be compared in the transform space.

Soundness Under Limited Sampling. Perfect soundness in tomographic systems is only guaranteed for infinite projections; for any finite number there arises the possibility that variations in the object may go undetected. These undetectable differences are called ghosts. Physically, ghosts can be thought of as a map of density variations that coincide spatially with ρ_{obj}, consisting of both positive and negative density-change values. A map g is a ghost if it has the property $\mathcal{K}(\rho_{\text{obj}}+g)=\mathcal{K}(\rho_{\text{obj}})$ for all projections considered. The objects $h=\rho_{\text{obj}}+g$ for each valid g constitute the complete set of all possible geometric hoax objects that are not the same as the authentic weapon, but which the system is unable to reject as mismatched. It has been shown that a smooth, continuous, and nontrivial g can be constructed for any finite set of projections (18). If one of the corresponding geometric hoaxes (h) can be practically manufactured, the system can be cheated.

Louis and Törnig showed that for the Radon transform, ghost functions on the unit disk can be described as a superposition of orthogonal Zernike polynomials (19). Importantly, the lowest-order polynomial in the set must have order H equal to or greater than the number of projections N (19). Therefore, whereas ghosts capable of spoofing the system always exist, if N is large the ghosts and corresponding hoax objects h will be highly oscillatory, geometrically complex, and eventually more difficult to manufacture than a true warhead. To illustrate this, example ghost functions with order $H=15$ and $H=78$ are shown in Fig. 3. Whereas it is impractical to test for the absence of all ghost objects, one can eliminate to an arbitrarily high level of confidence all of the low-order hoaxes that constitute the set of strategically useful cheats.

It is not necessary to measure $N>H$ projections to rule out all ghosts of polynomial order H. It is sufficient if the system can resolve $N>H$ projections and then sample only a few randomly selected projections. For a projection system limited to spherical rotations of the test object, the number of distinct projections can be approximated by the solid angle of circular cones as $N\approx 2/(1-\cos\delta\theta)$, where $\delta\theta$ is the effective angular resolution of the system. For example, a system able to resolve rotations of $\delta\theta=10°$ has $N\approx 132$ distinct projections. Given a system able to resolve N distinct projections, the probability that a hoax of minimum polynomial order H survives undetected by the system after k trials of randomly selected projections is bounded by Eq. **3**.

$$B\le \prod_{j=1}^{k}\left[(1-\alpha)\frac{H}{N}+\beta_j\left(1-\frac{H}{N}\right)\right], \qquad [3]$$

where α is the probability of erroneously rejecting a projection of a true warhead (type I error) because of statistical fluctuations, and β_j is the probability of failing to reject a nonmatching projection of a hoax because it could not be statistically differentiated from a matching projection (type II error) (*SI Appendix*, section 6.2). Note that β_j is specific to each hoax and projection, but in general it will increase as the hoax becomes more sophisticated in design, which also implies increasing H, and can be decreased by increasing measurement times, which also increases the number of resolvable projections N.

Sensitivity to Hoax Objects

Hoax objects were simulated to demonstrate the system's ability to detect several commonly suggested cheating scenarios:

i) Material replacement with a surrogate of similar density and atomic number. This type of hoax can spoof conventional X-ray and transmission-neutron measurements. In the example case, plutonium is replaced by depleted uranium.

ii) Replacing a material with another of the same element but with different isotopic ratios. In the example case, weapon-grade

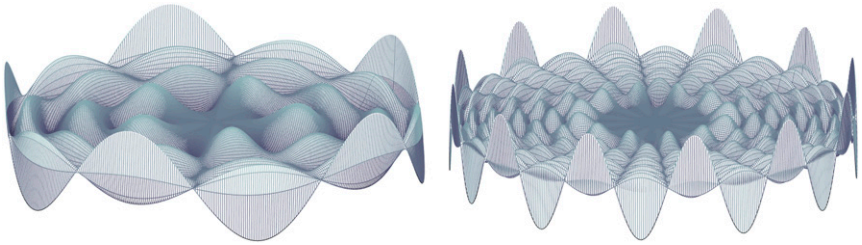

Fig. 3. Two-dimensional Zernike polynomials on the unit disk showing density modulations in a planar cross-section characteristic of a hoax that is able to be assured of escaping detection for 15 and 78 projections, respectively.

plutonium (WGPu) with 6% ^{240}Pu is replaced by fuel-grade plutonium (FGPu) with 14% ^{240}Pu.

iii) A geometric hoax designed to produce a spectrogram that matches the authentic object by arranging the correct materials in a nonwarhead geometry. In the example case, the hoax is a series of slabs that is able to confuse the system at two collinear angles of projection ($H = 2$).

Measurements from each cheating scenario are compared with the measurement of an authentic template. A simple statistical test was used in which the photon counts in four energy bins are compared, each corresponding to an NRF transition for an isotope of interest.

Test Thresholds. The discrepancy between the counts for energy bin i and projection j can be restated in SDs as $\Delta_{i,j} = (c_{\text{can},i,j} - c_{\text{tem},i,j})/\sqrt{\sigma^2_{\text{can},i,j} + \sigma^2_{\text{tem},i,j}}$, where $c_{\text{can},i,j}$ and $c_{\text{tem},i,j}$ are the counts in bin i for the candidate and template, respectively, and $\sigma^2_{\text{can},i,j}$ and $\sigma^2_{\text{tem},i,j}$ are their variances. The decision rule used here is to alarm if the absolute difference $|\Delta_{i,j}|$ for any of the chosen NRF bins exceeds the selected test threshold Δ_t. The probability that a true matching projection is rejected because of random fluctuation (the false-alarm rate) is $\alpha = 1 - [\Phi(\Delta_t) - \Phi(-\Delta_t)]^n$, where n is the number of NRF bins evaluated per projection and Φ is the cumulative distribution function of the normal distribution centered at zero with variance unity (or cumulative Skellam distribution if total counts are small). For the simulation examples below, $\Delta_t = 4.2\sigma$ and $n = 4$, giving $\alpha = 1 \times 10^{-4}$. A more considered protocol might aim for a higher false-alarm rate to intentionally exercise a procedure for occasionally retesting authentic weapons. This would minimize the political repercussions of a false alarm if an innocent mistake had been made, while simultaneously improving the probability of detecting a genuine hoax object.

With the alarm threshold chosen, the probability of failing to detect an unmatched projection can be estimated as $\beta_j = \min_{\forall i}(\Phi(\Delta_t - |\Delta_{i,j}|))$, where $\Delta_{i,j}$ are now the discrepancies in the units of SD for a particular hoax, at projection j, and NRF bin i. Values of β_j are reported for each hoax in Table 2. If the object is sampled at multiple projections, then the overall system probability of failing to reject a hoax after all k projections is B, and is bounded by Eq. **3**, where N and H are functions of the measurement time and the quality of the hoax, respectively. Note that the system is able to perform to arbitrarily high levels of accuracy for any hoax by measuring for an increased amount of time or by increasing the number of projections k.

Simulation Results for a Practical System. The results in Table 2 show what could be achieved with a practical system built using commercial off-the-shelf components: the Ion-Beam Applications Ltd. TT100 Rhodotron (an electron accelerator with beam energy adjustable up to 10 MeV at 4 mA; higher current models are also made), and an array of 30 idealized HPGe detectors [having 20% intrinsic efficiency for 2-MeV photons, better known as "100%" detectors based on their performance relative to the ANSI/IEEE-325–1996 standard for NaI (20); see also *SI Appendix*, section 8.3]. The results indicate that the described technique is able to reject all of the posited material-substitution hoaxes with greater than 99.9% probability in a 21-s measurement. See *SI Appendix*, section 5.4 on scaling to thinner foils.

To make a statement about the probability of detecting a geometric hoax, some estimate of the system's performance under random rotations of the hoax is required. A simulation was performed for a 10° rotation of a geometric hoax of polynomial order $H = 2$ and achieving $\beta_{10^\circ} = 6.82 \times 10^{-7}$ for one projection. This indicates it is conservative to assume that a $\delta\theta = 10^\circ$ rotation constitutes the smallest statistically discernible change in projection possible. Rotation through a larger angle would lead to a smaller β_j, which implies $\beta_j \le \beta_{10^\circ}$ for the $H = 2$ hoax when the system is restricted to a 10° minimum rotation. Under this assumption, and restricting projections to rotations around the unit sphere, $N \approx 2/(1 - \cos\delta\theta) = 132$. Using Eq. **3**, we can then bound the probability of cheating the example system with an $H = 2$ hoax after one, two, and three 21-s measurements at random angles as being less than 0.0152, 2.30×10^{-4}, and 3.48×10^{-7}, respectively.

Information Security

Eq. **1** states that, given a single NRF energy, the integrated density for each isotope in the test object ($\rho_{\text{obj},i}$) is impossible to infer without knowledge of the foil's composition. If the composition of the foil can be kept unknown to the inspector, then the system is underconstrained and is provably secure for each projection. A monochromatic photon source would selectively excite individual NRF resonances, providing a way to ensure that only one NRF transition is measured at a time. The corresponding isotope in the foil can be assumed to vary from tens of micrometers to centimeters, so in practice it is unknown only to within several orders of magnitude, but this is more than adequate to protect information about the warhead.

Although the foil's composition is not disclosed to the inspector, it is also important to prevent the inspector from building a credible statistical estimation of the foil. A foil should not be used for multiple projections, as doing so may weaken its secrecy, just as the multiple uses of a one-time pad can result in vulnerabilities. However, our findings suggest that only a few geometric projections will be necessary to achieve excellent rejection rates for possible hoax objects, and one or two foil changes would not be an arduous task.

Even under monochromatic excitation, nonresonant processes in the foil, namely secondary bremsstrahlung, multiple Compton scatter, and XRF, may produce measurable signals in other energy bins. The first two, although easily forward modeled using Monte Carlo simulations, have no closed-form expression. However, they are processes that are sensitive to the average elemental compositions of the warhead and foil and can, therefore, be rendered incapable of revealing information about the isotope of

Table 2. Detection potential for commonly suggested hoax objects for a single projection

	Difference in counts expressed in units of σ for select NRF bins at 21 array-seconds				Probability of failing to reject a hoax, β_j
Isotope Reference NRF line	U-238 2,176 keV	U-235 1,957 keV	Pu-239 2,431 keV	Pu-240 2,433 keV	
Template vs. authentic*	1.41	0.454	1.47	1.69	—
Template vs. WGPu → U-238	127	3.49	14.0	1.48	~0
Template vs. WGPu → FGPu	3.75	1.93	1.49	7.16	7.89×10^{-4}
Template vs. geometric hoax oriented to be undetectable	1.81	0.96	0.747	0.320	$1 - \alpha$ (by design)
Template vs. geometric hoax after 10° rotation	8.83	4.40	0.0796	0.964	6.82×10^{-7}

*For comparison, a second authentic object was also simulated by using a different Monte Carlo random seed. The discrepancy in counts for the authentic object and the matching geometric hoax are exaggerated by low simulation statistics (*SI Appendix*, section 5.3).

interest by adding material to the foil that will modulate their response independent of the isotope being examined. The third, XRF, will reveal some information about concentration of specific elements in the foil, but a modest tungsten filter, as illustrated in Fig. 1, will remove effectively all of the XRF photons before reaching the germanium detectors. The germanium detectors should also be separated from the warhead by heavy shielding so there is no possibility of measuring other photons that come directly from the warhead.

Much changes if a bremsstrahlung source is used. The ratios of NRF peaks can facilitate estimation of the foil's composition. The host can compensate by adding additional unknowns to the system (*SI Appendix*, section 7.1.2). Under bremsstrahlung illumination, the information content of the nonresonant continuum would also need to be examined for possible information content.

If the secret composition of the foil were somehow learned, useful information about the weapon would still be protected by the nonlinearity of the $\mathcal{K}$ transform. For example, with perfect knowledge of the foil, the inspector can estimate the line integral of the warhead's isotopic density for a single projection, but the nonlinearity of the transform ensures there is no one-to-one correspondence between the line-integrated density and the mass. The mass of an isotope can be recovered only if the inspector also knows, or correctly guesses, the geometric distribution of relative densities for the isotope in the warhead. The host can eliminate the reliability of such guesses by using a technique described in *SI Appendix*, section 7.1.3.

Finally, if an inspector can reliably and accurately discern the composition of all secret foils used for each projection, if inversion formula for the $\mathcal{K}$ transform is known, and if a large number of different projections of a warhead are sampled, then the inspector could discern the geometry of the warhead. However, our simulations show that only one or two randomly selected projections will be needed to achieve excellent rejection rates for the canonical hoaxes examined here, suggesting the last requirement of many projections need never be satisfied in a practical system.

Conclusions

Trends in nuclear disarmament indicate a need for a provably sound, complete, and information-secure method for authenticating nuclear warheads. The system proposed here seeks to meet these requirements with a mathematically quantifiable physical framework and logically sound set of axioms. By the virtue of being in the public domain, various aspects of the system can be freely studied and scrutinized by any participant in the arms reduction process. In addition to theoretical considerations, there may be important operational limitations to implementing such a system in existing national-security environments, which will require further study by government experts.

Soundness against strategically meaningful hoax warheads is achieved by careful attention to the logical connections that link observables to conclusions about warhead authenticity; by exploiting a process in nuclear physics that acts as a naturally occurring one-to-one function between energy and isotope identity; and through the introduction of a single-pixel tomographic transform that provides a unique measure of the warhead's geometry. Information is protected by the intrinsic physics of the system rather than relying on difficult-to-certify electronic circuits or software. Although electronics are used to operate the detectors and to measure signals leaving the shielded area, no sensitive information ever becomes experimentally observable.

Simulations for a bremsstrahlung-based system using available technologies and the simulated geometries suggest that two 21-s measurements should be able to reject a canonical set of hoaxes with greater than 99.9% probability, while falsely alarming on an authentic warhead about once in every 10,000 warheads. Realistic foil and warhead geometries will almost certainly require somewhat longer measurements than those simulated. Work remains to validate these simulations in an experimental setting. NRF cross-section data for some isotopes, mainly those not used here, are still sparse and may need to be measured if those isotopes are to be included in the verification process.

If a bremsstrahlung beam is used, several information security questions remain open. The information content of the continuum underneath the NRF signal is at present poorly understood; the simultaneous observation of multiple NRF transitions, which is unavoidable with bremsstrahlung, will require careful study of countermeasures to ensure the system remains underconstrained; and dose to the warhead will be high when using bremsstrahlung. Many of these problems can be avoided by using a tunable monochromatic photon source.

Finally, although the example system suggests excellent performance for canonical hoaxes and measurements on the order of minutes, a relationship between measurement time and the upper bound on the probability of failing to reject any possible hoax would be a useful future development. As these probabilities are inherently a function of hoax geometry, such a statement requires a description of the space of all possible hoax objects, which might first require the development of an inversion formula for the $\mathcal{K}$ transform. This remains as future work.

ACKNOWLEDGMENTS. The authors thank Richard Lanza, Zachary Hartwig, Bari Osmanov, and Gunther Uhlmann. This work is supported in part by the Consortium for Verification Technology under Department of Energy National Nuclear Security Administration Award DE-NA0002534 and by the Carnegie Corporation of New York.

1. Federation of American Scientists (2016) Status of world nuclear forces. Available at fas.org/issues/nuclear-weapons/status-world-nuclear-forces/. Accessed February 1, 2016.
2. US Department of Defense (2010) *Nuclear Posture Review* (US Department of Defense, Washington, DC).
3. US Arms Control and Disarmament Agency (1968) *Demonstrated Destruction of Nuclear Weapons: Final Report of Field Test FT-34* (US Arms Control and Disarmament Agency, Washington, DC).
4. Drell S, et al. (1993) Verification of Dismantlement of Nuclear Warheads and Controls on Nuclear Materials (JASON, MITRE Corporation, McLean, VA), Technical Report JSR-92-331.
5. UK Atomic Weapons Establishment (1998) Confidence, security and verification: The challenge of global nuclear weapons arms control (UK Atomic Weapons Establishment, Aldermaston, UK), Technical Report AWE/TR/2000/001.
6. US Department of Energy, Office of Nonproliferation Research and Engineering (2001) *Technology R&D for Arms Control* (Lawrence Livermore National Laboratory, Livermore, CA).
7. Kidder R, Sykes L, von Hippel FN (October 10, 1999) False fears about a test ban. *Washington Post*, p B07.
8. Malakoff D (July 31, 2002) Test ban treaty technically sound. *Sci News*. Available at www.sciencemag.org/news/2002/07/test-ban-treaty-technically-sound.
9. Close D, MacArthur D, Nicholas N (2001) Information barriers—a historical perspective (Los Alamos National Laboratory, Los Alamos, NM), Technical Report LA-UR-01-2180.
10. Brubaker E, et al. (2015) Information barriers for imaging: FY15 report (Pacific Northwest National Laboratory, Richland, WA), Technical Report PNNL-24782.
11. Cook SA, Reckhow RA (1979) The relative efficiency of propositional proof systems. *J Symb Log* 44(1):36–50.
12. Albert F, et al. (2014) Laser wakefield accelerator based light sources: Potential applications and requirements. *Plasma Phys Contr Fusion* 56(8):084015.
13. Metzger FR (1959) Resonance fluorescence in nuclei. *Prog Nucl Phys* 7:54–88.
14. Agostinelli S, et al. (2003) Geant4—a simulation toolkit. *Nucl Instrum Methods Phys Res, Sect A* 506(3):250–303.
15. Jordan DV, Warren GA (2007) Simulation of nuclear resonance fluorescence in Geant4. *Nuclear Science Symposium Conference Record* (IEEE, Piscataway, NJ), Vol 2, pp 1185–1190.
16. Fetter S, Cochran TB, Grodzins L, Lynch HL, Zucker MS (1990) Gamma-ray measurements of a Soviet cruise-missile warhead. *Science* 248(4957):828–834.
17. Sharafutdinov VA (1993) Uniqueness theorems for the exponential X-ray transform. *J Inverse Ill-Posed Probl* 1(4):355–372.
18. Derevtsov EY (1997) Ghost distributions in the cone-beam tomography. *J Inverse Ill-Posed Probl* 5(5):411–426.
19. Louis AK, Törnig W (1981) Ghosts in tomography - the null space of the radon transform. *Math Methods Appl Sci* 3(1):1–10.
20. IEEE (1997) *Standard Test Procedures for Germanium Gamma-Ray Detectors* (IEEE, New York).